DISCRETE STOCHASTIC ANALYSIS

NIKOS ZYGOURAS

Abstract. These are notes of series of lectures given at National Taiwan University and the University of Warwick.

Part of the classical stochastic analysis is devoted to the analysis of the so-called Wiener chaos, which is used to express L^2 random variables as a series expansion of iterated Wiener-Itô integrals. Theories like Malliavin calculus, hypercontractivity, Wick normalisation etc. play a significant role in the analysis of these expansions and associated gaussian spaces.

From the point of view of statistical mechanics of disordered systems or theoretical computer science and boolean functions, one is motivated to look at discrete analogues of Wiener chaos and develop tools that will allow to analyse these discrete structures. Furthermore, one is interested in scaling limits, which amounts to establishing convergence of the discrete structure to the continuum objects.

We use the term "Discrete Stochastic Analysis" to describe a set of tools that fall into this framework. The topics we will expose in these lectures cover

- general Lindeberg principles
- convergence of multilinear polynomials of random variables to Wiener chaoses
- Hoeffding decomposition
- the Fourth Moment Theorem
- elements of Malliavin calculus
- Stein's method
- discrete versions of general functional (such as Poincaré) inequalities
- applications

Contents

Date: March 15, 2021.

1991 *Mathematics Subject Classification.* Primary: 82B44.

Key words and phrases. Lindeberg theorem, Fourth moment theorem, Wiener chaos, multilinear chaos, disordered systems.

1. Lindeberg Theorems

1.1. Lindeberg's central limit theorem.

Theorem 1.1. *Let $(\omega_{n,j})_{1\leqslant j\leqslant n}$ be an i.i.d. triangular array satisfying*

$$\mathbb{E}\omega_{n,j}=0 \quad , \quad \sum_{i=1}^{n}\mathbb{E}\omega_{n,i}^2=1 \quad \textit{and} \tag{1.1}$$

$$\textit{for every } \varepsilon>0 \quad \sum_{i=1}^{n}\mathbb{E}\big[\omega_{n,i}^2\,;\,|\omega_{n,i}|>\varepsilon\big]\xrightarrow[n\to\infty]{}0. \tag{1.2}$$

Then $Z_n:=\omega_{n,1}+\cdots\omega_{n,n}$ converges in distribution to a standard Gaussian random variable.

Before getting into the proof, let us make a few remarks. First, let us see how this applies to the standard central limit theorem. In this situation the triangular array is $\omega_{n,i}=\omega_i/\sqrt{n}$. The Lindeberg condition writes as

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big[\omega_i^2\,;\,|\omega_i|\geqslant\varepsilon\sqrt{n}\big]=\mathbb{E}\big[\omega^2\,;\,|\omega|\geqslant\varepsilon\sqrt{n}\big],$$

which converges to zero by the assumption on finite second moment and dominated convergence.

The Lindeberg condition says that the CLT holds whenever *no* random variable in the sum is exceedingly large. So the Lindeberg condition is to be thought as a *uniform smallness* assumption. This is in contrast with the Poisson convergence where typically all variables are negligible except very few ones which are "large". For example if $(\omega_{n,i})$ are $\{0,1\}-$*valued* variables with

$$p_{n,i}=\mathbb{P}(\omega_{n,i}=1)=1-\mathbb{P}(\omega_{n,i}=0),\quad \sum_{i=1}^{n}p_{n,i}\xrightarrow[n\to\infty]{}\lambda>0 \quad\text{and}\quad \max_{1\leqslant i\leqslant n}p_{n,i}\xrightarrow[n\to\infty]{}0,$$

then $\sum_{i=1}^{n}\omega_{n,i}$ converges, as $n\to\infty$, to a Poisson random variable with parameter λ (check this as an exrecise).

Our final remark, which is also a preparation for the proof is that the Lindeberg theorem can be thought as a perturbation argument. This is to be understood as follows: the CLT is obvious when $(\omega_{n,i})$ are Gaussian variables (since the sum of Gaussian variables is Gaussian). Lindeberg's *uniform smallness* condition allows to say that for large n the limit that one has for general *i.i.d* variables is asymptotically the same if one had Gaussian variables.

Proof of Lindeberg's theorem. Let $f\in C_b(\mathbb{R})$ and for a triangular array $(\omega_{n,i})$ denote

$$f_n(\omega_{n,1},\ldots,\omega_{n,n}):=f\big(\omega_{n,1}+\cdots+\omega_{n,n}\big). \tag{1.3}$$

We will also consider the i.i.d. sequence of Gaussian variables $\xi_1,\xi_2,...$ and consider (1.3) but with $\omega_{n,1},\omega_{n,2},...$ replaced with $\xi_{n,i}:=n^{-1/2}\xi_i$ for $i=1,2,....$ By the definition of weak convergence, it suffices to show that

$$\mathbb{E}\big[f_n(\omega_{n,1},\dots,\omega_{n,n})\big] \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)\, e^{-\frac{x^2}{2}}\,\mathrm{d}x,$$

and since this limit is trivially valid for $\mathbb{E}\big[f_n(\xi_{n,1},\dots,\xi_{n,n})\big]$, it suffices to show that

$$\Big|\mathbb{E}\big[f_n(\omega_{n,1},\dots,\omega_{n,n})\big]-\mathbb{E}\big[f_n(\xi_{n,1},\dots,\xi_{n,n})\big]\Big| \xrightarrow[n\to\infty]{} 0. \tag{1.4}$$

The perturbation argument alluded to in the above remarks will be done through a telescoping argument, where we will successively change the array $(\omega_{n,1},...,\omega_{n,n})$ one by one, until we change all the array to $(\xi_{n,1},...,\xi_{n,n})$. In this way, we can bound the left hand side of (1.4) (the "bound" just amounts to simple use of the triangle inequality) by

$$\begin{aligned}\sum_{i=1}^{n}\Big|\ &\mathbb{E}\big[f_n(\,\xi_{n,1}\,,\dots,\,\xi_{n,i-1}\,,\,\xi_{n,i}\,,\,\omega_{n,i+1}\,,...,\,\omega_{n,n}\,)\big]\ -\\ &-\ \mathbb{E}\big[f_n(\,\xi_{n,1}\,,\dots,\,\xi_{n,i-1}\,,\,\omega_{n,i}\,,\,\omega_{n,i+1}\,,...,\,\omega_{n,n}\,)\big]\Big|,\end{aligned} \tag{1.5}$$

where we notice that in the above difference there is only a discrepancy at the i^{th} coordinate. We will Taylor expand in that coordinate. For this, let us introduce, for a sequence $x_1,...,x_n$, the function

$$h^x_{n,i}(y) := f(x_1,...,x_{i-1},y,x_{i+1},...,x_n).$$

The Taylor expansion is as follows:

$$h^x_{n,i}(y)=h^x_{n,i}(0)+\Big(\partial_y h^x_{n,i}(0)\Big)\,y+\frac{1}{2}\Big(\partial_y^2 h^x_{n,i}\Big)(0)\,y^2+R^x_{n,i}(y), \tag{1.6}$$

where the remainder term has the expression

$$R^x_{n,i}(y)=\frac{1}{2}\int_0^y\Big(\partial_y^3 h^x_{n,i}(t)\Big)\,(y-t)^2\mathrm{d}t, \tag{1.7}$$

and the following two bounds hold:

$$\big|\,R^x_{n,i}(y)\,\big| \leqslant \frac{1}{6}\|\,\partial_y^3 h^x_{n,i}\,\|_\infty\,|y|^3=\frac{1}{6}\|f'''\|_\infty|y|^3 \tag{1.8}$$

$$\big|\,R^x_{n,i}(y)\,\big| \leqslant \|\,\partial_y^2 h^x_{n,i}\,\|_\infty\,y^2=\|f''\|_\infty y^2. \tag{1.9}$$

The first bound follows by bounding $\partial_y^3 h^x_{n,i}$ in (1.7) by its supremum norm, while for the second bound we first perform an integration by parts and write the remainder as

$$R^x_{n,i}(y)=-\frac{1}{2}\partial_y^2 h^x_{n,i}(0)y^2+\int_0^y \partial_y^2 h^x_{n,i}(t)(y-t)\,\mathrm{d}t,$$

and then bound the $\partial_y^2 h^x_{n,i}$ by its supremum norm. Let us introduce the notation

$$[\xi,\omega]_i := (\xi_{n,1},...,\xi_{n,i-1},\omega_{n,i+1},...,\omega_{n,n}),$$

then each difference (1.5) writes as

$$\begin{aligned}h^{[\xi,\omega]_i}_{n,i}(\xi_i)-h^{[\xi,\omega]_i}_{n,i}(\omega_i)=&\left\{h^{[\xi,\omega]_i}_{n,i}(0)+\Big(\partial_y h^{[\xi,\omega]_i}_{n,i}(0)\Big)\,\xi_{n,i}+\frac{1}{2}\Big(\partial_y^2 h^{[\xi,\omega]_i}_{n,i}\Big)(0)\,\xi^2_{n,i}+R^x_{n,i}(\xi_{n,i})\right\}\\ &-\left\{h^{[\xi,\omega]_i}_{n,i}(0)+\Big(\partial_y h^{[\xi,\omega]_i}_{n,i}(0)\Big)\,\omega_{n,i}+\frac{1}{2}\Big(\partial_y^2 h^{[\xi,\omega]_i}_{n,i}\Big)(0)\,\omega^2_{n,i}+R^x_{n,i}(\omega_{n,i})\right\}\end{aligned}$$

Taking, at first, only expectation over the $\xi_{n,i}$, and $\omega_{n,i}$ variables, which we will denote by $\mathbb{E}_i$ and working with the easier assumption that $\xi_{n,i}$'s and $\omega_{n,i}$'s have matching first and second moments (we leave the details in the case of the more general condition (1.1) as an exercise), we have

$$\mathbb{E}_i\Big[h_{n,i}^{[\xi,\omega]_i}(\xi_i)\Big] - \mathbb{E}\Big[h_{n,i}^{[\xi,\omega]_i}(\omega_i)\Big] = \mathbb{E}_i\Big[R_{n,i}^{[\xi,\omega]_i}(\omega_{n,i})\Big] - \mathbb{E}_i\Big[R_{n,i}^{[\xi,\omega]_i}(\xi_{n,i})\Big]$$

So (1.5) is bounded by

$$\sum_{i=1}^n \mathbb{E}_i\Big[\Big|R_{n,i}^{[\xi,\omega]_i}(\omega_{n,i})\Big|\Big] + \sum_{i=1}^n \mathbb{E}_i\Big[\Big|R_{n,i}^{[\xi,\omega]_i}(\xi_{n,i})\Big|\Big].$$

We will estimate the first term, the second one being identical. For this, we denote by $\mathsf{C}_f := \max\{\|f''\|_\infty, \|f'''\|_\infty\}$ and we have by estimates (1.8), (1.9) that

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}_i\Big[\Big|R_{n,i}^{[\xi,\omega]_i}(\omega_{n,i})\Big|\Big] &\leqslant \mathsf{C}_f \sum_{i=1}^n \mathbb{E}\Big[\min\{\omega_{n,i}^2, \tfrac{1}{6}|\omega_{n,i}|^3\}\Big] \\
&= \mathsf{C}_f \sum_{i=1}^n \mathbb{E}\Big[\min\{\omega_{n,i}^2, \tfrac{1}{6}|\omega_{n,i}|^3\}\,;\, |\omega_{n,i}| \geqslant \varepsilon\Big] \\
&\qquad + \mathsf{C}_f \sum_{i=1}^n \mathbb{E}\Big[\min\{\omega_{n,i}^2, \tfrac{1}{6}|\omega_{n,i}|^3\}\,;\, |\omega_{n,i}| < \varepsilon\Big] \\
&\leqslant \mathsf{C}_f \sum_{i=1}^n \mathbb{E}\Big[\omega_{n,i}^2;\, |\omega_{n,i}| \geqslant \varepsilon\Big] + \tfrac{\varepsilon}{6}\mathsf{C}_f \sum_{i=1}^n \mathbb{E}\Big[|\omega_{n,i}|^2\Big]
\end{aligned}$$

and the first term converges to zero by the Lindeberg assumption, while the second can be made arbitrarily small by choosing ε small enough. $\square$

1.2. Efron-Stein inequality and applications. The Efron-Stein inequality is a discrete version of the Poincaré inequality. The latter states that if $f\colon \mathbb{R}^n \to \mathbb{R}$ is, let us assume, "smooth" and if $\xi := (\xi_1, ..., \xi_n)$ is a standard normal vector then

$$\mathbb{V}\mathrm{ar}(f(\xi)) \leqslant \mathbb{E}[\,|\nabla f(\xi)\,|^2].$$

The Efron-Stein inequality is as follows

Theorem 1.2 (Efron-Stein). *Let $\omega = (\omega_1, \omega_2, ..., \omega_n)$ be a vector of i.i.d. variables and $f\colon \mathbb{R}^n \to \mathbb{R}$. Let also T_i for $i = 1, ..., n$ be the operator which acts on ω by resampling independently the i^{th} coordinate of the vector ω. This means that*

$$\mathsf{T}_i\,\omega = (\omega_1, ..., \omega_{i-1}, \tilde{\omega}_i, \omega_{i+1}, ..., \omega_n),$$

where $\tilde{\omega}_i$ is a random variable independent of $\omega_1, ..., \omega_n$ but with the same distribution. Then the following inequality holds

$$\mathbb{V}\mathrm{ar}(f(\omega)) \leqslant \frac{1}{2}\sum_{i=1}^n \mathbb{E}\Big[\big(f(\omega) - f(\mathsf{T}_i\omega)\big)^2\Big]$$

Proof. The proof follows the same telescoping argument as used in the proof of Lindeberg's theorem. Let us recall the notation $\omega = (\omega_1, ..., \omega_n)$ and let us start by

$$\mathbb{V}\mathrm{ar}(f(\omega)) = \mathbb{E}\big[f(\omega)^2\big] - \mathbb{E}\big[f(\omega)\big]^2 = \mathbb{E}\big[f(\omega)^2\big] - \mathbb{E}\big[f(\omega)f(\tilde{\omega})\,\big] = \mathbb{E}\big[f(\omega)\big(f(\omega) - f(\tilde{\omega})\big)\,\big]$$

Let us now telescope the difference

$$f(\omega) - f(\tilde{\omega}) = \sum_{i=1}^{n} \Big(f(\mathsf{T}_1 \cdots \mathsf{T}_{i-1}\, \omega) - f(\mathsf{T}_1 \cdots \mathsf{T}_i\, \omega) \Big)$$

and write

$$f(\omega)\Big(f(\omega) - f(\tilde{\omega})\Big) = \sum_{i=1}^{n} f(\omega)\Big(f(\mathsf{T}_1 \cdots \mathsf{T}_{i-1}\, \omega) - f(\mathsf{T}_1 \cdots \mathsf{T}_i \omega) \Big).$$

and

$$\mathbb{E}\Big[\, f(\omega)\Big(f(\omega) - f(\tilde{\omega})\Big)\Big] = \sum_{i=1}^{n} \mathbb{E}\Big[\, f(\omega)\Big(f(\mathsf{T}_1 \cdots \mathsf{T}_{i-1}\, \omega) - f(\mathsf{T}_1 \cdots \mathsf{T}_i \omega) \Big)\Big]. \tag{1.10}$$

Notice now that

$$f(\omega)\Big(f(\mathsf{T}_1 \cdots \mathsf{T}_{i-1}\, \omega) - f(\mathsf{T}_1 \cdots \mathsf{T}_i\, \omega) \Big) =^d f(\mathsf{T}_i\, \omega)\Big(f(\mathsf{T}_1 \cdots \mathsf{T}_i\, \omega) - f(\mathsf{T}_1 \cdots \mathsf{T}_{i-1}\, \omega) \Big),$$

which is just a consequence of switching $\tilde{\omega}_i$ and ω_i. So (1.10) can be written as

$$\mathbb{E}\Big[\, f(\omega)\Big(f(\omega) - f(\tilde{\omega})\Big)\Big] = \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\Big[\, \Big(f(\omega) - f(\mathsf{T}_i\, \omega)\Big)\Big(f(\mathsf{T}_1 \cdots \mathsf{T}_{i-1}\, \omega) - f(\mathsf{T}_1 \cdots \mathsf{T}_i \omega) \Big)\Big]$$

Applying, now, Cauchy-Schwarz we bound this by

$$\begin{aligned} &\frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\Big[\, \Big(f(\omega) - f(\mathsf{T}_i\, \omega)\Big)^2\Big]^{1/2} \mathbb{E}\Big[\Big(f(\mathsf{T}_1 \cdots \mathsf{T}_{i-1}\, \omega) - f(\mathsf{T}_1 \cdots \mathsf{T}_i \omega) \Big)^2\Big]^{1/2} \\ &= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\Big[\, \Big(f(\omega) - f(\mathsf{T}_i\, \omega)\Big)^2\Big] \end{aligned}$$

where in the last step we replaced variables $\tilde{\omega}_1, ..., \tilde{\omega}_{i-1}$ by $\omega_1, ..., \omega_{i-1}$. The proof is now complete. □

We will now present an application of the Efron-Stein inequality to a model that has attracted much interest in probability and statistical mechanics called *first passage percolation.* The model is defined as follows: Consider $(\omega_{i,j})_{1 \leqslant i,j \leqslant n}$ to be an array of i.i.d. random variables with finite second moment. Then the first passage percolation time is defined to be

$$\tau_n := \min_{\pi:\ (1,1)\to(n.n)} \sum_{(i,j)\in\pi} \omega_{i,j},$$

where the minimum is taken over all nearest neighbour directed up-right paths from $(1,1)$ to (n,n).

Proposition 1.3. *Assume that $(\omega_{i,j})_{i,j \leqslant n}$ is a family of i.i.d., non negative random variables with finite second moments. Then there is a constant C such that*

$$\mathbb{V}\mathrm{ar}\,\big(\tau_n\big) \leqslant Cn.$$

Proof. Let us for notational convenience denote by $x = (i,j)$. Then by the Efron-Stein inequality we have that

$$\mathbb{V}\mathrm{ar}\,\big(\tau_n\big) \leqslant \frac{1}{2}\sum_{x} \mathbb{E}\Big[\big(\tau_n(\omega) - \tau_n(\mathsf{T}_x \omega)\big)^2\Big]. \tag{1.11}$$

Let $\tilde{\omega}_x$ be the resampled value of disorder at site x in $\mathsf{T}_x\omega$. By symmetry (and assuming without loss of generality that ω_x as a continuous distribution), we have that the right hand side of (1.11) equals

$$\begin{aligned}\frac{1}{2}\sum_x \mathbb{E}\Big[\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2\Big] &= \frac{1}{2}\sum_x \mathbb{E}\Big[\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2\,;\,\omega_x \leqslant \tilde{\omega}_x\Big]\\ &\quad + \frac{1}{2}\sum_x \mathbb{E}\Big[\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2\,;\,\omega_x \geqslant \tilde{\omega}_x\Big]\\ &= \sum_x \mathbb{E}\Big[\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2\,;\,\omega_x \leqslant \tilde{\omega}_x\Big]\end{aligned}$$

Denoting by $\pi_*(\omega)$ the optimal path, along which the min in $\tau_n(\omega)$ is achieved, we decompose

$$\begin{aligned}\sum_x \mathbb{E}\Big[\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2\,;\,\omega_x \leqslant \tilde{\omega}_x\Big] &= \sum_x \mathbb{E}\Big[\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2\,;\,\omega_x \leqslant \tilde{\omega}_x\,,\,\mathbb{1}_{x\in\pi_*(\omega)}\Big]\\ &\quad + \sum_x \mathbb{E}\Big[\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2\,;\,\omega_x \leqslant \tilde{\omega}_x\,,\,\mathbb{1}_{x\notin\pi_*(\omega)}\Big]\end{aligned}$$

On the intersection of the events that $x \notin \pi_*(\omega)$ and $\omega_x \leqslant \tilde{\omega}_x$, we have that $\tau_n(\omega) = \tau_n(\mathsf{T}_x\omega)$. This is because if an optimal (min) path in the environment ω does not pass through point x, then it will also not pass through when the environment at x is changed to a larger value. On the other hand, in any case we have the bound $\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2 \leqslant (\tilde{\omega}_x-\omega_x)^2$. Therefore, we have that

$$\begin{aligned}\sum_x \mathbb{E}\Big[\big(\tau_n(\omega)-\tau_n(\mathsf{T}_x\omega)\big)^2\,;\,\omega_x \leqslant \tilde{\omega}_x\Big] &\leqslant \sum_x \mathbb{E}\Big[\big(\tilde{\omega}_x-\omega_x)\big)^2\,;\,\omega_x \leqslant \tilde{\omega}_x\,,\,\mathbb{1}_{x\in\pi_*(\omega)}\Big]\\ &\leqslant \sum_x \mathbb{E}\Big[\tilde{\omega}_x^2\,;\,\omega_x \leqslant \tilde{\omega}_x\,,\,\mathbb{1}_{x\in\pi_*(\omega)}\Big]\\ &\leqslant \sum_x \mathbb{E}\Big[\tilde{\omega}_x^2\,;\,\mathbb{1}_{x\in\pi_*(\omega)}\Big]\\ &= \sum_x \mathbb{E}\Big[\tilde{\omega}_x^2\Big]\,\mathbb{E}\Big[\mathbb{1}_{x\in\pi_*(\omega)}\Big],\end{aligned}$$

where in the last we used the independence between ω and $\tilde{\omega}_x$. Noting also that $\mathbb{E}\Big[\tilde{\omega}_x^2\Big]$ is independent of x, we obtain the bound

$$\mathbb{V}\mathrm{ar}(\tau_n) \leqslant \mathbb{E}\big[\omega_\cdot^2\big]\,\mathbb{E}\Big[\sum_x \mathbb{1}_{x\in\pi_*(\omega)}\Big] = 2N\,\mathbb{E}\big[\omega_\cdot^2\big]$$

since the total length of the path is $2N$. □

We should remark that this bound is far from optimal. In dimension two the predicted order of the variance is $\mathbb{V}\mathrm{ar}(\tau_n) \approx n^{2/3}$. This is based on prediction emerging from the Kardar-Parisi-Zhang universality. The above proposition is due to Kesten [K93] and even though far from what expected it remained for very long time the best bound until the improvement of Benjamini-Kalai-Schramm [BKS03], where it was tinily but importantly improved, in the case of Bernoulli variables, to $N/\log N$ making use of Talagrand's improved Poincaré inequalities and averaging ideas. The extension of this bound to variables with general distributions was done by Benaim and Rossignol [BR08].

1.3. The notion of influence. In both the Lindeberg theorem and the Efron-Stein inequality an important feature was the "influence" that a single variable has on the overall random function. In other words, "how much" does the random function change if we change, e.g. by resampling, one of its (random) variables.

This motivates the need of putting the notion of *influence* in a mathematical context. We thus define

Definition 1.4. *Let $(\omega_x)_{x\in\mathsf{S}}$ be a family of i.i.d. real valued variables indexed by a countable set S and $f\colon \mathbb{R}^{\mathsf{S}} \to \mathbb{R}$ a function of this family of variables. The influence of entry $x \in \mathsf{S}$ is defined as*

$$\mathrm{Inf}_x(f) := \mathbb{E}\Big[\mathbb{V}\mathrm{ar}(f(\omega)\,\big|\,\{\omega_y\}_{y\neq x})\Big].$$

Let us look at some examples:

- **Central Limit Theorem.** In the CLT we consider a function
$$f(\omega_1,...,\omega_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_i.$$
The influence of entry $a \in \{1,...,n\}$ is easily computed to be $1/n$. Thus, each entry has an asymptotically negligible influence. This is consistent with the idea of Lindeberg, which says that the CLT should hold as long as all entries have negligible contribution.
- **Multilinear polynomials.** Multilinear polynomials will play an important role later on. Multilinear polynomials also go under the name *discrete chaos* due to their continuum counterpart called Wiener chaos. For a family of i.i.d. random variables $(\omega_x)_{x\in\mathsf{S}}$ indexed by a countable set S and a family of coefficients c_I, $I \subset \mathsf{S}$, we define multilinear polynomials as
$$f(\omega) = \sum_{I\subset\mathsf{S}} \mathsf{c}_I \prod_{x\in I}\omega_x,$$
where the sum is over all finite subsets of S and were we understand that all terms in the product are taken to be different. To compute the influence of an entry $y \in \mathsf{S}$, let us assume that the random variables have mean zero and variance one. We then write
$$f(\omega) = \omega_y \sum_{I\ni y}\mathsf{c}_I \prod_{x\in I, x\neq y}\omega_x + \sum_{I\not\ni y}\mathsf{c}_I\prod_{x\in I}\omega_x.$$
Since the second term does not depend on ω_y and the random variables are independent with mean zero and variance one, we can easily see that
$$\mathbb{V}\mathrm{ar}(f(\omega)\,\big|\,\{\omega_x\}_{x\neq y}) = \Big(\sum_{I\ni y}\mathsf{c}_I \prod_{x\in I, x\neq y}\omega_x\Big)^2.$$
and thus that
$$\mathrm{Inf}_y(f) := \mathbb{E}\Big[\mathbb{V}\mathrm{ar}(f(\omega)\,\big|\,\{\omega_x\}_{x\neq y})\Big] = \sum_{I\ni y}\mathsf{c}_I^2.$$
Actually, in the case of multilinear polynomials we can provide another expression for influence by noticing that
$$\sum_{I\ni y}\mathsf{c}_I^2 = \mathbb{V}\mathrm{ar}\Big(\frac{\partial f(\omega)}{\partial\omega_y}\Big).$$

This expression can also be recast in the format

$$\mathrm{Inf}_y(f) = \mathbb{V}\mathrm{ar}\Big(\frac{\partial f(\omega)}{\partial \omega_y}\Big),$$

which offers a more intuitive perspective on the idea of "how much the function f depends on its y-entry".

Let us at this point notice that the Efron-Stein inequality can also be written in terms of influences as

$$\mathbb{V}\mathrm{ar}\Big(f(\omega)\Big) \leqslant \sum_{a\in \mathsf{S}} \mathrm{Inf}_a(f).$$

To see this, we use the fact that if $X, \tilde{X}$ are two independent, identically distributed random variables, then for any measurable function f it holds that

$$\mathbb{V}\mathrm{ar}(f(X)) = \frac{1}{2}\mathbb{E}\Big[\big(f(X) - f(\tilde{X})\big)^2\Big].$$

and then we write that

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\Big(f(\omega)\Big) &\leqslant \sum_{a\in\mathsf{S}} \frac{1}{2}\mathbb{E}\Big[\big(f(\omega) - f(\mathsf{T}_a\,\omega)\big)^2\Big] \\
&= \sum_{a\in\mathsf{S}} \frac{1}{2}\mathbb{E}\bigg[\;\mathbb{E}\Big[\big(f(\omega) - f(\mathsf{T}_a\,\omega)\big)^2 \,\Big|\, \{\omega_x\}_{x\neq a}\Big]\bigg] \\
&= \sum_{a\in\mathsf{S}} \mathbb{E}\bigg[\;\mathbb{V}\mathrm{ar}\Big(f(\omega)\,\Big|\,\{\omega_x\}_{x\neq a}\Big)\bigg] \\
&= \sum_{a\in\mathsf{S}} \mathrm{Inf}_a(f).
\end{aligned}$$

1.4. Multilinear polynomials and hypercontractivity. Let us formally define multilinear polynomials as follows. Consider a family of i.i.d. random variables $\xi := (\xi_x)_{x\in\mathsf{S}}$ indexed by a countable set S Let $\mathcal{P}^{\mathrm{fin}}(\mathsf{S}) := \{I \subset \mathsf{S} \colon |I| < \infty\}$, the set of all finite subsets of S. Consider a (multi-index) function $\psi \colon \mathcal{P}^{\mathrm{fin}}(\mathsf{S}) \to \mathbb{R}$. Then a multilinear polynomial of disorder ξ, associated to ψ is defined as

$$\Psi(\xi) := \sum_{I\in\mathcal{P}^{\mathrm{fin}}(\mathsf{S})} \psi(I)\xi^I, \quad \text{where} \quad \xi^I := \prod_{a\in I}\xi_a, \quad \text{with} \quad \xi^{\varnothing} := 1. \tag{1.12}$$

Assuming that $\mathbb{E}[\xi_a] = 0$ and $\mathbb{V}\mathrm{ar}(\xi_a) = 1$, it is easy compute the variance of $\Psi(\xi)$ as

$$\mathbb{V}\mathrm{ar}(\Psi(\xi)) = \boldsymbol{\sigma}_\Psi^2 := \sum_{I\in\mathcal{P}^{\mathrm{fin}}(\mathsf{S}), I\neq\varnothing} \psi(I)^2.$$

As we have already said, the influence of entry $a \in \mathsf{S}$ in this case equals

$$\mathrm{Inf}_a\big(\Psi\big) = \sum_{I\ni a} \psi(I)^2.$$

Let us now discuss the notion of **hypercontractivity**. More details on hypercontractivity can be found in [S98] and for hypercontractivity on Gaussian spaces [J97]. As we saw the variance of a multilinear polynomial can be easily computed. This is, of course, not the case for higher (especially non integer) moments, which may and will arise naturally. In this situation hypercontractivity comes very handy as it allows to estimate higher than two moments of a multilinear polynomial in terms of just the second moment. The following definition of hypercontractivity actually captures this useful property that multi-linear polynomials have.

Definition 1.5. *Let $\Psi(\xi) := \sum_{I\in\mathcal{P}^{\mathrm{fin}}(\mathsf{S})} \psi(I)\xi^I$ be a multi-linear polynomial of the family of random variables $\xi = (\xi_a)_{a\in\mathsf{S}}$.*

For $\varrho > 0$, define the operator T_ϱ acting on the multilinear polynomial as

$$\big(T_\varrho\Psi\big)(\xi) := \sum_{I\in\mathcal{P}^{\mathrm{fin}}(\mathsf{S})} \varrho^{|I|}\,\psi(I)\,\xi^I,$$

where $|I|$ denotes the cardinality of the set $|I|$. For $\varrho \geqslant 1$ and $1 \leqslant p \leqslant q < \infty$, we will say that the family ξ is $(p, q, \frac{1}{\varrho})-$hypercontractive if

$$\|\Psi\|_q \leqslant \|\,T_\varrho\Psi\,\|_p.$$

for **all** *multi-linear polynomials Ψ.*

The question now is to classify when hypercontractivity holds. We will be mostly interested in the case of $(2, q, \frac{1}{\varrho})$-hypercontractivity. In the simplest case of linear polynomials, that is of the form $a + X$ where X is a real valued random variable, the above definition of hypercontractivity can be recast as that a random variable X is $(p, q, \frac{1}{\varrho})$-hypercontractive if

$$\|a + X\|_q \leqslant \|a + \varrho X\|_p, \quad \text{for all} \quad a \in \mathbb{R}.$$

It is not difficult to see that for $q > 2$, a random variable X is $(2, q, \frac{1}{\varrho})$-hypercontractive if and only if $\|X\|_q < \infty$.

Bernoulli variables which take the value ± 1 with probability $1/2$ turn out to be $(2, q, (q-1)^{1/2})$-hypercontractive [B70, B75]. The hypercontractivity bound for Bernoullis can be used to derive a hypercontractivity bound for general random variables with finite $q \geqslant 2$ moment. This is the content of the next proposition which is proved in [S98] Proposition 2.20 and [MOO10] Proposition 3.16.

Proposition 1.6. *Let X be a mean zero random variable with finite q^{th}-moment with $q \geqslant 2$. Then X is $(2, q, \frac{1}{\varrho_q})$-hypercontractive with $\varrho_q = 2(q-1)^{1/2}\ \|X\|_q/\|X\|_2$.*

Proof. The proof uses a symmetrisation trick, which then allows to obtain a hypercontractive estimate via the hypercontractivity of the Bernoulli variables. In particular, let us denote by $\tilde{X}$ an independent copy of the variable X and it symmetrised version $Y := X - \tilde{X}$. Since Y is symmetric, we have that Y has the same distribution as εY, where ε is a symmetric, Bernoulli ± 1 random variable.

The goal will be to show that for any number a it holds that

$$\big\|a + \tfrac{1}{\varrho_q}X\big\|_{L^q} \leqslant \big\|a + X\big\|_{L^2}.$$

Let us start from the left hand side and get by Jensen inequality that

$$\big\|a + \tfrac{1}{\varrho_q}X\big\|_{L^q} \leqslant \big\|a + \tfrac{1}{\varrho_q}Y\big\|_{L^q}. \tag{1.13}$$

This is not difficult to see since

$$\begin{aligned}\big\|a+\tfrac{1}{\varrho_q}Y\big\|_{L^q} &= \Big(\int\int \big|a+\tfrac{1}{\varrho_q}(X-\tilde X)\big|^q\, P(\mathrm{d}X)P(\mathrm{d}\tilde X)\Big)^{\frac1q}\\ &\geqslant \Big(\int\Big|\int \big(a+\tfrac{1}{\varrho_q}(X-\tilde X)\big)\, P(\mathrm{d}\tilde X)\Big|^q\, P(\mathrm{d}X)\,\Big)^{\frac1q}\\ &= \Big(\int \big|a+\tfrac{1}{\varrho_q}X\big|^q\, P(\mathrm{d}X)\,\Big)^{\frac1q}\\ &= \big\|a+\tfrac{1}{\varrho_q}X\big\|_{L^q},\end{aligned}$$

where in the second equality we used the fact that $\tilde X$ has mean zero. Continuing now from (1.13) and using the equality in distribution between Y and εY, followed by the $(2,q,(q-1)^{1/2})$ Bernoulli hypercontractivity, we have that

$$\begin{aligned}\big\|a+\tfrac{1}{\varrho_q}X\big\|_{L^q} &\leqslant \big\|a+\tfrac{1}{\varrho_q}Y\big\|_{L^q} = \big\|a+\tfrac{1}{\varrho_q}\varepsilon Y\big\|_{L^q(\mathrm{d}\varepsilon\,\mathrm{d}Y)}\\ &\leqslant \Big\|\,\big\|a+\sqrt{q-1}\tfrac{1}{\varrho_q}\varepsilon Y\big\|_{L^2(\mathrm{d}\varepsilon)}\Big\|_{L^q(\mathrm{d}Y)}\\ &= \Big\|\,a^2+(q-1)\tfrac{1}{\varrho_q^2}Y^2\Big\|^{1/2}_{L^{q/2}(\mathrm{d}Y)}\\ &\leqslant \sqrt{a^2+(q-1)\tfrac{1}{\varrho_q^2}\|Y\|^2_{L^q}},\end{aligned}$$

and inserting the value of $\varrho_q = 2(q-1)^{1/2}\|X\|_{L^q}/\|X\|_{L^2}$ we have that the above equals

$$\Big\{a^2+\big(\frac{\|Y\|_{L^q}}{2\|X\|_{L^q}}\big)^2\cdot\|X\|^2_{L^2}\Big\}^{1/2} \leqslant \Big\{a^2+\|X\|^2_{L^2}\Big\}^{1/2} = \|a+X\|_{L^2},$$

where we use that by triangle inequality $\|Y\|_{L^2}\leqslant 2\|X\|_{L^q}$. □

We should note that the hypercontractivity constant that appears in Proposition 1.6 is not optimal since when $q\to 1$ the constant converges to 2, while one would expect it to converge to 1. In [CSZ18b] it was shown that the optimal $(2,q,\frac{1}{\varrho_q^*})$-hypercontractivity constant for a random variable with finite q-moment does indeed converge to 1 when $q\to 2$.

Having a hypercontractivity bound for a single random variable, the question now is whether this bound can be "*tensorized*" to cover the multi-linear case. Indeed, this is the case and this tensorization is the subject of the next proposition

Proposition 1.7. *Let $\xi=(\xi_a)_{a\in\mathsf{S}}, \zeta=(\zeta_a)_{a\in\mathsf{S}}$ be two families of $(p,q,\frac{1}{\varrho})$-hypercontractive families. Then the concatenated family $\xi\sqcup\zeta := \{\xi_a\}_{a\in\mathsf{S}}\bigcup\{\zeta_a\}_{a\in\mathsf{S}}$ is also $(p,q,\frac{1}{\varrho})$-hypercontractive.*

Proof. Let us consider the multilinear polynomial $\Psi(\xi\sqcup\zeta)$ on the concatenated family $\xi\sqcup\zeta$. We have

$$\big\|\big(T_{1/\varrho}\Psi\big)(\xi\sqcup\zeta)\big\|_q = \Big\|\sum_{I,J}\varrho^{-|I|-|J|}\,\psi(I\cup J)\,\xi^I\,\zeta^J\Big\|_q = \Big\|T_{1/\varrho}\sum_J\Big(\sum_I \varrho^{-|I|}\,\psi(I\cup J)\,\xi^I\Big)\zeta^J\Big\|_q$$

where the $L^q := L^q(\mathrm{d}\zeta\,\mathrm{d}\,xi)$ norm is with respect to the product measure of the joint law of (ξ,ζ). We will now consider successively the expectations, first with respect to the law of ζ

and use the $(p,q,\frac{1}{\varrho})$-hypercontractivity for the ζ variables and then with respect to the law of ξ and use the $(p,q,\frac{1}{\varrho})$-hypercontractivity for the ζ variables. In this fashion we have that

$$\begin{aligned}\big\|\big(T_{1/\varrho}\Psi\big)(\xi \sqcup \zeta)\big\|_{L^q(\mathrm{d}\zeta\,\mathrm{d}\xi)} &= \Big\|\, \big\| T_{1/\varrho}\sum_J \Big(\sum_I \varrho^{-|I|}\,\psi(I\cup J)\,\xi^I\Big)\zeta^J \big\|_{L^q(\mathrm{d}\zeta)}\Big\|_{L^q(\mathrm{d}\xi)} \\ &\leqslant \Big\|\, \big\| \sum_J \Big(\sum_I \varrho^{-|I|}\,\psi(I\cup J)\,\xi^I\Big)\zeta^J \big\|_{L^p(\mathrm{d}\zeta)}\Big\|_{L^q(\mathrm{d}\xi)}, \end{aligned} \tag{1.14}$$

where we used the hypercontractivity with respect to the ζ variables. We will now use Minkowski's inequality (to interchange the norms with respect to the ζ and ξ variables and thus facilitate the application of the hypercontractivity with respect to ξ). We recall that the Minkowski inequality is an integral version of the triangle inequality for L^p spaces. More precisely, if $\mathcal{X},\mathcal{Y}$ two measure spaces with measures $\mu(\mathrm{d}x),\nu(\mathrm{d}y)$, respectively, then for $p\geqslant 1$, it holds that

$$\Big(\int_{\mathcal{Y}}\Big|\int_{\mathcal{X}} F(x,y)\,\mu(\mathrm{d}x)\Big|^p\nu(\mathrm{d}y)\Big)^{\frac{1}{p}} \leqslant \int_{\mathcal{X}}\Big(\int_{\mathcal{Y}}\big|F(x,y)\big|^p\,\nu(\mathrm{d}y)\Big)^{\frac{1}{p}}\mu(\mathrm{d}x).$$

Applying the Minkowski inequality to (1.14) and noting that in there $q\geqslant p$ (thus allowing the application of Minkowski on the $L^{q/p}(\mathrm{d}\xi)$ space), we have that (1.14) is bounded by

$$\Big\|\,\big\|\sum_J\Big(\sum_I \varrho^{-|I|}\,\psi(I\cup J)\,\xi^I\Big)\zeta^J\big\|_{L^q(\mathrm{d}\xi)}\Big\|_{L^p(\mathrm{d}\zeta)} = \Big\|\,\big\|\sum_I \varrho^{-|I|}\Big(\sum_J \psi(I\cup J)\,\zeta^J\Big)\xi^I\big\|_{L^q(\mathrm{d}\xi)}\Big\|_{L^p(\mathrm{d}\zeta)}$$

Applying the hypercontractivity of the ξ variables, we bound the above by

$$\Big\|\,\big\|\sum_I\Big(\sum_J \psi(I\cup J)\,\zeta^J\Big)\xi^I\big\|_{L^p(\mathrm{d}\xi)}\Big\|_{L^p(\mathrm{d}\zeta)} = \Big\|\sum_{I,J}\psi(I\cup J)\,\zeta^J\,\xi^I\Big\|_{L^p(\mathrm{d}\zeta\,\mathrm{d}\xi)}$$

establishing that for any multilinear polynomial Ψ it holds that

$$\big\|\big(T_{1/\varrho}\Psi\big)(\xi\sqcup\zeta)\big\|_q \leqslant \Big\|\sum_{I,J}\psi(I\cup J)\,\xi^I\,\zeta^J\Big\|_p = \|\Psi(\xi\sqcup\zeta)\|_p$$

which is equivalent to the definition of $(p,q,\frac{1}{\varrho})$-hypercontractivity by just setting Ψ to be $T_\varrho\Psi$. □

1.5. Lindeberg theorem for multilinear polynomials.

We have the following theorem

Theorem 1.8. *Let $\zeta=(\zeta_a)_{a\in\mathsf{S}}$ and $\xi=(\xi_a)_{a\in\mathsf{S}}$ be two families of independent random variables with mean zero, variance one and uniformly integrable second moment. Let $\Psi(\xi),\Psi(\zeta)$ be associated multilinear polynomials as defined in* (1.12) *and assume that $\boldsymbol{\sigma}_\Psi^2 := \sum_{\varnothing\neq I\in\mathcal{P}^{\mathrm{fin}}(\mathsf{S})}\psi(I)^2$ is finite.*

Then for every $f\in C_b^3(\mathbb{R})$ and any $\varepsilon>0$, there exists C_ε depending not only on ε but also on $\|f'\|_\infty,\|f''\|_\infty,\|f'''\|_\infty$ and $\boldsymbol{\sigma}_\Psi^2$, such that

$$\Big|\mathbb{E}\big[f\big(\Psi(\xi)\big)\big]-\mathbb{E}\big[f\big(\Psi(\zeta)\big)\big]\Big| \leqslant \varepsilon + C_\varepsilon\sqrt{\max_{a\in\mathsf{S}}\mathrm{Inf}_a(\Psi)}. \tag{1.15}$$

The above theorem was proved in [CSZ17a] and it is a sharp improvement of a theorem in [MOO10], where in the latter the Lindeberg principle for multilinear polynomials was proved under the assumption of finite third moments. The above theorem captures an optimal, in terms of moments, condition. In [CSZ17a] a more quantitative expression of the right hand side on (1.15) was provided. Moreover, in [CSZ17a] a statement of the Lindeberg principle for non mean zero variables was also stated.

A direct consequence of the above theorem is that if one has a sequence of multi-linear functionals Ψ_n for which it holds that

$$\max_{a\in \mathsf{S}} \mathrm{Inf}_a(\Psi_n) \xrightarrow[n\to\infty]{} 0,$$

then the asymptotic distribution of $\Psi_n(\xi)$ and $\Psi_n(\zeta)$ are the same assuming that the families ξ and ζ have mean zero, variance one (these two conditions can be relaxed to matching first and second moments) and uniformly integrable second moments.

Proof of Theorem 1.8. The proof of this theorem starts with the same telescoping procedure as done in the proof of the standard Lindeberg theorem. Without loss of generality we will assume that the index set S is finite and for notational simplicity we identify it with $\{1,...,n\}$. More crucially, we will assume that Ψ has degree ℓ which stays bounded in n, that is

$$\Psi(\xi) = \sum_{I\subset \mathsf{S}\,,\,|I|\leqslant \ell} \psi(I)\,\xi^I.$$

This assumption can be justified by a simple truncation argument (exercise). For a function $f\in C_b^3(\mathbb{R})$ we denote

$$g(x_1,...,x_n) := f(\Psi(x)),$$

and

$$h^{X^j}_{n,j}(y) := g(\zeta_1,...,\zeta_{j-1},y,\xi_{j+1},...,\xi_n), \quad \text{with} \quad X^j := (\zeta_1,...,\zeta_{j-1},\xi_{j+1},...,\xi_n), \tag{1.16}$$

and we have that

$$f(\Psi(\xi)) - f(\Psi(\zeta)) = \sum_{j=1}^n \big(h^{X^j}_{n,j}(\xi_j) - h^{X^j}_{n,j}(\zeta_j)\big).$$

We now perform the same Taylor expansion as done in (1.6) and employ the matching moment assumption to have the estimate

$$\begin{aligned}\Big|\mathbb{E}[f(\Psi(\xi))] - \mathbb{E}[f(\Psi(\zeta))]\Big| &= \Bigg|\sum_{j=1}^N \mathbb{E}\Big[R^{X^j}_{n,j}(\xi_j) - R^{X^j}_{n,j}(\zeta_j)\Big]\Bigg| \\ &\leqslant \sum_{j=1}^N \mathbb{E}\Big[|R^{X^j}_{n,j}(\zeta_j)|\Big] + \sum_{j=1}^N \mathbb{E}\Big[|R^{X^j}_{n,j}(\xi_j)|\Big].\end{aligned}$$

where we recall that the error $R^x_{n,j}(y)$ of the Taylor expansion, which was also defined in (1.7), has the form

$$R^x_{n,j}(y) = \frac{1}{2}\int_0^y \big(\partial^3_y h^x_{n,i}(t)\big)(y-t)^2 \mathrm{d}t,$$

and satisfies the bound

$$|\,R^x_{n,j}(y)\,| \leqslant \min\Big\{\frac{1}{6}\|\partial^3_y h^x_{n,i}\|_\infty\,|y|^3\,,\,\|\partial^2_y h^x_{n,i}\|_\infty\,y^2\,\Big\}.$$

The derivatives of $h_j^x(\cdot)$ are computed as:

$$\begin{aligned}\left(\partial_y^m h_j^x\right)(y) &= f^{(m)}\Big(\Psi(x_1,\dots,x_{j-1},y,x_{j+1},\dots,x_N)\Big)\Big(\,\partial_y\Psi(x_1,\dots,x_{j-1},y,x_{j+1},\dots,x_N)\,\Big)^m\\ &= f^{(m)}\Big(\Psi(x_1,\dots,x_{j-1},y,x_{j+1},\dots,x_N)\Big)\,\Big(\sum_{I\ni j}\psi(I)x^{I\setminus\{j\}}\Big)^m.\end{aligned}$$

Defining

$$L_j(x) := \sum_{I\ni j}\psi(I)x^I,$$

we obtain the bound

$$\sum_{j=1}^N \mathbb{E}\Big[|R_{n,j}^{X^j}(\zeta_j)|\Big] \leqslant C_f\sum_{j=1}^N \mathbb{E}\Big[\varphi(L_j(X^j))\Big],\quad \text{with} \tag{1.17}$$

$$C_f = \max\{\|f'\|_\infty, \|f^{(2)}\|_\infty, \|f^{(3)}\|_\infty\}\quad\text{and}\quad \varphi(x) := \min\Big\{\frac{|x|^3}{6}, |x|^2\Big\}. \tag{1.18}$$

To proceed with a sharp estimate on (1.17) under only the assumption of uniformly integrable second moments, we need to truncate the random variables in a way that also respects an orthogonality. The general truncation is described as follows:

Truncation procedure : Fix $M\in(0,\infty)$. We can decompose any real-valued random variable Y with zero mean and finite variance as

$$Y = Y^- + Y^+\,, \tag{1.19}$$

where Y^-, Y^+ are functions of Y and possibly of some extra randomness, such that

$$\begin{aligned} &\mathbb{E}[Y^-] = \mathbb{E}[Y^+] = 0\,, && Y^-Y^+ = 0\,,\\ &|Y^-| \leqslant |Y|\,\mathbb{1}_{\{|Y|\leqslant M\}}\,, && \mathbb{E}[(Y^+)^2] \leqslant 2\,\mathbb{E}[Y^2\mathbb{1}_{\{|Y|>M\}}]\,.\end{aligned} \tag{1.20}$$

We postpone the proof of the truncation properties (1.20) until the end of the proof of this theorem. Assuming these properties, we proceed by denoting by X^{j-} the vector X^j from (1.16) with all its entries truncated as above and also $X^{j+} := X^j - X^{j-}$. Noting the elementary inequality for ϕ (defined in (1.18))

$$\varphi(a+b) \leqslant 2a^2 + \frac{4}{3}|b|^3,\qquad \text{for real } a, b,$$

we have that the bound in (1.17) can be extended to

$$\mathbb{E}\big[\varphi(L_j(X^j))\big] \leqslant 4\,\mathbb{E}\Big[\big(L_j(X^j) - L_j(X^{j-})\big)^2\Big] + \frac{4}{3}\,\mathbb{E}\big[|L_j(X^{j-})|^3\big]. \tag{1.21}$$

Estimate on the first term in (1.21): To estimate the first term in (1.21) we write

$$\begin{aligned}L_j(X^j) - L_j(X^{j-}) &= \sum_{I\ni j}\psi(I)\left(X^{j-}+X^{j+}\right)^I - \sum_{I\ni j}\psi(I)\left(X^{j-}\right)^I\\ &= \sum_{I\ni j}\psi(I)\sum_{\Gamma\subseteq I,\,|\Gamma|\geqslant 1}(X^{j+})^\Gamma(X^{j-})^{I\setminus\Gamma}\,,\end{aligned}$$

where the second equality comes from a simple binomial expansion of the first term and a cancellation with the second one. By (1.20) the random variables $X_1^{j-}, X_1^{j+}, X_2^{j-}, X_2^{j+}, \dots$

are orthogonal. Setting $\sigma_{\pm,i}^2 := \mathbb{E}[(X_i^{j\pm})^2]$ and observing that $\sigma_{-,i}^2 + \sigma_{+,i}^2 = \mathbb{V}\mathrm{ar}(X_i^j) = 1$, we obtain

$$\begin{aligned}\mathbb{E}\big[\big(L_j(X^j) - L_j(X^{j-})\big)^2\big] &= \sum_{I \ni j} \psi(I)^2 \sum_{\Gamma \subseteq I,\, |\Gamma| \geqslant 1} (\sigma_+^2)^\Gamma (\sigma_-^2)^{I \setminus \Gamma} \\ &= \sum_{I \ni j} \psi(I)^2 \big(1 - (\sigma_-^2)^I\big) \leqslant \sum_{I \ni j} \psi(I)^2 \big(1 - (1 - \overline{\sigma}_+^2)^{|I|}\big)\,,\end{aligned} \tag{1.22}$$

where

$$\overline{\sigma}_+^2 := \max_{i=1,\ldots,N} \sigma_{+,i}^2 = \max_{i=1,\ldots,n} \mathbb{E}[(X_i^{j+})^2] \leqslant 2 \max_{i=1,\ldots,n} \mathbb{E}[(X_i^j)^2\, \mathbb{1}_{\{|X_i^j| > M\}}] \leqslant 2\, \boldsymbol{m}_2^{>M}\,,$$

having used (1.20) and having defined

$$\boldsymbol{m}_2^{>M} := \sup_{X \in \{\zeta_i, \xi_i\}_{i \geqslant 1}} \mathbb{E}\big[X^2 \mathbb{1}_{|X| \geqslant M}\big].$$

Using the estimate $1 - (1 - \overline{\sigma}_+^2)^{|I|} \leqslant |I|\, \overline{\sigma}_+^2$ in (1.22) we have that

$$\begin{aligned}\sum_{j=1}^n \mathbb{E}\big[\big(L_j(X^j) - L_j(X^{j-})\big)^2\big] \;&\leqslant\; 2\, \boldsymbol{m}_2^{>M} \sum_j \left(\sum_{I \ni j} |I|\, \psi(I)^2 \right) \\ &\leqslant 2\, \boldsymbol{m}_2^{>M}\, \ell^2 \sum_I \psi(I)^2\,,\end{aligned}$$

where we recall that ℓ is the degree of Ψ.

Estimate on the second term in (1.21)**:** For the second term we will use hypercontractivity bound from Propositions 1.6 and 1.7. This implies that

$$\|L_j(X^{j-})\|_3 \leqslant (B_3)^\ell\, \|L_j(X^{j-})\|_2\,, \tag{1.23}$$

where

$$B_3 := 2\sqrt{2} \max_{i \leqslant n} \frac{\|X_i^{j-}\|_3}{\|X_i^{j-}\|_2}\,.$$

Since for every i we have that $|X_i^{j-}| \leqslant |X_i^j|\, \mathbb{1}_{|X_i^{j-}| \leqslant M}$, by (1.20), we have

$$\|X_i^{j-}\|_3 \leqslant \mathbb{E}\big[\, |X_i^j|^3\, \mathbb{1}_{\{|X_i^j| \leqslant M\}}\,\big]^{1/3} \leqslant \big(\boldsymbol{m}_3^{\leqslant M}\big)^{1/3}\,,$$

with $\boldsymbol{m}_3^{\leqslant M}$ being the maximum truncated third moment of variables $\xi_i, \zeta_i,\ i \geqslant 1$. On the other hand, again by (1.20), we have that for every i

$$\begin{aligned}\|X_i^{j-}\|_2^2 &= \|X_i^j\|_2^2 - \|X_i^{j+}\|_2^2 = \mathbb{E}[(X_i^j)^2] - \mathbb{E}[(X_i^{j+})^2] \geqslant \mathbb{E}[(X_i^j)^2] - 2\mathbb{E}[(X_i^j)^2 \mathbb{1}_{\{|X_i^j| > M\}}] \\ &= 1 - 2\, \mathbb{E}[(X_i^j)^2 \mathbb{1}_{\{|X_i^j| > M\}}] \geqslant 1 - 2\boldsymbol{m}_2^{>M}\,,\end{aligned}$$

hence

$$B_3 \leqslant 2\sqrt{2} \frac{\big(\boldsymbol{m}_3^{\leqslant M}\big)^{1/3}}{\sqrt{1 - 2\boldsymbol{m}_2^{>M}}} \leqslant 4 \big(\boldsymbol{m}_3^{\leqslant M}\big)^{1/3}\,,$$

provided $\boldsymbol{m}_{>M}^{(2)} \leqslant \frac{1}{4}$, which can be achieved by choosing M large enough, thanks to the uniform integrability of the second moment. Therefore, (1.23) yields

$$\mathbb{E}\left[|L_j(X^{j-})|^3\right] \leqslant 64^\ell\, \big(\boldsymbol{m}_3^{\leqslant M}\big)^\ell\, \mathbb{E}\left[L_j(X^{j-})^2\right]^{3/2}\,.$$

Note that, since $\mathbb{E}[(X_i^{j-})^2] \leqslant \mathbb{E}[(X_i^j)^2] = 1$, we have

$$\mathbb{E}\left[L_j(X^{j-})^2\right] = \sum_{I \ni j} \psi(I)^2 \prod_{i \in I} \mathbb{E}[(X_i^{j-})^2] \leqslant \sum_{I \ni j} \psi(I)^2 = \mathrm{Inf}_j[\Psi] \,.$$

Therefore

$$\begin{aligned}\sum_{j=1}^N \mathbb{E}\left[|L_j(X^{j-})|^3\right] &\leqslant 64^\ell \; (\boldsymbol{m}_3^{\leqslant M})^\ell \left(\max_i \sqrt{\mathrm{Inf}_i[\Psi]}\right) \sum_j \sum_{I \ni j} \psi(I)^2 \\ &\leqslant \ell \, 64^\ell \, (\boldsymbol{m}_3^{\leqslant M})^\ell \; \left(\max_i \sqrt{\mathrm{Inf}_i[\Psi]}\right) \sum_{|I| \leqslant \ell} \psi(I)^2 \,.\end{aligned} \tag{1.24}$$

Proof of truncation properties (1.20)**.** Let $M > 0$. If $\mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y \leqslant M\}}] = 0$ we are done: just choose $Y^- := Y \mathbb{1}_{\{-M \leqslant Y \leqslant M\}}$ and $Y^+ := Y - Y^-$. If, on the other hand, $\mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y \leqslant M\}}] > 0$ (the strictly negative case is analogous), we set

$$\overline{T} := \sup\{T \in [0, M] : \; \mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y \leqslant T\}}] \leqslant 0\} \in [0, M] \,.$$

Note that $\mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y \leqslant \overline{T}\}}] \geqslant 0$, because $T \mapsto \mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y \leqslant T\}}]$ is right-continuous. If $\mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y \leqslant \overline{T}\}}] = 0$, defining $Y^- := Y \mathbb{1}_{\{-M \leqslant Y \leqslant \overline{T}\}}$ and $Y^+ := Y - Y^-$, all the properties in (1.20) are clearly satisfied, except the last one that will be checked below. Finally, we consider the case $\mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y \leqslant \overline{T}\}}] > 0$ (then necessarily $\overline{T} > 0$). Since $\mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y < \overline{T}\}}] \leqslant 0$ by definition of $\overline{T}$, we must have $\mathbb{P}(Y = \overline{T}) > 0$. Then take a random variable U uniformly distributed in $(0, 1)$ and independent of Y, and define

$$Y^- := Y\big(\mathbb{1}_{\{-M \leqslant Y < \overline{T}\}} + \mathbb{1}_{\{Y = \overline{T}, U \leqslant \varrho\}}\big) \,, \qquad \text{where} \qquad \varrho := \frac{-\mathbb{E}[Y \mathbb{1}_{\{-M \leqslant Y < \overline{T}\}}]}{\overline{T} \, \mathbb{P}(Y = \overline{T})} \in (0, 1) \,.$$

Setting $Y^+ := Y - Y^-$, all the properties (1.20) but the last one are clearly satisfied.

For the last property, we write

$$\mathbb{E}[(Y^+)^2] = \mathbb{E}[(Y^+)^2 \mathbb{1}_{\{|Y| > M\}}] + \mathbb{E}[(Y^+)^2 \mathbb{1}_{\{|Y| \leqslant M\}}] = \mathbb{E}[Y^2 \mathbb{1}_{\{|Y| > M\}}] + \mathbb{E}[(Y^+)^2 \mathbb{1}_{\{|Y| \leqslant M\}}] \,,$$

because $Y^+ = Y$ on the event $\{|Y| > M\}$. For the second term, since $0 \leqslant Y^+ \leqslant M$ on the event $\{|Y| \leqslant M\}$, we can write $(Y^+)^2 \leqslant M Y^+$ (no absolute value needed). Since $Y^- = Y^- \mathbb{1}_{\{|Y| \leqslant M\}}$ has zero mean by (1.20), we obtain

$$\begin{aligned}\mathbb{E}[(Y^+)^2 \mathbb{1}_{\{|Y| \leqslant M\}}] &\leqslant M \, \mathbb{E}[Y^+ \mathbb{1}_{\{|Y| \leqslant M\}}] = M \, \mathbb{E}[(Y^+ + Y^-) \mathbb{1}_{\{|Y| \leqslant M\}}] \\ &= M \, \mathbb{E}[Y \mathbb{1}_{\{|Y| \leqslant M\}}] = M \, (-\mathbb{E}[Y \mathbb{1}_{\{|Y| > M\}}]) \leqslant \mathbb{E}[Y^2 \mathbb{1}_{\{|Y| > M\}}] \,,\end{aligned}$$

where we have used the fact that $\mathbb{E}[Y] = 0$ by assumption, and Markov's inequality. The last relation in (1.20) is proved. □

2. Fourth moment theorems

A *fourth moment theorem* roughly says that certain projections of square integrable functions of many independent variables converge to normal distribution if their normalised fourth moment converges to 3. An important such class of function are multi-linear functions of independent variables.

2.1. Fourth moment theorems in the 20^{th} century. Fourth moment theorem turn out to have quite a long history dating back to the 1960s, although it seems that they were rediscovered in the 21^{st} century. For quadratic forms of gaussian variables it seems that a fourth moment theorem dates back to Sevast'yanov [S61] and then extended by Rotar' in [R74] to the case of i.i.d. variables.

Proposition 2.1. *Let $\xi = (\xi_i)_{i\geqslant 1}$ be a family of i.i.d. standard normal random variables and consider the quadratic polynomial $\Psi_n(\xi) = \sum_{1\leqslant i,j\leqslant n} \psi_n(i,j)\xi_i\xi_j$, associated to a symmetric kernel ψ_n. The multilinearity assumption forces the condition $\psi(i,i) = 0$ for all i. Denote by $\sigma_n^2 := \mathbb{V}\mathrm{ar}\big(\Psi(\xi)\big)$. If*

$$\frac{1}{\sigma_n^4}\mathbb{E}\Big[\big(\Psi_n(\xi)\big)^4\Big] \xrightarrow[n\to\infty]{} 3, \tag{2.1}$$

then $\frac{1}{\sigma_n}\Psi_n$ converges in distribution to a standard normal.

Proof. We will denote by $\boldsymbol{m}_{2r} = (2r-1)!!$ the $(2r)^{th}$ moment of a standard normal. Let us for simplicity also denote by ξ to be just the vector $(\xi_1, ..., \xi_n)$ and by $\boldsymbol{\Psi}_n$ the matrix $\boldsymbol{\Psi}_n := \big(\psi_n(i,j)\big)_{1\leqslant i,j,\leqslant n}$. We can then write the sum $\sum_{1\leqslant i,j\leqslant n} \psi_n(i,j)\xi_i\xi_j$ in the matrix form $\xi\boldsymbol{\Psi}_n\xi^{\mathsf{T}}$. Since $\boldsymbol{\Psi}_n$ is symmetric it can be diagonalised as $\boldsymbol{\Psi}_n = U_n \,\mathrm{diag}(\mu_1, ..., \mu_n)\, U_n^{\mathsf{T}}$, where $\mathrm{diag}(\mu_1, ..., \mu_n)$ is the diagonal matrix with entries the eigenvalues of $\boldsymbol{\Psi}_n$ and U_n is the associated orthogonal matrix. We then have that

$$\xi\, \boldsymbol{\Psi}_n\, \xi^{\mathsf{T}} = \xi U_n \,\mathrm{diag}(\mu_1, ..., \mu_n)\, U_n^{\mathsf{T}}\xi^{\mathsf{T}} = \big(\xi U_n\big)\,\mathrm{diag}(\mu_1, ..., \mu_n)\,\big(\xi U_n\big)^{\mathsf{T}}$$

Since ξ is a standard normal vector and U_n is orthogonal, the vector ξU_n is also distributed as a standard normal vector, which we denote by $(Y_1, ..., Y_n)$ and then we have that

$$\xi\, \boldsymbol{\Psi}_n\, \xi^{\mathsf{T}} =^d \sum_{i=1}^n \mu_i Y_i^2.$$

This is now in the standard central limit theorem form of a sum of independent random variables. Lindeberg's condition in Theorem (1.1) is implied by the condition

$$\frac{1}{\sigma_n^4}\sum_{i=1}^n \mu_i^4 \xrightarrow[n\to\infty]{} 0. \tag{2.2}$$

We will now check that the latter condition is satisfied if (2.1) holds. To this end, we first compute the variance

$$\sigma_n^2 = \mathbb{V}\mathrm{ar}\big(\sum_{i=1}^n \mu_i Y_i^2\big) = \sum_{i=1}^n \mu_i^2\, \mathbb{V}\mathrm{ar}(Y_i^2) = \sum_{i=1}^n \mu_i^2\big(\boldsymbol{m}_4 - 1\big) = 2\sum_{i=1}^n \mu_i^2.$$

and the fourth moment

$$\mathbb{E}\Big[\big(\xi\, \boldsymbol{\Psi}_n\, \xi^{\mathsf{T}}\big)^4\Big] = \mathbb{E}\Big[\big(\sum_{i=1}^n \mu_i Y_i^2\big)^4\Big] = \sum_{1\leqslant i,j,k,\ell\leqslant n} \mu_i\mu_j\mu_k\mu_\ell\, \mathbb{E}\big[Y_i^2\, Y_j^2\, Y_k^2\, Y_\ell^2\big]$$

which by considering all possible matchings of indices i, j, k, ℓ is written as

$$\mathbb{E}\Big[\big(\xi\,\boldsymbol{\Psi}_n\,\xi^{\mathsf{T}}\big)^4\Big] = \boldsymbol{m}_8\sum_i \mu_i^4 + 3\boldsymbol{m}_4^2\sum_{i\neq j}\mu_i^2\mu_j^2 + 6\boldsymbol{m}_4\sum_{i\neq j\neq k\neq i}\mu_i\mu_j\mu_k^2 \qquad (2.3)$$
$$+ 4\boldsymbol{m}_6\boldsymbol{m}_2\sum_{i\neq j}\mu_i^3\mu_j + \boldsymbol{m}_2^4\sum_{i\neq j\neq k\neq \ell\neq i}\mu_i\mu_j\mu_k\mu_\ell$$

We will use the fact that $\psi(i,i) = 0$ implies that the trace of matrix $\boldsymbol{\Psi}_n$ is zero and thus $\sum_{i=1}^n \mu_i = 0$. Thus,

$$\sum_{i\neq j\neq k\neq \ell\neq i}\mu_i\mu_j\mu_k\mu_\ell = \big(\sum_i \mu_i\big)^4 - \sum_i \mu_i^4 - 3\sum_{i\neq j}\mu_i^2\mu_j^2 - 6\sum_{i\neq j\neq k\neq i}\mu_i\mu_j\mu_k^2 - 4\sum_{i\neq j}\mu_i^3\mu_j$$
$$= -\sum_i \mu_i^4 - 3\sum_{i\neq j}\mu_i^2\mu_j^2 - 6\sum_{i\neq j\neq k\neq i}\mu_i\mu_j\mu_k^2 - 3\sum_{i\neq j}\mu_i^3\mu_j$$

and inserting into (2.3) we have that

$$\mathbb{E}\Big[\big(\xi\,\boldsymbol{\Psi}_n\,\xi^{\mathsf{T}}\big)^4\Big] = (\boldsymbol{m}_8 - 1)\sum_i \mu_i^4 + 3(\boldsymbol{m}_4^2 - 1)\sum_{i\neq j}\mu_i^2\mu_j^2 + 6(\boldsymbol{m}_4 - 1)\sum_{i\neq j\neq k\neq i}\mu_i\mu_j\mu_k^2$$
$$+ 4(\boldsymbol{m}_6\boldsymbol{m}_2 - 1)\sum_{i\neq j}\mu_i^3\mu_j$$

Removing the inequality constraints in the above summations, we can write (2.3) as

$$\mathbb{E}\Big[\big(\xi\,\boldsymbol{\Psi}_n\,\xi^{\mathsf{T}}\big)^4\Big] = (\boldsymbol{m}_8 - 1)\sum_i \mu_i^4 + 3(\boldsymbol{m}_4^2 - 1)\Big\{\big(\sum_i \mu_i^2\big)^2 - \sum_i \mu_i^4\Big\}$$
$$+ 6(\boldsymbol{m}_4 - 1)\Big\{(\sum_i \mu_i)^2\sum_k \mu_k^2 - 2\sum_i \mu_i\sum_k \mu_k^2 + 2\sum_i \mu_i^4 - \Big(\sum_i \mu_i^2\Big)^2\Big\}$$
$$+ 4(\boldsymbol{m}_6\boldsymbol{m}_2 - 1)\Big\{\sum_i \mu_i^3\sum_j \mu_j - \sum_i \mu_i^4\Big\}$$

and using again the fact that $\sum_i \mu_i = 0$ we have that for a specific constant $\mathsf{C} \neq 0$ (which can be explicitly computed)

$$\mathbb{E}\Big[\big(\xi\,\boldsymbol{\Psi}_n\,\xi^{\mathsf{T}}\big)^4\Big] = \mathsf{C}\sum_i \mu_i^4 + 3\,\big(2\sum_i \mu_i^2\big)^2 = \mathsf{C}\sum_i \mu_i^4 + 3\cdot\sigma_n^4$$

Dividing both sides by σ_n^2 we have that

$$\frac{\mathsf{C}}{\sigma_n^4}\sum_i \mu_i^4 = \frac{1}{\sigma_n^4}\mathbb{E}\Big[\big(\xi\,\boldsymbol{\Psi}_n\,\xi^{\mathsf{T}}\big)^4\Big] - 3,$$

which converges to zero, as $n\to\infty$, by (2.1) and thus Lindeberg's condition (2.2) is satisfied. □

The following theorem fourth moment theorem for *generalized quadratic forms* was proved by de Jong [dJ87]

Theorem 2.2. *Let $(\omega_i)_{i\geqslant 1}$ be a family of independent, real variables and $w_{i,j}^{(n)}(\cdot,\cdot)$ be a family of Borel measurable functions on $\mathbb{R}^2$ and consider the generalized quadratic form*

$$W_n(\omega) = \sum_{1\leqslant i,j\leqslant n} w_{i,j}^{(n)}(\omega_i,\omega_j) = \sum_{1\leqslant i\leqslant j\leqslant n} W_{i,j}, \qquad (2.4)$$

where, denoting by $W_{i,j} := w^{(n)}_{i,j}(\omega_i,\omega_j) + w^{(n)}_{j,i}(\omega_j,\omega_i)$*, we assume that*

$$\mathbb{E}\big[W_{i,j}\,\big|\,\omega_i\big] = 0, \qquad a.s. \qquad \textit{for all} \qquad i,j \leqslant n. \tag{2.5}$$

Denote by σ_n^2 *the variance of* $W_n(\omega)$ *and by* $\sigma_{i,j}^2$ *the variance of* $W_{i,j}$*. Then, the assumptions:*

$$\text{(a)} \qquad \frac{1}{\sigma_n^2}\max_{1\leqslant i\leqslant n}\sum_{1\leqslant j\leqslant n}\sigma_{i,j}^2 \xrightarrow[n\to\infty]{} 0, \tag{2.6}$$

$$\text{(b)} \qquad \frac{1}{\sigma_n^4}\mathbb{E}\big[\,W_n(\omega)^4\big] \xrightarrow[n\to\infty]{} 3, \tag{2.7}$$

imply that $\frac{1}{\sigma_n}W_n(\omega)$ *converges in distribution to a standard normal random variable.*

Let us make some remarks on the assumptions of the theorem.

- Assumption (2.5) is satisfied in the multilinear case when ω_i have mean zero.
- Assumption (2.6) is sort of a Lindeberg "uniform smallness" assumption. It says that no row in the array $(W_{i,j})_{1\leqslant i,j\leqslant n}$ has a dominant contribution to the total variance. For example it rules out situations of the form $W_n(\omega) = \sum_{1<j\leqslant n}\omega_1\omega_j$, where clearly the asymptotic distribution is determined by ω_1.
- The necessity of the fourth moment condition is seen by situation like $\sum_{1\leqslant i<j\leqslant n}\omega_i\omega_j$ where even though the Lindeberg condition (2.6) is satisfied the asymptotic distribution is chi-square. This can be seen by writing

$$2\sum_{1\leqslant i<j\leqslant n}\omega_i\omega_j = \big(\sum_i \omega_i\big)^2 - \sum_i \omega_i^2 \approx \big(\text{Gaussian}\big)^2 + \text{ l.o.t.}$$

The proof of Theorem 2.2 is based on the central limit theorem for martingales:

Theorem 2.3 (Martingale central limit theorem). *Let* $(X_n)_{n\geqslant 0}$ *be a sum of martingale differences* $X_n = \sum_{i=1}^n Y_i$ *with respect to a filtration* $\{\mathcal{F}_n\colon n \geqslant 0\}$*, satisfying*

$$\mathbb{E}\big[\,|Y_i|^{2+2\delta}\big] < \infty, \qquad \textit{for some } \delta \in (0,1].$$

Let

$$\sigma_i^2 := \mathbb{E}\big[Y_i^2\,|\,\mathcal{F}_{i-1}\big] \quad \textit{and} \quad s_n^2 := \sum_{i=1}^n \mathbb{E}[\sigma_i^2].$$

Then, the central limit theorem holds for $\frac{1}{s_n}X_n$*, if the following two conditions hold:*

$$(a) \qquad \tfrac{1}{s_n^{2+2\delta}}\sum_{i=1}^n \mathbb{E}\big[\,|Y_i|^{2+2\delta}\,\big] \xrightarrow[n\to\infty]{} 0, \tag{2.8}$$

$$(b) \qquad \mathbb{E}\Big[\,\big|\tfrac{1}{s_n^2}\sum_{i=1}^n Y_i^2 - 1\big|^{1+\delta}\,\Big] \xrightarrow[n\to\infty]{} 0, \tag{2.9}$$

In [HB70] conditions (2.8) and (2.9) are actually translated to quantitative bounds on the speed of convergence to the normal disitribution.

The proof of the Fourth Moment Theorem 2.2 proceeds by writing the generalized bilinear form $W_n(\omega)$ as a sum of martingale differences as:

$$W_n(\omega) = \sum_{j=1}^{n} U_{j,n}, \quad \text{with} \quad U_{j,n} := \tfrac{1}{\sigma_n} \sum_{1 \leqslant i < j} W_{i,j}.$$

and then applying the Martingale Central Limit Theorem 2.3 with $Y_k := U_{k,n}$ and δ chosen to be equal to 1. The idea is similar to that of the proof of Proposition 2.1, except that it does not make sense to employ a diagonalisation argument since on the one hand we do not have a quadratic form and on the other hand if the random variables are not normal their distribution will change after and orthogonal transformation. However, we can still proceed by expanding the fourth moment of $W_n(\omega) = \sum_{k=1}^{n} U_{k,n}$ and after grouping appropriately terms together, it turns out that conditions (2.6) and (2.7) imply conditions (2.8) and (2.9).

The computations required for the proof of Theorem 2.2 in the generalized quadratic case are quite more complicated than that of Proposition 2.1 and we refer the reader to [dJ87]. However, if we reduce ourselves to the standard quadratic case, then we can use Lindeberg's principle for multi-linear polynomials as of Theorem 1.8 to reduce to the situation of a quadratic forms of normal variables and thus get ourselves into the context of Proposition 2.1.

Theorem 2.2 was extended by de Jong [dJ90] to the case of what is known as *homogeneous sums*. To define this notion, let us consider a family random variables $(\omega_i)_{i \geqslant q}$ and denote by $\mathcal{F}_I$ the σ-algebra generated by $\{\omega_i \colon i \in I\}$ for $I \subset \{1, \dots, n\}$. We consider a family of $\mathcal{F}_I$-measurable random variables $\{W_I^{(n)}\}_{I \subset \{1,\dots,n\}}$ with the properties (for notational convenience we will drop the superscript (n)):

$$\mathbb{E}[\, W_I \,] = 0, \qquad \mathbb{E}[\, W_I^2 \,] =: \sigma_I^2 < \infty, \qquad \mathbb{E}[\, W_I W_J \,] = 0 \quad \text{if } I \neq J$$

A d-homogeneous sum associated to the family (W_I) and random variables (ω_i) is defined as

$$W_n(\omega) := \sum_{\substack{I \subset \{1,\dots,n\} \\ |I| = d}} W_I.$$

The fourth moment theorem for homogeneous sums reads as

Theorem 2.4. *Let $W_n(\omega)$ be a sequence of homogeneous sums of fixed degree d and let $\sigma_n^2 := \mathbb{V}\mathrm{ar}(W_n(\omega))$ and for $I \subset \{1, \dots, n\}$ let $\sigma_I^2 := \mathbb{V}\mathrm{ar}(W_I)$. Suppose that*

$$(a) \qquad \frac{1}{\sigma_n^2} \max_i \sum_{I \ni i} \sigma_I^2 \xrightarrow[n \to \infty]{} 0,$$

$$(b) \qquad \frac{1}{\sigma_n^4} \mathbb{E}\big[\, W_n(\omega)^4 \,\big] \xrightarrow[n \to \infty]{} 3,$$

then $\frac{1}{\sigma_n} W_n(\omega)$ converges in distribution to a standard normal.

The proof of Theorem 2.4 follows the same spirit as that of the fourth moment for generalized quadratic forms: on writes the homogeneous sum as a sum of martingale differences and then expands the fourth moment and regroups the summands so that the assumptions of the Martingale central limit theorem are satisfied.

We would like to close this subsection by commenting on how the proof of the martingale central limit theorem goes about. It makes use of the useful idea of *embedding theorems*, whose origins go back to Skorokhod and often called *Skorokhod embedding theorems.* Theorem 2.3 makes in particular use of a version of the embedding theorem for martingale differences due to Strassen [S67] (the original Skorokhod embedding was related to i.i.d. variables):

Theorem 2.5. *Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables such that, for every $n \geqslant 1$, it holds that*

$$\mathbb{E}\big[X_n^2 \,\big|\, X_{n-1},...,X_1\big] < \infty, \qquad \textit{and} \qquad \mathbb{E}\big[X_n \,\big|\, X_{n-1},...,X_1\big] = 0, \quad \textit{a.s.}$$

Then there exists a Brownian motion $B(\cdot)$ and a sequence of nonegative stopping times with respect to the σ-algebra generated by the Brownian motion, such that a.s.

$$\sum_{i=1}^{n} X_i = B\Big(\sum_{i=1}^{n} T_i\Big).$$

Moreover, it holds a.s. that

$$\mathbb{E}\big[T_n \,|\, \mathcal{F}_{n-1}\,\big] = \mathbb{E}\big[X_n^2 \,|\, X_{n-1},...,X_1\,\big].$$

2.2. Hoeffding's decomposition. Homogeneous sums are important as they form a sort of Fourier basis for expansion of square integrable random variables. In particular, we have Hoeffding's expansion [H48]:

Theorem 2.6. *Let $(\omega_i)_{i\geqslant 1}$ a family of independent random variables and let W be a square integrable random variable measurable with respect to $\{\omega_a\}_{a\in \mathsf{S}}$ indexed by a finite set S. Then W can be decomposed uniquely as*

$$W = \sum_{I\subset \mathsf{S}} W_I,$$

where $\{W_I\}_{I\subset \mathsf{S}}$ is defined as above, in particular is $\mathcal{F}_I$-measurable,

$$\begin{aligned} \mathbb{E}[\,W_I \,|\, \mathcal{F}_J] &= 0, \qquad \textit{a.s.} \qquad \textit{unless } I \subset J, \\ \textit{and} \quad \mathbb{E}\big[\,W_I W_J\,\big] &= 0, \qquad \textit{if } I \neq J. \end{aligned}$$

Proof. We will use the notation

$$\mathbb{E}\big[W \,|\, I\,\big] := \mathbb{E}\big[W \,\big|\, \{\omega_a\}_{a\in I}\,\big] \quad \text{and} \quad W_I := \sum_{A\subset I} (-1)^{|I|-|A|}\, \mathbb{E}\big[\,W \,|\, A\,\big].$$

We can represent W as $\sum_{I\subset \mathsf{S}} W_I$. Indeed, we have that

$$\begin{aligned} \sum_{I\subset \mathsf{S}} W_I &= \sum_{I\subset \mathsf{S}} \sum_{A\subset I} (-1)^{|I|-|A|}\, \mathbb{E}\big[\,W \,|\, A\,\big] = \sum_{A\subset \mathsf{S}} \mathbb{E}\big[\,W \,|\, A\,\big] \sum_{I\supset A} (-1)^{|I|-|A|} \\ &= W + \sum_{A\subset \mathsf{S}\,,\, A\neq \mathsf{S}} \mathbb{E}\big[\,W \,|\, A\,\big] \sum_{I\supset A} (-1)^{|I|-|A|} \\ &= W + \sum_{A\subset \mathsf{S}\,,\, A\neq \mathsf{S}} \mathbb{E}\big[\,W \,|\, A\,\big](-1+1)^{|\mathsf{S}\setminus A|} \\ &= W. \end{aligned}$$

We will now show that if I is not a subset of J then $\mathbb{E}[\,W_I \,|\, \mathcal{F}_J] = 0$. To see this, let us denote by $C := I \cap J$ and assume that $I\backslash C \neq \varnothing$. Then

$$\mathbb{E}[\,W_I \,|\, \mathcal{F}_J] = \sum_{A\subset I} (-1)^{|I|-|A|}\; \mathbb{E}\Big[\,\mathbb{E}\big[\,W \,|\, A\,\big] \Big|\, J\,\Big] = \sum_{A\subset I} (-1)^{|I|-|A|}\; \mathbb{E}\big[\,W \,|\, A\cap C\,\big]. \tag{2.10}$$

The second equality is because if $A = A_1 \cup A_2$ with $A_1 \subset C$ and $A_2 \subset I\backslash C$, then the conditional expectation $\mathbb{E}\big[\,\cdot\,|\, A\big]$ will fix the variables in $A = A_1 \cup A_2$ and average over the rest, but then the expectation $\mathbb{E}\big[\,\cdot\,|\, J\big]$ will average over the variables which are not included in J and in particular it will average the variables in A_2. So the only fixed variables will be those in $A_1 = A \cap C$. We can continue writing (2.10) as

$$\sum_{A_1\subset C\,,\,A_2\subset I\backslash C} (-1)^{|I|-|A_1|-|A_2|}\mathbb{E}\big[\,W\,|\,A_1\big] = \sum_{A_1\subset C} (-1)^{|I|-|A_1|}\mathbb{E}\big[\,W\,|\,A_1\big] \sum_{A_2\subset I\backslash C} (-1)^{|A_2|},$$

and the last sum equals $(-1+1)^{|I\backslash C|} = 0$ if $I\backslash C \neq \varnothing$.

From this fact, the orthogonality relation

$$\mathbb{E}\big[\,W_I W_J\,\big] = 0, \qquad \text{if } I \neq J,$$

follows easily by conditioning over either I or J. Since $I \neq J$ we cannot have both $I \subset J$ and $J \subset I$ and so one of the two conditional expectations will be zero. □

2.3. Fourth Moment Theorems in the 21^{st} centrury. In the 21^{st} century the fourth moment theorem was rediscovered in the setting of Wiener chaos by Nualart and Peccati [NP05]. The method used in [NP05] was actually the continuous analogue of de Jong's proof, which boils down to writing the iterated Itô integral as a martingale and then employing the Dambis-Dumbins-Schwarz theorem (instead of Strassen's theorem 2.5, which says that a continuous martingale $(M_t)_{t>0}$ is a time change of a Brownian motion, i.e. it can be written as $M_t = B_{\langle M\rangle_t}$, for some Brownian motion, with $\langle M\rangle_t$ being the quadratic variation of M_t. We will not expand on this approach, which parallels very much the approach exposed earlier but we would rather sketch new approaches that the rediscovery of the fourth moment theorem motivated via use of Malliavin calculus and Stein's method.

Basics of Stein's method. Let us start with the basic principles of Stein's method, which is a quantitative and robust method to prove central limit theorems. The starting point of Stein's method is that the normal distribution on $\mathbb{R}$ is the *only* distribution that satisfies the equation

$$\mathbb{E}\big[Zf(Z)\big] = \mathbb{E}\big[f'(Z)\big], \tag{2.11}$$

for every $f \in C_b^1(\mathbb{R})$. The fact that the normal distribution satisfies the above equation is an easy consequence of integration by parts. The idea of integration by parts will actually play an important role in what follows.

Stein's equation is the equation that given a (bounded) function h, asks for an *a.e.* differentiable function f such that

$$h(x) - \mathbb{E}[h(Z)] = f'(x) - xf(x), \qquad \text{for } x \in \mathbb{R},$$

where Z is a standard normal random variable. The fact that this equation has a solution, together with certain bounds is the content of what is called *Stein's lemma*, which we will not present here. In practice, one is interested in choosing $h(x)$ to be the indicator function $\mathbb{1}_{(-\infty,z)}(x)$, so that for a probability distribution μ on $\mathbb{R}$, one can estimate

$$\mu(-\infty,z) - \mathbb{P}(Z<z) = \int \big(\mathbb{1}_{(-\infty,z)}(x) - \mathbb{E}[\mathbb{1}_{(-\infty,z)}(Z)]\big)\,\mu(\mathrm{d}x) = \int \big(f'(x) - xf(x)\big)\,\mu(\mathrm{d}x),$$

so, in particular, we have that the total variation distance between a random variable X with distribution μ and a standard normal Z, which is defined as

$$2d_{TV}(X,Z) := \sup\Big\{\big|\mathbb{E}\big[u(X)-u(Z)\big]\big| : \|u\|_\infty \leqslant 1\Big\},$$

can be estimated as

$$d_{TV}(X,Z) \leqslant \sup_f \big|\mathbb{E}\big[f'(X) - Xf(X)\big]\big|, \tag{2.12}$$

where, as a consequence of Stein's lemma and the estimates therein, the supremum is taken over all piecewise continuously differentiable functions f, which are bounded by $\sqrt{\pi/2}$ and their first derivative is bounded by 2.

Basics of Wiener chaoses and statement of the Fourth Moment Theorem on Wiener chaoses. Let us start by defining the notion of *White noise* on $\mathbb{R}^n$.

Definition 2.7. *White $W(\cdot)$ defined on an abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a Gaussian process on $\mathbb{R}^d$, such that*

- *for every $A \subset \mathbb{R}^d$, the random variable $W(A)$ is distributed as a normal with mean zero and variance the Lebesque measure of A, denoted by $|A|$,*
- *for $A, B \subset \mathbb{R}^d$ such that $A \cap B = \varnothing$, the variables $W(A), W(B)$ are independent,*
- *for $A_1, A_2, \ldots$ disjoint Borel sets of $\mathbb{R}^d$ it holds a.s. that $W(\cup_{i=1}^\infty A_i) = \sum_{i=1}^\infty W(A_i)$.*

Given a symmetric function f in $L^2\big((\mathbb{R}^d)^q\big)$, that vanishes on the diagonals we define the iterated Wiener integral as

$$I_q(f) = \int_{(\mathbb{R}^d)^q} f(x_1, \ldots, x_q) W(\mathrm{d}x_1)\cdots W(\mathrm{d}x_q)$$

In the case that $d = 1$, these are the usual iterated Itô integrals.

It is known [J97, N06] that any function $F \in L^2(\mathbb{P})$ has a unique expansion of the form

$$F = \sum_{q=0}^\infty I_q(f_q). \tag{2.13}$$

This is called the Wiener chaos expansion. The space spanned by the q-fold iterated integrals $\{I_q(f) : f \in L^2\big((\mathbb{R}^d)^q\big)\}$ is called the *homogeneous chaos of order q*.

Let us notice that iterated integrals $I_q(f)$ are the continuous versions of the homogeneous multilinear polynomials considered in previous sections. In particular, a homogeneous multilinear polynomial

$$\Psi(\omega) := \sum_{I \subset (\mathbb{Z}^d)^q} \psi(I) \prod_{x\in I} \omega_x$$

with $(\omega_x)_{x\in\mathbb{Z}^d}$ being i.i.d. standard normal can be seen as the iterated Wiener integral

$$\frac{1}{q!}\int_{(\mathbb{R}^d)^q} \psi^{\mathrm{ext}}(x_1, \ldots, x_q) W(\mathrm{d}x_1)\cdots W(\mathrm{d}x_q)$$

by setting $\omega_x = |C_x|^{-1/2}\, W(C_x)$, where for $x \in \mathbb{Z}^d$ we denote by C_x the unit cube of $\mathbb{R}^d$ with "bottom-left" corner at $x \in \mathbb{Z}^d$ and $\psi^{\mathrm{ext}}(x_1, \ldots, x_q)$ the symmetric, piecewise constant function on $(\mathbb{R}^d)^k$ which takes the value $\psi(\{x_1, ..., x_q\})$ in $C_{x_1} \times \cdots C_{x_q}$. Moreover, the Wiener chaos expansion (2.13) should be viewed as the continuous analogue of Hoeffding's decomposition.

The Nualart-Peccati theorem [NP05] is as follows:

Theorem 2.8. *Consider a sequence $Z_n := I_q(f_n)$ of random variables in a fixed Wiener chaos of order q, such that $\mathbb{V}\mathrm{ar}(Z_n) \to 1$, as $n \to \infty$. Then Z_n converges in distribution to a standard normal if and only if $\mathbb{E}[Z_n^4] \to 3$.*

We should note that this type of theorem can be extended to random variables, which belong to *mixed chaoses* of bounded order.

We will present here a more quantitative version of this theorem due to Nourdin and Peccati [NP09]. Even though this theorem is stated for homogeneous chaoses of order 2, it can be extended to homogeneous chaoses of general order.

Theorem 2.9. *Let $(Z_n)_{n\geqslant 1}$ be a sequences of random variables belonging to the second homogeneous Wiener chaos and Z be a standard normal random variable. Then*

$$d_{TV}(Z_n, Z) \leqslant 2\sqrt{\frac{1}{6}\Big|\mathbb{E}\big[Z_n^4\big] - 3\Big| + \frac{3 + \mathbb{E}\big[Z_n^2\big]}{2}\,\Big|\mathbb{E}\big[Z_n^2\big] - 1\Big|}.$$

In particular, we see that if the sequence Z_n belongs to a fixed Wiener chaos and is such that the second moments converge to 1 and the third moments to 3, then the central limit theorem is valid and the above theorem provides also information on the speed of convergence.

Basics of Malliavin calculus. We will now present the essential notions of Malliavin calculus in the simplest case of iterated Wiener-Itô chaos expansions. These notions can be generalised to abstract Hilbert space settings. To keep things more direct we will not pay attention in detailing specific convergence assumptions and functional spaces and simply assume that things make sense in some L^2 space, which will be easy to identify.

We start with the basic notion of **Malliavin derivative**, which is denoted by D. If $F \in L^2((\Omega, \mathcal{F}, \mathbb{P}))$ has the Wiener chaos expansion $F = \sum_{q\geqslant 0} I_q(f_q)$, then DF can be identified as an element of $L^2(\Omega \times \mathbb{R}^d)$ defined as

$$D_a(F) := \sum_{q\geqslant 1} q I_{q-1}(f_q(\underbrace{\cdot, ..., \cdot}_{n-1}, a)). \tag{2.14}$$

For example, if

$$F = \int_{\mathbb{R}^2} f(x_1, x_2) W(\mathrm{d}x_1) W(\mathrm{d}x_2),$$

then

$$D_a F = 2\int_{\mathbb{R}} f(x, a) W(\mathrm{d}x) = \int_{\mathbb{R}} f(x, a) W(\mathrm{d}x) + \int_{\mathbb{R}} f(a, x) W(\mathrm{d}x),$$

where the last equality just highlights the fact that the factor 2 (or in general the factor q in (2.14)) comes from the symmetry assumption on the kernels f_q. The Malliavin derivative

D satisfies a **chain rule**: If $\phi\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function and $F_1, ..., F_n \in L^2(\mathbb{P})$, then

$$D\phi(F_1, ..., F_n) = \sum_{i=1}^{n} \partial_i \phi(F_1, ..., F_n)\, DF_i. \tag{2.15}$$

There are two more operators of central significance: The one is the **adjoint operator** of D (also called **Skorokhod integral)**, which is defined via

$$\mathbb{E}\big[F\,\delta(u)\big] = \mathbb{E}\big[\langle DF, u\rangle_{L^2(\mathbb{R}^d)}\big].$$

An easy exercise shows that if $F = \sum_{q\geqslant 0} I_q(f_q)$, then

$$\delta DF = -\sum_{q\geqslant 0} q I_q(f_q). \tag{2.16}$$

This motivates the second operator of significance

$$L := -\sum_{q\geqslant 0} q J_q,$$

where J_q denotes the projection operator on the q^{th}-homogeneous Wiener chaos; that is, if $F = \sum_{q\geqslant 0} I_q(f_q)$, then $J_q F := I_q(f_q)$. So relation (2.16) can be concisely written as

$$\delta D = -L. \tag{2.17}$$

Finally, we note that L has an inverse operator

$$L^{-1} = -\sum_{q\geqslant 1} q^{-1}\, J_q.$$

This is an inverse to L since it can be easily verified that

$$LL^{-1}Z = Z - \mathbb{E}[Z], \qquad \text{for any } Z \in L^2(\mathbb{P}). \tag{2.18}$$

The crux of the matter of the Malliavin calculus in relation to central limit and fourth moment theorems is the **product rule formula**, which reads as

$$\delta(Fu) = F\delta(u) - \langle DF, u\rangle_{L^2(\mathbb{R}^d)}, \tag{2.19}$$

which together with the chain rule (2.15) leads to the following **integration by parts formula**, which is to be thought of as an infinite dimensional version of the Gaussian integration by parts formula :

$$\mathbb{E}\big[\,Ff(F)\big] = \mathbb{E}\big[f'(F)\,\langle DF, -DL^{-1}F\rangle_{L^2(\mathbb{R}^d)}\,\big], \tag{2.20}$$

for $F \in L^2(\mathbb{P})$ mean zero and $f\colon \mathbb{R} \to \mathbb{R}$ smooth. Indeed, this can be easily verified:

$$\begin{aligned}
\mathbb{E}\big[f'(F)\,\langle DF, -DL^{-1}F\rangle_{L^2(\mathbb{R}^d)}\,\big] &= \mathbb{E}\big[\langle f'(F)DF, -DL^{-1}F\rangle_{L^2(\mathbb{R}^d)}\,\big] = \mathbb{E}\big[\langle Df(F), -DL^{-1}F\rangle_{L^2(\mathbb{R}^d)}\,\big] \\
&= \mathbb{E}\big[f(F)\,(-1)\delta DL^{-1}F\,\big] = \mathbb{E}\big[f(F)LL^{-1}F\,\big] \\
&= \mathbb{E}\big[f(F)F\,\big],
\end{aligned}$$

where the sequence of equalities follows: chain rule (second equality), definition of adjoint operator (third equality), property (2.17) (fourth equality) and property (2.18) (final equality).

Applications of Stein's method and Malliavin calculus to fourth moment theorems.

Theorem 2.10 (Theorem 3.1 and Proposition 3.2 in [NP09]). *Let Z be a standard normal variable and $F \in L^2(\mathbb{P})$ with $\mathbb{E}[F] = 0$. Then*

$$d_{TV}(F, Z) \leqslant 2\mathbb{E}\Big[\Big(1 - \langle DF, -DL^{-1}F\rangle_{L^2(\mathbb{R}^d)}\Big)^2\Big]^{1/2}.$$

Moreover, if $F = I_q(f)$, then

$$\mathbb{E}\Big[\Big(1 - \langle DF, -DL^{-1}F\rangle_{L^2(\mathbb{R}^d)}\Big)^2\Big] = \mathbb{E}\Big[\Big(1 - \tfrac{1}{q}\|DF\|^2_{L^2(\mathbb{R}^d)}\Big)^2\Big].$$

Proof. By (2.12) we have that

$$d_{TV}(X, Z) \leqslant \sup_f \big|\mathbb{E}\big[f'(X) - Xf(X)\big]\big|,$$

where the supremum is over functions f which are bounded by $\sqrt{\pi/2}$ and with first derivative bounded by 2. By the integration by parts formula we can write

$$\mathbb{E}\big[f'(X) - Xf(X)\big] = \mathbb{E}\Big[f'(X)\left(1 - \langle DF, -DL^{-1}LF\rangle_{L^2(\mathbb{R}^d)}\right)\Big],$$

and by Cauchy-Schwarz this is bounded by

$$\mathbb{E}\Big[f'(X)^2\Big]^{1/2}\mathbb{E}\Big[\left(1 - \langle DF, -DL^{-1}LF\rangle_{L^2(\mathbb{R}^d)}\right)^2\Big]^{1/2} \leqslant 2\mathbb{E}\Big[\left(1 - \langle DF, -DL^{-1}LF\rangle_{L^2(\mathbb{R}^d)}\right)^2\Big]^{1/2},$$

if $\|f'\|_\infty \leqslant 2$, giving the desired bound on the total variation distance.

Now, if $F = I_q(f)$, then $L^{-1}F = q^{-1}I_q(F) = q^{-1}F$ and so the second claim of the theorem follows immediately. □

Lemma 2.11 (Multiplication formula). *For two functions $f \in L^2((\mathbb{R}^d)^p)$ and $g \in L^2((\mathbb{R}^d)^q)$ we define the contraction to be*

$$\big(f \otimes_r g\big)(x_1, ..., x_{p+q-2r}) := \int_{(\mathbb{R}^d)^r} f(x_1, ..., x_{p-r}, y_1, ..., y_r)\, g(x_{p-r+1}, ..., x_{p+q-2r}, y_1, ..., y_r)\, \mathrm{d}y_1 \cdots \mathrm{d}y_r,$$

and we denote by $f \tilde{\otimes}_r g$ its canonical symmetrisation†*. Then the product $I_p(f)I_q(g)$ admits the Wiener chaos decomposition*

$$I_p(f)I_q(g) = \sum_{r=0}^{p\wedge q} r! \binom{p}{r}\binom{q}{r} I_{p+q-2r}(f \tilde{\otimes}_r g). \tag{2.21}$$

Let us just see how this formula writes for $p = q = 1$. In this case we have that

$$\left(f \otimes_0 g\right)(x_1, x_2) = f(x_1)g(x_2) \quad \text{and} \quad \left(f \tilde{\otimes}_0 g\right)(x_1, x_2) = \frac{1}{2!}\big(f(x_1)g(x_2) + f(x_2)g(x_1)\big)$$

and

$$\left(f \otimes_1 g\right) = \int f(y)g(y)\mathrm{d}y.$$

†for a function $f(x_1, ..., x_n)$ we define its canonical symmetrisation to be $\sum_\sigma \frac{1}{n!} f(x_{\sigma(1)}, ..., x_{\sigma(n)})$, with the sum running over all permutations

Then

$$\begin{aligned} I_1(f)I_1(g) &= I_2(f\,\tilde{\otimes}_2\,g) + I_0(f\,\otimes_1\,g) \\ &= \int\int_{\{x_1\neq x_2\}} f(x_1)g(x_2)\,W(\mathrm{d}x_1)W(\mathrm{d}x_2) + \int f(x)g(x)\mathrm{d}x. \end{aligned}$$

To understand what is behind this multiplication formula, it is instructive to view it in a (very) simple discrete setting. Suppose that we have two discrete, homogeneous chaoses of degree one:

$$I_1(f) := \sum_x f(x)\omega_x \qquad \text{and} \qquad I_1(g) := \sum_x g(x)\omega_x,$$

where $(\omega_x)_{x\in\mathsf{S}}$ is a family of i.i.d. mean zero, variance one random variables. Then we can write the product $I_1(f)I_1(g)$ as

$$I_1(f)I_1(g) = \sum_{x,y} f(x)g(y)\omega_x\omega_y.$$

The issue now is that when $x=y$ the term $\sum_x f(x)g(x)\omega_x^2$ is not a mean zero variable and homogeneous chaoses, of degree greater than zero, are supposed to be mean zero. To remedy this, we can centre and write the above as

$$\begin{aligned} I_1(f)I_1(g) &= \sum_{x,y} f(x)g(y)\big(\omega_x\omega_y - \mathbb{E}[\omega_x\omega_y]\big) + \sum_x f(x)g(x)\mathbb{E}[\omega_x^2] \\ &= \sum_{x,y} f(x)g(y)\big(\omega_x\omega_y - \mathbb{E}[\omega_x\omega_y]\big) + \sum_x f(x)g(x), \end{aligned}$$

where in the last step we used the assumption that $\mathbb{E}[\omega_x^2]=1$.

Proof of Theorem 2.9. For conciseness we will denote $\|\cdot\| = \|\cdot\|_{L^2(\mathbb{R}^d)}$ and also denote Z_n, for generic n, by $F = I_2(f)$ for some $f\in L^2((\mathbb{R}^d)^2)$, since it is assumed that Z_n belongs to the second homogeneous Wiener chaos. By Theorem 2.10 we need to estimate

$$\mathbb{E}\Big[\big(1-\tfrac12\|DF\|^2\big)^2\Big] = 1 - \mathbb{E}\big[\|DF\|^2\big] + \tfrac14\mathbb{E}\Big[\|DF\|^4\Big] \tag{2.22}$$

and so we need to compute the moments of DF. Using the multiplication formula (2.21), we have that

$$F^2 = I_2(f)^2 = I_4(f\otimes f) + 4I_2(f\otimes_1 f) + \mathbb{E}[F^2], \tag{2.23}$$

and from this we have that

$$L(F^2) = -4I_4(f\otimes f) - 8I_2(f\otimes_1 f). \tag{2.24}$$

We also have that

$$D_aF = 2I_1(f(\cdot,a)), \quad \text{for every } a\in\mathbb{R}^d,$$

and using again the multiplication formula we have that

$$\begin{aligned} \big(D_aF\big)^2 = 4I_1(f(\cdot,a))^2 &= 4I_2\big(f(\cdot,a)\otimes f(\cdot,a)\big) + 4\int\int f(x,a)^2\,\mathrm{d}x\mathrm{d}a \\ &= 4I_2\big(f(\cdot,a)\otimes f(\cdot,a)\big) + \mathbb{E}\big[\|DF\|^2\big] \end{aligned} \tag{2.25}$$

since

$$\mathbb{E}\big[\|DF\|^2\big] = 4\mathbb{E}\Big[\int \big(\int f(x,a)W(\mathrm{d}x)\big)^2 \mathrm{d}a\Big] = 4\int \mathbb{E}\Big[\big(\int f(x,a)W(\mathrm{d}x)\big)^2\Big]\mathrm{d}a$$
$$= 4\int\int f(x,a)^2 \mathrm{d}x\mathrm{d}a. \tag{2.26}$$

Let us now compute $\mathbb{E}[\|DF\|^4]$, which, to start with, we will write as $\mathbb{E}[\|DF\|^2\langle DF, DF\rangle]$ and we will use the product rule (2.19) to write

$$\mathbb{E}\big[\|DF\|^4\big] = \mathbb{E}\big[\|DF\|^2\langle DF, DF\rangle\big] = \mathbb{E}\big[\|DF\|^2(F\,\delta DF - \delta(FDF))\big]$$

and using $FDF = \frac{1}{2}D(F^2)$ from the general chain rule (2.15) we have that

$$\mathbb{E}\big[\|DF\|^4\big] = \mathbb{E}\big[\|DF\|^2(F\,\delta DF - \frac{1}{2}\delta(\,D(F^2)\,))\big].$$

Using now that $\delta DF = -LF = 2F$, with the second equality since F belongs to the second homogeneous chaos, and also using that $\delta DF^2 = -LF^2$, we get that

$$\mathbb{E}\big[\|DF\|^4\big] = 2\mathbb{E}\big[\|DF\|^2 F^2\big] + \frac{1}{2}\mathbb{E}\big[\|DF\|^2\, L(F^2)\big].$$

Inserting into this formula relations (2.23) and (2.24) and performing a simple algebra, taking also into account the orthogonality between chaoses of different order, we obtain that

$$\mathbb{E}\big[\|DF\|^4\big] = \mathbb{E}\big[\|DF\|^2 F^2\big] + \mathbb{E}\big[\,F^2\,\big]\mathbb{E}\big[\|DF\|^2\big]. \tag{2.27}$$

The penultimate step is to make use of the following formula

$$\mathbb{E}\big[F^s\|DF\|^2\,\big] = \frac{q}{s+1}\mathbb{E}\big[F^{s+2}\,\big],$$

which is valid for every F in a homogeneous chaos of order q and $s \in \mathbb{Z}_{\geqslant 0}$, and apply to the first term in (2.27) with $q = 2$ and $s = 2$ to get that

$$\mathbb{E}\big[\|DF\|^4\big] = \frac{2}{3}\mathbb{E}\big[F^4\big] + 2\mathbb{E}\big[\,F^2\,\big]^2. \tag{2.28}$$

Inserting (2.26) and (2.28) to (2.22) and after a simple algebra we have that

$$\mathbb{E}\Big[\big(1 - \tfrac{1}{2}\|DF\|^2\,\big)^2\Big] = \frac{1}{6}\big(\mathbb{E}\big[F^4\big] - 3\big) + \big(\mathbb{E}\big[F^2\big] - 1\big)\Big(\frac{1}{2}\mathbb{E}\big[F^2\big] - \frac{3}{2}\Big)$$
$$\leqslant \frac{1}{6}\big|\mathbb{E}\big[F^4\big] - 3\big| + \big|\mathbb{E}\big[F^2\big] - 1\big|\,\frac{\mathbb{E}[F^2] + 3}{2},$$

which completes the proof by inserting this estimate into the estimate of Theorem 2.10. □

3. Applications to disordered systems

We will now show some applications of the previously discussed Lindeberg principles and fourth moment central limit theorems in the study of scaling limits of disordered systems. The discussion here is based on works [CSZ17a, CSZ17b, CSZ18a, CSZ18b, CSZ18+].

3.1. Disorder relevance. Let us start with a description of the notion of a disordered system. Consider an open set $\Omega \subseteq \mathbb{R}^d$ and define the lattice $\Omega_\delta := (\delta\mathbb{Z})^d \cap \Omega$, for $\delta > 0$ which is the support of a *"random field"* $\sigma = (\sigma_x)_{x\in\Omega_\delta}$ whose law is determined by a probability measure, which we denote by $\mathrm{P}^{\mathrm{ref}}_{\Omega_\delta}$. Typically the field takes values $\sigma_x \in \{0,1\}$ or $\{\pm 1\}$. Even though it also sensible to consider fields that take non binary values, currently the treatment of such fields is out of the scope of the methods we will describe.

Some examples of such fields can be:

- *Random walks.* In this case, Ω_δ is typically $\mathbb{Z}^d \times \{0,1,...,N\}$ for $N \geqslant 1$ and δ is really to be thought of as being $1/N$. The field $\sigma = (\sigma_{n,x})_{n\leqslant N, x\in\mathbb{Z}^d}$ in this case is $\sigma_{n,x} = \mathbb{1}_{\{S_n = x\}}$, where $(S_n)_{n\geqslant 1}$ is the trajectory of a random walk, which may or may not be simple. $\mathrm{P}^{\mathrm{ref}}_{\Omega_\delta}$ represents the law of the random.
- *Ising models.* In this case, $\Omega_\delta := (\delta\mathbb{Z})^d \cap \Omega$ with δ representing the mesh of the grid on $\Omega \subset \mathbb{R}^d$ and $\sigma_x \in \{\pm 1\}$. The measure $\mathrm{P}^{\mathrm{ref}}_{\Omega_\delta}$ is the Ising measure given by
$$\mathrm{P}^{\mathrm{ref}}_{\Omega_\delta}(\sigma) = \frac{1}{Z^{\mathrm{ref}}_{\Omega_\delta}} e^{J\sum_{x\sim y}\sigma_x\sigma_y},$$
where $x \sim y$ means that sites x, y are nearest neighbour, i.e. connected by an edge of $\mathbb{Z}^d$, J is a coupling constant, which represents the strength of interaction between neighbouring values of the filed σ and
$$Z^{\mathrm{ref}}_{\Omega_\delta} = \sum_\sigma e^{J\sum_{x\sim y}\sigma_x\sigma_y},$$
is called the *partition function* and it is the normalisation needed to have a probability distribution.

A disordered system arises when on the lattice Ω_δ, on top of the reference field σ, there exists an additional randomness, $\omega := (\omega_x)_{x\in\Omega_\delta}$ modelled in the form of an i.i.d. collection, which typically is assumed to be of mean zero, variance one and having exponential moments (although it is sensible to relax the exponential moment assumption and consider heavy tailed fields, in which case new phenomena often arise, see for example [DZ16]). We will call the randomness ω *disorder* and we will denote its law by $\mathbb{P}$ and expectation with respect to it by $\mathbb{E}$.

Given a realisation of the disorder ω the *disordered model* is defined as the following probability measure $\mathrm{P}^\omega_{\Omega_\delta;\lambda,h}$ for the field $\sigma = (\sigma_x)_{x\in\Omega_\delta}$:
$$\mathrm{P}^\omega_{\Omega_\delta;\beta,h}(\sigma) := \frac{e^{\sum_{x\in\Omega_\delta}(\beta\omega_x + h)\sigma_x}}{Z^\omega_{\Omega_\delta;\beta,h}} \mathrm{P}^{\mathrm{ref}}_{\Omega_\delta}(\sigma)\,, \tag{3.1}$$
where now the partition function is defined by
$$Z^\omega_{\Omega_\delta;\beta,h} := \mathbb{E}^{\mathrm{ref}}_{\Omega_\delta}\big[e^{\sum_{x\in\Omega_\delta}(\beta\omega_x + h)\sigma_x}\big]\,. \tag{3.2}$$
and we remark that in this case it is itself a random variable, depending on the realisation ω.

A question of central interest in statistical physics but often very poorly understood is

Q. *" does an arbitrarily small amount of disorder change the statistical mechanics properties of the reference field ? "*

In the 70s A.B. Harris [H74] proposed the following criterion, which is known as *Harris criterion*:

Harris criterion: *If d is the dimension and γ is the correlation length exponent of the reference model, then if $\gamma < \frac{d}{2}$, then the model is* **disorder irrelevant**, *meaning that sufficiently small amount of disorder is not sufficient to change its statistical properties. If $\gamma > \frac{d}{2}$, then the model is* **disorder relevant**, *meaning that any arbitrarily small amount of disorder does change its statistical properties.*

Let us first define what we mean by a *correlation length exponent* here: Consider the (what is called) *k-point function* to be:

$$\psi_\delta^{(k)}(x_1, ..., x_k) := \mathbb{E}^{\mathrm{ref}}_{\Omega_\delta}[\sigma_{x_1} \cdots \sigma_{x_k}].$$

Then the correlation length exponent can be defined as the exponent γ such that

$$(\delta^{-\gamma})^k \, \psi^{(k)}_{\Omega_\delta}(x_1, \dots, x_k) \xrightarrow[\delta \downarrow 0]{} \boldsymbol{\psi}^{(k)}_{\Omega}(x_1, \dots, x_k). \tag{3.3}$$

where the limit is to be thought of as pointwise, although stronger forms such $L^2(\Omega^k)$ will be needed for the framework we will develop.

Even though very simple, actually rigorously verifying the Harris criterion in concrete examples is often difficult and requires a careful case by case analysis (although one can transfer some intuition and a set of "general principles" between different problems). In [CSZ17a] we proposed a different point of view for the Harris criterion using as a platform the Lindeberg principles for multilinear polynomials and focusing on the existence of *non-trivial* (i.e. random) scaling limits of the partition functions when β, h are suitably scaled to zero as $\delta \to 0$. The question can be phrased as:

Q. *Consider the partition function of a disordered model as defined in* (3.2)*. Can we choose $\beta = \beta_\delta$ and $h = h_\delta$, converging to zero as $\delta \to 0$, such that $Z^\omega_{\Omega_\delta;\beta_\delta,h_\delta}$ converges in distribution to a random (i.e. finite and not constant) random variable $\boldsymbol{Z}^W_{\Omega;\hat\beta,\hat h}$?*

Here W denotes white noise on $\mathbb{R}^d$ and we request that the limit should be a non trivial function of an underlying white noise.

We will now describe how we can answer this question using the Lindeberg principle for multilinear polynomials. Although it makes sense to consider a general value of h, we will for the purposes of this exposition restrain ourselves to the choice of $h = -\lambda(\beta)$, where $\lambda(\beta) := \log \mathbb{E}[e^{\beta\omega_x}]$. We denote the partition function associated to this choice by $Z^\omega_{\Omega_\delta;\beta}$.

The starting point is to write the partition function in the form of a multilinear polynomial. We do this via what is called in statistical mechanics *high temperature* or *Mayer expansion*, which goes by writing

$$Z^\omega_{\Omega_\delta;\beta} = \mathbb{E}^{\mathrm{ref}}_{\Omega_\delta}\Big[\prod_{x\in\Omega_\delta}\big(1+\beta\sigma_x\zeta_x\big)\Big], \qquad \text{where} \qquad \zeta_x := \frac{e^{(\beta\omega_x - \lambda(\beta))}-1}{\beta}. \tag{3.4}$$

and then by expanding the product and interchanging the (finite) summation with the expectation $\mathbb{E}^{\mathrm{ref}}_{\Omega_\delta}$, so that to have

$$Z^\omega_{\Omega_\delta;\beta} = 1 + \sum_{k=1}^\infty \beta^k \sum_{x_1,\dots,x_k\in\Omega_\delta} \mathbb{E}^{\mathrm{ref}}_{\Omega_\delta}\big[\sigma_{x_1}\cdots\sigma_{x_k}\big] \prod_{i=1}^k \zeta_{x_i},$$

where the inner sum is taken over k-tuples over *distinct* $x_1,...,x_k \in \Omega_\delta$ (and so the sum over k even though written as an infinite sum it is in fact finite). With $\psi_\delta^{(k)}(x_1,...,x_k) := \mathbb{E}^{\text{ref}}_{\Omega_\delta}[\sigma_{x_1}\cdots\sigma_{x_k}]$ we write

$$Z^\omega_{\Omega_\delta;\beta} = 1 + \sum_{k=1}^\infty (\beta\delta^\gamma)^k \sum_{x_1,...,x_k\in\Omega_\delta} (\delta^{-\gamma})^k\, \psi_\delta^{(k)}(x_1,...,x_k) \prod_{i=1}^k \zeta_{x_i}, \tag{3.5}$$

where we have inserted the scaling of the k-point correlation function. Note that the random variables (ζ_x) are mean zero precisely due to the choice of the parameter h to be equal to $-\lambda(\beta)$.

At this point the need of a Lindeberg principle comes: suppose that we can replace the random variables (ζ_x), from (3.4), by standard normal variables, which we denote by (ξ_x). If so, then we could *model* this new collection of i.i.d. normal via a White noise $W(\cdot)$ on $\mathbb{R}^d$ as

$$\xi_x = |C_{x,\delta}|^{-1/2}\, W\big(C_{x,\delta}\big),$$

where $C_{x,\delta}$ is the cube in $(\delta\mathbb{Z})^d$ with side length δ, "bottom-left" corner equal to x and volume $|C_{x,\delta}| = \delta^d$ and consider the partition function

$$Z^W_{\Omega_\delta;\beta} = 1 + \sum_{k=1}^\infty \big(\beta\delta^{\gamma-\frac{d}{2}}\big)^k \sum_{x_1,...,x_k\in\Omega_\delta} (\delta^{-\gamma})^k\, \psi_\delta^{(k)}(x_1,...,x_k) \prod_{i=1}^k W(C_{x_i,\delta}), \tag{3.6}$$

which can also be written as an iterated Wiener-Itô integral as

$$Z^W_{\Omega_\delta;\beta} = 1 + \sum_{k=1}^\infty \big(\beta\delta^{\gamma-\frac{d}{2}}\big)^k \int\cdots\int_{\Omega^k} (\delta^{-\gamma})^k\, \boldsymbol{\psi}_\delta^{(k,\text{ext})}(x_1,...,x_k) \prod_{i=1}^k W(\mathrm{d}x_i),$$

where $\boldsymbol{\psi}_\delta^{(k,\text{ext})}$ is the piecewise constant function on Ω^k, which takes the constant value $\psi_\delta^{(k)}$ on the cubes $C_{x_1,\delta}\times\cdots\times C_{x_k,\delta}$.

Choosing now

$$\beta = \beta_\delta = \hat\beta\,\delta^{\frac{d}{2}-\gamma}, \tag{3.7}$$

ones sees via an easy $L^2(\mathbb{P})$ estimate and using assumption (3.3) (strengthened to hold in an $L^2(\Omega^k)$ sense) that

$$Z^W_{\Omega_\delta;\hat\beta} \xrightarrow[\delta\downarrow 0]{L^2(\mathbb{P})} 1 + \sum_{k=1}^\infty \hat\beta^k \int\cdots\int_{\Omega^k} \boldsymbol{\psi}^{(k)}(x_1,...,x_k) \prod_{i=1}^k W(\mathrm{d}x_i). \tag{3.8}$$

We should also remark at the consistency with the Harris criterion: the scaling of β in (3.7) is consistent with the requirement that $\beta_\delta \to 0$ with $\delta\to 0$ (thus disorder is gradually smaller) if $\gamma < d/2$. In the case that $\gamma > d/2$ it turns out that any scaling of β tending to zero as $\delta\to 0$ will always lead to be a trivial, i.e. non random and in fact constant, limit.

As we see, the main point in obtaining the scaling limit of the disordered partition function is justifying the passage (in the limit $\delta\to 0$) from (3.5) to (3.6). This step is precisely achieved with the Lindeberg principle as of Theorem 1.8. Let us note that if h in (3.2) is taken to be different than $-\lambda(\beta)$, then the random variables (ζ_x) are not mean zero and in this case one needs to be more careful as one needs to handle issues of convergence of the series (when going to the limit) in (3.5). Moreover, one needs an extension of Lindeberg theorem 1.8 that will cover the situation of non-mean-zero variables. These issues were settled and suitable extensions of Theorem 1.8 were achieved in [CSZ17a].

3.2. Marginal relevance or critical dimension. We have seen that the Harris criterion classifies disorder as “relevant” or “irrelevant” according to whether $d/2 > \gamma$ or $d/2 < \gamma$. However, the Harris criterion is inconclusive when $d/2 = \gamma$. This case is called *marginal* and disorder can then be either relevant or irrelevant depending on the finer details of the system. A situation where the disorder relevant/irrelevant and marginally relevant/irrelelvant regimes have been succesfully classified (but only after a large number of works a small sample of which can be represented by [A08, AZ09, BL16, DGLT09, GLT11, T07]) is the case of the random pinning model an overview of which is contained in [G11].

The point of view described in the previous section, classifying relevance or irrelevance via the existence of a non trivial scaling limit of the partition function, also fails in the marginal case. The reason for this is that the natural candidate (3.8) which represents the scaling limit in the relevant case does not make sense at the marginal case. This is best manifested by looking at the case of the pinning model. The partition function for this model is

$$Z^{\mathrm{pin}}_{N,\beta} := \mathbb{E}\Big[e^{\sum_{n=1}^{N}(\beta\omega_n-\lambda(\beta))\mathbb{1}_{\{S_n=0\}}}\Big], \tag{3.9}$$

where $(S_n)_{n\geqslant 1}$ is a one dimensional, simple random walk. We note that it is sensible for this model to consider a more general class of Markov processes, so that the transition between relevance and irrelevance is observed while moving through this class. However, the simple random walk corresponds to the marginal case, in which disorder also turns out to be marginally relevant (but without exhausting the class of processes for which disorder is marginally relevant). In this case, if one adopted the point of view described in the previous section, taking into account that $\mathrm{P}(S_n = 0) \sim 1/\sqrt{2\pi n}$ for $n \to \infty$, one would guess the scaling limit

$$1 + \sum_{k\geqslant 1} \big(\tfrac{\beta}{2\pi}\big)^k \int\!\!\cdots\!\!\int_{0<t_1<\cdots<t_k<1} \frac{1}{\sqrt{t_1}\sqrt{t_2-t_1}\cdots\sqrt{t_k-t_{k-1}}} \prod_{i=1}^{k} W(\mathrm{d}t_i),$$

which clearly does not make sense since the kernel $\big(t_1(t_2-t_1)\cdots(t_k-t_{k-1})\big)^{-1/2}$ is not L^2 integrable.

To handle the marginal case a different point of view was necessary [CSZ17b] and in the course of implementing this new approach the fourth moment theorem played an important role.

But before turning to describe the approach let us point out that the notion of marginal disorder turns out to be identical to the notion of *criticality* in stochastic PDEs [H14, GIP15] and in renormalization theory [K14]. It turns out that the notion of *sub-criticality* for singular SPDEs (or super-renormalizability in renormalization theory) matches with the notion of disorder relevance, while *criticality* corresponds to the case where the effect of disorder is marginal. Let us illustrate this fact by looking at the example of the Stochastic Heat Equation (SHE), whose solution can be actually seen, via the stochastic Feynman-Kac formula, as the partition function of a continuum random polymer model. The stochastic heat equation writes as

$$\partial_t u = \frac{1}{2}\Delta u + \dot{W}\, u, \qquad t > 0, x \in \mathbb{R}^d \tag{3.10}$$

where $\dot{W} = W(\mathrm{d}t\mathrm{d}x)/\mathrm{d}t\mathrm{d}x$, with W being a space-time white noise. We note that due to the irregularity of the noise $\dot{W}$ the product $\dot{W}u$ is not well defined and giving a meaning to a solution to SHE is not straightforward. In dimension 1 this problem can be circumvented but in dimension 2 (which turns out to be critical) a proper notion of solution is not trivial.

To understand where the difficulty arises one resorts to a *renormalization* procedure. This is a standard first step (or heuristic) in the course of understanding regularity properties, and in the simplest case amounts to a change of variables such as

$$(t,x) = T_\varepsilon(\tilde{t},\tilde{x}) := (\varepsilon^2\,\tilde{t},\varepsilon\,\tilde{x})\,.$$

Using the gaussian scaling property of the white noise, it is not difficult to see that $\tilde{u}(\tilde{t},\tilde{x}) := u(T_\varepsilon(\tilde{t},\tilde{x}))$ formally solves the SPDE

$$\frac{\partial \tilde{u}}{\partial \tilde{t}} = \frac{1}{2}\Delta\tilde{u} + \beta\varepsilon^{1-\frac{d}{2}}\dot{\tilde{W}}\,\tilde{u}, \tag{3.11}$$

where $\tilde{W}$ is a new space-time White noise obtained from W via scaling. Therefore, space-time renormalization has the effect of changing the strength of the noise to $\varepsilon^{1-\frac{d}{2}}\beta$.

We now see that if $d < 2$, then, as $\varepsilon \to 0$, the strength of the noise in the renormalized equation goes to zero, which means that the noise will have a gradually decreasing effect on the regularity of the solution to the SHE and thus a solution can be suitably defined. On the other hand, for $d > 2$ the noise should crucially affect the solution as its strength after renormalization increases. Contrary to the previous cases, one sees that $d = 2$ is a critical dimension as the renormalization leaves the noise invariant and thus no conclusion can be drawn on the effect of noise to the existence and regularity of a solution.

In terms of disordered systems, one is interested in large scale effects and so it is meaningful to consider the reciprocal change of variables (renormalization) $(t,x) = T_{\varepsilon^{-1}}(\tilde{t},\tilde{x})$. This will result to a renormalized equation where now the strength of the noise is $\varepsilon^{\frac{d}{2}-1}\beta$. The conclusion in this case is that: for $d < 2$, then noise (disorder) has a prevailing effect (amounting to disorder relevance), while for $d > 2$ the effect of the noise vanishes, amounting to disorder irrelevance. However, again, when $d = 2$ the renormalization leaves the noise invariant and no conclusion can be drawn on the effect of noise. This is the marginal case.

To understand the structure in the marginal case (and how the fourth moment theorem enters), let us look at the example of the pinning partition function (3.9) and expand it in the form of a multilinear polynomial, as described in the previous section, as:

$$Z^{\mathrm{pin}}_{N,\beta} = 1 + \sum_{k\geqslant 1}\beta^k \sum_{1\leqslant n_1<\cdots<n_k\leqslant N}\prod_{i=1}^k q_{n_i-n_{i-1}}\prod_{i=1}^k \zeta_{n_i}, \tag{3.12}$$

where $\zeta_n := \beta^{-1}\big(e^{\beta\omega_n-\lambda(\beta)}-1\big)$ and $q_n := \mathrm{P}(S_n = 0)$.

To see what the suitable choice of β should be, we look at the variance of the first non-constant term:

$$\mathbb{V}\mathrm{ar}\Big(\beta\sum_{n=1}^N q_n\zeta_n\Big) = \beta^2\sum_{n=1}^N q_n^2\,\mathbb{V}\mathrm{ar}\left(\zeta_n\right) \approx \beta^2\sum_{n=1}^N q_n^2 \approx \beta^2\sum_{n=1}^N \frac{1}{2\pi n} \approx \frac{\beta^2}{2\pi}\log N \tag{3.13}$$

since, as it is easy to check, $\mathbb{V}\mathrm{ar}(\zeta_n) \approx 1$ for $\beta \approx 0$ and since for a one-dimensional simple random walk $q_n := \mathrm{P}(S_n = 0) \approx 1/\sqrt{2\pi n}$. This indicates that in order to ensure a non trivial limit, β should be chosen as $\hat{\beta}\sqrt{2\pi/\log N}$, so that the variance stays of order 1. This choice turns out to also ensure that each one of the remaining terms in the expansion is also of order 1.

A second information that this variance computation provides is on the correct time scale at which one should observe the system. To understand this, notice that the asymptotic variance (3.13) remains unchanged if we sum over a time horizon tN, for any arbitrary but fixed time variable t. On the other hand, if we considered a time horizon N^t with $t > 0$ fixed, then the asymptotic variance will indeed change when varying t. Therefore, the correct time scale is N^t. To incorporate this observation, we decompose the summations over $n_1, ..., n_k$ in the multilinear expansion (3.12) over intervals $n_j - n_{j-1} \in I_{i_j}$, with $I_{i_j} = \big(N^{\frac{i_j-1}{M}}, N^{\frac{i_j}{M}}\big]$, $i_j \in \{1, ..., M\}$ and with M being a coarse graining parameter (which will eventually tend to infinity). We can then rewrite the k-th term in the expansion (3.12) as

$$\frac{\hat{\beta}^k}{M^{k/2}} \sum_{1 \leqslant i_1, ..., i_k \leqslant M} \Theta^{N,M}_{i_1,...,i_k} \qquad \text{where} \tag{3.14}$$

$$\Theta^{N,M}_{i_1,...,i_k} := \left(\frac{2\pi M}{\log N}\right)^{k/2} \sum_{n_j - n_{j-1} \in I_{i_j} \text{ for } j=1,...,k} \prod_{j=1}^{k} q_{n_j - n_{j-1}} \, \zeta_{n_j}.$$

We now observe that if an index i_j is a running maximum for the k-tuple $\boldsymbol{i} := (i_1, ..., i_k)$, i.e. $i_j > \max\{i_1, ..., i_{j-1}\}$ then $\big(N^{\frac{i_j-1}{M}}, N^{\frac{i_j}{M}}\big] \ni n_j \gg n_r \in \big(N^{\frac{i_r-1}{M}}, N^{\frac{i_r}{M}}\big]$, for all $r < j$ †. This implies that $q_{n_j - n_{j-1}} \approx q_{n_j}$ for $n_j \in I_{i_j}$ and $n_{j-1} \in I_{i_{j-1}}$. Decomposing the sequence $\boldsymbol{i} := (i_1, ..., i_k)$ according to its running maxima, i.e. $\boldsymbol{i} = (\boldsymbol{i}^{(1)}, ..., \boldsymbol{i}^{(\mathfrak{m}(\boldsymbol{i}))})$ with $\boldsymbol{i}^{(r)} := (i_{\ell_r}, ..., i_{\ell_{r+1}} - 1)$ and with $i_1 = i_{\ell_1} < i_{\ell_2} < \cdots < i_{\ell_{\mathfrak{m}}}$ being the successive running maxima, it can be shown that (3.14) asymptotically factorizes for large N as

$$\frac{\hat{\beta}^k}{M^{\frac{k}{2}}} \sum_{\boldsymbol{i} \in \{1,...,M\}^k_\sharp} \Theta^{N;M}_{\boldsymbol{i}^{(1)}} \Theta^{N;M}_{\boldsymbol{i}^{(2)}} \cdots \Theta^{N;M}_{\boldsymbol{i}^{(\mathfrak{m})}}. \tag{3.15}$$

The heart of the matter is to show that all the $\Theta^{N;M}_{\boldsymbol{i}^{(j)}}$ converge jointly, when $N \to \infty$ to standard normal variables. This is where the fourth moment theorem is used.

Checking that the fourth moment of each of the $\Theta^{N;M}_{\boldsymbol{i}^{(j)}}$ actually converges to 3, reduces to a combinatorial problem: Expanding the fourth power of the summation defining $\Theta^{N;M}_{\boldsymbol{i}^{(j)}}$, see (3.14), produces a fourfold product $\prod_{i=1}^k \zeta_{a_j} \prod_{i=1}^k \zeta_{b_j} \prod_{i=1}^k \zeta_{c_j} \prod_{i=1}^k \zeta_{d_j}$ and when considering the expectation $\mathbb{E}[\cdot]$ then one looks at all possible combinatorial matchings among this list of ζ variables. It turns out (and towards this a crucial role is played by the logarithmic growth of the variance (3.13) as well as the exponential time scale N^t that it imposes) that the main contributions comes from *three* possible matchings: either all the ζ_a's will match with the ζ_b's (meaning that $a_i = b_i$ and thus $\zeta_{a_i} = \zeta_{b_i}$ for $i = 1, ..., k$) and the ζ_c's with the ζ_d's or all the ζ_a's will match with the ζ_c's and the ζ_b's with the ζ_d's or all the ζ_a's will match with the ζ_d's and the ζ_b's with the ζ_c's. These three main ways of matching lead to the value 3 for the asymptotic fourth moment of $\Theta^{N;M}_{\boldsymbol{i}^{(j)}}$. The details of this argument (in a more general setting than just the pinning model) is the subject of Proposition 5.2 in [CSZ17b].

---

†Strictly speaking, for this inequality to be valid uniformly, we need to restrict to values of $\boldsymbol{i} \in \{1, ..., M\}^k_\sharp := \{\boldsymbol{i} \in \{1, ..., M\}^k : |i_j - i_k| > 1\}$, but this is a minor technical point that be easily taken care of.

Having identified (with the help of the fourth moment theorem) the asymptotic behaviour of the building block $\Theta^{N;M}_{\boldsymbol{i}^{(j)}}$, we can identify the asymptotic behaviour of the partition function when $\beta = \beta_N = \hat{\beta}\sqrt{2\pi/\log N}$ via (3.12), (3.14) and (3.15). Let us remark that along the way certain re-summations are performed and for this it is important that $\hat{\beta} < 1$, a choice that also marks a phase transition.

We can summarise the final result in its general form, which also shows a universality in the behaviour of marginal models.

Theorem 3.1 ([CSZ17b]). *Let $Z^{\mathrm{marginal}}_{N,\beta_N}$ be a multilinear polynomial (typically a partition function) of the form*

$$Z^{\mathrm{marginal}}_{N,\beta_N} = 1 + \sum_{k=1}^{N} \beta_N^k \sum_{\substack{1 \leqslant n_1 < \cdots < n_k \leqslant N \\ x_1, \ldots, x_k \in \mathbb{Z}^d}} \prod_{i=1}^{k} q_{n_i - n_{i-1}}(x_i - x_{i-1})\, \zeta_{n_i, x_i},$$

where $(\zeta_{n,x})_{n \in \mathbb{N}, x \in \mathbb{Z}^d}$ is a collection of i.i.d. mean zero, variance one random variables with exponential moments and the kernel $(q_n(x))_{n \in \mathbb{N}, x \in \mathbb{Z}^d}$ satisfies that

$$R_N := \sum_{n=1}^{N} \sum_{x \in \mathbb{Z}^d} q_n(x)^2 \quad \text{grows to infinity as a slowly varying function.} \tag{3.16}$$

Then if $\beta_N := \hat{\beta}/\sqrt{R_N}$, it holds that

$$Z^{\omega}_{N,\beta_N} \xrightarrow[N \to \infty]{(d)} \boldsymbol{Z}_{\hat{\beta}} := \begin{cases} \exp\left(\sigma_{\hat{\beta}} \mathsf{X} - \frac{1}{2}\sigma^2_{\hat{\beta}}\right) & \text{if } \hat{\beta} < 1 \\ 0 & \text{if } \hat{\beta} \geqslant 1 \end{cases}. \tag{3.17}$$

where X is a standard normal variable with variance $\sigma^2_{\hat{\beta}} = \log(1 - \hat{\beta}^2)^{-1}$.

Condition (3.16), which is essentially derived via a computation of the form (3.13), can be used as a quantitative criterion for marginal relevance and it is what in statistical physics is called *overlap*. The kernel $q_n(x)$ may not have a dependence in $x \in \mathbb{Z}^d$, as that was the case in the pinning model. One can also possibly consider situations where $\mathbb{Z}^d$ is replaced by other lattices or more general sets. In the particular case that $d = 2$ and $q_n(x) = \mathrm{P}(S_n = x)$ with (S_n) being a two dimensional simple random walk, then $Z^{\mathrm{marginal}}_{N,\beta}$ corresponds to the partition function of a two-dimensional directed polymer and the above theorem can be used to give a meaning to the two-dimensional SHE, after mollification of the noise $\dot{W}^{\varepsilon}(t,x) := \int_{\mathbb{R}^2} j_{\varepsilon}(x-y)\dot{W}(y)\mathrm{d}y$ with $j_{\varepsilon}(x) = \varepsilon^{-2} j(x/\varepsilon)$ and proper renormalization

$$\partial_t u_{\varepsilon} = \frac{1}{2}\Delta u_{\varepsilon} + \hat{\beta}\sqrt{\frac{2\pi}{\log\frac{1}{\varepsilon}}}\, \dot{W}^{\varepsilon}\, u_{\varepsilon}, \qquad t > 0, x \in \mathbb{R}^2 \tag{3.18}$$

and fully characterise the limit of the solution u_{ε} in the subcritical regime $\hat{\beta} < 1$.

Via a Hopf-Cole transformation $h_{\varepsilon}(t,x) = \log u_{\varepsilon}(t,x)$ and a suitable approximation scheme this approach also leads to a characterisation of the renormalized two-dimensional KPZ:

$$\partial_t h_{\varepsilon} = \frac{1}{2}\Delta h_{\varepsilon} + \hat{\beta}\sqrt{\frac{2\pi}{\log\frac{1}{\varepsilon}}}\, |\nabla h_{\varepsilon}|^2 + \dot{W}^{\varepsilon},$$

which, for $\hat{\beta} < 1$, turns out to be a gaussian, log-correlated field, [CSZ18b].

References

[A08] K. S. Alexander, The effect of disorder on polymer depinning transitions, *Comm. Math. Phys.*, 279(1), 117-146, (2008).

[AZ09] K. S. Alexander, N. Zygouras, Quenched and annealed critical points in polymer pinning models, *Comm. Math. Phys.*, 291(3), 659-689, (2009).

[B75] W. Beckner, Inequalities in Fourier analysis. *Ann. Math.*, 159-182, (1975).

[BR08] M. Benaim, R. Rossignol, Exponential concentration for first passage percolation through modified Poincaré inequalities. *Annales Henri Poincaré, Prob. et Statistiques* Vol. 44, No. 3, pp. 544-573, (2008).

[BKS03] I. Benjamini, G. Kalai, O. Schramm, First passage percolation has sublinear distance variance, *Ann. Prob.*, 31(4), 1970-1978, (2003).

[BL16] Q. Berger, H. Lacoin, Pinning on a defect line: characterization of marginal disorder relevance and sharp asymptotics for the critical point shift, *Journal of the Institute of Mathematics of Jussieu*, 1-42, (2016).

[B70] A. Bonami, Étude dés coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(2), 335-402, (1970).

[CSZ17a] F. Caravenna, R. Sun, N. Zygouras, Polynomial chaos and scaling limits of disordered systems. *J. Eur. Math. Soc.* 19 (2017), 1–65.

[CSZ17b] F. Caravenna, R. Sun, N. Zygouras, Universality in marginally relevant disordered systems. *Ann. Appl. Probab.* 27 (2017), 3050–3112.

[CSZ18a] F. Caravenna, R. Sun, N. Zygouras, On the moments of the (2+1)-dimensional directed polymer and stochastic heat equation in the critical window. arXiv:1808.03586 (2018).

[CSZ18b] F. Caravenna, R. Sun, N. Zygouras, The two-dimensional KPZ equation in the entire subcritical regime, arxiv1812.03911

[CSZ18+] F. Caravenna, R. Sun, N. Zygouras, Sacling limits of disordered systems and disorder relevance *Proceedings of the XVIII International Congress of Mathematical Physics,* arxiv.1602.05825

[C09] S. Chatterjee, Fluctuations of eigenvalues and second order Poincaré inequalities, *Prob. Th. Rel. Fields*, 143, 1-40, (2009)

[DZ16] P.S. Dey, N.Zygouras; High temperature limits of directed polymers with heavy tail disorder, *Annals of Prob.*, 2016, Vol. 44, No. 6, 4006-4048,

[DGLT09] B. Derrida, G. Giacomin, H. Lacoin, F. L. Toninelli, Fractional moment bounds and disorder relevance for pinning models, *Comm. Math. Phys.*, 287(3), 867-887, (2009).

[G11] G. Giacomin, Disorder and Critical Phenomena Through Basic Probability Models: École d'Été de Probabilitś de Saint-Flour XL-2010 (Vol. 2025), *Springer*,(2011).

[GLT11] G. Giacomin, H. Lacoin, F. L. Toninelli, Disorder relevance at marginality and critical point shift, *Annales Inst. Henri Poincaré, Prob. Stat.* Vol. 47, No. 1, pp. 148-175, (2011).

[GIP15] M. Gubinelli, P. Imkeller, N. Perkowski. Paracontrolled distributions and singular PDEs. *Forum of Mathematics*, Pi (2015) Vol. 3, no. 6.

[H84] P. Hall, Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of multivariate analysis*, 14(1), 1-16, (1984).

[H14] M. Hairer. A theory of regularity structures. *Inventiones Math.* 198, 269–504, 2014.

[H74] A.B. Harris, Effect of random defects on the critical behaviour of Ising models. *Journal of Physics C: Solid State Physics*, 1974.

[HB70] C. C. Heyde, B.M. Brown, On the Departure from Normality of a Certain Class of Martingales, *Ann. of Mathematical Statistics*, 41(6), 2161-2165, (1970).

[H48] W. Hoeffding, A Class of Statistics with Asymptotically Normal Distribution, *The Annals of Mathematical Statistics*, 19(3), 293-325, (1948).

[dJ87] P. de Jong. A central limit theorem for generalized quadratic forms. *Prob. Th. Rel. Fields* 75, 261–277, 1987.

[dJ90] P. de Jong. A central limit theorem for generalized multilinear forms. *J. Multivariate Anal.* 34, 275–289, 1990.

[J97] S. Janson, Gaussian hilbert spaces, *Cambridge University Press*, Vol. 129, (1997).

[K93] H. Kesten, On the speed of convergence in first-passage percolation, *Ann. App. Prob.*, 296-338, (1993).

[K14] A. Kupiainen. Renormalization Group and Stochastic PDE's. arXiv:1410.3094, 2014.

[MOO10] E. Mossel, R. O'Donnell, K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Ann. Math* 171 (2010), 295–341.

[NP09] I. Nourdin, G. Peccati, Stein?s method on Wiener chaos, *Prob. Th. Rel. Fields*, 145(1-2), 75-118, (2009).

[NP05] D. Nualart, G. Peccati, Central limit theorems for sequences of multiple stochastic integrals, Ann. Prob., 33(1), 177-193, (2005).

[N06] D. Nualart, The Malliavin calculus and related topics, *Springer* (2006)

[O'D14] R. O'Donnell, Analysis of boolean functions, *Cambridge University Press*, (2014).

[R74] V. I. Rotar', Some limit theorems for polynomials of second degree, *Theory of Probability & Its Applications*, 18(3), 499-507, (1974).

[S61] B. A. Sevastyanov, A class of limit distributions for quadratic forms of normal stochastic variables *Theor. Probab. Appl,* 6, 337-340, (1961).

[S67] V. Strassen, Almost sure behaviour of sums independent random variables and martingales, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* 2, 315-343, (1967)

[S98] J. Szulga, Introduction to random chaos, *Chapman & Hall, London*, (1998).

[T07] F. L. Toninelli, A replica-coupling approach to disordered pinning models, *Comm. Math. Phys.* 280(2), 389-401, (2008).

[vZ84] W. R. Van Zwet, A Berry-Esseen bound for symmetric statistics, *Zeitschrift für Wahr. Verwandte Gebiete*, 66(3), 425-440, (1984).