# Regularity lemma for Gaussian polynomials

Chris Jones*

November 3, 2026

**Abstract**

We prove a regularity lemma for polynomials with fixed degree $d$ and any number of standard Gaussian input variables. Following modern approaches to the Szemerédi regularity lemma, we deduce our results from a general formulation of the regularity lemma using ultraproducts. Along the way, we give a direct combinatorial proof of the celebrated "fourth moment theorem" of Nualart and Peccati, which characterizes the asymptotically Gaussian polynomials inside degree-$d$ Gaussian chaos as those such that $\mathbb{E}\, p^2 \to 1$ and $\mathbb{E}\, p^4 \to 3$. We apply the regularity lemma to obtain a short proof of the robust Borell inequality of Mossel and Neeman.

## 1 Introduction

Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. How many "different" degree-$d$ polynomials $p(X_1, \dots, X_n)$ are there, in the sense of, say, the law of $p$ as a random variable? While $p$ seems like it could become arbitrarily complicated as $n \to \infty$, this is not the case. Some examples are:

1. $p(X_1) = \mathrm{He}_d(X_1)$ where $\mathrm{He}_d$ is the degree-$d$ probabilists' Hermite polynomial.
2. $p(X_1, \dots, X_n) = \frac{1}{\sqrt{n d!}} \sum_{i=1}^n \mathrm{He}_d(X_i)$. By the central limit theorem, the law of this random variable is asymptotically $\mathcal{N}(0,1)$ as $n \to \infty$.
3. $p(X_1, \dots, X_n) = \frac{1}{n} \sum_{1 \le i < j \le n} X_i X_j$. This is asymptotic to $\frac{1}{2}\mathrm{He}_2(Z)$ for $Z \sim \mathcal{N}(0,1)$.

Despite the limit $n \to \infty$ in examples 2 and 3, the law of $p$ approaches the law of a polynomial in just one standard Gaussian variable. Herry et al. [HMP24] proved that this is a general phenomenon: any sequence of degree-$d$ Gaussian polynomials which converges in law, converges in law towards another Gaussian polynomial (possibly of lower degree).

Here we characterize the asymptotic limits of Gaussian polynomials in much stronger ways, broadly termed "regularity lemmas for Gaussian polynomials". First, we prove a regularity lemma in the traditional sense of approximation by a structured constant-size object. For a set of variables $X$, let $\mathcal{H}_d(X)$ be the span of the degree-$d$ Hermite polynomials in $X$, and let $\mathcal{H}_{\le d}(X)$ be the span of all monomials in $X$ up to degree $d$. $\mathcal{H}_d(X)$ is also called the "degree-$d$ Gaussian chaos" for $X$.

**Theorem 1.1** (Regularity lemma, approximation version)**.** *For every $d \in \mathbb{N}$, $\varepsilon > 0$ there exists $M = M(d, \varepsilon)$ such that: for all $p \in \mathcal{H}_{\le d}(X)$ such that $\mathbb{E}\, p^2 \le 1$, there is $m \le M$, a polynomial $q(Z_1, \dots, Z_m), \deg(q) \le d$ and $Z_a \in \mathcal{H}_{\le d}(X)$ such that:*

*(i)* $\mathbb{E}(p - q)^2 \le \varepsilon$

*UC Davis, Department of Mathematics. chijones@ucdavis.edu

*(ii)* $W_2((Z_1,\dots,Z_m),\mathcal{N}(0,1)^{\otimes m}) \le \varepsilon$ *where* $W_2$ *is 2-Wasserstein distance.*

To prove Theorem 1.1, we will deduce it from a semantically stronger theorem characterizing the *ultraproduct* of Gaussian polynomials [ES12, Tao13]. Roughly speaking, our theorem states that the possible limits of Gaussian polynomials are "graded Gaussian polynomials".

At an intermediate level between the completely finitary Theorem 1.1 and the ultraproduct viewpoint, one can also view a regularity lemma as a "compactification". For example, the space of graphons compactifies the space of graphs under the cut metric [LS06, ES12, Lov12]. At this level, we prove that graded-degree-$d$ Gaussian polynomials are the closure of degree-$d$ Gaussian polynomials in distribution (and the compactification when the space is bounded, $\mathbb{E}\,p^2 \le 1$).

**Theorem 1.2** (Regularity lemma, compactification version)**.** *Let $X$ be a set of countably many variables, and let $Z$ be a set of graded variables with countably many variables of each grade. Let $\mathcal{H}_d^{\mathrm{gr}}(Z)$ be the span of graded-degree-$d$ Hermite polynomials in $Z$. Let $d \in \mathbb{N}$.*

*(i) For all $p \in \mathcal{H}_d^{\mathrm{gr}}(Z)$, there exists a sequence $p_n \in \mathcal{H}_d(X_1,\dots,X_n)$ such that $p_n \xrightarrow{d} p$.*

*(ii) If $p_n \in \mathcal{H}_d^{\mathrm{gr}}(Z)$ satisfies $p_n \xrightarrow{d} p$ for some $p$, then $p_n \xrightarrow{d} p'$ for some $p' \in \mathcal{H}_d^{\mathrm{gr}}(Z)$.*

Equivalently, $\{\mathrm{Law}(p) : p \in \mathcal{H}_d^{\mathrm{gr}}(Z)\}$ is the closure of $\{\mathrm{Law}(p) : p \in \mathcal{H}_d(X)\}$ in the space of probability measures on $\mathbb{R}$ with the weak topology. A consequence of (ii) is that $\{\mathrm{Law}(p) : p \in \mathcal{H}_d^{\mathrm{gr}}(Z), \mathbb{E}\,p^2 \le 1\}$ is compact in the weak topology.

In the main text, we will prove multivariate versions of these theorems. Using a hypercontractivity argument, we can also generically upgrade Theorem 1.2 from the weak topology to Wasserstein metrics, see Section A. Considering it to be a problem of independent interest, Herry et al. [HMP24] proved a variant of this theorem, which is that $\mathcal{H}_{\le d}(X)$ is closed in distribution when $X$ is a set of countably many standard Gaussian variables.

There is a large body of work on characterizing the Gaussian limits inside $\mathcal{H}_d(X)$. The famous "fourth moment theorem" of Nualart and Peccati [NP05, NOL08, NP15] states that, if $\mathbb{E}\,p_n^2 \to 1$, then $p_n \xrightarrow{d} \mathcal{N}(0,1)$ if and only if $\mathbb{E}\,p_n^4 \to 3$ if and only if all "contraction norms" of $p_n$ vanish. We identify a small extension: if and only if all proper flattenings of the coefficient tensor of $p_n$ have vanishing spectral norm. We use this to give a direct combinatorial proof of the fourth moment theorem and then to prove the regularity lemma.

Our idea for the regularity lemma was motivated by an application to the theory of the Grothendieck constant [HJM26]. In Section 4 we give another application: a much shorter proof of the "robust Borell inequality" of Mossel and Neeman [MN15]. Borell's inequality states that the maximum possible noise stability of a set $A \subseteq \mathbb{R}^n$ with fixed Gaussian measure is a halfspace. The robust version states that any near-maximizer is close to a halfspace in $L_2$. The regularity lemma gives robustness results like this almost for free, by a general strategy:

1. Suppose we want to determine which Gaussian polynomial $p \in \mathcal{H}_{\le d}(X)$ maximizes a target quantity. As long as the target is suitably continuous (such as noise stability), there exists a maximizer in the compactified space of graded Gaussian polynomials.

2. Identify the maximizers on the compactified space. Perturbative arguments can be used here (we can search for local optima, then select the largest, to identify the global optimum).

3. Robustness is free: by compactness and continuity, for every $\varepsilon > 0$, there exists $\delta > 0$ such that all $\delta$-near-optimal points are within $\varepsilon$ of global optima (in a suitable metric).

For the robust Borell inequality, we use a slight variant of this argument mainly because set indicator functions are not polynomial functions. Since noise stability smooths away high-degree components, approximation by a low-degree polynomial still works.

**Statement of AI use.** An initial draft of the regularity lemma was written in April 2026 using some assistance from GPT-5.4 (given the tensor decomposition strategy, it was able to suggest using the fourth moment theorem, although I could not understand its proof strategy). For the current version, GPT-6 Astra helped close technical gaps and extend the proofs. While the first version used an energy-increment argument, GPT-6 Astra suggested the ultraproduct approach (communicated by Giulio Malavolta). The author takes responsibility for the contents of the paper and any remaining errors.

## 2 Graded Gaussian polynomials

A graded variable is a variable with an integer label $r > 0$ called its "grade". We write $X_1, X_2, X_3, \ldots$ for "original" grade-one variables and $Z_1^{(r)}, Z_2^{(r)}, Z_3^{(r)}, \ldots$ for grade-$r$ variables, with the identification $X_i = Z_i^{(1)}$ at grade one. Sample $Z_i^{(r)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ independently across all variables regardless of grade.

The idea of the asymptotic limit is that an asymptotically Gaussian random variable in $\mathcal{H}_d(X)$ can be replaced in the limit by a single new grade-$d$ variable. In this way, we will represent the limit of a sequence of Gaussian polynomials by a single graded polynomial. When we do an operation on this limit object which depends on the degree (e.g. calculate the degree-$d$ weight) then we use the grade information to recover the "original" degree.

We use the following standard notions from Gaussian analysis. Let $\mathrm{He}_k$ be the unnormalized probabilists' Hermite polynomials. Let $\psi_k = \mathrm{He}_k / \sqrt{k!}$ be their normalizations. For a finitely-supported multi-index $\alpha \in \mathbb{N}^X$ let $X^\alpha = \prod_i X_i^{\alpha_i}$ and similarly for $\mathrm{He}_\alpha$ and $\psi_\alpha$. Let $|\alpha| = \sum_i \alpha_i$ and $\alpha! = \prod_i \alpha_i!$. Then $\{\psi_\alpha(X) : \alpha \in \mathbb{N}^X, |\alpha| < \infty\}$ is an orthonormal basis of $L_2(X)$. Let $\mathcal{H}_d(X)$ be the closed span of the degree-$d$ Hermite polynomials in $X$. Let $\mathcal{H}_{\le d}(X) = \oplus_{i=0}^d \mathcal{H}_i(X)$. Let $\langle f, g\rangle = \mathbb{E}\, fg$ and $\|f\|_2 = (\mathbb{E}\, f^2)^{1/2}$.

Polynomials in $Z$ have an additional notion of total grade, or graded degree. Let $\mathrm{gr}(\alpha) = \sum_{i,r} r\alpha_{ir}$ be the total grade of a multi-index $\alpha \in \mathbb{N}^Z$. The Hermite polynomial $\mathrm{He}_\alpha(Z)$ is defined as before, and $\psi_\alpha(Z)$ are still an orthonormal basis for $L_2(Z)$. Define graded-degree-$d$ chaos by

$$\mathcal{H}_d^{\mathrm{gr}}(Z) = \overline{\mathrm{span}}\,(\mathrm{He}_\alpha(Z) : \mathrm{gr}(\alpha) = d)\ . \tag{2.1}$$

The notation $\overline{\mathrm{span}}$ denotes the closure of the linear span in $L_2(Z)$.

The following notions on Gaussian space interact with degree, and so behave differently on graded Gaussian space. Let $\Pi_d$ be the operator on $L_2(Z)$ which projects to $\mathcal{H}_d^{\mathrm{gr}}(Z)$. Let $W_d(f) = \|\Pi_d f\|_2^2$. For variables with grade $r$, the Ornstein–Uhlenbeck noise operator at parameter $t \in [0,1]$ acts by sampling standard Gaussians $W$ with covariance $\mathbb{E}[W_i^{(r)} Z_i^{(r)}] = t^r$. Under these definitions, we still have the noise correlation and Parseval identities,

$$\mathop{\mathbb{E}}_{(W,Z)\sim t\text{-correlated}}[f(W)g(Z)] = \sum_{d\ge 0} t^d \langle \Pi_d f, \Pi_d g\rangle\,, \qquad \mathbb{E}[f(Z)^2] = \sum_{d\ge 0} W_d(f)\,. \tag{2.2}$$

We verify that all $p \in \mathcal{H}_d^{\mathrm{gr}}(Z)$ can be approximated by finite Gaussian polynomials, which is the first (easier) part of Theorem 1.2.

**Lemma 2.1.** *Let $d_k \in \mathbb{N}$ for $k \in \mathbb{N}$. For all $p_k \in \mathcal{H}^{\mathrm{gr}}_{d_k}(Z)$ there exist sequences $p_{n,k} \in \mathcal{H}_{d_k}(X_1, \dots, X_n)$ such that $(p_{n,k})_{k\in\mathbb{N}} \xrightarrow{d} (p_k)_{k\in\mathbb{N}}$.*

*Proof.* For each graded variable $Z_i^{(r)}$ introduce new standard Gaussians $X_{ir,j}$ for $j \in [N]$. Define approximants for $Z_i^{(r)}$ and $\mathrm{He}_m(Z_i^{(r)})$ by,

$$S_{ir} = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \psi_r(X_{ir,j}), \qquad U_{m,ir} = \frac{1}{N^{m/2}} \sum_{\substack{j_1,\dots,j_m \in [N] \\ j_1,\dots,j_m \text{ distinct}}} \psi_r(X_{ir,j_1})\psi_r(X_{ir,j_2}) \cdots \psi_r(X_{ir,j_m})\,.$$

We approximate $p_k \in \mathcal{H}^{\mathrm{gr}}_{d_k}(Z)$ by:

$$p_{N,k} = \sum_{\alpha} c_\alpha \prod_{i,r} U_{\alpha_{ir},ir} \qquad \text{where} \qquad p_k = \sum_{\alpha} c_\alpha \,\mathrm{He}_\alpha(Z) = \sum_{\alpha} c_\alpha \prod_{i,r} \mathrm{He}_{\alpha_{ir}}(Z_i^{(r)})\,.$$

All terms in $p_{N,k}$ have Hermite degree $d_k$ so $p_{N,k} \in \mathcal{H}_{d_k}(X)$. It remains to show $(p_{N,k})_{k\in\mathbb{N}} \xrightarrow{d} (p_k)_{k\in\mathbb{N}}$ as $N \to \infty$. This is equivalent to $(p_{N,1}, \dots, p_{N,k}) \xrightarrow{d} (p_1, \dots, p_k)$ for every finite $k$.

First assume each $p_k = \mathrm{He}_{\alpha^k}(Z)$ is a single Hermite polynomial. Consider $\tilde{p}_{N,k} = \mathrm{He}_{\alpha^k}(S)$. By the multivariate central limit theorem, $(\tilde{p}_{N,1}, \dots, \tilde{p}_{N,k}) \xrightarrow{d} (p_1, \dots, p_k)$. On the other hand,

$$\mathbb{E}(\tilde{p}_{N,k} - p_{N,k})^2 = O(N^{-1})\,. \tag{2.3}$$

This is a combinatorial moment calculation which we include in Lemma B.5.

Combine Eq. (2.3) and the convergence $(\tilde{p}_{N,1}, \dots, \tilde{p}_{N,k}) \xrightarrow{d} (p_1, \dots, p_k)$ with Slutsky's lemma to deduce $(p_{N,1}, \dots, p_{N,k}) \xrightarrow{d} (p_1, \dots, p_k)$. This completes the case of single Hermite polynomials.

The continuous mapping theorem gives $(p_{N,1}, \dots, p_{N,k}) \xrightarrow{d} (p_1, \dots, p_k)$ if every $p_k$ is a finite sum. For $p_k \in \mathcal{H}^{\mathrm{gr}}_{d_k}(Z)$ with an infinite number of terms, take a sequence of finite $L_2$ approximants, approximate each one, then take a diagonal sequence to obtain $(p_{N,1}, \dots, p_{N,k}) \xrightarrow{d} (p_1, \dots, p_k)$. □

# 3 Proof of the regularity lemma

## 3.1 Regularity lemma for degree 1 and 2

A degree-1 polynomial is of the form $p(X) = \sum_i c_i X_i$. The law of this polynomial is always Gaussian, so the characterization of degree-1 Gaussian polynomials is immediate.

As a warm-up, we prove a version of Theorem 1.1 for degree-2 Gaussian polynomials.

**Lemma 3.1.** *Let $\varepsilon > 0$. For all $p \in \mathcal{H}_2(X_1, \dots, X_n)$ such that $\mathbb{E}\,p^2 \le 1$, there exist $Z_0 \in \mathcal{H}_2(X_1, \dots, X_n)$ and $Z_1, \dots, Z_m \in \mathcal{H}_1(X_1, \dots, X_n)$ with $m \le 2\varepsilon^{-2}$ and coefficients $\lambda_a$ such that:*

*(i)* $p = Z_0 + \sum_{a=1}^{m} \lambda_a \,\mathrm{He}_2(Z_a)$.

*(ii)* $Z_0, Z_1, \dots, Z_m$ *are independent and* $W_2(Z_0, \mathcal{N}(0, s)) \le C\varepsilon$ *for an absolute constant* $C$ *where* $s = \mathbb{E}\,Z_0^2$.

*Proof.* Write $p$ in the Hermite basis as $p(X) = \sum_{i \neq j} P_{ij} X_i X_j + \sum_{i=1}^n P_{ii} \operatorname{He}_2(X_i)$ for a symmetric coefficient matrix $P \in \mathbb{R}^{n \times n}_{\text{sym}}$. Let $\sum_a \lambda_a v_a v_a^\intercal$ be the spectral decomposition of $P$. Then

$$p(X) = X^\intercal P X - \operatorname{tr}(P) = \sum_a \lambda_a \langle v_a, X \rangle^2 - \operatorname{tr}(P) = \sum_a \lambda_a (\langle v_a, X \rangle^2 - 1) = \sum_a \lambda_a \operatorname{He}_2(\langle v_a, X \rangle) \,. \tag{3.1}$$

By the orthogonality of the eigendecomposition, $\langle v_a, X \rangle$ are independent standard Gaussians.

We have $\frac{1}{2} \mathbb{E}\, p^2 = \sum_{i,j=1}^n P_{ij}^2 =: \|P\|_F^2$. On the other hand, $\|P\|_F^2 = \sum_a \lambda_a^2$. Since $\mathbb{E}\, p^2 \leq 1$, we deduce that $\sum_a \lambda_a^2 \leq \frac{1}{2}$.

Let $A = \{a : |\lambda_a| \geq \varepsilon/2\}$. By the above, $|A| \leq 2\varepsilon^{-2}$. Write

$$p = \sum_{a \in A} \lambda_a \operatorname{He}_2(\underbrace{\langle v_a, X \rangle}_{=: Z_a}) + \underbrace{\sum_{a \notin A} \lambda_a \operatorname{He}_2(\langle v_a, X \rangle)}_{=: Z_0} \,.$$

Since the $\operatorname{He}_2(\langle v_a, X \rangle)$ are independent and mean-zero, a central limit theorem implies $Z_0$ is approximately Gaussian if there are many summands. If there are few summands, $Z_0$ is small since $|\lambda_a| \leq \varepsilon/2$. Both cases are handled simultaneously by the CLT from [Rio09, Theorem 4.1] which yields for an absolute constant $C$, with $s = \mathbb{E}\, Z_0^2 = 2 \sum_{a \notin A} \lambda_a^2$,

$$W_2(Z_0, \mathcal{N}(0, s)) \leq \frac{C}{\sqrt{s}} \left( \sum_{a \notin A} \lambda_a^4 \right)^{1/2} \leq C \max_{a \notin A} |\lambda_a| \leq C\varepsilon \,. \qquad \square$$

### 3.2 Preliminaries on ultraproducts

A collection $\mathcal{U}$ of subsets of $\mathbb{N}$ is a *non-principal ultrafilter* if:

(i) $\mathcal{U}$ contains $\mathbb{N}$ but not $\emptyset$.

(ii) If $A \in \mathcal{U}$ and $A \subseteq B$, then $B \in \mathcal{U}$.

(iii) If $A, B \in \mathcal{U}$ then $A \cap B \in \mathcal{U}$

(iv) If $A \subseteq \mathbb{N}$ then exactly one of $A$ and $\mathbb{N} \setminus A$ is in $\mathcal{U}$.

(v) There are no finite sets in $\mathcal{U}$.

We say that $A \in \mathcal{U}$ is a *$\mathcal{U}$-large* set of natural numbers.

Fix any[1] non-principal ultrafilter $\mathcal{U}$ on $\mathbb{N}$. We can use $\mathcal{U}$ to "pick" the limit of a sequence of objects inside a compact Hausdorff space, as summarized in the next lemma.

**Lemma 3.2** ([Wil70, Theorems 13.7 and 17.4])**.**

*(i) (Existence and uniqueness of limits) Let $S$ be a compact Hausdorff space. For all sequences $x_n \in S$ there exists a unique $x \in S$ such that, for all open neighborhoods $V$ of $x$, we have $\{n : x_n \in V\} \in \mathcal{U}$. Write $x = \lim_{n \to \mathcal{U}} x_n$ or $x_n \xrightarrow[n \to \mathcal{U}]{} x$.*

[1] A non-principal ultrafilter $\mathcal{U}$ can be easily constructed using Zorn's lemma, which uses the axiom of choice (ZFC). We consider it all but certain that Theorems 3.8 and 3.9 can also be proven without choice.

*(ii) (Continuity of limits) Let $S, T$ be Hausdorff spaces and $f : S \to T$ continuous. If $x_n \underset{n\to\mathcal{U}}{\longrightarrow} x$ in $S$ then $f(x_n) \underset{n\to\mathcal{U}}{\longrightarrow} f(x)$ in $T$.*

*(iii) (Extraction of a subsequence) Let $x_n \underset{n\to\mathcal{U}}{\longrightarrow} x$ in a first-countable Hausdorff space $S$ (for example, a metrizable space). Then there exists an infinite subsequence such that $x_n \to x$ in $S$ along the subsequence. Moreover, the subsequence may be chosen inside any $A \in \mathcal{U}$.*

While Lemma 3.2 applies to compact spaces, sequences inside a non-compact space can still be studied using *ultraproducts*. We are specifically interested in sequences and limits of Gaussian polynomials. Define

$$B_d = \{(p_n)_{n\in\mathbb{N}} : p_n \in \mathcal{H}_{\le d}(X_1, \dots, X_n), \sup_n \mathbb{E}\, p_n^2 < \infty\}\,, \qquad N_d = \{(p_n)_{n\in\mathbb{N}} \in B_d : \lim_{n\to\mathcal{U}} \mathbb{E}\, p_n^2 = 0\}\,.$$

The scalar ultralimit in the definition of $N_d$ exists because $\mathbb{E}\, p_n^2$ is bounded and so $(\mathbb{E}\, p_n^2)_{n\in\mathbb{N}}$ lies in a compact subset of $\mathbb{R}$. By the triangle inequality, $N_d$ is a linear subspace of $B_d$. The *Hilbert ultraproduct* is the vector space

$$\mathscr{H}_{\le d} = B_d / N_d\,.$$

We write $[p_n]$ for elements of $\mathscr{H}_{\le d}$ which formally are equivalence classes of the form $[p_n] = \{(p_n + r_n)_{n\in\mathbb{N}} : (r_n)_{n\in\mathbb{N}} \in N_d\}$. A crucial property of $\mathscr{H}_{\le d}$ is that it inherits the Hilbert space structure from $\mathcal{H}_{\le d}(X)$.

**Lemma 3.3.** *$\mathscr{H}_{\le d}$ is a Hilbert space with inner product $\langle [p_n], [q_n]\rangle = \lim_{n\to\mathcal{U}} \mathbb{E}\, p_n q_n$.*

*Proof.* It is easy to check that $\langle [p_n], [q_n]\rangle$ defines a valid inner product on $\mathscr{H}_{\le d} = B_d/N_d$. What remains is to check that the space is complete.

Consider the space $B_d$ with the supremum norm $\|(p_n)_{n\in\mathbb{N}}\|_\infty = \sup_n \|p_n\|_2$. Since each coordinate space $\mathcal{H}_{\le d}(X_1, \dots, X_n)$ is complete, $B_d$ is complete in this norm ($B_d$ is a Banach space). Moreover, $(p_n)_{n\in\mathbb{N}} \mapsto \lim_{n\to\mathcal{U}} \|p_n\|_2$ is 1-Lipschitz in this norm, so the zero set $N_d$ is closed in this norm. By [Rud91, Theorem 1.41(d)], $B_d/N_d$ is complete in the quotient norm $\|[p_n]\|_{B_d/N_d} := \inf_{(r_n)_{n\in\mathbb{N}}\in N_d} \|(p_n + r_n)_{n\in\mathbb{N}}\|_\infty$.

It remains to show that the quotient norm agrees with the inner-product norm $\sqrt{\langle [p_n], [p_n]\rangle} = \lim_{n\to\mathcal{U}} \|p_n\|_2$ which will prove that $\mathscr{H}_{\le d}$ is complete in the inner-product norm. Fix $[p_n]$ and set $L = \lim_{n\to\mathcal{U}} \|p_n\|_2$. Every representative $(q_n)_{n\in\mathbb{N}} \in B_d$ of $[p_n]$ satisfies $\lim_{n\to\mathcal{U}} \|q_n\|_2 = L$. Hence $\sup_n \|q_n\|_2 \ge L$ by Lemma 3.2(iii) and therefore the quotient norm is at least $L$.

Conversely, for every $\varepsilon > 0$, the set $A = \{n \in \mathbb{N} : \|p_n\|_2 < L + \varepsilon\}$ belongs to $\mathcal{U}$ by Lemma 3.2(i). Define $q_n = p_n$ for $n \in A$ and $q_n = 0$ otherwise. Then $(p_n - q_n) \in N_d$ and $\sup_n \|q_n\|_2 \le L + \varepsilon$. Therefore the quotient norm is at most $L + \varepsilon$. Letting $\varepsilon \downarrow 0$ shows that it equals $L$, completing the proof. □

Let $\mathscr{H}_d = \{[p_n] : \forall n \in \mathbb{N}.\ p_n \in \mathcal{H}_d(X)\}$ be the degree-$d$ component. Since $\|\Pi_k p\|_2 \le \|p\|_2$ then $N_d$ vanishes under $\Pi_k$ and $\Pi_k$ is well-defined on $\mathscr{H}_{\le d}$. Hence, we maintain the decomposition $\mathscr{H}_{\le d} = \oplus_{j=0}^d \mathscr{H}_j$.

Two remarks on the purpose of introducing the ultraproduct:

1. A technical advantage of the ultralimit is that it consistently chooses a limit inside all compact spaces simultaneously. If $x_n$ are in a compact Hausdorff space (e.g. bounded real numbers)

then we can treat $\lim_{n\to\mathcal{U}} x_n$ like a normal limit and delay applying compactness, passing to subsequences, etc. until we are ready to apply Lemma 3.2(iii) to "exit" the ultraproduct.

2. For sequences in a non-compact space, ontologically, we can think of $[p_n] \in \mathscr{H}_{\le d}$ as a new type of object representing the limit of the sequence $p_n$. These objects inherit all of the usual operations of the original space, for example, given $[p_n], [q_n] \in \mathscr{H}_{\le d}$ we can select representatives $p_n, q_n$ and form $[p_n + q_n] \in \mathscr{H}_{\le d}$. Classifying the limits of Gaussian polynomials reduces to classifying the elements of $\mathscr{H}_{\le d}$ (starting with Lemma 3.3 which shows they form a Hilbert space).

   In the asymptotic Hilbert space, asymptotic approximations become exact relations. By Lemma 3.3,

$$\|x_n - y_n\|_2 \underset{n\to\mathcal{U}}{\longrightarrow} 0 \text{ if and only if } [x_n] = [y_n] \text{ in } \mathscr{H}_{\le d}$$
$$\langle x_n, y_n\rangle \underset{n\to\mathcal{U}}{\longrightarrow} 0 \text{ if and only if } [x_n] \perp [y_n] \text{ in } \mathscr{H}_{\le d}\,.$$

### 3.3 Orthogonal decomposition of the asymptotic Hilbert space

Since $\mathscr{H}_d$ is a Hilbert space, by Zorn's lemma it has an orthonormal basis, which is uncountable. The goal of this section is to construct an orthonormal basis of "graded Hermite polynomials" for $\mathscr{H}_d$ which will solve, in a way, the problem of describing $\mathscr{H}_d$.

For $p \in \mathcal{H}_r(X)$ and $q \in \mathcal{H}_s(X)$, define the *top product* (or *Wick product*)

$$p \diamond q = \Pi_{r+s}(pq).$$

It can be seen that $\diamond$ is commutative and associative. We can also lift the top product to $\mathscr{H}_d$.

**Lemma 3.4.** *$\diamond$ is a well-defined and continuous bilinear map $\mathscr{H}_r \times \mathscr{H}_s \to \mathscr{H}_{r+s}$.*

*Proof.* By contractivity of projection, Cauchy–Schwarz, and hypercontractivity (Lemma A.3),

$$\|p \diamond q\|_2 \le \|pq\|_2 \le \|p\|_4 \|q\|_4 \le 3^{(r+s)/2}\|p\|_2\|q\|_2.$$

Consequently, if $p_n \in \mathcal{H}_r(X)$ and $q_n \in \mathcal{H}_s(X)$ satisfy $\sup_n \mathbb{E}\, p_n^2 < \infty$ and $\sup_n \mathbb{E}\, q_n^2 < \infty$, then $\sup_n \mathbb{E}(p_n \diamond q_n)^2 < \infty$ and hence $(p_n \diamond q_n)_{n\in\mathbb{N}} \in B_{r+s}$. Lift $\diamond$ to $\mathscr{H}_r \times \mathscr{H}_s$ by $[p_n] \diamond [q_n] := [p_n \diamond q_n]$. To check that $\diamond$ is independent of the representatives, suppose $[p_n] = [p'_n]$ and $[q_n] = [q'_n]$. By bilinearity and the preceding bound,

$$\|p_n \diamond q_n - p'_n \diamond q'_n\|_2 \le 3^{(r+s)/2}\big(\|p_n - p'_n\|_2\|q_n\|_2 + \|p'_n\|_2\|q_n - q'_n\|_2\big).$$

The right-hand side tends to zero along $\mathcal{U}$, since $\|q_n\|_2, \|p'_n\|_2$ are bounded and the difference terms tend to zero along $\mathcal{U}$. Thus the product is well-defined.

Finally, taking ultralimits in the first inequality gives

$$\|u \diamond v\|_{\mathscr{H}_{r+s}} \le 3^{(r+s)/2}\|u\|_{\mathscr{H}_r}\|v\|_{\mathscr{H}_s}.$$

Hence $\diamond$ is a bounded bilinear map and therefore jointly continuous. □

Let $\mathscr{D}_d$ be the subspace of $\mathscr{H}_d$ generated by lower-degree polynomials,

$$\mathscr{D}_d = \overline{\operatorname{span}}\{u \diamond v : u \in \mathscr{H}_r,\ v \in \mathscr{H}_{d-r},\ 0 < r < d\}\,. \tag{3.2}$$

Since $\mathscr{H}_d$ is a Hilbert space, we can form the orthogonal decomposition $\mathscr{H}_d = \mathscr{D}_d \oplus \mathscr{P}_d$ where $\mathscr{P}_d$ is the orthogonal complement of $\mathscr{D}_d$. We call the elements of $\mathscr{P}_d$ *primitive*.

Select an arbitrary orthonormal basis $\{[e_{n,i}] : i \in I_d\}$ for $\mathscr{P}_d$ and let $I = \sqcup_{d\in\mathbb{N}} I_d$. We also normalize $\mathbb{E}\, e_{n,i}^2 = 1$ exactly. For finitely-supported $\alpha \in \mathbb{N}^I$ define $\mathrm{gr}(\alpha) = \sum_{d\in\mathbb{N}} \sum_{i\in I_d} d\alpha_i$. By the definition of $\mathscr{D}_d$ and $\mathscr{P}_d$ there is a dense spanning set for $\mathscr{H}_d$ given by

$$\left\{ \frac{1}{\sqrt{\alpha!}} \bigdiamond_{i\in I} [e_{n,i}^{\diamond\alpha_i}] : \alpha \in \mathbb{N}^I, \mathrm{gr}(\alpha) = d \right\}.$$

We will prove that all primitive elements are asymptotically Gaussian, and these top products are asymptotically orthonormal Hermite polynomials with respect to the primitive Gaussians. This will characterize the basis for $\mathscr{H}_d$ as the collection of graded Hermite polynomials in the primitive elements.

**Theorem 3.5** (Regularity lemma, ultrapower version)**.** *$\{[\psi_\alpha(e_n)] : \alpha \in \mathbb{N}^I, \mathrm{gr}(\alpha) = d\}$ is an orthonormal basis for $\mathscr{H}_d$. Furthermore, let $(Z_i)_{i\in I}$ be independent standard Gaussian variables such that $I_d$ is assigned grade $d$. Let $d_k \in \mathbb{N}$ for $k \in \mathbb{N}$. For any $[p_{n,k}] \in \mathscr{H}_{d_k}$ write the basis representation of $[p_{n,k}]$ and define $p_k \in \mathcal{H}_{d_k}^{\mathrm{gr}}(Z)$ by*

$$[p_{n,k}] = \sum_{\alpha\in\mathbb{N}^I : \mathrm{gr}(\alpha)=d_k} c_{k,\alpha}[\psi_\alpha(e_n)]\,, \qquad p_k = \sum_{\alpha\in\mathbb{N}^I : \mathrm{gr}(\alpha)=d_k} c_{k,\alpha}\psi_\alpha(Z)\,.$$

*Then $(p_{n,k})_{k\in\mathbb{N}} \xrightarrow[n\to\mathcal{U}]{d} (p_k)_{k\in\mathbb{N}}$.*

Note that the target $(p_k)_{k\in\mathbb{N}}$ of the convergence in distribution is independent of the choice of representatives $p_{n,k}$ for $[p_{n,k}]$ by Slutsky's lemma applied to the difference element from $N_d$.

### 3.3.1 Proof of Theorem 3.5

We use combinatorial arguments on the tensor representations of the polynomials. For an ordered tuple $\mathbf{i} = (i_1, \ldots, i_d)$, let $\alpha(\mathbf{i})_j = |\{s : i_s = j\}|$ and set $\mathrm{He}_{i_1\cdots i_d}(X) := \mathrm{He}_{\alpha(\mathbf{i})}(X)$. For $p \in \mathcal{H}_d(X_1, \ldots, X_n)$, the *Hermite coefficient tensor* $P \in (\mathbb{R}^n)^{\otimes d}_{\mathrm{sym}}$ is the symmetric order-$d$ tensor with dimension $n$ such that,

$$p(X) = \frac{1}{\sqrt{d!}} \sum_{i_1,\ldots,i_d=1}^{n} P_{i_1\cdots i_d} \mathrm{He}_{i_1\cdots i_d}(X)\,. \tag{3.3}$$

In Nualart and Peccati's notation, $p$ is realized as a $d$-dimensional multiple Wiener–Itô integral, $p = \frac{I_d(P)}{\sqrt{d!}}$ [NP05]. The entry normalization creates an important isometry between $\mathcal{H}_d(X_1, \ldots, X_n)$ and $(\mathbb{R}^n)^{\otimes d}_{\mathrm{sym}}$:

$$\mathbb{E}\, pq = \langle P, Q\rangle_F =: \sum_{i_1,\ldots,i_d} P_{i_1\cdots i_d} Q_{i_1\cdots i_d}, \qquad \mathbb{E}\, p^2 = \|P\|_F^2 := \sum_{i_1,\ldots,i_d} P_{i_1\cdots i_d}^2\,. \tag{3.4}$$

We call $P \in (\mathbb{R}^n)^{\otimes d}$ a non-symmetric Hermite coefficient tensor for $p$ if it satisfies Eq. (3.3). If at least one tensor $P, Q$ is symmetric in (3.4), the equality continues to hold since the other tensor can be symmetrized without loss of generality.

For $u \in \mathcal{H}_r(X)$ and $v \in \mathcal{H}_s(X)$ with Hermite coefficient tensors $U, V$ we observe that there is a non-symmetric Hermite coefficient tensor $W$ for $u \diamond v \in \mathcal{H}_{r+s}(X)$ given by

$$W = c_{r,s} U \otimes V, \qquad c_{r,s} = \sqrt{\frac{(r+s)!}{r!s!}}\,. \tag{3.5}$$

The factor $c_{r,s}$ accounts for normalization in (3.3).

Given a tensor $P$ and a partition of the indices $[d]$ into two ordered sets $(L, R)$, the *flattening* $P^{(L,R)} \in \mathbb{R}^{n^{|L|} \times n^{|R|}}$ is the matrix with entries $P^{(L,R)}_{i_L i_R} = P_{i_1 \cdots i_d}$ where $i_L = (i_\ell)_{\ell \in L}$ and $i_R = (i_r)_{r \in R}$. An order-$d$ symmetric tensor has only $d+1$ distinct flattenings. We write $P^{(k)} \in \mathbb{R}^{n^k \times n^{d-k}}$ for the $k$th flattening. vec denotes flattening the tensor into a column vector.

**Lemma 3.6.** *Let $[p_n] \in \mathscr{H}_d$. Let $P_n$ be the Hermite coefficient tensor of $p_n$. Then*

$$[p_n] \in \mathscr{P}_d \quad \iff \quad \lim_{n \to \mathcal{U}} \|P_n^{(r)}\| = 0 \quad \textit{for every } 0 < r < d.$$

*Proof.* Let $[p_n] \in \mathscr{H}_d(X), [u_n] \in \mathscr{H}_r(X), [v_n] \in \mathscr{H}_{d-r}(X)$ have Hermite coefficient tensors $P_n, U_n, V_n$. By the tensor isometry (3.4) and (3.5),

$$\begin{aligned} \mathbb{E}\, p_n(u_n \diamond v_n) &= c_{r,d-r} \langle P_n, U_n \otimes V_n \rangle = c_{r,d-r} \mathrm{vec}(U_n)^\intercal P_n^{(r)} \mathrm{vec}(V_n) \\ |\mathbb{E}\, p_n(u_n \diamond v_n)| &\le c_{r,d-r} \|P_n^{(r)}\| \|U_n\|_F \|V_n\|_F = c_{r,d-r} \|P_n^{(r)}\| \|u_n\|_2 \|v_n\|_2 \,. \end{aligned}$$

Suppose that $\lim_{n \to \mathcal{U}} \|P_n^{(r)}\| = 0$. Taking ultralimits and using that $\|u_n\|_2, \|v_n\|_2$ are bounded, then $\langle [p_n], [u_n \diamond v_n] \rangle_{\mathscr{H}_d} = \lim_{n \to \mathcal{U}} \mathbb{E}\, p_n(u_n \diamond v_n) = 0$. Therefore, $[p_n] \perp \mathscr{D}_d$ in this case.

Conversely, suppose $[p_n] \in \mathscr{P}_d$ and $\lim_{n \to \mathcal{U}} \|P_n^{(r)}\| > 0$. Choose unit-norm singular vectors $\tilde{U}_n, \tilde{V}_n$ for each $n$ such that

$$(\tilde{U}_n)^\intercal P_n^{(r)} \tilde{V}_n = \|P_n^{(r)}\| \,.$$

Unflatten $\tilde{U}_n, \tilde{V}_n$ into tensors $U_n, V_n$. Note that since $P_n$ is a symmetric tensor, its nonzero singular vectors also correspond to symmetric tensors. Then $U_n, V_n$ represent unit-norm $u_n \in \mathcal{H}_r(X)$, $v_n \in \mathcal{H}_{d-r}(X)$ with,

$$\mathbb{E}\, p_n(u_n \diamond v_n) = c_{r,d-r} \langle P_n, U_n \otimes V_n \rangle = c_{r,d-r} (\tilde{U}_n)^\intercal P_n^{(r)} \tilde{V}_n = c_{r,d-r} \|P_n^{(r)}\| \,.$$

Taking the ultralimit shows $\langle [p_n], [u_n] \diamond [v_n] \rangle > 0$ contrary to $[p_n] \perp \mathscr{D}_d$. □

**Lemma 3.7.** *For all finitely-supported $\alpha \in \mathbb{N}^I$,*

$$\frac{1}{\sqrt{\alpha!}} \bigdiamond_{i \in I} [e_{n,i}^{\diamond \alpha_i}] = [\psi_\alpha(e_n)] \,.$$

*Proof.* Let $h_{n,\alpha} = \frac{1}{\sqrt{\alpha!}} \bigdiamond_{i \in I} e_{n,i}^{\diamond \alpha_i}$. We have,

$$\mathbb{E}(\psi_\alpha(e_n) - h_{n,\alpha})^2 = \mathbb{E}\, \psi_\alpha(e_n)^2 - 2\, \mathbb{E}\, \psi_\alpha(e_n) h_{n,\alpha} + \mathbb{E}\, h_{n,\alpha}^2 = \mathbb{E}\, \psi_\alpha(e_n)^2 - \mathbb{E}\, h_{n,\alpha}^2 \,. \tag{3.6}$$

The last equality holds because $h_{n,\alpha} \in \mathcal{H}_{\mathrm{gr}(\alpha)}(X)$ exactly, so it is orthogonal to lower-degree terms in $\psi_\alpha(e_n)$, and the unique leading monomial of $\psi_\alpha(e_n)$ is the same as that of $h_{n,\alpha}$.

By Lemma 3.6, the norm of every proper flattening of every $e_{n,i}$ vanishes. By the fourth moment theorem (Theorem B.1), the $e_{n,i}$ supported in $\alpha$ are asymptotically jointly Gaussian (note that the proof of Theorem B.1 still holds when all convergence is along $\mathcal{U}$). Then we have the convergence of the variance $\mathbb{E}\, \psi_\alpha(e_n)^2 \underset{n \to \mathcal{U}}{\longrightarrow} 1$ by Lemma A.4. We have $\mathbb{E}\, h_{n,\alpha}^2 \underset{n \to \mathcal{U}}{\longrightarrow} 1$ by the combinatorial moment calculation in Lemma B.6. Therefore, Eq. (3.6) goes to 0 along $\mathcal{U}$ which proves the claim. □

*Proof of Theorem 3.5.* We observed above that a spanning set for $\mathscr{H}_d$ is

$$\left\{ \frac{1}{\sqrt{\alpha!}} \bigotimes_{i\in I}^{\diamond} [e_{n,i}^{\diamond \alpha_i}] : \alpha \in \mathbb{N}^I, \mathrm{gr}(\alpha) = d \right\}.$$

By Lemma 3.7, these equal $\{[\psi_\alpha(e_n)] : \alpha \in \mathbb{N}^I, \mathrm{gr}(\alpha) = d\}$. The Gaussian convergence in Theorem B.1 implies the $[\psi_\alpha(e_n)]$ are orthonormal in $\mathscr{H}_d$ like the usual Hermite basis.

For the second part of the theorem, if the basis representation of $[p_{n,k}]$ is finite, then convergence in distribution follows from Theorem B.1 and the continuous mapping theorem. It remains to handle the case where the basis representation of $[p_{n,k}]$ has countably many nonzero terms. The expression for $p_k$ has $\mathbb{E}\, p_k^2 < \infty$ (hence it is well-defined in $L_2(Z)$) because, by orthonormality in $\mathcal{H}_{d_k}^{\mathrm{gr}}(Z)$ and $\mathscr{H}_{d_k}$:

$$\mathbb{E}\, p_k^2 = \sum_{\alpha\in\mathbb{N}^I : \mathrm{gr}(\alpha)=d_k} c_{k,\alpha}^2 = \lim_{n\to\mathcal{U}} \mathbb{E}\, p_{n,k}^2 < \infty\,. \tag{3.7}$$

Fix $\varepsilon > 0$ and a finite set $F_k$ such that $\sum_{\alpha\notin F_k} c_{k,\alpha}^2 \le \varepsilon$ and define the truncated approximations,

$$[q_{n,k}] = \sum_{\alpha\in F_k} c_{k,\alpha}[\psi_\alpha(e_n)], \qquad q_k = \sum_{\alpha\in F_k} c_{k,\alpha}\psi_\alpha(Z)\,.$$

We have simultaneously:

$$\mathbb{E}(p_k - q_k)^2 = \sum_{\alpha\notin F_k} c_{k,\alpha}^2 \le \varepsilon \qquad\qquad \lim_{n\to\mathcal{U}} \mathbb{E}(p_{n,k} - q_{n,k})^2 = \sum_{\alpha\notin F_k} c_{k,\alpha}^2 \le \varepsilon \tag{3.8}$$

Furthermore, since $F_k$ is finite, we have $(q_{n,1},\dots,q_{n,k}) \xrightarrow[n\to\mathcal{U}]{d} (q_1,\dots,q_k)$ by the finite case.

To prove $(p_{n,k})_{k\in\mathbb{N}} \xrightarrow[n\to\mathcal{U}]{d} (p_k)_{k\in\mathbb{N}}$ it is equivalent to prove $(p_{n,1},\dots,p_{n,k}) \xrightarrow[n\to\mathcal{U}]{d} (p_1,\dots,p_k)$ for each $k$. Fix a bounded 1-Lipschitz test function $\varphi : \mathbb{R}^k \to \mathbb{R}$. Let $P_n = (p_{n,1},\dots,p_{n,k}), P = (p_1,\dots,p_k), Q_n = (q_{n,1},\dots,q_{n,k}), Q = (q_1,\dots,q_k)$. Non-asymptotically,

$$\begin{aligned}
&|\,\mathbb{E}\,\varphi(P_n) - \mathbb{E}\,\varphi(P)| \\
&\le |\,\mathbb{E}\,\varphi(P_n) - \mathbb{E}\,\varphi(Q_n)| + |\,\mathbb{E}\,\varphi(Q_n) - \mathbb{E}\,\varphi(Q)| + |\,\mathbb{E}\,\varphi(Q) - \mathbb{E}\,\varphi(P)| \\
&\le \left(\sum_{i=1}^k \mathbb{E}(p_{n,i} - q_{n,i})^2\right)^{1/2} + |\,\mathbb{E}\,\varphi(Q_n) - \mathbb{E}\,\varphi(Q)| + \left(\sum_{i=1}^k \mathbb{E}(p_i - q_i)^2\right)^{1/2} \qquad \text{(1-Lipschitz)}
\end{aligned}$$

The right term is at most $\sqrt{k\varepsilon}$ by Eq. (3.8). In the limit $n \to \mathcal{U}$, the left term is at most $\sqrt{k\varepsilon}$ and the middle term converges to 0 since $Q_n \xrightarrow[n\to\mathcal{U}]{d} Q$. We conclude, for every $\varepsilon > 0$,

$$\lim_{n\to\mathcal{U}} |\,\mathbb{E}\,\varphi(P_n) - \mathbb{E}\,\varphi(P)| \le 2\sqrt{k\varepsilon}\,.$$

Take $\varepsilon \downarrow 0$ to get $\lim_{n\to\mathcal{U}} |\,\mathbb{E}\,\varphi(P_n) - \mathbb{E}\,\varphi(P)| = 0$. Bounded 1-Lipschitz functions determine weak convergence by the Portmanteau theorem, so $P_n \xrightarrow[n\to\mathcal{U}]{d} P$. □

### 3.4 Proof of regularity lemmas

Let $p \in \mathcal{H}_d(X), \mathbb{E}\, p^2 \le 1$ with Hermite coefficient tensor $P$. The idea of the regularity lemma is, if a flattening $P^{(r)}$ has a large singular value $\sigma$, then write

$$\begin{aligned} P^{(r)} &= \sigma U^{(r)}(V^{(d-r)})^\intercal + P^{(r)}_{\mathrm{rem}}, && (U^{(r)} \in \mathbb{R}^{n^r}, V^{(d-r)} \in \mathbb{R}^{n^{d-r}}) \\ p &= \frac{\sigma}{c_{r,d-r}} u \diamond v + p_{\mathrm{rem}}, && (u \in \mathcal{H}_{\le r}(X), v \in \mathcal{H}_{\le d-r}(X)) \,. \end{aligned}$$

$U^{(r)}$ and $V^{(d-r)}$ are the (flattenings of the) Hermite coefficient tensors for lower-degree polynomials $u$ and $v$. Since $u$ and $v$ are lower degree, they can be decomposed by induction.

On the other hand, removing the singular direction decreases the squared norm,[2]

$$\mathbb{E}\, p_{\mathrm{rem}}^2 \le \|P^{(r)}_{\mathrm{rem}}\|_F^2 = \|P^{(r)}\|_F^2 - \sigma^2 = \|P\|_F^2 - \sigma^2 = \mathbb{E}\, p^2 - \sigma^2 \,.$$

Since initially $\mathbb{E}\, p^2 \le 1$, after removing at most $1/\varepsilon^2$ singular directions, all proper flattenings have spectral norm at most $\varepsilon$. Once all proper flattenings are small, by Theorem B.1, the remaining polynomial is Gaussian up to error depending on $\varepsilon$.

The ultrapower is essentially carrying out this approach implicitly, without having to worry about the recursion or the quantitative parameters. The energy measure $\mathbb{E}\, p^2 = \|P\|_F^2$ and the recursive decomposition are replaced by orthogonal decomposition in the Hilbert space $\mathscr{H}_d$. We now use this approach to prove the regularity lemmas.

**Theorem 3.8** (Regularity lemma, compactification version)**.** *Let $Z$ be a set of graded variables with countably many variables of each grade. Let $d_k \in \mathbb{N}$ for $k \in \mathbb{N}$. Let $p_{n,k} \in \mathcal{H}^{\mathrm{gr}}_{d_k}(Z)$ such that $\sup_n \mathbb{E}\, p_{n,k}^2 < \infty$ for each $k \in \mathbb{N}$. Then there exist $p_k \in \mathcal{H}^{\mathrm{gr}}_{d_k}(Z)$ and a subsequence of $n$ such that $(p_{n,k})_{k\in\mathbb{N}} \xrightarrow{d} (p_k)_{k\in\mathbb{N}}$ along the subsequence. Consequently, $\{\mathrm{Law}(p_1, \dots, p_k) : p_i \in \mathcal{H}^{\mathrm{gr}}_{d_i}(Z), \mathbb{E}\, p_i^2 \le 1\}$ is weakly compact for each $k$.*

*Proof.* First we show that, without loss of generality, we may assume $p_{n,k}$ are ordinary Gaussian polynomials (all variables have grade one). By Lemma 2.1 applied to $(p_{n,1}, \dots, p_{n,n})$, also using that bounded 1-Lipschitz test functions determine convergence in distribution by the Portmanteau theorem, there exist ordinary polynomials $q_{n,k}$ such that, for all $k \le n$,

$$|\mathbb{E}\, \varphi(p_{n,1}, \dots, p_{n,k}) - \mathbb{E}\, \varphi(q_{n,1}, \dots, q_{n,k})| \le \frac{1}{n}$$

for all bounded 1-Lipschitz functions $\varphi : \mathbb{R}^k \to \mathbb{R}$. So, if a subsequence of $q_{n,k}$ converges in distribution, then the same subsequence of $p_{n,k}$ converges in distribution to the same limit. By Lemma A.4, the approximants $q_{n,i}$ also satisfy $\sup_n \mathbb{E}\, q_{n,i}^2 < \infty$. Hence, it suffices to assume $p_{n,k} \in \mathcal{H}_{d_k}(X_1, \dots, X_n)$ are Gaussian polynomials such that $\sup_n \mathbb{E}\, p_{n,k}^2 < \infty$.

Applying Theorem 3.5, there exist graded Gaussian polynomials $p_k \in \mathcal{H}^{\mathrm{gr}}_{d_k}(Z)$ such that

$$(p_{n,k})_{k\in\mathbb{N}} \xrightarrow[n\to\mathcal{U}]{d} (p_k)_{k\in\mathbb{N}} \tag{3.9}$$

Each $p_k$ is a countable or finite sum of Hermite polynomials in the $Z$ variables, so the entire family $(p_k)_{k\in\mathbb{N}}$ can be represented using at most countably many graded Gaussian variables.

[2]The first inequality is not an equality because the Hermite coefficient tensor of $p_{\mathrm{rem}}$ is not $P_{\mathrm{rem}}$ but the symmetrization of $P_{\mathrm{rem}}$. Symmetrization cannot increase the Frobenius norm.

The weak topology on probability measures on $\mathbb{R}^{\mathbb{N}}$ is metrizable (by the sup over bounded 1-Lipschitz functions weighted by $2^{-i}$ in coordinate $i$), so we may use Lemma 3.2(iii) to extract a subsequence such that Eq. (3.9) holds ordinarily.

For the final assertion, apply this result to any sequence of admissible $(p_{n,1}, \dots, p_{n,k})$. The Portmanteau theorem gives, along the extracted subsequence,

$$\mathbb{E}\, p_i^2 \le \liminf_{n\to\infty} \mathbb{E}\, p_{n,i}^2 \le 1.$$

Hence $p_i$ is also admissible, and the set of admissible laws is compact. □

**Theorem 3.9** (Regularity lemma, approximation version)**.** *For every $k, D \in \mathbb{N}$, $\varepsilon > 0$ there exists $M = M(k, D, \varepsilon) \in \mathbb{N}$ such that the following holds. Let $d_1, \dots, d_k \le D$ (repeats allowed). Let $p_i \in \mathcal{H}_{d_i}(X)$ such that $\mathbb{E}\, p_i^2 \le 1$ for all $i = 1, \dots, k$. Then there exist $m \le M, g_a \le D$ and $Z_a \in \mathcal{H}_{g_a}(X)$ for $a \in [m]$, and $q_i \in \mathcal{H}_{d_i}^{\mathrm{gr}}(Z_1, \dots, Z_m)$ where $Z_a$ has grade $g_a$, such that:*

*(i)* $\sum_{i=1}^k \mathbb{E}(p_i - q_i)^2 \le \varepsilon$

*(ii)* $W_2((Z_1, \dots, Z_m), \mathcal{N}(0,1)^{\otimes m}) \le \varepsilon$

*Proof.* Suppose the assertion fails for some $k, D, \varepsilon$. Then for every $M$, there is a choice of polynomials which require at least $M$ variables to approximate within error $4\varepsilon$. Construct a sequence $p_{n,i} \in \mathcal{H}_{d_i}(X), \mathbb{E}\, p_{n,i}^2 \le 1$ such that the number of Gaussians required to approximate them within error $4\varepsilon$ grows with $n$. We will use the ultraproduct to approximate an infinite subsequence using a constant number of variables, yielding the contradiction.

Strictly speaking, the $p_{n,i}$ may not be polynomials since $X$ could be infinite. For each one, truncate the Hermite expansion to a finite approximant within $\varepsilon$ error. Then we obtain a sequence of bona fide polynomials $p_{n,i} \in \mathcal{H}_{d_i}(X_1, \dots, X_n), \mathbb{E}\, p_{n,i}^2 \le 1$ such that the number of Gaussians required to approximate them within error $\varepsilon$ grows with $n$. Embed $[p_{n,i}]$ into $\mathscr{H}_{d_i}$.

By Theorem 3.5, write the basis representation

$$[p_{n,i}] = \sum_{\mathrm{gr}(\alpha)=d_i} c_{i,\alpha} [\psi_\alpha(e_n)].$$

Choose finite sets $F_i$ of multi-indices of grade $d_i$ such that $\sum_{i=1}^k \sum_{\alpha \notin F_i} c_{i,\alpha}^2 < \varepsilon/2$. List the finitely many primitive elements $[e_{n,1}], \dots, [e_{n,m}]$ appearing in the union of the $F_i$ and for each one select a representative sequence of polynomials, written $Z_{n,a} \in \mathcal{H}_{g_a}(X_1, \dots, X_n)$.

Define the graded-degree-$d_i$ polynomial $q_i(Z_1, \dots, Z_m) = \sum_{\alpha \in F_i} c_{i,\alpha} \psi_\alpha(Z)$. By asymptotic orthonormality of the $\psi_\alpha(Z)$ in the ultraproduct,

$$\lim_{n\to\mathcal{U}} \sum_{i=1}^k \mathbb{E}\big(p_{n,i} - q_i(Z_{n,1}, \dots, Z_{n,m})\big)^2 = \sum_{i=1}^k \sum_{\substack{\mathrm{gr}(\alpha)=d_i \\ \alpha \notin F_i}} c_{i,\alpha}^2 < \varepsilon/2. \tag{3.10}$$

By Lemma 3.2(i), inequality (3.10) holds with error $\varepsilon$ for all $n$ in a $\mathcal{U}$-large set. Call this set $A$.

The second part of Theorem 3.5 gives $(Z_{n,1}, \dots, Z_{n,m}) \xrightarrow[n\to\mathcal{U}]{d} \mathcal{N}(0,1)^{\otimes m}$. We claim that the following set $B$ is also $\mathcal{U}$-large,

$$B = \{n : W_2((Z_{n,1}, \dots, Z_{n,m}), \mathcal{N}(0,1)^{\otimes m}) < \varepsilon\}.$$

Otherwise, $\mathbb{N} \setminus B \in \mathcal{U}$. By Lemma 3.2(iii) there is a subsequence of $n \in \mathbb{N} \setminus B$ such that $(Z_{n,1}, \dots, Z_{n,m}) \xrightarrow{d} \mathcal{N}(0,1)^{\otimes m}$ along the subsequence. By Lemma A.4, this also implies convergence in $W_2$, contradicting the definition of $B$. Thus $B \in \mathcal{U}$.

Then $A \cap B$ is also $\mathcal{U}$-large. Use Lemma 3.2(iii) to obtain a subsequence of $p_{n,i}$ which are approximated up to error $\varepsilon$ by at most $m$ variables, the desired contradiction. □

*Remark.* The proof of Theorem 3.9 gives a nonconstructive bound on $M(k, D, \varepsilon)$. GPT-6 Astra asserts that the strategy described at the start of this section can prove an Ackermann-type recursion for $M(k, D, \varepsilon)$, for example, $M(k, 2, \varepsilon) \le Ck^4\varepsilon^{-3}(k+1)^{Ck/\varepsilon}$, a tower-type bound for $D = 3$, and larger bounds for $D > 3$. We have not attempted to verify the proof.

## 4 Application: robust Borell inequality

For a parameter $s \in [-1, 1]$, the Ornstein–Uhlenbeck noise operator $T_s$ is defined on $L_2(X)$ by

$$(T_s f)(X) = \mathbb{E}[f(Y) \mid X], \qquad (X, Y) \sim s\text{-correlated Gaussians}.$$

Let $\mathrm{Stab}_s(f) = \langle f, T_s f\rangle = \mathbb{E}_{(X,Y)\sim s\text{-correlated}} f(X)f(Y)$ denote the noise stability of $f$. For a set $A \subseteq \mathbb{R}^n$ we also write $\mathrm{Stab}_s(A) = \mathrm{Stab}_s(1_A)$. A *halfspace* is a set of the form $H = \{X : \sum_i c_i X_i \ge c_0\}$ for some coefficients $c_i$ normalized to $\sum_i c_i^2 = 1$.

**Theorem 4.1** (Robust Borell inequality)**.** *For $a \in (0,1)$ let $H_a = (-\infty, \Phi^{-1}(a)) \subseteq \mathbb{R}$. For all $a \in (0,1), s \in (0,1), \varepsilon > 0$ there exists $\delta = \delta(a, s, \varepsilon) > 0$ such that: if $A \subseteq \mathbb{R}^n, \gamma(A) = a$ satisfies $\mathrm{Stab}_s(A) \ge \mathrm{Stab}_s(H_a) - \delta$ then there exists a halfspace $H \subseteq \mathbb{R}^n, \gamma_n(H) = a$ such that $\gamma_n(A \triangle H) \le \varepsilon$.*

Mossel and Neeman [MN15] originally proved Theorem 4.1, and subsequently Eldan improved the bound on $\delta$ [Eld15]. Here we prove the theorem with a nonquantitative bound on $\delta$, with the advantage that the argument is significantly simpler.

The idea of the proof is to show that all maximizers among the space of limits of sets $A \subseteq \mathbb{R}^n$ are halfspaces. To take the limit of a set $A$, we separately take the limit of each component $\Pi_k 1_A$ using the regularity lemma. There might be "mass escaping to infinity" which disappears in the limit, such as for $A_n = \{X : \prod_{i=1}^n X_i > 0\}$ which has no Hermite weight between degree 0 and $n$. But, this mass isn't seen by noise stability, so we still get $\mathrm{Stab}_s(A_n) \to \mathrm{Stab}_s(f)$ for the limit $f$.

**Lemma 4.2.** *Let $Z$ be a set of graded variables with countably many variables of each grade. Let $a \in [0,1]$. Let $A_n \subseteq \mathbb{R}^n, \gamma_n(A_n) = a$. Then there exists a subsequence of $A_n$ and $f \in L_2(Z)$ such that $0 \le f \le 1, \mathbb{E} f = a$, and for all $s \in (0,1)$, along the subsequence,*

$$(\Pi_k 1_{A_n})_{k\in\mathbb{N}} \xrightarrow{d} (\Pi_k f)_{k\in\mathbb{N}}, \qquad \lim_{n\to\infty} \mathrm{Stab}_s(A_n) = \mathrm{Stab}_s(f).$$

*(Recall that $\Pi_k, T_s, \mathrm{Stab}_s(f)$ are defined on the graded space according to grade, see (2.2).)*

*Proof.* Apply Theorem 3.8 to $(\Pi_k 1_{A_n})_{k\in\mathbb{N}}$. Then there exist graded polynomials $p_k \in \mathcal{H}_k^{\mathrm{gr}}(Z)$ such that $(\Pi_k 1_{A_n})_{k\in\mathbb{N}} \xrightarrow{d} (p_k)_{k\in\mathbb{N}}$. By Lemma A.4, convergence holds for all joint moments. Thus for all $D$,

$$\sum_{k=0}^{D} \|p_k\|_2^2 = \lim_{n\to\infty} \sum_{k=0}^{D} \|\Pi_k 1_{A_n}\|_2^2 \le \lim_{n\to\infty} \mathbb{E}\, 1_{A_n}^2 \le 1.$$

Thus we can define $f = \sum_{k\in\mathbb{N}} p_k$ in $L_2(Z)$ which has the desired components $\Pi_k f = p_k$. The mean of $f$ is, using that non-constant Hermite polynomials are mean-zero,

$$\mathbb{E}\, f = \mathbb{E}\,\Pi_0 f = \lim_{n\to\infty} \mathbb{E}\,\Pi_0 1_{A_n} = \lim_{n\to\infty} \gamma_n(A_n) = a\,.$$

Next, recall that the noise operator acts as $T_s f = \sum_{k\in\mathbb{N}} s^k \Pi_k f$. We have

$$|\mathrm{Stab}_s(A_n) - \mathrm{Stab}_s(f)| \le \sum_{k\in\mathbb{N}} s^k \left| \|\Pi_k 1_{A_n}\|_2^2 - \|\Pi_k f\|_2^2 \right|, \qquad \sum_{k\ge D} s^k \left| \|\Pi_k 1_{A_n}\|_2^2 - \|\Pi_k f\|_2^2 \right| \le 2s^D\,.$$

Therefore, for all $\varepsilon > 0$, by taking $D$ large enough then applying convergence in distribution to the components up to $D$, we get $|\,\mathrm{Stab}_s(A_n) - \mathrm{Stab}_s(f)| \le \varepsilon$ for all sufficiently large $n$. Hence $\lim_{n\to\infty} \mathrm{Stab}_s(A_n) = \mathrm{Stab}_s(f)$.

It remains to show $0 \le f \le 1$. We use the noise operator, proving first $T_s 1_{A_n} \xrightarrow{d} T_s f$. We have,

$$\left\| T_s 1_{A_n} - \sum_{k=0}^{D} s^k \Pi_k 1_{A_n} \right\|_2 \le s^{D+1}, \qquad \left\| T_s f - \sum_{k=0}^{D} s^k \Pi_k f \right\|_2 \le s^{D+1}\,.$$

Then for any bounded 1-Lipschitz function $\phi$,

$$\begin{aligned}
&|\,\mathbb{E}\,\phi(T_s 1_{A_n}) - \mathbb{E}\,\phi(T_s f)| \\
&\le \mathbb{E}\,|T_s 1_{A_n} - \textstyle\sum_{k=0}^{D} s^k \Pi_k 1_{A_n}| + |\,\mathbb{E}\,\phi(\sum_{k=0}^{D} s^k \Pi_k 1_{A_n}) - \mathbb{E}\,\phi(\sum_{k=0}^{D} s^k \Pi_k f)| + \mathbb{E}\,|\textstyle\sum_{k=0}^{D} s^k \Pi_k f - T_s f| \\
&\le 2s^{D+1} + |\,\mathbb{E}\,\phi(\textstyle\sum_{k=0}^{D} s^k \Pi_k 1_{A_n}) - \mathbb{E}\,\phi(\sum_{k=0}^{D} s^k \Pi_k f)|\,.
\end{aligned}$$

For any $\varepsilon > 0$, selecting $D$ large enough then applying joint convergence in distribution to the middle term makes the entire expression at most $\varepsilon$. Hence $T_s 1_{A_n} \xrightarrow{d} T_s f$.

Take $\phi$ to be a bounded continuous test function which is 0 on $[0,1]$ and positive elsewhere. By $T_s 1_{A_n} \xrightarrow{d} T_s f$ and the observation $T_s 1_A \in [0,1]$ for any set $A$,

$$\mathbb{E}\,\phi(T_s f) = \lim_{n\to\infty} \mathbb{E}\,\phi(T_s 1_{A_n}) = 0\,.$$

Since $\phi$ is nonnegative, we must have $T_s f \in [0,1]$ almost surely. Finally, since $T_s f$ converges to $f$ in $L_2$ as $s \uparrow 1$, then $f \in [0,1]$ almost surely. □

**Lemma 4.3** (Characterization of equality case)**.** *Let $a \in (0,1), s \in (0,1)$. Let $Z$ be a set of graded variables with countably many variables of each grade and let $f \in L_2(Z)$ such that $0 \le f \le 1$, $\mathbb{E}\, f = a$, and $\mathrm{Stab}_s(f) = \mathrm{Stab}_s(H_a)$. Up to a set of measure zero, then $f = 1_H$ for some $H = \{X : \sum_i c_i X_i \ge c_0\}$ with $\sum_i c_i^2 = 1$ supported on grade-1 variables.*

*Proof.* The noise stability is $\mathrm{Stab}_s(f) = \sum_{k\in\mathbb{N}} s^k \|\Pi_k f\|_2^2$. Graded variables in $f$ increase the graded-degree, and hence strictly worsen the noise stability: $\sum_{k\in\mathbb{N}} s^k \|\Pi_k f\|_2^2 \le \sum_{k\in\mathbb{N}} s^k \|\tilde{\Pi}_k f\|_2^2$ where $\tilde{\Pi}_k$ treats all variables as grade 1. Therefore maximizers can use only grade 1 variables.

For $0 \le f \le 1$, write $f = \mathbb{E}_\mu\, g$ for a mixture $\mu$ of 0/1-valued $g$. Since $\mathrm{Stab}_s(f)$ is strictly convex, maximizers must be 0/1-valued functions (i.e. sets). Strict convexity of $f \mapsto \mathrm{Stab}_s(f)$ follows from the observation that $\mathrm{Stab}_s(f) = \langle f, T_s f\rangle$ is a positive definite quadratic form for $s \in (0,1]$.

It remains to characterize the maximizers among $A \subseteq \mathbb{R}^X$ with a countable number of variables $X$. For a finite set of variables $X$ and Mossel and Neeman [MN15, Theorem 1.3] characterize the

maximizers as halfspaces $H = \{X : \sum_i c_i X_i \geq c_0\}$. The case of countably many variables follows by notational changes.[3] □

For a set $A \subseteq \mathbb{R}^n$, let $m_A$ denote its barycenter,

$$m_A := \int_{\mathbb{R}^n} x 1_A(x) d\gamma_n(x), \qquad (\Pi_1 1_A)(x) = \langle m_A, x \rangle .$$

The candidate halfspace approximation for $A$ is $H = \{x \in \mathbb{R}^n : \langle \frac{m_A}{\|m_A\|_2}, x \rangle \geq \Phi^{-1}(1-\gamma_n(A))\}$. The following lemma establishes an upper bound on the approximation error.

**Lemma 4.4.** *Let $a \in (0,1)$ and $A \subseteq \mathbb{R}^n, \gamma_n(A) = a$. Set $H = \{x \in \mathbb{R}^n : \langle x, \frac{m_A}{\|m_A\|_2} \rangle \geq \Phi^{-1}(1-a)\}$. Then*

$$\gamma_n(A \triangle H)^2 \leq \frac{4\sqrt{2}}{\sqrt{\pi}} \left( \|m_H\|_2 - \|m_A\|_2 \right) .$$

*Proof.* Let $u = \frac{m_A}{\|m_A\|_2}$ and $t = \Phi^{-1}(1-a)$ so $H = \{x : \langle u, x \rangle \geq t\}$. The unit vector $u$ points in the direction of both $m_A$ and $m_H$, so

$$\int_{\mathbb{R}^n} \langle u, x \rangle 1_A(x) d\gamma_n(x) = \langle u, m_A \rangle = \|m_A\|_2, \qquad \int_{\mathbb{R}^n} \langle u, x \rangle 1_H(x) d\gamma_n(x) = \langle u, m_H \rangle = \|m_H\|_2.$$

Therefore, using that $\int_{\mathbb{R}^n} (1_A(x) - 1_H(x)) d\gamma_n(x) = \gamma_n(A) - \gamma_n(H) = 0$,

$$\|m_H\|_2 - \|m_A\|_2 = \int_{\mathbb{R}^n} \left( \langle u, x \rangle - t \right) \left( 1_H(x) - 1_A(x) \right) d\gamma_n(x) = \int_{A \triangle H} |\langle u, x \rangle - t| d\gamma_n(x) . \tag{4.1}$$

In the rightmost expression, the worst case for $A \triangle H$ is if it is a slab around the plane $\langle u, x \rangle = t$. Therefore, we reduce to a one-dimensional integral,

$$\int_{A \triangle H} |\langle u, x \rangle - t| d\gamma_n(x) \geq \inf_I \int_I |x - t| d\gamma(x) \tag{4.2}$$

where the one-dimensional integral is Gaussian since $\langle u, x \rangle$ is standard Gaussian, and $I \ni t$ is an interval with $\gamma(I) = \gamma_n(A \triangle H) =: \delta$. Since the Gaussian PDF is pointwise upper bounded by $\frac{1}{\sqrt{2\pi}}$, $I$ must contain at least $\frac{\delta}{2}$ mass outside the interval $[t - \frac{\sqrt{\pi}}{2\sqrt{2}}\delta, t + \frac{\sqrt{\pi}}{2\sqrt{2}}\delta]$. All such points have $|x - t| \geq \frac{\sqrt{\pi}}{2\sqrt{2}}\delta$. Therefore,

$$\int_I |x - t| d\gamma(x) \geq \frac{\sqrt{\pi}}{2\sqrt{2}}\delta \cdot \frac{\delta}{2} = \frac{\sqrt{\pi}}{4\sqrt{2}}\delta^2 . \tag{4.3}$$

Combining (4.1), (4.2), (4.3) proves the claim. □

*Proof of Theorem 4.1.* Assume for the sake of contradiction that the assertion fails for some $a, s, \varepsilon$. Then there is a sequence of sets $A_n, \gamma(A_n) = a$ such that $\mathrm{Stab}_s(H_a) - o(1) \leq \mathrm{Stab}_s(A_n) \leq \mathrm{Stab}_s(H_a)$

[3] Except the line "$\nabla v_t(X) = \nabla w_t(Y)$ almost surely implies $\nabla v_t = \nabla w_t$ are almost surely equal to some constant, since the conditional distribution of $Y$ given $X$ is fully supported" ($X, Y$ are a pair of $\rho$-correlated Gaussians). Change this line to: We deduce $\nabla v_t(X) = \nabla w_t(Y) = \nabla v_t(X')$ almost surely where $X, X'$ are a pair of $\rho^2$-correlated Gaussians. It holds generally that, $\mathbb{E} \|\nabla v_t(X) - \nabla v_t(X')\|_2^2 \geq 2(1-\rho^2) \mathbb{E} \|\nabla v_t(X) - \mathbb{E} \nabla v_t(X)\|_2^2$. This implies $\nabla v_t(X) = \mathbb{E} \nabla v_t(X)$ almost surely. Hence $\nabla v_t(X)$ is constant, and $\nabla w_t(Y) = \nabla v_t(X)$ is the same constant.

but $\gamma_n(A_n \triangle H) > \varepsilon$ for all halfspaces $H$ of measure $a$. By Lemma 4.2, there exists $f \in L_2(Z)$ such that, along a subsequence,

$$(\Pi_k 1_{A_n})_{k\in\mathbb{N}} \xrightarrow{d} (\Pi_k f)_{k\in\mathbb{N}}, \qquad 0 \le f \le 1, \qquad \mathbb{E}\, f = a, \qquad \mathrm{Stab}_s(f) = \mathrm{Stab}_s(H_a)\,.$$

By Lemma 4.3, $f$ is a halfspace: $f = 1_H$ for some $H = \{X : \sum_i c_i X_i \ge c_0\}$. By the joint convergence in Lemma 4.2,

$$\lim_{n\to\infty} \|\Pi_1 1_{A_n}\|_2^2 = \|\Pi_1 1_H\|_2^2\,.$$

In the notation of Lemma 4.4, $\|\Pi_1 1_{A_n}\|_2^2 = \|m_{A_n}\|_2^2$ and $\|\Pi_1 1_H\|_2^2 = \|m_H\|_2^2$. Therefore, by Lemma 4.4, there are halfspaces $H_n$ such that $\gamma_n(A_n \triangle H_n) \to 0$, contradicting the assumption. □

# References


[Bil95] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, third edition, 1995. 17

[Bil99] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, second edition, 1999. 17

[Eld15] Ronen Eldan. A two-sided estimate for the gaussian noise stability deficit. *Inventiones mathematicae*, 201(2):561–624, 2015. 13

[ES12] Gábor Elek and Balázs Szegedy. A measure-theoretic approach to the theory of dense hypergraphs. *Advances in Mathematics*, 231(3-4):1731–1772, 2012. 2

[HJM26] Steven Heilman, Chris Jones, and Giulio Malavolta. Primal/dual method for the grothendieck constant. *arXiv preprint arXiv:2610.10477*, 2026. 2

[HMP24] Ronan Herry, Dominique Malicet, and Guillaume Poly. Limit distributions for polynomials with independent and identically distributed entries. *arXiv preprint arXiv:2412.06749*, 2024. 1, 2

[Jan97] Svante Janson. *Gaussian hilbert spaces*. Number 129. Cambridge university press, 1997. 18

[Lov12] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012. 2

[LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006. 2

[MN15] Elchanan Mossel and Joe Neeman. Robust optimality of gaussian noise stability. *Journal of the European Mathematical Society*, 17(2):433–482, 2015. 2, 13, 14

[NOL08] David Nualart and Salvador Ortiz-Latorre. Central limit theorems for multiple stochastic integrals and malliavin calculus. *Stochastic Processes and their Applications*, 118(4):614–628, 2008. 2, 18

[Nou11] Ivan Nourdin. Yet another proof of the nualart-peccati criterion. *arXiv:1107.3252*, 2011. 18

[NP05] David Nualart and Giovanni Peccati. Central limit theorems for sequences of multiple stochastic integrals. *The Annals of Probability*, 33(1):177–193, 2005. 2, 8, 18, 20

[NP15] Ivan Nourdin and Giovanni Peccati. The optimal fourth moment theorem. *Proceedings of the American Mathematical Society*, 143(7):3123–3133, 2015. 2

[PT05] Giovanni Peccati and Ciprian A Tudor. Gaussian limits for vector-valued multiple stochastic integrals. In *Séminaire de Probabilités XXXVIII*, pages 247–262. Springer, 2005. 18

[Rio09] Emmanuel Rio. Upper bounds for minimal distances in the central limit theorem. In *Annales de l'IHP Probabilités et statistiques*, volume 45:3, pages 802–817, 2009. 5

[Rud91] Walter Rudin. *Functional Analysis*. McGraw-Hill, 2 edition, 1991. 6

[Tao13] Terence Tao. Ultraproducts as a bridge between discrete and continuous analysis. *Blog, December*, 2013. 2

[Vil09] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009. 17

[Wil70] Stephen Willard. *General Topology*. Addison-Wesley, 1970. 5

# A Probability preliminaries

**Lemma A.1** ([Bil99, Example 2.4])**.** *Let $x_{n,k}$ be real-valued random variables. Endow $\mathbb{R}^{\mathbb{N}}$ with the product topology. The following are equivalent:*

*(i)* $(x_{n,k})_{k\in\mathbb{N}} \xrightarrow{d} (x_k)_{k\in\mathbb{N}}$

*(ii)* $(x_{n,1},\dots,x_{n,k}) \xrightarrow{d} (x_1,\dots,x_k)$ *for all* $k$

*(iii)* $(x_{n,j})_{j\in J} \xrightarrow{d} (x_j)_{j\in J}$ *for all finite sets* $J \subseteq \mathbb{N}$.

**Lemma A.2** (Method of moments [Bil95, Theorems 29.4, 30.1, 30.2])**.** *Let $x_{n,k}$ be real-valued random variables. Assume that $(x_1,\dots,x_k)$ have a jointly Gaussian distribution. Assume for all finitely supported $\alpha \in \mathbb{N}^k$,*

$$\mathbb{E}\prod_{i=1}^{k} x_{n,i}^{\alpha_i} \longrightarrow \mathbb{E}\prod_{i=1}^{k} x_i^{\alpha_i}\,.$$

*Then $(x_{n,1},\dots,x_{n,k}) \xrightarrow{d} (x_1,\dots,x_k)$.*

**Lemma A.3** (Gaussian hypercontractive inequality)**.** *For all $p \in \mathcal{H}_{\le d}(X)$ and $q \ge 2$,*

$$\|p\|_q \le (q-1)^{d/2}\|p\|_2\,.$$

**Lemma A.4.** *Let $d_1,\dots,d_k \in \mathbb{N}$ and $p_{n,i} \in \mathcal{H}_{\le d_i}(X)$. If $(p_{n,1},\dots,p_{n,k}) \xrightarrow{d} (p_1,\dots,p_k)$ then all mixed moments converge, and the convergence also holds in $W_p$ for $p \in [1,\infty)$. The same holds for graded polynomials $p_{n,i} \in \mathcal{H}^{\mathrm{gr}}_{d_i}(Z)$.*

*Proof.* Write $P_n = (p_{n,1},\dots,p_{n,k})$ and $P = (p_1,\dots,p_k)$. Since $P_n \xrightarrow{d} P$, each coordinate sequence $(p_{n,i})_{n\in\mathbb{N}}$ has tight laws. Recall a collection of probability measures on $\mathbb{R}^I$ is *tight* if for every $\varepsilon > 0$ there is compact $K \subseteq \mathbb{R}^I$ such that $\mu(K) \ge 1-\varepsilon$ for every measure $\mu$ in the collection.

By tightness, the Paley–Zygmund inequality, and Lemma A.3, we have for $p_{n,k} \neq 0$,

$$\Pr\left(|p_{n,k}| \ge \frac{\|p_{n,k}\|_2}{\sqrt{2}}\right) \ge \frac{\|p_{n,k}\|_2^4}{4\|p_{n,k}\|_4^4} \ge \frac{1}{4\cdot 9^{d_k}}.$$

Thus tightness implies $\sup_n \|p_{n,k}\|_2 < \infty$. Applying hypercontractivity again, for every $q \ge 2$,

$$\sup_n \|p_{n,k}\|_q \le (q-1)^{d_k/2} \sup_n \|p_{n,k}\|_2 < \infty.$$

Consequently, we deduce $\sup_n \mathbb{E}\,\|P_n\|^q < \infty$ and, by the Portmanteau theorem, $\mathbb{E}\,\|P\|^q < \infty$ for all $q \ge 1$. Then for every continuous function $\varphi$ of polynomial growth, the sequence $(\varphi(P_n))_{n\in\mathbb{N}}$ is uniformly integrable. The continuous mapping theorem and [Bil99, Theorem 3.5] imply

$$\mathbb{E}\,\varphi(P_n) \longrightarrow \mathbb{E}\,\varphi(P).$$

Thus we have convergence of all mixed moments. Taking $\varphi(x) = \|x\|^q$ gives convergence of the $q$th absolute moments. Together with weak convergence, this implies $W_q(\mathrm{Law}(P_n),\mathrm{Law}(P)) \to 0$ by [Vil09, Theorem 6.9]. Forgetting grades shows $\mathcal{H}^{\mathrm{gr}}_d(Z) \subseteq \mathcal{H}_{\le d}(Z)$ so the same argument applies to graded polynomials. □

# B Combinatorial proof of the fourth moment theorem

We use the tensor notation from Section 3.3.1. We adopt the Nualart and Peccati [NP05] notation that $P \otimes_r P$ equals $(P^{(r)})^\intercal P^{(r)}$ unflattened into a tensor. $P \otimes_r P$ is a non-symmetric tensor with order $2(d-r)$ when $P$ has order $d$.

**Theorem B.1** (Fourth moment theorem [NP05, PT05]). *Let $d_1, \dots, d_k \in \mathbb{N} \setminus \{0\}$ (repeats allowed). Let $p_{n,i} \in \mathcal{H}_{d_i}(X_1, \dots, X_n)$. Let $P_{n,i}$ be their Hermite coefficient tensors. Assume that $\lim_{n\to\infty} \mathbb{E}\, p_{n,i} p_{n,j} = \Sigma_{ij}$ for some matrix $\Sigma \in \mathbb{R}^{k\times k}$. The following are equivalent:*

*(i)* $(p_{n,1}, \dots, p_{n,k}) \xrightarrow{d} \mathcal{N}(0, \Sigma)$

*(ii)* $\mathbb{E}\, p_{n,i}^4 \to 3\Sigma_{ii}^2$ *for all* $i = 1, \dots, k$.

*(iii)* $\|P_{n,i} \otimes_r P_{n,i}\|_F \to 0$ *for all* $0 < r < d_i$ *and* $i = 1, \dots, k$.

*(iv)* $\|P_{n,i}^{(r)}\| \to 0$ *for all* $0 < r < d_i$ *and* $i = 1, \dots, k$.

Equivalence of (i),(ii),(iii) was proven by [NP05, NOL08]. It is easy to check that (iii) is equivalent to (iv) using the relations:

$$\|P^{(r)}\|^2 \le \|P \otimes_r P\|_F \le \|P^{(r)}\| \|P\|_F \,. \tag{B.1}$$

To see these relations, let $s_j$ be the singular values of $P^{(r)}$. Eq. (B.1) follows from the equations,

$$\|P^{(r)}\| = \max_j s_j, \qquad \|P \otimes_r P\|_F^2 = \sum_j s_j^4, \qquad \|P\|_F^2 = \sum_j s_j^2 \,.$$

Since $\|P_{n,i}\|_F^2 = \mathbb{E}\, p_{n,i}^2$ converges to the constant $\Sigma_{ii}$, then Eq. (B.1) implies $\|P \otimes_r P\|_F$ vanishes if and only if $\|P_{n,i}^{(r)}\|$ vanishes. Hence (iii) is equivalent to (iv).

We now give a direct proof of Theorem B.1 using the moment method (see also [Nou11] for a prior combinatorial proof). The moments of Hermite polynomials have natural combinatorial interpretations in terms of matchings.

**Lemma B.2** ([Jan97, Chapter 3]). *Let $\alpha^1, \dots, \alpha^\ell \in \mathbb{N}^I$ be finitely supported. Then:*

$$\mathbb{E}\left[\prod_{j=1}^{\ell} \mathrm{He}_{\alpha^j}(X)\right] = \sum_{M \in \mathcal{M}(\alpha^1, \dots, \alpha^\ell)} \prod_{uv \in M} \mathbf{1}_{u=v}$$

*where we take $\ell$ groups of objects, with the $j$th group being a multiset of size $|\alpha^j|$ with $\alpha_i^j$ copies of each $i \in I$. Let $\mathcal{M}(\alpha^1, \dots, \alpha^\ell)$ be the set of perfect matchings on the disjoint union of the groups such that no two objects are matched within a group. The equality $u = v$ is on $I$.*

A *tensor network* is a multigraph $\pi$ along with a collection of symmetric tensors $T_a \in (\mathbb{R}^n)^{\otimes \deg(a)}_{\mathrm{sym}}$ associated to the vertices of $\pi$. The value of the tensor network is:

$$\mathrm{val}_\pi(T) = \sum_{\varphi : E(\pi) \to [n]} \prod_{a \in V(\pi)} (T_a)_{\varphi(a)}$$

where $\varphi(a)$ is the multiset of labels on the edges incident to $a$ (the order of the incident edges does not matter since $T_a$ is assumed to be symmetric).

**Lemma B.3.** *Let $(T, \pi)$ be a tensor network such that $\pi$ has no self-loops. Then:*

$$|\operatorname{val}_\pi(T)| \le \prod_{a \in V(\pi)} \|T_a\|_F.$$

*Proof.* $\operatorname{val}_\pi(T)$ factors across connected components of $\pi$, so we may assume $\pi$ is connected. Use induction on $|V(\pi)|$. We also upgrade the statement to allow non-symmetric tensors (in this case, each vertex of $\pi$ comes with an ordering of the incident edges).

In the base case, $\pi$ is a single multiedge between two tensors $R, S$,

$$|\operatorname{val}_\pi(T)| = \left|\sum_{i \in [n]^d} R_i S_i\right| \le \left(\sum_{i \in [n]^d} R_i^2\right)^{1/2} \left(\sum_{i \in [n]^d} S_i^2\right)^{1/2} = \|R\|_F \|S\|_F \,.$$

In the induction step, let $a, b$ be the endpoints of a multiedge in $\pi$ with associated tensors $R, S$. Let $\pi'$ be $\pi$ with the multiedge contracted. Let $Q$ be the contraction of $R, S$ along the multiedges between $a, b$ and associate $Q$ to the contracted vertex in $\pi'$. Call this new tensor network $(T', \pi')$. Then $\operatorname{val}_\pi(T) = \operatorname{val}_{\pi'}(T')$ by definition. By induction,

$$|\operatorname{val}_{\pi'}(T')| \le \|Q\|_F \prod_{c \in V(\pi) \setminus \{a,b\}} \|T_c\|_F \,.$$

We can express $Q$ as a flattened matrix multiplication $R^{(L_a, R_a)} S^{(L_b, R_b)}$ followed by unflattening, where $R_a, L_b$ are the edges between $a, b$. Using $\|AB\|_F \le \|A\|_F \|B\|_F$ on the matrix multiplication,

$$\|Q\|_F \prod_{c \in V(\pi) \setminus \{a,b\}} \|T_c\|_F \le \|R\|_F \|S\|_F \prod_{c \in V(\pi) \setminus \{a,b\}} \|T_c\|_F = \prod_{a \in V(\pi)} \|T_a\|_F \,. \qquad \square$$

**Lemma B.4.** *Let $d_1, \dots, d_k \ge 1$. Let $p_a \in \mathcal{H}_{d_a}(X_1, \dots, X_n)$ such that $\mathbb{E}\, p_a^2 \le 1$ and let $P_a$ be their Hermite coefficient tensors. Assume that $\|P_a^{(r)}\| \le \delta$ for all $a \in [k]$ and $0 < r < d_a$. Let $Z = (Z_1, \dots, Z_k)$ be a centered Gaussian vector with the same covariance matrix as $(p_1, \dots, p_k)$, and let $D = \sum_{a=1}^k d_a$. If $D$ is even, then*

$$\left|\mathbb{E} \prod_{a=1}^k p_a - \mathbb{E} \prod_{a=1}^k Z_a\right| \le \frac{(D-1)!!}{\sqrt{d_1! \cdots d_k!}}\, \delta. \tag{B.2}$$

*If $D$ is odd, both expectations vanish.*

*Proof.* Take a set of $k$ vertices, where $a$ is incident to $d_a$ labeled half-edge endpoints. Let $\mathcal{D}$ be the set of perfect matchings of all $D$ half-edges such that no two half-edges at the same vertex are matched. $\pi \in \mathcal{D}$ forms a tensor network with $P_1, \dots, P_k$. We have,

$$\begin{aligned} \mathbb{E} \prod_{a=1}^k p_a &= \frac{1}{\sqrt{d_1! \cdots d_k!}} \sum_{i_a \in [n]^{d_a}} \prod_{a=1}^k (P_a)_{i_a}\, \mathbb{E}\left[\prod_{a=1}^k \operatorname{He}_{i_a}(X)\right] && \text{(Eq. (3.3))} \\ &= \frac{1}{\sqrt{d_1! \cdots d_k!}} \sum_{\pi \in \mathcal{D}} \operatorname{val}_\pi(P_1, \dots, P_k). && \text{(Lemma B.2)} \end{aligned} \tag{B.3}$$

Eq. (B.3) is a "Feynman diagram expansion" for the moment $\mathbb{E}\prod_{a=1}^{k} p_a$. When $D$ is odd, $\mathcal{D}$ is empty, and the expectation is zero as needed.

Call a diagram $\pi$ a *pair diagram* if every connected component has exactly two vertices. Such a component on vertices $a, b$ is possible only if $d_a = d_b = d$. Its $d!$ possible matchings all give $\langle P_a, P_b\rangle_F = \mathbb{E}[p_a p_b]$, and its normalization in (B.3) is $1/d!$. Thus the sum over pair diagrams equals

$$\sum_{\text{pairings } \eta \text{ of } \{1,\dots,k\}} \prod_{\{a,b\}\in\eta} \mathbb{E}[p_a p_b] = \mathbb{E}\prod_{a=1}^{k} Z_a.$$

Every remaining diagram has a connected component with at least three vertices. Choose adjacent vertices $a, b$ in this component, and let $r \geq 1$ be the number of edges between them. These edges cannot be all edges incident to both vertices, so, after interchanging $a, b$ if necessary, $0 < r < d_a$. As in the proof of Lemma B.3, form a new tensor network $\pi'$ by contracting the vertices $a, b$ and putting at the contracted vertex the tensor $Q$ which is the contraction of $P_a, P_b$. This defines $\mathrm{val}_\pi(P) = \mathrm{val}_{\pi'}(P, Q)$. By Lemma B.3,

$$|\mathrm{val}_{\pi'}(P, Q)| \leq \|Q\|_F \prod_{c\in[k]\setminus\{a,b\}} \|P_c\|_F \,.$$

We can express $Q$ as the unflattening of a matrix multiplication $(P_a^{(r)})^\intercal P_b^{(r)}$. Use $\|AB\|_F \leq \|A\|\|B\|_F$ on the matrix multiplication to deduce,

$$\|Q\|_F \prod_{c\in[k]\setminus\{a,b\}} \|P_c\|_F \leq \|P_a^{(r)}\| \prod_{b\in[k]\setminus\{a\}} \|P_b\|_F \leq \|P_a^{(r)}\| \leq \delta \,.$$

The second-to-last inequality uses $\|P_b\|_F^2 = \mathbb{E}\, p_b^2 \leq 1$. Thus $|\mathrm{val}_\pi(P)| \leq \delta$ for each non-pair diagram in Eq. (B.3). There are at most $(D-1)!!$ matchings of $D$ slots, which proves Eq. (B.2). □

*Proof of Theorem B.1.* (i) $\implies$ (ii). Convergence in distribution implies convergence of moments by Lemma A.4.

(ii) $\implies$ (iii). Let $p \in \mathcal{H}_d(X_1, \dots, X_n)$ have Hermite coefficient tensor $P$. The fourth cumulant $\kappa_4 := \mathbb{E}\, p^4 - 3(\mathbb{E}\, p^2)^2$ can be written as a sum-of-squares expression in terms of tensor contractions as [NP05, Proof of Proposition 3, (i) $\implies$ (ii)],

$$\kappa_4 = \sum_{r=1}^{d-1} \binom{d}{r}^2 \left( \|P \otimes_r P\|_F^2 + \binom{2d-2r}{d-r} \|\operatorname{Sym}(P \otimes_r P)\|_F^2 \right) . \tag{B.4}$$

Therefore, if $\kappa_4 \to 0$ as $n \to \infty$ then all of the tensor norms $\|P \otimes_r P\|_F$ must also vanish.

Here is a combinatorial interpretation of (B.4). Following Eq. (B.3),

$$\mathbb{E}\, p^4 = \frac{1}{(d!)^2} \sum_{\pi\in\mathcal{D}} \mathrm{val}_\pi(P, P, P, P)$$

where $\mathcal{D}$ is the set of perfect matchings on four vertices with $d$ incident half-edges such that no half-edges are matched at the same vertex.

Suppose $\pi$ has $r$ edges matched between vertex 1 and 2. Then there are also $r$ edges matched between vertex 3 and 4 and there are $\left(\binom{d}{r}^2 r!\right)^2$ ways to choose this part of $\pi$. We interpret the edges between 1,2 as contracted in $P \otimes_r P$ and the edges outgoing from 1,2 as the indices of $P \otimes_r P$.

There are $(2d-2r)!$ ways to choose the remaining part of $\pi$. If the remaining part of $\pi$ matches the half-edges at 1 with the half-edges at 3 (i.e. $\pi$ is a cycle graph), we obtain exactly $\mathrm{val}_\pi(P,P,P,P) = \|P \otimes_r P\|_F^2$. However, in general, the sum over $\pi$ treats all outgoing edges from $\{1,2\}$ in the same way, so the average value of $\mathrm{val}_\pi(P,P,P,P)$ is $\|\operatorname{Sym}(P \otimes_r P)\|_F^2$. In total,

$$\begin{aligned}\frac{1}{(d!)^2}\sum_{\pi\in\mathcal{D}} \mathrm{val}_\pi(P,P,P,P) &= \sum_{r=0}^{d} \frac{(2d-2r)!}{(d!)^2}\left(\binom{d}{r}^2 r!\right)^2 \|\operatorname{Sym}(P\otimes_r P)\|_F^2 \\ &= \sum_{r=0}^{d}\binom{d}{r}^2\binom{2d-2r}{d-r}\|\operatorname{Sym}(P\otimes_r P)\|_F^2 \end{aligned} \tag{B.5}$$

We re-analyze the terms $r=0$ and $r=d$ separately. The term $r=d$ equals $\|P\|_F^4 = (\mathbb{E}\,p^2)^2$. All $\pi$ in the term $r=0$ are cyclic graphs (or disconnected). Let $s$ be the number of 1-3 edges which equals the number of 2-4 edges. There are $\binom{d}{s}^4 (s!)^2((d-s)!)^2 = \binom{d}{s}^2 (d!)^2$ choices for $\pi$. For these cases, $\mathrm{val}_\pi(P,P,P,P) = \|P\otimes_s P\|_F^2$. Therefore,

$$\begin{aligned}\frac{1}{(d!)^2}\sum_{\pi\in\mathcal{D}} \mathrm{val}_\pi(P,P,P,P) &= \|P\|_F^4 + \sum_{s=0}^{d}\binom{d}{s}^2\|P\otimes_s P\|_F^2 + \sum_{r=1}^{d-1}\binom{d}{r}^2\binom{2d-2r}{d-r}\|\operatorname{Sym}(P\otimes_r P)\|_F^2 \\ &= 3\|P\|_F^4 + \sum_{s=1}^{d-1}\binom{d}{s}^2\|P\otimes_s P\|_F^2 + \sum_{r=1}^{d-1}\binom{d}{r}^2\binom{2d-2r}{d-r}\|\operatorname{Sym}(P\otimes_r P)\|_F^2\end{aligned}$$

Recall this equals $\mathbb{E}\,p^4$. Subtract $3\|P\|_F^4 = 3(\mathbb{E}\,p^2)^2$ to obtain $\kappa_4$.

(iii) $\Longrightarrow$ (iv). Proven immediately after the theorem statement.

(iv) $\Longrightarrow$ (i). Let $Z_n$ be a centered Gaussian vector with the same covariance matrix as $(p_{n,1},\dots,p_{n,k})$. The coordinatewise variances of $Z_n$ are uniformly bounded since $\mathbb{E}\,p_{n,i}^2 \to \Sigma_{ii}$, say by a constant $C \ge 1$. We may apply Lemma B.4 to get, for every $(m_1,\dots,m_k)\in\mathbb{N}^k$, letting $|m| = \sum_a m_a$ and $D = \sum_a m_a d_a$ and $\delta_n = \max_{a\in[k]}\max_{0<r<d_a}\|P_{n,a}^{(r)}\|$,

$$|\mathbb{E}\prod_{a=1}^{k} p_{n,a}^{m_a} - \mathbb{E}\prod_{a=1}^{k} Z_{n,a}^{m_a}| \le C^{|m|}(D-1)!!\,\delta_n \longrightarrow 0.$$

The covariances and hence also all moments of $Z_n$ converge to those of $\mathcal{N}(0,\Sigma)$. So by the method of moments in Lemma A.2, $(p_{n,1},\dots,p_{n,k}) \xrightarrow{d} \mathcal{N}(0,\Sigma)$. □

## B.1 Omitted moment calculations

**Lemma B.5.** *For $i,r\in\mathbb{N}$, let $S_{ir} = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\psi_r(G_{ir,j})$. For finitely-supported $\alpha\in\mathbb{N}^{\mathbb{N}\times\mathbb{N}}$, let*

$$U_{m,ir} = \frac{1}{N^{m/2}}\sum_{\substack{j_1,\dots,j_m\in[N]\\ j_1,\dots,j_m \text{ distinct}}} \psi_r(G_{ir,j_1})\psi_r(G_{ir,j_2})\cdots\psi_r(G_{ir,j_m}), \qquad U_\alpha = \prod_{i,r} U_{\alpha_{ir},ir}\,.$$

*Then for $G_{ir,j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$,*

$$\mathbb{E}(U_\alpha - \mathrm{He}_\alpha(S))^2 \le O_\alpha(N^{-1})\,.$$

*Proof.* Consider generic $S = N^{-1/2}\sum_{j=1}^{N}\psi_r(G_j)$ and $U_m = N^{-m/2}\sum_{j_1,\dots,j_m \text{ distinct}}\prod_{a=1}^{m}\psi_r(G_{j_a})$. Using Lemma B.2,

$$\mathbb{E}\,S^{2\ell} = \frac{1}{(Nr!)^\ell}\sum_{M\in\mathcal{M}(2\ell,r)}\sum_{\varphi:[2\ell]\to[N]}\prod_{uv\in M}\mathbf{1}_{\varphi(\text{group}(u))=\varphi(\text{group}(v))}\,.$$

where we have $2\ell$ groups of $r$ vertices each, $\mathcal{M}(2\ell,r)$ is the set of perfect matchings on the vertices such that no two vertices are matched within a group, and the indicator function checks whether all matched pairs use the same Gaussian $G_k$ on their group. The inner summation is $N^{c(M)}$ where $c(M)$ is the number of connected components of $M$ when the groups are contracted. Hence,

$$\mathbb{E}\,S^{2\ell} = \frac{1}{(Nr!)^\ell}\sum_{M\in\mathcal{M}(2\ell,r)}N^{c(M)} = (2\ell-1)!! + O_\alpha(N^{-1})\,. \tag{B.6}$$

The second equality separates the dominant terms where $\mathcal{M}$ is a perfect matching between groups from the remaining $O_\alpha(1)$ terms.

Express the Hermite polynomials in the monomial basis and use Eq. (B.6) to deduce $\mathbb{E}\,\mathrm{He}_m(S)^2 = \mathbb{E}\,\mathrm{He}_m(X)^2 + O_\alpha(N^{-1}) = m! + O_\alpha(N^{-1})$ for standard Gaussian variables $X$. We also have

$$\mathbb{E}\,\mathrm{He}_m(S)U_m = \mathbb{E}\,U_m^2 = m!\frac{(N)_m}{N^m} = m! + O_m(N^{-1})\,. \tag{B.7}$$

where $(N)_m$ is the falling factorial. Indeed, the first equality observes that the only component of $\mathrm{He}_m(S)$ which is not orthogonal to $U_m$ is its "leading monomial" $U_m$ itself. The second equality observes that nonzero terms in $\mathbb{E}\,U_m^2$ are perfect matchings between the indices in $U_m$. Therefore,

$$\mathbb{E}(\mathrm{He}_m(S) - U_m)^2 = \mathbb{E}\,\mathrm{He}_m(S)^2 - 2\,\mathbb{E}\,\mathrm{He}_m(S)U_m + \mathbb{E}\,U_m^2 = O_\alpha(N^{-1})\,.$$

Independence across the $i, r$ coordinates implies the same holds for the multivariate case. □

**Lemma B.6.** *Let $d_1,\dots,d_k \geq 1$ and $D = \sum_{a=1}^{k} d_a$. Let $p_a \in \mathcal{H}_{d_a}(X_1,\dots,X_n)$ such that $\mathbb{E}\,p_a^2 = 1$ and let $P_a$ be their Hermite coefficient tensors. Assume that $\|P_a^{(r)}\| \leq \delta$ for all $a \in [k]$ and $0 < r < d_a$. Then*

$$\left|\mathbb{E}\left(\bigdiamond_{a=1}^{k} p_a\right)^2 - \sum_{M\in\mathcal{M}(2,k)}\prod_{uv\in M}\mathbb{E}[p_u p_v]\right| \leq \frac{D!}{d_1!\cdots d_k!}\delta$$

*where $\mathcal{M}(2,k)$ is the set of perfect matchings between two copies of $[k]$.*

*Proof sketch.* The tensor $P = \sqrt{\frac{D!}{d_1!\cdots d_k!}}\otimes_{a=1}^{k} P_a$ represents the top product $\diamond_{a=1}^{k} p_a$. Then we write a diagrammatic expansion for $\mathbb{E}\left(\bigdiamond_{a=1}^{k} p_a\right)^2$ in the same way as Lemma B.4. Separating out the dominant and subdominant perfect matchings in the same way proves the claim. □