CASMA884 A TOPIC COURSE ON

Gaussian analysis with applications

(2025 FALL, BOSTON UNIVERSITY)

Guangqu ZHENG

(email: gzheng90@bu.edu)

Abstract. This is a preliminary set of lecture notes written for the topic course CASMA884 at Boston University. The course aims at familiarizing students with standard Gaussian objects (Gaussian processes, noises, etc), while equipping students with several powerful tools including Wiener chaos expansion, Malliavin calculus, Stein's method, etc. The applications in this course may include some of the following topics:

- (i) computations of Greeks for financial derivatives ✓
- (ii) a gentle touch on Kahane's theory of Gaussian multiplicative chaos ✓; ground-state energy of polymer in a Gaussian environment ✓.
- (iii) sliding window statistics, triangle counts in Erdős-Rényi ✓.
- (iii) random field solution for PAM/HAM with (time-independent) Gaussian noises ✓.
- (iv) random matrix theory, parameter estimation
- (v) zeros and/or level sets of random polynomials/Gaussian processes
- (vi) machine learning, and more? [**to be modified**]

∗ Frequently used notations

$\mathbb{N} = \{0, 1, 2, ...\}$, $\mathbb{N}_0 = \{1, 2, ...\}$, $\mathbb{R}_+ = [0, \infty)$, $[n] := \{1, ..., n\}$ for $n \in \mathbb{N}_0$, $\mathbb{Z} = \{0, -1, 1, -2, 2, ...\}$.

Throughout this notes, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability triplet, on which most probabilistic objects are defined unless specified otherwise.

$\|f\|_\infty$ denotes the essential sup-norm of a function f.

$\|X\|_p$ denotes the $L^p(\Omega)$-norm of a real random variable X. We use $L^p(\mathbb{P})$ and $L^p(\Omega)$ interchangeably.

For $a, b \in \mathbb{R}$, $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$.

iff is short for "if and only if"

i.i.d. is short for "identically and independently distributed"

a.e. is short for "almost everywhere"

a.s. is short for "almost surely"

LHS (RHS) is short for "left-hand side" ("right-hand side", respectively).

$\lfloor x \rfloor$ denotes the largest integer that is not bigger than x.

$\log x$ is the logarithm function: $y = \log x \Leftrightarrow x = e^y$.

$a \cdot b = \sum_{i=1}^n a_i b_i$ denotes the usual Euclidean inner product of $a, b \in \mathbb{R}^n$. The Euclidean norm of $x \in \mathbb{R}^d$ is denoted by $|x|$.

$\partial_j f(x)$ is short for $\partial_{x_j} f(x_1, ..., x_d)$; $\partial_j^2 = \partial_{x_j}\partial_{x_j}$; $\Delta f = \sum_{j=1}^d \partial_j^2 f$.

$\text{Hess} f(x) = (\partial_i \partial_j f(x_1, ..., x_d))_{i,j=1}^d \in \mathbb{R}^{d\times d}$. For $A, B \in \mathbb{R}^{d\times d}$,

$$\langle A, B\rangle_{\text{HS}} = \sum_{i,j=1}^d A_{ij} B_{ij}.$$

Contents

* **Frequently used notations** 1
1. **Gaussian integration by parts on $\mathbb{R}$** 3
1.1. Gaussian integration by parts formula 5
1.2. Hermite polynomials 7
1.3. Application I: two series representations of π (optional reading) 11
1.4. Application II: computation of Greek letters (optional reading) 13
2. **Gaussian analysis on $\mathbb{R}^d$** 16
2.1. Refresh on week 1 16
2.2. Week 1 extension to $\mathbb{R}^d$ 17
2.3. Application I: A gentle touch on GMC 22
2.4. Applications II: Maximum of Gaussians 27
2.5. Applications III: Gaussian random polymer 28
2.6. More on Ornstein-Uhlenbeck 30
3. **A pleasant detour: Stein's method of normal approximation** 33
3.1. What is Stein's method for? 33
3.2. Basics on Stein's method 33
3.3. Stein's method of exchangeable pairs 37
3.4. 2nd-order Gaussian Poincaré inequality 39
3.5. Stein's method via local dependency 41
4. **Redo the analysis in an isonormal framework** 47
4.1. Chaos expansion 50
4.2. Malliavin operators I: D & δ 53
4.3. (Generalized) chain rule 58
4.4. Malliavin operators $D, \delta, \mathcal{L} = -\delta D$, and $P_t = e^{t\mathcal{L}}$ 59
4.5. Ito v.s. Skorohod 60
4.6. Application: linear/NL SPDEs 60
5. **Fourth moment theorems and Nourdin-Peccati analysis** 61
6. How Stein meets Malliavin? 63
6.1. A methodological breakthrough by Nualart and Ortiz-Latorre 64
6.2. Application: parameter estimation 66
6.3. Application: GOE approximation of Wishart matrices 66
6.4. Application: zeros of random polynomials 66
7. **Regularity of Gaussian objects** 66
Appendix A. **Basic results in analysis** 67
Appendix B. **Basic results in probability** 69
Appendix C. **Answers to Questions** 70
References 71

1. **Gaussian integration by parts on $\mathbb{R}$**

In this section, we begin by introducing the one-dimensional version of the Malliavin calculus, or the stochastic analysis for random variables of the form $g(Z)$, with $Z \sim \mathcal{N}(0,1)$ a standard normal random variable. Let us give a brief review on the real Gaussian random variables. We say $Z \sim \mathcal{N}(0,1)$ is a standard normal if Z is a continuous random variable with the probability density function (PDF)

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \ z \in \mathbb{R}. \tag{1.1}$$

Throughout this course, we will use γ to denote the standard Gaussian measure on $\mathbb{R}$ unless specified otherwise:

$$\gamma(dz) = \phi(z)dz \quad \text{with } \phi \text{ as in (1.1)}.$$

For example, $L^2(\gamma)$ is the space of measurable functions from $\mathbb{R}$ to $\mathbb{R}$ that are square-integrable with respect to the standard Gaussian measure γ. Similarly, we write for $p \in [1,\infty)$,[1]

$$f \in L^p(\gamma) \Leftrightarrow \int_{\mathbb{R}} |f(z)|^p \gamma(dz) < \infty.$$

Let us first do some easy Gaussian computations. Using the above PDF (1.1), we can compute the fourth moment of $Z \sim \mathcal{N}(0,1)$ as follows:

$$\begin{aligned}
\mathbb{E}[Z^4] &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} z^4 dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} z^3 d(z^2/2) \\
&= -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} z^3 de^{-\frac{1}{2}z^2} = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz^3 \quad \text{(by integration by parts)} \\
&= 3\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} z^2 dz = 3\mathbb{E}[Z^2] = 3.
\end{aligned} \tag{1.2}$$

Exercise 1.1. *Verify the following moment formulae for $Z \sim \mathcal{N}(0,1)$:*

$$\mathbb{E}[Z^m] = \begin{cases} 0 & \text{if } m \in 2\mathbb{N}+1 \\ (2n-1)!! = \frac{(2n)!}{2^n n!} & \text{if } m = 2n \in 2\mathbb{N}. \end{cases} \tag{1.3}$$

Students are invited to think about the following warm-up questions.[2]

Question 1.2. *Let Z be a standard normal and $g : \mathbb{R} \to \mathbb{R}$ be a measurable function. Is the following equivalence true?*

"$g(Z)$ is normal distributed iff g is affine"

Question 1.2 appears to be straightforward, at least for the "if" direction. For the "only if" direction, it is somehow wrong in general. See the following example:

$$g(z) = z\mathbf{1}_{\{|z|\geq 1\}} - z\mathbf{1}_{\{|z|<1\}}.$$

What if we assume additionally that the function g is bijective and increasing?

Question 1.3. *Can you provide a direct proof of the following inequality?*

$$\|P_d\|_{L^p(\gamma)} \leq C_{p,d} \|P_d\|_{L^2(\gamma)} \tag{1.4}$$

for any polynomial P_d with degree at most d and for any $p \in [2,\infty)$, where $C_{p,d}$ is a constant that only depends on p and d.

[1] $f \in L^\infty(\gamma)$ iff f is bounded almost everywhere with respect to the Lebesgue measure.

[2] In this lecture notes, the Exercises are left for students to solve, while the Questions will be addressed in classes.

There are various reasons for the question

"*why the Gaussian objects are so important?*".

One of them is that the Gaussian distribution often arises as the limit of various important probabilistic/statistical objects. For example, the **law of large number** (LLN) asserts that (in the language of a undergraduate statistic course)

(LLN) for a random sample from a population distribution with mean zero and variance one, the sample mean statistic tends to zero[3] as the sample size tends to infinity.

If we assume further that the population distribution has finite second moment (or equivalently, finite variance), the classical central limit theorem asserts that after a proper normalization, the sample mean statistic admits *Gaussian fluctuation* as the sample size tends to infinity. Here, we see that the Gaussian distribution arises as the *universal* limiting distribution regardless of the population distribution. Results on asymptotic normality are always welcomed, e.g., in statistics, since one can use them for various statistical inferences (e.g., confidence interval estimation, hypothesis testing).

Theorem 1.4 (Classical central limit theorem)**.** *Let $\{X_i : i \in \mathbb{N}\}$ be i.i.d. random variables with mean zero and variance one. Then,*

$$V_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

converges in law/distribution to a standard normal distribution $\mathcal{N}(0,1)$ as $n \to \infty$. That is, for any $t \in \mathbb{R}$,

$$\mathbb{P}(V_n \le t) \xrightarrow{n\to\infty} \Phi(t) := \int_{-\infty}^{t} \phi(z)dz \tag{1.5}$$

with ϕ as in (1.1)*.*

The above central limit theorem (CLT) is of qualitative nature. The convergence in (1.5) is the point-wise convergence of the cumulative distribution function (CDF) of V_n to the standard normal CDF Φ. Under stronger assumptions (e.g., $\mathbb{E}[|X_1|^3] < \infty$), the famous **Berry-Esseen bound** in Theorem 1.5 gives a more quantitative result by specifying the rate at which this convergence takes place. The rate of convergence is measured by the Kolmogorov distance

$$d_{\mathrm{Kol}}(X,Y) := \sup_{t\in\mathbb{R}} \big|\mathbb{P}(X \le t) - \mathbb{P}(Y \le t)\big|. \tag{1.6}$$

There are other frequently used distributional metrics that measure the proximity of distributions (see Section 3).

Theorem 1.5 (Berry-Esseen bound)**.** *Let the assumptions in Theorem 1.4 hold and assume additionally that $\mathbb{E}[|X_1|^3] < \infty$. Then,*

$$d_{\mathrm{Kol}}(V_n, Z) \le \frac{C\mathbb{E}[|X_1|^3]}{\sqrt{n}},$$

where $Z \sim \mathcal{N}(0,1)$ and the absolute constant C does not depend on the population distribution nor the sample size n.[4] *We postpone the proof to Section 3.*

Let us now move to the core of this section.

[3]This convergence occurs almost surely and in $L^1(\Omega)$.

[4]References on the absolute constant **XXX**; Without assuming the finite absolute third moment, we can derive $d_{\mathrm{Kol}}(V_n, Z) \to 0$ as $n \to +\infty$.

1.1. **Gaussian integration by parts formula.**

Lemma 1.6. *Let $Z \sim \mathcal{N}(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function such that $f' \in L^1(\gamma)$. Then, $\mathbb{E}[|Zf(Z)|] < \infty$ and*

$$\mathbb{E}[f'(Z)] = \mathbb{E}[Zf(Z)] \tag{1.7}$$

or equivalently,

$$\int_{\mathbb{R}} f'(z)\gamma(dz) = \int_{\mathbb{R}} zf(z)\gamma(dz). \tag{1.8}$$

Proof. Let f, Z be given as above. To show the finiteness of $\mathbb{E}[|Zf(Z)|]$, it suffices to prove

$$\mathbb{E}[|Z[f(Z) - f(0)]|] < \infty. \tag{1.9}$$

Indeed, using the fundamental theorem of calculus (Theorem A.1), we can write

$$\begin{aligned}
\mathbb{E}[|Z[f(Z) - f(0)]|] &= \int_{-\infty}^{\infty} |z[f(z) - f(0)]|\phi(z)dz \\
&= \int_{-\infty}^{0} \left| z \int_{z}^{0} f'(t)dt \right| \phi(z)dz + \int_{0}^{\infty} \left| z \int_{0}^{z} f'(t)dt \right| \phi(z)dz \\
&\le \int_{-\infty}^{0} |z| \int_{z}^{0} |f'(t)|dt\phi(z)dz + \int_{0}^{\infty} |z| \int_{0}^{z} |f'(t)|dt\phi(z)dz \\
&= \int_{0}^{\infty} |f'(t)| \left(\int_{t}^{\infty} z\phi(z)dz \right) dt - \int_{-\infty}^{0} |f'(t)| \left(\int_{-\infty}^{t} z\phi(z)dz \right) dt.
\end{aligned} \tag{1.10}$$

Note that for $a \in \mathbb{R}_+$, one can easily verify that

$$\int_{a}^{\infty} z\phi(z)dz = -\int_{-\infty}^{-a} z\phi(z)dz = \phi(a). \tag{1.11}$$

It follows immediately that

$$\mathbb{E}[|Z[f(Z) - f(0)]|] \le \int_{-\infty}^{\infty} |f'(t)|\phi(t)dt < \infty.$$

That is, we just proved (1.9) and thus proved the finiteness of $\mathbb{E}[|Zf(Z)|]$. The proof of (1.8) can be done in the same way. Here is a cheap way. Removing the absolute signs in (1.10), we can get similar expressions with all equal signs, and particularly we have,

$$\mathbb{E}[Z[f(Z) - f(0)]] = \int_{0}^{\infty} f'(t) \left(\int_{t}^{\infty} z\phi(z)dz \right) dt - \int_{-\infty}^{0} f'(t) \left(\int_{-\infty}^{t} z\phi(z)dz \right) dt.$$

Note that the LHS coincides with $\mathbb{E}[Zf(Z)]$, and the RHS is exactly $\mathbb{E}[f'(Z)]$ in view of (1.11). This concludes our proof. □

• **Divergence operator δ.** For $f : \mathbb{R} \to \mathbb{R}$ absolutely continuous, we put

$$(\delta f)(x) = xf(x) - f'(x),$$

which is well defined for almost every $x \in \mathbb{R}$. From Lemma 1.6, we know that

$$\delta f \in L^1(\gamma) \text{ with } \mathbb{E}[\delta(f)(Z)] = 0, \text{ provided } f' \in L^1(\gamma). \tag{1.12}$$

Next, we present a generalization of Lemma 1.6.

Proposition 1.7. (i) [Gaussian Poincaré inequality] *Suppose f is absolutely continuous with f' in $L^2(\gamma)$. Then, $f \in L^2(\gamma)$ with*

$$\mathrm{Var}(f(Z)) \leq \mathbb{E}[|f'(Z)|^2].$$

The equality holds true only when f is affine, i.e., $f(z) = az + b$ for some $a, b \in \mathbb{R}$.
(ii) [Duality relation] *For f, g absolutely continuous with f', g' in $L^2(\gamma)$, we have*

$$\int_{\mathbb{R}} f'(z)g(z)\gamma(dz) = \int_{\mathbb{R}} f(z)\delta(g)(z)\gamma(dz). \tag{1.13}$$

Proof. Let us first prove part (i) using the fundamental theorem of calculus as in the proof of Lemma 1.6. Similarly, we write

$$\begin{aligned}
&\int_{\mathbb{R}} |f(z) - f(0)|^2 \phi(z)dz = \int_0^{\infty} \left|\int_0^z f'(t)dt\right|^2 \phi(z)dz + \int_{-\infty}^0 \left|\int_z^0 f'(t)dt\right|^2 \phi(z)dz \\
&\leq \int_0^{\infty} |z| \int_0^z |f'(t)|^2 dt\phi(z)dz + \int_{-\infty}^0 |z| \int_z^0 |f'(t)|^2 dt\phi(z)dz \\
&= \int_0^{\infty} dt|f'(t)|^2 \left(\int_t^{\infty} z\phi(z)dz\right) - \int_{-\infty}^0 dt|f'(t)|^2 \left(\int_{-\infty}^t z\phi(z)dz\right) \\
&= \int_{\mathbb{R}} dt|f'(t)|^2 \phi(t) < \infty,
\end{aligned} \tag{1.14}$$

where the last equality is due to (1.11). This proves $f \in L^2(\gamma)$. Moreover,

$$\mathrm{Var}(f(Z)) \leq \mathbb{E}[|f(Z) - f(0)|^2] \leq \int_{\mathbb{R}} dt|f'(t)|^2\phi(t) = \mathbb{E}[|f'(Z)|^2].$$

Note that the inequality in (1.14) becomes an equality only when

$$\left|\int_0^z f'(t)dt\right|^2 = z\int_0^z |f'(t)|^2 dt \quad \text{and} \quad \left|\int_y^0 f'(t)dt\right|^2 = |y|\int_y^0 |f'(t)|^2 dt$$

for any $y < 0 < z$. This occurs only when f' is a constant (*why?*). Therefore, the proof of part (i) is completed.

Now let us prove part (ii). Let g be any absolutely continuous function with $g' \in L^2(\gamma)$, then $g \in L^2(\gamma)$ and fg is absolutely continuous with

$$(fg)' = fg' + f'g \in L^1(\gamma).$$

See Theorem A.2. Therefore, Lemma 1.6 (see also (1.12)) indicates that $\delta(fg) \in L^1(\gamma)$ with

$$\begin{aligned}
0 = \int_{\mathbb{R}} \delta(fg)d\gamma &= \int_{\mathbb{R}} \big[zf(z)g(z) - f(z)g'(z) - f'(z)g(z)\big]\gamma(dz) \\
&= \int_{\mathbb{R}} f(z)\delta(g)(z)\gamma(dz) - \int_{\mathbb{R}} f'(z)g(z)\gamma(dz),
\end{aligned}$$

which is exactly the equality (1.13). Hence, the proof is completed. $\square$

Remark 1.8. Consider $f, g \in C_c^{\infty}(\mathbb{R})$, the integration by parts formula $\int_{\mathbb{R}} f'gdx = -\int_{\mathbb{R}} fg'dx$ in undergraduate calculus essentially follows from the fact that

$$\int_{\mathbb{R}} h'(x)dx = 0$$

for $h \in C_c^\infty(\mathbb{R})$. If we replace the Lebesgue measure by the standard Gaussian measure γ, we have

$$\int_{\mathbb{R}} h'(x)\gamma(dx) = \int_{\mathbb{R}} h'(x)\phi(x)dx = -\int_{\mathbb{R}} h(x)\phi'(x)dx = \int_{\mathbb{R}} xh(x)\phi(x)dx,$$

which is a particular case of the usual integration by parts. For this reason, we often call (1.13) a Gaussian integration by parts formula.

Let $\mathrm{dom}(\delta)$ be the set of absolutely continuous functions g with $g' \in L^2(\gamma)$. Let $\mathbb{D}$ be the set of absolutely continuous functions f with $f' \in L^2(\gamma)$. Note that $\mathrm{dom}(\delta)$ is the domain for the divergence operator and $\mathbb{D}$ is the domain for the derivative operator on $L^2(\gamma)$, and in our current simple setting, $\mathrm{dom}(\delta)$ coincides with $\mathbb{D}$. In a more general setting, these two domains are totally different. In Proposition 1.7, we can only see that $\delta(f) \in L^1(\gamma)$ when $f' \in L^2(\gamma)$. In fact, $f' \in L^2(\gamma)$ ensures $\delta(f) \in L^2(\gamma)$; see Remark 1.15.

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 1 (Sept. 2, 2025)** $\infty\infty\infty\infty\infty\infty\infty\infty$

Lecture 2.

Let us first introduce Hermite polynomials and then return for more discussions on the divergence operator δ and its domain $\mathrm{dom}(\delta)$.

1.2. **Hermite polynomials.**

Definition 1.9. *Put $H_0(x) = 1$, and $H_{p+1} = \delta(H_p)$ for $p \in \mathbb{N}$. H_p is called the Hermite polynomial of order p. The first few ones are given by*

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x, \dots$$

It is not difficult to verify by induction that

$$\frac{d}{dx} H_p(x) = pH_{p-1}(x) \quad \text{and} \quad H_{p+1}(x) = xH_p(x) - pH_{p-1}(x) \tag{1.15}$$

for any integer $p \geq 1$. In a compact manner, we can also define Hermite polynomials using Rodrigues' formula:

$$H_p(x) = (-1)^p \frac{1}{\phi(x)} \phi^{(p)}(x) \tag{1.16}$$

with ϕ as in (1.1). We leave the verification as an exercise to students. Another way to define the Hermite polynomials is via the Gaussian integral formula $H_p(x) = \mathbb{E}[(x+iZ)^p]$ with $Z \sim \mathcal{N}(0,1)$; see the expository article [NO25] for various applications.

Exercise 1.10. (i) *Verify the relations* (1.15) *and* (1.16).

(ii) *Show that any monomial can be expressed as a finite linear combination of Hermite polynomials.*

The family of Hermite polynomials play a crucial role in the Gaussian analysis. They are the orthogonal polynomials with respect to the standard Gaussian measure on $\mathbb{R}$.

Proposition 1.11. *The set of polynomials $\{H_p/\sqrt{p!} : p \in \mathbb{N}\}$ form an orthonormal basis for $L^2(\gamma)$. As a consequence, the following results hold true.*

(i) *With $Z \sim \mathcal{N}(0,1)$, we have*

$$\mathbb{E}[H_p(Z)H_q(Z)] = p!\mathbf{1}_{\{p=q\}}. \tag{1.17}$$

(ii) *For any $f \in L^2(\gamma)$, we have the following* **Hermite expansion***:*

$$f = \sum_{q\geq 0} c_q H_q \quad \text{with } c_q = \frac{1}{q!}\int_{\mathbb{R}} H_q(z)f(z)\gamma(dz),$$

where the above series converge in $L^2(\gamma)$ and $\sum_{q\geq 0} c_q^2 q! < \infty$.

Proof. First we prove that

$$\text{the linear span of the family } \{H_p/\sqrt{p!} : p \in \mathbb{N}\} \text{ is dense in } L^2(\gamma). \qquad (\#)$$

In view of Exercise 1.10, it is enough to show that the linear span of the family $\{x^p : p \in \mathbb{N}\}$ is dense in $L^2(\gamma)$. Suppose $f \in L^2(\gamma)$ satisfies

$$\int_{\mathbb{R}} f(z)z^k\gamma(dz) = 0, \ \forall k \in \mathbb{N}. \tag{1.18}$$

For any $t \in \mathbb{R}$, we have $\int_{\mathbb{R}} e^{2|tz|}\gamma(dz) < \infty$ and thus

$$\int_{\mathbb{R}} |f(z)| \sum_{k=0}^{\infty} \frac{|tz|^k}{k!}\gamma(dz) = \int_{\mathbb{R}} |f(z)|e^{|tz|}\gamma(dz) < \infty$$

by Cauchy-Schwarz inequality. This ensures that we can write

$$\int_{\mathbb{R}} f(z) \sum_{k=0}^{\infty} \frac{(itz)^k}{k!}\gamma(dz) = \sum_{k=0}^{\infty} \int_{\mathbb{R}} f(z)\frac{(itz)^k}{k!}\gamma(dz),$$

which is zero due to the condition (1.18). This implies in particular that $\int_{\mathbb{R}} f(z)e^{itz}\phi(z)dz = 0$ for any $t \in \mathbb{R}$. With $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$, we get $f = f^+ - f^-$ and

$$\int_{\mathbb{R}} e^{itz} f^+(z)\phi(z)dz = \int_{\mathbb{R}} e^{itz} f^-(z)\phi(z)dz. \tag{1.19}$$

If one of f^+ and f^- is a zero function (a.e.), then the other one has to be zero almost everywhere. Now assume otherwise so that (taking $t = 0$ in (1.19))

$$C := \int_{\mathbb{R}} f^+(z)\phi(z)dz = \int_{\mathbb{R}} f^-(z)\phi(z)dz \in (0, \infty),$$

In this case, we can deduce from (1.19) that $\frac{1}{C}f^+(z)\phi(z)$ and $\frac{1}{C}f^-(z)\phi(z)$ are two probability density functions with the same Fourier transform (i.e., characteristic function), and thus they must coincide with each other almost everywhere. That is, we have $f^+ = f^-$, or equivalently, $f = f^+ - f^- = 0$ almost everywhere. Now we can conclude that for a function $f \in L^2(\gamma)$ satisfying (1.18), $f = 0$ a.e., which shows the desired conclusion (#).

Next, we show the orthogonality relation (1.17). Suppose $p \leq q$ and recall the relation (1.15). We can deduce from the duality relation (Proposition 1.7) that

$$\begin{aligned}\mathbb{E}[H_p(Z)H_q(Z)] &= \mathbb{E}[H_p(Z)(\delta H_{q-1})(Z)] \\ &= \mathbb{E}[pH_{p-1}(Z)H_{q-1}(Z)] \\ &= p!\mathbb{E}[H_{q-p}(Z)] = p!\mathbf{1}_{\{p=q\}}.\end{aligned}$$

Hence, we just proved that $\{H_p/\sqrt{p!} : p \in \mathbb{N}\}$ form an orthonormal basis for $L^2(\gamma)$, and the Hermite expansion in part (ii) follows immediately. $\square$

Proposition 1.12 (Product formula for Hermite polynomials)**.** *Let p, q be positive integers. Then,*

$$H_p(x)H_q(x) = \sum_{r=0}^{p\wedge q} r!\binom{p}{r}\binom{q}{r} H_{p+q-2r}(x).$$

Proof. Note that $H_p(x)$ is a polynomial of degree p, then $H_p(x)H_q(x)$ is a polynomial of degree $p+q$. Thanks to Exercise 1.10, we can express $H_p(x)H_q(x)$ as a finite linear combination of Hermite polynomials:

$$H_p(x)H_q(x) = \sum_{r=0}^{p+q} c_r H_r(x)$$

with the coefficients c_r's to be determined. Since $\mathbb{E}[H_k(Z)] = 0$ with $Z \sim \mathcal{N}(0,1)$ for any $k \geq 1$ and $\frac{d^k}{dx^k} H_k(x) = k!$ (see (1.15)), we have

$$c_r r! = \mathbb{E}\Big(\frac{d^r}{dx^r}\big[H_p(x)H_q(x)\big]\big|_{x=Z}\Big). \tag{1.20}$$

By the general Leibniz rule, we have for $r \leq p+q$,

$$\begin{aligned}\frac{d^r}{dx^r}\big[H_p(x)H_q(x)\big] &= \sum_{k=0}^{r}\binom{r}{k}\big(\frac{d^k}{dx^k}H_p(x)\big)\big(\frac{d^{r-k}}{dx^{r-k}}H_q(x)\big)\\ &= \sum_{k=0}^{r}\binom{r}{k}\frac{p!}{(p-k)!}H_{p-k}(x)\frac{q!}{(q+k-r)!}H_{q+k-r}(x)\mathbf{1}_{\left\{\substack{k\leq p\\ r-k\leq q}\right\}}.\end{aligned} \tag{1.21}$$

Using the orthogonality relation in Proposition 1.11-(i), we have

$$\mathbb{E}\big[H_{p-k}(Z)H_{q+k-r}(Z)\big] = (p-k)!\mathbf{1}_{\{k=\frac{p-q+r}{2}\}}.$$

Therefore, it follows from (1.20) and (1.21) that

$$c_r r! = \sum_{k=0}^{r}\binom{r}{k}\frac{p!}{(p-k)!}\frac{q!}{(q+k-r)!}\mathbf{1}_{\left\{\substack{k\leq p\\ r-k\leq q}\right\}}(p-k)!\mathbf{1}_{\{k=\frac{p-q+r}{2}\}}.$$

Then, taking the above restrictions on k, r into account with $k = \frac{p-q+r}{2}$, we have

$$c_r = \frac{p!q!}{(\frac{p-q+r}{2})!(\frac{q+r-p}{2})!(\frac{p+q-r}{2})!}\mathbf{1}_{\left\{\substack{|p-q|\leq r\leq p+q\\ p+q+r \text{ is even}}\right\}}.$$

Then, making the change of variables $r' = (p+q-r)/2$ (thus $r = p+q-2r'$), we get

$$c_{p+q-2r'} = \frac{p!q!}{(p-r')!(q-r')!(r')!}\mathbf{1}_{\{0\leq r'\leq p\wedge q\}}.$$

Hence, this concludes our proof. $\square$

Question 1.13. *Use the same proof strategy to show the following "binomial formula" for Hermite polynomials:*

$$H_n(ax+by) = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}H_k(x)H_{n-k}(y), \tag{1.22}$$

where $a^2 + b^2 = 1$.

Proposition 1.14 (Gaussian Poincaré inequality revisited)**.** *Suppose $f \in \mathbb{D}$ and $Z \sim \mathcal{N}(0,1)$. Then,*

$$\mathrm{Var}(f(Z)) \leq \mathbb{E}[|f'(Z)|^2],$$

where the equality holds if and only if $f(x) = ax + b$ for some $a, b \in \mathbb{R}$.

Proof. In Proposition 1.7, we gave a proof of the Gaussian Poincaré inequality using the fundamental theorem of calculus. Now let us present a different proof of the Gaussian Poincaré inequality using the Hermite expansion.

Suppose $f \in L^2(\gamma)$ admits the following expansion $f = \sum_{q\in\mathbb{N}} c_q H_q$ with $\sum_{q\in\mathbb{N}} c_q^2 q! < \infty$. Then,

$$\mathrm{Var}(f(Z)) = \sum_{q\in\mathbb{N}_0} c_q^2 q!$$

and with $f' = \sum_{q\in\mathbb{N}_0} c_q q H_{q-1}$

$$\mathbb{E}[|f'(Z)|^2] = \sum_{q\in\mathbb{N}_0} c_q^2 q^2 (q-1)! = \sum_{q\in\mathbb{N}_0} c_q^2 q! q.$$

Thus,

$$\mathrm{Var}(f(Z)) \leq \mathbb{E}[|f'(Z)|^2]$$

with equality only when $c_q = 0$ for all $q \geq 2$. That is, the equality holds only when f is affine. $\square$

Remark 1.15. Let us return to the description of the domains for D and δ (i.e., $\mathbb{D} = \mathrm{dom}(\delta)$) and prove the anticipated result

$$\delta(f) \in L^2(\gamma) \text{ for } f \in \mathrm{dom}(\delta).$$

For $f \in L^2(\gamma)$ with the Hermite expansion

$$f = \sum_{q\geq 0} c_q H_q, \tag{1.23}$$

we have $\sum_{q\geq 0} c_q^2 q! < \infty$. The set $\mathbb{D}$ is then the set of functions f with the Hermite expansion as in (1.23) such that

$$\sum_{q\geq 0} c_q^2 q! q < \infty. \tag{1.24}$$

Indeed, $f' = \sum_{q\geq 1} c_q H_q' = \sum_{q\geq 1} c_q q H_{q-1} \in L^2(\gamma)$ implies

$$\sum_{q\geq 1} c_q^2 q^2 (q-1)! < \infty,$$

which is exactly the condition (1.24). For f as in (1.23) with (1.24), we deduce from the recurrence relation 1.15 that

$$xf(x) - f'(x) = \sum_{q=0}^{\infty} c_q [xH_q(x) - qH_{q-1}(x)] = \sum_{q=0}^{\infty} c_q H_{q+1}(x),$$

which is square-integrable with respect to the standard Gaussian measure, that is,

$$\delta(f) = xf(x) - f'(x) = \sum_{q\geq 1} c_q \delta(H_q) = \sum_{q=0}^{\infty} c_q H_{q+1} \in L^2(\gamma).$$

Question 1.16. (i) *Show that the derivative operator* ($i.e.$, $Df = f'$) *is* **closable** *meaning that if $f_n \in \mathbb{D}$ with $Df_n \to g$ in $L^2(\gamma)$ for some g and $f_n \to 0$ in $L^2(\gamma)$ as $n \to +\infty$, then $g = 0$. See also Proposition 4.8.*

(ii) *Show that δ (as the adjoint operator of D) is* **closed** *meaning that for $g_n \in \mathrm{dom}(\delta)$ with $\delta(g_n)$ convergent to h in $L^2(\gamma)$, one can find $g \in \mathrm{dom}(\delta)$ such that $h = \delta(g)$ and $g_n \to g$ in $L^2(\gamma)$.*

Exercise 1.17. *Let $\mathcal{A}$ be the set of functions $g \in L^2(\gamma)$ such that there is a constant $C = C_g$ such that*

$$\left| \int_{\mathbb{R}} f'(z)g(z)\gamma(dz) \right| \le C\|f\|_{L^2(\gamma)}$$

for any $f \in \mathbb{D}$. Prove that $\mathcal{A} = \mathrm{dom}(\delta)$.

Exercise 1.18. *Verify the duality relation in Proposition 1.7 by using the Hermite expansions.*

1.3. **Application I: two series representations of π (optional reading).**

$$\pi = \sum_{k=0}^{\infty} \frac{1}{2^{2k-1}(2k+1)} \binom{2k}{k} \qquad \text{Ramanujan's 1912 identity} \tag{1.25}$$

$$\pi = \sum_{k=0}^{\infty} \frac{3}{2^{4k}(2k+1)} \binom{2k}{k} \qquad \text{Newton's 1676 identity.} \tag{1.26}$$

We will prove the above two formulae using Hermite expansions and they appeared in a study on intransitivity of dice [HMRZ20].

• **Preparation.** Let $h_u(x) = \mathbf{1}_{\{x>u\}}$. We claim that

$$h_u = \mathbb{P}(Z > u) + \sum_{q=1}^{\infty} a_q H_q \quad \text{with } a_q = \frac{1}{q!}\phi(u)H_{q-1}(u). \tag{1.27}$$

Proof of claim (1.27). It is trivial that $h \in L^2(\gamma)$ and thus admits a Hermite expansion

$$h = \sum_{q\ge 0} a_q H_q,$$

where $a_0 = \mathbb{P}(Z > u)$ and for $q \ge 1$,

$$\begin{aligned} a_q :&= \frac{1}{q!}\int_{\mathbb{R}} H_q(z)h_u(z)\gamma(dz) = \frac{1}{q!}\int_u^{\infty} H_q(z)\phi(z)dz \\ &= \frac{1}{q!}\int_u^{\infty} (-1)^q \phi^{(q)}(z)dz \quad \text{using (1.16)} \\ &= \frac{1}{q!}(-1)^{q+1}\phi^{(q-1)}(u) = \frac{1}{q!}\phi(u)H_{q-1}(u) \quad \text{using (1.16) again.} \end{aligned}$$

This verifies the claim (1.27). □

Let us first present the formula for $H_q(0)$:

$$H_{2p-1}(0) = 0 \quad \text{and} \quad H_{2p}(0) = \frac{(-1)^p(2p)!}{2^p p!} \quad \text{for any } p \in \mathbb{N}_0. \tag{1.28}$$

This can be easily verified by induction (Exercise!). For general $x \in \mathbb{R}$, we have for $q \in \mathbb{N}_0$,

$$H_q(x) = \sum_{k=0}^{\lfloor q/2 \rfloor} \frac{q!(-1)^k}{k!(q-2k)!2^k} x^{q-2k}, \tag{1.29}$$

which can be derived by using (1.28) and (1.15) (Exercise!).

• *Proof of* (1.25). From the above discussions, we can first write, with $h_0(x) = \mathbf{1}_{\{x>0\}}$, that

$$\begin{aligned} h_0 &= \frac{1}{2} + \sum_{q=1}^{\infty} \frac{1}{q!\sqrt{2\pi}} H_{q-1}(0) H_q \\ &= \frac{1}{2} + \sum_{k=0}^{\infty} \frac{1}{(2k+1)!\sqrt{2\pi}} \frac{(-1)^k (2k)!}{2^k k!} H_{2k+1} \\ &= \frac{1}{2} + \sum_{k=0}^{\infty} d_{2k+1} H_{2k+1}, \end{aligned}$$

where

$$d_{2k+1} = \frac{(-1)^k}{2^k k!(2k+1)\sqrt{2\pi}}. \tag{1.30}$$

Then, using the orthogonality relation of Hermite polynomials, we have

$$\mathrm{Var}(h_0(Z)) = \sum_{k=0}^{\infty} d_{2k+1}^2 (2k+1)! = \sum_{k=0}^{\infty} \binom{2k}{k} \frac{1}{2^{2k+1}\pi(2k+1)}.$$

It is easy to get $\mathrm{Var}(h_0(Z)) = 1/4$, and hence we get (1.25). □

• *Proof of* (1.26). Recall Φ is the CDF of a standard normal $Z \sim \mathcal{N}(0,1)$. Then, by the probability integral transform $\Phi(Z)$ is uniformly distributed over $(0,1)$, so that $\mathrm{Var}(\Phi(Z)) = \frac{1}{12}$. On the other hand, we can write by using the above expansion of h_0 and the binomial formula (1.22),

$$\begin{aligned} \Phi(x) &= \mathbb{E}\big[h_0(\tfrac{Z}{\sqrt{2}} + \tfrac{x}{\sqrt{2}})\big] \\ &= \frac{1}{2} + \mathbb{E}\bigg(\sum_{k=0}^{\infty} d_{2k+1} H_{2k+1}(\tfrac{Z}{\sqrt{2}} + \tfrac{x}{\sqrt{2}})\bigg) \\ &= \frac{1}{2} + \mathbb{E}\bigg(\sum_{k=0}^{\infty} d_{2k+1} \sum_{r=0}^{2k+1} \binom{2k+1}{r} 2^{-k-\frac{1}{2}} H_r(Z) H_{2k+1-r}(x)\bigg). \end{aligned}$$

Since $\mathbb{E}[H_r(Z)] = 0$ for $r \geq 1$ and $H_0 \equiv 1$, we can simplify the above expression to

$$\Phi = \frac{1}{2} + \sum_{k=0}^{\infty} d_{2k+1} 2^{-k-\frac{1}{2}} H_{2k+1}.$$

Therefore, we get

$$\begin{aligned} \frac{1}{12} = \mathrm{Var}(\Phi(Z)) &= \sum_{k=0}^{\infty} d_{2k+1}^2 2^{-2k-1} (2k+1)! \\ &= \sum_{k=0}^{\infty} \binom{2k}{k} \frac{1}{2^{4k+2}\pi(2k+1)}, \end{aligned}$$

from which we can deduce (1.26). □

1.4. **Application II : computation of Greek letters (optional reading).** In the end of this section, we present a financial-math application of the Gaussian integration by parts formula. Let us first give a bit the financial context, which is by no means comprehensive, while we refer interested readers to the books [Sh04, Hull22] for the preliminaries in finance and relevant mathematics.

• **Basics of financial derivatives.** A *financial derivative* is a financial instrument whose value is derived from the value of an underlying asset, index, or rate. In this subsection, the underlying assets will be restricted to stocks whose prices follow the law of the classical geometric Brownian motion. Concretely, let S_t denote the stock price at time t, and it can be modeled by

$$S_t = S_0 e^{(r-\frac{1}{2}\sigma^2)t+\sigma W_t}, \tag{1.31}$$

where $S_0 > 0$ is the initial stock price, $r > 0$ is the risk-free interest rate, $\sigma > 0$ is the volatility, and $(W_t)_{t\geq 0}$ denotes the real-valued, standard Brownian motion.[5] Suppose a financial derivative's underlying asset is a stock with the above price (1.31) and its payoff at the terminal time $T > 0$ is described by $\psi(S_T)$, that is, the payoff of the financial derivative at time T is assumed to only depend on the terminal stock price. According to the principle of risk-neutral pricing (see, e.g., [Sh04, Chapter 5]), the initial price of such a financial derivative is given by

$$H = e^{-rT}\mathbb{E}[\psi(S_T)],$$

where e^{-rT} is the discounting factor of a future money value to the present value (continuous compounding, in the language of finance). In this simple example, it is not difficult to see that the initial value H depends on various parameters like r, σ, and the initial stock price S_0. Here is a natural and important question:

"how sensitive H is when the parameters r, σ, S_0 change?"

In financial markets, traders and portfolio managers use **Greek letters** to measure the sensitivity of prices of financial derivatives to various market parameters. These sensitivities help manage risk efficiently and construct effective hedging strategies.

The primary Greek letters are:
(i) Delta (Δ): Sensitivity of the option price to the underlying asset price;
(ii) Gamma (Γ): Rate of change of Delta with respect to the underlying price;
(iii) Vega ($\mathcal{V}$): Sensitivity to changes in volatility;
(iv) Rho (ρ): Sensitivity to interest rates.

They are defined by

$$\Delta := \frac{\partial}{\partial S_0}H, \quad \Gamma := \frac{\partial}{\partial S_0}\Delta = \frac{\partial^2}{\partial^2 S_0}H, \quad \mathcal{V} := \frac{\partial}{\partial \sigma}H, \quad \text{and} \quad \rho := \frac{\partial}{\partial r}H. \tag{1.32}$$

In what follows, we only present the formula for the Delta and leave the proof of other Greek letters for students.

Proposition 1.19 (formula of Delta)**.** *Consider the above framework and assume that $\psi : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous* (*see Footnote 39*)*. Then, we have*

$$\Delta = \frac{e^{-rT}}{\sigma S_0 T}\,\mathbb{E}\left[\psi(S_T)W_T\right]. \tag{1.33}$$

[5] Here is a quick definition of Brownian motion. $(W_t)_{t\geq 0}$ is a family of centered Gaussian random variables satisfying the following three conditions: (i) any finite linear combinations of W_t's is Gaussian (ii) $t \in \mathbb{R}_+ \mapsto W_t$ is a.s. continuous (iii) $\mathbb{E}[W_t W_s] = t \wedge s$ for any $s, t \in \mathbb{R}_+$. In particular, we have $W_t \sim \mathcal{N}(0, t)$ for any $t \in \mathbb{R}_+$. This is the only property we need for this section.

Proof. Recall from Footnote 5 that $W_T \sim \mathcal{N}(0,T)$, so we can write $W_T = \sqrt{T}\cdot Z$ with $Z \sim \mathcal{N}(0,1)$. Then, using Lemma 1.6, we get for any absolutely continuous function f with $\mathbb{E}[|f'(\sqrt{T}Z)|] < \infty$,

$$\mathbb{E}[W_T f(W_T)] = \mathbb{E}[\sqrt{T}Z \cdot f(\sqrt{T}Z)] = T\cdot\mathbb{E}[f'(\sqrt{T}Z)] = T\cdot\mathbb{E}[f'(W_T)],$$

that is, we have

$$\mathbb{E}[W_T f(W_T)] = T\mathbb{E}[f'(W_T)]. \tag{1.34}$$

Since

$$H = e^{-rT}\mathbb{E}[\psi(S_T)] \quad \text{with } S_T = S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma W_T},$$

We can write $S_T = h(W_T)$ as a function of W_T with $h(x) = S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma x}$. That is, we have

$$H = e^{-rT}\mathbb{E}[\psi\circ h(W_T)].$$

Note that the function $f = \psi\circ h$ is absolutely continuous with

$$f'(x) = \psi'(h(x))h'(x) = \sigma\psi'(h(x))h(x) \quad \text{and thus } f'(W_T) = \sigma S_T\psi'(S_T). \tag{1.35}$$

see, e.g., (A.1). Since ψ' is uniformly bounded (see, e.g., Footnote 39) and $h(W_T) = S_T$ has finite expectation, we see that the function f satisfy the assumptions for applying (1.34). Doing so leads to

$$\begin{aligned}
\Delta = \frac{\partial}{\partial S_0}H &= e^{-rT}\frac{\partial}{\partial S_0}\mathbb{E}\left[\psi\big(S_0 e^{(r-\frac{1}{2}\sigma^2)T+\sigma W_T}\big)\right] \\
&= e^{-rT}\,\mathbb{E}\left[\psi'(S_T)\cdot\frac{S_T\sigma}{S_0\sigma}\right] \qquad (1.36)\\
&= \frac{e^{-rT}}{\sigma S_0}\,\mathbb{E}\left[f'(W_T)\right] \qquad (1.37)\\
&= \frac{e^{-rT}}{\sigma S_0 T}\,\mathbb{E}\left[f(W_T)W_T\right] \quad \text{by (1.34)}.
\end{aligned}$$

Note that the step (1.36) follows from Lebesgue's differentiation theorem with ψ' uniformly bounded and $S_T \in L^1(\Omega)$, and the equality (1.37) is due to (1.35). Finally, the formula (1.33) is proved by noticing that $f(W_T) = \psi(S_T)$. $\square$

Remark 1.20. (i) The above derivation of the formula for Δ is quite elementary and only utilizes the Gaussian integration by parts formula in Lemma 1.6. However, the above proof strategy fails if the payoff does not just depend on the terminal stock price S_T, for example, when the payoff depends on $(S_t)_{t\in[0,T]}$, i.e., the whole stock price dynamics on the duration of the financial derivative. See Chapter 6 of [Nua06] for more discussions and in particular, the interested readers may want to compare our proof with the argument on [Nua06, page 332] that uses the full power of the Malliavin calculus. See also Chapter 2 of the book [MT06] by Malliavin and Thalmaier.

(ii) Consider the case where $\psi(x) = \max\{x-K, 0\}$ for some given $K > 0$. This corresponds to the payoff of a European call option with strike price K and terminal time T. The European call option gives its holder a right (not an obligation) to purchase the underlying asset (stock in our example) at the price K at time T. It is of fundamental importance in finance, and its price is given by the famous Black-Scholes-Merton formula:

$$H_{\text{call}} = S_0\Phi\big(d_1\big) - Ke^{-rT}\Phi\big(d_2\big)$$

with

$$d_1 = \frac{\log\left(S_0/K\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T}.$$

By directly differentiating the explicit formula of H_{call} with respect to S_0, we can get[6]

$$\Delta_{\text{call}} = \Phi(d_1).$$

Alternatively, one can first use (1.33) to write

$$\Delta_{\text{call}} = \frac{e^{-rT}}{\sigma S_0 T}\mathbb{E}\big[\max\{S_T - K, 0\}W_T\big]$$

and standard (although tedious) Gaussian computations will lead to the same formula $\Delta_{\text{call}} = \Phi(d_1)$. The formula (1.33) is advantageous when we do not have an explicit price formula.

Exercise 1.21. *Under the same assumptions as in Proposition 1.19, prove the following formulae:*

$$\Gamma = \frac{e^{-rT}}{\sigma^2 T S_0^2}\,\mathbb{E}\left[\left(\frac{W_T^2}{T} - 1\right)\psi(S_T)\right]; \qquad \mathcal{V} = \mathbb{E}\left[\left(\frac{W_T^2}{\sigma T} - W_T - \frac{1}{\sigma}\right)\psi(S_T)\right];$$

$$\rho = \frac{e^{-rT}}{\sigma}\mathbb{E}\big[\psi(S_T)W_T\big] - TH.$$

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 2 (Sept. 4, 2025)** $\infty\infty\infty\infty\infty\infty\infty\infty$

[6]Note that the expressions d_1, d_2 are involved with S_0 such that $S_0\Phi'(d_1) = Ke^{-rT}\Phi'(d_2)$.

Lecture 3.

2. Gaussian analysis on $\mathbb{R}^d$

In Section 1, we studied some basic Gaussian analysis that is involved with a single Gaussian random variable.

Let us briefly review Section 1 while introducing a few new things. Throughout this sections, we write

$$\gamma_d(dx) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}|x|^2} dx$$

for the standard Gaussian measure on $\mathbb{R}^d$.

2.1. **Refresh on week 1.** In Section 1, we introduced D, δ with a common domain

$$\mathbb{D}^{1,2}(\mathbb{R}) := \{f \in L^2(\gamma_1) : f \text{ is absolutely continuous and } f' \in L^2(\gamma_1)\}$$

The Gaussian Poincaré inequality (Proposition 1.14) ensures that $f \in L^2(\gamma_1)$ whenever $f' \in L^2(\gamma_1)$. We defined Hermite polynomials $\{H_p : p \in \mathbb{N}\}$, which are the orthogonal polynomials for γ_1. More precisely, for any $f \in L^2(\gamma_1)$, we can find $\{c_q : q \geq 0\}$ such that

$$f = \sum_{q\in\mathbb{N}} c_q H_q \quad \text{and} \quad \sum_{q\geq 0} c_q^2 q! < \infty.$$

With $Df = f'$ and $(\delta g)(x) = xg(x) - g'(x)$, we observe that

$$\mathcal{L}f := -\delta D f = f'' - xf'(x),$$

which gives us the generator of the real Ornstein-Uhlenbeck process described by the following stochastic differential equation (SDE)

$$\begin{cases} dX_t = -X_t dt + \sqrt{2} dB_t \\ X_0 = x \in \mathbb{R} \end{cases}$$

with B a standard real Brownian motion. One can easily solve the above SDE:

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

It is easy to verify that $X_t \sim \mathcal{N}(xe^{-t}, 1 - e^{-2t})$, which weakly converges (or converges in law) to the standard Gaussian measure γ_1 as $t \to +\infty$, regardless of the initial value $x \in \mathbb{R}$. The process $(X_t : t \in \mathbb{R}_+)$ is a Markov process and admits a semigroup (called Ornstein-Uhlenbeck semigroup)

$$(P_t f)(x) = \mathbb{E}[f(X_t)|X_0 = x] = \mathbb{E}[f(xe^{-t} + \sqrt{1 - e^{-2t}} Z)]$$

with $Z \sim \mathcal{N}(0, 1)$, and the second representation is known as Mehler's formula.

Exercise 2.1. (i) Show $P_{t+s} = P_t P_s$ for any $t, s \in \mathbb{R}_+$, $P_t H_q = e^{-qt} H_q$, $\mathcal{L}H_q = -qH_q$, and $DP_t = e^{-t} P_t D$.
(ii) Show that $\mathcal{L}f = \lim_{t\to 0} \frac{P_t f - f}{t}$ for every $f \in \mathrm{dom}(\mathcal{L})$ after you figure out $\mathrm{dom}(\mathcal{L})$.
(iii) Show that for every $f, g \in \mathbb{D}^{1,2}(\mathbb{R})$, with $Z \sim \mathcal{N}(0, 1)$, we have

$$\mathrm{Cov}(f(Z), g(Z)) = \int_0^\infty \langle Df(Z), DP_t g(Z)\rangle_{L^2(\gamma_1)} dt.$$

`Hint for` (iii): `Hermite expansion.`

2.2. **Week 1 extension to $\mathbb{R}^d$.** Now let us extend what we have learned to $\mathbb{R}^d$. We collect a few facts on Gaussian vectors in Appendix B, and students could refer to it when the need arises.

Definition 2.2. *Let $C^\infty_{\text{poly}}(\mathbb{R}^d)$ be the set of smooth functions $f : \mathbb{R}^d \to \mathbb{R}$ such that f and all its partial derivatives have at most polynomial growth at infinity.*

Proposition 2.3 (chaos expansion)**.** The Hilbert space $L^2(\gamma_d)$ can be decomposed into mutually orthogonal closed subspaces:

$$L^2(\gamma_d) = \bigoplus_{p\in\mathbb{N}} \mathbb{C}_p,$$

where $\mathbb{C}_p$ is the span of[7]

$$\Big\{(x_1, ..., x_d) \in \mathbb{R}^d \mapsto \prod_{j=1}^d H_{q_j}(x_j) : \sum_{j=1}^d q_j = p\Big\}.$$

That is, the following statements hold true.
(i) For any $p_i, q_i \in \mathbb{N}$, we have, with $\mathbf{G} \sim \mathcal{N}_d(0, \mathbf{I})$,

$$\mathbb{E}\bigg(\prod_{i=1}^d H_{p_i}(G_i)H_{q_i}(G_i)\bigg) = \mathbf{1}_{\{p_i=q_i,\forall i\}} \prod_{j=1}^d q_j!. \tag{2.1}$$

(ii) For any $f \in L^2(\gamma_d)$, there exist a family of coefficients $c_{q_1,...,q_d}$'s such that

$$f = \sum_{k=0}^\infty \underbrace{\sum_{\substack{q_1,...,q_d\in\mathbb{N}\\ q_1+...+q_d=k}} c_{q_1,...,q_d} \prod_{j=1}^d H_{q_j}(x_j)}_{\in\mathbb{C}_k} \quad \text{and} \quad \sum_{q_1,...,q_d\in\mathbb{N}} c^2_{q_1,...,q_d} \prod_{j=1}^d q_j! < \infty. \tag{2.2}$$

Proof. First note that (2.1) follows from the independence of different Gaussian entries and the orthogonality between different Hermite polynomials. Now let us prove (2.2) via an induction argument on $d \geq 1$. The 1D case is established in Proposition 1.11. Suppose that (2.2) holds true for any $d \leq n$ with $n \in \mathbb{N}_0$, and we will show it for $d = n + 1$.

Let $f \in L^2(\gamma_{n+1})$, i.e.,

$$\int_{\mathbb{R}^n} |f(x_1, ..., x_n, y)|^2 \gamma_d(dx)\gamma_d(dy) < \infty. \tag{2.3}$$

By Fubini's theorem, for γ_1-almost every $y \in \mathbb{R}$, the mapping

$$(x_1, ..., x_n) \in \mathbb{R}^n \mapsto f(x_1, ..., x_n, y) \in \mathbb{R}$$

is square-integrable with respect to γ_d. Then, the induction hypothesis ensures that one can find a family of (y-dependent!) coefficients $c_{q_1,...,q_d}(y)$ such that

$$f(x_1, ..., x_n, y) = \sum_{q_1,...,q_n\in\mathbb{N}} c_{q_1,...,q_n}(y) \prod_{j=1}^n H_{q_j}(x_j) \quad \text{in } L^2(\mathbb{R}^n, \gamma_n) \tag{2.4}$$

and the equation (2.3) is equivalent to

$$\sum_{q_1,...,q_n\in\mathbb{N}} \bigg(\prod_{j=1}^n q_j!\bigg) \int_{\mathbb{R}} c_{q_1,...,q_n}(y)^2 \gamma_1(dy) < \infty. \tag{2.5}$$

[7] $\mathbb{C}_0 \simeq \mathbb{R}$ is the set of constant random variables. Note that $\mathbb{C}_k$ is a closed subspace of $L^2(\gamma_d)$.

Then, applying induction hypothesis again (for $d = 1$), we can expand the function $y \in \mathbb{R} \mapsto c_{q_1,...,q_n}(y)$ as

$$c_{q_1,...,q_n}(y) = \sum_{q_{n+1}\geq 0} c_{q_1,...,q_n,q_{n+1}} H_{q_{n+1}}(y).$$

such that $\sum_{q_{n+1}\geq 0} c^2_{q_1,...,q_n,q_{n+1}} q_{n+1}! < \infty$. Together with (2.4) and (2.5), we can deduce that

$$f(x_1, ..., x_n, x_{n+1}) = \sum_{q_1,...,q_{n+1}\in\mathbb{N}} c_{q_1,...,q_{n+1}} \prod_{j=1}^{n+1} H_{q_j}(x_j) \quad \text{in } L^2(\gamma_{n+1})$$

and $\sum_{q_1,...,q_{n+1}\in\mathbb{N}} c^2_{q_1,...,q_{n+1}} \prod_{j=1}^{n} q_j! < \infty$. That is, (2.2) holds true for $d = n + 1$. As a result, we can conclude our proof by induction. □

Remark 2.4. The space $C^\infty_{\text{poly}}(\mathbb{R}^d)$ is dense in $L^2(\gamma_d)$. Indeed, for any $f \in L^2(\gamma_d)$ as in (2.2), we have

$$f_N = \sum_{k=0}^{N} \sum_{\substack{q_1,...,q_d\in\mathbb{N} \\ q_1+...+q_d=k}} c_{q_1,...,q_d} \prod_{j=1}^{d} H_{q_j}(x_j) \in C^\infty_{\text{poly}}(\mathbb{R}^d) \tag{2.6}$$

converges in $L^2(\gamma_d)$ to f.

• **Derivative operator** $D = \nabla$. When $f \in C^\infty_{\text{poly}}(\mathbb{R}^d)$,

$$Df = \nabla f = (\partial_1 f, ..., \partial_d f) : \mathbb{R}^d \to \mathbb{R}^d$$

is well defined. For f_N as in (2.6), we have

$$\partial_i f_N(x) = \sum_{k=1}^{N} \sum_{\substack{q_1,...,q_d\in\mathbb{N} \\ q_1+...+q_d=k}} c_{q_1,...,q_d} q_i H_{q_i-1}(x_i) \prod_{j=1}^{d} \mathbf{1}_{\{j\neq i\}} H_{q_j}(x_j)$$

with the convention $H_{-1}(x_i) = 0$. This implies that

$$\|\partial_i f_N\|^2_{L^2(\gamma_d)} = \sum_{k=1}^{N} \sum_{\substack{q_1,...,q_d\in\mathbb{N} \\ q_1+...+q_d=k}} c^2_{q_1,...,q_d} q_i \prod_{j=1}^{d} q_j!$$

and thus

$$\int_{\mathbb{R}^d} \|\nabla f_N(x)\|^2_{\mathbb{R}^d} \gamma_d(dx) = \sum_{i=1}^{d} \|\partial_i f_N\|^2_{L^2(\gamma_d)} = \sum_{k=1}^{N} k \sum_{\substack{q_1,...,q_d\in\mathbb{N} \\ q_1+...+q_d=k}} c^2_{q_1,...,q_d} \prod_{j=1}^{d} q_j!.$$

The **domain** of D, denoted by $\mathbb{D}^{1,2}(\mathbb{R}^d)$, should be the set of $f \in L^2(\gamma_d)$ absolutely continuous with $|\nabla f| \in L^2(\gamma_d)$. Then the above discussion suggests an equivalent definition, that is, for f as in (2.2), it belongs to $\mathbb{D}^{1,2}(\mathbb{R}^d)$ if and only if

$$\sum_{k=1}^{\infty} k \sum_{\substack{q_1,...,q_d\in\mathbb{N} \\ q_1+...+q_d=k}} c^2_{q_1,...,q_d} \prod_{j=1}^{d} q_j! < \infty. \tag{2.7}$$

• **Orthogonal projections**. Let J_k denote the orthogonal projection operator onto the k-th chaos $\mathbb{C}_k$. Then, for f as in (2.2), we can write $f = \sum_{k=0}^{\infty} J_k f$, with

$$J_k f = \sum_{\substack{q_1,...,q_d \in \mathbb{N} \\ q_1+...+q_d=k}} c_{q_1,...,q_d} \prod_{j=1}^{d} H_{q_j}(x_j). \tag{2.8}$$

Then, we have $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$ if and only if $\sum_{k\geq 1} k\|J_k f\|^2_{L^2(\gamma_d)} < \infty$.

Proposition 2.5 (Gaussian Poincaré inequality). *Let $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$ and $\mathbf{G} \sim \mathcal{N}_d(0, \mathbf{I})$. Then,*

$$\mathrm{Var}(f(\mathbf{G})) \leq \mathbb{E}[|\nabla f(\mathbf{G})|^2]$$

with equality only when $f(\mathbf{G}) = a_0 + \sum_{i=1}^{d} a_i G_i$ is a linear combination of the components, i.e., only when $f(\mathbf{G})$ is Gaussian.

Proof. First, we write $f = \sum_{k=0}^{\infty} J_k f$, then

$$\mathrm{Var}(f(\mathbf{G})) = \sum_{k=1}^{\infty} \|J_k f\|^2_{L^2(\gamma_d)} \leq \sum_{k=1}^{\infty} k\|J_k f\|^2_{L^2(\gamma_d)} = \mathbb{E}[|\nabla f(\mathbf{G})|^2].$$

It is clear that the above inequality becomes an equality only when $J_k f = 0$ for any $k \geq 2$. That is, $f(x) = a_0 + a \cdot x$ for some $a_0 \in \mathbb{R}$ and $a \in \mathbb{R}^d$. □

• **Divergence operator δ.** Let us now introduce the adjoint operator of D. Let $\mathrm{dom}(\delta)$ be the set of functions $g = (g_1, ..., g_d) : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\|g\|^2_{L^2(\gamma_d)} = \sum_{j=1}^{d} \|g_j\|^2_{L^2(\gamma_d)} < \infty$$

and for any $f \in C^{\infty}_{\mathrm{poly}}(\mathbb{R}^d)$,

$$\left|\mathbb{E}\langle \nabla f(\mathbf{G}), g(\mathbf{G})\rangle_{\mathbb{R}^d}\right| \leq C_g \|f\|_{L^2(\gamma_d)},$$

where $\mathbf{G} \sim \mathcal{N}_d(0, \mathbf{I})$ and $C_g \in (0, \infty)$ is a constant that depends only on g.

Note that $f \in C^{\infty}_{\mathrm{poly}} \mapsto \mathbb{E}\langle \nabla f(\mathbf{G}), g(\mathbf{G})\rangle_{\mathbb{R}^d}$ defines bounded linear functional on $C^{\infty}_{\mathrm{poly}}$, and thus extends to $L^2(\gamma_d)$ by density. Therefore, by Riesz's representation theorem (see Rudin's book), there exists a unique element in $L^2(\gamma_d)$, denoted by $\delta(g)$, such that

$$\langle \nabla f, g\rangle_{L^2(\gamma_d)} = \mathbb{E}\langle \nabla f(\mathbf{G}), g(\mathbf{G})\rangle_{\mathbb{R}^d} = \langle f, \delta(g)\rangle_{L^2(\gamma_d)} \tag{2.9}$$

for every $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$.

Remark 2.6. For $g = (g_1, ..., g_d) : \mathbb{R}^d \to \mathbb{R}^d$ with $g_j \in C^{\infty}_{\mathrm{poly}}(\mathbb{R}^d)$, we have the following explicit expression

$$\delta(g)(x) = x \cdot g - \mathrm{div} g \tag{2.10}$$

where $\mathrm{div} g = \sum_{j=1}^{d} \partial_j g_j$. Indeed, the LHS of (2.9) is equal to

$$\sum_{j=1}^{d} \mathbb{E}[g_j(\mathbf{G})\partial_j f(\mathbf{G})] = \sum_{j=1}^{d} \mathbb{E}\Big[\big(G_j g_j(\mathbf{G}) - \partial_j g_j(\mathbf{G})\big) f(\mathbf{G})\Big]$$

where the last equality follows from the 1D duality relation (1.13). This implies (2.10).

• **Ornstein-Uhlenbeck generator $\mathcal{L}$.** For $f \in C^\infty_{\text{poly}}(\mathbb{R}^d)$,

$$\mathcal{L}f = \Delta f - x \cdot \nabla f.$$

Note that $\mathcal{L} = -\delta Df$ for $f \in C^\infty_{\text{poly}}(\mathbb{R}^d)$. This suggests that the domain of $\mathcal{L}$ is

$$\text{dom}(\mathcal{L}) = \big\{ f \in \mathbb{D}^{1,2}(\mathbb{R}^d) : Df \in \text{dom}(\delta) \big\}.$$

Proposition 2.7. (i) *For $f \in \mathbb{C}_k$, we have $\mathcal{L}f = -kf$.* (ii) $f \in \text{dom}(\mathcal{L})$ *if and only if*

$$\sum_{k=1}^\infty k^2 \|J_k f\|^2_{L^2(\gamma_d)} < \infty,$$

where J_k denotes the projection operator onto $\mathbb{C}_k$ as in (2.8).

Proof. It suffices to consider $f(x_1, ..., x_d) = \prod_{i=1}^d H_{q_i}(x_i)$ with $q_1 + ... + q_d = k$. Then,[8]

$$\begin{aligned}
\mathcal{L}f &= \sum_{j=1}^d (\partial_j^2 - x_j \partial_j) f = \sum_{j=1}^d (\partial_j - x_j)\partial_j f \\
&= \sum_{j=1}^d q_j \underbrace{(\partial_j - x_j) H_{q_j - 1}(x_j)}_{= -H_{q_j}(x_j)} \prod_{i=1}^d \mathbf{1}_{\{i \neq j\}} H_{q_i}(x_i) \\
&= \Big(\sum_{j=1}^d q_j \Big) \prod_{i=1}^d H_{q_i}(x_i) = -kf,
\end{aligned}$$

where we noticed that $\partial_j g(x_j) - x_j g(x_j) = -\delta(g)(x_j)$ for $g : \mathbb{R} \to \mathbb{R}$. This proves (i).

Suppose $f \in \text{dom}(\mathcal{L})$, i.e., $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$ and $Df \in \text{dom}(\delta)$. In particular, we have $\mathcal{L}f = -\delta(Df) \in L^2(\gamma_d)$. Therefore, it follows from part (i) that

$$\|\mathcal{L}f\|^2_{L^2(\gamma_d)} = \|\mathcal{L}\sum_{k=0}^\infty J_k f\|^2_{L^2(\gamma_d)} = \|\sum_{k=0}^\infty \mathcal{L}J_k f\|^2_{L^2(\gamma_d)} = \sum_{k=1}^\infty k^2 \|J_k f\|^2_{L^2(\gamma_d)} < \infty.$$

Now let us assume that $\sum_{k=1}^\infty k^2 \|J_k f\|^2_{L^2(\gamma_d)} < \infty$ and we need to prove $f \in \text{dom}(\mathcal{L})$. It is clear that $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$ and

$$\|Df_N - Df\|^2_{L^2(\gamma_d)} \xrightarrow{N \to +\infty} 0,$$

where $f_N = \sum_{k \le N} J_k f$. It remains to show that $Df \in \text{dom}(\delta)$. Given $h \in C^\infty_{\text{poly}}(\mathbb{R}^d)$, we can write

$$\begin{aligned}
\langle Dh, Df \rangle_{L^2(\gamma_d)} &= \lim_{N \to +\infty} \langle Dh, Df_N \rangle_{L^2(\gamma_d)} \\
&= \lim_{N \to +\infty} \langle h, -\mathcal{L}f_N \rangle_{L^2(\gamma_d)} \quad \text{by duality,}
\end{aligned}$$

which is bounded by

$$\|h\|_{L^2(\gamma_d)} \sup_N \|\mathcal{L}f_N\|_{L^2(\gamma_d)} = \|h\|_{L^2(\gamma_d)} \sqrt{\sum_{k=1}^\infty k^2 \|J_k f\|^2_{L^2(\gamma_d)}}.$$

This shows $Df \in \text{dom}(\delta)$. Hence, $f \in \text{dom}(\mathcal{L})$ and our proof is completed. □

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 3 (Sept. 9, 2025)** $\infty\infty\infty\infty\infty\infty\infty$

[8] $x_j f(x)$ is viewed as the multiplication operator.

Lecture 4.

• **Ornstein-Uhlenbeck semigroups** $(P_t)_{t\in\mathbb{R}_+}$.

$$P_t = e^{tL} = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathcal{L}^k$$

Let $f \in L^2(\gamma_d)$, we deduce from Proposition 2.7 that

$$P_t J_q f = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathcal{L}^k J_q f = \sum_{k=0}^{\infty} \frac{t^k}{k!}(-q)^k J_q f = e^{-qt} J_q f.$$

Then, we have

$$P_t f = \sum_{q=0}^{\infty} e^{-qt} J_q f, \tag{2.11}$$

which is well defined for any $f \in L^2(\gamma_d)$. It is also evident that

$$\|P_t f\|^2_{L^2(\gamma_d)} = \sum_{q=0}^{\infty} e^{-2qt}\|J_q f\|^2_{L^2(\gamma_d)} \le \|f\|^2_{L^2(\gamma_d)},$$

which is known as the contraction property for the OU semigroup; see also Proposition 2.9.

Exercise 2.8. (i) *Show that for $f \in L^2(\gamma_d)$, $f \in \mathrm{dom}\mathcal{L}$ if and only if*

$$\lim_{t\downarrow 0} \frac{P_t f - f}{t}$$

exists in $L^2(\gamma_d)$ and it is equal to $\mathcal{L}f$.
(ii) *Verify that semigroup properties: $P_0 f = f$ and $P_{t+s} f = P_t P_s f$ for any $f \in L^2(\gamma_d)$.*
(iii) *Show that $DP_t f = e^{-t}P_t Df$ for any $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$ and $\partial_t P_t = \mathcal{L}P_t$ on $L^2(\gamma_d)$. Here, we should understand $P_t Df$ as a vector $(P_t\partial_1 f, ..., P_t\partial_d f)$.*

Proposition 2.9. (i) [Mehler's formula] For any $f \in L^2(\gamma_d)$, we have

$$(P_t f)(x) = \mathbb{E}\big[f(e^{-t}x + \sqrt{1-e^{-2t}}\mathbf{G})\big].$$

(ii) [contraction property] For any $f \in L^p(\gamma_d)$ with $p \in [2, \infty]$, we have

$$\|P_t f\|_{L^p(\gamma_d)} \le \|f\|_{L^p(\gamma_d)}.$$

Proof. To show Mehler's formula, it suffices to prove it for $f(x) = \prod_{j=1}^d H_{q_j}(x_j)$. and then the general result follows by linearity. In this case, $P_t f = e^{-(q_1+...+q_d)t} f$ and

$$\begin{aligned}\mathbb{E}\prod_{j=1}^d H_{q_j}(e^{-t}x_j + \sqrt{1-e^{-2t}}G_j) &= \prod_{j=1}^d \mathbb{E}H_{q_j}(e^{-t}x_j + \sqrt{1-e^{-2t}}G_j)\\ &= \prod_{j=1}^d e^{-q_j t}H_{q_j}(x),\end{aligned}$$

which follows from the binomial formula (see Question 1.13). Therefore, part (i) is proved.

Now let $f \in L^p(\gamma_d)$ for $p \in [2, \infty)$.[9] Then, the result in (i) ensures that we can write

$$(P_t f)(x) = \mathbb{E}\big[f(e^{-t}x + \sqrt{1-e^{-2t}}\mathbf{G})\big].$$

[9]The case $p = \infty$ is trivial by noting that $\|f\|_{L^\infty(\gamma_d)}$ coincides with the usual L^∞-norm of f.

Let $\mathbf{G}'$ be an independent copy of $\mathbf{G}$, we can write $\|P_t f\|^p_{L^p(\gamma_d)}$ as

$$\begin{aligned}\|P_t f\|^p_{L^p(\gamma_d)} &= \mathbb{E}\Big[\big|\mathbb{E}\big[f(e^{-t}\mathbf{G}' + \sqrt{1-e^{-2t}}\mathbf{G})|\mathbf{G}'\big]\big|^p\Big] \\ &\le \mathbb{E}\Big[\mathbb{E}\big[|f(e^{-t}\mathbf{G}' + \sqrt{1-e^{-2t}}\mathbf{G})|^p|\mathbf{G}'\big]\Big] \\ &= \mathbb{E}\Big[|f(e^{-t}\mathbf{G}' + \sqrt{1-e^{-2t}}\mathbf{G})|^p\Big],\end{aligned}$$

where the last two steps follow from Jensen's inequality and tower property for conditional expectation. Note that $e^{-t}\mathbf{G}' + \sqrt{1-e^{-2t}}\mathbf{G}$ has the same law as $\mathbf{G}$, so that we just proved $\|P_t f\|^p_{L^p(\gamma_d)} \le \|f\|^p_{L^p(\gamma_d)}$. □

Exercise 2.10. *Prove that the Mehler formula also holds for $f \in L^1(\gamma_d)$.*

Proposition 2.11 (Gaussian integration by parts)**.** (i) Let $Y \sim \mathcal{N}(0,\sigma^2)$ and $\mathbf{X} \sim \mathcal{N}_d(0,C)$ be jointly Gaussian. Then, for any $h \in C^\infty_{\text{poly}}(\mathbb{R}^d)$, it holds that

$$\mathbb{E}[Yh(\mathbf{X})] = \sum_{j=1}^d \mathbb{E}[YX_j]\mathbb{E}[\partial_j h(\mathbf{X})].$$

(ii) For $f \in C^2_{\text{poly}}(\mathbb{R}^d)$ and $\mathbf{X} \sim \mathcal{N}_d(0,C)$, we have

$$\mathbb{E}\big[\mathbf{X}\cdot\nabla f(\mathbf{X})\big] = \sum_{i,j=1}^d C_{ij}\mathbb{E}[\partial_{ij} f(\mathbf{X})] = \mathbb{E}\big[\langle C, \text{Hess} f(\mathbf{X})\rangle_{\text{HS}}\big]. \tag{2.12}$$

Proof. Let h be given as in the statement, and by scaling, we can assume $Y \sim \mathcal{N}(0,1)$. Rewriting

$$X_j = \mathbb{E}[YX_j]Y + (X_j - \mathbb{E}[YX_j]Y) =: \rho_j Y + W_j$$

as a sum of two independent Gaussians, we are able to transform our proof to the 1D case. Applying the integration by parts for Y yields

$$\begin{aligned}\mathbb{E}[Yh(\mathbf{X})] &= \mathbb{E}\big[Yh(\rho_1 Y + W_1, ..., \rho_d Y + W_d)\big] \\ &= \mathbb{E}\sum_{j=1}^d \rho_j \partial_j h(\rho_1 Y + W_1, ..., \rho_d Y + W_d) \\ &= \sum_{j=1}^d \mathbb{E}[YX_j]\mathbb{E}[\partial_j h(\mathbf{X})].\end{aligned}$$

This proves (i). To prove (ii), it suffices to apply (i) to obtain

$$\mathbb{E}\big[X_i\partial_i f(\mathbf{X})\big] = \sum_{j=1}^d \mathbb{E}[X_iX_j]\mathbb{E}\big[\partial_j\partial_i f(\mathbf{X})\big]$$

and conclude by linearity. □

2.3. **Application I: A gentle touch on GMC.** For an in-depth introduction to the theory of Gaussian multiplicative chaos (GMC), please refer to the survey [RV14] by Rhodes and Vargas.

Let $\mathcal{D} \subset \mathbb{R}^d$ be bounded and open and ν be a finite Radon measure on $\mathcal{D}$. A standard Gaussian multiplicative chaos (GMC) is a random measure on D formally defined as

$$M(A) = \int_A e^{\varphi(x) - \frac{1}{2}\text{Var}\varphi(x)}\nu(dx), \tag{2.13}$$

where $\{\varphi(x) : x \in \mathcal{D}\}$ is a centered Gaussian field with covariance kernel

$$C(x, y) = \mathbb{E}[\varphi(x)\varphi(y)].$$

The random measure in (2.13) is well defined when φ is a regular field meaning $C(x, y)$ is finite for every $x, y \in \mathcal{D}$, which excludes the case

$$C(x, y) = \max\{\log \tfrac{1}{|x-y|}, 0\} + g(x, y) \tag{2.14}$$

with g a bounded continuous function on $\mathcal{D} \times \mathcal{D}$. The corresponding Gaussian field (with covariance kernel in (2.14)) can only be interpreted in a generalized sense, and φ can not be defined pointwise. Hence, the exponential $e^{\varphi(x)}$ does not make any sense. In 1985, J.-P. Kahane introduced GMC in his paper [Kah85] to make Mandelbrot's initial model of energy dissipation in intermittent turbulence, the so-called Kolmogorov-Obukhov model [Man72].

Specifically, Kahane started from the covariance kernel C of σ-positive type, meaning

$$C(x, y) = \sum_{k=1}^{\infty} C_k(x, y), \tag{2.15}$$

where $\{C_k\}_k$ is a sequence of $[0, \infty)$-valued, continuous, positive-definite kernels:

$$\int_{\mathcal{D}^2} f(x)f(y)C_k(x, y)\nu(dx)\nu(dy) < \infty$$

for any 'test function' $f : \mathcal{D} \to \mathbb{R}$. Suppose we have an independent sequence of centered continuous Gaussian fields $\{\varphi_k\}_k$ with covariance kernels $\{C_k\}_k$. Then,

$$\Phi_n = \sum_{k=1}^{n} \varphi_k$$

is a centered Gaussian field with covariance kernel $\sum_{k=1}^n C_k$. Kahane proved that the random measure M_n on $\mathcal{D}$, with

$$A \in \mathcal{B}(\mathcal{D}) \mapsto M_n(A) := \int_A e^{\Phi_n(x) - \frac{1}{2}\mathrm{Var}\Phi_n(x)}\nu(dx),$$

converges almost surely in the space of Radon measure (equipped with the topology of weak convergence) towards a random measure M_∞. This random measure M_∞, formally written as

$$M_\infty(A) = \int_A e^{\Phi_\infty(x) - \frac{1}{2}\mathrm{Var}\Phi_\infty(x)}\nu(dx) \quad \text{or} \quad \int_A e^{\varphi(x) - \text{“}\infty\text{”}}\nu(dx)$$

is called Gaussian multiplicative chaos with kernel C acting on ν. Indeed, from the Doob's convergence theorem for positive martingales[10], we can deduce that for any Borel set $A \subset \mathcal{D}$, as $n \to +\infty$, $M_n(A)$ converges almost surely to some random variable $M_\infty(A)$ with finite expectation. Note that the construction of GMC seems to depend on the decomposition (2.15) of C. Kahane proved in his 1985 paper that this is not the case.

Theorem 2.12 (Uniqueness in law). *Let $\{\widetilde{\varphi}_k\}$ be another independent sequence of centered continuous Gaussian field with nonnegative covariance kernels $\{\widetilde{C}_k\}$ such that*

$$C(x, y) = \sum_{k=1}^{\infty} C_k(x, y) = \sum_{k=1}^{\infty} \widetilde{C}_k(x, y) \tag{2.16}$$

[10]You can find this classic result in most books on Ito calculus.

for any $(x,y) \in \mathcal{D}\times\mathcal{D}$. *Let* $\widetilde{M}_\infty$ *be the GMC constructed in the same way as* M_∞. *Then,* $\widetilde{M}_\infty = M_\infty$ *in law, meaning that for any* $f : \mathcal{D} \to \mathbb{R}$ *bounded continuous,*

$$\int_{\mathcal{D}} g(x)\widetilde{M}_\infty(dx) = \int_{\mathcal{D}} g(x)M_\infty(dx) \quad \textit{in law.} \tag{2.17}$$

We will prove Theorem 2.12 by using the following convexity inequality (or Gaussian comparison principle), which is a consequence of the Gaussian integration by parts formula (Proposition 2.11).

Proposition 2.13 (Kahane's convexity inequality)**.** (i) *Let* $(X_1,...,X_n)$ *and* $(Y_1,...,Y_n)$ *be two centered Gaussian vectors such that*

$$\mathbb{E}[X_iX_j] \le \mathbb{E}[Y_iY_j]\,, \quad \forall i,j\in\{1,...,n\}.$$

Then, for any $(p_1,...,p_n)\in\mathbb{R}^n_+$ *and for all convex functions* $F:\mathbb{R}_+\to\mathbb{R}$ *twice differentiable with at most polynomial growth at infinity,*[11] *we have*

$$\mathbb{E}\Big[F\Big(\sum_{i=1}^n p_i e^{X_i-\frac12\mathbb{E}[X_i^2]}\Big)\Big] \le \mathbb{E}\Big[F\Big(\sum_{i=1}^n p_i e^{Y_i-\frac12\mathbb{E}[Y_i^2]}\Big)\Big]. \tag{2.18}$$

(ii) *Let* $\mathcal{D}$ *be a bounded open subset of* $\mathbb{R}^d$ *and* ν *be a finite Radon measure on* $\mathcal{D}$.[12] *Let* $\{\Phi(x)\}_{x\in\mathcal{D}}$, $\{\widetilde{\Phi}(x)\}_{x\in\mathcal{D}}$ *be real centered continuous Gaussian fields on* $\mathcal{D}$ *with covariance kernel* $K,\widetilde{K}$ *respectively such that*

$$K(x,y)\le\widetilde{K}(x,y), \forall x,y\in\mathcal{D}.$$

Let F *be as in* (i). *Then, we have for any compact* $A\subset\mathcal{D}$

$$\mathbb{E}\Big[F\Big(\int_A e^{\Phi(x)-\frac12\mathrm{Var}\Phi(x)}\nu(dx)\Big)\Big] \le \mathbb{E}\Big[F\Big(\int_A e^{\widetilde{\Phi}(x)-\frac12\mathrm{Var}\widetilde{\Phi}(x)}\nu(dx)\Big)\Big]. \tag{2.19}$$

We will postpone the proof of Proposition 2.13 to the end of this section.

Proof of Theorem 2.12. To show (2.17), it suffices to consider the case where g is nonnegative and continuous with *compact* support (say, $\mathrm{supp}(g)=A$). In this case, the two random variables in (2.17) are nonnegative. Then, to show the equality in law, it suffices to show that they have the same Laplace transform. That is, for any $\lambda\in\mathbb{R}_+$,

$$\mathbb{E}\Big[\exp\Big(-\lambda\int_{\mathcal{D}} g(x)\widetilde{M}_\infty(dx)\Big)\Big] = \mathbb{E}\Big[\exp\Big(-\lambda\int_{\mathcal{D}} g(x)M_\infty(dx)\Big)\Big]. \tag{2.20}$$

Note that from (2.16), we can **claim** that given a compact set $A\subset\mathcal{D}$, for any $\varepsilon>0$ and for any $n\in\mathbb{N}$, there exists some large $m_{n,\varepsilon}\in\mathbb{N}$ such that for any $m\ge m_{n,\varepsilon}$,

$$\sum_{k=1}^n C_k(x,y) < \varepsilon + \sum_{k=1}^m \widetilde{C}_k(x,y) \tag{2.21}$$

for any $x,y\in A$. Indeed, letting F_m be the set of $(x,y)\in A\times A$ such that (2.21) fails, then F_m is closed due to the continuity of C_k and $\widetilde{C}_k$, and thus F_m is compact. It is also clear that $F_{m+1}\subset F_m$ for each m and $\bigcap_{m\ge1}F_m =$ due to (2.16). Therefore, we deduce from Heine-Borel theorem that F_m must be empty for all large m, that is, the desired claim is verified.

[11]The polynomial growth at infinity ensures that the expectations in (2.18) are finite.

[12]In particular, ν is inner regular meaning that for any open $U\subset\mathcal{D}$, $\nu(U)=\sup\{\nu(K): K\subset U \text{ and } K \text{ is compact}\}$. This ensures that the measure ν can be approximated by weighted sum of Dirac masses of the form $\sum_{j=1}^n p_{j,n}\delta_{x_{j,n}}$, with $p_j\in\mathbb{R}_+$ and $x_{j,n}\in\mathcal{D}$.

Now let us pick m large enough such that (2.21) holds and let $K(x,y)$ denote the LHS and $\widetilde{K}(x,y)$ denote the RHS. Then $K(x,y)$ is the covariance kernel of $\Phi_n = \varphi_1(x)+\ldots+\varphi_n(x)$ and $\widetilde{K}$ is the covariance kernel of $\sqrt{\varepsilon}Z+\widetilde{\Phi}_m = \sqrt{\varepsilon}Z+\widetilde{\varphi}_1(x)+\ldots+\widetilde{\varphi}_n(x)$, with $Z\sim\mathcal{N}_1(0,1)$ independent of $\widetilde{\Phi}_m$. Therefore, it follows from Proposition 2.13-(ii) (applied to the measure $g(x)\nu(dx)$ with compact support A) that

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(-\lambda\int_{\mathcal{D}} g(x)M_n(dx)\Big)\Big] &= \mathbb{E}\Big[\exp\Big(-\lambda\int_{\mathcal{D}} e^{\Phi_n(x)-\frac12\mathrm{Var}\Phi_n(x)}g(x)\nu(dx)\Big)\Big]\\ &\le \mathbb{E}\Big[\exp\Big(-\lambda\int_{\mathcal{D}} e^{\widetilde{\Phi}_m(x)-\frac12\mathrm{Var}\widetilde{\Phi}_m(x)+\sqrt{\varepsilon}Z-\frac12\varepsilon}g(x)\nu(dx)\Big)\Big]\\ &= \mathbb{E}\Big[\exp\Big(-\lambda e^{\sqrt{\varepsilon}Z-\frac12\varepsilon}\int_{\mathcal{D}} g(x)\widetilde{M}_m(dx)\Big)\Big] \end{aligned}\tag{2.22}$$

for any $m\ge m_{n,\varepsilon}$. Then, passing $m\to+\infty$ and then $\varepsilon\to 0$, $n\to+\infty$ yields

$$\mathbb{E}\Big[\exp\Big(-\lambda\int_{\mathcal{D}} g(x)M_\infty(dx)\Big)\Big]\le\mathbb{E}\Big[\exp\Big(-\lambda\int_{\mathcal{D}} g(x)\widetilde{M}_\infty(dx)\Big)\Big]\tag{2.23}$$

and by the same argument and switching the role of C and $\widetilde{C}$, we can get the reverse inequality, implying that the above inequality is indeed an equality, which is (2.20). □

$\infty\infty\infty\infty\infty\infty$ **End of Lecture 4 (Sept. 11, 2025)** $\infty\infty\infty\infty\infty\infty\infty\infty$

Lecture 5.

Proof of Proposition 2.13. Let us first prove (ii) by assuming (i).

Since ν is a Radon measure on $\mathcal{D}$, we can construct a sequence of measures ν_n of the form

$$\nu_n=\sum_{j=1}^n p_j\delta_{x_j}\quad\text{with } p_j\in\mathbb{R}_+ \text{ and } x_j\in\mathcal{D}$$

such that ν_n weakly converges to ν meaning that for any bounded continuous function $h:\mathcal{D}\to\mathbb{R}$,

$$\int_{\mathcal{D}} h(x)\nu_n(dx)=\sum_{j=1}^n p_j h(x_j)\xrightarrow{n\to+\infty}\int_{\mathcal{D}} h(x)\nu(dx).$$

Note that part (i) implies that

$$\mathbb{E}\Big[F\Big(\int_A e^{\Phi(x)-\frac12\mathrm{Var}\Phi(x)}\nu_n(dx)\Big)\Big]\le\mathbb{E}\Big[F\Big(\int_A e^{\widetilde{\Phi}(x)-\frac12\mathrm{Var}\widetilde{\Phi}(x)}\nu_n(dx)\Big)\Big].\tag{2.24}$$

Note that the two exponentials (as functions in x) in (2.24) are bounded continuous on the compact set A, and thus we have the almost sure convergence of the two integrals over A to the corresponding integrals with respect to ν. In other words, passing $n\to\infty$ in (2.24) yields (2.19).[13] This proves (ii).

Now let us prove (i) via the smart-path method. The basic idea of comparing the terms in (2.18) is to interpolate them via some "smart" path and then show certain monotonicity along

[13]Actually, a bit more work is needed. For example, we should also check the uniform integrability of $\{F(\int_A e^{\Phi(x)-\frac12\mathrm{Var}\Phi(x)}\nu_n(dx)\}_n$, which is ensured by the finiteness of $\sup_{x\in A}\mathbb{E}[e^{p\Phi(x)}]$ and $\nu(A)$.

such a path. Without losing any generality, we will assume that the two vectors $(X_1, ..., X_n)$ and $(Y_1, ..., Y_n)$ are independent and we define

$$G_j^t = tX_j + \sqrt{1-t^2}Y_j \quad \text{and} \quad v_j^t = \mathrm{Var}G_j^t = t^2\mathrm{Var}X_j + (1-t^2)\mathrm{Var}Y_j.$$

Define

$$\Psi(t) = \mathbb{E}\Big[F\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)\Big]$$

so that LHS of (2.18) is $\Psi(1)$ while RHS of (2.18) is $\Psi(0)$. Therefore,

$$\Psi(1) - \Psi(0) = \int_0^1 \frac{d}{dt}\Psi(t)dt \leq 0$$

when $\Psi'(t) = \frac{d}{dt}\Psi(t) \leq 0$ for all $t \in (0,1)$.

By Lebesgue's differentiation theorem, we have

$$\begin{aligned}\Psi'(t) &= \mathbb{E}\Big[F'\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)\sum_{j=1}^n p_j e^{G_j^t - \frac{1}{2}v_j^t}\frac{d}{dt}\big[G_j^t - \frac{1}{2}v_j^t\big]\Big] \\ &= \sum_{j=1}^n p_j \mathbb{E}\Big[F'\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)e^{G_j^t - \frac{1}{2}v_j^t}\Big[X_j - \frac{t}{\sqrt{1-t^2}}Y_j - t(\mathrm{Var}X_j - \mathrm{Var}Y_j)\Big]\Big].\end{aligned}$$

Applying the Gaussian integration by parts (Proposition 2.11-(i)), we have

$$\begin{aligned}&\mathbb{E}\Big[\overbrace{F'\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)e^{G_j^t - \frac{1}{2}v_j^t}}^{=h(G_1^t,...,G_t^n)}X_j\Big] \\ &= \sum_{i=1}^n \mathbb{E}[X_j G_i^t]\mathbb{E}\Big[\frac{\partial}{\partial G_i^t}h(G_1^t, ..., G_t^n)\Big] \\ &= \sum_{i=1}^n t\mathbb{E}[X_i X_j]\mathbb{E}\Big[F''\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)e^{G_j^t - \frac{1}{2}v_j^t}p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big] \\ &\quad + t\mathrm{Var}(X_j)\mathbb{E}\Big[F'\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)e^{G_j^t - \frac{1}{2}v_j^t}\Big]\end{aligned} \tag{2.25}$$

and similarly,

$$\begin{aligned}&\frac{t}{\sqrt{1-t^2}}\mathbb{E}\Big[F'\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)e^{G_j^t - \frac{1}{2}v_j^t}Y_j\Big] \\ &= \sum_{i=1}^n t\mathbb{E}[Y_i Y_j]\mathbb{E}\Big[F''\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)e^{G_j^t - \frac{1}{2}v_j^t}p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big] \\ &\quad + t\mathrm{Var}(Y_j)\mathbb{E}\Big[F'\Big(\sum_{i=1}^n p_i e^{G_i^t - \frac{1}{2}v_i^t}\Big)e^{G_j^t - \frac{1}{2}v_j^t}\Big].\end{aligned} \tag{2.26}$$

Therefore, it follows from (2.25), (2.26), and $F'' \geq 0$ that $\Psi'(t) \leq 0$. Hence, the proof is completed. $\square$

• **Gaussian Poincaré inequality** for $\mathbf{X} \sim \mathcal{N}_d(0, C)$. The Gaussian Poincaré inequality in Proposition 2.5 that for $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$, the variance of $f(\mathbf{G})$ with $\mathbf{G} \sim \mathcal{N}_d(0, \mathbf{I})$, is bounded by the corresponding second moment of $\|Df(\mathbf{G})\|_{\mathbb{R}^d}$. When we replace $\mathbf{G}$ by $\mathbf{X} \sim \mathcal{N}_d(0, C)$, the same inequality holds up to a constant.

Proposition 2.14. *Let $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$ (defined with respect to γ_d) and $\mathbf{X} \sim \mathcal{N}_d(0, C)$. Then,*

$$\mathrm{Var}(f(\mathbf{X})) \le \lambda_* \mathbb{E}\big[\|(Df)(\mathbf{X})\|_{\mathbb{R}^d}^2\big],$$

where λ_* denotes the largest eigenvalue of the covariance matrix C.

Proof. First, we write $\mathbf{X} = Q\mathbf{G}$ with $\mathbf{G} \sim \mathcal{N}_d(0, \mathbf{I})$ and $Q = \sqrt{C}$.[14] Then, Proposition 2.5 implies that

$$\begin{aligned} \mathrm{Var}(f(\mathbf{X})) &= \mathrm{Var}(f(Q\mathbf{G})) \le \mathbb{E}\langle Df(Q\mathbf{G}), Df(Q\mathbf{G})\rangle_{\mathbb{R}^d} \\ &= \mathbb{E}\langle Q \cdot (Df)(Q\mathbf{G}), Q \cdot (Df)(Q\mathbf{G})\rangle_{\mathbb{R}^d} = \mathbb{E}\langle (Df)(Q\mathbf{G}), C \cdot (Df)(Q\mathbf{G})\rangle_{\mathbb{R}^d} \qquad (2.27) \\ &\le \lambda_* \mathbb{E}\big[\|(Df)(\mathbf{X})\|_{\mathbb{R}^d}^2\big], \end{aligned}$$

where we used the chain rule in (2.27) and the fact that $\langle Qx, y\rangle_{\mathbb{R}^d} = \langle x, Qy\rangle_{\mathbb{R}^d}$ due to the symmetry of Q and $QQ = C$; in the last step, we used the fact that $\langle Cx, x\rangle \le \lambda_* \|x\|_{\mathbb{R}^d}^2$. □

2.4. **Applications II: Maximum of Gaussians.** In this section, we will look at the maximum of a centered Gaussian vector.

Proposition 2.15. *Let $\mathbf{X} \sim \mathcal{N}_d(0, C)$ and define $M_d = \max\{X_1, ..., X_d\}$. Then,*

$$\mathbb{E}[M_d] \le \sqrt{2C_* \log d} \quad \text{and} \quad \mathrm{Var}(M_d) \le C_*$$

with $C_* = \max\{C_{ii} : i = 1, ..., d\}$. In particular, when $\mathbf{X}$ is standard normal, we have $C_* = 1$.

Proof. Put $f(x) = \max\{x_1, ..., x_d\}$. Then, it is easy to verify that $|f(x) - f(y)| \le \max_i\{|x_i - y_i|\}$.[15] In particular, f is absolutely continuous (even Lipschitz continuous) with

$$\partial_i f(x) = \mathbf{1}_{\{x_i = f(x)\}} \qquad (2.28)$$

for almost every $x \in \mathbb{R}^d$. Therefore, it follows from the computations in (2.27) that

$$\begin{aligned} \mathrm{Var}(f(\mathbf{X})) &\le \mathbb{E}\langle (Df)(\mathbf{X}), C \cdot (Df)(\mathbf{X})\rangle_{\mathbb{R}^d} \\ &= \mathbb{E}\sum_{i=1}^d (\partial_i f)(\mathbf{X}) \sum_{j=1}^d C_{ij} (\partial_j f)(\mathbf{X}) = \mathbb{E}\sum_{i=1}^d C_{ii} |(\partial_i f)(\mathbf{X})|^2 \quad \text{by (2.28)} \\ &\le C_* \mathbb{E}\sum_{i=1}^d \mathbf{1}_{\{X_i = M_d\}} = C_*. \end{aligned}$$

Next, we prove the upper bound for the expectation. Note that for any $\beta > 0$,

$$\max\{x_1, ..., x_d\} \le \frac{1}{\beta} \log \sum_{i=1}^d e^{\beta x_i},$$

[14] This means that Q is a symmetric and positive-definite matrix such that $QQ = C$.

[15] Suppose $x_i = f(x)$ and $f(y) = y_j \ge x_i$, then $|f(x) - f(y)| = y_j - x_i \le y_j - x_j = |x_j - x_j| \le \max_i\{|x_i - y_i|\}$.

from which, we deduce

$$
\begin{aligned}
\mathbb{E}[M_d] &\le \frac{1}{\beta}\mathbb{E}\log\sum_{i=1}^d e^{\beta X_i} \\
&\le \frac{1}{\beta}\log\sum_{i=1}^d \mathbb{E}e^{\beta X_i} \quad \text{by Jensen's inequality} \\
&= \frac{1}{\beta}\log\sum_{i=1}^d e^{\frac{1}{2}\beta^2 C_{ii}} \le \frac{1}{\beta}\log\left(d\cdot e^{\frac{1}{2}\beta^2 C_*}\right) \\
&= \frac{\log d}{\beta} + \frac{1}{2}\beta C_*.
\end{aligned}
$$

An easy optimization (with $\beta = \sqrt{\frac{2}{C_*}\log d}$) leads to $\mathbb{E}[M_d] \le \sqrt{2C_*\log d}$. $\square$

A natural question after Proposition 2.15 is the sharpness of the bounds. In fact, the order of the expectation is optimal, but the order of the variance is not, as illustrated by the following.

Question 2.16. Let $\{g_i\}_{i\ge 1}$ be i.i.d. standard normal and define $M_n = \max\{g_1, ..., g_n\}$. Prove

$$\mathbb{P}\Big(\frac{M_n - a_n}{b_n} \le x\Big) \xrightarrow{n\to+\infty} \exp(-e^{-x})$$

for every $x \in \mathbb{R}_+$, where

$$a_n = \frac{1}{b_n} = \sqrt{2\log n}.$$

The above limit is the cumulative distribution function of the so-called Gumbel distribution. You may use the tail-ratio

$$\frac{1-\Phi(u+\frac{x}{u})}{1-\Phi(u)} \xrightarrow{u\to+\infty} e^{-x},$$

where Φ stands for the standard normal CDF.

Remark 2.17. (i) The above question tells us that the variance of M_n is of order $\frac{1}{\log n}$ and the expectation is of order $\sqrt{\log n}$. An application of the Gaussian Poincaré inequality only allows us to bound the variance by a constant (i.e., 1); see Proposition 2.15.

(ii) The famous L^1/L^2 variance inequality of M. Talagrand asserts that for $\mathbf{G} \sim \mathcal{N}_d(0, \mathbf{I})$ and $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$,

$$\mathrm{Var}(f(\mathbf{G})) \le 2\sum_{i=1}^d \frac{\|\partial_i f\|^2_{L^2(\gamma_d)}}{1+\log\left(\|\partial_i f\|_{L^2(\gamma_d)}/\|\partial_i f\|_{L^1(\gamma_d)}\right)}. \tag{2.29}$$

This inequality will allow us to obtain the optimal variance bound $\mathrm{Var}(M_n) \le \frac{C}{\log n}$ in the setting of Question 2.16 (*Exercise!*). We will prove (2.29) later via an application of the hypercontractivity property of the Ornstein-Uhlenbeck semigroups.

2.5. **Applications III: Gaussian random polymer.** For a generic path of a 1D random walk starting from (0,0), we can represent it as

$$(0,0) \to (1, a_1) \to (2, a_2) \to \cdots \to (n, a_n) \to \cdots,$$

where $a_i \in \mathbb{Z}$ and $|a_i - a_{i+1}| = 1$ with $a_0 = 0$. Let $\mathcal{P}_n$ denote the set of such paths with length n, then the cardinality of $\mathcal{P}_n$ is 2^n.

• **Gaussian environment.** Let $\{g_v : v \in \mathbb{Z}^2\}$ be i.i.d. real standard normal random variables modeling the random medium. A path $p \in \mathcal{P}_n$ immersed in this medium posses the "**energy**"

$$H_n(p) = -\sum_{v \in p} g_v,$$

where "$v \in p$" means that the vertex v belongs to the path p. The quantity H_n is also known as the **Hamiltonian**. The so-called **ground-state energy** is defined by

$$E_n = \min_{p \in \mathcal{P}_n} H_n(p),$$

that is, it is the minimum energy over all possible paths of length n.

The following result indicates that the variance of the ground-state energy is at most of linear order.

Proposition 2.18. *It holds that* $\mathrm{Var}(E_n) \leq n + 1$.

Proof. It is clear that E_n can be viewed as an absolutely continuous function of the Gaussian environment $\{g_v : v \in \mathbb{Z}^2\}$ (in fact only finitely many of g_v's). Let

$$p_* = \arg\min_{p \in \mathcal{P}_n} H_n(p).$$

By definition, p_* depends on the randomness inherited from the medium (and it is almost surely unique (*why?*)). It is also evident that if we perturbate a bit the Gaussian weight g_v assigned to the vertex v, E_n will vary if and only if v belongs to the optimal path p_*. As a result, we have

$$\frac{\partial}{\partial g_v} E_n = -\mathbf{1}_{\{v \in p_*\}}$$

and thus by Gaussian Poincaré inequality, we have

$$\begin{aligned}\mathrm{Var}(E_n) &\leq \mathbb{E} \sum_{v \in \mathbb{Z}^2} \left|\frac{\partial}{\partial g_v} E_n\right|^2 \\ &= \mathbb{E} \sum_{v \in \mathbb{Z}^2} \mathbf{1}_{\{v \in p_*\}} = n + 1,\end{aligned}$$

since there are $n + 1$ vertices in any path of length n. This concludes our proof. □

The above bound is not optimal and S. Chatterjee proved in 2008 that $\mathrm{Var}(E_n) \leq C \frac{n}{\log n}$, where the constant C does not depend on n; we refer interested students to [Cha14b] for more details.

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 5 (Sept. 16, 2025)** $\infty\infty\infty\infty\infty\infty\infty\infty$

Lecture 6.

2.6. **More on Ornstein-Uhlenbeck.** Recall that we can define the OU operator $\mathcal{L}$ by using the projection operators J_k:

$$\mathcal{L}f = \sum_{k=1}^{\infty} -kJ_k f$$

for any $f \in \mathrm{dom}(\mathcal{L})$, i.e., for $f \in L^2(\gamma_d)$ with $\sum_{k=1}^{\infty} k^2 \|J_k f\|^2_{L^2(\gamma_d)} < \infty$. Similarly,

$$P_t f = e^{t\mathcal{L}} f = \sum_{k=0}^{\infty} e^{-kt} J_k f$$

for $f \in L^2(\gamma_d)$.

• **Pseudo-inverse of the O-U operator.** For $f \in L^2(\gamma_d)$ with zero mean (i.e., $\langle f, 1\rangle_{L^2(\gamma_d)} = 0$), we define

$$\mathcal{L}^{-1} f = \sum_{k=1}^{\infty} -\frac{1}{k} J_k f.$$

It is clear that $\mathcal{L}^{-1} f \in \mathrm{dom}(\mathcal{L})$ with

$$\mathcal{L}\mathcal{L}^{-1} f = f$$

and

$$\mathcal{L}^{-1}\mathcal{L} g = g - \langle g, 1\rangle_{L^2(\gamma_d)}$$

for any $g \in \mathrm{dom}(\mathcal{L})$. Here we record a few facts of the operator $\mathcal{L}^{-1}$.

Lemma 2.19. *For any $f \in L^2(\gamma_d)$ with zero mean and $g \in \mathbb{D}^{1,2}(\mathbb{R}^d)$ with zero mean, we have*

$$-\mathcal{L}^{-1} f = \int_0^\infty P_t f dt, \tag{2.30}$$

$$-D\mathcal{L}^{-1} g = \int_0^\infty e^{-t} P_t D g dt. \tag{2.31}$$

Proof. By writing $f = \sum_{k\geq 1} J_k f$, we get (with probability one)

$$\int_0^\infty P_t f dt = \int_0^\infty \sum_{k\geq 1} e^{-kt} J_k f dt = \sum_{k\geq 1} \int_0^\infty e^{-kt} J_k f dt = \sum_{k=1}^{\infty} \frac{1}{k} J_k f = -\mathcal{L}^{-1} f,$$

where the second equality follows from Fubini together with

$$\begin{aligned} \int_0^\infty \sum_{k\geq 1} e^{-kt} |J_k f| dt &= \sum_{k\geq 1} \int_0^\infty e^{-kt} |J_k f| dt = \sum_{k\geq 1} \frac{1}{k} |J_k f| \\ &\leq \Big(\sum_{k\geq 1} \frac{1}{k^2}\Big)^{\frac{1}{2}} \Big(\sum_{k\geq 1} |J_k f|^2\Big)^{\frac{1}{2}}, \end{aligned}$$

which is finite almost surely. This proves the formula (2.30).

Let $v \in \mathrm{dom}(\delta)$, then we get

$$\begin{aligned}\langle v, -D\mathcal{L}^{-1}g\rangle_{L^2(\gamma_d)} &= \Big\langle \delta(v), \int_0^\infty P_t g dt\Big\rangle_{L^2(\gamma_d)} \quad \text{by duality and (2.30)} \\ &= \int_0^\infty \langle \delta(v), P_t g\rangle_{L^2(\gamma_d)} dt \quad \text{by Fubini} \qquad (2.32) \\ &= \int_0^\infty \langle v, DP_t g\rangle_{L^2(\gamma_d)} dt \quad \text{by duality} \\ &= \int_0^\infty e^{-t}\langle v, P_t Dg\rangle_{L^2(\gamma_d)} dt \\ &= \Big\langle v, \int_0^\infty e^{-t} P_t Dg dt\Big\rangle_{L^2(\gamma_d)}, \qquad (2.33)\end{aligned}$$

where in (2.32) we used $\|P_t g\|_{L^2(\gamma_d)} \le e^{-t}\|g\|_{L^2(\gamma_d)}$ and thus

$$\int_0^\infty \|\delta(v)\|_{L^2(\gamma_d)} \|P_t g\|_{L^2(\gamma_d)} dt \le \|\delta(v)\|_{L^2(\gamma_d)} \|g\|_{L^2(\gamma_d)} < \infty;$$

and similarly in the last step, we applied Fubini. Note that the above equality holds true for any $v \in \mathrm{dom}(\delta)$ and our proof is concluded by the density of $\mathrm{dom}(\delta)$ in $L^2(\mathbb{R}^d \to \mathbb{R}^d; \gamma_d)$. □

Exercise 2.20. (i) Justify the application of Fubini in the step (2.33).
(ii) $L^2(\mathbb{R}^d \to \mathbb{R}^d; \gamma_d)$ is the space of functions f from $\mathbb{R}^d$ to $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} \|f(x)\|^2_{\mathbb{R}^d} \gamma_d(dx) < \infty$. Show that $\mathrm{dom}(\delta)$ is dense in $L^2(\mathbb{R}^d \to \mathbb{R}^d; \gamma_d)$.

• A useful covariance formula.

Proposition 2.21. *Let $f, g \in \mathbb{D}^{1,2}(\mathbb{R}^d)$. Then,*

$$\mathrm{Cov}_{\gamma_d}(f, g) = \int_0^\infty e^{-t}\langle Df, P_t Dg\rangle_{L^2(\gamma_d)} dt.$$

Proof. Without losing any generality, we can assume g has zero mean. Then,

$$\begin{aligned}\mathrm{Cov}_{\gamma_d}(f, g) &= \langle f, g\rangle_{L^2(\gamma_d)} = \langle f, \mathcal{L}\mathcal{L}^{-1}g\rangle_{L^2(\gamma_d)} \\ &= \langle f, -\delta D\mathcal{L}^{-1}g\rangle_{L^2(\gamma_d)} = \langle Df, -D\mathcal{L}^{-1}g\rangle_{L^2(\gamma_d)}.\end{aligned}$$

Using (2.31), we continue to write

$$\begin{aligned}\mathrm{Cov}_{\gamma_d}(f, g) &= \Big\langle Df, \int_0^\infty e^{-t} P_t Dg dt\Big\rangle_{L^2(\gamma_d)} \\ &= \int_0^\infty e^{-t}\langle Df, P_t Dg\rangle_{L^2(\gamma_d)} dt,\end{aligned}$$

where the last step follows from (2.33) with $v = Df \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \gamma_d)$. □

• Hypercontractivity. The contraction property of the Ornstein-Uhlenbeck semigroup ensures that for any $p \in [1, \infty]$ and any $f \in L^p(\gamma_d)$,

$$\|P_t f\|_{L^p(\gamma_d)} \le \|f\|_{L^p(\gamma_d)}, \forall t \in \mathbb{R}_+,$$

which easily follows from the Mehler formula (Proposition 2.9). In fact, when the function f evolves along the O-U semigroup, more integrability can be gained. Such an improvement is known as the hypercontractivity property of the O-U semigroups.

Theorem 2.22 (Hypercontractivity). *Let $f \in L^p(\gamma_d)$ for $p \in [1,\infty)$ and $q_t = 1 + (p-1)e^{2t}$ for some $t \in \mathbb{R}_+$. Then,*

$$\|P_t f\|_{L^{q_t}(\gamma_d)} \le \|f\|_{L^p(\gamma_d)}. \tag{2.34}$$

Let us first state a corollary of Theorem 2.22, which is very useful in practice.

Corollary 2.23 (Wiener chaos estimate). *Let $f \in \bigoplus_{k\le M} \mathbb{C}_k$, i.e., f lives in the sum of first M chaoses. Then the following statements hold true.*
(i) *for any $q > p > 1$, and for $f \in \mathbb{C}_M$, we get*

$$\|f\|_{L^q(\gamma_d)} \le \big(\tfrac{q-1}{p-1}\big)^{\frac{M}{2}} \|f\|_{L^p(\gamma_d)}. \tag{2.35}$$

(ii) *for any $q \ge 2 \ge p > 1$, the inequality* (2.35) *holds for $f \in \bigoplus_{k\le M} \mathbb{C}_k$. As a consequence, $f \in L^q(\gamma_d)$ for any $q \in [2,\infty)$.*

In particular, the inequality in Question 1.3 holds true. Do you have any other proof (direct ones) to solve Question 1.3?

Let us first prove Corollary 2.23 and then conclude this section with the proof of Theorem 2.22.

Proof of Corollary 2.23. • case (i): Suppose $f \in \mathbb{C}_M$, then $P_t f = e^{-Mt} f$ so that

$$\|f\|_{L^q(\gamma_d)} = \|P_t e^{Mt} f\|_{L^q(\gamma_d)} \le \|e^{Mt} f\|_{L^p(\gamma_d)}$$

with $\frac{q-1}{p-1} = e^{2t}$. Then, with $e^{Mt} = \big(\frac{q-1}{p-1}\big)^{\frac{M}{2}}$, we get (2.35).
• case (ii): For $f = \sum_{k\le M} J_k f$, we define

$$\eta = \sum_{k=0}^{M} e^{kt} J_k f$$

so that $P_t \eta = f$. Then, with $\frac{q-1}{2-1} = e^{2t}$, we have

$$\begin{aligned} \|f\|_{L^q(\gamma_d)} &= \|P_t \eta\|_{L^q(\gamma_d)} \\ &\le \|\eta\|_{L^2(\gamma_d)} = \bigg(\sum_{k=0}^{M} e^{2kt} \|J_k f\|^2_{L^2(\gamma_d)}\bigg)^{\frac{1}{2}} \quad \text{by } L^2\text{-orthogonality} \\ &\le e^{Mt} \|f\|_{L^2(\gamma_d)} = (q-1)^{\frac{M}{2}} \|f\|_{L^2(\gamma_d)}. \end{aligned} \tag{2.36}$$

In particular, this implies that $\bigoplus_{k\le M} \mathbb{C}_k \subset L^q(\gamma_d)$ for any finite M and q.

For $p \in (1,2]$, we put $e^{2s} = \frac{1}{p-1}$, then by hypercontractivity and L^2-orthogonality, we can write

$$\begin{aligned} \|f\|_{L^2(\gamma_d)} &\le e^{Ms} \|P_t f\|_{L^2(\gamma_d)} \le e^{Ms} \|f\|_{L^p(\gamma_d)} \\ &= (p-1)^{-\frac{M}{2}} \|f\|_{L^p(\gamma_d)}. \end{aligned} \tag{2.37}$$

Therefore, (2.35) follows by combining (2.36) and (2.37). □

Question 2.24. What about the inequality (2.35) for (i) $q \ge p \ge 2$ and (ii) $1 < p \le q \le 2$?

• *Proof of Theorem 2.22.* to be added later.

□

∝∝∝∝∝∝ **End of Lecture 6 (Sept. 18, 2025)** ∝∝∝∝∝∝∝∝

Lecture 7.

3. A pleasant detour: Stein's method of normal approximation

3.1. **What is Stein's method for?** Let us start with the classic *central limit theorem*: suppose $\{X_j, j \geq 1\}$ is a sequence of *i.i.d* random variables with zero mean and unit variance, then

$$\frac{X_1 + \ldots + X_n}{\sqrt{n}} \quad \text{converges in law to } \mathcal{N}(0,1), \text{ as } n \to +\infty.$$

A standard proof consists of using the Fourier transform so as to establish the convergence of characteristic functions. However, this analytic proof strongly relies on the independence and leaves us no clue about how fast this distributional convergence happens. Moreover, this Fourier-based approach, in general, can not be applied to situations, where some dependence arises. The *Stein's method* is a very powerful toolbox of techniques that can be used not only to prove limit theorems, but also to provide rates of convergence in some chosen metrics, for instance, the *Berry-Esséen bound* in the *Kolmogorov distance*; see, e.g., [NP12, Theorem 3.7.1].

The Stein's method is named after Charles Stein, one of the greatest statisticians in the last century. In 1972, Charles Stein published his method concerning normal approximation in [St72], and after its first appearance, this method has been modified and further developed by Stein himself as well as many other mathematicians. We refer interested readers to the treatise [CGS11] and the lecture notes [Ros11].

3.2. **Basics on Stein's method.** We know from Lemma 1.6 that if $Z \sim \mathcal{N}(0,1)$, $f \in C_b^\infty(\mathbb{R})$, then $\mathbb{E}\big[Zf(Z)\big] = \mathbb{E}\big[f'(Z)\big]$. The converse statement is also true: if Z is integrable and $\mathbb{E}\big[f'(Z)\big] = \mathbb{E}\big[Zf(Z)\big]$ for all $f \in C_b^\infty(\mathbb{R})$, then Z must be distributed as the standard Gaussian. Indeed, considering $f(x) = \sin(\lambda x)$ and $f(x) = \cos(\lambda x)$, $\lambda \in \mathbb{R}$, we have $\mathbb{E}\big[Z\sin(\lambda Z)\big] = \lambda\mathbb{E}\big[\cos(\lambda Z)\big]$ and $\mathbb{E}\big[Z\cos(\lambda Z)\big] = -\lambda\mathbb{E}\big[\sin(\lambda Z)\big]$, from which we deduce that $\mathbb{E}\big[Ze^{\mathrm{i}\lambda Z}\big] = \mathrm{i}\lambda\mathbb{E}\big[e^{\mathrm{i}\lambda Z}\big]$ for any $\lambda \in \mathbb{R}$. Recognize that $\varphi_Z(\lambda) = \mathbb{E}\big[e^{\mathrm{i}\lambda Z}\big]$ is the characteristic function of Z and $\mathrm{i}\mathbb{E}\big[Ze^{\mathrm{i}\lambda Z}\big] = \varphi_Z'(\lambda)$, which gives us an ordinary differential equation $\varphi_Z'(\lambda) + \lambda\varphi_Z(\lambda) = 0$ subject to $\varphi_Z(0) = 1$. This ODE has a unique solution $\varphi_Z(\lambda) = \exp(-\lambda^2/2)$, which is the characteristic function of $\mathcal{N}(0,1)$.

The following result, known as *Stein's lemma*, summarizes the above discussion.

Lemma 3.1 (Stein's lemma). *Let Z be an integrable random variable, then $Z \sim \mathcal{N}(0,1)$ if and only if $\mathbb{E}\big[f'(Z)\big] = \mathbb{E}\big[Zf(Z)\big]$ for any continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ such that $f'(Z), Zf(Z) \in L^1(\mathbb{P})$, and f, f' have at most* polynomial growth at infinity[16].

Heuristically speaking, the above characterization hints that a real random variable Z is *close* to a standard Gaussian, whenever $\mathbb{E}\big[f'(Z) - Zf(Z)\big]$ is close to zero for every $f \in \mathscr{H}$, with $\mathscr{H}$ being some rich enough class of functions.

Then how to quantify the distance between two probability distributions?

Let $\mathscr{H}$ be a *separating* class[17] of real bounded measurable functions and we define, for two probability measures μ, ν, that $d_{\mathscr{H}}(\mu,\nu) := \sup_{h\in\mathscr{H}} \big| \int_{\mathbb{R}} h\, d\nu - \int_{\mathbb{R}} h\, d\mu \big|$. It is trivial that $d_{\mathscr{H}}$ is a metric on the set of probability distributions on $\mathbb{R}$ and we also write $d_{\mathscr{H}}(X,Y) = d_{\mathscr{H}}(\mu,\nu)$

[16] We say F has at most *polynomial growth at infinity*, if there are some universal constants $C > 0$ and $d \in N^*$ such that $|F(x)| \leq C + C|x|^d$ for each $x \in \mathbb{R}$. The proof of Lemma 3.1 is omitted here, for it is exactly the same as in previous discussion.

[17] that is, for any two different probability measures μ, ν on $\mathbb{R}$, there exists $f \in \mathscr{H}$ such that $\int_{\mathbb{R}} f\, d\nu \neq \int_{\mathbb{R}} f\, d\mu$.

if $X \sim \mu, Y \sim \nu$. For example, $\mathscr{H}_1 := \{\mathbf{1}_A : A \in \mathcal{B}(\mathbb{R})\}$, $\mathscr{H}_2 := \{\mathbf{1}_{(-\infty,x]} : x \in \mathbb{R}\}$ and $\mathscr{H}_3 := \{f : \|f\|_\infty + \|f'\|_\infty \leq 1\}$ are three separating classes; $d_{\mathscr{H}_1}$, $d_{\mathscr{H}_2}$, $d_{\mathscr{H}_3}$ are known as the *total-variation distance* d_{TV}, the *Kolmogorov distance* d_{Kol} and the *Fortet-Mourier distance* d_{FM} respectively. It is known from [Dud10, Section 11.3] that d_{FM} induces the same topology as the weak convergence of probability measures. It is clear that if μ_n, μ are probability measures on $\mathbb{R}$, then the following implications hold in view of the Portemanteau's theorem:

$$\lim_{n\to+\infty} d_{\mathrm{TV}}(\mu_n,\mu) = 0 \quad\Longrightarrow\quad \lim_{n\to+\infty} d_{\mathrm{Kol}}(\mu_n,\mu) = 0 \quad\Longrightarrow\quad \lim_{n\to+\infty} d_{\mathrm{FM}}(\mu_n,\mu) = 0.$$

Another important distance that we will use is the *Wasserstein distance* d_{W}, defined by $d_{\mathrm{W}}(\mu,\nu) := \sup\{\int_{\mathbb{R}} h\,d\nu - \int_{\mathbb{R}} h\,d\mu : \|h'\|_\infty \leq 1\}$ for two probability measures μ and ν on $\mathbb{R}$. Note that d_{FM} is sometimes called the "bounded Wasserstein distance".

Ingeniously, Charles Stein introduced the following equation:

$$f'(x) - xf(x) = h(x) - \mathbb{E}\big[h(N)\big]\,, \tag{3.1}$$

where $N \sim \mathcal{N}(0,1)$ and $h : \mathbb{R} \to \mathbb{R}$ is a measurable function satisfying $h(N) \in L^1(\mathbb{P})$. The equation (3.1) is known as the Stein's equation with unknown h. We call f a solution to (3.1), if it is *absolutely continuous* and one version of f' satisfies (3.1) everywhere.

Let us now stop to digest the ingenuity of (3.1): suppose that we can solve (3.1) and get nice properties of f in terms of those of h, then we replace the dummy variable x in (3.1) by the random variable Z, so that after taking expectation on both sides, we get

$$\mathbb{E}\big[f'(Z) - Zf(Z)\big] = \mathbb{E}\big[h(Z) - h(N)\big]\,. \tag{3.2}$$

That is to say, in order to get uniform control of the right-hand side, one can instead try to uniformly control the left-hand side, which fits the heuristic after Stein's lemma. In essence, Stein's method replaces the complex-valued characteristic function by the above real characterizing equation, and many coupling methods have been developed so far to deal with the left-hand side of (3.2); see aforementioned references.

The following lemma gives an explicit form for solutions to Stein's equation (3.1).

Lemma 3.2 (Stein's solution). *Let N, h be given as before, the solutions f to* (3.1) *are given by*

$$f(x) = ce^{x^2/2} + e^{x^2/2}\int_{-\infty}^x \Big[h(y) - \mathbb{E}\big(h(N)\big)\Big]e^{-y^2/2}\,dy\,, \quad x \in \mathbb{R}\,,$$

where $c \in \mathbb{R}$. In particular, the function

$$f_h(x) := e^{x^2/2}\int_{-\infty}^x \Big[h(y) - \mathbb{E}\big(h(N)\big)\Big]e^{-y^2/2}\,dy\,, \quad x \in \mathbb{R} \tag{3.3}$$

is the unique solution to (3.1) *verifying* $\lim_{|x|\to+\infty} \exp(-x^2/2)f(x) = 0$. *We call f_h the Stein's solution to* (3.1).

Proof. Noticing that $e^{x^2/2}\dfrac{d}{dx}\big(f(x)e^{-x^2/2}\big) = f'(x) - xf(x) = h(x) - \mathbb{E}\big[h(N)\big]$, we have[18]

$$f(x)e^{-x^2/2} - f(0) = \int_0^x \Big[h(y) - \mathbb{E}\big[h(N)\big]\Big]e^{-y^2/2}\,dy\,.$$

[18] We follow the convention $\int_0^x = -\int_x^0$ if $x < 0$

By the dominated convergence theorem and

$$\int_{\mathbb{R}} \Big(h(y) - \mathbb{E}\big[h(N)\big]\Big)\, e^{-y^2/2}\, dy = 0\,, \tag{3.4}$$

we have $c := \lim_{x\to-\infty} e^{-x^2/2} f(x) = -\lim_{x\to+\infty} e^{-x^2/2} f(x)$. Thus, for any $x \in \mathbb{R}$,

$$f(x)e^{-x^2/2} - c = \int_{-\infty}^{x} \Big[h(y) - \mathbb{E}\big[h(N)\big]\Big] e^{-y^2/2}\, dy\,,$$

which gives us the desired form for the solutions to (3.1). And the rest is trivial. $\square$

Now starting with the expression in (3.3), we study the properties of the Stein's solution f_h when h is 1-Lipschitz function or h is merely bounded measurable:

(i) Suppose that $h : \mathbb{R} \to [0, 1]$ is measurable, then it follows from Stein's equation that

$$|f_h'(x)| \le |h(x) - \mathbb{E}[h(N)]| + |xf_h(x)| \le 1 + |xf_h(x)|$$

and from (3.4) that $\int_{-\infty}^{x} \big(h(y) - \mathbb{E}[h(N)]\big)e^{-y^2/2}dy = -\int_{x}^{\infty} \big(h(y) - \mathbb{E}[h(N)]\big)e^{-y^2/2}dy$. Thus,

$|f_h(x)| \le e^{x^2/2} \int_{|x|}^{\infty} e^{-y^2/2}\, dy \le \sqrt{\pi/2}$, and $|xf_h(x)| \le |x| \cdot e^{x^2/2} \cdot \int_{|x|}^{\infty} e^{-y^2/2}\, dy \le 1$. So we can conclude that f_h is 2-Lipschitz and uniformly bounded by $\sqrt{\pi/2}$.

(ii) Suppose that $h : \mathbb{R} \to \mathbb{R}$ is K-Lipschitz for some $K \in (0, +\infty)$, then $\|f_h\|_\infty \le 2K$, $\|f_h'\|_\infty \le \sqrt{2/\pi}K$ and $\|f_h''\|_\infty \le 2K$. The proof for these bounds involves a careful analysis, here we just give a very sketchy one and refer interested readers to [CGS11, Lemma 2.4]: first denote the standard Gaussian density, distributional functions by ϕ, Φ respectively, one starts with $h(x) - \mathbb{E}[h(N)] = \int_{\mathbb{R}} [h(x) - h(u)]\phi(u)\, du$, then it follows from an application of Fubini's theorem that $h(x) - \mathbb{E}\big[h(N)\big] = \int_{-\infty}^{x} h'(t)\Phi(t)\, dt - \int_{x}^{\infty} h'(t)[1 - \Phi(t)]\, dt$, together with which (3.3) implies

$$\begin{aligned} f_h(\mathrm{w}) &= e^{\mathrm{w}^2/2} \int_{-\infty}^{\mathrm{w}} \left(\int_{-\infty}^{x} h'(t)\Phi(t)\, dt - \int_{x}^{\infty} h'(t)[1 - \Phi(t)]\, dt \right) e^{-x^2/2}\, dx \\ &= -\sqrt{2\pi}e^{\mathrm{w}^2/2}\big(1 - \Phi(\mathrm{w})\big) \int_{-\infty}^{\mathrm{w}} h'(t)\Phi(t)\, dt \\ &\qquad - \sqrt{2\pi}e^{\mathrm{w}^2/2}\Phi(\mathrm{w}) \int_{\mathrm{w}}^{\infty} h'(t)[1 - \Phi(t)]\, dt \end{aligned}$$

and

$$\begin{aligned} f_h'(\mathrm{w}) &= \mathrm{w}f_h(\mathrm{w}) + h(\mathrm{w}) - \mathbb{E}[h(N)] \\ &= \big\{1 - \sqrt{2\pi}\mathrm{w}e^{\mathrm{w}^2/2}[1 - \Phi(\mathrm{w})]\big\} \int_{-\infty}^{\mathrm{w}} h'(t)\Phi(t)\, dt \\ &\qquad - \big\{1 + \sqrt{2\pi}\mathrm{w}e^{\mathrm{w}^2/2}\Phi(\mathrm{w})\big\} \int_{\mathrm{w}}^{\infty} h'(t)[1 - \Phi(t)]\, dt. \end{aligned}$$

So the first two bounds follow from some standard Gaussian computations. Similar computations can be done for $\|f_h''\|_\infty$ once we derive $f_h''(\mathrm{w}) = (1 + \mathrm{w}^2)f_h(\mathrm{w}) + \mathrm{w}(h(\mathrm{w}) - \mathbb{E}[h(N)]) + h'(\mathrm{w})$ using (3.1).

(iii) Combining the above two points, we can also assert that if $h \in C_b^\infty(\mathbb{R})$ satisfies $0 \le h \le 1$, then $\|f_h\|_\infty \le \sqrt{\pi/2}$, $\|f_h'\|_\infty \le 2$ and $\|f_h''\|_\infty \le 2\|h'\|_\infty < +\infty$.

Now we give Stein's bounds based on the above discussions.

Proposition 3.3. *Let F be a real integrable random variable and $N \sim \mathcal{N}(0, 1)$, then*

(1) $d_{\mathrm{TV}}(F,N) \le \sup \big|\mathbb{E}[\varphi'(F) - F\varphi(F)]\big|$, *where the supremum runs over all* $\varphi \in C^2(\mathbb{R})$ *with* $\|\varphi\|_\infty \le \sqrt{\pi/2}$, $\|\varphi'\|_\infty \le 2$ *and* $\|\varphi''\|_\infty < +\infty$.

(2) $d_{\mathrm{W}}(F,N) \le \sup \big|\mathbb{E}[\varphi'(F) - F\varphi(F)]\big|$, *where the supremum runs over all* $\varphi \in C^2(\mathbb{R})$ *with* $\|\varphi\|_\infty \le 2$, $\|\varphi'\|_\infty \le \sqrt{2/\pi}$ *and* $\|\varphi''\|_\infty \le 2$.

Proof. It follows first from *Lusin's theorem*[19] that

$$\sup_{\substack{h \text{ measurable}\\ 0\le h\le 1}} \big|\mathbb{E}\big(h(N) - h(F)\big)\big| = \sup_{\substack{h\in C(\mathbb{R})\\ 0\le h\le 1}} \big|\mathbb{E}\big(h(N) - h(F)\big)\big|.$$

And for $h : \mathbb{R} \to [0,1]$ continuous, by standard mollification procedure, one can construct $h_\varepsilon \in C_b^\infty(\mathbb{R})$ such that $0 \le h_\varepsilon \le 1$ and $h_\varepsilon \to h$ pointwise, as $\varepsilon \downarrow 0$. So

$$d_{\mathrm{TV}}(F,N) = \sup_{\substack{h\in C_b^\infty(\mathbb{R})\\ 0\le h\le 1}} \big|\mathbb{E}\big(h(N) - h(F)\big)\big| \le \sup_{\substack{\|\varphi\|_\infty \le \sqrt{\pi/2}\\ \|\varphi'\|_\infty\le 2\\ \|\varphi''\|_\infty<+\infty}} \big|\mathbb{E}[\varphi'(F) - F\varphi(F)]\big|,$$

where the last inequality follows from point (iii) on last page and (3.2), this proves (1) while (2) follows from point (ii) on last page and (3.2). $\square$

Let us now look at an easy example.

Example 1. Let $\mathbf{Y} = (Y_i, i \in \mathbb{N})$ be i.i.d symmetric Bernoulli random variables, then $F_n := n^{-1/2}(Y_1 + \ldots + Y_n)$ converges in law to the standard Gaussian, according to the classic central limit theorem. Now fix $\varphi \in C^2(\mathbb{R})$ with $\|\varphi\|_\infty \le 2$, $\|\varphi'\|_\infty \le \sqrt{2/\pi}$ and $\|\varphi''\|_\infty \le 2$, then we have

$$\begin{aligned}
\mathbb{E}\big[F_n\varphi(F_n)\big] &= \sqrt{n}\,\mathbb{E}\big[Y_1\varphi(F_n)\big] \quad \text{using symmetry}\\
&= \frac{\sqrt{n}}{2}\mathbb{E}\Big[\varphi\Big(\frac{1 + Y_2 + \ldots + Y_n}{\sqrt{n}}\Big) - \varphi\Big(\frac{-1 + Y_2 + \ldots + Y_n}{\sqrt{n}}\Big)\Big] \quad \text{using independence}\\
&= \frac{\sqrt{n}}{2}\mathbb{E}\Big[\varphi\Big(\frac{1 + Y_2 + \ldots + Y_n}{\sqrt{n}}\Big) - \varphi(F_n) + \varphi(F_n) - \varphi\Big(\frac{-1 + Y_2 + \ldots + Y_n}{\sqrt{n}}\Big)\Big]\\
&= \mathbb{E}\big[\varphi'(F_n)\big] + \frac{\sqrt{n}}{2}R_n \quad \text{with } |R_n| \le 2\|\varphi''\|_\infty n^{-1},
\end{aligned} \tag{3.5}$$

where the last line follows from the usual Taylor expansion. Thus, it follows from Proposition 3.3 that[20] $d_{\mathrm{W}}(F_n, \mathcal{N}(0,1)) \le 2n^{-1/2}$. **Exercise**: work out the details for (3.5).

In the above toy-example, independence and symmetry within the structure contribute to the easy proof. Nevertheless, the strategy of combining an application of Taylor expansion with Stein's equation is usually effective in practice; see, e.g., our Proposition 3.4 and Theorem 3.10. Stein's ideas have contributed to many excellent results and solutions to numerous problems, such as local dependence, minimal spanning trees, concentration inequalities, machine learning, and many others; see, e.g., Barbour and Chen's Volumes [BC04, BC05] and S. Chatterjee's ICM survey [Cha14a].

$\infty\infty\infty\infty\infty\infty$ **End of Lecture 7 (Sept. 23, 2025)** $\infty\infty\infty\infty\infty\infty\infty\infty$

[19] Let $\mu = (\mathbb{P}\circ F^{-1} + \mathbb{P}\circ N^{-1})/2$, then for h measurable with values in $[0,1]$, by appropriate application of Lusin's Theorem (*e.g.* see the *Red Rudin*) one can find a sequence of continuous functions h_n such that $0 \le h_n \le 1$ and $h_n \to h$ μ-a.s., thus $h_n \to h$ $\mathbb{P}\circ F^{-1}$-a.s. and $\mathbb{P}\circ N^{-1}$−a.s..

[20] By explicit calculation, $\mathbb{E}[F_n^4] - 3 = -2n^{-1}$, so we obtain the **fourth moment bound** $d_{\mathrm{W}}(F_n, \mathcal{N}(0,1)) \le \sqrt{2(3 - \mathbb{E}[F_n^4])}$.

Lecture 8.

3.3. **Stein's method of exchangeable pairs.** The exchangeable pairs approach within Stein's method was first used in the paper [Dia77] which, however, attributed the method to Charles Stein. Later, this technique was presented in a systematic way in Stein's monograph [St86]. We recall that a pair (X, X') of random elements on a common probability space is said to be *exchangeable*, if (X, X') has the same distribution as (X', X). In the book [St86, Lecture III], it is highlighted that a given real random variable W is close in distribution to a standard normal variable N, whenever one can construct an exchangeable pair (W, W') such that W' is close to W in *some* sense and such that the *linear regression property*

$$\mathbb{E}\big[W' - W \,\big|\, \mathcal{G}\big] = -\lambda W \tag{3.6}$$

is satisfied for some small $\lambda > 0$ and $\mathrm{Var}\big(\frac{1}{2\lambda}\mathbb{E}[(W'-W)^2|\mathcal{G}]\big)$ is small, where $\mathcal{G}$ is a σ-algebra that contains $\sigma\{W\}$. Note the relation (3.6) forces W to be centered.

Theorem 3.4 (C. Stein, 1986)**.** *Let (W, W') be an exchangeable pair of random variables defined on a common probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ such that $\mathrm{Var}(W) = 1$, (3.6) holds for some $\lambda > 0$ and $W \in L^3(\mathbb{P})$, then we have*

$$d_{\mathrm{W}}\big(W, N\big) \le \sqrt{2/\pi}\sqrt{\mathrm{Var}\left(\frac{1}{2\lambda}\mathbb{E}\big[(W-W')^2|\mathcal{G}\big]\right)} + \frac{1}{2\lambda}\mathbb{E}\big[|W-W'|^3\big]\,,$$

where $N \sim \mathcal{N}(0,1)$.

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ belong to $C^2(\mathbb{R})$ with $\|f\|_\infty \le 2$, $\|f'\|_\infty \le \sqrt{2/\pi}$ and $\|f''\|_\infty \le 2$, then it is clear that $f(W')W$ and $f(W)W$ are integrable. It follows from the exchangeability that $\mathbb{E}\big[(W-W')f(W)\big] = \mathbb{E}\big[(W'-W)f(W')\big]$, from which we get $\mathbb{E}\big[(W-W')[f(W)+f(W')]\big] = 0$. So

$$\begin{aligned}
0 &= \mathbb{E}\big[(W-W')[-f(W)+f(W')]\big] + 2\mathbb{E}\big[(W-W')f(W)\big]\\
&= \mathbb{E}\big[(W-W')[-f(W)+f(W')]\big] + 2\mathbb{E}\big[f(W)\mathbb{E}(W-W'|\mathcal{G})\big]\\
&= \mathbb{E}\big[(W-W')[-f(W)+f(W')]\big] + 2\lambda\mathbb{E}\big[f(W)W\big] \quad \text{by (3.6)}
\end{aligned}$$

that is, $\mathbb{E}\big[(W'-W)[f(W')-f(W)]\big] = 2\lambda\mathbb{E}\big[f(W)W\big]$; and particularly, $\mathbb{E}[(W'-W)^2] = 2\lambda$. Therefore,

$$\begin{aligned}
&2\lambda\mathbb{E}\big[f(W)W - f'(W)\big] = \mathbb{E}\Big((W'-W)[f(W')-f(W)] - 2\lambda f'(W)\Big)\\
&= \mathbb{E}\Big(f'(W)\big[(W'-W)^2 - 2\lambda\big]\Big) + \mathbb{E}\left[(W'-W)^3\int_0^1 (1-t)f''(W+t(W'-W))\,dt\right]\\
&= \mathbb{E}\Big[f'(W)\Big(\mathbb{E}\big[(W'-W)^2|\mathcal{G}\big] - 2\lambda\Big)\Big] + \mathbb{E}\left[(W'-W)^3\int_0^1 (1-t)f''(W+t(W'-W))\,dt\right],
\end{aligned}$$

which together with Proposition 3.3 gives us the desired bound. □

In the following, we will apply Stein's bound of exchangeable pairs to obtain a rate of convergence for the CLT under the assumption of finite fourth moment.

Proposition 3.5 (Quantitative CLT via exchangeable pairs)**.** *Let $\{X_i\}_i$ be i.i.d. with mean zero and variance one such that $\mathbb{E}[|X_1|^4] < \infty$ Then with $V_n = \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)$,*

$$d_{\mathrm{W}}(V_n, \mathcal{N}(0,1)) \le \frac{C}{\sqrt{n}},$$

where $C = \left(\frac{\mathbb{E}[X_1^4]-1}{2\pi}\right)^{\frac{1}{2}} + 4\mathbb{E}[|X_1|^3]$.

Proof. Let us first construct the exchangeable pair:

$$V_n' = V_n + \frac{X_I' - X_I}{\sqrt{n}},$$

where $\{X_i'\}_i$ is an independent copy of $\{X_i\}_i$, and I is a uniform random variable with values in $\{1, ..., n\}$ such that I is independent from $\{X_i\}_i$ and $\{X_i'\}_i$. We claim that V_n and V_n' are exchangeable. Indeed, for any $t, s \in \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{P}(V_n \le t, V_n' \le s) &= \sum_{k=1}^n \mathbb{P}(V_n \le t, V_n' \le s, I = k) \\
&= \sum_{k=1}^n \mathbb{P}\Big(U_k + \frac{X_k}{\sqrt{n}} \le t, U_k + \frac{X_k'}{\sqrt{n}} \le s, I = k\Big) \quad \text{with } U_k = V_n - \frac{X_k}{\sqrt{n}} \\
&= \sum_{k=1}^n \frac{1}{n}\mathbb{P}\Big(U_k + \frac{X_k}{\sqrt{n}} \le t, U_k + \frac{X_k'}{\sqrt{n}} \le s\Big) \\
&= \sum_{k=1}^n \frac{1}{n}\mathbb{P}\Big(U_k + \frac{X_k'}{\sqrt{n}} \le t, U_k + \frac{X_k}{\sqrt{n}} \le s\Big)
\end{aligned}$$

by the independence among U_k, X_k, X_k'. Then, we have

$$\begin{aligned}
\mathbb{P}(V_n \le t, V_n' \le s) &= \sum_{k=1}^n \mathbb{P}\Big(U_k + \frac{X_k'}{\sqrt{n}} \le t, U_k + \frac{X_k}{\sqrt{n}} \le s, I = k\Big) \\
&= \mathbb{P}(V_n' \le t, V_n \le s).
\end{aligned}$$

That is, $(V_n, V_n') \overset{\text{(law)}}{=} (V_n', V_n)$. Next, we write, with $\mathcal{G} = \sigma\{X_1, ..., X_n\}$,

$$\begin{aligned}
\mathbb{E}[V_n' - V_n|\mathcal{G}] &= \frac{1}{\sqrt{n}}\mathbb{E}[X_I' - X_I|\mathcal{G}] = \frac{1}{\sqrt{n}}\sum_{k=1}^n \mathbb{E}[(X_k' - X_k)\mathbf{1}_{\{I=k\}}|\mathcal{G}] \\
&= \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{k=1}^n \mathbb{E}[X_k' - X_k|\mathcal{G}] \quad \text{as } I \text{ is independent from } \mathcal{G} \\
&= -\frac{1}{n}\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \quad \text{since } \mathbb{E}[X_k'|\mathcal{G}] = 0 \\
&= -\frac{1}{n}V_n.
\end{aligned}$$

That is, (3.6) holds true with $(W, W', \lambda) = (V_n, V_n', \frac{1}{n})$.

Next, by similar arguments, we can have

$$\begin{aligned}
\mathbb{E}\big[(V_n - V_n')^2|\mathcal{G}\big] &= \frac{1}{n}\mathbb{E}\big[(X_I - X_I')^2|\mathcal{G}\big] = \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\big[(X_k - X_k')^2|\mathcal{G}\big] \\
&= \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\big[|X_k|^2 - 2X_kX_k' + |X_k'|^2|\mathcal{G}\big] = \frac{1}{n^2}\sum_{k=1}^n (X_k^2 + 1).
\end{aligned}$$

Thus,

$$\mathrm{Var}\left(\frac{1}{2\lambda}\mathbb{E}\big[(V_n - V_n')^2|\mathcal{G}\big]\right) = \mathrm{Var}\left(\frac{1}{2n}\sum_{k=1}^n (1 + X_k^2)\right) = \frac{1}{4n}\big(\mathbb{E}[X_1^4] - 1\big).$$

Finally, let us look at the third moment of the difference $V_n' - V_n$:

$$\frac{1}{2\lambda}\mathbb{E}[|V_n' - V_n|^3] = \frac{1}{2\sqrt{n}}\mathbb{E}[|X_I - X_I'|^3] = \frac{1}{2\sqrt{n}}\sum_{k=1}^{n}\frac{1}{n}\mathbb{E}[|X_k - X_k'|^3]$$
$$= \frac{1}{2\sqrt{n}}\mathbb{E}[|X_1 - X_1'|^3] \le \frac{1}{2\sqrt{n}}\mathbb{E}\big[4|X_1|^3 + 4|X_1'|^3\big] = \frac{4}{\sqrt{n}}\mathbb{E}[|X_1|^3].$$

Combining the above bounds, we can get

$$d_{\mathrm{W}}(V_n, \mathcal{N}(0,1)) \le \Big\{\big(\frac{\mathbb{E}[X_1^4]-1}{2\pi}\big)^{\frac{1}{2}} + 4\mathbb{E}[|X_1|^3]\Big\} n^{-\frac{1}{2}}.$$

Hence, the proof is completed. □

3.4. **2nd-order Gaussian Poincaré inequality.** The Gaussian Poincaré inequality (Proposition 2.5) indicates that the variance of $F = f(\mathbf{G})$, with $f \in \mathbb{D}^{1,2}(\mathbb{R}^d)$ and $\mathbf{G} \sim \mathcal{N}_d(0, \mathbf{I})$, is controlled by the average magnitude of DF (i.e., $\mathbb{E}[\|DF\|^2_{\mathbb{R}^d}]$). That is, when the latter is small, the random variable F has small variance, i.e., small fluctuation from its mean. The bound in Proposition 2.5) becomes an exact equality only when $F = f(\mathbf{G})$ is a linear combination of the components $G_1, ..., G_d$, that is, when F is a Gaussian random variable. Naively, if f is close to an affine function, then F would be expected to be close to a Gaussian distribution (with matched mean and variance). Then, if we have a good control over the "second-order" derivatives of f (i.e., Hessf($\mathbf{G}$)), we should have a proximity of F to a Gaussian. It is S. Chatterjee who first came up with this idea and made it rigorous via Stein's method; see [Cha09]. Let us illustrate this idea more with the quadratic form. Let

$$F = \langle \mathbf{G}, A\mathbf{G}\rangle_{\mathbb{R}^d} = \sum_{i,j=1}^{d} A_{ij} G_i G_j,$$

where $A = (A_{ij})$ is a symmetric $d \times d$ matrix. By the orthogonal decomposition of a real symmetric matrix, we can write

$$A = Q^T \Lambda Q,$$

where $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_d)$ is the diagonal matrix with entries being the eigenvalues of A, Q is an orthogonal matrix (i.e., $QQ^T = \mathbf{I}$). Then, we can rewrite $F = \sum_{i=1}^d \lambda_i Y_i^2$, with $(Y_1, ..., Y_d) = Q\mathbf{G}$ again a standard Gaussian vector. By the CLT heuristic, F is approximately Gaussian if no eigenvalue dominates in the sum, or $\max \lambda_i^2$ is much smaller than $\sum_i \lambda_i^2$. Note that for $f(x) = \langle x, Ax\rangle_{\mathbb{R}^d}$, we have the Hessian matrix

$$\mathrm{Hess}f(x) = 2A \quad \text{so that the operator norm } \|\mathrm{Hess}f(x)\|_{\mathrm{op}} = 2\max_i |\lambda_i|,$$

while $\mathrm{Var}(F) = 2\sum_{i=1}^d \lambda_i^2$. This leads to the heuristic that F is approximately Gaussian if the typical value of $\|\mathrm{Hess}f(x)\|^2_{\mathrm{op}}$ is much smaller than the variance of F.

Theorem 3.6 (2nd-order GPI, [Cha09])**.** *Suppose $f \in C^2(\mathbb{R}^d)$ and $F = f(\mathbf{G})$ with $\mathbf{G} \sim \mathcal{N}_d(0, \mathbf{I})$. Suppose F has zero mean and variance $\sigma^2 \in (0, \infty)$. Then, we have*

$$d_{\mathrm{TV}}\big(F, \mathcal{N}(0,\sigma^2)\big) \le \frac{3}{\sigma^2}\mathbb{E}\Big[\|\mathrm{Hess}f(\mathbf{G})\|^4_{\mathrm{op}}\Big]^{\frac{1}{4}}\mathbb{E}\Big[\|\nabla f(\mathbf{G})\|^4_{\mathbb{R}^d}\Big]^{\frac{1}{4}}. \tag{3.7}$$

Proof. By the definition of total-variation distance, we have $d_{\mathrm{TV}}(F, \mathcal{N}(0,\sigma^2)) = d_{\mathrm{TV}}(\frac{F}{\sigma}, \mathcal{N}(0,1))$. Then, it is enough to consider the case where $\mathrm{Var}(F) = \sigma^2 = 1$. According to Stein's method, it suffices to prove a uniform bound for the expectation

$$\mathbb{E}[F\varphi(F)] - \mathbb{E}[\varphi'(F)] \ \text{ with } \|\varphi'\|_\infty \le 2.$$

In the rest of the proof, we assume that the RHS of (3.7) is finite, otherwise the proof is trivial. With F centered, we can write $F = \mathcal{L}\mathcal{L}^{-1}F$, and therefore with $\mathcal{L} = -\delta\nabla$ ($\nabla = D$),

$$\begin{aligned}\mathbb{E}[F\varphi(F)] &= \mathbb{E}[-\delta\nabla\mathcal{L}^{-1}F\varphi(F)] = \mathbb{E}[\langle -\nabla\mathcal{L}^{-1}F, \nabla\varphi(F)\rangle_{\mathbb{R}^d}] \quad \text{by duality relation}\\ &= \mathbb{E}[\langle -\nabla\mathcal{L}^{-1}F, \nabla F\rangle_{\mathbb{R}^d}\varphi'(F)] \quad \text{by chain rule,}\end{aligned}$$

and in particular, we have $1 = \mathbb{E}[F^2] = \mathbb{E}[\langle -\nabla\mathcal{L}^{-1}F, \nabla F\rangle_{\mathbb{R}^d}]$. Thus, the expression $\mathbb{E}[F\varphi(F)] - \mathbb{E}[\varphi'(F)]$ is uniformly bounded by $2\sqrt{\mathrm{Var}(\langle -\nabla\mathcal{L}^{-1}F, \nabla F\rangle_{\mathbb{R}^d})}$. From previous sections, we know that

$$-\nabla\mathcal{L}^{-1}F = \int_0^\infty e^{-t}P_t\nabla F dt \quad \text{and} \quad A\cdot(-\nabla\mathcal{L}^{-1}F) = \int_0^\infty e^{-t}A\cdot(P_t\nabla F)dt; \tag{3.8}$$

by the same arguments, we can get (with $\nabla^2 f = \mathrm{Hess} f \in \mathbb{R}^{d\times d}$)

$$-\nabla^2\mathcal{L}^{-1}F = \int_0^\infty e^{-2t}P_t\nabla^2 F dt \quad \text{and} \quad (-\nabla^2\mathcal{L}^{-1}F)\cdot y = \int_0^\infty e^{-2t}(P_t\nabla^2 F)\cdot y\, dt, \tag{3.9}$$

where $A\cdot y$ means that the matrix $A\in\mathbb{R}^{d\times d}$ acts on the vector $y\in\mathbb{R}^d$,

Applying the Gaussian Poincaré inequality, we have

$$\begin{aligned}\sqrt{\mathrm{Var}(\langle -\nabla\mathcal{L}^{-1}F, \nabla F\rangle_{\mathbb{R}^d})} &\le \sqrt{\mathbb{E}\big[\|\nabla\langle -\nabla\mathcal{L}^{-1}F, \nabla F\rangle_{\mathbb{R}^d}\|^2_{\mathbb{R}^d}\big]}\\ &\le \sqrt{\mathbb{E}\big[\|(-\nabla^2\mathcal{L}^{-1}F)\cdot(\nabla F)\|^2_{\mathbb{R}^d}\big]}\\ &\qquad + \sqrt{\mathbb{E}\big[\|(\nabla^2 F)\cdot(-\nabla\mathcal{L}^{-1}F)\|^2_{\mathbb{R}^d}\big]},\end{aligned} \tag{3.10}$$

where in the last step, we applied Minkowski's inequality. Now let us deal with the first term in (3.10): using (3.9) with Minkowski's inequality, we write

$$\begin{aligned}\Big\|\|(-\nabla^2\mathcal{L}^{-1}F)\cdot(\nabla F)\|_{\mathbb{R}^d}\Big\|_{L^2(\mathbb{P})} &\le \Big\|\int_0^\infty e^{-2t}\|P_t\nabla^2 F\|_{\mathrm{op}}\|\nabla F\|_{\mathbb{R}^d}dt\Big\|_{L^2(\mathbb{P})}\\ &\le \int_0^\infty e^{-2t}\Big\|\|P_t\nabla^2 F\|_{\mathrm{op}}\|\nabla F\|_{\mathbb{R}^d}\Big\|_{L^2(\mathbb{P})}dt\\ &\le \int_0^\infty e^{-2t}\big\|\|P_t\nabla^2 F\|_{\mathrm{op}}\big\|_{L^4(\mathbb{P})}\big\|\|\nabla F\|_{\mathbb{R}^d}\big\|_{L^4(\mathbb{P})}dt\\ &\le \int_0^\infty e^{-2t}\big\|\|\nabla^2 F\|_{\mathrm{op}}\big\|_{L^4(\mathbb{P})}\big\|\|\nabla F\|_{\mathbb{R}^d}\big\|_{L^4(\mathbb{P})}dt \quad (*)\\ &= \frac{1}{2}\big\|\|\nabla^2 F\|_{\mathrm{op}}\big\|_{L^4(\mathbb{P})}\big\|\|\nabla F\|_{\mathbb{R}^d}\big\|_{L^4(\mathbb{P})},\end{aligned}$$

where we used the contraction property of the O-U semigroup in the step $(*)$. The second term in (3.10) can be dealt with in the same way, so that the proof is concluded. □

Exercise 3.7. *Work out the details for the step $(*)$ and the details for the second term in (3.10).*

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 8 (Sept. 25, 2025)** $\propto\propto\propto\propto\propto\propto\propto\propto$

Lecture 9. A guest lecture by Nathan Ross

3.5. **Stein's method via local dependency.** In the last two lectures, we learned the basics of Stein's method, which is a powerful toolbox of establishing central limit theorems with rate of convergence. Let F be a real-valued random variable of interest, which is expected to be approximately normal with matched mean and matched variance. Say, $\mathbb{E}[F] = 0$ and $\mathrm{Var}(F) = 1$. The fundamental Stein's bounds (see, e.g., Proposition 3.3) tell us that in order to bound the distributional distance between F and a standard normal $Z \sim \mathcal{N}(0,1)$, it is enough for us to control the expectation $\mathbb{E}[f(F)F] - \mathbb{E}[f'(F)]$ with f in a suitable family of test functions. To achieve this goal, one often requires to utilize the particular structure of the random variable F, for example, independence and symmetry inherited from the construction of the statistic F (see Example 1), and the exchangeable pairs structure as in Theorem 3.4. As we have seen in Proposition 3.5, the construction of an exchangeable pair approach requires one to introduce the **auxiliary randomness**, which is quite standard within the Stein's method.

In what follows, we will introduce a notion of local dependency, with which we can combine the Stein's bounds as well. This notion goes back to Charles Stein and Louis H.Y. Chen [St72, Cha75] in the 70s, and it was extensively studied in the paper [CS04] by Chen and Shao under various dependence settings. See also [BR89, CR10] for more details. Let us start with a simple definition.

Definition 3.8. Given a collection of random variables $\mathbb{X} = \{X_i\}_{i\in\mathcal{V}}$ and a graph $\mathcal{G}$ with vertex set $\mathcal{V}$, we say $\mathcal{G}$ is a **dependency graph** for $\mathbb{X}$ if the following property holds:

$$\text{whenever } I, J \subset \mathcal{V} \text{ are disjoint with no edge between them,}$$
$$\{X_i\}_{i\in I} \text{ and } \{X_j\}_{j\in J} \text{ are independent.}$$

In particular, if there is no edge between $i, j \in \mathcal{V}$, then X_i is independent from X_j; in other words, the neighbors of the vertex i, together with i itself, form a **dependency neighborhood**.

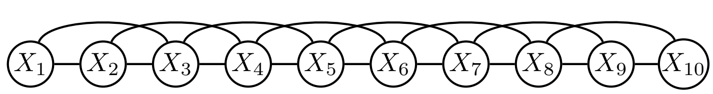
Here are a few examples of dependency graphs.

Example 2. (i) If $\mathbb{X} = \{X_1, ..., X_n\}$ is a family of independent random variables, any graph on $[n]$ is a dependency graph for $\mathbb{X}$. For example, the empty graph (i.e., empty edge set) and the complete graph (i.e., every two vertices are connected by an edge) are dependency graphs for $\mathbb{X}$. (*Which one is better?*)

(ii) [Sliding window statistics] Fix an integer $k \geq 1$. Let $Y_1, ..., Y_{n+k-1}$ be independent random variables with values in some measure space $(E, \mathcal{E}, \mu)$, and $\varphi_i : E^k \to \mathbb{R}$ be measurable. Define

$$X_i = \varphi_i(Y_i, Y_{i+1}, ..., Y_{i+k-1}),\ i = 1, ..., n.$$

When $k = 1$, we are back to the independent case as in (i). When $(k, n) = (3, 10)$, we have an example of dependency graph:



(edges connect variables with $|i-j| < 3$])

One can see that the maximum degree is 4 in the above picture.[21] The above is the smallest possible dependency graph.

[21] We write $i \sim j$ if the two (distinct) vertices $i, j \in [n]$ are directly connected by an edge. The degree of a vertex i is defined by $\deg(i) = \mathrm{card}\{j : j \sim i\}$, the number of vertices that are directly connected to i via an edge. The maximum degree is $\max\{\deg(i) : i \in [n]\}$.

Lemma 3.9. *Suppose $\{Z_v\}_{v\in\mathcal{V}}$ is a family of square-integrable random variables admitting a dependency graph with maximum degree D. Then,*

$$\mathrm{Var}\Big(\sum_{v\in\mathcal{V}} Z_v\Big) \le (D+1)\sum_{v\in\mathcal{V}}\mathbb{E}[Z_v^2].$$

Proof. We begin with

$$\mathrm{Var}\Big(\sum_{v\in\mathcal{V}} Z_v\Big) = \sum_{u,v\in\mathcal{V}}\mathrm{Cov}(Z_u,Z_v),$$

and note that for each u, there are at most $(D+1)$ vertices v such that $\mathrm{Cov}(Z_u,Z_v)\neq 0$. Using $|\mathrm{Cov}(Z_u,Z_v)|\le \|Z_u\|_2\|Z_v\|_2 \le \frac{1}{2}\big(\mathbb{E}[Z_v^2]+\mathbb{E}[Z_u^2]\big)$, we get

$$\begin{aligned}\mathrm{Var}\Big(\sum_{v\in\mathcal{V}} Z_v\Big) &\le \frac{1}{2}\sum_{u,v\in\mathcal{V}}\big(\mathbb{E}[Z_v^2]+\mathbb{E}[Z_u^2]\big)\mathbf{1}_{\{\mathrm{Cov}(Z_u,Z_v)\neq 0\}}\\ &= \sum_{u,v\in\mathcal{V}}\mathbb{E}[Z_v^2]\mathbf{1}_{\{\mathrm{Cov}(Z_u,Z_v)\neq 0\}} \quad \text{by symmetry}\\ &\le (D+1)\sum_{v\in\mathcal{V}}\mathbb{E}[Z_v^2].\end{aligned}$$

Hence, the proof is completed. □

In the following, we present a central limit theorem in case of the local dependency.

Theorem 3.10. *Let $\{X_i\}_{1\le i\le n}$ be a family of centered random variables with $\mathbb{E}[X_i^4]<\infty$ for each i. Assume that the associated dependency graph has maximum degree $\mathfrak{m}$, and*

$$\sigma := \sqrt{\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big)} > 0.$$

Then, with $F := \frac{1}{\sigma}\sum_{i=1}^n X_i$, we have

$$d_{\mathrm{W}}(F,\mathcal{N}(0,1)) \le (\mathfrak{m}+1)^2\underbrace{\Big(\sigma^{-3}\sum_{i=1}^n\mathbb{E}[|X_i|^3]\Big)}_{\text{1st term}} + \frac{2\sqrt{2/\pi}}{\sigma^2}(\mathfrak{m}+1)^{3/2}\sqrt{\sum_{i=1}^n\mathbb{E}[|X_i|^4]}. \tag{3.11}$$

The first term in (3.11) appears in the usual CLT, and the second term is often of comparable order. It is easy to see that when $\{X_i\}$ are i.i.d., then the above bound (3.11) is consistent with the one in Proposition 3.5.

Proof of Theorem 3.10. In view of Proposition 3.3, it suffices to bound the following expectation

$$\mathbb{E}[\varphi'(F)-F\varphi(F)]$$

for $\varphi\in C^2(\mathbb{R})$ with $\|\varphi\|_\infty\le 2$, $\|\varphi'\|_\infty\le\sqrt{2/\pi}$, and $\|\varphi''\|_\infty\le 2$.

Let $\{\mathcal{N}_i : i\in[n]\}$ be the dependency neighborhood, i.e., $\mathcal{N}_i=\{j : j\leftrightsquigarrow i\}$, where $j\leftrightsquigarrow i$ means $j=i$ or j,i are distinct and connected by an edge (denoted by $i\sim j$). Then, the cardinality $\mathrm{card}(\mathcal{N}_i)$ of $\mathcal{N}_i$ is bounded by $\mathfrak{m}+1$, and we also have the following equivalence:

$$j\in\mathcal{N}_i \iff i\leftrightsquigarrow j \iff i\in\mathcal{N}_j. \tag{3.12}$$

Put

$$F_i = F - \frac{1}{\sigma}\sum_{j\in\mathcal{N}_i} X_j$$

for $i \in [n]$. Then, F_i is independent from X_i. In particular, we have

$$\mathbb{E}[\varphi(F_i)X_i] = 0. \tag{3.13}$$

Now let us deal with the expectation $\mathbb{E}[F\varphi(F)]$ to derive some expression like $\mathbb{E}[\varphi'(F)\eta]$:

$$\begin{aligned}\mathbb{E}[F\varphi(F)] &= \frac{1}{\sigma}\sum_{i=1}^n \mathbb{E}[X_i\varphi(F)]\\ &= \frac{1}{\sigma}\sum_{i=1}^n \mathbb{E}\big[X_i\big(\varphi(F)-\varphi(F_i)\big)\big] \quad \text{by (3.13)} \qquad (3.14)\\ &= \Big(\frac{1}{\sigma}\sum_{i=1}^n \mathbb{E}\big[X_i\varphi'(F)(F-F_i)\big]\Big) - \frac{1}{2\sigma}\sum_{i=1}^n \mathbb{E}\big[X_i\varphi''(\widehat{F}_i)(F-F_i)^2\big],\end{aligned}$$

where in the last step, we did Taylor expansion for φ around F with $\widehat{F}_i$ some random variable with values between F and F_i. Let us first do some algebra and deduce from (3.14), with $\|\varphi'\|_\infty \le \sqrt{2/\pi}$, and $\|\varphi''\|_\infty \le 2$, that

$$\begin{aligned}\big|\mathbb{E}[F\varphi(F)] - \mathbb{E}[\varphi'(F)]\big| &\le \Big|\mathbb{E}\Big[\varphi'(F)\Big(-1+\frac{1}{\sigma^2}\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}X_jX_i\Big)\Big]\Big| + \frac{\|\varphi''\|_\infty}{2\sigma^3}\sum_{i=1}^n\mathbb{E}\Big|X_i\big(\sum_{j\in\mathcal{N}_i}X_j\big)^2\Big|\\ &\le \frac{\sqrt{2/\pi}}{\sigma^2}\,\mathbb{E}\Big|-\sigma^2+\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}X_jX_i\Big| + \frac{1}{\sigma^3}\sum_{i=1}^n\mathbb{E}\Big|X_i\big(\sum_{j\in\mathcal{N}_i}X_j\big)^2\Big|\\ &=: \frac{\sqrt{2/\pi}}{\sigma^2}\,\mathbf{T}_1 + \frac{1}{\sigma^3}\mathbf{T}_2. \qquad (3.15)\end{aligned}$$

• **bounding $\mathbf{T}_1$.** First we can deduce from (3.14) with $\varphi(x) = x$ that

$$\mathbb{E}\Big(\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}X_jX_i\Big) = \sigma^2.$$

For simplicity, we write $\sum_{i=1}^n\sum_{j\in\mathcal{N}_i} = \sum_{i\leftrightsquigarrow j}$ in view of the equivalence (3.12). Next, by Cauchy-Schwarz, we get

$$\big(\mathbf{T}_1\big)^2 \le \mathrm{Var}\Big(\sum_{i\leftrightsquigarrow j}X_iX_j\Big) = \sum_{i\leftrightsquigarrow j}\sum_{k\leftrightsquigarrow \ell}\mathrm{Cov}(X_iX_j, X_kX_\ell). \tag{3.16}$$

The above covariance is nonzero only when $k \in \mathcal{N}_i \cup \mathcal{N}_j$ or $\ell \in \mathcal{N}_i \cup \mathcal{N}_j$ for any given (i,j) with $i \leftrightsquigarrow j$. The number of pairs (k,ℓ) with the above conditions and subject to $k \leftrightsquigarrow \ell$ can be bounded as follows:

(1) given (i,j), the cardinality of $\mathcal{N}_i \cup \mathcal{N}_j$ is bounded by $2(\mathfrak{m}+1)$, then there are at most $2(\mathfrak{m}+1)$ vertices k in $\mathcal{N}_i \cup \mathcal{N}_j$;
(2) given (i,j) and fixing $k \in \mathcal{N}_i \cup \mathcal{N}_j$, we have at most $(\mathfrak{m}+1)$ vertices $\ell \leftrightsquigarrow k$
(3) given (i,j), if $k \notin \mathcal{N}_i \cup \mathcal{N}_j$, then we must have $\ell \in \mathcal{N}_i \cup \mathcal{N}_j$, there are at most $2(\mathfrak{m}+1)$ such vertices ℓ, and we have at most $(\mathfrak{m}+1)$ vertices $k \leftrightsquigarrow \ell$;
(4) thus, given (i,j) with $i \leftrightsquigarrow j$, there are at most $4(\mathfrak{m}+1)^2$ nonzero terms in (3.16).

Therefore, if we view $\{X_iX_j\}_{i\in[n],j\in\mathcal{N}_i}$ as a new family of random variables with a dependency graph, then the graph has maximum degree $\leq 4(\mathfrak{m}+1)^2-1$.[22] Thus, we deduce from Lemma 3.9 with (3.16), $ab\leq\frac{1}{2}(a^2+b^2)$, and (3.12) that

$$\begin{aligned}\big(\mathbf{T}_1\big)^2 &\leq 4(\mathfrak{m}+1)^2\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}\mathbb{E}[X_i^2X_j^2]\\ &\leq 4(\mathfrak{m}+1)^2\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}\frac{1}{2}\mathbb{E}[X_i^4+X_j^4]\\ &\leq 4(\mathfrak{m}+1)^3\sum_{i=1}^n\mathbb{E}[X_i^4].\end{aligned}\tag{3.17}$$

• **bounding $\mathbf{T}_2$.** Using the inequality of arithmetic and geometric means, we have

$$\mathbb{E}|X_iX_jX_k|\leq\frac{1}{3}\mathbb{E}\big(|X_i|^3+|X_j|^3+|X_k|^3\big).$$

Thus,

$$\begin{aligned}\mathbf{T}_2 &\leq \sum_{i=1}^n\sum_{k,j\in\mathcal{N}_i}\mathbb{E}|X_iX_jX_k|\leq\frac{1}{3}\sum_{i=1}^n\sum_{k,j\in\mathcal{N}_i}\mathbb{E}\big(|X_i|^3+|X_j|^3+|X_k|^3\big)\\ &=\left(\frac{1}{3}\sum_{i=1}^n\mathbb{E}\big(|X_i|^3\big)\sum_{k,j\in\mathcal{N}_i}1\right)+\frac{2}{3}\sum_{i=1}^n\sum_{k,j\in\mathcal{N}_i}\mathbb{E}\big(|X_k|^3\big)\\ &\leq\left(\frac{(\mathfrak{m}+1)^2}{3}\sum_{i=1}^n\mathbb{E}\big(|X_i|^3\big)\right)+\frac{2(\mathfrak{m}+1)}{3}\sum_{i=1}^n\sum_{k\in\mathcal{N}_i}\mathbb{E}\big(|X_k|^3\big)\\ &\leq(\mathfrak{m}+1)^2\sum_{i=1}^n\mathbb{E}\big(|X_i|^3\big).\end{aligned}$$

Together with (3.17) and (3.15), it leads to the desired conclusion (3.11). □

• **Application I: Headruns.** Fix an integer $k\geq1$. Let $Y_1,...,Y_{n+k-1}$ be i.i.d. Bernoulli random variables with success probability $p\in(0,1)$. Then

$$\prod_{j=i}^{i+k-1}Y_j\tag{3.18}$$

is equal to one if there is a *run* of ones of length k starting at the position i. The k-run statistic is a classical concept in probability and combinatorial statistics that counts the number of consecutive sequences (runs) of length k in a binary or categorical sequence. It has powerful applications in DNA sequence analysis, where nucleotides (A, C, G, T) form a symbolic sequence that can be modeled as a random string. Here is a short example of a DNA sequence – a real-like but synthetic one (not from any specific organism):

ATGCGTACCGTTAGCTGACCTGAAATTTGCGCGGCTAA,

where each letter represents one nucleotide base: A = Adenine, T = Thymine, C = Cytosine, and G = Guanine. In the above sequence (length =38), there are only two 3-runs. The statistics in

[22] −1 comes from the exclusion of a vertex itself when counting its degree.

(3.18) can model, for instance, the similarity between two DNA sequences. Here is an example:

sequence 1: ATG**CG**TACCGT**T**A**GC**T**G**ACC**T**G**A**AAT**T**TGCGC**G**GCTAA
sequence 2: GAT**CG**GTAACC**T**T**GC**A**G**GAT**T**A**A**GCG**T**CCGAT**G**CAAGC
matched? : $\underline{✓✓}$ ✓ $\underline{✓✓}$ ✓ ✓ ✓ ✓ ✓

we see that there are only two 2-runs of consecutive matches.

It is a natural question to ask about how many k-runs there are, and what is its asymptotic behavior as the number of observations tends to infinity. That is, letting

$$W_n = \sum_{i=1}^{n} \prod_{j=i}^{i+k-1} Y_j, \tag{3.19}$$

we are interested at the fluctuation of properly scaled W_n as $n \to +\infty$.

Proposition 3.11. *Let W_n be given as in* (3.19)*. Then,*

$$d_{\mathrm{W}}\big(\tfrac{1}{\sigma_n}(W_n - \mathbb{E}[W_n]), \mathcal{N}(0,1)\big) \lesssim \frac{k^2}{\sqrt{n}}, \tag{3.20}$$

where the above implicit constant depends on (p,k) *but not on* n*.*

Proof. Let $X_i = \prod_{j=i}^{i+k-1} Y_j$. Then, we are in the setting of Example 2-(ii). It is not difficult to see that the corresponding dependency graph for $\{X_1, ..., X_n\}$ has maximum degree $\mathfrak{m} \le 2(k-1)$. Then, it is enough to apply Theorem 3.10. Let us first do some moment estimates.

• **Basic moments.** Since $X_i := \prod_{j=i}^{i+k-1} Y_j = \mathbf{1}_{\{Y_i = \cdots = Y_{i+k-1} = 1\}}$, we have

$$\mathbb{E}[X_i] = p^k \quad \text{and} \quad \mathrm{Var}(X_i) = p^k(1-p^k).$$

It is also easy to verify that for $j \ge 1$, $\mathbb{E}[X_i X_{i+j}] = p^{k+j}\mathbf{1}_{\{1 \le j \le k-1\}} + p^{2k}\mathbf{1}_{\{j \ge k\}}$. Therefore,

$$\begin{aligned}
\sigma_n^2 := \mathrm{Var}(W_n) &= \sum_{i,j=1}^{n} \big(p^{k+|j-i|} - p^{2k}\big)\mathbf{1}_{\{|j-i| \le k\}} \\
&= np^k(1-p^k) + 2\sum_{1 \le i < j \le n} \big(p^{k+|j-i|} - p^{2k}\big)\mathbf{1}_{\{|j-i| \le k\}} \\
&\ge np^k(1-p^k),
\end{aligned} \tag{3.21}$$

where the second term in (3.21) is clearly strictly positive and of order n as $n \to \infty$ (for fixed p,k). That is, we have $\sigma_n \asymp \sqrt{n}$, although the above lower bound suffices for our purpose.

• **Dependency graph bound.** Set $\xi_i = X_i - \mathbb{E}X_i$. Then, $\{\xi_i\}$ satisfy the dependency graph condition with $\mathfrak{m} \le 2(k-1)$. Then, it follows from Theorem 3.10 that

$$d_{\mathrm{W}}\Big(\frac{1}{\sigma_n}\sum_{i=1}^{n} \xi_i, \mathcal{N}(0,1)\Big) \le (\mathfrak{m}+1)^2\Big(\sigma_n^{-3}\sum_{i=1}^{n}\mathbb{E}[|\xi_i|^3]\Big) + \frac{2\sqrt{2/\pi}}{\sigma_n^2}(\mathfrak{m}+1)^{3/2}\sqrt{\sum_{i=1}^{n}\mathbb{E}[|\xi_i|^4]}.$$

Since $|\xi_i| \le 1$, we have the trivial bounds: $\mathbb{E}|\xi_i|^3 \le 1$ and $\mathbb{E}|\xi_i|^4 \le 1$. Thus, the desired bound (3.20) follows immediately from (3.21). □

Remark 3.12. The bound in (3.20) holds for fixed (p,k), while as one can see from our proof, a finer estimate of the variance σ_n^2 could lead to an asymptotically negligible rate of convergence when k,p,n vary together in a certain way.

• **Application II: Triangles in Erdős-Rényi random graphs.** Erdős-Rényi random graphs mark the birth of modern random graph theory (1959-1961); see, e.g., the book [vdH17] for more details. An Erdős-Rényi random graph can be constructed as follows:

(i) Let K_n be a complete graph on the vertex set $[n]$, i.e., every two distinct vertices are connected by an edge;
(ii) let E_n be the edge set of K_n, and let $\{Y_e\}_{e\in E_n}$ be a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(Y_e = 1) = 1 - \mathbb{P}(Y_e = 0) = p \in (0, 1)$;
(iii) when $Y_e = 1$, we say the edge e is open; the subgraph of K_n with the vertex set $[n]$ and the open edges is called an Erdős-Rényi random graph, denoted by $\mathcal{G}(n, p)$.

Let T be the number of triangles formed in the Erdős-Rényi random graph $\mathcal{G}(n, p)$. The following result suggests that T is asymptotically normal in some regime where $p \to 0$ along $n \to \infty$; see, e.g., [BKR89, Rol22] for more results.

Proposition 3.13. *Let the above notations prevail. It holds for any $n \ge 3$ and $p \in (0, 1)$ that*

$$\sigma^2 := \mathrm{Var}(T) = \binom{n}{3} p^3[1 - p^3 + 3(n - 3)p^2(1 - p)]$$

and

$$\begin{aligned} d_{\mathrm{W}}\Big(\frac{1}{\sigma}\big(T - \mathbb{E}[T]\big),\ \mathcal{N}(0, 1)\Big) \le{}& \frac{(3n - 8)^2}{\sigma^3}\binom{n}{3} p^3(1 - p^3)[(1 - p^3)^2 + p^6] \\ &+ \frac{\sqrt{26}(3n - 8)^{\frac{3}{2}}}{\sigma^2\sqrt{\pi}}\sqrt{\binom{n}{3} p^3(1 - p^3)[(1 - p^3)^3 + p^9]}. \end{aligned} \tag{3.22}$$

In particular, if $p \sim n^{-\alpha}$ with $\alpha \in [0, \frac{2}{9})$, the bound in (3.22) *tends to zero.*

Proof. to be added later (not very soon). Or you can try to fill in the details. □

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 9 (Sept. 30, 2025)** $\infty\infty\infty\infty\infty\infty\infty\infty$

Lecture 10.

4. **Redo the analysis in an isonormal framework**

Previously, we have developed basic Gaussian analysis with finitely many Gaussian random variables. From now on, we will develop relevant stochastic analysis on infinitely many Gaussian random variables or a Gaussian process (e.g., Brownian motion, white noises). We will work on the isonormal framework.

Definition 4.1 (Isonormal Gaussian process)**.** Let $\mathfrak{H}$ be a real separable Hilbert space[23]. We say $W = \{W(h)\}_{h\in\mathfrak{H}}$ is an isonormal Gaussian process indexed by $\mathfrak{H}$ if W is a centered Gaussian family with $\mathbb{E}[W(f)W(g)] = \langle f, g\rangle_{\mathfrak{H}}$, $\forall f, g \in \mathfrak{H}$.

Remark 4.2. The mapping $W : \mathfrak{H} \to L^2(\Omega, \mathbb{P})$ is a bounded linear map: for any $\alpha, \beta \in \mathbb{R}$ and $f, g \in \mathfrak{H}$, we have

$$W(\alpha f + \beta g) = \alpha W(f) + \beta W(g) \quad \text{in } L^2(\mathbb{P}) \text{ and almost surely.}$$

Indeed, it is easy to verify by using the covariance structure of W that

$$\|W(\alpha f + \beta g) - \alpha W(f) - \beta W(g)\|^2_{L^2(\mathbb{P})} = 0,$$

which concludes the desired linearity. The operator norm of W is clearly one, since $\|W(h)\|_2 = \|h\|_{\mathfrak{H}}$ for any $h \in \mathfrak{H}$.

Proposition 4.3 (Existence of isonormal Gaussian process)**.** Let $\{e_i\}_{i\geq 1}$ be an orthonormal basis of $\mathfrak{H}$. Let $\{G_i\}_{i\geq 1}$ be i.i.d. standard normal random variables. Define

$$W(h) = \sum_{i=1}^{\infty} \langle h, e_i\rangle_{\mathcal{H}} G_i,$$

where the above series converges in $L^2(\mathbb{P})$. Then, W is an isonormal Gaussian process indexed by $\mathfrak{H}$.

Proof. From the orthonormality of $\{e_i\}$, we have

$$\sum_{i=1}^{\infty} \langle h, e_i\rangle^2_{\mathfrak{H}} = \|h\|^2_{\mathfrak{H}}.$$

In particular, we can see $W(h)$ as a $L^2(\Omega)$-limit of $\sum_{i=1}^{N} \langle h, e_i\rangle_{\mathcal{H}} G_i$ as $N \to +\infty$. Therefore, for any $h, g \in \mathfrak{H}$, we have

$$\begin{aligned}
\mathbb{E}[W(h)W(g)] &= \lim_{N\to+\infty} \mathbb{E}\bigg(\sum_{i,j\leq N} \langle h, e_i\rangle_{\mathfrak{H}} G_i \langle g, e_j\rangle_{\mathfrak{H}} G_j \bigg) \\
&= \lim_{N\to+\infty} \sum_{i,j\leq N} \langle h, e_i\rangle_{\mathfrak{H}} \langle g, e_j\rangle_{\mathfrak{H}} \mathbb{E}[G_i G_j] \\
&= \lim_{N\to+\infty} \sum_{i\leq N} \langle h, e_i\rangle_{\mathfrak{H}} \langle g, e_i\rangle_{\mathfrak{H}} = \langle h, g\rangle_{\mathfrak{H}}.
\end{aligned}$$

Hence, the covariance structure holds and the proof is completed. $\square$

In the following, we present a few examples of isonormal processes.

---

[23]By the basic Hilbert space theory, $\mathfrak{H}$ is isometric to the sequence space $\ell^2(\mathbb{N})$, which is the set of square-summable sequences of real numbers $a = \{a_j\}_{j\in\mathbb{N}}$ with $\sum_{i\in\mathbb{N}} a_i^2 < \infty$.

Example 3. (i) [Standard Gaussian vector on $\mathbb{R}^d$] Let $\mathbf{G} = (G_1, \dots, G_d) \sim \mathcal{N}_d(0, \mathbf{I})$ be a standard d-dimensional Gaussian vector, and take $\mathfrak{H} = \mathbb{R}^d$ equipped with the standard Euclidean inner product

$$\langle h, k \rangle_{\mathbb{R}^d} = \sum_{i=1}^d h_i k_i.$$

With $\{e_j\}_{j=1,\dots,d}$ the canonical basis of $\mathbb{R}^d$, we can realize $\mathbf{G}$ as an isonormal process indexed by $\mathfrak{H} = \mathbb{R}^d$:

$$G_j := W(e_j)$$

for $j = 1, ..., d$.

(ii) [General Gaussian vector on $\mathbb{R}^d$] Let $\mathbf{Z} \sim \mathcal{N}_d(0, C)$ be a centered Gaussian vector on $\mathbb{R}^d$ with covariance matrix C. With the decomposition $C = Q^2$ for some symmetric positive-definite matrix Q, we can write $\mathbf{Z} = Q\mathbf{G}$. Using (i) and with

$$h_i := \sum_{j=1}^d Q_{ij} e_j \in \mathbb{R}^d,$$

we have $Z_i = W(h_i)$ for every $i \in \{1, ..., d\}$.

(iii) [Continuous Gaussian process] Let $(G_t)_{t \geq 0}$ be a real-valued centered Gaussian process that is continuous in time almost surely. Define the following L^2-Gaussian Hilbert space

$$\mathcal{H}_G := L^2(\mathbb{P})\text{-closure of } \left\{ \sum_{i=1}^m \alpha_i G_{t_i} : \alpha_i \in \mathbb{Q}, t_i \in \mathbb{Q}_+, m \in \mathbb{N} \right\}.$$

By continuity, it is clear that G_t is contained in $\mathcal{H}_G$ for any $t \in \mathbb{R}_+$. By construction, $\mathcal{H}_G$ is a real separable Hilbert space, and thus isometric to some real separable Hilbert space $\mathfrak{H}$ (this is an existence argument, without further information on G, we can not really say more on $\mathfrak{H}$). Let W be the isometry mapping from $\mathfrak{H}$ to $\mathcal{H}_G$:

$$h \in \mathfrak{H} \mapsto W(h) \in \mathcal{H}_G \subset L^2(\Omega, \sigma\{G\}, \mathbb{P}),$$

which is the (abstract) isonormal Gaussian process generated by the Gaussian process G. Moreover, there exist a family $\{e_t\}_{t \geq 0} \subset \mathfrak{H}$ such that $G_t = W(e_t)$, $t \geq 0$. It is clear that

$$\mathbb{E}[W(e_s)W(e_t)] = \langle e_s, e_t \rangle_{\mathfrak{H}}.$$

Once we know more information on the covariance structure of G, we can say more on the associated inner product $\langle \bullet, \bullet \rangle_{\mathfrak{H}}$, with $\mathfrak{H}$ known as the **reproducing kernel Hilbert space** of G.

(iv) [Brownian motion] Let $(B_t)_{t \in \mathbb{R}_+}$ be a standard real Brownian motion on $[0, \infty)$. It is a real-valued, centered continuous Gaussian process with

$$\mathbb{E}[B_t B_s] = t \wedge s, \ \forall t, s \in \mathbb{R}_+.$$

For $f \in L^2(\mathbb{R}_+)$ (i.e., $\int_0^\infty f(s)^2 ds < \infty$), we can define the **Wiener integral**

$$W(f) := \int_0^\infty f(s) dB_s,$$

which is a centered Gaussian random variable with variance $\|f\|^2_{L^2(\mathbb{R}_+)}$. It is easy to see that $W = \{W(f)\}_{f \in L^2(\mathbb{R}_+)}$ is an isonormal process indexed by $\mathfrak{H} = L^2(\mathbb{R}_+)$. For any $t \in \mathbb{R}_+$, we have

$$B_t = W(\mathbf{1}_{[0,t]}) = \int_0^\infty \mathbf{1}_{[0,t]}(s) dB_s.$$

That is, the reproducing kernel Hilbert space for the Brownian motion is given by $\mathfrak{H} = L^2(\mathbb{R}_+)$.

(v) [white noise as derivative of Brownian motion] If we take the formal derivative of Brownian motion $\xi(t) := \frac{d}{dt}B_t$, we can write

$$W(f) = \int_0^\infty f(s)\xi(s)ds$$

and the (generalized) correlation kernel of ξ is given by

$$\mathbb{E}[\xi(t)\xi(s)] = \delta_0(t-s),$$

where δ_0 is the Dirac-delta function at zero.[24] Indeed, via a *formal* calculation, we have

$$\begin{aligned}\mathbb{E}[W(f)W(g)] &= \int_0^\infty \int_0^\infty g(t)f(s)\mathbb{E}[\xi(t)\xi(s)]dsdt \\ &= \int_0^\infty \int_0^\infty g(t)f(s)\delta_0(t-s)dsdt = \langle f, g\rangle_{L^2(\mathbb{R}_+)},\end{aligned} \tag{4.1}$$

and confirms the heuristic that the white noise on $\mathbb{R}_+$ is the formal derivative of the Brownian motion. Moreover, the generalized function δ_0 is the Fourier transform of constant power spectral density, and thus bears the name "white" noise:

$$\delta_0(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ix\xi}dx \tag{4.2}$$

That is, the corresponding **spectral measure** is a multiple of Lebesgue measure.

(vi) [Space-time white noise] The space-time white noise on $\mathbb{R}_+ \times \mathbb{R}^d$ generalizes the idea in (v) to generalized random fields indexed by both time and space. Formally, a **space-time white noise** ξ is a centered Gaussian family $\xi = \{\xi(t,x) : (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d\}$ with correlation structure

$$\mathbb{E}[\xi(t,x)\xi(s,y)] = \delta_0(t-s)\delta_0(x-y),$$

where the above two Dirac delta functions are defined on $\mathbb{R}_+$ and $\mathbb{R}^d$ respectively. Heuristically speaking, ξ is "very *rough/irregular*" and should be interpreted as a random generalized function, and thus it is only defined by paring with a nice test function: given any $f \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$, the pairing

$$\langle f, \xi\rangle := \int_{\mathbb{R}_+}\int_{\mathbb{R}^d} f(t,x)\xi(t,x)dtdx$$

is expected to a centered Gaussian random variable. Performing the same formal calculation as in (4.1), we have for any $f, g \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d)$,

$$\begin{aligned}\mathbb{E}[\langle f,\xi\rangle\langle g,\xi\rangle] &= \mathbb{E}\left[\left(\int f(t,x)\xi(t,x)dtdx\right)\left(\int g(s,y)\xi(s,y)dsdy\right)\right] \\ &= \mathbb{E}\left[\int_{\mathbb{R}_+\times\mathbb{R}^d}\int_{\mathbb{R}_+\times\mathbb{R}^d} f(t,x)g(s,y)\xi(t,x)\xi(s,y)dtdxdsdy\right] \\ &= \int_{\mathbb{R}_+\times\mathbb{R}^d}\int_{\mathbb{R}_+\times\mathbb{R}^d} f(t,x)g(s,y)\underbrace{\mathbb{E}[\xi(t,x)\xi(s,y)]}_{\delta_0(t-s)\delta_0(x-y)}dtdxdsdy \\ &= \ldots\ldots \\ &= \langle f, g\rangle_{L^2(\mathbb{R}_+\times\mathbb{R}^d)}.\end{aligned}$$

That is, we have defined a centered Gaussian family $\{W(f) : f \in C_c^\infty(\mathbb{R}_+ \times \mathbb{R}^d)\}$, whose existence is guaranteed by Kolmogorov's theorem. Then, by a standard density argument, we can extend the definition to the whole space $L^2(\mathbb{R}_+ \times \mathbb{R}^d)$. In this way, we have realized the space-time

[24] δ_0 is a generalized function defined by testing it with test functions in $C^\infty_{\mathrm{poly}}(\mathbb{R})$: $\langle \delta_0, \varphi\rangle = \varphi(0)$ for any $\varphi \in C^\infty_{\mathrm{poly}}(\mathbb{R})$.

white noise $\xi = \{\xi(f) = \langle f, \xi\rangle : f \in L^2(\mathbb{R}_+ \times \mathbb{R}^d)\}$ as an isonormal Gaussian process indexed by the Hilbert space $\mathfrak{H} = L^2(\mathbb{R}_+ \times \mathbb{R}^d)$. The space-time white noise often appears as the natural driving noise in stochastic partial differential equations (SPDEs). For example, the following stochastic heat equation with multiplicative white noise

$$\underbrace{\partial_t u(t,x) - \frac{1}{2}\Delta u(t,x)}_{\text{heat equation}} = \underbrace{u(t,x)\xi(t,x)}_{\text{random potential}}$$

is the well-known **parabolic Anderson model** (PAM). We will study the well-posedness of this PAM in a later section.

(vii) [Colored noise] As opposed to the white noise in (vi), the correlation kernel of a Gaussian colored noise is not Dirac delta, or equivalently, the corresponding spectral density is not constant. One popular family of Gaussian colored noises are the fractional Gaussian noises with Hurst index H. Formally, $\dot{W}^H$ is a fractional Gaussian noise on $\mathbb{R}$ with Hurst index H if

4.1. **Chaos expansion.**

Theorem 4.4 (Wiener–Itô chaos decomposition)**.** *Let $W = \{W(h)\}_{h\in\mathfrak{H}}$ be an isonormal Gaussian process. Then, the space $L^2(\Omega, \sigma(W), \mathbb{P})$ admits an orthogonal decomposition:*

$$L^2(\Omega, \sigma(W), \mathbb{P}) = \bigoplus_{n=0}^{\infty} \mathbb{C}_n^W, \tag{4.3}$$

where $\mathbb{C}_0^W$ is the set of constant random variables, $\mathbb{C}_k^W$ is called the k-th Wiener chaos and defined as

$$\mathbb{C}_k^W = L^2(\mathbb{P})\text{-closure of span}\left\{\prod_{j=1}^{\infty} H_{q_j}(W(e_j)) : \sum_{j=1}^{\infty} q_j = k\right\} \tag{4.4}$$

with H_q the q-th Hermite polynomial and $\{e_j\}_{j\geq1}$ an orthonormal basis of $\mathfrak{H}$.[25] *Moreover, we can further represent $\mathbb{C}_k^W$ as*

$$\mathbb{C}_k^W = L^2(\mathbb{P})\text{-closure of span}\big\{H_k(W(h)) : \|h\|_{\mathfrak{H}} = 1\big\}. \tag{4.5}$$

Proof. The k-th chaos $\mathbb{C}_k^W$ is a closed subspace of $L^2(\Omega, \sigma\{W\}, \mathbb{P})$, then we define J_k to be the projection operator onto $\mathbb{C}_k^W$. Then, it suffices to show that for all $F \in L^2(\Omega, \sigma(W), \mathbb{P})$, we can write

$$F = \sum_{k\geq0} J_k F.$$

Now let us fix an orthonormal basis $\{e_j\}_{j\geq1}$ of $\mathfrak{H}$ and set $G_j = W(e_j)$. Then $\{G_j\}_{j\geq1}$ are i.i.d. standard normal random variables and we can view F as an element in $L^2(\Omega, \sigma\{G_j : j \geq 1\}, \mathbb{P})$. For any $m \geq 1$, we define

$$\mathcal{F}_m = \sigma\{G_j : j \leq m\},\ m \geq 1,$$

which is a filtration on $(\Omega, \sigma\{W\}, \mathbb{P})$ with $\mathcal{F}_\infty = \sigma\{W\}$. For each $m \geq 1$, we define

$$F_m = \mathbb{E}[F|\mathcal{F}_m].$$

Therefore, we deduce from the martingale convergence theorem (Proposition B.6) that

$$F_m \xrightarrow[m\to+\infty]{\text{a.s. and in } L^2(\mathbb{P})} F.$$

[25] It is routine to check that the choice of the orthonormal basis is immaterial.

Since F_m is measurable with respect to $\sigma\{G_1,\dots,G_m\}$, there exists a measurable function $\varphi_m : \mathbb{R}^m \to \mathbb{R}$ such that

$$F_m = \varphi_m(G_1,\dots,G_m) \in L^2(\mathbb{P})$$

as a consequence of *Doob-Dynkin's lemma.* Next, using the chaos expansion for finitely many Gaussian random variables (Proposition 2.3), we can write

$$F_m = \sum_{k\geq 0} J_k F_m$$

with J_k the projection operator as in (2.8). Note that this argument also yields the orthogonality between different chaoses: suppose $U \in \mathbb{C}_p^W$ and $V \in \mathbb{C}_q^W$ with $p \neq q$, we define

$$U_m = \mathbb{E}[U|\mathcal{F}_m] \quad \text{and} \quad V_m = \mathbb{E}[V|\mathcal{F}_m],$$

which live in the p-th and q-th chaoses (respectively) generated by $G_1, ..., G_m$. Then, the orthogonality between U_m and V_m can be preserved in passing $m \to +\infty$. That is, $\mathbb{C}_p^W$ and $\mathbb{C}_q^W$ are mutually orthogonal for $p \neq q$. Finally, to conclude the orthogonal decomposition in (4.3), it suffices to pass $m \to \infty$ to obtain

$$F = \sum_{k\geq 0} J_k F \quad \text{in } L^2(\mathbb{P})$$

with $J_k F$ being the $L^2(\mathbb{P})$-limit of $J_k F_m$ as $m \to \infty$.

It is clear that J_k is a bounded linear operator on $L^2(\mathbb{P})$. Then,

$$\|J_k F_m - J_k F_n\|_{L^2(\mathbb{P})} \leq \|F_m - F_n\|_{L^2(\mathbb{P})},$$

that is, $\{J_k F_m\}_{m\geq 1}$ is Cauchy for each fixed k. Since $\mathbb{C}_k^W$ is a closed subspace of $L^2(\Omega, \sigma\{W\}, \mathbb{P})$, $J_k F_m$ converges in $L^2(\mathbb{P})$ to some $H_k \in \mathbb{C}_k^W$. We claim that

(i) the seires $\sum_{k\geq 0} H_k$ converge in $L^2(\mathbb{P})$,

(ii) as $m \to \infty$, F_m converges in $L^2(\mathbb{P})$ to $\sum_{k\geq 0} H_k$.

First we write

$$\sum_{k=0}^{\infty} \mathbb{E}[H_k^2] = \sum_{k=0}^{\infty} \lim_{m\to\infty} \mathbb{E}[(J_k F_m)^2] \overset{\text{Fatou}}{\leq} \liminf_{m\to\infty} \sum_{k=0}^{\infty} \mathbb{E}[(J_k F_m)^2] = \liminf_{m\to\infty} \mathbb{E}[F_m^2] = \mathbb{E}[F^2] < \infty,$$

which together with the orthogonality verifies the claim (i). To see the claim (ii), we can write

$$\Big\| F_m - \sum_{k=0}^{\infty} H_k \Big\|_{L^2(\mathbb{P})}^2 = \sum_{k=0}^{\infty} \underbrace{\|J_k F_m - H_k\|_{L^2(\mathbb{P})}^2}_{=:a_{k,m}}. \tag{4.6}$$

This is of the form $\sum_{k=0}^{\infty} a_{k,m}$, where

- $0 \leq a_{k,m} \xrightarrow{m\to\infty} 0$ for each $k \geq 0$;
- $a_{k,m} \leq b_{k,m}$ with $b_{k,m} = 2\|J_k F_m\|_2^2 + 2\|H_k\|_{L^2(\mathbb{P})}^2$;
- $\sum_{k\geq 0} b_{k,m} = 2\mathbb{E}[F_m^2] + 2\sum_{k=0}^{\infty} \mathbb{E}[H_k^2] \xrightarrow{m\to\infty} 2\mathbb{E}[F^2] + 2\sum_{k=0}^{\infty} \mathbb{E}[H_k^2] < \infty$.

By the generalized dominated convergence theorem (Proposition A.3), we see that the expression in (4.6) converges to zero as $m \to +\infty$. Hence, the proof of (4.3) is completed.

Next, we prove (4.5). Denote the RHS of (4.5) by $\mathcal{C}_k$. Then, we have

$$\bigoplus_{k\geq 0} \mathcal{C}_k \subset \bigoplus_{k\geq 0} \mathbb{C}_k^W = L^2(\Omega, \sigma\{W\}, \mathbb{P}).$$

Due to the orthogonality, it suffices to show that $\bigcup_{k\geq 0} \mathcal{C}_k$ is dense in $L^2(\Omega, \sigma\{W\}, \mathbb{P})$. That is, if $V \in L^2(\Omega, \sigma\{W\}, \mathbb{P})$ satisfies

$$\mathbb{E}\big[VH_k(W(h))\big] = 0 \tag{4.7}$$

for any unit vector $h \in \mathfrak{H}$ (i.e., $\|h\|_{\mathfrak{H}} = 1$) and for any $k \geq 0$, we must show that $V = 0$ almost surely.

Assume (4.7) holds for any unit vector $h \in \mathfrak{H}$ and for any $k \geq 0$. Since every monomial is a linear combination of Hermite polynomials, we have

$$\mathbb{E}\big[VW(h)^k\big] = 0,$$

and thus

$$\mathbb{E}\left[V\sum_{k=0}^{\infty}\frac{(it)^k W(h)^k}{k!}\right] = 0, \quad \text{i.e.,} \quad \mathbb{E}\big[Ve^{itW(h)}\big] = 0.$$

Therefore, with the orthonormal basis $\{e_j\}_{j\geq 1}$ of $\mathfrak{H}$, we can get

$$\mathbb{E}\Big[Ve^{i\sum_{j=1}^m \lambda_j W(e_j)}\Big] = 0$$

for any $\lambda_j \in \mathbb{R}$ and $m \geq 1$. Taking conditional expectation with respect to $\mathcal{F}_m$, we get

$$\mathbb{E}[V_m e^{i\sum_{j=1}^m \lambda_j W(e_j)}\Big] = 0$$

with $V_m = \mathbb{E}[V|\mathcal{F}_m] = \varphi_m(W(e_1), ..., W(e_m))$ for some $\varphi_m \in L^2(\gamma_m)$. By the same argument as in the proof of Proposition 1.11, we can conclude that $\varphi_m = 0$ almost everywhere, i.e., $V_m = 0$ almost surely. Then, another application of martingale convergence theorem yields $V = 0$ almost surely.

Hence, the proof of Theorem 4.4 is completed. □

Let $\mathcal{S}_c$ be the set of random variables of the form $f(W(h_1), ..., W(h_m))$ for some $f \in C_c^\infty(\mathbb{R}^m)$, $h_i \in \mathfrak{H}$, and $m \geq 1$. Similarly, we define $\mathcal{S}_b, \mathcal{S}_{\text{poly}}$ with C_c^∞ replaced by C_b^∞ and C_{poly}^∞ respectively. We also define $\mathcal{S}_0$ to be the set of random variables with finitely many chaos, i.e., $\mathcal{S}_0 = \bigcup_{k\geq 1}\bigoplus_{j\leq k}\mathbb{C}_j^W$.

Corollary 4.5. *It holds that $\mathcal{S}_c \subset \mathcal{S}_b \subset \mathcal{S}_{\text{poly}}$ are dense in $L^q(\Omega, \sigma\{W\}, \mathbb{P})$ for any $q \in [1, \infty)$. As a consequence, $\mathcal{S}_c(\mathfrak{H}) := \operatorname{span}\{Gh : G \in \mathcal{S}_c, h \in \mathfrak{H}\}$ is dense in $L^2(\Omega; \mathfrak{H})$.*

Proof. Given $F \in L^q(\Omega, \sigma\{W\}, \mathbb{P})$, we can proceed with the martingale convergence argument as in the proof of Theorem 4.4. For any $\varepsilon > 0$, there is some $m \geq 1$ such that

$$\|\mathbb{E}[F|\mathcal{F}_m] - F\|_q \leq \varepsilon/2.$$

Representing $\mathbb{E}[F|\mathcal{F}_m] = \varphi_m(W(e_1), ..., W(e_m))$ for some $\varphi_m \in L^q(\gamma_m)$ with γ_m the standard Gaussian measure on $\mathbb{R}^m$, one can then find some $\varphi_{m,c} \in C_c^\infty(\mathbb{R}^m)$ such that[26]

$$\|\varphi_m - \varphi_{m,c}\|_{L^q(\gamma_m)} \leq \varepsilon/2. \tag{4.8}$$

It follows that $\|\varphi_{m,c}(W(e_1), ..., W(e_m)) - F\|_q \leq \varepsilon$ with $\varphi_{m,c}(W(e_1), ..., W(e_m)) \in \mathcal{S}_c$. This proves the density of $\mathcal{S}_c$ in $L^q(\Omega, \sigma\{W\}, \mathbb{P})$ for any $q \in [1, \infty)$.

Next, we prove the density of $\mathcal{S}_c(\mathfrak{H})$ in $L^2(\Omega; \mathfrak{H})$. Let $V \in L^2(\Omega; \mathfrak{H})$ satisfy $\mathbb{E}[G\langle h, V\rangle_{\mathfrak{H}}] = 0$ for all $G \in \mathcal{S}_c$ and for all $h \in \mathfrak{H}$. Then, by density of $\mathcal{S}_c$ in $L^2(\Omega, \sigma\{W\}, \mathbb{P})$, we get $\langle h, V\rangle_{\mathfrak{H}} = 0$ almost surely for all $h \in \mathfrak{H}$. Then, $V = \sum_{j=1}^\infty \langle V, e_j\rangle_{\mathfrak{H}} e_j = 0$ almost surely, where $\{e_j\}_{j\geq 1}$ is an orthonormal basis of $\mathfrak{H}$. Hence, the proof is completed. □

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 10 (Oct.2, 2025)** $\infty\infty\infty\infty\infty\infty$

[26]The bound (4.8) is due to the density of $C_c^\infty(\mathbb{R}^m)$ in $L^q(\gamma_m)$. Can you write down a quick student-proof?

Lecture 11.

4.2. **Malliavin operators I:** D **&** δ**.** In this section, we will introduce several important operators in Malliavin calculus, which we have already studied in Section 2 with finitely many Gaussian random variables.

• **Malliavin derivative.** Recall that for a random variable $F = f(G_1, ..., G_d)$ for some $f \in C^\infty_{\text{poly}}(\mathbb{R}^d)$ and $(G_1, ..., G_d) \in \mathcal{N}_d(0, \mathbf{I})$, we have

$$\nabla f(G_1, ..., G_d) = \sum_{j=1}^d \partial_j f(G_1, ..., G_d)\mathfrak{e}_j \tag{4.9}$$

with $\{\mathfrak{e}_j : j = 1, ..., d\}$ denoting the canonical basis of $\mathbb{R}^d$.[27] In particular, we can realize the standard Gaussian vector in the isonormal framework as in Example 3. Putting $G_j = W(e_j)$ with $\{e_j\}_{j\geq 1}$ orthonormal basis of $\mathfrak{H}$, we can rewrite (4.9) as

$$DF = Df(W(e_1), ..., W(e_d)) = \sum_{j=1}^d \partial_j f(W(e_1), ..., W(e_d))e_j \in \mathfrak{H}.$$

This gives us the Malliavin derivative of $F = f(W(e_1), ..., W(e_d))$, which is a $\mathfrak{H}$-valued random variable (indeed, $DF \in L^2(\Omega; \mathfrak{H})$).

Now let us consider the case where $F' = g(W(h_1), ..., W(h_m))$ for some $g \in C^\infty_{\text{poly}}(\mathbb{R}^m)$ and $h_1, ..., h_m \in \mathfrak{H}$. Then, $\mathbf{X} = (W(h_1), ..., W(h_m))$ is a m-dimensional centered Gaussian vector with covariance matrix $C = QQ$ for some symmetric positive-definite matrix $Q \in \mathbb{R}^{m\times m}$. Assume $\mathbf{X} = Q\mathbf{G}$ with $\mathbf{G} = (W(e_1), ..., W(e_m))$ a standard Gaussian vector. Then, we have

$$\begin{aligned} DF' = Dg(Q\mathbf{G}) &= \sum_{j=1}^m \bigg(\sum_{i=1}^m \partial_i g(Q\mathbf{G})Q_{ij}\bigg)e_j \\ &= \sum_{i=1}^m \partial_i g\big(W(h_1), ..., W(h_m)\big)\sum_{j=1}^m Q_{ij}e_j \\ &= \sum_{i=1}^m \partial_i g\big(W(h_1), ..., W(h_m)\big)h_i, \end{aligned}$$

where the last equality follows from $W(h_j) = (Q\mathbf{G})_j = \sum_{j=1}^m Q_{ij}W(e_j) = W\big(\sum_{j=1}^m Q_{ij}e_j\big)$. This discussion suggests the following definition.

Definition 4.6. (i) *Suppose $F \in \mathcal{S}_{\text{poly}}$ can be expressed as $F = f(W(h_1), ..., W(h_m))$ with $f \in C^\infty_{\text{poly}}(\mathbb{R}^m)$, $h_1, ..., h_m \in \mathfrak{H}$, and $m \in \mathbb{N}_{\geq 1}$. The Malliavin derivative DF of F is defined by*

$$DF = \sum_{i=1}^m \partial_i f\big(W(h_1), ..., W(h_m)\big)h_i. \tag{4.10}$$

In particular, $DW(h) = h$.

(ii) *Let $\mathbb{D}^{1,2}$ be the closure of $\mathcal{S}_{\text{poly}}$ with respect to the following inner product*

$$\langle F, G\rangle_{\mathbb{D}^{1,2}} := \mathbb{E}[FG] + \langle DF, DG\rangle_{L^2(\Omega;\mathfrak{H})}.$$

In particular, $\mathbb{D}^{1,2}$ is a Hilbert space.

---

[27] $\mathfrak{e}_j \in \mathbb{R}^d$ is a unit vector with only one nonzero entry (i.e., +1) in the j-th coordinate.

The following question indicates that the definition (4.10) of Malliavin derivative DF is independent of the choices of representations.

Question 4.7. (i) *Given two representations of* $F = f(W(h_1),...,W(h_m)) = g(W(f_1),...,W(f_n))$*, prove that the following expressions are identical:*

$$\sum_{i=1}^{m} \partial_i f\big(W(h_1),...,W(h_m)\big)h_i = \sum_{j=1}^{n} \partial_j g\big(W(f_1),...,W(f_n)\big)f_j.$$

(ii) [Directional derivative] *Given* $F = f(W(h_1),...,W(h_m))$ *and* $h \in \mathfrak{H}$*. Verify that*

$$\langle DF, h\rangle_{\mathfrak{H}} = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big[-F + f\big(W(h_1)+\varepsilon\langle h_1,h\rangle_{\mathfrak{H}},...,W(h_m)+\varepsilon\langle h_m,h\rangle_{\mathfrak{H}}\big)\Big].$$

The quantity $\langle DF, h\rangle_{\mathfrak{H}}$ *is called the directional derivative of* F *in the direction* h*, which measures the sensitivity of the random variable* F *to the "small" amount of perturbations from* W *to* $W + \varepsilon h$*.*

Let us now recall a few basic properties of the Malliavin derivative operator.

Proposition 4.8. (i) *The derivative operator* D *is closable: if* $F_n \in \mathbb{D}^{1,2}$ *with* $DF_n \to V$ *in* $L^2(\Omega;\mathfrak{H})$ *for some* V *and* $F_n \to 0$ *in* $L^2(\mathbb{P})$ *as* $n \to +\infty$*, then* $V = 0$*.*
(ii) *Let* $\mathbb{C}_k^W$ *denote the* k*-Wiener chaos as in Theorem 4.4. Then,* $\mathbb{C}_k^W \subset \mathbb{D}^{1,2}$ *and*

$$\|DJ_kF\|^2_{L^2(\Omega;\mathfrak{H})} = k\|J_kF\|_2^2 \tag{4.11}$$

for any $F \in L^2(\Omega,\sigma\{W\},\mathbb{P})$*.*
(iii) *Given* $G \in L^2(\Omega,\sigma\{W\},\mathbb{P})$*, we have* $G \in \mathbb{D}^{1,2}$ *if and only if* $\sum_{k\geq 1} k\|J_kG\|_2^2 < \infty$*. In each case, we have* $\mathbb{E}[\|DG\|^2_{\mathfrak{H}}] = \sum_{k\geq 1} k\|J_kG\|_2^2$*. Here,* J_k *denotes the projection operator onto* $\mathbb{C}_k^W$*.*
(iv) [Gaussian Poincaré inequality] *For any* $F \in \mathbb{D}^{1,2}$*, we have*

$$\mathrm{Var}(F) \leq \mathbb{E}[\|DF\|^2_{\mathfrak{H}}],$$

where the equality holds only when $F \in \mathbb{C}_0^W \oplus \mathbb{C}_1^W$*.*

Proof. Given any $G \in \mathcal{S}_c, h \in \mathfrak{H}$, we can write

$$\begin{aligned}
\mathbb{E}[\langle V,h\rangle_{\mathfrak{H}}G] &= \lim_{n\to+\infty}\mathbb{E}[\langle DF_n,h\rangle_{\mathfrak{H}}G]\\
&= \lim_{n\to+\infty}\mathbb{E}[\langle D(F_nG) - F_nDG, h\rangle_{\mathfrak{H}}]\\
&= \lim_{n\to+\infty}\mathbb{E}[\langle F_nGW(h)] + \lim_{n\to+\infty}\mathbb{E}[F_n\langle DG,h\rangle_{\mathfrak{H}}]\\
&= 0,
\end{aligned} \tag{4.12}$$

where the second equality follows from the usual product rule $D(FG) = FDG + GDF$ for smooth random variables $F, G \in \mathcal{S}_{\mathrm{poly}}$,[28], the third equality follows from the Gaussian integration by parts formula in Proposition 2.11, and the last equality is due to the L^2-convergence of $F_n \to 0$ with $GW(h), DG, h\rangle_{\mathfrak{H}} \in L^2(\Omega)$. Recall from Corollary 4.5 that $\mathcal{S}_c(\mathfrak{H}) = \{Gh : G \in \mathcal{S}_c, h \in \mathfrak{H}\}$ is dense in $L^2(\Omega;\mathfrak{H})$. This implies that $V = 0$. That is, part (i) is proved.

Now let us prove (ii). Suppose $F \in \mathbb{C}_k^W$ has the following form

$$F = \sum_{\mathbf{q}\in\mathbb{N}^\infty} \mathbf{1}_{\{|\mathbf{q}|=\sum_{j\geq 1}q_j = k\}}c_{\mathbf{q}}\prod_{j=1}^{\infty}H_{q_j}(W(e_j))$$

[28]In view of Question (4.7), we can express F and G as two functions of a common Gaussian vector, so that the desired product rule is simply the usual product rule in calculus.

with $\{e_j\}_{j\geq 1}$ orthonormal basis of $\mathfrak{H}$. As in the proof of Theorem 4.4, we define

$$F_m = \sum_{q_1,\dots,q_m\geq 0} \mathbf{1}_{\{\sum_{j=1}^m q_j=k\}} c_{q_1,\dots,q_m} \prod_{j=1}^m H_{q_j}(W(e_j)),$$

which converges to F in $L^2(\Omega)$. Then, F_m is indeed a polynomial in independent Gaussians and belongs to $\mathcal{S}_{\text{poly}}$ with

$$DF_m = \sum_{q_1,\dots,q_m\geq 0} \mathbf{1}_{\{\sum_{j=1}^m q_j=k\}} c_{q_1,\dots,q_m} \sum_{i=1}^m q_i H_{q_i-1}(W(e_i))\bigg(\prod_{j\neq i} H_{q_j}(W(e_j))\bigg)e_i.$$

It follows from the results in Section 2 that

$$\mathbb{E}\big[\|DF_m\|_{\mathfrak{H}}^2\big] = k\mathbb{E}[F_m^2] \quad \text{and} \quad \mathbb{E}\big[\|D(F_m-F_n)\|_{\mathfrak{H}}^2\big] = k\mathbb{E}[(F_m-F_n)^2], \tag{4.13}$$

where we also used $\mathbb{E}[F_mF_n] = \mathbb{E}[F_m^2]$ and $\mathbb{E}[\langle DF_m, DF_n\rangle_{\mathfrak{H}}] = \mathbb{E}[\|DF_m\|_{\mathfrak{H}}^2]$ for $m\leq n$ to obtain the second equality. In particular, $\{DF_m\}_m$ is Cauchy in $L^2(\Omega;\mathfrak{H})$ with limit

$$DF = \sum_{\mathbf{q}\in\mathbb{N}^\infty} \mathbf{1}_{\{|\mathbf{q}|=k\}} c_{\mathbf{q}} \sum_{i=1}^\infty q_i H_{q_i-1}(W(e_i))\bigg(\prod_{j\neq i} H_{q_j}(W(e_j))\bigg)e_i.$$

It is clear that $DF\in L^2(\Omega;\mathfrak{H})$ and passing $m\to+\infty$ in (4.13) yields (4.11).

Next, we prove (iii). Given $G\in L^2(\Omega,\sigma\{W\},\mathbb{P})$, we define

$$G_N = \sum_{k\leq N} J_kG.$$

Then, by orthogonality relation and part (ii), we get

$$\|DG_N\|_{L^2(\Omega;\mathfrak{H})}^2 = \sum_{k\leq N} k\|J_kG\|_2^2.$$

If $G\in\mathbb{D}^{1,2}$, then due to (i) with $G_N \xrightarrow[L^2(\Omega)]{N\to+\infty} G$, we have $DG_N\to DG$ in $L^2(\Omega;\mathfrak{H})$. Therefore,

$$\sum_{k\geq 1} k\|J_kG\|_2^2 = \lim_{N\to+\infty} \|DG_N\|_{L^2(\Omega;\mathfrak{H})}^2 = \|DG\|_{L^2(\Omega;\mathfrak{H})}^2 < \infty.$$

Suppose $\sum_{k\geq 1} k\|J_kG\|_2^2 < \infty$. Then, $\{DG_N\}_N$ is Cauchy in $L^2(\Omega;\mathfrak{H})$, which together with $G_N \xrightarrow[L^2(\Omega)]{N\to+\infty} G$, implies that $G\in\mathbb{D}^{1,2}$ with DG the limit of DG_N in $L^2(\Omega;\mathfrak{H})$.

Finally, it is easy to see that part (iv) is an immediate consequence of part (iii). □

Remark 4.9. For $F\in\mathcal{S}_{\text{poly}}$, the definition of DF is independent of its representation; see Question 4.7-(i). The space $\mathbb{D}^{1,2}$ is defined by closing $\mathcal{S}_{\text{poly}}$ under the $\mathbb{D}^{1,2}$-norm, and the closability of D ensures that the definition of DF does not depend on the particular choice of approximating sequence $F_n\in\mathcal{S}_{\text{poly}}$. Part (iii) of Proposition 4.8 gives an equivalent definition of the space $\mathbb{D}^{1,2}$ using the projection operators.

In (4.12), we used the product rule for smooth random variables. It is not difficult to generalize it for more general random variables; see also [NP12, Exercise 2.3.10].

Proposition 4.10 (Product rule). *Suppose $F\in\mathbb{D}^{1,2}$ and $G\in\mathcal{S}_b$. Then, $FG\in\mathbb{D}^{1,2}$ and*

$$D(FG) = FDG + GDF. \tag{4.14}$$

Proof. Due to $F \in \mathbb{D}^{1,2}$ and $G \in \mathcal{S}_b$, we have the RHS of (4.14) belongs to $L^2(\Omega; \mathfrak{H})$ and can be represented as a limit of

$$F_n DG + GDF_n$$

with $F_n \in \mathcal{S}_{\text{poly}}$. For each n, by the usual product rule in calculus, we have $F_n DG + GDF_n = D(F_n G)$. Since $F_n G \in \mathcal{S}_{\text{poly}}$ converges to FG in $L^2(\Omega)$, we can conclude that its limit FG belongs to $\mathbb{D}^{1,2}$ and the equality (4.14) holds. □

• **Higher-order Malliavin derivatives.** For $k \in \mathbb{N}_{\geq 2}$, we can define the k-th Malliavin derivative D^k in a similar way.

(i) For $F \in \mathcal{S}_{\text{poly}}$ given by $F = f(W(h_1), ..., W(h_m))$, we define

$$D^k F = \sum_{i_1,\dots,i_k=1}^{m} \partial_{i_1\dots i_k} f(W(h_1), ..., W(h_m)) h_{i_1} \otimes \cdots \otimes h_{i_k},$$

which is clearly a random element in $\mathfrak{H}^{\odot k}$; here $\mathfrak{H}^{\odot k}$ denotes the k-th symmetric tensor product of $\mathfrak{H}$ (see Appendix (A) for more details).

(ii) Suppose D^{k-1} and its domain $\mathbb{D}^{k-1,2}$ have been defined. We define $\mathbb{D}^{k,2}$ to be the closure of $\mathcal{S}_{\text{poly}}$ under the following inner product

$$\begin{aligned}\langle F, G\rangle_{\mathbb{D}^{k,2}} &= \langle F, G\rangle_{\mathbb{D}^{k-1,2}} + \langle D^k F, D^k G\rangle_{L^2(\Omega;\mathfrak{H}^{\otimes k})} \\ &= \sum_{i=0}^{k} \mathbb{E}\big[\langle D^i F, D^i G\rangle_{\mathfrak{H}^{\otimes i}}\big] \quad \text{with } D^0 F = F.\end{aligned}$$

(iii) As in Proposition 4.8, one can show that D^k is closable and the following equivalence holds:

$$F \in \mathbb{D}^{k,2} \text{ if and only if } \sum_{p=1}^{\infty} p^k \|J_p F\|_2^2 < \infty. \tag{4.15}$$

We leave this as an exercise to interested students.

• **Divergence operator.** The divergence operator δ is the adjoint operator of D, which is characterized by the **duality relation**

$$\mathbb{E}[\langle DF, V\rangle_{\mathfrak{H}}] = \mathbb{E}[F\delta(V)]. \tag{4.16}$$

Let $\text{dom}(\delta)$ be the set of random vectors $V \in L^2(\Omega; \mathfrak{H})$ such that there is some finite constant $C = C_V$ such that $|\mathbb{E}[\langle DF, V\rangle_{\mathfrak{H}}]| \leq C\|F\|_2$ for any $F \in \mathcal{S}_{\text{poly}}$. In particular, the functional $F \mapsto \Psi(F) := \mathbb{E}[\langle DF, V\rangle_{\mathfrak{H}}]$ is bounded linear on $\mathcal{S}_{\text{poly}}$. Then, due to density of $\mathcal{S}_{\text{poly}}$, one can extend Ψ to a bounded linear function on the Hilbert space $L^2(\Omega, \sigma\{W\}, \mathbb{P})$, and thus Riesz's representation theorem indicates that there exists a unique element (denoted by $\delta(V)$) in $L^2(\Omega, \sigma\{W\}, \mathbb{P})$ such that (4.16) holds true. One can proceed in the same way to define the iterated divergence operator δ^k as the adjoint operator of D^k. We summarize the above discussion in the following.

Lemma 4.11. *For any $F \in \mathbb{D}^{k,2}$ and $V \in \text{dom}(\delta^k)$, we have*

$$\mathbb{E}[\langle D^k F, V\rangle_{\mathfrak{H}^{\otimes k}}] = \mathbb{E}[F\delta^k(V)].$$

Question 4.12. *Given a unit vector $h \in \mathfrak{H}$. Show that $\delta^k(h^{\otimes k}) = H_k(W(h))$.*

We conclude this section with a few properties of the divergence operator and its domain $\text{dom}(\delta)$.

Proposition 4.13. (i) *Recall* $\mathcal{S}_c(\mathfrak{H})$ *from Corollary 4.5 and define similarly* $\mathcal{S}_0(\mathfrak{H}) = \operatorname{span}\{Gh : h \in \mathfrak{H}, G \in \mathcal{S}_0\}$. *Then, both* $\mathcal{S}_0(\mathfrak{H})$ *and* $\mathcal{S}_c(\mathfrak{H})$ *are contained in* $\operatorname{dom}(\delta)$. *In particular,* $\operatorname{dom}(\delta)$ *is dense in* $L^2(\Omega;\mathfrak{H})$.
(ii) *Assume* $G \in \mathbb{D}^{1,2}$ *and* $V \in \operatorname{dom}(\delta)$ *satisfy*

$$\mathbb{E}\big[G^2\|V\|_{\mathfrak{H}}^2\big] + \mathbb{E}\big[G^2\delta(V)^2\big] + \mathbb{E}\big[\langle DG, V\rangle_{\mathfrak{H}}^2\big] < \infty. \tag{4.17}$$

Then, $GV \in \operatorname{dom}(\delta)$ *with*

$$\delta(GV) = G\delta(V) - \langle DG, V\rangle_{\mathfrak{H}}.$$

Proof. Suppose $F, G \in \mathcal{S}_{\text{poly}}$ and $h \in \mathfrak{H}$. Then, we can proceed as in (4.12)

$$\begin{aligned}\mathbb{E}\big[\langle DF, h\rangle_{\mathfrak{H}} G\big] &= \mathbb{E}\big[\langle D(FG), h\rangle_{\mathfrak{H}}\big] - \mathbb{E}\big[F\langle DG, h\rangle_{\mathfrak{H}}\big]\\ &= \mathbb{E}[FGW(h)] - \mathbb{E}\big[F\langle DG, h\rangle_{\mathfrak{H}}\big]\end{aligned} \tag{4.18}$$

using the product rule. Therefore, we deduce from Cauchy-Schwarz that

$$\big|\mathbb{E}\big[\langle DF, h\rangle_{\mathfrak{H}} G\big]\big| \le \|F\|_2\big(\|GW(h)\|_2 + \|DG\|_{L^2(\Omega;\mathfrak{H})}\|h\|_{\mathfrak{H}}\big). \tag{4.19}$$

Due to $\mathcal{S}_c \cup \mathcal{S}_0 \subset \mathbb{D}^{1,2}$, we can approximate any element in $\mathcal{S}_c \cup \mathcal{S}_0$ by a sequence of random variables in $\mathcal{S}_{\text{poly}}$ with respect to the norm $\| \bullet \|_{\mathbb{D}^{1,2}}$. That is, via a limiting argument (if necessary), we can obtain the bound (4.19) for any $G \in \mathcal{S}_c \cup \mathcal{S}_0$. This implies that $\mathcal{S}_0(\mathfrak{H}) \cup \mathcal{S}_c(\mathfrak{H}) \subset \operatorname{dom}(\delta)$ with

$$\delta(Gh) = GW(h) - \langle DG, h\rangle_{\mathfrak{H}} \tag{4.20}$$

for any $G \in \mathcal{S}_{\text{poly}} \cup \mathcal{S}_0$ in view of (4.18). Thus, the proof of part (i) is concluded by invoking Corollary 4.5.

Next, we prove part (ii), which generalizes (4.20). Let $F \in \mathcal{S}_b$, then we can deduce from the product rule (Proposition 4.10) and duality relation that

$$\begin{aligned}\mathbb{E}\big[\langle DF, GV\rangle_{\mathfrak{H}}\big] &= \mathbb{E}\big[\langle D(FG) - FDG, V\rangle_{\mathfrak{H}}\big]\\ &= \mathbb{E}\big[\langle D(FG), V\rangle_{\mathfrak{H}}\big] - \mathbb{E}\big[F\langle DG, V\rangle_{\mathfrak{H}}\big]\\ &= \mathbb{E}\big[FG\delta(V)\big] - \mathbb{E}\big[F\langle DG, V\rangle_{\mathfrak{H}}\big] \qquad \text{due to (4.17)}\end{aligned}$$

and thus, we get $GV\operatorname{dom}(\delta)$ and $\delta(GV) = G\delta(V) - \langle DG, V\rangle_{\mathfrak{H}}$. Note that we used the density of $\mathcal{S}_b$ in $L^2(\Omega, \sigma\{W\}, \mathbb{P})$ and $FG \in \mathbb{D}^{1,2}$ (a consequence of Proposition 4.10).

Hence, the proof of Proposition 4.13 is completed. $\square$

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 11 (Oct.7, 2025)** $\infty\infty\infty\infty\infty\infty\infty$

Lecture 12.

4.3. **(Generalized) chain rule.** XXX

• **Application: (local) Bouleau-Hirsch's criterion.**

Example 4 (Nonzero V with $\delta(V) = 0$)**.** Let $\{e_1, e_2\}$ be an orthonormal system in $\mathfrak{H}$ and define

$$f = e_1 \otimes e_2 - e_2 \otimes e_1 \in \mathfrak{H}^{\otimes 2}.$$

Then, $V := I_1(f) = W(e_1)\, e_2 \; - \; W(e_2)\, e_1$ is nonzero almost surely and

$$\delta(V) = I_2\big(\widetilde{f}\big) = 0.$$

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 12 (Oct.9, 2025)** $\infty\infty\infty\infty\infty\infty\infty$

Lecture 13.

4.4. **Malliavin operators II: divergence operator and Itô v.s. Skorohod.**

$\propto\propto\propto\propto\propto\propto$ **End of Lecture 13 (Oct.16, 2025)** $\infty\infty\infty\infty\infty\infty\infty$

4.5. **Ito v.s. Skorohod.**

4.6. **Application: linear/NL SPDEs.** optimal well-posedness for Anderson models with fractional rough noises.

5. **Fourth moment theorems and Nourdin-Peccati analysis**

Let us start with an example from Peccati and Yor's work [**?**] : let W be a standard real Brownian motion, then according to Jeulin's lemma[29],

$$\int_0^1 t^{-2} W_t^2\, dt = +\infty \quad \text{almost surely.}$$

The authors of [**?**] studied the fluctuation of $F_\varepsilon = \int_\varepsilon^1 W_t^2 t^{-2}\, dt$, after normalization, as ε goes to zero. And they obtained that

$$\widetilde{F_\varepsilon} := \frac{F_\varepsilon - \mathbb{E}[F_\varepsilon]}{\sqrt{\mathrm{Var} F_\varepsilon}} \xrightarrow[\varepsilon\to 0]{law} \mathcal{N}(0,1)\,. \tag{5.1}$$

The proof is roughly sketched as follows:

- It is easy to compute that $\sigma_\varepsilon := \sqrt{\mathrm{Var} F_\varepsilon} = \sqrt{4\log(1/\varepsilon) + 4\varepsilon - 4}$.
- Using Itô's formula and stochastic Fubini's theorem (see *e.g.* [**?**]), one can rewrite $\widetilde{F_\varepsilon}$ as follows:
$$\begin{aligned}\widetilde{F_\varepsilon} &= \frac{2}{\sigma_\varepsilon}\int_\varepsilon^1 \left(\int_0^1 t^{-2}\mathbf{1}_{(s\le t)} W_s dW_s\right) dt = \frac{2}{\sigma_\varepsilon}\int_0^1 \left(\int_\varepsilon^1 t^{-2}\mathbf{1}_{(s\le t)}\, dt\right) W_s dW_s \\ &= \frac{2}{\sigma_\varepsilon}\int_0^1 \left(\frac{1}{\varepsilon\vee s} - 1\right) W_s\, dW_s = \int_0^1 \varphi_s^\varepsilon\, dW_s\ ,\end{aligned}$$
with $\varphi_s^\varepsilon := 2\sigma_\varepsilon^{-1}\big[(\varepsilon\vee s)^{-1} - 1\big] W_s$, $s\in[0,1]$. It is also clear that $M_t^\varepsilon := \int_0^t \varphi_s^\varepsilon\, dW_s$, $t\in\mathbb{R}_+$, is a square-integrable continuous martingale such that its quadratic variation (at infinity) $\langle M^\varepsilon\rangle_\infty = +\infty$ almost surely[30].
- According to the theorem of Dambis and Dubins-Schwarz (see *e.g.* [**?**, Chapter V]), there exists a standard real Brownian motion β^ε such that $M_t^\varepsilon = \beta^\varepsilon_{\langle M^\varepsilon\rangle_t}$, $t\in\mathbb{R}_+$. As β_1^ε is a standard Gaussian random variable, in order to show the asymptotic normality of $\widetilde{F_\varepsilon} = M_1^\varepsilon$, it is enough to prove $\langle M^\varepsilon\rangle_1$ is close to 1 in $L^2(\P)$, in view of the Burkholder-Davis-Gundy inequality (see *e.g.* [**?**, Chapter IV]).
- It is easy to obtain $\|\langle M^\varepsilon\rangle_1 - 1\|_{L^2(\P)} = O(1/\sigma_\varepsilon)$, as $\varepsilon\downarrow 0$. This finishes our sketchy proof.

For more details, one can refer to the original paper [**?**] as well as the recent survey [**?**].

Later, Nualart and Peccati studied another Gaussian functional

$$F_{H,\varepsilon} := \int_\varepsilon^1 \frac{(W_t^H)^2}{t^{1+2H}}\, dt\,, \quad (0<\varepsilon<1) \tag{5.2}$$

where W^H is a fractional Brownian motion with Hurst parameter $H\in(0,1)$. The F_ε above corresponds to the case where $H = 1/2$, that is, when W^H is a standard Brownian motion. According to Jeulin's lemma, $F_{H,0} = +\infty$ a.s., so it is natural to see whether the Gaussian fluctuation of the renormalized $F_{H,\varepsilon}$ occurs for general H. Starting from this motivating example,

---

[29] Here is a simple version from the exercise book [**?**] by Chaumont and Yor: let $\{R_t, t\in(0,1]\}$ be a jointly measurable process with values in $\mathbb{R}_+$ such that $0<\mathbb{E}[R_1]<+\infty$ and $R_t \overset{\text{(law)}}{=} R_1$ for any $t\in(0,1]$, and let μ be a σ-finite measure on $(0,1]$. Then the following equivalence holds true: $\int_{(0,1]} R_t\,\mu(dt) < +\infty$ *almost surely* if and only if $\int_{(0,1]}\mu(dt)<+\infty$.

[30] One can first compute directly that $\sigma_\varepsilon^2\langle M^\varepsilon\rangle_\infty = 4\int_0^\infty \big[(\varepsilon\vee s)^{-1} - 1\big]^2 W_s^2\, ds \ge \int_2^\infty W_s^2 s^{-2}\, ds$, while applying Jeulin's lemma to the inverse-time Brownian motion, we get $\int_1^\infty W_s^2 s^{-2}\, ds = +\infty$ a.s., so we can conclude that $\langle M^\varepsilon\rangle_\infty = +\infty$ almost surely.

Nualart and Peccati published a very surprising result [?] in 2005, which is known as the "*fourth moment theorem*" nowadays.

Theorem 5.1 (D. Nualart and G. Peccati, 2005)**.** *Fix an integer $p \geq 2$. Let*

$$F_n \equiv I_p^W(f_n) := p! \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{p-1}} f_n(t_1, \dots, t_p)\, dW_{t_p} \cdots dW_{t_1} \tag{5.3}$$

be a sequence of pth-order multiple Wiener-Itô integrals. Assume that the kernel $f_n \in L^2\big([0,1]^p, dx\big)$ is symmetric almost everywhere such that $p!\|f_n\|^2_{L^2([0,1]^p)} \to 1$, as $n \to +\infty$. Then the following statements are equivalent:

(i) *F_n converges in law to a standard Gaussian, as $n \to +\infty$.*
(ii) *$\mathbb{E}\big[F_n^4\big]$ converges to 3, as $n \to +\infty$.*
(iii) *For each $r \in \{1, \dots, p-1\}$, $\|f_n \otimes_r f_n\|_{L^2([0,1]^{2p-2r})}$ converges to 0, as $n \to +\infty$, where the r-*contraction *$\otimes_r$ is defined as follows:*

$$\begin{aligned}&\big(f \otimes_r g\big)(x_1, \dots, x_{p+q-2r}) \\ &= \int_{[0,1]^r} f(x_1, \dots, x_{p-r}, s_1, \dots, s_r) g(x_{p-r+1}, \dots, x_{p+q-2r}, s_1, \dots, s_r) ds_1 \dots ds_r\,,\end{aligned} \tag{5.4}$$

for $f \in L^2([0,1]^p, dx)$, $g \in L^2([0,1]^q, dx)$ and[31] *$0 \leq r \leq p \wedge q$, see also (??).*

The name "fourth moment theorem" comes from the equivalence between (i) and (ii). The implication "(ii) $\Rightarrow$ (i)" is particularly surprising, as it simplifies to a great extent the usual way of proving the central limit theorem through the so-called method of moments and cumulants[32]. *The original proof of FMT follows essentially the same arguments as for* (5.1)*, that is, one sees F_n as the value of some continuous martingale $M^{(n)}$ at $t = 1$ and applies the random-time change:*

$$M_t^{(n)} := p! \int_0^t \left(\int_0^{t_1} \cdots \int_0^{t_{p-1}} f_n(t_1, \dots, t_p)\, dW_{t_p} \cdots dW_{t_2} \right) dW_{t_1} = \beta^{(n)}_{\langle M^{(n)} \rangle_t}\,,$$

where $\beta^{(n)}$ is some standard Brownian motion and $\langle M^{(n)} \rangle_\bullet$ denotes the quadratic variation process associated to $M^{(n)}$. At a handwaving level, if $\langle M^{(n)} \rangle_1$ is close to 1 in $L^2(\P)$, then by Burkholder-Davis-Gundy inequality, $F_n = M_1^{(n)}$ is close to $\beta_1^{(n)}$, which is standard Gaussian. Computing $\mathbb{E}\big[\langle M^{(n)} \rangle_1^2\big]$ involves an extensive use of the product formula (??) *for multiple Wiener-Itô integrals and this is where the contractions in (iii) come into the play, while after a lengthy computation of the fourth moment $\mathbb{E}\big[F_n^4\big]$ using the product formula* (??)*, one gets the equivalence between (ii) and (iii). Making these arguments rigorous is essentially what was done in the original article.*

By the isometry between any two real separable Hilbert spaces, the above FMT holds inside a general Gaussian Wiener chaos associated to an isonormal process, see [?]. Back to the Gaussian functional (5.2), one can treat $F_{H,\varepsilon} - \mathbb{E}[F_{H,\varepsilon}]$ as an element in the second Gaussian Wiener chaos; and according to the FMT, checking the Gaussian fluctuation reduces to computing the related contraction-norms, see [?, Section 3.1]. This application is just *a tip of the iceberg.* This univariate FMT together with its multivariate extension has produced a very efficient strategy for proving Gaussian limits on the Wiener space.

Let us first recall the multivariate FMT [?], established by Peccati and Tudor.

[31] Note $f \otimes_0 g = f \otimes g$ is the usual tensor product of two functions.
[32] See *e.g.* Section 30 in Billingsley's book [?].

Theorem 5.2 (G. Peccati and C. A. Tudor, 2005)**.** *Fix integers $d \geq 2$ and $1 \leq q_1 \leq \ldots \leq q_d$. Let $C = \big(C_{i,j}, 1 \leq i,j \leq d\big)$ be a $d \times d$ symmetric nonnegative definite matrix and $f_{n,i} \in L^2([0,1]^{q_i}, dx)$ be symmetric for any $n \geq 1$ and every $i \in [d] := \{1, \ldots, d\}$. Assume that the d-dimensional random vectors*

$$F^{(n)} = \big(F_1^{(n)}, \ldots, F_d^{(n)}\big)^T := \big(I^W_{q_1}(f_{n,1}), \ldots, I^W_{q_d}(f_{n,d})\big)^T$$

satisfy

$$\lim_{n\to+\infty} \mathbb{E}\big[F_i^{(n)} F_j^{(n)}\big] = C_{i,j} \text{ for every } i,j \in [d].$$

Then, as $n \to +\infty$, the following assertions are **equivalent***:*

(1) *The vector $F^{(n)}$ converges in law to a d-dimensional Gaussian vector $\mathcal{N}(0, C)$;*
(2) *for every $i \in [d]$, $F_i^{(n)}$ converges in law to a Gaussian random variable $\mathcal{N}\big(0, C_{i,i}\big)$;*
(3) *for every $i \in [d]$, $\mathbb{E}\big[(F_i^{(n)})^4\big] \to 3C_{i,i}^2$;*
(4) *for every $i \in [d]$ and each $1 \leq r \leq q_i - 1$, $\|f_{n,i} \otimes_r f_{n,i}\|_{L^2([0,1]^{2q_i-2r})} \to 0$.*

In fact, soon after the appearance of [**?**], Peccati and Tudor attempted to obtain the FMT for a sequence of random variables belonging to finitely many chaoses. Surprisingly, they obtained the above more interesting result: for random vectors with components in the fixed Gaussian Wiener chaos, the joint convergence to Gaussian (1) is ***equivalent*** to the component-wise convergence (2), and the latter is reduced to verifying (3) or (4) in view of Nualart and Peccati's univariate FMT. This observation provides practitioners with an alternative to the semimartingale approach (see *e.g.* [**?**]) to prove the central limit theorems on the Wiener space; roughly speaking, one can first decompose the random variable into Gaussian Wiener chaoses and then prove central convergence on each chaos. See, for example, [**?**, **?**, **?**] for applications to zeros of random polynomials; [**?**, **?**, **?**, **?**, **?**] for applications to statistical physics. Again, we invite interested readers to visit the website [**?**] for many more works around the FMTs.

Besides Peccati-Tudor's theorem, there have been several important inputs around the FMTs, notably

(i) Nualart and Ortiz-Latorre [**?**] gave another proof of FMT based on Malliavin calculus, which paved the road for *Stein to meet Malliavin* [**?**], see Section 6;

(ii) Ledoux took another insightful point-of-view for the FMT in [**?**] and he treated the Gaussian Wiener chaos as an eigenspace of a symmetric self-adjoint Markov diffusion operator, and then he employed Gamma calculus to derive the FMT in this setting. Such a direction has been further enriched in the papers [**?**, **?**, **?**]. Originally Ledoux's spectral viewpoint was taken in the diffusive setting and unexpectedly, after slight adjustment to the discrete case, it turns out to be a crucial ingredient for the obtention of the FMTs on the Poisson space and in the Rademacher setting, see Section **??**.

6. How Stein meets Malliavin?

This section tells the story of how Malliavin calculus was first combined with Stein's method. As mentioned before, we will first sketch Nualart and Ortiz-Latorre's methodological breakthrough, namely, linking the Malliavin operators to the study of limit theorems on the Wiener space. To illustrate better the ideas and also for the sake of later reference, let us now introduce some basics of Malliavin calculus and define the central object in this thesis, *Wiener chaos*. Readers, who have already been familiar with Malliavin calculus, could jump to Section 6.1.

The concept of chaos that we consider dates back to Norbert Wiener's 1938 paper [**?**], in which Wiener first introduced the notion of *multiple Wiener integral* calling it *polynomial chaos*. In 1947, Cameron and Martin [**?**] obtained the orthogonal development of nonlinear Brownian functionals in terms of *Hermite polynomials*[33], which motivated Itô's 1951 work [**?**].

In 1951, Itô modified the definition of multiple Wiener integral and made it more convenient for analysis in the point that the multiple Wiener integrals of different orders are orthogonal to each other, while Wiener's polynomial chaos does not possess such a property. Later in 1956, Itô generalized ideas from his previous paper and defined the multiple Wiener integral based on an *independently scattered* random measure. Note that the class of independently scattered random measures includes Gaussian and Poisson random measures. In the following section, we will define the Gaussian and Poisson Wiener chaoses in an unified manner.

From now on, we fix a σ-finite measure space $\big(\mathcal{Z}, \mathscr{Z}, \mu\big)$ such that $\{x\} \in \mathscr{Z}$ with $\mu(\{x\}) = 0$ for any $x \in \mathcal{Z}$, that is, μ is nonatomic. And we define $\mathscr{Z}_\mu := \{B \in \mathscr{Z} : \mu(B) < \infty\}$. All random objects in this thesis are defined on a common probability space $\big(\Omega, \mathcal{F}, \P\big)$.
The measure μ, appearing in Definition **??**, is called the *intensity measure*. The introduction of Gaussian and Poisson random measures can be treated as the randomization of the underlying intensity measure space $\big(\mathcal{Z}, \mathscr{Z}, \mu\big)$, and they constitute the basic building blocks of modern probability theory.

6.1. **A methodological breakthrough by Nualart and Ortiz-Latorre.** In 2008, Nualart and Ortiz-Latorre provided in [**?**] another equivalent condition for the FMT. More importantly, this paper contains a methodological breakthrough: the authors used Malliavin calculus tools, in particular the integration by parts formula on the Wiener space.

Theorem 6.1 (D. Nualart and S. Ortiz-Latorre, 2008)**.** *Let F_n be given as in* (5.3) *and have unit variance, then F_n converges in law to $\mathcal{N}(0,1)$ if and only if*

(iv) $\|DF_n\|^2$ converges in $L^2(\P)$ to p, as $n \to +\infty$.

Note $\|DF_n\|^2$ is somehow a reminiscence of the quadratic variation $\langle M^{(n)}\rangle_1$ from the random-time change, see page 4. Here we will just sketch the argument for the "if-part":

> By assumption, $F_n = p^{-1}\delta DF_n$ is bounded in $L^2(\P)$, hence admitting a subsequence F_{n_k} that converges in law to some random variable Z. It follows that $\mathbb{E}\big[e^{\mathrm{i}tF_{n_k}}F_{n_k}\big] = p^{-1}\mathbb{E}\big[e^{\mathrm{i}tF_{n_k}}\delta DF_{n_k}\big] = p^{-1}\mathbb{E}\big[\langle De^{\mathrm{i}tF_{n_k}}, DF_{n_k}\rangle\big] = p^{-1}\mathbb{E}\big[\mathrm{i}te^{\mathrm{i}tF_{n_k}}\langle DF_{n_k}, DF_{n_k}\rangle\big]$. If (iv) holds, then we have $\mathbb{E}\big[e^{\mathrm{i}tZ}Z\big] = \mathbb{E}\big[\mathrm{i}te^{\mathrm{i}tZ}\big]$. Denote by ϕ_Z the characteristic function of Z, then the previous equality suggests that $\phi_Z'(t) = -t\phi_Z(t)$, subject to the boundary condition $\phi_Z(0) = 1$. Solving this ODE gives us $Z \sim \mathcal{N}(0,1)$. This argument can be applied to any subsequence of (F_n), hence the original sequence indeed converges in law to $\mathcal{N}(0,1)$.

For the "only-if" part, one can first rewrite $\mathbb{E}\big[\big(\|DF_n\|^2 - p\big)^2\big] = \mathbb{E}\big[\|DF_n\|^4\big] - p^2$, which can be further expressed in terms of contractions, using the product formula (**??**). By condition (iii) in Nualart-Peccati's FMT, these contractions vanish asymptotically so that one can obtain the desired $L^2(\P)$-convergence.

[33]Hermite polynomials are orthogonal polynomials with respect to the Gaussian measure and it is one of the central ingredients in the Gaussian analysis. The first few Hermite polynomials are given by $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, and recursively, $H_{p+1}(x) := xH_p(x) - pH_{p-1}(x)$. Alternatively, one can define them via the *Rodrigues' formula* $H_p(x) = (-1)^p \exp(x^2/2)\frac{d^p}{dx^p}\exp(-x^2/2)$, $p \in \mathbb{N}$. See *e.g.* Section 1.4 in [**?**].

Remark 6.2. In 2009, Nourdin and Peccati [**?**] used exclusively results from Malliavin calculus to derive non-central extensions of the FMT. One of their main results states that for any sequence $\{f_k, k \geq 1\} \subset L^2_s(\mu^p)$ ($p \geq 2$ *even*) verifying $\lim_{k\to+\infty} p!\|f_k\|^2 = 2\nu > 0$, the following statements are equivalent, as $k \to +\infty$:

(a) $I_p^W(f_k)$ converges in law to $F(\nu)$, centered Gamma distribution[34] with parameter ν;

(b) $\mathbb{E}\big[I_p^W(f_k)^4\big] - 12\mathbb{E}\big[I_p^W(f_k)^3\big] \longrightarrow 12\nu^2 - 48\nu$;

(c) $\|DI_p^W(f_k)\|^2 - 2pI_p^W(f_k)$ converges in $L^2(\mathbb{P})$ to $2pv$.

This work exhibits again the power of Malliavin calculus.

A digression. The readers may have wondered "*why the fourth moment is so special for the asymptotic normality inside a fixed Gaussian Wiener chaos*". One immediate answer could be Theorem 6.1 and the following consequence of the product formula (**??**): for $F \in \mathbb{C}_q^W$,

$$\operatorname{Var}\big(q^{-1}\|DF\|^2\big) \leq \frac{q-1}{3q}\big(\mathbb{E}[F^4] - 3\mathbb{E}[F^2]^2\big) \leq (q-1)\operatorname{Var}\big(q^{-1}\|DF\|^2\big)\,, \tag{6.1}$$

see equation (5.2.7) in [**?**].

One can also answer this question using the classic method of moments and cumulants, as Nourdin explained in the short note [**?**]: for a normalized sequence $\{F_n, n \geq 1\} \subset \mathbb{C}_p^W$ such that $\mathbb{E}[F_n^4] \to 3$, one can prove, with some soft combinatorial arguments and the product formula, that every moment of F_n converges to the corresponding Gaussian moment. Here, let us provide a slightly different argument: as every monomial is a finite linear combination of Hermite polynomials[35], it suffices instead to show that $\mathbb{E}\big[H_k(F_n)\big] \to \mathbb{E}\big[H_k(Z)\big] = 0$ for[36] every $k \in \mathbb{N}$, $Z \sim \mathcal{N}(0,1)$. Note the expression $\mathbb{E}\big[H_k(X)\big]$ bears the particular name "*kth normal moment of X*" and is closely related to the χ^2-distance of Pearson, see [**?**, Proposition 8.2]. The first few normal moments of F_n are $\mathbb{E}[H_1(F_n)] = \mathbb{E}[F_n] = 0$, $\mathbb{E}[H_2(F_n)] = \mathbb{E}[F_n^2] - 1 = 0$, $\mathbb{E}[H_4(F_n)] = \mathbb{E}[F_n^4 - 6F_n^2 + 3] = \mathbb{E}[F_n^4] - 3$; for a generic $k \in \mathbb{N}$, we write $\mathbb{E}[H_{k+1}(F_n)] = \mathbb{E}[F_nH_k(F_n)] - \mathbb{E}[kH_{k-1}(F_n)]$ using the recursive relation of Hermite polynomials, and then we apply the integration by parts formula and chain rule as follows:

$$\mathbb{E}\big[F_nH_k(F_n)\big] = p^{-1}\mathbb{E}\big[\delta DF_nH_k(F_n)\big] = p^{-1}\mathbb{E}\big[\langle DF_n, DH_k(F_n)\rangle\big] = \mathbb{E}\big[p^{-1}\|DF_n\|^2kH_{k-1}(F_n)\big]\,,$$

thus,

$$\big|\mathbb{E}[H_{k+1}(F_n)]\big| = \big|\mathbb{E}\big[(p^{-1}\|DF_n\|^2 - 1)kH_{k-1}(F_n)\big]\big| \leq k\sqrt{\operatorname{Var}\big(p^{-1}\|DF_n\|^2\big)}\sqrt{\mathbb{E}\big[H_{k-1}(F_n)^2\big]}\,;$$

see also (**??**) for similar computations. The above inequality, together with (6.1), implies that the fourth moment controls other moments asymptotically.

Lastly, let us mention Rosiński's *intuitive* explanation [**?**] based on the independence on Wiener space. Let us first recall the seminal result of Üstünel and Zakai [**?**] that provides the necessary and sufficient condition for independence of two multiple Wiener-Itô integrals with respect to the same Gaussian random measure.

[34]*i.e.* $F(v) = 2\gamma_{\nu/2} - \nu$, where $\gamma_{\nu/2}$ has a Gamma law with parameter $\nu/2$.

[35]By definition (see footnote 8), Hermite polynomial is a finite linear combination of monomials, then inductively, one can easily verify that every monomial can be expressed as a finite linear combination of Hermite polynomials.

[36]Like to crack a walnut with a sledgehammer: Z is equal in law to $W(h)$ for some unit vector $h \in L^2(\mu)$, and Cameron-Martin's result [**?**] suggests that $H_k(W(h)) = I_k^W(h^{\otimes k})$, which is an element of kth Gaussian Wiener chaos, and this explains $\mathbb{E}[H_k(Z)] = 0$ for every $k \in \mathbb{N}$.

Theorem 6.3 (A.S. Üstünel and M. Zakai, 1989). *Fix $p, q \in \mathbb{N}$ and $f \in L^2_s(\mu^p)$, $g \in L^2_s(\mu^q)$, we set $F = I^W_p(f)$, $G = I^W_q(g)$. Then F, G are independent if and only if $f \otimes_1 g = 0$.*

Later, Rosiński and Samorodnitsky [**?**] made another interesting observation.

Proposition 6.4 (J. Rosiński and G. Samorodnitsky, 1999). *Fix $p, q \in \mathbb{N}$, then $\mathrm{Cov}(F^2, G^2) \geq 0$ for any $F \in \mathbb{C}^W_p$, $G \in \mathbb{C}^W_q$, with equality only when F and G are independent.*

Here comes Rosiński's nice heuristic arguments: suppose that $F \in \mathbb{C}^W_p$ satisfies $\mathbb{E}[F^2] = 1$ and[37] $\mathbb{E}[F^4] = 3$, let G be an independent copy of F, then $\mathrm{Cov}\big((F+G)^2, (F-G)^2\big) = 2\mathbb{E}[F^4] - 6\mathbb{E}[F^2]^2 = 0$, so it follows from Proposition 6.4 that $F + G$ is independent of $F - G$. Thus, by Bernstein's theorem, $F \sim \mathcal{N}(0, 1)$. In the statement of the FMT, $\mathbb{E}[F_n^4] = 3$ only asymptotically, it is natural to expect $F_n \to \mathcal{N}(0, 1)$. Such heuristics have led to several papers on asymptotic independence on the Wiener space [**?**, **?**], with applications to time series analysis [**?**].

The above ***digression*** around the importance of the fourth moment may also trigger one's desire of quantifying the CLT on Wiener chaos in terms of the fourth moment. It is known, for example from [**?**, Proposition C.3.2], that a sequence of random variables X_n converges in law to the standard Gaussian N if and only if $d_{\mathrm{Kol}}(X_n, N) \to 0$, where d_{Kol} denotes the Kolmogorov distance[38] appearing in the celebrated Berry-Esséen bound. So is it true that

$$(\#) \qquad d_{\mathrm{Kol}}\big(F, \mathcal{N}(0, \mathrm{Var}(F))\big) \leq \text{Constant} \times \sqrt{\mathbb{E}[F^4] - 3\mathbb{E}[F^2]^2} \text{ for any } F \in \mathbb{C}^W_p \text{ ?}$$

The answer is indeed positive and will be the highlight of the story told in the next section. It is also worth pointing out the paper [**?**] by Azmoodeh, Malicet, Mijoule and Poly, who provided a generalization of FMT by establishing the equivalence between asymptotic normality and convergence of *even* moments on the Gaussian Wiener chaos; see also (**??**).

6.2. **Application: parameter estimation.**

6.3. **Application: GOE approximation of Wishart matrices.**

6.4. **Application: zeros of random polynomials.**

7. **Regularity of Gaussian objects**

1. Kolmogorov theorems
2. analysis of PDEs.

[37] In fact, if $p \geq 2$, $\mathbb{E}[F^4] > 3$, see *e.g.* [**?**, Corollary 2] or our Chapter 3. So we view $\mathbb{E}[F^4] = 3$ as an asymptotic relation.

[38] $d_{\mathrm{Kol}}(X, N) := \sup\{|¶(X \leq t) - ¶(N \leq t)| : t \in \mathbb{R}\}$, see also Section 2.1.1.

Appendix A. Basic results in analysis

• Given an interval (a, b) with $a < b$. We say a function $f : (a, b) \to \mathbb{R}$ is **absolutely continuous** if for any $\varepsilon > 0$, there exists $\delta > 0$ such that whenever a *finite* sequence of *pairwise disjoint* sub-intervals (a_k, b_k) of (a, b) with $a_k < b_k$ and

$$\sum_k (b_k - a_k) < \varepsilon,$$

we have

$$\sum_k |f(b_k) - f(a_k)| < \varepsilon.$$

In particular, an absolutely continuous function is uniformly continuous. Note that an Hölder continuous function is uniformly continuous but in general not absolutely continuous.[39] Moreover, an absolutely continuous function is almost everywhere differentiable. This fact is contained in the following fundamental theorem of (Lebesgue integral) calculus.

Theorem A.1. *Let $a < b$ be two real numbers. Suppose $f : [a, b] \to \mathbb{R}$ is a continuous function. Then, the following statements are equivalent:*

(i) *f is absolutely continuous;*
(ii) *f has a derivative f' almost everywhere that is Lebesgue integrable, and*

$$f(x) = f(a) + \int_a^x f'(t)dt$$

for all $x \in [a, b]$; (the above integral is defined in Lebesgue's sense)
(iii) *there exists a real-valued and Lebesgue integrable function g on $[a, b]$ such that*

$$f(x) = f(a) + \int_a^x g(t)dt$$

for all $x \in [a, b]$.

If these equivalent conditions are satisfied, then necessarily any function g as in (iii) *satisfies $g = f'$ almost everywhere.*

As a consequence of Theorem A.1, one can easily derive the chain rule for absolutely continuous functions:

$$\frac{d}{dx}(f \circ h)(x) = f'(h(x))h'(x) \text{ a.e.} \tag{A.1}$$

for any absolutely continuous functions $f, g : \mathbb{R} \to \mathbb{R}$.

Theorem A.2 (Integration by parts)**.** *Let $f, g : [a, b] \to \mathbb{R}$ be absolutely continuous with $< -\infty < a < b < \infty$. Then, fg is absolutely continuous and*

$$\int_a^b f(t)g'(t)dt = \big[f(b)g(b) - f(a)g(a)\big] - \int_a^b f'(t)g(t)dt.$$

[39] We say $h : \mathbb{R} \to \mathbb{R}$ is Hölder continuous with exponent $\beta \in (0, 1]$ if there is some $C_\beta \in (0, \infty)$ such that

$$|h(x) - h(y)| \le C_\beta |x - y|^\beta \quad (*)$$

for any $x, y \in \mathbb{R}$. When $\beta = 1$, we call h a Lipschitz function and in this case, h is clearly absolutely continuous with f' uniformly bounded by the constant C_1 in $(*)$.

• Cauchy-Schwarz inequality:

$$\left|\sum_n a_n b_n\right| \le \Big(\sum_n a_n^2\Big)^{1/2}\Big(\sum_n b_n^2\Big)^{1/2}$$

$$\left|\int_X f(x)g(x)\mu(dx)\right| \le \Big(\int_X f(x)^2\mu(dx)\Big)^{1/2}\Big(\int_X g(x)^2\mu(dx)\Big)^{1/2}$$

• Hölder inequality, Jensen's inequality, Minkowski inequality, Young's inequality, **XXX**

Proposition A.3. (1) [Generalized dominated convergence] Suppose f_n, g_n, f, and g are real integrable functions defined on some measure space $(E, \mathcal{E}, \mu)$ such that

(i) for μ-almost every $x \in E$, $f_n(x) \to f(x)$ as $n \to +\infty$;
(ii) for μ-almost every $x \in E$, $g_n(x) \to g(x)$ as $n \to +\infty$;
(iii) $|f_n| \le g_n$ for each $n \ge 1$;
(iv) $\int_E g_n d\mu \to \int_E g d\mu$ as $n \to +\infty$.

Then, $\int_E f_n d\mu \to \int_E f d\mu$ as $n \to +\infty$. See, e.g., [Bas24, Exercise 7.5, p. 64].

(2) [Vitalli convergence theorem] Let $(E, \mathcal{E}, \mu)$ be a finite measure space and $p \in [1, \infty)$. Let $\{f_n\}$ be a sequence of functions in $L^p(E, \mathcal{E}, \mu)$ and let f be a real-valued and $\mathcal{E}$-measurable function. Then the following statements are equivalent:

(a) $f \in L^p(E, \mathcal{E}, \mu)$ and f_n converges to f in $L^p(E, \mathcal{E}, \mu)$.
(b) f_n converges in μ-measure to f and $\{|f_n|^p\}_{n\ge1}$ is uniformly integrable.

• **Basic Hilbert space notations.**

Appendix B. **Basic results in probability**

Definition B.1. Given a nonempty index set J, we say $\{X_j\}_{j\in J}$ is a real Gaussian family (or jointly Gaussian) if any finite linear combination

$$\sum_{j\in J_0} a_j X_j$$

with $a_j \in \mathbb{R}$ and J_0 any finite subset of J, is a real Gaussian random variable.[40] When J is a finite set with cardinality d, we view $\{X_j\}_{j\in J}$ as a random vector in $\mathbb{R}^d$ and call it a d-dimensional Gaussian vector.

It is well known in undergraduate probability that the law of a Gaussian family $\{X_j\}_{j\in J}$ can be uniquely determined by the mean function and covariance structure:

$$\text{mean} \quad m_j = \mathbb{E}[X_j] \quad \text{and} \quad \text{covariance} \quad C_{ij} = \mathbb{E}[X_i X_j] - m_i m_j$$

for any $i, j \in J$. When $m_j = 0$ for all j, we call $\{X_j\}_{j\in J}$ a center Gaussian family. In general, the family $\{X_j - m_j\}_{j\in J}$ is always centered.

In Section 2, we work with n-dimensional Gaussian vectors $\mathbf{G} \sim \mathcal{N}_d(0, C)$ i.e., $\mathbf{G} = (G_1, ..., G_d)$ is a centered Gaussian family with covariance matrix

$$\mathbb{E}[G_i G_j] = C_{ij}.$$

When $C = \mathbf{I}$ is the $d \times d$ identity matrix, $\mathbf{G}$ is known as the standard Gaussian vector on $\mathbb{R}^d$, whose law is called the standard Gaussian measure on $\mathbb{R}^n$:

$$\gamma_d(dx) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}|x|^2} dx.$$

Here, we record a few fundamental results on Gaussian vectors. For $X \sim \mathcal{N}_n(\mu, C)$, by definition, we have $a \cdot X = \sum_{i=1}^n a_i X_i$ is a real Gaussian random variable with mean $a \cdot \mu$ and variance $a \cdot (Ca) = \sum_{i,j} a_i C_{ij} a_j$. Thus,

$$\mathbb{E}[e^{ia\cdot X}] = e^{ia\cdot\mu - \frac{1}{2} a\cdot Ca}$$

for any $a \in \mathbb{R}^n$. This indeed tells us that the law of X is characterized by its mean and covariance matrix. Moreover, the following results follow immediately.

Lemma B.2. (i) *If $X \sim \mathcal{N}_n(\mu, C)$, then $AX + b \sim \mathcal{N}_m(A\mu + b, ACA^t)$ for any $b \in \mathbb{R}^m$ and A a $m \times n$ matrix.*

(ii) *Suppose $X \sim \mathcal{N}_n(\mu, C)$ and $k \in \{1, ..., n\}$. Assume that $I_1, ..., I_k$ are k disjoint subsets of $\{1, ..., n\}$ such that*

$$\mathrm{Cov}(X_i, X_j) = 0 \quad \textit{whenever } i, j \textit{ are in distinct } I_\ell\textit{'s}.$$

Then, the k families $\{X_i\}_{i\in I_1}$, ..., $\{X_i\}_{i\in I_k}$ are independent.

(iii) *Suppose $X \sim \mathcal{N}_n(0, C)$ with C invertible. Then, $C^{-\frac{1}{2}} X \sim \mathcal{N}_n(0, I_n)$.*

Lemma B.3 (Wick-Isserlis formula)**.** *Let $X \sim \mathcal{N}_n(0, C)$. Then,*

$$\mathbb{E}\bigg(\prod_{i=1}^n X_i\bigg) = \sum_\pi \prod_{\{i,j\}\in\pi} C_{ij}, \tag{B.1}$$

[40] Let X be a standard normal and B be a symmetric Bernoulli random variable (i.e., $\mathbb{P}(B = +1) = \mathbb{P}(B = -1) = 1/2$) such that B is independent of X, then BX is a standard normal but X and BX are not jointly Gaussian.

where the sum runs over all pair partitions of $\{1, ..., n\}$. *In particular, the above expectation vanishes when* n *is odd.*

Let us first explain the notion of pair partition of an even number of objects or essentially of the set $[2m] = \{1, ..., 2m\}$ for $m \in \mathbb{N}_0$. We say π is a *pair partition* of $\{1, ..., 2m\}$, denoted by $\pi \in \mathscr{P}_2([2m])$, if

$$\pi = \{\{i_1, i_2\}, ..., \{i_{2m-1}, i_{2m}\}\} \subset 2^{[2m]}$$

with $i_1, ..., i_{2m}$ mutually distinct. For example, $\{\{1, 3\}, \{2, 4\}\}$ is a pair partition of $[4]$ but not a pair partition of $[6]$.

Exercise B.4. Prove Lemma B.3 using the following identity at $t = 0$:

$$\mathbb{E}\Big[e^{t \cdot X} \prod_{j=1}^{n} X_j\Big] = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \mathbb{E}\big[e^{t \cdot X}\big] = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \exp\Big(\sum_{i,j=1}^{n} t_i t_j C_{ij}\Big).$$

Exercise B.5. *Let* $X \sim \mathcal{N}_2(0, C)$ *and* $(m_1, m_2) \in \mathbb{N}_0^2$. *Show that*

$$\mathbb{E}[X_1^{2m_1} X_2^{2m_2}] \geq \mathbb{E}[X_1^{2m_1}]\mathbb{E}[X_2^{2m_2}].$$

Proposition B.6. Let $(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m \geq 0}, \mathbb{P})$ be a filtered probability space with $\mathcal{F} = \sigma\{\mathcal{F}_m : m \geq 0\}$. Given $G \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for some $p \in [1, \infty)$, we define $G_m = \mathbb{E}[G | \mathcal{F}_m]$ for $m \geq 1$. Then, G_m converges almost surely and in $L^p(\Omega)$ to G as $m \to +\infty$.

Proof. The statement for $p = 1$ can be found, e.g., on page 264 in [LeG16]. The general case follows immediately from Vitalli's convergence theorem (Proposition A.3-(2)). $\square$

Appendix C. **Answers to Questions**

- *Answer to Question 1.2*

References

[BR89] P. Baldi, Y. Rinott, *On normal approximations of distributions in terms of dependency graphs,* Ann. Probab. 17 (1989), no. 4, 1646–1650.

[BC04] A.D. Barbour, L.H.Y. Chen, *An Introduction to Stein's Method,* Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore: Volume 4, 2004.

[BC05] A.D. Barbour, L.H.Y. Chen, *Stein's Method and Applications,* Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore: Volume 4, 2004.

[BKR89] A.D. Barbour, M. Karoński, A. Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs,* J. Combin. Theory Ser. B 47 (1989), no. 2, 125–145.

[Bas24] R.F. Bass, *Real analysis for graduate students*, Version 5.0, October 31, 2024; available on the author's webpage.

[Cha09] S. Chatterjee, *Fluctuations of eigenvalues and second order Poincaré inequalities,* Probab. Theory Related Fields 143 (2009), no. 1-2, 1–40.

[Cha14a] S. Chatterjee, *A short survey of Stein's method,* Proceedings of the ICM, 2014.

[Cha14b] S. Chatterjee, *Superconcentration and related topics*, Springer Monogr. Math. Springer, Cham, 2014. x+156 pp.

[Cha75] L.H.Y. Chen, *Poisson approximation for dependent trials,* Ann. Probability 3 (1975), no. 3, 534–545.

[CGS11] L.H.Y. Chen, L. Goldstein, Q.M. Shao, *Normal approximation by Stein's method,* Probability and Its Applications, 2011 Springer-Verlag Berlin Heidelberg.

[CR10] L.H.Y. Chen, A. Röllin, *Stein couplings for normal approximation,* arXiv:1003.6039v2 [math.PR].

[CS04] L.H.Y. Chen, Q.-M. Shao, *Normal approximation under local dependence,* Ann. Probab. 32 (2004), no. 3A, 1985–2028.

[Dia77] P. Diaconis, *The distribution of leading digits and uniform distribution mod 1*, Ann. Probability, 5(1):72–81, 1977.

[DdMN25] C.-P. Diez, L. da Maia, I. Nourdin, *Mathematical research with GPT-5: a Malliavin-Stein experiment*, arXiv:2509.03065v1 [math.PR].

[Dud10] R.M. Dudley, *Real Analysis and Probability,* 2nd edition, Cambridge University Press, 2010.

[HMRZ20] J. Hązła, E. Mossel, N. Ross, G. Zheng, *The probability of intransitivity in dice and close elections*, Probab. Theory Related Fields 178 (2020), no. 3-4, 951–1009.

[Hull22] J.C. Hull, *Options, futures, and other derivatives*, 11th edition, Pearson (2022).

[Kah85] J.-P. Kahane, *Sur le chaos multiplicatif,* Ann. Sci. Math. Québec, 9 no.2, 1985.

[LeCam86] L. Le Cam, *The central limit theorem around 1935,* Statist. Sci. 1 (1986), no. 1, 78–96.

[LeG16] J.-F. Le Gall, *Brownian motion, martingales, and stochastic calculus,* [Translated from the 2013 French edition], Grad. Texts in Math., 274 Springer, [Cham], 2016. xiii+273 pp.

[MT06] P. Malliavin, A. Thalmaier, *Stochastic calculus of variations in mathematical finance*, Springer Finance Springer-Verlag, Berlin, 2006. xii+142 pp.

[Man72] B.B. Mandelbrot, *A possible refinement of the lognormal hypothesis concerning the distribution of energu in intermittent turbulence,* Statistical models and turbulence, La Jolla, CA, Lecture Notes in Phys. no.12, Springer, (1972), 331-351.

[NO25] M. Nica, J. Ortmann, *A Gaussian integral formula for the Hermite polynomials: Combinatorics, Asymptotics and Applications,* arXiv:2508.13910 [math.PR].

[NP12] I. Nourdin and G. Peccati, *Normal approximations with Malliavin calculus: from Stein's method to universality*, Cambridge Tracts in Mathematics 192. Cambridge University Press, Cambridge, 2012, xiv+239.

[Nua06] D. Nualart, *The Malliavin calculus and related topics,* Second edition. Probability and its Applications (New York). Springer-Verlag, Berlin, 2006. xiv+382 pp.

[NN18] D. Nualart, E. Nualart, *Introduction to Malliavin calculus,* Institute of Mathematical Statistics Textbooks, 9. Cambridge University Press, Cambridge, 2018. xii+236 pp.

[RV14] R. Rhodes, V. Vargas, *Gaussian multiplicative chaos and applications: a review*, Probab. Surv. 11 (2014), 315–392.

[Rol22] A. Röllin, *Kolmogorov bounds for the normal approximation of the number of triangles in the Erdős-Rényi random graph,* Probab. Engrg. Inform. Sci. 36 (2022), no. 3, 747–773.

[Ros11] N. Ross, *Fundamentals of Stein's method,* Probability Surveys. Vol. 8, 210-293, 2011.

[Sh04] S.E. Shreve, *Stochastic calculus for finance. II. Continuous-time models,* Springer Finance Springer-Verlag, New York, 2004. xx+550 pp.

[St72] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Berkeley Symp. on Math. Statist. and Prob. Proc. Sixth Berkeley Symp. on Math. Statist. and Prob., Vol. 2 (Univ. of Calif. Press, 1972), 583-602.

[St86] C. Stein, *Approximate computation of expectations,* In Institute of Mathematical Statistics Lecture Notes - Monograph Series, volume 7. Institute of Mathematical Statistics, 1986.

[vdH17] R. van der Hofstad, *Random graphs and complex networks,* Vol. 1. Camb. Ser. Stat. Probab. Math., [43] Cambridge University Press, Cambridge, 2017. xvi+321 pp.