Generalized tensorization of $\boldsymbol{\rho}$-mixing

Rémi Peyre

February 13th, 2026

Abstract

The ρ-mixing coefficient (or "maximal correlation") between two random variables X and Y, denoted by $\rho(X;Y)$, is defined as the maximal Pearson correlation coefficient between real variables of the form resp. $f(X)$ and $g(Y)$; it is a way of quantifying to which extent X and Y are correlated or not. It is a classical result that, for (X_0, Y_0) and (X_1, Y_1) two independent couples of random variables, one has $\rho\big((X_0, X_1); (Y_0, Y_1)\big) \leqslant \max\big(\rho(X_0;Y_0), \rho(X_1;Y_1)\big)$. This monograph investigates how tensorization results of this kind can be obtained in the case where independence between (X_0, Y_0) and (X_1, Y_1) is only partial, in the sense that we have bounds on quantities like $\rho(X_0; Y_1)$ and $\rho(X_1; Y_0)$.

How to read this monograph

Chapter 1 is the introduction. In § 1.1 we recall the definition and some classical properties of the concept of ρ-mixing, which is at the core of the monograph; then in §§ 1.2 and 1.3 we expose two different sources of motivation for looking for a generalized tensorization property for ρ-mixing. In all this introductory chapter, no results are new.

Chapter 2 is devoted to the abstract generalized tensorization theorems, which are the main achievement of the monograph. After gathering a few technical lemmas in § 2.1, from § 2.2 to § 2.5 I state and prove my new tensorization results, by increasing order of technicality. The most important result is Theorem 39 in § 2.4; but it is strongly advised to first read § 2.2 (where I give an alternative proof of the classical "strict" tensorization theorem) and 2.3 (devoted to so-called "simple tensorization" results) to get acquainted with the techniques of the proofs of later Theorems 39 and 47; also, it is in § 2.3.1 that the concept of "$\rho^\ddagger$-mixing", which is crucial in the statement of all the main results, is introduced. In § 2.4 I state and prove Theorem 39, which is the centrepiece of the monograph. § 2.5 is devoted to Theorem 47 and Corollary 56, which refine Theorem 39 in the specific case of "parallel tensorization".—Note that the results of § 2.5 will not be actually needed for the applications of Chapter 3. I conclude Chapter 2 with § 2.6, where some refinements and variants of the generalized tensorization theorems are presented, and (in § 2.6.3) optimality of these theorems is discussed.

Chapter 3 is then devoted to applying the results of Chapter 2 to "real-life" models of statistical physics: a general machinery for such applications is first exposed in § 3.1; and then §§ 3.2, 3.3 and 3.4 each deal with one different type of model that our results can be applied to. (§ 3.4 is more specifically devoted to a peculiar kind of application, namely, using the generalized tensorization of ρ-mixing to prove results of *hypocoercivity* type).

In Chapter 4 I make a step to the side, turning my attention towards two topics not directly pertaining to tensorization, but which nevertheless keep the flavour of Chapter 2:

- In § 4.1, I tackle the question: "Do models that have "asymptotic independence" in the ρ-mixing sense satisfy spatial central limit theorems (like when one has actual independence)?". Actually this question is by no means new: the answer, known for long, is essentially: "Yes; but you have to require just a little extra to the ρ-mixing assumption". In my work I will look at a new declension of that question, viz.: "For models that satisfy the conditions to ensure ρ-mixing *by the way of generalized tensorization*, like the ones of Chapter 3, does spatial CLT hold?". The answer will turn out to be positive.

- In § 4.2, I tackle the question: "For models that satisfy the conditions to ensure ρ-mixing by the way of generalized tensorization, does one have spectral gap for the Glauber dynamics?". (Note that, here again, this is nothing more than a new declension of and old topic). Here too, the answer will turn out to be positive; however, this time, I will not use *directly* the tensorization results of

Chapter 2, but rather the *ideas* and *techniques* underlying the proofs of these results.

The interest of Chapter 4 should however not be overhyped: indeed, as regards the *actual* models of particle systems being studied by mathematicians, the results given by that chapter were already known for long, via different techniques… So, it is surely interesting to have an alternative (and somehow more general) frame for establishing spatial CLT and/or spectral gap; but it does not bring that much in practice.

Finally in Chapter 5 I sum up the achievements of the monograph and expose some open questions.

Contents

Notation

Below is some notation used throughout this paper.

Probability

- We shall always work on an implicit measurable space $(\Omega, \mathscr{B})$ endowed with a probability measure $\mathbb{P}$; all along the text, the symbol '$\mathbb{P}$' can be thought as meaning merely "probability". Sub-σ-algebras of $\mathscr{B}$ will be simply called "σ-algebras"; I will also often write "variable" for "random variable". Unless explicitly specified, variables on Ω may be valued in any set.

- When we need to refer to another probability measure for a few lines, this will be indicated by a superscript above the symbol '$\mathbb{P}$' (or '$\mathbb{E}$', 'Var', etc.), in which we tell which variables are assumed to follow which distribution under that measure, in such a way that it totally determines the expression considered. For instance, the formula

$$\operatorname{Var}^{X\sim P}(X) = \tfrac{1}{2}\,\mathbb{E}^{(X,Y)\sim P^{\otimes 2}}\big((Y-X)^2\big) \tag{AA}$$

  is to be read in the following way: "the variance of a random variable X following the probability distribution P is equal to half the expectation of $(Y-X)^2$ when X and Y are two variables following independently the distribution P."

- For X a variable on Ω, the σ-algebra generated by X (that is, the smallest σ-algebra w.r.t. which X is measurable) is denoted by $\sigma(X)$. For $\mathscr{F}$ and $\mathscr{G}$ σ-algebras, the σ-algebra generated by $\mathscr{F}$ and $\mathscr{G}$ (that is, the smallest σ-algebra containing both $\mathscr{F}$ and $\mathscr{G}$) is denoted by $\mathscr{F} \vee \mathscr{G}$; and for an arbitrary number of σ-algebras this notation extends into the ∞-ary operator $\bigvee$.

- An event $A \in \mathscr{B}$ is said to be *trivial* when $\mathbb{P}(A) \in \{0,1\}$. We denote by $\mathscr{N}$ the σ-algebra of all the trivial events. Two σ-algebras $\mathscr{F}$ and $\mathscr{G}$ will be said to be *almost-equal*, which we denote by "$\mathscr{F} \overset{\text{a.s.}}{=} \mathscr{G}$", when they coincide up to trivial events, i.e. when $\mathscr{F} \vee \mathscr{N} = \mathscr{G} \vee \mathscr{N}$. The σ-algebra $\{\varnothing, \Omega\}$, which we denote by $\mathscr{O}$, will be called the trivial σ-algebra; and more generally, a σ-algebra $\mathscr{F}$ will be said to be trivial when all its events are trivial, i.e. when $\mathscr{F} \overset{\text{a.s.}}{=} \mathscr{O}$.

- For T a totally ordered set endowed with a compatible topology, a filtration $(\mathscr{F}_t)_{t\in T}$ will be said to be *almost-right-continuous* when, for all $t \in T$ that is not right-isolated (i.e., that belongs to the adherence of $\{t' \in T \mid t' > t\}$), one has $\lim^{\searrow}_{t'\to t+} \mathscr{F}_{t'} \overset{\text{a.s.}}{=} \mathscr{F}_t$. (Note that the limit of a decreasing sequence of σ-algebras is merely their intersection, when seen as sets of events). In more compact terms, this is equivalent to requiring that the completed filtration $(\mathscr{F}_t \vee \mathscr{N})_{t\in T}$ is right-continuous.

- For f a real random variable, for $\mathscr{G}$ a σ-algebra, the conditional expectation of f w.r.t. $\mathscr{G}$ may be denoted in two ways: either by the classical notation "$\mathbb{E}(f \mid \mathscr{G})$", or by the alternative form "$f^{\mathscr{G}}$". In practice I will use "$\mathbb{E}(f \mid$

$\mathscr{G})$" when seeing conditional expectation as the expectation of f knowing the information of $\mathscr{G}$, while "$f^{\mathscr{G}}$" will be used when I see conditional expectation like the $\mathscr{G}$-measurable function best approximating f: so, both conventions may be used inside the same formula.[*]

- For $\boldsymbol{C}$ a positive-semidefinite matrix (possibly of dimension 1, in which case it will be identified with $\sigma^2 \in \mathbb{R}_+$), $\mathscr{N}(\boldsymbol{C})$ denotes the law of the centred Gaussian vector with covariance matrix $\boldsymbol{C}$. I shall write $\vec{m} + \mathscr{N}(\boldsymbol{C})$ to denote the non-centred Gaussian vector with mean $\vec{m}$ and variance $\boldsymbol{C}$.

Hilbertian and functional analysis

- All the functional spaces considered in this monograph shall be real, unless explicitly stated otherwise.
- When $\mathscr{F}$ is a σ-algebra, $\mathrm{L}^2(\mathscr{F})$ denotes the space of $\mathscr{F}$-measurable real variables (up to a.s. equality) which are square-integrable w.r.t. $\mathbb{P}$. If X is a random variable, $\mathrm{L}^2(X)$ is a shorthand for $\mathrm{L}^2(\sigma(X))$. If I is a countable set, $\ell^2(I)$ denotes the set of functions $f \colon I \to \mathbb{R}$ such that $\sum_{i\in I} f(i)^2 < \infty$. All these spaces are equipped with their natural Hilbertian product $\langle \bullet, \bullet \rangle_{\mathrm{L}^2}$ and the associated norm $\|\bullet\|_{\mathrm{L}^2}$.
- For $\mathscr{F}$ a σ-algebra, in $\mathrm{L}^2(\mathscr{F})$ the set of constant variables makes a line, which can be identified with $\mathbb{R}$; then, $\bar{\mathrm{L}}^2(\mathscr{F})$ denotes the quotient $\mathrm{L}^2(\mathscr{F})/\mathbb{R}$, equipped with its natural Hilbert structure: in other words, if $\bar{f} \in \bar{\mathrm{L}}^2(\mathscr{F})$ is the projection of $f \in \mathrm{L}^2(\mathscr{F})$, one has $\|\bar{f}\|_{\bar{\mathrm{L}}^2} \coloneqq \inf\{\|f-a\|_{\mathrm{L}^2} \mid a \in \mathbb{R}\} = \mathrm{Var}(f)$. Alternatively, $\bar{\mathrm{L}}^2(\mathscr{F})$ can also be seen as the subspace of centred functions of $\mathrm{L}^2(\mathscr{F})$, i.e. as $\{f \in \mathrm{L}^2(\mathscr{F}) \mid \langle f, 1\rangle_{\mathrm{L}^2} = 0\}$: throughout the paper we will implicitly switch between both interpretations.
- For $L \colon E \to F$ a linear operator between two Hilbert spaces with respective norms $\|\bullet\|_E$ and $\|\bullet\|_F$, the operator norm of f (defined as $\sup\{\|Lx\|_F \mid \|x\|_E = 1\}$) is denoted by $|||f|||_{E\to F}$.
- For $L \colon E \circlearrowleft$ a bounded linear operator on a Hilbert space, $\rho_{\mathrm{sp}}(L)$ denotes the spectral radius of L, defined as $\lim_{k\to\infty} |||L^k|||^{1/k}$ (this limit always exists). It is well known that one also has $\rho_{\mathrm{sp}}(L) = \sup\{|\lambda| \bigm| \lambda \in \mathrm{Spec}(L)\}$, where $\mathrm{Spec}(L)$ denotes the spectrum of L.
- For $L \colon H_1 \to H_2$ a linear operator between two Hilbert spaces, $L^* \colon H_2 \to H_1$ denotes the adjoint operator of L, characterized by the property that $\langle L^* y, x\rangle_{H_1} = \langle y, Lx\rangle_{H_2}$.
- A column vector $(a_i)_{i\in I}$ will automatically be identified with the corresponding element of $\ell^2(I)$; likewise, a matrix $\boldsymbol{A} = (\!(a_{ij})\!)_{(i,j)\in I\times J}$ will be identified with the corresponding linear operator from $\ell^2(J)$ to $\ell^2(I)$.

---

[*] By the way, one must not confuse $\mathrm{Var}(f \mid \mathscr{G})$, which is the variance of f under the law $\mathbb{P}(\bullet \mid \mathscr{G})$, with $\mathrm{Var}(f^{\mathscr{G}})$, which is the (unconditioned) variance of the random variable $f^{\mathscr{G}}$.

- In our computations we will often use the Cauchy–Schwarz inequality and its variants[†]: when using such an inequality, we shall indicate it by writing "CS" under the inequality sign concerned.

Miscellaneous

- For a, b real numbers, $a \wedge b$ denotes $\min(a, b)$, resp. $a \vee b$ denotes $\max(a, b)$.
- For $a \leqslant b$ integers, $[\![a, b]\!]$, $]\!]a, b[\![$, $[\![a, b[\![$, $]\!]a, b]\!]$ refer resp. to $[a, b] \cap \mathbb{Z}$, $(a, b) \cap \mathbb{Z}$, $[a, b) \cap \mathbb{Z}$, $(a, b] \cap \mathbb{Z}$.
- For A a matrix, $A^\top$ refers to the transpose of A. Also, $\mathbf{I}_N$ refers to the identity matrix in dimension N.
- For A a set, $\mathbf{1}_A$ denotes the indicator function of A; for B another set, $A \sqcup B$ means the same as $A \cup B$, but it stresses that A and B are disjoint in this case.
- If Θ is a set endowed with a metric *dist*, then for $I, J \subseteq \Theta$, $\mathit{dist}(I, J)$ denotes the distance between I and J, that is, $\mathit{dist}(I, J) \coloneqq \inf\{\mathit{dist}(i, j) \mid i \in I, j \in J\}$. For $i \in \Theta$, $J \subseteq \Theta$, one denotes likewise $\mathit{dist}(i, J) \coloneqq \mathit{dist}(\{i\}, J)$.
- The d-dimensional Lebesgue measure is denoted by $\mathrm{vol}_d(\bullet)$.
- Whenever I is a set and X a symbol, $\vec{X}_I$ will be a shorthand for $(X_i)_{i \in I}$.
- As is customary in physical literature, $\propto$ means "proportional to".
- I will sometimes write "$X \underset{\text{def}}{=} Y$" to stress that such an equality comes from the very definition of X (or of Y). This should not be confused with the notation "$X \coloneqq Y$", which means that X *is being defined* as equal to Y (nor with "$X \eqqcolon Y$", which means that the shorthand Y is being introduced as a substitute for X). Likewise, when I write "$\underset{\text{IH}}{=}$", this means "equal to by induction hypothesis".

[†] For example, the discrete form $\big(\sum_{i=1}^N a_i b_i\big) \leqslant \big(\sum_{i=1}^N a_i^2\big)^{1/2} \big(\sum_{i=1}^N b_i^2\big)^{1/2}$, the probabilistic form $|\mathrm{Cov}(f, g)| \leqslant \mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g)$, etc.

Chapter 1

Introduction

1.1 ρ-mixing

To start, let us recall the topic all this monograph is about, namely ρ-mixing.

Definition 1 (ρ-mixing coefficient, [1, § 1])**.** For $\mathscr{F}, \mathscr{G}$ σ-algebras, the *ρ-mixing coefficient* (or *maximal correlation*) between $\mathscr{F}$ and $\mathscr{G}$ is defined as

$$\rho(\mathscr{F};\mathscr{G}) \coloneqq \sup\{\operatorname{Corr}(f,g) \mid f \in \bar{\mathrm{L}}^2(\mathscr{F}) \setminus \{0\}, g \in \bar{\mathrm{L}}^2(\mathscr{G}) \setminus \{0\}\}, \tag{AB}$$

where $\operatorname{Corr}(f,g) \coloneqq \operatorname{Cov}(f,g) / \operatorname{Var}^{1/2}(f) \operatorname{Var}^{1/2}(g)$ is the classical Pearson correlation coefficient. In the case the supremum in (AB) is taken on an empty set (that is, when $\mathscr{F}$ or $\mathscr{G}$ is trivial), it is considered as being 0.

For X, Y random variables, $\rho(X;Y)$ will implicitly mean $\rho(\sigma(X);\sigma(Y))$. ♡

We now recall and state a couple of elementary properties of ρ-mixing:

Proposition 2 (Immediate properties)**.** *For all σ-algebras $\mathscr{F}$, $\mathscr{G}$ and $\mathscr{G}'$,*

1. $\rho(\mathscr{G};\mathscr{F}) = \rho(\mathscr{F};\mathscr{G})$*;*
2. $\mathscr{G} \subseteq \mathscr{G}' \Rightarrow \rho(\mathscr{F};\mathscr{G}) \leqslant \rho(\mathscr{F};\mathscr{G}')$*;*
3. $\rho(\mathscr{F};\mathscr{G}) \in [0,1]$*;*
4. $\rho(\mathscr{F};\mathscr{G}) = 0$ *if and only if $\mathscr{F}$ and $\mathscr{G}$ are independent;*
5. *If $\mathscr{F}$ is not trivial, then* $\rho(\mathscr{F};\mathscr{F}) = 1$*.* ◇♤

Proposition 3 (ρ-mixing coefficient in the complex setting)**.** *For f and g resp. $\mathscr{F}$- and $\mathscr{G}$-measurable* complex-*valued square-integrable variables, denoting in this context*

$$\operatorname{Var}(f) \coloneqq \mathbb{E}(|f - \mathbb{E}(f)|^2), \tag{AC}$$

resp. (provided neither f nor g is constant)

$$\operatorname{Corr}(f,g) \coloneqq \frac{\mathbb{E}\big((f - \mathbb{E}(f))^*(g - \mathbb{E}(g))\big)}{\operatorname{Var}^{1/2}(f)\operatorname{Var}^{1/2}(g)},^{[*]} \tag{AD}$$

one has, just like in the real case,

$$|\mathrm{Corr}(f,g)| \leqslant \rho(\mathscr{F};\mathscr{G}). \tag{AE}$$

◇

Proof. First, up to multiplying g by an appropriate unitary constant, we can always assume that $\mathrm{Corr}(f,g) = |\mathrm{Corr}(f,g)|$. Now, denote by resp. $f^{\mathfrak{R}}$ and $f^{\mathfrak{I}}$ the real and imaginary parts of f, and likewise for g. Then one has $\mathrm{Var}(f) = \mathrm{Var}(f^{\mathfrak{R}}) + \mathrm{Var}(f^{\mathfrak{I}})$, and likewise $\mathrm{Var}(g) = \mathrm{Var}(g^{\mathfrak{R}}) + \mathrm{Var}(g^{\mathfrak{I}})$; also, one has (using that we assumed $\mathrm{Corr}(f,g)$ to be real)

$$\mathbb{E}\big((f-\mathbb{E}(f))^*(g-\mathbb{E}(g))\big) = \mathrm{Cov}(f^{\mathfrak{R}},g^{\mathfrak{R}}) + \mathrm{Cov}(f^{\mathfrak{I}},g^{\mathfrak{I}}). \tag{AF}$$

Therefore, by the very definition of the ρ-mixing coefficient, we get that

$$\begin{aligned}&\big|\mathbb{E}\big((f-\mathbb{E}(f))^*(g-\mathbb{E}(g))\big)\big| \leqslant \\ &\qquad \rho(\mathscr{F};\mathscr{G})\,\mathrm{Var}^{1/2}(f^{\mathfrak{R}})\,\mathrm{Var}^{1/2}(g^{\mathfrak{R}}) + \rho(\mathscr{F};\mathscr{G})\,\mathrm{Var}^{1/2}(f^{\mathfrak{I}})\,\mathrm{Var}^{1/2}(g^{\mathfrak{I}}) \\ &\qquad \underset{\mathrm{CS}}{\leqslant} \rho(\mathscr{F};\mathscr{G})(\mathrm{Var}(f^{\mathfrak{R}})+\mathrm{Var}(f^{\mathfrak{I}}))^{1/2}(\mathrm{Var}(g^{\mathfrak{R}})+\mathrm{Var}(g^{\mathfrak{I}}))^{1/2} \\ &\qquad\qquad = \rho(\mathscr{F};\mathscr{G})\,\mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g),\end{aligned} \tag{AG}$$

which is what we wanted. ♤

Proposition 4 (ρ-mixing as an operator norm)**.** *Let $\mathscr{F},\mathscr{G}$ be σ-algebras. Obviously the application*

$$\begin{aligned}\pi_{\mathscr{G}\mathscr{F}}\colon \mathrm{L}^2(\mathscr{F}) &\to \mathrm{L}^2(\mathscr{G}) \\ f &\mapsto f^{\mathscr{G}}\end{aligned} \tag{AH}$$

factorizes[†] *into an application $\bar{\pi}_{\mathscr{G}\mathscr{F}}\colon \bar{\mathrm{L}}^2(\mathscr{F}) \to \bar{\mathrm{L}}^2(\mathscr{G})$. Then, the present proposition claims that*

$$\rho(\mathscr{F};\mathscr{G}) = |||\bar{\pi}_{\mathscr{G}\mathscr{F}}|||_{\bar{\mathrm{L}}^2(\mathscr{F})\to\bar{\mathrm{L}}^2(\mathscr{G})}. \tag{AI}$$

Moreover, defining $\pi_{\mathscr{F}\mathscr{G}}$, etc. in the same way as $\pi_{\mathscr{G}\mathscr{F}}$, etc., it is immediate to see that $\pi_{\mathscr{F}\mathscr{G}}$ is the adjoint operator of $\pi_{\mathscr{G}\mathscr{F}}$, so that $\bar{\pi}_{\mathscr{F}\mathscr{G}\mathscr{F}} \coloneqq \bar{\pi}_{\mathscr{F}\mathscr{G}} \circ \bar{\pi}_{\mathscr{G}\mathscr{F}}$ is self-adjoint (and positive-semidefinite); then, (AI) *can be restated into saying that*

$$\rho(\mathscr{F};\mathscr{G}) = \rho_{\mathrm{sp}}(\bar{\pi}_{\mathscr{F}\mathscr{G}\mathscr{F}})^{1/2}. \tag{AJ}$$

◇

Proof. Conditioning a random variable f w.r.t. $\mathscr{G}$ is tantamount to projecting it orthogonally onto $\mathrm{L}^2(\mathscr{G})$, so $\pi_{\mathscr{G}\mathscr{F}}$ is nothing but (the restriction to $\mathrm{L}^2(\mathscr{F})$ of the $\mathrm{L}^2(\mathscr{B})$-defined) orthogonal projection onto $\mathscr{G}$; and since $\mathbb{R}$ belongs both to $\mathrm{L}^2(\mathscr{F})$ and $\mathrm{L}^2(\mathscr{G})$, by factorizing, $\bar{\pi}_{\mathscr{G}\mathscr{F}}$ is nothing but the orthogonal projection from $\bar{\mathrm{L}}^2(\mathscr{F})$ onto $\bar{\mathrm{L}}^2(\mathscr{G})$

---

[*] In this formula, $\bullet^*$ denotes complex conjugation.

[†] If you prefer to see $\bar{\mathrm{L}}^2(\mathscr{F})$ and $\bar{\mathrm{L}}^2(\mathscr{G})$ as subspaces of $\mathrm{L}^2(\mathscr{F})$ and $\mathrm{L}^2(\mathscr{G})$ rather than as quotients, then just replace "factorizes into" by "restricts onto" in the sentence above.

(for the Hilbert structure of $\bar{\mathrm{L}}^2(\mathscr{B})$). Thus, its operator norm is the cosine of the angle between the subspaces $\bar{\mathrm{L}}^2(\mathscr{F})$ and $\bar{\mathrm{L}}^2(\mathscr{G})$, i.e.

$$\begin{aligned}\cos\inf\{\angle(f,g) \mid f\in\bar{\mathrm{L}}^2(\mathscr{F})\smallsetminus\{0\}, g\in\bar{\mathrm{L}}^2(\mathscr{G})\smallsetminus\{0\}\} \\ \underset{\text{def}}{=} \sup\left\{\frac{\langle f,g\rangle_{\bar{\mathrm{L}}^2}}{\|f\|_{\bar{\mathrm{L}}^2}\|g\|_{\bar{\mathrm{L}}^2}} \;\middle|\; f\in\bar{\mathrm{L}}^2(\mathscr{F})\smallsetminus\{0\}, g\in\bar{\mathrm{L}}^2(\mathscr{G})\smallsetminus\{0\}\right\} \\ \underset{\text{def}}{=} \sup\left\{\frac{\mathrm{Cov}(f,g)}{\mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g)} \;\middle|\; f\in\bar{\mathrm{L}}^2(\mathscr{F})\smallsetminus\{0\}, g\in\bar{\mathrm{L}}^2(\mathscr{G})\smallsetminus\{0\}\right\} \\ \underset{\text{def}}{=} \sup\{\mathrm{Corr}(f,g) \mid f\in\bar{\mathrm{L}}^2(\mathscr{F})\smallsetminus\{0\}, g\in\bar{\mathrm{L}}^2(\mathscr{G})\smallsetminus\{0\}\} \underset{\text{def}}{=} \rho(f;g). \quad ♤ \end{aligned} \tag{AK}$$

Corollary 5 (Sub-multiplicativity of ρ-mixing for Markov chains)**.** *If $X\to Y\to Z$ is a (not necessarily homogeneous) Markov chain, then*

$$\rho(X;Z)\leqslant\rho(X;Y)\,\rho(Y;Z). \tag{AL}$$

(Note that Equation (`AL`) *implies in particular that $\rho(X;Z)\leqslant\rho(X;Y)$).* ◇

Proof. Denote resp. $\sigma(X)=:\mathscr{F},\sigma(Y)=:\mathscr{G},\sigma(Z)=:\mathscr{H}$. It is a classical property (but we will justify it below anyway) that the Markov property for the chain $X\to Y\to Z$ actually translates into the following equality between linear operators from $\mathrm{L}^2(\mathscr{H})$ to $\mathrm{L}^2(\mathscr{F})$:

$$\pi_{\mathscr{F}\mathscr{H}}=\pi_{\mathscr{F}\mathscr{G}}\circ\pi_{\mathscr{G}\mathscr{H}}. \tag{AM}$$

That equality factorizes into "$\bar\pi_{\mathscr{F}\mathscr{H}}=\bar\pi_{\mathscr{F}\mathscr{G}}\circ\bar\pi_{\mathscr{G}\mathscr{H}}$": hence, by sub-multiplicativity of operator norms, $|\!|\!|\bar\pi_{\mathscr{F}\mathscr{H}}|\!|\!|_{\bar{\mathrm{L}}^2(\mathscr{F})\to\bar{\mathrm{L}}^2(\mathscr{H})}\leqslant|\!|\!|\bar\pi_{\mathscr{F}\mathscr{G}}|\!|\!|_{\bar{\mathrm{L}}^2(\mathscr{F})\to\bar{\mathrm{L}}^2(\mathscr{G})}|\!|\!|\bar\pi_{\mathscr{G}\mathscr{H}}|\!|\!|_{\bar{\mathrm{L}}^2(\mathscr{G})\to\bar{\mathrm{L}}^2(\mathscr{H})}$; and finally (`AL`) appears as a direct consequence of Proposition 4.

To check Equation (`AM`), consider an L^2 $\mathscr{H}$-measurable variable $h=:\eta(Z)$; and denote $g:=h^{\mathscr{G}}=:\gamma(Y)$, $f:=h^{\mathscr{F}}=:\varphi(X)$, $f':=g^{\mathscr{F}}=:\varphi'(X)$. We have to check that $f\equiv f'$, i.e. that $\varphi(x)=\varphi'(x)$ for (Law(X)-almost-)all x.

By definition of conditional expectation, $\gamma(y)=\mathbb{E}(\eta(Z)\mid Y=y)$ for (almost-) all y; likewise, $\varphi(x)=\mathbb{E}(\eta(Z)\mid X=x)$ and $\varphi'(x)=\mathbb{E}(\gamma(Y)\mid X=x)$ for all x. But, since the Markov property states that $\mathrm{Law}(Z\mid X=x \text{ and } Y=y)=\mathrm{Law}(Z\mid Y=y)$, we have the following chain of equalities:

$$\begin{aligned}\varphi(x)=\mathbb{E}(\eta(Z)\mid X=x)=\int_y\mathbb{E}(\eta(Z)\mid X=x\text{ and }Y=y)\,\mathbb{P}(Y\in\mathrm{d}y\mid X=x) \\ =\int_y\mathbb{E}(\eta(Z)\mid Y=y)\,\mathbb{P}(Y\in\mathrm{d}y\mid X=x) \\ =\int_y\gamma(Y)\,\mathbb{P}(Y\in\mathrm{d}y\mid X=x)=\mathbb{E}(\gamma(Y)\mid X=x)=\varphi'(x). \quad ♤\end{aligned} \tag{AN}$$

Proposition 6 (ρ-mixing in the finite-range case)**.** *Let X and Y be random variables with finite ranges resp. $[\![0,N[\![$ and $[\![0,M[\![$, and for all $x\in[\![0,N[\![, y\in[\![0,M[\![$, denote $p_x:=\mathbb{P}(X=x), p^y:=\mathbb{P}(Y=y), p_x^y:=\mathbb{P}(X=x$ and $Y=y)$. Then $\rho(X;Y)=|\!|\!|\boldsymbol{R}|\!|\!|$, where $\boldsymbol{R}$ is the $N\times M$ matrix with general entry*

$$\mathfrak{r}_{xy}:=\frac{p_x^y-p_xp^y}{(p_xp^y)^{1/2}}. \tag{AO}$$

◇

Remark 7. In particular, if both X and Y have range $\{0,1\}$, then, using the same notation as before, one has:

$$\uprho(X;Y) = \frac{|p_x^y - p_x p^y|}{(p_0 p_1 p^0 p^1)^{1/2}}, \tag{AP}$$

where the right-hand side of (AP) does not depend on the choice of $(x,y) \in \{0,1\}^2$. ♧

Proof of Proposition 6. By Proposition 4, $\uprho(X;Y)$ is the norm of the operator $\pi_{XY}\colon \bar{\mathrm{L}}^2(Y) \to \bar{\mathrm{L}}^2(X)$. Here it will be more convenient to work in L^2 spaces than in $\bar{\mathrm{L}}^2$ spaces, so we rather compute the norm of

$$\begin{aligned} \tilde{\pi}: \mathrm{L}^2(Y) &\to \mathrm{L}^2(X) \\ g &\mapsto g^{\sigma(X)} - \mathbb{E}(g), \end{aligned} \tag{AQ}$$

which is obviously the same as $\vert\!\vert\!\vert \pi_{XY} \vert\!\vert\!\vert$.

A function $g \in \mathrm{L}^2(Y)$ can be identified with a $\boldsymbol{M}$-dimensional vector also denoted by $\vec{g}$, and similarly $\tilde{\pi} g \in \mathrm{L}^2(X)$ can be identified with a $\boldsymbol{N}$-dimensional vector. Denote resp. $\boldsymbol{P} \coloneqq \llparenthesis p_x^y \rrparenthesis_{x,y} \in \mathbb{R}^{N \times M}$, $\boldsymbol{\Delta}_X \coloneqq \llparenthesis \updelta_{x=x'} p_x \rrparenthesis_{x,x'} \in \mathbb{R}^{N\times N}$, $\boldsymbol{\Delta}_Y \coloneqq \llparenthesis \updelta_{y=y'} p^y \rrparenthesis_{y,y'} \in \mathbb{R}^{M \times M}$, $\vec{1}_N \coloneqq 1^{\llbracket 0,N \llbracket} \in \mathbb{R}^N$. Applying Bayes' formula yields that

$$\tilde{\pi} g = \boldsymbol{\Delta}_X^{-1} \boldsymbol{P} \vec{g} - \vec{1}_N \vec{1}_N^{\top} \boldsymbol{P} \vec{g}. \tag{AR}$$

Now, $\|g\|_{\mathrm{L}^2(Y)} = \|\boldsymbol{\Delta}_Y^{1/2} \vec{g}\|$, resp. $\|\tilde{\pi} g\|_{\mathrm{L}^2(X)} = \|\boldsymbol{\Delta}_X^{1/2} \overrightarrow{\tilde{\pi} g}\|$, so:

$$\uprho(X;Y) = \sup_{\vec{g} \neq \vec{0}} \frac{\|(\boldsymbol{\Delta}_X^{-1/2} \boldsymbol{P} - \boldsymbol{\Delta}_X^{1/2} \vec{1}_N \vec{1}_N^{\top} \boldsymbol{P}) \vec{g}\|}{\|\boldsymbol{\Delta}_Y^{1/2} \vec{g}\|}. \tag{AS}$$

Performing the change of variables $\vec{h} = \boldsymbol{\Delta}_Y^{1/2} \vec{g}$, (AS) becomes

$$\uprho(X;Y) = \sup_{\vec{h} \neq \vec{0}} \frac{\|\boldsymbol{R} \vec{h}\|}{\|\vec{h}\|} = \vert\!\vert\!\vert \boldsymbol{R} \vert\!\vert\!\vert, \tag{AT}$$

with

$$\boldsymbol{R} \coloneqq \boldsymbol{\Delta}_X^{-1/2} \boldsymbol{P} \boldsymbol{\Delta}_Y^{-1/2} - \boldsymbol{\Delta}_X^{1/2} \vec{1}_N \vec{1}_N^{\top} \boldsymbol{P} \boldsymbol{\Delta}_Y^{-1/2}, \tag{AU}$$

which is Equation (AO) indeed. ♤

Lemma 8. *Let X, Y be two random variables with respective ranges $\mathscr{X}, \mathscr{Y}$ both of cardinality* 2*. Then, there exist $x \in \mathscr{X}, y \in \mathscr{Y}$ such that:*

$$\uprho(X;Y) \leqslant 1 - 4\,\mathbb{P}(X = x \text{ and } Y = y). \tag{AV}$$

◇

Proof. Let us use again the notation of Proposition 6. Since the difference $p_x^y - p_x p^y$ gets its sign changed whenever x, resp. y, changes, there are some x and y for which this value is nonpositive; moreover, denoting by $\{x, x'\}$ and $\{y, y'\}$ the respective ranges of X and Y, $p_{x'}^{y'} - p_{x'}p^{y'}$ is also nonpositive. Now one has

$$\frac{p_x p^y}{(p_x p_{x'} p^y p^{y'})^{1/2}} \times \frac{p_{x'} p^{y'}}{(p_x p_{x'} p^y p^{y'})^{1/2}} = 1, \tag{AW}$$

so that either $p_x p^y$ or $p_{x'} p^{y'}$ is $\leqslant (p_x p_{x'} p^y p^{y'})^{1/2}$. Up to changing notation we can assume that it is $p_x p^y$, and then we get that

$$\rho(X;Y) = \frac{|p_x^y - p_x p^y|}{(p_x p_{x'} p^y p^{y'})^{1/2}} = \frac{p_x p^y - p_x^y}{(p_x p_{x'} p^y p^{y'})^{1/2}} \leqslant 1 - \frac{p_x^y}{(p_x p_{x'} p^y p^{y'})^{1/2}} \leqslant 1 - 4p_x^y. \tag{AX}$$

♤

Proposition 9 (ρ-mixing in the Gaussian case, [2, Theorem 1])**.** *For $(\vec{X}, \vec{Y})$ a $(N+M)$-dimensional Gaussian vector whose covariance matrix writes blockwise*

$$\operatorname{Var}\big((\vec{X}, \vec{Y})\big) = \begin{pmatrix} \mathrm{I}_N & C \\ C^\top & \mathrm{I}_M \end{pmatrix}, \tag{AY}$$

one has $\rho(\vec{X}; \vec{Y}) = |||C|||$. *In particular, if (X, Y) is a 2-dimensional Gaussian vector,* $\rho(X;Y) = |\mathrm{Corr}(X,Y)|$. ◇♤

1.2 First motivation: On Ising's model

My initial motivation for this monograph was about models of statistical physics, like the celebrated Ising one. Ising's model represents a ferromagnetic material:

Definition 10 (Ising model, [3, § 1.3])**.** For d an integer, consider the lattice $\mathbb{Z}^d$ endowed with its usual graph structure (where each vertex has $2d$ neighbours) and with the graph distance. Define $\Sigma = \{\pm 1\}^{\mathbb{Z}^d}$, and for $\vec{\sigma} \in \Sigma$, set formally:

$$\mathscr{H}(\vec{\sigma}) \coloneqq -\frac{1}{2} \sum_{dist(i,j)=1} \sigma_i \sigma_j. \tag{AZ}$$

Then, for $\beta \geqslant 0$, the *(zero-field) Ising model on $\mathbb{Z}^d$ at reciprocal temperature β* is, formally, a probability measure $\mathbb{P}$ on Σ such that $\mathbb{P}(\vec{\sigma}) \propto \exp(-\beta \mathscr{H}(\vec{\sigma}))$. In rigorous terms, saying that $\mathbb{P}$ is an equilibrium measure for Ising's model means that for all $i \in \mathbb{Z}^d$, for all $\vec{s}_{\mathbb{Z}^d \setminus \{i\}} \in \{\pm 1\}^{\mathbb{Z}^d \setminus \{i\}}$,

$$\mathbb{P}(\sigma_i = s_i \mid \vec{\sigma}_{\mathbb{Z}^d \setminus \{i\}} = \vec{s}_{\mathbb{Z}^d \setminus \{i\}}) \propto \exp\Big(\beta \sum_{dist(i,j)=1} s_i s_j\Big). \tag{BA}$$

The variables σ_i of Ising's model will be called *spins*. ♡

Ising's model and the phase transition it exhibits have been the subject of dozens of works; see [3] for an overview. Here we will be interested in the subcritical regime:

Definition 11 (Subcritical regime)**.** It has been shown [4] that there is a $\beta_{\mathrm{c}} > 0$ (the "Curie reciprocal temperature") such that the solution of (BA) is unique as soon as $\beta < \beta_{\mathrm{c}}$. Then, one says that they are in the *subcritical regime.* ♡

An interesting feature of the subcritical regime is that for distant small subsets I and J, the random variables $\vec{\sigma}_I$ and $\vec{\sigma}_J$ are "almost independent", with exponential decay of correlations. That phenomenon, called *weak mixing property*, is stated by the following proposition:

Proposition 12 (Weak mixing property, [5, Theorem 2-(i)])**.** *For (zero-field) Ising's model on $\mathbb{Z}^d$ at some reciprocal temperature in the subcritical regime, there exist $\gamma > 0$ and $C < \infty$ such that for all disjoint $I, J \subseteq \mathbb{Z}^d$:*

$$\|\mathrm{Law}(\vec{\sigma}_{I\sqcup J}) - \mathrm{Law}(\vec{\sigma}_I) \otimes \mathrm{Law}(\vec{\sigma}_J)\|_{\mathrm{TV}} \leqslant C \sum_{(i,j)\in I\times J} \exp\bigl(-\gamma\, dist(i,j)\bigr), \tag{BB}$$

where $\|\bullet\|_{\mathrm{TV}}$ denotes the total variation norm. ◇♤

Nonetheless that phenomenon of exponential decay of correlations is not powerful enough to get decay of correlation between distant *infinite* bunches of spins, at least in the sense of "β-mixing":

Definition 13 (β-mixing coefficient)**.** For X and Y random variables valued in any sets, the *β-mixing coefficient* between X and Y is the total variation distance between the joint law $\mathrm{Law}(X, Y)$ and the product law $\mathrm{Law}(X) \otimes \mathrm{Law}(Y)$. (That distance actually only depends on $\sigma(X)$ and $\sigma(Y)$, so that one may rather speak of β-mixing between *σ-algebras*). The minimum possible value for this coefficient is 0, meaning that X and Y are independent, while the maximum possible value is 1, meaning that $\mathrm{Law}(X, Y)$ is singular to $\mathrm{Law}(X) \otimes \mathrm{Law}(Y)$. For more on β-mixing, see [1, § 1]. ♡

Proposition 14 (Absence of β-mixing)**.** *Let $d > 1$ and consider the (zero-field) Ising model on $\mathbb{Z}^d$. Then, for all $0 < \beta < \beta_{\mathrm{c}}$, for all $x > 0$, denoting $I \coloneqq \{0\} \times \mathbb{Z}^{d-1}$ and $J \coloneqq \{x\} \times \mathbb{Z}^{d-1}$, the joint law of $\vec{\sigma}_I$ and $\vec{\sigma}_J$ is singular to their product law. In other words, there is no β-mixing between hyperplanes for Ising's model.* ◇

Proof. Our goal is to show that $\mathrm{Law}(\vec{\sigma}_I, \vec{\sigma}_J)$ and $\mathrm{Law}(\vec{\sigma}_I) \otimes \mathrm{Law}(\vec{\sigma}_J)$ are mutually singular, i.e. that $\|\mathrm{Law}(\vec{\sigma}_I, \vec{\sigma}_J) - \mathrm{Law}(\vec{\sigma}_I) \otimes \mathrm{Law}(\vec{\sigma}_J)\|_{\mathrm{TV}} = 2$. Since obviously $\|\mathrm{Law}(\vec{\sigma}_I, \vec{\sigma}_J) - \mathrm{Law}(\vec{\sigma}_I) \otimes \mathrm{Law}(\vec{\sigma}_J)\|_{\mathrm{TV}} \geqslant \|\mathrm{Law}(\vec{\sigma}_{I'}, \vec{\sigma}_{J'}) - \mathrm{Law}(\vec{\sigma}_{I'}) \otimes \mathrm{Law}(\vec{\sigma}_{J'})\|_{\mathrm{TV}}$ for all $I' \subseteq I$, $J' \subseteq J$, we are looking for relevant I' and J' to bound their β-mixing coefficient from below. In the sequel, we will take I' and J' of the respective forms $I_N^{(\delta)}$ and $J_N^{(\delta)}$, for very large values of N and δ, where

$$I_N^{(\delta)} \coloneqq \{i_n^{(\delta)} \mid n \in [\![0, N[\![\}, \qquad i_n^{(\delta)} \coloneqq (0, \vec{0}_{d-2}, n\delta); \tag{BC}$$

$$J_N^{(\delta)} \coloneqq \{j_n^{(\delta)} \mid n \in [\![0, N[\![\}, \qquad j_n^{(\delta)} \coloneqq (x, \vec{0}_{d-2}, n\delta): \tag{BD}$$

so, informally, $i_n^{(\delta)}$ and $j_n^{(\delta)}$ are points of the respective hyperplanes facing each other, and all the pairs $(i_n^{(\delta)}, j_n^{(\delta)})$ are very distant from each other.

First, let us look at what happens at points $i_0^{(\delta)}$ and $j_0^{(\delta)}$: as these points do not actually depend on δ, we will denote them by merely resp. i_0 and j_0. Since in the subcritical regime the equilibrium measure is unique, by symmetry one has $\mathbb{P}(\sigma_{i_0} = -1) = \mathbb{P}(\sigma_{i_0} = +1) = \mathbb{P}(\sigma_{j_0} = -1) = \mathbb{P}(\sigma_{j_0} = +1) = 1/2$. Let $p \coloneqq \mathbb{P}(\sigma_{i_0} = \sigma_{j_0})$: interpretation of Ising's model as a random-cluster model (cf. [3, § 1.4]) ensures that $p > 1/2$. Let us also denote $\boldsymbol{P}_p \coloneqq \mathrm{Law}(\sigma_{i_0}, \sigma_{j_0})$, seen as a distribution on $\{\pm 1\}^2$.

By translation invariance, for each n, $\mathrm{Law}(\sigma_{i_n^{(\delta)}}, \sigma_{j_n^{(\delta)}}) = \boldsymbol{P}_p$. But now, considering N to be fixed, let $\delta \to \infty$: then, because of the weak mixing property of Ising's model (Proposition 12), the couples $(\sigma_{i_n^{(\delta)}}, \sigma_{j_n^{(\delta)}})$ tend to become all independent, so that

$$\mathrm{Law}(\vec{\sigma}_{I_N^{(\delta)}}, \vec{\sigma}_{J_N^{(\delta)}}) \underset{\text{def}}{=} \mathrm{Law}(\sigma_{i_0^{(\delta)}}, \sigma_{j_0^{(\delta)}}, \sigma_{i_1^{(\delta)}}, \sigma_{j_1^{(\delta)}}, \ldots, \sigma_{i_{N-1}^{(\delta)}}, \sigma_{j_{N-1}^{(\delta)}}) \overset{\delta\to\infty}{\to} \boldsymbol{P}_p^{\otimes N}, \tag{BE}$$

where the probability distributions in the above equations are considered as laws on $\{\pm 1\}^{2N}$, the latter being identified with either $(\{\pm 1\}^N)^2$ or $(\{\pm 1\}^2)^N$ depending on the context.

Another consequence of (BE) is that $\mathrm{Law}(\vec{\sigma}_{I_N^{(\delta)}})$ tends to $\mathrm{Unif}(\{\pm 1\})^{\otimes N}$ as $\delta \to \infty$; therefore, $\mathrm{Law}(\vec{\sigma}_{I_N^{(\delta)}}) \otimes \mathrm{Law}(\vec{\sigma}_{J_N^{(\delta)}}) \to \mathrm{Unif}(\{\pm 1\})^{\otimes 2N} \eqqcolon \boldsymbol{P}_{1/2}^{\otimes N}$.

But the law of $\sum_{n=0}^{N-1} \mathbf{1}_{\sigma_{i_n} = \sigma_{j_n}}$ is different depending on whether one lives under the law $\boldsymbol{P}_p^{\otimes N}$ or $\boldsymbol{P}_{1/2}^{\otimes N}$: indeed,

$$\mathrm{Law}^{(\sigma_{i_n}, \sigma_{j_n})_n \sim \boldsymbol{P}_p^{\otimes N}} \Bigl(\sum_{k=0}^{N-1} \mathbf{1}_{\sigma_{i_n} = \sigma_{j_n}} \Bigr) = \mathrm{Bin}^{\mathrm{ial}}(N, p), \tag{BF}$$

while

$$\mathrm{Law}^{(\sigma_{i_n}, \sigma_{j_n})_n \sim \boldsymbol{P}_{1/2}^{\otimes N}} \Bigl(\sum_{n=0}^{N-1} \mathbf{1}_{\sigma_{i_n} = \sigma_{j_n}} \Bigr) = \mathrm{Bin}^{\mathrm{ial}}(N, 1/2). \tag{BG}$$

As, for measures on a given finite set, the total variation distance is continuous w.r.t. the convergence of measures, it follows that

$$\varlimsup_{\delta\to\infty} \|\mathrm{Law}(\vec{\sigma}_{I_N^{(\delta)}}, \vec{\sigma}_{J_N^{(\delta)}}) - \mathrm{Law}(\vec{\sigma}_{I_N^{(\delta)}}) \otimes \mathrm{Law}(\vec{\sigma}_{J_N^{(\delta)}})\|_{\mathrm{TV}} \geqslant \|\mathrm{Bin}^{\mathrm{ial}}(N, p) - \mathrm{Bin}^{\mathrm{ial}}(N, 1/2)\|_{\mathrm{TV}}, \tag{BH}$$

hence

$$\|\mathrm{Law}(\vec{\sigma}_I, \vec{\sigma}_J) - \mathrm{Law}(\vec{\sigma}_I) \otimes \mathrm{Law}(\vec{\sigma}_J)\|_{\mathrm{TV}} \geqslant \|\mathrm{Bin}^{\mathrm{ial}}(N, p) - \mathrm{Bin}^{\mathrm{ial}}(N, 1/2)\|_{\mathrm{TV}}. \tag{BI}$$

But, fixing $p' \in (1/2, p)$ independently from N,

$$\|\mathrm{Bin}^{\mathrm{ial}}(N, p) - \mathrm{Bin}^{\mathrm{ial}}(N, 1/2)\|_{\mathrm{TV}} \geqslant 2 \bigl| \mathbb{P}\bigl(\mathrm{Bin}^{\mathrm{ial}}(N, p) > p'\bigr) - \mathbb{P}\bigl(\mathrm{Bin}^{\mathrm{ial}}(N, 1/2) > p'\bigr) \bigr|, \tag{BJ}$$

which tends to 2 as $N \to \infty$ by the law of large numbers; so finally one has necessarily that the left-hand side of (BI) is equal to 2, QED. ♤

So, Proposition 14 shows that there is no β-mixing at all between parallel hyperplanes for the subcritical Ising model. But on the other hand, if we measure decorrelation by the means of ρ-mixing, then we do get exponential decay of correlation between hyperplanes:

Proposition 15 (ρ-mixing for hyperplanes)**.** *Let* $d > 1$*. For (zero-field) Ising's model on* $\mathbb{Z}^d$ *in the subcritical regime, defining as before* $\boldsymbol{I} \coloneqq \{0\}\times\mathbb{Z}^{d-1}$ *and* $\boldsymbol{J} \coloneqq \{x\}\times\mathbb{Z}^{d-1}$ *for some* $x > 0$*, one has:*

$$\mathrm{\rho}(\vec{\sigma}_{\boldsymbol{I}};\vec{\sigma}_{\boldsymbol{J}}) \leqslant \mathrm{e}^{-\gamma x}, \tag{BK}$$

where γ *is the same as in Proposition 12.* ◇

Proof. Define the operator

$$\begin{aligned} \boldsymbol{P}\colon\ \bar{\mathrm{L}}^2(\vec{\sigma}_{\boldsymbol{I}}) &\to \bar{\mathrm{L}}^2(\vec{\sigma}_{\boldsymbol{J}}) \\ f &\mapsto f^{\sigma(\vec{\sigma}_{\boldsymbol{J}})}. \end{aligned} \tag{BL}$$

Then by Proposition 4, proving (BK) is equivalent to proving that $|||\boldsymbol{P}||| \leqslant \mathrm{e}^{-\gamma x}$. Now for all $t \in \{0,\ldots,x\}$, denote $\sigma_{(t)} \coloneqq \vec{\sigma}_{\{t\}\times\mathbb{Z}^{d-1}}$, and for all $t \in \{1,\ldots,x\}$, define

$$\begin{aligned} \pi_t :\ \bar{\mathrm{L}}^2(\sigma_{(t-1)}) &\to \bar{\mathrm{L}}^2(\sigma_{(t)}) \\ f &\mapsto f^{\sigma(\sigma_{(t)})}. \end{aligned} \tag{BM}$$

Due to the fact that the interactions in Ising's model have only range 1, $\sigma_{(0)} \to \sigma_{(1)} \to \cdots \to \sigma_{(x)}$ is a Markov chain, and therefore:

$$\boldsymbol{P} = \pi_x \circ \cdots \circ \pi_2 \circ \pi_1. \tag{BN}$$

Now, by horizontal translation all the $\bar{\mathrm{L}}^2(\sigma_{(t)})$ can be identified with a common Hilbert space $\boldsymbol{H}$. Then all the π_t's are identified with operators on $\boldsymbol{H}$, and by the translation invariance of the model all these operators are equal. $\boldsymbol{P}$ is also identified with an operator on $\boldsymbol{H}$, and (BN) becomes:

$$\boldsymbol{P} = \pi^x. \tag{BO}$$

But π is self-adjoint because, as the model is also invariant by reflection, the stationary Markov chain $\sigma_{(0)} \to \cdots \to \sigma_{(x)}$ is reversible. In particular π is a normal operator, and thus $|||\boldsymbol{P}||| = |||\pi|||^x$. So, proving that $|||\boldsymbol{P}||| \leqslant \mathrm{e}^{-\gamma x}$ is equivalent to proving that $|||\pi||| \leqslant \mathrm{e}^{-\gamma}$, which will be our new goal.

So, take $\boldsymbol{C} < \infty$ as in Proposition 12. For ℓ an integer, denote $\boldsymbol{I}_\ell$ and $\boldsymbol{J}_\ell$ to be resp. $\{0\}\times\{-\ell,\ldots,\ell\}^{d-1}$ and $\{x\}\times\{-\ell,\ldots,\ell\}^{d-1}$. Let f be a bounded function of $\bar{\mathrm{L}}^2(\vec{\sigma}_{\boldsymbol{I}_\ell})$ and denote $\boldsymbol{M} \coloneqq \|f\|_{\mathrm{L}^\infty}$. By translation, f can also be identified with a function of $\bar{\mathrm{L}}^2(\vec{\sigma}_{\boldsymbol{J}_\ell})$, which is bounded by $\boldsymbol{M}$ too. Now, since

$$\mathrm{Cov}(f(\vec{\sigma}_{\boldsymbol{I}_\ell}), f(\vec{\sigma}_{\boldsymbol{J}_\ell})) = \int f(\vec{\sigma}_{\boldsymbol{I}_\ell})f(\vec{\sigma}_{\boldsymbol{J}_\ell})\,\mathrm{d}(\mathrm{Law}(\vec{\sigma}_{\boldsymbol{I}_\ell\sqcup\boldsymbol{J}_\ell}) - \mathrm{Law}(\vec{\sigma}_{\boldsymbol{I}_\ell})\otimes\mathrm{Law}(\vec{\sigma}_{\boldsymbol{J}_\ell})), \tag{BP}$$

we can apply (BB) to $\boldsymbol{I}_\ell$ and $\boldsymbol{J}_\ell$ to obtain:

$$|\mathrm{Cov}(f(\vec{\sigma}_{\boldsymbol{I}_\ell}), f(\vec{\sigma}_{\boldsymbol{J}_\ell}))| \leqslant \boldsymbol{M}^2 \cdot 2\boldsymbol{C}(2\ell+1)^{2(d-1)}\mathrm{e}^{-\gamma x}. \tag{BQ}$$

In terms of operators, (BQ) means that

$$|\langle f, \boldsymbol{P}f\rangle_{\boldsymbol{H}}| \leqslant 2(2\ell+1)^{2(d-1)}\boldsymbol{M}^2\boldsymbol{C}\mathrm{e}^{-\gamma x}. \tag{BR}$$

As the value of x played no particular role in establishing (BR), that formula can be generalized into:

$$|\langle f, \pi^t f \rangle_H| \leqslant 2(2\ell + 1)^{2(d-1)} M^2 C \mathrm{e}^{-\gamma t} \tag{BS}$$

for all $t \in \mathbb{N}^*$. Letting $t \to \infty$, we thus obtain that for all ℓ, for all $f \in \bar{\mathrm{L}}^2(\vec{\sigma}_{I_\ell}) \cap \mathrm{L}^\infty$,

$$\overline{\lim_{t\to\infty}} |\langle f, \pi^t f \rangle|^{1/t} \leqslant \mathrm{e}^{-\gamma}. \tag{BT}$$

But $\bigcup_{\ell\in\mathbb{N}}(\bar{\mathrm{L}}^2(\vec{\sigma}_{I_\ell}) \cap \mathrm{L}^\infty)$ is a dense subset of H, so, since π is a normal operator, according to Lemma 161 in appendix this is enough to conclude that $|||\pi|||_H \leqslant \mathrm{e}^{-\gamma}$, which is what we wanted. ♤

So, we have seen that for the subcritical Ising model, there is some ρ-mixing between distant hyperplanes; we have also seen that this ρ-mixing property is all the more important that β-mixing does not hold. Then a natural question is, what is the level of generality of Proposition 15? In other words, which of its assumptions can be relaxed and which cannot?

To carry the above proof, actually we needed at least the following properties:

- To identify all the spaces $\bar{\mathrm{L}}^2(\sigma_{(t)})$, we used that I and J had the same shape and that one could tile $\mathbb{Z}^d$ (at least partly) with a sequence of tiles having that shape (namely, here, tiles of the form $I_t \coloneqq \{t\} \times \mathbb{Z}^{d-1}$).
- To introduce the Markov chain $\sigma_{(0)} \to \cdots \to \sigma_{(x)}$, we used that the interactions of our model had finite range, so that it was not possible get from one side of the tile I_t to the other side by a chain of interactions without crossing the tile I_t itself.
- To say that all the π_t were the same modulo identification, we used the translation invariance of the model.
- To state that the stationary Markov chain $\sigma_{(0)} \to \cdots \to \sigma_{(x)}$ was reversible, we used the reflection invariance of the model.
- To get a bound on $|||\pi|||$ from the asymptotics of $\langle f, \pi^t f \rangle$, we used that the decay of correlations was exponential.

All these points would make the proof we gave of Proposition 15 quite difficult to generalize. What, for instance, if we take I and J with arbitrary shapes, just requiring that $\mathit{dist}(I, J) \geqslant x$? What if we consider statistical physics models with infinite-range interactions? Etc. The above arguments would not work any more… Yet, we do not have the impression that the presence of ρ-mixing fundamentally relies on the peculiar symmetries of the case we have been treating!

So, my first motivation for this work was to establish ρ-mixing estimates by the means of general methods. Thanks to this approach, I will indeed be able to obtain fairly new decorrelation results in statistical physics: Theorem 105 will be a quite general concrete example of such a result. As you will see, I have achieved my goal by focusing on the properties of ρ-mixing "for itself", rather than on its links with other forms of decorrelation.

1.3 Second motivation: Tensorization of ρ-mixing

The following “tensorization property” of ρ-mixing, a proof of which we will recall, is well known:

Proposition 16 (“Strict” tensorization of ρ-mixing, [6, Theorem 6.2])**.** *Let I be a set and let $\big((X_i,Y_i)\big)_{i\in I}$ be independent couples of random variables.*[‡] *Then:*

$$\rho(\vec{X}_I;\vec{Y}_I) = \sup_{i\in I}\rho(X_i;Y_i). \tag{BU}$$

◇

Remark 17. The strength of Proposition 16 is that the right-hand side of (BU) is a *supremum*: so, we get a result of decorrelation between $\vec{X}_I$ and $\vec{Y}_I$ *which does not depend on the cardinality of I*. ♧

Proof of Proposition 16. This proposition is quite immediate to prove. Indeed, consider the projection operators $\pi_i\colon \mathrm{L}^2(X_i)\to\mathrm{L}^2(Y_i)$ defined as in Proposition 4. Then, independence between the X_i's implies that the space $\mathrm{L}^2(\vec{X}_I)$ may be seen as the tensor product $\bigotimes_{i\in I}\mathrm{L}^2(X_i)$, resp. independence between the Y_i's implies that $\mathrm{L}^2(\vec{Y}_I)=\bigotimes_i\mathrm{L}^2(Y_i)$; and independence between the couples (X_i,Y_i) means that the global operator $\pi_I\colon \mathrm{L}^2(\vec{X}_I)\to\mathrm{L}^2(\vec{Y}_I)$ is equal to the tensor product $\bigotimes_i\pi_i$, so that $\pi_I^*\pi_I=\bigotimes_i\pi_i^*\pi_i$. Therefore, considering spectra as multisets, $\mathrm{Spec}(\pi_I^*\pi_I)=\prod_i\mathrm{Spec}(\pi_i^*\pi_i)$. But the $\mathrm{Spec}(\pi_\bullet^*\pi_\bullet)$'s are included in $[0,1]$, and $\mathrm{Spec}(\bar{\pi}_\bullet^*\bar{\pi}_\bullet)=\mathrm{Spec}(\pi_\bullet^*\pi_\bullet)\smallsetminus\{1\}$ (where eigenvalue 1 is removed with multiplicity one), so that

$$\rho_{\mathrm{sp}}(\bar{\pi}_I^*\bar{\pi}_I)=\sup_{i\in I}\rho_{\mathrm{sp}}(\bar{\pi}_i^*\bar{\pi}_i), \tag{BV}$$

which yields the desired result by (AJ). ♤

Proposition 16 gives us a clue about why we have got ρ-mixing between hyperplanes for the Ising model. Indeed, facing hyperplanes may be considered as a set of facing pairs of spins. For each of these pairs correlation is weak, and two pairs of spins are asymptotically independent when the distance between them becomes large; therefore by Proposition 16, it is sensible that the correlation between hyperplanes remains about as weak as the correlation between individual spins. (Actually this is essentially the same reasoning as the argument to prove Proposition 14, except that in the previous case it was used to prove the *absence* of β-mixing, due to the fact that that form of mixing has bad tensorization properties).

Of course, the previous argument does not make a proof: first, because asymptotic independence is not independence *sensu stricto*; and above all, the problem is that to prove *presence* of mixing, we have to consider all the spins in hyperplanes I and J, and not only a subset of spins distant from each other…

Anyway, could one generalize Proposition 16 in a way which would allow us to handle this case? The proof we gave above leaves little hope for a simple generalization, as using tensor products strongly requires strict independence between the

---

[‡] But X_i and Y_i are not supposed to be independent.

$(\boldsymbol{X}_i, \boldsymbol{Y}_i)$'s... However, in the sequel we will introduce another proof of Proposition 16 (see § 2.2), and that proof will turn out to be more suitable for our goal. In the end, we will get generalized tensorization results which will also handle the case where "independence between the pairs is not complete": see in particular Theorem 39. Thanks to Theorem 39 and it siblings, we will be able to prove decorrelation results for statistical physics models by the mean of (generalized) tensorization.

Chapter 2

Tensorization results

Now let us turn to the core of this work, namely the tensorization results. Each section of this chapter will be devoted to a different tensorization theorem, those being more or less arranged in increasing order of technicality; the climax being Theorem 47 in § 2.5. Though our different sections are essentially independent, it might be wise for the reader not to skip directly to the most advanced results, but rather to read first a few simpler cases (e.g. Theorems 36 and 37 in § 2.3), in order to get acquainted with the notation and techniques involved in the proofs.

2.1 Toolbox on variance and covariance

The proofs of this chapter will make heavy use of a couple of elementary computational results about variance and covariance, which we state and prove in this section.

Definition 18 (Variable centred w.r.t. a σ-algebra)**.** Let f be a L^1 variable and $\mathscr{F}$ be a σ-algebra; we will say that f *is centred w.r.t.* $\mathscr{F}$ to mean that $\mathbb{E}(f \mid \mathscr{F}) \equiv 0$. ♡

Remark 19. Obviously if $\mathscr{F}' \subseteq \mathscr{F}$, then any variable centred w.r.t. $\mathscr{F}$ is also centred w.r.t. $\mathscr{F}'$. ♧

Lemma 20. *For $\mathscr{F}$ a σ-algebra, for all $f \in \mathrm{L}^2(\mathscr{F})$, $g \in \mathrm{L}^2(\mathscr{B})$:*

$$\mathrm{Cov}(f, g) = \mathrm{Cov}(f, g^{\mathscr{F}}). \tag{BW}$$

In particular, if g is centred w.r.t. $\mathscr{F}$,

$$\mathrm{Cov}(f, g) = 0. \tag{BX}$$

◇

Proof. For $f \in \mathrm{L}^2(\mathscr{F})$, $g \in \mathrm{L}^2(\mathscr{B})$,

$$\begin{aligned} \mathrm{Cov}(f, g) &= \mathbb{E}(fg) - \mathbb{E}(f)\,\mathbb{E}(g) \\ &= \mathbb{E}\big((fg)^{\mathscr{F}}\big) - \mathbb{E}(f)\,\mathbb{E}(g^{\mathscr{F}}) = \mathbb{E}(fg^{\mathscr{F}}) - \mathbb{E}(f)\,\mathbb{E}(g^{\mathscr{F}}) = \mathrm{Cov}(f, g^{\mathscr{F}}). \quad ♤ \end{aligned} \tag{BY}$$

Lemma 21. *Let f be a L^2 variable and $\mathscr{F}$ be a σ-algebra; then:*

$$\mathrm{Var}^{1/2}(f^{\mathscr{F}}) = \sup\left\{ \frac{|\mathrm{Cov}(f, g)|}{\mathrm{Var}^{1/2}(g)} \;\middle|\; g \in \mathrm{L}^2(\mathscr{F}) \setminus \{0\} \right\}. \tag{BZ}$$

◇

Proof. For $g \in \mathrm{L}^2(\mathscr{F})$, Equation (`BW`) of Lemma 20 applied with 'f' $\leftarrow g$, 'g' $\leftarrow f$, '$\mathscr{F}$' $\leftarrow \mathscr{F}$ yields that $\mathrm{Cov}(f, g) = \mathrm{Cov}(f^{\mathscr{F}}, g)$; therefore, by the Cauchy–Schwarz inequality,

$$|\mathrm{Cov}(f, g)| \leqslant \mathrm{Var}^{1/2}(f^{\mathscr{F}})\,\mathrm{Var}^{1/2}(g), \tag{\texttt{CA}}$$

proving the '$\geqslant$' sense of (`BZ`). For the reverse inequality, take $g = f^{\mathscr{F}}$. (If $f^{\mathscr{F}} \equiv 0$, there is nothing to prove). Then, applying Equation (`BW`) of Lemma 20 with 'f' $\leftarrow f^{\mathscr{F}}$, '$g$' $\leftarrow f$, '$\mathscr{F}$' $\leftarrow \mathscr{F}$ yields that $\mathrm{Cov}(f, f^{\mathscr{F}}) = \mathrm{Var}(f^{\mathscr{F}})$, so that

$$\frac{|\mathrm{Cov}(f, f^{\mathscr{F}})|}{\mathrm{Var}^{1/2}(f^{\mathscr{F}})} = \mathrm{Var}^{1/2}(f^{\mathscr{F}}), \tag{\texttt{CB}}$$

proving the '$\leqslant$' sense of (`BZ`). ♤

Remark 22. Actually only the '$\leqslant$' sense of (`BZ`) will be used in the sequel. ♧

Lemma 23. *Let $\mathscr{F}$ be a σ-algebra, let f be a L^2 variable centred w.r.t. $\mathscr{F}$ and let $g \in \mathrm{L}^2(\mathscr{B})$; then:*

$$\mathrm{Cov}(f, g) = \mathbb{E}\big(\mathrm{Cov}(f, g \mid \mathscr{F})\big). \tag{\texttt{CC}}$$

In particular:

$$\mathrm{Var}(f) = \mathbb{E}\big(\mathrm{Var}(f \mid \mathscr{F})\big). \tag{\texttt{CD}}$$

◇

Proof. Since f is centred conditionally to $\mathscr{F}$, one has $\mathrm{Cov}(f, g \mid \mathscr{F}) = \mathbb{E}(fg \mid \mathscr{F})$. Also, f being centred w.r.t. $\mathscr{F}$ implies it being centred w.r.t. $\mathscr{B}$, so one has also $\mathrm{Cov}(f, g) = \mathbb{E}(fg)$. So, by the law of total expectation,

$$\mathrm{Cov}(f, g) = \mathbb{E}(fg) = \mathbb{E}\big((fg)^{\mathscr{F}}\big) = \mathbb{E}\big(\mathrm{Cov}(f, g \mid \mathscr{F})\big). \tag{\texttt{CE}}$$

♤

Lemma 24. *Let $f_0, f_1 \in \mathrm{L}^2(\mathscr{B})$ and let $\mathscr{F}$ be a σ-algebra; then:*

$$\mathrm{Cov}(f_0, f_1) = \mathrm{Cov}(f_0^{\mathscr{F}}, f_1^{\mathscr{F}}) + \mathbb{E}\big(\mathrm{Cov}(f_0, f_1 \mid \mathscr{F})\big). \tag{\texttt{CF}}$$

In particular, for $f \in \mathrm{L}^2(\mathscr{B})$:

$$\mathrm{Var}(f) = \mathrm{Var}(f^{\mathscr{F}}) + \mathbb{E}\big(\mathrm{Var}(f \mid \mathscr{F})\big). \tag{\texttt{CG}}$$

◇

Proof. For $i \in \{0, 1\}$, denote $\tilde{f}_i \coloneqq f_i - \mathbb{E}(f_i^{\mathscr{F}})$, so that $f_i = f_i^{\mathscr{F}} + \tilde{f}_i$. Obviously the $f_i^{\mathscr{F}}$'s are $\mathscr{F}$-measurable while the $\tilde{f}_i$'s are centred w.r.t. $\mathscr{F}$; thus by Equation (`BX`) of Lemma 20, one has $\mathrm{Cov}(f_0^{\mathscr{F}}, \tilde{f}_1), \mathrm{Cov}(\tilde{f}_0, f_1^{\mathscr{F}}) = 0$. Therefore:

$$\mathrm{Cov}(f_0, f_1) = \mathrm{Cov}(f_0^{\mathscr{F}}, f_1^{\mathscr{F}}) + \mathrm{Cov}(\tilde{f}_0, \tilde{f}_1). \tag{\texttt{CH}}$$

Now $\tilde{f}_0$ and $\tilde{f}_1$ are centred w.r.t. $\mathscr{F}$, so by Equation (`CC`) of Lemma 23:

$$\mathrm{Cov}(\tilde{f}_0, \tilde{f}_1) = \mathbb{E}\big(\mathrm{Cov}(\tilde{f}_0, \tilde{f}_1 \mid \mathscr{F})\big). \tag{\texttt{CI}}$$

But, conditionally to $\mathscr{F}$, $\tilde{f}_i$ only differs from f_i by the "constant" $\mathbb{E}(f_i \mid \mathscr{F})$; so:

$$\mathrm{Cov}(\tilde{f}_0, \tilde{f}_1 \mid \mathscr{F}) = \mathrm{Cov}(f_0, f_1 \mid \mathscr{F}). \tag{\texttt{CJ}}$$

Combining (`CH`), (`CI`) and (`CJ`) finally yields (`CF`). ♤

Lemma 25. *For $f \in \mathrm{L}^2(\mathscr{B})$ and $\mathscr{F}$ a σ-algebra:*

$$\mathbb{E}\big(\mathrm{Var}(f \mid \mathscr{F})\big) \leqslant \mathrm{Var}(f). \tag{CK}$$

◇

Proof. This is a straightforward corollary of Equation (CG) of Lemma 24. ♤

Lemma 26 (Chain rule for covariance)**.** *Let $f_0, f_1 \in \mathrm{L}^2(\mathscr{B})$ be random variables and let $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}_N$ be σ-algebras; then:*

$$\mathrm{Cov}(f_0^{\mathscr{F}_N} - f_0^{\mathscr{F}_0}, f_1^{\mathscr{F}_N} - f_1^{\mathscr{F}_0}) = \sum_{i=0}^{N-1} \mathrm{Cov}(f_0^{\mathscr{F}_{i+1}} - f_0^{\mathscr{F}_i}, f_1^{\mathscr{F}_{i+1}} - f_1^{\mathscr{F}_i}). \tag{CL}$$

In particular, for $f \in \mathrm{L}^2(\mathscr{B})$:

$$\mathrm{Var}(f^{\mathscr{F}_N} - f^{\mathscr{F}_0}) = \sum_{i=0}^{N-1} \mathrm{Var}(f^{\mathscr{F}_{i+1}} - f^{\mathscr{F}_i}). \tag{CM}$$

◇

Proof. By induction it is enough to prove the case $N = 2$. Then, we observe that, for $i \in \{0,1\}$, $f_i^{\mathscr{F}_2} - f_i^{\mathscr{F}_0} = (f_i^{\mathscr{F}_1} - f_i^{\mathscr{F}_0}) + (f_i^{\mathscr{F}_2} - f_i^{\mathscr{F}_1})$. But the $(f_i^{\mathscr{F}_1} - f_i^{\mathscr{F}_0})$'s are $\mathscr{F}_1$-measurable while the $(f_i^{\mathscr{F}_2} - f_i^{\mathscr{F}_1})$'s are centred w.r.t. $\mathscr{F}_1$; so by Equation (BX) of Lemma 20, one has $\mathrm{Cov}(f_0^{\mathscr{F}_1} - f_0^{\mathscr{F}_0}, f_1^{\mathscr{F}_2} - f_1^{\mathscr{F}_1}) = 0$, resp. $\mathrm{Cov}(f_1^{\mathscr{F}_1} - f_1^{\mathscr{F}_0}, f_0^{\mathscr{F}_2} - f_0^{\mathscr{F}_1}) = 0$. Therefore,

$$\mathrm{Cov}(f_0^{\mathscr{F}_2} - f_0^{\mathscr{F}_0}, f_1^{\mathscr{F}_2} - f_1^{\mathscr{F}_0}) = \mathrm{Cov}(f_0^{\mathscr{F}_1} - f_0^{\mathscr{F}_0}, f_1^{\mathscr{F}_1} - f_1^{\mathscr{F}_0}) + \mathrm{Cov}(f_0^{\mathscr{F}_2} - f_0^{\mathscr{F}_1}, f_1^{\mathscr{F}_2} - f_1^{\mathscr{F}_1}), \tag{CN}$$

QED. ♤

Lemma 27. *For $\mathscr{F} \subseteq \mathscr{F}'$ σ-algebras, for all $f \in \mathrm{L}^2(\mathscr{B})$:*

$$\mathrm{Var}(f^{\mathscr{F}}) \leqslant \mathrm{Var}(f^{\mathscr{F}'}). \tag{CO}$$

◇

Proof. Applying the chain rule for variance [Equation (CM) of Lemma 26] with 'f' ← f, 'N' ← 2, '$\mathscr{F}_0$' ← $\mathscr{O}$, '$\mathscr{F}_1$' ← $\mathscr{F}$, '$\mathscr{F}_2$' ← $\mathscr{F}'$ yields that

$$\mathrm{Var}(f^{\mathscr{F}'}) = \mathrm{Var}(f^{\mathscr{F}}) + \mathrm{Var}(f^{\mathscr{F}'} - f^{\mathscr{F}}) \geqslant \mathrm{Var}(f^{\mathscr{F}}). \tag{CP}$$

♤

2.2 Strict tensorization revisited

We state here again the classical theorem on "strict" (i.e., independent) tensorization of ρ-mixing (Proposition 16 in § 1.3):

Theorem 28. *Let I be a set and let $\big((X_i, Y_i)\big)_{i \in I}$ be independent couples of random variables. Then:*

$$\rho(\vec{X}_I; \vec{Y}_I) = \sup_{i \in I} \rho(X_i; Y_i). \tag{CQ}$$

◇

Our goal in this section is to present a way to prove that theorem which is different from the proof given in § 1.3, and which will be suitable for generalization.

Proof of Theorem 28. To alleviate notation, in this proof 'x_i' will implicitly stand for an element in the range of X_i, resp. 'y_i' for an element in the range of Y_i; and $\mathbb{P}(\bullet \mid x_i)$, resp. $\mathbb{E}(\bullet \mid x_i)$, etc., will stand for $\mathbb{P}(\bullet \mid X_i = x_i)$, etc. We will also shorthand $\rho(X_i; Y_i) =: \rho_i$.

First, observe that the '$\geqslant$' inequality of (`CQ`) is trivial, so that we only have to prove the '$\leqslant$' inequality. Also, by a classical approximation argument we may assume that I is finite, say $I = \{0, \dots, N-1\}$ for some $N \in \mathbb{N}$. So, let f and g be resp. $\vec{X}_I$-measurable and $\vec{Y}_I$-measurable centred L^2 real variables; our goal is to bound above $|\mathrm{Cov}(f, g)|$ by $(\sup_{i\in I} \rho_i) \operatorname{Var}^{1/2}(f) \operatorname{Var}^{1/2}(g)$.

For $i \in \{0, \dots, N\}$, define the σ-algebra $\mathscr{F}_i \coloneqq \sigma(X_0, Y_0, X_1, Y_1, \dots, X_{i-1}, Y_{i-1})$. For $i \in \{0, \dots, N-1\}$, we define

$$f_i \coloneqq f^{\mathscr{F}_{i+1}} - \mathbb{E}(f \mid \mathscr{F}_i); \tag{CR}$$

$$g_i \coloneqq g^{\mathscr{F}_{i+1}} - \mathbb{E}(g \mid \mathscr{F}_i). \tag{CS}$$

Then obviously one has $f = \sum_{i\in I} f_i$, resp. $g = \sum_{i\in I} g_i$. Moreover, as $f^{\mathscr{F}_i} = (f^{\mathscr{F}_{i+1}})^{\mathscr{F}_i}$, the chain rule for variance [Equation (`CM`)] yields that $\operatorname{Var}(f_i) = \operatorname{Var}(f^{\mathscr{F}_{i+1}}) - \operatorname{Var}(f^{\mathscr{F}_i})$, so that (noticing that $f^{\mathscr{F}_N} = f$ and $f^{\mathscr{F}_0} = 0$)

$$\operatorname{Var}(f) = \sum_{i\in I} \operatorname{Var}(f_i), \tag{CT}$$

and similarly $\operatorname{Var}(g) = \sum_{i\in I} \operatorname{Var}(g_i)$. Also, when one expands $\mathrm{Cov}(f, g) = \sum_{(i,j)\in I\times I} \mathrm{Cov}(f_i, g_j)$, all the $\mathrm{Cov}(f_i, g_j)$'s for $i \neq j$ are zero: indeed, if, say, $i < j$, then f_i is $f^{\mathscr{F}_{i+1}}$-measurable while g_j is centred w.r.t. $\mathscr{F}_j \supseteq \mathscr{F}_{i+1}$, which implies zero covariance by Equation (`BX`) of Lemma 20. Thus,

$$\mathrm{Cov}(f, g) = \sum_{i\in I} \mathrm{Cov}(f_i, g_i). \tag{CU}$$

Now, remembering that f_i and g_i are centred and remain centred w.r.t. $\mathscr{F}_i$, we use the law of total expectation to see that

$$\begin{aligned}\mathrm{Cov}(f_i, g_i) = \mathbb{E}(f_i g_i) &= \int \mathbb{E}(f_i g_i \mid x_0, y_0, \dots, x_{i-1}, y_{i-1}) \,\mathrm{d}\mathbb{P}(x_0, y_0, \dots, x_{i-1}, y_{i-1}) \\ &= \int \mathrm{Cov}(f_i, g_i \mid x_0, y_0, \dots, x_{i-1}, y_{i-1}) \,\mathrm{d}\mathbb{P}(x_0, y_0, \dots, x_{i-1}, y_{i-1}).\end{aligned} \tag{CV}$$

The point now is that, under the law $\mathbb{P}(\bullet \mid x_0, y_0, \dots, x_{i-1}, y_{i-1})$, the variable f_i, which a priori depends on (X_i, Y_i), actually only depends on X_i (and similarly, g_i only depends on Y_i). To see that point, first observe that $f^{\mathscr{F}_i}$ only depends on $(X_0, \dots, X_{i-1})$ (and not on $(Y_0, \dots, Y_{i-1})$), since one can write

$$\begin{aligned}&f^{\mathscr{F}_i}(x_0, y_0, \dots, x_{i-1}, y_{i-1}) \\ &\quad = \int f(x_0, \dots, x_{i-1}, x_i, \dots, x_{N-1}) \,\mathrm{d}\mathbb{P}(x_i, \dots, x_{N-1} \mid x_0, y_0, \dots, x_{i-1}, y_{i-1}) \\ &\quad = \int f(x_0, \dots, x_{i-1}, x_i, \dots, x_{N-1}) \,\mathrm{d}\mathbb{P}(x_i, \dots, x_{N-1}),\end{aligned} \tag{CW}$$

where the last equality comes from the fact that all the (X_i, Y_i)'s are independent. From (`CW`) we get in a first time that f_i is actually a function of $(X_0, \dots, X_i)$; then, by the very same argument, we see that conditioning f_i w.r.t. $\mathscr{F}_i$ yields a variable only depending on $(X_0, \dots, X_{i-1})$. Combining the two previous observations, the claimed point follows. So, under $\mathbb{P}(\bullet \mid x_0, y_0, \dots, x_{i-1}, y_{i-1})$, f_i only depends on X_i and g_i only depends on Y_i. But under that law, (X_i, Y_i) has the same distribution as under the original law (because of the independence between the (X_i, Y_i)'s), so that, under the conditioned law, one still has $\mathrm{\rho}(X_i; Y_i) \leqslant \rho_i$; and thus, (recalling that f_i and g_i remain centred under the conditioned law),

$$\begin{aligned}
&|\mathrm{Cov}(f_i, g_i)| = \Big| \int \mathrm{Cov}(f_i, g_i \mid x_0, y_0, \dots, x_{i-1}, y_{i-1}) \,\mathrm{d}\mathbb{P}(x_0, y_0, \dots, x_{i-1}, y_{i-1}) \Big| \\
&\leqslant \rho_i \int \mathrm{Var}^{1/2}(f_i \mid x_0, y_0, \dots, x_{i-1}, y_{i-1}) \,\mathrm{Var}^{1/2}(g_i \mid \textit{idem}) \,\mathrm{d}\mathbb{P}(x_0, y_0, \dots, x_{i-1}, y_{i-1}) \\
&\underset{\mathrm{CS}}{\leqslant} \rho_i \Big(\int \mathrm{Var}(f_i \mid x_0, y_0, \dots, x_{i-1}, y_{i-1}) \,\mathrm{d}\mathbb{P}(x_0, y_0, \dots, x_{i-1}, y_{i-1}) \Big)^{1/2} \times \textit{(same for g)} \\
&\qquad\qquad\qquad\qquad \underset{(\mathtt{CK})}{\leqslant} \rho_i \,\mathrm{Var}^{1/2}(f_i) \,\mathrm{Var}^{1/2}(g_i).^{[*]} \qquad (\mathtt{CX})
\end{aligned}$$

Summing over i,

$$\begin{aligned}
|\mathrm{Cov}(f, g)| &\leqslant \sum_{i \in I} \rho_i \,\mathrm{Var}^{1/2}(f_i) \,\mathrm{Var}^{1/2}(g_i) \leqslant (\sup_{i \in I} \rho_i) \sum_{i \in I} \mathrm{Var}^{1/2}(f_i) \,\mathrm{Var}^{1/2}(g_i) \\
&\underset{\mathrm{CS}}{\leqslant} (\sup_{i \in I} \rho_i) \Big(\sum_{i \in I} \mathrm{Var}(f_i) \Big)^{1/2} \Big(\sum_{i \in I} \mathrm{Var}(g_i) \Big)^{1/2} = (\sup_{i \in I} \rho_i) \,\mathrm{Var}^{1/2}(f) \,\mathrm{Var}^{1/2}(g), \qquad (\mathtt{CY})
\end{aligned}$$

which is the desired bound. ♤

2.3 Simple tensorization

This section is devoted to understand what we can say about the ρ-mixing coefficient between a "vectorial" variable $\vec{X}_I$ and a "scalar" variable Y, when we have bounds on the $\mathrm{\rho}(X_i; Y)$'s for all the $i \in I$.[†] Actually to prove our theorems we will need a bit stronger assumption (which is fortunately satisfied in many natural situations, confer Chapter 3): introducing this assumption and the notation for it will be the purpose of § 2.3.1. Then, for pedagogical purposes, we will present two versions of our tensorization theorem: first (in § 2.3.2) a slightly suboptimal one, whose proof goes along the same lines as for Theorem 28; then (in § 2.3.3) a sharper one, for which we will actually give two different proofs, as each of these proofs involves some new ideas that will turn useful for proving later theorems.

2.3.1 $\rho^\ddagger$-mixing

Definition 29 ($\rho^\dagger$-mixing)**.** Let X, Y and Z be random variables. We define the *$\rho^\dagger$-mixing coefficient between X and Y w.r.t. Z*, which we denote by $\mathrm{\rho}^\dagger_Z(X; Y)$, as the

---

[*] The last inequality is actually an equality here, because f_i and g_i are centred w.r.t. $\mathscr{F}_i$.

[†] We call such a situation "simple tensorization" because only one of the variables is vectorial.

smallest ρ such that, for Law(Z)-almost-all z, one has $\rho(X;Y) \leqslant \rho$ *under the law* $\mathbb{P}(\bullet \mid Z = z)$. ♡

Remark 30. The previous definition can also be stated in terms of the σ-algebras spanned by the variables: if we denote $\mathscr{F} := \sigma(X)$, $\mathscr{G} := \sigma(Y)$, $\mathscr{H} := \sigma(Z)$, one has

$$\rho^{\dagger}_{Z}(X;Y) = \sup\left\{ \frac{\operatorname{Cov}(f,g)}{\operatorname{Var}^{1/2}(f)\operatorname{Var}^{1/2}(g)} \;\middle|\; \left\{\begin{matrix} f \in \mathrm{L}^2(\mathscr{F}\vee\mathscr{H})\smallsetminus\{0\}, \mathbb{E}(f \mid \mathscr{H}) \equiv 0 \\ g \in \mathrm{L}^2(\mathscr{G}\vee\mathscr{H})\smallsetminus\{0\}, \mathbb{E}(g \mid \mathscr{H}) \equiv 0 \end{matrix}\right.\right\}. \tag{CZ}$$

♧

Remark 31. For $\mathscr{H} \subseteq \mathscr{H}'$ σ-algebras, is it not true in general that $\rho^{\dagger}_{\mathscr{H}}(\mathscr{F};\mathscr{G}) \leqslant \rho^{\dagger}_{\mathscr{H}'}(\mathscr{F};\mathscr{G})$, nor is the converse inequality true; in particular, there is no general inequality between $\rho^{\dagger}_{\mathscr{H}}(\mathscr{F};\mathscr{G})$ and $\rho(\mathscr{F};\mathscr{G}) = \rho^{\dagger}_{\mathscr{O}}(\mathscr{F};\mathscr{G})$: see e.g. Examples 32 and 33 below. ♧

Example 32 (Case where $\rho = 0$ but $\rho^{\dagger} = 1$). Let X and Y be independent variables (so one shall have $\rho(X;Y) = 0$), both with uniform law on $\mathbb{R}\,/\,\mathbb{Z}$; and let $Z := X + Y$. Then $\rho^{\dagger}_{Z}(X;Y) = 1$: indeed, for all $z \in \mathbb{R}\,/\,\mathbb{Z}$, under $\mathbb{P}(\bullet \mid Z = z)$, $Y \equiv z - X$, so that Y is X-measurable (and nonconstant). (Note that getting $\rho^{\dagger}_{Z}(X;Y) = 1$ did not actually require to have a strong correlation between X and Y under $\mathbb{P}(\bullet \mid Z = z)$ for all $z \in \mathbb{R}\,/\,\mathbb{Z}$: only a nonzero-measure set of such z's would have sufficed). ♧

Example 33 (Case where $\rho = 1$ but $\rho^{\dagger} = 0$). Let α, β, γ be three independent random variables, all uniform on $\{0,1\}$; define $X := (\gamma,\alpha), Y := (\gamma,\beta)$ and $Z := \gamma$. Then, conditionally to $\{Z = 0\}$, X and Y are independent (with common law $\delta_0 \otimes \operatorname{Unif}(\{0,1\})$), and similarly X and Y are independent conditionally to $\{Z = 1\}$; so $\rho^{\dagger}_{Z}(X;Y) = 0$. (Here note that it was necessary to check independence of X and Y both under $\mathbb{P}(\bullet \mid Z = 0)$ and under $\mathbb{P}(\bullet \mid Z = 1)$). Yet X and Y are not independent at all, since for instance the events $\big\{X \in \{(0,0),(0,1)\}\big\}$ and $\big\{Y \in \{(0,0),(0,1)\}\big\}$ (which are non-trivial under $\mathbb{P}$) are equivalent (both being equivalent to $\{Z = 0\}$), so that $\rho(X;Y) = 1$. ♧

Definition 34 ($\rho^{\ddagger}$-mixing)**.** Let X and Y be random variables, K be a set, and $(Z_k)_{k\in K}$ be a family of random variables. We define the *$\rho^{\ddagger}$-mixing coefficient between X and Y w.r.t.* $(Z_k)_{k\in K}$, denoted by $\rho^{\ddagger}_{(Z_k)_k}(X;Y)$, as the supremum of the $\rho^{\dagger}_{\vec{Z}_{K'}}(X;Y)$'s for $K' \subseteq K$. ♡

Remark 35. In practice, the family $(Z_k)_{k\in K}$ will always be clear from the context: for instance, if we are looking at tensorization between the $(X_i)_{i\in I}$'s and the $(Y_j)_{j\in J}$'s, then we will take $K := I \sqcup J$ and $Z_{i\in I} := X_i$, resp. $Z_{j\in J} := Y_j$. So, in the sequel we will merely speak of "$\rho^{\ddagger}$-mixing", without specifying the family of variables of reference. ♧

2.3.2 Simple tensorization theorem, rough version

Theorem 36 (Rough simple tensorization)**.** *Let I be a set and let $(X_i)_{i\in I}, Y$ be random variables. Then:*

$$\rho(\vec{X}_I;Y) \leqslant \Big(\sum_{i\in I} \rho^{\ddagger}(X_i;Y)^2\Big)^{1/2}. \tag{DA}$$

◇

Proof. Like in the proof of Theorem 28, ‘x_i’ will implicitly stand for an element in the range of X_i; $\mathbb{P}(\bullet \mid x_i)$, resp. $\mathbb{E}(\bullet \mid x_i)$, etc., will stand for $\mathbb{P}(\bullet \mid X_i = x_i)$, etc.; and we will shorthand $\rho^\ddagger(X_i;Y) =: \rho_i$; we also shorthand $\big(\sum_{i\in I}\rho_i^2\big)^{1/2} =: \bar\rho$. Like in the previous proof again, by some approximation argument we may assume that $I = \{0,\dots,N-1\}$; we will shorthand $\vec{X}_{\{i'\in I \mid i'<i\}}$ into $\vec{X}_{<i}$, etc. So, let f and g be resp. $\vec{X}_I$-measurable and Y-measurable centred $\mathbb{L}^2$ real variables; our goal is to bound $|\mathrm{Cov}(f,g)|$ by $\bar\rho\,\mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g)$.

For $i\in\{0,\dots,N\}$, define the σ-algebra $\mathscr{F}_i \coloneqq \sigma(\vec{X}_{<i})$; for $i\in\{1,\dots,N\}$, we define

$$f_i \coloneqq f^{\mathscr{F}_{i+1}} - \mathbb{E}(f \mid \mathscr{F}_i). \tag{DB}$$

Then $f = \sum_{i\in I} f_i$; and by the chain rule for variance [Equation (`CM`)],

$$\mathrm{Var}(f) = \mathrm{Var}(f^{\mathscr{F}_N}) - \mathrm{Var}(f^{\mathscr{F}_0}) = \sum_{i\in I}\mathrm{Var}(f_i). \tag{DC}$$

By expanding, $\mathrm{Cov}(f,g) = \sum_{i\in I}\mathrm{Cov}(f_i,g)$; so we will try and bound the $|\mathrm{Cov}(f_i, g)|$’s. Recalling that f_i is centred w.r.t. $\mathscr{F}_i$, the law of total expectation gives:

$$\mathrm{Cov}(f_i,g) = \mathbb{E}(f_i g) = \int \mathbb{E}(f_i g \mid \vec{x}_{<i})\,\mathrm{d}\mathbb{P}(\vec{x}_{<i}) = \int \mathrm{Cov}(f_i, g \mid \vec{x}_{<i})\,\mathrm{d}\mathbb{P}(\vec{x}_{<i}). \tag{DD}$$

But under $\mathbb{P}(\bullet \mid \vec{x}_{<i})$, f_i is X_i-measurable[‡] while g is Y-measurable; and under this law one has $\rho(X_i;Y) \leqslant \rho_i$ (because of the $\rho^\ddagger$-mixing hypothesis); thus,

$$\big|\mathrm{Cov}(f_i, g \mid \vec{x}_{<i})\big| \leqslant \rho_i\,\mathrm{Var}^{1/2}(f_i \mid \vec{x}_{<i})\,\mathrm{Var}^{1/2}(g \mid \vec{x}_{<i}). \tag{DE}$$

It follows that

$$\begin{aligned}|\mathrm{Cov}(f_i,g)| &\leqslant \rho_i \int \mathrm{Var}^{1/2}(f_i \mid \vec{x}_{<i})\,\mathrm{Var}^{1/2}(g \mid \vec{x}_{<i})\,\mathrm{d}\mathbb{P}(\vec{x}_{<i}) \\ &\underset{\mathrm{CS}}{\leqslant} \rho_i\Big(\int \mathrm{Var}(f_i \mid \vec{x}_{<i})\,\mathrm{d}\mathbb{P}(\vec{x}_{<i})\Big)^{1/2} \times \Big(\int \mathrm{Var}(g \mid \vec{x}_{<i})\,\mathrm{d}\mathbb{P}(\vec{x}_{<i})\Big)^{1/2} \\ &\qquad\qquad\underset{(\mathtt{CK})}{\leqslant} \rho_i\,\mathrm{Var}^{1/2}(f_i)\,\mathrm{Var}^{1/2}(g);\end{aligned} \tag{DF}$$

and then, summing (`DF`) for all i:

$$\begin{aligned}|\mathrm{Cov}(f,g)| &\leqslant \sum_{i\in I}\rho_i\,\mathrm{Var}^{1/2}(f_i)\,\mathrm{Var}^{1/2}(g) \\ &\underset{\mathrm{CS}}{\leqslant} \Big(\sum_{i\in I}\rho_i^2\Big)^{1/2}\Big(\sum_{i\in I}\mathrm{Var}(f_i)\Big)^{1/2}\mathrm{Var}^{1/2}(g) = \bar\rho\,\mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g),\end{aligned} \tag{DG}$$

which is the announced bound. ♤

[‡] Contrary to the previous proof where we had to use independence between the (X_i,Y_i) to see that f_i was X_i-measurable under the conditioned law, here no specific argument is needed, as there is no question of possible dependence w.r.t. the “$Y_{i'}$”’s.

2.3.3 Simple tensorization, sharp version

Obviously Theorem 36 is not optimal, as the right-hand-side of (`DA`) may be larger than 1. The optimal result is actually the following one:

Theorem 37 (Sharp simple tensorization)**.** *Under the same hypotheses as in Theorem 36, one has:*

$$\rho(\vec{X}_I;Y) \leqslant \Big(1-\prod_{i\in I}\big(1-\rho^{\ddagger}(X_i;Y)^2\big)\Big)^{1/2}. \tag{DH}$$

◇

Proof. We re-use the notation of the previous proof; in particular, $I=\{0,\dots,N-1\}$ and $\mathscr{F}_i \coloneqq \sigma(\vec{X}_{<i})$; the only difference being that this time $\bar{\rho}$ denotes $\big(1-\prod_i(1-\rho_i^2)\big)^{1/2}$.

Our goal is to prove that $\rho(\vec{X}_I;Y)\leqslant\bar{\rho}$. By Equation (`AI`) of Proposition 4, this is tantamount to showing that, for any Y-measurable function g, one has

$$\mathrm{Var}^{1/2}(g^{\sigma(\vec{X}_I)}) \leqslant \bar{\rho}\,\mathrm{Var}^{1/2}(g). \tag{DI}$$

For $i\in\{0,\dots,N\}$, we define

$$g^i = g - \mathbb{E}(g \mid \mathscr{F}_i). \tag{DJ}$$

By the chain rule for variance, one has $\mathrm{Var}(g^i)=\mathrm{Var}(g)-\mathrm{Var}(g^{\mathscr{F}_i})$; so, applying this with $i=N$, (`DI`) is equivalent to proving that

$$\mathrm{Var}(g^N)\geqslant(1-\bar{\rho}^2)\,\mathrm{Var}(g)=\prod_i(1-\rho_i^2)\times\mathrm{Var}(g^0). \tag{DK}$$

Let us fix $i\in I$ for a while. Under the law $\mathbb{P}(\bullet\mid\vec{x}_{<i})$, g^i is Y-measurable, and $\rho(X_i;Y)\leqslant\rho_i$ (because of the $\rho^{\ddagger}$-mixing hypothesis), so that one has: (still under the conditioned law),

$$\mathrm{Var}^{1/2}((g^i)^{\sigma(X_i)})\leqslant\rho_i\,\mathrm{Var}^{1/2}(g^i). \tag{DL}$$

But what we called "$(g^i)^{\sigma(X_i)}$" under the conditioned law corresponds to $(g^i)^{\sigma(\vec{X}_{<i},X_i)}=(g^i)^{\mathscr{F}_{i+1}}$ under the reference law; and

$$(g^i)^{\mathscr{F}_{i+1}}=g^{\mathscr{F}_{i+1}}-(g^{\mathscr{F}_i})^{\mathscr{F}_{i+1}}=g^{\mathscr{F}_{i+1}}-g^{\mathscr{F}_i}=(g-g^{i+1})-(g-g^i)=g^i-g^{i+1}. \tag{DM}$$

So, reasoning again under $\mathbb{P}(\bullet\mid\vec{x}_{<i})$, (`DM`) means that $g^{i+1}=g^i-\mathbb{E}(g^i\mid X_i)$; therefore, by the chain rule for variance, (still under the conditioned law),

$$\mathrm{Var}(g^{i+1})=\mathrm{Var}(g^i)-\mathrm{Var}((g^i)^{\sigma(X_i)})\underset{(\mathtt{DL})}{\geqslant}(1-\rho_i^2)\,\mathrm{Var}(g^i). \tag{DN}$$

Re-stated under the reference law, (`DN`) means that for all $\vec{x}_{<i}$,

$$\mathrm{Var}(g^{i+1}\mid\vec{x}_{<i})\geqslant(1-\rho_i^2)\,\mathrm{Var}(g^i\mid\vec{x}_{<i}). \tag{DO}$$

Now, g^i and g^{i+1} are centred w.r.t. $\mathscr{F}_i$, so by the law of total variance (Equation (`CG`) of Lemma 24),

$$\mathrm{Var}(g^i)=\int\mathrm{Var}(g^i\mid x_0,\dots,x_{i-1})\,\mathrm{d}\mathbb{P}(x_0,\dots,x_{i-1}); \tag{DP}$$

$$\mathrm{Var}(g^{i+1})=\int\mathrm{Var}(g^{i+1}\mid x_0,\dots,x_{i-1})\,\mathrm{d}\mathbb{P}(x_0,\dots,x_{i-1}). \tag{DQ}$$

Thus, integrating (`DN`) against $\mathrm{d}\mathbb{P}(x_0, \dots, x_{i-1})$ still yields

$$\operatorname{Var}(g^{i+1}) \geqslant (1 - \rho_i^2) \operatorname{Var}(g^i), \tag{DR}$$

but this time to be understood under the reference law. Finally, combining (`DR`) for all $i \in \{N-1, N-2, \dots, 0\}$, we get Equation (`DK`), which is what we wanted. ♤

Actually it is also possible to get Theorem 37 from a refinement of the proof of Theorem 36. We also give that proof here, as the lines of its reasoning will be used again later in the proof of Theorem 47 (see in particular Lemmas 52—in which the role of index 'i' below is actually played by j—and 53).

Alternative proof of Theorem 37. Like in the previous proof, shorthand $\rho^{\ddagger}(X_i; Y) =: \rho_i$; assume $\boldsymbol{I} = \{0, \dots, N-1\}$; let f and g be centred L^2 $\vec{\boldsymbol{X}}$-measurable, resp. $\boldsymbol{Y}$-measurable, variables; and denote $\mathscr{F}_i := \sigma(\vec{X}_{<i})$, resp. $g^i := g - \mathbb{E}(g \mid \mathscr{F}_i)$. Also, like in the proof of Theorem 36, denote $f_i := f^{\mathscr{F}_{i+1}} - \mathbb{E}(f \mid \mathscr{F}_i)$.

As in the proof of Theorem 36, one has $\operatorname{Var}(f) = \sum_i \operatorname{Var}(f_i)$ and $\mathbb{E}(fg) = \sum_i \mathbb{E}(f_i g)$. Now observe that f_i is centred w.r.t. $\mathscr{F}_i$ while $g - g^i$ is $\mathscr{F}_i$-measurable, so by Lemma 20 $\mathbb{E}(f_i g) = \mathbb{E}(f_i g^i)$. But, conditionally to $\mathscr{F}_i$, f_i and g^i are both centred and resp. $\boldsymbol{X}_i$- and $\boldsymbol{Y}$-measurable, so

$$|\mathbb{E}(f_i g^i)| \leqslant \rho^{\dagger}_{\mathscr{F}_i}(X_i; Y) \operatorname{Var}^{1/2}(f_i) \operatorname{Var}^{1/2}(g^i) \leqslant \rho_i \operatorname{Var}^{1/2}(f_i) \operatorname{Var}^{1/2}(g^i). \tag{DS}$$

Now, for $i \in \{0, \dots, N-1\}$, denote

$$g_i := (g^i)^{\mathscr{F}_{i+1}}. \tag{DT}$$

Since $g^i = g_i + g^{i+1}$, where g_i is $\mathscr{F}_{i+1}$-measurable while g^i is centred w.r.t. $\mathscr{F}_{i+1}$, one has by the chain rule for variance:

$$\operatorname{Var}(g_i) = \operatorname{Var}(g^i) - \operatorname{Var}(g^{i+1}). \tag{DU}$$

Then, the key point consists in making the following observation: for $\operatorname{Var}(g^{i+1})$ to be large (that is, close to $\operatorname{Var}(g^i)$), $\operatorname{Var}(g_i)$ has to be small. But in that case $|\mathbb{E}(f_i g)|$ shall be small: one has indeed, since f_i is $\mathscr{F}_{i+1}$-measurable:

$$|\mathbb{E}(f_i g)| = |\mathbb{E}(f_i g^i)| = |\mathbb{E}(f_i (g^i)^{\mathscr{F}_{i+1}})| = |\mathbb{E}(f_i g_i)| \underset{\mathrm{CS}}{\leqslant} \operatorname{Var}^{1/2}(f_i) \operatorname{Var}^{1/2}(g_i). \tag{DV}$$

Let us sum up the relations obtained. One has, for all $i \in \{0, \dots, N-1\}$:

$$|\mathbb{E}(f_i g)| \leqslant \operatorname{Var}^{1/2}(f_i) \operatorname{Var}^{1/2}(g_i); \tag{DW}$$

$$\operatorname{Var}(g^{i+1}) = \operatorname{Var}(g^i) - \operatorname{Var}(g_i); \tag{DX}$$

$$|\mathbb{E}(f_i g)| \leqslant \rho_i \operatorname{Var}^{1/2}(f_i) \operatorname{Var}^{1/2}(g^i). \tag{DY}$$

Now define $\check{\rho}_i := |\mathbb{E}(f_i g)| \mathbin{/} \operatorname{Var}^{1/2}(f_i) \operatorname{Var}^{1/2}(g^i)$ (or $\check{\rho}_i := 0$ if the right-hand side is $0 \mathbin{/} 0$). Then (`DW`) means that $\operatorname{Var}^{1/2}(g_i) \geqslant \check{\rho}_i \operatorname{Var}^{1/2}(g^i)$, so that (`DX`) yields

$\mathrm{Var}^{1/2}(g^{i+1}) \leqslant \sqrt{1-\check{\rho}_i^2}\,\mathrm{Var}^{1/2}(g^i)$. Since $g^0 = g$, one has therefore by induction $\mathrm{Var}^{1/2}(g^i) \leqslant \prod_{i'=0}^{i-1}\sqrt{1-\check{\rho}_{i'}^2}\,\mathrm{Var}^{1/2}(g)$, so that the decomposition "$\mathbb{E}(fg) = \sum_i \mathbb{E}(f_i g)$" gives:

$$|\mathbb{E}(fg)| \leqslant \sum_{i=0}^{N-1}\Big(\check{\rho}_i \prod_{i'=0}^{i-1}\sqrt{1-\check{\rho}_{i'}^2}\Big)\,\mathrm{Var}^{1/2}(f_i)\,\mathrm{Var}^{1/2}(g). \tag{DZ}$$

By the Cauchy–Schwarz inequality, it follows that:

$$\begin{aligned}|\mathbb{E}(fg)| &\leqslant \sqrt{\sum_{i=0}^{N-1}\check{\rho}_i^2\prod_{i'=0}^{i-1}\big(1-\check{\rho}_{i'}^2\big)}\,\sqrt{\sum_{i=0}^{N-1}\mathrm{Var}(f_i)}\,\mathrm{Var}^{1/2}(g)\\ &= \sqrt{1-\prod_{0=1}^{N-1}\big(1-\check{\rho}_i^2\big)}\,\mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g). \end{aligned}\tag{EA}$$

Since (`DY`) forces to have $\check{\rho}_i \leqslant \rho_i\ \forall i$, obviously the maximal value for the right-hand side of (`EA`) is when $\check{\rho}_i = \rho_i$ for all i, then yielding (`DH`). ♤

2.4 Double tensorization

Theorem 36 (or Theorem 37), interesting as it may be, is however not powerful enough to yield Theorem 28 as a corollary. To bound the maximal correlation $\rho(\vec{X}_I;\vec{Y}_J)$ between two "vectorial" variables from bounds on the correlations between the "scalar" variables constituting them, the result is the following:

Theorem 39 (Double tensorization). *Let I and J be sets, and let $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ be random variables. Then,*

$$\rho(\vec{X}_I;\vec{Y}_J) \leqslant \Big|\!\Big|\!\Big|\Big(\big(\rho^{\ddagger}(X_i;Y_j)\big)\Big)_{i\in I, j\in J}\Big|\!\Big|\!\Big|_{\ell^2(J)\to\ell^2(I)}. \tag{EB}$$

◇

To prove Theorem 39, we will need the following

Lemma 40. *Let $J \coloneqq \{0,\ldots,M-1\}$; let X and $(Y_j)_{j\in J}$ be random variables, and assume that for all $j\in J$,*

$$\rho^{\ddagger}(X;Y_j) \leqslant \rho_j. \tag{EC}$$

For all $j \in J\cup\{M\}$, denote

$$\mathscr{G}_j \coloneqq \sigma(Y_0,\ldots,Y_{j-1}); \tag{ED}$$

$$\mathscr{G}'_j \coloneqq \mathscr{G}_j \vee \sigma(X). \tag{EE}$$

Let g be an $\bar{\mathrm{L}}^2(\vec{Y}_J)$ function; for $j\in J$, define

$$g_j \coloneqq g^{\mathscr{G}_{j+1}} - \mathbb{E}(g \mid \mathscr{G}_j); \tag{EF}$$

$$g'_j \coloneqq g^{\mathscr{G}'_{j+1}} - \mathbb{E}(g \mid \mathscr{G}'_j), \tag{EG}$$

and denote $\boldsymbol{W}_j \coloneqq \mathrm{Var}(g_j)$*, resp.* $\boldsymbol{W}'_j \coloneqq \mathrm{Var}(g'_j)$*. Then, for all* $j \in \boldsymbol{J}$*:*

$$\boldsymbol{W}'_j \geqslant (1-\rho_j^2)\boldsymbol{W}_j - 2\rho_j \boldsymbol{W}_j^{1/2} \sum_{j'>j} \rho_{j'} \boldsymbol{W}_{j'}^{1/2}. \tag{EH}$$

◇

Proof of Lemma 40. For $j \in \boldsymbol{J}$, define

$$\tilde{g}_j \coloneqq g - \mathbb{E}(g \mid \mathscr{G}_j); \tag{EI}$$

$$\tilde{g}'_j \coloneqq g - \mathbb{E}(g \mid \mathscr{G}'_j), \tag{EJ}$$

and denote $\tilde{\boldsymbol{W}}_j \coloneqq \mathrm{Var}(\tilde{g}_j)$, resp. $\tilde{\boldsymbol{W}}'_j \coloneqq \mathrm{Var}(\tilde{g}'_j)$. Then, by the chain rule for variance (`CM`),

$$\tilde{\boldsymbol{W}}_j = \sum_{j' \geqslant j} \boldsymbol{W}_{j'}; \tag{EK}$$

$$\tilde{\boldsymbol{W}}'_j = \sum_{j' \geqslant j} \boldsymbol{W}'_{j'}; \tag{EL}$$

consequently,

$$\boldsymbol{W}_j - \boldsymbol{W}'_j = (\tilde{\boldsymbol{W}}_j - \tilde{\boldsymbol{W}}'_j) - (\tilde{\boldsymbol{W}}_{j+1} - \tilde{\boldsymbol{W}}'_{j+1}). \tag{EM}$$

But one has

$$g'_j = g - g^{\mathscr{G}'_j} = g - g^{\mathscr{G}_j} + g^{\mathscr{G}_j} - g^{\mathscr{G}'_j} = \tilde{g}_j + (g^{\mathscr{G}_j})^{\mathscr{G}'_j} - g^{\mathscr{G}'_j} = \tilde{g}_j - (g - g^{\mathscr{G}_j})^{\mathscr{G}'_j} = \tilde{g}_j - \tilde{g}_j^{\mathscr{G}'_j}, \tag{EN}$$

so that by the chain rule for variance (applied with 'f' $\leftarrow \tilde{g}_j$, '$\boldsymbol{N}$' $\leftarrow 2$, '$\mathscr{F}_0$' $\leftarrow \mathscr{O}$, '$\mathscr{F}_1$' $\leftarrow \mathscr{G}'_j$, '$\mathscr{F}_2$' $\leftarrow \mathscr{B}$),

$$\tilde{\boldsymbol{W}}_j - \tilde{\boldsymbol{W}}'_j = \mathrm{Var}(\tilde{g}_j^{\mathscr{G}'_j}), \tag{EO}$$

and likewise $\tilde{\boldsymbol{W}}_{j+1} - \tilde{\boldsymbol{W}}'_{j+1} = \mathrm{Var}(\tilde{g}_{j+1}^{\mathscr{G}'_{j+1}})$. That suggests to try to bound $\mathrm{Var}(\tilde{g}_j^{\mathscr{G}'_j})$ by an expression involving $\mathrm{Var}(\tilde{g}_{j+1}^{\mathscr{G}'_{j+1}})$.

To bound $\mathrm{Var}(\tilde{g}_j^{\mathscr{G}'_j})$, we will use the characterization (`BZ`) given by Lemma 21. So, let f be any $\mathscr{G}'_j$-measurable function. Let us reason conditionally to $\mathscr{G}_j$ for a few lines. Under this conditioning, f is $\boldsymbol{X}$-measurable, $\tilde{g}_j$ is $\vec{\boldsymbol{Y}}$-measurable and centred, and

$$\tilde{g}_j = g_j + \tilde{g}_{j+1}, \tag{EP}$$

where g_j is $\boldsymbol{Y}_j$-measurable, and both g_j and $\tilde{g}_{j+1}$ are centred. Since $\rho(\boldsymbol{X};\boldsymbol{Y}_j) \leqslant \rho_j$ (due to the assumption that, under the reference law, $\rho^{\ddagger}(\boldsymbol{X};\boldsymbol{Y}_j) \leqslant \rho_j$), one has

$$\begin{aligned} |\mathrm{Cov}(f,\tilde{g}_j)| &\leqslant |\mathrm{Cov}(f,g_j)| + |\mathrm{Cov}(f,\tilde{g}_{j+1})| \\ &\leqslant \rho_j \mathrm{Var}^{1/2}(f)\mathrm{Var}^{1/2}(g_j) + \mathrm{Var}^{1/2}(f)\mathrm{Var}^{1/2}(\tilde{g}_{j+1}), \end{aligned} \tag{EQ}$$

which means that under the reference law (remember that (`EQ`) is stated under the law $\mathbb{P}(\bullet \mid \mathscr{G}_j)$),

$$\big|\mathrm{Cov}(f,\tilde{g}_j \mid \mathscr{G}_j)\big| \leqslant \mathrm{Var}^{1/2}(f \mid \mathscr{G}_j)\big(\rho_j \mathrm{Var}^{1/2}(g_j \mid \mathscr{G}_j) + \mathrm{Var}^{1/2}(g_{j+1} \mid \mathscr{G}_j)\big). \tag{ER}$$

Integrating over the possible values for $\vec{Y}_{<j}$, and using that f and $\tilde{g}_j$ are centred:

$$\begin{aligned}
|\mathrm{Cov}(f, \tilde{g}_j)| &\underset{(\mathtt{CF})}{\leqslant} \int \big|\mathrm{Cov}(f, \tilde{g}_j \mid \mathscr{G}_j)\big| \,\mathrm{d}\mathbb{P}(\vec{y}_{<j}) \\
&\leqslant \int \mathrm{Var}^{1/2}(f \mid \vec{y}_{<j})\big(\rho_j \mathrm{Var}^{1/2}(g_j \mid \vec{y}_{<j}) + \mathrm{Var}^{1/2}(\tilde{g}_{j+1} \mid \vec{y}_{<j})\big) \,\mathrm{d}\mathbb{P}(\vec{y}_{<j}) \\
&\underset{\mathrm{CS}}{\leqslant} \Big(\int \mathrm{Var}(f \mid \mathscr{G}_j) \,\mathrm{d}\mathbb{P}(\vec{y}_{<j})\Big)^{1/2} \times \\
&\qquad \Big(\rho_j \Big(\int \mathrm{Var}(g_j \mid \mathscr{G}_j) \,\mathrm{d}\mathbb{P}(\vec{y}_{<j})\Big)^{1/2} + \Big(\int \mathrm{Var}(\tilde{g}_{j+1} \mid \mathscr{G}_j) \,\mathrm{d}\mathbb{P}(\vec{y}_{<j})\Big)^{1/2}\Big) \\
&\underset{(\mathtt{CG})}{=} \mathrm{Var}^{1/2}(f)(\rho_j \mathrm{Var}^{1/2}(g_j) + \mathrm{Var}^{1/2}(\tilde{g}_{j+1})) = \mathrm{Var}^{1/2}(f)(\rho_j W_j^{1/2} + \mathrm{Var}^{1/2}(\tilde{g}_{j+1})). \qquad (\mathtt{ES})
\end{aligned}$$

Since this is valid for all $f \in \mathrm{L}^2(\mathscr{G}'_j)$, by Lemma 21 we conclude that

$$\mathrm{Var}^{1/2}(\tilde{g}_j^{\mathscr{G}'_j}) \leqslant \rho_j W_j^{1/2} + \mathrm{Var}^{1/2}(\tilde{g}_{j+1}^{\mathscr{G}'_{j+1}}). \qquad (\mathtt{ET})$$

Then, passing to squares,

$$\tilde{W}_j - \tilde{W}'_j \leqslant \tilde{W}_{j+1} - \tilde{W}'_{j+1} + \rho_j^2 W_j + 2\rho_j W_j^{1/2} \mathrm{Var}^{1/2}(\tilde{g}_{j+1}^{\mathscr{G}'_{j+1}}). \qquad (\mathtt{EU})$$

On the other hand, iterating successively (ET) for 'j' $\leftarrow j+1, j+2, \ldots, M-1$, and using that $\tilde{g}_M \equiv 0$, we have that

$$\mathrm{Var}^{1/2}(\tilde{g}_{j+1}^{\mathscr{G}'_{j+1}}) \leqslant \sum_{j' > j} \rho_{j'} W_{j'}^{1/2}. \qquad (\mathtt{EV})$$

Combining (EU) with (EV) finally yields the announced formula (EH). ♤

Now equipped with Lemma 40, we can move on to proving Theorem 39:

Proof of Theorem 39. First, thanks to a by now classical approximation argument, we may assume that $I = \{0, \ldots, N-1\}$ and $J = \{0, \ldots, M-1\}$; let us also denote $\bar{I} \coloneqq I \cup \{N\}$, resp. $\bar{J} \coloneqq J \cup \{M\}$. We denote $\rho_{ij} \coloneqq \uprho^{\ddagger}(X_i; Y_j)$ and $\boldsymbol{\rho} \coloneqq (\!(\rho_{ij})\!)_{i,j} \in \mathbb{R}^{I \times J}$.

Our goal is to prove that $\uprho(\vec{X}_I; \vec{Y}_J) \leqslant |\!|\!|\boldsymbol{\rho}|\!|\!|$. By Equation (AI) of Proposition 4, this is tantamount to showing that, for all $g \in \mathrm{L}^2(\vec{Y})$, one has $\mathrm{Var}^{1/2}(g^{\upsigma(\vec{X})}) \leqslant |\!|\!|\boldsymbol{\rho}|\!|\!| \mathrm{Var}^{1/2}(g)$; and since by the chain rule (CM) for variance (applied with 'f' $\leftarrow g$, 'N' $\leftarrow 2$, '$\mathscr{F}_0$' $\leftarrow \mathscr{O}$, '$\mathscr{F}_1$' $\leftarrow \upsigma(\vec{X})$, '$\mathscr{F}_2$' $\leftarrow \mathscr{B}$) one has $\mathrm{Var}(g^{\upsigma(\vec{X})}) = \mathrm{Var}(g) - \mathrm{Var}(g - g^{\upsigma(\vec{X})})$, this is equivalent to proving that

$$\mathrm{Var}(g - g^{\upsigma(\vec{X})}) \geqslant (1 - |\!|\!|\boldsymbol{\rho}|\!|\!|^2) \mathrm{Var}(g). \qquad (\mathtt{EW})$$

So, let us introduce an arbitrary $g \in \mathrm{L}^2(\vec{Y})$, and let us begin with introducing some notation. We define:

$$\mathscr{F}_i \coloneqq \upsigma(X_0, \ldots, X_{i-1}) \quad \text{for } i \in \bar{I}; \qquad (\mathtt{EX})$$

$$\mathscr{G}_j \coloneqq \upsigma(Y_0, \ldots, Y_{j-1}) \quad \text{for } j \in \bar{J} \qquad (\mathtt{EY})$$

and

$$g^i := g - \mathbb{E}(g \mid \mathscr{F}_i) \quad \text{for } i \in \bar{I}; \tag{EZ}$$

$$g^i_j := g^{\mathscr{F}_i \vee \mathscr{G}_{j+1}} - \mathbb{E}(g \mid \mathscr{F}_i \vee \mathscr{G}_j) \quad \text{for } i \in \bar{I}, j \in J. \tag{FA}$$

We also denote

$$\operatorname{Var}(g^i) =: W^i \quad \text{for } i \in \bar{I}; \tag{FB}$$

$$\operatorname{Var}(g^i_j) =: W^i_j \quad \text{for } i \in \bar{I}, j \in J. \tag{FC}$$

By the chain rule for variance (CM), one has, for all $i \in \bar{I}$,

$$W^i = \sum_{j \in J} W^i_j. \tag{FD}$$

Moreover, obviously $g = g^0$ and $g - g^{\sigma(\vec{X})} = g^N$, so the formula (EW) that we want to prove can be re-written as:

$$\sum_{j \in J} W^N_j \geqslant (1 - |||\rho|||^2) \sum_{j \in J} W^0_j. \tag{FE}$$

Now, Lemma 40 will give us the central relations between the W^i_j's. Applying that lemma conditionally to $\mathscr{F}_i$ with 'X' $\leftarrow X_i$, 'Y_j' $\leftarrow Y_j$, 'g' $\leftarrow g^i$ and 'ρ_j' $\leftarrow \rho_{ij}$, we have that 'g_j' $= g^i_j$ and 'g'_j' $= g^{i+1}_j$, so:

$$\operatorname{Var}(g^{i+1}_j \mid \mathscr{F}_i) \geqslant (1 - \rho_j^2) \operatorname{Var}(g^i_j \mid \mathscr{F}_i) - 2\rho_j \operatorname{Var}^{1/2}(g^i_j \mid \mathscr{F}_i) \sum_{j' > j} \rho_{j'} \operatorname{Var}^{1/2}(g^i_{j'} \mid \mathscr{F}_i). \tag{FF}$$

Now we can integrate that formula w.r.t. $\vec{X}_{<i}$, which yields, using the Cauchy–Schwarz inequality:

$$\begin{multline}\int \operatorname{Var}(g^{i+1}_j \mid \vec{x}_{<i}) \,\mathrm{d}\mathbb{P}(\vec{x}_{<i}) \geqslant (1 - \rho_j^2) \int \operatorname{Var}(g^i_j \mid \vec{x}_{<i}) \,\mathrm{d}\mathbb{P}(\vec{x}_{<i}) - \\ 2\rho_j \Big(\int \operatorname{Var}(g^i_j \mid \vec{x}_{<i}) \,\mathrm{d}\mathbb{P}(\vec{x}_{<i}) \Big)^{1/2} \sum_{j' > j} \rho_{j'} \Big(\int \operatorname{Var}(g^i_{j'} \mid \vec{x}_{<i}) \,\mathrm{d}\mathbb{P}(\vec{x}_{<i}) \Big)^{1/2}, \tag{FG}\end{multline}$$

which, as all the functions involved are centred w.r.t. $\mathscr{F}_i$, means by Equation (CD) of Lemma 23 that

$$W^{i+1}_j \geqslant (1 - \rho_{ij}^2) W^i_j - 2\rho_{ij} (W^i_j)^{1/2} \sum_{j' > j} \rho_{ij'} (W^i_{j'})^{1/2}. \tag{FH}$$

It will be more convenient to introduce a linearized version of Equation (FH), which we derive from it by the Cauchy–Schwarz inequality:

Corollary 41. *With the notation of this proof, for $(\alpha_j)_{j \in J}$ arbitrary positive coefficients, one has for all $i \in I, j \in J$:*

$$W^{i+1}_j \geqslant (1 - \rho_{ij}^2) W^i_j - \alpha_j^{-1} \rho_{ij} \Big(\sum_{j' > j} \alpha_{j'} \rho_{ij'} \Big) W^i_j - \alpha_j \rho_{ij} \sum_{j' > j} \alpha_{j'}^{-1} \rho_{ij'} W^i_{j'}. \tag{FI}$$

◇♤

Remark 42. (`FI`) is devised so that it coincides with (`FH`) if (and only if) $W^i_j \propto \alpha_j^2$. ♧

Thanks to all that work, our new goal will be to prove (`EW`) thanks to the relations (`FI`). This is a problem about linear operators in an L^1 setting. To handle it, let us first introduce a little notation:

Notation 43. For $(x_j)_j =: \vec{x} \in \mathbb{R}^{\boldsymbol{J}}$, we denote

$$\|\vec{x}\|_{\ell^1} := \sum_{j\in \boldsymbol{J}} |x_j|, \tag{FJ}$$

and we call "$\ell^1(\boldsymbol{J})$" the space $\mathbb{R}^{\boldsymbol{J}}$ endowed with that norm.

The dual of $\ell^1(\boldsymbol{J})$ may also be identified with $\mathbb{R}^{\boldsymbol{J}}$, via:

$$\langle (\xi_j)_j, (x_j)_j \rangle := \sum_{j\in \boldsymbol{J}} \xi_j x_j; \tag{FK}$$

then, the dual norm of $(\xi_j)_j =: \vec{\xi}$ is

$$\sup_{j\in \boldsymbol{J}} |\xi_j| =: \|\vec{\xi}\|_{\ell^\infty}, \tag{FL}$$

and we call "$\ell^\infty(\boldsymbol{J})$" the space $\mathbb{R}^{\boldsymbol{J}}$ endowed with that norm.

We also denote "$\vec{x} \geqslant 0$" to mean that all the x_j's are nonnegative, resp. "$\vec{\xi} \geqslant 0$" to mean that all the ξ_j's are nonnegative, the latter being actually equivalent to saying that $(\vec{x} \geqslant 0 \;\Rightarrow\; \langle \vec{\xi}, \vec{x} \rangle \geqslant 0)$. ♡

Now I claim the following lemma, whose proof is postponed:

Lemma 44. *Let $(\beta^i_{jj'})_{(i,j,j')\in \boldsymbol{I}\times\boldsymbol{J}\times\boldsymbol{J}}$ be nonnegative coefficients and let $(W^i_j)_{(i,j)\in\bar{\boldsymbol{I}}\times\boldsymbol{J}}$ be quantities such that, for all $i \in \boldsymbol{I}, j \in \boldsymbol{J}$:*

$$W^{i+1}_j \geqslant W^i_j - \sum_{j'\in\boldsymbol{J}} \beta^i_{jj'} W^i_{j'}. \tag{FM}$$

Define $\mathscr{L} \in \ell^\infty(\boldsymbol{J})$ by

$$\mathscr{L}\vec{x} := \sum_{\substack{i\in\boldsymbol{I}\\ j,j'\in\boldsymbol{J}}} \beta^i_{jj'} x_{j'}; \tag{FN}$$

then:

$$\sum_{j\in\boldsymbol{J}} W^N_j \geqslant (1 - \|\mathscr{L}\|_{\ell^\infty}) \sum_{j\in\boldsymbol{J}} W^0_j. \tag{FO}$$

◇

Here one can apply Lemma 44 with the actual $\boldsymbol{I}$, $\boldsymbol{J}$ and W^i_j's, and with

$$\beta^i_{jj'} := \mathbf{1}_{j'=j}\alpha_j^{-1}\rho_{ij} \sum_{j''\geqslant j} \alpha_{j''}\rho_{ij''} + \mathbf{1}_{j'>j}\alpha_j\alpha_{j'}^{-1}\rho_{ij}\rho_{ij'}; \tag{FP}$$

then, one has

$$\mathscr{L}_{j'} = \sum_{\substack{i\in\boldsymbol{I}\\ j\in\boldsymbol{J}}} \beta^i_{jj'} = \sum_{\substack{i\in\boldsymbol{I}\\ j''\geqslant j'}} \alpha_{j'}^{-1}\rho_{ij'}\alpha_{j''}\rho_{ij''} + \sum_{\substack{i\in\boldsymbol{I}\\ j<j'}} \alpha_j\alpha_{j'}^{-1}\rho_{ij}\rho_{ij'} = \sum_{\substack{i\in\boldsymbol{I}\\ j\in\boldsymbol{J}}} \alpha_j\alpha_{j'}^{-1}\rho_{ij}\rho_{ij'}, \tag{FQ}$$

so that $\|\mathscr{L}\|_\infty$ is equal to

$$\sup_{j\in J}\sum_{\substack{i\in I\\ j'\in J}} \alpha_j^{-1}\alpha_{j'}\rho_{ij}\rho_{ij'}. \tag{FR}$$

So, we have got a formula alike to the wished one (FE), except that instead of "$|||\boldsymbol{\rho}|||^2$" we have (FR). To get the wished expression, the last step will just consist in optimizing (FR) over all the possible values for $\vec{\alpha}_{\boldsymbol{J}}$. But (denoting "$\vec{\alpha} > 0$" to mean that all the α_j's are strictly positive):

$$\begin{aligned}\inf_{\vec{\alpha}>0}\sup_{j\in J}\sum_{\substack{i\in I\\ j'\in J}} \alpha_j^{-1}\alpha_{j'}\rho_{ij}\rho_{ij'} &= \inf_{\vec{\alpha}>0}\inf\Big\{\lambda \geqslant 0 \;\Big|\; \forall j \sum_{\substack{i\in I\\ j'\in J}} \alpha_{j'}\rho_{ij}\rho_{ij'} \leqslant \lambda\alpha_j\Big\}\\ &= \inf_{\vec{\alpha}>0}\inf\{\lambda\geqslant 0 \mid \boldsymbol{\rho}^*\boldsymbol{\rho}\vec{\alpha} \leqslant \lambda\vec{\alpha}\} = \inf\{\lambda \geqslant 0 \mid \exists\vec{\alpha}>0\ \boldsymbol{\rho}^*\boldsymbol{\rho}\vec{\alpha}\leqslant\lambda\vec{\alpha}\}\\ &= \mathrm{\rho}_{\mathrm{sp}}(\boldsymbol{\rho}^*\boldsymbol{\rho}) = |||\boldsymbol{\rho}|||^2, \end{aligned} \tag{FS}$$

where the last but one equality comes from Lemma 46 at the end of the section (applied to matrix $\boldsymbol{\rho}^*\boldsymbol{\rho}$), while the last equality comes from the fact that $\boldsymbol{\rho}^*\boldsymbol{\rho}$ is a positive normal operator, so that $\mathrm{\rho}_{\mathrm{sp}}(\boldsymbol{\rho}^*\boldsymbol{\rho}) = \sup_{\vec{x}\neq\vec{0}}(\vec{x}^{\mathsf{T}}\boldsymbol{\rho}^*\boldsymbol{\rho}\vec{x} \,/\, \|\vec{x}\|^2) = \sup_{\vec{x}\neq\vec{0}}(\|\boldsymbol{\rho}\vec{x}\| \,/\, \|\vec{x}\|)^2 = |||\boldsymbol{\rho}|||^2$. This ends the proof. ♤

Proof of Lemma 44. We will prove the lemma by induction on $\boldsymbol{N}$. The case $\boldsymbol{N}=0$ is trivial. Now, suppose $\boldsymbol{N}\geqslant 1$ and assume the result is true for $\boldsymbol{N}-1$. Assume that $\|\mathscr{L}\|_{\ell^\infty}\leqslant 1$, otherwise there is nothing to prove. We generalize the notation $\mathscr{L}$ by denoting, for $\bullet\in\{\textvisiblespace, 1, \mathrm{c}\}$:

$$\mathscr{L}^\bullet\vec{x} \coloneqq \sum_{\substack{i\in I^\bullet\\ j,j'\in J}} \beta^i_{jj'}x_{j'}, \tag{FT}$$

with $\boldsymbol{I}^1 \coloneqq \{0\}$, resp. $\boldsymbol{I}^{\mathrm{c}} = \boldsymbol{I}\smallsetminus\boldsymbol{I}^1$, so that $\mathscr{L} = \mathscr{L}^1 + \mathscr{L}^{\mathrm{c}}$. Note that $\|\mathscr{L}^{\mathrm{c}}\|_{\ell^\infty}\leqslant 1$ since $\|\mathscr{L}\|_{\ell^\infty}\leqslant 1$. By the induction hypothesis, we have:

$$\sum_{j\in J} W_j^N \geqslant \sum_{j\in J} W_j^1 - \mathscr{L}^{\mathrm{c}}((W_j^1)_{j\in J}). \tag{FU}$$

Now define, for all $j\in\boldsymbol{J}$,

$$\check{W}_j^1 \coloneqq W_j^0 - \sum_{j'\in J}\beta^0_{jj'}W^0_{j'}, \tag{FV}$$

which is the value that W_j^1 would take if there were equality in (FM) for $i=0$. With that notation, (FM) writes:

$$(W_j^1 - \check{W}_j^1)_{j\in J} \geqslant 0. \tag{FW}$$

Therefore, we have the following chain of inequalities:

$$\begin{aligned}
\sum_{j\in J} W_j^N &\underset{(\mathtt{FU})}{\geqslant} \sum_{j\in J} W_j^1 - \mathscr{L}^{\mathrm{c}}((W_j^1)_{j\in J}) \\
&\underset{(\mathtt{FW})}{=} \sum_{j\in J} \check{W}_j^1 + \|(W_j^1 - \check{W}_j^1)_{j\in J}\|_{\ell^1} - \mathscr{L}^{\mathrm{c}}((\check{W}_j^1)_{j\in J}) - \mathscr{L}^{\mathrm{c}}\big((W_j^1 - \check{W}_j^1)_{j\in J}\big) \\
&\underset{\|\mathscr{L}^{\mathrm{c}}\|_\infty \leqslant 1}{\geqslant} \sum_{j\in J} \check{W}_j^1 - \mathscr{L}^{\mathrm{c}}((\check{W}_j^1)_{j\in J}) = \sum_{j\in J} W_j^0 - \mathscr{L}^1((W_j^0)_{j\in J}) - \mathscr{L}^{\mathrm{c}}((\check{W}_j^1)_{j\in J}) \\
&\underset{\substack{\check{W}^1 \leqslant W^0 \\ \mathscr{L}^{\mathrm{c}} \geqslant 0}}{\geqslant} \sum_{j\in J} W_j^0 - \mathscr{L}^1((W_j^0)_{i\in I}) - \mathscr{L}^{\mathrm{c}}((W_j^0)_{j\in J}) = \sum_{j\in J} W_j^0 - \mathscr{L}((W_j^0)_{j\in J}), \qquad (\mathtt{FX})
\end{aligned}$$

so (`FO`) is true for N, whence the lemma by induction. ♤

Remark 45. Our proof of Theorem 39 handled the X_i's and the Y_j's in a fully non-symmetric way, since we began with putting orders on I and J, which orders played a crucial role in the decomposition of g. Yet the obtained bound (`EB`) is obviously symmetric by re-labelling the basic variables—and this is not due to having proceeded to any "re-symmetrization" step… To date I have no simple explanation for this "coincidence". ♧

To close this section, it just remains to state and prove the lemma that I invoked to get Equation (`FS`):

Lemma 46. *Denote $\vec{v} \geqslant 0$, resp. $\vec{v} > 0$, etc., to mean that the entries of a vector $\vec{v}$ are all nonnegative, resp. all strictly positive, etc.; also, in the context of this lemma, denote $\boldsymbol{A} \geqslant 0$ to mean that all the entries of a matrix $\boldsymbol{A}$ are nonnegative.*[§] *With that notation, the present lemma claims that, for $\boldsymbol{A}$ a square matrix with $\boldsymbol{A} \geqslant 0$:*

$$\inf\{\lambda \geqslant 0 \mid (\exists \vec{u} > 0)\ \boldsymbol{A}\vec{u} \leqslant \lambda\vec{u}\} = \rho_{\mathrm{sp}}(\boldsymbol{A}). \qquad (\mathtt{FY})$$

◇

Proof. The lemma is actually a corollary of the Perron–Frobenius theorem [7, Theorem 8.3.1], which asserts the existence of a $\rho_{\mathrm{sp}}(\boldsymbol{A})$-eigenvector $\vec{v}_0$ of $\boldsymbol{A}$ such that $\vec{v}_0 \geqslant 0$. In the sequel we assume such a $\vec{v}_0$ to be fixed.

We will prove separately the '$\leqslant$' and '$\geqslant$' senses of the equality. Let us begin with sense '$\leqslant$'. In the case $\vec{v}_0 > 0$, the value $\lambda \leftarrow \rho_{\mathrm{sp}}(\boldsymbol{A})$ is obviously admissible for the infimum in (`FY`) and then we are done. Otherwise if $\vec{v}_0 \not> 0$, up to a permutation of indices, $\vec{v}_0$ has the form $(\vec{0}_m; \vec{v}_2)$ for some $m \in \rrbracket 0, n \llbracket$, where n stands for the dimension of $\vec{v}_0$, $\vec{0}_m$ stands for the null vector of $\mathbb{R}^m$, and $\vec{v}_2$ is a strictly positive vector of $\mathbb{R}^{n-m}$. Then, all first m entries of $\boldsymbol{A}\vec{v}_0 = \rho_{\mathrm{sp}}(\boldsymbol{A})\vec{v}_0$ must be zero; and since $\vec{v}_2 > 0$ and $\boldsymbol{A} \geqslant 0$, this forces $\boldsymbol{A}$ to write blockwise

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & \mathbf{0}_{m\times(n-m)} \\ \cdots & \cdots \end{pmatrix} \qquad (\mathtt{FZ})$$

[§] On the other hand, in § 3.4 notation "$\boldsymbol{A} \geqslant \mathbf{0}$" will mean that $\boldsymbol{A}$ is the matrix of a *positive semidefinite quadratic form.*

for some nonnegative matrix $\boldsymbol{A}_1 \in \mathbb{R}^{m\times m}$. Using the fact that, for $\vec{u}_1$ a vector in $\mathbb{R}^m$, one has $\|(\vec{u}_1;\vec{0}_{n-m})\| = \|\vec{u}_1\|$ and

$$\|\boldsymbol{A}^k(\vec{u}_1;\vec{0}_{n-m})\| = \|(\boldsymbol{A}_1^k\vec{u}_1;\ldots)\| \geqslant \|\boldsymbol{A}_1^k\vec{u}_1\|, \tag{GA}$$

it is clear that $\rho_{\mathrm{sp}}(\boldsymbol{A}_1) \leqslant \rho_{\mathrm{sp}}(\boldsymbol{A})$. Now let $\varepsilon > 0$. Reasoning by induction, assume that we have already established the '$\leqslant$' sense of the lemma in dimension m: then there exists some $\vec{v}_1 > 0$ in $\mathbb{R}^m$ such that $\boldsymbol{A}_1\vec{v}_1 \leqslant (\rho_{\mathrm{sp}}(\boldsymbol{A}_1)+\varepsilon)\vec{v}_1$. Thus for $\eta > 0$, the n-vector $(\eta\vec{v}_1;\vec{v}_2)$ is strictly positive, and in the asymptotics $\eta \searrow 0$:

$$\begin{aligned}\boldsymbol{A}(\eta\vec{v}_1;\vec{v}_2) &= (\eta\boldsymbol{A}_1\vec{v}_1;\rho_{\mathrm{sp}}(\boldsymbol{A})\vec{v}_2 + \mathrm{O}(\eta))\\ &\leqslant \big((\rho_{\mathrm{sp}}(\boldsymbol{A}_1)+\varepsilon)\eta\vec{v}_1;\rho_{\mathrm{sp}}(\boldsymbol{A})\vec{v}_2 + \mathrm{O}(\eta)\big) \overset{\eta\searrow 0}{\leqslant} \big(\rho_{\mathrm{sp}}(\boldsymbol{A})+\varepsilon\big)(\eta\vec{v}_1;\vec{v}_2)^{[\P]};\end{aligned} \tag{GB}$$

so $\lambda \leftarrow \rho_{\mathrm{sp}}(\boldsymbol{A})+\varepsilon$ is admissible for the infimum in (FY), which ends the proof of the sense '$\leqslant$'.

For the sense '$\geqslant$', let $\lambda < \rho_{\mathrm{sp}}(\boldsymbol{A})$ and $\vec{u} > 0$; we want to prove that $\boldsymbol{A}\vec{u} \nleqslant \lambda\vec{u}$. Considering again the vector $\vec{v}_0$ given by the Perron–Frobenius theorem, there exists a (unique) $\beta \geqslant 0$ such that $\beta\vec{v}_0 \leqslant \vec{u}$ but $\beta\vec{v}_0 \nless \vec{u}$: for this β, one of the entries of $\beta\vec{v}_0$ and $\vec{u}$, say entry $\#i$, has to be identical for both vectors. Then, denoting $[\vec{v}]_i$ for entry $\#i$ of a vector $\vec{v}$:

$$\lambda[\vec{u}]_i < \rho_{\mathrm{sp}}(\boldsymbol{A})[\vec{u}]_i = \rho_{\mathrm{sp}}(\boldsymbol{A})\beta[\vec{v}_0]_i = [\boldsymbol{A}(\beta\vec{v}_0)]_i \leqslant [\boldsymbol{A}(\beta\vec{v}_0)]_i + [\boldsymbol{A}(\vec{u}-\beta\vec{v}_0)]_i = [\boldsymbol{A}\vec{u}]_i, \tag{GC}$$

so that $\boldsymbol{A}\vec{u} \nleqslant \lambda\vec{u}$. ♤

2.5 Parallel tensorization

Theorem 39 that we have been proving in the previous section is rather sharp: actually, as one will see later [Theorem 62 in § 2.6.3], it is even optimal at first order. However the bound (EB) is certainly not *exactly* optimal, for several steps in our proof let some space for improvements: in particular,

- In deriving (EQ), we neglected the fact that, because g_j and $\tilde{g}_{j+1}$ are orthogonal in $\bar{\mathrm{L}}^2$, it is not possible to have simultaneously nonzero correlation between f and g_j and full correlation between f and $\tilde{g}_{j+1}$. This is actually the same kind of observation that the rough simple tensorization theorem (Theorem 36) missed compared to the sharp simple tensorization theorem (Theorem 37).
- When linearizing (FH) into (FI), we did not allow the α_j's to depend on i. According to Remark 42, this will necessarily lead to some loss of optimality as soon as the sequences $(W_j^i)_j$'s are not all proportional to each other for different values of i—which has no reason to be the case!

[¶] Note that in the last inequality we used that $\vec{v}_2 > 0$ to ensure that $\mathrm{O}(\eta) \leqslant \varepsilon\vec{v}_2$ for η small enough.

However, given the look of the proof of Theorem 39, it seems unlikely that one could derive the strictly optimal bound in the general case. Nevertheless, it turns out that such an optimal bound <u>can</u> be computed when we introduce some extra symmetries in the problem. The corresponding result, which is the subject of this section, is the following:

Theorem 47 (Parallel tensorization)**.** *Let I and J be sets isomorphic to $\mathbb{Z}$, and let $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ be random variables such that, for all $i,j \in \mathbb{Z}$:*

$$\rho^\ddagger(X_i;Y_j) \leqslant \rho_{j-i} \tag{GD}$$

for some function $\rho\colon \mathbb{Z} \to [0,1]$. Then:

$$\rho(\vec{X}_I;\vec{Y}_J) \leqslant \bar{\rho}, \tag{GE}$$

where $\bar{\rho} \in [0,1]$ is characterized by

$$\arcsin \bar{\rho} \coloneqq \Big(\sum_{z\in\mathbb{Z}} \arcsin \rho_z\Big) \wedge \frac{\pi}{2}. \tag{GF}$$

◇

Remark 48. Applying Theorem 39 in this situation would yield "$\rho(\vec{X}_I;\vec{Y}_J) \leqslant \big(\sum_{z\in\mathbb{Z}} \rho_z\big) \wedge 1$", which is indeed always $\geqslant \bar{\rho}$, with strict inequality in general: so, when applicable, Theorem 47 is strictly stronger than Theorem 39. ♧

Proof of Theorem 47. Let f and g be resp. $\vec{X}_I$- and $\vec{Y}_J$-measurable L^2 variables; our goal is obviously to bound $\mathrm{Cov}(f,g)$ by $\bar{\rho}\,\mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g)$.

Before all, in order to avoid issues with infinite summations, we will assume by a harmless perturbation argument that there exist finite subsets $\check{I} \Subset I, \check{J} \Subset J$ such that f is $\vec{X}_{\check{I}}$-measurable, resp. g is $\vec{Y}_{\check{J}}$-measurable. Then, we also assume that for all $i \notin \check{I}$, resp. $j \notin \check{J}$, one has $X_i, Y_j \equiv \partial$, where ∂ is a cemetery value, so that conditioning by $X_{i\notin\check{I}}$ or by $Y_{j\notin\check{J}}$ will never have any effect: note that this modification of our problem keeps all the assumptions valid. We will denote by i_-, i_+, j_-, j_+ indices such that $\check{I} \subseteq \llbracket i_-, i_+ \llbracket$, resp. $\check{J} \subseteq \llbracket j_-, j_+ \llbracket$.

For $i \in \mathbb{Z}$, resp. $j \in \mathbb{Z}$, we denote as usual $\mathscr{F}_i \coloneqq \bigvee_{i'<i} \sigma(X_{i'})$, resp. $\mathscr{G}_j \coloneqq \bigvee_{j'<j} \sigma(Y_{j'})$; for $(i,j) \in \mathbb{Z}\times\mathbb{Z}$, we define

$$f_i^j \coloneqq f^{\mathscr{F}_{i+1}} - \mathbb{E}(f^{\mathscr{F}_{i+1}} \mid \mathscr{F}_i \vee \mathscr{G}_j) \tag{GG}$$

and[‖]

$$g_j^i \coloneqq g^{\mathscr{F}_i \vee \mathscr{G}_{j+1}} - \mathbb{E}(g \mid \mathscr{F}_i \vee \mathscr{G}_j). \tag{GH}$$

We finally denote

$$V_i^j \coloneqq \mathrm{Var}(f_i^j); \tag{GI}$$

$$W_j^i \coloneqq \mathrm{Var}(g_j^i); \tag{GJ}$$

$$C_{ij} \coloneqq \mathrm{Cov}(f_i^j, g_j^i). \tag{GK}$$

The quantities V_i^j, W_j^i, C_{ij} are linked to our essential quantities in the following way:

[‖] Beware: the definition of g_j^i is <u>not</u> analogous to the definition of f_i^j!

Proposition 49.

(i)

$$\mathrm{Var}(f) = \sum_i \overset{\nearrow}{\lim_{j\to-\infty}} V_i^j =: V. \tag{GL}$$

(ii)

$$\mathrm{Var}(g) = \overset{\nearrow}{\lim_{i\to-\infty}} \sum_j W_j^i =: W. \tag{GM}$$

(iii)

$$\mathrm{Cov}(f,g) = \sum_{i,j} C_{ij} =: C. \tag{GN}$$

◇

Remark 50. Due to the finiteness hypotheses that we have added at the beginning of the proof, one has $V_i^j = 0\ \forall i \notin \check{I}$, resp. $W_j^i = 0\ \forall j \notin \check{J}$, resp. $C_{ij} = 0\ \forall (i,j) \notin \check{I} \times \check{J}$; and V_i^j does not depend on j for $j \leqslant j_-$ nor for $j \geqslant j_+$, nor does W_j^i depend on i for $i \leqslant i_-$ or for $i \geqslant i_+$. As a consequence, all the sums above are actually finite, and all the limits are actually stationary. ♧

Proof of Proposition 49-(i). First, by the chain rule for variance (CM) applied with 'f' $\leftarrow f^{\mathscr{F}_{i+1}}$, '$N$' $\leftarrow 2$, '$\mathscr{F}_0$' $\leftarrow \mathscr{O}$, '$\mathscr{F}_1$' $\leftarrow \mathscr{F}_i \vee \mathscr{G}_j$ and '$\mathscr{F}_2$' $\leftarrow \mathscr{B}$, one has $V_i^j = \mathrm{Var}(f^{\mathscr{F}_{i+1}}) - \mathrm{Var}((f^{\mathscr{F}_{i+1}})^{\mathscr{F}_i \vee \mathscr{G}_j})$. Since the σ-algebras $\mathscr{F}_i \vee \mathscr{G}_j$ are increasing in j, $\mathrm{Var}((f^{\mathscr{F}_{i+1}})^{\mathscr{F}_i \vee \mathscr{G}_j})$ decreases as j decreases (cf. Lemma 27); thus V_i^j increases as j decreases. When $j \to -\infty$, the σ-algebra $\mathscr{F}_i \vee \mathscr{G}_j$ converges (and stationates) to $\mathscr{F}_i$, so f_i^j converges to $f^{\mathscr{F}_{i+1}} - \mathbb{E}(f^{\mathscr{F}_{i+1}} \mid \mathscr{F}_i) = f^{\mathscr{F}_{i+1}} - \mathbb{E}(f \mid \mathscr{F}_i) =: f_i$; and so V_i^j converges to $\mathrm{Var}(f_i)$.

But we can apply the chain rule for variance with 'f' $\leftarrow f$, 'N' $\leftarrow i_+ - i_-$, '$\mathscr{F}_0$' $\leftarrow \mathscr{F}_{i_-}$, '$\mathscr{F}_1$' $\leftarrow \mathscr{F}_{i_-+1}$, …, '$\mathscr{F}_N$' $\leftarrow \mathscr{F}_{i_+}$ to get that

$$\mathrm{Var}(f) = \mathrm{Var}(f^{\mathscr{F}_{i_+}} - f^{\mathscr{F}_{i_-}}) = \sum_{i=i_-}^{i_+-1} \mathrm{Var}(f_i) = \sum_{i\in\mathbb{Z}} \mathrm{Var}(f_i), \tag{GO}$$

where the first equality used that $f^{\mathscr{F}_{i_+}} = f$ and $f^{\mathscr{F}_{i_-}} \equiv 0$, and the last equality used that $f_i \equiv 0$ as soon as $i \notin \check{I}$. This proves (GL). ♤

Proof of Proposition 49-(ii). First, similarly to the second part of the proof of Proposition 49-(i), we apply the chain rule for variance (CM) with 'f' $\leftarrow g$, 'N' $\leftarrow j_+ - j_-$, '$\mathscr{F}_0$' $\leftarrow \mathscr{F}_i \vee \mathscr{G}_{j_-}$, …, '$\mathscr{F}_N$' $\leftarrow \mathscr{F}_i \vee \mathscr{G}_{j_+}$, getting that $\sum_{j\in\mathbb{Z}} W_j^i = \mathrm{Var}(g - g^{\mathscr{F}_i}) =: \mathrm{Var}(g^i)$. In a second time, similarly to the first part of the proof of Proposition 49-(i), we find that $\mathrm{Var}(g^i) = \mathrm{Var}(g) - \mathrm{Var}(g^{\mathscr{F}_i})$, and we observe that $\mathrm{Var}(g^{\mathscr{F}_i})$ decreases when i decreases and is zero for $i \leqslant i_-$; which proves (GM). ♤

Proof of Proposition 49-(iii). The first step of the proof consists in expanding $\mathrm{Cov}(f, g)$ as a sum indexed by i. Re-introduce the notation $f_i := f^{\mathscr{F}_{i+1}} - \mathbb{E}(f \mid \mathscr{F}_i)$ and

$g^i \coloneqq g - \mathbb{E}(g \mid \mathscr{F}_i)$ from the proofs of Items (i) and (ii). Since $\mathscr{F} = \mathscr{F}_{i_+}$ and $\mathscr{O} = \mathscr{F}_{i_-}$, one has the telescopic decomposition:

$$f - \mathbb{E}(f) = \sum_{i=i_-}^{i_+-1} f_i, \tag{GP}$$

whence

$$V = \sum_{i\in\mathbb{Z}} \mathrm{Cov}(f_i, g). \tag{GQ}$$

But f_i is centred w.r.t. $\mathscr{F}_i$, so by Equation (BW) of Lemma 20, $\mathrm{Cov}(f_i, g^{\mathscr{F}_i}) = 0$; and therefore, for all i:

$$\mathrm{Cov}(f_i, g) = \mathrm{Cov}(f_i, g - g^{\mathscr{F}_i}) = \mathrm{Cov}(f_i, g^i). \tag{GR}$$

Combining (GQ) and (GR) yields:

$$V = \sum_{i\in\mathbb{Z}} \mathrm{Cov}(f_i, g^i). \tag{GS}$$

The second step of the proof consists in expanding $\mathrm{Cov}(f_i, g^i)$ as a sum indexed by j. We start with observing that $g^i = g^{\mathscr{F}_i\vee\mathscr{G}_{j_+}} - g^{\mathscr{F}_i\vee\mathscr{G}_{j_-}}$, so that one has the telescopic decomposition:

$$g^i = \sum_{j\in\mathbb{Z}} g^i_j, \tag{GT}$$

whence

$$\mathrm{Cov}(f_i, g^i) = \sum_{j\in\mathbb{Z}} \mathrm{Cov}(f_i, g^i_j). \tag{GU}$$

But g^i_j is centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_j$, so by Lemma 20, $\mathrm{Cov}((f_i)^{\mathscr{F}_i\vee\mathscr{G}_j}, g^i_j) = 0$, and therefore, for all j:

$$\mathrm{Cov}(f_i, g^i_j) = \mathrm{Cov}(f_i - (f_i)^{\mathscr{F}_i\vee\mathscr{G}_j}, g^i_j) = \mathrm{Cov}(f^j_i, g^i_j) \underset{\mathrm{def}}{=} C_{ij}. \tag{GV}$$

Combining (GS) with (GU) and (GV) finally yields (GN). ♤

Now we are going to set certain relations between the V^j_i's, the W^i_j's and the C_{ij}'s: these correspond to the three lemmas below.

Lemma 51. *For all* $i \in I, j \in J$*:*

$$|C_{ij}| \leqslant \rho_{j-i} \sqrt{V^j_i W^i_j}\,. \tag{GW}$$

◇

Proof of Lemma 51. Conditionally to $\mathscr{F}_i \vee \mathscr{G}_j$, f^j_i is X_i-measurable and g^i_j is Y_j-measurable, while $\rho(X_i; Y_j) \leqslant \rho_{j-i}$ because of the $\rho^\ddagger$-mixing hypothesis; so, conditionally to $\mathscr{F}_i \vee \mathscr{G}_j$,

$$|\mathrm{Cov}(f^j_i, g^i_j)| \leqslant \rho_{j-i}\, \mathrm{Var}^{1/2}(f^j_i)\, \mathrm{Var}^{1/2}(g^i_j), \tag{GX}$$

which means in non-conditioned terms that

$$\left|\mathrm{Cov}(f_i^j, g_j^i \mid \mathscr{F}_i \vee \mathscr{G}_j)\right| \leqslant \rho_{j-i} \operatorname{Var}^{1/2}(f_i^j \mid \mathscr{F}_i \vee \mathscr{G}_j) \operatorname{Var}^{1/2}(g_j^i \mid \mathscr{F}_i \vee \mathscr{G}_j). \tag{GY}$$

But, when integrating over all the possible values of $(\vec{X}_{<i}, \vec{Y}_{<j})$, one has on the one hand that, as f_i^j and g_j^i are centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_j$, by Equation (`CC`) of Lemma 23:

$$\mathrm{Cov}(f_i^j, g_j^i) = \int \mathrm{Cov}(f_i^j, g_j^i \mid \vec{x}_{<i}, \vec{y}_{<j}) \,\mathrm{d}\mathbb{P}(\vec{x}_{<i}, \vec{y}_{<j}); \tag{GZ}$$

and on the other hand that:

$$\begin{aligned}\int \operatorname{Var}^{1/2}(f_i^j \mid \vec{x}_{<i}, \vec{y}_{<j}) \operatorname{Var}^{1/2}(g_j^i \mid \vec{x}_{<i}, \vec{y}_{<j}) \,\mathrm{d}\mathbb{P}(\vec{x}_{<i}, \vec{y}_{<j}) \underset{\mathrm{CS}}{\leqslant}& \\ \Big(\int \operatorname{Var}(f_i^j \mid \vec{x}_{<i}, \vec{y}_{<j}) \,\mathrm{d}\mathbb{P}(\vec{x}_{<i}, \vec{y}_{<j})\Big)^{1/2} \Big(\int \operatorname{Var}(g_j^i \mid \vec{x}_{<i}, \vec{y}_{<j}) \,\mathrm{d}\mathbb{P}(\vec{x}_{<i}, \vec{y}_{<j})\Big)^{1/2}& \\ \underset{(\mathtt{CK})}{\leqslant} \operatorname{Var}^{1/2}(f_i^j) \operatorname{Var}^{1/2}(g_j^i).&\end{aligned} \tag{HA}$$

So, integrating (`GY`) and using (`GZ`) and (`HA`), we get that

$$|\mathrm{Cov}(f_i^j, g_j^i)| \leqslant \rho_{j-i} \operatorname{Var}^{1/2}(f_i^j) \operatorname{Var}^{1/2}(g_j^i), \tag{HB}$$

i.e. (`GW`). ♤

Lemma 52. *For all $i \in \boldsymbol{I}, j \in \boldsymbol{J}$: (provided $W_j^i > 0$),*

$$V_i^{j+1} \leqslant V_i^j - (C_{ij})^2 / W_j^i. \tag{HC}$$

◇

Proof of Lemma 52. Since g_j^i is $(\mathscr{F}_i \vee \mathscr{G}_{j+1})$-measurable, by Equation (`BW`) of Lemma 20, $C_{ij} \underset{\mathrm{def}}{=} \mathrm{Cov}(f_i^j, g_j^i) = \mathrm{Cov}((f_i^j)^{\mathscr{F}_i \vee \mathscr{G}_{j+1}}, g_j^i)$, so by Cauchy–Schwarz:

$$|C_{ij}| \leqslant \operatorname{Var}^{1/2}((f_i^j)^{\mathscr{F}_i \vee \mathscr{G}_{j+1}}) \operatorname{Var}^{1/2}(g_j^i) = \operatorname{Var}^{1/2}((f_i^j)^{\mathscr{F}_i \vee \mathscr{G}_{j+1}}) \sqrt{W_j^i}\,. \tag{HD}$$

But one has $(f_i^j)^{\mathscr{F}_i \vee \mathscr{G}_{j+1}} = (f^{\mathscr{F}_{i+1}})^{\mathscr{F}_i \vee \mathscr{G}_{j+1}} - \mathbb{E}(f^{\mathscr{F}_{i+1}} \mid \mathscr{F}_i \vee \mathscr{G}_j)$; so we can apply the chain rule for variance (`CG`) with 'N' $\leftarrow$ 2, 'f' $\leftarrow$ $f^{\mathscr{F}_{i+1}}$, '$\mathscr{F}_0$' $\leftarrow$ $\mathscr{F}_i \vee \mathscr{G}_j$, '$\mathscr{F}_1$' $\leftarrow$ $\mathscr{F}_i \vee \mathscr{G}_{j+1}$, '$\mathscr{F}_2$' $\leftarrow$ $\mathscr{B}$ to get that

$$\operatorname{Var}((f_i^j)^{\mathscr{F}_i \vee \mathscr{G}_{j+1}}) \leqslant V_i^j - V_i^{j+1}. \tag{HE}$$

Combining (`HD`) with (`HE`) yields (`HC`). ♤

Lemma 53. *For all $i \in \boldsymbol{I}, j \in \boldsymbol{J}$: (provided $V_i^j > 0$),*

$$\sum_{j' \geqslant j} W_{j'}^{i+1} \leqslant \sum_{j' \geqslant j} W_{j'}^i - \Big(\sum_{j' \geqslant j} C_{ij'}\Big)^2 / V_i^j. \tag{HF}$$

◇

Proof of Lemma 53. For $j' \geqslant j$, first observe that $C_{ij'} \underset{\text{def}}{=} \operatorname{Cov}(f_i^{j'}, g_{j'}^i) = \operatorname{Cov}(f_i^j, g_{j'}^i)$: indeed, since $f_i^{j'} - f_i^j$ is $(\mathscr{F}_i \vee \mathscr{G}_{j'})$-measurable while $g_{j'}^i$ is centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_{j'}$, one has $\operatorname{Cov}(f_i^{j'} - f_i^j, g_{j'}^i) = 0$ by Equation (`BX`) of Lemma 20. Therefore:

$$\sum_{j' \geqslant j} C_{ij'} = \operatorname{Cov}\Big(f_i^j, \sum_{j' \geqslant j} g_{j'}^i\Big) = \operatorname{Cov}(f_i^j, g - g^{\mathscr{F}_i \vee \mathscr{G}_j}). \tag{HG}$$

But f_i^j is $(\mathscr{F}_{i+1} \vee \mathscr{G}_j)$-measurable, so $\operatorname{Cov}(f_i^j, g - g^{\mathscr{F}_i \vee \mathscr{G}_j}) = \operatorname{Cov}\big(f_i^j, (g - g^{\mathscr{F}_i \vee \mathscr{G}_j})^{\mathscr{F}_{i+1} \vee \mathscr{G}_j}\big)$ by Lemma 20; and then Cauchy–Schwarz yields:

$$\Big|\sum_{j' \geqslant j} C_{ij'}\Big| \leqslant \sqrt{V_i^j}\, \operatorname{Var}^{1/2}\big((g - g^{\mathscr{F}_i \vee \mathscr{G}_j})^{\mathscr{F}_{i+1} \vee \mathscr{G}_j}\big). \tag{HH}$$

But $(g - g^{\mathscr{F}_i \vee \mathscr{G}_j})^{\mathscr{F}_{i+1} \vee \mathscr{G}_j} = g^{\mathscr{F}_{i+1} \vee \mathscr{G}_j} - g^{\mathscr{F}_i \vee \mathscr{G}_j}$, so by chain rule for variance (`CM`) applied with 'f' $\leftarrow g$, 'N' $\leftarrow 2$, '$\mathscr{F}_0$' $\leftarrow \mathscr{F}_i \vee \mathscr{G}_j$, '$\mathscr{F}_1$' $\leftarrow \mathscr{F}_{i+1} \vee \mathscr{G}_j$, '$\mathscr{F}_2$' $\leftarrow \mathscr{B}$:

$$\operatorname{Var}\big((g - g^{\mathscr{F}_i \vee \mathscr{G}_j})^{\mathscr{F}_{i+1} \vee \mathscr{G}_j}\big) = \operatorname{Var}(g - g^{\mathscr{F}_i \vee \mathscr{G}_j}) - \operatorname{Var}(g - g^{\mathscr{F}_{i+1} \vee \mathscr{G}_j}). \tag{HI}$$

Finally, applying the chain rule for variance with 'f' $\leftarrow g$, 'N' $\leftarrow j_+ - j$, '$\mathscr{F}_0$' $\leftarrow \mathscr{F}_i \vee \mathscr{G}_j$, …, '$\mathscr{F}_N$' $\leftarrow \mathscr{F}_i \vee \mathscr{G}_{j_+}$, we get that

$$\operatorname{Var}(g - g^{\mathscr{F}_i \vee \mathscr{G}_j}) = \sum_{j' \geqslant j} W_j^i, \tag{HJ}$$

and likewise

$$\operatorname{Var}(g - g^{\mathscr{F}_{i+1} \vee \mathscr{G}_j}) = \sum_{j' \geqslant j} W_j^{i+1}. \tag{HK}$$

Combining (`HH`) with (`HI`) and using (`HJ`) and (`HK`), we get the announced result. ♤

So, we have transformed our initial probabilistic problem into the following analytic one. Let $\boldsymbol{A}$ be an array indexed by $\mathbb{Z} \times \mathbb{Z}$, each entry (i, j) of which contains three numbers $V_i^j \geqslant 0$, $W_j^i \geqslant 0$ and $C_{ij} \in \mathbb{R}$, satisfying the following conditions:

Compactness: There exist indices $i_- \leqslant i_+$, $j_- \leqslant j_+$ such that:

- V_i^j is zero as soon as $i \notin [\![i_-, i_+ [\![$, and does not depend on j for $j \leqslant j_-$, nor for $j \geqslant j_+$;
- W_j^i is zero as soon as $j \notin [\![j_-, j_+ [\![$, and does not depend on i for $i \leqslant i_-$, nor for $i \geqslant i_+$;
- C_{ij} is zero as soon as $(i, j) \notin [\![i_-, i_+ [\![\times [\![j_-, j_+ [\![$.

Correctness: For all $(i, j) \in \mathbb{Z} \times \mathbb{Z}$, Equations (`GW`), (`HC`) and (`HF`) are satisfied.

Define resp. $\boldsymbol{V}$, $\boldsymbol{W}$ and $\boldsymbol{C}$ from the V_i^j's, the W_j^i's and the C_{ij}'s, like in resp. (`GL`), (`GM`) and (`GN`); then our goal is to show that one has $|\boldsymbol{C}| \leqslant \bar{\rho}\sqrt{\boldsymbol{V}\boldsymbol{W}}$, $\bar{\rho}$ being defined by (`GF`).

To tackle that issue, we will first summarize our *two*-indices arrays into *one*-index objects, thanks to the translational symmetries of the problem. Here of how this is done. For $z \in \mathbb{Z}$, we define the *shift operator* $\boldsymbol{\sigma}^z$ on arrays indexed by $\mathbb{Z} \times \mathbb{Z}$ in the following way: if the entries of $\boldsymbol{A}$ at (i,j) are resp. V_i^j, W_j^i, C_{ij}, the entries of $\boldsymbol{\sigma}^z\boldsymbol{A}$ at (i,j) shall be resp. $V_{i+z}^{j+z}, W_{j+z}^{i+z}, C_{(i+z)(j+z)}$. Next, for $k \in \mathbb{N}$, we define the operator τ_k by:

$$\tau_k \boldsymbol{A} = \sum_{z=-k}^{k-1} \sigma^z \boldsymbol{A}, \tag{HL}$$

where summation of arrays is to be understood entry-wise. It turns out that $\tau_k\boldsymbol{A}$ inherits the properties of $\boldsymbol{A}$:

Lemma 54. *If the (compact) array $\boldsymbol{A}$ is correct, then $\tau_k\boldsymbol{A}$ is (compact and) correct too.* ◇

Proof of Lemma 54. To prove Lemma 54, we will prove that correctness is preserved both by translation operators and by summation. The first point is immediate. For the second point, consider $\acute{\boldsymbol{A}}$ and $\grave{\boldsymbol{A}}$ two correct (compact) arrays (whose entries are denoted by $\acute{V}_i^j, \acute{W}_j^i, \acute{C}_{ij}$ for $\acute{\boldsymbol{A}}$, resp. by $\grave{V}_i^j, \grave{W}_j^i, \grave{C}_{ij}$ for $\grave{\boldsymbol{A}}$); we want to prove that the (compact) array $\acute{\boldsymbol{A}} + \grave{\boldsymbol{A}}$ is correct too, i.e. that for all $(i,j) \in \mathbb{Z}^2$, one has:

$$|\acute{C}_{ij} + \grave{C}_{ij}| \leqslant \rho_{j-i}\sqrt{(\acute{V}_i^j + \grave{V}_i^j)}\sqrt{(\acute{W}_j^i + \grave{W}_j^i)}\,; \tag{HM}$$

$$|\acute{C}_{ij} + \grave{C}_{ij}| \leqslant \sqrt{(\acute{V}_i^j - \acute{V}_i^{j+1}) + (\grave{V}_i^j - \grave{V}_i^{j+1})}\sqrt{\acute{W}_j^i + \grave{W}_j^i}\,; \tag{HN}$$

$$\Big|\sum_{j'\geqslant j}\acute{C}_{ij'} + \sum_{j'\geqslant j}\grave{C}_{ij'}\Big| \leqslant \sqrt{\acute{V}_i^j + \grave{V}_i^j}\sqrt{\Big(\sum_{j'\geqslant j}\acute{W}_{j'}^i - \sum_{j'\geqslant j}\acute{W}_{j'}^{i+1}\Big) + \Big(\sum_{j'\geqslant j}\grave{W}_{j'}^i - \sum_{j'\geqslant j}\grave{W}_{j'}^{i+1}\Big)}. \tag{HO}$$

All these inequalities follow straightforwardly from Lemma 55, that we state and prove just below. ♤

Lemma 55. *Let $a_0, a_1, b_0, b_1 \geqslant 0$ and consider $c_0, c_1 \in \mathbb{R}$ such that $|c_0| \leqslant a_0^{1/2}b_0^{1/2}$, resp. $|c_1| \leqslant a_1^{1/2}b_1^{1/2}$; then:*

$$|c_0 + c_1| \leqslant (a_0 + a_1)^{1/2}(b_0 + b_1)^{1/2}. \tag{HP}$$

◇

Proof of Lemma 55. Obviously it is enough to show that $a_0^{1/2}b_0^{1/2} + a_1^{1/2}b_1^{1/2} \leqslant (a_0 + a_1)^{1/2}(b_0 + b_1)^{1/2}$. This inequality (which is actually a Brunn–Minkowski inequality) can be proved by observing that $\big((a_0 + a_1)^{1/2}(b_0 + b_1)^{1/2}\big)^2 - (a_0^{1/2}b_0^{1/2} + a_1^{1/2}b_1^{1/2})^2 = (a_0^{1/2}b_1^{1/2} - a_1^{1/2}b_0^{1/2})^2 \geqslant 0$. ♤

Now, the idea is to let k tend to infinity to make $\tau_k\boldsymbol{A}$ converge to some object needing only one index to describe its contents. Indeed, provided the array $\boldsymbol{A}$ is compact, when $k \to \infty$, the array $\tau_k\boldsymbol{A}$ converges entry-wise to an array that we denote by $\bar{\tau}\boldsymbol{A}$, whose entries, denoted by $\bar{V}_i^j, \bar{W}_j^i, \bar{C}_{ij}$, are respectively

$$\bar{V}_i^j = \sum\{V_{i'}^{j'} \mid j'-i' = j-i\}; \tag{HQ}$$

$$\bar{W}_j^i = \sum\{W_{j'}^{i'} \mid j'-i' = j-i\}; \tag{HR}$$

$$\bar{C}_{ij} = \sum\{C_{i'j'} \mid j'-i' = j-i\}. \tag{HS}$$

(Note that all these sums are actually finite, because of the compactness assumption). So the entries of the array $\bar{\tau}\boldsymbol{A}$ at (i,j) only depend on $j-i$; in other words, $\bar{\tau}\boldsymbol{A}$ is a *Toeplitz* array. (So we could call $\bar{\tau}$ the *Toeplitzation operator*). We denote $\bar{V}_i^j =: \bar{V}_{[j-i]}$, resp. $\bar{W}_j^i =: \bar{W}_{[j-i]}$, $\bar{C}_{ij} =: \bar{C}_{[j-i]}$. The array $\bar{\tau}\boldsymbol{A}$ too satisfies a compactness condition, adapted for Toeplitz arrays: namely, there exist indices $z_- = j_- - i_+$ and $z_+ = j_+ - i_-$ such that:

- $\bar{V}_{[z]}$ does not depend on z for $z \leqslant z_- + 1$, nor for $z \geqslant z_+$;
- $\bar{W}_{[z]}$ does not depend on z for $z \leqslant z_-$, nor for $z \geqslant z_+ - 1$;
- $\bar{C}_{[z]}$ is zero as soon as $z \notin \rrbracket z_-, z_+ \llbracket$.

Moreover, one can still get the values of V, W and C from the entries of $\bar{\tau}\boldsymbol{A}$: indeed, one sees easily that:

$$V = \bar{V}_{[z\leqslant z_-+1]}; \tag{HT}$$

$$W = \bar{W}_{[z\geqslant z_+-1]}; \tag{HU}$$

$$C = \sum_{z\in\mathbb{Z}} \bar{C}_{[z]}. \tag{HV}$$

Now, assume that $\boldsymbol{A}$ was correct. Then, what will be the consequences of the correctness constraints (GW), (HC) and (HF) on the entries of $\bar{\tau}\boldsymbol{A}$? As regards the first two constraints, the answer is straightforward: since the $\tau_k\boldsymbol{A}$'s are correct by Lemma 54, one gets by passing to the limit entry-wise in (GW) and (HC) that, for $z = j-i$ taking any value:

$$|\bar{C}_{[z]}| \leqslant \rho_z \sqrt{\bar{V}_{[z]}\bar{W}_{[z]}}\,, \tag{HW}$$

resp.

$$\bar{V}_{[z+1]} \leqslant \bar{V}_{[z]} - \bar{C}_{[z]}^2 \,/\, \bar{W}_{[z]}. \tag{HX}$$

Getting the constraint on $\bar{\tau}\boldsymbol{A}$ corresponding to (HF) is a bit trickier, because when one applies that formula to $\tau_k\boldsymbol{A}$, the sums of W_j^i's diverge as $k \to \infty$. However, we can re-write (HF) as:

$$\sum_{j'\geqslant j}(W_{j'}^i - W_{j'+1}^{i+1}) - W_j^{i+1} \geqslant \Big(\sum_{j'\geqslant j} C_{ij'}\Big)^2 \,/\, V_i^j, \tag{HY}$$

which applied to $\tau_k \boldsymbol{A}$ yields:

$$\sum_{j'\geqslant j}(W^{i-k}_{j'-k} - W^{i+k}_{j'+k}) - (\tau_k \boldsymbol{W})^{i+1}_j \geqslant \Big(\sum_{j'\geqslant j}(\tau_k \boldsymbol{S})_{ij'}\Big)^2 \,/\, (\tau_k \boldsymbol{V})^j_i. \tag{HZ}$$

But for $k \geqslant i - i_-$ one will always have $W^{i-k}_{j'-k} = W^{-\infty}_{j'-k}$, while for $k \geqslant j_+ - j$ one will always have $W^{i+k}_{j'+k} = 0$; so for k large enough, things simplify into:

$$\sum_{j'\geqslant j}(W^{i-k}_{j'-k} - W^{i+k}_{j'+k}) = \sum_{j'\geqslant j} W^{-\infty}_{j'-k} \overset{k\geqslant j-j_-}{=} \sum_{j'\in\mathbb{Z}} W^{-\infty}_{j'} = W. \tag{IA}$$

One can therefore pass to the limit in (HY) to finally get that, for $z = j - i$ taking any value:

$$\bar{W}_{[z-1]} \leqslant W - \Big(\sum_{z'\geqslant z} \bar{C}_{[z']}\Big)^2 \,/\, \bar{V}_{[z]}: \tag{IB}$$

this is the formula that we will take as counterpart to the correctness constraint (HF) when dealing with Toeplitz arrays.

So, now our goal is to prove that for a compact[**] Toeplitz array $\bar{\boldsymbol{A}}$ satisfying the correctness conditions (HW), (HX) and (IB), defining V, W, S by (HT)–(HV), one necessarily has $|S| \leqslant \bar{\rho}\sqrt{VW}$.

First, because of the compactness condition, we can safely assume that $\rho_z = 0$ for $z \notin \rrbracket z_-, z_+\llbracket$, which will avoid issues with infinite summations in the sequel. We also observe that, if $\bar{\boldsymbol{A}}$ is a correct (compact) Toeplitz array, we can safely replace the $\bar{V}_{[z]}$'s by new values $\bar{V}'_{[z]}$ defined by induction by

$$\begin{cases} \bar{V}'_{[z]} \coloneqq V & z \leqslant z_- + 1; \\ \bar{V}'_{[z+1]} \coloneqq \bar{V}'_{[z]} - \bar{C}^2_{[z]} \,/\, \bar{W}_{[z]} & z+1 > z_- + 1, \end{cases} \tag{IC}$$

so that there will always be equality in (HX): indeed, when modifying $\bar{\boldsymbol{A}}$ this way, the value of V (as defined by (HT)) remains unchanged, while $\bar{V}'_{[z]} \geqslant \bar{V}_{[z]}$ for all z, so that the correctness constraints (HW) and (IB) also remain valid. So in the sequel we assume that (HX) is actually an equality for all z.

Now, we define

$$\theta_z \coloneqq \arcsin\Big(\bar{C}_{[z]} \,/\, \sqrt{\bar{V}_{[z]}\bar{W}_{[z]}}\Big), \tag{ID}$$

noting that the constraint (HW) implies that $|\theta_z| \leqslant \arcsin \rho_z$. This way, Equation (HX) (with the '$\leqslant$' sign replaced by '$=$') becomes

$$\bar{V}_{[z+1]} = \cos^2\theta_z \times \bar{V}_{[z]}, \tag{IE}$$

so that, for all $z \in \mathbb{Z}$,

$$\bar{V}_{[z]} = \prod_{z'<z} \cos^2\theta_{z'} \times V; \tag{IF}$$

[**] Here "compact" is to be understood in the sense for Toeplitz arrays, cf. p. 43.

and then Equation (`IB`) becomes:

$$\bar{W}_{[z-1]} \,/\, W \leqslant 1 - \mathit{\Gamma}_z^2, \tag{IG}$$

where

$$\mathit{\Gamma}_z \coloneqq \sum_{z' \geqslant z} \Big(\sin\theta_{z'} \prod_{z \leqslant z'' < z'} \cos\theta_{z''} \times \sqrt{\bar{W}_{[z']} \,/\, W} \Big). \tag{IH}$$

The $\mathit{\Gamma}_z$'s satisfy the (downgoing) recursion equation

$$\mathit{\Gamma}_{z-1} = \sin\theta_{z-1} \sqrt{\bar{W}_{[z-1]} \,/\, W} + \cos\theta_{z-1} \mathit{\Gamma}_z; \tag{II}$$

so by (`IG`),

$$|\mathit{\Gamma}_{z-1}| \leqslant |\sin\theta_{z-1}| \sqrt{1 - \mathit{\Gamma}_z^2} + \cos\theta_{z-1} |\mathit{\Gamma}_z|, \tag{IJ}$$

i.e., denoting $\gamma_z \coloneqq \arcsin|\mathit{\Gamma}_z|$:

$$\sin\gamma_{z-1} \leqslant |\sin\theta_{z-1}| \cos\gamma_z + \cos\theta_{z-1} \sin\gamma_z = \sin(|\theta_{z-1}| + \gamma_z). \tag{IK}$$

We deduce by induction (observing that $\mathit{\Gamma}_z = 0$ for $z \geqslant z_+$) that, for all z:

$$\gamma_z \leqslant \Big(\sum_{z' \geqslant z} |\theta_{z'}| \Big) \wedge \frac{\pi}{2}, \tag{IL}$$

so in particular $|\mathit{\Gamma}_z| \leqslant \bar{\rho}$.

But, by construction of the $\mathit{\Gamma}_z$'s,

$$C \underset{\text{def}}{=} \sum_z \bar{C}_{[z]} = \mathit{\Gamma}_{z \leqslant z_-} \sqrt{VW}, \tag{IM}$$

so we get that $|C| \leqslant \bar{\rho} \sqrt{VW}$, *quod erat demonstrandum*. The proof of Theorem 47 is finally complete. ♤

Theorem 47 dealt specifically with structures indexed by $\mathbb{Z}$, but it can easily be generalized to the case where structures are indexed by $\mathbb{Z}^d$:

Corollary 56 (Metaparallel tensorization). *Let $d \in \mathbb{N}$; let I and J be sets isomorphic to $\mathbb{Z}^d$, and let $(X_i)_{i \in I}$ and $(Y_j)_{j \in J}$ be random variables such that, for all $i, j \in \mathbb{Z}^d$:*

$$\rho^{\ddagger}(X_i; Y_j) \leqslant \rho_{j-i} \tag{IN}$$

for some function $\rho \colon \mathbb{Z}^d \to [0,1]$. Then:

$$\rho(\vec{X}_I; \vec{Y}_J) \leqslant \bar{\rho}, \tag{IO}$$

where $\bar{\rho} \in [0,1]$ is characterized by

$$\arcsin\bar{\rho} \coloneqq \Big(\sum_{u \in \mathbb{Z}^d} \arcsin\rho_u \Big) \wedge \frac{\pi}{2}. \tag{IP}$$

◇

Proof. To alleviate notation, let us introduce the "arcsin-sum" as the binary operation $\tilde{+} \colon [0,1]^2 \to [0,1]$ defined by:

$$a \mathbin{\tilde{+}} b := \sin\Big((\arcsin a + \arcsin b) \wedge \tfrac{\pi}{2}\Big). \tag{IQ}$$

The operation $\tilde{+}$ is associative, commutative and nondecreasing in each variable, so it can be extended into an ∞-ary operator $\tilde{\sum}$; with this notation, (IP) merely means that $\bar{\rho} := \tilde{\sum}_{u \in \mathbb{Z}^d} \rho_u$.

Let $(\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{d-1})$ be a $\mathbb{Z}$-basis of $\mathbb{Z}^d$. For $0 \leqslant r \leqslant d$, we identify $\mathbb{Z}^r$ with $\mathbb{Z}\boldsymbol{e}_{d-r} \oplus \cdots \oplus \mathbb{Z}\boldsymbol{e}_{d-1}$; we also denote $\mathbb{Z}_r^{\perp} := \mathbb{Z}\boldsymbol{e}_0 \oplus \cdots \oplus \mathbb{Z}\boldsymbol{e}_{d-r-1}$. What we will prove is actually the following

Proposition 57. *For all* $r \in \{0, \ldots, d\}$, *for all* $x, y \in \mathbb{Z}_r^{\perp}$,

$$\rho^{\ddagger}(\vec{X}_{x+\mathbb{Z}^r}; \vec{Y}_{y+\mathbb{Z}^r}) \leqslant \tilde{\sum}\{\rho_u \mid u \in y - x + \mathbb{Z}^r\}.^{[\dagger\dagger]} \tag{IR}$$

◇

The statement of the corollary then corresponds to the proposition for $r = d$.

We prove Proposition 57 by induction on r. The case $r = 0$ holds by assumption. Now for $r > 0$ assume that Proposition 57 holds for $r - 1$; and take $x, y \in \mathbb{Z}_r^{\perp}$. We notice that

$$\vec{X}_{x+\mathbb{Z}^r} = (\vec{X}_{x+i\boldsymbol{e}_{d-r}+\mathbb{Z}^{r-1}})_{i\in\mathbb{Z}}, \tag{IS}$$

which we shorthand into "$\vec{X}_{x+\mathbb{Z}^r} = (\boldsymbol{X}_i)_{i\in\mathbb{Z}}$"; similarly we write, with obvious notation, $\vec{Y}_{y+\mathbb{Z}^r} =: (\boldsymbol{Y}_j)_{j\in\mathbb{Z}}$. By induction hypothesis one has for all $i, j \in \mathbb{Z}$ that

$$\rho^{\ddagger}(\boldsymbol{X}_i; \boldsymbol{Y}_j) \leqslant \tilde{\sum}\{\rho_u \mid u \in y - x + (j-i)\boldsymbol{e}_{d-r} + \mathbb{Z}^{r-1}\}. \tag{IT}$$

Since the right-hand side of (IT) only depends on $j - i$, we can apply Theorem 47[‡‡] to the $\boldsymbol{X}_i$'s and the $\boldsymbol{Y}_j$'s, which yields:

$$\begin{aligned}\rho^{\ddagger}(\vec{X}_{x+\mathbb{Z}^r}; \vec{Y}_{y+\mathbb{Z}^r}) &\leqslant \tilde{\sum_{z\in\mathbb{Z}}} \tilde{\sum}\{\rho_u \mid u \in y - x + z\boldsymbol{e}_{d-r} + \mathbb{Z}^{r-1}\} \\ &= \tilde{\sum}\{\rho_u \mid u \in y - x + \mathbb{Z}^r\},\end{aligned} \tag{IU}$$

i.e. (IR). ♤

2.6 Discussion on the tensorization theorems

In the previous sections we have established three main tensorization theorems, namely Theorems 37, 39 and 47 (plus Corollary 56). In this section we will discuss these results, by seeing what are actually the minimal assumptions required for them (in § 2.6.1), what happens if there are other variables involved than the X_i's and Y_j's (§ 2.6.2), and what can be said about optimality of our results (§ 2.6.3).

[††] In Equation (IR), the family of random variables w.r.t. which $\rho^{\ddagger}$-mixing is considered is made of $\vec{X}_{x+\mathbb{Z}^r}$, $\vec{Y}_{y+\mathbb{Z}^r}$, and all the X_i's and Y_j's for $i \notin x + \mathbb{Z}^r$, resp. $j \notin y + \mathbb{Z}^r$.

[‡‡] More precisely, here we use the version of Theorem 47 with third-party variables, cf. § 2.6.2.

2.6.1 Minimal assumptions

In our theorems we assumed $\rho^{\ddagger}$-mixing hypotheses. But a closer look at our proofs shows that only $\rho^{\dagger}$-mixing assumptions were actually needed. This is the following

Theorem 58 (Minimal assumptions for the tensorization theorems)**.**

(i) *In Theorem 37, if the set* $\boldsymbol{I}$ *is well-ordered,*[*] *"*$\rho^{\ddagger}(X_i;Y)$*" in the bound* (`DH`) *can be replaced by "*$\rho^{\dagger}_{\vec{X}_{<i}}(X_i;Y)$*". (Remember that '*$\vec{X}_{<i}$*' is a shorthand for* $\vec{X}_{\{i'\in I|i'<i\}}$*).*

(ii) *In Theorem 39, if the sets* $\boldsymbol{I}$ *and* $\boldsymbol{J}$ *are well-ordered, "*$\rho^{\ddagger}(X_i;Y_j)$*" in the bound* (`EB`) *can be replaced by "*$\rho^{\dagger}_{(\vec{X}_{<i},\vec{Y}_{<j})}(X_i;Y_j)$*".*

(iii) *In Theorem 47, provided the filtrations* $(\sigma(\vec{X}_{<i}))_{i\in\{-\infty\}\cup\mathbb{Z}}$ *and* $(\sigma(\vec{Y}_{<j}))_{j\in\{-\infty\}\cup\mathbb{Z}}$ *are almost-right-continuous, i.e. that* $\lim^{\searrow}_{i\to-\infty}\sigma(\vec{X}_{<i}), \lim^{\searrow}_{j\to-\infty}\sigma(\vec{Y}_{<j}) \overset{\text{a.s.}}{=} \mathscr{O}$*, it is possible to replace "*$\rho^{\ddagger}(X_i;Y_j)$*" in the assumption* (`GD`) *by "*$\rho^{\dagger}_{(\vec{X}_{<i},\vec{Y}_{<j})}(X_i;Y_j)$*".*

(iv) *In Corollary 56, denote* $\bar{\mathbb{Z}} \coloneqq \mathbb{Z}\cup\{-\infty\}$*, and endow* $\bar{\mathbb{Z}}^d$ *with the lexicographic ordering and the corresponding topology. Then, provided the filtrations* $(\sigma(\vec{X}_{<i}))_{i\in\bar{\mathbb{Z}}^d}$ *and* $(\sigma(\vec{Y}_{<j}))_{j\in\bar{\mathbb{Z}}^d}$ *are almost-right-continuous, "*$\rho^{\ddagger}(X_i;Y_j)$*" in the assumption* (`IN`) *can be replaced by "*$\rho^{\dagger}_{(\vec{X}_{<i},\vec{Y}_{<j})}(X_i;Y_j)$*".* ◇

Remark 59. The "minimal" hypotheses for Theorem 47 and Corollary 56 are not *technically* weaker than the "standard", original ones, as there are extra assumptions of filtration continuity… However, Theorem 47 can nevertheless be deduced from Theorem 58-(iii) by the following argument. Assume $\vec{X}_{\mathbb{Z}}$ and $\vec{Y}_{\mathbb{Z}}$ satisfy the standard hypotheses of Theorem 47; and, for arbitrarily large N, modify $\vec{X}$ and $\vec{Y}$ into respective versions $\vec{X}'$ and $\vec{Y}'$ which coincide with resp. $\vec{X}$ and $\vec{Y}$ on $[\![-N,\infty[\![$, but are such X'_i, $Y'_j :\equiv \partial$ for all $i,j < -N$ (∂ denoting some cemetery value). Then $\vec{X}'$ and $\vec{Y}'$ <u>do</u> satisfy the assumptions of Theorem 58-(iii): so the "minimal" theorem suffices to obtain that $\rho(\vec{X}';\vec{Y}') \leqslant \left\|\!\left|\left(\left(\rho^{\ddagger}(X_i;Y_j)\right)\right)_{i,j}\right|\!\right\|$. Therefore, as $\rho(\vec{X}';\vec{Y}')$ is arbitrarily close to $\rho(\vec{X};\vec{Y})$ provided N was taken large enough, we recover the conclusion of Theorem 47. A similar trick holds to deduce Corollary 56 from Theorem 58-(iv). ♧

Let us prove successively all four items of Theorem 58. As regards Theorem 58-(i), like in § 2.3, we will provide two proofs: first, one that mimics the first proof of Theorem (37) (i.e., the one starting on page 27), and then one that mimics the "alternative proof" starting on page 28. Like in § 2.3, the reason for giving two different proofs is that, later, the proofs of Theorems 58-(ii) and 58-(iii) shall involve reasonings coming from both proofs of Theorem 58-(i).

---

[*] This includes in particular the cases where I is finite and totally ordered, or isomorphic to $\mathbb{N}$, or isomorphic to some $\mathbb{N}^d$ endowed with lexicographic ordering.

First proof of Theorem 58-(i). In the case I is finite, the result just comes from a careful look at the (first) proof of Theorem 37. By, the way, note that the "finite I" case includes in particular the case $|I| = 2$.

In the case I is infinite, however, some technicalities do occur. What we will do is to prove the result by induction on the *initial segments* of I: what I mean by "initial segment of I" being a subset of the form $I_{<i_\star} := \{i \in I \mid i < i_\star\}$ for some $i_\star \in I \sqcup \{\infty\}$. (Here '$\infty$' stands for an extra element being strictly larger than any value of I, so that $I = I_{<\infty}$: thus, I can be considered as an initial segment of itself). Let us use notation 'I'' to generically denote an initial segment of I. As the set of the I''s is well-ordered (for inclusion), a reasoning by induction can be carried:

- The initialization case (corresponding to $I' = \varnothing$) is trivial.

- The successor case, i.e. the case where I' is of the form $I'' \sqcup \{i_1\}$ for i_1 the successor of initial segment I'', is obtained by applying the "$|I| = 2$" version of the result to 'I' $\leftarrow \{0, 1\}$, 'X_0' $\leftarrow \vec{X}_{I''}$, 'X_1' $\leftarrow X_{i_1}$ and 'Y' $\leftarrow Y$: assumptions of Theorem 58-(i) for this case being ensured by induction hypothesis.

- Finally, let us check the limit case, i.e. the case where I' appears as a union of different initial segments of I (all being strictly smaller than I' itself), for each of which the induction hypothesis has been checked. In this case, we just observe that, for I' any set, a variable $f \in \mathrm{L}^2(\vec{X}_{I'})$ can always approximated as close as one wants by a variable $f' \in \mathrm{L}^2(\vec{X}_{\tilde{I}'})$ for some finite $\tilde{I}' \subseteq I'$: therefore,

  $$\rho(\vec{X}_{I'}; Y) = \sup_{\substack{\tilde{I}' \subseteq I' \\ \tilde{I}' \text{ finite}}} \rho(\vec{X}_{\tilde{I}'}; Y). \tag{IV}$$

  Because of that, in the limit case, $\rho(\vec{X}_{I'}; Y)$ is at most equal to (and hence equal to) the supremum of the $\rho(\vec{X}_{I''}; Y)$'s for I'' initial segments of I being strictly shorter than I': but that supremum is, by induction hypothesis, equal to

  $$\sup_{I'' < I'} \Big(1 - \prod_{i \in I''} \big(1 - \rho^{\ddagger}(X_i; Y)^2\big)\Big)^{1/2}: \tag{IW}$$

  and, since the I'''s are totally ordered for inclusion and that their union is equal to I', this is actually equal to $\big(1 - \prod_{i \in I'} \big(1 - \rho^{\ddagger}(X_i; Y)^2\big)\big)^{1/2}$, as we wanted. ♤

Alternative Proof of Theorem 58-(i). The principle of this proof consists in following exactly the structure of the alternative proof of Theorem 37, except that I will no longer be of the form $\{0, \dots, N-1\}$. In this new context, note that the σ-algebra '$\mathscr{F}_{i+1}$' has to be re-interpreted as $\sigma(\vec{X}_{\leqslant i})$. Apart from that detail, the definitions of the f_i's, g^i's and g_i's remain unchanged, and the relations (DW)–(DY) remain perfectly valid—note in passing that here 'ρ_i' should be interpreted as $\rho_{\mathscr{F}_i}(X_i; Y)$, as we are working under minimal assumptions: but this does not cause any issues, since the proof does not require anything more that that version of 'ρ_i'.

What is <u>not</u> straightforward now is to ensure that one still has the relations $\mathrm{Var}(f) = \sum_i \mathrm{Var}(f_i)$, resp. $\mathrm{Var}^{1/2}(g^i) \leqslant \prod_{i'<i} \sqrt{1-\check{\rho}_{i'}^2}\, \mathrm{Var}^{1/2}(g)$, resp. $\mathbb{E}(fg) = \sum_i \mathbb{E}(f_i g)$[†].

- As regards the claim on $\mathrm{Var}(f)$, we can get it by proving by induction that for any $\boldsymbol{i} \in \boldsymbol{I} \sqcup \{\infty\}$, one has $\mathrm{Var}(f^{\mathscr{F}_i}) = \sum_{i'<i} \mathrm{Var}(f_{i'})$: it is the same kind of inductive reasoning as we did in the previous proof. By the way, note that a corollary of that fact is that only a countable number of the f_i's can be nonzero in $\mathrm{L}^2(\mathscr{B})$: hence there are only a countable number of $\mathbb{E}(f_i g)$'s which are nonzero, and also only a countable number of $\check{\rho}_i$'s which are nonzero.

- A similar induction works for the claim on the g^i's, except that this time is it not completely obvious that, for $\boldsymbol{i}$ a limit ordinal, $\mathrm{Var}(g^i)$ is actually the decreasing limit of the $\mathrm{Var}(g^{i'})$'s for $i' < i$: the reason why this is true is that $\mathrm{Var}(g^{\mathscr{F}_i})$ <u>is</u> the increasing limit of the $\mathrm{Var}(g^{\mathscr{F}_{i'}})$'s (by the same approximation argument as in the previous proof), and that one has, by orthogonality, $\mathrm{Var}(g^i) = \mathrm{Var}(g) - \mathrm{Var}(g^{\mathscr{F}_i})$, resp. $\mathrm{Var}(g^{i'}) = \mathrm{Var}(g) - \mathrm{Var}(g^{\mathscr{F}_{i'}})$.

- Then one proves by induction that $\mathbb{E}(f^{\mathscr{F}_i} g) = \sum_{i'<i} \mathbb{E}(f_{i'} g)$ for all $\boldsymbol{i} \in \boldsymbol{I} \sqcup \{\infty\}$. Only the limit cases of the induction are new compared to the proof of page 28. But for these, you observe that $f^{\mathscr{F}_{i'}}$ converges to $f^{\mathscr{F}_i}$ in $\mathrm{L}^2(\mathscr{B})$, and that $\sum_{i'<i} |\mathbb{E}(f_{i'} g)|$ is actually absolutely convergent by the Cauchy–Schwarz inequality (using the fact that $\sum_{i'<i} \mathrm{Var}(f_{i'}) = \mathrm{Var}(f^{\mathscr{F}_i})$, which we proved above): so one can pass the property to the limit.

Having ensured that all the summations remain meaningful and valid in this well-ordered context, we can then conclude in the same way as for the proof of page 28. ♤

Proof of Theorem 58-(ii). Again the idea is to try to follow the original proof of Theorem 39, but taking care of the technicalities due to $\boldsymbol{I}$ and $\boldsymbol{J}$ not being finite.

First of all, thanks to the well-ordered structure of $\boldsymbol{I}$ and $\boldsymbol{J}$, by induction the chain rule for variance remains valid in its infinite-size version: therefore, the core formulas stated in the proof of Theorem 39—in particular Equations (`FD`) and (`FH`)—remain valid here.

Another important point is that, for $\boldsymbol{i}$ a limit ordinal, one has $\boldsymbol{W}^i = \lim^{\searrow}_{i'\nearrow i} \boldsymbol{W}^{i'}$, and also $\lim^{\searrow}_{i'\nearrow i} \mathrm{Var}(g - g^{\mathscr{F}_{i'} \vee \mathscr{G}_j}) = \mathrm{Var}(g - g^{\mathscr{F}_i \vee \mathscr{G}_j})$, hence $\boldsymbol{W}_j^i = \lim_{i'\nearrow i} \boldsymbol{W}_j^{i'}$ for all j: checking these properties being done along the same lines as we did regarding the g^i's in the second proof of Theorem 58-(i). Thanks to these regularity properties, the use of Lemma 44 remains valid in the context where $\boldsymbol{I}$ is infinite (as well as $\boldsymbol{J}$).

Now it remains to show that the use of Lemma 46 can be adapted to the case where $\boldsymbol{I}$ and $\boldsymbol{J}$ are infinite: this is the most technical point of the present proof.

A first step consists in observing that it is always possible to restrict $\boldsymbol{I}$ and $\boldsymbol{J}$ to (finite or) countable subsets. As regards the operator norm of “matrix” $\boldsymbol{\rho}$, such

---

[†] Actually, at this point nothing guarantees that this sum is even meaningful, as it might lack convergence!

a restriction is harmless: indeed, denoting by $\boldsymbol{\rho}_{I'\times J'}$ the restriction of $\boldsymbol{\rho}$ to indices in $I' \times J'$, one has always that $|||\boldsymbol{\rho}_{I'\times J'}||| \leqslant |||\boldsymbol{\rho}|||$ (confer canonical embedding of $\mathrm{L}^2(I') \otimes \mathrm{L}^2(J')$ in $\mathrm{L}^2(I) \otimes \mathrm{L}^2(J)$), while a classical density argument conversely shows that $|||\boldsymbol{\rho}|||$ is not larger than the supremum of the $|||\boldsymbol{\rho}_{I'\times J'}|||$'s for I' and J' finite: therefore,

$$|||\boldsymbol{\rho}||| = \sup\{|||\boldsymbol{\rho}_{I',J'}||| \mid I', J' \text{ finite}\} = \sup\{|||\boldsymbol{\rho}_{I',J'}||| \mid |I'|, |J'| \leqslant \aleph_0\}. \tag{IX}$$

Moreover, for whichever variables $f \in \mathrm{L}^2(\vec{X})$ and $g \in \mathrm{L}^2(\vec{Y})$ that we are considering, only a countable number of indices i and j will play a role in the (generalized) proof of Theorem 39! Indeed, as regards the set I, there can only be a countable numbers of i's for which $\mathrm{Var}(g^{i+1})$[‡] is strictly less than $\mathrm{Var}(g^i)$; and for indices i such that $\mathrm{Var}(g^{i+1}) = \mathrm{Var}(g^i)$, one has actually $g^{i+1} = g^i$ (in $\mathrm{L}^2(\mathscr{B})$) and then everything for that i becomes trivial: so we can discard these i's and only consider a countable subset of I. But also, for any i, there can only be a countable number of j's such that $\mathrm{Var}(g^{\mathscr{F}_i \vee \mathscr{G}_{j+1}})$ is strictly larger than $\mathrm{Var}(g^{\mathscr{F}_i \vee \mathscr{G}_j})$: for all the other j's, one will have $g^i_j \not\equiv 0$ in $\mathrm{L}^2(\mathscr{B})$. And, since what happens for the j's such that $g^i_j \equiv 0$ for all i is trivial, we can finally restrict J to the set of j's such that $g^i_j \not\equiv 0$ for at least one i, which set is (at most) countable (once having made I countable).

Now that we have ensured that I and J may always be taken countable, the very final step of the proof will consist in checking that Lemma 46 can be extended to the case of "square matrices" of size $\aleph_0$: this is done by Lemma 162 in appendix. Actually in Lemma 162 I had to add the assumptions that the matrix "$\boldsymbol{A}$" is symmetric and bounded, but these extra assumptions are satisfied in our case: indeed, Lemma 162 is to be applied to the symmetric "matrix" $\boldsymbol{\rho}^*\boldsymbol{\rho}$, which can harmlessly be assumed to be bounded by 1, since $|||\boldsymbol{\rho}^*\boldsymbol{\rho}||| = |||\boldsymbol{\rho}|||^2$ and that the conclusion of Theorem 58-(ii) is automatic unless one has $|||\boldsymbol{\rho}||| < 1$. ♤

Proof of Theorem 58-(iii). Again, the idea will be to follow the proof of Theorem 47, but without the initial approximation step. Because we cannot limit ourselves any more to the case where only a finite number of indices i and j are involved, many parts of the proof shall be adapted in a more technical way: for instance, the compactness condition on $(\mathbb{Z}\times\mathbb{Z})$-indexed arrays will have to be replaced by summability and convergence conditions (which are indeed satisfied in the situation that we are considering), which are somehow trickier to handle... But, fundamentally, nothing of that changes the proof.

On the other hand, it is important to stress is that, if we had not imposed the regularity condition on σ-algebras, then the summation formulas used in the proof of Theorem 47 would not have remained valid! For instance, look at Equation (`GO`). In absence of any regularity condition for σ-algebras, it would still remain true that, for indices i_-, i_+ satisfying $i_- \leqslant i_+$, one would have $\mathrm{Var}(f^{\mathscr{F}_{i_+}} - f^{\mathscr{F}_{i_-}}) = \sum_{i=i_-}^{i_+} \mathrm{Var}(f_i)$. Then, it would seem natural to let i_+ tend to $+\infty$, resp. i_- tend to $-\infty$, to get that "$\mathrm{Var}(f) = \sum_{i\in\mathbb{Z}} \mathrm{Var}(f_i)$". But that would not be correct... Though it is true indeed that $f^{\mathscr{F}_{i_+}} \to f^{\mathscr{F}} = f$ as $i_+ \to \infty$, when $i_- \to -\infty$ however, despite the fact that the

---

[‡] Where, in this context, "g^{i+1}" actually has to be understood as $g - \mathbb{E}(g \mid \sigma(\vec{X}_{\leqslant i}))$.

set of indices $\{i \mid i \leqslant i_-\}$ shrinks down to $\varnothing$, there is no reason why $f^{\mathscr{F}_{i_-}}$ would tend to $f^{\mathscr{O}} \underset{\text{def}}{\equiv} \mathbb{E}(f)$![§] Instead, what we would actually get by passing to the limit is just that

$$\operatorname{Var}(f) = \operatorname{Var}(f^{\lim^{\searrow}_{i\to-\infty}\mathscr{F}_i}) + \sum_{i\in\mathbb{Z}} \operatorname{Var}(f_i), \tag{IY}$$

with a spurious term (in comparison with Equation (GO)) which is very annoying, since it somehow does not depend on any i: so, how would we control it…?!

A similar phenomenon would hold for Equation (GT): then, one would only get that

$$g^i = (g^{\lim^{\searrow}_{j\to-\infty}(\mathscr{F}_i\vee\mathscr{G}_j)} - g^{\mathscr{F}_i}) + \sum_{j\in\mathbb{Z}} g^i_j, \tag{IZ}$$

with likewise exhibits a spurious term involving no j in particular. This is the reason why I had to add regularity assumptions in the statement of Theorem 58-(iii): with these assumptions indeed, the spurious terms become zero, and then everything gets fine! ♤

Proof of Theorem 58-(iv). Theorem 58-(iv) gets deduced from Theorem 58-(iii) exactly the same way as we deduced Corollary 56 from Theorem 47. The only subtlety is that, in practice, Theorem 58-(iii) will have to be applied under some *conditional* probabilities $\mathbb{P}(\bullet \mid \vec{X}_{(<x)+\mathbb{Z}^r} = \vec{\xi}_{(<x)+\mathbb{Z}^r}, \vec{Y}_{(<y)+\mathbb{Z}^r} = \vec{\psi}_{(<y)+\mathbb{Z}^r})$, where "$(< x)$" denotes the set of $x' \in \mathbb{Z}_r^{\perp}$ preceding x for the lexicographic ordering, and likewise for "$(< y)$". But the "minimal assumptions" that we required when stating Theorem 58-(iv) precisely ensure that the assumptions of Theorem 58-(iii) will be satisfied under these conditional laws! ♤

2.6.2 Third-party variables

In practice, besides the variables $X_{i\in I}$ and $Y_{j\in J}$ one often has "third-party" variables $Z_{k\in K}$. Then we would like to get a *$\rho^\ddagger$-mixing result* (w.r.t. the Z_k's) between $\vec{X}_I$ and $\vec{Y}_J$ rather than just a ρ-mixing result. It is easy to check that this can indeed be done, provided the $\rho^\ddagger$-mixing assumptions between the X_i's and the Y_j's also take the Z_k's into account:

Theorem 60 (Tensorization theorems with third-party variables)**.** *In Theorems 37, 39, 47 and 56, assume that besides the variables $X_{i\in I}$ and $Y_{j\in J}$ one also has variables $Z_{k\in K}$, and that $\rho^\ddagger$-mixing coefficients between X_i and Y_i are considered w.r.t. the family of random variables $(Z_k)_{k\in\bar{K}}$, where one sets $\bar{K} \coloneqq K \sqcup I \sqcup J$ and $Z_i \coloneqq X_i$ for $i \in I$, resp. $Z_j \coloneqq Y_j$ for $j \in J$. Then $\mathrm{\rho}(\vec{X}_I; \vec{Y}_J)$ in the respective bounds* (DH), (EB) *and* (GE) *can be replaced by $\mathrm{\rho}^\ddagger(\vec{X}_I; \vec{Y}_J)$, where the family of random variables w.r.t. which $\rho^\ddagger$-mixing is considered is again $(Z_k)_{k\in\bar{K}}$.*[¶] ◇

[§] A very simple counterexample would consist in considering a nonconstant random variable $\bar{X} \in \mathrm{L}^2(\mathscr{B})$, and then to take $X_i \coloneqq \bar{X}$ for all i, resp. $f \coloneqq \bar{X}$.

[¶] In this context, the most natural family of random variables w.r.t. which $\rho^\ddagger$-mixing between $\vec{X}_I$ and $\vec{Y}_J$ would be considered should rather be $(Z_k)_{k\in K}$; however, stating the result w.r.t. the full family $(Z_k)_{k\in\bar{K}}$ is obviously stronger and can also be useful.

Proof. We only write down the proof in the case of Theorem 39, the cases of Theorems 37, 47 and 56 being exactly alike. Denote by ρ_{ij} the mixing coefficient $\rho^{\ddagger}_{(Z_k)_{k\in\bar{K}}}(X_i;Y_j)$ *under the reference law* $\mathbb{P}$. Proving that $\rho^{\ddagger}(\vec{X}_I;\vec{Y}_J)\leqslant |||\boldsymbol{\rho}||| \coloneqq |||\langle\!\langle\rho_{ij}\rangle\!\rangle_{i,j}|||$ means that for all $\boldsymbol{K'}\subseteq\bar{\boldsymbol{K}}$, for $\mathbb{P}$-almost-all $\vec{z}_{K'}$, the ρ-mixing coefficient between $\vec{X}_I$ and $\vec{Y}_J$ *under the law* $\mathbb{P}(\bullet\mid\vec{Z}_{K'}=\vec{z}_{K'})$ is bounded by $|||\boldsymbol{\rho}|||$. But, according to Theorem 39 itself, this will be the case provided that one has for all $\boldsymbol{i},\boldsymbol{j}$ that $\rho^{\ddagger}_{(Z_k)_{k\in I\sqcup J}}(X_i;Y_j)\leqslant\rho_{ij}$ *under the law* $\mathbb{P}(\bullet\mid\vec{Z}_{K'}=\vec{z}_{K'})$.

So, consider $\boldsymbol{K'}\subseteq\boldsymbol{K}$ and a $\boldsymbol{K'}$-uple of values $\vec{z}_{K'}$ for the $\boldsymbol{K'}$-uple of variables $\vec{Z}_{K'}$, and let us reason under the conditional law $\mathbb{P}(\bullet\mid\vec{Z}_{K'}=\vec{z}_{K'})$. What does it mean then that $\rho^{\ddagger}_{(Z_k)_{k\in I\sqcup J}}(X_i;Y_j)\leqslant\rho_{ij}$? Well, it means that for all $\boldsymbol{K''}\subseteq\boldsymbol{I}\sqcup\boldsymbol{J}$, for $\mathbb{P}(\bullet\mid\vec{Z}_{K'}=\vec{z}_{K'})$-almost-all $\vec{\zeta}_{K''}$, one has $\rho(X_i;Y_j)\leqslant\rho_{ij}$ *under the law* $\mathbb{P}(\bullet\mid\vec{Z}_{K''}=\vec{\zeta}_{K''})$. Note that, since we are reasoning conditionally to the fact that $\vec{Z}_{K'}=\vec{z}_{K'}$, the law "$\mathbb{P}(\bullet\mid\vec{Z}_{K''}=\vec{\zeta}_{K''})$" is actually, in unconditioned terms, the law $\mathbb{P}(\bullet\mid\vec{Z}_{K'}=\vec{z}_{K'}$ and $\vec{Z}_{K''}=\vec{\zeta}_{K''})$. But $\mathbb{P}(\bullet\mid\vec{Z}_{K'}=\vec{z}_{K'})$-almost-every $\vec{\zeta}_{K''}$ will be such that $\vec{\zeta}_{K'\cap K''}=\vec{z}_{K'\cap K''}$; therefore the condition "$\vec{Z}_{K'}=\vec{z}_{K'}$ and $\vec{Z}_{K''}=\vec{\zeta}_{K''}$" can be simplified into "$\vec{Z}_{K'\cup K''}=\vec{z}_{K'\cup K''}$", where one sets $z_k\coloneqq\zeta_k$ for $k\in K''\smallsetminus K'$. And the fact that $\rho(X_i;Y_j)\leqslant\rho_{ij}$ under the law $\mathbb{P}(\bullet\mid\vec{Z}_{K'\cup K''}=\vec{z}_{K'\cup K''})$ for $\mathbb{P}$-almost-all $\vec{z}_{K'\cup K''}$ precisely comes from the definition of ρ_{ij} as a $\rho^{\ddagger}$-mixing coefficient w.r.t. the variables $(Z_k)_{k\in\bar{K}}$. ♤

Remark 61. Obviously the improvements of this subsection can also be combined with those of the previous subsection on minimal assumptions: then, one gets results to bound a given $\rho^{\dagger}$-mixing coefficient between $\vec{X}_I$ and $\vec{Y}_J$ w.r.t. some variables $(Z_k)_{k\in K}$ by using only certain $\rho^{\dagger}$-mixing coefficients between the X_i's and the Y_j's. For the sake of lightness, we will not write down these results here, as their wordings are hopefully obvious to the reader. ♧

2.6.3 Optimality results

In this subsection we are going to show that the bounds given by our tensorization theorems are morally the best that one could expect. Let us begin with the following result, which states that the bound of Theorem 39 is essentially optimal at first order:

Theorem 62 (First-order optimality of Theorem 39). *Let* $\boldsymbol{I}$ *and* $\boldsymbol{J}$ *be finite sets. For all* $\boldsymbol{\rho}=\langle\!\langle\rho_{ij}\rangle\!\rangle_{i,j}\in\mathbb{R}_+^{I\times J}$*, denote*

$$\mathrm{Opt}^{\mathrm{d/s}}(\boldsymbol{\rho})\coloneqq\sup\{\rho(\vec{X}_I;\vec{Y}_J)\mid\forall i,j\ \ \rho^{\ddagger}(X_i;Y_j)\leqslant\rho_{ij}\},^{[|||]}\tag{JB}$$

where the supremum is taken over all the possible sets of variables X_i *and* Y_j *on all the possible probability spaces. This theorem claims that* $\mathrm{Opt}^{\mathrm{d/s}}(\bullet)$ *is equivalent to* $|||\bullet|||$ *when* $\boldsymbol{\rho}\to\boldsymbol{0}\coloneqq\vec{0}_{I\times J}$*, at least as long as* $\boldsymbol{\rho}$ *is no too close to the boundary of*

[|||] The superscript 'd/s' is for "double tensorization / standard assumptions".

$\mathbb{R}_+^{I\times J}$*: more precisely, equipping* $\mathbb{R}_+^{I\times J}$ *with its usual topology, we claim that for every closed cone* $\mathscr{C} \subseteq \mathbb{R}_+^{I\times J}$ *such that* $\mathscr{C} \setminus \{\mathbf{0}\} \subseteq (\mathbb{R}_+^*)^{I\times J}$*, one has on* $\mathscr{C}$ *that*

$$\mathrm{Opt}^{\mathrm{d/s}}(\boldsymbol{\rho}) \overset{\boldsymbol{\rho}\to\mathbf{0}}{\sim} |||\boldsymbol{\rho}|||, \tag{JC}$$

so that it this regime the bound of Theorem 39 cannot be improved at first order. ◇

Remark 63. The fact that we need a restriction to $\mathscr{C}$ comes from the fact that it may have deep consequences to impose that some, but not all, of the $\rho^\ddagger$-mixing coefficients are extremely low: in particular, it is not possible in general to build a Gaussian example in which only one given $\rho^\ddagger$-mixing coefficient would be zero while all the other ones would be nonzero… On the other hand, we do not have this problem if we consider just $\rho^\dagger$-mixing coefficients: as a consequence, as Theorem 65 will show a few pages below, the optimal bound *for minimal assumptions* (cf. Theorem 58-(ii)) $\mathrm{Opt}^{\mathrm{d/m}}(\bullet)$ is such that $\mathrm{Opt}^{\mathrm{d/m}}(\boldsymbol{\rho}) \overset{\boldsymbol{\rho}\to\mathbf{0}}{\sim} |||\boldsymbol{\rho}|||$ *on the whole* $\mathbb{R}_+^{I\times J}$. It is worth noticing in passing that comparison of Theorem 62 with Theorem 65 implies that, apart from the case where a few mixing coefficients are extremely low compared to the other ones, the optimal bound is the same, at first order, whether one considers the standard assumptions or the minimal assumptions: this suggests that in practice, using the standard assumptions instead of the minimal ones might be of little importance. ♧

Remark 64. In the case where $\boldsymbol{J}$ is a singleton, resp. where $\boldsymbol{I}, \boldsymbol{J} \cong \mathbb{Z}$ (or $\mathbb{Z}^d$) and that the bound on $\rho^\ddagger(X_i; Y_j)$ does only depend on $j - i$, the bound given by Theorem 39 is equivalent at first order to the bound given by Theorem 37, resp. to the bound given by Theorem 47 (or Corollary 56). Therefore, one has first-order optimality a fortiori for these theorems too. But actually in these cases there is even *exact* optimality, cf. Theorems 66 and 67 below. ♧

Proof of Theorem 62. On the space of matrices $\mathbb{R}^{I\times J}$, let us denote the ℓ^∞ norm by $\|\boldsymbol{\rho}\| \coloneqq \sup_{i,j}|\rho_{ij}|$. Fix $\varepsilon > 0$ small enough (in a sense to be made precise later) and consider $\boldsymbol{\rho} \in \mathbb{R}^{I\times J}$ such that $\|\boldsymbol{\rho}\| \leqslant \varepsilon$. Let $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ be real random variables such that $(\vec{X}_I, \vec{Y}_J)$ is a $(|I|+|J|)$-dimensional Gaussian vector whose covariance matrix is given by

$$\begin{aligned}\mathrm{Cov}(X_i, X_{i'}) &= \delta_{ii'},^{[**]}\\ \mathrm{Cov}(X_i, Y_j) &= \rho_{ij},\\ \mathrm{Cov}(Y_j, Y_{j'}) &= \delta_{jj'}:\end{aligned} \tag{JD}$$

this defines indeed a covariance matrix provided ε was chosen $\leqslant 1/(|I| \vee |J|) \eqqcolon c_1$. (Along this proof, we will denote by '$c_\bullet$' or '$C_\bullet$' explicit constants only depending on $|I|$ and $|J|$, but whose exact expression we do not want to bother with. We will use notation '$c_\bullet$' for constants which have to be strictly positive, resp. '$C_\bullet$' for constants which have to be finite).

Now we would like to compute $\rho^\ddagger(X_i; Y_j)$. Let us denote $K \coloneqq I \sqcup J$, and for $k \in K$, set $Z_k \coloneqq X_k$ if $k \in I$, resp. $Z_k \coloneqq Y_k$ if $k \in J$; then $\rho^\ddagger(X_i; Y_j)$ is by definition the supremum of the values $\rho(X_i; Y_j)$ under the laws $\mathbb{P}(\bullet \mid \vec{Z}_{K'} = \vec{z}_{K'})$, when K'

[**] Here 'δ' is Kronecker's delta.

describes all the possible subsets of K and $\vec{z}_{K'}$ describes (almost-)all the possible values of $\vec{Z}_{K'}$. We can restrict ourselves to the case where $i, j \notin K'$, since otherwise the conditional correlation is obviously 0. So, consider $K' \subseteq K \smallsetminus \{i, j\}$ and $\vec{z}_{K'} \in \mathbb{R}^{K'}$. Since $\vec{Z}_K$ is a Gaussian random vector, we know that under the conditional law $\mathbb{P}(\bullet \mid \vec{Z}_{K'} = \vec{z}_{K'})$, $\vec{Z}_K$ shall still be a (non-centred in general) Gaussian vector. The conditional covariance matrix of $\vec{Z}_K$ can be computed in the following way: denoting $K_0 \coloneqq K'$, $K_1 \coloneqq K \smallsetminus K'$, and re-ordering K as K_0 followed by K_1, say that the (unconditioned) covariance matrix of Z_K writes blockwise

$$\begin{pmatrix} V_0 & \Gamma \\ \Gamma^\top & V_1 \end{pmatrix}; \tag{JE}$$

then, the conditional covariance matrix of $\vec{Z}_{K_1}$ is $V_1 - \Gamma^\top V_0^{-1} \Gamma$, regardless of the value of $\vec{z}_{K'}$. (Here we assume that V_0 is invertible, which is indeed the case provided ε was chosen $< c_1$). So, by Proposition 9 on Gaussian variables, the ρ-mixing coefficient between X_i and Y_j under the conditional law is given by the absolute value of their Pearson correlation coefficient under that law, which can be read on the (conditional) covariance matrix of $\vec{Z}_{K_1}$.

Provided ε was chosen $\leqslant c_2 \coloneqq c_1 / 2$, one has $\|V_0^{-1}\| \leqslant (1 - c_2 / c_1)^{-1} \eqqcolon C_3$. Since moreover one has $\|\Gamma\| \leqslant \varepsilon$, it follows that $\|\Gamma^\top V_0^{-1} \Gamma\| \leqslant C_4 \varepsilon^2$ (with $C_4 \coloneqq (|I| + |J| - 2)^2 C_3$), and thus the covariance matrix of $\vec{Z}_{K_1}$ under $\mathbb{P}(\bullet \mid \vec{Z}_{K'} = \vec{z}_{K'})$ is within $C_4 \varepsilon^2$ of the covariance matrix of $\vec{Z}_{K_1}$ under the unconditioned law. Therefore, under the conditional law, and provided ε was chosen $\leqslant c_5 \coloneqq \frac{1}{2} C_4^{-1/2}$, the Pearson correlation coefficient between X_i and Y_j is at most, in absolute value,

$$\frac{\rho_{ij} + C_4 \varepsilon^2}{\sqrt{1 - C_4 \varepsilon^2}\,\sqrt{1 - C_4 \varepsilon^2}} \leqslant \rho_{ij} + C_6 \varepsilon^2 \tag{JF}$$

(with $C_6 \coloneqq (1 - C_4 c_5^2)^{-1} (1 + c_5) C_4$). Since all of that is valid independently from the choice of K', we deduce that

$$\rho^\ddagger(X_i; Y_j) \leqslant \rho_{ij} + C_6 \varepsilon^2. \tag{JG}$$

On the other hand, let us find a lower bound on $\rho(\vec{X}_I; \vec{Y}_J)$. Let $\vec{a} \eqqcolon (a_i)_i \in \mathbb{R}^I \smallsetminus \{\vec{0}_I\}$, resp. $\vec{b} \eqqcolon (b_j)_j \in \mathbb{R}^J \smallsetminus \{\vec{0}_J\}$; we define

$$f \coloneqq \sum_{i \in I} a_i X_i; \tag{JH}$$

$$g \coloneqq \sum_{j \in J} b_j Y_j, \tag{JI}$$

which are resp. $\vec{X}_I$-measurable and $\vec{Y}_J$-measurable real random variables. Using the covariance matrix of $(\vec{X}_I, \vec{Y}_J)$, one computes that

$$\mathrm{Var}(f) = \sum_{i\in I} a_i^2 = \|\vec{a}\|^2_{\ell^2(I)}; \tag{JJ}$$

$$\mathrm{Var}(g) = \sum_{j\in J} b_j^2 = \|\vec{b}\|^2_{\ell^2(J)}; \tag{JK}$$

$$\mathrm{Cov}(f,g) = \sum_{i,j} a_i b_j \rho_{ij} = \vec{a}^{\mathsf{T}} \boldsymbol{\rho} \vec{b}, \tag{JL}$$

so that

$$\rho(\vec{X}_I; \vec{Y}_J) \geqslant \frac{|\vec{a}^{\mathsf{T}} \boldsymbol{\rho} \vec{b}|}{\|\vec{a}\|_{\ell^2(I)} \|\vec{b}\|_{\ell^2(J)}}, \tag{JM}$$

where the supremum over $\vec{a}$ and $\vec{b}$ of the right-hand side is by definition $|||\boldsymbol{\rho}|||_{\ell^2(J)\to\ell^2(I)}$.

So, at this point, we have proved that for $\|\boldsymbol{\rho}\| \leqslant c_7 \coloneqq c_2 \wedge c_5$, denoting by $\mathbf{J}$ the $|I|$-by-$|J|$ matrix with all entries equal to 1, one has

$$\mathrm{Opt}^{\mathrm{d/s}}(\boldsymbol{\rho} + C_6 \|\boldsymbol{\rho}\|^2 \mathbf{J}) \geqslant |||\boldsymbol{\rho}|||. \tag{JN}$$

Now let $\mathscr{C}$ be a cone such as in the statement of the proposition. Then it is an easy geometric reasoning to show that there exists $\delta > 0$ such that, for all $\boldsymbol{\rho} \in \mathscr{C}$, one has $\rho_{ij} \geqslant \delta \|\boldsymbol{\rho}\|$ for all i, j. Therefore, for $\|\boldsymbol{\rho}\| < c_7 \wedge \delta C_6^{-1}$, if we define $\boldsymbol{\rho}' \coloneqq (1 - \delta^{-1} C_6 \|\boldsymbol{\rho}\|)\boldsymbol{\rho}$, we have that $\boldsymbol{\rho}' + C_6 \|\boldsymbol{\rho}'\|^2 \mathbf{J}$ is bounded entry-wise by $\boldsymbol{\rho}$, and hence:

$$\mathrm{Opt}^{\mathrm{d/s}}(\boldsymbol{\rho}) \geqslant |||\boldsymbol{\rho}'||| = (1 - \delta^{-1} C_6 \|\boldsymbol{\rho}\|) |||\boldsymbol{\rho}||| \overset{\boldsymbol{\rho}\to\mathbf{0}}{\sim} |||\boldsymbol{\rho}|||. \tag{JO}$$

Since on the other hand $\mathrm{Opt}^{\mathrm{d/s}}(\boldsymbol{\rho}) \leqslant |||\boldsymbol{\rho}|||$ by Theorem 39 itself, this shows that $\mathrm{Opt}^{\mathrm{d/s}}(\boldsymbol{\rho})$ is indeed equivalent to $|||\boldsymbol{\rho}|||$ when $\boldsymbol{\rho} \to \mathbf{0}$ on $\mathscr{C}$. ♤

Now we are going to show that, provided we are considering the statements with minimal assumptions (cf. § 2.6.1), the bound of Theorem 39 is optimal at first order on the whole $\mathbb{R}_+^{I\times J}$, and more importantly, the bounds of Theorems 37 and 47 are perfectly optimal:

Theorem 65 (First-order optimality of the minimal-assumption version of Theorem 39)**.** *Let I and J be finite (totally ordered) sets. For all $\boldsymbol{\rho} = (\!(\rho_{ij})\!)_{i,j} \in \mathbb{R}_+^{I\times J}$, denote*

$$\mathrm{Opt}^{\mathrm{d/m}}(\boldsymbol{\rho}) \coloneqq \sup\{\rho(\vec{X}_I; \vec{Y}_J) \mid \forall i,j\ \rho^{\dagger}_{(\vec{X}_{<i}, \vec{Y}_{<j})}(X_i; Y_j) \leqslant \rho_{ij}\}, \tag{JP}$$

where the supremum is taken over all the possible sets of variables X_i and Y_j on all the possible probability spaces. This theorem claims that $\mathrm{Opt}^{\mathrm{d/m}}(\bullet)$ is equivalent to $|||\bullet|||$ when $\boldsymbol{\rho} \to \mathbf{0}$ (without any cone restriction like in Theorem 62!), so that the bound of the minimal-assumption version of Theorem 39 cannot be improved (at all) at first order. ◇

Proof. Like in the proof of Theorem 62, we will consider a case where all the X_i's and Y_j's are real and $(\vec{X}_I, \vec{Y}_J)$ is Gaussian. Like in that proof too, $\|\boldsymbol{\rho}\|$ will stand for the ℓ^∞ norm of $\boldsymbol{\rho}$.

Here our construction is the following: let $(\xi_i)_{i\in I}$ and $(\psi_j)_{j\in J}$ be independent $\mathcal{N}(1)$ variables; and for $(c_{ij})_{i,j}$ deterministic real coefficients, let

$$X_i \coloneqq \xi_i + \sum_{j\in J} c_{ij}\psi_j \qquad \forall i \in I; \tag{JQ}$$

$$Y_j \coloneqq \psi_j \qquad \forall j \in J: \tag{JR}$$

this implies in particular that, denoting by $\boldsymbol{C}$ the matrix of the c_{ij}'s, the covariance matrix of $(\vec{X}, \vec{Y})$ writes blockwise

$$\begin{pmatrix} \mathbf{I}_I + \boldsymbol{C}\boldsymbol{C}^\top & \boldsymbol{C} \\ \boldsymbol{C}^\top & \mathbf{I}_J \end{pmatrix} \eqqcolon \boldsymbol{V} \tag{JS}$$

and is positive definite, since the construction of our Gaussian vector $(\vec{X}, \vec{Y})$ ensures it being nondegenerate.

Observe that for given i and j, due to the general properties of Gaussian vectors, the conditional joint law $\mathrm{Law}(X_i, Y_j \mid \vec{X}_{<i} = \vec{x}_{<i} \text{ and } \vec{Y}_{<j} = \vec{y}_{<j})$ is a (uncentred) normal distribution on $\mathbb{R}^2$ whose covariance matrix does not depend on $\vec{x}_{<i}$ nor on $\vec{y}_{<j}$: as a consequence, the conditional correlation coefficient $\mathrm{Corr}(X_i, Y_j \mid \mathscr{F}_i \vee \mathscr{G}_j)$ is constant and therefore can identified with a real number, which we will denote by r_{ij}; moreover, because of Proposition 9 giving the ρ-mixing coefficient in the Gaussian case, $\rho^\dagger_{(\vec{X}_{<i}, \vec{Y}_{<j})}(X_i; Y_j)$ is nothing but $|r_{ij}|$.

So, let us investigate further the r_{ij}'s—the matrix of which we will denote by $\boldsymbol{R}$. The way r_{ij} is obtained from $\boldsymbol{V}$ is a classical result (see also the previous proof): invert $\boldsymbol{V}$; extract its 2×2 submatrix corresponding to rows and columns i and j; invert that submatrix: this yields a positive definite symmetric 2×2 matrix, from which you obtain r_{ij} by dividing the (common) value of the non-diagonal entries by the geometric mean of the diagonal entries. This implies in particular that the function $\boldsymbol{C} \mapsto \boldsymbol{R}$ from $\mathbb{R}^{I\times J}$ to itself has C^1 regularity; moreover, first-order expansion of the procedure shows that, around the point $\boldsymbol{C} = \boldsymbol{0}$, one has

$$\frac{\partial r_{i'j'}}{\partial c_{ij}}\restriction_{\boldsymbol{C}=\boldsymbol{0}} = \delta_{ii'}\delta_{jj'}: \tag{JT}$$

in other words, at first order around $\boldsymbol{0}$, the map $\boldsymbol{C} \mapsto \boldsymbol{R}$ coincides with identity. Hence $\boldsymbol{C} \mapsto \boldsymbol{R}$ is a diffeomorphism from a open neighbourhood U of $\boldsymbol{0}$ in $\mathbb{R}^{I\times J}$ to another neighbourhood $\tilde{U}$ of $\boldsymbol{0}$: let us denote that diffeomorphism by φ. To bound $\mathrm{Opt}^{\mathrm{d/m}}(\boldsymbol{\rho})$ from below, our choice of parameters will be, for $\boldsymbol{\rho} \in \tilde{U}$, to take $\boldsymbol{C} \coloneqq \varphi^{-1}(\boldsymbol{\rho})$. Then one has $\boldsymbol{R} = \boldsymbol{\rho}$, so that our variables $\vec{X}$ and $\vec{Y}$ satisfy indeed the conditions in the supremum defining $\mathrm{Opt}^{\mathrm{d/m}}(\boldsymbol{\rho})$ (cf. Equation (`JP`)). Also, for $\boldsymbol{\rho} \to \boldsymbol{0}$ one will have $\boldsymbol{C} = \boldsymbol{\rho} + \mathrm{o}(\|\boldsymbol{\rho}\|)$, so in particular $|\!|\!|\boldsymbol{C}|\!|\!| \geqslant |\!|\!|\boldsymbol{\rho}|\!|\!| - \mathrm{o}(\|\boldsymbol{\rho}\|)$.

Now we want to find a lower bound for $\rho(\vec{X}; \vec{Y})$. By definition of the operator norm, there exist $\vec{a} \in \mathbb{R}^I, \vec{b} \in \mathbb{R}^J$ with $\|\vec{a}\|_{\ell^2(I)}, \|\vec{b}\|_{\ell^2(J)} = 1$ such that $\vec{a}^\top \boldsymbol{C}\vec{b} = |\!|\!|\boldsymbol{\rho}|\!|\!|$:

so let us fix such $\vec{a}$ and $\vec{b}$; and consider

$$f \coloneqq \sum_{i \in \boldsymbol{I}} a_i X_i; \tag{JU}$$

$$g \coloneqq \sum_{j \in \boldsymbol{J}} b_j Y_j. \tag{JV}$$

Since f and g are resp. $\vec{X}$- and $\vec{Y}$-measurable, $\operatorname{Corr}(f, g)$ shall be a lower bound for $\mathrm{\rho}(\vec{X}; \vec{Y})$. But one computes that, for $\boldsymbol{\rho} \to \mathbf{0}$ (hence also $\boldsymbol{C} \to \mathbf{0}$):

$$\operatorname{Var}(f) = \vec{a}^{\top}(\mathbf{I}_{\boldsymbol{I}} + \boldsymbol{C}\boldsymbol{C}^{\top})\vec{a} = \|\vec{a}\|_2^2 + \mathrm{O}(\|\boldsymbol{C}\|^2) = 1 + \mathrm{O}(\|\boldsymbol{\rho}\|^2); \tag{JW}$$

$$\operatorname{Var}(g) = \vec{b}^{\top}\mathbf{I}_{\boldsymbol{J}}\vec{b} = \|\vec{b}\|_2^2 = 1; \tag{JX}$$

$$\operatorname{Cov}(f, g) = \vec{a}^{\top}\boldsymbol{C}\vec{b} = |||\boldsymbol{C}||| \geqslant |||\boldsymbol{\rho}||| - \mathrm{o}(\|\boldsymbol{\rho}\|), \tag{JY}$$

so that

$$\operatorname{Opt}^{\mathrm{d/m}}(\boldsymbol{\rho}) \geqslant \operatorname{Corr}(f, g) \geqslant |||\boldsymbol{\rho}||| - \mathrm{o}(\|\boldsymbol{\rho}\|). \tag{JZ}$$

Since on the other hand Theorem 58-(ii) ensures that $\operatorname{Opt}^{\mathrm{d/m}}(\boldsymbol{\rho}) \leqslant |||\boldsymbol{\rho}|||$ for all $\boldsymbol{\rho}$, this proves Theorem 65. ♤

Theorem 66 (Exact optimality of Theorem 37)**.** *Let $\boldsymbol{I}$ be a well-ordered set; and for $\vec{\rho} \coloneqq (\rho_i)_i \in \mathbb{R}_+^{\boldsymbol{I}}$, denote*

$$\operatorname{Opt}^{\mathrm{s/m}}(\vec{\rho}) \coloneqq \sup\{\mathrm{\rho}(\vec{X}_{\boldsymbol{I}}; Y) \mid \forall i \;\; \mathrm{\rho}^{\dagger}_{\vec{X}_{<i}}(X_i; Y) \leqslant \rho\}, \tag{KA}$$

where the supremum is taken over all the possible sets of variables X_i and Y on all the possible probability spaces.

Then, for all $\vec{\rho} \in [0, 1]^{\boldsymbol{I}}$,

$$\operatorname{Opt}^{\mathrm{s/m}}(\vec{\rho}) = \Big(1 - \prod_{i \in \boldsymbol{I}}(1 - \rho_i^2)\Big)^{1/2}: \tag{KB}$$

in other words, the bound (DH) *of Theorem 37 is optimal (for the version with minimal assumptions).* ◇

Proof. Before all, observe that it is enough to prove Theorem 66 in the case where $\boldsymbol{I}$ is finite: indeed, for infinite $\boldsymbol{I}$, you can always find some finite $\check{\boldsymbol{I}} \Subset \boldsymbol{I}$ such that $\big(1 - \prod_{i \in \check{\boldsymbol{I}}}(1 - \rho_i^2)\big)^{1/2}$ is arbitrarily close to the right-hand side of (KB); then, taking $\vec{X}_{\check{\boldsymbol{I}}}$ and Y satisfying the assumptions of Theorem 58-(i) (stated for $\check{\boldsymbol{I}}$) such that $\mathrm{\rho}(\vec{X}_{\check{\boldsymbol{I}}}; Y)$ is arbitrarily close to the optimum, you can extend $\vec{X}$ to the whole $\boldsymbol{I}$ by setting $X_i \coloneqq \partial$ for all $i \in \boldsymbol{I} \setminus \check{\boldsymbol{I}}$ (∂ denoting some constant cemetery value): the $\vec{X}_{\boldsymbol{I}}$ and Y obtained this way will satisfy the assumptions of Theorem 58-(i) (stated for $\boldsymbol{I}$) and be such that $\mathrm{\rho}(\vec{X}_{\boldsymbol{I}}; Y) = \mathrm{\rho}(\vec{X}_{\check{\boldsymbol{I}}}; Y)$, thus being arbitrarily close to the wanted value. So in the sequel we assume that $\boldsymbol{I}$ is finite: without loss of generality we can take $\boldsymbol{I} = \{0, \dots, N - 1\}$.

Our proof will be divided into two steps: first, we will show that (KB) holds whenever $(\rho_i)_{i \in \boldsymbol{I}}$ are the $\rho^{\dagger}$-mixing coefficients of a Gaussian vector $(\vec{X}_{\boldsymbol{I}}, Y)$; then, we will check that any value of $\vec{\rho}$ in $[0, 1)^{\boldsymbol{I}}$ can indeed be obtained from a Gaussian vector.

First step: Equality in the Gaussian case Suppose that the X_i's and Y are real random variables such that $(X_0, \dots, X_{N-1}, Y)$ is a non-degenerate Gaussian vector. Then, for $i \in I, \vec{x}_{<i} \in \mathbb{R}^i$, under the law $\mathbb{P}(\bullet \mid \vec{X}_{<i} = \vec{x}_{<i})$, $(X_i, \dots, X_{N-1}, Y)$ is a non-degenerate Gaussian random vector whose covariance matrix does not depend on $\vec{x}_{<i}$; therefore the Pearson correlation coefficient between X_i and Y under that law does not depend on $\vec{x}_{<i}$ either: in the sequel, we will denote that constant conditional coefficient by $\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i)$.[††] Under the law $\mathbb{P}(\bullet \mid \vec{X}_{<i} = \vec{x}_{<i})$, (X_i, Y) is a 2-dimensional Gaussian vector, so by Proposition 9 we have that, under that law, $\rho(X_i; Y) = \left|\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i)\right|$; and since this is true for all $\vec{x}_{<i}$, we deduce that

$$\rho^\dagger_{\vec{X}_{<i}}(X_i; Y) = \left|\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i)\right|. \tag{KC}$$

Now define $g \coloneqq Y$ and take the notation of the proof of Theorem 37. We are going to prove that $\mathrm{Var}(g^{\sigma(\vec{X}_I)}) = \left(1 - \prod_{i \in I}\left(1 - \rho^\dagger_{\vec{X}_{<i}}(X_i; Y)^2\right)\right) \mathrm{Var}(g)$, which by Proposition 4 will prove the '$\geqslant$' sense of (KB) in the case where $\vec{\rho} = \left(\rho^\dagger_{\vec{X}_{<i}}(X_i; Y)\right)_{i \in I}$—the sense '$\leqslant$' being given by Theorem 37 itself, in its minimal version. Basically, what we will do is just to follow step by step the proof of Theorem 37 and to check that all the bounds of that proof are actually equalities when one handles Gaussian vectors. Let us see in particular how one derives the link between $\mathrm{Var}(g^i)$ and $\mathrm{Var}(g^{i+1})$. As g^i is centred w.r.t. $\mathscr{F}_i$, by Lemma 23 one has $\mathrm{Var}(g^i) = \mathbb{E}\left(\mathrm{Var}(g^i \mid \mathscr{F}_i)\right) = \mathbb{E}\left(\mathrm{Var}(g \mid \mathscr{F}_i)\right) \underset{\text{def}}{=} \mathbb{E}\left(\mathrm{Var}(Y \mid \mathscr{F}_i)\right)$; and as $\mathrm{Var}(Y \mid \mathscr{F}_i)$ is actually constant since $(X_0, \dots, X_{i-1}, Y)$ is a Gaussian vector, we can interpret $\mathrm{Var}(g^i)$ as a (constant) conditional variance:

$$\mathrm{Var}(g^i) = \mathrm{Var}(Y \mid \mathscr{F}_i). \tag{KD}$$

Likewise one has

$$\mathrm{Var}(g^{i+1}) = \mathrm{Var}(Y \mid \mathscr{F}_{i+1}). \tag{KE}$$

Now let us reason conditionally to $\mathbb{P}(\bullet \mid \vec{X}_{<i} = \vec{x}_{<i})$ for an arbitrary $\vec{x}_{<i} \in \mathbb{R}^i$. The variance decomposition formula (CG) tells us that, under that law:

$$\mathrm{Var}(Y) = \mathrm{Var}(Y^{\sigma(X_i)}) + \mathbb{E}\left(\mathrm{Var}(Y \mid \sigma(X_i))\right). \tag{KF}$$

In unconditional terms, the left-hand side of (KF) corresponds to $\mathrm{Var}(Y \mid \mathscr{F}_i)$, while in the right-hand side "$\mathrm{Law}(Y \mid \sigma(X_i))$" corresponds to $\mathrm{Law}(Y \mid \sigma(\vec{X}_{<i}) \vee \sigma(X_i)) = \mathrm{Law}(Y \mid \mathscr{F}_{i+1})$, so that the second term of the right-hand side is just $\mathrm{Var}(Y \mid \mathscr{F}_{i+1})$. As regards $Y^{\sigma(X_i)}$, under the conditional law (X_i, Y) is a 2-dimensional Gaussian vector, so (under the conditional law)

$$Y^{\sigma(X_i)} = \frac{\mathrm{Cov}(X_i, Y)}{\mathrm{Var}(X_i)} X_i, \tag{KG}$$

and therefore $\mathrm{Var}(Y^{\sigma(X_i)}) = \left(\mathrm{Cov}(X_i, Y) / \mathrm{Var}(X_i)\right)^2 \mathrm{Var}(X_i) = \mathrm{Corr}(X, Y_i)^2 \mathrm{Var}(Y)$, which in unconditional terms is equal to $\mathrm{Corr}(X, Y_i \mid \mathscr{F}_i)^2 \mathrm{Var}(Y \mid \mathscr{F}_i)$. So (KF) yields:

$$\mathrm{Var}(Y \mid \mathscr{F}_i) = \mathrm{Corr}(X, Y_i \mid \mathscr{F}_i)^2 \mathrm{Var}(Y \mid \mathscr{F}_i) + \mathrm{Var}(Y \mid \mathscr{F}_{i+1}), \tag{KH}$$

[††] So here $\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i)$ has to be understood as a mere number, not as a $\mathscr{F}_i$-measurable random variable.

which means by (KC), (KD) and (KE) that $\mathrm{Var}(g^{i+1}) = \big(1 - \rho^{\dagger}_{\vec{X}_{<i}}(X_i;Y)^2\big)\,\mathrm{Var}(g^i)$; combining that result for $i = 0, \dots, N-1$, we get that $\mathrm{Var}(g^{\sigma(\vec{X}_I)}) = \big(1-\prod_{i\in I}(1-\rho_i^2)\big) \times \mathrm{Var}(g)$, as wanted. So we have proved that the equality (KB) holds whenever $\vec{\rho}$ corresponds to a non-degenerate Gaussian vector.

Second step: Any $\vec{\rho}$ can be got from a Gaussian case Now we have to prove that any $\vec{\rho} \in [0,1)^I$ can be got from a Gaussian vector. Note that it is not necessary to deal with the case where some of the ρ_i's are equal to 1, since then (KB) is trivial; which is why we are assuming that $\rho_i < 1$ $\forall i$. Consider $N+1$ independent standard Gaussian random variables $\xi_0, \dots, \xi_{N-1}, \psi$; we are going to show that, for all possible $(r_i)_i \in (-1,1)^N$, it is possible to find coefficients $(c_i)_i \in \mathbb{R}^N$ such that the variables defined by

$$\begin{cases} X_i := \xi_i + c_i \psi \\ Y := \psi \end{cases} \tag{KI}$$

satisfy, for all i,

$$\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i) = r_i: \tag{KJ}$$

then, as we noticed in the first step of the proof, one will have $\rho^{\dagger}_{\vec{X}_{<i}}(X_i;Y) = |r_i|$ $\forall i$, so that (KB) will be proved for $\vec{\rho} = (|r_i|)_{i\in I}$.

In order to ensure the condition (KJ), we first observe that $\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i)$ only depends on the choice of $(c_0, \dots, c_i)$. Hence, our strategy will be, first to choose c_0 ensuring that $\mathrm{Corr}(X_0, Y \mid \mathscr{F}_0) = r_0$; then, that value of c_0 being fixed, to choose c_1 ensuring that $\mathrm{Corr}(X_1, Y \mid \mathscr{F}_1) = r_1$; and so on. But, for fixed values of $(c_0, \dots, c_{i-1})$, conditionally to $\mathscr{F}_i$, ψ is a Gaussian vector with some given non-zero variance $\mathrm{Var}(\psi \mid \mathscr{F}_i)$ (depending on the choice of $\vec{c}_{<i}$, but independent of the value of $\vec{X}_{<i}$),[‡‡] while ξ_i is still a standard normal variable independent of ψ; thus,

$$\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i) = \frac{c_i \,\mathrm{Var}(\psi \mid \mathscr{F}_i)}{\mathrm{Var}^{1/2}(\psi \mid \mathscr{F}_i)\big(1 + c_i^2 \,\mathrm{Var}^{1/2}(\psi \mid \mathscr{F}_i)\big)}. \tag{KK}$$

From Equation (KK), we see that the function $c_i \mapsto \mathrm{Corr}(X_i, Y \mid \mathscr{F}_i)$ is continuous and (strictly) increasing from $\mathbb{R}$ to $(-1,1)$, with $\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i) \to \pm 1$ as $c_i \to \pm\infty$; therefore, there exists indeed a (unique) choice of c_i such that $\mathrm{Corr}(X_i, Y \mid \mathscr{F}_i) = r_i$: this is what we wanted. ♤

Theorem 67 (Exact optimality of Theorem 47)**.** *For $\vec{\rho} := (\rho_z)_z \in \mathbb{R}_+^{\mathbb{Z}}$, denote*

$$\mathrm{Opt}^{\mathrm{p/m}}(\vec{\rho}) := \sup\left\{ \rho(\vec{X}_{\mathbb{Z}}; \vec{Y}_{\mathbb{Z}}) \;\middle|\; \left\{ \begin{array}{l} \overset{\searrow}{\lim\limits_{i\to-\infty}} \sigma(\vec{X}_{<i}) \overset{\text{a.s.}}{=} \mathcal{O}; \\ \overset{\searrow}{\lim\limits_{j\to-\infty}} \sigma(\vec{Y}_{<j}) \overset{\text{a.s.}}{=} \mathcal{O}; \\ \forall i,j \;\; \rho^{\dagger}_{\vec{X}_{<i},\vec{Y}_{<j}}(X_i;Y_j) \leqslant \rho_{j-i} \end{array} \right. \right\}, \tag{KL}$$

[‡‡] Actually one could compute that $\mathrm{Var}(\psi \mid \mathscr{F}_i) = (1 + \sum_{i'<i} c_{i'}^2)^{-1}$.

where the supremum is taken over all the possible sets of variables X_i and Y_j on all the possible probability spaces. Then, for all $\vec{\rho} \in [0,1]^{\mathbb{Z}}$,

$$\mathrm{Opt}^{\mathrm{p/m}}(\vec{\rho}) = \sin\Big(\Big(\sum_{z\in\mathbb{Z}} \arcsin \rho_z\Big) \wedge \frac{\pi}{2}\Big): \tag{KM}$$

in other words, the bound (`GE`) *of Theorem 47 is optimal (for the version with minimal assumptions).* ◇

Proof. Like the proof of Theorem 66, the proof of Theorem 67 will consist in showing that (essentially) any set of correlation coefficients can be realized by a Gaussian model, for which one will (essentially) have equality all along the proof of Theorem 47.

Before all, let us observe that it is enough to prove the theorem when only a finite number of ρ_z's are nonzero: indeed, the error made on the right-hand side of (`KM`) when replacing $(\rho_z)_{z\in\mathbb{Z}}$ by $(\mathbf{1}_{|z|\leqslant N}\rho_z)_{z\in\mathbb{Z}}$ tends to zero as N tends to infinity. So in the sequel we assume that $(\rho_z)_z$ has finite support, and denote by N a number such that $\rho_z = 0$ as soon as $|z| > N$. We can also assume safely that all the ρ_z's are < 1, since the result is trivial otherwise.

Let $(\theta_{-N}, \theta_{-N+1}, \dots, \theta_N) \in \big(-\frac{\pi}{2}, \frac{\pi}{2}\big)^{2N+1}$ and set $\theta_z := 0$ for $|z| > N$; introduce $\gamma_z := \sum_{z'\geqslant z} \theta_z$ for $z \in \mathbb{Z}$, resp. $\gamma_{-\infty} := \gamma_{z\leqslant N}$. The following technical lemma will be the key to build the set of random variables that we are looking for:

Lemma 68. *Take N, the θ_z's and the γ_z's as above; let $v, w \in \mathbb{R}_+^*$;*[*] *and let j_-, $j_+ \in \mathbb{Z}$ with $j_+ \geqslant j_- + 2N$, and $i \in [\![j_- + N, j_+ - N]\!]$. Consider $(Y_{j_-}, Y_{j_-+1}, \dots, Y_{j_+-1}) =: \vec{Y}$ an arbitrary non-degenerate centred Gaussian vector of dimension $j_+ - j_-$.*

*In the sequel, for $\vec{S}$ a centred Gaussian vector, we will say that a real random variable is $\vec{S}$-*linearly Gaussian *when it is a linear combination of the entries of $\vec{S}$.*[†]

In the first part of this lemma, we are going to build a real random variable g satisfying certain properties. Before that, we introduce some notation: we set $\mathscr{G} := \sigma(\vec{Y}_{[\![j_-, j_+ [\![})$, $\mathscr{G}_j := \sigma(\vec{Y}_{<j})$, and $g_j := g^{\mathscr{G}_{j+1}} - \mathbb{E}(g \mid \mathscr{G}_j)$. I claim that it is possible to find g such that:

(i) g is $\vec{Y}$-linearly Gaussian (and hence $\mathscr{G}$-measurable);

(ii) For all $j \in [\![j_-, j_+ [\![$, $\mathrm{Var}(g_j) = \cos^2 \gamma_{j-i+1} \times w$.

For the second part of this lemma, we assume that $i < j_+ - N$. We suppose that some g has been built according to the first part of the lemma, and we let ξ be a $\mathcal{N}(1)$ variable independent of $\vec{Y}$; in this context, we are going to build a real random variable f satisfying certain properties. Introducing the notation $f^j := f - \mathbb{E}(f \mid \mathscr{G}_j)$, I claim that it is possible to find f such that:

(iii) f is $(\xi, \vec{Y})$-linearly Gaussian;

(iv) For all $j \in [\![j_-, j_+]\!]$, $\mathrm{Var}(f^j) = \prod_{z<j-i} \cos^2 \theta_z \times v$;

---

[*] The quantities v and w are actually "homogeneity constants": we could have taken them equal to 1; but allowing for arbitrary values will give a better match with the proof of Theorem 47.

[†] So note that linearly Gaussian variables shall always be centred.

(v) *For all* $j \in [\![j_-, j_+ [\![$, $\operatorname{Cov}(f^j, g_j) = \prod_{z<j-i} \cos\theta_z \times \sin\theta_{j-i} \cos\gamma_{j-i+1} \times \sqrt{vw}$*;*

(vi) *For all* $j \in [\![j_-, j_+ [\![$, $\rho^{\dagger}_{\vec{Y}_{<j}}(f; Y_j) = |\sin\theta_{j-i}|$.

In the third part of this lemma, still for $i < j_+ - N$*, we state certain further properties of the* f *and* g *that we have been building. Before that, we introduce some notation: set* $\mathscr{F} \coloneqq \sigma(f)$, $\mathscr{G}' \coloneqq \mathscr{F} \vee \mathscr{G}$, $g' \coloneqq g - \mathbb{E}(g \mid \mathscr{F})$, $\mathscr{G}'_j \coloneqq \mathscr{F} \vee \mathscr{G}_j$, $g'_j \coloneqq g^{\mathscr{G}'_{j+1}} - \mathbb{E}(g \mid \mathscr{G}'_j)$*. We have that:*

(vii) $\vec{Y} - \mathbb{E}(\vec{Y} \mid \mathscr{F}) \eqqcolon \vec{Y}'$ *is a nondegenerate centred Gaussian vector, independent of* $\mathscr{F}$*;*

(viii) g' *is* $\vec{Y}'$*-linearly Gaussian;*

(ix) *For all* $j \in [\![j_-, j_+ [\![$, $\operatorname{Var}(g'_j) = \cos^2\gamma_{j-i} \times w$. ◇

Proof of the first part of Lemma 68. Since the covariance matrix of the Y_j's is nondegenerate, there exists a (unique) lower triangular matrix $\boldsymbol{L}$ with strictly positive diagonal coefficients (and hence invertible) such that $\boldsymbol{L}^{-1}\vec{Y} \eqqcolon \vec{\psi}$ is a $\mathcal{N}(\mathbf{I}_{j_+ - j_-})$ Gaussian vector. (This is a classical application of the well-known Cholesky decomposition theorem). As $\boldsymbol{L}$ has been taken lower triangular, one has moreover that $\mathscr{G}_j \underset{\text{def}}{=} \sigma(\vec{Y}_{<j}) = \sigma(\vec{\psi}_{<j})$. Then, writing g as a linear combination $\sum_j b_j \psi_j$ of the ψ_j's rather than of the Y_j's, we observe that $g_j = b_j \psi_j$; and therefore we can build the wanted variable g by choosing $b_j = \pm\cos\gamma_{j-i+1} \times \sqrt{w}\,\psi_j$ for all j. (So there are exactly $2^{j_+ - j_-}$ ways to find such a g). ♤

Proof of the second part of Lemma 68. We are looking for f under the form $f = a\xi + \sum_j c_j \psi_j$. Then one will have $f^j = a\xi + \sum_{j' \geqslant j} c_{j'} \psi_{j'}$, and hence $\operatorname{Var}(f^j) - \operatorname{Var}(f^{j+1}) = c_j^2$, which implies we have to choose $c_j = \pm \prod_{z<j-i} \cos\theta_z \times \sin\theta_{j-i} \times \sqrt{v}$. Now, remembering that $g_j = \pm\cos\gamma_{j-i+1} \times \sqrt{w}\,\psi_j$, one will have $\operatorname{Cov}(f^j, g_j) = \pm c_j \cos\gamma_{j-i+1} \times \sqrt{w}$: therefore, Item (v) will be satisfied by choosing the correct sign for c_j. To satisfy Item (iv), given that our choice for the c_j's has already ensured the wished value for $\operatorname{Var}(f^{j+1}) - \operatorname{Var}(f^j)$, it only remains to ensure that $\operatorname{Var}(f^{j_+}) = \prod_{z\in\mathbb{Z}} \cos^2\theta_z \times v$. But $f^{j_+} = a\xi$; hence we have two possibilities to ensure that, viz. by choosing $a = \pm \prod_{z\in\mathbb{Z}} \cos\theta_z \times \sqrt{v}$. (So there are exactly two ways to build an f satisfying the wished properties, as the c_j's are imposed and a is determined up to a free sign).

As regards Item (vi), we are going to show that it is actually automatically implied by the previous points. First observe that, conditionally to $\mathscr{G}_j$, g_j is Y_j-measurable; since everything is Gaussian, this implies that g_j is of the form $\alpha Y_j + b(\vec{Y}_{<j})$, where α does not depend on $\vec{Y}_{<j}$. Assume for the sake of simplicity that $\alpha \neq 0$.[‡] Then, conditionally to $\mathscr{G}_j$, Y_j and g_j are in bijection and hence generate the same σ-algebra,

---

[‡] If $\alpha = 0$, one has to use that $\rho^{\dagger}_{\vec{Y}_{<j}}(f; Y_j) = \rho^{\dagger}_{\vec{\psi}_{<j}}(f^j; \psi_j)$ and go back to the way f was constructed from ξ and the ψ_j's: this works just fine, but these computations are much less enlightening in my opinion.

so that $\rho^{\dagger}_{\vec{Y}_{<j}}(f;Y_j)=\rho^{\dagger}_{\vec{Y}_{<j}}(f;g_j)=\rho^{\dagger}_{\vec{Y}_{<j}}(f^j;g_j)$ (where the last equality comes from the fact that f and f^j only differ by a constant under the conditional law). But (under the conditional law) (f^j,g_j) is a 2-dimensional Gaussian vector, so by Proposition 9 one has

$$\rho^{\dagger}_{\vec{Y}_{<j}}(f;Y_j)=\frac{\left|\operatorname{Cov}(f^j,g_j\mid\mathscr{G}_j)\right|}{\operatorname{Var}^{1/2}(f^j\mid\mathscr{G}_j)\operatorname{Var}^{1/2}(g_j\mid\mathscr{G}_j)}{}^{[\S]}\tag{KN}$$

(where the $\operatorname{Corr}(\bullet,\bullet\mid\mathscr{G}_j)$ and $\operatorname{Var}(\bullet\mid\mathscr{G}_j)$'s are actually constants, since everything is Gaussian). Since by construction f^j and g_j are centred w.r.t. $\mathscr{G}_j$, Lemma 23 gives that $\operatorname{Cov}(f^j,g_j\mid\mathscr{G}_j)=\operatorname{Cov}(f^j,g_j)$, resp. $\operatorname{Var}(f^j\mid\mathscr{G}_j)=\operatorname{Var}(f^j)$, $\operatorname{Var}(g_j\mid\mathscr{G}_j)=\operatorname{Var}(g_j)$; so, combining Items (v), (iv) and (ii), we get the announced value of $\rho^{\dagger}_{\vec{Y}_{<j}}(f;Y_j)$. ♤

Proof of the third part of Lemma 68. Item (vii) is a standard property of Gaussian vectors. Item (viii) comes from the fact that g was a linear combination of the Y_j's and that conditional expectation is linear. Finally, Item (ix) will be proved along the same lines as how we derived (HF), except that here we will have equality since everything is Gaussian. Viz.:

- First, denoting $\tilde{g}_j:=g-\mathbb{E}(g\mid\mathscr{G}_j)$, we note that $\sum_{j'\geqslant j}\operatorname{Cov}(f^{j'},g_{j'})=\sum_{j'\geqslant j}\operatorname{Cov}(f^j,g_{j'})=\operatorname{Cov}(f^j,\tilde{g}_j)=\operatorname{Cov}(f^j,\tilde{g}_j^{\mathscr{G}'_j})$, by using successively (BX) applied to $\mathscr{G}_{j'}$, telescopic decomposition, and (BW) applied to $\mathscr{G}'_j$.

- Next we observe that, conditionally to $\mathscr{G}_j$, both f^j and $\tilde{g}_j^{\mathscr{G}'_j}$ are $\mathscr{F}$-measurable; since everything is Gaussian and that $\mathscr{F}$ is spanned by a unique real Gaussian variable, this implies that (still conditionally to $\mathscr{G}_j$) there is an affine relation between f^j and $\tilde{g}_j^{\mathscr{G}'_j}$, and therefore that $\operatorname{Cov}(f^j,\tilde{g}_j^{\mathscr{G}'_j}\mid\mathscr{G}_j)^2=\operatorname{Var}(f^j\mid\mathscr{G}_j)\operatorname{Var}(\tilde{g}_j^{\mathscr{G}'_j}\mid\mathscr{G}_j)$, where $\operatorname{Cov}(f^j,\tilde{g}_j^{\mathscr{G}'_j}\mid\mathscr{G}_j)$, $\operatorname{Var}(f^j\mid\mathscr{G}_j)$ and $\operatorname{Var}(\tilde{g}_j^{\mathscr{G}'_j}\mid\mathscr{G}_j)$ are actually constants since everything is Gaussian. Observing that f^j and $\tilde{g}_j^{\mathscr{G}'_j}$ are centred w.r.t. $\mathscr{G}_j$, Lemma 23 simplifies that relation into "$\operatorname{Cov}(f^j,\tilde{g}_j^{\mathscr{G}'_j})^2=\operatorname{Var}(f^j)\operatorname{Var}(\tilde{g}_j^{\mathscr{G}'_j})$".

- Then we note that, denoting $\tilde{g}'_j:=g-\mathbb{E}(g\mid\mathscr{G}'_j)$, one has that $\operatorname{Var}(\tilde{g}_j^{\mathscr{G}'_j})=\operatorname{Var}(g^{\mathscr{G}'_j}-g^{\mathscr{G}_j})=\operatorname{Var}(\tilde{g}_j)-\operatorname{Var}(\tilde{g}'_j)$, the last equality coming from Lemma 26.

- Fourth, using the chain rule for variance, we compute that $\operatorname{Var}(\tilde{g}_j)=\sum_{j'\geqslant j}\operatorname{Var}(g_{j'})$ and $\operatorname{Var}(\tilde{g}'_j)=\sum_{j'\geqslant j}\operatorname{Var}(g'_{j'})$.

- Putting all the previous points together, we get that

$$\sum_{j'\geqslant j}\operatorname{Var}(g_{j'})-\sum_{j'\geqslant j}\operatorname{Var}(g'_{j'})=\frac{\left(\sum_{j'\geqslant j}\operatorname{Cov}(f^{j'},g_{j'})\right)^2}{\operatorname{Var}(f^j)}.\tag{KO}$$

[§] Note that both f^j and g_j are non-constant conditionally to $\mathscr{G}_j$, so this formula is well-defined.

In (`KO`), we compute that

$$\begin{aligned}\sum_{j'\geqslant j}\mathrm{Cov}(f^{j'},g_{j'}) &= \sum_{j'\geqslant j}\Big(\prod_{z<j'-i}\cos\theta_z\times\sin\theta_{j'-i}\cos\gamma_{j'-i+1}\Big)\times\sqrt{vw}\\ &= \prod_{z<j-i}\cos\theta_z\times\sin\gamma_{j-i}\times\sqrt{vw}\,, \qquad (\mathtt{KP})\end{aligned}$$

as one checks easily by downgoing induction on j.

- Therefore we finally get that

$$\begin{aligned}\mathrm{Var}(g'_j) &= \mathrm{Var}(g_j)-(\mathrm{Var}(g_j)-\mathrm{Var}(g'_j))\\ &= \mathrm{Var}(g_j)-\Big(\sum_{j'\geqslant j}\mathrm{Var}(g_{j'})-\sum_{j'\geqslant j}\mathrm{Var}(g'_{j'})\Big)+\Big(\sum_{j'\geqslant j+1}\mathrm{Var}(g_{j'})-\sum_{j'\geqslant j+1}\mathrm{Var}(g'_{j'})\Big)\\ &= w\cos^2\gamma_{j-i+1}-\frac{vw\sin^2\gamma_{j-i}\prod_{z<j-i}\cos^2\theta_z}{v\prod_{z<j-i}\cos^2\theta_z}+\frac{vw\sin^2\gamma_{j-i+1}\prod_{z<j-i+1}\cos^2\theta_z}{v\prod_{z<j-i+1}\cos^2\theta_z}\\ &= \big(\cos^2\gamma_{j-i+1}-\sin^2\gamma_{j-i}+\sin^2\gamma_{j-i+1}\big)w=(1-\sin^2\gamma_{j-i})w=\cos^2\gamma_{j-i}\times w. \quad \spadesuit \qquad (\mathtt{KQ})\end{aligned}$$

Thanks to Lemma 68, for any range of indices $[\![j_-,j_+[\![$ such that $j_+-j_-\geqslant 2N$, denoting $i_-\coloneqq j_-+N$, resp. $i_+\coloneqq j_+-N$, one can build sets of random variables $(X_i)_{i_-\leqslant i<i_+}$ and $(Y_j)_{j_-\leqslant j<j_+}$, and resp. $\vec{X}$-measurable and $\vec{Y}$-measurable real variables f and g, such that, taking the notation of the proof of Theorem 47, one has for all i,j:

$$\rho^\dagger_{(\vec{X}_{<i},\vec{Y}_{<j})}(X_i;Y_j)=|\sin\theta_{j-i}|; \qquad (\mathtt{KR})$$

$$V_i^j=\prod_{z<j-i}\cos^2\theta_z\times v; \qquad (\mathtt{KS})$$

$$W_j^i=\cos^2\gamma_{j-i-1}\times w; \qquad (\mathtt{KT})$$

$$C_{ij}=\prod_{z<j-i}\cos\theta_z\times\sin\theta_{j-i}\cos\gamma_{j-i-1}\times\sqrt{vw}\,. \qquad (\mathtt{KU})$$

How do we do this? We start from an arbitrary non-degenerate Gaussian random vector $\vec{Y}_{[\![j_-,j_+[\![}\eqqcolon\vec{Y}$; and we pick a $\vec{Y}$-linearly Gaussian variable g such that, for all j, $\mathrm{Var}(g_j)\underset{\text{def}}{=}\mathrm{Var}(g_j^{i_-})=\cos^2\gamma_{j-i_-+1}\times w$: this is possible thanks to the first step of Lemma 68. Next, we introduce a $\mathscr{N}(1)$ variable ξ_{i_-} independent of $\vec{Y}$; and we pick a $(\xi_{i_-},\vec{Y})$-linearly Gaussian variable f_{i_-} such that $\mathrm{Var}(f_{i_-}^j)=\prod_{z<j-i_-}\cos^2\theta_z\times v$, $C_{i_-j}=\prod_{z<j-i_-}\cos\theta_z\times\sin\theta_{j-i_-}\cos\gamma_{j-i_-+1}\times\sqrt{vw}$, and $\rho^\dagger_{\vec{Y}_{<j}}(f_{i_-};Y_j)=|\sin\theta_{j-i_-}|$: this is possible thanks to the second step of the lemma. We set $X_{i_-}\coloneqq f_{i_-}$. Now to get X_{i_-+1}, we reason conditionally to $\mathscr{F}_{i_-+1}$: under this conditioning, the third step of the lemma states that $\vec{Y}$ has become a new non-degenerate Gaussian vector (possibly non-centred), and that g^{i_-+1} is a linear combination the $(Y_j-\mathbb{E}(Y_j))$'s, such that

$\mathrm{Var}(g_j^{i_-+1}) = \cos^2\gamma_{j-i_-} \times w$. Then, still conditionally to $\mathscr{F}_{i_-+1}$, we introduce a $\mathscr{N}(1)$ variable ξ_{i_-+1}; and then the second step of the lemma allows us to find a $(\xi_{i_-+1}, \vec{Y} - \mathbb{E}(\vec{Y}))$-linearly Gaussian variable f_{i_-+1} such that $\mathrm{Var}(f_{i_-+1}^j) = \prod_{z<j-i_--1} \cos^2\theta_z \times v$, $C_{(i_-+1)j} = \prod_{z<j-i_--1} \cos\theta_z \times \sin\theta_{j-i_--1} \cos\gamma_{j-i_-} \times \sqrt{vw}$, and $\mathrm{\rho}^\dagger_{(X_{i_-}, \vec{Y}_{<j})}(f_{i_-+1}; Y_j) = |\sin\theta_{j-i_--1}|$. We set $X_{i_-+1} \coloneqq f_{i_-+1}$. In the next step we reason conditionally to $\mathscr{F}_{i_-+2}$, where g^{i_-+2} satisfies the wanted assumptions; we introduce an independent $\mathscr{N}(1)$ variable ξ_{i_-+2} independent of $\vec{Y}$; we find a $(\xi_{i_-+2}, \vec{Y} - \mathbb{E}(\vec{Y}))$-linearly Gaussian f_{i_-+2} satisfying the wanted assumptions; we set $X_{i_-+2} \coloneqq f_{i_-+2}$; and we continue alternating the use of the third and second steps of Lemma 68, until we have built all the f_i's. Then we define $f \coloneqq \sum_i f_i$, and check that this re-gives indeed the f_i^j's that we have been computing.[¶]

Now, also define $X_i :\equiv \partial$ for $i \notin [\![i_-, i_+[\![$, resp. $Y_j :\equiv \partial$ for $j \notin [\![j_-, j_+[\![$. Then $(X_i)_{i\in\mathbb{Z}}$ and $(Y_j)_{j\in\mathbb{Z}}$ satisfy the minimal version of the assumptions of Theorem 47, with $\rho_z = |\sin\theta_z|$. One computes that

$$\mathrm{Var}(f) = \sum_i V_i^{j_-} = \sum_{i=i_-}^{i_+-1} v = (i_+ - i_-)v; \tag{KV}$$

$$\mathrm{Var}(g) = \sum_j W_j^{i_-} = \sum_{j=j_-}^{j_+-1} (\cos^2\gamma_{j-i_-+1} \times w) \leqslant (j_+ - j_-)w; \tag{KW}$$

and lastly

$$\begin{aligned}\mathrm{Cov}(f,g) = \sum_{i,j} C_{ij} = \sum_{i=i_-}^{i_+-1}\sum_{j=j_-}^{j_+-1}\Big(\prod_{z<j-i} \cos\theta_z \times \sin\theta_{j-i}\cos\gamma_{j-i+1} \times \sqrt{vw}\Big)\\ = (i_+ - i_-)\sin\gamma_{-\infty} \times \sqrt{vw}\,, \end{aligned} \tag{KX}$$

where in (KX) we used the identity "$\sum_{z\in[\![-N,N]\!]}\big(\prod_{z'<z}\cos\theta_{z'} \times \sin\theta_z\cos\gamma_{z+1}\big) = \sin\gamma_{-\infty}$" that we had established in (KP).

From that we deduce that

$$\begin{aligned}\mathrm{Opt}^{\mathrm{p/m}}(\vec{\rho}) \geqslant \mathrm{\rho}(\vec{X};\vec{Y}) \geqslant \frac{|\mathrm{Cov}(f,g)|}{\mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g)}\\ \geqslant \left(\frac{i_+ - i_-}{j_+ - j_-}\right)^{1/2} |\sin\gamma_{-\infty}| = \left(1 - \frac{2N}{j_+ - j_-}\right)^{1/2} |\sin\gamma_{-\infty}|; \end{aligned} \tag{KY}$$

[¶] Here some subtle point has to be made clear. When we introduced the "f_i"'s during our proof, the 'i' in "f_i" was just an ordinary index with no particular signification, and "f_i^j" just meant $(f_i)^j$, with "$\bullet^j$" having its meaning from Lemma 68. Then we built a function f, for which one could define "f_i"'s and "f_i^j"'s *with "$\bullet_i$" and "$\bullet_i^j$" having their meanings from the proof of Theorem 47.* When I say "this re-gives the f_i^j's that we have been computing", what I mean is that both ways of interpreting "f_i^j" (and "f_i") do define the same variables indeed—which, in retrospect, justifies these notational choices! ◡̈ and since $j_+ - j_-$ can be chosen arbitrarily large, we deduce by passing to the limit that $\mathrm{Opt}^{\mathrm{p/m}}(\vec{\rho}) \geqslant |\sin \gamma_{-\infty}|$.

Now, for given $\vec{\rho} \in [0,1)^{\mathbb{Z}}$ (with finitely many nonzero entries), if we take $\theta_z := \arcsin \rho_z$, one has indeed that $\rho_z = |\sin \theta_z|$; and, provided that $\gamma_\infty \leqslant \frac{\pi}{2}$, $|\sin \gamma_{-\infty}|$ is exactly the right-hand side of (`KM`): this proves the '$\geqslant$' sense of (`KM`) in the case where $\sum_z \arcsin \rho_z \leqslant \frac{\pi}{2}$. In the case where $\sum_z \arcsin \rho_z > \frac{\pi}{2}$, we use that $\mathrm{Opt}^{\mathrm{p/m}}(\vec{\rho})$ is trivially increasing in each of the ρ_z's to get the '$\geqslant$' sense of (`KM`) from the previous case. And the '$\leqslant$' sense is just (the minimal version of) Theorem 47 itself. ♤

Theorem 69 (Exact optimality of Corollary 56)**.** *For $\boldsymbol{d} \in \mathbb{N}$, for $\vec{\rho} := (\rho_u)_u \in \mathbb{R}_+^{\mathbb{Z}^d}$, denote*

$$\mathrm{Opt}^{\mathrm{m/m}}(\vec{\rho}) := \sup\left\{ \rho(\vec{X}_{\mathbb{Z}^d}; \vec{Y}_{\mathbb{Z}^d}) \;\middle|\; \begin{cases} (\sigma(\vec{X}_{<i}))_{i \in (\{-\infty\} \cup \mathbb{Z})^d} \textit{ is almost-right-continuous;} \\ (\sigma(\vec{Y}_{<j}))_{j \in (\{-\infty\} \cup \mathbb{Z})^d} \textit{ is almost-right-continuous;} \\ \forall i, j \;\; \rho^{\dagger}_{\vec{X}_{<i}, \vec{Y}_{<j}}(X_i; Y_j) \leqslant \rho_{j-i} \end{cases} \right\}, \tag{KZ}$$

where the supremum is taken over all the possible sets of variables $\boldsymbol{X}_i$ and $\boldsymbol{Y}_j$ on all the possible probability spaces. Then, for all $\vec{\rho} \in [0,1]^{\mathbb{Z}^d}$,

$$\mathrm{Opt}^{\mathrm{m/m}}(\vec{\rho}) = \sin\Bigl(\Bigl(\sum_{u \in \mathbb{Z}^d} \arcsin \rho_u \Bigr) \wedge \frac{\pi}{2} \Bigr). \tag{LA}$$

◇

Proof. In essence, the proof of Theorem 69 is the same as the proof of Theorem 67, except that the order structure to be considered is now the lexicographic ordering on $\mathbb{Z}^d$. However, as that ordering is more subtle than the ordering on $\mathbb{Z}$, I chose to present this proof from a different angle: namely, here I will give directly explicit formulae for the variables involved, the rationale for these formulae coming from the previous proof. Also, due to the order structure on $\mathbb{Z}^d$, the way inductive reasoning is carried here shall be slightly different. On a less essential level, note that here there will not be any more "v"'s and "w"'s, as I took these quantities equal to 1 for the sake of simplicity.

First note that it is enough to prove the case where $\vec{\rho}$ has finite support: indeed, if we replace $(\rho_u)_{u \in \mathbb{Z}^d}$ by $(\mathbf{1}_{|u| \leqslant N} \rho_u)_{u \in \mathbb{Z}^d}$ (where $|u|$ stands for the ℓ^∞ norm of u), then the replaced right-hand side of (`LA`) will tend to the original one as $N \to \infty$. So in the sequel we assume that $\vec{\rho}$ is finitely supported, and we let N be large enough so that $\rho_u = 0$ as soon as $|u| > N$. For $u \in \mathbb{Z}^d$ we denote $\theta_u := \arcsin \rho_u$. Since it is clearly sufficient to prove the theorem in the case $\sum_{u \in \mathbb{Z}^d} \theta_u < \frac{\pi}{2}$, we will make that assumption in the sequel.

We pick $j_-, j_+ \in \mathbb{Z}$ with $j_+ - j_- \geqslant 2N$ and set $i_- := j_- + N$, resp. $i_+ := j_+ - N$. Let $\boldsymbol{I} := [\![i_-, i_+ [\![^d$, resp. $\boldsymbol{J} := [\![j_-, j_+ [\![^d$; also let $\bar{\boldsymbol{I}} := [\![i_-, i_+ [\![^{d-1} \times [\![i_-, i_+]\!]$ and $\bar{\boldsymbol{J}} := [\![j_-, j_+ [\![^{d-1} \times [\![j_-, j_+]\!]$. Denote $\boldsymbol{i}_- := (i_-, \ldots, i_-)$ and $\boldsymbol{i}_+ := (i_+ - 1, \ldots, i_+ - 1, i_+)$, with similar definitions for $\boldsymbol{j}_-$ and $\boldsymbol{j}_+$; also denote $\boldsymbol{N} := (N, \ldots, N)$. The set $\mathbb{Z}^d$ is endowed with lexicographic ordering. For $\boldsymbol{i} \in \boldsymbol{I}$, we denote $\mathrm{S}\boldsymbol{i} := \boldsymbol{i} + (0, \ldots, 0, 1) \in \bar{\boldsymbol{I}}$; also, provided $\mathrm{S}\boldsymbol{i} \neq \boldsymbol{i}_+$, we denote $\tilde{\mathrm{S}}\boldsymbol{i} := \min\{\boldsymbol{i}' \in \boldsymbol{I} \mid \boldsymbol{i}' > \boldsymbol{i}\}$, i.e. $\tilde{\mathrm{S}}\boldsymbol{i}$ is the next element in $\boldsymbol{I}$ after $\boldsymbol{i}$. We define similarly $\mathrm{S}\boldsymbol{j}$ and $\tilde{\mathrm{S}}\boldsymbol{j}$ for $\boldsymbol{j} \in \boldsymbol{J}$.

We will build an example of $\vec{X}_{\mathbb{Z}^d}$ and $\vec{Y}_{\mathbb{Z}^d}$ satisfying the conditions of Equation (`KZ`), and for which $\rho(\vec{X}_{\mathbb{Z}^d};\vec{Y}_{\mathbb{Z}^d})$ shall be arbitrarily close to the right-hand side of (`LA`). In this construction, all the X_i's and Y_j's will be real, with $X_i \equiv 0$ for all $\boldsymbol{i} \notin \boldsymbol{I}$, resp. $Y_j \equiv 0$ for all $j \notin \boldsymbol{J}$ (so that the right-continuity conditions in (`KZ`) are automatically satisfied), and $(\vec{X}_{\boldsymbol{I}}, \vec{Y}_{\boldsymbol{J}})$ shall be a Gaussian vector. Our construction will rely on $|\boldsymbol{I}|+|\boldsymbol{J}|$ independent $\mathscr{N}(1)$ real random variables, denoted by resp. $(\xi_i)_{i\in \boldsymbol{I}}$ and $(\psi_j)_{j\in \boldsymbol{J}}$, of which the X_i's and the Y_j's shall be linear combinations. More generally, in the sequel we will speak of *linear random variables* to refer to variables being linear combinations of the ξ_i's and the ψ_j's. We will also say that a σ-algebra is *linear* when it is spanned (as a σ-algebra) by linear random variables.

In the sequel we set $\gamma_u \coloneqq \sum_{u'\geqslant u} \theta_u$, and denote by $\gamma_{-\infty}$ the common value of the γ_u's for $u \leqslant -\boldsymbol{N}$. We define the variable g in the following way:

$$g \coloneqq \sum_{j\in \boldsymbol{J}} \cos \gamma_{\mathrm{S}j-\boldsymbol{i}\_} \psi_j. \tag{LB}$$

We also define $Y_j \coloneqq \psi_j$ and $\mathscr{G}_j \coloneqq \sigma(\vec{Y}_{<j})$. Observe that, since all the $\cos\gamma_u$'s are nonzero, one has also $\mathscr{G}_j = \sigma(\vec{\psi}_{<j})$.

The definition of f is much more tricky: we define recursively the following:

$$\mathscr{F}_i \coloneqq \sigma(\vec{X}_{<i}) \qquad \text{for } i \in \bar{\boldsymbol{I}}; \tag{LC}$$

$$g^i_j \coloneqq g^{\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}} - \mathbb{E}(g \mid \mathscr{F}_i \vee \mathscr{G}_j) \qquad \text{for } j \in \boldsymbol{J}; \tag{LD}$$

$$\psi^i_j \coloneqq g^i_j \,/\, \mathrm{Var}^{1/2}(g^i_j) \qquad \text{(in the case } g^i_j \equiv 0,^{[‖]} \text{ see Remark 74);} \tag{LE}$$

$$f_i \coloneqq \sum_{j\in \boldsymbol{J}} \Big(\prod_{u<j-i} \cos\theta_u \times \sin\theta_{j-i}\Big)\psi^i_j + \prod_{u\in\mathbb{Z}^d} \cos\theta_u \times \xi_i \qquad \text{for } i \in \boldsymbol{I}; \tag{LF}$$

$$X_i \coloneqq f_i, \tag{LG}$$

and we set $f \coloneqq \sum_{i\in \boldsymbol{I}} f_i$. We also define $f^j_i \coloneqq f_i - \mathbb{E}(f_i \mid \mathscr{F}_i \vee \mathscr{G}_j)$ for $\boldsymbol{J} \in \bar{\boldsymbol{J}}$, $g^i \coloneqq g - \mathbb{E}(g \mid \mathscr{F}_i)$ for $i \in \bar{\boldsymbol{I}}$, $g^{ij} \coloneqq g - \mathbb{E}(g \mid \mathscr{F}_i \vee \mathscr{G}_j)$ for $i \in \bar{\boldsymbol{I}}, j \in \bar{\boldsymbol{J}}$; and we denote $\mathscr{G}^i_j \coloneqq \bigvee_{j'<j} \sigma(\psi^i_{j'})$ for $i \in \boldsymbol{I}, j \in \bar{\boldsymbol{J}}$.

Remark 70. ♧

Before tackling the core of the proof, we have to make two very important remarks and to prove one technical proposition:

Remark 71 (Induction via the S operator). Our definitions ensure that one always has $\mathscr{F}_{\tilde{\mathrm{S}}i} = \mathscr{F}_{\mathrm{S}i}$, $\mathscr{G}_{\tilde{\mathrm{S}}j} = \mathscr{G}_{\mathrm{S}j}$, $\mathscr{G}^i_{\tilde{\mathrm{S}}j} = \mathscr{G}^i_{\mathrm{S}j}$; and $f_i^{\tilde{\mathrm{S}}j} = f_i^{\mathrm{S}j}$, $g^{\tilde{\mathrm{S}}i} = g^{\mathrm{S}i}$, $g_j^{\tilde{\mathrm{S}}i} = g_j^{\mathrm{S}i}$, $g^{(\tilde{\mathrm{S}}i)j} = g^{(\mathrm{S}i)j}$, $g^{i(\tilde{\mathrm{S}}j)} = g^{i(\mathrm{S}j)}$, $\psi_j^{\tilde{\mathrm{S}}i} = \psi_j^{\mathrm{S}i}$. (Note that notation was chosen so that, as regards variables, you can always make the substitution of S̃ by S in *superscripts*). As a consequence, in the sequel of the proof, when we want to prove something by induction on $\boldsymbol{i} \in \boldsymbol{I}$ (resp. on $j \in \boldsymbol{J}$), we will do the following:

[‖] It will turn out that one has never $g^i_j \equiv 0$ for $j \in \boldsymbol{J}$ (cf. (`LK`) in Proposition 78); however, it is better to suppose a priori that such a case could be possible and to tell how one would define ψ^i_j in that case.

1. Proving the result for $\boldsymbol{i}_-$;

2. Proving that, if the result stands for $\boldsymbol{i}$, then it also stands for $\mathrm{S}\boldsymbol{i}$;

(3). *Implicitly* using that getting the result for $\mathrm{S}\boldsymbol{i}$ is tantamount to getting it for $\tilde{\mathrm{S}}\boldsymbol{i}$. ♧

Remark 72. If $\mathscr{H}$ is a linear σ-algebra generated by the (real) linear variables $(\varphi_k)_{k\in K}$, then, due to the Gaussian structure of our model, the linear variables which are $\mathscr{H}$-measurable are exactly the linear combinations of the φ_k's. For that reason, we may speak of the *dimension of a linear σ-algebra*, which actually refers to the dimension of the vector space made by the real linear variables being measurable w.r.t. that σ-algebra. The next proposition precisely states a dimensional result on certain linear σ-algebras. ♧

Proposition 73. *For $i \in \bar{\boldsymbol{I}}, j \in \bar{\boldsymbol{J}}$:*

$$\dim(\mathscr{F}_i \vee \mathscr{G}_j) = \big|\{i' \in \boldsymbol{I} \mid i' < i\}\big| + \big|\{j' \in \boldsymbol{J} \mid j' < j\}\big|. \tag{LH}$$

◇

Proof of Proposition 73. We prove the result by induction on $\boldsymbol{i}$. For $\boldsymbol{i} = \boldsymbol{i}_-$, one has $\mathscr{F}_{\boldsymbol{i}_-} \vee \mathscr{G}_j = \mathscr{O} \vee \mathscr{G}_j = \mathscr{G}_j$; and since we have noticed that $\mathscr{G}_j = \sigma(\vec{\psi}_{<j})$, the result follows in this case, as the family $(\xi_{i'})_{i'\in\boldsymbol{I}} \sqcup (\psi_{j'})_{j'\in\boldsymbol{J}}$ is linearly free by construction.

Now for $\boldsymbol{i} \in \boldsymbol{I}$, $\mathscr{F}_{\mathrm{S}i} \vee \mathscr{G}_j \underset{\mathrm{def}}{=} (\mathscr{F}_i \vee \sigma(X_i)) \vee \mathscr{G}_j = (\mathscr{F}_i \vee \mathscr{G}_j) \vee \sigma(X_i)$; so, since X_i is real, when passing from $\boldsymbol{i}$ to $\mathrm{S}\boldsymbol{i}$ the dimension increases by at most 1. But on the other hand, X_i cannot be $(\mathscr{F}_i \vee \mathscr{G}_j)$-measurable. Indeed, when you look at the decomposition of X_i in terms of $\xi_{i'}$'s and ψ'_{j},s, it has a nonzero ξ_i coefficient (because all the $\cos\gamma_u$'s are nonzero); while on the other hand, all the $Y_{j'}$'s can be expressed as linear combinations of the $\psi_{j'}$'s, and all the $X_{i'}$'s for $i' < i$ can be expressed as linear combinations of the $(\xi_{i'})_{i'<i}$ and the $(\psi_{j'})_{j'\in J}$, so that all those variables have zero ξ_i coefficients. In the end, we get that dimension increases exactly by 1 when stepping from $\boldsymbol{i} \in \boldsymbol{I}$ to $\mathrm{S}\boldsymbol{i}$; hence the result is valid for $\mathrm{S}\boldsymbol{i}$; which is enough to conclude by induction thanks to Remark 71. ♤

One of the interests of Proposition 73 is that is allows for a fully rigorous definition of ψ^i_j:

Remark 74. A more rigorous definition for ψ^i_j, also working when $g^i_j \equiv 0$, is the following. By Proposition 73, the σ-algebra $\mathscr{F}_i \vee \mathscr{G}_j$ is *strictly* included in $\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}$; so, since these σ-algebras are linear, there exists a nonzero linear variable which is $(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j})$-measurable and centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_j$; up to dividing this variable by its standard deviation, we may even assume that it has unit variance (hence $\mathscr{N}(1)$ law). Then we have the following alternative: either $g^i_j \equiv 0$ and then we take arbitrarily ψ^i_j to be <u>any</u> such $\mathscr{N}(1)$ variable; or $g^i_j \not\equiv 0$ and then we take $\psi^i_j \coloneqq g^i_j \,/\, \mathrm{Var}^{1/2}(g^i_j)$. The important point is that in both cases one will have $g^i_j = \mathrm{Var}^{1/2}(g^i_j)\times\psi^i_j$, and that ψ^i_j is a $\mathscr{N}(1)$ linear variable which is $(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j})$-measurable and centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_j$. ♧

Now let us turn to the core of the proof of Theorem 69. We will show that a certain number of properties hold for all i, j.[**] First, the following properties follow immediately from the general properties of Gaussian vectors:

Proposition 75.

(i) For all i, j, the σ-algebras $\mathscr{F}_i$, $\mathscr{G}_j$, $\mathscr{G}_j^i$ and $\mathscr{F}_i \vee \mathscr{G}_j$ all are linear; and the variables ψ_j^i, f, f_i, f_i^j, g, g^i, g_j^i, g^{ij}, X_i, Y_j all are linear.

(ii) For φ a (possibly multidimensional) linear variable, for $\mathscr{H}$ a linear σ-algebra, if φ is centred w.r.t. $\mathscr{H}$, then φ is independent from $\mathscr{H}$. So in particular, in the case φ is real, $\mathrm{Var}(\varphi \mid \mathscr{H})$ *is actually the constant* $\mathrm{Var}(\varphi)$.

(iii) For all i, j, conditionally to $\mathscr{F}_i \vee \mathscr{G}_j$, the variables ξ_i, f_i^j, g_j^i, g^{ij} and ψ_j^i are centred. ◇♤

The following property is not very complicated, but deserves a bit of attention:

Proposition 76.

(iv) For all i, the ψ_j^i's and ξ_i are $\mathscr{N}(1)$ independent variables, jointly independent from $\mathscr{F}_i$. ◇

Proof of Proposition 76. The fact that the ψ_j^i's and ξ_i all have $\mathscr{N}(1)$ laws follows from their construction; they being Gaussian follows from Proposition 75-(i); they being (jointly) independent from $\mathscr{F}_i$ follows from they being centred w.r.t. $\mathscr{F}_i$ (cf. Proposition 75-(iii), observing that $\mathscr{F}_i \vee \mathscr{G}_j \supseteq \mathscr{F}_i$), joint with the observation of Proposition 75-(ii). To prove that the ψ_j^i's are independent, observe that for $j < j'$, ψ_j^i is $(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j})$-measurable (hence a fortiori $(\mathscr{F}_i \vee \mathscr{G}_{j'})$-measurable), while $\psi_{j'}^i$ is centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_{j'}$: by Lemma 20 this implies that $\mathrm{Cov}(\psi_j^i, \psi_{j'}^i) = 0$, which is enough to prove independence in this Gaussian context. Finally, ξ_i being independent from the ψ_j^i's is trivial by construction. ♤

Now we state some identities between the variables that we have introduced:

Proposition 77. *For all i, j:*

(v) $f_i = f^{\mathscr{F}_{\mathrm{S}i}} - \mathbb{E}(f \mid \mathscr{F}_i)$.

(vi)

$$f_i^j = \sum_{j' \geqslant j} \Big(\prod_{u < j' - i} \cos\theta_u \times \sin\theta_{j'-i} \Big) \psi_{j'}^i + \prod_{u \in \mathbb{Z}^d} \cos\theta_u \times \xi_i. \qquad (\mathtt{LI})$$

(This is actually the same formula as (`LF`), *but where the sum over $\boldsymbol{J}$ was truncated to indices $j' \geqslant j$).*

(vii) $g_j^{i_-} = \cos\gamma_{\mathrm{S}j-i_-}\psi_j$*: in other words, the summands in* (`LB`) *are actually the $g_j^{i_-}$'s.*

[**] In the sequel I will merely say "for i" (resp. "for j") rather than specifying whether I mean "for $i \in I$" or "for $i \in \bar{I}$" (resp. …): the set to which i (resp. j) belongs shall always be clear from the context.

(viii) $g^{ij} = \sum_{j' \geqslant j} g^i_{j'}$. ◇

Proof of Proposition 77. Let us prove Item (v). Expanding the definition of f, one has

$$f^{\mathscr{F}_{\mathrm{S}i}} - \mathbb{E}(f \mid \mathscr{F}_i) = \sum_{i'} \big((f_{i'})^{\mathscr{F}_{\mathrm{S}i}} - \mathbb{E}(f_{i'} \mid \mathscr{F}_i) \big). \tag{LJ}$$

But $f_{i'}$ is $\mathscr{F}_{\mathrm{S}i'}$-measurable by construction; and it is centred w.r.t. $\mathscr{F}_{i'}$ since all the $\psi^{i'}_j$'s and $\xi_{i'}$ are (cf. Proposition 76). Therefore, one has $(f_i)^{\mathscr{F}_{\mathrm{S}i}} - \mathbb{E}(f_i \mid \mathscr{F}_i) = f_i - 0 = f_i$; while for $i' < i$ one has $(f_{i'})^{\mathscr{F}_{\mathrm{S}i}} - (f_{i'})^{\mathscr{F}_i} = f_{i'} - f_{i'} = 0$; and for $i' > i$ one has $\mathbb{E}(f_{i'} \mid \mathscr{F}_{\mathrm{S}i}) - \mathbb{E}(f_{i'} \mid \mathscr{F}_i) = 0 - 0 = 0$: so in the end, the right-hand side of (`LJ`) reduces to f_i, as announced.

Now let us prove Item (vi). By definition $f^j_i = f_i - \mathbb{E}(f_i \mid \mathscr{F}_i \vee \mathscr{G}_j)$; but we have defined f_i as a linear combination of the $\psi^i_{j'}$'s and ξ_i, so f^j_i will be a linear combination of the $\big(\psi^i_{j'} - \mathbb{E}(\psi^i_{j'} \mid \mathscr{F}_i \vee \mathscr{G}_j)\big)$'s and $\big(\xi_i - \mathbb{E}(\xi_i \mid \mathscr{F}_i \vee \mathscr{G}_j)\big)$. Now, for $j' < j$, by construction $\psi^i_{j'}$ is $(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j'})$-measurable, hence a fortiori $(\mathscr{F}_i \vee \mathscr{G}_j)$-measurable, so in this case $\psi^i_{j'} - \mathbb{E}(\psi^i_{j'} \mid \mathscr{F}_i \vee \mathscr{G}_j)$ is zero. On the other hand, for $j' \geqslant j$, $\psi^i_{j'}$ is centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_{j'}$ (cf. Proposition 75-(iii)), hence a fortiori w.r.t. $\mathscr{F}_i \vee \mathscr{G}_j$, so in that case $\psi^i_{j'} - \mathbb{E}(\psi^i_{j'} \mid \mathscr{F}_i \vee \mathscr{G}_j)$ is just $\psi^i_{j'}$. As regards ξ_i, it is the same case as for $\psi^i_{j' \geqslant j}$. So finally we have got Equation (`LI`).

Now let us turn to Item (vii). One has $\mathscr{F}_{\boldsymbol{i}_-} = \mathscr{O}$, so that $g^{\boldsymbol{i}_-}_j = g^{\mathscr{G}_{\mathrm{S}j}} - \mathbb{E}(g \mid \mathscr{G}_j)$. But remember that we noticed that $\mathscr{G}_j = \bigvee_{j' < j} \sigma(\psi_{j'})$, resp. $\mathscr{G}_{\mathrm{S}j} = \bigvee_{j' \leqslant j} \sigma(\psi_{j'})$, and that all the $\psi_{j'}$'s are independent: therefore, the result follows directly from the definition (`LB`) of g.

As regards Item (viii), by telescopic sum one has $\sum_{j' \geqslant j} g^i_{j'} = g^{\mathscr{F}_i \vee \mathscr{G}_{\boldsymbol{j}_+}} - \mathbb{E}(g \mid \mathscr{F}_i \vee \mathscr{G}_j)$; but g is $\mathscr{G}_{\boldsymbol{j}_+}$-measurable by construction, hence a fortiori $(\mathscr{F}_i \vee \mathscr{G}_{\boldsymbol{j}_+})$-measurable; so $g^{\mathscr{F}_i \vee \mathscr{G}_{\boldsymbol{j}_+}} = g$ and thus $g^{ij} = \sum_{j' \geqslant j} g^i_{j'}$. ♤

Now that the above propositions have enlightened our understanding of the structure of the model, let us turn to computational results:

Proposition 78. *For all* i, j*:*

$$\operatorname{Var}(g^i_j) = \cos^2 \gamma_{\mathrm{S}j-i}; \tag{LK}$$

$$\operatorname{Var}(f^j_i) = \prod_{u < j-i} \cos^2 \theta_u; \tag{LL}$$

$$\operatorname{Cov}(f^j_i, g^i_j) = \prod_{u < j-i} \cos \theta_u \times \sin \theta_{j-i} \times \cos \gamma_{\mathrm{S}j-i}. \tag{LM}$$

◇

Proof of Proposition 78. We shall prove the proposition *globally* by induction on $\boldsymbol{i}$. What I mean by this is that first we will establish (`LK`), (`LL`) and (`LM`) (in this order) for $i = \boldsymbol{i}_-$; then, using that we have (`LK`), (`LL`) and (`LM`) for $i = \boldsymbol{i}_-$, we will prove (`LK`), (`LL`) and (`LM`) (again in this order) for $i = \mathrm{S}\boldsymbol{i}_-$; then for $i = \mathrm{S}\,\tilde{\mathrm{S}}\boldsymbol{i}_-$; etc.

First let us observe in the case $i = \boldsymbol{i}_-$, (`LK`) follows directly from Proposition 77-(vii).

Let us turn to (`LL`). From (`LI`) in Proposition 77, and from the fact that the ψ^i_j's and ξ_i are independent $\mathcal{N}(1)$ variables (cf. Proposition 76), we deduce that

$$\mathrm{Var}(f^j_i) = \sum_{j' \geqslant j} \Big(\prod_{u < j'-i} \cos^2 \theta_u \times \sin^2 \theta_{j'-i} \Big) + \prod_{u \in \mathbb{Z}^d} \cos^2 \theta_u. \tag{LN}$$

Let us show that this quantity is actually equal to $\prod_{u<j-i} \cos^2 \theta_u$. We proceed by downgoing induction on j (i being supposed fixed). The result is clear for $j = \boldsymbol{j}_+$. Now assume that we have proved the result for Sj. In this case,

$$\begin{aligned}
\mathrm{Var}(f^j_i) &= \sum_{j' \geqslant j} \Big(\prod_{u < j'-i} \cos^2 \theta_u \times \sin^2 \theta_{j'-i} \Big) + \prod_{u \in \mathbb{Z}^d} \cos^2 \theta_u \\
&= \prod_{u<j-i} \cos^2 \theta_u \times \sin^2 \theta_{j-i} + \sum_{j' \geqslant \mathrm{S}j} \Big(\prod_{u < j'-i} \cos^2 \theta_u \times \sin^2 \theta_{j'-i} \Big) + \prod_{u \in \mathbb{Z}^d} \cos^2 \theta_u \\
&\underset{\mathrm{IH}}{=} \prod_{u<j-i} \cos^2 \theta_u \times \sin^2 \theta_{j-i} + \prod_{u < \mathrm{S}j-i} \cos^2 \theta_u \\
&= \Big(\prod_{u<j-i} \cos^2 \theta_u \Big) \times (\sin^2 \theta_{j-i} + \cos^2 \theta_{j-i}) = \prod_{u<j-i} \cos^2 \theta_u.
\end{aligned} \tag{LO}$$

Now let us turn to (`LM`). We can write f^j_i and g^i_j as explicit linear combinations of the $\psi^i_{j'}$'s and ξ_i:

$$f^j_i = \sum_{j' \geqslant j} \Big(\prod_{u<j'-i} \cos \theta_u \times \sin \theta_{j'-i} \Big) \psi^i_{j'} + \prod_{u \in \mathbb{Z}^d} \cos \theta_u \times \xi_i; \tag{LP}$$

$$g^i_j = \cos \gamma_{\mathrm{S}j-i} \psi^i_j, \tag{LQ}$$

where (`LP`) is just Proposition 77-(vi), and (`LQ`) comes from the fact that g^i_j and ψ^i_j are nonnegatively collinear by construction, joint with Equation (`LK`) for $\mathrm{Var}(g^i_j)$. But since we know that the ψ^i_j's and ξ_i are independent $\mathcal{N}(1)$ variables (cf. Proposition 76), expanding via the above equations yields the wished formula for $\mathrm{Cov}(f^j_i, g^i_j)$.

Now let us turn back to (`LK`), but this time in the case 'i' is of the form Si. To begin with, let us observe that $g^{\mathrm{S}i}_j = (g^{(\mathrm{S}i)j})^{\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}}$ and that $g^{(\mathrm{S}i)j} - g^{\mathrm{S}i}_j = g^{(\mathrm{S}i)(\mathrm{S}j)}$, so that, by the chain rule for variance:

$$\mathrm{Var}(g^{\mathrm{S}i}_j) = \mathrm{Var}(g^{(\mathrm{S}i)j}) - \mathrm{Var}(g^{(\mathrm{S}i)(\mathrm{S}j)}), \tag{LR}$$

and likewise $\mathrm{Var}(g^i_j) = \mathrm{Var}(g^{ij}) - \mathrm{Var}(g^{i(\mathrm{S}j)})$. Thanks to that we can investigate the value of $\mathrm{Var}(g^{\mathrm{S}i}_j)$ via values of the form $\mathrm{Var}(g^{i'j'})$. In particular, we will get interested in $\mathrm{Var}(g^{(\mathrm{S}i)j})$.

Let us observe that it is a direct consequence of the definitions that

$$g^{(\mathrm{S}i)j} = g^{ij} - \mathbb{E}(g^{ij} \mid \mathscr{F}_{\mathrm{S}i} \vee \mathscr{G}_j). \tag{LS}$$

Now, let us reason conditionally to some $\{\vec{X}_{<i} \equiv \vec{x}_{<i}$ and $\vec{Y}_{<j} \equiv \vec{y}_{<j}\}$ for a few lines: in this context (`LS`) translates into saying that $g^{(\mathrm{S}i)j} = g^{ij} - \mathbb{E}(g^{ij} \mid \sigma(X_i))$. By the chain rule for variance, this implies that

$$\mathrm{Var}(g^{(\mathrm{S}i)j}) = \mathrm{Var}(g^{ij}) - \mathrm{Var}((g^{ij})^{\sigma(X_i)}). \tag{LT}$$

But due to the Gaussian context, and remembering that $X_i \underset{\text{def}}{=} f_i$ (and that f_i is nonzero by construction), one has

$$\mathrm{Var}((g^{ij})^{\sigma(X_i)}) = \left(\frac{\mathrm{Cov}(g^{ij}, f_i)}{\mathrm{Var}(f_i)}\right)^2, \tag{LU}$$

so in the end

$$\mathrm{Var}(g^{(\mathrm{S}i)j}) = \mathrm{Var}(g^{ij}) - \left(\frac{\mathrm{Cov}(g^{ij}, f_i)}{\mathrm{Var}(f_i)}\right)^2. \tag{LV}$$

But remember that the above formula was derived under conditioning by $\{\vec{X}_{<i} \equiv \vec{x}_{<i}$ and $\vec{Y}_{<j} \equiv \vec{y}_{<j}\}$. If we want to go back to unconditional formalism, using Proposition 75-(ii) we just have to replace $g^{(\mathrm{S}i)j}$, g^{ij} and f_i by their recentred-under-$(\mathscr{F}_i \vee \mathscr{G}_j)$ versions (i.e., "φ" has to be replaced by "$\varphi - \mathbb{E}(\varphi \mid \mathscr{F}_i \vee \mathscr{G}_j)$"): the corresponding variables are resp. $g^{(\mathrm{S}i)j}$ and g^{ij} themselves (cf. Proposition 75-(iii)), and f_i^j. So, in unconditioned terms:

$$\mathrm{Var}(g^{(\mathrm{S}i)j}) = \mathrm{Var}(g^{ij}) - \left(\frac{\mathrm{Cov}(g^{ij}, f_i^j)}{\mathrm{Var}(f_i^j)}\right)^2. \tag{LW}$$

In this formula we already know $\mathrm{Var}(f_i^j)$ by induction hypothesis; and it will turn out that the value of $\mathrm{Var}(g^{ij})$ itself does not matter. As regards $\mathrm{Cov}(g^{ij}, f_i^j)$, using that $g^{ij} = \sum_{j' \geqslant j} g^i_{j'}$ (cf. Proposition 77-(viii)):

$$\mathrm{Cov}(g^{ij}, f_i^j) = \sum_{j' \geqslant j} \mathrm{Cov}(g^i_{j'}, f_i^j) = \sum_{j' \geqslant j} \mathrm{Cov}(g^i_{j'}, f_i^{j'}), \tag{LX}$$

where the last equality comes from the fact that $g^i_{j'}$ is centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_{j'}$ (cf. Proposition 75-(iii)) while $f_i^{j'} - f_i^j = \mathbb{E}(f_i \mid \mathscr{F}_i \vee \mathscr{G}_j) - \mathbb{E}(f_i \mid \mathscr{F}_i \vee \mathscr{G}_{j'})$ is $(\mathscr{F}_i \vee \mathscr{G}_{j'})$-measurable. Hence, using (`LM`) in the "i" case:

$$\mathrm{Cov}(g^{ij}, f_i^j) = \sum_{j' \geqslant j} \Bigl(\prod_{u < j'-i} \cos\theta_u \times \sin\theta_{j'-i} \cos\gamma_{\mathrm{S}j'-i} \Bigr). \tag{LY}$$

Now I claim that the right-hand side of (`LY`) is actually equal to

$$\prod_{u < j-i} \cos\theta_u \times \sin\gamma_{j-i}. \tag{LZ}$$

We prove that by downgoing induction on j. For $j = \boldsymbol{j}_+$, both the right-hand side of (`LY`) and (`LZ`) are zero (because for all $\boldsymbol{i} \in \boldsymbol{I}$ one has $\boldsymbol{j}_+ - \boldsymbol{i} > \boldsymbol{N}$). Now assume we have the result for $\mathrm{S}j$; then:

$$\sum_{j' \geqslant j} \Big(\prod_{u < j'-i} \cos\theta_u \times \sin\theta_{j'-i} \cos\gamma_{\mathrm{S}j'-i} \Big) =$$

$$\prod_{u<j-i} \cos\theta_u \times \sin\theta_{j-i} \cos\gamma_{\mathrm{S}j-i} + \sum_{j' \geqslant \mathrm{S}j} \Big(\prod_{u<j'-i} \cos\theta_u \times \sin\theta_{j'-i} \cos\gamma_{\mathrm{S}j'-i} \Big)$$

$$\underset{\mathrm{IH}}{=} \prod_{u<j-i} \cos\theta_u \times \sin\theta_{j-i} \cos\gamma_{\mathrm{S}j-i} + \prod_{u<\mathrm{S}j-i} \cos\theta_u \times \sin\gamma_{\mathrm{S}j-i}$$

$$= \prod_{u<j-i} \cos\theta_u \times \underbrace{(\sin\theta_{j-i} \cos\gamma_{\mathrm{S}j-i} + \cos\theta_{j-i} \sin\gamma_{\mathrm{S}j-i})}_{= \sin(\theta_{j-i} + \gamma_{\mathrm{S}j-i}) \underset{\mathrm{def}}{=} \sin(\gamma_{j-i})} = \prod_{u<j-i} \cos\theta_u \times \sin\gamma_{j-i}, \qquad (\mathtt{MA})$$

so we have indeed the wanted identity for all j.

Thanks to that identity, (`LW`) yields that

$$\mathrm{Var}(g^{(\mathrm{S}i)j}) = \mathrm{Var}(g^{ij}) - \sin^2\gamma_{j-i}, \qquad (\mathtt{MB})$$

and likewise $\mathrm{Var}(g^{(\mathrm{S}i)(\mathrm{S}j)}) = \mathrm{Var}(g^{i(\mathrm{S}j)}) - \sin^2\gamma_{\mathrm{S}j-i}$. It ensues that

$$\begin{aligned}
\mathrm{Var}(g_j^{\mathrm{S}i}) \underset{(\mathtt{LR})}{=} {} & \mathrm{Var}(g^{(\mathrm{S}i)j}) - \mathrm{Var}(g^{(\mathrm{S}i)(\mathrm{S}j)}) \\
& \underset{(\mathtt{MB})}{=} \mathrm{Var}(g^{ij}) - \sin^2\gamma_{j-i} - \mathrm{Var}(g^{i(\mathrm{S}j)}) + \sin^2\gamma_{\mathrm{S}j-i} \\
& = \underbrace{(\mathrm{Var}(g^{ij}) - \mathrm{Var}(g^{i(\mathrm{S}j)}))}_{\underset{(\mathtt{LR})}{=} \mathrm{Var}(g_j^i) \underset{\mathrm{IH}}{=} \cos^2\gamma_{\mathrm{S}j-i}} + \sin^2\gamma_{\mathrm{S}j-i} - \sin^2\gamma_{j-i} \\
& = 1 - \sin^2\gamma_{j-i} = \cos^2\gamma_{j-i} = \cos^2\gamma_{\mathrm{S}j-\mathrm{S}i}, \qquad (\mathtt{MC})
\end{aligned}$$

which is the wanted result. ♤

The next proposition shows that our model satisfies indeed the wanted assumptions concerning $\rho^\dagger$-mixing:

Proposition 79. *For all* i, j,

$$\mathrm{\rho}^\dagger_{(\vec{X}_{<i}, \vec{Y}_{<j})}(X_i; Y_j) = \rho_{j-i}. \qquad (\mathtt{MD})$$

◇

Proof of Proposition 79. We will actually prove that for <u>all</u> $\vec{x}_{<i}, \vec{y}_{<j}$, one has exactly

$$\mathrm{\rho}(X_i; Y_j \mid \vec{X}_{<i} \equiv \vec{x}_{<i} \text{ and } \vec{Y}_{<j} = \vec{y}_{<j}) = \rho_{j-i}. \qquad (\mathtt{ME})$$

For the sequel, let us shorthand $\mathbb{P}(\bullet \mid \vec{X}_{<i} \equiv \vec{x}_{<i} \text{ and } \vec{Y}_{<j} = \vec{y}_{<j}) =: \mathbb{P}_{\mathrm{cond}}(\bullet)$. First, I claim that under $\mathbb{P}_{\mathrm{cond}}$, X_i is in bijection with f_i^j, resp. Y_j is in bijection with g_j^i, so that $\mathrm{\rho}_{\mathrm{cond}}(X_i; Y_j) = \mathrm{\rho}_{\mathrm{cond}}(f_i^j; g_j^i)$. As regards X_i being in bijection with f_i^j, this

is fairly obvious: remember indeed that $\boldsymbol{X}_i \underset{\text{def}}{=} f_i$ and that $f_i^j \underset{\text{def}}{=} f_i - \mathbb{E}(f_i \mid \mathscr{F}_i \vee \mathscr{G}_j)$, so that under $\mathbb{P}_{\text{cond}}$ the variables $\boldsymbol{X}_i$ and f_i^j just differ by an additive constant.

Now we want to prove that $\boldsymbol{Y}_j$ is in bijection with g_j^i under $\mathbb{P}_{\text{cond}}$: actually we will rather prove that it is in bijection with ψ_j^i (which is in bijection with g_j^i by construction, as (`LK`) in Proposition 78 established that $\operatorname{Var}(g_j^i) \neq 0$): proving such a thing, for all the values of $\vec{x}_{<i}$ and $\vec{y}_{<j}$, is tantamount to proving that $(\mathscr{F}_i \vee \mathscr{G}_j) \vee \sigma(\boldsymbol{Y}_j) = (\mathscr{F}_i \vee \mathscr{G}_j) \vee \sigma(\psi_j^i)$. But that in turn will be a consequence of the identity

$$\mathscr{F}_i \vee \mathscr{G}_j^i = \mathscr{F}_i \vee \mathscr{G}_j: \tag{MF}$$

indeed, $(\mathscr{F}_i \vee \mathscr{G}_j) \vee \sigma(\boldsymbol{Y}_j) = \mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}$ by definition, while (`MF`) ensures that $(\mathscr{F}_i \vee \mathscr{G}_j) \vee \sigma(\psi_j^i) \underset{(\mathtt{MF})}{=} \mathscr{F}_i \vee \mathscr{G}_j^i \vee \sigma(\psi_j^i) \underset{\text{def}}{=} \mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}^i \underset{(\mathtt{MF})}{=} \mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}$.

We prove (`MF`) by induction on j. The result is automatic for $j = \boldsymbol{j}_-$ since then $\mathscr{G}_{\boldsymbol{j}_-}^i = \mathscr{G}_{\boldsymbol{j}_-} = \mathscr{O}$. For '$j$' of the form $\mathrm{S}j$, observe first that one has automatically by construction that $\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}^i \subseteq \mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}$: so, proving that $\dim(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}^i) \geqslant \dim(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j})$ will suffice. Since by induction hypothesis one has $\mathscr{F}_i \vee \mathscr{G}_j^i = \mathscr{F}_i \vee \mathscr{G}_j$, and that by Proposition 73 one has $\dim(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}) = \dim(\mathscr{F}_i \vee \mathscr{G}_j) + 1$, it just remains to prove that $\dim(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}^i) > \dim(\mathscr{F}_i \vee \mathscr{G}_j^i)$. Now $\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}^i \underset{\text{def}}{=} \mathscr{F}_i \vee \mathscr{G}_j^i \vee \sigma(\psi_j^i) \underset{\text{IH}}{=} (\mathscr{F}_i \vee \mathscr{G}_j) \vee \sigma(\psi_j^i)$; so to prove that $\dim(\mathscr{F}_i \vee \mathscr{G}_{\mathrm{S}j}^i) > \dim(\mathscr{F}_i \vee \mathscr{G}_j^i)$, it finally suffices to prove that ψ_j^i is not $(\mathscr{F}_i \vee \mathscr{G}_j)$-measurable: but this is a consequence from ψ_j^i being centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_j$ while having nonzero variance!

So at this point we have established that $\rho_{\text{cond}}(\boldsymbol{X}_i; \boldsymbol{Y}_j) = \rho_{\text{cond}}(f_i^j; g_j^i)$. But f_i^j and g_j^i are linear random variables which are centred w.r.t. $\mathscr{F}_i \vee \mathscr{G}_j$ (cf. Proposition 75-(iii)), so by Proposition 75-(ii) $\operatorname{Law}_{\text{cond}}(f_i^j, g_j^i) = \operatorname{Law}(f_i^j, g_j^i)$, and therefore $\rho_{\text{cond}}(f_i^j; g_j^i) = \rho(f_i^j; g_j^i)$. But f_i^j and g_j^i are real variables forming a Gaussian vector, so by Proposition 9

$$\rho(f_i^j; g_j^i) = \frac{|\operatorname{Cov}(f_i^j, g_j^i)|}{\operatorname{Var}^{1/2}(f_i^j)\operatorname{Var}^{1/2}(g_j^i)}, \tag{MG}$$

which by Proposition 78 is equal to $\sin\theta_{j-i}$, that is, to ρ_{j-i}. ♤

The final bunch of intermediate results consists in computing the variances and covariance of f and g:

Proposition 80.

$$\operatorname{Var}(f) = |\boldsymbol{I}|; \tag{MH}$$

$$\operatorname{Var}(g) \leqslant |\boldsymbol{J}|; \tag{MI}$$

$$\operatorname{Cov}(f, g) = |\boldsymbol{I}| \sin\gamma_{-\infty}. \tag{MJ}$$

◇

Proof of Proposition 80. First let us observe that $f_i = f^{\mathscr{F}_{\mathrm{S}i}} - \mathbb{E}(f \mid \mathscr{F}_i)$. Because of Proposition 77-(v), the chain rule for variance yields that $\operatorname{Var}(f) = \sum_{i \in \boldsymbol{I}} \operatorname{Var}(f_i)$; but for all $\boldsymbol{i}$ one has $\mathrm{Var}(f_i) = 1$ because of (`LL`) in Proposition 78 applied with $j = \boldsymbol{j}_-$; so (`MH`) follows.

As regards (`MI`), the very definition of g ensures that $\mathrm{Var}(g) = \sum_{j \in J} \cos^2 \gamma_{\mathrm{S}j-\boldsymbol{i}_-}$, where obviously each $\cos^2 \gamma_{\mathrm{S}j-\boldsymbol{i}_-}$ is less than 1.

As regards (`MJ`), using again Proposition 77-(v), the chain rule for covariance yields that

$$\mathrm{Cov}(f, g) = \sum_{i \in I} \mathrm{Cov}(f_i, g^{\mathscr{F}_{\mathrm{S}i}} - g^{\mathscr{F}_i}) = \sum_{i \in I} \mathrm{Cov}(f_i, g^i), \tag{MK}$$

where the last equality comes from the fact that f_i is $\mathscr{F}_{\mathrm{S}i}$-measurable while $g^i - g^{\mathscr{F}_{\mathrm{S}i}} + g^{\mathscr{F}_i} \underset{\mathrm{def}}{=} g - g^{\mathscr{F}_{\mathrm{S}i}}$ is centred w.r.t. $\mathscr{F}_{\mathrm{S}i}$. But f_i and g^i coincide by construction with resp. $f_i^{\boldsymbol{j}_-}$ and $g^{i\boldsymbol{j}_-}$; so we can use the computation of $\mathrm{Cov}(f_i^j, g^{ij})$ that we did in the proof of Proposition 78 (see (`LY`) and the following equations) to get that

$$\mathrm{Cov}(f_i, g^i) = \prod_{u < \boldsymbol{j}_- - i} \cos \theta_u \times \sin \gamma_{\boldsymbol{j}_- - i}, \tag{ML}$$

which is always equal to $\sin \gamma_{-\infty}$, since for all $\boldsymbol{i} \in \boldsymbol{I}$ one has $\boldsymbol{j}_- - \boldsymbol{i} \leqslant -\boldsymbol{N}$. ♤

So, since f and g are resp. $\vec{X}_{\mathbb{Z}^d}$- and $\vec{Y}_{\mathbb{Z}^d}$-measurable, Proposition 80 implies that

$$\mathrm{Opt}^{\mathrm{m/m}}(\vec{\rho}) \underset{(\mathtt{MD})}{\geqslant} \rho(\vec{X}_{\mathbb{Z}^d}; \vec{Y}_{\mathbb{Z}^d}) \geqslant \frac{\mathrm{Cov}(f, g)}{\mathrm{Var}^{1/2}(f)\,\mathrm{Var}^{1/2}(g)} \geqslant \sqrt{|\boldsymbol{I}| \,/\, |\boldsymbol{J}|}\, \sin \gamma_{-\infty}. \tag{MM}$$

In (`MM`), the right-hand side converges to $\sin \gamma_{-\infty}$ as $j_+ - j_-$ tends to infinity, since $|\boldsymbol{I}| = (j_+ - j_- - 2N)^d$ while $|\boldsymbol{J}| = (j_+ - j_-)^d$. So we have proved that $\mathrm{Opt}^{\mathrm{m/m}}(\vec{\rho}) \geqslant \sin \gamma_{-\infty}$, which concludes the proof of Theorem 69. ♤

Chapter 3

Examples of applications

Now we are going to see how the results of Chapter 2 can be applied to various concrete models of particle systems. We will tackle three such model families:

- First, as a response to § 1.2 in the introduction, in § 3.2 we will establish fairly general ρ-mixing properties for Ising-like models: what we shall prove, in terms of the vocabulary used in the specialized literature (see Definition 82), is that these Ising-like models satisfy *interlaced ρ-mixing* at a certain speed, with moreover "$\rho^*(1) < 1$".
- In § 3.3 we will also look at certain models of statistical mechanics with *continuous* state space, for which we will similarly prove interlaced ρ-mixing properties.
- Finally, from a quite different perspective, in § 3.4 we will turn to establishing ρ-mixing properties for separation along the time dimension: more precisely, investigating a toy-model situation, we will show how tensorization of ρ-mixing can help to prove some results of *hypocoercivity* type.

These three sections will be preceded in § 3.1 by an explanation on general mechanisms to apply tensorization of ρ-mixing for such statistical mechanics models.

Remark 81. In this chapter we will often encounter $\rho^\ddagger$-mixing coefficients. For such coefficients, in accordance with Remark 35, we will not in general specify explicitly the family of random variables w.r.t. which $\rho^\ddagger$-mixing is considered: indeed, every time, that family will merely be the set of the "elementary" random variables that the particle system is made of (for instance, individual spins for Ising-like models; individual positions and momenta at a given time for the Langevin dynamics; etc.). ♧

3.1 A general machinery for statistical mechanics systems

3.1.1 Vocabulary of ρ-mixing systems

In this section, we are considering a family of random variables $(\sigma_i)_{i\in\mathbb{V}}$ (which variables we will call *spins*, by analogy with the Ising model) indexed by some "physical space" $\mathbb{V}$. The physical space is endowed with some distance *dist*. First, let us introduce some vocabulary. The first definition is customary in specialized literature:

Definition 82 (Interlaced ρ-mixing, [8, Equation 2.7])**.** For $\delta > 0$, the *interlaced ρ-mixing coefficient*—or *ρ^*-mixing coefficient*—at distance δ for the system, denoted by $\rho^*(\delta)$, is defined as

$$\rho^*(\delta) \coloneqq \sup\{\mathrm{\rho}(\vec{\sigma}_{\boldsymbol{I}}; \vec{\sigma}_{\boldsymbol{J}}) \mid \mathit{dist}(\boldsymbol{I}, \boldsymbol{J}) \geqslant \delta\}, \tag{MN}$$

where $\boldsymbol{I}$ and $\boldsymbol{J}$ refer to subsets of $\mathbb{V}$. When $\rho^*(\delta) \to 0$ as $\delta \to \infty$, the system is said to be *ρ^*-mixing*.

Also, in the context of this monograph, it will be natural to introduce a *conditional* version of $\rho^*(\delta)$:

$$\rho^{*\ddagger}(\delta) \coloneqq \sup\{\mathrm{\rho}^{\ddagger}(\vec{\sigma}_{\boldsymbol{I}}; \vec{\sigma}_{\boldsymbol{J}}) \mid \mathit{dist}(\boldsymbol{I}, \boldsymbol{J}) \geqslant \delta\}. \tag{MO}$$

♡

The following definitions will be handy to state the results of this chapter:

Definition 83 (Polynomial and exponential ρ-mixing)**.** For $\gamma > 0$, we will say that a function $f \colon \mathbb{R}_+ \to \mathbb{R}$ is *γ-polynomially decaying*, resp. *γ-exponentially decaying*, when $f(\delta) = \mathrm{O}(\delta^{-\gamma+\mathrm{o}(1)})$ as $\delta \to \infty$, resp. when $f(\delta) = \mathrm{O}(\mathrm{e}^{-(\gamma-\mathrm{o}(1))\delta})$. The function will be said to be *polynomially* (resp. *exponentially*) *decaying* when it is γ-polynomially (resp. γ-exponentially) decaying for some $\gamma > 0$.

Our particle system will be called *(γ-)polynomially* (resp. *exponentially*) *ρ^*-mixing* when its function $\rho^*(\bullet)$ decays that way. ♡

Definition 84 (Immediate ρ-mixing)**.** We will say that our system is *immediately ρ^*-mixing* when $\rho^*(\delta) < 1$ as soon as $\delta > 0$. ♡

The next and final definition will be handy to state the assumption necessary to get a ρ^*-mixing result by tensorization:

Definition 85 (Pairwise ρ-mixing)**.** In the same context as above, the *pairwise $\rho^\ddagger$-mixing coefficient*—or *$\tilde{\rho}^\ddagger$-mixing coefficient*—at distance δ for the system is defined as

$$\tilde{\rho}^{\ddagger}(\delta) \coloneqq \sup\{\mathrm{\rho}^{\ddagger}(\sigma_i; \sigma_j) \mid |j - i| \geqslant \delta\}, \tag{MP}$$

where i and j are *elements* (not subsets) of $\mathbb{V}$. As above, we will also speak of polynomial, exponential or immediate pairwise ρ-mixing. ♡

3.1.2 Geometry of the physical space

The results of the current section will require a certain assumption on the geometry of $\mathbb{V}$:

Assumption 86. We suppose that any ball of $\mathbb{V}$ of radius $\delta < \infty$ only contains a finite number of points, which number is bounded uniformly by some value only depending on δ. Moreover, we suppose that $\mathbb{V}$ is uniformly discrete, i.e. that it is ε-separated for some $\varepsilon > 0$ (which means that for any distinct $i, j \in \mathbb{V}$ one has $\mathit{dist}(i, j) \geqslant \varepsilon$). ♡

Assumption 86 allows us to introduce two functions $\boldsymbol{n}(\bullet)$ and $\boldsymbol{\delta}_\bullet$ which somehow "sum up" the geometric properties of $\mathbb{V}$:

Notation 87.

- For $\delta \in \mathbb{R}_+$, we let
$$\boldsymbol{n}(\delta) \coloneqq \sup_{i\in\mathbb{V}} \bigl|\{j \in \mathbb{V} \mid \mathit{dist}(i,j) < \delta\}\bigr| : \tag{MQ}$$
in other words, $\boldsymbol{n}(\delta)$ counts the maximal number of points in an open ball of $\mathbb{V}$ of radius δ. By Assumption 86, one has $\boldsymbol{n}(\delta) < \infty$ for all δ.

- For $n \in \mathbb{N}$, we let
$$\boldsymbol{\delta}_n \coloneqq \inf\{\delta \in \mathbb{R}_+ \mid \boldsymbol{n}(\delta) > n\} \in \bar{\mathbb{R}}_+, \tag{MR}$$
which is the minimum possible distance for the $\boldsymbol{n}$-th closest neighbour of a point of $\mathbb{V}$ (counting the point itself as its 0-th closest neighbour). One has obviously $\boldsymbol{\delta}_0 = 0$; and Assumption 86 ensures that $\boldsymbol{\delta}_1 > 0$ and that $\boldsymbol{\delta}_n \to \infty$ as $n \to \infty$. ♡

Assumption 86 is satisfied by most of the natural geometries that one would like to consider:

Examples 88.

(i) Any finite metric space satisfies Assumption 86.

(ii) For $d \geqslant 1$, the space $\mathbb{Z}^d$ (endowed e.g. with the ℓ^1 graph distance) satisfies Assumption 86, with $\boldsymbol{n}(\delta) = \Theta(\delta^d)$ as $\delta \to \infty$, hence $\boldsymbol{\delta}_n = \Theta(n^{1/d})$ as $n \to \infty$.

(iii) It is straightforward to see that the product of two metric spaces satisfying Assumption 86 still satisfies Assumption 86. Therefore, as a corollary of the two previous items, for $d \in \mathbb{N}^*$ and $\mathscr{C}$ finite, any $\mathbb{Z}^d \times \mathscr{C}$ satisfies Assumption 86: essentially, this means that "flat geometries" satisfy this assumption.

(iv) For $k \in \rrbracket 1, \infty \llbracket$, the $(k+1)$-regular tree satisfies Assumption 86, with $\boldsymbol{n}(\delta) = \Theta(k^\delta)$, hence $\boldsymbol{\delta}_n = \log_k(n) + \mathrm{O}(1)$.

(v) For $d \in \rrbracket 1, \infty \llbracket$, $\varepsilon > 0$, any ε-separated subset of the hyperbolic space $\mathbf{H}^d$ satisfies Assumption 86, with $\boldsymbol{n}(\delta) = \exp(\mathrm{O}(\delta))$, hence $\boldsymbol{\delta}_n = \Omega(\log(n))$.

(vi) In addition to all the previous cases, it is straightforward to see that Assumption 86 is preserved by taking a subspace. ♧

Assumption 86 has two main implications:

Lemma 89. *For $\mathbb{V}$ satisfying Assumption 86, using the notation of Definition 87, the following holds: if $\rho\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a decreasing function, then for all $i \in \mathbb{V}$,*
$$\sum_{j\in\mathbb{V}} \rho\bigl(\mathit{dist}(i,j)\bigr) \leqslant \sum_{n\in\mathbb{N}} \rho(\boldsymbol{\delta}_n). \tag{MS}$$
◇

Proof. Order the points of $\mathbb{V}$ by increasing distance to $\boldsymbol{i}$: then if $\boldsymbol{j}$ is point $\#\boldsymbol{n}$, $\mathit{dist}(\boldsymbol{i}, \boldsymbol{j}) \geqslant \boldsymbol{\delta}_n$, hence $\rho\big(\mathit{dist}(\boldsymbol{i}, \boldsymbol{j})\big) \leqslant \rho(\boldsymbol{\delta}_n)$. ♤

Lemma 90. *If $\mathbb{V}$ satisfies Assumption 86, then for any $\delta < \infty$ it is possible to find a finite partition $\mathbb{V} = \mathbb{V}_0 \sqcup \mathbb{V}_1 \sqcup \ldots \sqcup \mathbb{V}_{n-1}$ such that all the $\mathbb{V}_c$'s are δ-separated (i.e., two distinct points of $\mathbb{V}_c$ are always at distance $\geqslant \delta$ from each other).* ◇

Proof. Consider the graph obtained by connecting two distinct points of $\mathbb{V}$ whenever their distance is $< \delta$: this graph has valency bounded by $\boldsymbol{n}(\delta) - 1$, so it admits a $\boldsymbol{n}(\delta)$-colouring: then, partition $\mathbb{V}$ according to the colours of the vertices. ♤

3.1.3 The abstract result for interlaced ρ-mixing

All that vocabulary being set up, let us state the main result this section is about:

Lemma 91 (Machinery for interlaced mixing)**.** *For $\mathbb{V}$ satisfying Assumption 86, let $(\sigma_i)_{i\in\mathbb{V}}$ be a system of random variables.*

(i) Assume that $\sum_{n\in\mathbb{N}} \tilde{\rho}^{\ddagger}(\boldsymbol{\delta}_n) < \infty$. Then the system has interlaced $\rho^{\ddagger}$-mixing, with

$$\rho^{*\ddagger}(\delta) \leqslant \sum_{n \geqslant \boldsymbol{n}(\delta)} \tilde{\rho}^{\ddagger}(\boldsymbol{\delta}_n). \tag{MT}$$

In particular, in the case $\mathbb{V} = \mathbb{Z}^d$, pairwise $\rho^{\ddagger}$-mixing with γ-polynomial decay (for $\gamma > d$), resp. with γ-exponential decay (for $\gamma > 0$), implies interlaced ρ-mixing with $(\gamma - d)$-polynomial decay, resp. with γ-exponential decay.

(ii) Moreover, if, in addition to the previous assumption, the pairwise $\rho^{\ddagger}$-mixing behaviour of the system is immediate, then so is its interlaced $\rho^{\ddagger}$-mixing behaviour. ◇

Remark 92. The reason why I claimed that (`MT`) establishes interlaced $\rho^{\ddagger}$-mixing is because its right-hand side tends to 0 as $\delta \to \infty$. The rationale behind that fact depends on whether $\mathbb{V}$ is finite or infinite: in the case $\mathbb{V}$ is infinite, one must have $\boldsymbol{n}(\delta) \to \infty$ as $\delta \to \infty$, and thus convergence to 0 of the right-hand side of (`MT`) comes from convergence of $\sum_{n\in\mathbb{N}} \tilde{\rho}^{\ddagger}(\boldsymbol{\delta}_n)$; while in the (not very interesting) case $\mathbb{V}$ is finite, one has $\boldsymbol{n}(\delta) = |\mathbb{V}|$ for δ large enough, and since $\boldsymbol{\delta}_{|\mathbb{V}|}$ is obviously infinity, the right-hand side of (`MT`) is zero in that regime. ♧

Proof of Lemma 91. It will be enough for our goal to prove *unconditioned* ρ^*-mixing results: then indeed, the $\rho^{*\ddagger}$-mixing results can be deduced by appropriate conditioning. Let us first prove Assertion (i). Let $\delta > 0$ and let $\boldsymbol{I}, \boldsymbol{J} \subseteq \mathbb{V}$ with $\mathit{dist}(\boldsymbol{I}, \boldsymbol{J}) \geqslant \delta$: since $\delta > 0$, $\boldsymbol{I}$ and $\boldsymbol{J}$ are disjoint. Now, denoting by ∂ a cemetery value, for $\boldsymbol{k} \in \mathbb{V}$ we define

$$X_k := \begin{cases} \sigma_k & \text{if } \boldsymbol{k} \in \boldsymbol{I}; \\ \partial & \text{otherwise,} \end{cases} \qquad Y_k := \begin{cases} \sigma_k & \text{if } \boldsymbol{k} \in \boldsymbol{J}; \\ \partial & \text{otherwise.} \end{cases} \tag{MU}$$

Then, for $i, j \in \mathbb{V}$, one has

$$\rho^{\ddagger}(X_i; Y_j) \leqslant \begin{cases} \tilde{\rho}^{\ddagger}(dist(i,j)) & \text{if } i \in I \text{ and } j \in J; \\ 0 & \text{otherwise:} \end{cases} \tag{MV}$$

in particular, $\rho^{\ddagger}(X_i; Y_j) \leqslant \mathbf{1}_{dist(i,j) \geqslant \delta} \tilde{\rho}^{\ddagger}(dist(i,j))$. Therefore, using Theorem 39:

$$\rho(\vec{\sigma}_I; \vec{\sigma}_J) = \rho(\vec{X}_{\mathbb{V}}; \vec{Y}_{\mathbb{V}}) \leqslant \left|\!\left|\!\left| \left(\left(\mathbf{1}_{dist(i,j) \geqslant \delta} \tilde{\rho}^{\ddagger} \big(dist(i,j) \big) \right) \right)_{i,j \in \mathbb{V}} \right|\!\right|\!\right|. \tag{MW}$$

To bound the right-hand side of (MW), we introduce a random (sub-)walk $(W_t)_{t \in \mathbb{N}}$ on $\mathbb{V}$ such that at each step, a particle at i either jumps onto some distant vertex j or dies, the probability of jumping from i to j being

$$\mathbb{P}(W_{t+1} = j \mid W_t = i) = Z^{-1} \mathbf{1}_{dist(i,j) \geqslant \delta} \tilde{\rho}^{\ddagger}(dist(i,j)), \tag{MY}$$

with

$$Z \coloneqq \sum_{n \geqslant \boldsymbol{n}(\delta)} \tilde{\rho}^{\ddagger}(\boldsymbol{\delta}(n)) \tag{MZ}$$

—note that this choice of Z ensures that $\sum_{j \in \mathbb{V}} \mathbb{P}(W_{t+1} = j \mid W_t = i) \leqslant 1$. Then, the counting measure on $\mathbb{V}$ is a super-equilibrium for the random walk: that is, applying the sub-Markov kernel of the walk to the counting measure on $\mathbb{V}$ yields a smaller measure.

But, the operator in the right-hand side of (MW) is nothing but Z times the kernel of the above walk. And, as the counting measure on $\mathbb{V}$ is a super-equilibrium measure for that walk, the operator norm of the kernel for the $\mathrm{L}^2(\mathbb{V})$ norm is necessarily $\leqslant 1$[*]; so the right-hand side of (MW) is $\leqslant Z$, hence the announced ρ^*-mixing result.

Now let us turn to Assertion (ii). We take back the reasoning above at the point where we introduced the X_k's and the Y_k's; and now we use the following lemma (whose proof is postponed):

Lemma 93. *Let $\mathbb{V}$ satisfy Assumption 86 and let $\rho \colon \mathbb{R}_+ \to \mathbb{R}_+$ be a decreasing function such that $\rho(0) < 1$ and $\sum_{n \in \mathbb{N}} \rho(\boldsymbol{\delta}_n) < \infty$. Then there exists a constant $\boldsymbol{R} < 1$ (only depending on $\mathbb{V}$ and $\rho(\bullet)$) such that, for any system of variables $(\vec{X}_{\mathbb{V}}, \vec{Y}_{\mathbb{V}})$ satisfying $\rho^{\ddagger}(X_i; Y_j) \leqslant \rho\big(dist(i,j)\big)$ $\forall i, j \in \mathbb{V}$, one has $\rho(\vec{X}_{\mathbb{V}}; \vec{Y}_{\mathbb{V}}) \leqslant \boldsymbol{R}$.* ◇

Applying Lemma 93 with '$\rho(\delta)$' $\leftarrow \mathbf{1}_{\delta > 0} \tilde{\rho}^{\ddagger}(\delta)$, we directly get the wished immediate ρ^*-mixing result, $\rho^*(1)$ being the '$\boldsymbol{R}$' given by the lemma. This completes the proof of Lemma 91, modulo proving Lemma 93. ♤

Proof of Lemma 93. As a preliminary remark, note that one can safely assume that $\rho(\delta) \to 0$ as $\delta \to \infty$: indeed, either $\mathbb{V}$ is infinite and then it is an automatic consequence of the fact that $\sum_{n \in \mathbb{N}} \rho(\boldsymbol{\delta}_n)$ converges, or $\mathbb{V}$ is finite and then you can modify $\rho(\delta)$ freely for $\delta > \operatorname{diam}(\mathbb{V})$.

Denote $\tilde{\boldsymbol{R}}_0 \coloneqq \rho(0)$: our assumptions imply that $\sup_{\delta \in \mathbb{R}_+} \rho(\delta) \leqslant \tilde{\boldsymbol{R}}_0 < 1$. Since $\sum_{n \in \mathbb{N}} \rho(\boldsymbol{\delta}_n)$ converges, by dominated convergence there exists some $\delta^{\diamond} < \infty$ such that

---

[*] When the physical space is negatively curved, it is actually possible to do better: see Remark 94 infra.

$\sum_{n\in\mathbb{N}} \rho(\boldsymbol{\delta}_n \vee \delta^\diamond) < 1 - \tilde{\boldsymbol{R}}_0$: take such a $\delta^\diamond$. By Lemma 90, we can partition $\mathbb{V}$ into a finite number $(\mathbb{V}_c)_{c\in\mathscr{C}}$ of $2\delta^\diamond$-separated sets. Then, for all $i \in \mathbb{V}$, for all $c' \in \mathscr{C}$, there is at most one $j \in \mathbb{V}_{c'}$ such that $\mathit{dist}(i,j) < \delta^\diamond$.

Now let $(\vec{X}_{\mathbb{V}}, \vec{Y}_{\mathbb{V}})$ be a system of variables like in the lemma's statement. For any $c, c' \in \mathscr{C}$, one has by double tensorization (Theorem 39):

$$\rho(\vec{X}_{\mathbb{V}_c}; \vec{Y}_{\mathbb{V}_{c'}}) \leqslant \left|\!\left|\!\left| \Big(\big(\rho\big(\mathit{dist}(i,j)\big)\big)\Big)_{i\in\mathbb{V}_c, j\in\mathbb{V}_{c'}} \right|\!\right|\!\right|. \tag{NA}$$

To bound above the right-hand side of (NA), we observe that the same technique as at the end of the proof of Lemma 91-(ii)[†] allows to bound it above by

$$\sup_{i\in\mathbb{V}_c} \sum_{j\in\mathbb{V}_{c'}} \rho\big(\mathit{dist}(i,j)\big). \tag{NB}$$

The sum $\sum_{j\in\mathbb{V}_{c'}} \rho(\mathit{dist}(i,j))$ can be split depending on whether $\mathit{dist}(i,j) < \delta^\diamond$—which shall occur at most once, as we noted above—or not, so that

$$\rho(\vec{X}_{\mathbb{V}_c}; \vec{Y}_{\mathbb{V}_{c'}}) \leqslant \tilde{\boldsymbol{R}}_0 + \sum_{n\in\mathbb{N}} \rho(\boldsymbol{\delta}_n \vee \delta^\diamond) =: \boldsymbol{R}_0 < 1. \tag{NC}$$

Note that the value of $\boldsymbol{R}_0$ only depends on ($\mathbb{V}$ and) $\rho(\bullet)$.

The previous reasoning could also have be carried conditionally: so one has, for all $c, c' \in \mathscr{C}$:

$$\rho^\ddagger(\vec{X}_{\mathbb{V}_c}; \vec{Y}_{\mathbb{V}_{c'}}) \leqslant \boldsymbol{R}_0. \tag{ND}$$

But now, we can consider $\vec{Y}_{\mathbb{V}}$ as the collection $(\vec{Y}_{\mathbb{V}_{c'}})_{c'\in\mathscr{C}}$; so, applying *simple* tensorization (Theorem 37) to the $\vec{Y}_{\mathbb{V}_{c'}}$'s, we get that for all c, $\rho^\ddagger(\vec{X}_{\mathbb{V}_c}; \vec{Y}_{\mathbb{V}}) \leqslant \big(1 - (1 - \boldsymbol{R}_0^2)^{|\mathscr{C}|}\big)^{1/2} =: \boldsymbol{R}_1 < 1$. Then, applying again simple tensorization, but this time to the $\vec{X}_{\mathbb{V}_c}$'s, we finally get that $\rho(\vec{X}_{\mathbb{V}}; \vec{Y}_{\mathbb{V}}) \leqslant \big(1 - (1 - \boldsymbol{R}_1^2)^{|\mathscr{C}|}\big)^{1/2} =: \boldsymbol{R}_2 < 1$: $\boldsymbol{R}_2$ is the wished constant '$\boldsymbol{R}$'. ♤

Remark 94. The random sub-walk that we introduced to bound the right-hand side of (MW) should morally be understood as an actual random walk (i.e., having no mass loss): this is technically the case, in particular, as soon as the physical space $\mathbb{V}$ is transitive—that is, when all its vertices are indistinguishable as regards the geometric structure.

In many practical situations, the kernel of such a random walk has its operator norm (in $\mathrm{L}^2(\mathbb{V})$) worth *exactly* 1: this is in particular the case when the physical space has a "flat" geometry like $\mathbb{Z}^d$ (provided the jumps of the walk have some finite moments), and more generally as soon as the growth rate of $\mathbb{V}$ (i.e., the growth behaviour of the function $\delta \mapsto \boldsymbol{n}(\delta)$) is sub-exponential. However, *negatively curved* spaces (e.g., transitive discrete subsets of the hyperbolic space $\mathbf{H}^d$ for $d > 1$) do have exponential growth; and in that case, the kernel of the associated random walks

---

[†] There is however a little difference here, because you have to consider a *one-step* random sub-walk $(\boldsymbol{W}_t)_{t\in\{0,1\}}$, jumping from $\mathbb{V}_c$ to $\mathbb{V}_{c'}$. As a consequence, instead of saying that "the counting measure is a super-equilibrium measure", here you have to say that "starting from the counting measure on $\mathbb{V}_c$, after one step one gets bounded above by the counting measure on $\mathbb{V}_{c'}$".

may be *strictly contracting* in L^2! For instance, this is the case when you consider the standard random walk on the vertices of the order-7 triangular tiling in $\mathbf{H}^2$ (cf. Figure 3.1).

More generally, when the physical space is transitive, the kernel of our random walk can be seen as a mixture of elementary random walks of the type "choose randomly one of your neighbours among those at distance exactly δ"; and provided you avoid some "pathological" cases,[‡] these elementary random walks shall be strictly contracting in L^2 as soon as one has negative curvature. Hence, in negative curvature (and provided one avoids certain pathologies), the right-hand side of (`MT`) can be improved into something of the form

$$\sum_{n \geqslant \boldsymbol{n}(\delta)} \lambda(\boldsymbol{\delta}_n)\tilde{\rho}^{\ddagger}(\boldsymbol{\delta}_n), \tag{NE}$$

where the $\lambda(\delta)$'s only depend on the geometric structure of $\mathbb{V}$, and are all < 1 for $\delta > 0$. ♧

Remark 95. Note that to prove Assertion (ii), we have made a crucial use of the full $\rho^{\ddagger}$-mixing assumption: in the sense that, contrary to what we did in Theorem 58, here the $\rho^{\ddagger}$-mixing coefficients that appear in the assumptions could not have been replaced by appropriate $\rho^{\dagger}$-mixing coefficients! And no smarter proof could circumvent that limitation: indeed, Theorem 67 shows that it is possible to have all the pairwise $\rho^{\dagger}$-mixing coefficients < 1 while still having a global correlation coefficient of 1. ♧

3.2 First application: Ising-like models

3.2.1 Original Ising model

In the introduction, we said that we would like to get more general ρ-mixing results for Ising's model. This is what this section is going to achieve, through generalized tensorization. However, to apply tensorization of ρ-mixing, we will need the model to be in a regime somehow stronger than weak mixing:

Proposition 98 (Complete analyticity, [9, Theorem 2.2-$\mathscr{A}$]). *There is a $\beta'_{\mathrm{c}} > 0$ such that, as soon as $\beta < \beta'_{\mathrm{c}}$, the (zero-field) Ising model is* completely analytical, *i.e. there exist $\gamma' > 0$ and $\boldsymbol{C}' < \infty$ such that the following variant of* (`BB`) *holds: for all $\boldsymbol{K} \subseteq \mathbb{Z}^d$, for any boundary condition $\vec{s}_{\boldsymbol{K}} \in \{\pm 1\}^{\boldsymbol{K}}$, one has, for all disjoint $\boldsymbol{I}, \boldsymbol{J} \subseteq \mathbb{Z}^d$:*

$$\|\mathrm{Law}(\vec{\sigma}_{\boldsymbol{I}\sqcup\boldsymbol{J}} \mid \vec{\sigma}_{\boldsymbol{K}} = \vec{s}_{\boldsymbol{K}}) - \mathrm{Law}(\vec{\sigma}_{\boldsymbol{I}} \mid \vec{\sigma}_{\boldsymbol{K}} = \vec{s}_{\boldsymbol{K}}) \otimes \mathrm{Law}(\vec{\sigma}_{\boldsymbol{J}} \mid \vec{\sigma}_{\boldsymbol{K}} = \vec{s}_{\boldsymbol{K}})\|_{\mathrm{TV}} \leqslant \boldsymbol{C}' \sum_{(i,j) \in \boldsymbol{I} \times \boldsymbol{J}} \exp\bigl(-\gamma'\, \mathit{dist}(i,j)\bigr). \quad \diamondsuit\spadesuit \tag{NJ}$$

[‡] In particular, $\mathbb{V}$ might appear as the disjoint union of some finite clusters of vertices, these clusters being quite distant from each other: in which case a random walk that remains restrained to a single cluster would obviously not be contracting.

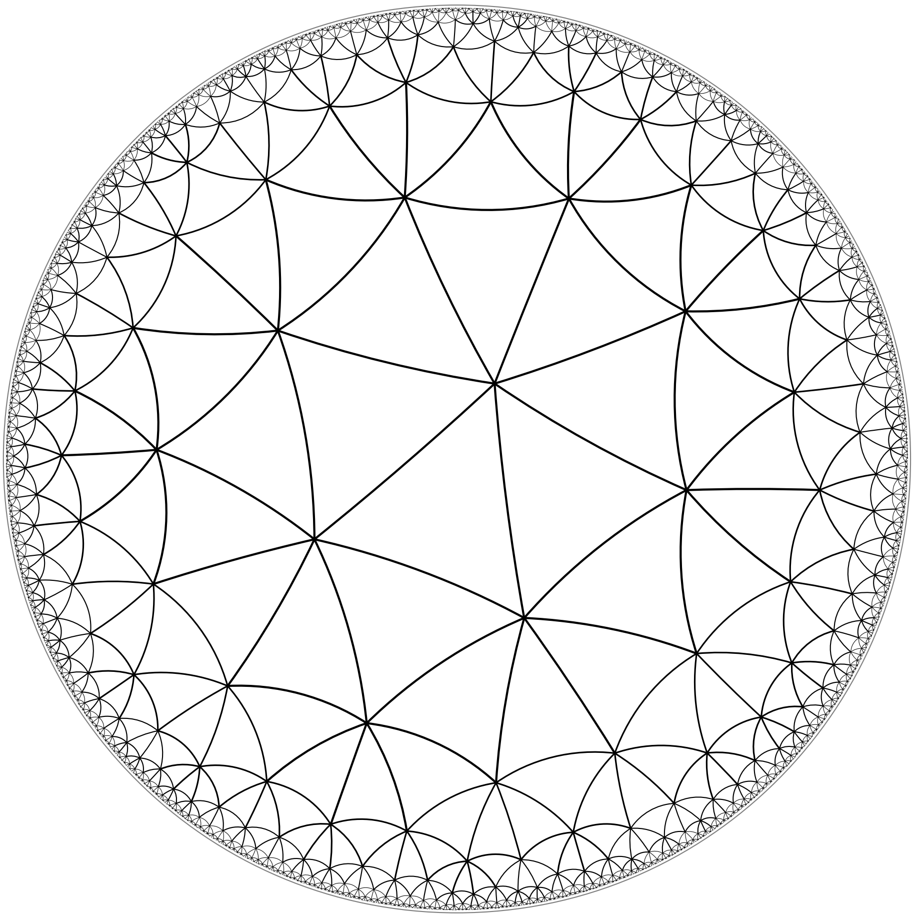

Figure 3.1: This drawing represents, via the Poincaré disk convention, the regular tiling of the hyperbolic plane $\mathbf{H}^2$ corresponding to the Schläfli symbol $\{3, 7\}$. (Note that all the segments of the figure do actually have the same length in $\mathbf{H}^2$). Denoting by $\mathbb{V}$ the set of vertices of this graph, the nearest-neighbour random walk on $\mathbb{V}$ exhibits the following remarkable behaviour: despite admitting the uniform (counting) measure as an equilibrium, its operator norm in $\mathrm{L}^2(\mathbb{V})$ is strictly less than 1! Such a behaviour is actually typical of negatively curved geometries. [The drawing was traced by adapting a public code due to Jeremy Tan Jie Rui, a.k.a. Parcly Taxel].

Remark 99. In general, complete analyticity is a strictly stronger condition than weak mixing, as was first exemplified by Shlosman [10]. As regards the specific case of the zero-field Ising model, to the best of my knowledge, it remains an open question whether $\beta'_{\mathrm{c}} = \beta_{\mathrm{c}}$ or $\beta'_{\mathrm{c}} < \beta_{\mathrm{c}}$. (The answer might actually depend on dimension). For a deeper discussion on these issues, see e.g. [11, § 2.3] and the references therein. ♧

Armed with the notion of complete analyticity, we are now able to state our general ρ-mixing result for Ising's model:

Theorem 100 (Immediate ρ^*-mixing for Ising's model)**.** *In the completely analytical regime, Ising's model is immediately $\rho^{*\ddagger}$-mixing with γ'-exponential decay, where γ' is the same as in Proposition 98.* ◇

Remark 101. So, now we have got four decorrelation results on Ising's model: namely, on the one hand, Propositions 12 (weak mixing), 15 (ρ-mixing for hyperplanes) and 98 (complete analyticity), which were already known in the literature; and on the other hand, Theorem 100 (interlaced ρ-mixing), which is our newcomer. It is interesting to see the strengths and weaknesses of the latter with respect to the three other:

- Like Proposition 15, Theorem 100 quantifies decorrelation in terms of ρ-mixing, while Propositions 12 and 98 spoke the language of β-mixing. As such, this is neither a strength nor a weakness, given that there is no general link between ρ- and β-mixing [12, Ch. 3]. But the advantage of ρ-mixing for such models of statistical physics is that it allows for decorrelation results which are robust in the size of the bunches of spins considered (which is indeed the case for Theorems 15 and 100), which would be unattainable by β-mixing (cf. Proposition 14). Note by the way that both ρ-mixing behaviours given by Theorems 15 and 100 are immediate, which was not granted a priori.[§]

- Most importantly, the result of Theorem 100 makes no assumption at all on the shape of the bunches of spins to decorrelate, which is a real improvement compared to Proposition 15. Also, it allows for a conditional version, which was not the case for the introduction's proposition.

- However, a serious drawback of Theorem 100 is that it is only valid in the completely analytical regime; moreover, the exponential speed γ' that it yields for decorrelations is a priori weaker than the speed γ in Proposition 15. ♧

Proof of Theorem 100. Thanks to Lemma 91, it will be enough for us to prove that there exist constants $A < \infty$ and $R < 1$ (only depending on d and $\beta < \beta'_{\mathrm{c}}$) such that, for all $i, j \in \mathbb{Z}^d$:

$$\mathrm{\rho}^{\ddagger}(\sigma_i ; \sigma_j) \leqslant A\mathrm{e}^{-\gamma' |j-i|} \wedge R. \tag{NK}$$

So, let $K \subseteq \mathbb{Z}^d$ and $\vec{s}_K \in \{\pm 1\}^K$ be some arbitrary boundary condition, and denote $\mathbb{P}_{\mathrm{cond}}(\bullet) \coloneqq \mathbb{P}(\bullet \mid \vec{\sigma}_K = \vec{s}_K)$: our goal is to show that, for all distinct $i, j \in \mathbb{Z}^d \setminus K$, one

---

[§] As regards Propositions 12 and 98, they are not sufficient as such to prove immediacy of β-mixing; but for such results dealing with individual spins, it would actually be fairly easy to show that β-mixing is in fact immediate.

has $\rho_{\mathrm{cond}}(\sigma_i;\sigma_j) \leqslant A\mathrm{e}^{-\gamma'|j-i|} \wedge R$—where the subscript “cond” means that ρ-mixing is considered under the conditioned probability law.

What Proposition 98 gives us is, under the law $\mathbb{P}_{\mathrm{cond}}$, a $\underline{\beta}$-mixing result between σ_i and σ_j. To link β-mixing with ρ-mixing, we will use that σ_i and σ_j have finite ranges: in such a case indeed, by Remark 7, $\rho_{\mathrm{cond}}(\sigma_i;\sigma_j)$ can be explicitly computed to be

$$\frac{|\mathbb{P}_{\mathrm{cond}}(\sigma_i = s_i \text{ and } \sigma_j = s_j) - \mathbb{P}_{\mathrm{cond}}(\sigma_i = s_i)\,\mathbb{P}_{\mathrm{cond}}(\sigma_j = s_j)|}{\big(\mathbb{P}_{\mathrm{cond}}(\sigma_i = -1)\,\mathbb{P}_{\mathrm{cond}}(\sigma_i = +1)\,\mathbb{P}_{\mathrm{cond}}(\sigma_j = -1)\,\mathbb{P}_{\mathrm{cond}}(\sigma_j = +1)\big)^{1/2}}, \tag{NL}$$

for s_i and s_j arbitrary values in $\{\pm 1\}$. The numerator of (NL) is actually equal to $\frac{1}{4}\|\mathrm{Law}_{\mathrm{cond}}(\sigma_i,\sigma_j) - \mathrm{Law}_{\mathrm{cond}}(\sigma_i) \otimes \mathrm{Law}_{\mathrm{cond}}(\sigma_j)\|_{\mathrm{TV}}$, which by Proposition 98 is bounded by $\frac{1}{4}C'\mathrm{e}^{-\gamma'|j-i|}$: so, to get the exponential part of the bound (NK), it is enough to show that the denominator of (NL) is bounded from below independently from i, j and $\vec{s}_K$.

We will just show how one can bound $\mathbb{P}_{\mathrm{cond}}(\sigma_i = -1)$ from below, the method for the three other factors being exactly the same. Let N denote the set of the neighbours of i (of cardinality $2d$). Since Ising’s model’s interactions are local, for $\vec{s}'_N \in \{\pm 1\}^N$, one has $\mathbb{P}_{\mathrm{cond}}(\sigma_i = -1 \mid \vec{\sigma}_N = \vec{s}'_N) \underset{\mathrm{def}}{=} \mathbb{P}(\sigma_i = -1 \mid \vec{\sigma}_K = \vec{s}_K \text{ and } \vec{\sigma}_N = \vec{s}'_N) = \mathbb{P}(\sigma_i = -1 \mid \vec{\sigma}_N = \vec{s}'_N)$; thus

$$\begin{aligned}\mathbb{P}_{\mathrm{cond}}(\sigma_i = -1) = \sum_{\vec{s}'_N \in \{\pm 1\}^N} \underbrace{\mathbb{P}(\vec{\sigma}_N \equiv \vec{s}'_N)}_{\sum_{\vec{s}'_N}(\bullet) = 1} \underbrace{\mathbb{P}_{\mathrm{cond}}(\sigma_i = -1 \mid \vec{\sigma}_N \equiv \vec{s}'_N)}_{= \mathbb{P}(\sigma_i = -1 \mid \vec{\sigma}_N \equiv \vec{s}'_N)} \\ \geqslant \min_{\vec{s}'_N \in \{\pm 1\}^N} \mathbb{P}(\sigma_i = -1 \mid \vec{\sigma}_N \equiv \vec{s}'_N).\end{aligned} \tag{NM}$$

But, given the Hamiltonian (AZ) of the model, the minimum possible value for $\mathbb{P}(\sigma_i = -1 \mid \vec{\sigma}_N \equiv \vec{s}'_N)$ is attained for $\vec{s}'_N \equiv +1$ and is then worth $1/(\mathrm{e}^{4d\beta} + 1)$, which is indeed a strictly positive constant independent from i and $\vec{s}_K$!

So we have got the “$\rho^{\ddagger}(\sigma_i;\sigma_j) \leqslant A\mathrm{e}^{-\gamma'|j-i|}$” part of (NK), with $A = (\mathrm{e}^{4d\beta} + 1)^2 C'/4$. As regards the “$\rho^{\ddagger}(\sigma_i;\sigma_j) \leqslant R$” part, by examining the Hamiltonian of the Ising model (as we did a few lines above to bound $\mathbb{P}_{\mathrm{cond}}(\sigma_i = -1)$ from below), we observe that

$$\mathbb{P}_{\mathrm{cond}}(\sigma_i = s_i \text{ and } \sigma_j = s_j) \geqslant 1/(\mathrm{e}^{4d\beta} + 1)^2: \tag{NO}$$

so, thanks to Lemma 8, one can take

$$R = 1 - 4\,/\,(\mathrm{e}^{4d\beta} + 1)^2, \tag{NP}$$

which is indeed < 1 and does not depend on $\vec{s}_K$ nor on (i,j). ♤

3.2.2 Generalization of Ising’s model

In this subsection we will illustrate how the ρ-mixing tensorization argument is independent from the facts that Ising’s model has finite-range interactions and that it is invariant by translation—both of which properties were used crucially to derive Proposition 15 in the introduction. More precisely, we are going to establish interlaced ρ-mixing for a class of *spin glass* models with *infinite range* interactions. Let us introduce the model first:

Definition 103 (Infinite-range spin glass model)**.** Our *infinite-range spin glass model* is defined as the model of statistical mechanics where the state space is the same as for Ising's model (namely, it is $\{\pm 1\}^{\mathbb{V}}$ for a physical space $\mathbb{V} \coloneqq \mathbb{Z}^d$, the latter being endowed with the ℓ^1 graph distance $|\bullet - \bullet|$) and where the Hamiltonian of a state $\vec{\sigma}_{\mathbb{V}}$ is formally defined by

$$\mathscr{H}(\vec{\sigma}_{\mathbb{V}}) \coloneqq -\frac{1}{2} \sum_{i \neq j} J_{ij} \sigma_i \sigma_j \tag{NQ}$$

for some symmetric function $J_{\bullet\bullet} \colon \mathbb{V} \times \mathbb{V} \to \mathbb{R}$,[¶] J being supposed to be such that the function

$$\mathbb{R}_+^* \ni \delta \mapsto \hat{J}(\delta) \coloneqq \sup_{|j-i| \geqslant \delta} |J_{ij}| \tag{NR}$$

decays either exponentially or γ-polynomially for some $\gamma > d$. (Apart from that, J may be completely arbitrary). ♡

For this model, we have an analogue of the complete analyticity of Ising's model, whose proof is postponed at the end of the subsection:

Proposition 104 ("Complete mixing" for the infinite-range spin glass model)**.** *For the infinite-range spin glass model, there exists a $\beta_{\mathrm{c}}' > 0$ such that, for $\beta < \beta_{\mathrm{c}}'$, the following holds: there is a unique equilibrium probability for the model; and one can find a function $\tau \colon \mathbb{R}_+^* \to \mathbb{R}_+$ that is either exponentially decaying (in the case J is exponentially decaying) or γ-polynomially decaying (in the case J is polynomially decaying), such that, for all $K \subseteq \mathbb{V}$, for any boundary condition $\vec{s}_K \in \{\pm 1\}^K$, for all distinct $i, j \in \mathbb{V} \smallsetminus K$:*

$$\|\mathrm{Law}(\sigma_i, \sigma_j \mid \vec{\sigma}_K = \vec{s}_K) - \mathrm{Law}(\sigma_i \mid \vec{\sigma}_K = \vec{s}_K) \otimes \mathrm{Law}(\sigma_j \mid \vec{\sigma}_K = \vec{s}_K)\|_{\mathrm{TV}} \leqslant \tau(|j-i|). \tag{NS}$$

◇

With Proposition 104 at hand, applying the very same techniques as in the previous subsection, we get a ρ^*-mixing result for the infinite-range spin glass model. This is the main result of this subsection:

Theorem 105 (ρ^*-mixing for the infinite-range spin glass model)**.** *For $\beta < \beta_{\mathrm{c}}'$, the infinite-range spin glass model is $\rho^{*\ddagger}$-mixing, with either exponential decay (in the exponential case) or $(\gamma - d)$-polynomial decay (in the polynomial case).* ◇♤

Also, as the spins of our spin glass model take values in a finite space, and J is bounded, the same techniques as in the previous subsection show that $\tilde{\rho}^{\ddagger}$-mixing is immediate:

[¶] Note that J_{ij} may be negative.

Proposition 106. *For $\beta < \beta'_c$, there exists $R < 1$ such that for all $i \neq j$, $\mathrm{\rho}^{\ddagger}(\sigma_i; \sigma_j) \leqslant R$.* ◇♤

Therefore, Theorem 105 can be reinforced to include immediacy of mixing:

Theorem 107 (Immediacy of ρ^*-mixing for the infinite-range spin glass model)**.** *The $\rho^{*\ddagger}$-mixing behaviour given by Theorem 105 is immediate.* ◇♤

Now let us prove Proposition 104. Note that the flavour of the ideas in this proof will be used again for the model of § 3.3.

Proof of Proposition 104. For the time being, let us not bother with the issue of equilibrium's unique existence: this will actually be proved "in passing" in Corollary 113. So, let us pretend that we are just interested in establishing (`NS`).

The first step of the proof consists in proving an estimate of "strong mixing" type for our model: this is the following

Proposition 108 (Strong mixing estimate for the infinite-range spin glass model)**.** *For β small enough, there exists a function $\tau\colon \mathbb{R}_+^* \to \mathbb{R}_+$ (the same as in* (`NS`)[‖]*), either exponentially decaying (exponential case) or γ-polynomially decaying (polynomial case), such that, for all cofinite $K \subseteq \mathbb{V}$, for all $\vec{s}_K \in \{\pm 1\}^K$, for $j \in \mathbb{V} \smallsetminus K$, $I \subseteq \mathbb{V} \smallsetminus K \smallsetminus \{j\}$:*

$$\|\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K, \sigma_j = +1) - \mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K, \sigma_j = -1)\|_{\mathrm{TV}} \leqslant \sum_{i \in I} \tau(|j - i|). \tag{NT}$$

◇

So, let us prove Proposition 108. In the sequel, we consider $j \in \mathbb{V} \smallsetminus K$ to be fixed. Our strategy will rely on considering the *Glauber dynamics* for the system conditioned by $\{\vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = -1\}$, resp. by $\{\vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = +1\}$. Let us first recall what the Glauber dynamics is:

Definition 109 (Glauber dynamics, [11])**.** For a statistical mechanics system made of a finite number of "spins" $(\sigma_i)_{i \in \Lambda}$, the *Glauber dynamics* is the process of (continuous) temporal evolution of the system according to the following rules:

- To each spin is attached a clock ringing in a Poissonian way (independently of everything else) at rate 1;
- When the clock of spin i rings at time t, that spin is re-drawn randomly according to the conditional law $\mathbb{P}_{\mathrm{eq}}(\sigma_i \mid \vec{\sigma}_{\Lambda \smallsetminus \{i\}} \equiv \vec{\sigma}_{\Lambda \smallsetminus \{i\}}(t-))$, where $\mathbb{P}_{\mathrm{eq}}(\bullet)$ denotes the equilibrium measure of the system and $\vec{\sigma}(t-)$ is the state of the system just before the clock rang. ♡

Remark 110. In the cases that we will consider, $\mathbb{P}_{\mathrm{eq}}(\bullet)$ will be of the form $\mathbb{P}(\bullet \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = s_j)$: so, since the values of the spins in $K \sqcup \{j\}$ are fixed, we can take $\Lambda \coloneqq \mathbb{V} \smallsetminus K \smallsetminus \{j\}$, which is finite indeed. Moreover, "$\mathbb{P}_{\mathrm{eq}}(\sigma_i \mid \vec{\sigma}_{\Lambda \smallsetminus \{i\}} \equiv \vec{\sigma}_{\Lambda \smallsetminus \{i\}}(t-))$" in Definition 109 actually corresponds to $\mathbb{P}_{\mathrm{eq}}(\sigma_i \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = s_j \text{ and } \vec{\sigma}_{\mathbb{V} \smallsetminus K \smallsetminus \{i,j\}} \equiv \vec{\sigma}_{\mathbb{V} \smallsetminus K \smallsetminus \{i,j\}}(t-))$ for the global system. ♧

[‖] More precisely, if some function $\tau(\bullet)$ fits for (`NT`), then it will also fit for (`NS`).

The following well-known property, which is immediate, is the reason why Glauber dynamics is interesting to study the properties of an equilibrium measure:

Proposition 111 (Glauber dynamics preserves equilibrium measure, [13])**.** *The Glauber dynamics of a system preserves its equilibrium measure.* ◇♤

Our strategy to prove (NT) will consist in *coupling* the Glauber dynamics of the two conditioned systems that interest us. I.e., we will define, on the same probability space, two processes $(\vec{\sigma}^-(t))_t$ and $(\vec{\sigma}^+(t))_t$ following the Glauber dynamics associated to the respective measures $\mathbb{P}(\bullet \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = -1)$ and $\mathbb{P}(\bullet \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = +1)$. The coupling shall be done in a clever way so that $\vec{\sigma}^-(t)$ and $\vec{\sigma}^+(t)$ get "close" enough to each other as time passes.

Here is the definition of our coupled process $(\vec{\sigma}^-(t), \vec{\sigma}^+(t))_t$. First, each of the Glauber dynamics starts from the equilibrium measure of the corresponding system (so this measure will be preserved along time), the initial states of each Glauber dynamics being coupled arbitrarily (e.g. independently). We denote by $\mathscr{F}_t$ the σ-algebra of everything that happened up to time t (here t may actually refer to a *stopping* time), with the convention that, if t corresponds to a time when a clock rings, we consider that at time t, one only knows that the clock has rung, but not yet what the new state of the corresponding spin will be. The coupled evolution follows the following rules:

- The clocks for each Glauber dynamics ring at the exact same times, for the exact same spins;
- When the clock for vertex i rings at time t, the states of spins σ_i^- and σ_i^+ just after that time (which I denote by resp. $\sigma_i^-(t+)$ and $\sigma_i^+(t+)$) are re-drawn in such a way that $\mathbb{P}(\sigma_i^-(t+) \neq \sigma_i^+(t+) \mid \mathscr{F}_t)$ is minimal, i.e. equal to $\frac{1}{2}\|\mathrm{Law}(\sigma_i^+(t+) \mid \mathscr{F}_t) - \mathrm{Law}(\sigma_i^-(t+) \mid \mathscr{F}_t)\|_{\mathrm{TV}}$.

Since the distance $\|\mathrm{Law}(\sigma_i^+(t+) \mid \mathscr{F}_t) - \mathrm{Law}(\sigma_i^-(t+) \mid \mathscr{F}_t)\|_{\mathrm{TV}}$ plays a crucial role in the coupled evolution, let us bound it from above. The Hamiltonian of the system yields that

$$\frac{\mathbb{P}(\sigma_i^-(t+) = +1 \mid \mathscr{F}_t)}{\mathbb{P}(\sigma_i^-(t+) = -1 \mid \mathscr{F}_t)} = \frac{\exp\bigl(\beta \sum_{j' \in \mathbb{V} \smallsetminus \{i\}} J_{ij'} \sigma_{j'}^-(t-)\bigr)}{\exp\bigl(-\beta \sum_{j' \in \mathbb{V} \smallsetminus \{i\}} J_{ij'} \sigma_{j'}^-(t-)\bigr)}, \tag{NU}$$

whence

$$\mathbb{P}(\sigma_i^-(t+) = \pm 1 \mid \mathscr{F}_t) = \frac{1}{2} \pm \tfrac{1}{2} \tanh\Bigl(\beta \sum_{j' \in \mathbb{V} \smallsetminus \{i\}} J_{ij'} \sigma_{j'}^-(t-)\Bigr); \tag{NV}$$

likewise, $\mathbb{P}(\sigma_i^+(t+) = \pm 1 \mid \mathscr{F}_t) = 1/2 \pm \frac{1}{2} \tanh\bigl(\beta \sum_{j' \in \mathbb{V} \smallsetminus \{i\}} J_{ij'} \sigma_{j'}^+(t-)\bigr)$. Hence:

$$\|\mathrm{Law}(\sigma_i^+(t+) \mid \mathscr{F}_t) - \mathrm{Law}(\sigma_i^-(t+) \mid \mathscr{F}_t)\|_{\mathrm{TV}} = \\ \Bigl|\tanh\Bigl(\beta \sum_{j' \in \mathbb{V} \smallsetminus \{i\}} J_{ij'} \sigma_{j'}^+(t-)\Bigr) - \tanh\Bigl(\beta \sum_{j' \in \mathbb{V} \smallsetminus \{i\}} J_{ij'} \sigma_{j'}^-(t-)\Bigr)\Bigr|. \tag{NW}$$

For the sequel, let us denote

$$\varDelta(t) \coloneqq \{j' \in \mathbb{V} \smallsetminus K \mid \sigma^-_{j'}(t) \neq \sigma^+_{j'}(t)\}. \tag{NX}$$

Splitting the sums in (NW) depending on whether $j' \in \varDelta(t-)$ and using the elementary inequality $|\tanh(b) - \tanh(a)| \leqslant |b - a|$, one gets the simplified bound:

$$\begin{aligned} \|\mathrm{Law}(\sigma^+_i(t+) \mid \mathscr{F}_t) - \mathrm{Law}(\sigma^-_i(t+) \mid \mathscr{F}_t)\|_{\mathrm{TV}} &\leqslant \Big|\beta \sum_{j' \in \varDelta(t-) \smallsetminus \{i\}} J_{ij'} (\sigma^+_{j'}(t-) - \sigma^-_{j'}(t-))\Big| \\ &\leqslant 2\beta \sum_{j' \in \varDelta(t-) \smallsetminus \{i\}} |J_{ij'}|. \end{aligned} \tag{NY}$$

So, the probability that σ^-_i and σ^+_i become (or remain) different after their (common) clock rings cannot be higher than $\beta \sum_{j' \in \varDelta(t-) \smallsetminus \{i\}} |J_{ij'}|$. Moreover, even when $\varDelta(t-)$ is large, the probability that σ^-_i and σ^+_i become (or remain) identical after their clock rings can never be too low: indeed, by (NW),

$$\begin{aligned} \mathbb{P}(\sigma^-_i(t+) = \sigma^+_i(t+) \mid \mathscr{F}_t) &\underset{\mathrm{def}}{=} 1 - \tfrac{1}{2} \|\mathrm{Law}(\sigma^+_i(t+) \mid \mathscr{F}_t) - \mathrm{Law}(\sigma^-_i(t+) \mid \mathscr{F}_t)\|_{\mathrm{TV}} \\ &\geqslant 1 - \tanh\Big(\beta \sum_{j' \in \mathbb{V} \smallsetminus \{i\}} |J_{ij'}|\Big), \end{aligned} \tag{NZ}$$

where one can bound

$$\sum_{j' \in \mathbb{V} \smallsetminus \{i\}} |J_{ij'}| \leqslant \sum_{u \in \mathbb{Z}^d \smallsetminus \{0\}} \hat{J}(|u|) \eqqcolon \bar{J}, \tag{OA}$$

which is $< \infty$ due to the decay assumption on $\hat{J}(\bullet)$. We will denote $r(\beta) \coloneqq 1 - \tanh(\beta \bar{J})$, which is positive and only depends on β (and on $J_{\bullet\bullet}$ of course).

Thanks to the estimates (NY) and (OA), we can define a Markovian process $(\hat{\varDelta}(t))_{t \geqslant 0}$ valued in $\mathfrak{P}(\mathbb{V} \smallsetminus K \smallsetminus \{j\})$, coupled with the processes $(\vec{\sigma}^-(t))_t$ and $(\vec{\sigma}^+(t))_t$, which will be such that $\hat{\varDelta}(t) \supseteq \varDelta(t)$ for all t. Below we just describe the law for the evolution of $\hat{\varDelta}(\bullet)$, the way it is coupled with $\vec{\sigma}^-(\bullet)$ and $\vec{\sigma}^+(\bullet)$ being hopefully obvious. This evolution is the following:

- Initially one has $\hat{\varDelta}(t = 0) \coloneqq \mathbb{V} \smallsetminus K \smallsetminus \{j\}$.
- For each $(i, j') \in (\mathbb{V} \smallsetminus K \smallsetminus \{j\}) \times \mathbb{V}$ with $i \neq j'$, there is a clock ringing in a Poissonian way at rate $\beta |J_{ij'}|$ (such a clock is called a *decoupling clock*). All the decoupling clocks are independent.
- For each $i \in \mathbb{V} \smallsetminus K \smallsetminus \{j\}$, there is a clock ringing in a Poissonian way at rate $r(\beta)$ (such a clock is called a *recoupling clock*). All the recoupling clocks are independent, and independent from the decoupling clocks.
- If, at time t, the decoupling clock for pair (i, j') rings $\underline{\text{and}}$ $\hat{\varDelta}(t-) \ni j'$, then the process $\hat{\varDelta}$ jumps to $\hat{\varDelta}(t+) \coloneqq \hat{\varDelta}(t-) \cup \{i\}$. (Note that in the case where one already had $\hat{\varDelta}(t-) \ni i$, nothing actually happens).

- If, at time t, the recoupling clock for spin i rings, then the process $\hat{\Delta}$ jumps to $\hat{\Delta}(t+) \coloneqq \hat{\Delta}(t-) \smallsetminus \{i\}$. (In the case one had $\hat{\Delta}(t-) \not\ni i$, nothing actually happens).

Because of the general properties of Markov processes on finite state spaces, the process $\hat{\Delta}(t)$ converges to a (unique) equilibrium measure as $t \to \infty$; so we can define

$$\theta(i) \coloneqq \lim_{t\to\infty} \mathbb{P}(\hat{\Delta}(t) \ni i). \tag{0B}$$

The function $\theta(\bullet)$ is what we need to get a strong mixing estimate for the system: indeed, since, for all t, $\vec{\sigma}^-(t)$ and $\vec{\sigma}^+(t)$ follow the respective distributions $\mathrm{Law}(\vec{\sigma}_{\mathbb{V}\smallsetminus K\smallsetminus\{j\}} \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = -1)$ and $\mathrm{Law}(\vec{\sigma}_{\mathbb{V}\smallsetminus K\smallsetminus\{j\}} \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = +1)$, and since the points where $\vec{\sigma}^-(t)$ and $\vec{\sigma}^+(t)$ differ do necessarily belong to $\hat{\Delta}(t)$, we get that for $I \subseteq \mathbb{V} \smallsetminus K \smallsetminus \{j\}$:

$$\begin{aligned}&\|\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = +1) - \mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = -1)\|_{\mathrm{TV}} \leqslant \\ &\qquad 2\,\mathbb{P}(\vec{\sigma}_I^-(t) \neq \vec{\sigma}_I^+(t)) \leqslant 2\,\mathbb{P}(\hat{\Delta}(t) \cap I \neq \varnothing) \leqslant 2\sum_{i\in I} \mathbb{P}(\hat{\Delta}(t) \ni i) \overset{t\to\infty}{\to} 2\sum_{i\in I}\theta(i),\end{aligned} \tag{0C}$$

which is the kind of result that we are looking for.

Now, since the law of $\hat{\Delta}(t)$ converges to equilibrium, as $t \to \infty$, the recoupling events (which, for site i, arrive at asymptotic rate $r(\beta)\theta(i)$) have to compensate exactly the decoupling events (which arrive at rate at most $\beta \sum_{j'\in\mathbb{V}\smallsetminus K\smallsetminus\{i\}} |J_{ij'}|\theta(j')$). Therefore, the function $\theta(\bullet)$ has to satisfy the following system:

$$\begin{cases} r(\beta)\theta(i) \leqslant \beta \displaystyle\sum_{\substack{j'\in\mathbb{V}\smallsetminus K\\ j'\neq i}} \hat{J}(|j'-i|)\theta(j') & \text{for } i \in \mathbb{V} \smallsetminus K \smallsetminus \{j\};\\ \theta(k) = \mathbf{1}_{k=j} & \text{for } k \in K \sqcup \{j\}.\end{cases} \tag{0D}$$

Introducing the convolution kernel $\kappa(\bullet)$ on $\mathbb{Z}^d$ defined by

$$\begin{cases} \kappa(0) = 1;\\ \kappa(u) = -r(\beta)^{-1}\beta\hat{J}(|u|) & \text{for } u \neq 0,\end{cases} \tag{0E}$$

the first line of (0D) (i.e. the “bulk inequality”) implies that $\kappa * \theta \leqslant 0$ on $\mathbb{V} \smallsetminus K \smallsetminus \{j\}$. Also, given the decay assumption on $\hat{J}$ and the way that $r(\beta)$ depends on β, one has $\|\delta_0 - \kappa\|_{\ell^1(\mathbb{Z}^d)} < 1$ for β small enough: so we will assume that this is the case in the sequel. Then, the convolution operator $\kappa*\bullet$ can be decomposed as “$\kappa*\vartheta = \mu\vartheta + \tilde{\kappa}*\vartheta$”, where $\mu \coloneqq 1 - \|\delta_0 - \kappa\|_{\ell^1(\mathbb{Z}^d)}$ is positive and the “$\kappa * \bullet$” operator is of (discrete, fractional) elliptic type, so that (0D) means that $\theta(\bullet)$ has to be a *sub-solution* of some difference equation of *elliptic* type.

By analogy with the theory of partial differential equations, we are going to look for a *barrier function* for that difference equation, i.e. for some function $\hat{\theta}(\bullet)$ satisfying

$$\begin{cases} \big(\kappa * \hat{\theta}\big)(i) = 0 & \text{for } i \in \mathbb{V} \smallsetminus K \smallsetminus \{j\};\\ \hat{\theta}(k) \geqslant \mathbf{1}_{k=j} & \text{for } k \in K \sqcup \{j\}:\end{cases} \tag{0F}$$

then, by the (discrete) maximum principle, it will follow that $\theta \leqslant \hat{\theta}$ everywhere. To find the desired barrier function, we observe that, since $\|\delta_0 - \kappa\|_{\ell^1} < 1$, κ is invertible for the convolution product, with inverse

$$\kappa^{-*} = \sum_{k=0}^{\infty} (\delta_0 - \kappa)^{*k} := \delta_0 + (\delta_0 - \kappa) + (\delta_0 - \kappa) * (\delta_0 - \kappa) + \cdots ; \tag{OG}$$

also, since the kernel $\delta_0 - \kappa$ is nonnegative, κ^{-*} is nonnegative everywhere, with $\kappa^{-*}(0) > 0$. Therefore, we can take the following barrier function:

$$\hat{\theta}(i) := \frac{\kappa^{-*}(i-j)}{\kappa^{-*}(0)}. \tag{OH}$$

So we have just proved that, for all $i \in \mathbb{V} \smallsetminus K \smallsetminus \{j\}$, one has $\theta(i) \leqslant \kappa^{-*}(i-j)/\kappa^{-*}(0)$. But, given that $\hat{J}(\bullet)$ was assumed to decay at either exponential or γ-polynomial speed, we have an either exponential or γ-polynomial control on the decay of $\delta_0 - \kappa$; and therefore, by either Lemma 174 or Lemma 176 in appendix, $\kappa^{-*}(\bullet)$ decays[**] exponentially or γ-polynomially too. Via Equation (`OC`), this finally proves (`NT`) with $\tau(\delta) := 2 \sup_{|u| \leqslant \delta} |\kappa^{-*}(u)| \,/\, \kappa^{-*}(0)$, ending the proof of Proposition 108.

To get Proposition 104 from Proposition 108, the idea is the following: as $\mathrm{Law}(\sigma_i \mid \vec{\sigma}_K \equiv \vec{s}_K)$ is a barycentre of $\mathrm{Law}(\sigma_i \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = -1)$ and $\mathrm{Law}(\sigma_i \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = +1)$, applying (`NT`) with '$I$' ← $\{i\}$, we have that, for all $s_j \in \{\pm 1\}$:

$$\|\mathrm{Law}(\sigma_i \mid \vec{\sigma}_K \equiv \vec{s}_K) - \mathrm{Law}(\sigma_i \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = s_j)\|_{\mathrm{TV}} \leqslant \tau(|j-i|); \tag{OI}$$

from which, integrating over the value of s_j, we get (`NS`).

Technically however, the above only works provided K was cofinite… To get (`NS`) also in the case K where is not cofinite, we need the following lemma, whose proof we postpone for now:

Lemma 112. *For $K \subseteq \mathbb{V}$, $\vec{s}_K \in \mathcal{S}^K$, for $I \subseteq \mathbb{V} \smallsetminus K$ finite, let $(\Lambda_n)_n$ be an increasing sequence of finite sets, all containing I, converging to $\mathbb{V}$. Then:*

$$\operatorname{diam}\big\{\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \vec{\sigma}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K} \equiv \vec{s}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K}) \bigm| \vec{s}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K} \in \{\pm1\}^{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K}\big\} \overset{n\to\infty}{\to} 0, \tag{OJ}$$

where "$\operatorname{diam} \mathscr{M}$" denotes the diameter, for the total variation distance, of a set $\mathscr{M}$ of probability measures. ◇

Lemma 112 has the following immediate corollary, since $\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K)$ is a barycentre of the $\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \vec{\sigma}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K} \equiv \vec{s}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K})$'s:

Corollary 113. *For $K \subseteq \mathbb{V}$, $\vec{s}_K \in \mathcal{S}^K$, there is a unique way[††] to define consistently $\mathrm{Law}(\vec{\sigma}_{\mathbb{V}\smallsetminus K} \mid \vec{\sigma}_K \equiv \vec{s}_K)$; and it has the following characterization: for all $\vec{s}_{\mathbb{V}\smallsetminus K} \in \{\pm1\}^{\mathbb{V}\smallsetminus K}$, for any finite $I \subseteq \mathbb{V} \smallsetminus K$, for $(\Lambda_n)_n$ like in Lemma 112:*

$$\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_{K\cup(\mathbb{V}\smallsetminus\Lambda_n)} \equiv \vec{s}_{K\cup(\mathbb{V}\smallsetminus\Lambda_n)}) \overset{n\to\infty}{\to} \mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K). \tag{OK}$$

[**] Technically, we should rather say that $\delta \mapsto \sup_{|u|\leqslant\delta} |\kappa^{-*}(u)|$ decays exponentially (resp. γ-polynomially), since we only defined exponential (resp. polynomial) decay for functions on $\mathbb{R}_+$.

[††] The important point here is uniqueness: existence is indeed immediate, by a compactness argument.

(Note that $\vec{\sigma}_I$ *has values in a finite space, so there is no ambiguity in the notion of convergence of its law).*

Incidentally, when applied to $K = \varnothing$*, this corollary also ensures uniqueness of equilibrium for the (unconditioned) system.* ◇♤

Thanks to Corollary 113, for K not necessarily cofinite, we can approximate arbitrarily well $\mathrm{Law}(\sigma_i \mid \vec{\sigma}_K = \vec{s}_K)$ and $\mathrm{Law}(\sigma_i \mid \vec{\sigma}_K = \vec{s}_K$ and $\sigma_j = s_j)$ by something where K is cofinite, then getting by Proposition 108 that the total variation between our approximations is $\leqslant \tau(|j-i|)$. Since that bound does not depend on the level of approximation chosen, we deduce that the bound $\tau(|j-i|)$ in (`OI`), and hence in (`NS`), is actually valid for any general K. ♤

It finally remains to prove Lemma 112:

Proof of Lemma 112. Proposition 108 already shows the following: if $\vec{s}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K}$ and $\vec{s}'_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K}$ are two (partial) boundary conditions differing by just <u>one</u> spin at j, then

$$\|\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \vec{\sigma}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K} \equiv \vec{s}'_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K}) - \mathrm{Law}(\vec{\sigma}_I \mid \vec{s}_K \text{ and } \vec{s}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K})\|_{\mathrm{TV}} \leqslant \\ \sum_{i\in I} \tau(|j-i|) \leqslant |I|\tau\big(\mathit{dist}(I,j)\big). \quad (\texttt{OM})$$

Then, the idea consists in saying that one can go from any boundary condition $\vec{s}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K}$ to any other boundary condition $\vec{s}'_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K}$ by switching all the needed spins of $\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K$ one by one, which will have a global impact in total variation bounded by

$$|I| \sum_{j\in\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K} \tau\big(\mathit{dist}(I,j)\big), \qquad (\texttt{ON})$$

which indeed tends to 0 as $n \to \infty$, given the decay speed of τ.

That idea is actually correct; however a little care has to be taken, because there are an infinite number of spins to switch... To get over that technicality, we observe that under $\mathbb{P}(\bullet \mid \vec{\sigma}_{K\cup(\mathbb{V}\smallsetminus\Lambda_n)} \equiv \vec{s}_{K\cup(\mathbb{V}\smallsetminus\Lambda_n)})$, the law of $\vec{\sigma}_{\Lambda_n\smallsetminus K}$ can be described by the effective Hamiltonian

$$\mathscr{H}_{\mathrm{eff}}(\vec{\sigma}_{\Lambda_n\smallsetminus K}) \coloneqq -\tfrac{1}{2} \sum_{\substack{i,j\in\Lambda_n\smallsetminus K\\ i\neq j}} J_{ij}\sigma_i\sigma_j - \sum_{i\in\Lambda_n\smallsetminus K}\Big(\sum_{j\in K\cup(\mathbb{V}\smallsetminus\Lambda_n)} J_{ij}s_j\Big)\sigma_i, \qquad (\texttt{OO})$$

where for all i, regardless of the value of $\vec{s}_{K\cup(\mathbb{V}\smallsetminus\Lambda_n)}$, the sum over j is absolutely convergent due to the decay property of $J_{\bullet\bullet}$. From (`OO`) you see that, when switching all the (needed) spins of $\vec{s}_{\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K}$ one by one, the effective Hamiltonian, and hence the law of $\vec{\sigma}_{\Lambda_n\smallsetminus K}$ (which variable takes values in a finite space), converge to the situation with all the spins flipped: this makes legit to pass to the limit of flipping an infinite number of spins in (`ON`).[‡‡] ♤

[‡‡] When I say that one "passes to the limit of flipping an infinite number of spins", you have to understand that the spins are getting flipped according to some $\mathbb{N}$-isomorphic ordering of $\mathbb{V}\smallsetminus\Lambda_n\smallsetminus K$, so that flipping all the spins appears as the limit of flipping a finite number of spins.

3.3 Second application: Lattice of nonlinear particles

In this section we will consider a model with *continuous* variables. The model is the following. One has a lattice of particles indexed by $\mathbb{Z}^d =: \mathbb{V}$ ($\mathbb{V}$ being equipped with the ℓ^1 graph distance, denoted by $|\bullet - \bullet|$), and the state of the particle at site i is described by its "polarization" $\sigma_i \in \mathbb{R}^\ell =: \mathcal{S}$. Each particle is submitted to a *pinning force* deriving from a potential V, and to *interaction forces* with its $2d$ neighbours, the interaction deriving from a (symmetric) potential W on the difference between polarizations: in other words, the Hamiltonian is formally

$$\mathscr{H}(\vec{\sigma}_{\mathbb{V}}) := \sum_{i\in\mathbb{V}} V(\sigma_i) + \frac{1}{2} \sum_{\substack{i,j\in\mathbb{V}\\ |j-i|=1}} W(\sigma_j - \sigma_i). \tag{OP}$$

We make the following assumptions:

Assumption 115. Both V and W are convex, with V being uniformly strictly convex and the Hessian of W being bounded; i.e., there exist constants $\alpha > 0$ and $B < \infty$ such that for all $s \in \mathcal{S}$, $\alpha\mathbf{I}_\ell \leqslant \nabla\nabla V(s)$ and $\mathbf{0}_\ell \leqslant \nabla\nabla W(s) \leqslant B\mathbf{I}_\ell$ in the sense of quadratic forms (where '$\nabla\nabla$' refers to Hessian matrix). ♡

We are interested in the equilibrium state of the system at some inverse temperature $\beta \in (0, \infty)$, which we fix arbitrarily for all the sequel. Then, the probability distribution of the system is formally described by the following: ($\vec{s}$ standing for a point of $\mathcal{S}^{\mathbb{V}}$ and "$\mathrm{d}\vec{s}$" for an infinitesimal neighbourhood of $\vec{s}$),

$$\mathbb{P}(\vec{\sigma} \in \mathrm{d}\vec{s}) \propto \exp(-\beta\mathscr{H}(\vec{s}))\,\mathrm{vol}(\mathrm{d}\vec{s}), \tag{OQ}$$

where vol($\bullet$) is the formal "infinite-dimensional Lebesgue measure". Because of our convexity assumptions on V and W, morally, $\mathscr{H}$ is uniformly strictly convex and hence (OQ) defines a unique probability distribution.[*] The goal of this section is to prove the following

Theorem 116 (ρ^*-mixing for our lattice of nonlinear particles)**.** *The model that we have been defining is exponentially $\rho^{*\ddagger}$-mixing.* ◇

To prove Theorem 116, thanks to Lemma 91, it will suffice to prove a pairwise ρ-mixing result:

Proposition 117 (Pairwise ρ-mixing for our model)**.** *There exists an exponentially decaying function $\tau\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that, for all $i, j \in \mathbb{V}$:*

$$\rho^\ddagger(\sigma_i; \sigma_j) \leqslant \tau(|j-i|). \tag{OR}$$

◇

---

[*] In reality, the infinite number of particles makes things more complicated: there are actually several Gibbs measures for the system, but one of them arises as canonical, which is the measure that we will consider in the sequel. (That canonical equilibrium may be defined as the limit of equilibria for finite boxes with free boundary: see § A.3.1 in appendix for more detail).

To prove Proposition 117, we are going to use some concepts of optimal transportation. Let us recall the definition of the W_2 *Wasserstein distance* between probability measures:

Definition 118 (W_2 distance, [14, Definition 6.1])**.** For μ_0, μ_1 two probability measures on some metric space $(\mathscr{X}, \mathit{dist})$, $\mathrm{W}_2(\mu_0, \mu_1)$ is defined as the infimum of the values $\mathbb{E}\big(\mathit{dist}(X_0, X_1)^2\big)^{1/2}$ over the class of joint random variables[†] (X_0, X_1) that satisfy $\mathrm{Law}(X_0) = \mu_0, \mathrm{Law}(X_1) = \mu_1$. Provided $\mathscr{X}$ is Polish, $\mathrm{W}_2(\bullet, \bullet)$ defines a distance (possibly taking infinite values) on the set of probability measures on $\mathscr{X}$. ♡

The W_2 distances satisfy the following easy lemma, which will be useful in the sequel:

Lemma 119. *For P_0, P_1 probability distributions on a metric space $\mathscr{X}$ and $f\colon \mathscr{X} \to \mathbb{R}$ a k-Lipschitz function, one has $|\mathbb{E}^{X\sim P_1}(f(X)) - \mathbb{E}^{X\sim P_0}(f(X))| \leqslant k\,\mathrm{W}_2(P_0, P_1)$.* ♢

Proof. For X_0, X_1 joint variables following resp. the laws P_0 and P_1:

$$\begin{aligned}|\mathbb{E}^{X\sim P_1}(f(X)) - \mathbb{E}^{X\sim P_0}(f(X))| = |\mathbb{E}(f(X_1)) - \mathbb{E}(f(X_0))| \leqslant \mathbb{E}(|f(X_1) - f(X_0)|)\\ \leqslant k\,\mathbb{E}\big(\mathit{dist}(X_0, X_1)\big) \underset{\text{CS}}{\leqslant} k\,\mathbb{E}\big(\mathit{dist}(X_0, X_1)^2\big)^{1/2}, \quad \texttt{(OS)}\end{aligned}$$

where you can let the right-hand side tend to $\mathrm{W}_2(P_0, P_1)$ by definition of the latter. ♤

That vocabulary being set, here is the key result to prove Proposition 117:

Proposition 120 (Continuity estimate on the impact of conditioning spins)**.** *There exists an exponentially decaying function $\tau\colon \mathbb{R}_+ \to \mathbb{R}_+$ (the same as in* (`OR`)*) such that the following holds: for all $K \subseteq \mathbb{V}$, for $\mathbb{P}$-almost-all $\vec{s}_K \in \mathscr{S}^K$, for all distinct i, $j \in \mathbb{V} \smallsetminus K$:*

$$\begin{aligned}&\forall s_j^0, s_j^1 \in \mathscr{S}\\ &\quad \mathrm{W}_2\big(\mathrm{Law}(\sigma_i \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = s_j^0), \mathrm{Law}(\sigma_i \mid \vec{s}_K \text{ and } s_j^1)\big) \leqslant \tau(|j - i|)|s_j^1 - s_j^0|,\\ &\hfill \texttt{(OT)}\end{aligned}$$

where $|\bullet|$ stands for the standard Euclidean norm on $\mathscr{S}$. ♢

In a first time, let us show how Proposition 120 implies Proposition 117 (and hence Theorem 116):

Proof of Proposition 117 from Proposition 120. First, we need to introduce a certain norm for σ_i-measurable variables:

[†] Here it is implicit that we are working on a probability space which is rich enough so that any possible joint law for (X_0, X_1) can be obtained.

Definition 121 (Lip(σ_i) norm)**.** For $f\colon \mathcal{S} \to \mathbb{R}$ a Lipschitz function, denote by $\|f\|_{\mathrm{Lip}}$ its optimal Lipschitz constant. Then, for $i \in \mathbb{V}$, for φ a $\upsigma(\sigma_i)$-measurable[‡] real variable, we define

$$\|\varphi\|_{\mathrm{Lip}(\sigma_i)} \coloneqq \inf\{\|f\|_{\mathrm{Lip}} \mid \varphi \equiv f(\sigma_i) \text{ a.-e.}\}, \tag{OU}$$

which defines a seminorm (possibly taking infinite values) on $\mathrm{L}^2(\upsigma(\sigma_i))$. As obviously $\|\varphi\|_{\mathrm{Lip}(\sigma_i)} = \|\varphi + c\|_{\mathrm{Lip}(\sigma_i)}$ for $c \in \mathbb{R}$, notation $\|\bullet\|_{\mathrm{Lip}(\sigma_i)}$ can be quotiented to get defined on $\bar{\mathrm{L}}^2(\upsigma(\sigma_i))$, on which it is a norm. ♡

That notation being set, let us turn to Proposition 117 itself. The proposition means that for all $K \subseteq \mathbb{V}$, for almost-all $\vec{s}_K \in \mathcal{S}^K$, denoting $\mathbb{P}_{\mathrm{cond}}(\bullet) \coloneqq \mathbb{P}(\bullet \mid \vec{\sigma}_K \equiv \vec{s}_K)$, for all distinct $i, j \in \mathbb{V} \setminus K$ one has

$$\uprho_{\mathrm{cond}}(\sigma_i; \sigma_j) \leqslant \tau(|j-i|), \tag{OV}$$

where "$\uprho_{\mathrm{cond}}$" means that we are considering ρ-mixing under the law $\mathbb{P}_{\mathrm{cond}}$. So, let us fix $\vec{s}_K \in \mathcal{S}^K$ outside the null set of "pathological" boundary conditions,[§] and denote $\mathbb{P}_{\mathrm{cond}}$ as above. Now, like in Proposition 4, let us define $\pi_{\sigma_j\sigma_i}\colon \bar{\mathrm{L}}^2_{\mathrm{cond}}(\sigma_i) \to \bar{\mathrm{L}}^2_{\mathrm{cond}}(\sigma_j)$ by

$$\pi_{\sigma_j\sigma_i}(f) \coloneqq f^{\upsigma(\sigma_j)} \tag{OW}$$

(here of course conditioning is done under the ambient law $\mathbb{P}_{\mathrm{cond}}$); define likewise $\pi_{\sigma_i\sigma_j}$, and set $\pi_{\sigma_i\sigma_j\sigma_i} \coloneqq \pi_{\sigma_i\sigma_j} \circ \pi_{\sigma_j\sigma_i}$. Proposition 4 tells us that, regarding $\pi_{\sigma_i\sigma_j\sigma_i}$ as an operator on $\bar{\mathrm{L}}^2_{\mathrm{cond}}(\sigma_i)$:

$$\uprho_{\mathrm{cond}}(\sigma_i; \sigma_j) = \uprho_{\mathrm{sp}}(\pi_{\sigma_i\sigma_j\sigma_i})^{1/2}. \tag{OX}$$

But operator $\pi_{\sigma_j\sigma_i}$ can also be characterized in the following way: for $f(\sigma_i) \eqqcolon \varphi$ a random variable, $\pi_{\sigma_j\sigma_i}(\varphi)$ is equal to $g(\sigma_j)$, with

$$g(s_j) \coloneqq \mathbb{E}_{\mathrm{cond}}(f(\sigma_i) \mid \sigma_j = s_j) \underset{\mathrm{def}}{=} \mathbb{E}(f(\sigma_i) \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = s_j). \tag{OY}$$

With the notation of (OY), for $s_j^0, s_j^1 \in \mathcal{S}$, Proposition 120 implies that $|g(s_j^1) - g(s_j^0)| \leqslant \tau(|j-i|)\|f\|_{\mathrm{Lip}}|s_j^1 - s_j^0|$ by Lemma 119; so, when seen as an operator from $\mathrm{Lip}(\sigma_i)$[¶]

[‡] Here notation may be a bit confusing: the first symbol '$\upsigma$' (in roman type) refers to generating a $\upsigma$-algebra, while the second 'σ' (in italic type) is (part of) the name of the random variable describing the state of some particle! So, "$\upsigma(\sigma_i)$" is actually the $\upsigma$-algebra of the random variables that are measurable w.r.t. the polarization of particle i.

[§] The "pathological" boundary conditions of $\mathcal{S}^K$ are those for which the Langevin dynamics would not converge. But fortunately, the probability (under the unconditioned equilibrium measure) of such boundary conditions is zero: see § A.3.3 in appendix for more details.

[¶] Here in full rigour I should speak of "$\mathrm{Lip}_{\mathrm{cond}}(\sigma_i)$" (with a 'cond' subscript) all along (and likewise of "$\mathrm{Lip}_{\mathrm{cond}}(\sigma_j)$"); however the spaces $\mathrm{Lip}(\sigma_i)$ and $\mathrm{Lip}_{\mathrm{cond}}(\sigma_i)$ are essentially the same: indeed, giving a variable φ in either $\mathrm{Lip}(\sigma_i)$ or $\mathrm{Lip}_{\mathrm{cond}}(\sigma_i)$ is in both cases "almost" equivalent to giving a Lipschitz f on $\mathcal{S}$ such that $\varphi \equiv f(\sigma_i)$. There is actually some subtlety due to the fact that the equality "$\varphi \equiv f(\sigma_i)$" only has to stand almost-everywhere (and that the notion of "almost-everywhere" is not the same for $\mathbb{P}$ as for $\mathbb{P}_{\mathrm{cond}}$); however that subtlety does not matter for the current proof; so let us not bother with the 'cond' subscript here.

to $\mathrm{Lip}(\sigma_j)$, $\pi_{\sigma_j\sigma_i}$ is $\tau(|j-i|)$-contracting.[⦀] Likewise, $\pi_{\sigma_i\sigma_j}$ is $\tau(|j-i|)$-contracting from $\mathrm{Lip}(\sigma_j)$ to $\mathrm{Lip}(\sigma_i)$; hence $\pi_{\sigma_i\sigma_j\sigma_i}$ is $\tau(|j-i|)^2$-contracting when seen as an operator on $\mathrm{Lip}(\sigma_i)$.

Now we observe that, because of the convexity properties of $\mathscr{H}$, $\mathrm{Law}_{\mathrm{cond}}(\sigma_i)$ is log-concave[**] (see § A.3.2 in appendix for more justification); so in particular it has finite second moment (cf. e.g. [15, Corollary 1]). As a consequence, $\mathrm{Lip}(\sigma_i)$ is continuously embedded into $\bar{\mathrm{L}}^2_{\mathrm{cond}}$ by inclusion, since for all k-Lipschitz $f\colon \mathscr{S} \to \mathbb{R}$, one has

$$\begin{aligned}\|f(\sigma_i)\|^2_{\bar{\mathrm{L}}^2_{\mathrm{cond}}} &\underset{\mathrm{def}}{=} \mathrm{Var}_{\mathrm{cond}}(f(\sigma_i)) = \tfrac{1}{2}\,\mathbb{E}^{(\sigma,\sigma')\sim\mathrm{Law}_{\mathrm{cond}}(\sigma_i)^{\otimes 2}}\big((f(\sigma')-f(\sigma))^2\big) \\ &\leqslant \tfrac{1}{2}\,\mathbb{E}^{(\sigma,\sigma')\sim\mathrm{Law}_{\mathrm{cond}}(\sigma_i)^{\otimes 2}}(k^2|\sigma'-\sigma|^2) = \underbrace{(\mathbb{E}_{\mathrm{cond}}(|\sigma_i|^2)-|\mathbb{E}_{\mathrm{cond}}(\sigma_i)|^2)}_{<\infty}\times k^2. \qquad (\mathtt{OZ})\end{aligned}$$

Therefore, denoting by $\boldsymbol{C}$ the embedding constant from $\mathrm{Lip}(\sigma_i)$ into $\bar{\mathrm{L}}^2_{\mathrm{cond}}$, the $\tau(|j-i|)^2$-contractivity of $\pi_{\sigma_i\sigma_j\sigma_i}$ in $\mathrm{Lip}(\sigma_i)$ implies that, for all $f \in \mathrm{Lip}(\sigma_i), k \in \mathbb{N}$:

$$\begin{aligned}|\langle f, \pi^k_{\sigma_i\sigma_j\sigma_i} f\rangle_{\bar{\mathrm{L}}^2_{\mathrm{cond}}}| &\underset{\mathrm{CS}}{\leqslant} \|f\|_{\bar{\mathrm{L}}^2_{\mathrm{cond}}} \|\pi^k_{\sigma_i\sigma_j\sigma_i} f\|_{\bar{\mathrm{L}}^2_{\mathrm{cond}}} \\ &\leqslant C^2 \|f\|_{\mathrm{Lip}(\sigma_i)} \|\pi^k_{\sigma_i\sigma_j\sigma_i} f\|_{\mathrm{Lip}(\sigma_i)} \leqslant C^2 \tau(|j-i|)^{2k} \|f\|^2_{\mathrm{Lip}(\sigma_i)}, \qquad (\mathtt{PA})\end{aligned}$$

so that

$$\varlimsup_{k\to\infty} |\langle f, \pi^k_{\sigma_i\sigma_j\sigma_i} f\rangle_{\bar{\mathrm{L}}^2_{\mathrm{cond}}}|^{1/k} \leqslant \tau(|j-i|)^2. \qquad (\mathtt{PB})$$

But, since $\mathrm{Lip}(\sigma_i)$ is a dense subset of $\bar{\mathrm{L}}^2_{\mathrm{cond}}(\sigma_i)$ and $\pi_{\sigma_i\sigma_j\sigma_i}$ is a self-adjoint operator on $\bar{\mathrm{L}}^2_{\mathrm{cond}}(\sigma_i)$, (PB) implies that $\rho_{\mathrm{sp}}(\pi_{\sigma_i\sigma_j\sigma_i}) \leqslant \tau(|j-i|)^2$ on $\bar{\mathrm{L}}^2_{\mathrm{cond}}(\sigma_i)$, by Lemma 161 in appendix (which we already used to prove Proposition 15 in Chapter 1). In view of (OX), this proves (OV). ♤

Now let us prove Proposition 120.

Proof of Proposition 120. Fix $\boldsymbol{K} \subseteq \mathbb{V}, j \in \mathbb{V} \setminus \boldsymbol{K}, \vec{s}_K \in \mathscr{S}^K, s^0_j, s^1_j \in \mathscr{S}$. The principle of our proof will consist in using two Langevin dynamics to construct variables following (approximately) the respective distributions $\mathrm{Law}(\vec{\sigma}_{\mathbb{V}} \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = s^0_j)$ and $\mathrm{Law}(\vec{\sigma}_{\mathbb{V}} \mid \vec{\sigma}_K \equiv \vec{s}_K \text{ and } \sigma_j = s^1_j)$, and then to couple the Langevin dynamics associated to resp. s^0_j and s^1_j to get a good W_2 coupling between the variables "σ_i" corresponding to each case.

For our system, the *(overdamped) Langevin dynamics* associated to the conditioning event $\{\vec{\sigma}_{K\sqcup\{j\}} \equiv \vec{s}_{K\sqcup\{j\}}\}$ is the process evolving on $\mathscr{S}^{\mathbb{V}}$ such that one has the

---

[⦀] Here it may actually occur that $\tau(|j-i|) \geqslant 1$, in which case "$\tau(|j-i|)$-contracting" should be understood as an abuse of language.

[**] Let us remind that a probability distribution on (an affine subspace of) $\mathbb{R}^\ell$ is said to be *log-concave* when it admits a density whose logarithm (extended by $-\infty$ wherever density is zero) is concave.

boundary condition $\vec{\sigma}_{K\sqcup\{j\}}(t) \equiv \vec{s}_{K\sqcup\{j\}}$ for all times, and that for each $i \in \mathbb{V} \smallsetminus \boldsymbol{K} \smallsetminus \{j\}$, the motion of polarization σ_i is

$$\mathrm{d}\sigma_i(t) \coloneqq \Big(-\nabla V(\sigma_i(t)) + \sum_{|i'-i|=1} \nabla W(\sigma_{i'}(t) - \sigma_i(t)) \Big)\, \mathrm{d}t + \sqrt{2\beta^{-1}}\, \mathrm{d}B^i_t, \tag{PC}$$

where the $(\mathrm{d}B^i_t)_{t\in\mathbb{R}_+}, i \in \mathbb{V}$ are independent ℓ-dimensional white noises. One sets moreover the initial condition $\vec{\sigma}_{\mathbb{V}\smallsetminus K\smallsetminus\{j\}}(t = 0) \equiv 0$. This dynamics is devised so that, provided $\vec{s}_K$ does not belong to some "pathological" subset of $\mathcal{S}^K$ of null probability (for the model's law), as $t \to \infty$, the law of $\vec{\sigma}_{\mathbb{V}}(t)$ (for the dynamics) tends to $\mathrm{Law}(\vec{\sigma}_{\mathbb{V}} \mid \vec{\sigma}_{K\sqcup\{j\}} \equiv \vec{s}_{K\sqcup\{j\}})$ (for the "static" equilibrium measure): see Proposition 171 in appendix.

In the sequel, assume that $\vec{s}_K$ is non-pathological, and let us denote by $(\vec{\sigma}^0_{\mathbb{V}}(t))_t$ and $(\vec{\sigma}^1_{\mathbb{V}}(t))_t$ the respective Langevin dynamics corresponding to fixing σ_j at resp. s^0_j and s^1_j. We will use the same noise process for both dynamics, henceforth *coupling* them on some common probability space in a non-trivial way.

For $i \in \mathbb{V}$, let us denote $\eta_i(t) \coloneqq \sigma^1_i(t) - \sigma^0_i(t) \in \mathbb{R}^\ell$; then $\vec{\eta}_{\mathbb{V}}(t)$ evolves according to the following equation: for all $i \in \mathbb{V} \smallsetminus K \smallsetminus \{j\}$,

$$\mathrm{d}\eta_i(t) = \Big(\nabla V(\sigma^0_i(t)) - \nabla V(\sigma^1_i(t)) + \sum_{|i'-i|=1} \big(\nabla W(\sigma^1_{i'}(t) - \sigma^1_i(t)) - \nabla W(\sigma^0_{i'}(t) - \sigma^0_i(t)) \big) \Big)\, \mathrm{d}t; \tag{PD}$$

moreover, for all $t \in \mathbb{R}_+$, one has the boundary conditions:

$$\vec{\eta}_K(t) \equiv 0; \qquad\qquad \eta_j(t) = s^1_j - s^0_j. \tag{PE}$$

Now, observing that one can write

$$\nabla V(\sigma^1_i) - \nabla V(\sigma^0_i) = \Big(\int_{u=0}^1 \nabla\nabla V\big(\sigma^0_i + u(\sigma^1_i - \sigma^0_i)\big)\, \mathrm{d}u \Big) \cdot \underbrace{(\sigma^1_i - \sigma^0_i)}_{=\eta_i}, \tag{PF}$$

resp.

$$\begin{gathered} \nabla W(\sigma^1_{i'} - \sigma^1_i) - \nabla W(\sigma^0_{i'} - \sigma^0_i) = \\ \Big(\int_{u=0}^1 \nabla\nabla W\big(\sigma^0_{i'} - \sigma^0_i + u(\sigma^1_{i'} - \sigma^1_i - \sigma^0_{i'} + \sigma^0_i)\big)\, \mathrm{d}u \Big) \cdot \underbrace{(\sigma^1_{i'} - \sigma^1_i - \sigma^0_{i'} + \sigma^0_i)}_{=\eta_{i'} - \eta_i}, \end{gathered} \tag{PG}$$

we can rewrite (PD) as

$$\mathrm{d}\eta_i(t) = \Big(-\boldsymbol{V}(i,t) \cdot \eta_i(t) + \sum_{|i'-i|=1} \boldsymbol{W}(\{i,i'\},t) \cdot (\eta_{i'}(t) - \eta_i(t)) \Big)\, \mathrm{d}t, \tag{PH}$$

for some symmetric matrices $\boldsymbol{V}(i,t), \boldsymbol{W}(\{i,i'\},t) \in \mathbb{R}^{\ell\times\ell}$ obtained by averaging certain values of resp. $\nabla\nabla V$ and $\nabla\nabla W$, with $\boldsymbol{W}(\{i,i'\},t) = \boldsymbol{W}(\{i',i\},t)\ \forall i,i',t,$[††] that

---

[††] Hence the notation for the first argument of $\boldsymbol{W}$.

symmetry property coming from the change of variables $u \leftarrow 1-u$ in the formula defining $\boldsymbol{W}$ joint with the symmetry of potential $W(\bullet)$. Because of Assumption 115, $\boldsymbol{V}(i,t)$ and $\boldsymbol{W}(\{i,i'\},t)$ always satisfy the following inequalities:

$$\alpha\mathbf{I}_\ell \leqslant \boldsymbol{V}(i,t); \qquad\qquad \mathbf{0}_\ell \leqslant \boldsymbol{W}(\{i,i'\},t) \leqslant \boldsymbol{B}\mathbf{I}_\ell. \tag{PI}$$

So, $(\vec{\eta}_{\mathbb{V}}(t))_t$ is the solution of some "discrete damped heat equation", whose coefficients may vary along time, but always have to satisfy (`PI`).

Now we will control the behaviour of the solutions of (`PH`) thanks to the following Lyapunov functional:

$$\mathcal{U}(\vec{\eta}_{\mathbb{V}\setminus\boldsymbol{K}\setminus\{j\}}(t)) \coloneqq \sum_{i\in\mathbb{V}\setminus\boldsymbol{K}\setminus\{j\}} \mathrm{e}^{2\gamma|j-i|}|\eta_i(t)|^2, \tag{PJ}$$

for $\gamma > 0$ a small constant. In the sequel, it will be convenient to shorthand $\mathrm{e}^{2\gamma|j-i|} \eqqcolon w_i$; moreover we will make the dependence in t implicit, and also denote merely "$\vec{\eta}$" for "$\vec{\eta}_{\mathbb{V}\setminus\boldsymbol{K}\setminus\{j\}}$".

Assuming that $\mathcal{U}(\vec{\eta}) < \infty$, we can estimate its temporal derivative by using Equation (`PH`): then, $\dot{\mathcal{U}}(\vec{\eta})$ appears as the sum of several contributions. As regards the "$-\boldsymbol{V}(i,t)\cdot\eta_i\,\mathrm{d}t$" terms of (`PH`), they contribute to the temporal derivative of the Lyapunov functional by

$$-\sum_{i\in\mathbb{V}\setminus\boldsymbol{K}\setminus\{j\}} w_i\eta_i^\top\cdot\boldsymbol{V}(i,t)\cdot\eta_i \underset{(\texttt{PI})}{\leqslant} -\sum w_i\alpha|\eta_i|^2 = -\alpha\mathcal{U}(\vec{\eta}). \tag{PK}$$

As regards the contribution of the interaction terms "$\boldsymbol{W}(\{i,i'\},t)\cdot(\eta_{i'}-\eta_i)\,\mathrm{d}t$", we will decompose it by edges, i.e. group together the terms for (i,i') and (i',i). Depending on where the endpoints of edge $\{i,i'\}$ lie, three cases are to be distinguished:

First case: Nor i nor i' are on the boundary $\boldsymbol{K}\sqcup\{j\}$. Then the contribution of the interaction terms between i and i' is

$$\begin{aligned} w_i\eta_i^\top\cdot\boldsymbol{W}(\{i,i'\})\cdot(\eta_{i'}-\eta_i) + w_{i'}\eta_{i'}^\top\cdot\boldsymbol{W}(\{i,i'\})\cdot(\eta_i-\eta_{i'}) \\ = -\frac{w_i+w_{i'}}{2}\underbrace{(\eta_{i'}-\eta_i)^\top\cdot\boldsymbol{W}(\{i,i'\})\cdot(\eta_{i'}-\eta_i)}_{\geqslant 0 \text{ by } (\texttt{PI})} \\ +\frac{w_{i'}-w_i}{2}\big(\eta_i^\top\cdot\boldsymbol{W}(\{i,i'\})\cdot\eta_i - \eta_{i'}^\top\cdot\boldsymbol{W}(\{i,i'\})\cdot\eta_{i'}\big). \end{aligned} \tag{PL}$$

We can upper-bound further the last line of (`PL`). To do so, assume to fix ideas that $w_{i'} \geqslant w_i$. Then the positiveness of $\boldsymbol{W}$ implies that $\eta_{i'}^\top\cdot\boldsymbol{W}(\{i,i'\})\cdot\eta_{i'}$ is positive, while the last inequality of (`PI`) implies that $\eta_i^\top\cdot\boldsymbol{W}(\{i,i'\})\cdot\eta_i$ is bounded above by $\boldsymbol{B}|\eta_i|^2$; so, summing everything together, and observing that one has $w_{i'} \leqslant \mathrm{e}^{2\gamma}w_i$ (by the definition of $w_\bullet$ and the triangle inequality), we finally get a total contribution of the interaction between edges i and i' that is bounded above by

$$\frac{\mathrm{e}^{2\gamma}w_i - w_i}{2}\boldsymbol{B}|\eta_i|^2 \leqslant \frac{(\mathrm{e}^{2\gamma}-1)\boldsymbol{B}}{2}(w_i|\eta_i|^2 + w_{i'}|\eta_{i'}|^2), \tag{PM}$$

where the final bounding step was added to get a result where i and i' play the same role. When we finally sum all the edges $\{i,i'\}$ corresponding to this first case, any vertex $i \in \mathbb{V} \setminus \boldsymbol{K} \setminus \{j\}$ appears at most $2d$ times: so, the global contribution of these edges is bounded above by

$$(\mathrm{e}^{2\gamma}-1)dB\mathscr{U}(\vec{\eta}). \tag{PN}$$

Second case: Both i and i' are on the boundary. This case is trivial: neither η_i nor $\eta_{i'}$ evolve, so the contribution is zero.

Third case: Exactly one of the endpoints is on the boundary. To fix ideas, assume that i is the endpoint outside the boundary while i' belongs to the boundary. Then the contribution for edge $\{i,i'\}$ is

$$w_i\eta_i^\top\cdot\boldsymbol{W}(\{i,i'\},t)\cdot(\eta_{i'}-\eta_i) = w_i\eta_i^\top\cdot\boldsymbol{W}(\{i,i'\},t)\cdot\eta_{i'} - w_i\underbrace{\eta_i^\top\cdot\boldsymbol{W}(\{i,i'\},t)\cdot\eta_i}_{\geqslant 0} \leqslant w_iB|\eta_i||\eta_{i'}|, \tag{PO}$$

where the last equality follows from $\boldsymbol{W}$ being bounded by $B\mathbf{I}_\ell$. Here one has two subcases: either $i' \in \boldsymbol{K}$, in which case $\eta_{i'} = 0$ and thus the contribution is globally negative—so that we can disregard this situation—; or $i' = j$, in which case we just write that $|w_i\eta_i|^2 \leqslant w_i\mathscr{U}(\vec{\eta}) = \mathrm{e}^{2\gamma}\mathscr{U}(\vec{\eta})$ (using that $|j-i| = |i'-i| = 1$ in that situation), so that by Cauchy–Schwarz the contribution of the edge is bounded above by $B\mathrm{e}^{\gamma}\mathscr{U}(\vec{\eta})^{1/2}|\eta_j|$. When we finally sum all the contributions belonging to (the second subcase of) this third case, there at most $2d$ of them: so the global contribution for this case is bounded above by

$$2dB\mathrm{e}^{\gamma}\mathscr{U}(\vec{\eta})^{1/2}|\eta_j|. \tag{PP}$$

Summing up all the contributions (the pinning force plus all three cases of interaction forces), we finally get that

$$\dot{\mathscr{U}}(\vec{\eta}) \leqslant -\big(\alpha-(\mathrm{e}^{2\gamma}-1)dB\big)\mathscr{U}(\vec{\eta}) + 2dB\mathrm{e}^{\gamma}|s_j^1-s_j^0|\mathscr{U}(\vec{\eta})^{1/2}, \tag{PQ}$$

which I will shorthand as "$\dot{\mathscr{U}}(\vec{\eta}) \leqslant -C_1\mathscr{U}(\vec{\eta}) + C_2\mathscr{U}(\vec{\eta})^{1/2}$" for a few lines. (Beware that C_1 and C_2 implicitly depend on γ and $|s_j^1-s_j^0|$). So, the time evolution of $\mathscr{U}(\vec{\eta}(t))$ is a solution of the ordinary differential inequation "$\dot{y} \leqslant -C_1y + C_2y^{1/2}$". But, provided γ was chosen small enough, C_1 is strictly positive, so that the equality case "$\dot{y} = -C_1y + C_2y^{1/2}$" admits the positive constant solution $(C_2\,/\,C_1)^2$. And since the actual Lyapunov functional $\mathscr{U}(\vec{\eta}(t))$ starts from zero at time zero, a comparison argument implies that it always remains below the constant solution:

$$\text{Almost-surely,}\quad \forall t\in\mathbb{R}_+ \qquad \mathscr{U}(\vec{\eta}(t)) \leqslant (C_2\,/\,C_1)^2 = \left(\frac{\mathrm{e}^{\gamma}}{\alpha\,/\,B-(\mathrm{e}^{2\gamma}-1)d}\right)^2 |s_j^1-s_j^0|^2. \tag{PR}$$

But, for a given i, our coupling between the Langevin dynamics implies that the squared W_2 distance between the law of $\sigma_i(t)$ under the first dynamics and the law of $\sigma_i(t)$ under the second dynamics is bounded above by the expectation of $|\eta_i(t)|^2$, which in turn is bounded by $\mathbb{E}\big(w_i^{-1}\mathcal{U}(\vec{\eta}(t))\big) \underset{\text{def}}{=} \mathrm{e}^{-2\gamma|j-i|}\,\mathbb{E}\big(\mathcal{U}(\vec{\eta}(t))\big)$, which is less than $(C_2\,/\,C_1)^2\mathrm{e}^{-2\gamma|j-i|}$ by (PR). Now, we know that the law of $\sigma_i(t)$ under either Langevin dynamics converges to the equilibrium law of σ_i under the corresponding boundary condition: but the W_2 distance is lower-semicontinuous for the weak topology (see e.g. [14, Remark 6.12]); therefore by passing to the limit the left-hand side of (OT) is bounded above by

$$\frac{\mathrm{e}^{-\gamma}}{\alpha\,/\,B-(\mathrm{e}^{2\gamma}-1)d}\mathrm{e}^{-\gamma|j-i|}|s_j^1-s_j^0|, \tag{PS}$$

which has the wanted form. ♤

Remark 122. With the above proof, the function $\tau(\bullet)$ that we get in bound (OT) decays exponentially fast indeed, but does not satisfy $\tau(1) < 1$ in general: for that reason, nothing ensures that ρ-mixing in Proposition 117 (hence ρ^*-mixing in Theorem 116) should be immediate! I tried to find whether there would be another way to prove immediacy of ρ-mixing for the class of models of this section, or whether there are some models in this class for which ρ-mixing would not be immediate; but I did not manage to succeed in either direction… So I leave this here as an open question.

My personal guess is that there should be immediate ρ-mixing anyway, since it would seem very bizarre that a behaviour of such a qualitative nature as "being fully correlated in the sense of ρ-mixing" would hold for small distances, but suddenly vanish from a certain point on, while the model only shows interactions between neighbouring particles… (Yet that intuition may be misleading: for instance, for the half-space Ising model, it is shown in [16] that a layer of unstable spins can appear over the substrate, without extending to infinite distance!).

However, even in the case ρ-mixing would indeed be immediate, it is much possible that the optimal $\tau(\bullet)$ in Equation (OT) would not satisfy $\tau(\delta) < 1$ as soon as $\delta > 0$. Indeed, while immediate ρ-mixing consists in showing that the "conditioning operator"

$$\begin{gathered}\bar{\mathrm{L}}^2(\sigma_i) \to \bar{\mathrm{L}}^2(\sigma_j)\\ (s_i \mapsto f(s_i)) \mapsto \big(s_j \mapsto \mathbb{E}(f(\sigma_i)\ |\ \sigma_j = s_j)\big)\end{gathered} \tag{PT}$$

is (strictly) contracting as soon as $i \neq j$, getting $\tau(1) < 1$ for (OT) would essentially require to get the same contracting behaviour *when the conditioning operator is seen as an operator from* $\mathrm{Lip}(\sigma_i)$ *to* $\mathrm{Lip}(\sigma_j)$*!* But, qualitatively, there is an importance difference between considering the conditioning operator in $\bar{\mathrm{L}}^2$ spaces or in Lip spaces. Indeed, in $\bar{\mathrm{L}}^2$ spaces, it is automatic that the conditioning operator is non-expanding: so, showing contractivity just means that the norm of that operator does not take its "maximal value". On the other hand, as regards Lip spaces, there is no reason why the operator norm should automatically be $\leqslant 1$: so, there would be no blatant paradox between having some qualitatively neat decorrelation between spins σ_i and σ_j while still having a conditioning operator whose norm from $\bar{\mathrm{L}}^2(\sigma_i)$ to $\bar{\mathrm{L}}^2(\sigma_j)$ would be larger than 1… To put it in other words, there is no reason why immediate ρ-mixing, provided it holds, could necessarily be obtained through transportation distance arguments.[‡‡]

♧

3.4 Third application: A hypocoercive system of interacting particles

3.4.1 Motivation

In the examples of the two previous sections, decorrelation between random variables occurred when the particles were far apart *spatially*. However, ρ-mixing between infinite bunches of variables is also well-suited to study the *temporal* relaxation of a large- or infinite-dimensional stochastic system: indeed, considering time as an extra coordinate leads to a situation analogous to parallel hyperplanes for Ising's model (cf. Propositions 14 and 15 in § 1.2). This was actually my main motivation to study generalized tensorization of ρ-mixing initially.

When the temporal dynamics of the system considered is reversible, tensorization is actually of little use to study relaxation to equilibrium: then indeed, ρ-mixing corresponds to a spectral property of some self-adjoint operator; and we saw in the proof of Proposition 15 how one can get $\mathbf{L}^2$-type results (i.e. ρ-mixing) from $\mathbf{L}^1$-type results (i.e. β-mixing) in such a case. However, *non-reversible* evolutions are much trickier to treat. The archetypical example of temporal relaxation for a non-reversible evolution is the *underdamped* Langevin dynamics, that is, when you have a Hamiltonian evolution perturbed by some noise *acting on momenta*: in terms of partial differential equations, this corresponds to the probability density of the system evolving according to a *kinetic Fokker–Planck equation*. The study of such dynamics is complicated by the fact that diffusion only occurs along certain directions of the state space, so that non-reversibility of the evolution is actually essential to ensure convergence to equilibrium.

Of course, understanding the temporal relaxation of non-reversible systems is by no means a new question! In particular, from the PDE point of view, in [17] Villani developed the so-called *hypocoercivity* methods to tackle such problems. My goal in this section will be to show that probabilistic methods also have their say in this context, and that generalized tensorization of ρ-mixing, in particular, is a natural and useful tool when the setting is infinite-dimensional.

However, it turned out that, when the system looked at is complicated enough to make Villani-like techniques fail, trying to apply generalized tensorization of ρ-mixing leads to highly technical issues: tackling these (assuming it is possible indeed) would likely require a whole other monograph… In view of that, I lowered my ambitions and just looked at a quite elementary example (for which even hypocoercivity tools are actually unneeded to get the wanted results): what I will show is that, at least for that example, it is possible indeed to get the generalized tensorization machinery work to prove temporal relaxation to equilibrium, in a way that is both dimension-

---

[‡‡] Of course, I also looked for a way to get ρ-mixing without relying on transportation distances: but I was no more successful at it…

independent[*] and functioning for arbitrarily small times. So, the next subsection is just a "proof of concept" aiming at convincing you that it should be relevant to try and use generalized tensorization techniques for more complicated models.

3.4.2 The model studied

Remark 123. In order to keep things simple, below I took both the underlying physical space and the particle state space to be 1-dimensional. However, this is not an essential feature for what follows: everything would work the same for multidimensional physical and/or state spaces, just at the price of heavier notation. ♧

Definition 124 (Underdamped Langevin dynamics for a chain of coupled linear oscillators)**.** For real parameters $m, \omega, c, \beta, \lambda > 0$,[†] we consider a system of particles indexed by $\mathbb{Z}$, the particle indexed by x being described by its momentum $p_x \in \mathbb{R}$ and its position $q_x \in \mathbb{R}$. We consider the (formal) Hamiltonian

$$\mathscr{H}(\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}}) \coloneqq \frac{1}{2m} \sum_{x\in\mathbb{Z}} p_x^2 + \frac{m\omega^2}{2} \sum_{x\in\mathbb{Z}} q_x^2 + \frac{mc^2}{2} \sum_{x\in\mathbb{Z}} (q_{x+1} - q_x)^2. \tag{PU}$$

Then, the system $(\vec{p}_{\mathbb{Z}}(u), \vec{q}_{\mathbb{Z}}(u))_{u\geqslant 0}$ evolves according to the Hamiltonian $\mathscr{H}$, plus a "volumic Langevin heat bath at temperature β^{-1}": the latter consists, on each p_x, of a white noise of quadratic variation $\mathrm{d}[p_x] \coloneqq 2\lambda\beta^{-1} m\,\mathrm{d}u$, plus a friction force $F_x \coloneqq -\lambda p_x$ which dissipates the energy brought by the white noise, the white noises for the different x's being independent. In other words, being given independent white noises $(\mathrm{d}B_u^x)_{u\geqslant 0}$ for each $x \in \mathbb{Z}$, the evolution of the system is given by

$$\begin{cases} \mathrm{d}p_x \coloneqq \big(-m\omega^2 q_x + mc^2(q_{x-1} + q_{x+1} - 2q_x) - \lambda p_x\big)\,\mathrm{d}u + \sqrt{2\lambda\beta^{-1}m}\,\mathrm{d}B_u^x \\ \mathrm{d}q_x \coloneqq m^{-1} p_x\,\mathrm{d}u. \end{cases} \tag{PV}$$

♡

We will consider this system under its (canonical) equilibrium probability distribution: this corresponds to taking the initial state of the system so that, formally, the density of $(\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})$ is proportional to $\exp(-\beta\mathscr{H}(\bullet))$; which density will then be preserved by the dynamics. Now fix an arbitrary time $t > 0$. The goal of this section is to prove the following

---

[*] Here "dimension" does not refer to the dimension of the physical space "$\mathbb{V}$" (which in this section will be $\mathbb{Z}$), nor to the dimension of the space "$\mathscr{S}$" in which each particle lives (which will be $\mathbb{R}$), but to the dimension of the global state space "$\mathscr{S}^{\mathbb{V}}$" of the system: so, when I say that we are in the "infinite-dimensional" case, I actually mean that the number of particles considered is infinite.

[†] In physical terms, m represents the mass of each particle; ω is the pulsation corresponding to the pinning potential; c is morally the speed of sound (which reflects the strength of the interaction between neighbouring particles); β is the inverse temperature; and λ is the relaxation constant governing the thermal bath. For those not interested in the physical interpretation of the model, you may just drop the homogeneity constants by taking $m, \omega, \beta = 1$. By the way, actually some of the computations in the sequel will be physically inhomogeneous, in particular when introducing the L^2 norm on $\mathbb{R}^{\mathbb{Z}\times\{\mathtt{p},\mathtt{q}\}}$; but this will just be a kind of notational abuse, which will be without actual consequences: as regards the important formulae, they $\underline{\text{will}}$ in fact be physically homogeneous!

Theorem 125 (Hypocoercive behaviour for our Langevin dynamics)**.** *For the model defined above:*

$$\rho\big((\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=0); (\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=t)\big) < 1. \tag{PW}$$

◇

3.4.3 Proving hypocoercivity via generalized tensorization of ρ-mixing

Before turning to the proof of Theorem 125, let us to introduce quite a bit of notation:

Notation 126. In the sequel, we denote the state of particle x at time $u=0$ by (p_x, q_x), resp. its state at time $u=t$ by (p'_x, q'_x). The set of indices needed to describe the state of the system at some given time, namely $\mathbb{Z}\times\{\mathrm{p},\mathrm{q}\}$ (as each particle is described by its momentum and position), will be shorthanded as "$\mathbb{Z}\times 2$". Likewise, the set of indices $\mathbb{Z}\times\{\mathrm{p},\mathrm{q},\mathrm{p}',\mathrm{q}'\}$ needed to describe jointly the initial and final states of the system will be shorthanded as "$\mathbb{Z}\times 4$". The global states of the system at times resp. 0 and t will be denoted by resp. $\vec{\xi}$ and $\vec{\xi}'$. (These are $\mathbb{R}^{\mathbb{Z}\times 2}$-valued variables). When considering $\rho^{\ddagger}$-mixing, the p_x's, q_x's, p'_x's and q'_x's will be considered as the "fundamental" variables that the system is made of: in this context, we may also denote these variables by $(X_i)_{i\in\mathbb{Z}\times 4}$.

Also, in the sequel we will deal with a lot of (infinite) square matrices indexed by $\mathbb{Z}\times 4$ or one of its subsets. For $\boldsymbol{M}_{\mathrm{A}}$ such a matrix, the entry of $\boldsymbol{M}_{\mathrm{A}}$ associated with (say) indices $(\mathrm{p}_x, \mathrm{q}'_y)$ will be denoted by $\mathbf{m}^{\mathrm{A}}_{\mathrm{p}_x\mathrm{q}'_y}$; the same notational convention applies, mutatis mutandis, for other matrices names. Also, for $\boldsymbol{M}$ a matrix and $\vec{v}$ a vector, the quantity $\vec{v}^{\mathsf{T}}\boldsymbol{M}\vec{v}$ (that is, the application to $\vec{v}$ of the quadratic form encoded by $\boldsymbol{M}$) may also be denoted "$\boldsymbol{M}\cdot\vec{v}^{\otimes 2}$". (In other words, '$\cdot$' may refer to the Hilbert–Schmidt product). ♡

Proof of Theorem 125. First, observe that we can freely assume that t was taken small enough: indeed, if there exists some $t_1 > 0$ such that Theorem 125 is valid as long as $t \leqslant t_1$, then for $t > t_1$, we can use that $(\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=0) \to (\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=t_1) \to (\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=t)$ is a Markov chain, so that by Corollary 5, $\rho\big((\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=0); (\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=t)\big) \leqslant \rho\big((\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=0); (\vec{p}_{\mathbb{Z}}, \vec{q}_{\mathbb{Z}})(u=t_1)\big) < 1$. That being said, thanks to Lemma 91, Theorem 125 will be a consequence of Proposition 127 stated just below. ♤

Proposition 127 (Pairwise ρ-mixing for our model)**.** *Provided t is small enough, there exists an exponentially decaying function $\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+$ with $\varepsilon(\delta) < 1$ for all $\delta \in \mathbb{R}_+$,*[‡] *such that for all $x, y \in \mathbb{Z}$ (possibly identical), one has* $\rho^{\ddagger}(p_x; p'_y)$, $\rho^{\ddagger}(p_x; q'_y)$, $\rho^{\ddagger}(q_x; p'_y)$, $\rho^{\ddagger}(q_x; q'_y) \leqslant \varepsilon(|y-x|)$. ◇

The proof of Proposition 127 will involve a lot of estimates on (infinite) square matrices. Let us introduce some vocabulary to that end:

[‡] Note that here one must also have $\varepsilon(\delta = 0) < 1$!

Definition 128 (Coercivity, boundedness and exponential decay of square matrices)**.** For $\boldsymbol{M}$ a square matrix indexed by some $\mathbb{V} \subseteq \mathbb{Z}\times 4$, we will say that $\boldsymbol{M}$ is

- *coercive* if there exists $\alpha > 0$ such that $\boldsymbol{M} \geqslant \alpha \mathbf{I}_{\mathbb{V}}$ in the sense of quadratic forms;
- *bounded* if there exists $B < \infty$ such that $\boldsymbol{M} \leqslant B \mathbf{I}_{\mathbb{V}}$ in the sense of quadratic forms;[§]
- *exponentially decaying* if there exists an exponentially decaying function $\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+$ such that, for all $x, y \in \mathbb{Z}$,

$$\underbrace{|\mathfrak{m}|_{\mathrm{p}_x\mathrm{p}_y}, |\mathfrak{m}|_{\mathrm{p}_x\mathrm{q}_y}, |\mathfrak{m}|_{\mathrm{p}_x\mathrm{p}'_y}, \ldots, |\mathfrak{m}|_{\mathrm{q}'_x\mathrm{q}'_y}}_{\text{all 16 possibilities, as long as indices are valid}} \leqslant \varepsilon(|y-x|). \tag{PX}$$

(Note that an exponentially decaying matrix is always bounded). ♡

Proof of Proposition 127. Denote by $\boldsymbol{Q}_{\mathrm{H}}$ [‘H’ for “Hamiltonian”] the symmetric bilinear form such that $\mathscr{H}(\vec{\xi}) =: \frac{1}{2}\boldsymbol{Q}_{\mathrm{H}} \cdot \vec{\xi}^{\otimes 2}$, i.e.:

$$\mathfrak{q}^{\mathrm{H}}_{\mathrm{p}_x\mathrm{p}_x} := m^{-1}; \tag{PY}$$

$$\mathfrak{q}^{\mathrm{H}}_{\mathrm{q}_x\mathrm{q}_x} := m(\omega^2 + 2c^2); \tag{PZ}$$

$$\mathfrak{q}^{\mathrm{H}}_{\mathrm{q}_x\mathrm{q}_{x\pm 1}} := -mc^2, \tag{QA}$$

all the other entries of $\boldsymbol{Q}_{\mathrm{H}}$ being zero. This matrix governs the distribution of the initial state $\vec{\xi}$: more precisely, by the Maxwell–Boltzmann statistics and the properties of Gaussian variables, $\vec{\xi}$ is distributed according to the centred Gaussian law with covariance matrix $\beta^{-1}\boldsymbol{Q}_{\mathrm{H}}^{-1}$.[¶]

Now let us introduce a certain semigroup of linear endomorphisms on $\mathbb{R}^{\mathbb{Z}\times 2}$, denoted by $(\boldsymbol{T}^u)_{u\geqslant 0}$, which describes the evolution of the system *in absence of noise*. The definition of this semigroup is the following: if we removed the Brownian noise for the system, but still keeping the friction driven by parameter λ, then the state of the system at time u would be a deterministic linear function of the state at time 0: this is the function that we denote by $\boldsymbol{T}^u$ (when seen as a matrix). So, if there were no noise, one would have “$\vec{\xi}' = \boldsymbol{T}^t\vec{\xi}$”.

In practice, due to the noise, one has actually $\vec{\xi}' = \boldsymbol{T}^t\vec{\xi} + \vec{\nu}$, where the random variable $\vec{\nu} \in \mathbb{R}^{\mathbb{Z}\times 2}$ represents the effect of noise. As our model is fully linear, $\vec{\nu}$ is a centred normal vector and is independent of $\vec{\xi}$. Let us denote by $\beta^{-1}\boldsymbol{C}_{\mathrm{N}}$ [‘N’ for “noise”] the covariance matrix of $\vec{\nu}$, and set $\boldsymbol{Q}_{\mathrm{N}} := \boldsymbol{C}_{\mathrm{N}}^{-1}$ (though for the time being it is not clear that $\boldsymbol{Q}_{\mathrm{N}}$ actually exists). Since $\vec{\nu}$ is independent from $\vec{\xi}$, the global state $(\vec{\xi}; \vec{\xi}')$ of the system is distributed according to the (formal) density proportional to

---

[§] Of course the notions of coercivity and boundedness are primarily relevant for *symmetric* matrices.

[¶] The fact that $\boldsymbol{Q}_{\mathrm{H}}$ is actually invertible derives from the fact that it is a coercive symmetric matrix; but we will not actually need that point here.

$\exp\bigl(-\frac{1}{2}\beta\boldsymbol{Q}_{\mathrm{E}}\cdot(\vec{\xi};\vec{\xi}')^{\otimes 2}\bigr)$, where $\boldsymbol{Q}_{\mathrm{E}}$ ['E' for "evolution"] is the following symmetric bilinear form:

$$\boldsymbol{Q}_{\mathrm{E}}\cdot(\vec{\xi};\vec{\xi}')^{\otimes 2} \coloneqq \boldsymbol{Q}_{\mathrm{H}}\cdot\vec{\xi}^{\otimes 2} + \boldsymbol{Q}_{\mathrm{N}}\cdot(\vec{\xi}' - \boldsymbol{T}^t\vec{\xi})^{\otimes 2}. \tag{QB}$$

The matrix $\boldsymbol{Q}_{\mathrm{E}}$ satisfies two crucial properties, whose proofs are postponed:

Proposition 129. *The matrix $\boldsymbol{Q}_{\mathrm{E}}$ is coercive.* ◇

Proposition 130. *The matrix $\boldsymbol{Q}_{\mathrm{E}}$ is exponentially decaying (hence also bounded).* ◇

Now that we have a better view on the probability distribution of $(\vec{\xi};\vec{\xi}')$, we can turn to our primary goal, namely, bounding $\rho^{\ddagger}(\boldsymbol{X}_i;\boldsymbol{X}_j)$ for distinct $i,j\in\mathbb{Z}\times 4$. (Remember that, depending on what the type of index i is, notation "$\boldsymbol{X}_i$" may refer to any variable of the type $\boldsymbol{p}_x$, $\boldsymbol{q}_x$, $\boldsymbol{p}'_x$ or $\boldsymbol{q}'_x$). Denote by x and y the respective spatial indices that indices i and j refer to: i.e., i is of the form p_x, q_x, p'_x or q'_x, while j is of the form p_y, q_y, p'_y or q'_y.[‖] (Note that, contrary to i and j, x and y may be identical). For $\boldsymbol{K}\subseteq\mathbb{Z}\times 4$ not containing i nor j, our goal is to prove a bound of the form

$$\rho^{\dagger}_{\vec{X}_K}(X_i;X_j) \leqslant \varepsilon(|y-x|), \tag{QC}$$

where the function $\varepsilon(\bullet)$ must not depend on i, j nor $\boldsymbol{K}$, and has to be exponentially decaying, with $\varepsilon(\delta)<1$ for all δ.

$\rho^{\dagger}_{\vec{X}_K}(X_i;X_j)$ can be computed in the following way: let $\boldsymbol{Q}_{\mathrm{R}}$ ['R' for "restriction"] be the restriction of $\boldsymbol{Q}_{\mathrm{E}}$ to indices in $\mathbb{Z}\times 4\smallsetminus\boldsymbol{K}$; then $\beta^{-1}\boldsymbol{Q}_{\mathrm{R}}^{-1} \eqqcolon \beta^{-1}\boldsymbol{C}_{\mathrm{R}}$ is the covariance matrix of $\vec{\boldsymbol{X}}_{\mathbb{Z}\times 4\smallsetminus K}$ conditionally to fixing any value for $\vec{\boldsymbol{X}}_{\boldsymbol{K}}$; so that, by the properties of Gaussian vectors (Proposition 9):

$$\rho^{\dagger}_{\vec{X}_K}(X_i;X_j) = \frac{|\mathfrak{c}^{\mathrm{R}}_{ij}|}{\sqrt{\mathfrak{c}^{\mathrm{R}}_{ii}\mathfrak{c}^{\mathrm{R}}_{jj}}}. \tag{QD}$$

But, thanks to Propositions 129 and 130, we have constants $\alpha>0, B<\infty$ and an exponentially decaying function $\varepsilon_1\colon\mathbb{R}_+\to\mathbb{R}_+$ such that

$$\alpha\mathbf{I}_{\mathbb{Z}\times 4} \leqslant \boldsymbol{Q}_{\mathrm{E}} \leqslant B\mathbf{I}_{\mathbb{Z}\times 4}; \tag{QE}$$

$$\forall x,y \qquad \underbrace{|\mathfrak{q}^{\mathrm{E}}_{\mathrm{p}_x\mathrm{p}_y}|, |\mathfrak{q}^{\mathrm{E}}_{\mathrm{p}_x\mathrm{q}_y}|, \ldots, |\mathfrak{q}^{\mathrm{E}}_{\mathrm{q}'_x\mathrm{q}'_y}|}_{\text{16 possibilities}} \leqslant \varepsilon_1(|y-x|), \tag{QF}$$

so by restriction $\boldsymbol{Q}_{\mathrm{R}}$ satisfies the same properties (modulo taking into account that matrix indices are restricted to $\mathbb{Z}\times 4\smallsetminus\boldsymbol{K}$). Inverting $\boldsymbol{Q}_{\mathrm{R}}$, we find that

$$B^{-1}\mathbf{I}_{\mathbb{Z}\times 4\smallsetminus K} \leqslant \boldsymbol{C}_{\mathrm{R}} \leqslant \alpha^{-1}\mathbf{I}_{\mathbb{Z}\times 4\smallsetminus K}; \tag{QG}$$

$$\forall x,y \qquad \underbrace{|\mathfrak{c}^{\mathrm{R}}_{\mathrm{p}_x\mathrm{p}_y}|, |\mathfrak{c}^{\mathrm{R}}_{\mathrm{p}_x\mathrm{q}_y}|, \ldots, |\mathfrak{c}^{\mathrm{R}}_{\mathrm{q}'_x\mathrm{q}'_y}|}_{\text{for indices not in } K} \leqslant \varepsilon_2(|y-x|), \tag{QH}$$

[‖] Actually to prove Proposition 127 one only needs to consider the case where $i\in\mathbb{Z}\times\{\mathrm{p},\mathrm{q}\}$ and $j\in\mathbb{Z}\times\{\mathrm{p}',\mathrm{q}'\}$; but our proof does not care.

for an exponentially decaying function $\boldsymbol{\varepsilon}_2(\bullet)$ which can be determined directly from $\boldsymbol{\alpha}$ and $\boldsymbol{\varepsilon}_1(\bullet)$ (cf. Lemma 185 in appendix), and in particular does not depend on $\boldsymbol{K}$.

From (`QH`), we have got an exponential bound on $|\mathfrak{c}^{\mathrm{R}}_{ij}|$. And from (`QG`), by restriction we deduce that

$$\boldsymbol{B}^{-1}\mathbf{I}_2 \leqslant \begin{pmatrix} \mathfrak{c}^{\mathrm{R}}_{ii} & \mathfrak{c}^{\mathrm{R}}_{ij} \\ \mathfrak{c}^{\mathrm{R}}_{ij} & \mathfrak{c}^{\mathrm{R}}_{jj} \end{pmatrix} \leqslant \boldsymbol{\alpha}^{-1}\mathbf{I}_2, \tag{QI}$$

whence, on the one hand, $\mathfrak{c}^{\mathrm{R}}_{ii}, \mathfrak{c}^{\mathrm{R}}_{jj} \geqslant \boldsymbol{B}^{-1}$; and on the other hand, $\mathfrak{c}^{\mathrm{R}}_{ii}, \mathfrak{c}^{\mathrm{R}}_{jj} \leqslant \boldsymbol{\alpha}^{-1}$ and $\mathfrak{c}^{\mathrm{R}}_{ii}\mathfrak{c}^{\mathrm{R}}_{jj} - (\mathfrak{c}^{\mathrm{R}}_{ij})^2 \geqslant \boldsymbol{B}^{-2}$, hence $(\mathfrak{c}^{\mathrm{R}}_{ij})^2 \leqslant (1 - \boldsymbol{\alpha}^2 / \boldsymbol{B}^2)\mathfrak{c}^{\mathrm{R}}_{ii}\mathfrak{c}^{\mathrm{R}}_{jj}$.[**] So, in the end:

$$\frac{|\mathfrak{c}^{\mathrm{R}}_{ij}|}{\sqrt{\mathfrak{c}^{\mathrm{R}}_{ii}\mathfrak{c}^{\mathrm{R}}_{jj}}} \leqslant \boldsymbol{B}\boldsymbol{\varepsilon}_2(|y-x|) \wedge \left(1 - \frac{\boldsymbol{\alpha}^2}{\boldsymbol{B}^2}\right)^{1/2}, \tag{QJ}$$

which is an $\boldsymbol{\varepsilon}(|y-x|)$ of the wished form. ♤

Let us now check Propositions 129 and 130:

Proof of Proposition 129. Let $(\vec{\xi};\vec{\xi}') \underset{\text{def}}{=} (\vec{p}_{\mathbb{Z}};\vec{q}_{\mathbb{Z}};\vec{p}'_{\mathbb{Z}};\vec{q}'_{\mathbb{Z}}) \in \mathrm{L}^2(\mathbb{Z}\times 4)$. We have obviously $\|(\vec{\xi};\vec{\xi}')\|^2_{\mathrm{L}^2(\mathbb{Z}\times 4)} = \|\vec{\xi}\|^2_{\mathrm{L}^2(\mathbb{Z}\times 2)} + \|\vec{\xi}'\|^2_{\mathrm{L}^2(\mathbb{Z}\times 2)}$. Denoting by $\boldsymbol{Q}_{\mathrm{F}}$ ['F' for "free"] the matrix defined by the same formulae as (`PY`)–(`QA`) except that "$\boldsymbol{c}$" is replaced by $\mathbf{0}$ (which corresponds to the Hamiltonian of free oscillators) and noticing that $\boldsymbol{Q}_{\mathrm{H}} \geqslant \boldsymbol{Q}_{\mathrm{F}}$ in the sense of quadratic forms (because the interaction terms have positive energy), we observe that

$$\boldsymbol{Q}_{\mathrm{E}} \cdot (\vec{\xi};\vec{\xi}')^{\otimes 2} \underset{(\mathtt{QB})}{\geqslant} \boldsymbol{Q}_{\mathrm{H}} \cdot \vec{\xi}^{\otimes 2} \geqslant \boldsymbol{Q}_{\mathrm{F}} \cdot \vec{\xi}^{\otimes 2} \geqslant (m^{-1} \wedge m\omega^2)\|\vec{\xi}\|^2_{\mathrm{L}^2(\mathbb{Z}\times 2)}. \tag{QK}$$

From (`QK`), we also deduce that

$$\boldsymbol{Q}_{\mathrm{E}} \cdot (\vec{\xi};\vec{\xi}')^{\otimes 2} \geqslant (m^{-1} \wedge m\omega^2)\|\vec{\xi}'\|^2_{\mathrm{L}^2(\mathbb{Z}\times 2)}: \tag{QL}$$

indeed, despite Langevin dynamics not being reversible in the mathematical sense, under its stationary distribution it is invariant *in law* by *physical time reversal*, i.e. when you replace $\boldsymbol{u}$ by $-\boldsymbol{u}$ $\underline{\text{and } p_x \text{ by } -p_x}$ [18, Lemma 6.2]: and when applied to the distribution of $(\vec{\xi};\vec{\xi}')$, physical time reversibility gives that $\boldsymbol{Q}_{\mathrm{E}} \cdot (\vec{p}_{\mathbb{Z}};\vec{q}_{\mathbb{Z}};\vec{p}'_{\mathbb{Z}};\vec{q}'_{\mathbb{Z}})^{\otimes 2} = \boldsymbol{Q}_{\mathrm{E}} \cdot (-\vec{p}'_{\mathbb{Z}};\vec{q}'_{\mathbb{Z}};-\vec{p}_{\mathbb{Z}};\vec{q}_{\mathbb{Z}})^{\otimes 2}$, which combined with (`QK`) yields (`QL`). Averaging (`QK`) with (`QL`) finally gives the wished coercivity inequality, with coercivity constant $\frac{1}{2}(m^{-1} \wedge m\omega^2)$. ♤

Proof of Proposition 130. Since the sum, resp. the product of two exponentially decaying matrices are themselves exponentially decaying (cf. Lemma 184 in appendix), given Equation (`QB`) defining $\boldsymbol{Q}_{\mathrm{E}}$, in order to get exponential decay for $\boldsymbol{Q}_{\mathrm{E}}$ it will be sufficient to prove exponential decays for $\boldsymbol{Q}_{\mathrm{H}}$, $\boldsymbol{Q}_{\mathrm{N}}$ and $\boldsymbol{T}^t$.

$\boldsymbol{Q}_{\mathrm{H}}$'s exponential decay is trivial from Equations (`PY`)–(`QA`): actually it has even "compact support decay".

[**] Actually one could even prove that $(\mathfrak{c}^{\mathrm{R}}_{ij})^2 \leqslant (\frac{\boldsymbol{B}-\boldsymbol{\alpha}}{\boldsymbol{B}+\boldsymbol{\alpha}})^2\mathfrak{c}^{\mathrm{R}}_{ii}\mathfrak{c}^{\mathrm{R}}_{jj}$.

As regards $\boldsymbol{T}^t$, first observe that one has by definition $\boldsymbol{T}^t = \exp(t\boldsymbol{A})$, where the generator $\boldsymbol{A}$ (defined as the operator such that one would have "$(\dot{\vec{p}}_{\mathbb{Z}}(u); \dot{\vec{q}}_{\mathbb{Z}}(u)) = \boldsymbol{A}(\vec{p}_{\mathbb{Z}}(u); \vec{q}_{\mathbb{Z}}(u))$" in absence of noise) is the matrix having the following nonzero entries:

$$\mathfrak{a}_{\mathrm{p}_x\mathrm{p}_x} := -\lambda; \tag{QM}$$

$$\mathfrak{a}_{\mathrm{p}_x\mathrm{q}_x} := -m\omega^2 - 2mc^2; \tag{QN}$$

$$\mathfrak{a}_{\mathrm{p}_x\mathrm{q}_{x\pm1}} := mc^2; \tag{QO}$$

$$\mathfrak{a}_{\mathrm{q}_x\mathrm{p}_x} := m^{-1}, \tag{QP}$$

all the other entries being zero. The matrix $\boldsymbol{A}$ has compact support decay, hence exponential decay; therefore so has $\boldsymbol{T}^t$, since the exponential of an exponentially decaying matrix is itself exponentially decaying (cf. Lemma 184 again).

Lastly, as regards the matrix $\boldsymbol{Q}_{\mathrm{N}}$, remember that it was defined as the inverse of $\boldsymbol{C}_{\mathrm{N}}$. But the latter satisfies the following properties (which we will prove a few lines below):

Proposition 131. *The matrix $\boldsymbol{C}_{\mathrm{N}}$ is coercive and exponentially decaying.* ◇

Since the inverse of an exponentially decaying coercive matrix (exists and) is itself exponentially decaying (cf. Lemma 185 in appendix), we deduce that $\boldsymbol{Q}_{\mathrm{N}}$ (exists and) is exponentially decaying, which ends the proof of Proposition 130. ♤

Proof of Proposition 131. An expression of the covariance matrix due to the noise, hence of $\boldsymbol{C}_{\mathrm{N}}$, can be obtained by integrating the quadratic variation of the white noise (since the elementary noises for different instants are independent), taking the deterministic flow into account:

$$\boldsymbol{C}_{\mathrm{N}} = 2\lambda m \int_0^t \boldsymbol{T}^{t-u}\mathbf{I}_{\mathrm{p}}(\boldsymbol{T}^{t-u})^{\top}\,\mathrm{d}u \underset{u\leftarrow t-u}{=} 2\lambda m \int_0^t \boldsymbol{T}^{u}\mathbf{I}_{\mathrm{p}}(\boldsymbol{T}^{u})^{\top}\,\mathrm{d}u, \tag{QQ}$$

where $\mathbf{I}_{\mathrm{p}}$ refers to the diagonal matrix indexed by $\mathbb{Z}\times 2$ whose entries are 1 at indices of the form p_x and 0 at indices q_x.

For u small, one can compute formal limited expansions for all the entries of $\boldsymbol{T}^u$. One finds that, at least formally:

$$\mathfrak{t}^u_{\mathrm{p}_x\mathrm{p}_y} = \frac{c^{2|y-x|}}{(2|y-x|)!}u^{2|y-x|} + \mathrm{O}\left(\frac{c^{2|y-x|}}{(2|y-x|)!}u^{2|y-x|+1}\right); \tag{QR}$$

$$\mathfrak{t}^u_{\mathrm{p}_x\mathrm{q}_x} = -m(\omega^2 + 2c^2)u + \mathrm{O}(u^2); \tag{QS}$$

$$\mathfrak{t}^u_{\mathrm{p}_x\mathrm{q}_{y\neq x}} = \frac{mc^{2|y-x|}}{(2|y-x|-1)!}u^{2|y-x|-1} + \mathrm{O}\left(\frac{c^{2|y-x|}}{(2|y-x|-1)!}u^{2|y-x|}\right); \tag{QT}$$

$$\mathfrak{t}^u_{\mathrm{q}_x\mathrm{p}_y} = \frac{m^{-1}c^{2|y-x|}}{(2|y-x|+1)!}u^{2|y-x|+1} + \mathrm{O}\left(\frac{c^{2|y-x|}}{(2|y-x|+1)!}u^{2|y-x|+2}\right); \tag{QU}$$

$$\mathfrak{t}^u_{\mathrm{q}_x\mathrm{q}_x} = 1 + \mathrm{O}(u^2); \tag{QV}$$

$$\mathfrak{t}^u_{\mathrm{q}_x\mathrm{q}_{y\neq x}} = \frac{c^{2|y-x|}}{(2|y-x|)!}u^{2|y-x|} + \mathrm{O}\left(\frac{c^{2|y-x|}}{(2|y-x|)!}u^{2|y-x|+1}\right). \tag{QW}$$

I claim that these formal limited developments can be turned into rigorous estimates, in the following sense: there exist constants $t_1 > 0$ and $C < \infty$[††] such that, provided $u \leqslant t_1$, one has

$$\left|\mathbf{t}^u_{\mathrm{p}_x\mathrm{p}_y} - \frac{c^{2|y-x|}}{(2|y-x|)!}u^{2|y-x|}\right| \leqslant C\frac{(|y-x|+1)c^{2|y-x|}}{(2|y-x|+1)!}u^{2|y-x|+1}; \tag{QX}$$

$$\left|\mathbf{t}^u_{\mathrm{p}_x\mathrm{q}_x} + m(\omega^2+2c^2)u\right| \leqslant C\frac{m(\omega^2+2c^2)}{2}u^2; \tag{QY}$$

$$\left|\mathbf{t}^u_{\mathrm{p}_x\mathrm{q}_{y\neq x}} - \frac{mc^{2|y-x|}}{(2|y-x|-1)!}u^{2|y-x|-1}\right| \leqslant C\frac{|y-x|mc^{2|y-x|}}{(2|y-x|)!}u^{2|y-x|}; \tag{QZ}$$

$$\left|\mathbf{t}^u_{\mathrm{q}_x\mathrm{p}_y} - \frac{m^{-1}c^{2|y-x|}}{(2|y-x|+1)!}u^{2|y-x|+1}\right| \leqslant C\frac{|y-x|m^{-1}c^{2|y-x|}}{(2|y-x|+2)!}u^{2|y-x|+2}; \tag{RA}$$

$$|\mathbf{t}^u_{\mathrm{q}_x\mathrm{q}_x} - 1| \leqslant C\frac{\lambda^{-1}(\omega^2+2c^2)}{2}u^2; \tag{RB}$$

$$\left|\mathbf{t}^u_{\mathrm{q}_x\mathrm{q}_{y\neq x}} - \frac{c^{2|y-x|}}{(2|y-x|)!}u^{2|y-x|}\right| \leqslant C\frac{|y-x|c^{2|y-x|}}{(2|y-x|+1)!}u^{2|y-x|+1}. \tag{RC}$$

To check Equations (QX)–(RC), let us mimic the proof of Grönwall's lemma. As an exponential semigroup, $(\boldsymbol{T}^u)_{u\leqslant t_1}$ can be obtained by iterating infinitely many times (starting from $\mathbf{0}$) the map

$$(\boldsymbol{T}^u)_{u\leqslant t_1} \mapsto \left(\mathbf{I}_{\mathbb{Z}\times 2} + \int_{v=0}^{u} \boldsymbol{A}\boldsymbol{T}^v\,\mathrm{d}v\right)_{u\leqslant t_1}: \tag{RD}$$

so it will be enough for us to check that, whenever Equations (QX)–(RC) are satisfied for a given $(\boldsymbol{T}^u)_{u\leqslant t_1}$, they remain satisfied after iterating (RD). Clearly it is actually sufficient to check the equations in the case $u \leftarrow t_1$: so, starting from a trajectory $(\boldsymbol{T}^u)_{u\leqslant t_1}$ satisfying (QX)–(RC), we want to check that (provided t_1 and C were chosen in a clever way) these equations are still valid when replacing '$\boldsymbol{T}^u$' by $\mathbf{I}_{\mathbb{Z}\times 2} + \int_{u=0}^{t_1} \boldsymbol{A}\boldsymbol{T}^u\,\mathrm{d}u$ (and 'u' by t_1).

Below I will only detail how you check that Equation (QZ) remains satisfied. (It is the most complicated case: as regards the other equations, the treatment would be similar, and slightly simpler). So, let $(\boldsymbol{T}^u)_{u\leqslant t_1}$ satisfy (QX)–(RC); our goal is to check that (QZ) is also satisfied for the iteration of $(\boldsymbol{T}^u)_{u\leqslant t_1}$ by (RD) (taken at time t_1). To fix ideas, say that we are in the case $y > x$, and let $z \coloneqq y - x$.

Given the sparse structure of matrix $\boldsymbol{A}$, the $(\mathrm{p}_x, \mathrm{q}_y)$ entry of $\boldsymbol{A}\boldsymbol{T}^u$ is merely

$$\begin{aligned}[\boldsymbol{A}\boldsymbol{T}^u]_{\mathrm{p}_x\mathrm{q}_y} &= \mathfrak{a}_{\mathrm{p}_x\mathrm{p}_x}\mathbf{t}^u_{\mathrm{p}_x\mathrm{q}_y} + \mathfrak{a}_{\mathrm{p}_x\mathrm{q}_{x-1}}\mathbf{t}^u_{\mathrm{q}_{x-1}\mathrm{q}_y} + \mathfrak{a}_{\mathrm{p}_x\mathrm{q}_x}\mathbf{t}^u_{\mathrm{q}_x\mathrm{q}_y} + \mathfrak{a}_{\mathrm{p}_x\mathrm{q}_{x+1}}\mathbf{t}^u_{\mathrm{q}_{x+1}\mathrm{q}_y} \\ &= -\lambda\mathbf{t}^u_{\mathrm{p}_x\mathrm{q}_y} + mc^2\mathbf{t}^u_{\mathrm{q}_{x-1}\mathrm{q}_y} - m(\omega^2+2c^2)\mathbf{t}^u_{\mathrm{q}_x\mathrm{q}_y} + mc^2\mathbf{t}^u_{\mathrm{q}_{x+1}\mathrm{q}_y}. \end{aligned} \tag{RE}$$

In the right-hand side of (RE), the dominant term will be the last one, which according to (QW) (or (QV) in the case $z = 1$) is approximately equal to $mc^{2z}u^{2z-2} \,/\, (2z-2)!$:

[††] Here note that there is just <u>one</u> constant C, common to all expressions!

more precisely, if (`QX`)–(`RC`) are satisfied, one has the following control:[‡‡]

$$\begin{aligned}\left|[\boldsymbol{AT}^u]_{\mathrm{p}_x\mathrm{q}_y} - \frac{mc^{2z}}{(2z-2)!}u^{2z-2}\right| &\leqslant \frac{\lambda mc^{2z}}{(2z-1)!}u^{2z-1} + \frac{mc^{2z+4}}{(2z+2)!}u^{2z+2} + \frac{m(\omega^2+2c^2)c^{2z}}{(2z)!}u^{2z} \\ &+ C\left(\frac{z\lambda mc^{2z}}{(2z)!}u^{2z} + \frac{(z+1)mc^{2z+4}}{(2z+3)!}u^{2z+3} + \frac{zm(\omega^2+2c^2)c^{2z}}{(2z+1)!}u^{2z+1} + \frac{(z-1)mc^{2z}}{(2z-1)!}u^{2z-1}\right) \\ &\leqslant \frac{\lambda mc^{2z}}{(2z-1)!}u^{2z-1} + C\frac{(z-1)mc^{2z}}{(2z-1)!}u^{2z-1} + \frac{m(\omega^2+2c^2)c^{2z}}{(2z)!}u^{2z} + \frac{mc^{2z+4}}{(2z+2)!}u^{2z+2} \\ &\qquad + \tfrac{1}{2}Cu\left(\frac{\lambda mc^{2z}}{(2z-1)!}u^{2z-1} + \frac{m(\omega^2+2c^2)c^{2z}}{(2z)!}u^{2z} + \frac{mc^{2z+4}}{(2z+2)!}u^{2z+2}\right). \qquad (\mathtt{RF})\end{aligned}$$

The last step is now to show that, provided t_1 and C were chosen cleverly, the right-hand-side of (`RF`) is bounded above by $Czmc^{2z}u^{2z-1} / (2z-1)!$ (with the same value of C as before!): and then, it will just remain to integrate from 0 to t_1 to get (`QZ`) for the iterated matrix. To achieve the wanted bound on the right-hand side of (`RF`), let $\varepsilon > 0$ be a small value (to be fixed later). If one ensures that $t_1 \leqslant \varepsilon c^{-1}$, the expression $(\varepsilon^{-1}cu)^k / k!$ is a nonincreasing function of k; therefore one has

$$\begin{aligned}&\left|[\boldsymbol{AT}^u]_{\mathrm{p}_x\mathrm{q}_y} - \frac{mc^{2z}}{(2z-2)!}u^{2z-2}\right| \leqslant \\ &\Big(Cz - C + \lambda\big(1+\tfrac{1}{2}Ct_1\big) + (\omega^2c^{-1}+2c)\big(1+\tfrac{1}{2}Ct_1\big)\varepsilon + c(1+\tfrac{1}{2}Ct_1)\varepsilon^3\Big) \times \frac{mc^{2z}u^{2z-1}}{(2z-1)!}. \qquad (\mathtt{RG})\end{aligned}$$

Provided $t_1 \leqslant 2C^{-1}$, this is in turn bounded above by

$$\Big(Cz - C + 2\lambda + (2\omega^2c^{-1}+4c)\varepsilon + 2c\varepsilon^3\Big) \times \frac{mc^{2z}u^{2z-1}}{(2z-1)!}: \qquad (\mathtt{RH})$$

so, it suffices to choose successively $C \geqslant 3\lambda$, ε small enough so that $(2\omega^2c^{-1}+4c)\varepsilon + 2c\varepsilon^3 \leqslant \lambda$, and $t_1 \leqslant \varepsilon c^{-1} \wedge 2C^{-1}$; and this will ensure that the wished bound is satisfied.[*]

In conclusion of the reasoning started one page ago, we have established that it is true indeed that (`QR`)–(`QW`) hold; and moreover, the "O(•)"'s in these estimates can be made uniform in x, y and u (for u small enough), in the sense that there exist constants $t_1 > 0$ and $C < \infty$ such that, provided $u \leqslant t_1$, each of the expressions of the form "$\mathrm{O}\big(g(x,y,u)\big)$" in (`QR`)–(`QW`) can be bounded in absolute value by $Cg(x,y,u)$, with the same constant C for all expressions.

[‡‡] In the case $z = 1$, at one point one would have to use (`RB`) instead of (`RC`), which would lead to replace the "$C(z-1)mc^{2z}u^{2z-1} / (2z-1)!$" term in (`RF`) by "$C\lambda^{-1}m(\omega^2+2c^2)c^2u^2 / 2$": but this would not be a problem, as it is a term of higher order.

[*] Here there are two important points to be noted. The first one is that our choice of t_1 and C did not depend on z. The second one is that, in practice, one may have extra constraints of the same type to be satisfied by C, ε and t_1, due to the necessity to satisfy all six equations (`QX`)–(`RC`): but, given the way that we picked successively C, ε and t_1, satisfying a finite number of such constraints at each step would not be a problem! ◡̈

Plugging (QR)–(QW) into (QQ), after a few computations (a flavour of which is given just below) one finds that

$$\mathfrak{c}^{\mathrm{N}}_{\mathrm{p}_x\mathrm{p}_x} = 2\lambda m t + \mathrm{O}(t^2); \tag{RI}$$

$$\mathfrak{c}^{\mathrm{N}}_{\mathrm{p}_x\mathrm{p}_{y\neq x}} = \mathrm{O}\biggl(\frac{(2c)^{2|y-x|}}{(2|y-x|+1)!}t^{2|y-x|+1}\biggr); \tag{RJ}$$

$$\mathfrak{c}^{\mathrm{N}}_{\mathrm{p}_x\mathrm{q}_x} = \mathfrak{c}^{\mathrm{N}}_{\mathrm{q}_x\mathrm{p}_x} = \lambda t^2 + \mathrm{O}(t^3); \tag{RK}$$

$$\mathfrak{c}^{\mathrm{N}}_{\mathrm{p}_x\mathrm{q}_{y\neq x}} = \mathfrak{c}^{\mathrm{N}}_{\mathrm{q}_{y\neq x}\mathrm{p}_x} = \mathrm{O}\biggl(\frac{(2c)^{2|y-x|}}{(2|y-x|+2)!}t^{2|y-x|+2}\biggr); \tag{RL}$$

$$\mathfrak{c}^{\mathrm{N}}_{\mathrm{q}_x\mathrm{q}_x} = \tfrac{2}{3}\lambda m^{-1}t^3 + \mathrm{O}(t^4); \tag{RM}$$

$$\mathfrak{c}^{\mathrm{N}}_{\mathrm{q}_x\mathrm{q}_{y\neq x}} = \mathrm{O}\biggl(\frac{(2c)^{2|y-x|}}{(2|y-x|+3)!}t^{2|y-x|+3}\biggr), \tag{RN}$$

where again all the “$\mathrm{O}(\bullet)$”’s are uniform in x, y and t small enough.

Let me detail how, for instance, (the uniform version of) (RL) was derived. (This is arguably the most complicated case). The $(\mathrm{p}_x, \mathrm{q}_y)$ entry of $\boldsymbol{T}^u\mathbf{I}_{\mathrm{p}}(\boldsymbol{T}^u)^{\mathsf{T}}$ (which matrix appears in Equation (QQ) relating $\boldsymbol{C}_{\mathrm{N}}$ to the $\boldsymbol{T}^u$’s) expands into

$$[\boldsymbol{T}^u\mathbf{I}_{\mathrm{p}}(\boldsymbol{T}^u)^{\mathsf{T}}]_{\mathrm{p}_x\mathrm{q}_y} \underset{\mathrm{def}}{=} \sum_{w\in\mathbb{Z}} \mathfrak{t}^u_{\mathrm{p}_x\mathrm{p}_w}\mathfrak{t}^u_{\mathrm{q}_y\mathrm{p}_w}. \tag{RO}$$

From the uniform versions of (QR) and (QU), one can find a constant $C < \infty$ (independent from x, y and u—provided that u is taken lower than some t_1 independent from x and y) such that (RO) is bounded above, in absolute value, by

$$\begin{aligned}\sum_{w\in\mathbb{Z}}\biggl(C\frac{c^{2|w-x|}}{(2|w-x|)!}u^{2|w-x|} \times C\frac{m^{-1}c^{2|w-y|}}{(2|w-y|+1)!}u^{2|w-y|+1}\biggr) \\ = C'\sum_{w\in\mathbb{Z}}\frac{c^{2|w-x|+2|w-y|+1}}{(2|w-x|)!(2|w-y|+1)!}u^{2|w-x|+2|w-y|+1},\end{aligned} \tag{RP}$$

where in the right-hand-side we set $C' \coloneqq C^2m^{-1}c^{-1}$.

Now, assume to fix ideas that $x < y$, and let $z \coloneqq y - x$. I decompose the sum in the right-hand side of (RP) according to three possibilities: either $w < x$, or $w \in [\![x, y]\!]$, or $w > y$. Let us start with the “$w < x$” part. Since we are only interested in short times, we may assume that one has always $u \leqslant c^{-1} / 2$, so that the function $k \mapsto (2c)^k u^k / k!$ is nonincreasing. Then, one has

$$\begin{aligned}&\sum_{w=-\infty}^{x-1}\frac{c^{2|w-x|+2|w-y|+1}}{(2|w-x|)!(2|w-y|+1)!}u^{2|w-x|+2|w-y|+1} \\ &\quad= \sum_{w=-\infty}^{x-1}\biggl(\frac{c^{2(x-w)}}{\bigl(2(x-w)\bigr)!}u^{2(x-w)} \times \frac{c^{2(y-w)+1}}{\bigl(2(y-w)+1\bigr)!}u^{2(y-w)+1}\biggr) \\ &\quad\leqslant \sum_{w=-\infty}^{x-1}\biggl(2^{-2(x-w)} \times 2^{-2(x-w)}\frac{c^{2(y-x)+1}}{\bigl(2(y-x)+1\bigr)!}u^{2(y-x)+1}\biggr) = \frac{1}{3}\frac{c^{2z+1}}{(2z+1)!}u^{2z+1}.\end{aligned} \tag{RQ}$$

Likewise, the "$w > y$" part is bounded by $c^{2z+1}u^{2z+1} / 3(2z+1)!$ too.

As regards the "$w \in [\![x, y]\!]$" part, in this case $2|w-x| + 2|w-y| + 1$ is always equal to $2z+1$; therefore, making the change of variables "$k \leftarrow 2(w-x)$" (so that k will run on the *even* values of $[\![0, 2z+1]\!]$), one has

$$\begin{aligned}\sum_{w=x}^{y} \frac{c^{2|w-x|+2|w-y|+1}}{(2|w-x|)!(2|w-y|+1)!} u^{2|w-x|+2|w-y|+1} &= \sum_{k\,\mathrm{even}} \frac{c^{2z+1}}{k!(2z+1-k)!} u^{2z+1} \\ &= \sum_{k\,\mathrm{even}} \binom{2z+1}{k} \frac{c^{2z+1}}{(2z+1)!} u^{2z+1} = \frac{1}{2} \frac{(2c)^{2z+1}}{(2z+1)!} u^{2z+1}, \qquad (\mathtt{RR})\end{aligned}$$

where we used the well-known fact that the even-indexed values of row number n in Pascal's triangle add up to 2^{n-1}.

In the end, we have a bound of $|[\boldsymbol{T}^u \mathbf{I}_{\mathrm{p}} (\boldsymbol{T}^u)^\top]_{\mathrm{p}_x \mathrm{q}_y}|$ of the form $C(2c)^{2z+1}u^{2z+1}/(2z+1)!$ (for a value of C different from the original one, but still independent from x, y and u), which, after injection in (QQ) and integration, yields indeed a uniform version of (RL). The other estimates in (RI)–(RN) can be obtained in a similar fashion.

Thanks to (RI)–(RN), $\boldsymbol{C}_{\mathrm{N}}$ can be seen as a perturbation of the matrix $\boldsymbol{C}_{\mathrm{L}}$ ['L' for "linearized"], which we define by Equations (RI)–(RN) where all the "$\mathrm{O}(\bullet)$" terms are replaced by 0. Let us denote $\boldsymbol{D} \coloneqq \boldsymbol{C}_{\mathrm{N}} - \boldsymbol{C}_{\mathrm{L}}$. To cope with the fact that, for fixed x, y, the values $\boldsymbol{\mathfrak{d}}_{\mathrm{p}_x\mathrm{p}_y}$, $\boldsymbol{\mathfrak{d}}_{\mathrm{p}_x\mathrm{q}_y}$, $\boldsymbol{\mathfrak{d}}_{\mathrm{q}_x\mathrm{p}_y}$ and $\boldsymbol{\mathfrak{d}}_{\mathrm{q}_x\mathrm{q}_y}$ do not have the same order of magnitude, let us introduce the "homogeneity matrix" $\boldsymbol{H} \coloneqq \mathbf{I}_{\mathrm{p}} + m^{-1}t\mathbf{I}_{\mathrm{q}}$ (i.e. the diagonal matrix whose entries are 1 at indices p_x, resp. $m^{-1}t$ at indices q_x); and denote $\tilde{\boldsymbol{D}} \coloneqq \boldsymbol{H}^{-1}\boldsymbol{D}\boldsymbol{H}^{-1}$, resp. $\tilde{\boldsymbol{C}}_{\mathrm{N}} \coloneqq \boldsymbol{H}^{-1}\boldsymbol{C}_{\mathrm{N}}\boldsymbol{H}^{-1}$, etc. From the estimates (RI)–(RN) we have the following control on the entries of $\tilde{\boldsymbol{D}}$:

$$\tilde{\boldsymbol{\mathfrak{d}}}_{\mathrm{p}_x\mathrm{p}_x}, \tilde{\boldsymbol{\mathfrak{d}}}_{\mathrm{p}_x\mathrm{q}_x}, \tilde{\boldsymbol{\mathfrak{d}}}_{\mathrm{q}_x\mathrm{p}_x}, \tilde{\boldsymbol{\mathfrak{d}}}_{\mathrm{q}_x\mathrm{q}_x} = \mathrm{O}(t^2); \qquad (\mathtt{RS})$$

$$\tilde{\boldsymbol{\mathfrak{d}}}_{\mathrm{p}_x\mathrm{p}_{y\neq x}}, \tilde{\boldsymbol{\mathfrak{d}}}_{\mathrm{p}_x\mathrm{q}_{y\neq x}}, \tilde{\boldsymbol{\mathfrak{d}}}_{\mathrm{q}_x\mathrm{p}_{y\neq x}}, \tilde{\boldsymbol{\mathfrak{d}}}_{\mathrm{q}_x\mathrm{q}_{y\neq x}} = \mathrm{O}\left(\frac{(2c)^{2|y-x|}}{(2|y-x|+1)!} t^{2|y-x|+1} \right), \qquad (\mathtt{RT})$$

with again "$\mathrm{O}(\bullet)$" terms uniform in x, y and t small enough. In particular, we readily see that $\tilde{\boldsymbol{D}}$ is exponentially decaying; therefore, using Lemma 184 twice again, so is $\tilde{\boldsymbol{C}}_{\mathrm{N}} = \tilde{\boldsymbol{C}}_{\mathrm{L}} + \tilde{\boldsymbol{D}}$ (since $\tilde{\boldsymbol{C}}_{\mathrm{L}}$ is compact-support decaying), and thus so is $\boldsymbol{C}_{\mathrm{N}} = \boldsymbol{H}\tilde{\boldsymbol{C}}_{\mathrm{N}}\boldsymbol{H}$ (since $\boldsymbol{H}$ is a bounded diagonal matrix).

Also, (RS)-(RT) imply that, in the sense of quadratic forms, $-\tilde{\boldsymbol{D}}$ is bounded in the following way:

$$-\tilde{\boldsymbol{D}} \leqslant 2\left(\mathrm{O}(t^2) + \sum_{z \in \mathbb{Z} \smallsetminus \{0\}} \mathrm{O}\left(\frac{(2c)^{2|z|}}{(2|z|+1)!} t^{2|z|+1} \right) \right) \mathbf{I}_{\mathbb{Z}\times 2} \overset{t\to 0}{\leqslant} Bt^2 \mathbf{I}_{\mathbb{Z}\times 2} \qquad (\mathtt{RU})$$

for some constant $B < \infty$. On the other hand, given that $\tilde{\boldsymbol{C}}_{\mathrm{L}}$ is "blockwise scalar", one computes that

$$\tilde{\boldsymbol{C}}_{\mathrm{L}} \geqslant \tfrac{4-\sqrt{13}}{3} \lambda m t \mathbf{I}_{\mathbb{Z}\times 2} \geqslant \tfrac{1}{8} \lambda m t \mathbf{I}_{\mathbb{Z}\times 2}, \qquad (\mathtt{RV})$$

where the factor $(4-\sqrt{13})/3$ comes as the smallest eigenvalue of the positive-definite 2×2 matrix $(\!(2, 1; 1, 2/3)\!)$. But for t small enough, the "Bt^2" in the right-hand side of (`RU`) is strictly smaller than the "$\frac{1}{8}\boldsymbol{\lambda m t}$" in the right-hand side of (`RV`): and then by subtraction $\tilde{\boldsymbol{C}}_{\mathrm{N}} = \tilde{\boldsymbol{C}}_{\mathrm{L}} + \tilde{\boldsymbol{D}}$ is $\left(\frac{1}{8}\boldsymbol{\lambda m t} - \boldsymbol{B}\boldsymbol{t}^2\right)$-coercive, hence $\boldsymbol{C}_{\mathrm{N}} = \boldsymbol{H}\tilde{\boldsymbol{C}}_{\mathrm{N}}\boldsymbol{H}$ is $(1 \wedge \boldsymbol{m}^{-1}\boldsymbol{t})^2\left(\frac{1}{8}\boldsymbol{\lambda m t} - \boldsymbol{B}\boldsymbol{t}^2\right)$-coercive, which ends our proof. ♤

Chapter 4

Other applications of tensorization techniques

In the two previous chapters we have been seeing how ρ-mixing assumptions between pairs of random variables could yield “global” results on an arbitrary number of variables, by splitting functions of several variables into relevant telescopic sums. I used the word “tensorization” to qualify these results, as the conclusions were of the same nature as the assumptions.

But the techniques of Chapter 2 can also be applied to get other types of results. In this chapter I am going to show how, from ρ-mixing assumptions, one can get results on some classical features of particle systems which are not linked with ρ-mixing at first sight.

I will deal with two such features. First, in § 4.1 I will look at the implications of ρ-mixing on the existence of a *central limit theorem*—more precisely, a *spatial* central limit theorem, since I am more interested in random fields than in sequences (i.e., in variables indexed by $\mathbb{Z}^d$ rather than by $\mathbb{Z}$). Very sharp results concerning this issue are already known; however, I found it interesting to show how it goes via generalized tensorization of ρ-mixing: this approach takes indeed a quite different way to do the job, which may be neater by certain sides.

Next, in § 4.2 I will look at the question of *spectral gap* for *Glauber dynamics*. Though this point has already been thoroughly studied in the β-mixing paradigm, the present work, to the best of my knowledge, is the first to show how ρ-mixing can be used to tackle it.

My main goal here is just to show <u>how</u> the techniques of ρ-mixing tensorization can be applied to the problems of spatial central limit theorem and convergence of the Glauber dynamics: accordingly, I have favoured simplicity of the proofs over refinement of the results.

4.1 Spatial central limit theorem

4.1.1 Introduction

The central limit theorem (CLT) is one of the most fundamental results in probability theory. In its standard statement, the CLT assumes complete independence between the random variables involved: it is then very natural to wonder whether that assumption can be relaxed into some kind of asymptotic independence. The paradigm of ρ-mixing is a natural frame for such a generalization, since the CLT already takes place in an L^2 setting.

The viewpoint that I will follow here is motivated by statistical physics. Let $\mathbb{Z}^d$ be a lattice, on each vertex i of which there is a random "spin" X_i ranged in some space $\mathscr{X}$. We assume that the law of the system is translation invariant, i.e. that for all $u \in \mathbb{Z}^d$, $(X_{i+u})_{i\in\mathbb{Z}^d}$ has the same law as $(X_i)_{i\in\mathbb{Z}^d}$. Then, for all $u \in \mathbb{Z}^d$, we set

$$\rho_u \coloneqq \mathrm{\rho}^{\ddagger}(X_i; X_{i+u}). \tag{RW}$$

We are interested in situations where the ρ_u's are sufficiently "rapidly decreasing" as $|u| \to \infty$ so that $\sum_{u\in\mathbb{Z}^d} \rho_u < \infty$.

Under these conditions, let $f \colon \mathscr{X} \to \mathbb{R}$ be a function such that $f(X_0)$ is square-integrable and centred.[*] The question is, does one get a CLT when summing $f(X_i)$ for i in a large subset of $\mathbb{Z}^d$, i.e., does the sum grow as the square root of the number of its terms and have asymptotically normal distribution? For instance, we would like the law of the variable

$$\frac{1}{\sqrt{L^d}} \sum_{i\in[\![0,L[\![^d} f(X_i) \tag{RX}$$

to (weakly) converge, when $L \to \infty$, to some (centred) Gaussian distribution.

Remark 132. Note that, in such a context, the limiting distribution (if it exists) will not, in general, have its variance equal to $\mathrm{Var}(f(X_0))$, because of the nonzero covariance between $f(X_i)$ and $f(X_j)$ for i, j close spins. ♧

In the case $d = 1$, extremely sharp results on this topic have been known from long: let us cite, among many others, [19, 20, 21, 22]. For $d \geqslant 2$, similar results do also exist: see e.g. [21, Theorem 5] for an example of such a statement, or [12, Ch. 29] for a more general survey of the topic. In all these results, the notion of *uniform integrability* plays a key role in the proofs: as a consequence, the theorems' assumptions naturally require some more-than-L^2 integrability of the variables involved and/or some control on the decay speed of the ρ-mixing coefficients.

On the other hand, the reasoning that I will expose below mimics Lévy's proof of the CLT via characteristic functions, hence needing no uniform integrability at all. Note however that this does not imply that the assumptions of my own theorems would be *intrinsically* weaker than those of the classical ones, because my CLTs are

---

[*] Actually we will also cover the case where "$f(X_i)$" is replaced by a function depending on X_i *and on a finite number of neighbouring spins*: for instance, in the case the X_i's are real-valued, it could be something like "$X_i X_{i+e_0} X_{i+e_1}$", where $\boldsymbol{e}_0$ and $\boldsymbol{e}_1$ denote the vectors of $\mathbb{Z}^d$ with respective coordinates $(1, 0, \dots, 0)$ and $(0, 1, 0, \dots, 0)$.

stated in a slightly different vocabulary: indeed, while the classical results were stated under assumptions of ρ^*-mixing type (i.e., you directly assume that the ρ-mixing coefficient between *bunches* of spins tends to zero, in a way that does not depend on the size of the bunches, but just on their distance), here, in accordance with the idea of tensorization, my assumptions will rather be of $\rho^\ddagger$-mixing type: i.e., you just look at the ρ-mixing coefficient between *pairs* of spins, but that coefficient has to remain controlled even when conditioning by some other spins... We have seen in Lemma 91 that summability of the $\rho^\ddagger$-mixing coefficients implies ρ^*-mixing; so, the conditions in my CLTs are actually slightly stronger than in the classical results: but as a counterpart, the statements of my results get quite neater! ☺

All in all, it is unclear whether the results presented here do improve the state of the art; nevertheless, it is at least likely that, when turning to quantitative versions of these results, the difference between the usual methods and mine would yield some difference in the corresponding non-asymptotic bounds.

4.1.2 Product of weakly coupled variables

The key tool to prove my spatial CLT will be the following

Lemma 133. *Let $d \in \mathbb{N}^*$ and let $(X_i)_{i\in\mathbb{Z}^d}$ be random variables. For all $u \in \mathbb{Z}^d$, denote $\rho_u \coloneqq \sup_{i\in\mathbb{Z}^d} \mathrm{\rho}^\ddagger(X_i; X_{i+u})$ (where $\rho^\ddagger$-mixing is to be considered relative to the family of the X_i's), so that $\mathrm{\rho}^\ddagger(X_i; X_j) \leqslant \rho_{j-i}$ for all $i, j \in \mathbb{Z}^d$. Assume that the sum of the ρ_u's is convergent, and denote $\bar{\rho} \coloneqq \sum_{u\in\mathbb{Z}^d\smallsetminus\{0\}} \rho_u < \infty$.*

In this context, if $(\Phi_i)_{i\in\mathbb{Z}^d}$ is a family of complex-valued random variables such that each Φ_i is $\sigma(X_i)$-measurable, and such that all the Φ_i's have the same law with $\|\Phi_i\|_{\mathrm{L}^\infty} \leqslant 1$, then one has, for all finite subsets $I \subseteq \mathbb{Z}^d$:

$$\Big|\mathbb{E}\Big(\prod_{i\in I}\Phi_i\Big) - \mathbb{E}(\Phi_0)^{|I|}\Big| \leqslant |I| \times (\exp(\bar{\rho}) - 1) \times \tfrac{1}{2}(1 - |\mathbb{E}(\Phi_0)|^2). \tag{RY}$$

◇

Proof. Consider some finite $I \subset \mathbb{Z}^d$. Let us introduce a decreasing sequence $\mathbb{Z}^d \eqqcolon Z_0 \supset Z_1 \supset Z_2 \supset \ldots$ of subgroups of $\mathbb{Z}^d$, such that $|Z_n/Z_{n+1}| = 2$ for all n and that $\bigcap_{n\in\mathbb{N}} Z_n = \{0\}$.[†] Our proof will consist in showing that, for all $n \in \mathbb{N}$ and all $u \in \mathbb{Z}^d/Z_n$, the result actually holds when replacing 'I' by $I \cap (Z_n + u)$ and '$\bar{\rho}$' by $\bar{\rho}_n \coloneqq \sum_{u\in Z_n} \rho_u$.[‡]. The strategy of the proof will consist in downgoing induction on n. (That is to say, we will prove by downgoing induction on n that the proposition "the result holds for this value of n, for all $u \in \mathbb{Z}^d/Z_n$" is always valid).

For n large enough, all the $I \cap (Z_n + u)$'s have cardinality at most 1, so that the result is trivial. Now let us show that, if (RY) holds for all the sets of the form $I \cap (Z_{n+1} + u)$ (with '$\bar{\rho}$' replaced by $\bar{\rho}_{n+1}$), then it also holds for all the sets of the form $I \cap (Z_n + u)$ (with '$\bar{\rho}$' replaced by $\bar{\rho}_n$). As the structure of the reasoning is the

---

[†] Such a sequence does exist: for instance, denoting by $n \,\%\, d$ the canonical residue of n modulo d, you can define Z_n as being the set of d-uples $(z_0, \ldots, z_{d-1})$ such that all the z_k's are divisible by $2^{\lfloor n/d\rfloor}$ and that moreover all the $z_{k'}$'s for $k' < n \,\%\, d$ are divisible by $2^{\lfloor n/d\rfloor+1}$.

[‡] The rationale for replacing $\bar{\rho}$ by $\bar{\rho}_n$ comes from first identifying $Z_n{+}u$ with $\mathbb{Z}^d$ through isomorphy.

same for all n's and all u's, we are only going to write down the case $n = 0$ (for which there is only one possible u), in order to alleviate notation.

Denote $I_1 \coloneqq I \cap Z_1$, $Z_1' \coloneqq Z_0 \setminus Z_1$ (which is indeed of the form $Z_1 + u$, because Z_1 is a subgroup of Z_0 of index 2), and $I_1' \coloneqq I \cap Z_1'$. Also, to alleviate notation, denote $\mu \coloneqq \mathbb{E}(\Phi_0)$ and $\sigma^2 \coloneqq 1 - |\mu|^2$. By induction hypothesis, one has, for all $\bullet \in \{\sqcup, '\}$:

$$\Big|\mathbb{E}\Big(\prod_{i \in I_1^\bullet} \Phi_i\Big) - \mu^{|I_1^\bullet|}\Big| \leqslant |I_1^\bullet|(\mathrm{e}^{\bar\rho_1} - 1)\frac{\sigma^2}{2}. \tag{RZ}$$

Now denote $\Psi \coloneqq \prod_{i \in I} \Phi_i$, resp. $\Psi_1 \coloneqq \prod_{i \in I_1} \Phi_i$, resp. $\Psi_1' \coloneqq \prod_{i \in I_1'} \Phi_i$. Because one has $I = I_1 \sqcup I_1'$, one has $\Psi = \Psi_1 \Psi_1'$ and $|I| = |I_1| + |I_1'|$. Our goal is to bound above $|\mathbb{E}(\Psi) - \mu^{|I|}| = |\mathbb{E}(\Psi_1\Psi_1') - \mu^{|I_1|}\mu^{|I_1'|}|$. We start from the following decomposition:

$$\begin{aligned} &\mathbb{E}(\Psi_1\Psi_1') - \mu^{|I_1|}\mu^{|I_1'|} = \mathbb{E}(\Psi_1)\,\mathbb{E}(\Psi_1') - \mu^{|I_1|}\mu^{|I_1'|} + \mathrm{Cov}(\Psi_1, \Psi_1') \\ &= (\mathbb{E}(\Psi_1) - \mu^{|I_1|})\,\mathbb{E}(\Psi_1') + (\mathbb{E}(\Psi_1') - \mu^{|I_1'|})\mu^{|I_1|} + \mathrm{Corr}(\Psi_1^*, \Psi_1')\,\mathrm{Var}^{1/2}(\Psi_1)\,\mathrm{Var}^{1/2}(\Psi_1')^{[\S]} \end{aligned} \tag{SA}$$

(where Ψ_1^* denotes the complex conjugate of Ψ_1). Since one has obviously $|\mathbb{E}(\Psi_1')| \leqslant 1$, resp. $|\mu^{|I_1|}| \leqslant 1$, and that Ψ_1^* and Ψ_1' are resp. $\vec{X}_{Z_1}$- and $\vec{X}_{Z_1'}$-measurable, this decomposition implies that

$$|\mathbb{E}(\Psi) - \mu^{|I|}| \leqslant |I_1|(\mathrm{e}^{\bar\rho_1} - 1)\frac{\sigma^2}{2} + |I_1'|(\mathrm{e}^{\bar\rho_1} - 1)\frac{\sigma^2}{2} + \rho(\vec{X}_{Z_1}; \vec{X}_{Z_1'})\,\mathrm{Var}^{1/2}(\Psi_1)\,\mathrm{Var}^{1/2}(\Psi_1'). \tag{SB}$$

Now, we bound $\rho(\vec{X}_{Z_1}; \vec{X}_{Z_1'})$ in the following way: as Z_1 and Z_1' are parallel copies of Z_1 (which is itself isomorphic to $\mathbb{Z}^d$), and as the set of the $(j - i)$'s for $(i, j) \in Z_1 \times Z_1'$ is precisely Z_1' (since Z_1 is a subgroup of Z_0 of index 2), Corollary 56[¶] ensures that

$$\rho(\vec{X}_{Z_1}; \vec{X}_{Z_1'}) \leqslant \sum_{u \in Z_1'} \rho_u \eqqcolon \bar\rho_1'. \tag{SC}$$

Note in passing that one has $\bar\rho_1' \underset{\text{def}}{=} \bar\rho - \bar\rho_1$.

To bound above $\mathrm{Var}^{1/2}(\Psi_1)\,\mathrm{Var}^{1/2}(\Psi_1')$, we first use Young's inequality:

$$\mathrm{Var}^{1/2}(\Psi_1)\,\mathrm{Var}^{1/2}(\Psi_1') \leqslant \tfrac{1}{2}\,\mathrm{Var}(\Psi_1) + \tfrac{1}{2}\,\mathrm{Var}(\Psi_1'). \tag{SD}$$

Next, we observe that $\mathrm{Var}(\Psi_1)$ (resp. $\mathrm{Var}(\Psi_1')$) cannot be too large. Indeed, one has $\mathrm{Var}(\Psi_1) \underset{\text{def}}{=} \mathbb{E}(|\Psi_1^2|) - |\mathbb{E}(\Psi_1)|^2$; but since $\|\Psi_1\|_\infty \leqslant 1$, this is bounded above by $1 - |\mathbb{E}(\Psi_1)|^2$. On the other hand, by induction hypothesis one has

$$|\mathbb{E}(\Psi_1)| \geqslant |\mu|^{|I_1|} - |I_1|(\mathrm{e}^{\bar\rho_1} - 1)\frac{\sigma^2}{2}: \tag{SE}$$

[§] Here variance and Pearson correlation have to be understood in the *complex* sense, confer Proposition 3. As observed in Proposition 3, the bound "$|\mathrm{Corr}(f(X), g(Y))| \leqslant \rho(X; Y)$" remains valid in this complex-valued context.

[¶] Together with the elementary fact that, in Equation (IP), the value '$\bar\rho$' can be bounded above by "$\sum_{u \in \mathbb{Z}^d} \rho_u$".

therefore, using the elementary property

$$\forall x \in \mathbb{R}_+, y \in [0,1], z \in \mathbb{R}_+ \qquad x \geqslant y - z \implies 1 - x^2 \leqslant 1 - y^2 + 2z, \tag{SF}$$

it follows that

$$\operatorname{Var}(\Psi_1) \leqslant 1 - |\mu|^{2|I_1|} + |I_1|(\mathrm{e}^{\bar\rho_1} - 1)\sigma^2 \leqslant |I_1|\mathrm{e}^{\bar\rho_1}\sigma^2, \tag{SG}$$

where the last inequality comes from the following elementary bound: $1 - |\mu|^{2|I_1|} \leqslant |I_1|(1 - |\mu|^2) = |I_1|\sigma^2$.

Putting (`SC`), (`SD`) and (`SG`) together, we finally get that

$$\rho(\vec X_{Z_1}; \vec X_{Z_1'}) \operatorname{Var}^{1/2}(\Psi_1) \operatorname{Var}^{1/2}(\Psi_1') \leqslant |I| \bar\rho_1' \mathrm{e}^{\bar\rho_1} \frac{\sigma^2}{2}, \tag{SH}$$

which combined with (`SB`) finally yields

$$|\mathbb{E}(\Psi) - \mu^{|I|}| \leqslant |I|(\mathrm{e}^{\bar\rho_1} - 1)\frac{\sigma^2}{2} + |I|\bar\rho_1'\mathrm{e}^{\bar\rho_1}\frac{\sigma^2}{2}. \tag{SI}$$

But $\bar\rho_1' \leqslant \mathrm{e}^{\bar\rho_1'} - 1$, so $\bar\rho_1'\mathrm{e}^{\bar\rho_1} \leqslant \mathrm{e}^{\bar\rho} - \mathrm{e}^{\bar\rho_1}$, and in the end we have got that

$$|\mathbb{E}(\Psi) - \mu^{|I|}| \leqslant |I|(\mathrm{e}^{\bar\rho} - 1)\frac{\sigma^2}{2}, \tag{SJ}$$

which is exactly what we wanted! This ends the proof of the lemma. ♤

4.1.3 Spatial CLT on cubes

In a first time, I will state and prove a CLT on cubes. For the purposes of this subsection, it will be convenient to introduce some notation for cubes of $\mathbb{Z}^d$ centred around $\mathbf{0}$:

Notation 134. For $\ell \in \mathbb{N}$, we denote

$$\mathcal{Q}_\ell \coloneqq [\![-\ell, \ell]\!]^d, \tag{SK}$$

which is a set of cardinality $(2\ell+1)^d$. ♡

That notation being set, the result of this subsection is the following:

Theorem 135 (Spatial CLT on cubes). *Consider a translation-invariant spin model on $\mathbb{Z}^d$, whose spins are denoted by X_i, valued in some space $\mathcal{X}$; for $u \in \mathbb{Z}^d$, define ρ_u by (`RW`); assume that $\sum_{u\in\mathbb{Z}^d} \rho_u < \infty$.*

Under these conditions, let $k < \infty$ and let $f\colon \mathcal{X}^{\mathcal{Q}_k} \to \mathbb{R}$ be some measurable function; we then define the "local variables" (which all have the same law)

$$F_i \coloneqq f((X_{i+u})_{u\in\mathcal{Q}_k}). \tag{SL}$$

The present theorem claims that, if the law of F_0 is centred and square-integrable, then there exists a constant $\sigma < \infty$ such that

$$\operatorname{Law}\Big(\frac{1}{\sqrt{|\mathcal{Q}_L|}} \sum_{i\in\mathcal{Q}_L} F_i\Big) \overset{L\to\infty}{\to} \mathcal{N}(\sigma^2). \tag{SM}$$

◇

Proof. Before all, let us observe that the control on the correlations between the X_i's induces some control on the correlations between the F_i's. Using for instance Theorem 39, we have indeed that

$$\mathrm{\rho}^{\ddagger}(F_i;F_j) \leqslant |||(\!(\rho_{(j+u')-(i+u)})\!)_{u,u'\in\mathcal{Q}_k}||| \leqslant \sum_{u,u'\in\mathcal{Q}_k} \rho_{(j+u')-(i+u)}: \tag{SN}$$

therefore, denoting

$$\hat{\rho}_u \coloneqq \sum_{v,v'\in\mathcal{Q}_k} \rho_{u+v'-v}, \tag{SO}$$

we get that $\mathrm{\rho}^{\ddagger}(F_i;F_j) \leqslant \hat{\rho}_{j-i}$. An important point here is to note that, given the way the $\hat{\rho}_u$'s were defined, the summability assumption on the ρ_u's implies that $\sum_u \hat{\rho}_u < \infty$.

In the sequel, the random variable whose law is considered in the left-hand side of (SM) will be denoted by by $G^{(L)}$.

A preliminary step to derive (SM) is to identify the adequate value for σ. As one expects, the variance of the limit of $\mathrm{Law}(G^{(L)})$ will turn out to be the limit of the variances:

$$\sigma^2 = \lim_{L\to\infty} \mathrm{Var}(G^{(L)}). \tag{SP}$$

Let us check that the above limit exists, and give a limitless expression for it. By bilinearity of covariance, one has

$$\mathrm{Var}(G^{(L)}) = \frac{1}{|\mathcal{Q}_L|}\sum_{i\in\mathcal{Q}_L}\sum_{j\in\mathcal{Q}_L}\mathrm{Cov}(F_i,F_j) = \frac{1}{|\mathcal{Q}_L|}\sum_{i\in\mathcal{Q}_L}\sum_{j\in\mathcal{Q}_L}\mathrm{Cov}(F_0,F_{j-i}), \tag{SQ}$$

where the last equality comes from the translation invariance of the model. But for all $i\in\mathbb{Z}^d$, the sum $\sum_{j\in\mathcal{Q}_L}\mathrm{Cov}(F_0,F_{j-i})$ converges to $\sum_{u\in\mathbb{Z}^d}\mathrm{Cov}(F_0,F_u)$—moreover, the latter sum is absolutely convergent, since $\mathrm{Cov}(F_0,F_u)$ is bounded by $\hat{\rho}_u\,\mathrm{Var}(F_0)$. This suggests that one should have

$$\mathrm{Var}(G^{(L)}) \overset{L\to\infty}{\to} \sum_{u\in\mathbb{Z}^d}\mathrm{Cov}(F_0,F_u): \tag{SR}$$

this is indeed true, as one deduces from the fact that $\sum_{u\in\mathbb{Z}^d}\mathrm{Cov}(F_0,F_u)$ converges absolutely.[|||] So, in the sequel we will define σ^2 to be the right-hand side of (SR). By the way, we will also denote

$$\hat{\sigma}^2 \coloneqq \Big(\sum_u \hat{\rho}_u\Big)\|F_0\|^2_{\mathrm{L}^2}, \tag{SS}$$

which is a finite upper bound for σ^2.

[|||] The argument is the following. For $\varepsilon > 0$ arbitrarily small, let $M < \infty$ be such $\sum_{u\notin\mathcal{Q}_M}|\mathrm{Cov}(F_0,F_u)| < \varepsilon$; and define $\hat{\sigma}^2$ by (SS). Then you have $\big|\sum_{j\in\mathcal{Q}_L}\mathrm{Cov}(F_0,F_{j-i}) - \sigma^2\big| \leqslant \varepsilon$ as soon as $i\in\mathcal{Q}_{L-M}$, resp. $\big|\sum_{j\in\mathcal{Q}_L}\mathrm{Cov}(F_0,F_{j-i})\big| \leqslant \hat{\sigma}^2$ for all $i\in\mathcal{Q}_L\smallsetminus\mathcal{Q}_{L-M}$. But when $L\to\infty$, $|\mathcal{Q}_{L-M}|\,/\,|\mathcal{Q}_L|$ tends to 1, so that you get that $\overline{\lim}_{L\to\infty}|\mathrm{Var}(G^{(L)})-\sigma^2| \leqslant \varepsilon$.

Now let us turn to proving convergence of $\mathrm{Law}(\boldsymbol{G}^{(L)})$. Fix some arbitrarily small $\varepsilon > 0$. Since $\sum \hat{\rho}_u < \infty$, there exists $\ell_0 < \infty$ such that

$$\sum \{ \hat{\rho}_u \mid u \notin \mathcal{Q}_{\ell_0} \} \leqslant \varepsilon; \tag{ST}$$

and by (SP), we can also fix $\ell_1 < \infty$ such that

$$|\mathrm{Var}(\boldsymbol{G}^{(\ell_1)}) - \sigma^2| \leqslant \varepsilon. \tag{SU}$$

Our strategy will consist in "tiling" the cube of size $\boldsymbol{L}$ into a "patchwork" made of cubes of size $2\ell_1$, being set at distance $(\ell_0 + 1)$ from each other, plus some "scrap" (see Figure 4.1). Each of the $2\ell_1$-sized patches will be called a "tile". Technically speaking, for $\boldsymbol{i} \in \mathbb{Z}^d$, we define the tile $\mathcal{T}_i$ by

$$\mathcal{T}_i \coloneqq \{ (2\ell_1 + \ell_0 + 1)i + u \mid u \in \mathcal{Q}_{\ell_1} \}, \tag{SV}$$

and we define *Tiles* to be the set of the $\boldsymbol{i}$'s such that $\mathcal{T}_i \subseteq \mathcal{Q}_L$; while

$$\mathit{Scrap} \coloneqq \mathcal{Q}_L \setminus \bigcup_{i \in \mathit{Tiles}} \mathcal{T}_i. \tag{SW}$$

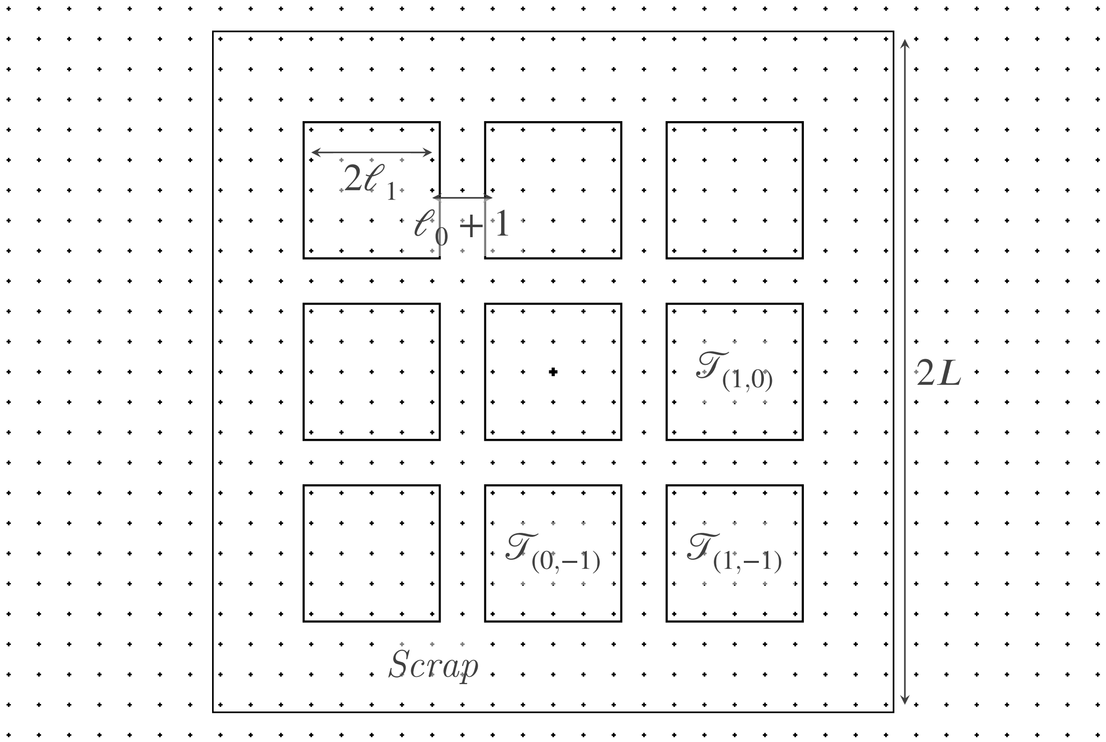

Figure 4.1: How the cube $\mathcal{Q}_L$ is patched with tiles of size $2\ell_1$, illustrated in dimension $d = 2$. Each small cross represents one spin in the lattice, the larger cross at the centre corresponding to the origin of $\mathbb{Z}^d$. On this drawing I took $L = 11, \ell_0 = 1, \ell_1 = 2$; one then has $|\mathit{Tiles}| = 9$ and $|\mathit{Scrap}| = 304$.

We can then decompose $\boldsymbol{G}^{(L)}$ into a part due to the tiles, denoted by $\boldsymbol{G}^{(L)}_{\square}$, plus a part due to the scrap, denoted by $\boldsymbol{G}^{(L)}_{*}$:

$$\boldsymbol{G}^{(L)}_{\square} \coloneqq \frac{1}{|\mathbb{Q}_L|^{1/2}} \sum_{i \in \mathit{Tiles}} \sum_{i' \in \mathscr{T}_i} \boldsymbol{F}_{i'}; \tag{SX}$$

$$\boldsymbol{G}^{(L)}_{*} \coloneqq \frac{1}{|\mathbb{Q}_L|^{1/2}} \sum_{i' \in \mathit{Scrap}} \boldsymbol{F}_{i'}. \tag{SY}$$

The variable $\boldsymbol{G}^{(L)}_{\square}$ can itself be decomposed according to the tiles: we will denote by $\grave{\boldsymbol{G}}_i$ the contribution of tile i, i.e.,

$$\grave{\boldsymbol{G}}_i \coloneqq \frac{1}{|\mathbb{Q}_L|^{1/2}} \sum_{i' \in \mathscr{T}_i} \boldsymbol{F}_{i'}. \tag{SZ}$$

Here observe that all the $\grave{\boldsymbol{G}}_i$'s have the same law, with

$$\mathrm{Law}(\grave{\boldsymbol{G}}_0) = \mathrm{Law}\biggl(\frac{|\mathbb{Q}_{\ell_1}|^{1/2}}{|\mathbb{Q}_L|^{1/2}} \boldsymbol{G}^{(\ell_1)} \biggr). \tag{TA}$$

Now let us observe that, using e.g. Corollary 56 (or more precisely its "third-party" version given by Theorem 60), the $\rho^{\ddagger}$-mixing coefficient between tiles $\mathscr{T}_i$ and $\mathscr{T}_j$ is bounded above[**] by

$$\begin{aligned} &\sum \{ \hat{\rho}_{u'} \mid \exists (i', j') \in \mathscr{T}_i \times \mathscr{T}_j \quad u' = j' - i' \} \\ &\qquad = \sum \{ \hat{\rho}_{u'} \mid u' \in (2\ell_1 + \ell_0 + 1)(j - i) + \mathbb{Q}_{2\ell_1} \} \eqqcolon \grave{\rho}_{j-i}. \end{aligned} \tag{TB}$$

But, for all $u' \in \mathbb{Z}^d$, there are at most 2^d possible $u \in \mathbb{Z}^d$ such that $u' \in (2\ell_1 + \ell_0 + 1)u + \mathbb{Q}_{2\ell_1}$; and moreover, if $u' \in \mathbb{Q}_{\ell_0}$, the only possibility is that $u = 0$: therefore, the definition of the $\grave{\rho}_u$'s implies that

$$\sum_{u \in \mathbb{Z}^d \setminus \{0\}} \grave{\rho}_u \leqslant 2^d \sum_{u' \notin \mathbb{Q}_{\ell_0}} \hat{\rho}_{u'} \leqslant 2^d \varepsilon. \tag{TC}$$

Thanks to what precedes, we are in position of controlling the characteristic function of $\mathrm{Law}(\boldsymbol{G}^{(L)}_{\square})$ via the use of Lemma 133. For $\lambda \in \mathbb{R}$, denote $\boldsymbol{\Phi}^{(\lambda)}_i \coloneqq \exp(\mathrm{i}\lambda \grave{\boldsymbol{G}}_i)$[††], and set $\mu(\lambda) \coloneqq \mathbb{E}(\boldsymbol{\Phi}^{(\lambda)}_i)$ (which does not depend on i). Then the $\boldsymbol{\Phi}^{(\lambda)}_i$'s satisfy the assumptions of Lemma 133 (where the role of '$\bar{\rho}$' is played by $\sum_{u \in \mathbb{Z}^d \setminus \{0\}} \grave{\rho}_u$), with $\prod_{i \in \mathit{Tiles}} \boldsymbol{\Phi}^{(\lambda)}_i = \exp(\mathrm{i}\lambda \boldsymbol{G}^{(L)}_{\square})$, so that

$$\Bigl| \mathbb{E}\bigl(\exp(\mathrm{i}\lambda \boldsymbol{G}^{(L)}_{\square}) \bigr) - \mu(\lambda)^{|\mathit{Tiles}|} \Bigr| \leqslant |\mathit{Tiles}| \bigl(\exp(2^d \varepsilon) - 1 \bigr) \tfrac{1}{2} (1 - |\mu(\lambda)|^2). \tag{TD}$$

[**] The bound is got by applying Corollary 56 with '$\boldsymbol{X}_{i'}$' $\leftarrow \boldsymbol{F}_{i'}$ for $i' \in \mathscr{T}_i$ and '$\boldsymbol{X}_{i'}$' $\leftarrow \partial$ for $i' \notin \mathscr{T}_i$, resp. '$\boldsymbol{Y}_{j'}$' $\leftarrow \boldsymbol{F}_{j'}$ for $j' \in \mathscr{T}_j$ and '$\boldsymbol{Y}_{j'}$' $\leftarrow \partial$ for $j' \notin \mathscr{T}_j$, using that the ρ-mixing coefficient between a constant variable and anything else is always 0. (Also, we use that the "arcsin-sum" [cf. (`IQ`)] is bounded above by the ordinary sum).

[††] Beware of not confusing the imaginary unit 'i' (typeset in roman face) with the index $i \in \mathbb{Z}^d$ (typeset in italic face) used to identify tile $\mathscr{T}_i$.

Let us look at the asymptotics of Equation (TD) when $L \to \infty$. First, one obviously has

$$|\mathbb{Q}_L| \overset{L\to\infty}{\sim} (2L)^d; \qquad\qquad |Tiles| \overset{L\to\infty}{\sim} \frac{(2L)^d}{(2\ell_1+\ell_0+1)^d}. \tag{TE}$$

Also, since one has

$$\mu(\lambda) \underset{\text{def}}{=} \mathbb{E}\big(\exp(\mathrm{i}\lambda G_0)\big) \underset{(\mathtt{TA})}{=} \mathbb{E}\big(\exp(\mathrm{i}\lambda |\mathbb{Q}_{\ell_1}|^{1/2} G^{(\ell_1)} \,/\, |\mathbb{Q}_L|^{1/2})\big) \tag{TF}$$

and that $G^{(\ell_1)}$ is a centred random variable, one has (using the connection between the moments of a random variable and the local expansion at 0 of its characteristic function):

$$\mu(\lambda) \overset{L\to\infty}{=} 1 - \frac{\lambda^2}{2}\frac{(2\ell_1+1)^d}{(2L)^d}\operatorname{Var}(G^{(\ell_1)}) + \mathrm{o}(L^{-d}). \tag{TG}$$

Therefore, one has the following asymptotics in (TD):

$$\mu(\lambda)^{|Tiles|} \overset{L\to\infty}{\to} \exp\biggl(-\frac{\lambda^2}{2}\frac{(2\ell_1+1)^d}{(2\ell_1+\ell_0+1)^d}\operatorname{Var}(G^{(\ell_1)})\biggr); \tag{TH}$$

$$|Tiles|\tfrac{1}{2}(1-|\mu(\lambda)|^2) \overset{L\to\infty}{\to} \frac{\lambda^2}{2}\frac{(2\ell_1+1)^d}{(2\ell_1+\ell_0+1)^d}\operatorname{Var}(G^{(\ell_1)}); \tag{TI}$$

moreover, provided ℓ_1 was taken large enough (for a fixed value of ℓ_0), the respective right-hand sides of (TH) and (TI) can be made arbitrarily close to resp. $\exp(-\lambda^2\sigma^2/2)$ and $\lambda^2\sigma^2/2$.

So, now we have a good estimate for $\mathbb{E}\big(\exp(\mathrm{i}\lambda G_\square^{(L)})\big)$, which corresponds to the contribution of the tiles; but it remains to control the contribution of the scrap spins… To do that, our starting point will be the following immediate lemma:

Lemma 136. *Let X and H be real random variables with $\|H\|_{\mathrm{L}^1} < \infty$. Then, for $\lambda \in \mathbb{R}$:*

$$|\mathbb{E}(\mathrm{e}^{\mathrm{i}\lambda(X+H)}) - \mathbb{E}(\mathrm{e}^{\mathrm{i}\lambda X})| \leqslant |\lambda|\|H\|_{\mathrm{L}^1}. \tag{TJ}$$

◇♤

From Lemma 136, we deduce that

$$\Big|\mathbb{E}\big(\exp(\mathrm{i}\lambda G^{(L)})\big) - \mathbb{E}\big(\exp(\mathrm{i}\lambda G_\square^{(L)})\big)\Big| \leqslant |\lambda|\|G_*^{(L)}\|_{\mathrm{L}^1}; \tag{TK}$$

but, since $G_*^{(L)}$ is centred, $\|G_*^{(L)}\|_{\mathrm{L}^2}^2$ is nothing but $\operatorname{Var}(G_*^{(L)})$: therefore,

$$\|G_*^{(L)}\|_{\mathrm{L}^1} \leqslant \|G_*^{(L)}\|_{\mathrm{L}^2} = \operatorname{Var}^{1/2}(G_*^{(L)}). \tag{TL}$$

Next, to bound above $\mathrm{Var}(\boldsymbol{G}_*^{(L)})$, we observe that, by elementary computations,[‡‡]

$$|\mathit{Scrap}| \leqslant \left(1 - \frac{(2\ell_1+1)^d}{(2\ell_1+\ell_0+1)^d} + \frac{(4\ell_1+2\ell_0)d}{2L+1}\right)|\mathscr{Q}_L|: \tag{TM}$$

we then deduce that

$$\begin{aligned}\mathrm{Var}(\boldsymbol{G}_*^{(L)}) &= \sum_{i\in \mathit{Scrap}}\sum_{j\in \mathit{Scrap}} \frac{1}{|\mathscr{Q}_L|}\,\mathrm{Cov}(\boldsymbol{F}_i,\boldsymbol{F}_j)\\ &\leqslant \frac{1}{|\mathscr{Q}_L|}\sum_{i\in \mathit{Scrap}}\sum_{j\in \mathit{Scrap}} \hat\rho_{j-i}\,\mathrm{Var}^{1/2}(\boldsymbol{F}_i)\,\mathrm{Var}^{1/2}(\boldsymbol{F}_j) = \frac{1}{|\mathscr{Q}_L|}\sum_{i\in \mathit{Scrap}}\sum_{j\in \mathit{Scrap}} \hat\rho_{j-i}\,\mathrm{Var}(\boldsymbol{F}_0)\\ &\leqslant \frac{1}{|\mathscr{Q}_L|}\sum_{i\in \mathit{Scrap}}\sum_{j\in \mathbb{Z}^d} \hat\rho_{j-i}\,\mathrm{Var}(\boldsymbol{F}_0) = \frac{|\mathit{Scrap}|}{|\mathscr{Q}_L|}\hat\sigma^2\\ &\leqslant \left(1 - \frac{(2\ell_1+1)^d}{(2\ell_1+\ell_0+1)^d} + \frac{(4\ell_1+2\ell_0)d}{2L+1}\right)\hat\sigma^2, \end{aligned}\tag{TN}$$

so that

$$\overline{\lim_{L\to\infty}}\ \mathrm{Var}(\boldsymbol{G}_*^{(L)}) \leqslant \left(1 - \frac{(2\ell_1+1)^d}{(2\ell_1+\ell_0+1)^d}\right)\hat\sigma^2, \tag{TO}$$

which can be made arbitrarily close to 0 provided ℓ_1 was taken large enough (for a fixed value of ℓ_0).

Finally, let us put everything together: using the decomposition

$$\begin{aligned}\overline{\lim_{L\to\infty}}\left|\mathbb{E}\big(\exp(\mathrm{i}\lambda \boldsymbol{G}^{(L)})\big) - \mathrm{e}^{-\lambda^2\sigma^2/2}\right| \leqslant{}& \overline{\lim_{L\to\infty}}\left|\mu(\lambda)^{|\mathit{Tiles}|} - \mathrm{e}^{-\lambda^2\sigma^2/2}\right|\\ &+ \overline{\lim_{L\to\infty}}\left|\mathbb{E}\big(\exp(\mathrm{i}\lambda \boldsymbol{G}_\Box^{(L)})\big) - \mu(\lambda)^{|\mathit{Tiles}|}\right|\\ &+ \overline{\lim_{L\to\infty}}\left|\mathbb{E}\big(\exp(\mathrm{i}\lambda \boldsymbol{G}^{(L)}\big) - \mathbb{E}\big(\exp(\mathrm{i}\lambda \boldsymbol{G}_\Box^{(L)})\big)\right|, \end{aligned}\tag{TP}$$

we have shown that $\overline{\lim}_{L\to\infty}\left|\mathbb{E}\big(\exp(\mathrm{i}\lambda \boldsymbol{G}^{(L)})\big) - \mathrm{e}^{-\lambda^2\sigma^2/2}\right|$ is bounded above by

$$\begin{aligned}&\left|\exp\!\left(-\frac{\lambda^2}{2}\frac{(2\ell_1+1)^d}{(2\ell_1+\ell_0+1)^d}\,\mathrm{Var}(\boldsymbol{G}^{(\ell_1)})\right) - \mathrm{e}^{-\lambda^2\sigma^2/2}\right|\\ &+ \big(\exp(2^d\varepsilon)-1\big)\frac{\lambda^2}{2}\frac{(2\ell_1+1)^d}{(2\ell_1+\ell_0+1)^d}\,\mathrm{Var}(\boldsymbol{G}^{(\ell_1)})\\ &+ |\lambda|\left(1 - \frac{(2\ell_1+1)^d}{(2\ell_1+\ell_0+1)^d}\right)\hat\sigma^2, \end{aligned}\tag{TQ}$$

[‡‡] The precise computations are not very important, as only the general shape of Equation (`TM`) actually matters here. For the record, I proceeded as follows. On a very large scale, the proportion of space occupied by tiles is $(2\ell_1+1)^d\,/\,(2\ell_1+\ell_0+1)^d$. Moreover, if we remove from $\mathscr{Q}_L$ an "shell" of appropriate width (the width of that shell being at most $(2\ell_1+\ell_0)$), we obtain a smaller cube for which the most external tiles touch the boundary of the cube: thus, inside that smaller cube, the proportion occupied by the tiles is *at least* $(2\ell_1+1)^d\,/\,(2\ell_1+\ell_0+1)^d$. But the total number of spins in the shell cannot be more than $(2\ell_1+\ell_0)$ times the size of the boundary of the larger cube, which size is (disregarding multiple counts) $2d\times(2L+1)^{d-1} = 2d\,/\,(2L+1)\times|\mathscr{Q}_L|$.

where, by taking ℓ_1 large enough, the respective terms of (`TQ`) can be made arbitrarily close to resp. 0, $\big(\exp(2^d\varepsilon)-1\big)\lambda^2\sigma^2/2$ and 0. So,

$$\varlimsup_{L\to\infty}\left|\mathbb{E}\big(\exp(\mathrm{i}\lambda G^{(L)})\big)-\mathrm{e}^{-\lambda^2\sigma^2/2}\right|\leqslant\big(\exp(2^d\varepsilon)-1\big)\frac{\lambda^2\sigma^2}{2};\tag{TR}$$

and in the end, since ε could be taken arbitrarily close to 0, we have got that

$$\forall\lambda\in\mathbb{R}\qquad\mathbb{E}\big(\exp(\mathrm{i}\lambda G^{(L)})\big)\overset{L\to\infty}{\to}\mathrm{e}^{-\lambda^2\sigma^2/2}.\tag{TS}$$

By Lévy's continuity theorem, this is tantamount to saying that $\mathrm{Law}(G^{(L)})$ converges to $\mathcal{N}(\sigma^2)$: this ends the proof of the central limit theorem for cubes. ♤

4.1.4 More general CLTs

In the previous subsection, I proved that spatial CLT holds when one considers the sum *on a cube* of some centred square-integrable local function of the spins. However, the restriction to that specific case was merely motivated by the sake of convenience, in order to moderate the level of technicality in the proof of Theorem 135: but in fact, once we are able to prove CLT on cubes, we can easily deduce more general forms of spatial CLT! In this subsection I will give two examples of such broader results.

Theorem 137 (Spatial CLT for general shapes)**.** *Assume the same assumptions as in Theorem 135; and consider a sequence $(U_n)_{n\in\mathbb{N}}$ of open subsets of $\mathbb{R}^d$ that get more and more regular, in the following sense: one can find a sequence $(r_n)_n$ with $r_n\to\infty$ such that, for all n, the set U_n can be written as a union of Euclidean balls of radius r_n.*[*] *Also, assume that each of the U_n's is bounded, so that its Lebesgue volume $\mathrm{vol}_d(U_n)$ is finite. Then,*

$$\mathrm{Law}\left(\frac{1}{\sqrt{\mathrm{vol}_d(U_n)}}\sum_{i\in U_n\cap\mathbb{Z}^d}F_i\right)\overset{n\to\infty}{\to}\mathcal{N}(\sigma^2),\tag{TT}$$

where σ^2 is the same as in Theorem 135. ◇

Remark 138. The regularity assumption on the U_n's actually ensures that $\mathrm{vol}_d(U_n)\to\infty$ and $|U_n\cap\mathbb{Z}^d|\sim\mathrm{vol}_d(U_n)$ as $n\to\infty$. ♧

Remark 139. *Strictly speaking*, Theorem 137 does not generalize Theorem 135; however, it is possible to approximate the cube $\mathcal{Q}_L$ by some open subset $U_L\subset\mathbb{R}^d$ such that U_L gets more and more regular as $L\to\infty$ and that the symmetric difference $(U_L\cap\mathbb{Z}^d)\vartriangle\mathcal{Q}_L$ has cardinality asymptotically $\ll|\mathcal{Q}_L|$: then, one would prove easily (by the same techniques as to control $\mathrm{Var}(G_*^{(L)})$ in the proof of Theorem 135) that, as $L\to\infty$, the difference between the random variables in the respective left-hand-sides of (`SM`) and (`TT`) tends to 0 in L^2 norm; so that in the end one deduces Theorem 135 from Theorem 137 anyway. ♧

---

[*] Equivalently, for all $x\in U_n$, there exists $y\in U_n$ such that the open ball of radius r_n centred at y contains x and is entirely included in U_n.

Remark 140. Theorem 137 has the advantage of not requiring any kind of convergence on the shape of the U_n's: you just need them to be more and more regular; but apart from that, they may be completely different from each other! ♧

Sketch of proof of Theorem 137. In fact, the proof is basically the same as for the CLT on cubes: patching U_n with tiles; using Lemma 133 to control the characteristic function of the contribution of tiles; showing that the contribution of scrap does not matter too much; and concluding by Lévy's continuity theorem. Moreover, the tiles will be defined exactly in the same way as in the proof of Theorem 135: confer Figure 4.1, just replacing $\mathcal{Q}_L$ by U_n. The only actual difference with the previous proof lies in how you control the contribution of the scrap spins: while previously we could bound above the number of scrap spins by explicit computations (cf. Equation (`TM`)), here one would use more general geometric arguments to show that the difference

$$\left| \frac{|Scrap|}{|U_n \cap \mathbb{Z}^d|} - \left(1 - \frac{(2\ell_1 + 1)^d}{(2\ell_1 + \ell_0 + 1)^d}\right) \right| \tag{TU}$$

can be bounded above by something that only depends on r_n and that tends to zero as the latter tends to infinity. ♤

An even more general spatial CLT statement would consist in showing that, when you look at the random distribution

$$\frac{1}{L^{d/2}} \sum_{i \in \mathbb{Z}^d} F_i \delta_{i/L} \tag{TV}$$

(in other words, you put a mass F_i at each $i \in \mathbb{Z}^d$, and you "zoom out" things to get a large-scale picture, with the adequate scaling factor), then you get something whose law tends, as $L \to \infty$, to some Gaussian white noise. The theorem below shows that this is indeed the case:

Theorem 141 (White noise limit at large scales)**.** *Assume the same assumptions as in Theorems 135 and 137. Then, for $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ a compactly supported continuous function, one has*

$$\mathrm{Law}\left(\frac{1}{L^{d/2}} \sum_{i \in \mathbb{Z}^d} \varphi(L^{-1} i) F_i \right) \overset{L\to\infty}{\to} \mathcal{N}\left(\sigma^2 \int_{\mathbb{R}^d} \varphi(x)^2 \, \mathrm{vol}_d(\mathrm{d}x)\right), \tag{TW}$$

where σ^2 is again the same as in Theorems 135 and 137. ◇

Sketch of proof. Let $R < \infty$ be such that $\varphi(\bullet)$ is zero outside $[-R, R]^d$. We will patch $[-R, R]^d$ with a large, but *fixed* number of tiles $(\mathscr{U}_i)_{i \in I}$ [beware, this time $\mathscr{U}_i$ is a subset of $\mathbb{R}^d$], each tile having size $\asymp \varepsilon$ and being separated from each other by some strictly positive distance $\eta \ll \varepsilon$. On each tile $\mathscr{U}_i$, denoting by x_i the centre of the tile, $\varphi(\bullet)$ is nearly equal to the constant $\varphi(x_i)$: more precisely, the error is bounded above by some modulus of continuity $\omega(\varepsilon)$, which does not depend on the tile and can be made arbitrarily close to 0 by taking ε small enough. Therefore, the random variable in the left-hand side of (`TW`) can be well approximated by the sum

$$\frac{1}{L^{d/2}} \sum_{i\in I} \varphi(x_i) \sum \{F_{i'} \mid i' \in L\mathcal{U}_i \cap \mathbb{Z}^d\}: \tag{TX}$$

more precisely, one can control the L^2 error of this approximation (using the same techniques as to control $\mathrm{Var}(G_*^{(L)})$ in the proof of Theorem 135) by something only depending on $\omega(\varepsilon)$ and of η / ε, that we can make arbitrarily small by choosing ε and η appropriately.

But the limiting law of (`TX`) is rather easy to derive! Indeed, for each $i \in I$, one can apply Theorem 135 to the cube $L\mathcal{U}_i \cap \mathbb{Z}^d$, getting

$$\mathrm{Law}\left(\frac{1}{\sqrt{|L\mathcal{U}_i \cap \mathbb{Z}^d|}} \sum \{F_{i'} \mid i' \in L\mathcal{U}_i \cap \mathbb{Z}^d\}\right) \overset{L\to\infty}{\to} \mathcal{N}(\sigma^2); \tag{TY}$$

and since $|L\mathcal{U}_i \cap \mathbb{Z}^d| \sim \mathrm{vol}_d(\mathcal{U}_i) L^d$ as $L \to \infty$, we deduce that

$$\mathrm{Law}\left(\frac{1}{L^{d/2}} \varphi(x_i) \sum \{F_{i'} \mid i' \in L\mathcal{U}_i \cap \mathbb{Z}^d\}\right) \overset{L\to\infty}{\to} \mathcal{N}(\varphi(x_i)^2 \,\mathrm{vol}_d(\mathcal{U}_i) \sigma^2). \tag{TZ}$$

Moreover, by the $\rho^\ddagger$-mixing assumption, the variables $\sum_{i'\in L\mathcal{U}_i \cap \mathbb{Z}^d} F_{i'}$ become asymptotically all independent (because the distance between the sets $L\mathcal{U}_i \cap \mathbb{Z}^d$ is $\asymp L\eta$, which tends to infinity): therefore, by summation,[†] we get that

$$\mathrm{Law}\left(\frac{1}{L^{d/2}} \sum_{i\in I} \varphi(x_i) \sum \{F_{i'} \mid i' \in L\mathcal{U}_i \cap \mathbb{Z}^d\}\right) \overset{L\to\infty}{\to} \mathcal{N}\Big(\sigma^2 \sum_{i\in I} \varphi(x_i)^2 \,\mathrm{vol}_d(\mathcal{U}_i)\Big). \tag{UA}$$

But in that formula, provided we have taken ε and η / ε small enough, the discrete sum $\sum_{i\in I} \varphi(x_i)^2 \,\mathrm{vol}_d(\mathcal{U}_i)$ can be made arbitrarily close to $\int_{x\in\mathbb{R}^d} \varphi(x)^2 \,\mathrm{vol}_d(\mathrm{d}x)$: so we have got what we wanted! ♤

4.2 Spectral gap for the Glauber dynamics

4.2.1 Introduction of the section

In this section we are once again looking at a probabilistic system made of a large number of "elementary" random variables $(X_i)_{i\in I}$, each X_i being called a "spin". The main estimates (Theorems 146 and 150) will only be stated in the case where I is finite, the infinite case (cf. §§ 4.2.3 and 4.2.4) being got by passing to the limit.

We are getting interested in the *Glauber dynamics* of our system. The notion of Glauber dynamics was already encountered earlier in this monograph [Definition 109]; but let us recall what it is anyway, for the sake of self-consistency (and adapting the definition to the notation that we are using here):

[†] Here note that we are summing a *fixed* number of variables, so that, to get convergence for the law of the sum, we only need asymptotic independence *in the sense of convergence in law*: and this gets easily deduced from the fact that all the $\rho^\ddagger$-mixing coefficients between pairs of tiles tend to zero.

Definition 142 (Glauber dynamics, bis)**.** Denoting by Ω the state space of $\vec{X}_I$, let $\mathbb{P}$ be a probability measure on Ω. The *Glauber dynamics* [13, 23] associated to $\mathbb{P}$ is the Markov process on Ω having the following law: on each $i \in I$ there is an alarm clock, which rings at times following a Poisson point process on $\mathbb{R}_+$ with constant intensity 1, all the clocks being independent. When a clock rings, the state of spin X_i—and only it—is "flipped" in such a way that the state of X_i immediately after the flip follows the law $\mathrm{Law}^{\mathbb{P}}(X_i \mid \vec{X}_{I\smallsetminus\{i\}})$. In more formal terms, the Glauber dynamics is the Markov process whose generator $\mathscr{L}$ on $\mathrm{L}^\infty(\Omega)$ is defined by:

$$\big(\mathscr{L}f\big)(\vec{x}_I) = \sum_{i\in I} \mathbb{E}(f(\vec{X}_I) - f(\vec{x}_I) \mid \vec{X}_{I\smallsetminus\{i\}} \equiv \vec{x}_{I\smallsetminus\{i\}}). \tag{UB}$$

♡

Let us recall some basic facts on the Glauber dynamics (see [24, Chapter IV] for more details). By construction $\mathbb{P}$ is a reversible equilibrium measure for the dynamics, so $\mathscr{L}$ is self-adjoint on $\mathrm{L}^2(\mathbb{P})$. Since obviously $\mathscr{L}1 \equiv 0$, one can also define $\mathscr{L}$ on $\bar{\mathrm{L}}^2(\mathbb{P})$, on which it is self-adjoint too. This leads to the following definition:

Definition 143 (Dirichlet energy)**.** The *Dirichlet energy* of $f \in \bar{\mathrm{L}}^2(\mathbb{P})$ is

$$\mathscr{E}(f,f) = \langle \mathscr{L}f, f\rangle. \tag{UC}$$

♡

The following immediate identity shows that $\mathscr{E}$ is always a nonnegative quadratic form:

Lemma 144.

$$\mathscr{E}(f,f) = \int_\Omega \mathbb{P}(\mathrm{d}\vec{x}_I) \sum_{i\in I} \mathrm{Var}(f \mid \vec{X}_{I\smallsetminus\{i\}} = \vec{x}_{I\smallsetminus\{i\}}). \tag{UD}$$

◇♤

In this section, the specific concept that interests us will be the *spectral gap* of the Glauber dynamics. It is defined in the following way:

Definition 145 (Spectral gap)**.** For $\lambda > 0$, the Glauber dynamics is said to have *spectral gap* $\geqslant \lambda$ if, for all $f \in \bar{\mathrm{L}}^2(\mathbb{P})$,

$$\mathscr{E}(f,f) \geqslant \lambda\, \mathrm{Var}(f). \tag{UE}$$

♡

What makes spectral gap interesting is that its strict positiveness is equivalent to exponential convergence to 0 of the semigroup $(\mathrm{e}^{-t\mathscr{L}})_{t\geqslant 0}$ on $\bar{\mathrm{L}}^2(\mathbb{P})$, the rate of convergence being equal to the width of the spectral gap (see e.g. [25, § 4.2]). As the Glauber dynamics is one of the easiest ways to simulate the law $\mathbb{P}$ for complicated models, the stake of having such an exponential convergence is evident.

Many works have been done on the spectral gap of Glauber dynamics, see for instance Martinelli's St-Flour course [11]. In particular, several results (e.g. [26, 27, 28]) state that, morally, the less the spins are correlated, the larger the spectral gap is. However, the researchers who work on this issue usually express decorrelation between spins in terms of *β-mixing* (cf. Definition 13): while, in my opinion, it would me more natural to look at spin decorrelation in terms of *ρ-mixing*, as Equation (`UE`) stating the spectral gap criterion takes place in a Hilbertian frame!

Thus, my goal here will be to find a bound on the spectral gap of the Glauber dynamics that is expressed in terms of ρ-mixing conditions: as these conditions look to be the minimal frame to study the spectral gap, hopefully the results yielded by such a method will be sharp. Another noticeable feature of my approach is that it will remain at a quite abstract level: no symmetry property on $\boldsymbol{I}$ nor $\mathbb{P}$ need be assumed, all the work essentially consisting in manipulating relevant quadratic forms.

4.2.2 A lower bound for the spectral gap

The core result of this whole section is the following:

Theorem 146 (Lower bound for the spectral gap)**.** *Take $\boldsymbol{I} = \llbracket 0, N \llbracket$. Suppose that for all distinct $i, j \in \boldsymbol{I}$ one has $\rho^\ddagger(X_i; X_j) \leqslant \rho_{ij} < 1$. (We will make the costless assumption that $\rho_{ji} = \rho_{ij}$). For $i, j \in \boldsymbol{I}$ with $i < j$, denote*

$$\tilde{\rho}_{ij} \coloneqq \frac{\rho_{ij}}{\prod_{j' \in \rrbracket i, j \rrbracket} (1 - \rho_{ij'}^2)}, \tag{UF}$$

and for $i \in \boldsymbol{I}$, denote

$$\tilde{1}_i \coloneqq \frac{1}{\prod_{j > i} (1 - \rho_{ij}^2)}. \tag{UG}$$

Then the Glauber dynamics has spectral gap at least $|||\boldsymbol{M}|||^{-2}$, where $\boldsymbol{M}$ is the $N \times N$ matrix defined by

$$\boldsymbol{M} = \begin{pmatrix} 1 & -\tilde{\rho}_{01} & -\tilde{\rho}_{02} & \cdots & -\tilde{\rho}_{0(N-1)} \\ 0 & 1 & -\tilde{\rho}_{12} & \cdots & -\tilde{\rho}_{1(N-1)} \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -\tilde{\rho}_{(N-2)(N-1)} \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{1}_0 & 0 & 0 & \cdots & 0 \\ 0 & \tilde{1}_1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \tilde{1}_{N-1} \end{pmatrix}. \tag{UH}$$

◇

Remark 147. Since all the ρ_{ij}'s were supposed < 1, all the $\tilde{\rho}_{ij}$'s and the $\tilde{1}_i$'s are finite; moreover, the upper-diagonal shape of the first matrix in the right-hand side of (`UH`) ensures that it is invertible: therefore, the lower bound $|||\boldsymbol{M}|||^{-2}$ is strictly positive. ♧

Proof of Theorem 146. Let f be a centred square-integrable variable on $(\boldsymbol{\Omega}, \mathbb{P})$. For $\boldsymbol{I}' \subseteq \boldsymbol{I}$, denote $\mathscr{F}_{\boldsymbol{I}'} \coloneqq \sigma(\vec{X}_{\boldsymbol{I}'})$. For $i \in \boldsymbol{I}$, $\boldsymbol{I}' \subseteq \boldsymbol{I} \setminus \{i\}$, denote

$$f_i^{\boldsymbol{I}'} \coloneqq f^{\mathscr{F}_{\boldsymbol{I}' \sqcup \{i\}}} - \mathbb{E}(f \mid \mathscr{F}_{\boldsymbol{I}'}); \tag{UI}$$

define moreover

$$f_i^{\neq} \coloneqq f_i^{\boldsymbol{I} \setminus \{i\}}; \tag{UJ}$$

$$f_i^{<} \coloneqq f_i^{\llbracket 0, i \llbracket}. \tag{UK}$$

Then by Proposition 144, one has

$$\mathscr{E}(f,f) = \sum_{i\in I} \operatorname{Var}(f_i^{\neq}), \tag{UL}$$

while the usual telescopic argument (confer the proofs of Chapter 2) shows that

$$\operatorname{Var}(f) = \sum_{i\in I} \operatorname{Var}(f_i^{<}). \tag{UM}$$

So, to prove the theorem, we will have to establish links between the different values $\operatorname{Var}(f_i^{I'})$. To that end, it will be convenient to introduce the shorthands $\operatorname{Var}^{1/2}(f_i^{I'}) \eqqcolon \varDelta_i^{I'}$, resp. $\operatorname{Var}^{1/2}(f_i^{\neq}) \eqqcolon \varDelta_i^{\neq}$, $\operatorname{Var}^{1/2}(f_i^{<}) \eqqcolon \varDelta_i^{<}$. That notation being set, our main technical tool will be the following

Lemma 148. *For $I' \subseteq I$ and $i, j \in I \smallsetminus I'$ with $j \neq i$,*

$$\varDelta_i^{I'} \leqslant \varDelta_i^{I' \sqcup \{j\}} + \rho_{ij} \varDelta_j^{I'}. \tag{UN}$$

$\diamondsuit$

Proof of Lemma 148. Assume in a first time that $I' = \varnothing$, and denote $f_i \coloneqq f_i^{\varnothing}$, $f_j \coloneqq f_j^{\varnothing}$, $f_i^j \coloneqq f_i^{\{j\}}$ and $\mathscr{F}_i \coloneqq \mathscr{F}_{\{i\}}$. Projecting the decomposition "$f_{\{i,j\}}^{\varnothing} = f_{\{j\}}^{\varnothing} + f_{\{i\}}^{\{j\}}$" onto $\bar{\mathrm{L}}^2(\mathscr{F}_i)$, one has $f_i = (f_j)^{\mathscr{F}_i} + (f_i^j)^{\mathscr{F}_i}$, whence by the triangle inequality:

$$\operatorname{Var}^{1/2}(f_i) \leqslant \operatorname{Var}^{1/2}((f_i^j)^{\mathscr{F}_i}) + \operatorname{Var}^{1/2}((f_j)^{\mathscr{F}_i}). \tag{UO}$$

Here one has trivially $\operatorname{Var}^{1/2}((f_i^j)^{\mathscr{F}_i}) \leqslant \operatorname{Var}^{1/2}(f_i^j)$; also, since f_j is X_j-measurable, by Proposition 4 one has $\operatorname{Var}^{1/2}((f_j)^{\mathscr{F}_i}) \leqslant \mathrm{\rho}(\sigma(X_j); \mathscr{F}_i) \operatorname{Var}^{1/2}(f_j) \leqslant \rho_{ij} \operatorname{Var}^{1/2}(f_j)$: so in the end, (UO) becomes

$$\operatorname{Var}^{1/2}(f_i) \leqslant \operatorname{Var}^{1/2}(f_i^j) + \rho_{ij} \operatorname{Var}^{1/2}(f_j), \tag{UP}$$

which is (UN) for $I' = \varnothing$.

In the case $I' \neq \varnothing$, the same reasoning can be performed, except that one has to work conditionally to $\mathscr{F}_{I'}$. Then, taking $f_i \coloneqq f_i^{I'}$, $f_j \coloneqq f_j^{I'}$, $f_i^j \coloneqq f_i^{I' \sqcup \{j\}}$ (and letting the role of '$\mathscr{F}_i$' be played by $\mathscr{F}_{I' \sqcup \{i\}}$), one gets

$$\operatorname{Var}^{1/2}(f_i \mid \mathscr{F}_{I'}) \leqslant \operatorname{Var}^{1/2}(f_i^j \mid \mathscr{F}_{I'}) + \rho_{ij} \operatorname{Var}^{1/2}(f_j \mid \mathscr{F}_{I'}). \tag{UQ}$$

Now

$$\operatorname{Var}(f_i) = \int \operatorname{Var}(f_i \mid \vec{X}_{I'} = \vec{x}_{I'})\, \mathbb{P}(\mathrm{d}\vec{x}_{I'}), \tag{UR}$$

with similar formulas for f_j and f_i^j, since all these functions are centred w.r.t. $\mathscr{F}_{I'}$. Therefore, squaring (UQ), integrating over $\vec{x}_{I'}$, applying Minkowski's inequality to the right-hand side, and using the (UR)-like formulas, we get

$$\operatorname{Var}^{1/2}(f_i) \leqslant \operatorname{Var}^{1/2}(f_i^j) + \rho_{ij} \operatorname{Var}^{1/2}(f_j), \tag{US}$$

i.e. (UN). ♤

Having proved Lemma 148, let us carry on with the proof of Theorem 146. For $i \in I, j \in \rrbracket i, N \rrbracket$, denote

$$\mathit{\Delta}_i^{[j]} \coloneqq \mathit{\Delta}_i^{\llbracket 0, j \llbracket \setminus \{i\}}. \tag{UT}$$

Lemma 148 will be used through the following corollary:

Proposition 149. *For all $i, j \in I$ with $i < j$,*

$$\mathit{\Delta}_i^{[j]} \leqslant \frac{1}{1 - \rho_{ij}^2} (\mathit{\Delta}_i^{[j+1]} + \rho_{ij} \mathit{\Delta}_j^{<}). \tag{UU}$$

◇

Proof of Proposition 149. We have to bound $\mathit{\Delta}_i^{[j]}$, which we rather denote as “$\mathit{\Delta}_a^{[b]}$” here to avoid confusion with the notation of Lemma 148. Applying Lemma 148 with ‘I'’ $\leftarrow \llbracket 0, b \llbracket \setminus \{a\}$, ‘$i$’ $\leftarrow a$ and ‘j’ $\leftarrow b$, one has

$$\mathit{\Delta}_a^{[b]} \underset{\text{def}}{=} \mathit{\Delta}_a^{\llbracket 0, b \llbracket \setminus \{a\}} \leqslant \mathit{\Delta}_a^{\llbracket 0, b \rrbracket \setminus \{a\}} + \rho_{ab} \mathit{\Delta}_b^{\llbracket 0, b \llbracket \setminus \{a\}} \underset{\text{def}}{=} \mathit{\Delta}_a^{[b+1]} + \rho_{ab} \mathit{\Delta}_b^{\llbracket 0, b \llbracket \setminus \{a\}}. \tag{UV}$$

But applying again Lemma 148, this time with ‘I'’ $\leftarrow \llbracket 0, b \llbracket \setminus \{a\}$, ‘$i$’ $\leftarrow b$ and ‘j’ $\leftarrow a$, one has

$$\mathit{\Delta}_b^{\llbracket 0, b \llbracket \setminus \{a\}} \leqslant \mathit{\Delta}_b^{\llbracket 0, b \llbracket} + \rho_{ab} \mathit{\Delta}_a^{\llbracket 0, b \llbracket \setminus \{a\}} \underset{\text{def}}{=} \mathit{\Delta}_b^{<} + \rho_{ab} \mathit{\Delta}_a^{[b]}. \tag{UW}$$

We then get (UU) by combining (UV) and (UW). ♤

Finally, let us show how Proposition 149 implies the theorem. If, say, $\underline{N = 10}$, iterating Proposition 149 yields, for instance:

$$\begin{aligned}
\mathit{\Delta}_3^{<} \underset{\text{def}}{=} \mathit{\Delta}_3^{[4]} &\leqslant \frac{\rho_{34}}{1 - \rho_{34}^2} \mathit{\Delta}_4^{<} + \frac{1}{1 - \rho_{34}^2} \mathit{\Delta}_3^{[5]} \underset{\text{def}}{=} \tilde{\rho}_{34} \mathit{\Delta}_4^{<} + \frac{\mathit{\Delta}_3^{[5]}}{1 - \rho_{34}^2} \\
&\leqslant \tilde{\rho}_{34} \mathit{\Delta}_4^{<} + \frac{\rho_{35}}{(1 - \rho_{34}^2)(1 - \rho_{35}^2)} \mathit{\Delta}_5^{<} + \frac{1}{(1 - \rho_{34}^2)(1 - \rho_{35}^2)} \mathit{\Delta}_3^{[6]} \underset{\text{def}}{=} \tilde{\rho}_{34} \mathit{\Delta}_4^{<} + \tilde{\rho}_{35} \mathit{\Delta}_5^{<} + \frac{\mathit{\Delta}_3^{[6]}}{(1 - \rho_{34}^2)(1 - \rho_{35}^2)} \\
&\leqslant \tilde{\rho}_{34} \mathit{\Delta}_4^{<} + \tilde{\rho}_{35} \mathit{\Delta}_5^{<} + \tilde{\rho}_{34} \mathit{\Delta}_6^{<} + \frac{\mathit{\Delta}_3^{[7]}}{(1 - \rho_{34}^2)(1 - \rho_{35}^2)(1 - \rho_{36}^2)} \\
&\leqslant \cdots \leqslant \tilde{\rho}_{34} \mathit{\Delta}_4^{<} + \tilde{\rho}_{35} \mathit{\Delta}_5^{<} + \tilde{\rho}_{36} \mathit{\Delta}_6^{<} + \tilde{\rho}_{37} \mathit{\Delta}_7^{<} + \tilde{\rho}_{38} \mathit{\Delta}_8^{<} + \tilde{\rho}_{39} \mathit{\Delta}_9^{<} + \tilde{1}_3 \mathit{\Delta}_3^{[10]} = \sum_{j > 3} \tilde{\rho}_{3j} \mathit{\Delta}_j^{<} + \tilde{1}_3 \mathit{\Delta}_3^{\neq};
\end{aligned} \tag{UX}$$

and more generally, for any i, one has

$$\mathit{\Delta}_i^{<} = \tilde{1}_i \mathit{\Delta}_i^{\neq} + \sum_{j > i} \tilde{\rho}_{ij} \mathit{\Delta}_j^{<}. \tag{UY}$$

Equation (UY) can itself be iterated, as follows. I will just detail the reasoning for $\underline{N = 4}$, hoping that generalizing is obvious then. First, one has trivially

$$\mathit{\Delta}_3^{<} = \tilde{1}_3 \mathit{\Delta}_3^{\neq}; \tag{UZ}$$

also, applying (`UY`) with $i \leftarrow 2$, one has

$$\Delta_2^< \leqslant \tilde{1}_2 \Delta_2^{\neq} + \tilde{\rho}_{23} \Delta_3^< \underset{(\mathtt{UZ})}{=} \tilde{1}_2 \Delta_2^{\neq} + \tilde{1}_3 \tilde{\rho}_{23} \Delta_3^{\neq}. \tag{VA}$$

Now we apply (`UY`) with $i \leftarrow 1$:

$$\Delta_1^< \leqslant \tilde{1}_1 \Delta_1^{\neq} + \tilde{\rho}_{12} \Delta_2^< + \tilde{\rho}_{13} \Delta_3^< \leqslant \tilde{1}_1 \Delta_1^{\neq} + \tilde{1}_2 \tilde{\rho}_{12} \Delta_2^{\neq} + \tilde{1}_3 (\tilde{\rho}_{12} \tilde{\rho}_{23} + \tilde{\rho}_{13}) \Delta_3^{\neq}, \tag{VB}$$

where the last inequality comes from (`VA`) and (`UZ`). Finally, applying (`UY`) with $i \leftarrow 0$:

$$\begin{aligned} &\Delta_0^< \leqslant \tilde{1}_0 \Delta_0^{\neq} + \tilde{\rho}_{01} \Delta_1^< + \tilde{\rho}_{02} \Delta_2^< + \tilde{\rho}_{03} \Delta_3^< \\ &\leqslant \tilde{1}_0 \Delta_0^{\neq} + \tilde{1}_1 \tilde{\rho}_{01} \Delta_1^{\neq} + \tilde{1}_2 (\tilde{\rho}_{01} \tilde{\rho}_{12} + \tilde{\rho}_{02}) \Delta_2^{\neq} + \tilde{1}_3 (\tilde{\rho}_{01} \tilde{\rho}_{12} \tilde{\rho}_{23} + \tilde{\rho}_{01} \tilde{\rho}_{13} + \tilde{\rho}_{02} \tilde{\rho}_{23} + \tilde{\rho}_{03}) \Delta_3^{\neq}, \end{aligned} \tag{VC}$$

where the last inequality comes from (`VB`), (`VA`) and (`UZ`).

One can sum up Equations (`UZ`)–(`VC`) into the matricial expression

$$\begin{pmatrix} \Delta_0^< \\ \Delta_1^< \\ \Delta_2^< \\ \Delta_3^< \end{pmatrix} \leqslant \begin{pmatrix} 1 & \tilde{\rho}_{01} & \tilde{\rho}_{02} + \tilde{\rho}_{01} \tilde{\rho}_{12} & \tilde{\rho}_{03} + \tilde{\rho}_{02} \tilde{\rho}_{23} + \tilde{\rho}_{01} \tilde{\rho}_{13} + \tilde{\rho}_{01} \tilde{\rho}_{12} \tilde{\rho}_{23} \\ 0 & 1 & \tilde{\rho}_{12} & \tilde{\rho}_{13} + \tilde{\rho}_{12} \tilde{\rho}_{23} \\ 0 & 0 & 1 & \tilde{\rho}_{23} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{1}_0 \Delta_0^{\neq} \\ \tilde{1}_1 \Delta_1^{\neq} \\ \tilde{1}_2 \Delta_2^{\neq} \\ \tilde{1}_3 \Delta_3^{\neq} \end{pmatrix}, \tag{VD}$$

where, looking back at how the square matrix in (`VD`) was constructed, we observe that

$$\begin{pmatrix} \textit{square} \\ \textit{matrix} \\ \textit{in} \\ (\mathtt{VD}) \end{pmatrix} = \sum_{k=0}^{\infty} \begin{pmatrix} 0 & \tilde{\rho}_{01} & \tilde{\rho}_{02} & \tilde{\rho}_{03} \\ 0 & 0 & \tilde{\rho}_{12} & \tilde{\rho}_{13} \\ 0 & 0 & 0 & \tilde{\rho}_{23} \\ 0 & 0 & 0 & 0 \end{pmatrix}^k = \begin{pmatrix} 1 & -\tilde{\rho}_{01} & -\tilde{\rho}_{02} & -\tilde{\rho}_{03} \\ 0 & 1 & -\tilde{\rho}_{12} & -\tilde{\rho}_{13} \\ 0 & 0 & 1 & -\tilde{\rho}_{23} \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1}. \tag{VE}$$

So, in the end we have obtained that

$$\begin{pmatrix} \Delta_0^< \\ \Delta_1^< \\ \Delta_2^< \\ \Delta_3^< \end{pmatrix} \leqslant \boldsymbol{M} \begin{pmatrix} \Delta_0^{\neq} \\ \Delta_1^{\neq} \\ \Delta_2^{\neq} \\ \Delta_3^{\neq} \end{pmatrix}, \tag{VF}$$

where $\boldsymbol{M}$ is given by (`UH`). Then the very definition of operator norm yields that $\mathrm{Var}(f) = \sum_i (\Delta_i^<)^2 \leqslant |||\boldsymbol{M}|||^2 \sum_i (\Delta_i^{\neq})^2 = |||\boldsymbol{M}|||^2 \mathscr{E}(f, f)$. This ends the proof of Theorem 146. ♤

So, with Theorem 146 we have got a lower bound for the spectral gap of the Glauber dynamics. Before closing this subsection, let us give a variant of that result, with a slightly looser, but much nicer formula:

Corollary 150 (Simplified lower bound for the spectral gap). *With the assumptions of Theorem 146, denote*

$$\boldsymbol{R} \coloneqq \begin{pmatrix} 0 & \rho_{01} & \rho_{02} & \cdots & \rho_{0(N-1)} \\ \rho_{01} & 0 & \rho_{12} & \cdots & \rho_{1(N-1)} \\ \rho_{02} & \rho_{12} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_{(N-2)(N-1)} \\ \rho_{0(N-1)} & \rho_{1(N-1)} & \cdots & \rho_{(N-2)(N-1)} & 0 \end{pmatrix}. \tag{VG}$$

Then, provided $|||\boldsymbol{R}||| < 1$*, the spectral gap of the Glauber dynamics is at least* $(1 - |||\boldsymbol{R}|||)^2$. ◇

Remark 151. Note that, contrary to the result of Theorem 146, the bound of Corollary 150 is symmetric by re-labelling the indices of $\boldsymbol{I}$. ♧

Remark 152. In concrete situations, where N is very large (or even infinite) and that all the $\sum_{j\neq i} \rho_{ij}$'s are (approximately) equal to some common value $\bar{\rho} \ll 1$, one will have $(1-|||\boldsymbol{R}|||)^2 \approx 1-2\bar{\rho} \approx |||\boldsymbol{M}|||^{-2}$ at order one, so that the formula of Corollary 150 turns out to be nearly as good as that of Theorem 146. ♧

Proof of Corollary 150. Assume that $|||\boldsymbol{R}||| < 1$, otherwise there is nothing to prove. Observe that, then, $(1 - |||\boldsymbol{R}|||)^2$ can also be written as $|||\boldsymbol{M}'|||^{-2}$, with

$$\boldsymbol{M}' \coloneqq (\boldsymbol{I}_N - \boldsymbol{R})^{-1}: \tag{VH}$$

this comes from the facts that, for symmetric matrices, diagonalization properties ensure that the operator norm coincides with the spectral radius, and that, since $\boldsymbol{R}$ has nonnegative entries, its spectral radius is attained for a nonnegative real eigenvalue (cf. Perron–Frobenius theorem). Thanks to that re-writing of $(1 - |||\boldsymbol{R}|||)^2$, what we actually want to get is that the result of Theorem 146 still holds when replacing $\boldsymbol{M}$ by $\boldsymbol{M}' \coloneqq (\boldsymbol{I}_N - \boldsymbol{R})^{-1}$. To prove that, our strategy will be to show that each entry of $\boldsymbol{M}$ is actually bounded above by the corresponding entry of $\boldsymbol{M}'$: indeed, as both $\boldsymbol{M}$ and $\boldsymbol{M}'$ have nonnegative entries (as will appear clearly below), this will imply that $|||\boldsymbol{M}'||| \geqslant |||\boldsymbol{M}|||$.[‡]

So, let us compare the entries of $\boldsymbol{M}'$ with the corresponding entries of $\boldsymbol{M}$. To do so, we will "expand" the entries of $\boldsymbol{M}$, resp. $\boldsymbol{M}'$. First, notice that $1/(1 - \rho_{ij}^2)$ can be expanded into $1 + \rho_{ij}\rho_{ji} + \rho_{ij}\rho_{ji}\rho_{ij}\rho_{ji} + \cdots$, so that one has the expansions

$$\tilde{\rho}_{ij} = \sum_{k\in\mathbb{N}} \sum \Big\{ \Big(\prod_{\ell=0}^{k-1} \rho_{ij_\ell}\rho_{j_\ell i} \Big) \rho_{ij} \;\Big|\; i < j_0 \leqslant \cdots \leqslant j_{k-1} \leqslant j \Big\} \tag{VI}$$

and

$$\tilde{1}_i = \sum_{k\in\mathbb{N}} \sum \Big\{ \prod_{\ell=0}^{k-1} \rho_{ij_\ell}\rho_{j_\ell i} \;\Big|\; i < j_0 \leqslant \cdots \leqslant j_{k-1} \Big\}. \tag{VJ}$$

[‡] Indeed, for a square matrix $\boldsymbol{A}$ with nonnegative entries, the supremum of the ratios $\|\boldsymbol{A}\vec{v}\| / \|\vec{v}\|$ can be attained by just considering vectors $\vec{v}$ having nonnegative entries.

Then, using the inversion formula $(\mathbf{I}_N - \boldsymbol{A})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{A}^k$ (which is valid here since we apply it in a case where '$\boldsymbol{A}$' is a strictly triangular array), one obtains that (denoting by $\mathfrak{m}_{ij}$ the entry of $\boldsymbol{M}$ at index (i,j)),

$$\mathfrak{m}_{ij} = \sum_{k\in\mathbb{N}} \sum \Big\{ \prod_{\ell=0}^{k-1} \rho_{i_\ell i_{\ell+1}} \;\Big|\; (i_0, i_1, \dots, i_k) \text{ satisfies the \textit{first condition} w.r.t. } (i,j) \Big\}, \tag{VK}$$

where the "first condition" is the following:

Definition 153 (First condition)**.** A sequence $(i_0, \dots, i_k)$ is said to satisfy the *first condition* w.r.t. indices i and j when:

(i) $i_0 = i$;

(ii) One has always $i_\ell \neq i_{\ell+1}$ (provided $i_{\ell+1}$ is defined, i.e. for $\ell < k$);

(iii) $i_1 > i_0$ (provided i_1 is defined, i.e. for $k \geqslant 1$);

(iv) If $i_{\ell+1} < i_\ell$, then $i_{\ell+1} = i_{\ell-1}$;[§]

(v) If $i_{\ell+1} < i_\ell$, then $i_{\ell+2} \geqslant i_\ell$ (provided $i_{\ell+2}$ is defined, i.e. for $\ell \leqslant k-2$);

(vi) $i_k = j$. ♡

On the other hand, using the expansion $(\mathbf{I}_N - \boldsymbol{R})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{R}^k$ (which is convergent here, since we assumed that $|||\boldsymbol{R}||| < 1$), one has a similar formula for $\boldsymbol{M}'$:

$$\mathfrak{m}'_{ij} = \sum_{k\in\mathbb{N}} \sum \Big\{ \prod_{\ell=0}^{k-1} \rho_{i_\ell i_{\ell+1}} \;\Big|\; (i_0, i_1, \dots, i_k) \text{ satisfies the \textit{second condition} w.r.t. } (i,j) \Big\}, \tag{VL}$$

where

Definition 154 (Second condition)**.** A sequence $(i_0, \dots, i_k)$ is said to satisfy the *second condition* w.r.t. indices i and j when it satisfies Conditions (i), (ii) and (vi) of Definition 153. ♡

Since the *second condition* is obviously weaker than the *first condition*, one has $\mathfrak{m}_{ij} \leqslant \mathfrak{m}'_{ij}$ for all i, j, so that $|||\boldsymbol{M}'||| \geqslant |||\boldsymbol{M}|||$: and thus, as Theorem 146 bounds below the spectral gap by $|||\boldsymbol{M}|||^{-2}$, the gap is a fortiori bounded below by $|||\boldsymbol{M}'|||^{-2}$! ♤

4.2.3 Application to concrete models

The point of having got a lower bound for the spectral gap in the previous subsection is that, when considering *infinite* systems of particles, that bound may remain strictly positive, hence ensuring positiveness of the spectral gap for the infinite system. In particular, Corollary 150 yields immediately the following result:

[§] Note that in this case $i_{\ell-1}$ will always be defined, thanks to condition (iii).

Corollary 155. *Regardless of whether I is finite or not, provided*

$$\bar{\rho} \coloneqq \sup_{i\in I} \sum_{j\in I\smallsetminus\{i\}} \rho_{ij} < 1, \tag{VM}$$

the spectral gap of the Glauber dynamics is strictly positive (and bounded from below by $(1-\bar{\rho})^2$). ◇

So, in particular, the respective models of §§ 3.2.2 and 3.3 do have spectral gap for their Glauber dynamics, provided that the bounds of the form "$\rho^\ddagger(\sigma_i,\sigma_j) \leqslant \tau(|j-i|)$" that one gets for such models (confer Equations (`NS`)[¶] and (`OR`)) decay fast enough so that

$$\sum_{u\in\mathbb{Z}^d\smallsetminus\{0\}} \tau(|u|) < 1. \tag{VN}$$

However, the constraint (`VN`) looks somehow "artificial", as in fact, in §§ 3.2.2 and 3.3 we have managed to prove ρ^*-mixing just under the assumption that the sum of the $\tau(|u|)$'s was *finite*: so, it looks natural that the same condition would actually be sufficient to get positiveness of the spectral gap! This is indeed the case, as the next subsection will show: actually it will go along the same principles as what we did in § 3.1.3 to get immediate ρ^*-mixing even when the right-hand side of (`EB`) was not always < 1.

4.2.4 Avoiding the artificial phase transition

The following result shows that, in Corollary 155, bounding the left-hand side of (`VN`) by 1 was not actually necessary to get positiveness of the spectral gap: just having it finite suffices:

Theorem 156 (Spectral gap under general $\rho^\ddagger$-mixing)**.** *In the $I = \mathbb{Z}^d$ case, suppose that for all $i,j \in \mathbb{Z}^d$ one has $\rho^\ddagger(X_i;X_j) \leqslant \rho(j-i)$ for some symmetric function $\rho\colon \mathbb{Z}^d \to [0,1]$ such that $\rho(u) < 1$ as soon as $u \neq 0$. Then, if $\sum_{u\in\mathbb{Z}^d} \rho(u) < \infty$, the spectral gap of the Glauber dynamics is strictly positive.* ◇

Proof. The assumption on $\sum_u \rho(u)$ allows us to take $n < \infty$ large enough so that

$$\sum_{u\in n\mathbb{Z}^d\smallsetminus\{0\}} \rho(u) < 1. \tag{VO}$$

We split $\mathbb{Z}^d$ into a partition of $n^d \eqqcolon N$ "sublattices" $Z_0,\ldots,Z_{N-1}$, each sublattice Z_c being of the form $n\mathbb{Z}^d + u_c$ for some $u_c \in \mathbb{Z}^d/n\mathbb{Z}^d$. Then we define an auxiliary dynamics:

Definition 157 (Sublattice Glauber dynamics)**.** The *sublattice Glauber dynamics* is the Glauber dynamics for $\vec{X}_{\mathbb{Z}^d}$ *considered as the finite-dimensional vector* $(\vec{X}_{Z_0},$

[¶] Beware that the "$\tau(\bullet)$" function in (`NS`) (which dealt with a $\underline{\beta}$-mixing estimate) is not exactly the same as the "$\tau(\bullet)$" of which I spoke below: in fact, the β-mixing bound would have to first be converted into a ρ-mixing bound, like we did for the original Ising model in § 3.2.1.

$\dots, \vec{X}_{Z_{N-1}})$. In other words, for each $c \in \{0, \dots, N-1\}$ there is an independent alarm clock ringing according to a unit-intensity Poisson point process; and when clock c rings, the state of the whole $\vec{X}_{Z_c}$ is flipped in one shot according to $\mathrm{Law}(\vec{X}_{Z_c} \mid \vec{X}_{\mathbb{Z}^d \setminus Z_c})$. ♡

Now let $f \in \bar{\mathrm{L}}^2(\Omega)$. In addition to the notation of the proof of Theorem 146, we introduce the following definition:

Notation 158. For $c \in \{0, \dots, N-1\}$, we denote

$$f^{\neq}_{(c)} \coloneqq f - \mathbb{E}(f \mid \vec{X}_{\mathbb{Z}^d \setminus Z_c}). \tag{VP}$$

♡

Remark 159. The $f^{\neq}_{(c)}$'s are the equivalents, for the sublattice Glauber dynamics, of the $f^{\neq}_i$'s introduced in the proof of Theorem 146. ♧

Fixing some "boundary condition" $\vec{x}_{\mathbb{Z}^d \setminus Z_c}$ on $\mathbb{Z}^d \setminus Z_c$, one can apply Corollary 150 to the Glauber dynamics for $\vec{X}_{Z_c}$ under the law $\mathbb{P}(\bullet \mid \vec{X}_{\mathbb{Z}^d \setminus Z_c} = \vec{x}_{\mathbb{Z}^d \setminus Z_c})$: after integrating, we get that

$$\mathrm{Var}(f^{\neq}_{(c)}) \leqslant (1 - |||\boldsymbol{R}|||)^{-2} \sum_{i \in Z_c} \mathrm{Var}(f^{\neq}_i), \tag{VQ}$$

where $\boldsymbol{R}$ is the operator on $\mathrm{L}^2(n\mathbb{Z}^d)$ defined by

$$\big(\boldsymbol{R}g\big)(i) = \sum_{u \in n\mathbb{Z}^d \setminus \{0\}} \rho(u) g(i+u), \tag{VR}$$

whose norm is obviously bounded by $\sum_{u \in n\mathbb{Z}^d \setminus \{0\}} \eqqcolon \zeta < 1$. Then, summing (VQ) for all c:

$$\sum_{c=0}^{N-1} \mathrm{Var}(f^{\neq}_{(c)}) \leqslant (1-\zeta)^{-2}\, \mathscr{E}(f,f). \tag{VS}$$

Now, let us apply Theorem 146 to the sublattice Glauber dynamics [Definition 157]. It yields that

$$\mathrm{Var}(f) \leqslant |||\boldsymbol{M}|||^2 \sum_{c=0}^{N-1} \mathrm{Var}(f^{\neq}_{(c)}), \tag{VT}$$

where $\boldsymbol{M}$ is some $N \times N$ matrix depending on the $\rho^{\ddagger}(\vec{X}_{Z_c}; \vec{X}_{Z_{c'}})$'s. But by Lemma 91-(ii), $\rho^{\ddagger}(\vec{X}_{Z_c}; \vec{X}_{Z_{c'}}) < 1$ for all $c \neq c'$; thus $|||\boldsymbol{M}||| < \infty$ by Remark 147. Combining (VS) and (VT), we finally get that the spectral gap of the Glauber dynamics for $\vec{X}_{\mathbb{Z}^d}$ is bounded below by $|||\boldsymbol{M}|||^{-2} \times (1-\zeta)^2 > 0$. ♤

Remark 160. Theorem 156 could actually be stated in the more general frame of "abstract" geometric spaces satisfying Assumption 86, like we did for Lemma 91 in § 3.1.3. ♧

Chapter 5

Conclusion and open perspectives

Summary of the monograph So, in this monograph I have explored how, from "partial independence" between individual random variables, this partial independence being stated in the vocabulary of ρ-mixing, one can deduce non-trivial partial independence results (still in terms of ρ-mixing) between bunches of these variables *of arbitrarily large size*: this is what was shown by Theorem 39 (whose proof is far from trivial!), which bounds above the correlation between the bunches by the operator norm of some matrix of individual correlations (Equation (`EB`)). Theorem 39 also came together with a few "siblings", viz. Theorems 37, 47 and 56, which slightly refine it in some particular cases. In § 2.6.3 I have shown that the bounds of all these theorems are (essentially) optimal. Theorem 39 can be seen as an extension of the *tensorization* property of ρ-mixing (Proposition 16)—hence the title "generalized tensorization"—, but with the huge advantage that the assumption of strict independence required for "ordinary" tensorization, which was preventing from applications to interesting "real-life" models, is not needed any more.

In Chapter 3 of the monograph, I have shown that Theorem 39 is indeed applicable to a various set of natural models in statistical mechanics. As a first result, not only we recover (in the completely analytical regime) the well-known property of ρ-mixing between parallel hyperplanes for the Ising model (Proposition 15), but we also extend it to bunches of spins of arbitrary shapes (Theorem 100). Most importantly, generalized tensorization allows to get ρ-mixing results for a large variety of other models (see e.g. Theorems 105 and 116), which, to the best of my knowledge, were inaccessible until now! One other application that I find particularly interesting is that generalized tensorization of ρ-mixing can be used to get $\mathbf{L}^2$ contractivity for evolution systems that do not have a spectral gap (a.k.a. *hypocoercivity* property) for large- or infinite-dimensional models: see Theorem 125 in § 3.4.

In Chapter 4 I also explored other uses of the ideas involved in the proof of Theorem 39, showing that, with similar techniques, it is possible to get *spatial central limit theorem*, as well as *spectral gap for the Glauber dynamics*, for a large class of systems in statistical mechanics. However, as regards applications to classical models, in practice these methods only provide alternative proofs of already known results.

In spite of the achievements reported above, at least two directions for major improvements (at least to my eyes) still remain open. So I would like to end my monograph with explaining these perspectives:

Relaxing the requirement of complete analyticity My generalized tensorization theorems require some assumptions of "$\rho^\ddagger$-mixing": viz., ρ-mixing must also hold under conditioning by fixing arbitrarily the values of an arbitrary number of variables. In the context of spin models with finite-range interactions, this assumption corresponds to the so-called *complete analyticity* criterion [9] (a.k.a. "strong mixing for arbitrary shapes"), which is a very strong requirement: in particular, for many models (including the very important *Ising model with external field*), complete analyticity is known to fail for some ranges of temperature where exponential decay of correlations nevertheless holds [16]. On the other hand, the case of parallel hyperplanes for Ising's model (cf. Proposition 15) strongly suggests that ρ-mixing between infinite bunches of spins *should* still hold as soon as exponential decay of correlation holds… Is this indeed the case? If yes, can it be proved by some generalized tensorization argument? These remain frustratingly open questions in my opinion.

Actually, the "classical" works studying decorrelation for models of spins (with finite-range interactions) (e.g. [29, 30, 31, 32]) have already stressed the importance of some mixing criteria which are strictly weaker than complete analyticity. One of these criteria is weak mixing, which we already encountered in § 1.2; another important one is the so-called property of *strong mixing for cubes* (often merely called "strong mixing") [27, § 2]. These conditions seem to capture the real essence of what causes exponential decay of correlations for such systems: in particular, for Ising's model with external field, weak mixing is known to hold as soon as one is not on the critical line [33];[*] also, in dimension 2, weak and strong mixing are known to be equivalent for spins models [36]… So, it would be an interesting challenge to try to adapt the proof of Theorem 39 (maybe by some clever ordering of the spins?…) so that it would only require, say, strong mixing for cubes!

Another annoying feature of the $\rho^\ddagger$-mixing condition is that the $\rho^\ddagger$-mixing coefficient is defined as a (essential) supremum, so that one must have a *uniform* control for conditioning under (almost-)<u>all</u> the possible values for the boundary conditions: this is a quite "rigid" constraint; which I do not find quite satisfactory…[†] Maybe, rather

---

[*] Conversely, on the critical line, either the equilibrium measure is not unique (which is a kind of extreme correlation!), or one expects correlation between spins to have only polynomial decay [34]: so, from this point of view, long-distance correlations in Ising's model (with external field) exhibit a neat dichotomy, depending on whether the parameter lies on the critical line or not. But, for that model, physicists expect that only <u>one</u> phase transition should occur [35], and therefore that most correlation phenomena should exhibit the same dichotomy: so, this suggests that strong mixing for cubes, resp. ρ-mixing for arbitrary shapes, would also hold as soon as one lies outside the critical line, even if one is not in the completely analytical regime.

[†] Actually this observation also stands for the classical weak- and strong mixing criteria. In practice however, weak- and strong mixing criteria were introduced to study models in which individual spins belong to some finite state space, in which case it is not as a big issue to handle uniformity assumptions as when one deals with continuous variables (as explained in the main text a few lines below). In this monograph, the goal of my work was to show how general a tool it is to tensorize ρ-mixing: in view of which, it would obviously be most natural to try to get rid of uniformity assumptions!

than requiring that conditional $\mathbf{L}^2$ mixing coefficients are bounded uniformly in the boundary condition, one could imagine requirements like, say, that some conditional $\mathbf{L}^4$ mixing coefficient would have to be an $\mathbf{L}^4$-integrable function of the boundary condition… Personally I did not investigate that trail, given the already high technicality of the proof of Theorem 39; however, such a result, if it were true, would be interesting, especially for models with continuous variables: indeed, for such models, one would like to allow for some slight "irregularities" (for instance, in the model of § 3.3, it would be nice not to impose the force potentials to be convex *everywhere* as we did in Assumption 115); but such irregularities make it hard or even impossible to satisfy assumptions *uniformly*: and in that case, less rigid hypotheses could be very welcome! ◡̈

Hypocoercivity for nonlinear models As already mentioned above, I find it quite of interest that one may apply generalized tensorization of ρ-mixing to derive hypocoercivity results for large or infinite systems in statistical mechanics: in particular, it might provide a way to show how large systems of particles, in presence of "weak" noise, could nevertheless have "fast" relaxation to equilibrium, which would be very interesting from a physical point of view.[‡] Applying generalized tensorization of ρ-mixing to hypocoercivity seems all the more promising that it is apparently a quite new idea, so that it might solve questions out of the scope of the current analytic techniques: that is why I devoted § 3.4 to a proof of concept for such an application. However, the example of § 3.4 is an extremely specific toy model, because it is linear; and I made heavy use of its linearity to check the pairwise mixing assumptions in my hypocoercivity proof… So, it remains an open challenge to see if one could prove the assumptions required by generalized tensorization of ρ-mixing for more general, nonlinear models (say, for something akin to the model of § 3.3),

---

[‡] Let me elaborate a bit on that, though I have to warn as a disclaimer that the following thoughts are still very speculative. In classical physics, it often occurs that one deals with a system which is fully *deterministic* (e.g. a gas of rigid spherical atoms, or a lattice of coupled anharmonic oscillators), but which is expected to behave, at large scales, according to a certain *non-reversible* dynamics (e.g. the Boltzmann equation, resp. the Fourier law), that macroscopic dynamics being derived informally from the microscopic model by *statistical* arguments, assuming as an ansatz that the system will satisfy some *local equilibrium* behaviour at any time. But proving the long-time stability of that local equilibrium is very difficult when the only source of randomness lies in the initial condition (see e.g. [37] as regards the Boltzmann equation): and actually, it looks out of reach of today's mathematical techniques! However, one might try to circumvent that issue by introducing a small, conservative noise in the system: provided that that noise could be taken arbitrarily small at the macroscopic scale, this would not be that much of "cheating" with the reality of physics. But even that approach remains quite hard to handle: for instance, in the works of Rezakhanlou [38] (who investigates this trail for the Boltzmann equation) or of Olla & al. [39, 40] (for the Fourier law), the kind of perturbations that you need introduce to make things work properly do not look quite satisfactory from a physical point of view… Considering all that, getting able to make use of small perturbations of (conservative) Langevin type would be a really natural goal: and in this perspective, obtaining fast relaxation to local equilibrium would be more or less equivalent to showing that some hypocoercivity property holds. Far-fetched as the above argument may look, I nevertheless really believe that understanding the relaxation of particle evolution systems with a weak, degenerate, volumic noise would be of genuine interest for classical physics!

in order to get hypocoercivity for such models: and solving this challenge would be of great interest in my opinion!

Appendix A

Auxiliary results

A.1 On the norm of self-adjoint operators

Lemma 161. *Let L be a self-adjoint operator on a real Hilbert space H, and let $C < \infty$. Then, to prove that $|||L||| \leqslant C$, it suffices to ensure that the set*

$$\{x \in H \mid \overline{\lim_{k\to\infty}} |\langle L^k x, x\rangle|^{1/k} \leqslant C\} \tag{VU}$$

is dense in H. ◇

Proof. Reasoning by contraposition, we will to show that, for L a self-adjoint operator on H, for all $C < |||L|||$, the set of the $x \in H$ such that

$$\overline{\lim_{k\to\infty}} |\langle L^k x, x\rangle|^{1/k} > C \tag{VV}$$

contains a nonempty open subset of H.

Since L is self-adjoint, by the spectral theorem [41, Theorem 7.18], it is unitarily equivalent to the "multiplication by identity" operator M on a space $\bigoplus_{\alpha\in A} \mathrm{L}^2(\mu_\alpha)$, for A some set and μ_α some Radon measures on $\mathbb{R}$, that is: [see the footnote for precisions about the notation used in the equation],

$$M\Big(\bigoplus_{\alpha\in A} f_\alpha(\lambda)\Big) = \bigoplus_{\alpha\in A} \lambda f_\alpha(\lambda)^{[*]}. \tag{VW}$$

So we will assume L is of that form.

One has obviously:

$$|||L||| = \sup\{\lambda \geqslant 0 \mid \exists \alpha \in A \quad \mu_\alpha(\mathbb{R} \smallsetminus [\pm\lambda]) > 0\}; \tag{VX}$$

moreover, for all $f =: \bigoplus_\alpha f_\alpha \in H$,

$$\langle L^k f, f\rangle = \sum_\alpha \int_{\mathbb{R}} \lambda^k |f_\alpha(\lambda)|^2 \, \mathrm{d}\mu_\alpha(\lambda), \tag{VY}$$

---

[*] In this equation, we used the summation symbol '$\bigoplus$' instead of '$\sum$' to mean implicitly that each f_α leaves in the appropriate subspace $\mathrm{L}^2(\mu_\alpha)$ of which H is the direct sum. Also, 'λ' just stands for the free variable, so that "$f(\lambda)$" is synonymous with f and "$\lambda f(\lambda)$" denotes the pointwise multiplication of f with the identity operator.

so that (observing that, for k even, $\lambda^k \geqslant 0\ \forall\lambda$),

$$\varlimsup_{k\to\infty} |\langle L^k f, f\rangle|^{1/k} = \sup\Big\{ \lambda \geqslant 0 \ \Big|\ \exists\alpha\in A \quad \int_{\mathbb{R}\setminus[\pm\lambda]} |f_\alpha(\lambda')|^2 \, \mathrm{d}\mu_\alpha(\lambda') > 0 \Big\}. \tag{VZ}$$

Now, for $C < |||L|||$, the set

$$U := \Big\{ f\in H \ \Big|\ \exists\alpha\in A \quad \int_{\mathbb{R}\setminus[\pm C]} |f_\alpha(\lambda)|^2 \, \mathrm{d}\mu_\alpha(\lambda) > 0 \Big\} \tag{WA}$$

is open because $\int_{\mathbb{R}\setminus[\pm C]} |f_\alpha(\lambda)|^2 \, \mathrm{d}\mu_\alpha(\lambda)$ is a continuous function of f, and it is nonempty by (VX). But (VV) is satisfied for all $x\in U$ by (VZ), so U fulfills our quest. ♤

A.2 A lemma à la Perron–Frobenius for infinite matrices

The goal of this section is to extend Lemma 46 to the case where the "matrix" $\boldsymbol{A}$ has countably infinite size. Here we will consider a set of indices with no particular structure: for that reason, our set of indices shall be denoted by $\aleph_0$. In this context, by "vector", I will mean an element of $\mathrm{L}^2(\aleph_0)$, i.e. an $\aleph_0$-indexed square-summable family of real numbers. Like in Lemma 46, I denote $\vec{v} \geqslant 0$, resp. $\vec{v} > 0$, to mean that all entries of $\vec{v}$ are nonnegative, resp. strictly positive. By "square matrix", I will mean an array $\boldsymbol{A}$ indexed by $\aleph_0\times\aleph_0$; and (like in Lemma 46 again) I denote $\boldsymbol{A}\geqslant 0$ to mean that all the entries of $\boldsymbol{A}$ are nonnegative. If $\boldsymbol{A}$ is a square matrix with $\boldsymbol{A}\geqslant 0$, then for all $\vec{v}\geqslant 0$, one defines $\boldsymbol{A}\vec{v}$ by the ordinary matrix-vector product rule, which gives an $\aleph_0$-indexed family of values of $\mathbb{R}_+ \sqcup \{\infty\}$. If that family is square-integrable for all vectors $\vec{v}\geqslant 0$ in $\mathrm{L}^2(\aleph_0)$, then the matrix-vector product rule may be extended (with absolute convergence) to all vectors in $\mathrm{L}^2(\aleph_0)$, and moreover in this case the exists some $C<\infty$ such that $\|\boldsymbol{A}\vec{v}\|_{\mathrm{L}^2(\aleph_0)} \leqslant C\|\vec{v}\|_{\mathrm{L}^2(\aleph_0)}$ for all $\vec{v}$: in such a case, I will say that $\boldsymbol{A}$ is *bounded*. When the matrix $\boldsymbol{A}$ is bounded, it can also be interpreted as a bounded linear endomorphism on the Hilbert space $\mathrm{L}^2(\aleph_0)$; and then one can define its operator norm and its spectral radius in the usual way.

That being set, the result of this section is the following:

Lemma 162. *Assume $\boldsymbol{A}$ is a bounded nonnegative square "matrix" indexed by $\aleph_0$. Assume moreover that $\boldsymbol{A}$ is symmetric. Then,*

$$\inf\{\lambda\geqslant 0 \mid (\exists\vec{u}>0)\ \boldsymbol{A}\vec{u}\leqslant\lambda\vec{u}\} = \rho_{\mathrm{sp}}(\boldsymbol{A}). \tag{WB}$$

(By the way, since $\boldsymbol{A}$ is assumed to be symmetric, here $\rho_{\mathrm{sp}}(\boldsymbol{A})$ is actually the same thing as $|||\boldsymbol{A}|||$). ◇

Remark 163. The boundedness assumption on $\boldsymbol{A}$ is rather natural here (as one speaks of the spectral radius of $\boldsymbol{A}$ seen as an operator), all the more that it was automatically satisfied in the finite case (i.e. in Lemma 46). The symmetry assumption, on the other hand, is unexpected—and quite strong! Yet removing that assumption would make

Lemma 162 invalid, as shown by the following counterexample: take $\boldsymbol{A}$ to be the left shift operator on $\mathbb{N}$ (which is nonnegative and bounded); then $\rho_{\mathrm{sp}}(\boldsymbol{A}) = 1$, while for all $\lambda > 0$, the vector $(\lambda^i)_{i\in\mathbb{N}}$ (which is in $\mathrm{L}^2(\mathbb{N})$ provided $\lambda < 1$) does satisfy $\boldsymbol{A}\vec{u} = \lambda\vec{u}$. ♧

Proof of Lemma 162. For the '$\leqslant$' sense of Equation (`WB`), consider any $\lambda > \rho_{\mathrm{sp}}(\boldsymbol{A})$: our goal will be to find a strictly positive $\vec{u}$ such that $\boldsymbol{A}\vec{u} \leqslant \lambda\vec{u}$. To this end, let $\vec{u}_0$ be any strictly positive vector of $\mathrm{L}^2(\aleph_0)$ (such a vector exists, since $\aleph_0$ is countable). For all $k \in \mathbb{N}$, one has

$$\|\boldsymbol{A}^k\vec{v}\|_{\mathrm{L}^2} \leqslant |||\boldsymbol{A}^k||| \|\vec{v}\|_{\mathrm{L}^2}; \tag{WC}$$

and since $\lim_{k\to\infty}|||\boldsymbol{A}^k|||^{1/k} \underset{\mathrm{def}}{=} \rho_{\mathrm{sp}}(\boldsymbol{A}) < \lambda$, it ensues that the series $\sum_{k\in\mathbb{N}} \lambda^{-k}\boldsymbol{A}^k\vec{u}_0$ is convergent in $\mathrm{L}^2(\aleph_0)$.

So, let us denote

$$\vec{u} \coloneqq \sum_{k\in\mathbb{N}} \lambda^{-k}\boldsymbol{A}^k\vec{u}_0. \tag{WD}$$

As $\vec{u}$ is the sum of $\vec{u}_0$ (which is > 0) and of the $\lambda^{-k}\boldsymbol{A}^k\vec{u}_0$ for $k \neq 0$ (which all are $\geqslant 0$, since $\boldsymbol{A} \geqslant 0$), one has $\vec{u} > 0$. Moreover, the very construction of $\vec{u}$ ensures that $\lambda\vec{u} = \lambda\vec{u}_0 + \boldsymbol{A}\vec{u}$, where one has $\lambda\vec{u}_0 > 0$: so, $\boldsymbol{A}\vec{u} \leqslant \lambda\vec{u}$. This ends the '$\leqslant$' sense of the proof.

Now let us prove the '$\geqslant$' sense of Equation (`WB`). We assume that $\rho_{\mathrm{sp}}(\boldsymbol{A}) > 0$, otherwise there is nothing to prove. Let $\lambda < \rho_{\mathrm{sp}}(\boldsymbol{A})$ (with $\lambda \geqslant 0$): our goal will be to prove that for all $\vec{u} > 0$, one has $\boldsymbol{A}\vec{u} \nleqslant \lambda\vec{u}$. Since we assumed $\boldsymbol{A}$ to be symmetric, one has $\rho_{\mathrm{sp}}(\boldsymbol{A}) = |||\boldsymbol{A}|||$; and that property is also true for all the square submatrices of $\boldsymbol{A}$ [a "square submatrix of $\boldsymbol{A}$" being the matrix obtained from $\boldsymbol{A}$ when one restricts the set of indices to some subset]. But $|||\boldsymbol{A}|||$ is the supremum of the $|||\boldsymbol{A}'|||$'s for $\boldsymbol{A}'$ square submatrices of $\boldsymbol{A}$ of finite size [this comes from a density argument in $\mathrm{L}^2(\aleph_0)$]: therefore, the assumption that $\lambda < \rho_{\mathrm{sp}}(\boldsymbol{A})$ implies that $\lambda < \rho_{\mathrm{sp}}(\boldsymbol{A}')$ for some square submatrix of $\boldsymbol{A}$ of finite size. Let us denote by $\boldsymbol{I}$ the set of indices of that submatrix. We can apply the original Lemma 46 to matrix $\boldsymbol{A}'$: denoting by $\vec{u}_{\restriction\boldsymbol{I}}$ the restriction of $\vec{u}$ to $\mathrm{L}^2(\boldsymbol{I})$ (which is > 0 in $\mathrm{L}^2(\boldsymbol{I})$), it then yields that $\boldsymbol{A}'\vec{u}_{\restriction\boldsymbol{I}} \nleqslant \lambda\vec{u}_{\restriction\boldsymbol{I}}$. But by positiveness of $\boldsymbol{A}$, $(\boldsymbol{A}\vec{u})_{\restriction\boldsymbol{I}} \geqslant \boldsymbol{A}'\vec{u}_{\restriction\boldsymbol{I}}$: so, one has a fortiori that $(\boldsymbol{A}\vec{u})_{\restriction\boldsymbol{I}} \nleqslant \lambda\vec{u}_{\restriction\boldsymbol{I}}$, which finally implies that $\boldsymbol{A}\vec{u} \nleqslant \lambda\vec{u}$. ♤

A.3 Technical checks on the system of § 3.3

In § 3.3, we investigated the system of statistical physics defined formally by Equation (`OP`) in order to establish its ρ^*-mixing properties. In that investigation, we used a couple of classical arguments of statistical physics: in particular, we defined the probability distribution on our system via its Hamiltonian; we viewed the conditional distributions of the system as the equilibria of some Langevin dynamics; and we used convexity properties of the Hamiltonian to ensure that everything worked properly. However, in full rigour, all these arguments are only guaranteed to work for *finite-dimensional* systems, which ours was not… So, this appendix section gathers the technical arguments needed to ensure that all the claims that we made on the behaviour of our infinite-dimensional system were actually valid.

Remember that our system is made of continuous-valued spins laid on an infinite lattice, submitted to pinning and nearest-neighbour interactions: it is formally defined by Hamiltonian (`OP`), where the pinning potential $V(\bullet)$ and the interaction potential $W(\bullet)$ have to satisfy Assumption 115 regarding convexity. The "polarization space" that each spin lives in is $\mathbb{R}^\ell =: \mathcal{S}$, which is endowed with its usual Euclidean norm $|\bullet|$; while the lattice according to which the spins are laid out is $\mathbb{Z}^d =: \mathbb{V}$, which is endowed with the ℓ^1 graph distance $|\bullet - \bullet|$, also denoted by $\mathit{dist}(\bullet, \bullet)$. The system is studied at some fixed inverse temperature $\beta \in (0, \infty)$. In this appendix we will refer to that system as the "LoNP system" (for "Lattice of Nonlinear Particles").

A.3.1 Giving a precise meaning to the LoNP system

The goal of this first subsection is to give a precise meaning to the law of our LoNP system. To do so, we will first establish estimates for finite-dimensional versions of the system; then, thanks to these estimates, we will be able to pass to the limit in a coherent way.

LoNP conditioned at the boundary of a box A case where everything is defined without ambiguity is when we consider the system defined on the whole $\mathbb{V}$, but that we only look at what happens inside Λ, and that we have fixed a *boundary condition* on the boundary $\partial\Lambda$ of Λ, that "boundary" being defined as the set of vertices of $\mathbb{V} \setminus \Lambda$ that are connected to at least one vertex in Λ. Indeed, the structure of the (full) Hamiltonian of the system implies that what happens outside Λ may only interfere with what happens inside Λ through what happens on $\partial\Lambda$, so that, once you consider a conditioning of the form $\{\vec{\sigma}_{\partial\Lambda} \equiv \vec{s}_{\partial\Lambda}\}$ (for some $\vec{s}_{\partial\Lambda} \in \mathcal{S}^{\partial\Lambda}$), it does not matter whether you are considering the system on the whole $\mathbb{V}$ (which, a priori, might be ill-defined) or if you are considering its restriction to some finite $\Lambda' \supseteq \Lambda \cup \partial\Lambda$ (for which everything is well-defined). In view of which, one can define the conditional law for the full system as the conditional law for the system restricted to Λ': this law does indeed not depend (as regards what happens in Λ) on the choice of Λ'. (This is actually nothing but the principle of a Gibbs measure!).

Also, note that in such a case the conditional distribution $\mathrm{Law}(\vec{\sigma}_\Lambda \mid \vec{\sigma}_{\partial\Lambda} \equiv \vec{s}_{\partial\Lambda})$ will be the (unique) equilibrium distribution for the overdamped Langevin dynamics associated to the conditioned system, and that that dynamics will converge to that equilibrium: indeed, under the conditions that we have required, the conditioned Langevin dynamics is nothing but a Fokker–Planck diffusion in some convex potential, for which all of that is well known (see e.g. [18, § 4.5]).

Until now, we have just recalled some elementary facts. Now let us turn to a first non-trivial lemma, which tells that the law of the system on Λ depends on the boundary condition in a continuous way:

Lemma 164. *There exist constants $\gamma > 0$ and $C < \infty$ such that the following holds: for Λ a finite subset of $\mathbb{V}$, for $\vec{s}^{\,0}_{\partial\Lambda}, \vec{s}^{\,1}_{\partial\Lambda}$ two boundary conditions on $\partial\Lambda$, for all $I \subseteq \Lambda$,* *the conditional distributions on I corresponding to the respective boundary conditions are close to each other in terms of Wasserstein distance,*[†] *with the following control:*

$$\mathrm{W}_2\big(\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_{\partial\Lambda} \equiv \vec{s}^{\,0}_{\partial\Lambda}), \mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_{\partial\Lambda} \equiv \vec{s}^{\,1}_{\partial\Lambda})\big)^2 \leqslant C\mathrm{e}^{-2\gamma\, dist(I,\partial\Lambda)} \times \sum_{k\in\partial\Lambda} |s^1_k - s^0_k|^2. \tag{WE}$$

◇

Proof. This lemma is actually completely similar to Proposition 120, and gets proved in exactly the same way. The only differences are the following: the vertex i gets replaced by the set of vertices I; the vertex j gets replaced by the boundary $\partial\Lambda$; and there is no set K any more. As a consequence, here the Lyapunov functional would become

$$\mathcal{U}(\vec{\eta}_\Lambda) \coloneqq \sum_{i\in\Lambda} \mathrm{e}^{2\gamma\, dist(i,\partial\Lambda)} |\eta_i(t)|^2; \tag{WF}$$

but apart from these modifications, everything works just the same! ♤

Finite LoNP with free boundary Besides the situation of a finite box with fixed boundary condition, one may also consider the restriction of the LoNP system with free boundary condition. This corresponds, for finite $\Lambda \subseteq \mathbb{V}$, to considering the following Hamiltonian:

$$\mathcal{H}_\Lambda(\vec{\sigma}_\Lambda) \coloneqq \sum_{i\in\Lambda} V(\sigma_i) + \tfrac{1}{2} \sum_{\substack{i,j\in\Lambda \\ |j-i|=1}} W(\sigma_j - \sigma_i). \tag{WG}$$

Here, Equation (WG) is a finite sum, and the state space $\mathcal{S}^\Lambda$ concerned by $\mathcal{H}_\Lambda$ is finite-dimensional; so there is absolutely no ambiguity in defining the Gibbs measure at inverse temperature β for that Hamiltonian: we will denote it by $\mathbb{P}_\Lambda(\bullet)$, and refer to it as “the LoNP system restricted to Λ”. Also, convexity of the Hamiltonian ensures that the (overdamped) Langevin dynamics associated to this restricted system (is well-defined and) converges to the Gibbs measure as times goes to infinity, for any starting state.

Now, our next lemma will consist in controlling the polarization of individual spins for such finite-dimensional versions of the LoNP system:

Lemma 165. *There exists a constant $C < \infty$*[‡] *such that, for all finite $\Lambda \subseteq \mathbb{V}$, when considering the LoNP system restricted to Λ (with free boundary), for all $i \in \Lambda$, one has the following L^2 control on the polarization of spin i:*

$$\mathbb{E}_\Lambda(|\sigma_i|^2) \leqslant C. \tag{WH}$$

◇

Proof. Let $i \in \Lambda$; and let us consider the Langevin dynamics $(\vec{\sigma}_\Lambda(t))_{t\in\mathbb{R}_+}$ for our restricted system, starting from state $\vec{\sigma}_\Lambda(t=0) \equiv 0$ (i.e., initially all the spins have

---

[†] As regards the W_2 Wasserstein distance, see Definition 118. Here the distance considered on $\mathcal{S}^I$ is the “meta-Euclidian” one: $dist(\vec{s}^{\,0}_I, \vec{s}^{\,1}_I) \coloneqq \sum_{i\in I} |s^1_i - s^0_i|^2$.

[‡] The important point here is that C only depends on $V(\bullet)$, $W(\bullet)$ and β, but not on Λ nor on i.

polarization 0).[§] We will show that this dynamics cannot go too "wild" by considering the following Lyapunov functional:

$$\mathscr{U}(\vec{\sigma}_\Lambda(t)) \coloneqq \sum_{j\in\Lambda} \mathrm{e}^{-2\gamma|j-i|} |\sigma_j(t)|^2. \tag{WI}$$

Below we will shorthand $\mathrm{e}^{-2\gamma|j-i|} \eqqcolon w_j$.

Let us remind that the Langevin dynamics is defined by

$$\mathrm{d}\sigma_j(t) \coloneqq \Big(-\nabla V(\sigma_j(t)) + \sum_{|j'-j|=1} \nabla W(\sigma_{j'}(t) - \sigma_j(t))\Big)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}\boldsymbol{B}_t^j, \tag{WJ}$$

where the $(\mathrm{d}\boldsymbol{B}_t^j)_{t\in\mathbb{R}_+}$, $j \in \Lambda$ are independent ℓ-dimensional white noises. From (WJ), we get that the evolution of $\mathscr{U}(t)$ is made up of four terms:

- Evolution due to the pinning forces, contributing to the temporal derivative $\dot{\mathscr{U}}$ by $-\sum_{j\in\Lambda} w_j\sigma_j \cdot \nabla_{\upharpoonright\sigma_j} V$. Using that $\nabla_{\upharpoonright\sigma_j} V = \nabla_{\upharpoonright 0} V + \nabla\nabla_{\upharpoonright\sigma'} V \cdot \sigma_j$ for some σ' in the segment $[0,1]\sigma_j$, and that $\nabla\nabla_{\upharpoonright\sigma'} V \geqslant \alpha \mathbf{I}_\ell$, we get that this term is bounded above by

  $$\begin{aligned} &-\sum_{j\in\Lambda} w_j(\sigma_j \cdot \nabla_{\upharpoonright 0} V + \alpha|\sigma_j^2|) \\ &\leqslant \frac{1}{2}\sum_{j\in\Lambda} w_j(\delta|\sigma_j|^2 + \delta^{-1}|\nabla_{\upharpoonright 0}V|^2) - \alpha\mathscr{U}(t) = -(\alpha-\delta/2)\mathscr{U}(t) + C_1\delta^{-1}|\nabla_{\upharpoonright 0}V|^2/2, \end{aligned} \tag{WK}$$

  where $\delta \in \mathbb{R}_+^*$ is a constant that we will have to choose small enough, and C_1 is the finite constant $\sum_{u\in\mathbb{Z}^d} \mathrm{e}^{-2\gamma|u|}$.

- Evolution due to interaction forces. The contribution of the interaction forces to $\dot{\mathscr{U}}$ appears as a sum of terms indexed by the couples (j, j') such that $|j'-j| = 1$. We will group these terms by edges, i.e. by *pairs* $\{j, j'\}$ of neighbouring vertices. Let us look at what happens regarding such an edge: it contributes to $\dot{\mathscr{U}}$ by

  $$(w_j\sigma_j - w_{j'}\sigma_{j'}) \cdot \nabla_{\upharpoonright\sigma_{j'}-\sigma_j} W. \tag{WL}$$

  We decompose $w_j\sigma_j - w_{j'}\sigma_{j'}$ into $\frac{1}{2}(w_j + w_{j'})(\sigma_j - \sigma_{j'}) + \frac{1}{2}(w_j - w_{j'})(\sigma_j + \sigma_{j'})$, which yields a decomposition of (WL) into two terms. The first term is nonpositive because of the convexity assumption on $W(\bullet)$. Assuming to fix ideas that $w_{j'} \geqslant w_j$, the second term is bounded above by

  $$\begin{aligned} \frac{1}{2}(w_{j'} - w_j)|\sigma_j + \sigma_{j'}|B|\sigma_{j'} - \sigma_j| &\leqslant \frac{(\mathrm{e}^{2\gamma}-1)B}{2} w_j(|\sigma_j|^2 + |\sigma_{j'}|^2) \\ &\leqslant \frac{(\mathrm{e}^{2\gamma}-1)B}{2}(w_j|\sigma_j|^2 + w_{j'}|\sigma_{j'}|^2). \end{aligned} \tag{WM}$$

[§] Here we are assuming implicitly that polarization 0 is actually allowed for our spins, i.e. that $V(s=0) < \infty$. At first sight this looks restrictive, because there are some interesting cases that do not abide by this property: think e.g. of $V(s) \leftarrow |s|^2 + \mathbf{1}_{|s-u|>1/2} \cdot \infty$ for some unit vector u. However, we can always change the origin of space $\mathscr{S}$ to ensure that $V(\bullet)$ is finite on a neighbourhood of 0 (as it is implicitly assumed, in any case, that $V(\bullet)$ is finite on a nonvoid open set): so this is not a real issue.

So, we have got an upper bound on the contribution of each edge to $\dot{\mathscr{U}}$; and by summing on all edges, we get that the total contribution of the interaction forces to $\dot{\mathscr{U}}$ is bounded above by $d(\mathrm{e}^{2\gamma}-1)B\mathscr{U}(t)$.

- Stochastic term. This term is $\sum_{j\in\Lambda} 2\sqrt{2\beta^{-1}}\, w_j\sigma_j \cdot \mathrm{d}B_t^j$; so it is a white noise with variance by unit of time equal to
$$8\beta^{-1}\sum_{j\in\Lambda} w_j^2|\sigma_j|^2 \leqslant 8\beta^{-1}\mathscr{U}(t) \tag{WN}$$
(using that all the w_j's are $\leqslant 1$).
- Itô term. This one is just $\sum_{j\in\Lambda} 2\beta^{-1}\ell w_j\,\mathrm{d}t \leqslant 2\ell C_1\beta^{-1}\,\mathrm{d}t$.

All in all, we have the following control on the evolution of the Lyapunov functional:

$$\begin{aligned}\mathrm{d}\mathscr{U}(t) \leqslant\ & -\big(\alpha-(\mathrm{e}^{2\gamma}-1)Bd-\tfrac{1}{2}\delta\big)\mathscr{U}(t)\,\mathrm{d}t+\big(\tfrac{1}{2}\delta^{-1}|\nabla_{\restriction 0}V|^2+2\ell\beta^{-1}\big)C_1\,\mathrm{d}t\\ &+2\sqrt{2\beta^{-1}}\,\mathscr{U}(t)^{1/2}\,\mathrm{d}B_t,\end{aligned} \tag{WO}$$

for $(\mathrm{d}B_t)_{t\in\mathbb{R}_+}$ some white noise (adapted to the natural filtration of the process).

Assume that we have chosen γ small enough so that $(\mathrm{e}^{2\gamma}-1)Bd<\alpha$. Then, we can choose δ small enough so that the factor in front of $\mathscr{U}(t)\,\mathrm{d}t$ in (WO) is strictly negative. So let us assume that we did so; denote that strictly negative factor by $-C_2$; and also denote $\frac{1}{2}\delta^{-1}|\nabla_{\restriction 0}V|^2+2\ell\beta^{-1} =: C_3$. By an immediate comparison argument, $\mathscr{U}(t)$ is then bounded (at all times, almost-surely) by a process $\tilde{\mathscr{U}}(t)$ that satisfies the stochastic differential equation

$$\mathrm{d}\tilde{\mathscr{U}}(t) = -C_2\tilde{\mathscr{U}}(t)\,\mathrm{d}t + C_1C_3\,\mathrm{d}t + 2\sqrt{2\beta^{-1}}\,\mathscr{U}(t)^{1/2}\,\mathrm{d}B_t, \tag{WP}$$

starting from the initial condition $\tilde{\mathscr{U}}(t=0)\equiv 0$. Then, taking expectations formally [see Footnote [¶] for a justification of why it is legit here], we get that

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}(\tilde{\mathscr{U}}(t)) = -C_2\,\mathbb{E}(\tilde{\mathscr{U}}(t)) + C_1C_3, \tag{WQ}$$

from which, given the initial condition, one gets by a comparison argument that $\mathbb{E}(\tilde{\mathscr{U}}(t)) \leqslant C_1C_3\,/\,C_2$ at all times, and therefore that $\mathbb{E}(|\sigma_i(t)|^2) \leqslant \mathbb{E}(\mathscr{U}(t)) \leqslant C_1C_3\,/\,C_2$ at all times. But, passing to the limit as $t\to\infty$, the law of $\sigma_i(t)$ tends to the law of σ_i for the LoNP process restricted to Λ; and therefore we have proved the wanted result, with $C \leftarrow C_1C_3\,/\,C_2$. ♤

[¶] It is not clear a priori that $\mathbb{E}(\tilde{\mathscr{U}}(t))$ would be well-defined and that the expectation of the "$\mathscr{U}(t)^{1/2}\,\mathrm{d}B_t$" term in (WP) would really be zero… However, what actually interests us is not Equation (WQ) itself, but just the estimate "$\mathbb{E}(\tilde{\mathscr{U}}(t)) \leqslant C_1C_3\,/\,C_2$" that gets deduced from it: and the latter can be proved by classical stopping time arguments (which were kindly explained to me by S. MAZZONETTO), as follows. For $M<\infty$, let τ_M be the first time when $\tilde{\mathscr{U}}(t)$ attains some value M. Considering the stopped processes $\mathscr{U}(t\wedge\tau_M)$ and $\tilde{\mathscr{U}}(t\wedge\tau_M)$, everything becomes bounded, and then one can rigorously integrate (WP), take expectations and apply Grönwall's lemma to obtain that $\mathbb{E}\big(\tilde{\mathscr{U}}(t\wedge\tau_M)\big) \leqslant C_1C_3\,/\,C_2$ at all times: but since that bound is independent of M, one finally gets the announced result by letting $M\to\infty$.

LoNP as a limit of finite-dimensional systems Now that we are armed with Lemmas 164 and 165, we can combine them to get the following

Proposition 166 (Convergence of restricted laws to a limit). *As Λ grows to the whole $\mathbb{V}$, the law of the restricted system $\vec{\sigma}_\Lambda$ converges to a limit, in the following technical sense: for any finite $I \subseteq \mathbb{V}$, denoting by $\mathscr{L}_n$ the set of all finite "boxes" $\Lambda \subseteq \mathbb{V}$ containing all the vertices at distance $\leqslant n$ from I [so, $\mathscr{L}_n$ is a set of "large enough" boxes], denoting by $\mathbb{P}_\Lambda(\bullet)$ the law of the system restricted to Λ (with free boundary condition), one has*

$$\operatorname{diam}\{\mathrm{Law}_\Lambda(\vec{\sigma}_I) \mid \Lambda \in \mathscr{L}_n\} \overset{n\to\infty}{\to} 0, \tag{WR}$$

where the distance considered between the laws on $\mathcal{S}^I$ is the W_2 *Wasserstein distance.*

◇♤

Finally, via Proposition 166 we can define "the"[‖] law of the infinite-dimensional system in a rigorous way: namely, we consider it to be the law whose finite-dimensional marginals $\mathrm{Law}(\vec{\sigma}_I)$ are the respective limits of $\mathrm{Law}_\Lambda(\vec{\sigma}_I)$ when $\Lambda \to \mathbb{V}$! (where "$\Lambda \to \mathbb{V}$" means that, for all n, one has ultimately $\Lambda \in \mathscr{L}_n$).

By the way, by passing to the limit in Lemma 165, we also have the following control on the individual spins for that infinite-dimensional law:

Corollary 167 (L^2 control on the law of spins). *There exists a constant $C < \infty$ (the same as in Lemma 165) such that, for the equilibrium distribution on the* global *system $\vec{\sigma}_\mathbb{V}$, for all $i \in \mathbb{V}$:*

$$\mathbb{E}(|\sigma_i|^2) \leqslant C. \tag{WS}$$

◇♤

A.3.2 Log-concavity of the LoNP marginals

Now that we have given a precise meaning to the LoNP system, let us turn to a second part of this section, viz., proving that the conditional finite-dimensional laws of the LoNP system are log-concave (in other words, that the logarithm of their density w.r.t. the Lebesgue measure is concave).

Actually, if the system were *globally* finite-dimensional—say, if we were considering the system restricted to Λ—, log-concavity would be fairly easy to check. The reasoning would go along the following lines: since $\mathrm{Law}(\vec{\sigma}_{\Lambda\smallsetminus K} \mid \vec{\sigma}_K \equiv \vec{s}_K)$ can be described by a convex Hamiltonian, it has log-concave density w.r.t. the Lebesgue measure on $\mathcal{S}^{\Lambda\smallsetminus K}$; and thus, by the Prékopa-Leindler inequality [42, Theorem 8], for $I \subseteq \Lambda \smallsetminus K$, the law $\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K)$ (which is the pushforward, by the canonical projection from $\mathcal{S}^{\Lambda\smallsetminus K}$ to $\mathcal{S}^I$, of $\mathrm{Law}(\vec{\sigma}_{\Lambda\smallsetminus K} \mid \vec{\sigma}_K \equiv \vec{s}_K)$) is a log-concave measure too.

However, in the actual situation we are in infinite-dimensional context; and the notion of being "log-concave" in such a context is ill-defined… Hence, we need to check formally that the log-concavity property actually remains valid in that case. This is the following result:

[‖] In practice, there may (and will) exist other Gibbs measures; but the one that we have been defining is canonical.

Lemma 168. *For finite $I \subseteq \mathbb{V}$, for (not necessarily cofinite) $K \subseteq \mathbb{V} \setminus I$, for* $\mathrm{Law}(\vec{\sigma}_K)$*-almost-all $\vec{s}_K \in \mathcal{S}^K$, the law* $\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K)$ *is log-concave.* ◇

Proof. The finite-dimensional argument explained at the beginning of this subsection shows that, for Λ a finite box containing I, one has log-concavity for the law of $\vec{\sigma}_I$ under the conditional law where the spins of $\partial\Lambda \setminus K$ would be set to zero, i.e. for the law

$$\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K, \vec{\sigma}_{\partial\Lambda\setminus K} \equiv 0) =: \mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_I \mid \vec{\sigma}_{\partial\Lambda\setminus K} \equiv 0). \tag{WT}$$

The point will now consist in getting convergence, as $\Lambda \to \mathbb{V}$, of the above law to the actual law $\mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_I)$, in a way that is compatible with preservation of log-concavity. In the sequel we will more specifically consider boxes Λ of the form $\{i' \in \mathbb{V} \mid \mathit{dist}(I, i') < \delta\}$, so that, for $i \in I, j \notin \Lambda$, one has always $\mathit{dist}(I, j) \geqslant \mathit{dist}(I, \partial\Lambda)$.

The same argument as in Proposition 120 and Lemma 164, considering in this case the Lyapunov functional

$$\mathcal{U}(\vec{\eta}) := \sum_{i\in\Lambda\setminus K} \mathrm{e}^{2\gamma\,\mathit{dist}(i,\partial\Lambda)} |\eta_i(t)|^2, \tag{WU}$$

shows that, for any boundary condition $\vec{s}_{\mathbb{V}\setminus K\setminus\Lambda}$, one has the following control in W_2 distance (where $C < \infty$ and $\gamma > 0$ are some fixed constants):

$$\mathrm{W}_2\big(\mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_I \mid \vec{\sigma}_{\partial\Lambda\setminus K} \equiv 0), \mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_I \mid \vec{s}_{\partial\Lambda\setminus K})\big)^2 \leqslant C\mathrm{e}^{-2\gamma\,\mathit{dist}(I,\partial\Lambda)} \sum_{j\in\partial\Lambda\setminus K} |s_j|^2. \tag{WV}$$

But $\mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_I)$ is a mixture of the laws $\mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_I \mid \vec{\sigma}_{\partial\Lambda\setminus K} \equiv \vec{s}_{\partial\Lambda\setminus K})$ (obtained for $\vec{s}_{\partial\Lambda\setminus K}$ following the distribution $\mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_{\partial\Lambda\setminus K})$); so by convexity of optimal transportation cost [14, Theorem 4.8] it follows that

$$\begin{aligned}\mathrm{W}_2\big(\mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_I \mid \vec{\sigma}_{\partial\Lambda\setminus K} \equiv 0), \mathrm{Law}_{\mathrm{cond}}(\vec{\sigma}_I)\big)^2 &\leqslant C\mathrm{e}^{-2\gamma\,\mathit{dist}(I,\partial\Lambda)} \sum_{j\in\partial\Lambda\setminus K} \mathbb{E}^{s_j\sim\mathrm{Law}_{\mathrm{cond}}(\sigma_j)}(|s_j|^2)\\ &= C\mathrm{e}^{-2\gamma\,\mathit{dist}(I,\partial\Lambda)} \sum_{j\in\partial\Lambda\setminus K} \mathbb{E}_{\mathrm{cond}}(|\sigma_j|^2) \leqslant C \sum_{j\in\mathbb{V}\setminus K\setminus\Lambda} \mathrm{e}^{-2\gamma\,\mathit{dist}(I,j)}\, \mathbb{E}_{\mathrm{cond}}(|\sigma_j|^2).\end{aligned} \tag{WW}$$

But, because of the a priori bound on $\mathbb{E}(|\sigma_j|^2)$ given by Corollary 167, we have that

$$\begin{aligned}\int_{\mathcal{S}^K} \Big(\sum_{j\in\mathbb{V}\setminus K\setminus\Lambda} \mathrm{e}^{-2\gamma\,\mathit{dist}(I,j)}\, \mathbb{E}(|\sigma_j^2| \mid \vec{\sigma}_K \equiv \vec{s}_K)\Big)\, \mathrm{d}\mathbb{P}(\vec{s}_K) &= \sum_{j\in\mathbb{V}\setminus K\setminus\Lambda} \mathrm{e}^{-2\gamma\,\mathit{dist}(I,j)}\, \mathbb{E}(|\sigma_j|^2)\\ &\leqslant \sup_{j\in\mathbb{V}} \mathbb{E}(|\sigma_j|^2) \times \sum_{j\in\mathbb{V}} \mathrm{e}^{-2\gamma\,\mathit{dist}(I,j)} < \infty,\end{aligned} \tag{WX}$$

so that, for $\mathrm{Law}(\vec{\sigma}_K)$-almost-all $\vec{s}_K$, the series $\sum_{j\in\mathbb{V}\setminus K\setminus\Lambda} \mathrm{e}^{-2\gamma\,\mathit{dist}(I,j)}\, \mathbb{E}(|\sigma_j^2| \mid \vec{\sigma}_K \equiv \vec{s}_K)$ is converging. Therefore, by letting Λ tend to $\mathbb{V}$, we get that, for almost-all $\vec{s}_K$, $\mathrm{Law}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K)$ is the limit, for the W_2 distance, of a sequence of log-concave probability distributions. One then concludes by Lemma 169 below. ♤

Lemma 169. *The set of log-concave probability measures on $\mathbb{R}^n$ is closed w.r.t. the* W_2 *Wasserstein distance.* ◇

Proof. For μ a probability distribution on $\mathbb{R}^n$, as explained in [42, Theorem 2], the log-concavity of μ is actually equivalent to saying that, for all $x, y \in \mathbb{R}^n, \varepsilon > 0$, one has the following inequality between the mass given to closed balls:

$$\mu\left(\bar{\mathrm{B}}\left(\frac{x+y}{2}, \varepsilon\right)\right) \geqslant \mu\big(\bar{\mathrm{B}}(x,\varepsilon)\big)^{1/2} \mu\big(\bar{\mathrm{B}}(y,\varepsilon)\big)^{1/2}. \tag{WY}$$

Because of that, if μ is not log-concave, one can find $x, y \in \mathbb{R}^n$ and $0 < \varepsilon < \varepsilon' < \infty$ such that: (denoting 'B°' for an *open* ball),

$$\mu\left(\mathrm{B}^\circ\left(\frac{x+y}{2}, \varepsilon'\right)\right) < \mu\big(\bar{\mathrm{B}}(x,\varepsilon)\big)^{1/2} \mu\big(\bar{\mathrm{B}}(y,\varepsilon)\big)^{1/2}. \tag{WZ}$$

But, the set of measures μ satisfying (WZ) is open for the W_2 distance, because the maps of the form $\mu \mapsto \mu\big(\mathrm{B}^\circ(z,r)\big)$ (resp. $\mu \mapsto \mu\big(\bar{\mathrm{B}}(z,r)\big)$) are lower-semicontinuous (resp. upper-semicontinuous) w.r.t. that distance.[**] Thus, since Equation (WZ) is a sufficient condition for failing to be log-concave, the set of measures failing to be log-concave is open for the W_2 distance. This is what we wanted. ♤

Remark 170. Before leaving this subsection, let me observe that Lemma 168 was not actually *needed* for this work: indeed, in § 3.3 log-concavity was used only to ensure L^2 integrability, which could have been got directly from Corollary 167. However, unlike Corollary 167, Lemma 168 had the advantage of being *qualitatively* obvious: which is why I felt it better to use log-concavity arguments in § 3.3. ♧

A.3.3 Conditional laws as time limits of Langevin dynamics

In this third subsection we will turn to another property of the LoNP system that we "expect" to be valid by analogy with the finite-dimensional case, but whose validity needs to be checked in infinite dimension: viz., the fact that the conditional equilibria of the form $\mathbb{P}(\bullet \mid \vec{\sigma}_K \equiv \vec{s}_K)$ can be obtained as the time limit of the associated overdamped Langevin dynamics, starting from zero. (In the case of co-finite $\boldsymbol{K}$ this would nothing but be classical stuff; but here we also want to consider the case where $\mathbb{V} \setminus \boldsymbol{K}$ may be infinite). The result is the following:

Proposition 171 (Convergence to equilibrium for conditioned Langevin dynamics). *Let $\boldsymbol{K} \subseteq \mathbb{V}$ be any set of vertices. For $\vec{s}_K \in \mathcal{S}^K$, say that the boundary condition $\vec{s}_K$ is* non-pathological *when both of the following conditions hold:*

(i) One can meaningfully define the conditional equilibrium probability $\mathbb{P}_{\mathrm{eq}}(\bullet \mid \vec{\sigma}_K \equiv \vec{s}_K)$;

[**] Indeed, for $r < r'$, if μ and μ' are two probability measures such that $\mu\big(\mathrm{B}^\circ(z,r')\big) \leqslant \mu\big(\bar{\mathrm{B}}(z, r)\big) - \varepsilon$, any transportation plan from μ to μ' will have to ensure that a mass at least ε goes from $\bar{\mathrm{B}}(z,r)$ to the complement of $\mathrm{B}^\circ(z,r')$ (which requires to move by at least $r'-r$), so that in such a case one necessarily has $\mathrm{W}_2(\mu,\mu') \geqslant \varepsilon^{1/2}(r'-r)$. From that observation, the claimed semicontinuity properties follow easily.

(ii) The overdamped Langevin dynamics with boundary condition $\vec{s}_K$ on K, when starting from the initial condition $\vec{\sigma}_{\mathbb{V}\setminus K} \equiv 0$, converges in law to that conditional equilibrium when time tends to infinity.[††]

With that vocabulary, the present proposition states that the set of pathological boundary conditions on K has zero measure under the equilibrium distribution $\mathrm{Law}_{\mathrm{eq}}(\vec{\sigma}_K)$ *of the system: in other words,* $\mathrm{Law}_{\mathrm{eq}}(\vec{\sigma}_K)$*-almost-all values of $\vec{s}_K$ are non-pathological, and hence satisfy Items (i) and (ii) above.* ◇

Remark 172. Note that, provided $d > 1$, it would not be very difficult to show that the set of pathological boundary conditions is non-empty, by devising boundary conditions for which there is no convergence for the Langevin dynamics. Morally, this is because the influence of a boundary condition at some vertex k takes some time to propagate to vertex 0, and does so in a diffusive way. Using that observation, the kind of "pathological boundary condition" that I have in mind is the following. Consider $d = 2$ and take $K \coloneqq \{k_0, k_1, k_2, \dots\}$ with $k_n \coloneqq (2^n, 0) \in \mathbb{Z}^2$; consider spins to be real (i.e., "$\ell = 1$") and potentials $V(\bullet)$ and $W(\bullet)$ to be exactly quadratic (and strictly convex); and set $s_{k_n} \coloneqq (-C)^n$ for some very large constant C. (How large exactly C needs to be will depend on the ratio between the convexities of W and V). The point is the following: the boundary condition at spin k_n tends to "contaminate" the other spins (up to some damping effect, which is overcome here by the fact that C was chosen large enough) to make them close to $(-C)^n$. However, that contamination needs some time to take effect: is particular, if you want to see the effect of that boundary condition on spin $(0, 0)$, the time needed before the effect is really felt[‡‡] is of order of magnitude $(2^n)^2 = 4^n$. Therefore, the state of spin $(0, 0)$ will tend to "oscillate" in the following way: when time is of order 4^n, it feels mostly the effect of spin k_n (because the spins $k_{m<n}$ are much lower, while the spins $k_{m>n}$ have not come into effect yet), so that we get (very roughly) $\sigma_{(0,0)}(t = 4^n) \approx (-C)^n$. This oscillating behaviour implies that $\sigma_{(0,0)}(t)$ never converges, hence neither does the whole $\vec{\sigma}_{\mathbb{V}\setminus K}$. ♧

Proof of Proposition 171. First, it is obvious that if we choose our boundary condition $\vec{s}_K$ according to the *actual* distribution $\mathrm{Law}(\vec{\sigma}_K)$, and that we take the initial condition for the Langevin dynamics according to the *actual* conditional distribution $\mathrm{Law}(\vec{\sigma}_{\mathbb{V}\setminus K} \mid \vec{\sigma}_K \equiv \vec{s}_K)$ (which, by the basic properties of conditioning, can be defined for almost-all $\vec{s}_K$), there is convergence to equilibrium: because we are *already* at equilibrium! So, as we want to check that the Langevin dynamics *starting from* $\vec{0}_{\mathbb{V}\setminus K}$ also converges to equilibrium, it will be sufficient to show that the initial condition gets asymptotically forgotten, in the sense that the dynamics gets closer and closer to the dynamics starting from an initial condition drawn according to the actual $\mathrm{Law}(\vec{\sigma}_{\mathbb{V}\setminus K} \mid \vec{\sigma}_K \equiv \vec{s}_K)$.

[††] In the sense that for all finite $I \subseteq \mathbb{V} \setminus K$, the law of $\vec{\sigma}_I(t)$ for that Langevin dynamics converges to the conditional equilibrium law $\mathbb{P}_{\mathrm{eq}}(\vec{\sigma}_I \mid \vec{\sigma}_K \equiv \vec{s}_K)$.

[‡‡] Here note that the effect of spin k_n *shall* attain spin $(0, 0)$, despite the fact that we have fixed the spins k_m for $m < n$, because we are in dimension 2: so, the influence of spin k_n can "bypass" the spins k_m. This would not have worked in dimension 1, where the boundary condition at k_0 would have "screened" all the other ones.

Here again, we do so by *coupling* our two Langevin dynamics, taking the same noise for both. Note however that, contrary to what we did in the proofs of Proposition 120 and Lemma 164, here the *boundary* conditions are the same, but the *initial* conditions differ. This leads us to introduce a different Lyapunov functional: denoting by $\boldsymbol{I}$ a finite set of vertices for which we want to show the convergence of $\mathrm{Law}(\vec{\sigma}_{\boldsymbol{I}}(t))$, our Lyapunov functional will be

$$\mathscr{U}(\vec{\eta}_{\mathbb{V}\smallsetminus K}) \coloneqq \sum_{i\in\mathbb{V}\smallsetminus K} \mathrm{e}^{-2\gamma\, dist(I,i)} |\eta_i(t)|^2 \tag{XA}$$

for some small $\gamma > 0$. So, what matters this time is not the distance to the boundary, but the distance to $\boldsymbol{I}$; and also, the weight of η_i is decreasing[*] as one goes further from $\boldsymbol{I}$.

By Corollary 167 combined with exponential decrease of the weights of the η_i's, one has that

$$\mathbb{E}\big(\mathscr{U}\big(\vec{\eta}(t=0)\big)\big) < \infty. \tag{XB}$$

In Equation (XB), expectation has to be understood in the "annealed" sense: that is, not only you consider the initial condition of the reference Langevin dynamics to be random, but beforehand, you also take the boundary condition $\vec{\sigma}_K$ as random, according to the equilibrium distribution of the unconditioned system. (So, in this case, the "annealed" law of the initial condition of the reference Langevin dynamics is nothing but the unconditioned law of the global system $\vec{\sigma}_{\mathbb{V}}$). The fact that (XB) holds under the annealed law implies that it also holds under almost-all "quenched" laws: that is, for almost-all boundary conditions, when you fix that boundary condition and that only the initial condition of the reference Langevin dynamics remains random, one still has finite expectation for the initial value of the Lyapunov functional.

Now we have to study how the Lyapunov functional evolves along time. We again mimic the proof of Proposition 120, but with a key difference: indeed, since here we are considering identical boundary conditions, the analogue of Equation (PO) just becomes

$$-w_i \eta_i^{\mathsf{T}} \cdot \boldsymbol{W}(\{i,i'\},t) \cdot \eta_i, \tag{XC}$$

which is nonpositive.

Because of the above, the analogue of Equation (PQ) becomes merely

$$\dot{\mathscr{U}}(\vec{\eta}) \leqslant -\big(\alpha - (\mathrm{e}^{2\gamma} - 1) d B\big) \mathscr{U}(\vec{\eta}). \tag{XD}$$

From (XD), you get by Grönwall's lemma that $\mathbb{E}(\mathscr{U}(\vec{\eta}))$ tends to zero, and thus that one has convergence to 0 of the W_2 distance between the actual distribution $\mathrm{Law}(\vec{\sigma}_{\boldsymbol{I}} \mid \vec{\sigma}_K \equiv \vec{s}_K)$ and the distribution that you get at time t for the Langevin dynamics when starting from the initial condition "$\vec{\sigma}_{\mathbb{V}\smallsetminus K} \equiv 0$". This is what we wanted. ♤

---

[*] Actually, on this point the general philosophy remains unchanged: viz., the weight of η_i gets larger towards the "inside", resp. smaller towards the "outside".

A.4 Stability of fast decay properties

All the lemmas in this section will have the same general philosophy: to show that, when applying certain operations to (infinite-dimensional) square matrices indexed by $\mathbb{Z}$ or $\mathbb{Z}^d$, if the entries of the input matrices go quickly enough to zero as you get further from the diagonal, then the output of the operation shall again be a matrix whose coefficients go quickly to zero as you get further from the diagonal.

A.4.1 Stability by inverse convolution

In this subsection, we are not interested in matrices *stricto sensu*, but rather in convolution of functions: however, function convolution may be seen as a special case of matrix multiplication, cf. Remark 183 on page 158. The results that we are going to expose were used in § 3.2.2 to prove mixing between spins in the infinite-range spin-glass model: more specifically, we used these results to control the tails of the inverse convolution $\boldsymbol{\kappa}^{-*}$ of some rapidly decaying function $\boldsymbol{\kappa}\colon \mathbb{Z}^d \to \mathbb{R}$—hence the name of this subsection. In the sequel, what was called '$\boldsymbol{\kappa}$' in the proof of Proposition 104 will correspond to what we call "$\boldsymbol{\delta}_0 - \boldsymbol{\lambda}$" below. Recall that, for $\boldsymbol{f}$ a function of $\ell^1(\mathbb{Z}^d)$ and $k \in \mathbb{N}$, "$\boldsymbol{f}^{*k}$" stands for the k-th convolution power of $\boldsymbol{f}$. (That notation will also be used for $\mathfrak{f} \in \mathrm{L}^1(\mathbb{R}^d)$).

Remark 173. As there will be many different symbols used in this subsection, I have adopted the following convention to make things clearer: real constants and points of $\mathbb{Z}^d$ or $\mathbb{R}^d$ will be referred to by light font symbols; while *functions* will be typeset in bold fonts: either Roman symbols for functions defined on $\mathbb{Z}^d$, or Fraktur symbols for functions defined on $\mathbb{R}^d$. ♧

A few prolegomena are required before starting. First, as we observed in page 90's footnote [**] on functions' decay, the notion of "polynomial (resp. exponential) decay" has to be adapted for functions defined on $\mathbb{Z}^d$ or $\mathbb{R}^d$, in the following way: here, when we say that (for instance) a function $\mathfrak{f}\colon \mathbb{R}^d \to \mathbb{R}$ is γ-polynomially decaying, what we actually mean is that one controls the tails of $\mathfrak{f}$ in a γ-polynomially decaying way: i.e., for the norm $|\bullet|$ on $\mathbb{R}^d$ that one is considering,

$$\sup_{|x|\geqslant\delta} |\mathfrak{f}(x)| \overset{\delta\to\infty}{=} \mathrm{O}(\delta^{-\gamma+\mathrm{o}(1)}). \tag{XE}$$

As regards the norm being used, in this subsection it will be convenient to consider that "$|\bullet|$" actually refers to the *Euclidean* norm on $\mathbb{R}^d$ or $\mathbb{Z}^d$ (the latter being seen as a subset of $\mathbb{R}^d$). Note that the fact of being γ-polynomially decaying does not depend on the choice of the norm being considered; on the other hand, as regards exponential decay, the *quantitative* rate of decay (i.e., the constant "γ" in Definition 83) may depend on the norm, but the *qualitative* property of "being exponentially decaying" does not.

All of that being said, let us first treat the case of exponential decay, which is easier:

Lemma 174. *For $d \in \mathbb{N}^*$, let $\boldsymbol{\lambda}\colon \mathbb{Z}^d \to \mathbb{R}_+$ be an exponentially decaying function; and assume that $\|\boldsymbol{\lambda}\|_{\ell^1(\mathbb{Z}^d)} < 1$. Then, the function $\sum_{k=0}^{\infty} \boldsymbol{\lambda}^{*k}$ (which is the sum of a convergent series in $\ell^1(\mathbb{Z}^d)$) is exponentially decaying too.* ◇

Proof. We start with observing that, actually, "$\boldsymbol{\lambda}$ is exponentially decaying" is equivalent to "the set of linear forms $\langle \varphi, \bullet \rangle$ on $\mathbb{R}^d$ for which $\sum_{v\in\mathbb{Z}^d} \mathrm{e}^{\langle \varphi, v\rangle} |\boldsymbol{\lambda}(v)| < \infty$ contains a neighbourhood of zero (in the space of linear forms on $\mathbb{R}^d$)".[†] Moreover, in the case $\|\boldsymbol{\lambda}\|_{\ell^1(\mathbb{Z}^d)} < 1$, by dominated convergence this is also equivalent to the fact that linear forms close enough to zero satisfy $\sum_{v\in\mathbb{Z}^d} \mathrm{e}^{\langle \varphi, v\rangle} |\boldsymbol{\lambda}(v)| < 1$.

But, provided there is convergence, by multiplicativity of the exponential one has that $\sum_{v\in\mathbb{Z}^d} \mathrm{e}^{\langle \varphi, v\rangle} \boldsymbol{\lambda}^{*k}(v) = \big(\sum_{v\in\mathbb{Z}^d} \mathrm{e}^{\langle \varphi, v\rangle} \boldsymbol{\lambda}(v)\big)^k$. Therefore, if $\boldsymbol{\lambda}$ is exponentially decaying, for $\langle \varphi, \bullet \rangle$ close enough to zero so that $\sum_{v\in\mathbb{Z}^d} \mathrm{e}^{\langle \varphi, v\rangle} |\boldsymbol{\lambda}(v)| < 1$, the values $\sum_{v\in\mathbb{Z}^d} \mathrm{e}^{\langle \varphi, v\rangle} |\boldsymbol{\lambda}^{*k}(v)|$ will decay exponentially in k, and thus they will be summable: so, one will have

$$\sum_{v\in\mathbb{Z}^d} \mathrm{e}^{\langle \varphi, v\rangle} \Big| \sum_{k\in\mathbb{N}} \boldsymbol{\lambda}^{*k}(v) \Big| < \infty, \tag{XF}$$

proving that $\sum_{k\in\mathbb{N}} \boldsymbol{\lambda}^{*k}$ is exponentially decaying. ♤

Now let us turn to the case of polynomial decay, which is much trickier. Before starting, here it will be more convenient to introduce a slightly more restrictive notion of γ-polynomial decay:

Definition 175 (Decay in $\mathrm{O}(|\bullet|^{-\gamma})$)**.** We say that the function $\boldsymbol{f}$ (defined on $\mathbb{Z}^d$ or $\mathbb{R}^d$) *decays in* $\mathrm{O}(|\bullet|^{-\gamma})$ when one has

$$\sup_{|x|\geqslant\delta} |\boldsymbol{f}(x)| \overset{\delta\to\infty}{=} \mathrm{O}(\delta^{-\gamma}). \tag{XG}$$

♡

With that definition, "to be γ-polynomially decaying" is actually synonymous with "to decay in $\mathrm{O}(|\bullet|^{-\tilde{\gamma}})$ for all $\tilde{\gamma} < \gamma$".

With that vocabulary at hand, the second lemma of this subsection is the following:

Lemma 176. *For $d \in \mathbb{N}^*$, let $\boldsymbol{\lambda}\colon \mathbb{Z}^d \to \mathbb{R}_+$ be a function that decays in* $\mathrm{O}(|\bullet|^{-\gamma})$ *for some $\gamma > d$; and assume that $\|\boldsymbol{\lambda}\|_{\ell^1(\mathbb{Z}^d)} < 1$. Then, the function $\sum_{k=0}^{\infty} \boldsymbol{\lambda}^{*k}$ decays in* $\mathrm{O}(|\bullet|^{-\gamma})$ *too (with the same value of γ).* ◇

Remark 177. Given the link between "decaying in $\mathrm{O}(|\bullet|^{-\gamma})$" and "$\gamma$-polynomially decaying" [confer a few lines above], the lemma obviously remains valid if one replaces simultaneously both occurrences of "decays in $\mathrm{O}(|\bullet|^{-\gamma})$" by "is γ-polynomially decaying"; also, by the triangle inequality, the lemma still holds if you allow $\boldsymbol{\lambda}$ to take negative values.

When in the proof of Proposition 104 I referred to "Lemma 176", I actually implied "generalized according to Remark 177". ♧

[†] Too see that, observe that you can bound above the function $\mathrm{e}^{\langle \varphi, \bullet\rangle}$ by $\mathrm{e}^{|\varphi||\bullet|}$, and that conversely you can bound above $\mathrm{e}^{\gamma|\bullet|}$ by a linear combination of the form $\sum_i \mathrm{e}^{\langle \varphi_i, \bullet\rangle}$, with the φ_i's proportional to γ.

Proof of Lemma 176. Let $\boldsymbol{\lambda}$ be a function like in the statement of the lemma. The proof will rely on the following key intermediate result, whose proof is postponed:

Lemma 178. *Let $d \in \mathbb{N}^*$ and $\gamma > d$. Then there exists a function $\boldsymbol{s} \colon \mathbb{Z}^d \to \mathbb{R}_+$ that satisfies all the following properties:*

- *Normalization: $\boldsymbol{s}$ is a probability mass function, i.e. $\sum_{u \in \mathbb{Z}^d} \boldsymbol{s}(u) = 1$.*
- *Decay: $\boldsymbol{s}$ decays in $\mathrm{O}(|\bullet|^{-\gamma})$.*
- *Lower bound: There exists some $c > 0$ such that $\boldsymbol{s}(u) \geqslant c|u|^{-\gamma}$ as soon as $|u|$ is large enough.*
- *Regularity: There exists a nondecreasing function $W \colon \mathbb{R}_+ \to \mathbb{R}_+$, that increases at most polynomially[‡], such that*
$$\forall u, v \in \mathbb{Z}^d \qquad \boldsymbol{s}(v) \leqslant W(|v-u|)\boldsymbol{s}(u). \tag{XH}$$
- *Stability: There exists a function $\boldsymbol{P} \colon \mathbb{N} \to \mathbb{R}_+$, that increases at most polynomially, such that*
$$\forall k \in \mathbb{N} \quad \forall u \in \mathbb{Z}^d \qquad \boldsymbol{s}^{*k}(u) \leqslant \boldsymbol{P}(k)\boldsymbol{s}(u). \tag{XI}$$

Moreover, as regards the lower bound property, one can prescribe c to take any finite value fixed in advance, however large it may be. ◇

Remark 179. Morally, Lemma 178 ensures the existence of some kind of "$(\gamma - d)$-stable probability distribution on $\mathbb{R}^d$"—except that stability only holds in an approximate way. ♧

Thanks to Lemma 178, we are able to introduce functions $\boldsymbol{b}, \boldsymbol{s} \colon \mathbb{Z}^d \to \mathbb{R}_+$ and constants $\mu, \eta \in \mathbb{R}_+$ such that all the following hold (see below for a justification):

- $\boldsymbol{\lambda} \leqslant \mu\boldsymbol{b} + \eta\boldsymbol{s}$ (i.e., one has $\boldsymbol{\lambda}(u) \leqslant \mu\boldsymbol{b}(u) + \eta\boldsymbol{s}(u)$ for all $u \in \mathbb{Z}^d$);
- $\boldsymbol{b}$ and $\boldsymbol{s}$ are probability mass functions;
- $\boldsymbol{b}$ has bounded support;
- $\boldsymbol{s}$ satisfies all the properties stated in Lemma 178;
- $\mu + \eta < 1$.

How can we build such functions? Well, first we take μ and η such that $\mu \geqslant \|\boldsymbol{\lambda}\|_{\ell^1(\mathbb{Z}^d)}$, $\eta > 0$ and $\mu + \eta < 1$: this is possible, as we assumed $\|\boldsymbol{\lambda}\|_{\ell^1} < 1$. Also, by the decay assumption on $\boldsymbol{\lambda}$, there exists a constant $C < \infty$ such that $\boldsymbol{\lambda}(u) \leqslant C|u|^{-\gamma}$ as soon as $|u|$ is large enough: therefore, applying Lemma 178 with imposing 'c' $\leftarrow C / \eta$ for the lower bound property, we obtain a function $\boldsymbol{s}$ such that $\boldsymbol{\lambda}(u) \leqslant \eta\boldsymbol{s}(u)$ as soon as $|u|$ is large enough. Now it just remains to consider the mass function $\tilde{\boldsymbol{b}}$ defined by

---

[‡] I.e., there exists $\kappa < \infty$ such that $W(\delta) = \mathrm{O}(\delta^\kappa)$ as $|\delta| \to \infty$.

$\tilde{\boldsymbol{b}}(u) \coloneqq (\boldsymbol{\lambda}(u) - \eta \boldsymbol{s}(u)) \vee 0$, which has compact support and whose total value is obviously $\leqslant \|\boldsymbol{\lambda}\|_{\ell^1}$, and to normalize it (assuming harmlessly that it is not identically zero) into a probability mass function $\boldsymbol{b}$, for which one will have $\mu \boldsymbol{b}(u) \underset{\text{def}}{=} (\mu / \|\tilde{\boldsymbol{b}}\|_{\ell^1}) \tilde{\boldsymbol{b}}(u) \geqslant (\mu / \|\boldsymbol{\lambda}\|_{\ell^1})(\boldsymbol{\lambda}(u) - \eta \boldsymbol{s}(u)) \geqslant \boldsymbol{\lambda}(u) - \eta \boldsymbol{s}(u)$ indeed.

Now that we have written $\boldsymbol{\lambda}$ as a linear combination of $\boldsymbol{b}$ and $\boldsymbol{s}$, by using the bilinearity and symmetry of the convolution product, we get that, for all $k \in \mathbb{N}$:

$$\boldsymbol{\lambda}^{*k} \leqslant \sum_{k'=0}^{k} \binom{k}{k'} \mu^{k'} \eta^{k-k'} \boldsymbol{b}^{*k'} * \boldsymbol{s}^{*(k-k')} \leqslant \sum_{k'=0}^{k} \binom{k}{k'} \mu^{k'} \eta^{k-k'} P(k-k') \boldsymbol{b}^{*k'} * \boldsymbol{s}. \tag{XJ}$$

Now a key observation is that, denoting by $\boldsymbol{R}$ some finite radius such that $\operatorname{supp} \boldsymbol{b}$ in included in the ball $\mathrm{B}(0, \boldsymbol{R})$ (such a radius does exist indeed), the function $\boldsymbol{b}^{*k'}$ is the density of a probability distribution with support contained in $\mathrm{B}(0, k'\boldsymbol{R})$. But then, given the regularity property of $\boldsymbol{s}$ provided by Lemma 178, for all $u \in \mathbb{Z}^d$ we have the bound

$$\begin{aligned}\big(\boldsymbol{b}^{*k'} * \boldsymbol{s}\big)(u) &= \sum_{h \in \mathbb{Z}^d} \boldsymbol{b}^{*k'}(h) \boldsymbol{s}(u-h) = \sum_{|h| \leqslant k'R} \boldsymbol{b}^{*k'}(h) \boldsymbol{s}(u-h) \\ &\leqslant \underbrace{\Big(\sum_{h \in \mathbb{Z}^d} \boldsymbol{b}^{*k'}(h)\Big)}_{=1} \underbrace{\sup\{\boldsymbol{s}(u') \mid u' \in \mathrm{B}(u, k'\boldsymbol{R})\}}_{\leqslant W(k'R)\boldsymbol{s}(u)} \leqslant W(k'\boldsymbol{R}) \boldsymbol{s}(u),\end{aligned} \tag{XK}$$

whence

$$\boldsymbol{\lambda}^{*k} \leqslant \sum_{k'=0}^{k} \binom{k}{k'} W(k'\boldsymbol{R}) \mu^{k'} P(k-k') \eta^{k-k'} \boldsymbol{s}. \tag{XL}$$

Here we observe that, since $W(\bullet)$ increases at most polynomially, then so does $k' \mapsto W(k'\boldsymbol{R})$. And since any polynomial increases slower than any exponential, it follows that, choosing arbitrary constants $\tilde{\mu} > \mu$, $\tilde{\eta} > \eta$ with $\tilde{\mu} + \tilde{\eta} < 1$, there exists a constant $C < \infty$ such that

$$\forall k, k' \qquad W(k'\boldsymbol{R}) \mu^{k'} \leqslant C \tilde{\mu}^{k'}, \quad P(k-k') \eta^{k-k'} \leqslant C \tilde{\eta}^{k-k'}. \tag{XM}$$

Therefore, combining (`XL`) with (`XM`), the binomial expansion yields that, for all $k \in \mathbb{N}$:

$$\boldsymbol{\lambda}^{*k} \leqslant C^2 (\tilde{\mu} + \tilde{\eta})^k \boldsymbol{s}. \tag{XN}$$

And finally, summing (`XN`) on all values of k, we get that

$$\Big(\sum_{k=0}^{\infty} \boldsymbol{\lambda}^{*k}\Big)(u) \leqslant \frac{C^2}{1 - \tilde{\mu} - \tilde{\eta}} \boldsymbol{s}(u), \tag{XO}$$

which (as a function of u) deceases in $\mathrm{O}(|\bullet|^{-\gamma})$ indeed, because of the decay property of $\boldsymbol{s}$ provided by Lemma 178. This ends the proof of Lemma 176. ♤

It however remains to prove Lemma 178. We will actually see that lemma as a corollary of a variant dealing with functions defined on $\mathbb{R}^d$ (which are easier to handle). The statement of this variant is the following:

Lemma 180. *Let $d \in \mathbb{N}^*$ and $\gamma > d$. Then there exists a function $\mathfrak{s} \colon \mathbb{R}^d \to \mathbb{R}_+$ that satisfies all the following properties:*

- *Normalization: $\mathfrak{s}$ is a probability* density *function, i.e. $\int_{x\in\mathbb{R}^d} \mathfrak{s}(x)\,\mathrm{vol}_d(\mathrm{d}x) = 1$.*
- *Decay: $\mathfrak{s}$ decays in $\mathrm{O}(|\bullet|^{-\gamma})$.*
- *Lower bound: There exists some $\hat{c} > 0$ such that $\mathfrak{s}(x) \geqslant \hat{c}|x|^{-\gamma}$ as soon as $|x|$ is large enough.*
- *Regularity: There exists a nondecreasing function $\hat{W} \colon \mathbb{R}_+ \to \mathbb{R}_+$, that increases at most polynomially, such that*

$$\forall x, y \in \mathbb{R}^d \qquad \mathfrak{s}(y) \leqslant \hat{W}(|y-x|)\mathfrak{s}(x). \tag{XP}$$

- *Stability: There exists a function $\hat{P} \colon \mathbb{N}^* \to \mathbb{R}_+$, that increases at most polynomially, such that*

$$\forall k \in \mathbb{N}^* \quad \forall x \in \mathbb{R}^d \qquad \mathfrak{s}^{*k}(x) \leqslant \hat{P}(k)\mathfrak{s}(x). \tag{XQ}$$

Moreover, as regards the lower bound property, one can prescribe $\hat{c}$ to take any finite value fixed in advance, however large it may be. ◇

Proof. Let $\alpha \coloneqq \gamma - d$, which is > 0 by assumption. We define $\mathfrak{s}$ in the following way: we set

$$\mathfrak{s}(x) \coloneqq (1-\mathrm{e}^{-\alpha}) \times \sum_{n=0}^{\infty} \mathrm{e}^{-n\alpha}\mathfrak{n}(\mathrm{e}^{2n}; x),^{[\S]} \tag{XR}$$

where $\mathfrak{n}(\sigma^2; x)$ is the density, at x, of the d-dimensional multivariate normal distribution with variance matrix $\sigma^2\mathbf{I}_d$:

$$\mathfrak{n}(\sigma^2; x) \coloneqq \frac{1}{(\sqrt{2\pi}\,\sigma)^d}\exp\left(-\frac{|x|^2}{2\sigma^2}\right). \tag{XS}$$

Now, let us check the properties stated by the lemma:

First, the normalization property is a direct consequence of the fact that normal distributions integrate to 1 and that $(1-\mathrm{e}^{-\alpha}) \times \sum_{n=0}^{\infty} \mathrm{e}^{-n\alpha} = 1$.

To get the decay property, let us consider, for all $x \in \mathbb{R}^d \setminus \{0\}$, the following extended sum (which is obviously an upper bound for $\mathfrak{s}(x)\,/\,(1-\mathrm{e}^{-\alpha})$):

$$\sum_{n=-\infty}^{\infty} \mathrm{e}^{-n\alpha}\mathfrak{n}(\mathrm{e}^{2n}; x) =: \bar{\mathfrak{s}}(x). \tag{XT}$$

Obviously $\bar{\mathfrak{s}}(x)$ is of the form $\bar{s}(|x|)$, since all the terms in (XT) only depend on x via $|x|$. Elementary convergence properties show that $\bar{s}(\delta)$ is finite for all $\delta \in (0,\infty)$;

[§] Here I have chosen to take a ratio of e between the scales of two successive normal densities; but we could have replaced e by any constant $r \in (1,\infty)$, up to replacing the coefficient $\mathrm{e}^{-n\alpha}$ by $r^{-n\alpha}$, resp. the normalization constant "$1-\mathrm{e}^{-\alpha}$" by $1-r^{-\alpha}$.

moreover, the function $\bar{s}(\bullet)$ is obviously decreasing (because all the terms in the sum are); and lastly, an immediate index shift on n shows that $\bar{s}(\mathrm{e}\delta) = \mathrm{e}^{-\gamma}\bar{s}(\delta)$ for all δ. Taken together, these three properties imply that $\bar{s}(\bullet)$ decays in $\mathrm{O}(|\bullet|^{-\gamma})$: therefore so does $\bar{\mathfrak{s}}$, and hence $\mathfrak{s}$.

As regards the lower bound property, for $|x| > 1/\mathrm{e}$, take $n \leftarrow \lceil \ln x \rceil$; then one has $\mathrm{e}^n \in [|x|, \mathrm{e}|x|]$, so that

$$\mathfrak{n}(\mathrm{e}^{2n}; x) \geqslant \frac{\mathrm{e}^{-1/2}}{(\sqrt{2\pi}\,\mathrm{e}|x|)^d}; \tag{XU}$$

and thus, by just taking into account the contribution of that value of n, one has:

$$\mathfrak{s}(x) \geqslant (1-\mathrm{e}^{-\alpha}) \times (\mathrm{e}|x|)^{-\alpha} \frac{|x|^{-d}}{\sqrt{2\pi}^d\,\mathrm{e}^{d+1/2}} = \frac{(1-\mathrm{e}^{-\alpha})\mathrm{e}^{-\alpha}}{\sqrt{2\pi}^d\,\mathrm{e}^{d+1/2}}|x|^{-\gamma}, \tag{XV}$$

which is indeed the kind of lower bound that we were looking for.

Now let us check the regularity property. We observe that one has $\mathfrak{n}(\sigma^2; x) \leqslant 1/(\sqrt{2\pi}\,\sigma)^d$ for all x, so that

$$\forall x \in \mathbb{R}^d \qquad \mathfrak{s}(x) \leqslant (1-\mathrm{e}^{-\alpha}) \sum_{n=0}^{\infty} \frac{\mathrm{e}^{-n\alpha}}{(\sqrt{2\pi}\,\mathrm{e}^n)^d} = \frac{1-\mathrm{e}^{-\alpha}}{\sqrt{2\pi}^d\,(1-\mathrm{e}^{-\gamma})} < \infty; \tag{XW}$$

moreover, $\mathfrak{s}(x)$ is obviously a decreasing function of $|x|$. From that, together with the estimates on $\mathfrak{s}$ already established, we deduce that there exist constants $0 < c \leqslant C < \infty$ such that

$$\forall x \in \mathbb{R}^d \qquad c(|x| \wedge 1)^{-\gamma} \leqslant \mathfrak{s}(x) \leqslant C(|x| \wedge 1)^{-\gamma}: \tag{XX}$$

the regularity property is then an easy consequence of (XX), with $\hat{W}(\delta) \leftarrow (C/c)^\gamma(1+\delta)^\gamma$ (the worst case to check corresponding to the situation where $|x| > 1$ and $y = x/|x|$).

Now let us turn to the stability property, which is the most technical part of this proof. Let $k \in \mathbb{N}^*$. We expand $\mathfrak{s}^{*k}$ according to the summation series (XR), using notation '$\vec{n}$' to refer to the k-uple of values $(n_0, \dots, n_{k-1})$ that indexes this expansion. Then, denoting $w(\vec{n}) \coloneqq \sum_{\ell=0}^{k-1} n_\ell$, resp. $\sigma(\vec{n}) \coloneqq \big(\sum_\ell \mathrm{e}^{2n_\ell}\big)^{1/2}$, the convolution properties of normal densities yield that

$$\forall x \in \mathbb{R}^d \qquad \mathfrak{s}^{*k}(x) = (1-\mathrm{e}^{-\alpha})^k \sum_{\vec{n}\in\mathbb{N}^k} \mathrm{e}^{-w(\vec{n})\alpha}\,\mathfrak{n}(\sigma(\vec{n})^2; x). \tag{XY}$$

We will decompose the above sum according to the value of $\max(n_0, \dots, n_{k-1})$ (hereafter shorthanded into "$\max(\vec{n})$"). In this perspective, we start with observing that one has always

$$\mathrm{e}^{\max(\vec{n})} \leqslant \sigma(\vec{n}) \leqslant \sqrt{k}\,\mathrm{e}^{\max(\vec{n})}. \tag{XZ}$$

Also, by covering $\big[1, \sqrt{k}\big]$ with intervals of the form $[\mathrm{e}^{m'-1}, \mathrm{e}^{m'}]$, and using the fact that $\sigma' \in [\sigma/\mathrm{e}, \sigma] \implies \mathfrak{n}((\sigma')^2; x) \leqslant \mathrm{e}^d\,\mathfrak{n}(\sigma^2; x)$ (as a trivial computation shows), we get the following

Lemma 181. *For $k \in \mathbb{N}^*$, denoting $L(k) \coloneqq \left\lceil \frac{1}{2}\ln k \right\rceil \vee 1$, for $m \in \mathbb{N}$, for all $\sigma \in \left[\mathrm{e}^m, \sqrt{k}\,\mathrm{e}^m\right]$, one has*

$$\forall x \in \mathbb{R}^d \qquad \mathfrak{n}(\sigma^2; x) \leqslant \mathrm{e}^d \sum_{m'=m+1}^{m+L(k)} \mathfrak{n}(\mathrm{e}^{2m'}; x). \tag{YA}$$

◇

From (XZ) and Lemma 181, partitioning expansion (XY) according to the value of $\max(\vec{n})$, we get the following bound:

$$\forall x \in \mathbb{R}^d \qquad \mathfrak{s}^{*k}(x) \leqslant (1-\mathrm{e}^{-\alpha})^k \mathrm{e}^d \sum_{m=0}^{\infty} \Bigg(\sum_{\substack{\vec{n}\in\mathbb{N}^k \\ \max(\vec{n})=m}} \mathrm{e}^{-w(\vec{n})\alpha} \times \sum_{m'=m+1}^{m+L(k)} \mathfrak{n}(\mathrm{e}^{2m'}; x) \Bigg). \tag{YB}$$

To bound above $\sum\{\mathrm{e}^{-w(\vec{n})\alpha} \mid \max(\vec{n}) = m\}$, we partition the set $\{\vec{n} \in \mathbb{N}^k \mid \max(\vec{n}) = m\}$ into d subsets, depending on which entry of $(n_0, \dots, n_{k-1})$ the maximum is attained at. As each of these subsets is included in a set of the form $\{m\} \times \mathbb{N}^{k-1}$ (up to permuting indices), we then get (remembering that $w(\vec{n})$ is actually the sum of the entries of $\vec{n}$) that:

$$\begin{aligned} \sum\{\mathrm{e}^{-w(\vec{n})\alpha} \mid \max(\vec{n}) = m\} &\leqslant d \times \sum\Bigg\{ \exp\Bigg(-\Bigg(m + \sum_{\ell=1}^{k-1} n_\ell\Bigg)\alpha\Bigg) \;\Bigg|\; (n_1, \dots, n_{k-1}) \in \mathbb{N}^{k-1} \Bigg\} \\ &= d\mathrm{e}^{-m\alpha} \sum\Bigg\{ \prod_{\ell=1}^{k-1} \mathrm{e}^{-n_\ell\alpha} \;\Bigg|\; \vec{n}_{[\![1,k-1]\!]} \in \mathbb{N}^{k-1} \Bigg\} = d\mathrm{e}^{-m\alpha} \prod_{\ell=1}^{k-1} \sum_{n_\ell\in\mathbb{N}} \mathrm{e}^{-n_\ell\alpha} = \frac{d\mathrm{e}^{-m\alpha}}{(1-\mathrm{e}^{-\alpha})^{k-1}}. \end{aligned} \tag{YC}$$

Using the bound (YC) and grouping indices by m' rather than by m in (YB), we finally get that

$$\mathfrak{s}^{*k}(x) \leqslant d(1-\mathrm{e}^{-\alpha})\mathrm{e}^d \sum_{m'=1}^{\infty} \Bigg(\sum_{m=m'-L(k)}^{m'-1} \mathrm{e}^{-m\alpha} \Bigg) \mathfrak{n}(\mathrm{e}^{2m'}; x). \tag{YD}$$

But one has

$$\sum_{m=m'-L(k)}^{m'-1} \mathrm{e}^{-m\alpha} \leqslant L(k) \times \max_{m=m'-L(k)}^{m'-1} \mathrm{e}^{-m\alpha} = L(k)\mathrm{e}^{L(k)\alpha}\mathrm{e}^{-m'\alpha} \leqslant \mathrm{e}^{\alpha}k^{\alpha/2}L(k)\mathrm{e}^{-m'\alpha} \tag{YE}$$

(where the last inequality comes from the fact that $L(k) \underset{\text{def}}{\leqslant} \frac{1}{2}\ln k + 1$), so that in the end:

$$\mathfrak{s}^{*k}(x) \leqslant d\mathrm{e}^d\mathrm{e}^{\alpha}k^{\alpha/2}L(k)(1-\mathrm{e}^{-\alpha}) \sum_{m'=1}^{\infty} \mathrm{e}^{-m'\alpha}\mathfrak{n}(\mathrm{e}^{2m'}; x) \underset{\text{def}}{\leqslant} d\mathrm{e}^{\gamma}k^{\alpha/2}L(k)\mathfrak{s}(x): \tag{YF}$$

and since $k^{\alpha/2}L(k)$ grows (not faster than) polynomially, we have indeed obtained a bound of the wanted type for the stability property.

The final step of the proof is to check that the property that "$\mathfrak{s}(x) \geqslant \hat{c}|x|^{-\gamma}$ for $|x|$ large enough" can actually be imposed for any constant $\hat{c}$ fixed in advance. To do so, we start from the density (`XR`) that we have been building above, for which we will use indices '1'; and we proceed to a scale change: denoting, for $\rho \in \mathbb{R}_+^*$, the density

$$\mathfrak{s}_\rho(x) \coloneqq \rho^{-d}\mathfrak{s}_1(x \,/\, \rho), \tag{YG}$$

it is immediate to deduce from the properties of $\mathfrak{s}_1$ that the function $\mathfrak{s}_\rho$ also satisfies the properties of the lemma, with $\hat{P}_\rho(\bullet) \leftarrow \hat{P}_1(\bullet)$, $\hat{W}_\rho(\bullet) \leftarrow \hat{W}_1(\bullet/\rho)$ and $\hat{c}_\rho \leftarrow \rho^\alpha \hat{c}_1$: so, it suffices to take ρ large enough to get the lemma with $\hat{c}$ as large as one wishes. ♤

Now that we have Lemma 180, let us show how Lemma 178 can be obtained from it:

Proof of Lemma 178. Take $\mathfrak{s} \colon \mathbb{R}^d \to \mathbb{R}_+$ as given by Lemma 180. For $\boldsymbol{u} \in \mathbb{Z}^d$, let us define $\boldsymbol{s}(\boldsymbol{u})$ as the average of $\mathfrak{s}$ on the cube $\boldsymbol{u} + [-1/2, 1/2]^d$. In the sequel we will denote by $\varDelta$ the maximum value of $|\boldsymbol{h}|$ for $\boldsymbol{h} \in [-1/2, 1/2]^d$—that is, $\varDelta \underset{\text{def}}{=} \frac{1}{2}\sqrt{d}$.

Let us check that the properties of $\mathfrak{s}$ are indeed inherited by $\boldsymbol{s}$.

The normalization property comes from the fact that the cubes $\boldsymbol{u}+[-1/2, 1/2]^d$ for $\boldsymbol{u} \in \mathbb{Z}^d$ make a partition of $\mathbb{R}^d$ (up to a zero-measure set) and all have unit volume. The regularity property holds with $\boldsymbol{W}(\bullet) \leftarrow \hat{W}(\bullet)$, as one has

$$\begin{aligned}\boldsymbol{s}(v) \underset{\text{def}}{=} \int_{h\in[-1/2,1/2]^d} \mathfrak{s}(v+h)\,\mathrm{vol}_d(\mathrm{d}h) &\leqslant \int_{[-1/2,1/2]^d} \hat{W}\big((v+h)-(u+h)\big)\mathfrak{s}(u+h)\,\mathrm{vol}_d(\mathrm{d}h)\\ &= \hat{W}(v-u)\int_{[-1/2,1/2]^d} \mathfrak{s}(u+h)\,\mathrm{vol}_d(\mathrm{d}h) \underset{\text{def}}{=} \hat{W}(v-u)\boldsymbol{s}(u).\end{aligned} \tag{YH}$$

As regards the "decay" and "lower bound" properties, by the regularity property of $\mathfrak{s}$, one has $\mathfrak{s}(\boldsymbol{u}+\boldsymbol{h}) \in [\mathfrak{s}(\boldsymbol{u}) \,/\, \hat{W}(\varDelta), \mathfrak{s}(\boldsymbol{u})\hat{W}(\varDelta)]$ for all $\boldsymbol{h} \in [-1/2, 1/2]^d$: therefore, averaging yields that $\boldsymbol{s}(\boldsymbol{u}) \in [\mathfrak{s}(\boldsymbol{u}) \,/\, \hat{W}(\varDelta), \mathfrak{s}(\boldsymbol{u})\hat{W}(\varDelta)]$ for all $\boldsymbol{u} \in \mathbb{Z}^d$, so that the decay and lower bound properties of $\mathfrak{s}$ get readily inherited by $\boldsymbol{s}$; moreover, as regards the lower bound property, one can take 'c' $\leftarrow \hat{c} \,/\, \hat{W}(\varDelta)$, which can be made arbitrarily large as $\hat{c}$ can.

The stability property is slightly more tricky. The case $k \leftarrow 0$ (which we did not consider in the continuous case) does not cause any difficulty: one just has to observe that $\boldsymbol{s}(0) > 0$, which is immediate. Now let us consider $k \in \mathbb{N}^*$. We introduce k i.i.d. $\mathbb{R}^d$-valued random variables $\boldsymbol{X}_0, \dots, \boldsymbol{X}_{k-1}$, each having density $\mathfrak{s}$; and we denote by $\boldsymbol{Y}_\ell$ the closest point to $\boldsymbol{X}_\ell$ in $\mathbb{Z}^d$.[¶] Then, $\boldsymbol{s}^{*k}(\boldsymbol{u})$ is the probability that the sum of the $\boldsymbol{Y}_\ell$'s is equal to $\boldsymbol{u}$. But one has always $|\boldsymbol{X}_\ell - \boldsymbol{Y}_\ell| \leqslant \varDelta$, so that the event $\{\sum_\ell \boldsymbol{Y}_\ell = \boldsymbol{u}\}$ is included in the event $\{\sum_\ell \boldsymbol{X}_\ell \in \mathrm{B}(\boldsymbol{u}, k\varDelta)\}$. And the probability of the latter event is bounded above by

$$\begin{aligned}&\mathrm{vol}_d\big(\mathrm{B}(u, k\varDelta)\big) \times \sup\{\mathfrak{s}^{*k}(x) \mid x \in \mathrm{B}(u, k\varDelta)\}\\ &\leqslant (k\varDelta)^d\,\mathrm{vol}_d\big(\mathrm{B}(0,1)\big) \times \hat{P}(k) \sup_{x\in\mathrm{B}(u,k\varDelta)} \mathfrak{s}(x) \leqslant \mathrm{vol}_d\big(\mathrm{B}(0,1)\big)(k\varDelta)^d\hat{P}(k) \times \hat{W}(k\varDelta)\mathfrak{s}(u)\\ &\qquad\leqslant \mathrm{vol}_d\big(\mathrm{B}(0,1)\big)(k\varDelta)^d\hat{P}(k)\hat{W}(k\varDelta)\hat{W}(\varDelta) \times \boldsymbol{s}(u),^{[\|]}\end{aligned} \tag{YI}$$

[¶] Since Law($\boldsymbol{X}_\ell$) has a density, almost-surely there will be exactly one closest point to $\boldsymbol{X}_\ell$ in $\mathbb{Z}^d$.

where, given the properties of $\hat{\boldsymbol{W}}(\bullet)$ and $\hat{\boldsymbol{P}}(\bullet)$, the factor in front of $\boldsymbol{s}(u)$ increases (at most) polynomially in k. This shows that the stability property remains satisfied by $\boldsymbol{s}$ indeed. ♤

A.4.2 Stability by matrix operations

In this subsection we will deal with square matrices indexed by $\mathbb{Z}$ (or by $\mathbb{Z}\times 2$ or $\mathbb{Z}\times 4$): in this context, the relevant notion of "being exponentially decaying" has been given by Definition 128. In the sequel, we will only deal with the case of matrices indexed by $\mathbb{Z}$: indeed, one may identify $\mathbb{Z}\times 2$ or $\mathbb{Z}\times 4$ with $\mathbb{Z}$ via the ordering "$\ldots < \mathrm{p}_{-1} < \mathrm{q}_{-1} < \mathrm{p}'_{-1} < \mathrm{q}'_{-1} < \mathrm{p}_0 < \mathrm{q}_0 < \mathrm{p}'_0 < \mathrm{q}'_0 < \ldots$", and that identification does not impact the exponential decay property.

As Toeplitz matrices will play an import role in this subsection, let us introduce some notation for them:

Notation 182. For $f \in \ell^1(\mathbb{Z})$ a function, we denote by $\boldsymbol{T}_f$ the $\mathbb{Z}$-indexed square matrix $(\!(\mathbf{t}^{(f)}_{xy})\!)_{x,y}$ defined by

$$\mathbf{t}^{(f)}_{xy} \coloneqq f(y-x). \tag{YJ}$$

♡

Remark 183. It is a direct observation that, in the Toeplitz case, the product rule for matrices corresponds to the convolution product of $\ell^1(\mathbb{Z})$ functions:

$$\boldsymbol{T}_f \boldsymbol{T}_g = \boldsymbol{T}_{f*g}. \tag{YK}$$

In view of that, the inverse convolution properties studied in § A.4.1 and the matrix operations of the current subsection appear as two flavours of the same general idea: which explains why I have grouped these into the same section. ♧

Now let us turn to theorems. We start with a general stability result for exponentially decaying matrices:

Lemma 184. *If $\boldsymbol{M}$ and $\boldsymbol{M}'$ are exponentially decaying square matrices (indexed by $\mathbb{Z}$), then $\boldsymbol{M}+\boldsymbol{M}'$, $\boldsymbol{M}'\boldsymbol{M}$ and $\exp(\boldsymbol{M})$ are exponentially decaying too.* ◇

Proof. For $\gamma > 0$, denote by $e_\gamma(\bullet)$ the function $z \mapsto \mathrm{e}^{-\gamma|z|}$, which belongs to $\ell^1(\mathbb{Z})$. A key observation is that, for $0 < \gamma < \gamma'$, there exists a constant $K(\gamma,\gamma') < \infty$ such that

$$\forall z \in \mathbb{Z} \qquad \big(e_{\gamma'} * e_\gamma\big)(z) \leqslant K(\gamma,\gamma') e_\gamma(z): \tag{YL}$$

this can be checked by direct computation, writing that $\big(e_{\gamma'} * e_\gamma\big)(z)$ is equal to $\sum_{z'\in\mathbb{Z}} \mathrm{e}^{-\gamma'|z'|}\mathrm{e}^{-\gamma|z-z'|}$, and decomposing that sum according to the respective signs of z' and $z-z'$ to obtain that

$$\big(e_\gamma * e_{\gamma'}\big)(z) = \left(\frac{1-\mathrm{e}^{-(\gamma'-\gamma)|z|}}{1-\mathrm{e}^{-(\gamma'-\gamma)}} + \frac{\mathrm{e}^{-(\gamma+\gamma')}+\mathrm{e}^{-(\gamma'-\gamma)|z|}}{1-\mathrm{e}^{-(\gamma+\gamma')}}\right)\mathrm{e}^{-\gamma|z|}, \tag{YM}$$

[|||] The last inequality comes from the fact that one always has $\boldsymbol{s}(u) \geqslant \mathfrak{s}(u)/\hat{W}(\varDelta)$, as was established a few lines above.

where the parenthesis in factor of $\mathrm{e}^{-\gamma|z|}$ is bounded uniformly by $1/(1-\mathrm{e}^{-(\gamma'-\gamma)}) + (1+\mathrm{e}^{-(\gamma+\gamma')})\,/\,(1-\mathrm{e}^{-(\gamma+\gamma')})$.

Equation (`YL`) can then be translated in terms of Toeplitz matrices, confer Remark 183: denoting $\boldsymbol{T}_{e_\gamma} =: \boldsymbol{E}_\gamma$, one gets the entry-wise comparison

$$\boldsymbol{E}_{\gamma'}\boldsymbol{E}_\gamma \leqslant K(\gamma,\gamma')\boldsymbol{E}_\gamma. \tag{YN}$$

Those prolegomena being set, let us turn to proving the lemma itself. Exponential decay of $\boldsymbol{M}+\boldsymbol{M}'$ is trivial. As regards multiplication, denote by resp. $\mathfrak{m}_{xy}$ and $\mathfrak{m}'_{xy}$ the entries of resp. $\boldsymbol{M}$ and $\boldsymbol{M}'$. The exponential decay assumption implies that there exist constants $\boldsymbol{B}, \boldsymbol{B}' < \infty$ and $\gamma, \gamma' > 0$ such that

$$\forall x,y \in \mathbb{Z} \qquad |\mathfrak{m}_{xy}| \leqslant \boldsymbol{B}\mathrm{e}^{-\gamma|y-x|}, |\mathfrak{m}'_{xy}| \leqslant \boldsymbol{B}'\mathrm{e}^{-\gamma'|y-x|}. \tag{YO}$$

Up to worsening the constant γ, we may assume that $\gamma < \gamma'$. Then, denoting by $\mathfrak{p}_{xy}$ the entries of $\boldsymbol{M}'\boldsymbol{M}$, one has

$$|\mathfrak{p}_{xy}| = \Big|\sum_{z\in\mathbb{Z}} \mathfrak{m}'_{xz}\mathfrak{m}_{zy}\Big| \leqslant \sum_{z\in\mathbb{Z}} |\mathfrak{m}'_{xz}||\mathfrak{m}_{zy}| \leqslant \boldsymbol{B}'\boldsymbol{B} \sum_{z\in\mathbb{Z}} \mathrm{e}^{-(\gamma'|z-x|+\gamma|y-z|)}: \tag{YP}$$

but $\sum_{z\in\mathbb{Z}} \mathrm{e}^{-(\gamma'|z-x|+\gamma|y-z|)}$ is nothing but the (x,y) entry of $\boldsymbol{E}_{\gamma'}\boldsymbol{E}_\gamma$, which (`YN`) bounds above by $\boldsymbol{K}(\gamma,\gamma')\mathrm{e}^{-\gamma|y-x|}$: thus, $|\mathfrak{p}_{xy}|$ is bounded above by $\boldsymbol{B}'\boldsymbol{B}\boldsymbol{K}(\gamma,\gamma')\mathrm{e}^{-\gamma|y-x|}$, which is indeed an exponentially decaying function of $|y-x|$; so $\boldsymbol{M}'\boldsymbol{M}$ is exponentially decaying.

Now let us turn to controlling $\exp(\boldsymbol{M})$. To do that, we will use iteratively the control that we have been showing on matrix products. Let us fix arbitrarily $\tilde{\gamma} \in (0,\gamma)$. Applying (`YP`) with the role of '$\boldsymbol{M}'$' played by $\boldsymbol{M}$, we get that the entries of $\boldsymbol{M}^2$ are bounded (in absolute value) by $\boldsymbol{B}^2\boldsymbol{K}(\tilde{\gamma},\gamma)\mathrm{e}^{-\tilde{\gamma}|y-x|}$. Then, applying again (`YP`), but this time with '$\boldsymbol{M}$' $\leftarrow \boldsymbol{M}^2$ and '$\boldsymbol{M}'$' $\leftarrow \boldsymbol{M}$, we get that the entries of $\boldsymbol{M}^3$ are bounded by $\boldsymbol{B}^3\boldsymbol{K}(\tilde{\gamma},\gamma)^2\mathrm{e}^{-\tilde{\gamma}|y-x|}$. And more generally, iterating again and again, we find that for all $k \in \mathbb{N}^*$, denoting by $\mathfrak{p}^{(k)}_{xy}$ the entries of $\boldsymbol{M}^k$, one has

$$|\mathfrak{p}^{(k)}_{xy}| \leqslant \boldsymbol{B}^k\boldsymbol{K}(\tilde{\gamma},\gamma)^{k-1}\mathrm{e}^{-\tilde{\gamma}|y-x|}. \tag{YQ}$$

This allows us, thanks to the series expansion "$\exp(\boldsymbol{M}) = \sum_{k\in\mathbb{N}} \frac{1}{k!}\boldsymbol{M}^k$", to bound the (x,y) entry of $\exp(\boldsymbol{M})$ by

$$\mathbf{1}_{x=y} + \left(\boldsymbol{B}\sum_{k=1}^{\infty} \frac{\big(\boldsymbol{B}\boldsymbol{K}(\tilde{\gamma},\gamma)\big)^{k-1}}{k!}\right)\mathrm{e}^{-\tilde{\gamma}|y-x|}, \tag{YR}$$

in which the sum is converging (to $(\mathrm{e}^{\boldsymbol{B}\boldsymbol{K}(\tilde{\gamma},\gamma)}-1)\,/\,\boldsymbol{B}\boldsymbol{K}(\tilde{\gamma},\gamma)$): so, the wanted control is achieved, as (`YR`) is indeed an exponentially decaying function of $|y-x|$! ♤

The second result of this subsection deals with matrix inversion:

Lemma 185. *Let* $\boldsymbol{I} \subseteq \mathbb{Z}$ *and let* $\boldsymbol{M} =: (\!(\mathfrak{m}_{xy})\!)_{x,y\in I}$ *be a symmetric square matrix. Assume that, when seen as a quadratic form on* $\mathrm{L}^2(\boldsymbol{I})$*, one has* $\boldsymbol{M} \geqslant \alpha\mathbf{I}_I$ *for some* $\alpha > 0$*, and that there exist constants* $\boldsymbol{B} < \infty$ *and* $\gamma > 0$ *such that*

$$\forall x, y \in \boldsymbol{I} \qquad |\mathfrak{m}_{xy}| \leqslant \boldsymbol{B}\mathrm{e}^{-\gamma|y-x|}. \tag{YS}$$

Then $\boldsymbol{M}$ *is invertible, and there exist constants* $\boldsymbol{B}' < \infty$ *and* $\gamma' > 0$*, which are explicit functions of* $\alpha, \gamma, \boldsymbol{B}$ *(in particular, they do* <u>not</u> *depend on* $\boldsymbol{I}$*), such that, denoting* $\boldsymbol{M}^{-1} =: (\!(\mathfrak{n}_{xy})\!)_{x,y\in I}$*, one has the following control on the entries of* $\boldsymbol{M}^{-1}$*:*

$$\forall x, y \in \boldsymbol{I} \qquad |\mathfrak{n}_{xy}| \leqslant \boldsymbol{B}'\mathrm{e}^{-\gamma'|y-x|}. \tag{YT}$$

◇

Proof. Before all, let us observe that the exponential decay assumption on the matrix $\boldsymbol{M}$ implies that it is bounded: indeed, denoting by $\boldsymbol{M}_{[z]}$ the matrix with general entry $\mathbf{1}_{y-x=z}\mathfrak{m}_{xy}$, one has $|||\boldsymbol{M}_{[z]}||| \leqslant \sup\{|\mathfrak{m}_{xy}| \mid y - x = z\} \leqslant \boldsymbol{B}\mathrm{e}^{-\gamma|z|}$, whence $|||\boldsymbol{M}||| \leqslant \sum_{z\in\mathbb{Z}} \boldsymbol{B}\mathrm{e}^{-\gamma|z|} = \boldsymbol{B}\coth(\gamma / 2)$. As a consequence, up to multiplying $\boldsymbol{B}$ by $\coth(\gamma / 2)$, we can assume that $|||\boldsymbol{M}||| \leqslant \boldsymbol{B}$. Also, as $\boldsymbol{M}$ is the matrix of a coercive continuous bilinear form, the Lax-Milgram theorem ensures that it is invertible indeed.

Now let us turn to bounding the $\mathfrak{n}_{xy}$'s. First, up to multiplying $\boldsymbol{M}$ by a scalar, we can alleviate computations by assuming that $\boldsymbol{B} = 1$. Next, introducing $\boldsymbol{H} := \mathbf{I}_I - \boldsymbol{M}$, one has the following converging series expansion for $\boldsymbol{M}^{-1}$:

$$\boldsymbol{M}^{-1} = \sum_{k=0}^{\infty} \boldsymbol{H}^k. \tag{YU}$$

The fact that the above series is converging comes from the fact that $|||\boldsymbol{H}||| \leqslant 1-\alpha < 1$: indeed, since $\boldsymbol{M}$ is semidefinite positive symmetric with $|||\boldsymbol{M}||| \leqslant 1$, one has $\boldsymbol{M} \leqslant \mathbf{I}_I$ (in the sense of quadratic forms) by the spectral theorem, so that (as one also has $\boldsymbol{M} \geqslant \alpha\mathbf{I}_I$ by assumption) $\mathbf{0}_I \leqslant \boldsymbol{H} \leqslant (1-\alpha)\mathbf{I}_I$, which implies in turn that $|||\boldsymbol{H}||| \leqslant 1-\alpha$ by the spectral theorem again.

Now, up to replacing $\boldsymbol{B}$ by $\boldsymbol{B} + 1$, we have the same entry-wise control on $\boldsymbol{H}$ as on $\boldsymbol{M}$. Then, exactly in the same way as we derived (YQ), we deduce the following entry-wise control on $\boldsymbol{H}^k$ (where '$\mathfrak{h}^{(k)}_{xy}$' refers to the general entry of $\boldsymbol{H}^k$, $\tilde{\gamma}$ is an arbitrary value in $(0, \gamma)$, and we assume harmlessly that $\boldsymbol{K}(\tilde{\gamma}, \gamma) \geqslant 1$):

$$\forall x, y \in \boldsymbol{I} \qquad |\mathfrak{h}^{(k)}_{xy}| \leqslant \boldsymbol{B}^k\boldsymbol{K}(\tilde{\gamma}, \gamma)^{k-1}\mathrm{e}^{-\tilde{\gamma}|y-x|} \leqslant \big(\boldsymbol{B}\boldsymbol{K}(\tilde{\gamma}, \gamma)\big)^k\mathrm{e}^{-\tilde{\gamma}|y-x|}. \tag{YV}$$

(Here note that, as planned, the set $\boldsymbol{I}$ does not appear in the expression of the constants).

However, Equation (YV) is not enough yet to get the desired entry-wise control on $\boldsymbol{M}^{-1}$, as $\boldsymbol{B}\boldsymbol{K}(\tilde{\gamma}, \gamma) =: \tilde{\boldsymbol{B}}$ is larger than 1. But now, observe that the bound $|||\boldsymbol{H}^k||| \leqslant$

$(1-\alpha)^k$ implies that all the $\mathfrak{h}_{xy}^{(k)}$'s are bounded by $(1-\alpha)^k$; thus:

$$
\begin{aligned}
|\mathfrak{n}_{xy}| &\leqslant \sum_{k=0}^{\infty} \bigl(\mathrm{e}^{-\tilde{\gamma}|y-x|}\tilde{\boldsymbol{B}}^k \wedge (1-\alpha)^k\bigr) \\
&\leqslant \sum \{\mathrm{e}^{-\tilde{\gamma}|y-x|}\tilde{\boldsymbol{B}}^k \mid \mathrm{e}^{-\tilde{\gamma}|y-x|}\tilde{\boldsymbol{B}}^k \leqslant (1-\alpha)^k\} + \sum \{(1-\alpha)^k \mid (1-\alpha)^k < \mathrm{e}^{-\tilde{\gamma}|y-x|}\tilde{\boldsymbol{B}}^k\} \\
\leqslant \sum_{k'=-\infty}^{0} &\tilde{\boldsymbol{B}}^{k'} \times \max\{\mathrm{e}^{-\tilde{\gamma}|y-x|}\tilde{\boldsymbol{B}}^k \mid \leqslant (1-\alpha)^k\} + \sum_{k'=0}^{\infty} (1-\alpha)^{k'} \times \max\{(1-\alpha)^k \mid < \mathrm{e}^{-\tilde{\gamma}|y-x|}\} \\
&\leqslant \left(\frac{\tilde{\boldsymbol{B}}}{\tilde{\boldsymbol{B}}-1} + \frac{1}{\alpha}\right) \sup_{k\in\mathbb{R}} \bigl(\mathrm{e}^{-\tilde{\gamma}|y-x|}\tilde{\boldsymbol{B}}^k \wedge (1-\alpha)^k\bigr) = \bigl(\tfrac{\tilde{\boldsymbol{B}}}{\tilde{\boldsymbol{B}}-1} + \tfrac{1}{\alpha}\bigr) \exp\bigl(-\tfrac{|\ln(1-\alpha)|\tilde{\gamma}}{|\ln(1-\alpha)|+\ln\tilde{\boldsymbol{B}}}|y-x|\bigr),
\end{aligned}
\tag{YW}
$$

from which you read suitable values for $\boldsymbol{B}'$ and γ'. ♤

Remark 186. Actually Lemma 185 remains valid even if you remove the assumption that $\boldsymbol{M}$ is symmetric. In that case, one uses the fact that, if $\boldsymbol{M}$ satisfies $\alpha\mathbf{I}_I \leqslant (\boldsymbol{M}+\boldsymbol{M}^\top)\,/\,2$ (in the sense of quadratic forms) and $|||\boldsymbol{M}||| \leqslant \boldsymbol{B}$, then, for $\varepsilon \in (0, 2\alpha\,/\,\boldsymbol{B}^2)$, the matrix $\mathbf{I}_I - \varepsilon\boldsymbol{M}$ is strictly contracting (as an endomorphism in L^2). From that observation, you can get an expansion of $\boldsymbol{M}^{-1}$ as a sum of the $(\mathbf{I}_I - \varepsilon\boldsymbol{M})^k$'s; and then everything works the same as above—just at the price of slightly more complicated formulas. ♧

Bibliography

[1] Richard C. Bradley. Basic properties of strong mixing conditions. In *Dependence in probability and statistics (Oberwolfach, 1985)*, volume 11 of *Progr. Probab. Statist.*, pages 165–192. Birkhäuser, Boston, 1986.

[2] Andrey N. Kolmogorov and Jurij A. Rozanov. On a strong mixing condition for stationary Gaussian processes. *Teor. Verojatnost. i Primenen.*, 5:222–227, 1960.

[3] Geoffrey Grimmett. *The random-cluster model*, volume 333 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, 2006.

[4] Rudolf Peierls. On Ising's model of ferromagnetism. *Proc. Cambridge Philos. Soc.*, 32:477–481, 1936.

[5] Yasunari Higuchi. Coexistence of infinite $(*)$-clusters. II. Ising percolation in two dimensions. *Probab. Theory Related Fields*, 97(1-2):1–33, 1993.

[6] Péter Csáki and János Fischer. On the general notion of maximal correlation. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8:27–51, 1963.

[7] Roger A. Horn and Charles R. Johnson. *Matrix analysis.* Cambridge University Press, 1990. Corrected reprint of the 1985 original.

[8] Richard C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144 (electronic), 2005. Update of the 1986 original.

[9] Roland L. Dobrushin and Senya B. Shlosman. Completely analytical Gibbs fields. In J. Fritz, A. Jaffe, and D. Szász, editors, *Statistical Physics and Dynamical Systems: Rigorous Results*, pages 371–403. Birkhäuser, Boston, MA, 1985.

[10] Senya B. Shlosman. Uniqueness and half-space nonuniqueness of Gibbs states in Czech models. *Theor. Math. Phys.*, 66(3), 1986.

[11] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, 1999.

[12] Richard C. Bradley. *Introduction to strong mixing conditions*, volumes 1–3. Kendrick Press, Heber City, UT, 2007.

[13] Roy J. Glauber. Time-dependent statistics of the Ising model. *J. Math. Phys.*, 4:294–307, 1963.

[14] Cédric Villani. *Optimal transport. Old and new*, volume 338 of *Grundlehren Math. Wiss.* Springer, Berlin, 2009.

[15] Mark Y. An. Log-concave probability distributions: Theory and statistical testing. Working paper 95-03, Department of Economics, Duke University, 1995.

[16] Alexander G. Basuev. Ising model in half-space: A series of phase transitions in low magnetic fields. *Theor. Math. Phys.*, 153(2):1539–1574, 2007.

[17] Cédric Villani. *Hypocoercivity*, volume 950 of *Mem. Am. Math. Soc.* American Mathematical Society (AMS), Providence, RI, 2009.

[18] Grigorios A. Pavliotis. *Stochastic processes and applications. Diffusion processes, the Fokker–Planck and Langevin equations*, volume 60 of *Texts Appl. Math.* Springer, New York, NY, 2014.

[19] Ildar A. Ibragimov. A note on the central limit theorems for dependent random variables. *Theory of Probability and its Applications*, 20(1):135–141, 1975.

[20] Magda Peligrad. On the central limit theorem for ρ-mixing sequences of random variables. *Ann. Probab.*, 15(4):1387–1394, 1987.

[21] Richard C. Bradley. On the spectral density and asymptotic normality of weakly dependent random fields. *J. Theoret. Probab.*, 5(2):355–373, 1992.

[22] Qiman Shao. An invariance principle for stationary ρ-mixing sequences with infinite variance. *Chin. Ann. Math., Ser. B*, 14(1):27–42, 1993.

[23] Roland L. Dobrušin. Markov processes with a large number of locally interacting components: The invertible case and certain generalizations. *Problemy Peredači Informacii*, 7(3):57–66, 1971.

[24] Thomas M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, 1985.

[25] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren Math. Wiss.* Springer, Cham, 2014.

[26] Daniel W. Stroock and Boguslaw Zegarlinski. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Commun. Math. Phys.*, 149(1):175–193, 1992.

[27] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I: The attractive case. *Commun. Math. Phys.*, 161(3):447–486, 1994.

[28] Anne-Séverine Boudou, Pietro Caputo, Paolo Dai Pra, and Gustavo Posta. Spectral gap estimates for interacting particle systems via a Bochner-type identity. *J. Funct. Anal.*, 232(1):222–258, 2006.

[29] Michael Aizenman, David J. Barsky, and Roberto Fernández. The phase transition in a general class of Ising-type models is sharp. *Journal of Statistical Physics*, 47:343–374, 1987.

[30] Roberto H. Schonmann and Senya B. Shlosman. Complete analyticity for 2D Ising completed. *Communications in Mathematical Physics*, 170(2):453–482, 1995.

[31] Kenneth S. Alexander. Mixing properties and exponential decay for lattice systems in finite volumes. *Ann. Probab.*, 32(1A):441–487, 2004.

[32] Raimundo Briceño. The topological strong spatial mixing property and new conditions for pressure approximation. *Ergodic Theory Dyn. Syst.*, 38(5):1658–1696, 2016.

[33] Sébastien Ott. Weak mixing and analyticity of the pressure in the Ising model. *Commun. Math. Phys.*, 377(1):675–696, 2020.

[34] Michael J. Buckingham and James D. Gunton. Correlations at the critical point of the Ising model. *Phys. Rev.*, 178(2):848–853, 1969.

[35] Tsing-Dao Lee and Chen-Ning Yang. Statistical theory of equations of state and phase transitions. II: Lattice gas and Ising model. *Phys. Rev., II. Ser.*, 87:410–419, 1952.

[36] F. Martinelli, E. Olivieri, and Roberto H. Schonmann. For 2-D lattice spin systems weak mixing implies strong mixing. *Commun. Math. Phys.*, 165(1):33–47, 1994.

[37] Thierry Bodineau, Isabelle Gallagher, Laure Saint-Raymond, and Sergio Simonella. Dynamics of dilute gases: A statistical approach. In *International congress of mathematicians 2022, ICM 2022, Helsinki, Finland, virtual, July 6–14, 2022. Volume 2. Plenary lectures*, pages 750–795. European Mathematical Society (EMS), Berlin, 2023.

[38] Fraydoun Rezakhanlou. Boltzmann–Grad limits for stochastic hard sphere models. *Commun. Math. Phys.*, 248(3):553–637, 2004.

[39] Giada Basile, Cédric Bernardin, Milton Jara, Tomasz Komorowski, and Stefano Olla. Thermal conductivity in harmonic lattices with random collisions. In Stefano Lepri, editor, *Thermal Transport in Low Dimensions: From Statistical Physics to Nanoscale Heat Transfer*, pages 215–237. Springer International Publishing, Cham, 2016.

[40] Tomasz Komorowski, Joel L. Lebowitz, Stefano Olla, and Marielle Simon. On the conversion of work into heat: Microscopic models and macroscopic equations. In *Probability and Statistical Mechanics. Papers in honor of Errico Presutti*, volume 38 of *Ensaios Matemáticos*, pages 325–341. Sociedade Brasileira de Matemática, Rio de Janeiro, 2023.

[41] Joachim Weidmann. *Linear operators in Hilbert spaces. Transl. by Joseph Szücs*, volume 68 of *Grad. Texts Math.* Springer, Cham, 1980.

[42] András Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math.*, 34:335–343, 1973.