# Tensorizing maximal correlations

Rémi Peyre

August 19, 2010

## Abstract

For $X$ and $Y$ two random variables, the *maximal correlation coefficient* between $X$ and $Y$, denoted by $\{X:Y\}$, is the supremum value of $|\mathrm{Corr}(f(X),g(Y))|$ for real measurable functions $f$ and $g$, where "Corr" denotes the Pearson correlation coefficient. It is well known that for independent pairs of variables $(X_i,Y_i)_{i\in I}$, $\{\vec{X}_I:\vec{Y}_I\}$ is the supremum of the $\{X_i:Y_i\}$; the main goal of this monograph will be to get such tensorization results when independence between the $(X_i,Y_i)$ is only partial. More generally, for $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ random variables, we will look for a bound on $\{\vec{X}_I:\vec{Y}_J\}$ from bounds on the $\{X_i:Y_j\}$, $i\in I, j\in J$.

Our tensorization theorems shall provide new results about decorrelation for models of statistical mechanics exhibiting asymptotic independence, like the subcritical Ising model. Namely, we shall prove that in such models, two distant bunches of spins are decorrelated *regardless of their sizes and shapes* in the sense of maximal correlation: if $I$ and $J$ are sets of spins with $dist(i,j)\geqslant d$ for all $i\in I, j\in J$, then one has a non-trivial bound on $\{\vec{X}_I:\vec{Y}_J\}$ only depending on $d$.

This work will also get interested in using maximal decorrelations to get spatial central limit theorems for models like subcritical Ising's, as well as positiveness of the spectral gap of the Glauber dynamics. Again, that shall be performed thanks to tensorization techniques.

Last but not least, this monograph will present a new criterion to control $\{\mathscr{F}:\mathscr{G}\}$, $\mathscr{F}$ and $\mathscr{G}$ being two $\sigma$-algebras, from a bound on the $|\mathbf{P}[A\cap B]-\mathbf{P}[A]\mathbf{P}[B]|/\sqrt{\mathbf{P}[A]\mathbf{P}[B]}$ for all $A\in\mathscr{F}, B\in\mathscr{G}$. Similar criteria were already known, but ours improves them and can even be shown to be optimal.

# Introduction

## Overview of the monograph

This monograph is devoted to the study of maximal decorrelations, in particular to showing how this concept can be 'tensorized' to yield new results about systems of statistical mechanics exhibiting asymptotic independence. I have divided it into six chapters:

- The first chapter, numbered "0", aims at motivating the study of maximal correlations and their tensorization. In this chapter, I will recall some classical results on the subcritical Ising model, which is a classical model showing asymptotic independence for pairs of spins. When one gets interested in very large 'bunches' of spins, it is known that asymptotic independance cannot be captured by $\beta$-mixing any more, but that, in certain cases at least, it still holds in terms of $\rho$-mixing. The techniques used so far to establish $\rho$-mixing for bunches of spins are strongly limited by technical assumptions looking somehow artificial, which will motivate studying $\rho$-mixing 'for itself' and trying to tensorize it.
- In Chapter 1, I shall recall the definition of the maximal correlation coefficient (also known as "$\rho$-mixing" or "Hilbertian correlation"); I shall also recall some classical facts about this concept and give some examples. This chapter can be seen as a 'crash course' on $\rho$-mixing for the non-specialist reader: almost nothing in it is new.
- In Chapter 2, I shall give some new criteria to bound Hilbertian correlation between two $\sigma$-algebras, which criteria assume bounds on the $(\mathbf{P}[A\cap B]-\mathbf{P}[A]\mathbf{P}[B])$ for events $A$ and $B$ belonging to these respective $\sigma$-algebras. My "strong event sufficient condition", which improves previous results of several authors, shall even be shown to be optimal.
- Chapter 3 is the core of this monograph: in it I will handle tensorization of Hilbertian decorrelations. This chapter begins with a refined version of the concept of correlation, called "subjective correlation", which is necessary to write the subsequent tensorization results. Then I shall state and prove my three main tensorization theorems: Theorem 3.2.2 ('$N$ against 1' theorem) bounds the correlation between a 'simple' and a 'vector' variable; Theorem 3.3.1 ('$N$ against $M$' theorem) deals with correlation between two vector variables, and Theorem 3.3.10 ('$\mathbb{Z}$ against $\mathbb{Z}$' theorem) refines the previous one in the case where certain symmetries are present. Then, I will discuss some refinement and optimality statements about these theorems; in § 3.9, I will also present a geometric corollary of tensorization results which underlines quite well the Hilbertian aspect of maximal correlations.

- In Chapter 4, I will continue to use the tensorization techniques of Chapter 3, but this time instead of proving tensorization results *stricto sensu* I will turn to different types of results, namely the spatial central limit theorem and the presence of spectral gap for the Glauber dynamics.
- Finally, Chapter 5 will present some concrete applications of the results of this monograph. For instance, I shall prove new results about decorrelation between distant bunches of spins in Ising's model [see Theorem 5.1.1]; I will also give results of the same type for quite general models of statistical mechanics [see e.g. Theorems 5.2.10 and 5.3.7], also proving spatial CLT and spectral gap for the Glauber dynamics for these models. I will also show how tensorization of Hilbertian correlations can be used to get 'hypocoercivity' results [Theorem 5.4.6].

# Conventions and notation

This monograph is written assuming that the reader has a graduate knowledge of probability theory; whenever areas of mathematics less familiar to a probabilist are involved, references on these areas shall be provided. In particular, a good deal of analysis will be used, especially about linear operators between Hilbert spaces: a good reference on that topic is [45].

Notation will not always be perfectly rigorous: to make reading easier, it may occur sometimes that formalism is slightly loose, or that some writing conventions or assumptions are implicit. However this shall only be done in situations where adding the missing information by the reader is (hopefully) obvious.

Here is some notation used throughout this text:

### Miscellaneous

- The symbol $\mathbb{N}$ denotes the set of nonnegative integers, including 0. The set of positive integers $\mathbb{N} \smallsetminus \{0\}$ is denoted by $\mathbb{N}^*$.
- For $a,b$ real numbers, $a \wedge b$ denotes $\min\{a,b\}$, resp. $a \vee b$ denotes $\max\{a,b\}$; $a_+$ denotes the positive part of $a$, i.e. $a \vee 0$.
- For $A$ a set, $A^{\mathsf{c}}$ denotes the complement set of $A$ (the set of reference shall always be clear); $\mathbf{1}_A$ denotes the indicator function of $A$, that is, the function being 1 on $A$ and 0 on $A^{\mathsf{c}}$.
- For $A$,$B$ sets, $A \triangle B$ denotes the symmetric difference of $A$ and $B$, i.e. $(A \smallsetminus B) \uplus (B \smallsetminus A)$, where "$\uplus$" means the same as "$\cup$", but with underlining that the union is disjoint.
- The identity matrix in dimension $n$ will be denoted by $\mathbf{I}_n$. The transpose of a matrix $A$ will be denoted by $A^\top$.
- If $\Theta$ is a set endowed with a metric *dist*, then for $I,J \subset \Theta$, $dist(I,J)$ denotes the distance between $I$ and $J$, that is, $dist(I,J) := \inf\{dist(i,j) \colon\ i \in I, j \in J\}$.
- As is customary in physical literature, $\propto$ means "proportional to".
- Whenever $I$ is a set and $X$ a symbol, $\vec{X}_I$ will be a shorthand for "$(X_i)_{i \in I}$".

### Probability

- We will always work on an implicit probability space $(\Omega, \mathscr{B})$ equipped with a probability measure $\mathbf{P}$. Sub-$\sigma$-algebras of $\mathscr{B}$ will be merely called "$\sigma$-algebras"; I will also often write "variable" for "random variable". Unless explicitly specified, variables on $\Omega$ can be valued in any set.
- If $f$ is a real random variable, the expectation of $f$ is denoted by $\mathbf{E}[f]$; its variance is denoted $\mathrm{Var}(f)$; its standard deviation is denoted $\mathrm{Sd}(f) := \sqrt{\mathrm{Var}(f)}$; if $g$ is another real variable, the covariance between $f$ and $g$ is denoted by $\mathrm{Cov}(f,g) := \mathbf{E}[fg] - \mathbf{E}[f]\mathbf{E}[g]$. All that notation extends to the case where $f$ and $g$ are valued in some vector space $\mathbb{R}^N$, except that in this case it refers to vectors or matrices.
- If $B$ is an event with $\mathbf{P}[B] > 0$, then $\mathbf{P}[A|B]$, $\mathbf{E}[f|B]$, $\mathrm{Var}(f|B)$, ... stand resp. for the probability of $A$, the expectation of $f$, the variance of $f$, ... under the conditional law $d\mathbf{P}[\cdot|B] := \mathbf{1}_B\, d\mathbf{P}[\cdot]/\mathbf{P}[B]$. Similarly, if $\mathscr{F}$ is a $\sigma$-algebra, $\mathbf{P}[A|\mathscr{F}]$, $\mathbf{E}[f|\mathscr{F}]$, ... stand for the conditional probability of $A$, the conditional expectation of $f$, ... w.r.t. $\mathscr{F}$.
- Concerning conditional expectations, I will actually use two different conventions: for $\mathscr{G}$ a $\sigma$-algebra, $\mathbf{E}[f|\mathscr{G}]$ can also be denoted by $f^{\mathscr{G}}$. Both conventions can be used inside the same formula.[*][†]
- If $X$ is a variable on $\Omega$, the $\sigma$-algebra generated by $X$ (that is, the smallest $\sigma$-algebra w.r.t. which $X$ is measurable) is denoted by $\sigma(X)$. If $\mathscr{F}$ and $\mathscr{G}$ are $\sigma$-algebras, the $\sigma$-algebra generated by $\mathscr{F}$ and $\mathscr{G}$ (that is, the smallest $\sigma$-algebra containing both $\mathscr{F}$ and $\mathscr{G}$) is denoted by $\mathscr{F} \vee \mathscr{G}$, and this notation extends into the $\infty$-ary operator $\bigvee$ for an arbitrary number of $\sigma$-algebras.
- An event $A \in \mathscr{B}$ is said to have trivial probability, or to be trivial, if $\mathbf{P}[A] \in \{0,1\}$. A $\sigma$-algebra is said to be trivial if all its events are trivial. The $\sigma$-algebra $\{\emptyset, \Omega\}$, which is trivial under any law $\mathbf{P}$, will be denoted by $\mathscr{O}$ and refered to as "the" trivial sigma-algebra.
- The Lebesgue measure on $\mathbb{R}^n$ will be denoted by $dx$, "$x$" being the name of the integration variable. For a Borel set $A \subset \mathbb{R}^n$, $\int_{x\in A} dx$ will sometimes be denoted by $|A|$.
- For $C$ a positive-semidefinite matrix (possibly of dimension 1, in which case it is identified with $\sigma^2 \in \mathbb{R}_+$), $\mathscr{N}(C)$ denotes the law of the centered Gaussian vector with covariance matrix $C$. I will write $\mathscr{N}(C) + m$ to denote the non-centered Gaussian vector with variance $C$ and mean $m$.

### Functional analysis

- Unless otherwise specified, all the functional spaces considered in this monograph will be real.
- For $I$ an open interval of $\mathbb{R}$ and $k \in \mathbb{N} \cup \{\infty\}$, $\mathscr{C}_0^k(I)$ denotes the subset of functions of $\mathscr{C}^k(I)$ with compact support.
- If $\mu$ is a nonnegative measure on some measurable space $(\Omega, \mathscr{B})$, $L^2(\mu)$ denotes the set of measurable functions $f$ (up to $\mu$-a.e. equality) such that $\int_\Omega f(\omega)^2\, d\mu(\omega) <$

[*]The use of the first or the second convention will depend on the way we prefer to see the conditional expectation of $f$ w.r.t. $\mathscr{G}$: if it is rather seen as the expectation of $f$ knowing the information of $\mathscr{G}$, notation $\mathbf{E}[f|\mathscr{G}]$ will be chosen, while if it is more seen like the $\mathscr{G}$-measurable function best approximating $f$, we will use the notation $f^{\mathscr{G}}$.

[†]One must not confuse $\mathrm{Var}(f|\mathscr{G})$, which is the variance of $f$ under the law $\mathbf{P}[\cdot|\mathscr{G}]$, with $\mathrm{Var}(f^{\mathscr{G}})$ which is the (unconditioned) variance of the random variable $f^{\mathscr{G}}$. One has the well-known identity $\mathrm{Var}(f) = \mathrm{Var}(f^{\mathscr{G}}) + \mathbf{E}[\mathrm{Var}(f|\mathscr{G})]$, which I shall refer to as *associativity of variance*.

$\infty$. If $I$ is a countable set, $L^2(I)$ denotes the set of functions $f : I \to \mathbb{R}$ such that $\sum_{i \in I} f(i)^2 < \infty$. If $\mathscr{F}$ is a $\sigma$-algebra, $L^2(\mathscr{F})$ denotes the space of $\mathscr{F}$-measurable functions (up to a.s. equality) which are square-integrable w.r.t. $\mathbf{P}$. All these spaces are equipped with their natural Hilbertian product $\langle \cdot, \cdot \rangle$ and the associated norm $\|\cdot\|$.
- For $\mu$ a finite measure, in $L^2(\mu)$ the constant functions make a line which can be identified with $\mathbb{R}$; then, $\bar{L}^2(\mu)$ will denote the quotient $L^2(\mu)/\mathbb{R}$, equipped with its natural Hilbert structure. In other words, if $\bar{f} \in \bar{L}^2(\mu)$ is the projection of $f \in L^2(\mu)$, $\|\bar{f}\|_{\bar{L}^2} := \inf\{\|f - a\|_{L^2} : a \in \mathbb{R}\} = (\|f\|^2_{L^2} - \langle f, 1/\|1\|\rangle^2_{L^2})^{1/2}$. $\bar{L}^2(\mu)$ can also be seen as the subspace of centered functions of $L^2(\mu)$, i.e. as $\{f \in L^2(\mu) : \langle f, 1\rangle = 0\}$; throughout the monograph we will implicitly switch between both interpretations.
- If $L : H_1 \to H_2$ is a linear operator between two Hilbert spaces, then $L^* : H_2 \to H_1$ denotes the adjoint operator of $L$, characterized by the relationship $\langle L^* y, x\rangle_{H_1} = \langle y, Lx\rangle_{H_2}$.
- If $L : E \to F$ is a linear operator between two Banach spaces (not necessarily Hilbert) with respective norms $\|\cdot\|_E$ and $\|\cdot\|_F$, the operator norm of $f$, denoted by $|\!|\!| f |\!|\!|$, is defined as $\sup\{\|Lx\|_F : \|x\|_E = 1\}$.
- If $L : E \to E$ is a linear operator on a Banach space, then $\rho(L)$ denotes the spectral radius of $f$, that is, $\rho(L) := \lim_{k\to\infty} |\!|\!| L^k |\!|\!|^{1/k}$ —that limit always exists.
- A column vector $(a_i)_{i \in I}$ will automatically be identified with the corresponding element of $L^2(I)$. Likewise, a matrix $A = ((a_{ij}))_{(i,j) \in I \times J}$ will be identified with the corresponding linear operator from $L^2(J)$ to $L^2(I)$.
- 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. Similarly, "IP" under an equality sign will mean that this equality follows from an integration by parts.


# Acknowledgements

This work owes to many people's help. First of all I must mention V. Beffara, who launched me on this topic incidentally. His relevant comments, as well of those of (among others) Y. Ollivier, C. Villani, Y. Velenik, T. Bodineau, C. Shalizi and R. Bradley, have also been the origin for several important improvements of this monograph.

Several colleagues provided me with some help on mathematical topics I was not familiar with. In particular, Y. Ollivier suggested to me the use of Lipschitz spaces to prove Lemma 2.2.10; S. Martineau pointed out how Gershgorin's Lemma solved a technical point in the complete proof of Theorem 3.5.2; V. Calvez had the idea of using Laplace transform to prove Lemma 5.5.5. Over the Internet, F. Martinelli and S. Shlosman also gave me precious bibliographic references on the state of the art about weak and strong mixing in statistical mechanics.

Most of the drawings in this monograph were made thanks to the excellent LaTeX


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

extension Ti*k*Z, combined with computations in C language. The dice of Figure 2 have been kindly drawn for me by A. Alvarez, using POV-Ray.

# Contents

# Chapter 0

# Motivation

## 0.1 Some results on Ising's model

In this subsection we recall the definition of Ising's model and give two classical results on it, namely Theorems 0.1.8 and 0.1.9. In § 0.2, considerations on these results will serve as a motivation to the whole monograph.

### 0.1.a Ising's model

Ising's celebrated model is a basic model of equilibrium thermodynamics, which represents a ferromagnetic material:

**0.1.1 Definition.** For $n$ an integer, consider the lattice $\mathbb{Z}^n$ endowed with its usual graph structure (each vertex has $2n$ neighbours), and denote by *dist* the graph distance. Define $\Omega = \{\pm 1\}^{\mathbb{Z}^n}$, and for $\vec{\omega} \in \Omega$, set formally:

$$H(\vec{\omega}) = -\frac{1}{2} \sum_{dist(i,j)=1} \omega_i \omega_j. \tag{1}$$

Then, for $T \geqslant 0$, the *Ising model on $\mathbb{Z}^n$ at temperature $T$* is, formally, a probability measure $\mathbf{P}$ on $\Omega$ such that $\mathbf{P}[\vec{\omega}] \propto \exp(-T^{-1} H(\vec{\omega}))$. In rigorous terms, saying that $\mathbf{P}$ is an equilibrium measure for Ising's model means that for all $i \in \mathbb{Z}^n$, for all $\vec{\hat{\omega}}_{\{i\}^c} \in \{\pm 1\}^{\{i\}^c}$,

$$\mathbf{P}\big[\omega_i = \hat{\omega}_i \big| \vec{\omega}_{\{i\}^c} = \vec{\hat{\omega}}_{\{i\}^c}\big] \propto \exp\Big(T^{-1} \sum_{dist(i,j)=1} \hat{\omega}_i \hat{\omega}_j\Big). \tag{2}$$

Ising's model and the phase transition it exhibits have been the subject of dozens of works; see [21] for an overview. Here we are interested in the subcritical regime:

**0.1.2 Theorem** (Subcritical regime, [32])**.** *There is a $T_c < \infty$ (the 'Curie temperature') such that the solution of (2) is unique for $T > T_c$.*

For $T > T_c$ one says that they are in the *subcritical regime*. An interesting feature of this regime is that for distant $i$ and $j$, the random variables $\omega_i$ and $\omega_j$ are 'almost independent'. That phenomenon, called *exponential decay of correlations*, is stated by the following theorem:

**0.1.3 Theorem** (Exponential decay of correlations, [2]). *For Ising's model on $\mathbb{Z}^n$ in the subcritical regime,*

*(i) For all $i \in \mathbb{Z}^n$, $\mathbf{P}[\omega_i = -1] = \mathbf{P}[\omega_i = 1] = 1/2$.*
*(ii) There exists $\psi > 0$ and $C < \infty$ such that for all $i, j \in \mathbb{Z}^n$,*

$$|\mathbf{E}[\omega_i \omega_j]| \leqslant C \exp(-\psi \mathit{dist}(i,j)). \tag{3}$$

## 0.1.b Absence of $\beta$-mixing

Theorem 0.1.3 states that two distant spins $i$ and $j$ are exponentially decorrelated. However, it does not inform us about the dependence of 'bunches' of spins. The question is the following: if $I$ and $J$ are two disjoint, distant subsets of $\mathbb{Z}^n$, to what extent are $\vec{\omega}_I$ and $\vec{\omega}_J$ independent?

To answer such a question, the first thing to do is to define a way of measuring independence between 'complicated' variables like $\vec{\omega}_I$ and $\vec{\omega}_J$, having an arbitrarily large range. The most common choice is the $\beta$-mixing coefficient:

**0.1.4 Definition.**

(i) Recall that for $\mu$, $\nu$ two probability measures on the same measurable space $(\Omega, \mathscr{F})$, the *total variation distance* between $\mu$ and $\nu$ is the total mass of both the positive and the negative parts of the signed measure $\nu - \mu$, that is, $\mathit{dist}_{\mathrm{TV}}(\mu,\nu) = \sup_{A \subset \mathscr{F}} |\nu(A) - \mu(A)|$.
(ii) If $X$ and $Y$ are two random variables (with arbitrary ranges) defined on the same space, then one defines the *$\beta$-mixing coefficient* between $X$ and $Y$ as

$$\beta(X,Y) := \mathit{dist}_{\mathrm{TV}}(\mathit{Law}_X \otimes \mathit{Law}_Y, \mathit{Law}_{(X,Y)}). \tag{4}$$

Notice that $\beta(X,Y)$ actually only depends on the $\sigma$-algebras $\sigma(X)$ and $\sigma(Y)$ [7, Formula (1.5)]. The following proposition is immediate:

**0.1.5 Proposition.**

*(i) One has always $\beta(X,Y) \in [0,1]$, and*
*(ii) $\beta(X,Y) = 0$ if and only if $X$ and $Y$ are independent ;*
*(iii) $\beta(X,Y) = 1$ if and only if $\mathit{Law}_X \otimes \mathit{Law}_Y$ and $\mathit{Law}_{(X,Y)}$ are mutually disjoint.*
*(iv) If $X'$ is $X$-measurable and $Y'$ is $Y$-measurable, then $\beta(X',Y') \leqslant \beta(X,Y)$.*

So, $\beta(X,Y)$ is a measure of 'how much $X$ and $Y$ are correlated'.

With that tool at hand, decorrelation of bunches of spins in statistical physics models has already been thoroughly studied. Concerning Ising's model, there are two well-known great results:

**0.1.6 Theorem** (Weak mixing property, [29]). *For Ising's model on $\mathbb{Z}^n$ in the subcritical regime, there exists $\psi > 0$ and $C < \infty$ (the same as in Theorem 0.1.3) such that for all disjoint $I, J \subset \mathbb{Z}^n$:*

$$\beta(\vec{\omega}_I, \vec{\omega}_J) \leqslant C \sum_{(i,j) \in I \times J} \exp(-\psi \mathit{dist}(i,j)). \tag{5}$$

**0.1.7 Theorem** (Complete analyticity, [15])**.** *There exists some $T_c \leqslant T'_c < \infty$*[*] *such that for $T > T'_c$, Ising's model is* completely analytical*, i.e. there exists $\psi' > 0$ and $C' > 0$ such that the following holds: for all $K \subset \mathbb{Z}^n$, for all 'boundary' condition $\vec{\hat{\omega}}_K \in \{\pm 1\}^K$, denoting $\mathbf{P}_{\vec{\hat{\omega}}_K} := \mathbf{P}[\cdot|\vec{\omega}_K = \vec{\hat{\omega}}_K]$, Formula (5) holds with Law replaced by $Law_{\vec{\hat{\omega}}_K}$ and $\psi, C$ replaced resp. by $\psi'$ and $C'$.*

Thanks to Theorem 0.1.6, we get an exponential decay of correlation between two bunches of spins of fixed size when the distance between these bunches increases. However, we cannot say much about decorrelation between bunches of *variable* size which are at fixed distance from each other. For instance, consider $n = 2$; for $x > 0$, define $I_l := \{(0, y) : |y| \leqslant l\}$, resp. $J_l := \{(x, y) : |y| \leqslant l\}$. Then Theorem 0.1.6 cannot give us anything better than something like:

$$\beta(\vec{\omega}_{I_l}, \vec{\omega}_{J_l}) \lesssim Cle^{-\psi x}. \tag{6}$$

But recall that a $\beta$-mixing coefficient is always bounded by 1; so, for $l \gtrsim e^{\psi x}/C$, (6) tell us absolutely nothing about the decorrelation between $I_l$ and $J_l$.

Though the bound (5) is not completely optimal, the previous point is an *intrinsic* shortcoming of $\beta$-mixing coefficients, in the sense that it can be proved that bounds like (6) *must* become trivial when $l \to \infty$:

**0.1.8 Theorem.** *For all $T_c < T < \infty$, for all $x > 0$, denoting $I := \{0\} \times \mathbb{Z}$ and $J := \{x\} \times \mathbb{Z}$, one has*

$$\beta(\vec{\omega}_I, \vec{\omega}_J) = 1. \tag{7}$$

*Proof.* Denote $i_0 := (0,0)$, resp. $j_0 := (x,0)$. As we told in Theorem 0.1.3-(i), $\mathbf{E}[\omega_{i_0}], \mathbf{E}[\omega_{j_0}] = 0$. Interpretation of Ising's model as a random-cluster model [21, § 1.4] shows that $\mathbf{P}[\omega_{i_0} = \omega_{j_0}] > 1/2$, so we define

$$\gamma := \mathbf{E}[\omega_{i_0}\omega_{j_0}] > 0. \tag{8}$$

Now, let $N$ be some large integer, fixed for the time being. Let $p$ be some large integer and define $i_1, \ldots, i_N$, resp. $j_1, \ldots, j_N$, by $i_k := (0, kp)$, resp. $j_k := (x, kp)$; by translation invariance, for each $k$, $(\omega_{i_k}, \omega_{j_k})$ has the same law as $(\omega_{i_0}, \omega_{j_0})$. Now denote by $P_{N,p}$ the joint law of $(\omega_{i_1}, \ldots, \omega_{i_N}, \omega_{j_1}, \ldots, \omega_{j_N})$. By Theorem 0.1.6, when $p \to \infty$ $P_{N,p}$ tends to the law $P_{N,\infty}$[†] where all the $(\omega_{i_k}, \omega_{j_k})$ are independent—with the same law as previous, namely,

$$P_{N,\infty}[\omega_{i_k} = \eta \text{ and } \omega_{j_k} = \theta] = \frac{1 + \gamma\eta\theta}{4}. \tag{9}$$

Therefore, the value of $\beta((\omega_{i_k})_k, (\omega_{j_k})_k)$ under the law $P_{N,p}$, which is known by Proposition 0.1.5-(iv) to be a lower bound for $\beta(\vec{\omega}_I, \vec{\omega}_J)$, tends to its value under $P_{N,\infty}$ when $p \to \infty$. This is summed up by the following formula:

$$\beta(\vec{\omega}_I, \vec{\omega}_J) \geqslant \beta_{P_{N,\infty}}((\omega_{i_k})_{1 \leqslant k \leqslant N}, (\omega_{j_k})_{1 \leqslant k \leqslant N}). \tag{10}$$

[*]It is not known whether $T'_c = T_c$ today, but in general situations, weak mixing does not always imply complete analyticity. A classical counterexample is Ising's model with external field, cf. [30, § 2].

[†]Note that $P_{N,p}$ takes its values in a space of finite dimension, so there is no ambiguity when speaking of its convergence.

To end the proof, we will bound the right-hand side of (10) below by a quantity which tends to 1 when $N \to \infty$. Denote by $\tilde{P}_{N,\infty}$ the product of two the marginals of $P_{N,\infty}$ relative resp. to the $(\omega_{i_k})_k$ and the $(\omega_{j_k})_k$, so that $\beta_{P_{N,\infty}}((\omega_{i_k})_k,(\omega_{j_k})_k) = dist_{\mathrm{TV}}(P_{N,\infty},\tilde{P}_{N,\infty})$ by the very definition of the $\beta$-mixing coefficient. Obviously the expression of $\tilde{P}_{N,\infty}$ is the same as the expression of $\tilde{P}_{N,\infty}$, but with $\gamma$ replaced by 0 in (9). Under $P_{N,\infty}$, $(\omega_{i_k}\omega_{j_k})_{1\leqslant k\leqslant N}$ is a sequence of i.i.d. random variables having a certain law with mean $\gamma > \gamma/2$, so that by the law of large numbers,

$$P_{N,\infty}\Big[N^{-1}\sum_{k=1}^{N}\omega_{i_k}\omega_{j_k} \leqslant \frac{\gamma}{2}\Big] \overset{N\to\infty}{\to} 0. \tag{11}$$

Similarly, since $0 < \gamma/2$,

$$P_{N,\infty}\Big[N^{-1}\sum_{k=1}^{N}\omega_{i_k}\omega_{j_k} \leqslant \frac{\gamma}{2}\Big] \overset{N\to\infty}{\to} 1, \tag{12}$$

so that

$$dist_{\mathrm{TV}}(P_{N,\infty},\tilde{P}_{N,\infty}) \geqslant \Big|\tilde{P}_{N,\infty}\Big[N^{-1}\sum_{k=1}^{N}\omega_{i_k}\omega_{j_k} \leqslant \frac{\gamma}{2}\Big] - P_{N,\infty}[\textit{the same}]\Big| \overset{N\to\infty}{\to} 1, \tag{13}$$

which proves our point. □

## 0.1.c Presence of $\rho$-mixing

So, Theorem 0.1.8 tells us that, for Ising's model on $\mathbb{Z}^2$, there is a 'full' correlation between $\vec{\omega}_I$ and $\vec{\omega}_J$ in the sense of $\beta$-mixing. Yet it is well known too that Theorem 0.1.6 nevertheless implies a Hilbertian form of decorrelation (called "$\rho$-mixing", cf. Remark 1.1.2) between these variables:

**0.1.9 Theorem.** *For Ising's model on $\mathbb{Z}^2$ in the subcritical regime, defining as before $I = \{0\}\times\mathbb{Z}$ and $J = \{x\}\times\mathbb{Z}$ for some $x > 0$, one has for all $f \in \bar{L}^2(\vec{\omega}_I)$ and $g \in \bar{L}^2(\vec{\omega}_J)$:*

$$|\mathbf{E}[fg]| \leqslant e^{-\psi x}\,\mathrm{Sd}(f)\,\mathrm{Sd}(g), \tag{14}$$

*where $\psi$ is the same as in Theorem 0.1.6.*

*Proof.* Define the operator

$$\begin{array}{rccc} P: & \bar{L}^2(\vec{\omega}_I) & \to & \bar{L}^2(\vec{\omega}_J) \\ & f & \mapsto & f^{\sigma(\vec{\omega}_J)}. \end{array} \tag{15}$$

(Recall that $f^{\sigma(\vec{\omega}_J)}$ is an alternative notation for $\mathbf{E}[f|\vec{\omega}_J]$, insisting on the its being a $\sigma(\vec{\omega}_J)$-measurable function). Then (14) is equivalent to proving that $|||P||| \leqslant e^{-\psi x}$ (see § 1.1.c). Now for all $t \in \{0,\ldots,x\}$, denote $\omega_{(t)} := \vec{\omega}_{\{t\}\times\mathbb{Z}}$, and for all $t \in \{1,\ldots,x\}$,

$$\begin{array}{rccc} \pi_t: & \bar{L}^2(\omega_{(t-1)}) & \to & \bar{L}^2(\omega_{(t)}) \\ & f & \mapsto & f^{\sigma(\omega_{(t)})}. \end{array} \tag{16}$$

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

$$P = \pi_x \circ \cdots \circ \pi_2 \circ \pi_1. \tag{17}$$

Now, by horizontal translation all the $\bar{L}^2(\omega_{(t)})$ can be identified with a common Hilbert space $H$. Then all the $\pi_t$ are identified with operators on $\bar{L}^2(H)$, and by the translation invariance of the model all these operators are actually the same. $P$ is also identified with an operator on $\bar{L}^2(H)$, and (17) becomes:

$$P = \pi^x. \tag{18}$$

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

Take $C < \infty$ like in Theorem 0.1.6. For $l$ an integer, denote $I_l$ and $J_l$ to be resp. $\{0\} \times \{-l, \ldots, l\}$ and $\{x\} \times \{-l, \ldots, l\}$. Let $f$ be a bounded[‡] function of $\bar{L}^2(\vec{\omega}_{I_l})$ and denote $M := \|f\|_{L^\infty}$. By translation, $f$ can also be identified with a function of $\bar{L}^2(\vec{\omega}_{J_l})$, which is also bounded by $M$. Now, since

$$\mathbf{E}[f(\vec{\omega}_{I_l})f(\vec{\omega}_{J_l})] = \mathrm{Cov}(f(\vec{\omega}_{I_l}), \vec{\omega}_{J_l}) = \int f(\vec{\omega}_{I_l})f(\vec{\omega}_{J_l})\, d\big(Law(\vec{\omega}_{I_l \uplus J_l}) - Law(\vec{\omega}_{I_l}) \otimes Law(\vec{\omega}_{J_l})\big), \tag{19}$$

we can apply (5) to $I_l$ and $J_l$ to obtain:

$$|\mathbf{E}[f(\vec{\omega}_{I_l})f(\vec{\omega}_{J_l})]| \leqslant M^2 \cdot 2C(2l+1)^2 e^{-\psi x}. \tag{20}$$

In terms of operators, (20) means that

$$|\langle f, Pf \rangle_{\bar{L}^2(H)}| \leqslant 2(2l+1)^2 M^2 C e^{-\psi x}. \tag{21}$$

As the value of $x$ played no particular role to establish (21), that formula can be generalized into

$$|\langle f, \pi^t f \rangle_{\bar{L}^2(H)}| \leqslant 2(2l+1)^2 M^2 C e^{-\psi t} \tag{22}$$

for all $t \in \mathbb{N}^*$. Letting $t$ tend to infinity, we obtain that for all $l$, for all $f \in \bar{L}^2(\vec{\omega}_{I_l}) \cap L^\infty$,

$$\overline{\lim_{t \to \infty}} \left( \log |\langle f, \pi^t f \rangle| \right)^{1/t} \leqslant e^{-\psi}. \tag{23}$$

But $\bigcup_{l \in \mathbb{N}} (\bar{L}^2(\vec{\omega}_{I_l}) \cap L^\infty)$ is a dense subset of $\bar{L}^2(\vec{\omega}_I)$, so by Lemma 0.3.1 set in appendix, we conclude that $|||\pi|||_{\bar{L}^2(H)} \leqslant e^{-\psi}$, which is what we wanted. □

*0.1.10 Remark.* Claim 0.1.8 and Theorem 0.1.9 adapt straighforwardly, with similars proof, to any $n \geqslant 2$, replacing $I$ by $\{0\} \times \mathbb{Z}^{n-1}$ and $J$ by $\{x\} \times \mathbb{Z}^{n-1}$.

## 0.2 Problematics

Thanks to Theorem 0.1.9, we see that the Hilbertian concept of $\rho$-mixing can reveal some independence between infinite bunches of lowly correlated variables in situations where the $\beta$-mixing coefficient does not show any independence at all. In the proof we gave, $\rho$-mixing appeared as a corollary of $\beta$-mixing decorrelation results. What additional hypotheses did we need to get our corollary? We used at least the following:

[‡]In fact here it is superfluous to impose that $f$ is bounded since $\vec{\omega}_{I_l}$ can only take a finite number of values. I wrote the proof like this just to underline that the finiteness of the range of the $\omega_i$ does not play any role in the proof.

- To introduce the Markov chain $\omega_{(0)} \to \cdots \to \omega_{(x)}$, we used that the interactions of our model had finite range.
- To identify all the spaces $\bar{L}^2(\omega_{(t)})$, we used that $I$ and $J$ were shaped such that one could tile up $\mathbb{Z}^2$ with a sequence of tiles of the same shape (namely, here, tiles of the form $\{t\} \times \mathbb{Z}$).
- To say that all the $\pi_t$ were the same modulo that identification, we used the translation invariance of the model.
- To state that the stationary Markov chain $\omega_{(0)} \to \cdots \to \omega_{(x)}$ was reversible, we used the reflection invariance of the model.
- To use Lemma 0.3.1, we used the exponential decay of correlations.

All these points make the proof of Theorem 0.1.9 we gave in § 0.1 quite difficult to generalize. What, for instance, if we take $I$ and $J$ with arbitrary shapes, just requiring that $dist(I,J) \geqslant x$? 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 $\rho$-mixing relies fundamentally on the peculiar symmetries of the case we treated...

So, here will be the goal of this monograph: *establishing $\rho$-mixing estimates by general methods*. To achieve this goal, I shall try to concentrate on the properties of $\rho$-mixing 'for itself' rather than to its links with other forms of decorrelation. I will carry out a thorough study of the $\rho$-mixing coefficient, also called "maximal correlation", in order to get $\rho$-mixing results for 'complicated' variables from decorrelation results *of the same type* for more 'basic' variables; in other words, I will *tensorize* maximal decorrelations. It turns out that tensorization for such type of decorrelations gives results which are quite robust as the size of bunches of variables increases. Thanks to this methods, I shall obtain fairly new decorrelation theorems for various models of statistical physics.

This work is intended to be complete in some sense. I mean, besides the core of this monograph—namely, tensorization results—, I have tried to answer several other questions which appeared natural to me concerning maximal decorrelation. This includes studying many examples, finding sharp criteria for maximal decorrelation, looking at the optimality issues in the tensorization results or showing other applications of the tensorization techniques. Though these topics were initially thought as 'side-work', some of them may be quite interesting for themselves.

## 0.3 Appendix: On the norm of self-adjoint operators

In this appendix we prove the following

**0.3.1 Lemma.** *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*

$$\{x \in H \colon \varlimsup_{k\to\infty} |\langle L^k x, x\rangle|^{1/k} \leqslant C\} \tag{24}$$

*is a dense subset of $H$.*

*Proof.* Reasoning by contraposition, we have to show that, for $L$ a self-adjoint operator on $H$, for all $C < |||L|||$, the set of the $x \in H$ such that

$$\overline{\lim_{k\to\infty}} |\langle L^k x, x\rangle|^{1/k} > C \tag{25}$$

contains a non-empty open subset of $H$.

Since $L$ is self-adjoint, by the spectral theorem [45, Theorem 7.18], it is unitarily equivalent to the "multiplication by identity" operator $M$ on a space $\bigoplus_{\alpha\in A} L^2(\rho_\alpha)$, for $A$ some set and $\rho_\alpha$ some Radon measures on $\mathbb{R}$, that is (in the following equation, the variable $\lambda$ is free, so that $f(\lambda)$ is synonymous with $f$):

$$M\Big(\sum_\alpha f_\alpha(\lambda)\Big) = \sum_\alpha \lambda f_\alpha(\lambda). \tag{26}$$

So we will assume $L$ is of that form.

One has obviously:

$$|||L||| = \sup\{\lambda \geqslant 0 : \ (\exists \alpha \in A)\ \big(\rho_\alpha([-\lambda,\lambda]^c) > 0\big)\}; \tag{27}$$

moreover, for all $f \in H$, $f = \sum_{\alpha\in A} f_\alpha$ with $f_\alpha \in L^2(\rho_\alpha)$,

$$\langle L^k f, f\rangle = \sum_{\alpha\in A} \int_{\mathbb{R}} \lambda^k |f_\alpha(\lambda)|^2 \, d\rho_\alpha(\lambda), \tag{28}$$

so that (observing that, for $k$ even, $\lambda^k \geqslant 0 \ \ \forall \lambda$)

$$\overline{\lim_{k\to\infty}} |\langle L^k f, f\rangle|^{1/k} = \sup\Big\{\lambda \geqslant 0 : \ (\exists \alpha \in A)\ \Big(\int_{[-\lambda,\lambda]^c} |f_\alpha(\lambda')|^2 \, d\rho_\alpha(\lambda') > 0\Big)\Big\}. \tag{29}$$

Now, for $C < |||L|||$, the set

$$U = \Big\{f \in H : \ (\exists \alpha \in A)\ \Big(\int_{[-C,C]^c} |f_\alpha(\lambda)|^2 \, d\rho_\alpha(\lambda) > 0\Big)\Big\} \tag{30}$$

is open because $\int_{[-C,C]^c} |f_\alpha(\lambda)|^2 \, d\rho_\alpha(\lambda)$ is a continuous function of $f$, and it is non-empty by (27). But (25) is satisfied for all $x \in U$ by (29), so $U$ fulfills our quest. $\square$

# Chapter 1

# A first approach to maximal correlations

## 1.1 Definition and first properties

### 1.1.a Equivalent definitions

**1.1.1 Definition.** Let $(\Omega, \mathscr{B}, \mathbf{P})$ be a probability space. For $\mathscr{F}, \mathscr{G}$ two sub-$\sigma$-algebras of $\mathscr{B}$, the *maximal correlation coefficient* (or merely "correlation") between $\mathscr{F}$ and $\mathscr{G}$ is defined as

$$\{\mathscr{F} : \mathscr{G}\} := \sup_{\substack{f \in \bar{L}^2(\mathscr{F}) \smallsetminus \{0\} \\ g \in \bar{L}^2(\mathscr{G}) \smallsetminus \{0\}}} \frac{|\mathbf{E}[fg]|}{\mathrm{Sd}(f)\,\mathrm{Sd}(g)}. \tag{31}$$

If the supremum in (31) is taken over an empty set, that is, if $\mathscr{F}$ or $\mathscr{G}$ is trivial, we define this supremum to be 0.

*1.1.2 Remark.* $\{\mathscr{F} : \mathscr{G}\}$ is often called "$\rho$-mixing coefficient" between $\mathscr{F}$ and $\mathscr{G}$ and denoted by $\rho(\mathscr{F}, \mathscr{G})$, cf. [7].

*1.1.3 Remark.* In other words, $\{\mathscr{F} : \mathscr{G}\}$ is the best $k \in \mathbb{R}_+$ such that the following refined Cauchy–Schwarz inequality holds in the Hilbert space $\bar{L}^2(\mathscr{B})$:

$$\forall f \in \bar{L}^2(\mathscr{F}) \;\; \forall g \in \bar{L}^2(\mathscr{G}) \qquad |\langle f, g\rangle| \leqslant k \|f\| \|g\|. \tag{32}$$

Yet another formulation is that $\{\mathscr{F} : \mathscr{G}\}$ is the cosine of the angle between $\bar{L}^2(\mathscr{F})$ and $\bar{L}^2(\mathscr{G})$, seen as subspaces of $\bar{L}^2(\mathscr{B})$ —this angle being defined as the infimum angle between any two non-zero vectors of these respective subspaces.

If we speak in terms of $L^2$ spaces rather than $\bar{L}^2$ spaces, $\{\mathscr{F} : \mathscr{G}\}$ is the best $k \in \mathbb{R}_+$ such that for all non-constant square-integrable $f, g$ resp. $\mathscr{F}$ and $\mathscr{G}$-measurable,

$$|\mathrm{Corr}(f, g)| \leqslant k, \tag{33}$$

where $\mathrm{Corr}(f, g) := \mathrm{Cov}(f, g)/\mathrm{Sd}(f)\mathrm{Sd}(g)$ is the Pearson correlation coefficient between $f$ and $g$.

**1.1.4 Definition.** We say that $\mathscr{F}$ and $\mathscr{G}$ are *$\varepsilon$-decorrelated*, resp. *$\varepsilon$-correlated*, if $\{\mathscr{F}:\mathscr{G}\} \leqslant \varepsilon$, resp. $\{\mathscr{F}:\mathscr{G}\} \geqslant \varepsilon$.

**1.1.5 Definition.** If $X$ and $Y$ are random variables (with arbitrary range), then $\{X:Y\}$ denotes $\{\sigma(X):\sigma(Y)\}$.

*1.1.6 Remark.* One can rewrite Definition 1.1.5 as

$$\{X:Y\} = \sup_{f,g} \frac{\mathrm{Cov}(f(X),g(Y))}{\mathrm{Sd}(f(X))\,\mathrm{Sd}(g(Y))}, \tag{34}$$

where it is implied that $f$ and $g$ have to be measurable, real, and such that $0 < \mathrm{Sd}(f(X)), \mathrm{Sd}(g(Y)) < \infty$.

☛ *More generally, all the questions relative to maximal correlations may be handled either in terms of $\sigma$-algebras or in terms of random variables. In the sequel, we will frequently switch implicitly between these two paradigms.*

It is natural to enquire what happens if one deals with *complex* $\bar{L}^2$ spaces. In fact it does not change anything:

**1.1.7 Proposition** ([46, Theorem 1.1])**.** *Let $\mathscr{F}$ and $\mathscr{G}$ be two $\sigma$-algebras and let $f,g$ be two complex centered $L^2$ variables, measurable w.r.t. resp. $\mathscr{F}$ and $\mathscr{G}$. Then, with* $\mathrm{Sd}(f)$ *meaning* $\sqrt{\mathbf{E}[|f-\mathbf{E}[f]|^2]}$*, one has:*

$$|\mathbf{E}[fg]| \leqslant \{\mathscr{F}:\mathscr{G}\}\,\mathrm{Sd}(f)\,\mathrm{Sd}(g). \tag{35}$$

*Proof.* I recall the proof for the sake of completeness. Up to multiplying $g$ by a well-chosen unit complex number, we can assume that $\mathbf{E}[fg] \in \mathbb{R}_+$. Then we can apply Definition 1.1.1 to the *real* $\bar{L}^2$ variables $\mathfrak{Re}\, f$ and $\mathfrak{Re}\, g$, resp. $\mathfrak{Im}\, f$ and $\mathfrak{Im}\, g$, getting:

$$\begin{aligned} |\mathbf{E}[fg]| = \mathfrak{Re}\,\mathbf{E}[fg] &= \mathbf{E}[\mathfrak{Re}\, f\, \mathfrak{Re}\, g] - \mathbf{E}[\mathfrak{Im}\, f\, \mathfrak{Im}\, g] \\ &\leqslant \{\mathscr{F}:\mathscr{G}\}\big(\mathrm{Sd}(\mathfrak{Re}\, f)\mathrm{Sd}(\mathfrak{Re}\, g) + \mathrm{Sd}(\mathfrak{Im}\, f)\mathrm{Sd}(\mathfrak{Im}\, g)\big) \\ &\underset{\mathrm{CS}}{\leqslant} \{\mathscr{F}:\mathscr{G}\}\sqrt{\mathrm{Var}(\mathfrak{Re}\, f) + \mathrm{Var}(\mathfrak{Im}\, f)}\sqrt{\mathrm{Var}(\mathfrak{Re}\, g) + \mathrm{Var}(\mathfrak{Im}\, g)} \\ &= \{\mathscr{F}:\mathscr{G}\}\,\mathrm{Sd}(f)\,\mathrm{Sd}(g). \end{aligned} \tag{36}$$

□

Now we turn to a different way of seeing correlation levels.

**1.1.8 Definition.** For $\mathscr{F},\mathscr{G}$ two $\sigma$-algebras, we denote by $\pi_{\mathscr{G}\mathscr{F}}$ the 'projection' operator

$$\begin{array}{rccc} \pi_{\mathscr{G}\mathscr{F}}: & \bar{L}^2(\mathscr{F}) & \to & \bar{L}^2(\mathscr{G}) \\ & f & \mapsto & f^{\mathscr{G}}. \end{array} \tag{37}$$

For $\mathscr{F},\dots,\mathscr{Z}$ $\sigma$-algebras, we denote $\pi_{\mathscr{Z}\mathscr{Y}\mathscr{X}\dots\mathscr{G}\mathscr{F}} := \pi_{\mathscr{Z}\mathscr{Y}} \circ \pi_{\mathscr{Y}\mathscr{X}} \circ \cdots \circ \pi_{\mathscr{G}\mathscr{F}}$.

With this vocabulary at hand,

**1.1.9 Proposition.** *For $\mathscr{F},\mathscr{G}$ two $\sigma$-algebras,* $\{\mathscr{F}:\mathscr{G}\} = |||\pi_{\mathscr{G}\mathscr{F}}|||$.

*Proof.* $\pi_{\mathscr{G}\mathscr{F}}$ is the orthogonal projection from $\bar{L}^2(\mathscr{F})$ to $\bar{L}^2(\mathscr{G})$ in the Hilbert space $\bar{L}^2(\mathscr{B})$, so its norm is the cosine of the angle between $\bar{L}^2(\mathscr{F})$ and $\bar{L}^2(\mathscr{G})$, i.e. $\{\mathscr{F}:\mathscr{G}\}$. □

*1.1.10 Remark.* One has $\pi_{\mathscr{F}\mathscr{G}} = \pi^*_{\mathscr{G}\mathscr{F}}$, since $\langle \pi_{\mathscr{G}\mathscr{F}} f, g\rangle = \mathbf{E}[fg] = \langle f, \pi_{\mathscr{F}\mathscr{G}} g\rangle$. Therefore the expression $|||\pi_{\mathscr{G}\mathscr{F}}|||$ in Proposition 1.1.9 can be rewritten into $\sqrt{|||\pi_{\mathscr{F}\mathscr{G}\mathscr{F}}|||}$, which is also $\rho(\pi_{\mathscr{F}\mathscr{G}\mathscr{F}})$ since $\pi_{\mathscr{F}\mathscr{G}\mathscr{F}}$ is self-adjoint.

## 1.1.b Immediate properties

Having defined Hilbertian correlations, it is now time to study their behaviour.

The following properties are trivial from Definition 1.1.1:

**1.1.11 Proposition.** *For all σ-algebras $\mathscr{F}$, $\mathscr{G}$ and $\mathscr{G}'$,*

*(i)* $\{\mathscr{G}:\mathscr{F}\} = \{\mathscr{F}:\mathscr{G}\}$*;*
*(ii)* $\mathscr{G} \subset \mathscr{G}' \ \Rightarrow\ \{\mathscr{F}:\mathscr{G}\} \leqslant \{\mathscr{F}:\mathscr{G}'\}$*;*
*(iii)* $\{\mathscr{F}:\mathscr{G}\} \in [0,1]$*;*
*(iv)* $\{\mathscr{F}:\mathscr{G}\} = 0$ *if and only if $\mathscr{F}$ and $\mathscr{G}$ are independent;*
*(v) If $\mathscr{F}$ is not trivial, then* $\{\mathscr{F}:\mathscr{F}\} = 1$.

When one is concerned by correlation between variables, it often occurs that some of these variables are vector-valued. The following proposition means that it suffices to know the behaviour of finite-length vectors to understand the behaviour of all vectors:

**1.1.12 Proposition.** *Let $I, J$ be possibly infinite sets and let $\vec{X}_I, \vec{Y}_J$ be vector-valued variables. Then, denoting "$I' \Subset I$" to mean that $I'$ is a finite subset of $I$,*

$$\{\vec{X}_I : \vec{Y}_J\} = \sup_{I'\Subset I, J'\Subset J} \{\vec{X}_{I'} : \vec{Y}_{J'}\}. \tag{38}$$

*Proof.* This is because $\bigcup_{I'\Subset I} \bar{L}^2(\vec{X}_{I'})$, resp. $\bigcup_{J'\Subset J} \bar{L}^2(\vec{X}_{J'})$, is a dense subset of $\bar{L}^2(\vec{X}_I)$, resp. $\bar{L}^2(\vec{Y}_J)$. That property follows by classical approximation arguments like in the proof of [42, Theorem 3.14]. See [8, Theorem 3.16(II-3)] for a more detailed proof. □

## 1.1.c Operator interpretation

**1.1.13 Proposition.** *If $X \to Y \to Z$ is a Markov chain, then* $\{X:Z\} \leqslant \{X:Y\}\{Y:Z\}$.

*Proof.* The Markov chain property is equivalent to meaning that $\pi_{ZX} = \pi_{ZYX}$, so the result is a consequence of the submultiplicativity of operator norms. See also [41, § VII-4]. □

There is a refined version of Proposition 1.1.13 which is particularly interesting for reversible chains:

**1.1.14 Proposition.** *If $X \to Y \to Z$ is a Markov chain, then* $\{X:Z\} = \sqrt{\rho(\pi_{YZY} \circ \pi_{YXY})}$.

*Proof.* Because of the Markov chain property, $\pi_{XZ} = \pi_{XY} \circ \pi_{YZ}$ and $\pi_{ZX} = \pi_{ZY} \circ \pi_{YX}$. Using that for any pair of operators $\pi: H_1 \to H_2$ and $\tau: H_2 \to H_1$, one has $\rho(\pi \circ \tau) = \rho(\tau \circ \pi)$, we get that $\{X : Z\}^2 = \rho(\pi_{XY}) = \rho(\pi_{XYZYX}) = \rho(\pi_{XY} \circ \pi_{YZYX}) = \rho(\pi_{YZYX} \circ \pi_{XY}) = \rho(\pi_{YZY} \circ \pi_{YXY})$. □

**1.1.15 Corollary.** *If $\cdots \to X_{-1} \to X_0 \to X_1 \to \cdots$ is a stationary Markov chain so that $\pi_{X_1X_0}$ and $\pi_{X_0X_1}$ commute*[*]*, then for all $k \in \mathbb{Z} \smallsetminus \{0\}$, $\{X_0 : X_k\} = \{X_0 : X_1\}^{|k|}$.*

*Proof.* Since the chain is stationary, all the $X_n$ have the same law and thus all the $\bar{L}^2(X_n)$ can be identified; then the stationarity property is equivalent to saying that $\pi_{X_{n+1}X_n} = \pi_{X_1X_0}$ for all $n \in \mathbb{Z}$. Thanks to the commutation hypothesis, one can write for $k > 0$:

$$\{X_0 : X_k\} = \sqrt{\rho(\pi_{X_0X_kX_0})} = \sqrt{\rho(\pi^k_{X_0X_1} \circ \pi^k_{X_1X_0})} = \sqrt{\rho(\pi^k_{X_0X_1X_0})} = \sqrt{\rho(\pi_{X_0X_1X_0})^k} = \{X_0 : X_1\}^k. \tag{39}$$

For the case $k < 0$, we use that $\{X_0 : X_k\} = \{X_{-k} : X_0\}$. □

## 1.1.d First criteria for decorrelation

### Density sufficient condition

**1.1.16 Proposition.** *Let $X$ and $Y$ be two random variables resp. valued in $E$ and $F$. Suppose that $Law(X,Y)$ has a density $h$ w.r.t. the product probability $Law(X) \otimes Law(Y)$. Then*

$$\{X : Y\} \leqslant \Big( \int_{E\times F} (h-1)^2 \, dLaw_X \, dLaw_Y \Big)^{1/2}. \tag{40}$$

*1.1.17 Remark.* The integral expression in (40) is nothing but the bilinearized version of the mutual information

$$I(X;Y) := \int_{E\times F} h \log h \, dLaw_X \, dLaw_Y. \tag{41}$$

Yet Example 1.3.2 shows that one does not have $\{X : Y\} \leqslant I(X;Y)$ in general.

*Proof.* To alleviate notation, denote resp. $\mathbf{P}_X, \mathbf{P}_Y, \mathbf{P}_{(X,Y)}$ for $Law(X)$, $Law(Y)$, $Law(X,Y)$. Let $f$ and $g$ be centered $L^2$ functions being resp. $X$- and $Y$-measurable. Observe first that

$$\int_{E\times F} f(x)g(y) \, d\mathbf{P}_X[x] d\mathbf{P}_Y[y] = \Big( \int_E f \, d\mathbf{P}_X \Big) \Big( \int_F g \, d\mathbf{P}_Y \Big) = 0 \times 0 = 0, \tag{42}$$

so that

$$\mathbf{E}[fg] = \int fg \, d\mathbf{P}_{(X,Y)} = \int hfg \, d\mathbf{P}_X d\mathbf{P}_Y = \int (h-1) fg \, d\mathbf{P}_X d\mathbf{P}_Y \tag{43}$$

and thus

$$|\mathbf{E}[fg]| \leqslant \Big( \int (h-1)^2 \, d\mathbf{P}_X d\mathbf{P}_Y \Big)^{1/2} \Big( \int f^2 g^2 \, d\mathbf{P}_X d\mathbf{P}_Y \Big)^{1/2} \tag{44}$$

by the Cauchy–Schwarz inequality. But the last factor in the right-hand side of (44) is

$$\Big( \int f^2(x) g^2(y) \, d\mathbf{P}_X[x] d\mathbf{P}_Y[y] \Big)^{1/2} = \Big( \int f^2 d\mathbf{P}_X \Big)^{1/2} \Big( \int g^2 d\mathbf{P}_Y \Big)^{1/2} = \mathrm{Sd}(f)\mathrm{Sd}(g), \tag{45}$$

so that (40) is proved.

You may also see [9, Theorem 2.5] for an analogous result. □

[*]Reversible chains always satisfy this condition since then $\pi_{X_1X_0} = \pi_{X_0X_1}$.

**Event necessary condition**

**1.1.18 Proposition** (event necessary condition)**.** *Let $\mathscr{F}$ and $\mathscr{G}$ be two $\sigma$-algebras. If $\{\mathscr{F}:\mathscr{G}\} \leqslant \varepsilon$, then for all events $A \in \mathscr{F}$ and $B \in \mathscr{G}$ with respective probabilities $p$ and $q$,*

$$|\mathbf{P}[A \cap B] - pq| \leqslant \varepsilon\sqrt{p(1-p)q(1-q)}. \tag{46}$$

*In particular, if there exists two non-trivial events $A \in \mathscr{F}, B \in \mathscr{G}$ which are equivalent (in the sense that $\mathbf{P}[A \triangle B] = 0$), then $\{\mathscr{F}:\mathscr{G}\} = 1$.*[†]

*Proof.* It follows from (33) applied to $\mathbf{1}_A$ and $\mathbf{1}_B$. □

## 1.1.e Independent tensorization

Now we are turning to the basic tensorization theorem, which will motivate § 3:

**1.1.19 Theorem** ([13, Theorem 6.2])**.** *Let $I$ be a set and let $\vec{X}_I, \vec{Y}_I$ be vectors of variables. Suppose all the pairs $(X_i, Y_i)$, $i \in I$ are independent, then*

$$\{\vec{X}:\vec{Y}\} = \sup_{i\in I}\{X_i : Y_i\}. \tag{47}$$

*Proof.* The simplest proof of Theorem 1.1.19 relies on the operator interpretation of correlations, see e.g. the proof of [47, Theorem 1]. Here however I shall give a proof based on decomposing functions of several variables into telescopic sums, for this kind of arguments will be used again in the proofs of the more general tensorization theorems of § 3.

First, observe that the "$\geqslant$" inequality (47) is trivial, so we only have to prove the "$\leqslant$" inequality. We denote $\varepsilon_i := \{X_i : Y_i\}$, and to alleviate notation, $x_i$ will implictly stand for an element in the range of $X_i$, resp. $y_i$ for an element in the range of $Y_i$.

By Proposition 1.1.12, we may assume that $I$ is finite, say $I = \{1,\dots,N\}$ for some $N \in \mathbb{N}$. Let $f$ and $g$ be resp. $\vec{X}_I$-measurable and $\vec{Y}_I$-measurable centered $L^2$ real functions; our goal is to bound above $|\mathbf{E}[fg]|$.

For $i \in \{0,\dots,N\}$, define $\mathscr{F}_i = \bigvee_{j\leqslant i}\sigma(X_j,Y_j)$. I claim that, because of the independence hypothesis, $f^{\mathscr{F}_i}$ only depends on the values of $X_1,\dots,X_i$ and not on $Y_1,\dots,Y_i$, and similarly that $g^{\mathscr{F}_i}$ only depends on the values of $Y_1,\dots,Y_i$: one can write indeed (in the case of $f$):

$$\begin{aligned} f^{\mathscr{F}_i}(x_1,y_1,\dots,x_i,y_i) &= \int f(x_1,\dots,x_i,x_{i+1},\dots,x_n)\,d\mathbf{P}[x_{i+1},\dots,x_n|x_1,y_1,\dots,x_i,y_i] \\ &= \int f(x_1,\dots,x_i,x_{i+1},\dots,x_n)\,d\mathbf{P}[x_{i+1},\dots,x_n]. \end{aligned} \tag{48}$$

[†]The converse is not true: it can occur that $\{\mathscr{F}:\mathscr{G}\} = 1$ while no non-trivial events of $\mathscr{F}$ and $\mathscr{G}$ are equivalent. A counterexample is the following: let $(X_n)_{n\in\mathbb{N}}$ be independent *Bernoulli*(1/2) variables, and define independently $Y_n = 1-X_n$ with probability $\varepsilon_n$ and $Y_n = X_n$ otherwise, where $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of numbers such that $0 < \varepsilon_n \leqslant 1/2$ for all $n$ and $\varepsilon_n \xrightarrow{n\to\infty} 0$. Then the vectorial variables $\vec{X}$ and $\vec{Y}$ obviously satisfy $\{\vec{X}:\vec{Y}\} = 1$, but it is not hard to prove that no $\vec{X}$-measurable non-trivial event is equivalent to a $\vec{Y}$-measurable one.

Now, for $i \in \{1,\dots,N\}$, define

$$f_i(x_1,\dots,x_i) = f^{\mathscr{F}_i}(x_1,\dots,x_i) - \mathbf{E}[f|x_1,\dots,x_{i-1}], \tag{49}$$

with a similar definition for $g$. One has $f = \sum_i f_i$, resp. $g = \sum_i g_i$, and $f_i$ and $g_i$ are $\mathscr{F}_i$-measurable and centered w.r.t. $\mathscr{F}_{i-1}$ (that is, $(f_i)^{\mathscr{F}_{i-1}}, (g_i)^{\mathscr{F}_{i-1}} \equiv 0$), so

$$\operatorname{Var} f = \sum_i \operatorname{Var} f_i, \tag{50}$$

resp. $\operatorname{Var} g = \sum_i \operatorname{Var} g_i$.

We expand:

$$\mathbf{E}[fg] = \sum_{(i,j)\in I\times I} \mathbf{E}[f_i g_j]. \tag{51}$$

In the right-hand side of (51), if $i \neq j$ then $\mathbf{E}[f_i g_j] = 0$ since if, say, $i < j$, $f_i$ is $\mathscr{F}_i$-measurable while $g_j$ is centered w.r.t. $\mathscr{F}_{j-1} \supset \mathscr{F}_i$, so (51) turns into:

$$\mathbf{E}[fg] = \sum_{i\in I} \mathbf{E}[f_i g_i]. \tag{52}$$

Writing the law of total expectation,

$$\mathbf{E}[f_i g_i] = \int \mathbf{E}[f_i g_i | x_1, y_1, \dots, x_{i-1}, y_{i-1}] d\mathbf{P}[x_1, y_1, \dots, x_{i-1}, y_{i-1}]. \tag{53}$$

But, as we noticed before, under $\mathbf{P}[\cdot|x_1,y_1,\dots,x_{i-1},y_{i-1}]$, $f_i$ only depends on $X_i$ and $g_i$ only depends on $Y_i$. Moreover, because of the independence property the law of $(X_i, Y_i)$ is the same under $\mathbf{P}[\cdot|x_1,y_1,\dots,x_{i-1},y_{i-1}]$ as under $\mathbf{P}$, so under $\mathbf{P}[\cdot|x_1,y_1,\dots,x_{i-1},y_{i-1}]$, $f_i$ and $g_i$ are centered and $\varepsilon_i$-independent. Thus

$$\begin{aligned}
|\mathbf{E}[f_i g_i]| &\leqslant \varepsilon_i \int \operatorname{Sd}(f_i|x_1,y_1,\dots,x_{i-1},y_{i-1}) \operatorname{Sd}(g_i|x_1,y_1,\dots,x_{i-1},y_{i-1}) d\mathbf{P}[x_1,y_1,\dots,x_{i-1},y_{i-1}] \\
&\underset{\text{CS}}{\leqslant} \varepsilon_i \sqrt{\int \operatorname{Var}(f_i|x_1,y_1,\dots,x_{i-1},y_{i-1}) d\mathbf{P}[x_1,y_1,\dots,x_{i-1},y_{i-1}]} \sqrt{\textit{the same for } g} \\
&= \varepsilon_i \operatorname{Sd}(f_i) \operatorname{Sd}(g_i),
\end{aligned} \tag{54}$$

the last equality following from the fact that $f_i$ and $g_i$ are centered w.r.t. $\mathscr{F}_{i-1}$. Summing over $i$,

$$\begin{aligned}
|\mathbf{E}[fg]| &\leqslant \sum_{i\in I} \varepsilon_i \operatorname{Sd}(f_i) \operatorname{Sd}(g_i) \leqslant \sup_{i\in I} \varepsilon_i \cdot \sum_{i\in I} \operatorname{Sd}(f_i) \operatorname{Sd}(g_i) \\
&\underset{\text{CS}}{\leqslant} \sup_{i\in I} \varepsilon_i \cdot \sqrt{\sum_{i\in I} \operatorname{Var}(f_i)} \sqrt{\sum_{i\in I} \operatorname{Var}(g_i)} = \sup_{i\in I} \varepsilon_i \cdot \operatorname{Sd}(f) \operatorname{Sd}(g),
\end{aligned} \tag{55}$$

which is the desired bound. □

## 1.2 Examples

### 1.2.a Finite-ranged variables

**1.2.1 Proposition.** *Let $X, Y$ be random variables with finite ranges resp. $\{1,\dots,N\}$ and $\{1,\dots,M\}$, and denote $p_a := \mathbf{P}[X = a], p^b := \mathbf{P}[Y = b], p_a^b := \mathbf{P}[X = a$ and $Y = b]$.*

*Then* $\{X:Y\} = |||\Pi|||$, *where* $\Pi$ *is the* $N\times M$ *matrix with general entry*

$$\Pi_{ab} = \frac{p_a^b - p_a p^b}{\sqrt{p_a p^b}}. \tag{56}$$

*1.2.2 Remark.* In particular, if both $X$ and $Y$ have range $\{1,2\}$, using the same notation as before, one has

$$\{X:Y\} = \frac{|p_a^b - p_a p^b|}{\sqrt{p_1 p_2 p^1 p^2}}, \tag{57}$$

where the right-hand side of (57) does not depend on the choice of $a,b\in\{1,2\}$.

*Proof of Proposition 1.2.1.* By Proposition 1.1.9, $\{X:Y\}$ is the norm of the operator $\pi_{XY}:\ \bar{L}^2(Y)\to\bar{L}^2(X)$. Here it will be more convenient to work in $L^2$ spaces than in $\bar{L}^2$ spaces, so we rather compute the norm of

$$\begin{array}{rccc} \tilde{\pi}: & L^2(Y) & \to & L^2(X) \\ & g & \mapsto & g^X - \mathbf{E}[g], \end{array} \tag{58}$$

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

A function $g\in L^2(Y)$ can be identified with a $M$-dimensional vector also denoted by $g$, and similarly $\tilde{\pi}g\in L^2(X)$ can be identified with a $N$-dimensional vector. Denote $P := ((p_a^b))_{a,b}\in\mathbb{R}^{N\times M}$, $I_X := ((\delta_{aa'}p_a))_{a,a'}\in\mathbb{R}^{N\times N}$, $I_Y := ((\delta_{bb'}p^b))_{b,b'}\in\mathbb{R}^{M\times M}$, $1_N := 1^{\{1,\dots,N\}}\in\mathbb{R}^N$. Applying Bayes' formula yields that

$$\tilde{\pi}g = I_X^{-1}Pg - 1_N(1_N)^\top Pg. \tag{59}$$

Now, $\|g\|_{L^2(Y)} = \|I_Y^{1/2}g\|$, resp. $\|\tilde{\pi}g\|_{L^2(X)} = \|I_X^{1/2}(\tilde{\pi}g)\|$, so:

$$\{X:Y\} = \sup_{g\neq 0}\frac{|||\,(I_X^{-1/2}P - I_X^{1/2}1_N(1_N)^\top P)g\,|||}{\|\,I_Y^{1/2}g\,\|}. \tag{60}$$

Performing the change of variables $h = I_Y^{1/2}g$, (60) becomes $\{X:Y\} = \sup_{h\neq 0}\|\Pi h\|/\|h\| = |||\Pi|||$, with

$$\Pi = I_X^{-1/2}PI_Y^{-1/2} - I_X^{1/2}1_N(1_N)^\top PI_Y^{-1/2}, \tag{61}$$

which is Equation (56) indeed. □

*1.2.3 Remark.* With the same kind of proof, one has even a similar proposition to calculate $\{X,Y\}$ if *either* $X$ or $Y$ has finite range, provided you know (in the case it is $X$ which has finite range) all the $\mathbf{P}[X=x]$ and all the

$$\int_y \frac{d\mathbf{P}[Y=y|X=x]\,d\mathbf{P}[Y=y|X=x']}{d\mathbf{P}[Y=y]}. \tag{62}$$

*1.2.4 Remark.* In the case $X$ or $Y$ has range of cardinality 2, applying Proposition (1.2.1) yields that $\{X:Y\}^2$ depends smoothly on $Law(X,Y)$. Yet this is not the case in general: in fact, maximal correlations are nothing more than a particular case of operator norms (cf. § 1.1.c), and thus they have the same behaviour—they are a continuous function of the parameters, but they have some $\mathscr{C}^1$ singularity. The following example exhibits such a singularity.

*1.2.5 Example.* Suppose both $X$ and $Y$ have range $\{1,2,3\}$ and

$$((\mathbf{P}[X=a \text{ and } Y=b]))_{a,b} = \begin{pmatrix} 2/9 & 1/18 & 1/18 \\ 1/18 & 2/9+\alpha & 1/18-\alpha \\ 1/18 & 1/18-\alpha & 2/9+\alpha \end{pmatrix} \tag{63}$$

for a parameter $\alpha \in [-2/9, 1/18]$. Then the matrix $\Pi$ defined by (56) is

$$\Pi = \begin{pmatrix} 1/3 & -1/6 & -1/6 \\ -1/6 & 1/3+3\alpha & -1/6-3\alpha \\ -1/6 & -1/6-3\alpha & 1/3+3\alpha \end{pmatrix} = U \begin{pmatrix} 1/2+6\alpha & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix} U^{-1}, \tag{64}$$

with

$$U = \begin{pmatrix} 0 & -2/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \end{pmatrix} \tag{65}$$

being orthogonal. So by Proposition 1.2.1, $\{X:Y\} = 1/2 + 6\alpha_+$.

### 1.2.b Gaussian variables

The following theorem, which I will frequently use in the sequel, computes exactly the maximal correlation between two jointly Gaussian variables:

**1.2.6 Theorem** ([27, 26])**.** *Let $(\vec{X},\vec{Y})$ be an $(N+M)$-dimensional Gaussian vector whose covariance matrix writes blockwise*

$$\operatorname{Var}(\vec{X},\vec{Y}) = \begin{pmatrix} \mathbf{I}_N & C \\ C^\top & \mathbf{I}_M \end{pmatrix}, \tag{66}$$

*then $\{\vec{X}:\vec{Y}\} = |||C|||$.*

*1.2.7 Remark.* In other words, Theorem 1.2.6 tells that in the Gaussian case, the supremum in (34) defining $\{\vec{X}:\vec{Y}\}$ can be restricted to *linear* functions $f$ and $g$.

*1.2.8 Remark.* By a linear change of variables, Theorem 1.2.6 actually allows us to compute $\{\vec{X}:\vec{Y}\}$ for *any* Gaussian vector $(\vec{X},\vec{Y})$.

*Proof of Theorem 1.2.6.* I recall (a sketch of) the proof for the sake of completeness. By the properties of Gaussian vectors, the law of $Y$ knowing that $X=x$ [I dropped the vector arrows] is the normal law $\mathcal{N}(\mathbf{I}_M - C^\top C) + C^\top x$, and similarly the law of $X$ knowing that $Y=y$ is the normal law $\mathcal{N}(\mathbf{I}_N - CC^\top) + Cy$. Consequently, the operator $\pi_{XYX}$ is the generator of the following random walk on $\mathbb{R}^N$ (whose equilibrium measure is the standard Gaussian law): when one is at $x$, they jump to a point distributed according to the normal law $\mathcal{N}(\mathbf{I}_N - CC^\top CC^\top) + CC^\top x$. This walk is a multidimensional AR(1)-process (see. [34, § 2.6]), whose properties are perfectly known; in particular, the eigenvalue $f$ of $\pi_{XYX}$ responsible for its spectral radius will be a linear function, so we only have to consider linear $f$ in the supremum (34). For such $f$, the optimal $g$ will also be linear by the Gaussian nature of the system, so in the end $\{\vec{X}:\vec{Y}\}$ is equal to $|||C|||$. $\square$

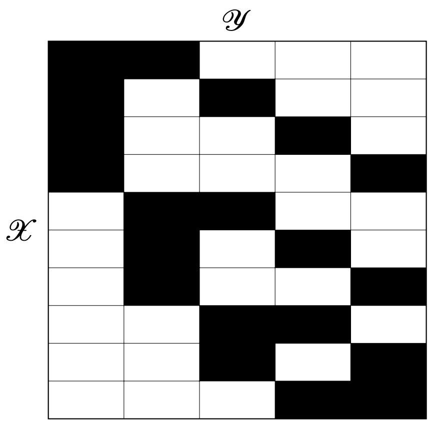



Figure 1: Schematic representation of Example 1.2.9 for $n = 5$ and $p = 2$.

### 1.2.c Miscellaneous examples

**Random conditional laws**

*1.2.9 Example.* Let $0 < p < n$ be integers. $(X,Y)$ is a random variable such that $Y$ has range $\mathscr{Y} := \{1,\dots,n\}$ and $X$ has range $\mathscr{X} := \mathfrak{P}_p(\mathscr{Y})$, the set of subsets $y \subset \mathscr{Y}$ with cardinality $p$ —so, $\#\mathscr{X} = \binom{n}{p}$ and $\#\mathscr{Y} = n$ —, and we take the law of $(X,Y)$ uniform on the pairs $(x,y)$ such that $y \in x$: see Figure 1.

When considered as operators on $L^2$ spaces, it is obvious that $\pi_{XY}$ and $\pi_{YX}$ are characterized by $(\pi_{XY} f)(x) = p^{-1} \sum_{y \in x} f(y)$, resp. $(\pi_{YX} g)(y) = \binom{n-1}{p-1}^{-1} \sum_{y \in x} g(x)$, so that

$$(\pi_{YXY} f)(y) = \frac{1}{p} f(y) + \sum_{y' \neq y} \frac{p-1}{p(n-1)} f(y'). \tag{67}$$

Thus, on $\bar{L}^2(Y)$, $\pi_{YXY}$ is nothing but the scalar operator $\frac{n-p}{p(n-1)}\mathbf{I}$, and therefore

$$\{X : Y\} = \sqrt{\frac{n-p}{p(n-1)}} \tag{68}$$

by Proposition 1.1.9 and Remark 1.1.10.

**Weakly coupled particles**

**1.2.10 Proposition.** *Let $V_1$ and $V_2$ be potentials on $\mathbb{R}^n$, $n \geqslant 1$, i.e. the $V_i$ are real-valued measurable functions on $\mathbb{R}^n$ with $\int_{\mathbb{R}^n} e^{-V_i(x)}\,dx < \infty$. For $i \in \{1,2\}$, denote by $\mathbf{P}_i$ the probability measure on $\mathbb{R}^n$ proportional to $e^{-V_i(x)}dx$, which is to be thought as the law of the position $X_i$ of a particle $i$ subjected to the potential $V_i$. Denote $\mathbf{P}_\otimes := \mathbf{P}_1 \otimes \mathbf{P}_2$, which is the joint law of $(X_1,X_2)$ in absence of interaction.*

*Now let $W$ be an interaction potential on $(\mathbb{R}^n)^2$, such that $e^{-[V_1(x_1)+V_2(x_2)+W(x_1,x_2)]}$ is integrable; denote by $\mathbf{P}$ the probability measure on $(\mathbb{R}^n)^2$ proportional to $e^{-[V_1+V_2+W]}dx_1dx_2$, which is the joint law of $(X_1,X_2)$ in presence of interaction potential $W$.*

*Then, under the law $\mathbf{P}$,*

$$\{X_1 : X_2\} \leqslant \frac{\mathrm{Sd}_\otimes(e^{-W})}{\mathbf{E}_\otimes[e^{-W}]}. \tag{69}$$

*Proof.* The law $\mathbf{P}$ has density $h = e^{-W}/\mathbf{E}_{\otimes}[e^{-W}]$ w.r.t. $\mathbf{P}_{\otimes}$, whence the result by Proposition 1.1.16.
□

*1.2.11 Remark.* Proposition 1.2.10 gives a rigorous sense to the intuition that two weakly coupled particles must have nearly independent positions. This is valid in a quite general setting, in particular, $W$ does not have to be bounded.

**Non-reversible Markov chain**

*1.2.12 Example.* Here is a example showing that the inequality in Proposition 1.1.13 is strict in general. Consider the stationary Markov chain on $\{1,2,3\}$ defined by

$$P = ((\mathbf{P}[X_{k+1} = a | X_k = b]))_{ab} = \begin{pmatrix} 0 & 1/2 & 1 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix}, \tag{70}$$

which has equilibrium measure $(2/5, 2/5, 1/5)$. Diagonalizing $P$ shows that

$$P^t = \begin{pmatrix} 2/5 & 2/5 & 2/5 \\ 2/5 & 2/5 & 2/5 \\ 1/5 & 1/5 & 1/5 \end{pmatrix} + O(2^{-t/2}), \tag{71}$$

whence $\{X_k : X_{k+t}\} = O(2^{-t/2})$ when $t \to +\infty$ by Proposition 1.2.1. Yet $\{X_k : X_{k+1}\} = 1$, since the non-trivial events $\{X_k = 1\}$ and $\{X_{k+1} = 2\}$ are equivalent (cf. Proposition 1.1.18).

**Hyperplanes in Ising's model** As I told in Chapter 0, the initial motivation of this monograph was to understand the presence $\rho$-mixing in Ising's model (cf. § 0.1.a); in particular, I intended to re-get a result similar to Theorem 0.1.9 by a more 'natural' method. That will be achieved indeed in § 5.1:

**1.2.13 Theorem** (Theorem 5.1.1-(i))**.** *For Ising's model on $\mathbb{Z}^n$ in the completely analytical regime, for all disjoint $I, J \subset \mathbb{Z}^n$,*

$$\{\vec{\omega}_I : \vec{\omega}_J\} \leqslant \exp\big[-(\psi' + o(1)) dist(I,J)\big], \tag{72}$$

*where $\psi'$ is the same as in Theorem 0.1.7 and where the "$o(1)$" (to be understood "as $dist(I,J) \to \infty$") is uniform in $I, J$.*

If we apply this result to the case of parallel 'hyperplanes' of $\mathbb{Z}^n$ (I mean, sets of the form $\{t\} \times \mathbb{Z}^{n-1}$), Formula (72) looks far less neat than Formula (14) in Theorem 0.1.9.

This bound can however be improved by using Proposition 1.1.14. Indeed, as we noticed in § 0.1.c, the states of two parallel hyperplanes are elements of some reversible stationary Markov chain. Then, applying Corollary 1.1.15 (in which we let $k \to \infty$), we get a result exactly similar to (14), except that we have to replace $\psi$ by $\psi'$ —recall that it is not known whether $\psi' = \psi$.

## 1.3 Comparing $\rho$-mixing to other measures of dependence

The material of this section is classical; most of it can be found for instance in [8, §§ 3 & 5]. Here we will say that a sequence of pairs $(\mathscr{F}^n, \mathscr{G}^n)$ of $\sigma$-algebras is *$\rho$-mixing* to mean that $\{\mathscr{F}^n : \mathscr{G}^n\} \overset{n\to\infty}{\to} 0$.

### 1.3.a $\alpha$-mixing

**1.3.1 Definition.** The *$\alpha$-mixing* coefficient of two $\sigma$-algebras $\mathscr{F}$ and $\mathscr{G}$ is

$$\alpha(\mathscr{F}, \mathscr{G}) := \sup_{\substack{A\in\mathscr{A} \\ B\in\mathscr{B}}} |\mathbf{P}[A\cap B] - \mathbf{P}[A]\mathbf{P}[B]|. \tag{73}$$

Proposition 1.1.18 shows that '$\rho$-mixing implies $\alpha$-mixing', in the sense that one has $\alpha(\mathscr{F}, \mathscr{G}) \leqslant A(\{\mathscr{F} : \mathscr{G}\})$ for some universal function $A\colon\ [0,1] \to [0,1]$ with $A(\rho) \overset{\rho\to 0}{\to} 0$. The following example shows that the converse is not true:

*1.3.2 Example.* For $\varepsilon \in (0, 1/2]$, define $(X^\varepsilon, Y^\varepsilon)$ in the following way:

- With probability $\varepsilon$, one samples $X^\varepsilon$ and $Y^\varepsilon$ independently with common uniform law on $[0, \varepsilon]$;
- With probability $(1-\varepsilon)$, one samples $X^\varepsilon$ and $Y^\varepsilon$ independently with common law uniform on $[\varepsilon, 1]$.

Then for all $\varepsilon > 0$ one has $\{X^\varepsilon : Y^\varepsilon\} = 1$, since the non-trivial events $\{X^\varepsilon \leqslant \varepsilon\}$ and $\{Y^\varepsilon \leqslant \varepsilon\}$ are equivalent (cf. Proposition 1.1.18). However is is easy to show that $\alpha(X^\varepsilon, Y^\varepsilon) = \varepsilon - \varepsilon^2 \overset{\varepsilon\to 0}{\to} 0$.

*1.3.3 Remark.* Saying that the correlation of two variables tends to 0 means that their joint law tends in some sense to the product law. When variables take place in a Polish space, a common notion of convergence is weak convergence, that is, convergence against all bounded continuous function. [3, Theorem 2.2] states that weak convergence is implied by $\alpha$-mixing, hence by $\rho$-mixing. The precise statement is the following: if $(X^n, Y^n)_{n\in\mathbb{N}}$ is a sequence of pairs of random variables such that all the $X_n$ (resp. $Y_n$) have the same law $Law(X)$ (resp. $Law(Y)$) in some Polish space $E$ (resp. $F$), then $\big(\alpha(X^n, Y^n) \overset{n\to\infty}{\to} 0\big) \Rightarrow \big(Law(X^n, Y^n) \overset{n\to\infty}{\rightharpoonup} Law(X) \otimes Law(Y)\big)$.

### 1.3.b $\beta$-mixing

Recall the definition of the $\beta$-mixing coefficient from the previous chapter [Definition 0.1.4].

*1.3.4 Example.* For $\varepsilon \in (0, 1)$, consider two random sequences $(X_i)_{i\in\mathbb{N}}$ and $(Y_i)_{i\in\mathbb{N}}$ defined in the following way: $(X_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. variables with uniform law on $\{\pm 1\}$, and for each $i \in \mathbb{N}$, independently, one sets $Y_i = X_i$ with probability $\varepsilon$, and with probability $(1-\varepsilon)$ one chooses $Y_i$ uniformly on $\{\pm 1\}$. Then all the $(X_i, Y_i)$ are i.i.d. with $\mathbf{P}[X_i = \eta \text{ and } Y_i = \theta] = (1 + \eta\theta\varepsilon)/4$ for all $\eta, \theta \in \{\pm 1\}$, thus $\{X_i : Y_i\} = \varepsilon$ by Remark 1.2.2, whence $\{\vec{X}:\vec{Y}\}=\varepsilon$ by Theorem 1.1.19. Yet $Law(\vec{X},\vec{Y})$ and $Law(\vec{X})\otimes Law(\vec{Y})$ are mutually singular for all $\varepsilon>0$.

This shows that $\rho$-mixing does not imply $\beta$-mixing, and *a fortiori* that there can be no kind of converse to Proposition 1.1.16.

### 1.3.c Mutual information

Recall the definition (41) of mutual information. [8, Theorem 5.3(III)] states that mutual information controls the $\beta$-mixing coefficient, so Example 1.3.4, which shows that $\rho$-mixing does not imply $\beta$-mixing in general, shows that it does not imply mutual information to tend to 0 either.

Proposition 1.1.16 suggests that, on the other hand, maximal correlation could be controlled by mutual information, but it is not true either: in Example 1.3.2 indeed, $\{X^\varepsilon:Y^\varepsilon\}=1$ for all $\varepsilon>0$, but

$$I(X^\varepsilon;Y^\varepsilon)=\varepsilon\log(\varepsilon^{-1})+(1-\varepsilon)\log\big((1-\varepsilon)^{-1}\big)\overset{\varepsilon\to 0}{\to}0. \tag{74}$$

Mutual information measures the quantity of information shared by two random variables, which explains intuitively the following property ([12, Theorem 2.5.2]): if $X\to Y\to Z$ is a Markov chain, then $I(Y;X,Z)\leqslant I(X;Y)+I(Y;Z)$. Does a similar inequality hold for Hilbertian correlations? In the Gaussian case, the answer is "yes" thanks to Theorem 1.2.6: one gets that

$$\{Y:X,Z\}^2\leqslant 1-\frac{(1-\{X:Y\}^2)(1-\{Y:Z\}^2)}{1-\{X:Y\}^2\{Y:Z\}^2}\leqslant\{X:Y\}^2+\{Y:Z\}^2. \tag{75}$$

But that property does not hold in general, as the following example shows:

*1.3.5 Example.* Consider a Markov chain $X\to Y\to Z$, where $(Y,X)$ and $(Y,Z)$ have the same law, which is the joint law described in Example 1.2.9—the role of "$Y$" in that example being played here by $Y$ in both cases. Fix $y\in\mathscr{Y}$; define event $A$ as "$Y=y$" and event $B$ as "$y\in X\cap Z$". Then one computes that $\mathbf{P}[A]=n^{-1}$, while

$$\mathbf{P}[B]=\frac{1}{n}+\frac{(p-1)^2}{n(n-1)}. \tag{76}$$

Since $A\subset B$, Proposition 1.1.18 then yields that

$$\{Y:X,Z\}\geqslant\sqrt{\mathbf{P}[B^c]\mathbf{P}[A]/\mathbf{P}[A^c]\mathbf{P}[B]}=\left(\frac{(n-1)^2-(p-1)^2}{(n-1)^2+(n-1)(p-1)^2}\right)^{1/2}. \tag{77}$$

Comparing (68) and (77), one sees that taking $n\gg 1$ and $1\ll p\ll n^{1/2}$ makes $\{X:Y\}$ and $\{Y:Z\}$ arbitrarily close to 0 while $\{Y:X,Z\}$ gets arbitrarily close to 1.

*1.3.6 Remark.* Lack of stationarity or reversibility has nothing to do with Example 1.3.5: there are indeed similar examples, though more complicated, in the case $X,Y,Z$ are three successive states of some reversible continuous-time Markov process [36].

# Chapter 2

# Event sufficient conditions

In § 1.1.d we saw that the maximal correlation coefficient $\{\mathscr{F}:\mathscr{G}\}$ controls the difference between $\mathbf{P}[A\cap B]$ and $\mathbf{P}[A]\mathbf{P}[B]$ for $A$ and $B$ two events resp. $\mathscr{F}$- and $\mathscr{G}$-measurable. A natural question is whether the opposite is true, i.e. whether saying that $\mathbf{P}[A\cap B]$ is always close in some sense to $\mathbf{P}[A]\mathbf{P}[B]$ implies a control on $\{\mathscr{F}:\mathscr{G}\}$. We saw in § 1.3.a that $\alpha$-mixing does not fit, but maybe stronger conditions of the same type would work.

In § 2.1 I will present a simple such condition [Theorem 2.1.3]. This condition demands $|\mathbf{P}[A\cap B]-\mathbf{P}[A]\mathbf{P}[B]|$ to be bounded uniformly by $\zeta(\mathbf{P}[A])\theta(\mathbf{P}[B])$ for functions $\zeta,\theta\colon [0,1]\to\mathbb{R}_+$ sufficiently well behaved. This result, whose proof is rather simple, is apparently new.

Proposition 1.1.18, however, suggests that the natural condition on events would be a uniform control on $|\mathbf{P}[A\cap B]-\mathbf{P}[A]\mathbf{P}[B]|/\sqrt{\mathbf{P}[A^c]\mathbf{P}[A]}\sqrt{\mathbf{P}[B^c]\mathbf{P}[B]}$, which is out of the scope of Theorem 2.1.3. Bradley [5] proved in 1983 that that condition was indeed sufficient to get $\rho$-mixing. His result was improved in the next few years (see for instance the bound of [10]), but the optimal bound remained unknown, though its value was conjectured. In § 2.2, I will prove this optimal bound. My method, different from the techniques of [5, 10], relies on the analysis of the spectral properties of a Markov process which I call the "Chogosov process", whose study is proceeded to in § 2.2.b.

## 2.1 Weak event sufficient condition

To state our next result we need some functional analysis reminders first:

**2.1.1 Definition.** On the space $\mathscr{C}_0^\infty(0,1)$ of compactly supported fuctions of $\mathscr{C}^\infty(0,1)$, one defines the scalar product

$$\langle\varphi,\psi\rangle_{H_0^1}=\int_0^1\varphi'(x)\psi'(x)\,dx. \tag{78}$$

$\mathscr{C}_0^\infty(0,1)$ endowed with $\langle\cdot,\cdot\rangle_{H_0^1}$ is a prehilbertian space; its completion is denoted by $H_0^1(0,1)$.

Recall that elements of $H_0^1(0,1)$ may be seen as ordinary functions:

**2.1.2 Lemma** (Sobolev, [1, Theorem 4.12])**.** *Any element $f \in H_0^1(0,1)$ can be identified with a unique function $\bar{f} \in \mathscr{C}_0^0[0,1]$, the space of continuous functions on $[0,1]$ with $\bar{f}(0), \bar{f}(1) = 0$. Conversely, a function $\bar{f} \in \mathscr{C}_0^0[0,1]$ corresponds to an element of $H_0^1(0,1)$ if and only if*

$$\sup_{g \in \mathscr{C}_0^\infty(0,1)} \frac{|\int_0^1 f(x) g''(x) dx|}{\sqrt{\int_0^1 g'(x)^2 dx}} \tag{79}$$

*is finite, and then there is a unique $f \in H_0^1(0,1)$ associated to $\bar{f}$, whose norm is (79).*

In accordance with Lemma 2.1.2, we will identify functions of $\mathscr{C}_0^0[0,1]$ with elements of $H_0^1(0,1)$ whenever it is possible. If $f \in \mathscr{C}_0^0[0,1]$ does not correspond to an element of $H_0^1(0,1)$, then we set $\|f\|_{H_0^1} = +\infty$.

Now we can state the

**2.1.3 Theorem** (Weak event sufficient condition)**.** *Let $\mathscr{F}$ and $\mathscr{G}$ be two $\sigma$-algebras such that, for all $A \in \mathscr{F}$ and $B \in \mathscr{G}$ with respective probabilities $p$ and $q$,*

$$\mathbf{P}[A \cap B] - pq \leqslant \zeta(p)\theta(q)^{[*]} \tag{80}$$

*for some $\zeta, \theta \in \mathscr{C}_0^0[0,1]$. Then,*

$$\{\mathscr{F} : \mathscr{G}\} \leqslant \|\zeta\|_{H_0^1} \|\theta\|_{H_0^1}. \tag{81}$$

*Proof.* We begin with the following formula for covariance:

**2.1.4 Lemma.** *For $f$ and $g$ two real $L^2$ functions,*

$$\mathrm{Cov}(f,g) = \int_{\mathbb{R}\times\mathbb{R}} \big(\mathbf{P}[f \leqslant x \text{ and } g \leqslant y] - \mathbf{P}[f \leqslant x]\mathbf{P}[g \leqslant y]\big)\, dx dy. \tag{82}$$

*Proof of Lemma 2.1.4.* Suppose in a first time that $f$ and $g$ are nonnegative. A classical Fubini argument (see [4, Problem 21.6]) shows that

$$\mathbf{E}[f] = \int_{\mathbb{R}_+} \mathbf{P}[f > x]\, dx, \tag{83}$$

with a similar formula for $g$. By the same method,

$$\mathbf{E}[fg] = \int_{\mathbb{R}_+\times\mathbb{R}_+} \mathbf{P}[f > x \text{ and } g > y]\, dx dy, \tag{84}$$

so that, using the computational formula $\mathrm{Cov}(f,g) = \mathbf{E}[fg] - \mathbf{E}[f]\mathbf{E}[g]$,

$$\mathrm{Cov}(f,g) = \int_{\mathbb{R}_+\times\mathbb{R}_+} \big(\mathbf{P}[f > x \text{ and } g < y] - \mathbf{P}[f > x]\mathbf{P}[g > y]\big)\, dx dy. \tag{85}$$

Observing that the integrand is also $(\mathbf{P}[f \leqslant x \text{ and } g \leqslant y] - \mathbf{P}[f \leqslant x]\mathbf{P}[g \leqslant y])$ and that it is zero for $(x,y) \notin \mathbb{R}_+ \times \mathbb{R}_+$, we get (82) in the nonnegative case. By translation invariance, the formula remains true for all $f, g$ bounded below, and then by approximation for all $f, g \in L^2$. □

[*]Note that there is no need to put absolute values in the left-hand side of (80).

Now, let $f$ and $g$ be $L^2$ variables resp. $\mathscr{F}$- and $\mathscr{G}$-mesurable, and denote by $F$ and $G$ the respective distribution functions of $f$ and $g$. Up to a slight perturbation, $F$ and $G$ may be supposed to be diffeomorphisms from $\mathbb{R}$ onto $(0,1)$; denote by $\alpha$ and $\beta$ their respective inverse maps. Then a change of variables in (82) yields:

$$\operatorname{Cov}(f,g) = \int_{(0,1)^2} \big(\mathbf{P}[f \leqslant \alpha(p) \text{ and } g \leqslant \beta(q)] - pq\big)\alpha'(p)\beta'(q)\,dp\,dq, \tag{86}$$

so by assumption (80):

$$\operatorname{Cov}(f,g) \leqslant \Big(\int_0^1 \zeta(p)\alpha'(p)dp\Big)\cdot\Big(\int_0^1 \theta(q)\beta'(q)dq\Big). \tag{87}$$

Then our theorem becomes equivalent to the claim stated and proved just below. □

**2.1.5 Claim.** *If $f$ is a random variable whose repartition function $F$ is a diffeomorphism of inverse $\alpha$, then for $\zeta \in \mathscr{C}_0^0[0,1]$:*

$$\Big|\int_0^1 \zeta(p)\alpha'(p)dp\Big| \leqslant \|\zeta\|_{H_0^1} \operatorname{Sd}(f). \tag{88}$$

*Proof.* First note that, replacing $g$ by $f$ in (86), one has:

$$\operatorname{Var}(f) = \int_{(0,1)^2} [p(1-q) \wedge q(1-p)]\alpha'(p)\alpha'(q)\,dp\,dq. \tag{89}$$

In fact, one can define a scalar product[†] $\langle\cdot,\cdot\rangle_V$ on $\mathscr{C}^0(0,1)$ by setting

$$\langle\varphi,\psi\rangle_V = \int_{(0,1)^2} [p(1-q) \wedge q(1-p)]\varphi(p)\psi(q)\,dp\,dq, \tag{90}$$

so that if $\alpha$ is the inverse distribution fuction of a variable $f$, $\operatorname{Var}(f) = \|\alpha'\|_V^2$.

So, we are considering three scalar products on some subspaces of $\mathscr{C}^0(0,1)$: the ordinary $L^2$ product, which we denote by $\langle\cdot,\cdot\rangle_{L^2}$, the $H_0^1(0,1)$ product $\langle\cdot,\cdot\rangle_{H_0^1}$ and the variance product $\langle\cdot,\cdot\rangle_V$. Our goal is to show that for all $\varphi \in H_0^1(0,1), \psi \in \mathscr{C}^0(0,1)$,

$$|\langle\varphi,\psi\rangle_{L^2}| \leqslant \|\varphi\|_{H_0^1}\|\psi\|_V. \tag{91}$$

By approximation we can suppose that $\varphi \in \mathscr{C}_0^\infty(0,1)$. A direct computation shows that

$$\langle\varphi,\psi\rangle_V = \langle L\varphi,\psi\rangle_{L^2}, \tag{92}$$

where the operator $L\colon\ \mathscr{C}_0^\infty(0,1) \to \mathscr{C}^2(0,1)$ is defined by:

$$(L\varphi)(x) = x\int_0^x (1-y)\varphi(y)dy + (1-x)\int_x^1 y\varphi(y)dy. \tag{93}$$

For the sequel we need a left inverse of operator $L$. We notice that for $\varphi \in \mathscr{C}_0^2(0,1)$, by double integration by parts,

$$L(-\varphi'') = \varphi. \tag{94}$$

So,

$$|\langle\varphi,\psi\rangle_{L^2}| = |\langle L(-\varphi''),\psi\rangle_{L^2}| = |\langle-\varphi'',\psi\rangle_V| \underset{\mathrm{CS}}{\leqslant} \|\varphi''\|_V\|\psi\|_V, \tag{95}$$

where

$$\|\varphi''\|_V^2 = \langle\varphi'',\varphi''\rangle_V = \langle L\varphi'',\varphi''\rangle_{L^2} = -\langle\varphi,\varphi''\rangle_{L^2} \underset{\mathrm{IP}}{=} \|\varphi'\|_{L^2}^2 = \|\varphi\|_{H_0^1}^2, \tag{96}$$

whence (91). □

[†]The positivity of $\langle\cdot,\cdot\rangle_V$ follows from the identity: $\|\varphi\|_V^2 = \int_{p<q}\big(\int_p^q \varphi(r)dr\big)^2 dp\,dq$.

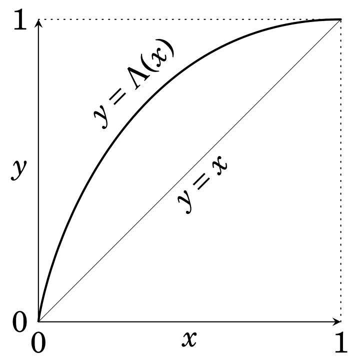



Figure 1: The function $\Lambda$.

## 2.2 Strong event sufficient condition

### 2.2.a The strong event sufficient condition

A natural choice of functions $\zeta$ and $\theta$ in Theorem 2.1.3 would be to $\zeta(p) = \theta(p) = \varepsilon^{1/2}\sqrt{p(1-p)}$, since that would give a converse to Formula (46) of Proposition 1.1.18. Unfortunately $\|\sqrt{p(1-p)}\|_{H_0^1} = +\infty$, so Theorem 2.1.3 does not work in this case. There is however a specific result then:

**2.2.1 Theorem** (Strong event sufficient condition)**.** *Let $\mathscr{F}$ and $\mathscr{G}$ be two $\sigma$-algebras such that, for all A and B resp. in $\mathscr{F}$ and $\mathscr{G}$ with respective probabilities p and q,*

$$\mathbf{P}[A \cap B] - pq \leqslant \varepsilon\sqrt{p(1-p)q(1-q)} \tag{97}$$

*for some $\varepsilon \in [0,1]$. Then*

$$\{\mathscr{F} : \mathscr{G}\} \leqslant \Lambda(\varepsilon), \tag{98}$$

*where $\Lambda\colon\ [0,1] \to \mathbb{R}_+$ is defined by*

$$\Lambda(\varepsilon) = \begin{cases} \varepsilon(1+|\log\varepsilon|) & \text{if } \varepsilon > 0, \\ 0 & \text{if } \varepsilon = 0. \end{cases} \tag{99}$$

*2.2.2 Remark.* The function $\Lambda$ is increasing on $[0,1]$ and satisfies $\Lambda(0) = 0$, $\Lambda(1) = 1$, and $\Lambda(\varepsilon) > \varepsilon$ for all $\varepsilon \in (0,1)$. Moreover it is continuous, in particular $\Lambda(\varepsilon) \searrow 0$ as $\varepsilon \searrow 0$ (see Figure 1).

*2.2.3 Remark.* I called Theorems 2.1.3 and 2.2.1 resp. “weak” and “strong” event sufficient conditions, yet that vocabulary is a bit misleading, since the strong condition does not imply the weak one *stricto sensu*: with the hypotheses of Theorem 2.1.3, Theorem 2.2.1 only implies that

$$\{\mathscr{F} : \mathscr{G}\} \leqslant \Lambda\big(\|\zeta\|_{H_0^1}\|\theta\|_{H_0^1}\big). \tag{100}$$

But the right-hand side of (100) tends to 0 as soon the right-hand side of (81) does, so it is relevant to say that Theorem 2.2.1 is ‘qualitatively stronger’ than Theorem 2.1.3.

*2.2.4 Remark.* With the same informal vocabulary, Theorem 2.2.1 is a ‘qualitative converse’ of Proposition 1.1.18: maximal decorrelation is ‘qualitatively equivalent’ to decorrelation of events as defined by Formula (46).

*2.2.5 Remark.* One can prove that the bound $\Lambda(\varepsilon)$ in (98) is the best possible: see § 2.2.c.

*Proof.* The principle of the proof is the same as for Theorem 2.1.3, except that we first perform a tricky refinement of the hypothesis: observing that, for $A$ and $B$ with respective probabilities $p$ and $q$, one has the trivially $\mathbf{P}[A \cap B] \leqslant p \wedge q$, the bound (97) can be strengthened into:

$$\mathbf{P}[A \cap B] \leqslant \left(pq + \varepsilon\sqrt{p(1-p)q(1-q)}\right) \wedge p \wedge q. \tag{101}$$

The right-hand side of (101) will be denoted by $Z_\varepsilon(p,q)$.

Now, like in the proof of Theorem 2.1.3, if (97) is satisfied, for $f$ and $g$ are $L^2$ real variables resp. $\mathscr{F}$- and $\mathscr{G}$-measurable, having respective distribution functions $F$ and $G$ with respective inverses maps $\alpha$ and $\beta$,

$$\mathrm{Cov}(f,g) \leqslant \int_{(0,1)^2} (Z_\varepsilon(p,q) - pq)\alpha'(p)\beta'(q)\,dp\,dq. \tag{102}$$

Call $\langle \alpha', \beta' \rangle_{Z_\varepsilon}$ the right-hand side of (102).

To bound $\langle \alpha', \beta' \rangle_{Z_\varepsilon}$, this time we are remaining on a random variable paradigm. Define the

**2.2.6 Definition.** The *Chogosov process*[‡] is the (unique in law) $(0,1)$-valued Markov chain $(r_i)_{i\in\mathbb{Z}}$ such that, for all $i \in \mathbb{Z}$,

$$\mathbf{P}\left(r_i \leqslant p \text{ and } r_{i+1} \leqslant q\right) = Z_\varepsilon(p,q). \tag{103}$$

It will be proved in § 2.2.b that the Chogosov process actually exists. The Chogosov process is obviously stationary with uniform equilibrium measure on $(0,1)$; let $\mathscr{L}$ denote its generator. The very definition of $\langle\cdot,\cdot\rangle_{Z_\varepsilon}$ yields:

$$\langle \varphi, \psi \rangle_{Z_\varepsilon} = \mathrm{Cov}(\alpha, \mathscr{L}\beta), \tag{104}$$

where by writing "$\mathrm{Cov}(\alpha, \mathscr{L}\beta)$" I consider functions $\alpha$ and $\mathscr{L}\beta$ as real random variables on the probability space $(0,1)$ endowed with the uniform measure.

By the Cauchy–Schwarz inequality, it is then enough to prove that $\mathrm{Var}(\mathscr{L}\beta) \leqslant \Lambda(\varepsilon)\mathrm{Var}(\beta)$, i.e. that the operator norm of $\mathscr{L}$ on $\bar{L}^2(0,1)$ is bounded above by $\Lambda(\varepsilon)$. That work is achieved by Lemma 2.2.10 in the next subsection. □

### 2.2.b The Chogosov process

This subsection deals with the "Chogosov process", which we introduced in the proof of Theorem 2.2.1.

☛ *Throughout this subsection we suppose $\varepsilon \in (0,1)$ fixed and we write $\Lambda$ for $\Lambda(\varepsilon)$, resp. $Z$ for $Z_\varepsilon$. The drawings will be made for $\varepsilon = 1/2$.*

Recall Definition 2.2.6 of the Chogosov process. To check that the Chogosov process actually exists, we have to prove the

[‡]So called in honour of my dear friend M. K. Chogosov.

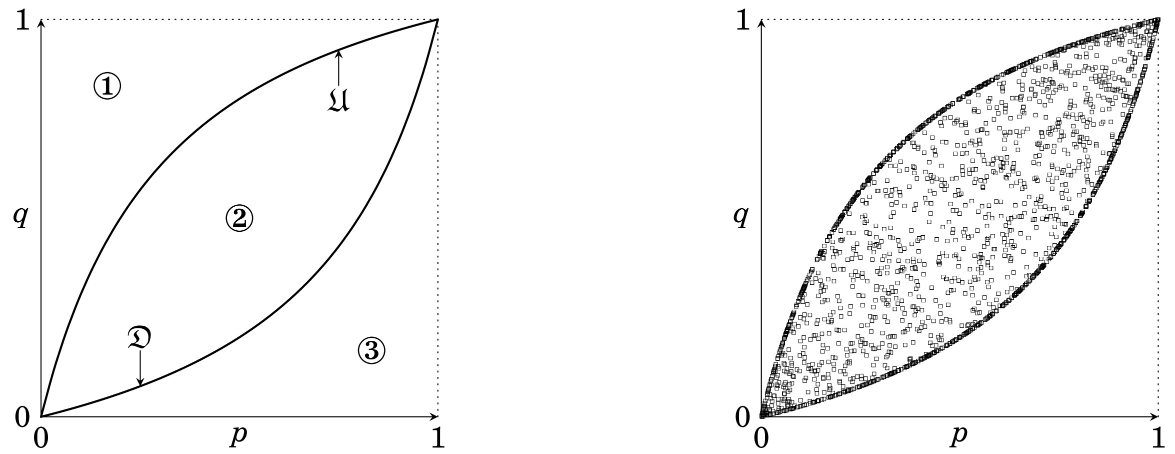



Figure 2: The measure $\mu$. On the left are drawn the different zones relative to the support of the measure; on the right is a cloud of 2,048 independent points with law $\mu$.

**2.2.7 Claim.**

*(i) There exists a (unique) probability measure $\mu$ on $(0,1)^2$ such that*

$$\forall p,q \in [0,1] \quad \mu\big[\{(x_1,x_2)\in(0,1)^2 :\ x_1 \leqslant p \text{ and } x_2 \leqslant q\}\big] = Z(p,q), \tag{105}$$

*where we recall that*

$$Z(p,q) = \big(pq + \varepsilon\sqrt{p(1-p)q(1-q)}\big) \wedge p \wedge q. \tag{106}$$

*(ii) Both marginals of $\mu$ are uniform on $(0,1)$.*

*Proof.* Provided (i) is true, (ii) is immediate since $Z(p,1) \equiv p$, resp. $Z(1,q) \equiv q$.

Concerning (i), (105) means that the density of $\mu$ on $(0,1)^2$ is equal to the distribution $\partial^2_{x_1x_2}Z$; the non-trivial point consists in proving that that distribution is nonnegative.

☛ *From now on in this subsection elements of $(0,1)^2$ will be automatically denoted by $(p,q)$. Moreover, we will denote $\bar{p} := 1-p$ and $\tilde{p} := p-1/2$, resp. $\bar{q} := 1-q$ and $\tilde{q} := q-1/2$.*

The analytic formula defining $Z(p,q)$ depends on the zone of $(0,1)^2$ in which $(p,q)$ lies (see Figure 2):

- If $p\bar{q}/q\bar{p} \leqslant \varepsilon^2$, then $Z(p,q) = p$ and we will say that we are in zone ①;
- If $\varepsilon^2 \leqslant p\bar{q}/q\bar{p} \leqslant \varepsilon^{-2}$, then $Z(p,q) = pq + \sqrt{p\bar{p}q\bar{q}}$ and we will say that we are in zone ②;
- If $\varepsilon^{-2} \leqslant p\bar{q}/q\bar{p}$, then $Z(p,q) = q$ and we will say that we are in zone ③.

So the expression of $\partial_p Z$ depends on the zone where one lies: in ① it is "1", in ② it is "$q - \varepsilon\tilde{p}\sqrt{q\bar{q}/p\bar{p}}$", and in ③ it is "0". Anyway it is defined and finite evereryewhere, just having jumps at the borders between the zones, which borders we will denote respectively $\mathfrak{U}$ for the border between ① and ②, and $\mathfrak{D}$ for the border between ② and ③ (see Figure 2). To prove that the distribution $\partial^2_{pq}Z$ is nonnegative, we have to show that $\partial_p Z$ is increasing in $q$ at $p$ fixed. Let us check it:

- In ① and ③, $\partial_p Z$ is differentiable with $\partial_q(\partial_p Z) = 0 \geqslant 0$;

- In ②, $\partial_p Z$ is differentiable with $\partial_q(\partial_p Z) = 1 + \varepsilon\tilde{p}\tilde{q}/\sqrt{p\bar{p}q\bar{q}}$. Denoting by $\rho(p,q)$ that expression, let us prove that $\rho(p,q)$ is nonnegative (and even positive) in ②: either $\tilde{p}$ and $\tilde{q}$ have the same sign and then $\rho(p,q)$ is trivially $\geqslant 1$, or $\tilde{p}$ and $\tilde{q}$ have opposite signs. In the latter case, say for instance that ($\tilde{p} \geqslant 0$ and $\tilde{q} \leqslant 0$). Then $p \geqslant 1/2$ and $q \leqslant 1/2$, so $|\tilde{p}| = p - 1/2 < p$ and $|\tilde{q}| = 1/2 - q < \bar{q}$, which implies that
$$\varepsilon \frac{|\tilde{p}\tilde{q}|}{\sqrt{p\bar{p}q\bar{q}}} < \varepsilon\sqrt{\frac{p\bar{q}}{q\bar{p}}} \overset{②}{\leqslant} \varepsilon\sqrt{\varepsilon^{-2}} = 1, \tag{107}$$
so that $\rho(p,q) > 0$.
- on $\mathfrak{D}$, $\partial_p Z$ makes a jump. Denote by $q_p^{\mathfrak{D}}$ the unique $q$ such that $(p,q) \in \mathfrak{D}$. When $q$ tends to $q_p^{\mathfrak{D}}$ by lower values, $(p,q)$ is in ③, so $\partial_p Z(p, q_p^{\mathfrak{D}}-) = 0$, while when $q$ tends to $q_p^{\mathfrak{D}}$ by upper values, $(p,q)$ is in ②, so $\partial_p Z(p, q_p^{\mathfrak{D}}+) = q_p^{\mathfrak{D}} - \varepsilon\tilde{p}\sqrt{q_p^{\mathfrak{D}}\bar{q}_p^{\mathfrak{D}}/p\bar{p}}$. But on $\mathfrak{D}$, $q\bar{p} = \varepsilon^2 p\bar{q}$, so
$$q_p^{\mathfrak{D}} - \varepsilon\tilde{p}\sqrt{\frac{q_p^{\mathfrak{D}}\bar{q}_p^{\mathfrak{D}}}{p\bar{p}}} = q_p^{\mathfrak{D}} - \varepsilon\tilde{p}\sqrt{\frac{(q_p^{\mathfrak{D}})^2}{\varepsilon^2 p^2}} = q_p^{\mathfrak{D}}\Big(1 - \frac{\tilde{p}}{p}\Big) = \frac{q_p^{\mathfrak{D}}}{2p} > 0, \tag{108}$$
so that the jump of $\partial_p Z(p,\cdot)$ at $q_p^{\mathfrak{D}}$ occurs in the increasing sense.
- Similarly we find that on $\mathfrak{U}$, with obvious notation, $\partial_p Z(p, q_p^{\mathfrak{U}}+) - \partial_p Z(p, q_p^{\mathfrak{U}}-) = \bar{q}_p^{\mathfrak{U}}/2\bar{p} > 0$.

So we have proved that $\partial_p Z(p,q)$ is increasing in $q$, which is what we wanted. □

*2.2.8 Remark.* The measure $\mu$ has a rather complicated structure: it is supported by zone ②; it has density $1 + \varepsilon\tilde{p}\tilde{q}/\sqrt{p\bar{p}q\bar{q}}$ w.r.t. the Lebesgue measure in the interior of that zone, and on its boundaries it has a *linear* density giving a mass $(q/2p)dp$ to the infinitesimal part of $\mathfrak{D}$ of abscissa $p$, resp. a mass $(\bar{q}/2\bar{p})dp$ to the infinitesimal part of $\mathfrak{U}$ of abscissa $p$. See Figure 2.

Now that its existence is ensured, we notice an immediate property of the Chogosov process:

**2.2.9 Proposition.** *The Chogosov process is stationary and reversible for the uniform measure on* $(0,1)$*.*

*Proof.* The equilibrium measure of the process is the common value of the marginals of $\mu$, which is uniform by Claim 2.2.7-(ii). Reversibility follows from $\mu$ being invariant under exchange of $p$ and $q$, which itself is due to the symmetry of $Z(p,q)$. □

Now we can turn to the main result of this subsection:

**2.2.10 Lemma.** *The operator norm of the generator $\mathscr{L}$ of the Chogosov process, regarded as an operator on* $\bar{L}^2(0,1)$*, is bounded above by* $\Lambda$.

*Proof.* Let $\eta \in (0,1/2)$, devised to tend to 0, and define the distance $d_\eta$ on $(0,1)$ by:
$$\forall p_1 < p_2 \qquad d_\eta(p_1,p_2) := \int_{p_1}^{p_2} (p\bar{p})^{-3/2+\eta}\, dp. \tag{109}$$

For continuous $f\colon (0,1)\to\mathbb{R}$, define

$$\|f\|_{Lip(\eta)} := \sup_{p_1\neq p_2} \frac{|f(p_2)-f(p_1)|}{d_\eta(p_1,p_2)}, \tag{110}$$

and denote by $Lip(\eta)$ the set of functions $f$ with $\|f\|_{Lip(\eta)}<\infty$. $Lip(\eta)$ is obviously complete for $\|\cdot\|_{Lip(\eta)}$, yet that semi-norm is not definite since it is zero for any constant function. We thus define $\overline{Lip}(\eta)$ as $Lip(\eta)/\mathbb{R}$, which is actually a Banach space. I claim that

**2.2.11 Claim.** *$\overline{Lip}(\eta)$ is continuously imbedded in $\bar{L}^2(0,1)$, i.e. there exists some $C<\infty$ (depending on $\eta$) such that for all $f\in Lip(\eta)$, $\mathrm{Sd}(f)\leqslant C\|f\|_{Lip(\eta)}$.*

*Proof of Claim 2.2.11.* Fix some arbitrary $p_0\in(0,1)$. For $f\in Lip(\eta)$, denoting $y_0 := f(p_0)$, one has, for all $p\in(0,1)$,

$$|f(p)-y_0| \leqslant \|f\|_{Lip(\eta)} \Big|\int_{p_0}^{p} (q\bar{q})^{-3/2+\eta}dq\Big|, \tag{111}$$

whence:

$$\mathrm{Sd}(f) \leqslant \sqrt{\int_0^1 |f(p)-y_0|^2 dp} \leqslant \|f\|_{Lip(\eta)} \sqrt{\int_0^1 \Big(\int_{p_0}^{p} (q\bar{q})^{-3/2+\eta}dq\Big)^2 dp}. \tag{112}$$

Since $\eta>0$, the integral in the right-hand side of (112) is finite, which proves the claim. □

Now, the cruxpoint is the following claim, whose proof is postponed:

**2.2.12 Claim.** *(i) There exists a constant $\Lambda_\eta<\infty$ such that for all $f\in\overline{Lip}(\eta)$, $\|\mathscr{L}f\|_{Lip(\eta)}\leqslant\Lambda_\eta\|f\|_{Lip(\eta)}$.*
*(ii) It is possible to choose $\Lambda_\eta$ so that $\varliminf_{\eta\searrow 0}\Lambda_\eta\leqslant\Lambda$.*

Using Claims 2.2.11 and 2.2.12, one has for all $f\in\overline{Lip}(\eta)$, for $n\in\mathbb{N}$,

$$\mathrm{Sd}(\mathscr{L}^n f) \leqslant C\|\mathscr{L}^n f\|_{Lip(\eta)} \leqslant C\Lambda_\eta^n \|f\|_{Lip(\eta)} \overset{n\to\infty}{=} O(\Lambda_\eta^n). \tag{113}$$

Now, $\mathscr{L}$ is self-adjoint on $\bar{L}^2(0,1)$ since the Chogosov process is reversible, and $\overline{Lip}(\eta)$ is a dense subset of $\bar{L}^2(0,1)$, so by Lemma 0.3.1, (113) implies that $|||\mathscr{L}|||_{\bar{L}^2(0,1)}\leqslant\Lambda_\eta$. Making $\eta\searrow 0$, $|||\mathscr{L}|||_{\bar{L}^2(0,1)}\leqslant\Lambda$, QED. □

*Proof of Claim 2.2.12.* The proof relies on a technique of Markov chains coupling, which itself relies on the monotone rearrangement of measures (cf. [43, p. 75]). For $p\in(0,1)$, $\omega\in[0,1]$, define

$$Q(p,\omega) = \inf\{q\in(0,1)\colon\ \partial_p Z(p,q)\geqslant\omega\}, \tag{114}$$

so that $Q(p,\omega)$ is nondecreasing in $\omega$ and that, for uniform $\omega\in(0,1)$, the law of $Q(p,\omega)$ is the conditional law knowing $p$ of the measure $\mu$ defined by (105)—see Figure 3. Thanks to the function $Q$, we have got a new way of building the Chogosov process

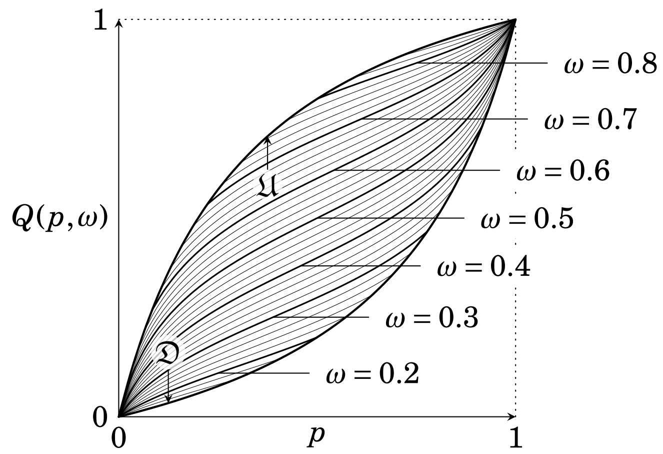



Figure 3: The function $Q(p,\omega)$. This drawing plots the functions $Q(\cdot,\omega)$ for values of $\omega$ running from 0 to 1 with step 0.02. Note that all these functions are defined on the whole $(0,1)$: in fact the graph of $Q$ 'merges' with $\mathfrak{D}$ beyond a certain point for $\omega < 1/2$, resp. it merges with $\mathfrak{U}$ below a certain point for $\omega > 1/2$. For $\omega < \varepsilon^2/2$, resp. $\omega > 1-\varepsilon^2/2$ (which corresponds here to $\omega < 0.125$, resp. $\omega > 0.875$), the whole graph of $Q(p,\cdot)$ is actually equal to the curve $\mathfrak{D}$, resp. $\mathfrak{U}$.

on $\mathbb{N}$: let $p$ be uniform on $(0,1)$ and let $(\omega_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. uniform variables on $[0,1]$, independent of $p$, then

$$p,\ Q(p,\omega_0),\ Q(Q(p,\omega_0),\omega_1),\ Q\big(Q(Q(p,\omega_0),\omega_1),\omega_2\big),\ \ldots \tag{115}$$

is a Chogosov process.

Therefore, if $\omega$ is a random variable uniform on $[0,1]$, one has

$$(\mathscr{L}f)(p) = \mathbf{E}[f(Q(p,\omega))]^{[\S]}, \tag{116}$$

whence the following 'coupling formula':

$$(\mathscr{L}f)(p_2) - (\mathscr{L}f)(p_1) = \mathbf{E}[f(Q(p_2,\omega)) - f(Q(p_1,\omega))]. \tag{117}$$

From (117) we deduce that

$$|(\mathscr{L}f)(p_2) - (\mathscr{L}f)(p_1)| \leqslant \|f\|_{Lip(\eta)}\mathbf{E}[d_\eta(Q(p_1,\omega),Q(p_2,\omega))]. \tag{118}$$

So, if we can prove that for all $p_1 < p_2$,

$$\mathbf{E}[d_\eta(Q(p_1,\omega),Q(p_2,\omega))] \leqslant \Lambda_\eta d_\eta(p_1,p_2), \tag{119}$$

then we are done.

Now I claim (it will be checked later) that $Q$ is absolutely continuous w.r.t. $p$, i.e. that there exists an integrable function $Q'\colon\ (0,1)\times[0,1]\to\mathbb{R}$ such that for all $\omega, p_1, p_2$ one has $Q(p_2,\omega) = Q(p_1,\omega) + \int_{p_1}^{p_2} Q'(p,\omega)dp$. Introducing that function, (119) becomes:

$$\mathbf{E}\Big[\Big|\int_{p_1}^{p_2}(Q(p,\omega)\bar{Q}(p,\omega))^{-3/2+\eta}Q'(p,\omega)dp\Big|\Big] \leqslant \Lambda_\eta \mathbf{E}\Big[\int_{p_1}^{p_2}(p\bar{p})^{-3/2+\eta}dp\Big], \tag{120}$$

[§]Beware that here the expectation is not taken w.r.t. $p$ but w.r.t. $\omega$.

so by Fubini's theorem (which is legal here since, as we will see later, $Q'$ is bounded), proving (120) for all $p_1 < p_2$ is tantamount to proving that, for all $p \in (0,1)$,

$$\mathbf{E}\big[(Q(p,\omega)\bar{Q}(p,\omega))^{-3/2+\eta}|Q'(p,\omega)|\big] \leqslant \Lambda_\eta(p\bar{p})^{-3/2+\eta}. \tag{121}$$

So we have to compute $Q'(p,\omega)$. Using the structure of the law $\mu$ (cf. Remark 2.2.8), we find the following (see Figure 3):

- First if $\omega < q_p^{\mathfrak{D}}/2p$, then $Q(p,\omega) = q_p^{\mathfrak{D}}$, whence $Q'(p,\omega) = dq_p^{\mathfrak{D}}/dp$. Differentiating the equality $q\bar{p} = \varepsilon^2 p\bar{q}$ defining $\mathfrak{D}$, one finds that $dq_p^{\mathfrak{D}}/dp = (q_p^{\mathfrak{D}} + \varepsilon^2\bar{q}_p^{\mathfrak{D}})/(\bar{p}+\varepsilon^2 p)$, which simplifies into $q_p^{\mathfrak{D}}\bar{q}_p^{\mathfrak{D}}/p\bar{p}$ using once again that $q\bar{p} = \varepsilon^2 p\bar{q}$.
- Similarly if $\omega > 1-\bar{q}_p^{\mathfrak{U}}/2\bar{p}$, one has $Q'(p,\omega) = q_p^{\mathfrak{U}}\bar{q}_p^{\mathfrak{U}}/p\bar{p}$.
- If $q_p^{\mathfrak{D}}/2p < \omega < 1-\bar{q}_p^{\mathfrak{U}}/2\bar{p}$, then $\partial_p Z(p,q) = q - \varepsilon\tilde{p}\sqrt{q\bar{q}/p\bar{p}}$, thus differentiating the equality $\partial_p Z(p,Q(p,\omega)) = \omega$, we get:

$$Q'(p,\omega) = \frac{\varepsilon\sqrt{Q(p,\omega)\bar{Q}(p,\omega)}}{4\sqrt{p\bar{p}}^3\Big(1+\varepsilon\frac{\tilde{p}\tilde{Q}(p,\omega)}{\sqrt{p\bar{p}Q(p,\omega)\bar{Q}(p,\omega)}}\Big)}. \tag{122}$$

- Finally in the critical cases $\omega = q_p^{\mathfrak{D}}/2p, 1-\bar{q}_p^{\mathfrak{U}}/2\bar{p}$, there is no canonical value for $Q'(\omega)$ since at these points $Q(\cdot,\omega)$ is not $\mathscr{C}^1$, but that does not matter.

*2.2.13 Remark.* Note that one always has $Q'(p,\omega) > 0$, i.e. $Q(\cdot,\omega)$ is increasing. In other words, if one couples two Chogosov processes $(p_i)_{i\in\mathbb{N}}$ and $(p'_i)_{i\in\mathbb{N}}$ by constructing them thanks to the same $\omega_i$, then for almost-all realizations of the coupled processes, the relative order of $p_i$ and $p'_i$ is the same for all $i$.

We have computed $Q'(p,\omega)$, so now we can tackle (121): we have to bound

$$\int_0^1 \Big(\frac{p\bar{p}}{Q(p,\omega)\bar{Q}(p,\omega)}\Big)^{3/2-\eta} Q'(p,\omega)d\omega, \tag{123}$$

uniformly in $p$. We begin with noticing that

**2.2.14 Claim.** *For all $p \in (0,1)$, all $q \in [q_p^{\mathfrak{D}}, q_p^{\mathfrak{U}}]$, one has $q\bar{q}/p\bar{p} \leqslant \varepsilon^{-2}$.*

*Proof of Claim 2.2.14.* The condition $q \in [q_p^{\mathfrak{D}}, q_p^{\mathfrak{U}}]$ means that $\varepsilon^2 \leqslant p\bar{q}/q\bar{p} \leqslant \varepsilon^{-2}$. Then we distinguish two cases:

- If $p \leqslant q$, then $q\bar{q}/p\bar{p} = (\bar{q}/\bar{p})^2 q\bar{p}/p\bar{q} \leqslant q\bar{p}/p\bar{q} = (p\bar{q}/q\bar{p})^{-1} \leqslant \varepsilon^{-2}$;
- If $p \geqslant q$, then $q\bar{q}/p\bar{p} = (q/p)^2 p\bar{q}/q\bar{p} \leqslant p\bar{q}/q\bar{p} \leqslant \varepsilon^{-2}$.

□

Thanks to Claim 2.2.14, we bound (123) by

$$\varepsilon^{-2\eta}\int_0^1 \Big(\frac{p\bar{p}}{Q(p,\omega)\bar{Q}(p,\omega)}\Big)^{3/2} Q'(p,\omega)d\omega, \tag{124}$$

which we shorthand into $\varepsilon^{-2\eta}\lambda(p)$. Splitting the integral in (124) according to the value of $\omega$ (resp. for $\omega \in (0, q_p^{\mathfrak{D}}/2p)$, $(q_p^{\mathfrak{D}}/2p, 1-\bar{q}_p^{\mathfrak{U}}/2\bar{p})$ and $(1-\bar{q}_p^{\mathfrak{U}}/2\bar{p}, 1)$), one finds:

$$
\begin{aligned}
\lambda(p) &= \frac{q_p^{\mathfrak{D}}}{2p}\Big(\frac{p\bar{p}}{q_p^{\mathfrak{D}}\bar{q}_p^{\mathfrak{D}}}\Big)^{3/2}\frac{q_p^{\mathfrak{D}}\bar{q}_p^{\mathfrak{D}}}{p\bar{p}} && (125)\\
&+ \int_{q_p^{\mathfrak{D}}/2p}^{1-\bar{q}_p^{\mathfrak{U}}/2\bar{p}}\Big(\frac{p\bar{p}}{Q(p,\omega)\bar{Q}(p,\omega)}\Big)^{3/2}\frac{\varepsilon\sqrt{Q(p,\omega)\bar{Q}(p,\omega)}}{4\sqrt{p\bar{p}}^3\Big(1+\varepsilon\frac{\tilde{p}\tilde{Q}(p,\omega)}{\sqrt{p\bar{p}Q(p,\omega)\bar{Q}(p,\omega)}}\Big)}d\omega && (126)\\
&+ \frac{\bar{q}_p^{\mathfrak{U}}}{2\bar{p}}\Big(\frac{p\bar{p}}{q_p^{\mathfrak{U}}\bar{q}_p^{\mathfrak{U}}}\Big)^{3/2}\frac{q_p^{\mathfrak{U}}\bar{q}_p^{\mathfrak{U}}}{p\bar{p}}. && (127)
\end{aligned}
$$

(125) simplifies into $\frac{1}{2}\sqrt{q_p^{\mathfrak{D}}\bar{p}/p\bar{q}_p^{\mathfrak{D}}} = \frac{1}{2}\sqrt{\varepsilon^2} = \varepsilon/2$; similarly (127) $= \frac{1}{2}\sqrt{p\bar{q}_p^{\mathfrak{U}}/q_p^{\mathfrak{U}}\bar{p}} = \varepsilon/2$. Concerning term (126), we make the change of variables $q = Q(p,\omega)$, for which $d\omega = (1+\varepsilon\tilde{p}\tilde{q}/\sqrt{p\bar{p}q\bar{q}})dq$ because of the expression of the density of $\mu$ in zone ② (cf. Remark 2.2.8). One gets:

$$
(126) = \frac{\varepsilon}{4}\int_{q_p^{\mathfrak{D}}}^{q_p^{\mathfrak{U}}}\frac{1}{q\bar{q}}dq = \frac{\varepsilon}{4}\Big[\log\frac{q}{\bar{q}}\Big]_{q_p^{\mathfrak{D}}}^{q_p^{\mathfrak{U}}} = \frac{\varepsilon}{4}\Big(\log\frac{\bar{p}}{\varepsilon^2 p} - \log\frac{\varepsilon^2\bar{p}}{p}\Big) = \frac{\varepsilon}{4}\log\frac{1}{\varepsilon^4} = \varepsilon|\log\varepsilon|. \quad (128)
$$

So in the end we have $\lambda(p) = \varepsilon/2+\varepsilon/2+\varepsilon|\log\varepsilon| = \Lambda$ for all $p$, thus $\Lambda_\eta \leqslant \varepsilon^{-2\eta}\Lambda$ (hence (i)), which tends to $\Lambda$ as $\eta \searrow 0$ (hence (ii)). □

*2.2.15 Remark.* The simplifications in the computation of $\lambda(p)$ look rather miraculous… *A priori* I only expected that $\lambda(p) \leqslant \Lambda$ on $(0,1)$ with $\lambda(p) \overset{p\to 0,1}{\to} \Lambda$). That I found the exact quasi-eigenvector associated to the quasi-eigenvalue $\Lambda$ (cf. Remark 2.2.16) is purely fortuitous; I have no explanation for why things work so well.

*2.2.16 Remark.* $\mathscr{L}$ is self-adjoint, hence normal, so its operator norm is also its spectral radius. Therefore there is some (*eigenvalue*,*eigenvector*) pair, or more precisely (since here the spectral radius of $\mathscr{L}$ is due to its continuous spectrum) some 'quasi-eigenvalue' and its 'quasi-eigenvector' (cf. [39, § 4]), which are responsible for the value of the operator norm.

Tracking this quasi-eigenvector throughout the proof of Lemma 2.2.10, we find that $\Lambda$ is a quasi-eigenvalue of $\mathscr{L}$ and that the quasi-eigenvector associated is:

$$
f_\Lambda\colon\ p \mapsto \int_{1/2}^{p}(p'\bar{p}')^{-3/2}dp'. \quad (129)
$$

Obviously $f_\lambda$ is not in $L^2$, so it is not a true eigenvector, however one can perturb it slightly to get an element $\tilde{f}_\Lambda \in \bar{L}^2(0,1)\smallsetminus\{0\}$ such that $\langle\mathscr{L}\tilde{f}_\Lambda,\tilde{f}_\Lambda\rangle_{\bar{L}^2(0,1)}/\|\tilde{f}_\Lambda\|^2_{\bar{L}^2(0,1)}$ is arbitrarily close to $\Lambda$.

*2.2.17 Remark.* An interesting feature of $f_\Lambda$ is that its '$L^2$ mass' is concentrated about 0 and 1, so that one needs only look at what happens near 0 and 1 to understand how $f_\Lambda$ contributes to the operator norm of $\mathscr{L}$.

When one 'zooms' more and more to the point $(0,0)$ —the same behaviour would happen about $(1,1)$ —, $\mu$ 'looks more and more like' the measure $\mu^*$ on $(0,\infty)^2$ defined by (see Figure 4):

$$
\forall p,q \in [0,\infty)^2 \quad \mu^*[\{(x_1,x_2)\in(0,\infty)^2\colon\ x_1 \leqslant p \text{ and } x_2 \leqslant q\}] = \varepsilon\sqrt{pq}\wedge p\wedge q, \quad (130)
$$

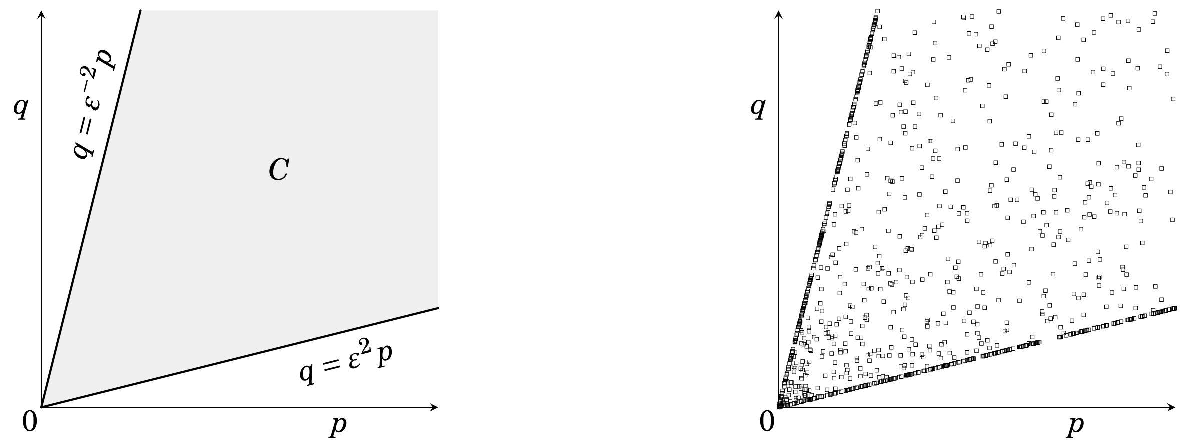



Figure 4: The measure $\mu^*$. On the left, the different zones for the measure; on the right, a Poisson cloud of points with density $\mu^*$. The scale and density of the cloud are consistent with Figure 2.

i.e.

$$d\mu^*(p,q) = \mathbf{1}_{\{\varepsilon^2 p < q < \varepsilon^{-2}p\}} \frac{\varepsilon}{4\sqrt{pq}} dp dq + \mathbf{1}_{\{q=\varepsilon^2 p\}} \frac{\varepsilon^2}{2} dp + \mathbf{1}_{\{q=\varepsilon^{-2}p\}} \frac{1}{2} dp. \qquad (131)$$

So, near 0, $\mathscr{L}$ behaves like the operator $\mathscr{L}^*$ on $L^2(0,\infty)$ defined by:

$$(\mathscr{L}f)(p) = \int_{\varepsilon^2 p}^{\varepsilon^{-2}p} \frac{\varepsilon}{4\sqrt{pq}} dq + \frac{\varepsilon^2}{2} f(\varepsilon^2 p) + \frac{1}{2} f(\varepsilon^{-2}p). \qquad (132)$$

$\mathscr{L}^*$ has scale invariance properties which make it easy to study. One finds that $\mathscr{L}^*$ is self-adjoint, that its spectral radius is $\Lambda$, and that it has $\Lambda$ as a quasi-eigenvalue, associated with the quasi-eigenvector $(p \mapsto 1/\sqrt{p})$. So, you see that it suffices to study the 'local' operator $\mathscr{L}^*$ to compute the spectral radius of the 'global' operator $\mathscr{L}$; in other words, there is a phenomenon of 'localization of the spectral radius' for $\mathscr{L}$.

## 2.2.c Optimality of the strong event sufficient condition

Now I will prove that Theorem 2.2.1 is optimal:

**2.2.18 Theorem.** *The factor $\Lambda(\varepsilon)$ in (98) cannot be improved. In other words, for all $\Lambda' < \Lambda(\varepsilon)$ it is possible to find $\sigma$-fields $\mathscr{F}$ and $\mathscr{G}$ satisfying*

$$\forall A \in \mathscr{F}, B \in \mathscr{G} \quad \mathbf{P}[A \cap B] - \mathbf{P}[A]\mathbf{P}[B] \leqslant \varepsilon \sqrt{\mathbf{P}[A]\mathbf{P}[A^c]\mathbf{P}[B]\mathbf{P}[B^c]}, \qquad (133)$$

*but such that $\{\mathscr{F} : \mathscr{G}\} \geqslant \Lambda'$.*

Actually I will rather prove the following statement, which is equivalent to the theorem by continuity of the function $\Lambda(\cdot)$:

**2.2.19 Claim.** *For all $\varepsilon' > \varepsilon$ it is possible to find $\sigma$-fields $\mathscr{F}$ and $\mathscr{G}$ satisfying $\{\mathscr{F} : \mathscr{G}\} \geqslant \Lambda(\varepsilon)$, but such that*

$$\forall A \in \mathscr{F}, B \in \mathscr{G} \quad \mathbf{P}[A \cap B] - \mathbf{P}[A]\mathbf{P}[B] \leqslant \varepsilon' \sqrt{\mathbf{P}[A]\mathbf{P}[A^c]\mathbf{P}[B]\mathbf{P}[B^c]}. \qquad (134)$$

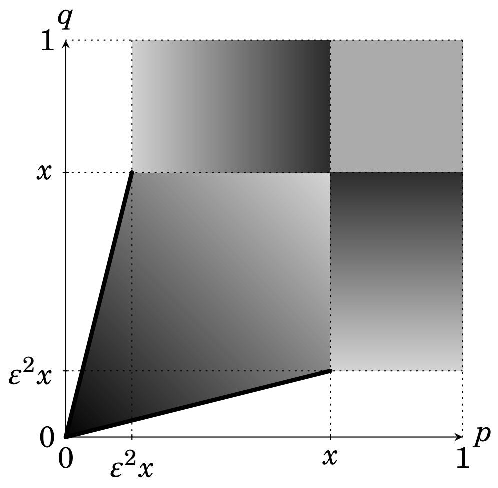



Figure 5: A schematic representation of the measure $\nu$.

*Proof.* According to the proof of Lemma 2.2.10 in § 2.2.b, the 'natural' proof would be to take for space $(\Omega, \mathscr{B}, \mathbf{P})$ the set $(0,1)^2$ equipped with its Borel $\sigma$-field and endowed with the measure $\mu$ defined by (105), and to set $\mathscr{F} = \sigma(p)$ and $\mathscr{G} = \sigma(q)$. Though it seems to be true that that system satisfies (97), the complicated structure of $\mu$ makes existence of a short proof for this property unlikely. Therefore I will rather adapt the previous idea to the nicer measure $\mu^*$ defined by (130), or more precisely to a 'truncation' of it.

My system is the following: $(\Omega, \mathscr{B}, \mathbf{P})$ is the set $(0,1)^2$ equipped with its Borel $\sigma$-field and endowed with a certain measure $\nu$ (specified just after), and I take $\mathscr{F} = \sigma(p)$, resp. $\mathscr{G} = \sigma(q)$. The measure $\nu$, which depends on some parameter $x \in (0,1)$ morally close to 0, is a measure on $(0,1)^2$ having uniform marginals, which coincides with $\mu^*$ on $(0,x]^2$ and which is 'as uniform as possible' outside $(0,x]^2$ (see Figure 5). Technically:

$$\nu[A \times B] = \begin{cases} \mu^*(A \times B) & \text{if } A \subset (0,x] \text{ and } B \subset (0,x]; \\ 0 & \text{if } A \subset (0,\varepsilon^2 x] \text{ and } B \subset (x,1); \\ 0 & \text{if } A \subset (x,1) \text{ and } B \subset (0,\varepsilon^2 x]; \\ [\int_A (1 - \frac{\varepsilon}{2}\sqrt{\frac{x}{p}})dp]|B|/(1-x) & \text{if } A \subset (\varepsilon^2 x, x] \text{ and } B \subset (x,1); \\ |A|[\int_B (1 - \frac{\varepsilon}{2}\sqrt{\frac{x}{q}})dq]/(1-x) & \text{if } A \subset (x,1) \text{ and } B \subset (\varepsilon^2 x, x]; \\ [1-(2-\varepsilon)x]|A||B|/(1-x)^2 & \text{if } A \subset (x,1) \text{ and } B \subset (x,1). \end{cases} \tag{135}$$

*First step: Proof that* $\{\mathscr{F} : \mathscr{G}\} \geqslant \Lambda$. Let $\Lambda' < \Lambda$. Since $\Lambda$ is in the spectrum of the self-adjoint operator $\mathscr{L}^*$ on $L^2(0,\infty)$ (see (132) and the lines surrounding it), there exists $f \in L^2(0,\infty) \setminus \{0\}$ such that $\langle \mathscr{L}^* f, f \rangle / \|f\|_{L^2(0,\infty)} > \Lambda'$. By a standard truncation argument, we can assume that $f$ has bounded support, say that $f$ is zero outside $(0,Y]$. Dividing $f$ by its norm we can also assume that $\|f\|_{L^2} = 1$.

Now, for $y \in (0,x]$ define the function $f_y$ by:

$$f_y(p) = \sqrt{\frac{Y}{y}}\, f\Big(\frac{Y}{y} p\Big). \tag{136}$$

$f_y$ is zero outside $(0,y]$; it satisfies $\|f_y\|_{L^2} = 1$ and

$$\langle \mathscr{L}^* f_y, f_y \rangle = \langle \mathscr{L}^* f, f \rangle > \Lambda' \tag{137}$$

by the scale invariance properties of $\mathscr{L}^*$.

Denote $m := \int f(p)dp/\sqrt{Y}$, which is finite since $f$ is $L^2$ with compact support; one has $\int f_y(p)dp = \sqrt{y}m$, so the projection of $f_y$ on $\bar{L}^2(0,1)$ is the function $\bar{f}_y = f_y - \sqrt{y}m$. One has $\|\bar{f}_y\|_{\bar{L}^2(0,1)} \leqslant \|f_y\|_{L^2} = 1$, and for $y \in (0,x]$,

$$\mathbf{E}[\bar{f}_y(p)\bar{f}_y(q)] = \langle \mathscr{L}^* f_y, f_y\rangle - m^2 y > \Lambda' - m^2 y, \tag{138}$$

so that $\{\mathscr{F}:\mathscr{G}\} > \Lambda' - m^2 y$. Making $y \to 0$ and then $\Lambda' \to \Lambda$, one finally gets $\{\mathscr{F}:\mathscr{G}\} \geqslant \Lambda$.

*Second step: Proof of (134).* Let $\varepsilon' > \varepsilon$; we want to prove that, provided $x$ is small enough, (134) is satisfied.

Let $A$ and $B$ be resp. $\mathscr{F}$- and $\mathscr{G}$-measurable events. One can assume safely that $|A| \leqslant 1/2$, since replacing simultaneously $A$ by $A^{\mathrm{c}}$ and $B$ by $B^{\mathrm{c}}$ leaves both sides of (134) unchanged. One can also assume that $|B| < 1/(1+\varepsilon^2)$, since for $|B| \geqslant 1/(1+\varepsilon^2)$, (134) comes 'freely' by writing

$$\mathbf{P}[A\cap B] - \mathbf{P}[A]\mathbf{P}[B] \leqslant \mathbf{P}[A](1-\mathbf{P}[B]) \leqslant \sqrt{\mathbf{P}[A]\mathbf{P}[A^{\mathrm{c}}]} \times \varepsilon\sqrt{\mathbf{P}[B]\mathbf{P}[B^{\mathrm{c}}]}. \tag{139}$$

☛ *In the sequel of this proof we indentify A and B with Borel subsets of* $(0,1)$*, rewriting the p-measurable event A into the set* $A\times(0,1)$*, resp. the q-measurable event B into the set* $(0,1)\times B$*. Since both marginals of* $\nu$ *are uniform on* $(0,1)$*, one then has* $\mathbf{P}[A] = |A|$*, resp.* $\mathbf{P}[B] = |B|$*, so that our goal becomes proving:*

$$\nu[A\times B] - |A||B| \leqslant \varepsilon'\sqrt{|A||B||A^{\mathrm{c}}||B^{\mathrm{c}}|}. \tag{140}$$

Denote $\breve{A} := A\cap(0,x]$, resp. $\breve{B} := B\cap(0,x]$. Provided $x \leqslant \varepsilon/2$, the signed measure $d\nu(p,q) - dpdq$ is nonpositive on $(0,x]\times(x,1)\cup(x,1)\times(0,x]$, so that

$$\nu[A\times B] - |A||B| \leqslant \nu[\breve{A}\times\breve{B}] - |\breve{A}||\breve{B}| + \nu[(A\smallsetminus\breve{A})\times(B\smallsetminus\breve{B})] - |A\smallsetminus\breve{A}||B\smallsetminus\breve{B}|. \tag{141}$$

Now let us bound above the right-hand side of (141):

- The second term is obviously nonpositive.
- The third term is $[1-(2-\varepsilon)x]|A\smallsetminus\breve{A}||B\smallsetminus\breve{B}|/(1-x)^2$, so the sum of the two last terms is $(\varepsilon x - x^2)|A\smallsetminus\breve{A}||B\smallsetminus\breve{B}|/(1-x)^2 \leqslant (\varepsilon x - x^2)|A||B|/(1-x)^2$. Since $|A| \leqslant 1/2$ and $|B| \leqslant 1/(1+\varepsilon^2)$, that quantity is in turn bounded by $\frac{\varepsilon x - x^2}{\varepsilon(1-x)^2} \times \sqrt{|A||B||A^{\mathrm{c}}||B^{\mathrm{c}}|}$.
- For the first term, by Lemma 2.2.20 stated just below, one has $\nu[\breve{A}\times\breve{B}] = \mu^*[\breve{A}\times\breve{B}] \leqslant \varepsilon\sqrt{|\breve{A}||\breve{B}|} \leqslant \varepsilon\sqrt{|\breve{A}||\breve{A}^{\mathrm{c}}||\breve{B}||\breve{B}^{\mathrm{c}}|}/(1-x)$, in which, provided $x \leqslant \varepsilon^2/(1+\varepsilon^2)$, one has $\sqrt{|\breve{B}||\breve{B}^{\mathrm{c}}|} \leqslant \sqrt{B|B^{\mathrm{c}}|}$ (because then $|\breve{B}| \leqslant |B|\wedge x$ and $|B| \leqslant 1-x$), and similarly $\sqrt{|\breve{A}||\breve{A}^{\mathrm{c}}|} \leqslant \sqrt{A|A^{\mathrm{c}}|}$, so that in the end the first term is bounded by $\frac{\varepsilon}{1-x}\sqrt{|A||B||A^{\mathrm{c}}||B^{\mathrm{c}}|}$.

Summing things up, we get:

$$\nu[A\times B] - |A||B| \leqslant \Big(\frac{\varepsilon}{1-x} + \frac{\varepsilon x - x^2}{\varepsilon(1-x)^2}\Big)\sqrt{|A||B||A^{\mathrm{c}}||B^{\mathrm{c}}|}. \tag{142}$$

Taking $x$ sufficiently close to 0, the first factor of the right-hand side of (142) is $\leqslant \varepsilon'$, whence the second step of the proof. $\square$

**2.2.20 Lemma.** *For all $A, B \subset (0,\infty)^2$ with Lebesgue measures $|A|, |B| < \infty$, $\mu^*[A \times B] \leqslant \varepsilon\sqrt{|A||B|}$.*

*Proof of Lemma 2.2.20.* Recall that $\mu^*$ is the Radon measure on $(0,\infty)^2$ having density $\varepsilon/4\sqrt{pq}$ w.r.t. the Lebesgue measure inside the cone $C = \{(p,q):\ \varepsilon^2 p < q < \varepsilon^{-2} p\}$, being zero outside $C$, and giving to the borders of $C$ a lineic mass defined by $\mu^*\{(p, \varepsilon^2 p):\ p \in A\} = \varepsilon^2|A|/2$, resp. $\mu^*\{(p, \varepsilon^{-2} p):\ p \in A\} = |A|/2$ (see Figure 4). $\mu^*$ is invariant by switching $p$ and $q$, and its marginals are both the Lebesgue measure on $(0,\infty)$. Let $A, B \subset (0,\infty)$ be Borel; our goal is to show that $\mu^*[A \times B] \leqslant \varepsilon\sqrt{|A||B|}$.

***Step 1.*** If $|A| \leqslant \varepsilon^2|B|$ the result is trivially true, since then $\mu^*[A \times B] \leqslant \mu^*[A \times (0,\infty)] = |A| \leqslant \varepsilon\sqrt{|A||B|}$. Similarly the result is true if $|B| \leqslant \varepsilon^2|A|$. Therefore in our proof we will always assume that $\varepsilon^2|A| < |B| < \varepsilon^{-2}|A|$.

***Step 2.*** As for the measure $\mu$, decompose the support of $\mu^*$ into three parts $\mathfrak{U}$, ② and $\mathfrak{D}$, corresponding resp. to the line "$p = \varepsilon^2 q$", the cone $C$ and the line "$p = \varepsilon^{-2} q$" (see Figure 4). Write $\mu^*[A \times B] = m_U + m_2 + m_D$, where $m_U = \mu^*[(A \times B) \cap \mathfrak{U}]$, etc..

Denote by $\mu_q^*$ the 'conditional law' of $\mu$ knowing $q$, i.e. the measure such that

$$\mu^*[X] = \int_0^\infty \mu_q^*[\{p:\ (p,q) \in X\}]\, dq, \tag{143}$$

which can be computed explicitly to be:

$$d\mu_q^*[p] = \mathbf{1}_{\{p = \varepsilon^2 q\}} \frac{\varepsilon^2}{2} + \mathbf{1}_{\{\varepsilon^2 q < p < \varepsilon^{-2} q\}} \frac{\varepsilon}{4\sqrt{pq}} dp + \mathbf{1}_{\{p = \varepsilon^{-2} q\}} \frac{1}{2}. \tag{144}$$

The three terms of the right-hand side of (144) are respectively due to $\mathfrak{U}$, ② and $\mathfrak{D}$, so that, integrating, one finds:

$$m_U = \int_B \frac{\varepsilon^2}{2} \mathbf{1}_{\{A \ni \varepsilon^2 q\}} dq \leqslant \frac{\varepsilon^2}{2}|B|. \tag{145}$$

Switching the roles of $p$ and $q$, one has similarly $m_D \leqslant \varepsilon^2|A|/2$. Then it only remains to bound $m_2$.

***Step 3.*** Let us study further the measures $\mu_q^*$. If $q \in \varepsilon^2 A$, then $A \ni \varepsilon^{-2} q$ and thus $\mu_q^*[A] \geqslant \mu^*[\{\varepsilon^{-2} q\}] = 1/2$, and conversely if $q \notin \varepsilon^2 A$, then $A \not\ni \varepsilon^{-2} q$ and thus $\mu_q^*[A] \leqslant 1 - \mu^*[\{\varepsilon^{-2} q\}] = 1/2$. So, $\mu_q^*[A]$ is never smaller if $q \in \varepsilon^2 A$ than if $q \notin \varepsilon^2 A$.

As a consequence, let us show that we can always assume that $\varepsilon^2 A \subset B$. Since $|B| > \varepsilon^2|A|$, $|B \smallsetminus \varepsilon^2 A| > |\varepsilon^2 A \smallsetminus B|$, so we can fix some $B^- \subset B \smallsetminus \varepsilon^2 A$ such that $|B^-| = |\varepsilon^2 A \smallsetminus B|$. One has:

$$\mu^*[A \times B^-] = \int_{B^-} \mu_q^*(A)\, dq \leqslant \frac{|B^-|}{2} = \frac{|\varepsilon^2 A \smallsetminus B|}{2} \leqslant \int_{\varepsilon^2 A \smallsetminus B} \mu_q^*(A)\, dq = \mu^*[A \times (\varepsilon^2 A \smallsetminus B)]. \tag{146}$$

Shorthanding "$(B \smallsetminus B^-) \cup \varepsilon^2 A$" into "$B'$", (146) implies that replacing $B$ by $B'$ —which does not modify the value of $|B|$ —cannot make $\mu^*[A \times B]$ decrease. Consequently, if we prove that $\mu^*[A \times B'] \leqslant \varepsilon\sqrt{|A||B'|}$, then we will also have proved that $\mu^*[A \times B] \leqslant \varepsilon\sqrt{|A||B|}$. As $\varepsilon^2 A \subset B'$, we have thus demonstrated the statement at the beginning of this paragraph: one can always assume that $\varepsilon^2 A \subset B$.

☛ *Switching the roles of $p$ and $q$, we will rather impose, instead of $\varepsilon^2 A \subset B$, that $\varepsilon^2 B \subset A$.*

***Step 4.*** Call $\mu^\circ$ the measure $\mu^*$ restricted to $C$, i.e. $d\mu^\circ = \mathbf{1}_C d\mu^*$, so that $m_2 = \mu^\circ[A \times B]$. $\mu^\circ$ is absolutely continuous w.r.t. the Lebesgue measure; denote by $\mu_q^\circ$ its 'conditional measure' for fixed $q$, i.e. the measure such that

$$\mu^\circ[X] = \int_0^\infty \mu_q^\circ[\{p:\ (p,q) \in X\}]\,dq, \tag{147}$$

which has the following explicit density w.r.t. the Lebesgue measure:

$$d\mu_q^\circ[p] = \mathbf{1}_{\{\varepsilon^2 q < p < \varepsilon^{-2} q\}} \frac{\varepsilon}{4\sqrt{pq}}\,dp. \tag{148}$$

We perform a change of variables: for $y \in (0,|B|)$, define

$$\beta(y) = \inf\{q \in (0,\infty):\ |B \cap (0,q)| \geqslant y\}; \tag{149}$$

so that the push-forward $\beta \# dq$ of the Lebesgue measure on $(0,|B|)$ by the map $\beta$ is equal to $\mathbf{1}_B dq$, the Lebesgue measure restricted to $B$; then

$$m_{②} = \int_B \mu_q^\circ[A]\,dq = \int_0^{|B|} \mu_{\beta(y)}^\circ[A]\,dy. \tag{150}$$

Our strategy will consist in bounding $\mu_{\beta(y)}^\circ[A]$ for all $y$.

First, we observe that some portion of $A$ does not contribute to $\mu_{\beta(y)}^\circ[A]$. Denote indeed $A_y = \{\varepsilon^2 q:\ q \in B \cap (0,\beta(y))\}$; by definition of $\beta$, $|A_y| = \varepsilon^2 y$, and one has $A_y \subset \varepsilon^2 B \subset A$. But $A_y \subset (0,\varepsilon^2\beta(y))$, so $\mu_{\beta(y)}^\circ[A_y] = 0$, and thus $\mu_{\beta(y)}^\circ[A] = \mu_{\beta(y)}^\circ[A \smallsetminus A_y]$, where $|A \smallsetminus A_y| = |A| - |A_y| = |A| - \varepsilon^2 y$.

Now, for $q \in (0,\infty)$, the density of $\mu_q^\circ$ is zero for $p \leqslant \varepsilon^2 q$ and it is nonincreasing for $p > \varepsilon^2 q$, so an immediate coupling argument shows that the maximal value of $\mu_q^\circ[X]$ under the constraint "$|X| = x$" is attained for $X = (\varepsilon^2 q, \varepsilon^2 q + x)$. Applying that result to the conclusion of the previous paragraph, we get that:

$$\mu_{\beta(y)}^\circ[A] \leqslant \mu_{\beta(y)}^\circ[(\varepsilon^2\beta(y),\, \varepsilon^2\beta(y) + |A| - \varepsilon^2 y)]. \tag{151}$$

But for $x \geqslant 0$, the quantity $\mu_q^\circ[(\varepsilon^2 q\,,\, \varepsilon^2 q + x)]$ can be computed explicitly to be

$$\mu_q^\circ[(\varepsilon^2 q\,,\, \varepsilon^2 q + x)] = \begin{cases} (1-\varepsilon^2)/2 & \text{if } q \leqslant x/(\varepsilon^{-2} - \varepsilon^2); \\ (\varepsilon\sqrt{\varepsilon^2 + x/q} - \varepsilon^2)/2 & \text{if } q > x/(\varepsilon^{-2} - \varepsilon^2). \end{cases} \tag{152}$$

In particular, that quantity is a nonincreasing function of $q$. Since, by definition of $\beta$, one always has $\beta(y) \geqslant y$, it follows that (151) can be improved into:

$$\mu_{\beta(y)}^\circ[A] \leqslant \mu_y^\circ[(\varepsilon^2 y\,,\, |A|)] = \begin{cases} (1-\varepsilon^2)/2 & \text{if } y \leqslant \varepsilon^2 |A|; \\ (\varepsilon\sqrt{|A|/y} - \varepsilon^2)/2 & \text{if } y > \varepsilon^2 |A|. \end{cases} \tag{153}$$

Integrating, one finds finally:

$$\begin{aligned} m_2 &\leqslant \int_0^{\varepsilon^2|A|} \frac{1-\varepsilon^2}{2}\,dy + \int_{\varepsilon^2|A|}^{|B|} \Big(\frac{\varepsilon\sqrt{|A|}}{2\sqrt{y}} - \frac{\varepsilon^2}{2}\Big)dy \\ &= \frac{(1-\varepsilon^2)\varepsilon^2|A|}{2} + \Big[\varepsilon\sqrt{|A|y} - \frac{\varepsilon^2 y}{2}\Big]_{\varepsilon^2|A|}^{|B|} = \varepsilon\sqrt{|A||B|} - \frac{\varepsilon^2}{2}(|A| + |B|). \end{aligned} \tag{154}$$

***Step 5.*** We put our bounds together to get the lemma:

$$\mu^*[A\times B] \leqslant m_D + m_U + m_2 = \frac{\varepsilon^2}{2}(|A|+|B|) + \varepsilon\sqrt{|A||B|} - \frac{\varepsilon^2}{2}(|A|+|B|) = \varepsilon\sqrt{|A||B|}. \quad (155)$$

$\square$

*2.2.21 Remark.* A careful reading of the previous proof shows that the maximal value of $\mu^*[A\times B]$ is attained for $A=(0,|A|), B=(0,|B|)$, in which case, provided $\varepsilon^2|A| \leqslant |B| \leqslant \varepsilon^{-2}|A|$, one has equality in Lemma 2.2.20.

# Chapter 3

# Tensorization

## 3.1 Subjective correlation

In this chapter we will need more advanced definitions for decorrelation.

**3.1.1 Definition.** Let $X$, $Y$ and $Z$ be random variables. For $\varepsilon \geqslant 0$, one says that $X$ and $Y$ are *subjectively $\varepsilon$-decorrelated w.r.t. $Z$* (or *$\varepsilon$-decorrelated seen from $Z$*) if $X$ and $Y$ are $\varepsilon$-decorrelated under the law $Law(X,Y|Z=z)$ for $Law(Z)$-almost-all $z$[*].

The smallest $\varepsilon$ such that $X$ and $Y$ are $\varepsilon$-decorrelated seen from $Z$ will be called the *subjective correlation level between $X$ and $Y$ w.r.t. $Z$* (or *correlation level between $X$ and $Y$ seen from $Z$*); we denote it $\{X:Y\}_Z$.

In § 1.1, we had given the definitions in terms of $\sigma$-algebras rather than random variables. Of course there is also a $\sigma$-algebra definition for subjective correlation, though I find it harder to understand:

**3.1.2 Definition.** Let $\mathscr{F}$, $\mathscr{G}$ and $\mathscr{H}$ be $\sigma$-algebras. For $\varepsilon \in [0,1]$, the expression "$\{\mathscr{F}:\mathscr{G}\}_{\mathscr{H}} \leqslant \varepsilon$" means that for all $f \in \bar{L}^2(\mathscr{F} \vee \mathscr{H})$ and all $g \in \bar{L}^2(\mathscr{G} \vee \mathscr{H})$ satisfying $\mathbf{E}[f|\mathscr{H}] \equiv 0$, resp. $\mathbf{E}[g|\mathscr{H}] \equiv 0$, one has:

$$|\mathbf{E}[fg]| \leqslant \varepsilon \,\mathrm{Sd}(f)\,\mathrm{Sd}(g). \tag{156}$$

We let the reader check that with that definition, for $X$, $Y$ and $Z$ random variables, $\{X:Y\}_Z = \{\sigma(X):\sigma(Y)\}_{\sigma(Z)}$.

*3.1.3 Remark.* The ordinary correlation can be seen as a particular case of subjective correlation, since $\{\mathscr{F}:\mathscr{G}\} = \{\mathscr{F}:\mathscr{G}\}_{\mathscr{O}}$ for $\mathscr{O} = \{\emptyset,\Omega\}$ the trivial $\sigma$-field.

*3.1.4 Remark.* Warning! Writing that $\{X:Y\}_Z \leqslant \varepsilon$ does *not* imply that for all subset $C$ of the range of $Z$, $X$ and $Y$ are $\varepsilon$-decorrelated under $Law(X,Y|Z \in C)$: see Examples 3.1.8 and 3.1.9 below.

*3.1.5 Remark.* Warning again! There is no general inequality between $\{X:Y\}$ and $\{X:Y\}_Z$: see Examples 3.1.7 and 3.1.8 below.

[*]The conditional laws $Law(\cdot|Z=z)$ are only defined up to $Law(Z)$-a.e. equality, whence the need to specify "for $Law(Z)$-almost-all $z$".

*3.1.6 Example.* Let $f:\ \mathbb{R}\to\mathbb{R}_+$ be a nonnegative continuous function with $\int_{\mathbb{R}} f(x)\,dx = 1$ and let $(X,Y,Z)$ be a variable on $\mathbb{R}^3$ with density

$$d\mathbf{P}[(X,Y,Z)=(x,y,z)] = \frac{1}{2\pi} f(z)\exp\big(\sinh z\cdot xy - \tfrac{1}{2}\cosh z\cdot(x^2+y^2)\big)\,dxdydz. \tag{157}$$

Then, conditionally to "$Z=z$", $(X,Y)$ is a Gaussian vector with covariance matrix $\begin{pmatrix}\cosh z & \sinh z\\ \sinh z & \cosh z\end{pmatrix}$, so by Theorem 1.2.6, under the law $\mathbf{P}[\cdot|Z=z]$ one has $\{X:Y\} = |\tanh z|$. Consequently $\{X:Y\}_Z = \sup\{|\tanh z|:\ f(z)>0\}$.

The three following examples show that subjective correlation may behave rather wildly, especially when one changes the $\sigma$-field of reference:

*3.1.7 Example.* Let $X$ and $Y$ be independent variables with uniform law on $\mathbb{R}/\mathbb{Z}$ and let $Z = X+Y$; then $\{X:Y\}_Z = 1$: under $\mathbf{P}[\cdot|Z=z]$ indeed $Y$ is $X$-measurable (and not constant), since $Y \equiv z - X$.

*3.1.8 Example.* Let $\alpha,\beta,\gamma$ be three independent random variables 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 uniform on $\{(0,0),(0,1)\}$, and similarly $X$ and $Y$ are independent conditionally to "$Z=1$", so $\{X:Y\}_Z = 0$. Yet $X$ and $Y$ are not independent since the events "$X\in\{(0,0),(0,1)\}$" and "$Y\in\{(0,0),(0,1)\}$", which are non-trivial under $\mathbf{P}$, are equivalent, so that $\{X:Y\}=1$.

*3.1.9 Example.* Let $X=(X_1,X_2)$ and $Y=(Y_1,Y_2)$ be independent with uniform laws on $\{0,1\}^2$ and define $Z=(X_1,Y_1)$; then one easily checks that $\{X:Y\}_Z = 0$. Now let $Z' = \mathbf{1}_{X_1=Y_1}$, which is $Z$-measurable; one has $\{X:Y\}_{Z'}=1$ since, for instance, under "$Z'=1$" the events "$X_1=0$" and "$Y_1=0$" are non-trivial and equivalent.

Now we define a more restrictive concept of subjective correlation.

**3.1.10 Definition.** A *$\sigma$-metalgebra $\mathscr{M}$* is a set $\{\mathscr{H}:\ \mathscr{H}\in\mathscr{M}\}$ of $\sigma$-algebras which is stable under the "$\bigvee$" operator, i.e. such that for any $\mathscr{M}'\subset\mathscr{M}$, $\bigvee_{\mathscr{H}\in\mathscr{M}'}\mathscr{H}\in\mathscr{M}$.

One can speak of the '$\sigma$-metalgebra spanned by some set of $\sigma$-algebras', as states the following immediate proposition:

**3.1.11 Proposition.** *If $(\mathscr{H}_k)_{k\in K}$ is a set of $\sigma$-algebras, then there is a smallest $\sigma$-metalgebra containing all the $\mathscr{H}_k$, which is*

$$\mathscr{M} = \Big\{\bigvee_{k\in K'}\mathscr{H}_k\ ;\ K'\subset K\Big\}. \tag{158}$$

When one deals with random variables rather than $\sigma$-algebras, one has the following variant of Proposition 3.1.11:

**3.1.12 Proposition.** *Let $(Z_k)_{k\in K}$ be a set of random variables, then the $\sigma$-metalgebra spanned by $\{\sigma(Z_k):\ k\in K\}$ is $\{\sigma(\vec{Z}_{K'}):\ K'\subset K\}$.*

**3.1.13 Definition.** Let $\mathscr{F}$ and $\mathscr{G}$ be $\sigma$-algebras and $\mathscr{M}$ be a $\sigma$-metalgebra. We define the *correlation between $\mathscr{F}$ and $\mathscr{G}$ seen from $\mathscr{M}$* by:

$$\{\mathscr{F}:\mathscr{G}\}_{\mathscr{M}} = \sup_{\mathscr{H}\in\mathscr{M}}\{\mathscr{F}:\mathscr{G}\}_{\mathscr{H}}. \tag{159}$$

*3.1.14 Remark.* Speaking in terms of random variables, if $X$, $Y$ and $(Z_k)_{k\in K}$ are variables, denoting by $\mathscr{M}$ the $\sigma$-metalgebra spanned by the $Z_k$, then $\{X \underset{\cdot}{:} Y\}_{\mathscr{M}}$ is the supremum[†] of the $\{X : Y\}$ when taken under all the laws of kind $\mathbf{P}[\cdot|\vec{Z}_{K'} = \vec{z}_{K'}]$ for $K'$ a subset of $K$ and $z_k, k \in K'$ elements of the respective ranges of the $Z_k$.

Finally, the following proposition gathers some easy properties of relative correlation w.r.t. a $\sigma$-metalgebra:

**3.1.15 Proposition.**

*(i) Call $\mathscr{M}_{\emptyset}$ the trivial $\sigma$-metalgebra, that is, $\mathscr{M}_{\emptyset} = \{\mathscr{O}\}$; then for all $\sigma$-algebras $\mathscr{F}$ and $\mathscr{G}$, $\{\mathscr{F} : \mathscr{G}\} = \{\mathscr{F} : \mathscr{G}\}_{\mathscr{M}_{\emptyset}}$.*

*(ii) If $\mathscr{M} \subset \mathscr{M}'$, then $\{\mathscr{F} : \mathscr{G}\}_{\mathscr{M}} \leqslant \{\mathscr{F} : \mathscr{G}\}_{\mathscr{M}'}$.*

*(iii) Let $\mathscr{F}$ and $\mathscr{G}$ be $\sigma$-algebras, let $\mathscr{M}$ be a $\sigma$-metalgebra, and call $\tilde{\mathscr{M}}$ the $\sigma$-metalgebra spanned by $\mathscr{M}$, $\mathscr{F}$ and $\mathscr{G}$; then $\{\mathscr{F} : \mathscr{G}\}_{\mathscr{M}} = \{\mathscr{F} : \mathscr{G}\}_{\tilde{\mathscr{M}}}$.*

**3.1.16 Definition.** In the sequel, the probabilistic systems which we shall consider will often be made of some 'elementary' variables, say $(X_i)_{i\in I}$. In this case, the so-called *natural $\sigma$-metalgebra* of the system will mean the $\sigma$-metalgebra spanned by the $X_i$.

## 3.2 Simple tensorization

Now we turn to tensorization. First let us deal with 'simple' tensorization, by which I mean that tensorization is performed on only one variable. The main result of this section will be the '$N$ against 1' theorem (Theorem 3.2.2).

The problem considered is the following: Let $I$ be a set and $(X_i)_{i\in I}, Y$ be random variables; call $\mathscr{M}$ the natural $\sigma$-metalgebra of this system, that is, the $\sigma$-metalgebra spanned by the $X_i$ and $Y$ (cf. Definition 3.1.16). Suppose we have bounds $\{X_i : Y\}_{\mathscr{M}} \leqslant \varepsilon_i$ for all $i$; the question is, can we deduce from them a bound on $\{\vec{X}_I : Y\}$? We shall prove that the answer is "yes", and moreover the bound (169) we will give is optimal in some way (see § 3.5).

For pedagogical purpose, let us first state and prove a weaker but easier proposition:

**3.2.1 Proposition.** *With the notation above,*

$$\{\vec{X}_I : Y\} \leqslant \sqrt{\sum_{i\in I} \varepsilon_i^2}. \tag{160}$$

*Proof.* By Proposition 1.1.12 we may assume $I = \{1,\dots,N\}$. Let $f$ and $g$ be centered $L^2$ $\vec{X}$-measurable, resp. $Y$-measurable, functions; our goal is to bound $|\mathbf{E}[fg]|$.

For all $i \in \{0,\dots,N\}$, denote

$$\mathscr{F}_i = \sigma(X_1,\dots,X_i), \tag{161}$$

and for all $i \in \{1,\dots,N\}$,

$$f_i = f^{\mathscr{F}_i} - \mathbf{E}[f|\mathscr{F}_{i-1}]. \tag{162}$$

[†]More precisely it is a *true* supremum (over $K'$) of *essential* suprema (over $\vec{z}_{K'}$).

Then $f=\sum_i f_i$, where each $f_i$ is $\mathscr{F}_i$-measurable and centered w.r.t. $\mathscr{F}_{i-1}$ (i.e., $\mathbf{E}[f_i|\mathscr{F}_{i-1}]\equiv 0$). Consequently, for all $i_0<i_1$ one has $\mathbf{E}[f_{i_0}f_{i_1}]=0$ (since $f_{i_0}$ is $\mathscr{F}_{i_0}$-measurable while $f_{i_1}$ is centered w.r.t. $\mathscr{F}_{i_1-1}\supset\mathscr{F}_{i_0}$) and thus when one expands $\mathrm{Var}(f)=\mathbf{E}\big[(\sum_i f_i)^2\big]$ all the non-diagonal terms vanish, yielding:

$$\mathrm{Var}(f)=\sum_{i=1}^{N}\mathrm{Var}(f_i). \tag{163}$$

Now, the decomposition "$f=\sum_i f_i$" yields

$$\mathbf{E}[fg]=\sum_{i=1}^{N}\mathbf{E}[f_ig], \tag{164}$$

so let us bound the $|\mathbf{E}[f_ig]|$. The law of total expectation gives:

$$\mathbf{E}[f_ig]=\int\mathbf{E}\big[f_ig|X_1=x_1,\dots,X_{i-1}=x_{i-1}\big]\,d\mathbf{P}[x_1,\dots,x_{i-1}]. \tag{165}$$

But under $d\mathbf{P}[\cdot|x_1,\dots,x_{i-1}]$, $f_i$ is $X_i$-measurable and centered while $g$ is $Y$-measurable, moreover under this law $\{X_i:Y\}\leqslant\{X_i:Y\}_{\mathscr{F}_{i-1}}\leqslant\{X_i:Y\}_{\mathscr{M}}\leqslant\varepsilon_i$, so:

$$\big|\mathbf{E}\big[f_ig|Y_1=y_1,\dots,Y_{i-1}=y_{i-1}\big]\big|\leqslant\varepsilon_i\,\mathrm{Sd}\big(f_i|x_1,\dots,x_{i-1}\big)\mathrm{Sd}\big(g|x_1,\dots,x_{i-1}\big). \tag{166}$$

Using the bound $\mathrm{Sd}(h)\leqslant\sqrt{\mathbf{E}[h^2]}$, it follows that:

$$\begin{aligned}|\mathbf{E}[f_ig]|&\leqslant\varepsilon_i\int\sqrt{\mathbf{E}[f_i^2|x_1,\dots,x_{i-1}]}\sqrt{\mathbf{E}[g^2|x_1,\dots,x_{i-1}]}\,d\mathbf{P}[x_1,\dots,x_{i-1}]\\&\underset{\mathrm{CS}}{\leqslant}\varepsilon_i\sqrt{\int\mathbf{E}[f_i^2|x_1,\dots,x_{i-1}]\,d\mathbf{P}[x_1,\dots,x_{i-1}]}\sqrt{\textit{the same for }g}=\varepsilon_i\,\mathrm{Sd}(f)\mathrm{Sd}(g_i).\end{aligned} \tag{167}$$

So, summing (167) for all $i$:

$$|\mathbf{E}[fg]|\leqslant\sum_{i=1}^{N}\varepsilon_i\,\mathrm{Sd}(f_i)\mathrm{Sd}(g)\underset{\mathrm{CS}}{\leqslant}\sqrt{\sum_i\varepsilon_i^2}\sqrt{\sum_i\mathrm{Var}(f_i)}\,\mathrm{Sd}(g)=\sqrt{\sum_i\varepsilon_i^2}\,\mathrm{Sd}(f)\mathrm{Sd}(g). \tag{168}$$

Since (168) is true for all $f,g$, (160) is proved. □

It is striking in Proposition 3.2.1 that the right-hand side of (160) may be greater than 1, which is never the case for a correlation level. Actually there is some 'loss of optimality' in the proof of the proposition when we bound above $\mathrm{Var}(f_i|\mathscr{F}_{i-1})$ by $\mathbf{E}[f_i^2|\mathscr{F}_{i-1}]$, since $\mathbf{E}[f_i^2|\mathscr{F}_{i-1}]-\mathrm{Var}(f_i|\mathscr{F}_{i-1})=\mathbf{E}[f_i|\mathscr{F}_{i-1}]^2$ may be different to 0. We will use a technique for 'recycling' that loss to get the following result, which § 3.5 shall prove to be optimal:

**3.2.2 Theorem** ('$N$ against 1' theorem)**.** *Take the same hypotheses as in Proposition 3.2.1: $\forall i\in I\quad\{X_i:Y\}_{\mathscr{M}}\leqslant\varepsilon_i$, where $\mathscr{M}$ is the natural $\sigma$-metalgebra of the system. Then:*

$$\{\vec{X}_I:Y\}\leqslant\sqrt{1-\prod_{i\in I}(1-\varepsilon_i^2)}. \tag{169}$$

*3.2.3 Remark.* The right-hand side of (169) is the $\bar{\varepsilon}\in[0,1]$ characterized by $1-\bar{\varepsilon}^2=\prod_i(1-\varepsilon_i)^2$.

*3.2.4 Remark.* The right-hand side of (169) is bounded above by $\sqrt{\sum_i \varepsilon_i^2}$, so Theorem 3.2.2 gives back Proposition 3.2.1 as a corollary.

*Proof.* As in the proof of Proposition 3.2.1, let $f$ and $g$ be centered $L^2$ $\vec{X}$-measurable, resp. $Y$-measurable, functions. Assume $I = \{1,\dots,N\}$; denote $\mathscr{F}_i := \sigma(X_1,\dots,X_i)$ and $f_i := f^{\mathscr{F}_i} - \mathbf{E}[f|\mathscr{F}_{i-1}]$. Also denote, for $i \in \{0,\dots,N\}$,

$$g^i := g - \mathbf{E}[g|\mathscr{F}_i]. \tag{170}$$

As before, one has $\mathrm{Var}(f) = \sum_i \mathrm{Var}(f_i)$ and $\mathbf{E}[fg] = \sum_{i=1}^N \mathbf{E}[f_i g]$. But $f_i$ is centered w.r.t. $\mathscr{F}_{i-1}$ while $(g - g^{i-1})$ is $\mathscr{F}_{i-1}$-measurable, so $\mathbf{E}[f_i g] = \mathbf{E}[f_i g^{i-1}]$. Since, conditionally to $\mathscr{F}_{i-1}$, $f_i$ and $g^{i-1}$ are both centered and resp. $X_i$- and $Y$-measurable, the fact that $\{X_i : Y\}_{\mathscr{M}} \leqslant \varepsilon_i$ implies, by the same argument as in the previous proof, that

$$|\mathbf{E}[f_i g^{i-1}]| \leqslant \varepsilon_i \,\mathrm{Sd}(f_i)\mathrm{Sd}(g^{i-1}). \tag{171}$$

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

$$\bar{g}^i = \mathbf{E}[g^{i-1}|\mathscr{F}_i]. \tag{172}$$

Since $g^{i-1} = \bar{g}^i + g^i$, where $\bar{g}^i$ is $\mathscr{F}_i$-measurable while $g^i$ is centered w.r.t. $\mathscr{F}_i$, one has:

$$\mathrm{Var}(g^i) = \mathrm{Var}(g^{i-1}) - \mathrm{Var}(\bar{g}^i). \tag{173}$$

Then, the points consist in making the following observation: for $\mathrm{Var}(g^i)$ to be large (that is, close to $\mathrm{Var}(g^{i-1})$), $\mathrm{Var}(\bar{g}^i)$ has to be small. But in that case $|\mathbf{E}[f_i g]|$ will be small: one has indeed, since $f_i$ is $\mathscr{F}_i$-measurable,

$$|\mathbf{E}[f_i g]| = |\mathbf{E}[f_i g^{i-1}]| = \big|\mathbf{E}\big[f_i (g^{i-1})^{\mathscr{F}_i}\big]\big| = \big|\mathbf{E}\big[f_i \bar{g}^i\big]\big| \underset{\mathrm{CS}}{\leqslant} \mathrm{Sd}(f_i)\mathrm{Sd}(\bar{g}^i). \tag{174}$$

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

$$|\mathbf{E}[f_i g]| \leqslant \varepsilon_i \,\mathrm{Sd}(f_i)\mathrm{Sd}(g^{i-1}); \tag{175}$$

$$\mathrm{Sd}(g^i) = \sqrt{\mathrm{Sd}(g^{i-1})^2 - \mathrm{Sd}(\bar{g}^i)^2}; \tag{176}$$

$$|\mathbf{E}[f_i g]| \leqslant \mathrm{Sd}(f_i)\mathrm{Sd}(\bar{g}^i). \tag{177}$$

Now define $\hat{\varepsilon}_i = |\mathbf{E}[f_i g]|/\mathrm{Sd}(f_i)\mathrm{Sd}(g^{i-1})$, or $\hat{\varepsilon}_i = 0$ if the right-hand side is 0/0. Then (175) ensures that $\hat{\varepsilon}_i \leqslant \varepsilon_i$, and (177) means that $\mathrm{Sd}(\bar{g}^i) \geqslant \hat{\varepsilon}_i \,\mathrm{Sd}(g^{i-1})$, so that (176) yields $\mathrm{Sd}(g^i) \leqslant \sqrt{1 - \hat{\varepsilon}_i^2}\,\mathrm{Sd}(g^{i-1})$. Since $g^0 = g$, one has therefore by induction $\mathrm{Sd}(g^i) \leqslant \prod_{i'=1}^{i-1} \sqrt{1 - \hat{\varepsilon}_{i'}^2}\,\mathrm{Sd}(g)$, so that the decomposition "$\mathbf{E}[fg] = \sum_i \mathbf{E}[f_i g]$" gives:

$$|\mathbf{E}[fg]| \leqslant \sum_{i=1}^N \Big(\hat{\varepsilon}_i \prod_{i'=1}^{i-1} \sqrt{1 - \hat{\varepsilon}_{i'}^2}\Big) \mathrm{Sd}(f_i)\mathrm{Sd}(g). \tag{178}$$

By the Cauchy–Schwarz inequality, it follows that:

$$|\mathbf{E}[fg]| \leqslant \sqrt{\sum_{i=1}^N \hat{\varepsilon}_i^2 \prod_{i'=1}^{i-1} (1 - \hat{\varepsilon}_{i'}^2)}\ \mathrm{Sd}(f)\mathrm{Sd}(g) = \sqrt{1 - \prod_{i=1}^N (1 - \hat{\varepsilon}_i^2)}\ \mathrm{Sd}(f)\mathrm{Sd}(g). \tag{179}$$

Obviously the maximal value for the right-hand side of (179) is when $\hat{\varepsilon}_i = \varepsilon_i$ for all $i$, then yielding (169). □

There is an alternative proof, which is less intuitive but whose reasoning shall be used again in the proof of Theorem 3.3.1:

*Alternative proof of Theorem 3.2.2.* We use the same notation as in the previous proof. As $f$ is $\mathscr{F}$-measurable, $\mathbf{E}[fg] = \mathbf{E}[fg^{\mathscr{F}}]$, so by the Cauchy–Schwarz inequality:

$$|\mathbf{E}[fg]| \leqslant \mathrm{Sd}(f)\,\mathrm{Sd}(g^{\mathscr{F}}). \tag{180}$$

Now, by associativity of variance $\mathrm{Sd}(g^{\mathscr{F}}) = \sqrt{\mathrm{Var}(g) - \mathrm{Var}(g - g^{\mathscr{F}})}$, so by (180) it suffices to prove that

$$\mathrm{Var}\big(g - g^{\mathscr{F}}\big) \geqslant \prod_{i=1}^{N} (1 - \varepsilon_i^2)\,\mathrm{Var}(g). \tag{181}$$

With our notation, $g - g^{\mathscr{F}} = g^N$ and $g = g^0$; we will prove that for all $i \in \{1, \ldots, N\}$,

$$\mathrm{Var}(g^i) \geqslant (1 - \varepsilon_i^2)\,\mathrm{Var}(g^{i-1}). \tag{182}$$

Since $g^{i-1}$ and $g^i$ are centered w.r.t. $\mathscr{F}_{i-1}$, one has

$$\mathrm{Var}(g^{i-1}) = \int \mathrm{Var}(g^{i-1} | x_1, \ldots, x_{i-1})\, d\mathbf{P}[x_1, \ldots, x_{i-1}], \tag{183}$$

with a similar decomposition for $\mathrm{Var}(g^i)$, so that it suffices to prove (182) conditionally to $\mathscr{F}_{i-1}$.

Conditionally to $\mathscr{F}_{i-1}$, $g^{i-1}$ is centered and $Y$-measurable. Moreover, $g^i = g^{i-1} - (g^{i-1})^{\sigma(X_i)}$, so by associativity of variance $\mathrm{Var}(g^i) = \mathrm{Var}(g^{i-1}) - \mathrm{Var}\big((g^{i-1})^{\sigma(X_i)}\big)$, and therefore (182) is equivalent to

$$\mathrm{Var}\big((g^{i-1})^{\sigma(X_i)}\big) \geqslant \varepsilon_i^2\,\mathrm{Var}(g^{i-1}), \tag{184}$$

which follows directly from the assumption "$\{X_i : Y\}_{\mathscr{F}_{i-1}} \leqslant \varepsilon_i$". □

## 3.3 Double tensorization

Simple tensorization as itself is already interesting since it gives an $L^2$-type bound for the correlation between $X$ and $\vec{Y}$, which is better than the $L^1$-type bounds typically obtained by total variation methods. Yet it does not exhaust the full potential of maximal correlations concerning tensorization, since obviously it does not contain results like independent tensorization (cf. § 1.1.e).

The aim of this section is to get sharp tensorization results where we perform tensorizing on *both* sides, without having to assume complete independence like in Theorem 1.1.19. The price to pay is that the techniques involved, though similar in their spirit, will be much more tricky, moreover the bounds obtained will not be completely optimal (see § 3.5).

### 3.3.a 'N against M' tensorization

The following theorem may be considered as the main result of this monograph. As will be explained in § 3.5.b, it 'contains' qualitatively all the other tensorization theorems (i.e. Theorems 1.1.19, 3.2.2 and 3.3.10).

**3.3.1 Theorem** ('$N$ against $M$' theorem). *Let $I$ and $J$ be sets, and let $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ be random variables, the $\sigma$-metalgebra they generate being denoted by $\mathscr{M}$. Suppose for any $i,j$, $\{X_i : Y_j\}_{\mathscr{M}} \leqslant \varepsilon_{ij}$ for some $\varepsilon_{ij} \geqslant 0$, and define the operator*

$$\begin{array}{rccl} \boldsymbol{\varepsilon}: & L^2(J) & \to & L^2(I) \\ & (a_j)_{j\in J} & \mapsto & \left(\sum_{j\in J} \varepsilon_{ij} a_j\right)_{i\in I}, \end{array} \tag{185}$$

*then:*

$$\{\vec{X}_I : \vec{Y}_J\} \leqslant |||\boldsymbol{\varepsilon}||| \wedge 1. \tag{186}$$

*3.3.2 Remark.* On $(\mathbb{R}_+)^{I\times J}$, $|||\boldsymbol{\varepsilon}|||$ is a nondecreasing function of each $\varepsilon_{ij}$.

☛ *As the proof of Theorem 3.3.1 is rather technical, I found it useful to write down how it goes on a concrete example. This is performed in Appendix 3.7, which I suggest the reader to look at in parallel with the proof as a complement.*

To prove Theorem 3.3.1, we will need the following

**3.3.3 Lemma.** *Let $X_1, X_2, \ldots, X_N$ and $Y$ be random variables, call $\mathscr{M}$ their natural $\sigma$-metalgebra, and assume that for all $i \in \{1,\ldots,N\}$,*

$$\{X_i : Y\}_{\mathscr{M}} \leqslant \varepsilon_i. \tag{187}$$

*Let $f$ be an $\vec{L}^2(\vec{X})$ function. For all $0 \leqslant i \leqslant N$, denote $\mathscr{F}_i := \sigma(X_1,\ldots,X_i)$, resp. $\mathscr{F}_i^* := \sigma(X_1,\ldots,X_i,Y)$, and for all $1 \leqslant i \leqslant N$, define*

$$f_i := f^{\mathscr{F}_i} - \mathbf{E}[f|\mathscr{F}_{i-1}], \tag{188}$$

$$\textit{resp.}\quad f_i^* := f^{\mathscr{F}_i^*} - \mathbf{E}[f|\mathscr{F}_{i-1}^*], \tag{189}$$

*and denote by $V_i$ and $V_i^*$ their respective variances. Then, for all $1 \leqslant i \leqslant N$,*

$$V_i^* \geqslant (1-\varepsilon_i^2)V_i - 2\varepsilon_i\sqrt{V_i}\Big(\sum_{i'>i} \varepsilon_{i'}\sqrt{V_{i'}}\Big). \tag{190}$$

*Proof.* For $0 \leqslant i \leqslant N$, define

$$\tilde{f}_i := f - \mathbf{E}[f|\mathscr{F}_i], \tag{191}$$

$$\text{resp.}\quad \tilde{f}_i^* := f - \mathbf{E}[f|\mathscr{F}_i^*], \tag{192}$$

and call $\tilde{V}_i$ and $\tilde{V}_i^*$ their respective variances. One has $\tilde{f}_i = \sum_{i'>i} f_{i'}$, resp. $\tilde{f}_i^* = \sum_{i'>i} f_{i'}^*$. Moreover, by the same argument as in the proof of Proposition 3.2.1, all the $f_i$ are orthogonal (that is, $i_0 \neq i_1 \Rightarrow \mathbf{E}[f_{i_0} f_{i_1}] = 0$), thus

$$\tilde{V}_i = \sum_{i'>i} V_{i'}; \tag{193}$$

similarly,

$$\widetilde{V}_i^* = \sum_{i'>i} V_{i'}^*. \tag{194}$$

In a first step, we observe that for all $i$, $\widetilde{f}_i - \widetilde{f}_i^* = (f - f^{\mathscr{F}_i}) - (f - f^{\mathscr{F}_i^*}) = f^{\mathscr{F}_i^*} - f^{\mathscr{F}_i} = f^{\mathscr{F}_i^*} - (f^{\mathscr{F}_i})^{\mathscr{F}_i^*} = (f - f^{\mathscr{F}_i})^{\mathscr{F}_i^*} = (\widetilde{f}_i)^{\mathscr{F}_i^*}$, which by associativity of variance yields the following

**3.3.4 Claim.**

$$\widetilde{V}_i - \widetilde{V}_i^* = \mathrm{Var}\big((\widetilde{f}_i)^{\mathscr{F}_i^*}\big). \tag{195}$$

Now, the following claim will be the main tool for proving the lemma:

**3.3.5 Claim.** *For all* $1 \leqslant i \leqslant N$*,*

$$\widetilde{V}_{i-1} - \widetilde{V}_{i-1}^* \leqslant \big(\varepsilon_i \sqrt{V_i} + \sqrt{\widetilde{V}_i - \widetilde{V}_i^*}\big)^{2\,[\ddagger]}. \tag{196}$$

Admit temporarily Claim 3.3.5. Since $\widetilde{V}_N = \widetilde{V}_N^* = 0$, (196) applied with $i = N$ gives $\widetilde{V}_{N-1} - \widetilde{V}_{N-1}^* \leqslant \varepsilon_N^2 V_N$, which in turn we can use in (196) with $i = N-1$, and so on, to finally prove by finite (decreasing) induction that, for all $i$,

$$\widetilde{V}_i - \widetilde{V}_i^* \leqslant \Big(\sum_{i'>i} \varepsilon_{i'} \sqrt{V_{i'}}\Big)^2. \tag{197}$$

Now to get (190), we note that $V_i = \widetilde{V}_{i-1} - \widetilde{V}_i$, resp. $V_i^* = \widetilde{V}_{i-1}^* - \widetilde{V}_i^*$, so, using successively the inequalities (196) and (197),

$$\begin{aligned} V_i - V_i^* &= (\widetilde{V}_{i-1} - \widetilde{V}_{i-1}^*) - (\widetilde{V}_i - \widetilde{V}_i^*) \\ &\leqslant \big(\varepsilon_i \sqrt{V_i} + \sqrt{\widetilde{V}_i - \widetilde{V}_i^*}\big)^2 - (\widetilde{V}_i - \widetilde{V}_i^*) = \varepsilon_i^2 V_i + 2\varepsilon_i \sqrt{V_i}\sqrt{\widetilde{V}_i - \widetilde{V}_i^*} \\ &\leqslant \varepsilon_i^2 V_i + 2\varepsilon_i \sqrt{V_i}\Big(\sum_{i'>i} \varepsilon_{i'} \sqrt{V_{i'}}\Big), \end{aligned} \tag{198}$$

which is equivalent to (190). □

*Proof of Claim 3.3.5.* Thanks to Claim 3.3.4, what we have to prove is:

$$\mathrm{Var}\big((\widetilde{f}_{i-1})^{\mathscr{F}_{i-1}^*}\big) \leqslant \big(\varepsilon_i \sqrt{V_i} + \sqrt{\widetilde{V}_i - \widetilde{V}_i^*}\big)^2. \tag{199}$$

By the definition of conditional expectation and the equality case in the Cauchy–Schwarz inequality, (199) is equivalent to saying that for all $\bar{L}^2(\mathscr{F}_{i-1}^*)$ function $g$,

$$|\mathbf{E}[\widetilde{f}_{i-1} g]| \leqslant \big(\varepsilon_i \sqrt{V_i} + \sqrt{\widetilde{V}_i - \widetilde{V}_i^*}\big)\,\mathrm{Sd}(g). \tag{200}$$

So let $g$ be a centered $L^2$ $\mathscr{F}_{i-1}^*$-measurable real function. Since $\widetilde{f}_{i-1} = f_i + \widetilde{f}_i$, $\mathbf{E}[\widetilde{f}_{i-1} g] = \mathbf{E}[f_i g] + \mathbf{E}[\widetilde{f}_i g]$, which two terms we shall bound separately.

For the first term, under $\mathbf{P}[\cdot|\mathscr{F}_{i-1}]$, $f_i$ is centered and only depends on $X_i$, and $g$ only depends on $Y$. Since $\{X_i : Y\}_{\mathscr{F}_{i-1}} \leqslant \{X_i : Y\}_{\mathscr{M}} \leqslant \varepsilon_i$, it follows that

$$|\mathbf{E}[f_i g|\mathscr{F}_{i-1}]| \leqslant \varepsilon_i\, \mathrm{Sd}(f_i|\mathscr{F}_{i-1})\, \mathrm{Sd}(g|\mathscr{F}_{i-1}), \tag{201}$$

[‡]Taking the square root of $(\widetilde{V}_i - \widetilde{V}_i^*)$ is allowed, since that quantity is nonnegative by Claim 3.3.4.

which yields upon integrating:

$$
\begin{aligned}
|\mathbf{E}[f_i g]| \leqslant \varepsilon_i \int \mathrm{Sd}(f_i|\mathscr{F}_{i-1})\mathrm{Sd}(g|\mathscr{F}_{i-1})\,d\mathbf{P}& \\
\underset{\mathrm{CS}}{\leqslant} \varepsilon_i \sqrt{\int \mathrm{Var}(f_i|\mathscr{F}_{i-1})\,d\mathbf{P}}\sqrt{\int \mathrm{Var}(g|\mathscr{F}_{i-1})\,d\mathbf{P}}& \\
= \varepsilon_i\sqrt{V_i}\sqrt{\mathrm{Var}(g)-\mathrm{Var}(g^{\mathscr{F}_{i-1}})} \leqslant \varepsilon_i \sqrt{V_i}\,\mathrm{Sd}(g).& \quad (202)
\end{aligned}
$$

For the second term, under $\mathbf{P}[\cdot|\mathscr{F}_i]$, $g$ only depends on $Y$, and $\mathbf{E}[\tilde{f}_i|Y] \equiv \tilde{f}_i - \tilde{f}_i^*$ as we noticed just before Claim 3.3.4, so $\mathbf{E}[\tilde{f}_i g|\mathscr{F}_i] = \mathbf{E}[(\tilde{f}_i - \tilde{f}_i^*)g|\mathscr{F}_i]$, which yields upon integrating:

$$
|\mathbf{E}[\tilde{f}_i g]| = \left|\mathbf{E}\left[(\tilde{f}_i - \tilde{f}_i^*)g\right]\right| \underset{\mathrm{CS}}{\leqslant} \mathrm{Sd}(\tilde{f}_i - \tilde{f}_i^*)\mathrm{Sd}(g) = \sqrt{\tilde{V}_i - \tilde{V}_i^*}\,\mathrm{Sd}(g), \quad (203)
$$

the last equality coming from Claim 3.3.4. Then it just remains to combine (202) and (203) to get (200). □

*Proof of Theorem 3.3.1.* First, thanks to a by now classical approximation argument we may assume that $I = \{1,\dots,N\}$ and $J = \{1,\dots,M\}$. Denote $\mathscr{F} := \sigma(\vec{X}_I)$, resp. $\mathscr{G} := \sigma(\vec{Y}_J)$; our goal is to prove that for all $f \in \bar{L}^2(\mathscr{F})$, all $g \in \bar{L}^2(\mathscr{G})$, one has $|\mathbf{E}[fg]| \leqslant (|\!|\!|\boldsymbol{\varepsilon}|\!|\!| \wedge 1)\mathrm{Sd}(f)\mathrm{Sd}(g)$. We will use the same trick as in our alternative proof of Theorem 3.2.2: by the definition of conditional expectation and the Cauchy–Schwarz inequality, proving the inequality above is equivalent to showing that for all $f \in \bar{L}^2(\mathscr{F})$,

$$
\mathrm{Var}(f^{\mathscr{G}}) \leqslant (|\!|\!|\boldsymbol{\varepsilon}|\!|\!|^2 \wedge 1)\mathrm{Var}(f), \quad (204)
$$

which, by associativity of variance, is in turn equivalent to:

$$
\mathrm{Var}(f - f^{\mathscr{G}}) \geqslant (1 - |\!|\!|\boldsymbol{\varepsilon}|\!|\!|^2)_+ \mathrm{Var}(f). \quad (205)
$$

For $0 \leqslant i \leqslant N$, resp. $0 \leqslant j \leqslant M$, define $\mathscr{F}_i = \sigma(X_1,\dots,X_i)$, resp. $\mathscr{G}_j = \sigma(Y_1,\dots,Y_j)$. For all $0 \leqslant j \leqslant M$, define

$$
f^j := f - \mathbf{E}[f|\mathscr{G}_j], \quad (206)
$$

and for all $1 \leqslant i \leqslant N$, define moreover

$$
f_i^j := f^{\mathscr{G}_j \vee \mathscr{F}_i} - \mathbf{E}[f|\mathscr{G}_j \vee \mathscr{F}_{i-1}]. \quad (207)
$$

Denote $V^j := \mathrm{Var}(f^j)$, resp. $V_i^j := \mathrm{Var}(f_i^j)$. For fixed $j$, the $f_i^j$ are pairwise orthogonal (again by the argument in the proof of Proposition 3.2.1) and their sum is equal to $f^j$, so:

$$
V^j = \sum_{i=1}^N V_i^j. \quad (208)
$$

Thus, with this notation our goal (205) becomes:

$$
\sum_{i=1}^N V_i^M \geqslant (1 - |\!|\!|\boldsymbol{\varepsilon}|\!|\!|^2)_+ \sum_{i=1}^N V_i^0. \quad (209)
$$

The main tool to prove (209) will be Lemma 3.3.3. Actually the rough formula (190) is quite impratical, so we introduce a *linearized* version of it: for each $1 \leqslant i \leqslant N$ take some $\alpha_i > 0$ (which for the time being is arbitrary), then by the Cauchy–Schwarz inequality, (190) implies that:

$$V_i^* \geqslant (1-\varepsilon_i^2)V_i - \frac{\varepsilon_i V_i}{\alpha_i}\sum_{i'>i}\varepsilon_{i'}\alpha_{i'} - \varepsilon_i\alpha_i\sum_{i'>i}\frac{\varepsilon_{i'}V_{i'}}{\alpha_{i'}}. \tag{210}$$

*3.3.6 Remark.* (210) is devised so that its right-hand side is exactly the same as in (190) if $V_i \propto \alpha_i^2\ \ \forall i$.

Let us reason conditionally to $\mathscr{G}_{j-1}$ for a few lines. Under this conditioning, call $\dot{\mathscr{F}}_i := \sigma(X_1,\ldots,X_i)$, resp. $\dot{\mathscr{F}}_i^* := \sigma(X_1,\ldots,X_i,Y_j)$, and $\dot{f} := f^{j-1}$. Then $\dot{f}$ is an $\bar{L}^2(\vec{X})$ function, so we are in situation of applying Lemma 3.3.3 to the functions

$$\begin{aligned} \dot{f}_i &:= \dot{f}^{\dot{\mathscr{F}}_i} - \mathbf{E}[\dot{f}|\dot{\mathscr{F}}_{i-1}] && (211)\\ \text{and}\qquad \dot{f}_i^* &:= \dot{f}^{\dot{\mathscr{F}}_i^*} - \mathbf{E}[\dot{f}|\dot{\mathscr{F}}_{i-1}^*]. && (212)\end{aligned}$$

But in fact we already know these functions: namely, $\dot{f}_i = f_i^{j-1}$ and $\dot{f}_i^* = f_i^j$. Then, applying the linearized version (210) of Lemma 3.3.3

$$\operatorname{Var}(f_i^j|\mathscr{G}_{j-1}) \geqslant \Big(1-\varepsilon_{ij}^2 - \frac{\varepsilon_{ij}}{\alpha_i}\sum_{i'>i}\varepsilon_{i'j}\alpha_{i'}\Big)\operatorname{Var}(f_i^{j-1}|\mathscr{G}_{j-1}) - \varepsilon_{ij}\alpha_i\sum_{i'>i}\frac{\varepsilon_{i'j}\operatorname{Var}(f_{i'}^{j-1}|\mathscr{G}_{j-1})}{\alpha_{i'}}, \tag{213}$$

whence upon integrating:

$$V_i^j \geqslant (1-\varepsilon_{ij}^2)V_i^{j-1} - \big(\sum_{i'>i}\varepsilon_{i'j}\alpha_{i'}\big)\frac{\varepsilon_{ij}V_i^{j-1}}{\alpha_i} - \varepsilon_{ij}\alpha_i\sum_{i'>i}\frac{\varepsilon_{i'j}V_{i'}^{j-1}}{\alpha_{i'}}. \tag{214}$$

By Equation (214), we have transformed our initial problem into a purely abstract operator problem, *posed in an $L^1$ setting*. To handle it, we need a little notation. Call $L^1(I)$ the set of real functions on $I$, endowed with the $L^1$ norm

$$\|(v_i)_{i\in I}\|_1 := \sum_{i\in I}|v_i|. \tag{215}$$

The dual space of $L^1(I)$ is made of the linear forms $l\colon\ (v_i)_{i\in I}\mapsto \sum l_i v_i$, equipped with the $L^\infty$ norm

$$\|l\|_\infty := \sup_{i\in I}|l_i|. \tag{216}$$

We shall write "$L^1(I) \ni v \geqslant 0$" to mean that all the entries of $v$ are nonnegative, and "$(L^1(I))' \ni l \geqslant 0$" to mean that $(v\geqslant 0) \Rightarrow (lv \geqslant 0)$, which is equivalent to say that all the $l_i$ are nonnegative. Now I claim the following lemma, whose proof is postponed:

**3.3.7 Lemma.** *Suppose given some nonnegative numbers $V_i^j$ for $(i,j)\in\{1,\ldots,N\}\times\{0,\ldots,M\}$, such that Equation (214) is satisfied for all $i,j$. Call $\mathscr{L}$ the nonnegative linear form on $L^1(I)$ defined by*

$$\mathscr{L}v = \sum_{\substack{j\in J\\ i,i'\in I}}\frac{\alpha_{i'}}{\alpha_i}\varepsilon_{ij}\varepsilon_{i'j}v_i, \tag{217}$$

*and assume $\|\mathscr{L}\|_\infty \leqslant 1$, then:*

$$\sum_{i=1}^N V_i^M \geqslant \sum_{i=1}^N V_i^0 - \mathscr{L}((V_i^0)_{i\in I}). \tag{218}$$

Lemma 3.3.7 has the following immediate

**3.3.8 Corollary.** *Suppose given some nonnegative numbers $V_i^j$ for $(i,j) \in \{1,\dots,N\} \times \{0,\dots,M\}$, such that Equation (214) is satisfied for all $i,j$, then:*

$$\sum_{i=1}^N V_i^M \geqslant \Big(1 - \sup_{i\in I} \sum_{\substack{j\in J\\ i'\in I}} \frac{\alpha_{i'}}{\alpha_i} \varepsilon_{ij}\varepsilon_{i'j}\Big)_+ \sum_{i=1}^N V_i^0. \tag{219}$$

Now we finish the proof of Theorem 3.3.1: thanks to Corollary 3.3.8 we have proved that (219) stands true in our situation for *any* choice of positive $(\alpha_i)_{i\in I}$. The last step then consists in optimizing that choice. Denote "$\alpha > 0$" to mean that all the $\alpha_i$ are positive. One has:

$$\begin{aligned}\inf_{\alpha>0} \sup_{i\in I} \sum_{\substack{j\in J\\ i'\in I}} \frac{\alpha_{i'}}{\alpha_i} \varepsilon_{ij}\varepsilon_{i'j} &= \inf\{\lambda \geqslant 0 :\ (\exists \alpha > 0)(\forall i)(\sum_{\substack{j\in J\\ i'\in I}} \varepsilon_{ij}\varepsilon_{i'j}\alpha_{i'} \leqslant \lambda \alpha_i)\} \\ &= \inf\{\lambda \geqslant 0 :\ (\exists \alpha > 0)(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^* \alpha \leqslant \lambda\alpha)\}.\end{aligned} \tag{220}$$

But $\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^*$ is a nonnegative operator on $L^2(I)$ (I mean, when seen as a matrix all its entries are nonnegative), so by Lemma 3.8.1 in appendix:

$$\inf\{\lambda \geqslant 0 :\ (\exists \alpha > 0)(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^* \alpha \leqslant \lambda\alpha)\} = \rho(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^*) = |||\boldsymbol{\varepsilon}|||^2. \tag{221}$$

This ends the proof of Theorem 3.3.1. □

*Proof of Lemma 3.3.7.* We prove Lemma 3.3.7 by induction on $M$. The case $M = 0$ is trivial. Suppose $M \geqslant 1$ and assume the result is true for $(M-1)$. We generalize the notation $\mathscr{L}$ by defining, for $\bullet \in \{\text{␣}, 1, *\}$,

$$\mathscr{L}^\bullet v = \sum_{\substack{j\in J^\bullet\\ i,i'\in I}} \frac{\alpha_{i'}}{\alpha_i} \varepsilon_{ij}\varepsilon_{i'j} v_i, \tag{222}$$

with $J^1 = \{1\}$, resp. $J^* = \{2,\dots,M\}$, so that $\mathscr{L} = \mathscr{L}^1 + \mathscr{L}^*$. Notice that $\|\mathscr{L}^*\|_\infty \leqslant 1$ since $\|\mathscr{L}\|_\infty \leqslant 1$. For all $i \in I$, define

$$\check{V}_i^1 = (1-\varepsilon_{i1}^2)V_i^0 - \frac{\varepsilon_{i1}V_i^0}{\alpha_i} \sum_{i'>i} \varepsilon_{i'1}\alpha_{i'} - \varepsilon_{i1}\alpha_i \sum_{i'>i} \frac{\varepsilon_{i'1}V_{i'}^0}{\alpha_{i'}}, \tag{223}$$

which is the value that $V_i^1$ would take if there were equality in (214) for $j = 1$. With that notation, (214) writes

$$(V_i^1 - \check{V}_i^1)_{i\in I} \geqslant 0, \tag{224}$$

and by induction hypothesis we have:

$$\sum_{i=1}^N V_i^M \geqslant \sum_{i=1}^N V_i^1 - \mathscr{L}^*((V_i^1)_{i\in I}). \tag{225}$$

Introducing the $\check{V}_i^1$, we have therefore the following chain of inequalities:

$$\begin{aligned}
\sum_{i=1}^N V_i^M \underset{(225)}{\geqslant} \sum_{i=1}^N V_i^1 - \mathscr{L}^*((V_i^1)_{i\in I}) \\
&\underset{(224)}{=} \sum_{i=1}^N \check{V}_i^1 + \|(V_i^1 - \check{V}_i^1)_{i\in I}\|_1 - \mathscr{L}^*((\check{V}_i^1)_{i\in I}) - \mathscr{L}^*((V_i^1 - \check{V}_i^1)_{i\in I}) \\
&\underset{\|\mathscr{L}^*\|_\infty \leqslant 1}{\geqslant} \sum_{i=1}^N \check{V}_i^1 - \mathscr{L}^*((\check{V}_i^1)_{i\in I}) = \sum_{i=1}^N V_i^0 - \mathscr{L}^1((V_i^0)_{i\in I}) - \mathscr{L}^*((\check{V}_i^1)_{i\in I}) \\
&\underset{\substack{\check{V}^1 \leqslant V^0 \\ \mathscr{L}^* \geqslant 0}}{\geqslant} \sum_{i=1}^N V_i^0 - \mathscr{L}^1((V_i^0)_{i\in I}) - \mathscr{L}^*((V_i^0)_{i\in I}) = \sum_{i=1}^N V_i^0 - \mathscr{L}((V_i^0)_{i\in I}), \qquad (226)
\end{aligned}$$

so (218) is true for $M$, whence the lemma by induction. □

*3.3.9 Remark.* Our proof of Theorem 3.3.1 handled the $X_i$ and the $Y_j$ 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 $f$. Yet the bound (186) obtained *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'.

### 3.3.b 'ℤ against ℤ' tensorization

The proof of the '$N$ against $M$' theorem was quite more technical than that of the '$N$ against 1' theorem; because of that, in order to get tractable computations we had to use suboptimal inequalities at two places:

- Claim 3.3.5 is suboptimal: it has indeed the same shortcoming as Proposition 3.2.1 exhibited compared to Theorem 3.2.2, namely, it does not 'recycle the losses' occurring when one makes $g$ covariate with both $f_i$ and $\tilde{f}_i$ (cf. the discussion on page 49, just after the proof of Proposition 3.2.1).
- Our linearization technique is suboptimal in general, even after optimizing the $\alpha_i$. In fact, as we said before, Inequality (210) is optimal if and only if one has $V_i \propto \alpha_i$; thus, for (214) to be always optimal, one has to have $V_i^j \propto \alpha_i$ for *all* $j$, with the *same* values for the $\alpha_i$. This would imply that all the sequences $(V_i^j)_{0\leqslant i<n}$ are proportional, which is not true in general.

So, Theorem 3.3.1 is certainly not optimal[§]—this is confirmed by the example of § 3.7. Nonetheless, there is one particular case in which an alternative reasoning yields an optimal bound[¶]. This case is when some symmetries in the decorrelation hypotheses allow us to transform the original two-parameter problem (indexed by $I \times J$) into a one-parameter problem (indexed by ℤ). Let us state and prove the corresponding result:

**3.3.10 Theorem** ('ℤ against ℤ' theorem)**.** *Let $I$ and $J$ be sets isomorphic to ℤ, and let $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ be random variables such that, $\mathscr{M}$ denoting the σ-metalgebra*

[§]Though, as we will see in § 3.5.b, it is 'asymptotically optimal'.

[¶]The bound's being optimal shall be proved by Theorem 3.5.3.

*they generate, one has for all $i,j\in\mathbb{Z}$*

$$\{X_i : Y_j\}_{\mathscr{M}} \leqslant \varepsilon(j-i) \tag{227}$$

*for some function $\varepsilon\colon\ \mathbb{Z}\to[0,1]$.*

*Then*

$$\{\vec{X}_I : \vec{Y}_J\} \leqslant \bar{\varepsilon}, \tag{228}$$

*where $\bar{\varepsilon}\in[0,1]$ is characterized by:*

$$\operatorname{Arcsin}\bar{\varepsilon} = \Big(\sum_{z\in\mathbb{Z}} \operatorname{Arcsin}\varepsilon(z)\Big)\wedge\frac{\pi}{2}. \tag{229}$$

*3.3.11 Remark.* If we apply Theorem 3.3.1 to the situation above, we find $\{\vec{X}_I : \vec{Y}_J\} \leqslant \big(\sum_{z\in\mathbb{Z}}\varepsilon(z)\big)\wedge 1$ [cf. § 3.6.b]. The latter expression is always $\geqslant\bar{\varepsilon}$ because of the concavity of the function $\sin(\cdot\wedge\frac{\pi}{2})$ on $\mathbb{R}_+$, and even $>\bar{\varepsilon}$ if $\bar{\varepsilon}\neq 0,1$; so, when it is applicable, Theorem 3.3.10 is strictly stronger than Theorem 3.3.1.

*Proof.* Let $f$ and $g$ be resp. $\vec{X}_I$- and $\vec{Y}_J$-measurable $\bar{L}^2$ functions. Denote $\mathscr{F} := \sigma(\vec{X})$, resp. $\mathscr{G} := \sigma(\vec{Y})$, and for $i\in\mathbb{Z}$, resp. $j\in\mathbb{Z}$, denote $\mathscr{F}_i := \bigvee_{i'\leqslant i}\sigma(X_{i'})$, resp. $\mathscr{G}_j := \bigvee_{j'\leqslant j}\sigma(Y_{j'})$. For $(i,j)\in\mathbb{Z}\times\mathbb{Z}$, define

$$f_i^j := f^{\mathscr{G}_j\vee\mathscr{F}_i} - \mathbf{E}[f|\mathscr{G}_j\vee\mathscr{F}_{i-1}] \tag{230}$$

and[||]

$$g_j^i := g^{\mathscr{G}_j} - \mathbf{E}[g^{\mathscr{G}_j}|\mathscr{G}_{j-1}\vee\mathscr{F}_i]. \tag{231}$$

Denote $V := \operatorname{Var}(f)$, $W := \operatorname{Var}(g)$, $V_i^j := \operatorname{Var}(f_i^j)$, $W_j^i := \operatorname{Var}(g_j^i)$; also denote

$$S_{ij} := \mathbf{E}[f_i^{j-1} g_j^{i-1}]. \tag{232}$$

Our auxiliary functions were devised so that

**3.3.12 Claim.** *Provided the sum in the right-hand side is absolutely convergent,*

$$\mathbf{E}[fg] = \sum_{i,j} S_{ij}. \tag{233}$$

*Proof of Claim 3.3.12.* First define $\bar{f} := f^{\mathscr{G}}$, so that $\bar{f}$ is $\mathscr{G}$-measurable and $\mathbf{E}[fg] = \mathbf{E}[\bar{f}g]$. For $j\in\mathbb{Z}$, define $g_j := g^{\mathscr{G}_j} - \mathbf{E}[g|\mathscr{G}_{j-1}]$, resp. $\bar{f}_j := \bar{f}^{\mathscr{G}_j} - \mathbf{E}[\bar{f}|\mathscr{G}_{j-1}]$: we have $g = \sum_j g_j$ and $\bar{f} = \sum_j \bar{f}_j$, which are the respective decompositions of $g$ and $\bar{f}$ on the same basis of orthogonal subspaces of $\bar{L}^2(\mathscr{G})$, so $\mathbf{E}[fg] = \sum_j \mathbf{E}[\bar{f}_j g_j]$. The terms of the right-hand side of that formula are unchanged upon replacing $\bar{f}_j$ by $f^{j-1} := f - \mathbf{E}[f|\mathscr{G}_{j-1}]$, since $\mathbf{E}[(f^{j-1}-\bar{f}_j)g_j]$ is zero—the function $(f^{j-1}-\bar{f}_j)$ is indeed equal to $(f-\mathbf{E}[f|\mathscr{G}_j])$, which is centered conditionally to $\mathscr{G}_j$, while $g_j$ is $\mathscr{G}_j$-measurable. In the end we have:

$$\mathbf{E}[fg] = \sum_j \mathbf{E}[f^{j-1}g_j]. \tag{234}$$

[||]Beware: the definition of $g_j^i$ is *not* analogous to the definition of $f_i^j$ !

So in a first step we have decomposed $\mathbf{E}[fg]$ into a sum indexed by $j$. Now we decompose each term of that sum into a sum indexed by $i$. Let us reason conditionally to $\mathscr{G}_{j-1}$. Then $f^{j-1}$ is an $\bar{L}^2(\mathscr{F})$ function and $g_j$ is in $\bar{L}^2(Y_j)$. We compute $\mathbf{E}[f^{j-1}g_j]$ as in the first step of this proof: first we replace $g_j$ by $\bar{g}_j := (g_j)^{\mathscr{F}}$; then we decompose $f^{j-1} = \sum_i f_i^{j-1}$ and $\bar{g}_j = \sum_i \bar{g}_{ji}$, with $f_i^{j-1} := (f^{j-1})^{\mathscr{F}_i} - \mathbf{E}[f^{j-1}|\mathscr{F}_{i-1}]$[*], resp. $\bar{g}_{ji} := \bar{g}_j^{\mathscr{F}_i} - \mathbf{E}[\bar{g}_j|\mathscr{F}_{i-1}]$, and by orthogonal decomposition we get $\mathbf{E}[f^{j-1}g_j] = \sum_i \mathbf{E}[f_i^{j-1}\bar{g}_{ji}]$; then we conclude by saying that $\mathbf{E}[f_i^{j-1}\bar{g}_{ji}]$ is actually equal to $\mathbf{E}[f_i^{j-1}g_j^{i-1}]$, since $(g_j^{i-1} - \bar{g}_{ji})$ is centered conditionally to $\mathscr{F}_i$ while $f_i^{j-1}$ is $\mathscr{F}_i$-measurable. In the end we have obtained

$$\mathbf{E}[f^{j-1}g_j] = \sum_i \mathbf{E}[f_i^{j-1}g_j^{i-1}], \tag{235}$$

which combined with (234) yields (233). □

So we have expressed $\mathbf{E}[fg]$ as a function of the $S_{ij}$. It is also possible to 'read' the values of $V$ and $W$ from the $V_i^j$, resp. from the $W_j^i$, *via* the formulas:

$$V = \lim_{j\to-\infty}^{\nearrow} \Big(\sum_i V_i^j\Big); \tag{236}$$

$$W = \sum_j \Big(\lim_{i\to-\infty}^{\nearrow} W_j^i\Big). \tag{237}$$

Now we are looking for relations between the $V_i^j$, the $W_j^i$ and the $S_{ij}$. The first relation comes from the decorrelation hypothesis: conditionally to $\mathscr{G}_{j-1} \vee \mathscr{F}_{i-1}$, $f_i^{j-1}$ is in $\bar{L}^2(X_i)$, resp. $g_j^{i-1}$ is in $\bar{L}^2(Y_j)$, and $\{X_i : Y_j\} \leqslant \varepsilon(j-i)$, so:

$$|S_{ij}| \leqslant \varepsilon(j-i)\sqrt{V_i^{j-1}W_j^{i-1}}. \tag{238}$$

The second relation means that a large value of $|S_{ij}|$ forces $W_j^i$ to diminish. To state it, we observe that, since $f_i^{j-1}$ is $(\mathscr{G}_{j-1} \vee \mathscr{F}_i)$-measurable, $S_{ij} = \mathbf{E}[f_i^{j-1}(g_j^{i-1})^{\mathscr{G}_{j-1}\vee\mathscr{F}_i}]$, so by the Cauchy–Schwarz inequality $|S_{ij}| \leqslant \mathrm{Sd}(f_i^{j-1})\mathrm{Sd}\big((g_j^{i-1})^{\mathscr{G}_{j-1}\vee\mathscr{F}_i}\big)$. Moreover, since $g_j^{i-1} - (g_j^{i-1})^{\mathscr{G}_{j-1}\vee\mathscr{F}_i} = g_j^i$, one has by orthogonality $\mathrm{Var}\big((g_j^{i-1})^{\mathscr{G}_{j-1}\vee\mathscr{F}_i}\big) = \mathrm{Var}(g_j^{i-1}) - \mathrm{Var}(g_j^i)$, so our inequality becomes

$$|S_{ij}| \leqslant \sqrt{V_i^{j-1}}\sqrt{W_j^{i-1} - W_j^i} \tag{239}$$

(where it is understood that $W_j^i \leqslant W_j^{i-1}$), or more eloquently

$$W_j^i \leqslant W_j^{i-1} - (S_{ij})^2/V_i^{j-1} \tag{240}$$

provided $V_i^{j-1} > 0$.

[*]Notation is consistent: this $f_i^{j-1}$ is indeed the same as the $f_i^{j-1}$ defined by (230), since we are reasoning conditionally to $\mathscr{G}_{j-1}$.

The third and last relation means, on the other hand, that a large value of $|\sum_{i'>i} S_{i'j}|$ forces $\sum_{i'>i} V_i^j$ to diminish. To state it, we denote

$$\tilde{f}_i^j := f - f^{\mathscr{G}_j \vee \mathscr{F}_i} = \sum_{i'>i} f_{i'}^j, \tag{241}$$

whose variance is $\mathrm{Var}(\tilde{f}_i^j) = \sum_{i'>i} \mathrm{Var}(f_{i'}^j)$ since the $f_{i'}^j$ are pairwise orthogonal. One has

$$\sum_{i'>i} S_{i'j} = \sum_{i'>i} \mathbf{E}[f_{i'}^{j-1} g_j^{i'-1}] = \sum_{i'>i} \mathbf{E}[f_{i'}^{j-1} g_j^i] = \mathbf{E}[\tilde{f}_i^{j-1} g_j^i] = \mathbf{E}[(\tilde{f}_i^{j-1})^{\mathscr{G}_j \vee \mathscr{F}_i} g_j^i], \tag{242}$$

so by the Cauchy–Schwarz inequality,

$$\Big|\sum_{i'>i} S_{i'j}\Big| \leqslant \mathrm{Sd}\big((\tilde{f}_i^{j-1})^{\mathscr{G}_j \vee \mathscr{F}_i}\big)\mathrm{Sd}(g_j^i). \tag{243}$$

Since $\tilde{f}_i^{j-1} - (\tilde{f}_i^{j-1})^{\mathscr{G}_j \vee \mathscr{F}_i} = \tilde{f}_i^j$, one has by orthogonality

$$\mathrm{Var}\big((\tilde{f}_i^{j-1})^{\mathscr{G}_j \vee \mathscr{F}_i}\big) = \mathrm{Var}(\tilde{f}_i^{j-1}) - \mathrm{Var}(\tilde{f}_i^j), \tag{244}$$

so our inequality becomes

$$\Big|\sum_{i'>i} S_{i'j}\Big| \leqslant \sqrt{\sum_{i'>i} V_{i'}^{j-1} - \sum_{i'>i} V_{i'}^j}\sqrt{W_j^i}, \tag{245}$$

or more eloquently:

$$\sum_{i'>i} V_{i'}^j \leqslant \sum_{i'>i} V_{i'}^{j-1} - \Big(\sum_{i'>i} S_{i'j}\Big)^2 / W_j^i. \tag{246}$$

So, we have transformed our initial probabilistic problem into the following analytic one: let $\mathbf{A}$ 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 $S_{ij}$, satisfying (238), (239) and (245)—we will say such an array is *correct*. We define $V$ by (236) and $W$ by (237), and we set $S = \sum_{i,j} S_{ij}$ (provided it makes sense); our goal is to get a bound of the form "$|S| \leqslant \bar{\varepsilon}\sqrt{VW}$", with $\bar{\varepsilon}$ only depending on $\varepsilon(\cdot)$.

Note that *A priori* some problems of summability can arise from $\mathbf{A}$'s being infinite, for instance to check (246) or to define $S$. However, in the situations which are of interest to us, we can restrict to cases in which $\mathbf{A}$ is of nice particular form. To do this, we first approximate $f$ in $\bar{L}^2(\vec{X})$, resp. $g$ in $\bar{L}^2(\vec{Y})$, by a function depending only on a finite number of $X_i$, resp. of $Y_j$ —say, we assume $f$ is $\vec{X}_{\dot{I}}$-measurable and $g$ is $\vec{Y}_{\dot{J}}$-measurable for finite $\dot{I} \subset I, \dot{J} \subset J$. Then, we define a new model $(\tilde{X}_i)_{i\in\mathbb{Z}}, (\tilde{Y}_j)_{j\in\mathbb{Z}}$ by $\tilde{X}_i = X_i$ for $i \in \dot{I}$, resp. $\tilde{Y}_j = Y_j$ for $j \in \dot{J}$, and $\tilde{X}_i, \tilde{Y}_j = \partial$ for $i \notin \dot{I}, j \notin \dot{J}$, $\partial$ being some cemetery point. This new model still gives a correct array, for which $S/\sqrt{VW}$ is arbitrarily close to the initial value of $\mathbf{E}[fg]/\mathrm{Sd}(f)\mathrm{Sd}(g)$; and the new array is of the following form, which we will call *compact*, for which all the quantities of interest are well defined:

- $V_i^j$ is zero as soon as $i \notin \dot{I}$, and it does not depend on $j$ for $j < \min \dot{J}$, nor for $j \geqslant \max \dot{J}$;
- Similarly, $W_j^i$ is zero as soon as $j \notin \dot{J}$, and it does not depend on $i$ for $i < \min \dot{I}$, nor for $i \geqslant \max \dot{I}$;

- $S_{ij}$ is zero as soon as $(i,j) \notin \dot{I} \times \dot{J}$. (This condition automatically follows from the first two if the array is correct).

We define the following operations on arrays:

**3.3.13 Definition.**

- For $z \in \mathbb{Z}$, we define the *translation operator* $\tau^z$ on arrays such that, if the entries of $\mathbf{A}$ at $(i,j)$ are $V_i^j, W_j^i, S_{ij}$, the entries of $\tau^z\mathbf{A}$ at $(i,j)$ are $V_{i+z}^{j+z}, W_{j+z}^{i+z}, S_{(i+z)(j+z)}$.
- For $\grave{\mathbf{A}}$ and $\acute{\mathbf{A}}$ two arrays with entries $\grave{V}_i^j, \grave{W}_j^i, \grave{S}_{ij}$, resp. $\acute{V}_i^j$,etc., for $\alpha, \beta$ two real numbers, we define the linear combination $\alpha\grave{\mathbf{A}} + \beta\acute{\mathbf{A}}$ as the array with entries $\alpha\grave{V}_i^j + \beta\acute{V}_i^j, \alpha\grave{W}_j^i + \beta\acute{W}_j^i$,etc..

**3.3.14 Lemma.** *Correct arrays are stable by translations and by nonnegative linear combinations, i.e., if* $\mathbf{A}$ *and* $\mathbf{B}$ *are correct arrays, then for all* $z \in \mathbb{Z}$ *and* $\alpha, \beta \geqslant 0$, $\tau^z\mathbf{A}$ *and* $\alpha\mathbf{A} + \beta\mathbf{B}$ *are correct too.*

*Proof of Lemma 3.3.14.* Recall that being correct means satisfying (238), (239) and (245). These conditions are trivially stable by multiplication by a nonnegative constant and by translation[†]. It remains to see that they are stable by addition. The technique being the same for all three inequalities, we just treat the case of (239). Stability of this condition by addition is a consequence of the following inequality (which is in fact a particuliar case of the *Brunn–Minkowski inequality*, see [17]):

**3.3.15 Lemma.** *For all* $a_1, b_1, a_2, b_2 \geqslant 0$,

$$\sqrt{(a_1 + a_2)(b_1 + b_2)} \geqslant \sqrt{a_1 b_1} + \sqrt{a_2 b_2}. \tag{247}$$

*Proof of Lemma 3.3.15.* Take squares on both sides of (247) and notice that $(a_1 + a_2)(b_1 + b_2) - (\sqrt{a_1 b_1} + \sqrt{a_2 b_2})^2 = a_1 b_2 + a_2 b_1 - 2\sqrt{a_1 b_1 a_2 b_2} = (\sqrt{a_1 b_2} - \sqrt{a_2 b_1})^2 \geqslant 0$.
□

For $\grave{\mathbf{A}}$ and $\acute{\mathbf{A}}$ two correct arrays satisfying (239), applying (247) with $a_1 = \acute{V}_i^{j-1}, a_2 = \grave{V}_i^{j-1}, b_1 = \acute{W}_j^{i-1} - \acute{W}_j^i, b_2 = \grave{W}_j^{i-1} - \grave{W}_j^i$, we get:

$$\begin{aligned} |\grave{S}_{ij} + \acute{S}_{ij}| \leqslant |\grave{S}_{ij}| + |\acute{S}_{ij}| &\leqslant \sqrt{\grave{V}_i^{j-1}}\sqrt{\grave{W}_j^{i-1} - \grave{W}_j^i} + \sqrt{\acute{V}_i^{j-1}}\sqrt{\acute{W}_j^{i-1} - \acute{W}_j^i} \\ &\leqslant \sqrt{\grave{V}_i^{j-1} + \acute{V}_i^{j-1}}\sqrt{(\grave{W}_j^{i-1} + \acute{W}_j^{i-1}) - (\grave{W}_j^i + \acute{W}_j^i)}, \end{aligned} \tag{248}$$

so (239) is still valid for $(\grave{\mathbf{A}} + \acute{\mathbf{A}})$. □

Now, thanks to Lemma 3.3.14 we will reduce our problem on $(\mathbb{Z} \times \mathbb{Z})$-arrays into a problem on $\mathbb{Z}$-arrays. Suppose $\mathbf{A}$ is a correct array with certain values of $V$, $W$ and $S$. Then, for $k \geqslant 0$, the array

$$\mathbf{A}_k = \frac{1}{2k+1}\sum_{z=-k}^{k} \tau^z\mathbf{A} \tag{249}$$

is correct too, with the same values of $V$, $W$ and $S$ as $\mathbf{A}$. Now when $k \to \infty$, $\mathbf{A}_k$ 'looks more and more like a Toeplitz array', that is, an array whose entries at $(i,j)$ only depend on $(j - i)$. To state it rigorously, we need some definitions:

[†]Getting stability of Condition (238) by translation is actually the only place where the symmetries of the problem are used.

**3.3.16 Definition.**

- Here, a *Toeplitz array* will mean a $\mathbb{Z}\times\mathbb{Z}$ array whose entries at $(i,j)$ only depend on $(j-i)$. For such an array, for $z\in\mathbb{Z}$ we denote by $V_{(z)},W_{(z)},S_{(z)}$ the quantities characterized by $V_i^j=V_{(j-i)}$, etc..
- Actually we can always assume our Toeplitz array is *Toeplitz compact*, which means that there exists some $z^-\leqslant z^+$ such that:
  - $V_{(z)}$ does not depend on $z$ for $z<z^-$, nor for $z\geqslant z^+$;
  - $W_{(z)}$ does not depend on $z$ for $z\leqslant z^-$, nor for $z>z^+$;
  - $S_{(z)}$ is zero as soon as $z<z^-$ or $z>z^+$.
- For a compact Toeplitz array, we define $v,w,s$ as 'renormalized versions' of $V,W,S$:

$$\begin{aligned} v &:= V_{(z<z^-)}; &(250)\\ w &:= W_{(z>z^+)}; &(251)\\ s &:= \sum_{z\in\mathbb{Z}} S_{(z)}. &(252)\end{aligned}$$

- A Toeplitz array is said to be *correct* if it is correct when seen as an ordinary array. For a Toeplitz array, Equations (238), (240) and (246) become respectively[‡]:

$$\begin{aligned} |S_{(z)}| &\leqslant \varepsilon(z)\sqrt{V_{(z-1)}W_{(z+1)}}; &(253)\\ W_{(z)} &\leqslant W_{(z+1)}-S_{(z)}^2/V_{(z-1)}; &(254)\\ V_{(z-1)} &\leqslant v-\Big(\sum_{z'<z} S_{(z')}\Big)^2/W_{(z)}. &(255)\end{aligned}$$

With that vocabulary, our informal statement can be made precise: let $\mathbf{A}$ be a compact correct array with entries $V_i^j,W_j^i,S_{ij}$, and associated quantities $V,W,S$, and define the arrays $\mathbf{A}_k$ by (249). Then when $k\to\infty$ one has $(2k+1)\mathbf{A}_k\to\bar{\mathbf{A}}$ (in the sense that each entry of $(2k+1)\mathbf{A}_k$ converges to the corresponding entry of $\bar{\mathbf{A}}$), where $\bar{\mathbf{A}}$ is the Toeplitz array with entries $\bar{V}_i^j,\bar{W}_j^i,\bar{S}_{ij}$ defined by:

$$\begin{aligned} \bar{V}_{(z)} &= \sum_{j-i=z} V_i^j; &(256)\\ \bar{W}_{(z)} &= \sum_{j-i=z} W_j^i; &(257)\\ \bar{S}_{(z)} &= \sum_{j-i=z} S_{ij}. &(258)\end{aligned}$$

This array $\bar{\mathbf{A}}$ is Toeplitz compact with $z^-=\min\dot{J}-\max\dot{I}$, resp. $z^+=\max\dot{J}-\min\dot{I}$, and the quantities (250)–(252) for $\bar{\mathbf{A}}$ are:

$$\begin{aligned} \bar{v} &= V; &(259)\\ \bar{w} &= W; &(260)\\ \bar{s} &= S. &(261)\end{aligned}$$

Moreover $\bar{\mathbf{A}}$ is correct, because all the $(2k+1)\mathbf{A}_k$ are, and being correct is clearly conserved by array convergence.

[‡]Note that the way (255) follows from (246) is rather tricky, because it appears a difference between two infinite quantities, which has to be 'renormalized' in the convenient way.

The consequence of this statement is the following claim, which achieves the reduction to a '$\mathbb{Z}$-indexed' problem I alluded to a few lines above:

**3.3.17 Claim.** *The supremum of $|S|/\sqrt{VW}$ for correct arrays is not greater than the supremum of $|s|/\sqrt{vw}$ for correct Toeplitz arrays.*

So we have to study (compact) correct Toeplitz arrays. Consider such an array. Denote $\theta(z) := \operatorname{Arcsin} \varepsilon(z)$; then (253) can be rewritten:

$$\exists \hat{\theta}(z) \in [\pm\theta(z)] \qquad S_{(z)} = \sin\hat{\theta}(z) \cdot \sqrt{V_{(z-1)} W_{(z+1)}}. \tag{262}$$

Now, notice that for fixed values of the $V_{(z)}$, the $S_{(z)}$ and $w$, if we have values $W_{(z)}$ such that (253)–(255) are satisfied, we can modify those $W_{(z)}$ so that (254) becomes an equality for all $z$, an operation which keeps (253) and (255) true since it can only make the $W_{(z)}$ increase. So we can suppose that (254) actually is an equality, i.e. that for all $z \in \mathbb{Z}$,

$$W_{(z)} = w \prod_{z' \geqslant z} \cos^2 \hat{\theta}(z'). \tag{263}$$

Then it remains to integrate (255). For $z \in \mathbb{Z}$, denote

$$\Gamma(z) := \sum_{z'<z} \Big( \sin\hat{\theta}(z') \cdot \prod_{z'<z''<z} \cos\hat{\theta}(z'') \cdot \sqrt{V_{(z'-1)}} \Big), \tag{264}$$

so that (255) becomes:

$$V_{(z-1)} \leqslant v - \Gamma(z)^2. \tag{265}$$

$\Gamma(\cdot)$ satisfies the recursion equation

$$\Gamma(z+1) = \sin\hat{\theta}(z)\sqrt{V_{(z-1)}} + \cos\hat{\theta}(z)\Gamma(z), \tag{266}$$

so by (265):

$$|\Gamma(z+1)| \leqslant \sin|\hat{\theta}(z)| \sqrt{v - \Gamma(z)^2} + \cos\hat{\theta}(z) |\Gamma(z)|. \tag{267}$$

From (267), we will now prove that for all $z \in \mathbb{Z}$:

$$|\Gamma(z)| \leqslant \sin\Big(\frac{\pi}{2} \wedge \sum_{z'<z} \theta(z')\Big) \sqrt{v}. \tag{268}$$

Indeed, (268) is equivalent to saying that there exists some $\eta(z) \in [0, \sum_{z'<z} \theta(z')]$ such that $|\Gamma(z)| = \sin\eta(z)\sqrt{v}$, which we prove by induction. First, since our Toeplitz array was supposed compact, $\forall z < z^-$ $\hat{\theta}(z) = 0$, so the formula is true for $z \leqslant z^-$ with $\eta(z) = 0$. Next, if the formula is true for $z$, then (267) yields

$$|\Gamma(z)| \leqslant \big(\sin|\hat{\theta}(z)| \cos\eta(z) + \cos|\hat{\theta}(z)| \sin\eta(z)\big)\sqrt{v} = \sin\big(\eta(z) + |\hat{\theta}(z)|\big)\sqrt{v}, \tag{269}$$

where $\eta(z) + |\hat{\theta}(z)| \leqslant \sum_{z'<z} \theta(z') + \theta(z) = \sum_{z'<z+1} \theta(z')$, so the formula is true for $(z+1)$, which ends the induction.

To conclude, we write that $s = \sum_z S_{(z)} = \Gamma(z > z^+)\sqrt{w}$. But by (268), $|\Gamma(z > z^+)| \leqslant \sin\bar{\varepsilon} \cdot \sqrt{v}$, so in the end:

$$|s| \leqslant \sin\bar{\varepsilon} \cdot \sqrt{vw}, \tag{270}$$

*quod erat demonstrandum.* □

**3.3.18 Corollary** ('$\mathbb{Z}^n$ against $\mathbb{Z}^n$' theorem). *Let $n \geqslant 1$; let $(X_x)_{x\in\mathbb{Z}^n}$ and $(Y_y)_{y\in\mathbb{Z}^n}$ be random variables, and assume there exists a function $\varepsilon\colon \mathbb{Z}^n \to [0,1]$ such that for all $x,y \in \mathbb{Z}^n$,*

$$\{X_x : Y_y\}_{\mathcal{M}} \leqslant \varepsilon(y-x), \tag{271}$$

*$\mathcal{M}$ being the natural $\sigma$-metalgebra of the system. Then $\{\vec{X} : \vec{Y}\} \leqslant \bar{\varepsilon}$, where $\bar{\varepsilon}$ the number in $[0,1]$ such that*

$$\operatorname{Arcsin}(\bar{\varepsilon}) = \Big(\sum_{v\in\mathbb{Z}^n} \operatorname{Arcsin}\varepsilon(v)\Big) \wedge \frac{\pi}{2}. \tag{272}$$

*Proof.* To alleviate notation, we define the 'arcsin-sum' as the binary operation $\tilde{+}$: $[0,1]^2 \to [0,1]$ defined by:

$$a \tilde{+} b = \sin\Big(\big(\operatorname{Arcsin} a + \operatorname{Arcsin} b\big) \wedge \frac{\pi}{2}\Big). \tag{273}$$

$\tilde{+}$ is associative, commutative and nondecreasing, so it can be extended into an $\infty$-ary operator $\tilde{\sum}$; with this notation, (272) merely writes $\bar{\varepsilon} = \tilde{\sum}_{v\in\mathbb{Z}^n} \varepsilon(v)$.

Let $(\mathbf{e}_1,\dots,\mathbf{e}_n)$ be a $\mathbb{Z}$-basis of $\mathbb{Z}^n$. For $1 \leqslant r \leqslant n$, we identify $\mathbb{Z}^r$ with $\mathbb{Z}\mathbf{e}_1 \oplus \mathbb{Z}\mathbf{e}_2 \oplus \cdots \oplus \mathbb{Z}\mathbf{e}_r$; we also denote $\mathbb{Z}_r^{\perp} := \mathbb{Z}\mathbf{e}_{r+1} \oplus \cdots \oplus \mathbb{Z}\mathbf{e}_n$. What we will prove is actually the following

**3.3.19 Claim.** *For all $r \in \{1,\dots,n\}$, all $x,y \in \mathbb{Z}_r^{\perp}$,*

$$\{\vec{X}_{x+\mathbb{Z}^r} : \vec{Y}_{y+\mathbb{Z}^r}\}_{\mathcal{M}} \leqslant \tilde{\sum}_{v\in\mathbb{Z}^r} \varepsilon(y-x+v). \tag{274}$$

The statement of the lemma then corresponds to the claim for $r = n$.

We prove Claim 3.3.19 by induction on $r$. The case $r = 1$ is merely Theorem 3.3.10[§]. Now let us show how to go from the case $r-1$ to the case $r$ for $r > 1$:

Take $x,y \in \mathbb{Z}_r^{\perp}$. We notice that

$$\vec{X}_{x+\mathbb{Z}^r} = (\vec{X}_{x+i\mathbf{e}_r+\mathbb{Z}^{r-1}})_{i\in\mathbb{Z}}, \tag{275}$$

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

$$\{\mathbf{X}_i : \mathbf{Y}_j\}_{\mathcal{M}} \leqslant \tilde{\sum}_{v\in\mathbb{Z}^{r-1}} \varepsilon\big(y-x+(j-i)\mathbf{e}_r+v\big). \tag{276}$$

Since the right-hand side of (276) only depends on $(j-i)$, we can apply Theorem 3.3.10 to the $\mathbf{X}_i$ and the $\mathbf{Y}_j$, which yields

$$\{\vec{X}_{x+\mathbb{Z}^r} : \vec{Y}_{y+\mathbb{Z}^r}\}_{\mathcal{M}} \leqslant \tilde{\sum}_{z\in\mathbb{Z}}\Big(\tilde{\sum}_{v\in\mathbb{Z}^{r-1}} \varepsilon\big(y-x+z\mathbf{e}_r+v\big)\Big) = \tilde{\sum}_{v\in\mathbb{Z}^r} \varepsilon(y-x+v), \tag{277}$$

i.e. (274). □

[§]More precisely, it is the subjective version of that theorem, cf. § 3.4.b.

## 3.4 Generalizations of the tensorization results

### 3.4.a Minimal Hypotheses

When reading the proofs of the tensorization theorems, you may have noticed that taking the decorrelation hypotheses w.r.t. the whole $\sigma$-metalgebra of the system was a needlessly strong assumption. Actually each decorrelation hypothesis can be stated relatively to only *one* $\sigma$-algebra, in the following way:

- For Theorem 3.2.2, one needs only assume that for all $i \in I$, $X_i$ and $Y$ are $\varepsilon_i$-decorrelated when seen from $\sigma\big((X_{i'})_{i'<i}\big)$;
- For Theorems 3.3.1 and 3.3.10, one needs only assume that $X_i$ and $Y_j$ are $\varepsilon_{ij}$-decorrelated (or $\varepsilon(j-i)$-decorrelated) when seen from $\sigma\big((X_{i'})_{i'<i},(Y_{j'})_{j'<j}\big)$.

In practice it is rare that one can bound above $\{X_i : Y\}_{\vec{X}_{\{i'<i\}}}$ or $\{X_i : Y_j\}_{(\vec{X}_{\{i'<i\}},\vec{Y}_{\{j'<j\}})}$ more sharply than $\{X : Y_i\}_{\mathscr{M}}$, resp. $\{X_i : Y_j\}_{\mathscr{M}}$; yet it is worth remembering that the 'genuine' decorrelation hypotheses are weaker than those we wrote, especially when one gets interested in optimality issues (cf. § 3.5).

*3.4.1 Remark.* In our tensorization proofs we took $I$ and $J$ finite; yet those proofs, and therefore everything in this subsection, remain valid if we take for $I$ or $J$ any (countable) well-ordered set, in particular if $I$ or $J$ is $\mathbb{N}$.

### 3.4.b Subjective versions of the theorems

In the tensorization theorems I stated, the decorrelation hypotheses were given with regard to the natural $\sigma$-metalgebra $\mathscr{M}$ of the system, while the results were given in terms of 'objective' (I mean, not subjective) decorrelations. Yet actually it can be shown that our results are still valid w.r.t. $\mathscr{M}$ —or even w.r.t. any sharper $\sigma$-metalgebra $\mathscr{N} \supset \mathscr{M}$, provided decorrelation hypotheses are stated w.r.t. $\mathscr{N}$. As an example, let us state and prove the subjective result corresponding to Theorem 3.2.2:

**3.4.2 Corollary.** *Let $X$, $(Y_i)_{i\in I}$ and $(Z_\theta)_{\theta\in\Theta}$ be random variables, and call $\mathscr{N}$ the $\sigma$-metalgebra they span. Suppose we have bounds $\{X : Y_i\}_{\mathscr{N}} \leqslant \varepsilon_i$ for all $i \in I$; then:*

$$\{X : \vec{Y}_I\}_{\mathscr{N}} \leqslant \sqrt{1 - \prod_{i\in I}(1-\varepsilon_i^2)}. \tag{278}$$

*Proof.* Up to making up copies of $I$ and $\Theta$, we can assume that $\{0\}$, $I$ and $\Theta$ are disjoint, which allows us to denote $Z_0 := X$ and $Z_i := Y_i$ for $i \in I$, so that $\mathscr{N}$ is the $\sigma$-metalgebra spanned by the $Z_\theta$ for $\theta \in \bar{\Theta} := \{0\} \uplus I \uplus \Theta$. Then (278) means that for all $\Xi \subset \bar{\Theta}$, for (almost-)all $\vec{z}_\Xi$, one must have:

$$\{X : \vec{Y}_I\} \leqslant \sqrt{1 - \prod_{i\in I}(1-\varepsilon_i^2)} \quad \text{under the law } \mathbf{P}[\cdot|\vec{Z}_\Xi = \vec{z}_\Xi]. \tag{279}$$

So, Corollary 3.4.2 will ensue from Theorem 3.2.2 provided we can prove that, denoting by $\mathscr{M}$ the $\sigma$-metalgebra spanned by $X$ and the $Y_i$, one has for all $i \in I$:

$$\{X : Y_i\}_{\mathscr{M}} \leqslant \varepsilon_i \quad \text{under the law } \mathbf{P}[\cdot|\vec{Z}_\Xi = \vec{z}_\Xi]. \tag{280}$$

But under a law $P$, saying that $\{X:Y_i\}_{\mathcal{M}} \leqslant \varepsilon_i$ means that for all $\Upsilon \subset \{0\} \uplus I$, for (almost-)all $\vec{z}'_\Upsilon$, one has $\{X:Y_i\} \leqslant \varepsilon_i$ under the law $P[\cdot|\vec{Z}_\Upsilon = \vec{z}'_\Upsilon]$. So, for $P = \mathbf{P}[\cdot|\vec{Z}_\Xi = \vec{z}_\Xi]$, (280) means that, for all $\vec{z}'_\Upsilon$:

$$\{X:Y_i\} \leqslant \varepsilon_i \quad \text{under the law } \mathbf{P}[\cdot|\vec{Z}_\Xi = \vec{z}_\Xi \text{ and } \vec{Z}_\Upsilon = \vec{z}'_\Upsilon]. \tag{281}$$

In Formula (281) we can assume that $z_\theta$ and $z'_\theta$ coincide for all $\theta \in \Xi \cap \Upsilon$, since otherwise the event "$\vec{Z}_\Xi = \vec{z}_\Xi$ and $\vec{Z}_\Upsilon = \vec{z}'_\Upsilon$" would be empty and there would be nothing to say. Then "$\vec{Z}_\Xi = \vec{z}_\Xi$ and $\vec{Z}_\Upsilon = \vec{z}'_\Upsilon$" is of the form "$\vec{Z}_{\Xi\cup\Upsilon} = \vec{z}_{\Xi\cup\Upsilon}$", where $\Xi \cup \Upsilon \subset \bar{\Theta}$, so that (281) follows directly from the hypothesis $\{X:Y_I\}_{\mathcal{N}} \leqslant \varepsilon_i$. □

# 3.5 Optimality

## 3.5.a Exact Optimality

With the minimal hypotheses stated in § 3.4.a, Theorems 3.2.2 and 3.3.10 are optimal:

**3.5.1 Theorem.** *The bound (169) in Theorem 3.2.2 is optimal, in the following sense: for any integer $N$, for all $(\varepsilon_i)_{1\leqslant i\leqslant N}$ in $[0,1]^N$, one can find random variables $X_1,\dots,X_N,Y$ such that for all $i \in \{1,\dots,N\}$,*

$$\{X_i:Y\}_{\vec{X}_{\{i'<i\}}} = \varepsilon_i \tag{282}$$

*and*

$$\{\vec{X}:Y\} = \sqrt{1-\prod_i(1-\varepsilon_i^2)}. \tag{283}$$

**3.5.2 Theorem.** *The bound (229) in Theorem 3.3.10 is optimal, in the following sense: for any integer $N$, for all $(\varepsilon(z))_{-N\leqslant z\leqslant N} \in [0,1]^{\{-N,\dots,N\}}$, one can find random variables $(X_i)_{i\in\mathbb{Z}}$ and $(Y_j)_{j\in\mathbb{Z}}$ such that for all $i,j \in \mathbb{Z}$,*

$$\{X_i:Y_j\}_{(\vec{X}_{\{i'<i\}},\vec{Y}_{\{j'<j\}})} = \begin{cases} \varepsilon(j-i) & \text{if } |j-i| \leqslant N; \\ 0 & \text{if } |j-i| > N \end{cases} \tag{284}$$

*and $\{\vec{X}:\vec{Y}\} = \bar{\varepsilon}$, with $\bar{\varepsilon}$ defined by:*

$$\operatorname{Arcsin}\bar{\varepsilon} = \sum_{z=-N}^{N} \operatorname{Arcsin}\varepsilon(z) \wedge \frac{\pi}{2}. \tag{285}$$

Actually, as proving Theorem 3.5.2 for *all* the $(\varepsilon(z))_{-N\leqslant z\leqslant N}$ involves some heavy technicalities [35], I will only prove the slightly weaker following

**3.5.3 Theorem.** *For any integer $N$, the exists a neighbourhood $U$ of $\vec{0}$ in $[0,1]^{\{-N,\dots,N\}}$ such that, for all $(\varepsilon(z))_{-N\leqslant z\leqslant N} \in U$, one can find random variables $(X_i)_{i\in\mathbb{Z}}$ and $(Y_j)_{j\in\mathbb{Z}}$ satisfying (284) and (285)*[¶].

*3.5.4 Remark.* On the other hand, Theorem 3.3.1 is obviously not optimal since, as we pointed out, its bound is strictly weaker than that of Theorem 3.3.10.

[¶]Notice that in the neighbourhood of $\vec{0}$, one can drop the "$\wedge\frac{\pi}{2}$" in the right-hand side of (285).

The proof of Theorem 3.5.1 relies on the following important result:

**3.5.5 Lemma.** *Let* $(X_1,\ldots,X_N,Y)$ *be an* $(N+1)$*-dimensional Gaussian vector. For all* $1\leqslant i\leqslant N$*, define*

$$e_i := \{X_i : Y\}_{\vec{X}_{\{i'<i\}}}, \tag{286}$$

*then one has exactly:*

$$\{\vec{X} : Y\} = \sqrt{1-\prod_i(1-e_i^2)}. \tag{287}$$

*3.5.6 Remark.* Maximal correlation, as I told in § 1, is fundamentally a Hilbertian concept. When one deals with Gaussian vectors, the Hilbert spaces involved actually have finite dimensions, so that Lemma 3.5.5 about decorrelations can also be seen as a result about Euclidian spaces. In Appendix 3.9, I will present an unexpected corollary of this lemma, stating a geometric property of the 3-dimensional Euclidian space.

*Proof of Lemma 3.5.5.* To alleviate notation, we denote $\mathscr{F}_{i-1} := \sigma(\vec{X}_{\{i'<i\}})$. Since $(\vec{X},Y)$ is Gaussian, the law of $(X_i,Y)$ under $\mathbf{P}[\cdot|x_1,\ldots,x_{i-1}]$ is Gaussian and only depends on $(x_1,\ldots,x_{i-1})$ through an additive constant; consequently, we can speak of '*the* maximal correlation between $X_i$ and $Y$ conditionally to $\mathscr{F}_{i-1}$', which is $e_i$, and also of '*the* conditional variance of $X_i$ w.r.t. $\mathscr{F}_{i-1}$', resp. '*the* conditional variance of $Y$', resp. '*the* conditional covariance of $(X_i,Y)$', which we denote resp. $\mathrm{Var}(X_i|\mathscr{F}_{i-1})$, $\mathrm{Var}(Y|\mathscr{F}_{i-1})$, $\mathrm{Cov}(X_i,Y|\mathscr{F}_{i-1})$. By Theorem 1.2.6, one has:

$$\mathrm{Cov}(X_i,Y|\mathscr{F}_{i-1}) = \pm e_i\,\mathrm{Sd}(X_i|\mathscr{F}_{i-1})\,\mathrm{Sd}(Y|\mathscr{F}_{i-1}). \tag{288}$$

Now take $g(Y)=Y$ and $f(X)=\sum_{i=1}^N \beta_i X_i$, for some $\beta_i\in\mathbb{R}$ to be chosen later. Then $g^{i-1}$ is equal to $Y-\mathbf{E}[Y|\mathscr{F}_{i-1}]$ and $f_i$ is proportional to $X_i-\mathbf{E}[X_i|\mathscr{F}_{i-1}]$, thus, by (288) and our model's being Gaussian, all the inequalities until (177) in the proof of Theorem 3.2.2 actually are equalities for $\varepsilon_i=e_i$. If moreover $\mathrm{Cov}(f_i,g^{i-1}|\mathscr{F}_{i-1})\geqslant 0$ for all $i$, then we can drop the absolute values in their left-hand sides, and thus (178) will also be an equality. Then, to get an equality in (179), it just remains to ensure that the final Cauchy–Schwarz equality is an equality, i.e. to ensure that one has, for all $i$:

$$\mathrm{Var}(f_i) \propto e_i^2 \prod_{i'=1}^{i-1}(1-e_{i'}^2). \tag{289}$$

If all of that is satisfied, then one will have exactly $\mathbf{E}[fg]=\sqrt{1-\prod_i(1-e_i^2)}\,\mathrm{Sd}(f)\,\mathrm{Sd}(g)$, so that $\{\vec{X}:Y\}\geqslant\sqrt{1-\prod_i(1-e_i^2)}$. The converse inequality being obviously true by (the minimal version of) Theorem 3.2.2, the result will follow.

So, we have to check that the choice of the $\beta_i$ can be performed so that (289) is satisfied, with $\mathrm{Cov}(f_i,g^{i-1}|\mathscr{F}_{i-1})$ of the good sign. To do this, we will choose successively relevant values for $\beta_N,\beta_{N-1},\ldots,\beta_1$.

We observe that, if $\beta_N,\ldots,\beta_{i+1}$ have already been fixed, then $\beta_i\mapsto\mathrm{Cov}(f_i,g^{i-1}|\mathscr{F}_{i-1})$ is an affine function with slope

$$\pm e_i\,\mathrm{Sd}(Y|\mathscr{F}_{i-1})\frac{\mathrm{Sd}(X_i|\mathscr{F}_{i-1})}{\mathrm{Sd}(X_i)}. \tag{290}$$

Moreover, $\mathrm{Var}(f_i) = \mathrm{Var}(f_i|\mathscr{F}_{i-1})$ as $f_i$ is centered w.r.t. $\mathscr{F}_{i-1}$; so, since $f_i \propto X_i - \mathbf{E}[X_i|\mathscr{F}_{i-1}]$, (288) implies:

$$\mathrm{Var}(f_i) = \frac{\mathrm{Cov}(f_i, g^{i-1}|\mathscr{F}_{i-1})^2}{e_i^2 \mathrm{Var}(Y|\mathscr{F}_{i-1})}. \tag{291}$$

So, provided $e_i$, $\mathrm{Sd}(Y|\mathscr{F}_{i-1})$ and $\mathrm{Sd}(X_i|\mathscr{F}_{i-1})$ are nonzero, there exists a (unique) $\beta_i$ satisfying our assumptions.

Now if $\mathrm{Sd}(Y|\mathscr{F}_{i-1})$ is zero, this means that $Y$ is $\mathscr{F}_{i-1}$-measurable; then one of the $e_{i'}$ has to be 1 and thus the result is trivial. Next if $\mathrm{Sd}(X_i|\mathscr{F}_{i-1})$ is zero, this means that $X_i$ is $\mathscr{F}_{i-1}$-measurable; then $e_i = 0$ and $f_i \equiv 0$, so that our assumptions are automatically satisfied. Finally if $e_i = 0$ and $\mathrm{Var}(X_i|\mathscr{F}_{i-1}) > 0$, then there exists a (unique) $\beta_i$ such that $f_i \equiv 0$, for which our assumptions are satisfied. So all those particular cases actually work fine too. □

*Proof of Theorem 3.5.1.* For technical reasons, we begin with noticing that the theorem is immediate if some $e_i$ is equal to 1, so that we can assume that all the $e_i$ are $< 1$. Thanks to Lemma 3.5.5, it suffices to prove that for any sequence of $\varepsilon_i \in [0,1)$ it is possible to build a Gaussian vector $(X, \vec{Y})$ for which $e_i = \varepsilon_i\ \forall i$. To do this, let $\xi, \zeta_1, \ldots, \zeta_N$ be i.i.d. $\mathscr{N}(1)$ variables, and take $Y = \xi$ and $X_i = \sqrt{1-\alpha_i}\zeta_i + \sqrt{\alpha_i}\xi$ for some parameters $\alpha_i \in [0,1)$. We want to choose the $\alpha_i$ such that $\vec{e}(\vec{\alpha}) = \vec{\varepsilon}$; this is always possible, by the following method:

- First we compute $\alpha_1$: By Theorem 1.2.6, one can write down the equation linking $\alpha_1$ and $e_1$. It is clear without knowing the precise form of that equation (actually, $e_1 = \sqrt{\alpha_1}$) that $e_1$ is a continuous increasing function of $\alpha_1$ with $e_1 = 0$ for $\alpha_1 = 0$ and $e_1 = 1$ for $\alpha_1 = 1$. Therefore there is a unique $\alpha_1$ such that $e_1 = \varepsilon_1$.
- Then we compute $\alpha_2$: As we already know the value of $\alpha_1$, we can treat it as a constant and look for the equation linking $\alpha_2$ and $e_2$, which we compute by Theorem 1.2.6 again. That equation, though more complicated than in the previous case (actually, $e_2 = \sqrt{\alpha_2}\sqrt{1-\alpha_1}/\sqrt{1-\alpha_1\alpha_2}$), exhibits the same behaviour: $e_2$ is a continuous increasing function of $\alpha_2$ with $e_2(\alpha_2 = 0) = 0$ and $e_2(\alpha_2 = 1) = 1$. Therefore there is a unique $\alpha_2$ such that $e_2 = \varepsilon_2$.
- We carry on this process until having determined all the $\alpha_i$.

□

*Proof of Theorem 3.5.3.* Again, the principle of the proof will consist in showing how the optimal bound can be attained for relevant Gaussian vectors and linear functions of them.

We consider independent $\mathscr{N}(1)$ variables $(\xi_j)_{j\in\mathbb{Z}}$ and $(\omega_{ij})_{(i,j)\in\mathbb{Z}\times\mathbb{Z}}$. For all $i$ we set:

$$X_i = \sum_{z=-N}^{N} \omega_{i(i+z)}, \tag{292}$$

resp. for all $j$:

$$Y_j = \xi_j + \sum_{z=-N}^{N} \alpha_z \omega_{(j-z)j} \tag{293}$$

for some real parameters $(\alpha_z)_{-N\leqslant z\leqslant N}$ to be fixed later. This model is obviously invariant by translation of the indexes. For $z\in\mathbb{Z}$, define

$$\dot{e}_z := \frac{\mathrm{Cov}(X_i, Y_{i+z}|\mathscr{F}_{i-1}\vee\mathscr{G}_{i+z-1})}{\mathrm{Sd}(X_i|\mathscr{F}_{i-1}\vee\mathscr{G}_{i+z-1})\cdot\mathrm{Sd}(Y_{i+z}|\mathscr{F}_{i-1}\vee\mathscr{G}_{i+z-1})}, \tag{294}$$

where the choice of $i$ does not matter. Since our model is Gaussian, by Theorem 1.2.6,

$$\{X_i : Y_{i+z}\}_{(\vec{X}_{\{i'<i\}},\vec{Y}_{\{j'<i+z\}})} = |\dot{e}_z|. \tag{295}$$

By the properties of Gaussian vectors, it is possible to write down explicitly the equations linking the $\dot{e}_z$ to the $\alpha_z$. Though these equations may be quite horrendous, some of their properties can be easily established:

**3.5.7 Claim.**

*(i)* *For $|z|>N$, $\dot{e}_z=0$ (for any choice of the $\alpha_z$);*

*(ii)* *The map $(\alpha_{-N},\ldots,\alpha_N)\mapsto(\dot{e}_{-N},\ldots,\dot{e}_N)$ is of class $\mathscr{C}^1$ on the neighbourhood of $(0,\ldots,0)$, with:*

$$\left(\frac{\partial\dot{e}_z}{\partial\alpha_y}\right)(\vec{0}) = \frac{\mathbf{1}_{y=z}}{\sqrt{2N+1}}. \tag{296}$$

By the inverse function theorem, one can therefore find neighbourhoods $V$ and $U$ of $\vec{0}$ in $\mathbb{R}^{\{-N,\ldots,N\}}$ such that the map $\vec{\alpha}\mapsto\vec{e}$ is a $\mathscr{C}^1$-diffeomorphism from $V$ onto $U$. In particular, for $\vec{\varepsilon}$ in such an $U$ we can always fix the $\alpha_z$ of our model such that $\forall z$ $\dot{e}_z=\mathbf{1}_{|z|\leqslant N}\varepsilon(z)$, so that (284) is satisfied.

Now we have to choose $f$ and $g$. Morally[‖] we have to take the functions $f$ and $g$ having maximal Pearson correlation. Since the model is Gaussian, these functions will be linear, and since the model is invariant by translation, they will likely be invariant by translation too. So we would like to take, formally, $f(\vec{X})=\sum_{i\in\mathbb{Z}}X_i$ and $g(\vec{Y})=\sum_{j\in\mathbb{Z}}Y_j$. As such functions are not properly defined, we will rather consider $f[k](\vec{X})=\sum_{i=-k}^{k}X_i$, resp. $g[k](\vec{Y})=\sum_{j=-k}^{k}Y_j$, and then we will let $k$ tend to infinity.

For these $f[k]$ and $g[k]$, define the $V[k]_i^j$, the $W[k]_j^i$ and the $S[k]_{ij}$ as in the proof of Theorem 3.3.10, which are gathered into the array $\mathbf{A}[k]$. The following properties of the $\mathbf{A}[k]$ follow easily from the structure of our model:

**3.5.8 Claim.**

*(i)* *All the $V[k]_i^j, W[k]_j^i, S[k]_{ij}$ are bounded uniformly in $i,j,k$.*

*(ii)*
- *$V[k]_i^j$ is zero as soon as $i\notin\{-k-2N,\ldots,k\}$;*
- *$W[k]_j^i$ is zero as soon as $j\notin\{-k,\ldots,k\}$.*

*(iii)* *$S[k]_{ij}$ is zero as soon as $|j-i|>N$.*

*(iv)*
- *For $-k\leqslant i\leqslant k-2N$, $V[k]_i^j$ only depends on $(j-i)$, even when $k$ varies. We denote its value by $V_{(j-i)}$.*
- *For $-k\leqslant j\leqslant k$, $W[k]_j^i$ only depends on $(j-i)$, even when $k$ varies. We denote its value by $W_{(j-i)}$.*
- *For $-k\leqslant i\leqslant k-2N$ and $-k\leqslant j\leqslant k$, $S[k]_{ij}$ only depends on $(j-i)$, even when $k$ varies. We denote its value by $S_{(j-i)}$.*

[‖]I say "morally" because nothing ensures that the supremum (34) would actually be a maximum here.

*(v)* • $V_{(z)}$ *has some constant value* $v$ *for* $z < -N$;
• $W_{(z)}$ *has some constant value* $w$ *for* $z > N$.

By Claim 3.5.8, $\mathbf{A}[k]$ converges pointwise to some compact Toeplitz array $\mathbf{A}$, whose entries are the $V_{(z)}, W_{(z)}, S_{(z)}$ introduced at Item (iv) of the claim, whose values $v$ and $w$ are those introduced at Item (v), and whose value $s$ is $\sum_{z=-N}^{N} S_{(z)}$. All the arrays $\mathbf{A}[k]$ are obviously correct since they correspond to true functions, so by passing to the limit $\mathbf{A}$ is correct too.

Since our model is Gaussian, all the inequalities (238), (239) and (245) are actually equalities for the arrays $\mathbf{A}[k]$; moreover, since the $\dot{\varepsilon}_z$ are nonnegative, the $S[k]_{ij}$ are nonnegative. By letting $k$ tend to infinity, it follows that all the inequalities (253)–(255) are actually equalities for the array $\mathbf{A}$, with the $S_{(z)}$ nonnegative. Consequently in (262) one has $\widehat{\theta}(z) = \theta(z)$, and all the further inequalities are actually equalities, so that in the end (270) becomes:

$$\frac{s}{\sqrt{vw}} = \bar{\varepsilon}. \tag{297}$$

Now, defining $V[k]$, $W[k]$ and $S[k]$ by resp. (236), (237) and (233) for the arrays $\mathbf{A}[k]$, Claim 3.5.8 shows that, when $k \to \infty$, $V[k] \sim 2kv$, resp. $W[k] \sim 2kw$, resp. $S[k] \sim 2ks$, so (297) implies that $S[k]/\sqrt{V[k]W[k]} \to \bar{\varepsilon}$. But recall that $V[k]$, $W[k]$ and $S[k]$ are the respective variances and covariance of the functions $f[k] \in \bar{L}^2(\vec{X})$ and $g[k] \in \bar{L}^2(\vec{Y})$, so by the very definition (34) of maximal correlations,

$$\{\vec{X} : \vec{Y}\} \geqslant \frac{S[k]}{\sqrt{V[k]W[k]}}. \tag{298}$$

Making $k \to \infty$, it follows that $\{\vec{X} : \vec{Y}\} \geqslant \bar{\varepsilon}$; the converse inequality being obviously true by (the minimal version of) Theorem 3.3.10, this proves Theorem 3.5.3. □

*3.5.9 Example.* In this example we will carry out explicit computations for a Gaussian model close to the model presented in the proof above. We take independent $\mathscr{N}(1)$ variables $\ldots, \zeta_{-1}, \zeta_0, \zeta_1, \ldots, \ldots, \xi_{-1/2}, \xi_{1/2}, \xi_{3/2} \ldots, \ldots, \omega_{-1/4}, \omega_{1/4}, \omega_{3/4}, \ldots$, and we set

$$X_i = \zeta_i + \sqrt{\alpha}(\omega_{i-1/4} + \omega_{i+1/4}), \tag{299}$$

$$\text{resp.} \quad Y_j = \xi_j + \sqrt{\alpha}(\omega_{j-1/4} + \omega_{j+1/4}) \tag{300}$$

for all integer $i$, resp. all half-integer $j$, where $\alpha$ is some arbitrary nonnegative parameter. We are going to show that for this system (229) is actually an equality, in accordance with the proof of Theorem 3.5.3.

For half-integer $z$ denote

$$e_z := \{X_i : Y_{i+z}\}_{(\vec{X}_{\{i'<i\}}, \vec{Y}_{\{j'<i+z\}})}, \tag{301}$$

where the choice of $i$ does not matter by translation invariance. Clearly $e_{-z} = e_z$ for all $z$ and $e_z = 0$ for $|z| > 1/2$, so to know all the $e_z$ the only nontrivial computation is computing $e_{1/2}$. Let us perform it.

Since everything is Gaussian, by Theorem 1.2.6, $e_{1/2}$ is the value, under the law $\mathbf{P}[\cdot | \vec{X}_{\{i<0\}}, \vec{Y}_{\{j<1/2\}} \equiv 0]$, of

$$|\mathbf{E}[X_0 Y_{1/2}]| / \mathrm{Sd}(X_0)\mathrm{Sd}(Y_{1/2}). \tag{302}$$

Under the law $\mathbf{P}[\cdot|\vec{X}_{\{i<0\}},\vec{Y}_{\{j<1/2\}}\equiv 0]$, it is clear that $\zeta_0,\omega_{1/4},\xi_{1/2},\omega_{3/4},\ldots$ have exactly the same (joint) law as under $\mathbf{P}$, and that $\omega_{-1/4}$ is still independent of these (joint) variables, though its variance shall have diminished. So we need only compute

$$v := \mathrm{Var}(\omega_{-1/4}|\vec{X}_{\{i<0\}},\vec{Y}_{\{j<1/2\}}\equiv 0). \tag{303}$$

Denote $\vec{L}_{\mathrm{r}} := (\ldots,X_{-2},Y_{-3/2},X_{-1},Y_{-1/2})$, resp. $\vec{L}_{\mathrm{l}} := (\ldots,X_{-2},Y_{-3/2},X_{-1})$. We write that (formally)

$$d\mathbf{P}[\vec{L}_{\mathrm{r}}\equiv 0 \text{ and } \omega_{-1/4}=x]\propto e^{-x^2/2v}dx, \tag{304}$$

and also $d\mathbf{P}[\vec{L}_{\mathrm{l}}\equiv 0 \text{ and } \omega_{-3/4}=y]\propto e^{-y^2/2v}dy$ by translation invariance. But under $\mathbf{P}[\cdot|\vec{L}_{\mathrm{l}}\equiv 0 \text{ and } \omega_{-3/4}=y]$, the law of $(\xi_{-1/2},\omega_{1/4})$ is the same as under $\mathbf{P}$, so one has:

$$\begin{aligned}
e^{-x^2/2v} &\propto d\mathbf{P}[\vec{L}_{\mathrm{r}}\equiv 0 \text{ and } \omega_{-1/4}=x]\\
&= \int_y dy\, d\mathbf{P}[\vec{L}_{\mathrm{l}}\equiv 0 \text{ and } \omega_{-3/4}=y]d\mathbf{P}[Y_{-1/2}=0 \text{ and } \omega_{-1/4}=x|\vec{L}_{\mathrm{l}}\equiv 0 \text{ and } \omega_{-3/4}=y]\\
&\propto \int_y d\mathbf{P}[Y_{-1/2}=0 \text{ and } \omega_{-1/4}=x|\vec{L}_{\mathrm{l}}\equiv 0 \text{ and } \omega_{-3/4}=y]e^{-y^2/2v}\,dy\\
&= \int_y d\mathbf{P}[\xi_{-1/2}=-\sqrt{\alpha}(x+y) \text{ and } \omega_{-1/4}=x|\vec{L}_{\mathrm{l}}\equiv 0 \text{ and } \omega_{-3/4}=y]e^{-y^2/2v}\,dy\\
&\propto \int_y e^{-\alpha(x+y)^2/2}e^{-x^2/2}e^{-y^2/2v}\,dy \propto \exp\Big\{\Big(1+\alpha-\frac{\alpha^2}{\alpha+1/v}\Big)\frac{x^2}{2}\Big\},
\end{aligned} \tag{305}$$

so that $v$ must satisfy:

$$1+\alpha-\frac{\alpha^2}{\alpha+1/v}=\frac{1}{v}, \tag{306}$$

whose only nonnegative solution is

$$v=\frac{\sqrt{1+4\alpha}-1}{2\alpha}. \tag{307}$$

So one has $\mathrm{Sd}(X_0|\vec{L}_{\mathrm{l}}\equiv 0)=\sqrt{1+\alpha v+\alpha}=(\sqrt{1+4\alpha}+1)/2$, $\mathrm{Sd}(Y_{1/2}|\vec{L}_{\mathrm{l}}\equiv 0)=\sqrt{1+2\alpha}$ and $\mathbf{E}[X_0Y_{1/2}|\vec{L}_{\mathrm{l}}\equiv 0]=\alpha$, so that in the end (302) yields:

$$e_{1/2}=\frac{\sqrt{1+4\alpha}-1}{2\sqrt{1+2\alpha}}. \tag{308}$$

With this value, Theorem 3.3.10 states that one has necessarily

$$\{\vec{X}:\vec{Y}\}\leqslant \sin(2\operatorname{Arcsin} e_{1/2})^{[*]}=2e_{1/2}\sqrt{1-e_{1/2}^2}=\frac{2\alpha}{1+2\alpha}. \tag{309}$$

We show that (309) is actually an equality: take indeed $f[k](\vec{X}):=\sum_{i=1}^k X_k$, resp. $g[k](\vec{Y}):=\sum_{j=1/2}^{k-1/2} Y_k$, then $\mathrm{Var}(f[k])=\mathrm{Var}(g[k])=k(1+2\alpha)$ and $\mathbf{E}[fg]=(2k-1)\alpha$, so that

$$\{\vec{X}:\vec{Y}\}\geqslant\frac{(2k-1)\alpha}{k(1+2\alpha)}\overset{k\to\infty}{\to}\frac{2\alpha}{1+2\alpha}. \tag{310}$$

[*]As here one always has $e_{1/2}\leqslant 1/\sqrt{2}$, we can drop the "$\wedge\frac{\pi}{2}$" of Formula (229).

*3.5.10 Remark.* One can formally set $\alpha = +\infty$ in the previous example, which actually means that one takes $X_i = \omega_{i-1/4} + \omega_{i+1/4}$, resp. $Y_j = \omega_{j-1/4} + \omega_{j+1/4}$. In this case, both Formulas (308) and (310) 'pass to the limit', yielding $e_{1/2} = 1/\sqrt{2}$ and $\{\vec{X} : \vec{Y}\} = 1$. This shows that it is possible indeed that the $e_z$ have 'mild' values and that yet $\vec{X}$ and $\vec{Y}$ are fully correlated. In other words, the "$\wedge\frac{\pi}{2}$" in (229) is not an 'artifact' of the proof of Theorem 3.3.10[†], but the expression of a real 'phase transition' phenomenon[‡]. Such a phase transition did not occur for the simple tensorization formula (169), which shows that double tensorization in intrinsically more complicated than simple tensorization.

## 3.5.b Asymptotic optimality

In the previous subsection we saw that (the minimal versions of) Theorems 3.2.2 and 3.3.10 were optimal, while Theorem 3.3.1 was not. However it turns out that that result is nevertheless 'asymptotically optimal', in the sense that the bound it gives is equivalent to the optimal bound when the correlations between the variables become weak. Here is a precise statement:

**3.5.11 Theorem.** *Let $I = \{1,\dots,N\}$ and $J = \{1,\dots,M\}$ be finite sets, and define the function $Opt$: $[0,1]^{I\times J} \to [0,1]$ by*

$$Opt(\vec{\varepsilon}_{I\times J}) := \sup\big\{\{\vec{X}_I : \vec{Y}_J\}\ ;\ (\forall (i,j)\in I\times J)(\{X_i : Y_j\}_{(\vec{X}_{\{i'<i\}},\vec{Y}_{\{j'<j\}})} \leqslant \varepsilon_{ij})\big\}; \tag{311}$$

*then, when $\vec{\varepsilon}_{I\times J} \to \vec{0}$, one has:*

$$Opt(\vec{\varepsilon}) \sim |||\boldsymbol{\varepsilon}|||. \tag{312}$$

*3.5.12 Remark.* In the same way, the simple bound (160) of Proposition 3.2.1 is asymptotically equivalent to the optimal bound (169) of Theorem 3.2.2.

*Proof.* Take $(M+NM)$ i.i.d. $\mathcal{N}(1)$ variables $\xi_1,\dots,\xi_M,\omega_{11},\dots,\omega_{NM}$. For $((\alpha_{ij}))_{i,j} \in \mathbb{R}^{N\times M}$, set

$$\begin{cases} X_i &=& \sum_j \omega_{ij}; \\ Y_j &=& \xi_j + \sum_i \alpha_{ij}\omega_{ij}. \end{cases} \tag{313}$$

Denote

$$e_{ij} := \{X_i : Y_j\}_{(\vec{X}_{\{i'<i\}},\vec{Y}_{\{j'<j\}})}, \tag{314}$$

and define $\dot{e}_{ij}$ as the Pearson correlation coefficient of $X_i$ and $Y_j$ under the law $\mathbf{P}[\cdot| \vec{X}_{\{i'<i\}},\vec{Y}_{\{j'<j\}} \equiv 0]$. Then, as in the proof of Theorem 3.5.3, one has $e_{ij} = |\dot{e}_{ij}|$, and the function $\vec{\alpha} \mapsto \vec{\dot{e}}$ is $\mathscr{C}^1$ around $\vec{0}$, with

$$\vec{\dot{e}} = \frac{1}{\sqrt{M}}\vec{\alpha} + O(\|\vec{\alpha}\|^2) \qquad \text{when } \vec{\alpha} \to \vec{0}. \tag{315}$$

By the inverse function theorem, $\vec{\alpha} \mapsto \vec{\dot{e}}$ is therefore a diffeomorphism from some neighbourhood $V$ of $\vec{0}$ onto some neighbourhood $U$ of $\vec{0}$, whose inverse function is such that

$$\vec{\alpha} = \sqrt{M}\,\vec{\dot{e}} + O(\|\vec{\dot{e}}\|^2) \qquad \text{when } \vec{\dot{e}} \to \vec{0}. \tag{316}$$

[†]On the other hand, it is possible that the "$\wedge 1$" in (186) is such an artifact, since Theorem 3.3.1 is not optimal.

[‡]There exist indeed situations going 'beyond the phase transition', i.e. for which $\sum_{z\in\mathbb{Z}} \operatorname{Arcsin}(e_z) > \pi/2$, though this is not the case for Example 3.5.9.

Now let $\vec{\varepsilon} \in (\mathbb{R}_+)^{N\times M} \cap U$. Take $\vec{\alpha} \in V$ such that $\vec{\dot{e}}(\vec{\alpha}) = \vec{\varepsilon}$, so that the condition of (311) is satisfied. For $\varphi \in \mathbb{R}^N, \psi \in \mathbb{R}^M$ with $\|\varphi\|, \|\psi\| = 1$, set

$$\begin{cases} f(\vec{X}) & := & \sum_i \varphi_i X_i; \\ g(\vec{Y}) & := & \sum_j \psi_j Y_j. \end{cases} \tag{317}$$

One has

$$\mathrm{Var}(f) = M, \tag{318}$$

$$\mathrm{Var}(g) = 1 + O(\|\vec{\alpha}\|^2) = 1 + O(\|\vec{\varepsilon}\|^2) \tag{319}$$

and

$$\mathbf{E}[fg] = \sum_{i,j} \alpha_{ij}\varphi_i\psi_j = \langle \varphi, \boldsymbol{\varepsilon}\psi \rangle + O(\|\vec{\varepsilon}\|^2), \tag{320}$$

where the constants implicit in the "$O(\|\vec{\varepsilon}\|^2)$" are uniform in $(\varphi, \psi)$. So one has

$$Opt(\vec{\varepsilon}) \geqslant \{\vec{X} : \vec{Y}\} \geqslant \frac{|\mathbf{E}[fg]|}{\mathrm{Sd}(f)\mathrm{Sd}(g)} = |\langle \varphi, \boldsymbol{\varepsilon}\psi \rangle| + O(\|\vec{\varepsilon}\|^2), \tag{321}$$

whence after taking supremum over $(\varphi, \psi)$:

$$Opt(\vec{\varepsilon}) \geqslant |||\boldsymbol{\varepsilon}||| + O(\|\vec{\varepsilon}\|^2) \overset{\vec{\varepsilon}\to\vec{0}}{\sim} |||\boldsymbol{\varepsilon}|||. \tag{322}$$

Since on the other hand $Opt(\vec{\varepsilon}) \leqslant |||\boldsymbol{\varepsilon}|||$ by Theorem 3.3.1, the proposition follows. □

*3.5.13 Remark.* If we state decorrelation hypotheses w.r.t. the *whole* $\sigma$-metalgebra of the system (denoted by "$*$"), no quantity analogous to $\dot{e}_{ij}$ shall exist any more; one can only write, denoting $e'_{ij} := \{X_i : Y_j\}_*$:

$$e'_{ij}(\vec{\alpha}) = \frac{|\alpha_{ij}|}{\sqrt{M}} + O(\|\vec{\alpha}\|^2). \tag{323}$$

So, to see how the correlations depend on the parameters, we have to study the map $\vec{\alpha} \mapsto \vec{e}'$, which is approximated by a homothety *only on the cone* $\mathbb{R}_+^{N\times M}$ —and which moreover is no better than continuous here. So we shall replace the inverse function theorem by an alternative technique, which will yield the slightly weaker theorem stated just below.

**3.5.14 Theorem.** *Define*

$$Opt'((\varepsilon_{ij})_{(i,j)\in I\times J}) := \sup\big\{\{\vec{X}_I : \vec{Y}_J\}\ ;\ (\forall (i,j) \in I \times J)\big(\{X_i : Y_j\}_* \leqslant \varepsilon_{ij}\big)\big\}; \tag{324}$$

*then for any closed cone $C$ of $\mathbb{R}^{N\times M}$ contained in $(\mathbb{R}_+^*)^{N\times M} \cup \{0\}$, on $C$, one has:*

$$Opt'(\vec{\varepsilon}) \overset{\vec{\varepsilon}\to\vec{0}}{\sim} |||\boldsymbol{\varepsilon}|||. \tag{325}$$

## 3.6 Machinery for using the tensorization theorems

Up to now we stated the tensorization theorems in a rather 'theoretical' form. To apply these results to 'concrete' situations, some additional techniques may be needed. This section gives such techniques, which we will use later for the applications of Chapter 5.

☛ *In this section, all the probability systems considered will be endowed with their natural $\sigma$-metalgebras, cf. Definition 3.1.16. To alleviate notation, I will give no names to these $\sigma$-metalgebras, but will plainly denote $\{X : Y\}_*$ to mean "the subjective decorrelation between $X$ and $Y$ seen from the natural $\sigma$-metalgebra of the underlying system".*

## 3.6.a The 'doubling-up' technique

**3.6.1 Definition.** For $I$ a set and $\mathscr{R}$ a binary relation on $I$, $J_1, J_2 \subset I$, we will say that "$J_2$ is $\mathscr{R}$-disjoint to $J_1$" if $(i,j) \in J_1 \times J_2 \Rightarrow i \not\mathscr{R} j$.

**3.6.2 Lemma** ('Doubling-up' lemma)**.** *Let $I$ be a (countable) set and let $(X_i)_{i\in I}$ be random variables such that for all $i,j \in I$, $\{X_i : X_j\}_* \leqslant \varepsilon_{ij}$ for a certain family of $\varepsilon_{ij} \in [0,1]$.*

*Let $\mathscr{R}$ be a binary relation on $I$; for $i,j \in I$, denote $\varepsilon^{\mathscr{R}}_{ij} := \mathbf{1}_{i\not\mathscr{R}j}\varepsilon_{ij}$.*

*Define $\mathbf{I} = I_1 \uplus I_2$ to be a disjoint union of two copies of $I$; denote by $(i_1)_{i\in I}$, resp. $(j_2)_{j\in I}$, the elements of $I_1$, resp. $I_2$. Assume that the following holds for a certain $\varepsilon \in [0,1]$: "if $(Y_{i_\kappa})_{i_\kappa\in\mathbf{I}}$ are random variables such that $\forall i,j \in I$ $\{Y_{i_1} : Y_{j_2}\}_* \leqslant \varepsilon^{\mathscr{R}}_{ij}$, then $\{\vec{Y}_{I_1} : \vec{Y}_{I_2}\} \leqslant \varepsilon$".*

*Then for all $J_1, J_2 \subset I$ such that $J_2$ is $\mathscr{R}$-disjoint to $J_1$, $\{\vec{X}_{J_1} : \vec{X}_{J_2}\}_* \leqslant \varepsilon$.*

*3.6.3 Remark.* The interest of Lemma 3.6.2 is that, by proving *one* tensorization result on $\{\vec{Y}_{I_1} : \vec{Y}_{I_2}\}$, one gets tensorization results on *all* the $\{\vec{X}_{J_1} : \vec{X}_{J_2}\}$ for $J_2$ $\mathscr{R}$-disjoint to $J_1$.

*3.6.4 Example.*

1. If you take for $\mathscr{R}$ the equality relation, then Lemma 3.6.2 gives a decorrelation result for all disjoint $J_1$ and $J_2$.
2. If $I$ is equipped with a distance *dist* and if you take $(i\mathscr{R}j) \Leftrightarrow (\mathit{dist}(i,j) < d_1)$, then you get a decorrelation result for all $J_1$ and $J_2$ such that $\mathit{dist}(J_1,J_2) \geqslant d_1$.

*Proof.* Assume that the hypotheses of the lemma hold and let $J_1, J_2 \subset I$ with $J_2$ $\mathscr{R}$-disjoint to $J_1$. For $i_\kappa \in \mathbf{I}$, define

$$Y_{i_\kappa} = \begin{cases} X_i & \text{if } (\kappa = 1 \text{ and } i \in J_1) \text{ or } (\kappa = 2 \text{ and } i \in J_2); \\ \partial & \text{otherwise,} \end{cases} \tag{326}$$

for $\partial$ some cemetery point in the range of none of the $X_i$. Since a constant variable is always independent of any variable, the hypothesis "$\{X_i : X_j\}_* \leqslant \varepsilon_{ij}$" for all $i,j \in I$ implies that $\{Y_{i_1} : Y_{j_2}\}_* \leqslant \varepsilon^{\mathscr{R}}_{ij}$, so, by the assumption of the lemma, $\{\vec{Y}_{I_1} : \vec{Y}_{I_2}\} \leqslant \varepsilon$. But $\vec{X}_{J_1}$ is $\vec{Y}_{I_1}$-measurable, resp. $\vec{X}_{J_2}$ is $\vec{Y}_{I_2}$-measurable, hence $\{\vec{X}_{J_1} : \vec{X}_{J_2}\} \leqslant \varepsilon$.

Getting the subjective result w.r.t. "$*$" is just a variant of that reasoning, cf. § 3.4.b. □

## 3.6.b A practical result on $\mathbb{Z}^n$

In Chapter 5, the situations we will handle shall always be of the following form:

**3.6.5 Assumption.** For some $n \in \mathbb{N}^*$, the system is made of random variables $X_i$, $i \in \mathbb{Z}^n$, which satisfy the condition

$$\forall i,j \in \mathbb{Z}^n \qquad \{X_i : X_j\}_* \leqslant \varepsilon(j-i) \tag{327}$$

for some symmetric function $\varepsilon : \mathbb{Z}^n \to [0,1]$.

For systems satisfying Assumption 3.6.5, one has the following practical synthetic result:

**3.6.6 Lemma.** *Consider a norm $|\cdot|$ on $\mathbb{R}^n$, the associated distance on the affine $\mathbb{R}^n$ being denoted by dist. Then for a system satisfying Assumption 3.6.5, for all $J_1, J_2 \subset I$:*

$$\{\vec{X}_{J_1} : \vec{X}_{J_2}\} \leqslant \Big(\sum_{\substack{z\in\mathbb{Z}^n \\ |z|\geqslant dist(J_1,J_2)}} \varepsilon(z)\Big) \wedge 1. \tag{328}$$

*Proof.* To alleviate notation, denote $d := dist(J_1, J_2)$. Applying Lemma 3.6.2, taking for "$\mathscr{R}$" the relation "be at distance $< d$" [cf. Example 3.6.4-2], our goal becomes the following: supposing $(Y_{i_\kappa})_{i_\kappa \in \mathbb{Z}_1^n \uplus \mathbb{Z}_2^n}$ are random variables such that $\{Y_{i_1} : Y_{j_2}\}_* \leqslant \mathbf{1}_{|j-i|\geqslant d}\,\varepsilon(j-i)$, we want to bound above $\{\vec{Y}_{\mathbb{Z}_1^n} : \vec{Y}_{\mathbb{Z}_2^n}\}$.

To do this we apply Theorem 3.3.1, and we get that $\{\vec{Y}_{\mathbb{Z}_1^n} : \vec{Y}_{\mathbb{Z}_2^n}\}$ is bounded by $|||\boldsymbol{\varepsilon}||| \wedge 1$, where $\boldsymbol{\varepsilon}$ is the following operator:

$$\begin{array}{rrcl} \boldsymbol{\varepsilon}: & L^2(\mathbb{Z}) & \circlearrowleft & \\ & (g(j))_{j\in\mathbb{Z}} & \mapsto & \big(\sum_{j\in\mathbb{Z}} \mathbf{1}_{|j-i|\geqslant d}\,\varepsilon(j-i) g(j)\big)_{i\in\mathbb{Z}}. \end{array} \tag{329}$$

To compute $|||\boldsymbol{\varepsilon}|||$, we split $\boldsymbol{\varepsilon}$ as $\sum_{z\in\mathbb{Z}} \mathbf{1}_{|z|\geqslant d}\,\varepsilon(z) M_z$, where $M_z$ is the operator

$$\begin{array}{rrcl} M_z: & L^2(\mathbb{Z}) & \circlearrowleft & \\ & (g(j))_{j\in\mathbb{Z}} & \mapsto & (g(i+z))_{i\in\mathbb{Z}}. \end{array} \tag{330}$$

Obviously $|||M_z||| = 1$, thus $|||\boldsymbol{\varepsilon}||| \leqslant \sum_{|z|\geqslant d} \varepsilon(z)$ —actually there is even equality—, which ends the proof of Lemma 3.6.6. □

*3.6.7 Remark.* Instead of Theorem 3.3.1, here we could have used Theorem 3.3.18, which would yield a better result; yet that would be very specific to $\mathbb{Z}^n$ [cf. § 3.6.d], and the result would actually be almost equivalent to (328) [cf. § 3.5.b].

## 3.6.c Avoiding the artificial phase transition

Let us look again at Formula (328): the "$\wedge 1$" in it is not really relevant since a correlation level is *always* bounded by 1. In fact the situation is dichotomic: denoting $d := dist(I, J)$, either $\sum_{|z|\geqslant d} \varepsilon(z)$ is $< 1$ and then (328) is a true decorrelation result, or it is $\geqslant 1$ and then (328) tells us actually nothing. In other words, our result has a 'phase transition' depending on the relative values of $\sum_{|z|\geqslant d} \varepsilon(z)$ and 1, similar to the phenomenon we discussed in Remark 3.5.10.

However, as I pointed out in Footnote [†] on page 72, it is not clear whether the phase transition we are dealing with is a real phenomenon: maybe it is rather an artifact due to Theorem 3.3.1's bound's being non-optimal, which could be avoided by a cleverer reasoning. We are strengthened in that thought by observing that, if $\sum_{z\in\mathbb{Z}^n} \varepsilon(z) < \infty$, then for $d$ large enough one has $\sum_{|z|\geqslant d} \varepsilon(d) < 1$, so that there is no phase transition for long distances; why would a transition appear all of a sudden for short distances?

This subsection will show that, indeed, phase transitions can be avoided in the situations we deal with.

**3.6.8 Lemma.** *For a system satisfying Assumption 3.6.5 with $\varepsilon(z) < 1$ as soon as $z \neq 0$ and $\sum_{z\neq 0} \varepsilon(z) < \infty$, there exists a constant $k < 1$ such that, for all disjoint $J_1, J_2 \subset I$, one has $\{\vec{X}_{J_1} : \vec{X}_{J_2}\} \leqslant k$.*

*Proof.* As before, using Lemma 3.6.2 we have to bound above $\{\vec{Y}_{\mathbb{Z}_1^n} : \vec{Y}_{\mathbb{Z}_2^n}\}$ in the relevant doubled-up model. Our plan to avoid the phase transition will consist in reducing to the 'long distance' case.

For some $l \in \mathbb{N}^*$, we split $\mathbb{Z}_1^n$, resp. $\mathbb{Z}_2^n$, into a partition of $l^n =: N$ sublattices $Z_1^{(1)}, \ldots, Z_1^{(N)}$, resp. $Z_2^{(1)}, \ldots, Z_2^{(N)}$, each lattice $Z_\kappa^{(u)}$ being of the form $l\mathbb{Z}^n + z_u$ for some $z_u \in \mathbb{Z}^n/l\mathbb{Z}^n$. I claim two fundamental properties of these sublattices:

**3.6.9 Claim.** *For all $u, v \in \{1, \ldots, N\}$,*

$$\{\vec{Y}_{\mathbb{Z}_1^{(u)}} : \vec{Y}_{\mathbb{Z}_2^{(v)}}\}_* \leqslant \sum_{z \equiv z_v - z_u} \mathbf{1}_{z\neq 0}\, \varepsilon(z). \tag{331}$$

*Proof.* It is analogous to the proof of Lemma 3.6.6. □

**3.6.10 Claim.** *Provided $l$ is large enough, the right-hand side of (331) is (strictly) less than 1 for all the possible values of $u, v$.*

*Proof.* Denote $\zeta := \sup_{z\neq 0} \varepsilon(z)$; notice that our assumptions imply that $\zeta < 1$. Since $\sum_{z\in\mathbb{Z}^n} \varepsilon(z)$ converges, there exists some $d_1 < \infty$ such that $\sum_{|z|>d_1} \varepsilon(z) < 1 - \zeta$. Now, denoting $d_0 := \min\{|z| : z \in \mathbb{Z}^n \smallsetminus \{0\}\}$, for $l > 2d_1/d_0$, for all $u, v$ there is at most one $z$ congruent to $z_v - z_u$ [mod. $l$] such that $|z| \leqslant d_1$, whence the following uniform bound for the right-hand side of (331):

$$\sum_{z\equiv z_v - z_u} \mathbf{1}_{z\neq 0}\, \varepsilon(z) \leqslant \underbrace{\sum_{|z|>d_1} \varepsilon(z)}_{<1-\zeta} + \underbrace{\sum_{\substack{|z|\leqslant d_1\\ z\equiv z_v - z_u\\ z\neq 0}} \varepsilon(z)}_{\substack{\leqslant \zeta \text{ because}\\ \text{the sum has at most}\\ \text{one term, being } \leqslant \zeta}} < 1. \tag{332}$$

□

Now, suppose $l$ large enough so that Claim 3.6.10 works. We apply *simple* tensorization (Theorem 3.2.2) to the $\vec{Y}_{\mathbb{Z}_2^{(v)}}$: writing that $\vec{Y}_{\mathbb{Z}_2^n} = (\vec{Y}_{\mathbb{Z}_2^{(1)}}, \ldots, \vec{Y}_{\mathbb{Z}_2^{(N)}})$, we get that, for any $u \in \{1, \ldots, N\}$,

$$\{\vec{Y}_{\mathbb{Z}_1^{(u)}} : \vec{Y}_{\mathbb{Z}_2^n}\}_* \leqslant \sqrt{1 - \prod_{v=1}^{N} \left(1 - \{\vec{Y}_{\mathbb{Z}_1^{(u)}} : \vec{Y}_{\mathbb{Z}_2^{(v)}}\}_*^2\right)} < 1. \tag{333}$$

Now we write $\vec{Y}_{\mathbb{Z}_1^n} = (\vec{Y}_{\mathbb{Z}_1^{(1)}}, \ldots, \vec{Y}_{\mathbb{Z}_1^{(N)}})$ and we apply simple tensorization again—this time to the $\vec{Y}_{\mathbb{Z}_1^{(u)}}$ —to get:

$$\{\vec{Y}_{\mathbb{Z}_1^n} : \vec{Y}_{\mathbb{Z}_2^n}\} \leqslant \sqrt{1 - \prod_{u=1}^{N} \left(1 - \{\vec{Y}_{\mathbb{Z}_1^{(u)}} : \vec{Y}_{\mathbb{Z}_2^n}\}_*^2\right)} < 1. \tag{334}$$

Bound (334) achieves our goal. □

*3.6.11 Remark.* With that proof, the way $k$ depends on $\varepsilon(\cdot)$ is rather complicated; in particular, you cannot express $k$ as a function of only $\sum_{z\neq 0}\varepsilon(z)$ and $\sup_{z\neq 0}\varepsilon(z)$.

*3.6.12 Remark.* In the case $n = 1$, at first sight Lemma 3.6.8 seems to contradict Theorem 3.5.3, in which we told that Theorem 3.3.10, which does have a phase transition, was optimal. The explanation for this paradox stands in the slight difference between the assumptions of Lemma 3.6.8 and Theorem 3.3.10: while in Lemma 3.6.8 we really imposed that $\{X_i : X_j\}_* \leqslant \varepsilon\big(\eth(i,j)\big)$, with "$*$" denoting the *full* natural $\sigma$-metalgebra of the system, in Theorem 3.3.10—more precisely, in the version of Theorem 3.3.10 Theorem 3.5.3 proved to be optimal, which was the *minimal* version of this theorem [cf. § 3.4.a]—the conditions on subjective decorrelations were a bit looser. That difference makes all the trick when one performs the steps of simple tensorization in the proof of Lemma 3.6.8, because these steps require subjective decorrelations w.r.t. the $\vec{Y}_{I_\kappa^{(u)}}$, which the sole assumptions of Theorem 3.5.3 do not provide.

## 3.6.d Non-flat geometries

It is natural to ask what we one can do when the basic variables $X_i$ are not indexed by $\mathbb{Z}^n$, but by the vertices of a more general graph, for instance a tree or a finitely generated group. This shall occur indeed if the physical space one works in exhibits some curvature—though Chapter 5 will not handle such situations.

Actually for general graphs there are results analogous to those of §§ 3.6.b and 3.6.c, whith similar (though more technical) proofs. Here I will only give the statements of these results; the proofs can be found in an earlier version of this monograph [37].

In this subsection the situation will be the following:

**3.6.13 Assumption.** The system is made of random variables $(X_i)$ indexed by a (countable) set $I$. There is a group $G$ acting transitively on $I$, and $I$ is endowed with a symmetric map $\eth\colon\ I \times I \to \mathfrak{D}$, called the 'abstract distance', which is preserved by the action of $G$. We assume that one has

$$\forall i,j \in I \quad \{X_i : X_j\}_* \leqslant \varepsilon\big(\eth(i,j)\big) \tag{335}$$

for some function $\varepsilon\colon\ \mathfrak{D} \to [0,1]$.

**3.6.14 Definition.** For $d \in \mathfrak{D}$, we define $\mathrm{val}(d) := \#\{i \in I :\ \eth(o,i) = d\}$, where the choice of $o \in I$ does not matter.

Then the analogous to Lemma 3.6.6 is the

**3.6.15 Lemma.** *For $\mathfrak{D}' \subset \mathfrak{D}$, for all $J_1, J_2 \subset I$ such that $(i \in J_1, j \in J_2) \Rightarrow \eth(i,j) \in \mathfrak{D}'$,*

$$\{\vec{X}_{J_1} : \vec{X}_{J_2}\} \leqslant \Big(\sum_{d\in\mathfrak{D}'} \mathrm{val}(d)\varepsilon(d)\Big) \wedge 1. \tag{336}$$

The analogous of Lemma 3.6.8 is the

**3.6.16 Lemma.** *Assume that Assumption 3.6.13 is satisfied; denoting by* 0 *the (common) value of the* $\mathfrak{d}(i,i)$, *also assume that* $\mathrm{val}(0) = 1$ *and that* $\varepsilon(d) < 1$ *as soon as* $d \neq 0$. *Assume that* $\sum_{d \in \mathfrak{D}} \mathrm{val}(d)\varepsilon(d) < \infty$.

*Moreover, assume that the action of G on I is* profinite *(cf. [25, Definition 1.1]), i.e. that there is a subset* $\mathcal{N} \subset \mathbb{N}^*$ *such that for each* $N \in \mathcal{N}$, *there is a subgroup* $G_N \leqslant G$ *such that:*

(i) *The action of* $G_N$ *splits I into exactly N orbits* $I^{(1)}, \ldots, I^{(N)}$;
(ii) $G_N$ *is normal, so that the partition of I into the* $I^{(u)}$ *is stable by the action of G;*
(iii) *Any two distinct points of I are ultimately separated by the partitions induced by the* $G_N$, *i.e.:*

$$\varlimsup_{\substack{N \in \mathcal{N} \\ N \to \infty}} (G_N \cdot o) = \{o\}. \tag{337}$$

*Then there exists a constant* $k < 1$ *such that, for all disjoint* $J_1, J_2 \subset I$, *one has* $\{\vec{X}_{J_1} : \vec{X}_{J_2}\} \leqslant k$.

*3.6.17 Example.* For $I = \mathbb{Z}^n$ on which $G = \mathbb{Z}^n$ acts by translation, equipped with the abstract distance $\mathfrak{d}(x,y) = \{\pm(y-x)\}$, the assumptions of Lemmas 3.6.15 and 3.6.16 are checked, and these lemmas re-give resp. Lemmas 3.6.6 and 3.6.8.

*3.6.18 Example.* For $I$ the modular group $PSL_2(\mathbb{Z})$ acting by left multiplication on itself, equipped with its natural abstract distance (i.e., $\mathfrak{d}(i,j) = \{i^{-1}j, j^{-1}i\}$), the assumptions of Lemmas 3.6.15 and 3.6.16 are also checked—to see that the action of $G$ on $I$ is profinite, take for the $G_{N(l)}$ the principal congruence subgroups $\Gamma(l)$ of the modular group [38]. Notice that $PSL_2(\mathbb{Z})$ is an example of graph having negative curvature [18].

## 3.7 Appendix: Illustration of the proof of Theorem 3.3.1

☛ *This subsection is devised for the readers who would like to understand better the proof of Theorem 3.3.1 by seeing how it works on a concrete example. It only contains pedagogical material, and thus can be skipped safely.*

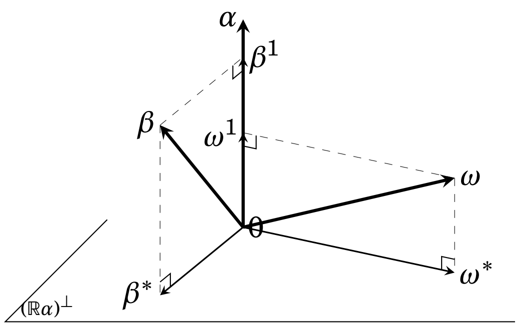
### 3.7.a A Gaussian system of variables

In this illustration we take $N = 2, M = 1$ —since $M = 1$, $Y_1$ will merely be denoted by $Y$ —, and we take $(X_1, X_2, Y)$ Gaussian (and centered), whose law is described through a $3 \times 3$ matrix *via* writing that, for some standard Gaussian vector $(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3$,

$$\begin{pmatrix} X_1 \\ X_2 \\ Y \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \omega_1 & \omega_2 & \omega_3 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}. \tag{338}$$

We denote the matrix appearing in (338) by $\mathbf{M}$. The rows of $\mathbf{M}$ will be denoted by $\alpha, \beta, \gamma \in \mathbb{R}^3$, and $(\xi_1, \xi_2, \xi_3)$ will be denoted by $\xi \in \mathbb{R}^3$. On $\mathbb{R}^3$ we will use the Euclidian scalar product "$\cdot$" and the associated norm "$\|\cdot\|$".

The advantage of this model is that, by of the general properties of Gaussian vectors (in particular Theorem 1.2.6), all the quantities of interest are computable exactly.



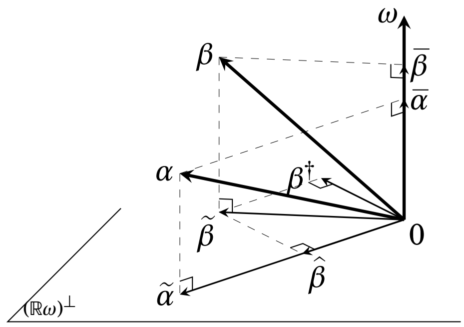




Figure 1: Visual definitions of the vectors derived from $\alpha$, $\beta$ and $\omega$: the left drawing shows how to build $\beta^1, \omega^1, \beta^*, \omega^*$; the right drawing (with different values for $\alpha, \beta, \omega$) explains the construction of $\overline{\alpha}, \overline{\beta}, \tilde{\alpha}, \tilde{\beta}, \hat{\beta}, \beta^\dagger$.

First we compute the correlation levels: by Theorem 1.2.6,

$$\{X_1 : Y\} = \frac{|\alpha \cdot \omega|}{\|\alpha\| \, \|\omega\|}, \tag{339}$$

similarly $\{X_2 : Y\} = |\beta \cdot \omega| / \|\beta\| \, \|\omega\|$; and

$$\{\vec{X} : Y\} = \sqrt{1 - \frac{|\omega \cdot (\alpha \vec{\times} \beta)|^2}{\|\omega\|^2 \, \|\alpha \vec{\times} \beta\|^2}}, \tag{340}$$

where "$\vec{\times}$" denotes the cross product on $\mathbb{R}^3$. Concerning the conditional quantities, denote by $\beta^1$, resp. $\omega^1$, the (orthogonal) projections of $\beta$, resp. $\omega$, on $\mathbb{R}\alpha$, and $\beta^*$, resp. $\omega^*$, the projections of the same vectors on $(\mathbb{R}\alpha^\perp)$, i.e. (assuming that $\alpha \neq 0$):

$$\beta^1 := (\beta \cdot \alpha / \|\alpha\|^2)\, \alpha, \quad \omega^1 := (\omega \cdot \alpha / \|\alpha\|^2)\, \alpha; \tag{341}$$

$$\beta^* := \beta - \beta^1, \quad \omega^* := \omega - \omega^1 \tag{342}$$

[see Figure 1]. Then one has $\mathbf{E}[X_2|X_1] = \beta^1_1 \xi_1 + \beta^1_2 \xi_2 + \beta^1_3 \xi_3 = \beta^1 \cdot \xi$, resp. $\mathbf{E}[Y|X_1] = \omega^1 \cdot \xi$, thus $X_2 - \mathbf{E}[X_2|X_1] = \beta^* \cdot \xi$, resp. $Y - \mathbf{E}[Y|X_1] = \omega^* \cdot \xi$. As $(X_1, X_2, Y)$ is Gaussian, the law of $(X_2 - \mathbf{E}[X_2|X_1], Y - \mathbf{E}[Y|X_1])$ under $\mathbf{P}[\cdot|X_1 = x]$ does not depend on the value of $x$; therefore we know all the conditional laws of $(X_2, Y)$ under the $\mathbf{P}[\cdot|X_1 = x]$, and for all these laws $\{X_2 : Y\}$ is equal by Theorem 1.2.6 to $|\beta^* \cdot \omega^*| / \|\beta^*\| \, \|\omega^*\|$, so in the end:

$$\{X_2 : Y\}_{X_1} = \frac{|\beta^* \cdot \omega^*|}{\|\beta^*\| \, \|\omega^*\|}. \tag{343}$$

$\{X_1 : Y\}_{X_2}$ can be computed by a similar formula.

Now let us 'dissect' the proof of Theorem 3.3.1 on our example. We take $f$ linear, namely

$$f(X_1, X_2) := X_1 + X_2, \tag{344}$$

so that all the computations shall again be tractable exactly.

Let us start with computing the quantities linked to $f^0$: one has

$$\begin{aligned} f^0 = f &= (\alpha + \beta) \cdot \xi; && (345) \\ f^0_1 = f^{\sigma(X_1)} &= X_1 + (X_2)^{\sigma(X_1)} = (\alpha + \beta^1) \cdot \xi; && (346) \\ f^0_2 = f - f^{\sigma(X_1)} &= \beta^* \cdot \xi, && (347) \end{aligned}$$

whence respectively

$$
\begin{aligned}
V = V^0 &= \|\alpha+\beta\|^2 = \|\alpha\|^2 + \|\beta\|^2 + 2\alpha\cdot\beta; && (348)\\
V_1^0 &= \|\alpha+\beta^1\|^2 = \|\alpha\|^2 + 2\alpha\cdot\beta + \frac{(\alpha\cdot\beta)^2}{\|\alpha\|^2}; && (349)\\
V_2^0 &= \|\beta^*\|^2 = \|\beta\|^2 - \frac{(\alpha\cdot\beta)^2}{\|\alpha\|^2}. && (350)
\end{aligned}
$$

By the way we check that, as claimed by Formula (208), $V^0 = V_1^0 + V_2^0$.

Now we turn to the quantities linked to $f^1$. First we have to compute the conditional laws of $(X_1, X_2)$ under the events "$Y = y$". The technique is the same as for computing $\{X_2 : Y\}_{X_1}$ a few lines above: denoting by $\bar{\alpha}$, resp. by $\bar{\beta}$, the projections of $\alpha$, resp. $\beta$, on $\mathbb{R}\omega$, and $\tilde{\alpha}$, resp. $\tilde{\beta}$, the projections of the same vectors on $(\mathbb{R}\omega^\perp)$, i.e. [see Figure 1]

$$\bar{\alpha} := (\alpha\cdot\omega/\|\omega\|^2)\omega, \quad \bar{\beta} := (\beta\cdot\omega/\|\omega\|^2)\omega; \tag{351}$$

$$\tilde{\alpha} := \alpha - \bar{\alpha}, \quad \tilde{\beta} := \beta - \bar{\beta}, \tag{352}$$

one has $\mathbf{E}[X_1|Y] = \bar{\alpha}\cdot\xi$, resp. $\mathbf{E}[X_2|Y] = \bar{\beta}\cdot\xi$, thus $X_1 - \mathbf{E}[X_1|Y] = \tilde{\alpha}\cdot\xi$, resp. $X_2 - \mathbf{E}[X_2|Y] = \tilde{\beta}\cdot\xi$; and $(X_1 - \mathbf{E}[X_1|Y], X_2 - \mathbf{E}[X_2|Y])$ has the same law under all the $\mathbf{P}[\cdot|Y = y]$. So we can compute the quantities linked to $f^1$ in the same way as we computed those linked to $f^0$: denoting

$$\hat{\beta} := \frac{\tilde{\beta}\cdot\tilde{\alpha}}{\|\tilde{\alpha}\|^2}\tilde{\alpha}; \tag{353}$$

$$\beta^\dagger := \tilde{\beta} - \hat{\beta} \tag{354}$$

[see Figure 1], one finds

$$
\begin{aligned}
f^1 &= (\tilde{\alpha}+\tilde{\beta})\cdot\xi; && (355)\\
f_1^1 &= (\tilde{\alpha}+\hat{\beta})\cdot\xi; && (356)\\
f_2^1 &= \beta^\dagger\cdot\xi, && (357)
\end{aligned}
$$

whence respectively:

$$
\begin{aligned}
V^1 &= \|\tilde{\alpha}+\tilde{\beta}\|^2; && (358)\\
V_1^1 &= \|\tilde{\alpha}+\hat{\beta}\|^2; && (359)\\
V_2^1 &= \|\beta^\dagger\|^2. && (360)
\end{aligned}
$$

As for $f^0$, we check that $V^1 = V_1^1 + V_2^1$, since $\tilde{\alpha}+\tilde{\beta}$ is the orthogonal sum of $\tilde{\alpha}+\hat{\beta}$ and $\beta^\dagger$. Moreover one always has $V^1 \leqslant V^0$, resp. $V_2^1 \leqslant V_2^0$: the first inequality follows indeed from $(\tilde{\alpha}+\tilde{\beta})$'s being the projection of $(\alpha+\beta)$ on $(\mathbb{R}\omega)^\perp$, and the second one from $\beta^\dagger$'s being the projection of $\beta^*$ on $(\mathbb{R}\omega+\mathbb{R}\alpha)^\perp$. These inequalities are consistent with the following corollary of Claim 3.3.4, obtained by applying the claim conditionally to $\mathscr{G}_{j-1}$ with the role of "$f$" played by $f^{j-1}$ and the role of "$Y$" played by $Y_j$:

**3.7.1 Proposition.** *For all $1 \leqslant j \leqslant M$, all $0 \leqslant i \leqslant N$,*

$$\sum_{i'>i} V_{i'}^j \leqslant \sum_{i'>i} V_{i'}^{j-1}. \tag{361}$$

### 3.7.b Numerical computations

Now let us see a numerical example. Our parameters will be chosen so that the function $f$ defined by (344) is optimal in the supremum (34) defining the maximal correlation coefficient $\{\vec{X}:Y\}$; other than that, the behaviour of our example will be generic:

$$\mathbf{M} = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}. \tag{362}$$

For that $\mathbf{M}$ the calculations of the previous subsection give:

| $\chi$ | $\alpha$ | $\beta$ | $\gamma$ | $\alpha\vec{\times}\beta$ | $\beta^1$ | $\omega^1$ | $\beta^*$ | $\omega^*$ | $\bar{\alpha}$ | $\bar{\beta}$ | $\tilde{\alpha}$ | $\tilde{\beta}$ | $\hat{\beta}$ | $\beta^\dagger$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\chi_1$ | 4 | 1 | 1 | −3 | 2 | 2 | −1 | −1 | 1/2 | 1/2 | 7/2 | 1/2 | 7/6 | −2/3 |
| $\chi_2$ | 1 | 4 | 1 | −3 | 1/2 | 1/2 | 7/2 | 1/2 | 1/2 | 1/2 | 1/2 | 7/2 | 1/6 | 10/3 |
| $\chi_3$ | 1 | 1 | 4 | 15 | 1/2 | 1/2 | 1/2 | 7/2 | 2 | 2 | −1 | −1 | −1/3 | −2/3 |

whence $\{X_1:Y\} = 1/2$ and $\{X_1:Y\}_{X_2} = 1/3$, thus $\{X_1:Y\}_{\mathcal{M}} = 1/2$; and similarly $\{X_2 : Y\}_{\mathcal{M}} = 1/2$.

Then Theorem 3.2.2 yields:

$$\{\vec{X}:Y\} \leqslant 1/\sqrt{2} = 0.707\ldots, \tag{363}$$

and even, according to the refinements of § 3.4.a:

$$\{\vec{X}:Y\} \leqslant \sqrt{13}/6 = 0.600\ldots; \tag{364}$$

on the other hand, the true result is:

$$\{\vec{X}:Y\} = 1/\sqrt{3} = 0.577\ldots. \tag{365}$$

So here the bound (186) is (fortunately!) correct, and even rather sharp.

Now, as the proof of Theorem 3.3.1 consists in studying the relations between the $V_i^j$, let us see what these quantities look like here. One computes:

$$\begin{pmatrix} V_1^0 & V_2^0 \\ V_1^1 & V_2^1 \end{pmatrix} = \begin{pmatrix} 40\frac{1}{2} & 13\frac{1}{2} \\ 24 & 12 \end{pmatrix}. \tag{366}$$

As a first consequence, we can check the conclusions of Proposition 3.7.1: $V_2^1 = 12 \leqslant V_2^0 = 13\frac{1}{2}$, resp. $V_1^1 + V_2^1 = 36 \leqslant V_1^0 + V_2^0 = 54$. Going further, we check the conclusions of Claim 3.3.5, which forbids the differences $V_2^0 - V_2^1$ and $(V_1^0 + V_2^0) - (V_1^1 + V_2^1)$ to be too large: for the first difference, one has $V_2^0 - V_2^1 = 1\frac{1}{2} \leqslant \varepsilon_2^2 V_2^0 = 3\frac{3}{8}$[§], and for the second one, $(V_1^0 + V_2^0) - (V_1^1 + V_2^1) = 18 \leqslant (\varepsilon_1 V_1^0 + \sqrt{V_2^0 - V_2^1})^2 = 19.419\ldots$.

[§]According to § 3.4.a, one can replace $\varepsilon_2 = 1/2$ by $\{X_2:Y\}_{X_1} = 1/3$ in this inequality. Then the inequality even becomes an equality: this is linked to the optimality of certain tensorization results for Gaussian variables, cf. § 3.5.

### 3.7.c Some traps to avoid

To finish with this appendix, I would like to comment on what is true or not about the $V_i^j$ in general situations. Proposition 3.7.1 pointed out that for all $\hat{\imath} \in \{0,\ldots,N\}$, $\sum_{i>\hat{\imath}} V_i^j$ is a nonincreasing function of $j$; in particular, when one looks at the table of the $V_i^j$, the last term ($\hat{\imath} = N-1$), resp. the total ($\hat{\imath} = 0$) of line $j$ can only decrease. Moreover, if in some line $j$ all the $V_i^j$ are zero from some position $\hat{\imath}+1$, then this property remains true in all the lower lines $j' > j$. This can be explained very simply, since saying that all the $V_i^j$ are zero from position $\hat{\imath}+1$ means indeed that $f$ is $(\mathscr{G}_j \vee \mathscr{F}_{\hat{\imath}})$-measurable, hence *a fortiori* $(\mathscr{G}_{j'} \vee \mathscr{F}_{\hat{\imath}})$-measurable. The following example, in which $f$ turns out to be $2X_1$, illustrates this phenomenon:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \qquad \Rightarrow \qquad \begin{pmatrix} V_1^0 & V_2^0 \\ V_1^1 & V_2^1 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 2 & 0 \end{pmatrix}. \tag{367}$$

However, keep careful: almost anything else you would like to say about the table of the $V_i^j$ would be false! In particular, for $i < N$, $V_i^j$ is not a nonincreasing function of $j$ in general; it is not even true that $V_i^j = 0 \Rightarrow V_i^{j'>j} = 0$, as shown by the following example:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \qquad \Rightarrow \qquad \begin{pmatrix} V_1^0 & V_2^0 \\ V_1^1 & V_2^1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1/6 & 1/2 \end{pmatrix}. \tag{368}$$

It is not true either that, if $V^j$ remains unchanged from one line to another (that is, the total of the $V_i^j$ remains unchanged), then all the $V_i^j$ are unchanged. In fact, that $V^{j+1}$ is equal to $V^j$ means that, conditionally to $\mathscr{G}_j$, $f^j$ is centered w.r.t. $Y_{j+1}$, and then $f^{j+1} = f^j$. However, the way $f^{j+1}$ decomposes into a sum of $f_i^{j+1}$ may be different to the way $f^j$ decomposed into a sum of $f_i^j$, because conditioning w.r.t. $Y_{j+1}$ may make the law of the $X_i$ change! That is what happens in the following example:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \Rightarrow \qquad \begin{pmatrix} V_1^0 & V_2^0 \\ V_1^1 & V_2^1 \end{pmatrix} = \begin{pmatrix} 1/2 & 3/2 \\ 1 & 1 \end{pmatrix}. \tag{369}$$

## 3.8 Appendix: A corollary of the Perron–Frobenius theorem

In this appendix I handle a lemma used in the proof of Theorem 3.3.1. We are working on the vector space $\mathbb{R}^N$ for some $N > 0$; a vector or a matrix is said to be $> 0$ if all its entries are positive, resp. $\geqslant 0$ if all its entries are nonnegative. Then the *Perron–Frobenius theorem* [23, Theorem 8.3.1] states that if a square matrix $A$ is $\geqslant 0$, then $A$ has some $\geqslant 0$ eigenvector for the eigenvalue $\rho(A)$. Our goal here is prove the following corollary:

**3.8.1 Lemma.** *Let $A \geqslant 0$ be a square matrix, then:*

$$\inf\{\lambda \geqslant 0 \colon (\exists u > 0)(Au \leqslant \lambda u)\} = \rho(A). \tag{370}$$

*Proof.* We prove separately each sense of the equality. Let us begin with sense "$\leqslant$". Let $v \geqslant 0$ be some eigenvector of $A$ for the eigenvalue $\rho(A)$. If $v > 0$, then the value $\lambda = \rho(A)$ checks the condition in the infimum and we are done. Otherwise if $v \not> 0$, up to a permutation of indices it has the form $(0,\ldots,0,v'_{n+1},\ldots,v'_N)$ with $0 < n < N$ and all the $v'_i$ positive. Reasoning by induction, assume that we have proved the sense "$\leqslant$" of the lemma for all $n < N$. The form of the eigenvector $v$ forces $A$ to write blockwise

$$A = \begin{pmatrix} \tilde{A} & 0 \\ * & * \end{pmatrix} \tag{371}$$

with $\mathbb{R}^{n\times n} \ni \tilde{A} \geqslant 0$. I claim that $\rho(\tilde{A}) \leqslant \rho(A)$, since if $\tilde{v}$ is an eigenvector of $\tilde{A}$ for the eigenvalue $\rho(\tilde{A})$, then for $t \geqslant 0$

$$A^t(\tilde{v},0,\ldots,0) = (\rho(\tilde{A})^t\tilde{v},*,\ldots,*), \tag{372}$$

so

$$\overline{\lim_{t\to\infty}}\, \rho(\tilde{A})^{-t}|A^t(\tilde{v},0,\ldots,0)| > 0^{[¶]} \tag{373}$$

and consequently $\rho(A) \geqslant \rho(\tilde{A})$. Now let $\varepsilon > 0$. By induction hypothesis there exists some $\mathbb{R}^n \ni w > 0$ such that $\tilde{A}w \leqslant (\rho(\tilde{A})+\varepsilon)w$. Thus for $\eta > 0$, $\mathbb{R}^N \ni (\eta w, v') > 0$ and

$$A(\eta w,v') = (\eta\tilde{A}w, \rho(A)v' + O(\eta)) \leqslant (\eta(\rho(\tilde{A})+\varepsilon)w, \rho(A)v' + O(\eta)) \overset{\eta\searrow 0}{\leqslant} (\rho(A)+\varepsilon)(\eta w, v'). \tag{374}$$

So $(\rho(A)+\varepsilon)$ checks the condition in the right-hand side of the infimum, which ends the proof of the sense "$\leqslant$" of (370).

For the sense "$\geqslant$", consider any $\mathbb{R}^N \ni u > 0$ and let again $v \geqslant 0$ be some eigenvector of $A$ for the eigenvalue $\rho(A)$. Then there exists a (unique) $\beta \geqslant 0$ such that $u - \beta v \geqslant 0$ but $u - \beta v \not> 0$. For this $\beta$, one of the entries of $\beta v$ and $u$ is the same, say $\beta v_{i_0} = u_{i_0}$. So if $\lambda < \rho(A)$,

$$\lambda u_{i_0} < \rho(A)u_{i_0} = \rho(A)\beta v_{i_0} = \big(A(\beta v)\big)_{i_0} \leqslant \big(A(\beta v)\big)_{i_0} + \big(A(u-\beta v)\big)_{i_0} = (Au)_{i_0}, \tag{375}$$

thus $Au \not\leqslant \lambda u$. That relation being true for any $u > 0$, $\lambda$ does not check the condition in the infimum, which proves the sense "$\geqslant$" of (370). □

## 3.9 Appendix: A geometric consequence of results on correlations

As I pointed out in Remark 3.5.6, for Gaussian vectors maximal correlations can be interpreted in terms of Euclidian spaces. In this appendix I will present a funny corollary of Lemma 3.5.5 following from this interpretation. That result itself is actually more or less a pretext: the real goal of this appendix is in fact to show in an eloquent way the geometric meaning of maximal correlations and the Hilbertian frame that underlies them.

First we need some vocabulary about Euclidian spaces:

[¶]Equation (373) is meaningless if $\rho(\tilde{A}) = 0$, but in that case there is nothing to prove.

**3.9.1 Definition.**

1. For $L_1,L_2$ two vector lines in the Euclidian space $\mathbb{R}^2$, or more generally in any Hilbert space, we call *geometric angle* between $L_1$ and $L_2$, denoted by $\widehat{L_1L_2}$, their "angle" in the elementary sense: for arbitrary $\vec{a} \in L_1 \smallsetminus \{0\}, \vec{b} \in L_2 \smallsetminus \{0\}$,

$$\widehat{L_1L_2} = \operatorname{Arccos} \frac{|\vec{a}\cdot\vec{b}|}{\|\vec{a}\|\,\|\vec{b}\|} \in [0,\pi/2]. \tag{376}$$

2. For $L_1,L_2$ and $L_3 \neq L_1,L_2$ three vector lines in the Euclidian space $\mathbb{R}^3$ (or any Hilbert space), we call *apparent angle between $L_1$ and $L_2$ seen from $L_3$* the geometric angle that an observer located somewhere on $L_3 \smallsetminus \{0\}$ would have the impression, due to perspective, that $L_1$ and $L_2$ make [see Figure 2]: technically, it is the geometric angle $\widehat{L_1'L_2'}$, where $L_1'$ and $L_2'$ are the respective orthogonal projections of $L_1$ and $L_2$ onto the plane $(L_3)^\perp$.

Then one has the following corollary of Lemma 3.5.5:

**3.9.2 Theorem.** *Let $L_1,L_2,L_3$ be three distinct vector lines of $\mathbb{R}^3$. Denote $\widehat{A} := \widehat{L_2L_3}, \widehat{B} := \widehat{L_3L_1}, \widehat{\Omega} := \widehat{L_1L_2}$, and denote by $\widehat{A}'$ the apparent angle between $L_2$ and $L_3$ seen from $L_1$, resp. $\widehat{B}'$ the apparent angle between $L_3$ and $L_1$ seen from $L_2$, etc.. Then the relative order of $\widehat{A}$ and $\widehat{A}'$ is the same as the relative order of $\widehat{B}$ and $\widehat{B}'$ and as the relative order of $\widehat{\Omega}$ and $\widehat{\Omega}'$, i.e., "$\widehat{A}' < \widehat{A}$" (resp. "$\widehat{A}' = \widehat{A}$", resp. "$\widehat{A}' > \widehat{A}$") is equivalent to "$\widehat{B}' < \widehat{B}$" (resp. "$\widehat{B}' = \widehat{B}$", resp. "$\widehat{B}' > \widehat{B}$"), etc..*

*3.9.3 Remark.* I found Theorem 3.9.2 by chance, one day that I was looking for a situation where one would have $\widehat{B}' > \widehat{B}$ but $\widehat{A}' < \widehat{A}$, in order to build a 'nice' example for § 3.7.b. I thought that such a situation would be generic, but after having looked for it without success, I realized that it was actually impossible, and that the explanation had a simple interpretation in terms of correlations.

*Proof.* Fix three arbitrary nonzero vectors $\alpha,\beta,\omega$ of resp. $L_1,L_2,L_3$; and consider the Gaussian system (338) of § 3.7 for these vectors. Then the correlation coefficients between $X_1$, $X_2$ and $Y$ can be interpreted as angles between $L_1$, $L_2$ and $L_3$; more precisely, one has the following correspondance:

**3.9.4 Proposition.**

*(i) $\{X_1 : X_2\}$ is the cosine of the geometric angle between $L_1$ and $L_2$;*
*(ii) $\{X_1 : X_2\}_Y$ is the cosine of the apparent angle between $L_1$ and $L_2$ seen from $L_3$.*

*Proof.* (i) is nothing but the Euclidian interpretation of Theorem 1.2.6. (ii) follows from the fact that, in the vector space spanned by jointly Gaussian real random variables, conditional expectation corresponds to orthogonal projection and independence corresponds to orthogonality.
□

By Proprosition 3.9.4, in our situation Lemma 3.5.5 gives:

$$\{(X_1,X_2) : Y\} = \sqrt{1 - \sin^2\widehat{B}\,\sin^2\widehat{A}'}. \tag{377}$$

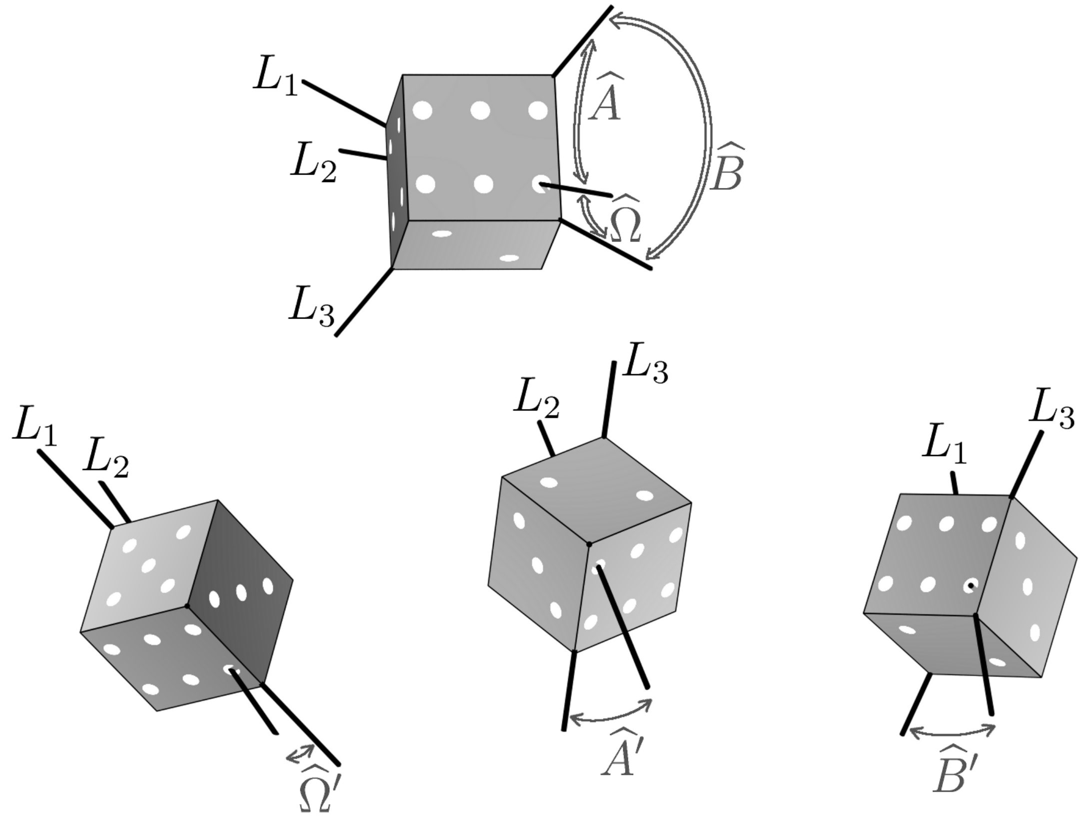



Figure 2: This figure shows four different views of the same 3-dimensional object. What interests us actually is only the three concurrent lines $L_1$,$L_2$,$L_3$, but we added a die centered at their point of concurrency to see depth better on the pictures [Recall that on a die, the total number of points on two opposite faces is always 7.]. On the top picture, the die is shown in generic position. We represent the angles $\widehat{A}$, $\widehat{B}$ and $\widehat{\Omega}$; these angles are 3-dimensional angles, which we underline by drawing them with double strokes. On each of the bottom pictures, the die is viewed from the direction of one of the lines (from the left to the right, $L_3$, $L_1$ and $L_2$), so that this line appears completely foreshortened. We represent the angles $\widehat{\Omega}'$, $\widehat{A'}$ and $\widehat{B}'$ made by the two other lines *as they appear on the drawing*; we underline that these angles are 2-dimensional by drawing them with simple strokes [Note that in the case of $\widehat{B}'$, the angular sector representing $\widehat{B}'$ is not the projection of the angular sector representing $\widehat{B}$, but its supplementary—otherwise $\widehat{B}'$ would be greater than $\pi/2$, which would contradict our 'geometric' definition of angles.]. On this example one has $\widehat{A} \simeq 58°$,$\widehat{B} \simeq 71°$,$\widehat{\Omega} \simeq 15°$ and $\widehat{A'} \simeq 30°$,$\widehat{B}' \simeq 34°$,$\widehat{\Omega}' \simeq 9°$; so, perspective makes angles appear smaller than they are really for all three pairs of lines, which is in accordance with Theorem 3.9.2.

Obviously the roles of $X_1$ and $X_2$ can be interchanged in the above argument, yielding:

$$\{(X_2, X_1) : Y\} = \sqrt{1 - \sin^2 \widehat{A} \sin^2 \widehat{B}'}. \tag{378}$$

But $(X_1, X_2)$ and $(X_2, X_1)$ generate the same $\sigma$-algebra, so $\{(X_1, X_2) : Y\} = \{(X_2, X_1) : Y\}$, and thus, comparing (377) and (378):

$$\frac{\sin \widehat{A'}}{\sin \widehat{A}} = \frac{\sin \widehat{B'}}{\sin \widehat{B}}. \tag{379}$$

This implies in particular that $\sin \widehat{A}, \sin \widehat{A'}$ and $\sin \widehat{B}, \sin \widehat{B}'$ have the same relative order, so also do $\widehat{A}, \widehat{A'}$ and $\widehat{B}, \widehat{B}'$. A cyclic permutation of $L_1$, $L_2$ and $L_3$ shows that the result is still valid for $\widehat{\Omega}, \widehat{\Omega}'$. $\square$

*3.9.5 Example.* See Figure 2 on page 85.

# Chapter 4

# Other applications of tensorization techniques

In the previous chapter we have been seeing how maximal decorrelation hypotheses between pairs of 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 hypotheses.

But the techniques of § 3 can also be applied to get other types of results. In this chapter I am going to show how, from maximal decorrelation hypotheses, one can get results on some classical features of particle systems which are not linked with maximal correlations *a priori*.

I will deal with two such features. First, I will look at the implications of $\rho$-mixing on the existence of a central limit theorem—more precisely, of a *spatial* central limit theorem, since I am more interested in random fields than in sequences (variables indexed by $\mathbb{Z}^n$ rather than by $\mathbb{Z}$). Very sharp results concerning this issue are already known; however, I find interesting to show how it goes with my 'tensorization-like' approach: this approach takes indeed a quite different way to do the job, which may be neater by certain sides. Moreover, the results are stated with a slighlty different vocabulary—namely, subjective correlations.

Next, I will look at the question of spectral gap for Glauber dynamics. Though this point has already been thouroughly studied in a $\beta$-mixing paradigm, this work, to the best of my knowledge, is the first to show how $\rho$-mixing can be used to tackle this issue.

My main goal here is just to show *how* the techniques of this monograph may be applied to the problems of spatial central limit theorem and convergence of the Glauber dynamics. Accordingly, I favoured the simplicity on proofs against the refinement of the results.

# 4.1 Spatial central limit theorem

## 4.1.a Introduction

A fundamental result in probability theory is the central limit theorem (CLT), which, in its standard statement, requires an assumption of complete independence. It is natural to wonder whether that assumption can be relaxed into an hypothesis of 'near independence'. Maximal decorrelations are a natural frame for such a generalization, since the CLT already takes place in an $L^2$ setting.

Our point of view is motivated by statistical physics. Let $\mathbb{Z}^n$ be a lattice, on each vertex $i$ of which there is a random 'spin' $X_i$ ranged in some space $\mathscr{X}$ not depending on $i$. We assume that the law of the system is translation invariant, i.e. that for all $z \in \mathbb{Z}^n$, $(X_{i+z})_{i\in\mathbb{Z}^n}$ has the same law as $\vec{X}_{\mathbb{Z}^n}$. Then, for all $z \in \mathbb{Z}^n$, we denote

$$\varepsilon_z = \{X_i : X_{i+z}\}_*. \tag{380}$$

We are interested in situations where the $\varepsilon_z$ are sufficiently 'rapidly decreasing' as $|z| \to \infty$ so that $\sum_{z\in\mathbb{Z}^n} \varepsilon_z < \infty$.

Let $f\colon\ \mathscr{X} \to \mathbb{R}$ be a function such that $f(X_0)$ is square-integrable and centered. The question is, does one get a CLT when summing $f(X_i)$ for $i$ in a large subset of $\mathbb{Z}^n$, 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

$$\frac{1}{\sqrt{l^n}} \sum_{\substack{i\in\mathbb{Z}^n \\ 0\leqslant i_1,\dots,i_n<l}} f(X_i) \tag{381}$$

to weakly converge, when $l \to \infty$, to some Gaussian distribution.

*4.1.1 Remark.* Note that the limit distribution, if it exists, will have to be centered, but its variance will not be equal to $\mathrm{Var}\big(f(X_0)\big)$ in general.

In the case $n = 1$, extremely sharp results for this topic have been known from long; let us cite, among many others, [40, 24, 33, 6]. For $n \geqslant 2$, similar results also exist; see e.g. [6, Theorem 5] for such a result, and [8, § 29] for a survey of the topic. All these proofs relie on some 'coupling' between (bunches of) the spins and other convenient variables which are close to them, but which are *actually* independent, so as to deduce the CLT for the former from the CLT for the latter. On the other hand, my proof will mimick Lévy's proof of the CLT, hence needing no coupling argument.

*A priori* the results presented here do not improve the state of the art; however, when turning to quantitative versions of these results, it is likely that the difference between the usual method and mine would yield a difference in the corresponding non-asymptotic bounds obtained.

## 4.1.b Product of weakly coupled variables

My results relie on the following

**4.1.2 Lemma.** *Let $N \geqslant 1$ and let $\dot{\mathscr{F}}_1,\dots,\dot{\mathscr{F}}_N$ be σ-algebras with $\{\dot{\mathscr{F}}_i : \dot{\mathscr{F}}_j\} \leqslant \varepsilon_{ij}$, and denote*

$$\bar{\varepsilon} = \sup_i \sum_{j\neq i} \varepsilon_{ij}. \tag{382}$$

*Let $\Phi_1,\dots,\Phi_N$ be complex-valued random variables with $|\Phi_i| \leqslant 1$ a.s., such that $\Phi_i$ is $\dot{\mathscr{F}}_i$-measurable for all $i$, with all the $\Phi_i$ having the same distribution. Then, denoting by $\varphi$ the common value of the $\mathbf{E}[\Phi_i]$,*

$$\left|\mathbf{E}\Big[\prod_i \Phi_i\Big] - \varphi^N\right| \leqslant N\bar{\varepsilon}(1+\bar{\varepsilon})(1-|\varphi|^2). \tag{383}$$

*Proof.* Denote $\delta := \mathrm{Sd}(\Phi)$. Since $\mathbf{E}[|\Phi^2|] \leqslant 1$, the definition of (complex) variance ensures that $\delta \leqslant \sqrt{1-|\varphi|^2}$.

For all $i \in \{0,\dots,N\}$, denote $\mathscr{F}_i := \bigvee_{i'\leqslant i} \dot{\mathscr{F}}_i$ ; denote $\Psi^{(i)} := \prod_{i'\leqslant i} \Phi_{i'}$; define

$$\Psi^{(i)}_j := (\Psi^{(i)})^{\mathscr{F}_j} - \mathbf{E}[\Psi^{(i)}|\mathscr{F}_{j-1}] \tag{384}$$

and denote $\Delta^{(i)}_j := \mathrm{Sd}(\Psi^{(i)}_j)$. Also denote

$$\Phi_{i,j} := (\Phi_i)^{\mathscr{F}_j} - \mathbf{E}[\Phi_i|\mathscr{F}_{j-1}]. \tag{385}$$

Usual manipulation on conditioning shows that, for $i \geqslant 1$,

$$\Psi^{(i)}_j = \Psi^{(i-1)}_j (\Phi_i)^{\mathscr{F}_j} + \Psi^{(i-1)}_{j-1}\Phi_{i,j}. \tag{386}$$

Since $\|\Phi_i\|_{L^\infty} \leqslant 1$, one has also $\|(\Phi_i)^{\mathscr{F}_j}\|_{L^\infty} \leqslant 1$, hence

$$\mathrm{Sd}\big(\Psi^{(i-1)}_j (\Phi_i)^{\mathscr{F}_j}\big) \leqslant \mathrm{Sd}(\Psi^{(i-1)}_j) = \Delta^{(i-1)}_j. \tag{387}$$

Similarly, it is obvious that $\|\Psi^{(i-1)}\|_{L^\infty} \leqslant 1$, whence

$$\mathrm{Sd}\big(\Psi^{(i-1)}_{j-1}\Phi_{i,j}\big) \leqslant \mathrm{Sd}(\Phi_{i,j}). \tag{388}$$

Now, I claim that

**4.1.3 Claim.**

$$\mathrm{Sd}(\Phi_{i,j}) \leqslant \varepsilon_{ij}\delta. \tag{389}$$

Indeed, $\Phi_{i,j}$ is the part relative to $\dot{\mathscr{F}}_j$ of the $\dot{\mathscr{F}}_i$-measurable function $\Phi_i$, whose standard deviation is $\delta$.

In the end, we got that

$$\Delta^i_j \leqslant \Delta^{(i-1)}_j + \varepsilon_{ij}\delta. \tag{390}$$

Since $\Delta^0_j = 0$, one has therefore:

$$\forall i,j \qquad \Delta^i_j \leqslant (1+\bar{\varepsilon})\delta. \tag{391}$$

Now, denoting $\psi^{(i)} := \mathbf{E}[\Psi^{(i)}]$, one has

$$|\psi^{(i)} - \varphi\psi^{(i-1)}| = \Big|\sum_{j<i} \mathbf{E}\big[\Psi^{(i-1)}_j \Phi_{i,j}\big]\Big| \leqslant \sum_{j<i} \Delta^{(i-1)}_j \mathrm{Sd}(\Phi_{i,j}) \leqslant \bar{\varepsilon}(1+\bar{\varepsilon})\delta^2, \tag{392}$$

and finally

$$|\psi^{(N)} - \varphi^N| = \sum_{i=1}^N |\varphi|^{N-i}|\psi^{(i)} - \varphi\psi^{(i-1)}| \leqslant \sum_{i=1}^N |\psi^{(i)} - \varphi\psi^{(i-1)}| \leqslant N\bar{\varepsilon}(1+\bar{\varepsilon})\delta^2, \tag{393}$$

which is (383) if you recall that $\delta^2 \leqslant 1-|\varphi|^2$. □

### 4.1.c A spatial CLT

First I state and prove a CLT on cubes:

**4.1.4 Theorem.** *Consider a translation-invariant spin model on a lattice $\mathbb{Z}^n$ and define $\varepsilon_z$ by (380). Assume that $\sum_{z\in\mathbb{Z}^n}\varepsilon_z<\infty$. Then for any centered square-summable function $f\colon\ \mathscr{X}\to\mathbb{R}$, there exists a constant $\sigma<\infty$ such that*

$$\frac{1}{\sqrt{l^n}}\sum_{\substack{i\in\mathbb{Z}^n\\ 0\leqslant i_1,\dots,i_n<l}} f(X_i)\overset{l\to\infty}{\rightharpoonup}\mathcal{N}(\sigma^2), \tag{394}$$

*where "$\rightharpoonup$" denotes convergence in law.*

*Proof.* Denote by $F(l)$ —or merely $F$ —the left-hand side of (394).

What will be the value of $\sigma$? Clearly we must have

$$\sigma^2=\lim_{l\to\infty}\operatorname{Var}(F(l)), \tag{395}$$

which yields

$$\sigma=\sqrt{\sum_{z\in\mathbb{Z}}\mathbf{E}\left[f(X_0)f(X_z)\right]}, \tag{396}$$

where the expression under the root sign, which is necessarily nonnegative, is finite because $|\mathbf{E}[f(X_0)f(X_z)]|\leqslant\varepsilon_z\operatorname{Sd}(f(X_0))\operatorname{Sd}(f(X_z))=\varepsilon_z\operatorname{Var}(f)$. By the way, we will denote

$$\sigma_*^2:=\Big(\sum_z\varepsilon_z\Big)\|f\|_{L^2}^2. \tag{397}$$

Fix some arbitrary $\eta>0$. The assumption that $\sum\varepsilon_z<\infty$ implies the existence of an $l_0<\infty$ such that

$$\sum_{|z|_\infty>l_0}\varepsilon_z\leqslant\eta, \tag{398}$$

where $|z|_\infty$ denotes $\max(|z_1|,\dots,|z_n|)$. By (395), we can also fix an $l_1<\infty$ such that

$$|\operatorname{Var}(F(l))-\sigma^2|\leqslant\eta. \tag{399}$$

Now we will 'tile' the cube of size $l$ into a 'patchwork' made of cubes of size $l_1$ which I call "tiles", each tile being at distance at least $l_0$ from the others, plus some "scrap". I denote by $\widetilde{F}$ the part of $F$ due to the tiles and by $F^*$ the part of $F$ due to the scrap.

Index the tiles by $\{1,\dots,N\}$, with $N:=\lfloor(l+l_0)/(l_1+l_0)\rfloor^n$. We write, with obvious notation, $\widetilde{F}=:F_1+\cdots+F_N$. For $\lambda\in\mathbb{R}$, denote

$$\Psi(\lambda,l):=\exp(i\lambda\widetilde{F}),\qquad \psi(\lambda,l):=\mathbf{E}[\Psi(\lambda,l)]; \tag{400}$$

$$\Phi_j(\lambda,l):=\exp(i\lambda F_j),\qquad \varphi(\lambda,l):=\mathbf{E}[\Phi_j(\lambda,l)]. \tag{401}$$

Then we are exactly in situation of applying Lemma 4.1.2, which yields:

$$|\psi(\lambda,l)-\varphi(\lambda,l)^N|\leqslant N\eta(1+\eta)(1-|\varphi(\lambda,l)|^2). \tag{402}$$

Let us look at the asymptotics of Formula (402) when $l \to \infty$. We observe that, denoting

$$F_{\mathrm{t}} := \frac{1}{\sqrt{l_1^n}} \sum_{i \in \textit{fixed tile}} f(X_i), \tag{403}$$

one has

$$F_j \stackrel{\text{law}}{=} \frac{\sqrt{l_1^n}}{\sqrt{l^n}} F_{\mathrm{t}}. \tag{404}$$

Since $F_{\mathrm{t}}$ is centered, its Fourier transform satisfies $\hat{F}_{\mathrm{t}}(0) = 1$, $\hat{F}'_{\mathrm{t}}(0) = 0$ and $\hat{F}''_{\mathrm{t}} = \mathrm{Var}(F_{\mathrm{t}})$, so that

$$\lim_{l \to \infty} l^n (1 - \varphi(\lambda, l)) = \frac{\lambda^2}{2} l_1^n \mathrm{Var}(F_{\mathrm{t}}), \tag{405}$$

where, denoting $\sigma_{l_1}^2 := \mathrm{Var}(F_{\mathrm{t}})$, we recall that $l_1$ has been taken sufficiently large so that $|\sigma_{l_1}^2 - \sigma^2| \leqslant \eta$. Then, since $N \stackrel{l \to \infty}{\sim} l^n/(l_1 + l_0)^n$, one has the following asymptotics for (402):

$$\varphi(\lambda, l)^N \quad \stackrel{l \to \infty}{\to} \quad \exp\left[ -\sigma_{l_1}^2 \Big( \frac{l_1}{l_1 + l_0} \Big)^n \frac{\lambda^2}{2} \right]; \tag{406}$$

$$N(1 - |\varphi(\lambda, l)|^2) \quad \stackrel{l \to \infty}{\to} \quad \sigma_{l_1}^2 \Big( \frac{l_1}{l_1 + l_0} \Big)^n \lambda^2. \tag{407}$$

It remains to control the contribution of $F^*$.

**4.1.5 Claim.** *There are at most*

$$\left[ 1 - \Big( \frac{l_1}{l_1 + l_0} \Big)^n \right] l^n + n l_1 l^{n-1} \tag{408}$$

*scrap spins.*

By Claim 4.1.5,

$$\|F^*\|_{L^1} \leqslant \mathrm{Var}(F^*) \leqslant \left[ 1 - \Big( \frac{l_1}{l_1 + l_0} \Big)^n + n \frac{l_1}{l} \right] \sigma_*^2; \tag{409}$$

then the contribution of $F^*$ is controlled using the following immediate

**4.1.6 Lemma.** *Let $X$ and $H$ be real random variables with $\|H\|_{L^1} < \infty$. Then, for $\lambda \in \mathbb{R}$,*

$$|\mathbf{E}[e^{i\lambda(X+H)}] - \mathbf{E}[e^{i\lambda X}]| \leqslant |\lambda| \|H\|_{L^1}. \tag{410}$$

In the end, putting everything together we get:

$$\overline{\lim_{l \to \infty}} |\psi(\lambda, l) - e^{-\sigma^2 \lambda^2 / 2}| \leqslant$$
$$|e^{-(\sigma^2 - \eta)[l_1/(l_1 + l_0)]^n \lambda^2/2} - e^{-\sigma^2 \lambda^2/2}| + \eta(1+\eta) \Big( \frac{l_1}{l_1 + l_0} \Big)^n (\sigma^2 + \eta) \lambda^2 + \sqrt{1 - \Big( \frac{l_1}{l_1 + l_0} \Big)^n} \sigma_*. \tag{411}$$

Since there were no upper restriction on the value of $l_1$, we can assume that we have taken it such that $[l_1/(l_1+l_0)]^n \geqslant 1-\eta$. Then (411) becomes:

$$\overline{\lim_{l\to\infty}} |\psi(\lambda,l) - e^{-\sigma^2\lambda^2/2}| \leqslant \left|e^{-(\sigma^2-\eta)(1-\eta)\lambda^2/2} - e^{-\sigma^2\lambda^2/2}\right| + \eta(1+\eta)(1-\eta)(\sigma^2+\eta)\lambda^2 + \sqrt{\eta}\sigma_*. \tag{412}$$

The right-hand side of (412) can be made arbitrarily close to 0 by taking $\eta$ small enough, so we have proved that

$$\forall\lambda\in\mathbb{R} \quad \mathbf{E}[e^{i\lambda F(l)}] \overset{l\to\infty}{\to} e^{-\sigma^2\lambda^2/2}. \tag{413}$$

By Lévy's theorem on characteristic functions, this is tantamount to saying that $F(l)$ converges in law to $\mathcal{N}(\sigma^2)$. □

The CLT should remain valid for other shapes than a cube, since morally the random field $f(X_i)$ should look like a Gaussian white noise at large scales. Indeed, the same proof as above yields a CLT for general shapes, where moreover convergence is uniform in the shape considered in some way. Let us give a precise statement:

**4.1.7 Definition.** An open subset $U \subset \mathbb{R}^n$ (not necessarily connected) is said to be $\mathscr{C}^2$ if its boundary $M$ is a $\mathscr{C}^2$ submanifold of $\mathbb{R}^n$ (of codimension 1). We define the *smoothness* of $U$, denoted by $\kappa(U)$, as

$$\kappa(U) := \sup_{x\in M} \|\!| \mathrm{II}(x) |\!\|, \tag{414}$$

where $\mathrm{II}(\cdot)$ denotes the shape tensor of $M$ [22, Chapter 10], which measures the local deviation of $M$ from being flat. Also, the Lebesgue measure of $U$ will be denoted by $\mathrm{vol}(U)$.

**4.1.8 Theorem.** *Consider a translation-invariant spin model on a lattice $\mathbb{Z}^n$ and define $\varepsilon_z$ by (380). Assume that $\sum_{z\in\mathbb{Z}^n}\varepsilon_z < \infty$. Then for any centered square-summable function $f\colon \mathscr{X} \to \mathbb{R}$, if $(U_k)_{k\in\mathbb{N}}$ is a sequence of $\mathscr{C}^2$ bounded subsets of $\mathbb{R}^n$ with $\sup_k \kappa(U_k) < \infty$ and $(l_k)_{k\in\mathbb{N}}$ is a sequence of positive numbers tending to infinity,*

$$\frac{1}{\sqrt{l_k^n\,\mathrm{vol}(U_k)}} \sum_{i\in l_kU_k\cap\mathbb{Z}^n} f(X_i) \overset{k\to\infty}{\rightharpoonup} \mathcal{N}(\sigma^2), \tag{415}$$

*where $\sigma^2$ is the same as in Theorem 4.1.4.*

*Proof.* Just copy the proof of Theorem 4.1.4. The only difference lies in proving the analoguous of Claim 4.1.5, which is where one needs the $\kappa(U_k)$ to be bounded. Observe that we use the non-asymptotic form of our intermediate bounds to get a result independent of the precise shape of the $U_k$.
□

*4.1.9 Remark.* Another generalization of the CLT, still based on the idea that the field $f(X_i)$ looks like a Gaussian white noise at large scales, is the statement that for $\varphi$ a continuous function with compact support,

$$\frac{1}{\sqrt{l^n}} \sum_{i\in\mathbb{Z}^n} \varphi(X_i/l) f(X_i) \overset{l\to\infty}{\rightharpoonup} \mathcal{N}(\sigma^2 \textstyle\int_{\mathbb{R}^n} \varphi(x)^2\,dx). \tag{416}$$

This can be proved with the same methods as before.

## 4.2 Spectral gap for the Glauber dynamics

### 4.2.a Introduction

In this section we are looking at a probabilistic system made of a large number of 'elementary' random variables $(X_i)_{i\in I}$ — $I$ may be seen as lattice and $X_i$ as the state of the particle being at site $i$. As is customary by now, theorems will only be stated in the case where $I$ is finite, the infinite case being got by passing to the limit.

**4.2.1 Definition.** Denoting by $\Omega$ the states space of $\vec{X}_I$, let $\mathbf{P}$ be a probability measure on $\Omega$. The *Glauber dynamics* [20, 16] associated to $\mathbf{P}$ is the Markov process on $\Omega$ having the following law: on each $i \in I$ there is an alarm clock, all the clocks being independent and ringing with law *Poisson*(1). When a clock rings, the state of spin $X_i$ —and only it—is flipped so that the state of $X_i$ immediately after the flip follows the law $\mathbf{P}(X_i|\vec{X}_{I\smallsetminus\{i\}})$.

In formal terms, the Glauber dynamics is the Markov process whose generator $\mathscr{L}$ on $L^\infty(\Omega)$ is defined by:

$$(\mathscr{L}f)(\vec{x}_I) = \sum_{i\in I} \mathbf{E}\big[f(\vec{X}_I) - f(\vec{x}_I)|\vec{X}_{I\smallsetminus\{i\}} = \vec{x}_{I\smallsetminus\{i\}}\big]. \tag{417}$$

Let us recall some basic facts on the Glauber dynamics (see [28, Chapter IV] for more details). By construction $\mathbf{P}$ is a reversible equilibrium measure for the dynamics, so $\mathscr{L}$ is self-adjoint on $L^2(\mathbf{P})$. Since obviously $\mathscr{L}1 \equiv 0$, one can also define $\mathscr{L}$ on $\bar{L}^2(\mathbf{P})$, on which it is self-adjoint too. This leads to the following definition:

**4.2.2 Definition.** The *energy* of $f \in \bar{L}^2(\mathbf{P})$ is

$$\mathscr{E}(f,f) = \langle Lf, f\rangle \tag{418}$$

The following immediate identity shows that $\mathscr{E}$ is always a nonnegative bilinear form:

**4.2.3 Proposition.**

$$\mathscr{E}(f,f) = \int_\Omega d\mathbf{P}[\vec{x}_I] \sum_i \mathrm{Var}\big(f|\vec{X}_{I\smallsetminus\{i\}} = \vec{x}_{I\smallsetminus\{i\}}\big). \tag{419}$$

**4.2.4 Definition.** For $\lambda > 0$, the Glauber dynamics is said to have *spectral gap* $\geqslant \lambda$ if, for all $f \in \bar{L}^2(\mathbf{P})$,

$$\mathscr{E}(f,f) \geqslant \lambda \mathrm{Var}(f). \tag{420}$$

What makes spectral gap interesting is that its positiveness is equivalent to exponential convergence to 0 of the semigroup $(e^{-t\mathscr{L}})_{t\geqslant 0}$ on $\bar{L}^2(\mathbf{P})$, the rate of convergence being equal to the width of the spectral gap. As the Glauber dynamics is one of the easiest ways to simulate the law $\mathbf{P}$ for complicated models, the stake of having exponential convergence for it is evident.

Many works have been done on the spectral gap of the Glauber dynamics, see for instance Martinelli's St-Flour course [29]. Several results state that, the less spins are correlated, the larger the spectral gap is. Yet the researchers who work on this topic generally express the decorrelation between the spins in terms of $\beta$-mixing (cf. Definition 0.1.4), while it seems be more natural to look at them in terms of maximal decorrelations, since the formula (420) stating the spectral gap problem takes place in a Hilbertian frame itself. Thus my goal here will be to find a control on the spectral gap expressed in terms of $\rho$-mixing conditions. Since maximal correlations look to be the minimal frame to study the spectral gap for the Glauber dynamics, hopefully the bounds yielded by this method will be sharp.

Another noticeable feature of my approach is that it remains at a quite abstract level: no symmetry property of $I$ or $\mathbf{P}$ need be assumed, all the work essentially consisting in manipulating relevant quadratic forms.

### 4.2.b A lower bound for the spectral gap

The central theorem of this section is the following:

**4.2.5 Theorem.** *Take $I = \{1,\ldots,N\}$. Suppose that for all distinct $i,j \in I$ one has $\{X_i : X_j\}_* \leqslant \varepsilon_{ij} < 1$ —we will make the costless assumption that $\varepsilon_{ji} = \varepsilon_{ij}$. For $i \in I$, denote*

$$\tilde{1}_i := \frac{1}{\prod_{i<j\leqslant N}(1-\varepsilon_{ij}^2)} = \frac{\tilde{\varepsilon}_{iN}}{\varepsilon_{iN}}, \tag{421}$$

*and for $i<j$, denote*

$$\tilde{\varepsilon}_{ij} := \frac{\varepsilon_{ij}}{\prod_{i<j'\leqslant j}(1-\varepsilon_{ij'}^2)}. \tag{422}$$

*Then the Glauber dynamics has spectral gap at least $|||M|||^{-2}$, where $M$ is the $(N\times N)$ matrix defined by*

$$M = \begin{pmatrix} 1 & -\tilde{\varepsilon}_{12} & \cdots & -\tilde{\varepsilon}_{1N} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\tilde{\varepsilon}_{(N-1)N} \\ 0 & \cdots & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{1}_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \tilde{1}_N \end{pmatrix}. \tag{423}$$

*4.2.6 Remark.* The form of the first matrix in the right-hand side of (423) ensures that it is invertible. Since moreover all the $\varepsilon_{ij}$ were supposed $<1$, all the $\tilde{\varepsilon}_{ij}$ and the $\tilde{1}_i$ are finite; thus, the lower bound $|||M|||^{-2}$ is strictly positive.

*Proof.* Let $f$ be a centered square-integrable function on $(\Omega,\mathbf{P})$. For $I'\subset I$, denote $\mathscr{F}_{I'} := \sigma(\vec{X}_{I'})$. For $i\in I$, $I'\subset I\smallsetminus\{i\}$, denote

$$f_i^{I'} := f^{\mathscr{F}_{I'\uplus\{i\}}} - \mathbf{E}[f|\mathscr{F}_{I'}]; \tag{424}$$

define moreover

$$f_i^{\neq} := f_i^{I\smallsetminus\{i\}}; \tag{425}$$

$$f_i^{<} := f_i^{\{1,\ldots,i-1\}}. \tag{426}$$

Then by Proposition 4.2.3, one has

$$\mathscr{E}(f,f) = \sum_i \mathrm{Var}(f_i^{\neq}), \tag{427}$$

while the usual telescopic argument shows that

$$\mathrm{Var}(f) = \sum_i \mathrm{Var}(f_i^{<}). \tag{428}$$

So to prove the theorem, we have to establish links between the different values $\mathrm{Var}(f_i^{I'})$. It will be convenient to introduce the shorthands $\Delta_i^{I'} = \mathrm{Sd}(f_i^{I'})$. One has the following

**4.2.7 Claim.** *For $I' \subset I$ and $i,j \in I \smallsetminus I'$ with $j \neq i$,*

$$\Delta_i^{I'} \leqslant \Delta_i^{I' \uplus \{j\}} + \varepsilon_{ij} \Delta_j^{I'}. \tag{429}$$

*Proof.* Assume in a first time that $I' = \emptyset$, and denote $f_i := f_i^{\emptyset}$, $f_j := f_j^{\emptyset}$, $f_i^j := f_i^{\{j\}}$ and $\mathscr{F}_i := \mathscr{F}_{\{i\}}$. Projecting the decomposition "$f_i = f_i^j + (f_i - f_i^j)$" on $L^2(\mathscr{F}_i)$, one has $f_i = (f_i^j)^{\mathscr{F}_i} + (f_j)^{\mathscr{F}_i}$, whence by the Cauchy–Shwarz inequality:

$$\mathrm{Sd}(f_i) \leqslant \mathrm{Sd}\big((f_i^j)^{\mathscr{F}_i}\big) + \mathrm{Sd}\big((f_j)^{\mathscr{F}_i}\big). \tag{430}$$

One has trivially $\mathrm{Sd}\big((f_i^j)^{\mathscr{F}_i}\big) \leqslant \mathrm{Sd}(f_i^j)$; on the other hand, $f_j$ is $X_j$-measurable, so $\mathrm{Sd}\big((f_j)^{\mathscr{F}_i}\big) \leqslant \varepsilon_{ij}\,\mathrm{Sd}(f_j)$. In the end, (430) becomes

$$\mathrm{Sd}(f_i) \leqslant \mathrm{Sd}(f_i^j) + \varepsilon_{ij}\,\mathrm{Sd}(f_j), \tag{431}$$

which is (429) for $I' = \emptyset$.

In the case $I' \neq \emptyset$, the same reasoning can be performed, except that one have to work conditionally to $\mathscr{F}_{I'}$. Then, taking $f_i = f_i^{I'}$, $f_j = f_j^{I'}$, $f_i^j = f_i^{I' \uplus \{j\}}$, $\mathscr{F}_i = \mathscr{F}_{I' \uplus \{i\}}$, one gets

$$\mathrm{Sd}(f_i | \mathscr{F}_{I'}) \leqslant \mathrm{Sd}(f_i^j | \mathscr{F}_{I'}) + \varepsilon_{ij}\,\mathrm{Sd}(f_j | \mathscr{F}_{I'}). \tag{432}$$

Now

$$\mathrm{Sd}(f_i) = \sqrt{\int \mathrm{Sd}\big(f_i | \vec{X}_{I'} = \vec{x}_{I'}\big)^2 d\mathbf{P}[\vec{x}_{I'}]}, \tag{433}$$

with similar formulas for $f_j$ and $f_i^j$, since all these functions are centered w.r.t. $\mathscr{F}_{I'}$. Therefore, integrating (432) and applying Minkowski's inequality yields:

$$\mathrm{Sd}(f_i) \leqslant \mathrm{Sd}(f_i^j) + \varepsilon_{ij}\,\mathrm{Sd}(f_j), \tag{434}$$

i.e. (429). □

For $i \leqslant j$, let us denote

$$\Delta_i^{[j]} = \Delta_i^{\{1,\dots,j\} \smallsetminus \{i\}}. \tag{435}$$

Claim 4.2.7 will be used through the following corollary:

**4.2.8 Claim.** *For all $i<j$,*

$$\Delta_i^{[j-1]} \leqslant \frac{1}{1-\varepsilon_{ij}^2}(\Delta_i^{[j]} + \varepsilon_{ij}\Delta_j^<). \tag{436}$$

*Proof.* We have to bound $\Delta_i^{[j-1]}$, which here we rather denote $\Delta_a^{[b-1]}$ to avoid confusion with the notation of Claim 4.2.7. Applying Claim 4.2.7 with $I' = \{1,\dots,b-1\} \smallsetminus \{a\}$, $i=a$ and $j=b$, one has

$$\Delta_a^{[b-1]} = \Delta_a^{\{1,\dots,b-1\}\smallsetminus\{a\}} \leqslant \Delta_a^{\{1,\dots,b\}\smallsetminus\{a\}} + \varepsilon_{ab}\Delta_b^{\{1,\dots,b-1\}\smallsetminus\{a\}} = \Delta_a^{[b]} + \varepsilon_{ab}\Delta_b^{\{1,\dots,b-1\}\smallsetminus\{a\}}. \tag{437}$$

But applying again Claim 4.2.7, this time with $I' = \{1,\dots,b-1\} \smallsetminus \{a\}$, $i=b$ and $j=a$, one has

$$\Delta_b^{\{1,\dots,b-1\}\smallsetminus\{a\}} \leqslant \Delta_b^{\{1,\dots,b-1\}} + \varepsilon_{ab}\Delta_a^{\{1,\dots,b-1\}\smallsetminus\{a\}} = \Delta_b^< + \varepsilon_{ab}\Delta_a^{[b-1]}. \tag{438}$$

Combining (437) and (438) then yields (436). □

Now let us show how Claim 4.2.8 implies the theorem. To avoid heavy formalism, I will detail the computations for $I = \{1,2,3,4\}$ (rather denoted by $I = \{a,b,c,d\}$ here to avoid confusions with "1" and "2" taken as numbers), hoping that generalizing is obvious then.

First, note that

$$\Delta_d^< = \Delta_d^{\neq}. \tag{439}$$

Now, by a direct use of Claim 4.2.8,

$$\Delta_c^< = \Delta_c^{[c]} \leqslant \frac{1}{1-\varepsilon_{cd}^2}(\Delta_c^{[d]} + \varepsilon_{cd}\Delta_d^<) = \tilde{1}_c\Delta_c^{\neq} + \tilde{\varepsilon}_{cd}\Delta_d^{\neq}. \tag{440}$$

To bound $\Delta_b^<$, we have to iterate Claim 4.2.8 twice:

$$\begin{aligned}
\Delta_b^< = \Delta_b^{[b]} &\leqslant \frac{1}{1-\varepsilon_{bc}^2}\Delta_b^{[c]} + \tilde{\varepsilon}_{bc}\Delta_c^< \\
&\leqslant \frac{1}{(1-\varepsilon_{bc}^2)(1-\varepsilon_{bd}^2)}(\Delta_b^{[d]} + \varepsilon_{cd}\Delta_d^<) + \tilde{\varepsilon}_{bc}\Delta_c^< = \tilde{1}_b\Delta_b^{\neq} + \tilde{\varepsilon}_{bd}\Delta_d^{\neq} + \tilde{\varepsilon}_{bc}\Delta_c^< \\
&\underset{(440)}{\leqslant} \tilde{1}_b\Delta_b^{\neq} + \tilde{1}_c\tilde{\varepsilon}_{bc}\Delta_c^{\neq} + (\tilde{\varepsilon}_{bd} + \tilde{\varepsilon}_{bc}\tilde{\varepsilon}_{cd})\Delta_d^{\neq}.
\end{aligned} \tag{441}$$

Last, bounding $\Delta_a^<$ requires iterating Claim 4.2.8 three times:

$$\begin{aligned}
\Delta_a^< = \Delta_a^{[a]} &\leqslant \frac{1}{1-\varepsilon_{ab}^2}\Delta_a^{[b]} + \tilde{\varepsilon}_{ab}\Delta_b^< \\
&\leqslant \frac{1}{(1-\varepsilon_{ab}^2)(1-\varepsilon_{ac}^2)}\Delta_a^{[c]} + \tilde{\varepsilon}_{ac}\Delta_c^< + \tilde{\varepsilon}_{ab}\Delta_b^< \leqslant \tilde{1}_a\Delta^{\neq a} + \tilde{\varepsilon}_{ad}\Delta_d^< + \tilde{\varepsilon}_{ac}\Delta_c^< + \tilde{\varepsilon}_{ab}\Delta_b^< \\
&\underset{(440,441)}{\leqslant} \tilde{1}_a\Delta^{\neq a} + \tilde{1}_b\tilde{\varepsilon}_{ab}\Delta_b^{\neq} + \tilde{1}_c(\tilde{\varepsilon}_{ac} + \tilde{\varepsilon}_{ab}\tilde{\varepsilon}_{bc})\Delta_c^{\neq} + (\tilde{\varepsilon}_{ad} + \tilde{\varepsilon}_{ac}\tilde{\varepsilon}_{cd} + \tilde{\varepsilon}_{ab}\tilde{\varepsilon}_{bd} + \tilde{\varepsilon}_{ab}\tilde{\varepsilon}_{bc}\tilde{\varepsilon}_{cd})\Delta_d^{\neq}.
\end{aligned} \tag{442}$$

One can sum up Equations (439)–(442) into the matricial expression

$$\begin{pmatrix}\Delta_a^<\\ \Delta_b^<\\ \Delta_c^<\\ \Delta_d^<\end{pmatrix} \leqslant \begin{pmatrix}1 & \tilde\varepsilon_{ab} & \tilde\varepsilon_{ac}+\tilde\varepsilon_{ab}\tilde\varepsilon_{bc} & \tilde\varepsilon_{ad}+\tilde\varepsilon_{ac}\tilde\varepsilon_{cd}+\tilde\varepsilon_{ab}\tilde\varepsilon_{bd}+\tilde\varepsilon_{ab}\tilde\varepsilon_{bc}\tilde\varepsilon_{cd}\\ 0 & 1 & \tilde\varepsilon_{bc} & \tilde\varepsilon_{bd}+\tilde\varepsilon_{bc}\tilde\varepsilon_{cd}\\ 0 & 0 & 1 & \tilde\varepsilon_{cd}\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\tilde 1_a\Delta_a^{\neq}\\ \tilde 1_b\Delta_b^{\neq}\\ \tilde 1_c\Delta_c^{\neq}\\ \Delta_d^{\neq}\end{pmatrix}. \tag{443}$$

If we look back at how the square matrix in (443) has been constructed, we find that

$$\begin{pmatrix}\textit{square}\\ \textit{matrix}\\ \textit{in}\\ \textit{(443)}\end{pmatrix} = \sum_{k=0}^{\infty}\begin{pmatrix}0 & \tilde\varepsilon_{ab} & \tilde\varepsilon_{ac} & \tilde\varepsilon_{ad}\\ 0 & 0 & \tilde\varepsilon_{bc} & \tilde\varepsilon_{bd}\\ 0 & 0 & 0 & \tilde\varepsilon_{cd}\\ 0 & 0 & 0 & 0\end{pmatrix}^k = \begin{pmatrix}1 & -\tilde\varepsilon_{ab} & -\tilde\varepsilon_{ac} & -\tilde\varepsilon_{ad}\\ 0 & 1 & -\tilde\varepsilon_{bc} & -\tilde\varepsilon_{bd}\\ 0 & 0 & 1 & -\tilde\varepsilon_{cd}\\ 0 & 0 & 0 & 1\end{pmatrix}^{-1}, \tag{444}$$

so in the end we obtain that

$$\begin{pmatrix}\Delta_a^<\\ \Delta_b^<\\ \Delta_c^<\\ \Delta_d^<\end{pmatrix} \leqslant M\begin{pmatrix}\Delta_a^{\neq}\\ \Delta_b^{\neq}\\ \Delta_c^{\neq}\\ \Delta_d^{\neq}\end{pmatrix}, \tag{445}$$

where $M$ is given by (423). Then it is immediate that $\mathrm{Var}(f) = \sum_i (\Delta_i^<)^2 \leqslant \|\!|M|\!\|^2 \sum_i (\Delta_i^{\neq})^2 = \|\!|M|\!\|^2 \mathscr{E}(f,f)$, QED. □

The bound we have obtained for the spectral gap is not symmetric by permutation of the indexes in $I$. It can however can be bounded by a simpler expression, which is nearly as good as the original one in concrete situations:

**4.2.9 Corollary.** *In Theorem 4.2.5, M can be replaced by the matrix*

$$M' = \begin{pmatrix}1 & -\varepsilon_{12} & \cdots & -\varepsilon_{1N}\\ -\varepsilon_{12} & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & -\varepsilon_{(N-1)N}\\ -\varepsilon_{1N} & \cdots & -\varepsilon_{(N-1)N} & 1\end{pmatrix}^{-1}, \tag{446}$$

*provided* $\rho(\mathbf{I}_N - M') < 1$.

*Proof.* Each entry of $M$ is actually bounded by the corresponding entry of $M'$. To see it, we 'expand' the entries of $M$, resp. $M'$. First, notice that $1/(1-\varepsilon_{ij}^2)$ can be expanded into $1+\varepsilon_{ij}\varepsilon_{ji}+\varepsilon_{ij}\varepsilon_{ji}\varepsilon_{ij}\varepsilon_{ji}+\cdots$, so that one has the expansions

$$\tilde 1_i = \sum_{i<j_1\leqslant\cdots\leqslant j_k}\prod_{l=1}^{k}\varepsilon_{ij_l}\varepsilon_{j_l i} \tag{447}$$

and

$$\tilde\varepsilon_{ij} = \sum_{i<j_1\leqslant\cdots\leqslant j_k\leqslant j}\Big(\prod_{l=1}^{k}\varepsilon_{ij_l}\varepsilon_{j_l i}\Big)\varepsilon_{ij}. \tag{448}$$

Then, using the inversion formula $(\mathrm{I}-A)^{-1} = \sum_{k=0}^{\infty} A^k$ for triangular arrays, one obtains that

$$M_{ij} = \sum_{\substack{(i_0,i_1,\ldots,i_k)\\ \textit{first condition}}}\prod_{l=0}^{k-1}\varepsilon_{i_l i_{l+1}}, \tag{449}$$

where the meaning of "*first condition*" is given by the following

**4.2.10 Definition.** A sequence $(i_0,\dots,i_k)$ is said to satisfy the *first condition* if:

(i) $i_0 = i$ and $i_k = j$;
(ii) $i_l \neq i_{l+1}$ for all $l$;
(iii) $i_{l+1} < i_l$ only if $l \geqslant 1$ and $i_{l+1} = i_{l-1}$;
(iv) If $i_{l+1} < i_l$ and $l \leqslant k-2$, then $i_{l+2} \geqslant i_l$.

One has a similar formula for $M'$:

$$M'_{ij} = \sum_{\substack{(i_0,i_1,\dots,i_k)\\ \textit{second condition}}} \prod_{l=0}^{k-1} \varepsilon_{i_l i_{l+1}}, \tag{450}$$

where

**4.2.11 Definition.** A sequence $(i_0,\dots,i_k)$ is said to satisfy the *second condition* if it satisfies Conditions (i) and (ii) of Definition 4.2.10.

Since the *second condition* is obviously weaker than the *first condition*, one has $M_{ij} \leqslant M'_{ij}$. □

There is a still weaker but even simpler formula:

**4.2.12 Corollary.** *Defining*

$$\begin{array}{rccc} \boldsymbol{\varepsilon}: & L^2(I) & \to & L^2(I) \\ & (a_i)_{i\in I} & \mapsto & (\sum_{j\neq i} \varepsilon_{ij} a_j)_{i\in I}, \end{array} \tag{451}$$

*the spectral gap of the Glauber dynamics is at least*

$$(1 - |||\boldsymbol{\varepsilon}|||)_+^2. \tag{452}$$

*Proof.* One has $M' = (\mathrm{I} - \boldsymbol{\varepsilon})^{-1}$, so, provided $|||A||| < 1$,

$$|||M'||| = |||(\mathrm{I}-A)^{-1}||| = \Big|\Big|\Big| \sum_{k=0}^{\infty} A^k \Big|\Big|\Big| \leqslant \sum_{k=0}^{\infty} |||A|||^k = (1 - |||A|||)^{-1}. \tag{453}$$

In the case $|||A||| \geqslant 1$, (452) is trivial. □

## 4.2.c Avoiding the articial phase transition

A common situation in which we would like to apply the previous results is when $I = \mathbb{Z}^n$ and $\varepsilon_{ij}$ is of the form $\varepsilon(j-i)$ for some symmetric function $\varepsilon\colon\ \mathbb{Z}^n \to [0,1]$. Then Corollary 4.2.12 tells that the Glauber dynamics has a (strictly) positive spectral gap as soon as $\sum_{z\neq 0} \varepsilon(z) < 1$. But like in § 3.6.c, we are going to prove that that bound is somehow 'artificial' and that it can be relaxed into the neater condition "$\sum_{z\neq 0} \varepsilon(z) < \infty$":

**4.2.13 Theorem.** *Suppose that $I = \mathbb{Z}^n$ and that for all $i,j \in \mathbb{Z}^n$ one has $\{X_i : X_j\}_* \leqslant \varepsilon(j-i)$ for some symmetric function $\varepsilon\colon\ \mathbb{Z}^n \to [0,1]$ such that $\varepsilon(z) < 1$ as soon as $z \neq 0$. Then if $\sum_{z\in\mathbb{Z}^n} \varepsilon(z) < \infty$, the spectral gap of the Glauber dynamics is positive.*

*Proof.* The assumption on $\sum_z \varepsilon(z)$ allows us to take $l<\infty$ large enough so that

$$\sum_{z\in l\mathbb{Z}^n\smallsetminus\{0\}} \varepsilon(z)<1. \tag{454}$$

We split $\mathbb{Z}^n$ into a partition of $l^n =: N$ sublattices $Z_1,\dots,Z_N$, each lattice $Z_u$ being of the form $l\mathbb{Z}^n+z_u$ for some $z_u\in\mathbb{Z}^n/l\mathbb{Z}^n$. Then we define an auxiliary dynamics:

**4.2.14 Definition.** The *sublattice Glauber dynamics* is the Glauber dynamics for $\vec{X}_{\mathbb{Z}^n}$ *considered as the finite-dimensional vector* $(\vec{X}_{Z_1},\dots,\vec{X}_{Z_N})$. In other words, on each $u\in\{1,\dots,N\}$ there is an independent *Poisson*(1) alarm clock, and when clock $u$ rings, the state of the whole $\vec{X}_{Z_u}$ is flipped in one shot according to $\mathbf{P}(X_{Z_u}|\vec{X}_{\mathbb{Z}^n\smallsetminus Z_u})$.

Now let $f\in\bar{L}^2(\Omega)$. In addition to the notation of the proof of Theorem 4.2.5, we introduce the following definition:

**4.2.15 Definition.** For $u\in\{1,\dots,N\}$, we define

$$f^{\neq}_{(u)} := f-\mathbf{E}[f|\vec{X}_{\mathbb{Z}^n\smallsetminus Z_u}]. \tag{455}$$

*4.2.16 Remark.* The $f^{\neq}_{(u)}$ are the equivalent of the $f^{\neq}_i$ for the sublattice Glauber dynamics.

Fixing some 'boundary condition' $\vec{x}_{\mathbb{Z}^n\smallsetminus Z_u}$ on $\mathbb{Z}^n\smallsetminus Z_u$, we can apply Corollary 4.2.12 to the Glauber dynamics for $\vec{X}_{Z_u}$ under the law $\mathbf{P}[\cdot|\vec{X}_{\mathbb{Z}^n\smallsetminus Z_u}=\vec{x}_{\mathbb{Z}^n\smallsetminus Z_u}]$. After integrating, one gets that

$$\operatorname{Var}(f^{\neq}_{(u)})\leqslant(1-|||\boldsymbol{\zeta}|||)^{-2}\sum_{i\in Z_u}\operatorname{Var}(f^{\neq}_i), \tag{456}$$

where $\boldsymbol{\zeta}$ is the operator on $L^2(l\mathbb{Z}^n)$ defined by

$$(\boldsymbol{\zeta}g)(i)=\sum_{z\in l\mathbb{Z}^n\smallsetminus\{0\}}\varepsilon(z)g(i+z), \tag{457}$$

whose norm is obviously bounded by $\sum_{z\in l\mathbb{Z}^n\smallsetminus\{0\}} =: \zeta<1$. Then, summing (456) for all $u$:

$$\sum_{u=1}^N \operatorname{Var}(f^{\neq}_{(u)})\leqslant(1-\zeta)^{-2}\mathscr{E}(f,f). \tag{458}$$

Now, let us apply Theorem 4.2.5 to the sublattice Glauber dynamics [Definition 4.2.14]. It yields that

$$\operatorname{Var}(f)\leqslant|||M|||^2\sum_{u=1}^N\operatorname{Var}(f^{\neq}_{(u)}), \tag{459}$$

where $M$ is some $(N\times N)$ matrix depending on the $\{\vec{X}_{Z_u}:\vec{X}_{Z_v}\}_*$. But by Theorem 3.6.8, $\{\vec{X}_{Z_u}:\vec{X}_{Z_v}\}_*<1$ for all $u\neq v$, thus $|||M|||<\infty$ by Remark 4.2.6. Combining (458) and (459), we finally get that the spectral gap of the Glauber dynamics for $\vec{X}_{\mathbb{Z}^n}$ is bounded below by $|||M|||^{-2}\times(1-\zeta)^2>0$. □

*4.2.17 Remark.* Like Theorem 3.6.8, Theorem 4.2.13 could actually be stated in the general case of 'abstract' metric spaces on which some group acts profinitely.

# Chapter 5

# Concrete examples

It is now time to see what the results of Chapters 3 and 4 yield for concrete models of statistical physics. I will try to give rather different types of examples, so as to illustrate the advantages of working with maximal correlations: this frame is indeed quite general, as it requires little structure on the models considered.

In § 5.1 we will look back at Ising's model, seeing how tensorization of maximal decorrelations improves the results of § 0.1, and what other results are given by the theorems of § 4. We will also consider two kinds of generalizations, namely when the range of interactions becomes infinite and when the strength of the interactions is random [*spin glasses*]. In the two next sections we will look at models with continuous states spaces: first a quite general class of linear models [§ 5.2], then a family of nonlinear models [§ 5.3]. Finally in § 5.4 we will see how one can consider time as a supplementary dimension of the system to get contractivity results for non-reversible Markov chains [*hypocoercivity*] on an infinite system of particles.

☛ *In this chapter, all the probability systems considered will be endowed with their natural $\sigma$-metalgebras, cf. Definition 3.1.16. To alleviate notation, I will give no names to these $\sigma$-metalgebras, but will plainly write "$\{X:Y\}_*$" to mean "the subjective decorrelation between $X$ and $Y$ seen from the natural $\sigma$-metalgebra of the underlying system".*

## 5.1 Back to Ising's model

### 5.1.a Standard Ising's model

In all this section, we work on the lattice $\mathbb{Z}^n$ equipped with its natural distance *dist*; accordingly $|\cdot|$ will denote the $l^1$ norm on $\mathbb{R}^n$. Recall the definition of Ising's model and the related notation that we introduced in § 0.1, and Theorem 0.1.7 on the existence of a completely analytical regime.

The following theroem states that Ising's model in completely analytical regime is $\rho$-mixing, i.e. that two distant bunches of spins are little correlated in the sense of maximal correlation:

**5.1.1 Theorem.** *For Ising's model on $\mathbb{Z}^n$ in the completely analytical regime,* *(i) There exists some $\psi' > 0$ (the same as in Theorem 0.1.7) such that for all disjoint $I, J \subset \mathbb{Z}^n$, one has when dist$(I,J) \to \infty$ that*

$$\{\vec{\omega}_I : \vec{\omega}_J\} \leqslant \exp\left[-(\psi' + o(1))\, dist(I,J)\right], \tag{460}$$

*where the "$o(1)$" can be easily computed as an explicit function of dist$(I,J)$, $n$, $T$, $\psi'$ and the $C'$ appearing in Theorem 0.1.7.*

*(ii) There exists some $k < 1$ such that for all disjoint $I, J \subset \mathbb{Z}^n$,*

$$\{\vec{\omega}_I : \vec{\omega}_J\} \leqslant k. \tag{461}$$

*(iii) Points (i) and (ii) remain valid uniformly under any law of the form $\mathbf{P}[\cdot|\vec{\omega}_K = \vec{\hat{\omega}}_K]$, for $K \subset \mathbb{Z}^n$ and $\vec{\hat{\omega}}_K \in \{\pm 1\}^K$ a 'boundary condition' on $K$.*

*5.1.2 Remark.* Let us compare Theorem 5.1.1 with Theorem 0.1.7. Both theorems state decorrelation between distant bunches of spins above temperature $T'_{\mathrm{c}}$; the difference relies in using maximal correlations rather than $\beta$-mixing to quantify dependence between the bunches in Theorem 5.1.1.

Both results give an exponential decay of correlations, with the same exponential constant $\psi'$, but Theorem 5.1.1 is more powerful in the sense that the bound (460) is uniform in the size of $I$ and $J$ while (5) was not. Moreover, thanks to Point (ii) we get a non-trivial result for any choice of disjoint $I$ and $J$, which was not the case beforehand. Recall that the drawbacks of Theorem 0.1.7 were inherent to $\beta$-mixing, as Theorem 0.1.8 shew.

Both result remain valid under conditioning. However, if one takes a *random* boundary condition—that is, if one works under the law $\mathbf{P}[\cdot|\vec{\omega}_K \in C]$ for some non-singleton $C \subset \{\pm 1\}^K$ —, then Point (iii) of Theorem 5.1.1 fails (cf. Remark 3.1.4), while (5) is still valid by convexity of the total variation norm.

*5.1.3 Remark.* Let us compare Theorem 5.1.1 with Theorem 0.1.9. The result of Theorem 0.1.9 can be rewritten:

$$\{\vec{\omega}_{\{0\}\times\mathbb{Z}^{n-1}} : \vec{\omega}_{\{x\}\times\mathbb{Z}^{n-1}}\} \leqslant e^{-\psi x}. \tag{462}$$

Theorem 5.1.1 can be seen as a generalization of that result to the case where $I$ and $J$ have arbitrary shapes.[*] Moreover, Point (iii) also gives the existence of a conditional version, which we did not have before.

There is however a price to pay for this greater generality, since we had to require complete analyticity rather than just weak mixing, which can be really more restrictive in some cases (cf. Footnote [$*$] on page 12).

*5.1.4 Remark.* Continuing the previous remark, a natural open question is whether one can tensorize maximal decorrelation under assumptions of weak mixing type. In the case of Ising's model at least, I expect $\rho$-mixing to remain true—even for arbitrary shapes—as soon as $T > T_{\mathrm{c}}$, because on the one hand Theorem 0.1.9 proves $\rho$-mixing

[*]Note that in the case $I$ and $J$ are hyperplanes, we shew on page 26 that (460) could be improved into

$$\{\vec{\omega}_{\{0\}\times\mathbb{Z}^{n-1}} : \vec{\omega}_{\{x\}\times\mathbb{Z}^{n-1}}\} \leqslant e^{-\psi' x}. \tag{463}$$

between parallel hyperplanes, while on the other hand $\rho$-mixing seems to hold also in the 'opposite extreme case' when $I$ and $J$ make a check pattern.

By the way, it is likely that the natural condition should not be weak mixing itself but rather something like *strong mixing for cubes* (often called merely *strong mixing*[†], which means that when a boundary condition is fixed outside a cube of arbitrary edge, changing one spin on the boundary has an effect in total variation which decreases exponentially with the distance to the spin changed. In fact it has been proved [31] that in dimension 2, weak mixing is equivalent to strong mixing.

*Proof of Theorem 5.1.1.* Theorem 5.1.1 will be a direct consequence of the work of Chapter 3 as soon as we show that, denoting by $*$ the natural $\sigma$-metalgebra of the system (i.e. the $\sigma$-metalgebra generated by the $\omega_i$), for all distinct $i, j \in \mathbb{Z}^n$, one has

$$\{\omega_i : \omega_j\}_* \leqslant c_0 C' e^{-\psi' dist(i,j)} \wedge k_0 \tag{464}$$

for some explicit $c_0 < \infty$ and $k_0 < 1$ only depending on $n$ and $T$. Then indeed, Proposition 3.6.6 yields

$$\begin{aligned}\{\vec{\omega}_I : \vec{\omega}_J\} &\leqslant \sum_{\substack{\delta\in\mathbb{Z}^n\\|\delta|\geqslant dist(I,J)}} c_0 C' e^{-\psi'|\delta|} = c_0 C' \sum_{d=dist(I,J)}^{\infty} \#\{\delta\in\mathbb{Z}^n:\ |\delta| = d\} e^{-\psi' d} \\ &\overset{dist(I,J)\to\infty}{\sim} c_0 C' \sum_{d=dist(I,J)}^{\infty} \frac{2^n d^{n-1}}{(n-1)!} e^{-\psi' d} \sim \frac{c_0 C' 2^n}{(n-1)!} dist(I,J)^{n-1} e^{-\psi' dist(I,J)} \\ &= e^{-(\psi'+o(1)) dist(I,J)},\end{aligned} \tag{465}$$

whence Point (i). Moreover, since

$$\sum_{\substack{\delta\in\mathbb{Z}^n\smallsetminus\{0\}\\|\delta|\leqslant dist(I,J)}} c_0 C' e^{-\psi' dist(i,j)} < \infty, \tag{466}$$

Point (ii) follows from Lemma 3.6.8, and finally (iii) is a consequence of § 3.4.b about subjective results.

So, we have to prove (464). Let $\vec{\hat{\omega}}_K \in \{\pm1\}^K$, $K \subset \mathbb{Z}^n$, be some arbitrary boundary condition, and denote by $\mathbf{P}_{\text{con}}$ the associated law, that is, $\mathbf{P}_{\text{con}} = \mathbf{P}[\cdot|\vec{\omega}_K = \vec{\hat{\omega}}_K]$; our goal is to show that under $\mathbf{P}_{\text{con}}$, for all distinct $i, j \in \mathbb{Z}^n$, one has $\{\omega_i : \omega_j\} \leqslant c_0 C' e^{-\psi' dist(i,j)} \wedge k_0$.

The result is immediate if $i \in K$, resp. $j \in K$ (since then $\omega_i$, resp. $\omega_j$, is constant and thus independent of everything), so we assume $i, j \notin K$. We begin with observing that if $K$ is the set $N(i)$ of all the neighbours of $i$, equilibrium at $i$ implies that, whatever the boundary condition may be:

$$\mathbf{P}_{\text{con}}[\omega_i = -1], \mathbf{P}_{\text{con}}[\omega_i = +1] \geqslant (e^{4n/T} + 1)^{-1} \tag{467}$$

—the extremal cases being when $\vec{\hat{\omega}}_{N(i)} \equiv +1$, resp. $\vec{\hat{\omega}}_{N(i)} \equiv -1$. Now in the general case $K \subset \mathbb{Z}^n \smallsetminus \{i\}$, $Law_{\text{con}}[\omega_i]$ is an average of laws of the form $Law(\omega_i|\vec{\omega}_{N(i)} = \vec{\hat{\omega}}_{N(i)})$, so that (467) remains valid. Similarly, equilibrium on $\{i, j\}$ gives that for all $a, b \in \{\pm1\}$,

$$\mathbf{P}_{\text{con}}[\omega_i = a \text{ and } \omega_j = b] \geqslant (e^{8n/T} + 2e^{(4n+2)/T} + 1)^{-1}. \tag{468}$$

[†]Strong mixing *stricto sensu* is actually the same as complete analyticity, so that mathematicians have got used to undermeaning "for cubes"—but strong mixing for cubes is strictly weaker than complete analyticity! [30, § 2].

Now, recall that the correlation level between two two-ranged variables can be computed by Formula (57), where $|p_a^b - p_a p^b|$ is also $\beta(X,Y)/2$. Thus the bound "$\{\omega_i : \omega_j\} \leqslant C_0 e^{-\psi' dist(i,j)}$" is a direct consequence of Theorem 0.1.7, with

$$c_0 = \frac{1/2}{(e^{4n/T}+1)^{-1}(1-(e^{4n/T}+1)^{-1})} = \tanh(4n/T)+1. \tag{469}$$

It remains to prove the bound "$\{\omega_i : \omega_j\} \leqslant k_0$". We will use the following corollary of (57):

**5.1.5 Lemma.** *With the notation of Remark 1.2.2, there exists $a,b$ in the respective ranges of $X,Y$ such that*

$$\{X:Y\} \leqslant 1-4p_a^b. \tag{470}$$

*Proof of Lemma 5.1.5.* The difference $p_a^b - p_a p^b$ gets its sign changed whenever $a$, resp. $b$, changes, so there are some $a$ and $b$ for which this value is nonpositive; moreover, denoting by $\{a,a'\}$ and $\{b,b'\}$ the respective ranges of $X$ and $Y$, $p_{a'}^{b'} - p_{a'} p^{b'}$ is also nonpositive. Now one has

$$\frac{p_a p^b}{\sqrt{p_a p_{a'} p^b p^{b'}}} \times \frac{p_{a'} p^{b'}}{\sqrt{p_a p_{a'} p^b p^{b'}}} = 1, \tag{471}$$

so that either $p_a p^b$ or $p_{a'} p^{b'}$ is $\leqslant \sqrt{p_a p_{a'} p^b p^{b'}}$. Up to changing notation we can assume that it is $p_a p^b$, and then

$$\{X:Y\} = \frac{|p_a^b - p_a p^b|}{\sqrt{p_a p_{a'} p^b p^{b'}}} = \frac{p_a p^b - p_a^b}{\sqrt{p_a p_{a'} p^b p^{b'}}} \leqslant 1 - \frac{p_a^b}{\sqrt{p_a p_{a'} p^b p^{b'}}} \leqslant 1-4p_a^b. \tag{472}$$

□

Combining Lemma 5.1.5 with (468), we then get the desired bound, with

$$k_0 = 1-4(e^{8n/T}+2e^{(4n+2)/T}+1)^{-1} < 1. \tag{473}$$

□

Formula (464) is also what we need to apply the results of Chapter 4. Indeed, denoting $\varepsilon(z) := \{X_i : X_{i+z}\}_*$, it gives that $\sum_{z\in\mathbb{Z}^n} \varepsilon(z) < \infty$ with $\varepsilon(z) < 1$ as soon as $z \neq 0$, so that Theorems 4.1.8 and 4.2.13 yield respectively:

**5.1.6 Theorem.** *In completely analytical regime, the spins Ising's model satisfies the central limit theorem, in the sense that the conclusions of Theorem 4.1.8 hold for them.*

**5.1.7 Theorem.** *In completely analytical regime, the Glauber dynamics for Ising's model has a (strictly) positive spectral gap, and this remains valid uniformly if one fixes a 'boundary condition' on the spins of some $K \subset \mathbb{Z}^n$.*

*5.1.8 Remark.* As I told in Chapter 4, results of these kinds have already been studied by other methods (see e.g. [6, 14] for the CLT and [29] for the spectral gap). For the standard Ising model in completely analytical regime, which is "very nice", these previous works apply well, so the two theorems above are not new. They are interesting however because of the new method used to prove them, which is quite direct and likely to apply to a broader class of models. Such models will be presented in the sequel of this chapter.

## 5.1.b Generalizations of Ising's model

The previous results can be adapted to several kinds of generalizations of Ising's model. Let us expose some of them.

### Long-range Ising models

A physically important case is the *long-range Ising models* on $\mathbb{Z}^n$. In these models, the states space is unchanged, but the Hamiltonian $H$ becomes

$$H(\vec{\omega}) = -\frac{1}{2}\sum_{i\neq j} J(j-i)\omega_i\omega_j, \tag{474}$$

where $J\colon\ \mathbb{Z}^n \smallsetminus \{0\} \to \mathbb{R}$ is some symmetric function with non-compact support such that $J(z) \overset{|z|\to\infty}{=} O(|z|^{-(n+\alpha)})$ for some $\alpha > 0$.

Let us state a decorrelation result for this class of models. The frame of the proof of the following proposition will work as well for the other generalizations of Ising's model.

**5.1.9 Proposition.** *There exists an temperature $T_1 < \infty$ such that, provided $T \geqslant T_1$:*

*(i) Equilibrium for the long-range Ising model is unique;*

*(ii) Uniformly in $i,j$, $\{\omega_i : \omega_j\}_* \overset{|j-i|\to\infty}{=} O(|j-i|^{-(n-\alpha)})$;*

*(iii) There exists some $k_0 < 1$ such that for all $i \neq j$, $\{\omega_i : \omega_j\}_* \leqslant k_0$.*

*Proof.* The principle of the proof consists in coupling two Glauber dynamics with different initial conditions. Recall that the Glauber dynamics is defined as follows: each spin has an independent clock ringing with rate 1, and when the clock of a spin rings, this spin is flipped so that its final state is drawn according to its equilibrium measure conditionnally to the state of all other spins. Namely, if the clock of spin $i$ rings at time $t$, denoting as usual $\beta = T^{-1}$,

$$\mathbf{P}[\omega_i(t+) = +1] = \frac{\exp\big(\beta\sum_{j\neq i} J(j-i)\omega_j(t)\big)}{2\cosh\big(\beta\sum_{j\neq i} J(j-i)\omega_j(t)\big)} \tag{475}$$

and $\mathbf{P}[\omega_i(t+) = -1] = 1 - \mathbf{P}[\omega_i(t+) = +1]$.

To couple the Glauber dynamics, we will assume that, rather than just "ringing" at time $t$, the clock of $i$ is a Poisson process on $\mathbb{R}_+ \times (0,1)$, points of which are denoted by $(t,y)$. Then, if at time $t$ the clock of spin $i$ has a point $(t,y)$, spin $i$ flips to $+1$ if $y < \mathbf{P}[\omega_i(t+) = +1]$, resp. to $-1$ if $y \geqslant \mathbf{P}[\omega_i(t+) = +1]$.

Now, consider two Glauber dynamics $\vec{\omega}^-$ and $\vec{\omega}^+$ having the same Poisson process, but starting with different initial conditions. It will be convenient[‡] to assume that $\vec{\omega}^-(t=0) \leqslant \vec{\omega}^+(t=0)$ almost-surely: then, as we will see, for the coupled dynamics one has (a.s.) $\vec{\omega}^-(t) \leqslant \vec{\omega}^+(t)\ \forall t$. At time $t$, denote by $\Theta(t)$ the set of points where $\vec{\omega}^-$ and $\vec{\omega}^+$ differ:

$$\Theta(t) = \{i \in \mathbb{Z}^n :\ \big(\omega_i^-(t), \omega_i^+(t)\big) = (-1,+1)\}. \tag{476}$$

[‡]In the cases where interactions can be antiferromagnetic ($J < 0$), monotonicity does not stand any more; the proof however remains valid with a heavier formalism, replacing ">" by "$\neq$" and putting absolute values at the right places.

When the clock at spin $i$ rings at time $t$, three cases have to be distinguished:

1. If $y < \exp(\beta\sum_{j\neq i} J(j-i)\omega_j^-(t))/2\cosh(\beta\sum_{j\neq i} J(j-i)\omega_j^-(t))$, then both $\omega_i^+$ and $\omega_i^-$ flip into state $+1$;
2. If $y > \exp(\beta\sum_{j\neq i} J(j-i)\omega_j^+(t))/2\cosh(\beta\sum_{j\neq i} J(j-i)\omega_j^+(t))$, then both $\omega_i^+$ and $\omega_i^-$ flip into state $-1$;
3. If $\frac{\exp(\beta\sum_{j\neq i} J(j-i)\omega_j^-(t))}{2\cosh(\beta\sum_{j\neq i} J(j-i)\omega_j^-(t))} < y < \frac{\exp(\beta\sum_{j\neq i} J(j-i)\omega_j^+(t))}{2\cosh(\beta\sum_{j\neq i} J(j-i)\omega_j^+(t))}$, then $\omega_i^+$ flips into state $+1$ while $\omega_i^-$ flips into state $-1$.

Denoting

$$\mathscr{J} := \sum_{z\in\mathbb{Z}^n\smallsetminus\{0\}} J(z), \tag{477}$$

which is always finite by the assumption on $J$, the probability of each of the two first cases is bounded below by $e^{-\beta\mathscr{J}}/2\cosh(\beta\mathscr{J})$. The probability of the third case is

$$\frac{\sinh(2\beta\sum_{j\in\Theta(t)\smallsetminus\{i\}} J(j-i))}{2\cosh(\beta\sum_{j\neq i} J(j-i)\omega_j^-(t))\cosh(\beta\sum_{j\neq i} J(j-i)\omega_j^+(t))}, \tag{478}$$

which is bounded above by $\beta\sum_{j\in\Theta(t)\smallsetminus\{i\}} J(j-i)$ thanks to the following computational

**5.1.10 Lemma.** *For $a \leqslant b$ two real numbers,*

$$\sinh(b-a) \leqslant (b-a)\cosh a\cosh b. \tag{479}$$

*Proof.* Making the change of variables $x = (a+b)/2, t = (b-a)/2$, we have to prove that for $x\in\mathbb{R}, t\geqslant 0$, one has:

$$\sinh(2t) \leqslant 2t\cosh(x-t)\cosh(x+t). \tag{480}$$

If we consider the right-hand side of (480) as a function of $x$, it is symmetric (since cosh is symmetric) and its logarithm is convex (since $\log\circ\cosh$ is convex, its derivative being the increasing function tanh), so its minimum is attained for $x = 0$; thus it suffices to prove (480) in that case, i.e. to prove that $\sinh(2t) \leqslant 2t\cosh^2 t$ for all $t \geqslant 0$. But $\sinh(2t) = 2\sinh t\cosh t$, so we can simplify both sides by $2\cosh t$, and then it suffices to prove that $\sinh t \leqslant t\cosh t$, which is true since $\tanh t \leqslant t$ for all $t \geqslant 0$. □

Thanks to these estimates, we can define a process Markovian $\Theta^*(t)$ on $\mathfrak{P}(\mathbb{Z}^n)$ such that almost-surely, $\Theta^*(t) \supset \Theta(t)\ \forall t$. This process has the following law:

**5.1.11 Definition.** The law of $\Theta^*$ is defined thanks to independent Poissonian clocks indexed by $(\mathbb{Z}^n)^2$. For $i \neq j$ the clock $(i,j)$ has rate $\beta J(j-i)$, while the clock $(i,i)$ has rate $e^{-\beta\mathscr{J}}/\cosh(\beta\mathscr{J})$. At $t = 0$ one has $\Theta^*(0) = \Theta(0)$. If at time $t$ the clock $(i,j)$ rings, with $j \neq i$, then:

- Either $i \in \Theta^*(t-)$ and then $\Theta^*$ changes so that $\Theta^*(t+) = \Theta^*(t-)\cup\{j\}$[§];
- Or $i \notin \Theta^*(t-)$ and then $\Theta^*$ does not change.

On the other hand, if at time $t$ the clock $(i,i)$ rings, then $\Theta^*$ changes so that $\Theta^*(t+) = \Theta^*(t-)\smallsetminus\{i\}$.

[§]Of course, if $j\in\Theta^*(t-)$ then $\Theta^*$ does actually not change.

Let $\lambda := \beta\mathscr{J} - (e^{-\beta\mathscr{J}}/\cosh(\beta\mathscr{J}))$. If we take $\mathbf{E}[\#\Theta(t=0)] < \infty$[¶], it is immediate that $\#\Theta^*(t)/e^{\lambda t}$ is a supermartingale. So, provided $T$ is large enough so that $\lambda < 0$, i.e.

$$\beta\mathscr{J} < \frac{e^{-\beta\mathscr{J}}}{\cosh(\beta\mathscr{J})}, \tag{481}$$

the two processes $\vec{\omega}^-(t)$ and $\vec{\omega}^+(t)$ tend to be equal when $t \to \infty$; in particular they have the same equilibrium. That proves Point (i) of the Lemma, since any initial condition stands between the 'extreme' conditions $\vec{\omega}^-(t=0) \equiv -1$ and $\vec{\omega}^+(t=0) \equiv +1$.

Observe that the previous reasoning remains entirely valid if one reasons conditionally to some boundary condition of the form "$\vec{\omega}_K = \vec{\hat{\omega}}_K$", with the same condition on $T$.

Now we are turning to the correlation between two distant spins. Let $i \in \mathbb{Z}^n$ and let $\vec{\hat{\omega}}_K$ be some boundary condition on some $K \subset \mathbb{Z}^n \setminus \{i\}$. Suppose $T$ satisfies (481); I want to compare the Glauber dynamics corresponding to the boundary condition "$\vec{\omega}_{K\uplus\{i\}} = (\vec{\hat{\omega}}_K, (+1)^{\{i\}})$"—where $(\vec{\hat{\omega}}_K, (+1)^{\{i\}})$ stands for the function on $K \uplus \{i\}$ which is equal to $\hat{\omega}$ on $K$ and to $+1$ at $i$ —with the Glauber dynamics corresponding to the boundary condition "$\vec{\omega}_{K\uplus\{i\}} = (\vec{\hat{\omega}}_K, (-1)^{\{i\}})$". In this frame, one defines the process $\Theta^*$ as previously, except that one imposes that $\Theta^*(t) \cap K = 0$ and $i \in \Theta^*(t)$ for all $t$. This time, it is the equilibrium behaviour of $\Theta^*$ which interests us. Denote by $\mathbf{P}_{\mathrm{eq}}$ the equilibrium law of $\Theta^*$; for $j' \in \mathbb{Z}^n \setminus K$, denote $\theta(j) := \mathbf{P}_{\mathrm{eq}}[j \in \Theta^*]$. Then $\theta$ satisfies the following discrete subelliptic equation with Dirichlet boundary conditions:

$$\begin{cases} \forall j \notin K \uplus \{i\} & \frac{e^{-\beta\mathscr{J}}}{\cosh(\beta\mathscr{J})}\theta(j) \leqslant \beta \sum_{\substack{i' \in \mathbb{Z}^n \setminus K \\ i' \neq j}} J(i'-j)\theta(i'); \\ \forall k \in K & \theta(k) = 0; \qquad \theta(i) = 1. \end{cases} \tag{482}$$

Define the convolution kernel $a$ on $\mathbb{Z}^n$ by

$$\begin{cases} a(0) = 1; \\ \forall z \neq 0 \quad a(z) = -\frac{\cosh(\beta\mathscr{J})}{e^{-\beta\mathscr{J}}}\beta J(z), \end{cases} \tag{483}$$

so that (482) writes in the bulk:

$$a * \theta \leqslant 0. \tag{484}$$

Writing $a =: \delta_0 - \tilde{a}$, Condition (481) ensures that $\|\tilde{a}\|_{l^1} < 1$. Since $l^1(\mathbb{Z}^n)$ is a Banach algebra for the convolution operator $*$, with neutral element $\delta_0$, it follows that $a$ is invertible with inverse

$$a^{-*} = \delta_0 + \tilde{a} + \tilde{a} * \tilde{a} + \tilde{a} * \tilde{a} * \tilde{a} + \cdots. \tag{485}$$

Since $\tilde{a} \geqslant 0$, $a^{-*}$ is nonnegative everywhere with $a^{-*}(0) > 0$. Therefore the function $F := (a^{-*}(0))^{-1}\delta_i * a^{-*}$ satisfies:

$$\begin{cases} \forall j \notin K \uplus \{i\} & (a * F)(j) = 0; \\ \forall k \in K & F(k) \geqslant 0; \qquad F(i) = 1. \end{cases} \tag{486}$$

[¶]The general case where $\Theta^*$ can be infinite can be got from the finite case by passing to the limit, despite some technicalities of little interest.

Comparing (482) with (486), since (482) is subelliptic, we can apply a maximum principle to it[‖], which yields that $\theta \leqslant F$ everywhere. But $J(z) = O(|z|^{-(n+\alpha)})$, so by Lemma 5.5.7 in appendix, $F(j) = O(|j-i|^{-(n+\alpha)})$, and therefore

$$\mathbf{P}[\omega_j = +1|\vec{\omega}_K = \vec{\hat{\omega}}_K, \omega_i = 1] - \mathbf{P}[\omega_j = -1|\vec{\omega}_K = \vec{\hat{\omega}}_K, \omega_i = -1] = O(|j-i|^{-(n+\alpha)}), \quad (487)$$

uniformly in $i, j, K, \vec{\hat{\omega}}_K$.

The end of the proof, namely deducing Point (ii) from (487) and proving Point (iii), is then performed in the same way as to establish (464) in the proof of Theorem 5.1.1. □

Thanks to Proposition 5.1.9, we can apply the results of Chapters 3 and 4. One gets the following

**5.1.12 Theorem.** *For the long-range Ising model on $\mathbb{Z}^n$ at $T \geqslant T_1$,*

*(i) For all disjoint $I, J \subset \mathbb{Z}^n$, uniformly in $I, J$, one has an estimate*

$$\{\vec{\omega}_I : \vec{\omega}_J\} \leqslant O(dist(I,J)^{-\alpha}), \quad (488)$$

*where the $O(\cdot)$ can be turned into an explicit constant only depending on $\mathcal{J}$ and $T$. Moreover, there exists some $k < 1$ (still explicit and only depending on $\mathcal{J}$ and $T$) such that for all disjoint $I, J \subset \mathbb{Z}^n$,*

$$\{\vec{\omega}_I : \vec{\omega}_J\} \leqslant k. \quad (489)$$

*(ii) The spins satisfies the central limit theorem, in the sense that the conclusions of Theorem 4.1.8 hold for them.*

*(iii) The Glauber dynamics has a positive spectral gap.*

*(iv) Points (i) and (iii) remain valid uniformly under any law of the form $\mathbf{P}[\cdot|\vec{\omega}_K = \vec{\hat{\omega}}_K]$, for $K \subset \mathbb{Z}^n$ and $\vec{\hat{\omega}}_K \in \{\pm 1\}^K$ a 'boundary condition' on $K$.*

*Proof.* The proof is the same as the work done in the previous subsection. The only difference is to prove (488), which follows from the following computation: denoting $D := dist(I,J)$, one has that, when $D \to \infty$,

$$\sum_{\substack{z \in \mathbb{Z}^n \\ |z| \geqslant D}} \frac{1}{|z|^{n+\alpha}} \leqslant \sum_{d=D}^{\infty} \frac{\#\{z \in \mathbb{Z}^n : \ |z| = d\}}{d^{n+\alpha}} = \sum_{d=D}^{\infty} \frac{O(d^{n-1})}{|d^{n+\alpha}|} = O\Big(\sum_{d=D}^{\infty} \frac{1}{d^{1+\alpha}}\Big) = O(D^{-\alpha}). \quad (490)$$

□

## Spin glasses

*Spin glasses* are another generalization of Ising's model. In these models, the interaction constants are not invariant by translation any longer. The Hamiltonian writes

$$H(\vec{\omega}) = -\frac{1}{2} \sum_{i \neq j} J(i,j) \omega_i \omega_j \quad (491)$$

(with $J(j,i) = J(i,j)$), where the $J(i,j)$ themselves are random. We make the following assumptions on the interaction constants:

[‖]The maximum principle is generally stated in a PDE context, see for instance [19, § 3.1], but it works exactly the same for discrete equations.

**5.1.13 Assumption.** For distinct unordered pairs $\{i,j\}$, all the $J(i,j)$ are independent. Moreover, $J(i,j)$ is distributed according to some law $P_J^{(j-i)}$ only depending on $(j-i)$[*]. We will assume that all the $P_J^{(z)}$ have bounded support, and we denote by $J_\infty(z)$ the smallest number such that $P_J^{(z)}[|J| \leqslant J_\infty(z)] = 1$.

*5.1.14 Remark.* Here the $J(i,j)$ can be negative, which corresponds to antiferromagnetic interactions.

☛ *In spin glass models, there are two levels of randomness: first to fix the $J(i,j)$, next to take $\vec{\omega}$ according to the Gibbs measure associated to $H$. When both levels of randomness are taken into consideration, one speaks of* annealed *law. Here I am only interested in the* quenched *laws, which deal with the second level of randomness for fixed $J(i,j)$. I will write sentences beginning with "for almost-all quenched systems", which mean that what follows is valid for almost-all Gibbs measures when the $J(i,j)$ are taken randomly according to Assumption 5.1.13.*

The machinery exposed above still works for spin glass models. We obtain the

**5.1.15 Theorem.** *Suppose that when $|z| \to \infty$, $J_\infty(z)$ decreases at least as fast as $O(|z|^{-(n+\alpha)})$ for some $\alpha > 0$. Then there is a $T_1 < \infty$ such that, for the spin glass model on $\mathbb{Z}^n$ at $T \geqslant T_1$, for almost-all quenched systems,*

*(i) If $J_\infty(z) = O(|z|^{-(n+\alpha)})$, then for all disjoint $I, J \subset \mathbb{Z}^n$, uniformly in $I, J$, one has an estimate*

$$\{\vec{\omega}_I : \vec{\omega}_J\} \leqslant O\big(dist(I,J)^{-\alpha}\big). \tag{492}$$

*If moreover $J_\infty(z)$ has exponential decay (see Definition 5.5.4 in the appendix), then the right-hand side of (492) can even be replaced by "$\theta\big(dist(I,J)\big)$" for some function $\theta(\cdot)$ with exponential decay.*
*In both cases, there exists some $k < 1$ such that for all disjoint $I, J \subset \mathbb{Z}^n$,*

$$\{\vec{\omega}_I : \vec{\omega}_J\} \leqslant k. \tag{493}$$

*(ii) Points (ii)–(iv) of Theorem 5.1.12 hold.*

## Synthetic vocabulary

For all the models considered in this section, the techniques used and the results stated walked along the same lines. First, one establishes a bound $\{X_i : X_j\}_* \leqslant \varepsilon(j-i) \wedge k_0$ for all $i \neq j$, for some sufficiently rapidly decreasing function $\varepsilon\colon\ \mathbb{Z}^n \to [0,1]$ and some $k_0 < 1$. Then, one applies the results of Chapters 3 and 4, which yield maximal decorrelation for distant bunches of spins (which is sometimes called *(interlaced) $\rho^*$-mixing*) with uniformly non-full correlation between any two disjoint bunches of spins (which is sometimes denoted "$\rho^*(1) < 1$"), central limit theorem, and spectral gap for the Glauber dynamics.

Since this method will be used again in the following sections, it will be convenient to introduce some synthetic vocabulary:

[*]Observe that one has necessarily $P_J^{(-z)} = P_J^{(z)}$ for all $z$; in particular the function $J_\infty\colon\ \mathbb{Z}^n \smallsetminus \{0\} \to \mathbb{R}_+$ shall always be symmetric.

**5.1.16 Definition.** If a spin model (spins can have arbitrary range) $\vec{X}$ on $\mathbb{Z}^n$ satisfies some bound "$\{X_i : X_j\}_* \leqslant \varepsilon(j-i) \wedge k_0$" for all distinct $i, j \in \mathbb{Z}^n$, with $\sum_{z \in \mathbb{Z}^n \smallsetminus \{0\}} \varepsilon(z) < \infty$ and $k_0 < 1$, we say that this model is *well-ρ-mixing*. According to our results, for such a model one has $\rho^*$-mixing with $\rho^*(1) < 1$, CLT and spectral gap.

Moreover,

(i) If $\varepsilon(z) = O(|z|^{-(n+\alpha)})$ when $|z| \to \infty$, then we say that the model is *α-polynomially ρ-mixing*. According to our results, in this case $\rho^*$-mixing is polynomial with rate $\alpha$, i.e. Formula (488) holds.

(ii) If $\varepsilon(z)$ has exponential decay (cf. Definition 5.5.4), then we say that the model is *exponentially ρ-mixing*. According to our results, in this case $\rho^*$-mixing has an exponential speed of decay (but not with the same rate as $\varepsilon(\cdot)$, cf. Remark 5.5.6), i.e. a formula similar to (460) holds.

## 5.2 Quadratic models

☛ *In this subsubsection, an arbitrary norm $|\cdot|$ on $\mathbb{Z}^n$ is fixed.*

**5.2.1 Definition.** In our *quadratic model*, the states space is $\Omega = \mathbb{R}^{\mathbb{Z}^n}$ for some $n \in \mathbb{N}^*$. For $\vec{\omega}_{\mathbb{Z}^n} \in \Omega$, $i \in \mathbb{Z}^n$, the real number $\omega_i$ will be called the *polarization of particle $i$*. Each particle $i$ is submitted to two types of forces:

- A *pinning force*, preventing the particle from having a too large polarization, which derives from the quadratic potential $\omega_i^2/2$;
- *Interaction forces*: each particle $j \neq i$ exerts a force on $i$ which tends to make the polarizations of particles $i$ and $j$ equal; this force derives from a quadratic potential $\gamma_{j-i}(\omega_j - \omega_i)^2/2$.

In other words, the Hamiltonian of the system is formally defined by

$$H(\vec{\omega}) = \frac{1}{2} \sum_{i \in \mathbb{Z}^n} \omega_i^2 + \frac{1}{4} \sum_{i \neq j} \gamma_{j-i}(\omega_j - \omega_i)^2, \tag{494}$$

where the $\gamma_z$, for $z \in \mathbb{Z}^n \smallsetminus \{0\}$, are nonnegative numbers which we impose to satisfy the symmetry condition $\gamma_z = \gamma_{-z}$ for all $z$. Moreover we impose the that the sum of the $\gamma_z$ is convergent, and we denote

$$\Gamma := \sum_{z \in \mathbb{Z}^n \smallsetminus \{0\}} \gamma_z < \infty. \tag{495}$$

The Hamiltonian $H$ is a quadratic function of $\vec{\omega}$, so at fixed parameter $\beta$ the (infinite-dimensional) random vector $\vec{\omega}$ will be Gaussian (and centered). Let us compute its covariance: the probability density of $\vec{\omega}$ w.r.t. the 'Lebesgue measure' on $\Omega$ is formally defined by

$$\frac{d\mathbf{P}_\beta[\vec{\omega}]}{\prod_{i \in \mathbb{Z}^n} d\omega_i} \propto \exp\left(\tfrac{1}{2}\omega^\top(\beta Q)\omega\right), \tag{496}$$

where $Q$ is the (infinite-dimensional) symmetric matrix defined by

$$\begin{cases} Q_{ij} := -\gamma_{j-i} & \text{for } i \neq j; \\ Q_{ii} := 1 + \Gamma & \text{on the diagonal,} \end{cases} \tag{497}$$

thus the covariance matrix of $\vec{\omega}$ is $(\beta Q)^{-1}$. So we have to compute $Q^{-1}$, the inverse matrix of $Q$. Since $Q$ is a Toeplitz matrix (with $n$-dimensional indexes)[†], $Q^{-1}$ —if it exists—will be of the same form. Now, knowing that it is a Toeplitz matrix, $Q$ is described by the function $a_Q : \mathbb{Z}^n \to \mathbb{R}$ such that for all $i,j$, $Q_{ij} = a_Q(j-i)$. With this notation, (497) rewrites:

$$\forall z \in \mathbb{Z}^n \qquad a_Q(z) = \mathbf{1}_{z=0}(1+\Gamma) - \mathbf{1}_{z\neq 0}\gamma_z. \tag{498}$$

When coded by functions like $a_Q$, the multiplication of Toeplitz matrices becomes the convolution product:

$$\forall M,N \text{ Toeplitz} \qquad a_{MN} = a_M * a_N. \tag{499}$$

So, $Q^{-1}$ will be the Toeplitz matrix whose $a_{Q^{-1}}$ is the inverse of $a_Q$ for the convolution product. Thanks to Condition (495), such an inverse always exists: indeed we can write $a_Q = (1+\Gamma)(\delta_0 - \tilde{a}_Q)$, where $\tilde{a}_Q$ is a nonnegative function with $\|\tilde{a}_Q\|_{l^1} = \Gamma/(1+\Gamma) < 1$, so that $a_Q$ is invertible with

$$a_Q^{-*} = (1+\Gamma)^{-1}(\delta_0 + \tilde{a}_Q + \tilde{a}_Q * \tilde{a}_Q + \tilde{a}_Q * \tilde{a}_Q * \tilde{a}_Q + \cdots). \tag{500}$$

In the end, at parameter $\beta > 0$ the covariance matrix of $\vec{\omega}$ has entries:

$$\mathrm{Cov}(\omega_i, \omega_j) = \frac{a_{Q^{-1}}(j-i)}{\beta}. \tag{501}$$

*5.2.2 Remark.* All the entries of $\mathrm{Cov}(\vec{\omega})$ are nonnegative, which reflects the fact that all the interaction forces are attractive.

*5.2.3 Remark.* Since $\mathrm{Cov}(\vec{\omega})$ depends on $\beta$ only through a constant factor, the behaviour of the system is exactly the same, up to a multiplicative constant, for all $\beta > 0$. Hence the study of correlations will not depend on $\beta$.

☛ *In the sequel, we fix arbitrarily $\beta = 1$ and we denote $\mathbf{P}$ for $\mathbf{P}_{\beta=1}$.*

Since the model is Gaussian, by (501) and Theorem 1.2.6 one has for all $i \neq j$:

$$\{\omega_i : \omega_j\} = \frac{a_{Q^{-1}}(j-i)}{a_{Q^{-1}}(0)}. \tag{502}$$

Now we have the following claim, with an immediate key corollary:

**5.2.4 Claim.** *For all $i \neq j$, for all $K \subset \mathbb{Z} \smallsetminus \{i,j\}$,*

$$\{\omega_i : \omega_j\}_{\vec{\omega}_K} \leqslant \{\omega_i : \omega_j\}. \tag{503}$$

**5.2.5 Corollary.** *Denoting by $*$ the natural $\sigma$-metalgebra of the system, for all $i \neq j$,*

$$\{\omega_i : \omega_j\}_* = \{\omega_i : \omega_j\} = \frac{a_{Q^{-1}}(j-i)}{a_{Q^{-1}}(0)}. \tag{504}$$

*Proof.* The proof of Claim 5.2.4 relies on the following claims:

[†]Recall that saying that matrix $Q$ is *Toeplitz* means that its entries $Q_{ij}$ only depend on $(j-i)$.

**5.2.6 Claim.** *Up to an additive constant, $Law(\vec{\omega}_{\mathbb{Z}^n}|\vec{\omega}_K = \vec{\hat{\omega}}_K)$ is the same for all $\vec{\hat{\omega}}_K \in \mathbb{R}^K$, i.e. there exists a vector-valued function $\vec{\hat{\omega}}_K \mapsto offset(\vec{\hat{\omega}}_K) \in \mathbb{R}^{\mathbb{Z}^n}$ such that the law of $\vec{\omega}_{\mathbb{Z}^n}$ under $\mathbf{P}[\cdot|\vec{\omega}_K = \vec{\hat{\omega}}_K]$ is the same as the law of $\vec{\omega}_{\mathbb{Z}^n} + offset(\vec{\hat{\omega}}_K)$ under $\mathbf{P}[\cdot|\vec{\omega}_K \equiv 0]$.*

**5.2.7 Lemma.** *For $(X,Y)$ a two-dimensional centered Gaussian vector with $X$ and $Y$ non-degenerate,*

$$\{X:Y\} = \sqrt{\frac{\mathbf{E}[X^2]}{\mathbf{E}[Y^2]}}\,|\mathbf{E}[Y|X=1]|. \tag{505}$$

**5.2.8 Claim.** *For $K \subset \mathbb{Z}^n$, the function offset defined in Claim 5.2.6 is nondecreasing, in the sense that each of the entries of $offset(\vec{\hat{\omega}}_K)$ is a nondecreasing function of each $\hat{\omega}_k$ for $k \in K$.*

**5.2.9 Claim.** *For $i \in \mathbb{Z}^n$, $K \subset \mathbb{Z}^n \smallsetminus \{i\}$:*

$$offset(1^{\{i\}}, 0^K) \leqslant offset(1^{\{i\}}), \tag{506}$$

*where $(1^{\{i\}}, 0^K)$ stands for the function on $K \uplus \{i\}$ which is equal to $1$ at $i$ and to $0$ on $K$, resp. $1^{\{i\}}$ stands for the function on $\{i\}$ mapping $i$ to $1$.*

Admit temporarily the claims. Let $i,j$ be distinct points of $\mathbb{Z}^n$, let $K \subset \mathbb{Z}^n \smallsetminus \{i,j\}$ and let $\vec{\hat{\omega}}_K \in \mathbb{R}^K$; our goal is to compute $\{\omega_i : \omega_j\}$ under $\mathbf{P}[\cdot|\vec{\omega}_K = \vec{\hat{\omega}}_K]$. First, by Claim 5.2.6 we can suppose that $\vec{\hat{\omega}}_K \equiv 0$. Now under $\mathbf{P}[\cdot|\vec{\omega}_K \equiv 0]$, $(\omega_i, \omega_j)$ is still Gaussian by the properties of Gaussian vectors, and it is centered by symmetry, therefore by Lemma 5.2.7, $\{\omega_i : \omega_j\}$ is equal to

$$\sqrt{\frac{\mathbf{E}[\omega_i^2|\vec{\omega}_K \equiv 0]}{\mathbf{E}[\omega_j^2|\vec{\omega}_K \equiv 0]}}\,|\mathbf{E}\left[\omega_j|\vec{\omega}_K \equiv 0 \text{ and } \omega_i = 1\right]| = \sqrt{\frac{\mathbf{E}[\omega_i^2|\vec{\omega}_K \equiv 0]}{\mathbf{E}[\omega_j^2|\vec{\omega}_K \equiv 0]}}(offset(1^{\{i\}}, 0^K)\cdot j) \tag{507}$$

—one has indeed $offset(1^{\{i\}}, 0^K)\cdot j \geqslant 0$, since by Claim 5.2.8, $offset(1^{\{i\}}, 0^K) \geqslant offset(0^{\{i\}\uplus K}) \equiv 0$.

Now, taking $K = \emptyset$ in (507), we find that under the law $\mathbf{P}$:

$$\{\omega_i : \omega_j\} = \sqrt{\frac{\mathbf{E}[\omega_i^2]}{\mathbf{E}[\omega_j^2]}}(offset(1^{\{i\}})\cdot j), \tag{508}$$

which is $\geqslant \sqrt{\mathbf{E}[\omega_i^2]/\mathbf{E}[\omega_j^2]}(offset(1^{\{i\}}, 0^K)\cdot j)$ by Claim 5.2.9. But up to switching the roles of $i$ and $j$, we can assume that $\mathbf{E}[\omega_i^2]/\mathbf{E}[\omega_j^2] \geqslant \mathbf{E}[\omega_i^2|\vec{\omega}_K \equiv 0]/\mathbf{E}[\omega_j^2|\vec{\omega}_K \equiv 0]$, thus getting the desired result:

$$\{\omega_i : \omega_j\} \geqslant \sqrt{\frac{\mathbf{E}[\omega_i^2|\vec{\omega}_K \equiv 0]}{\mathbf{E}[\omega_j^2|\vec{\omega}_K \equiv 0]}}(offset(1^{\{i\}}, 0^K)\cdot j) = \{\omega_i : \omega_j\}_{\vec{\omega}_K}. \tag{509}$$

□

*Proof of the claims.*

*Claim 5.2.6* – It is a well-known property of Gaussian vectors, which here is stated in an infinite-dimensional setting.

*Claim 5.2.7* – Since $(X,Y)$ is centered Gaussian, $Y^{\sigma(X)}$ is the orthogonal projection of the $L^2$ variable $Y$ on $\mathbb{R}X$, so $\mathbf{E}[Y|X=x] \propto x$. Thus one has:

$$\mathbf{E}[XY] = \int x\,\mathbf{E}[Y|X=x]\,d\mathbf{P}[X=x] = \int x^2\,\mathbf{E}[Y|X=1]\,d\mathbf{P}[X=x] = \mathbf{E}[Y|X=1]\,\mathbf{E}[X^2]. \tag{510}$$

But for such a Gaussian vector, Theorem 1.2.6 gives that

$$\{X:Y\} = \frac{|\mathrm{Cov}(X,Y)|}{\mathrm{Sd}(x)\,\mathrm{Sd}(y)} = \frac{|\mathbf{E}[XY]|}{\sqrt{\mathbf{E}[X^2]\,\mathbf{E}[Y^2]}}, \tag{511}$$

which combined with (510) gives (505).

*Claim 5.2.8* – First, notice that $\mathbf{E}[\vec{\omega}_{\mathbb{Z}^n}|\vec{\omega}_K \equiv 0] = \vec{0}$, so that

$$\mathit{offset}(\vec{\hat{\omega}}_K) = \mathbf{E}\big[\vec{\omega}|\vec{\omega}_K = \vec{\hat{\omega}}_K\big]. \tag{512}$$

Now, allowing temporarily $\beta$ to vary again, by the properties of Gaussian vectors, the vector-valued variable $\mathbf{E}_\beta[\vec{\omega}_{\mathbb{Z}^n}|\vec{\omega}_K = \hat{\omega}_K]$ is Gaussian with constant expectation and covariance matrix proportional to $\beta$. Therefore, the common expectation of all these laws is equal to the constant value of $\vec{\omega}_{\mathbb{Z}^n}$ for $\beta = 0$, which is the $\vec{\omega}$ minimising $H$ under the constraint "$\vec{\omega}_K = \vec{\hat{\omega}}_K$":

$$\mathit{offset}(\vec{\hat{\omega}}_K) = \operatorname*{arg\,min}_{\vec{\omega}_K = \vec{\hat{\omega}}_K} H(\vec{\omega}). \tag{513}$$

Since it minimizes energy, the state $\mathit{offset}(\vec{\omega}_K)$ is at equilibrium outside $K$. In other words, it is the solution of the following subelliptic system:

$$\begin{cases} \forall i \in \mathbb{Z}^n \smallsetminus K \quad -\omega_i + \sum_{j\neq i} \gamma_{j-i}(\omega_j - \omega_i) = 0; \\ \forall i \in K \quad \omega_i = \hat{\omega}_i. \end{cases} \tag{514}$$

(That system is clearly subelliptic because the pinning and interaction forces all are attractive). By the maximum principle [19, § 3.1], the solution of (514) is an increasing function of the boundary condition, which was our claim.

*Claim 5.2.9* – Denote $\vec{\omega}^1_{\mathbb{Z}^n} = \mathit{offset}(1^{\{i\}})$. Since $\mathit{offset}(0^{\{i\}}) = 0^{\mathbb{Z}^n}$, by Claim 5.2.8 one has $\vec{\omega}^1_{\mathbb{Z}^n} \geqslant 0^{\mathbb{Z}^n}$. Since obvisouly $\vec{\omega}^1_i = 1$, one has even $\vec{\omega}^1_{\mathbb{Z}^n} \geqslant (1^{\{i\}}, 0^{\mathbb{Z}^n \smallsetminus \{i\}})$. In particular, $\vec{\omega}^1_{\{i\}\uplus K} \geqslant (1^{\{i\}}, 0^K)$; therefore, using again Claim 5.2.8,

$$\mathit{offset}(\vec{\omega}^1_{\{i\}\uplus K}) \geqslant \mathit{offset}(1^{\{i\}}, 0^K). \tag{515}$$

Now, we defined $\vec{\omega}^1_{\mathbb{Z}^n}$ as $\mathit{offset}(1^{\{i\}})$, so by Formula (514) it satisfies

$$-\omega^1_{i'} + \sum_{j\neq i'} \gamma_{j-i'}(\omega_j - \omega'_i) = 0 \tag{516}$$

for all $i' \in \mathbb{Z}^n \smallsetminus \{i\}$, hence *a fortiori* for all $i' \in \mathbb{Z}^n \smallsetminus (\{i\} \uplus K)$. Since moreover $\vec{\omega}^1$ obviously coincides with $\vec{\omega}^1_{\{i\}\uplus K}$ on $\{i\} \uplus K$, this implies, by Formula (514) again, that

$$\mathit{offset}(\vec{\omega}^1_{\{i\}\uplus K}) = \vec{\omega}^1_{\mathbb{Z}^n} = \mathit{offset}(1^{\{i\}}). \tag{517}$$

So, (515) becomes "$\mathit{offset}(1^{\{i\}}) \geqslant \mathit{offset}(1^{\{i\}}, 0^K)$", what we wanted. □

Thanks to Corollary 5.2.5 our tensorization theorems give decorrelation results for the quadratic model:

**5.2.10 Theorem.** *Provided Condition (495) holds:*

*(i) The quadratic model is well-ρ-mixing, cf. Definition 5.1.16. If* $\Gamma < 1$*, one can be more specific about the property "*$\rho^*(1) < 1$*": for all disjoint* $I, J \subset \mathbb{Z}^n$*,* $\{\vec{\omega}_I : \vec{\omega}_J\} \leqslant \Gamma$.

*(ii) Moreover, if there is polynomial decay of interactions* $\gamma_z = O(1/|z|^{n+\alpha})$*, then the model is α-polynomially ρ-mixing, and if* $\gamma_z$ *has exponential decay, then the model is exponentially ρ-mixing (but not with the same rate as* $\gamma_z$ *in general).*

*Proof.* To prove Point (i), we have to show that

$$\sum_{z \in \mathbb{Z}^n \smallsetminus \{0\}} \frac{a_{Q^{-1}}(z)}{a_{Q^{-1}}(0)} \leqslant \Gamma \tag{518}$$

—recall that we assumed $\Gamma < \infty$. We write:

$$\sum_{z \in \mathbb{Z}^n \smallsetminus \{0\}} \frac{a_{Q^{-1}}(z)}{a_{Q^{-1}}(0)} = \frac{\sum_{z \in \mathbb{Z}^n} a_{Q^{-1}}(z)}{a_{Q^{-1}}(0)} - 1. \tag{519}$$

There, $\sum_{z \in \mathbb{Z}^n} a_{Q^{-1}}(z)$ is equal to 1: indeed, $a_{Q^{-1}}$ is the convolution inverse of $a_Q$, so by Fubini's theorem:

$$\sum_{z \in \mathbb{Z}^n} a_{Q^{-1}}(z) = \Big( \sum_{z \in \mathbb{Z}^n} a_Q(z) \Big)^{-1}, \tag{520}$$

where

$$\sum_{z \in \mathbb{Z}^n} a_Q(z) = \sum_{z \in \mathbb{Z}^n \smallsetminus \{0\}} \big( -\gamma(z) \big) + (1 + \Gamma) = -\Gamma + 1 + \Gamma = 1. \tag{521}$$

Now, by (500), $a_{Q^{-1}}(0)$ is obviously bounded below by $(1+\Gamma)^{-1}$, so in the end:

$$\sum_{z \in \mathbb{Z}^n \smallsetminus \{0\}} \frac{a_{Q^{-1}}(z)}{a_{Q^{-1}}(0)} \leqslant \frac{1}{(1+\Gamma)^{-1}} - 1 = \Gamma. \tag{522}$$

To prove Point (ii), we have to show that polynomial decay of $\gamma_z$ implies polynomial decay of $a_{Q^{-1}}$ with the same exponent, resp. that exponential decay of $\gamma_z$ implies exponential decay of $a_{Q^{-1}}$. This is achieved resp. by Lemmas 5.5.5 and 5.5.7 in the appendix. □

## 5.3 Nonlinear lattice of particles

In this section we will consider a model with continuous spins, but where interactions are nonlinear, so that we cannot use the properties of Gaussian variables. One has a lattice of particles indexed by $\mathbb{Z}^n$ (equipped with its $l^1$ graph structure), each particle $i$ being described by its "polarization" $\omega_i \in \mathbb{R}$. Each particle is submitted to a pinning force deriving from a potential $V$, and to interaction forces with its neighbours, the interactions deriving from a potential $W$. In other words, the Hamiltonian is formally

$$H(\vec{\omega}) = \sum_{i \in \mathbb{Z}} V(\omega_i) + \frac{1}{2} \sum_{i \sim j} W(\omega_j - \omega_i). \tag{523}$$

We make the following assumptions:

**5.3.1 Assumption.** Both $V$ and $W$ are convex; moreover $V$ is uniformly strictly convex and the Hessian of $W$ is bounded, i.e. there exist constants $v_* > 0$ and $w_* < \infty$ such that for all $x \in \mathbb{R}$, $v_* \leqslant V''(x)$ and $W''(x) \leqslant w^*$.

We are interested in the equilibrium state of the system at some inverse temperature $0 < \beta < \infty$. (In the sequel we suppose that $\beta$ is fixed).

Let $i \neq j \in \mathbb{Z}$, $K \subset \mathbb{Z} \smallsetminus \{i,j\}$ and $\vec{\hat{\omega}}_K \in \mathbb{R}^K$; we want to study the law of $(\omega_i, \omega_j)$ under the law $\mathbf{P}[\cdot|\vec{\omega}_K = \vec{\hat{\omega}}_K]$. Then, the probability distribution of the system is formally described by

$$d\mathbf{P}(\omega_i, \omega_j, \vec{\omega}_{K^c \smallsetminus \{i,j\}}) \propto \exp\big(-\beta H(\omega_i, \omega_j, \vec{\omega}_{K^c \smallsetminus \{i,j\}}, \vec{\hat{\omega}}_K)\big). \tag{524}$$

Our assumptions ensure that the function $H(\cdot,\cdot,\cdot,\vec{\hat{\omega}}_K)$ is uniformly convex, so that the equilibrium exists and is unique.

For the sequel, we need to recall the definition of the $W_\infty$ Wasserstein distance:

**5.3.2 Definition** (see also [11]). For $\mu_1, \mu_2$ two measures on some metric space $(X,d)$, "$W_\infty(\mu_1,\mu_2) \leqslant \varepsilon$" means that there exists a probability measure $\gamma$ on $E^2$ such that the two respective marginals of $\gamma$ are $\mu_1$ and $\mu_2$ and such that $d(x_1,x_2) \leqslant \varepsilon$ $\gamma$-a.s.. This defines a (possibly infinite) distance on the probability measures on $E$.

The fundamental lemma of this subsection is the following

**5.3.3 Claim.** *For $\hat{\omega}_j \in \mathbb{R}$, denote by $\mu(\hat{\omega}_j)$ the law of $\omega_i$ under $\mathbf{P}[\cdot|\vec{\omega}_{K \uplus \{j\}} = (\vec{\hat{\omega}}_K, \hat{\omega}_j)]$. There exists a function $\varepsilon\colon\ \mathbb{Z} \to [0,1]$ with $\varepsilon(d) < 1$ as soon as $d > 0$ and $\varepsilon(d) \overset{d\to\infty}{\leqslant} Ce^{-\psi d}$ for some $\psi > 0$ and $C < \infty$, such that*

$$\forall \hat{\omega}_j^1, \hat{\omega}_j^2 \in \mathbb{R} \qquad W_\infty(\mu(\hat{\omega}_j^1), \mu(\hat{\omega}_j^2)) \leqslant \varepsilon(|j-i|)|\hat{\omega}_j^2 - \hat{\omega}_j^1|. \tag{525}$$

*Proof.* The proof relies on the 'explicit' construction of a coupling measure $\gamma$ between $\mu(\hat{\omega}_j^1)$ and $\mu(\hat{\omega}_j^2)$. To do that, we will construct $\mathbf{P}[\cdot|\vec{\omega}_{K \uplus \{j\}} = (\vec{\hat{\omega}}_K, \hat{\omega}_j)]$ thanks to the Glauber dynamics of the system, and then couple the Glauber dynamics for $\hat{\omega}_j^1$ and $\hat{\omega}_j^2$.

We define the Glauber dynamics thanks to independent white noises $(dB_t^{i'})_{t\in\mathbb{R}}$ for $i' \in \mathbb{Z}^n \smallsetminus (K \uplus \{j\})$. The motion of point $i'$ is defined by:

$$d\omega_{i'} = -\beta\big(V'(\omega_{i'}) + \sum_{i''\sim i'} W'(\omega_{i'} - \omega_{i''})\big) + \sqrt{2}\, dB_t^{i'}, \tag{526}$$

with the boundary condition $\vec{\omega}_{K \uplus \{j\}} = \vec{\hat{\omega}}_{K \uplus \{j\}}$ for all times. Coupling then consists in taking the same noise for the two processes. The initial condition is not very important since it is asymptotically forgotten, so we will suppose that the two systems have been coupled for an infinite time, so that at any time both systems follow their equilibrium law. We denote by $\vec{\omega}^1(t)$ the system correponding to the boundary condition "$\vec{\omega}_{K \uplus \{j\}} = (\vec{\hat{\omega}}_K, \hat{\omega}_j^1)$", resp. by $\vec{\omega}^2(t)$ the system correponding to the other boundary condition. We denote $\Delta_{i'}(t) := \omega_{i'}^2(t) - \omega_{i'}^1(t)$. Then when the dynamics are coupled, $\vec{\Delta}$ evolves according to the following equation:

$$d(\Delta_{i'}) = -\beta\Big[\big(V'(\omega_{i'}^2) - V'(\omega_{i'}^1)\big) + \sum_{i''\sim i'} \big(W'(\omega_{i'}^2 - \omega_{i''}^2) - W'(\omega_{i'}^1 - \omega_{i''}^1)\big)\Big]. \tag{527}$$

Obviously the right-hand side is not a deterministic function of $\vec{\Delta}(t)$, but it can nonetheless be written as

$$-\beta[v(i',t)\Delta_{i'}(t)+w(i',i'',t)(\Delta_{i'}(t)-\Delta_{i''}(t))], \tag{528}$$

for some $v(i',t)$ and $w(i',i",t)$ satisfying

$$\begin{aligned} v(i',t) &\geqslant v_* \quad \text{and} && (529)\\ 0\leqslant w(i',i",t) &\leqslant w^* && (530) \end{aligned}$$

by Assumption 5.3.1. Moreover, one has the boundary conditions:

$$\forall t \qquad \left\{ \begin{array}{rcl} \vec{\Delta}_K & \equiv & 0; \\ \Delta_j & = & \hat{\omega}_j^2-\hat{\omega}_j^1. \end{array} \right. \tag{531}$$

So, $\vec{\Delta}$ is the solution of some discrete 'damped heat equation', whose coefficients can vary along time though having to satisfy bounds (529) and (530). Such an equation has no stationary solution *stricto sensu*; however there exists some $\vec{\Delta}^+$ such that

$$\vec{\Delta}(t)\leqslant\vec{\Delta}^+ \qquad \Rightarrow \qquad \forall t'\geqslant t \quad \vec{\Delta}(t')\leqslant\vec{\Delta}^+; \tag{532}$$

namely, this $\vec{\Delta}^+$ is defined as the solution of the following system of equations: $\vec{\Delta}^+_K\equiv 0$, $\Delta_j=\hat{\omega}_j^2-\hat{\omega}_j^1$, and for all $i'\notin K\uplus\{j\}$,

$$0=-v_*\Delta^+_{i'}+\sum_{i''\sim i'}\mathbf{1}_{\Delta^+_{i''}\geqslant\Delta^+_{i'}}w^*(\Delta^+_{i''}-\Delta^+_{i'}). \tag{533}$$

One has similarly that

$$\vec{\Delta}(t)\geqslant\vec{0} \qquad \Rightarrow \qquad \forall t'\geqslant t \quad \vec{\Delta}(t')\geqslant\vec{0}. \tag{534}$$

Consequently, I claim that for all $t$ one has

$$\vec{0}\leqslant\vec{\Delta}(t)\leqslant\vec{\Delta}^+ : \tag{535}$$

indeed if the initial condition of the system satisfies (535), then that property remains valid for all subsequent times; now, as I told, initial conditions are asymptotically forgotten, so in fact (535) is always satisfied.

One has the following control on $\vec{\Delta}^+$:

**5.3.4 Claim.** *There exists a function $\varepsilon\colon\ \mathbb{Z}\to[0,1]$ with $\varepsilon(d)<1$ as soon as $d>0$ and $\varepsilon(d)\overset{d\to\infty}{\leqslant}Ce^{-\psi d}$ for some $\psi>0$ and $C<\infty$, such that*

$$\forall i\in\mathbb{Z}^n \qquad \Delta^+_i\leqslant\varepsilon(|j-i|). \tag{536}$$

*Moreover, the function $\varepsilon$ does not depend on $K$ nor on $j$.*

Combining (535) with Claim 5.3.4 ends the proof of Claim 5.3.3. □

*Proof of Claim 5.3.4.* First, notice that Equation (533) satisfies a maximum principle, so we know in advance that $\Delta^+$ is uniquely defined with $0 \leqslant \Delta^+ \leqslant 1$ everywhere.

For $i' \sim i''$, denote $w_{i'}(i'') := \mathbf{1}_{\Delta^+_{i''} \geqslant \Delta^+_{i'}} w^*$. Then (533) can be rewritten into:

$$\Delta_{i'} = \sum_{i'' \sim i'} \frac{w_{i'}(i'')}{v_* + \sum_{i''\sim i'} w_{i'}(i'')} \times \Delta_{i''} + \frac{v_*}{v_* + \sum_{i''\sim i'} w_{i'}(i'')} \times 0. \tag{537}$$

Now I define the following Markov chain on $\mathbb{Z}^n \uplus \{\partial\}$, $\partial$ denoting a cemetery point:

**5.3.5 Definition.**

- If at some time the particle is on some point $i'$ of $\mathbb{Z}^n \smallsetminus (K \uplus \{j\})$, at next time it jumps onto the neighbour $i''$ of $i'$ with probability $w_{i'}(i'')/(v_* + \sum_{i''\sim i'} w_{i'}(i''))$, and it jumps onto $\partial$ with probability $v_*/(v_* + \sum_{i''\sim i'} w_{i'}(i''))$;
- If the particle is somewhere in $K \uplus \{\partial, j\}$ at some time, then it does not move any more.

Call $(X_t)_{t\in\mathbb{N}}$ such a Markov chain and denote by $\mathscr{L}$ its generator. It is clear that with probability one, $X_t$ eventually remains at some point of $K \uplus \{\partial, i\}$. Extend $\Delta^+$ to $\mathbb{Z}^n \uplus \{\partial\}$ by setting $\Delta^+_\partial = 0$; then, (537) merely means that $\Delta^+$ is $\mathscr{L}$-harmonic, and it follows that

$$\Delta^+_i = \mathbf{E}[f(X_\infty) | X_0 = i]. \tag{538}$$

Thus, to bound above $\Delta^+_i$ I write that

$$\begin{aligned}
\mathbf{E}[f(X_\infty)|X_0 = i] &= \sum_{\substack{i=i_0\sim\cdots\sim i_t=j \\ i_1,\ldots,i_{t-1}\notin K\uplus\{j\}}} \prod_{u=0}^{t-1} \frac{w_{i_u}(i_{u+1})}{v_* + \sum_{i''\sim i_u} w_{i'}(i'')} \\
&\leqslant \sum_{\substack{i=i_0\sim\cdots\sim i_t=j \\ i_1,\ldots,i_{t-1}\neq j}} \prod_{u=0}^{t-1} \frac{w_{i_u}(i_{u+1})}{\sum_{i''\sim i_u} w_{i'}(i'')} \Big(\frac{2dw^*}{2dw^* + v_*}\Big)^t \\
&\leqslant \Big(\frac{2dw^*}{2dw^* + v_*}\Big)^{|j-i|} \underbrace{\sum_{\substack{i=i_0\sim\cdots\sim i_t=j \\ i_1,\ldots,i_{t-1}\neq j}} \prod_{u=0}^{t-1} \frac{w_{i_u}(i_{u+1})}{\sum_{i''\sim i_u} w_{i'}(i'')}}_{\leqslant 1} \leqslant \Big(\frac{2dw^*}{2dw^* + v_*}\Big)^{|j-i|}.
\end{aligned} \tag{539}$$

□

From Claim 5.3.3, we take the following

**5.3.6 Corollary.** *For a Lipschitzian function $f\colon \mathbb{R} \to \mathbb{R}$, denote by $\|f\|_{Lip}$ the optimal Lipschitz constant for $f$. On $\overline{L}^2(\omega_i)$, define the (possibly infinite) norm $\|\cdot\|_{Lip}$ such that $\|f(\omega_i)\|_{Lip} = \|f\|_{Lip}$*[‡]*; denote by $\overline{Lip}(\omega_i)$ the corresponding Banach space.*

*Then under the law $\mathbf{P}[\cdot|\vec{\omega}_K = \vec{\hat{\omega}}_K]$, the map $\pi_{\omega_j\omega_i}$ defined by (37) is $\varepsilon(|j-i|)$-contracting when seen as an application from $\overline{Lip}(\omega_i)$ into $\overline{Lip}(\omega_j)$.*

[‡]This definition can be unambiguous if the support of $\omega_i$ is not the whole $\mathbb{R}$; in that case, just add an infimum in the definition.

Consequently, the map $\pi_{\omega_i\omega_j\omega_i}\colon \overline{Lip}(\omega_i) \to \overline{Lip}(\omega_i)$ is $\varepsilon(|j-i|)^2$-contracting. But the canonical embedding $\overline{Lip}(\omega_i) \mapsto \bar{L}^2(\omega_i)$ is continuous as our hypotheses ensure that $Law(\omega_i)$ is uniformly log-concave, therefore for all $f \in \overline{Lip}(\omega_i)$ one has

$$\overline{\lim_{k\to\infty}} |\langle \pi^k_{\omega_i\omega_j\omega_i} f, f\rangle_{\bar{L}^2(\omega_i)}|^{1/k} \leqslant \varepsilon(|j-i|)^2. \tag{540}$$

Since $\pi_{\omega_i\omega_j\omega_i}$ is self-adjoint in $\bar{L}^2(\omega_i)$ and $\overline{Lip}(\omega_i)$ is a dense subset of $\bar{L}^2(\omega_i)$, it follows by Lemma 0.3.1 that $\pi_{\omega_i\omega_j\omega_i}$ is $\varepsilon(|j-i|)^2$-contracting also in $\bar{L}^2(\omega_i)$. This, by Remark 1.1.10, is equivalent to saying that

$$\{\omega_i : \omega_j\}_{\vec{\omega}_K} \leqslant \varepsilon(|j-i|). \tag{541}$$

(541) is what we need to apply Lemma 3.6.8; in the end, we get the

**5.3.7 Theorem.** *The model (523) is exponentially ρ-mixing.*

## 5.4 A hypocoercive system of interacting particles

For the time being we have only been dealing with *spatial* decorrelations. Yet I have had the idea that the ability of Hilbertian decorrelations to get tensorized for infinite sets could be well adapted to the study of *temporal* relaxation of an infinite stochastic system: one can consider indeed time as an extra dimension for the particle system, which leads to a situation analogous to the parallel hyperplanes of § 0.1.c. In the reversible case, we saw that spectral techniques make it possible to get $L^2$ results from $L^1$ results, cf. Theorem 0.1.9. Here I will show how Hilbertian decorrelations can be used for a *non-reversible* particle stochastic system.

The system which we will study here as an example is governed by the *Vlasov-Fokker-Planck* equation. This equation, which arises naturally in physics, corresponds to a Hamiltonian evolution perturbed by some noise *acting on speeds*. The study of such systems is made complicated by the fact that diffusion is only performed along certain directions of the states space, so that the non-reversibility of the evolution is essential to ensure convergence to equilibrium. In [44], Villani proves $L^2$ convergence for such systems in situations where the state of the system lives in a finite-dimensional manifold. Here we will use tensorization of Hilbertian decorrelations in a fundamental way to get a result valid in an infinite-dimensional setting. Moreover, we will get non-trivial bounds for arbitrary small times, which is a new feature compared to [44].

**5.4.1 Definition.** For real parameters $m, \omega, c, T, \lambda > 0$[§], we consider a system of particles $i$ indexed by $\mathbb{Z}$, each particle being described by its momentum $p_i \in \mathbb{R}$ and its position $q_i \in \mathbb{R}$. We consider the Hamiltonian

$$H(\vec{p},\vec{q}) = m^{-1}\sum_{i\in\mathbb{Z}} \frac{p_i^2}{2} + m\omega^2 \sum_{i\in\mathbb{Z}} \frac{q_i^2}{2} + mc^2 \sum_{i\in\mathbb{Z}} \frac{(q_{i+1}-q_i)^2}{2}. \tag{542}$$

[§] $m$ is the mass of each particle, $\omega$ is the frequency corresponding to the pinning potential, $c$ is more or less the speed of sound, expressed in inter-atomic distances by unit of time, $T$ is the temperature and $\lambda$ is the relaxation constant of the friction. Physical homogeneity of these constants are resp. $[\mathsf{M}], [\mathsf{T}^{-1}], [\mathsf{T}^{-1}], [\mathsf{ML}^2\mathsf{T}^{-2}], [\mathsf{T}^{-1}]$.

Then the system $(\vec{p}(u), \vec{q}(u))$ evolves according to the Hamiltonian $H$, plus a white noise independent on each $p_i$, plus a friction force $F_i = -\lambda p_i$ on each $i$ which dissipates the energy brought by the white noise, friction being adjusted to the noise so that their association constitutes a (volumic) thermal bath at temperature $T$. One computes that this means that the quadratic variation on $p_i$ is given by $d[p_i] = 2T\lambda m\, du$.

In other words, if $(W_i(u))_{i\in\mathbb{Z}}$ denotes a family of independent brownian motions, the evolution of the system is given by

$$\left\{ \begin{array}{rcl} dp_i &=& (-m\omega^2 q_i + mc^2(q_{i-1}+q_{i+1}-2q_i) - \lambda p_i)\, du + \sqrt{2T\lambda m}\, dW_i \\ dq_i &=& m^{-1} p_i\, du. \end{array} \right. \tag{543}$$

*5.4.2 Remark.* The system of Definition 5.4.1 is to be thought as a toy model for a large class of similar systems obtained by generalizing it in several ways. A first example, which would change almost nothing but complicating the formalism, is to replace the states space $\mathbb{R}\times\mathbb{R}$ of each particle by $\mathbb{R}^n\times\mathbb{R}^n$, or to replace the lattice $\mathbb{Z}$ by $\mathbb{Z}^n$. A trickier generalization is to consider the case of non-harmonic interactions: then I expect the results stated below to remain qualitatively true, but proving them might be far more difficult since one cannot use the properties of Gaussian vectors any more. Also, if one allows for infinite-ranged interactions, which speed of decay is required to get temporal decorrelations?

All these questions look quite worthwhile to me, though answering them is out of the scope of this monograph. Here I will only show how maximal correlations make everything work fine for the toy model, hoping that it shall be useful for the general situation.

Let us consider the equilibrium dynamics of our system. We fix an arbitrary time $0 < t < \infty$. Denote by $(p_i, q_i)$ the state of particle $i$ at time $u = 0$, resp. by $(p'_i, q'_i)$ the state of particle $i$ at time $u = t$. We have to prove the

**5.4.3 Claim.** *Provided $t$ is small enough, for all $i, j \in \mathbb{Z}$ (possibly identical), one has $\{p_i : p'_j\}_*, \{p_i, q'_j\}_*, \{q_i : p'_j\}_*, \{q_i, q'_j\}_* < 1$, uniformly in $i, j$. Moreover, still uniformly in $i, j$, these quantities are bounded by $O(e^{-\gamma|j-i|})$ for some $\gamma > 0$.*

*Proof.* We denote by $\eta$ (resp. $\eta'$) $\in \mathbb{R}^{\mathbb{Z}\times\{p,q\}}$ the global state $(p_i, q_i)_{i\in\mathbb{Z}}$ (resp. $(p'_i, q'_i)_{i\in\mathbb{Z}}$) at time 0 (resp. $t$). We also denote by $(\varphi^u)_{u\geqslant 0}$ the semigroup of operators on $\mathbb{R}^{\mathbb{Z}\times\{p,q\}}$ corresponding to the evolution of the system in absence of noise, but with the friction remaining. Since the system is linear, the $\varphi^u$ are linear operators.

By the work of § 5.2, we know that $\eta$ is distributed according to the centered Gaussian law with covariance matrix $T^{-1}\check{C}$, where $\check{C}$ is defined as $\check{Q}^{-1}$, the matrix $\check{Q}$ being in turn defined by:

$$\check{Q}_{p_i p_i} := m^{-1}; \tag{544}$$

$$\check{Q}_{q_i q_i} := m(\omega^2 + 2c^2); \tag{545}$$

$$\check{Q}_{q_i q_{i\pm 1}} := -mc^2, \tag{546}$$

the other entries of $\check{Q}$ being zero. Observe that, as the matrix of a quadratic form, $\check{Q}$ is bounded (this is obvious from (544)–(546)); moreover, $\check{Q}^{-1}$ (actually exists and) is also bounded: that follows from $\check{Q}$'s being bounded below by the matrix having the same expression with $c$ replaced by 0, which we denote by $\check{Q}^\circ$, which is a strictly positive 'scalar' matrix (modulo some homogeneity constant).

Because of the linear nature of the system, we have moreover that, conditionally to $\eta$, the law of $\eta'$ is some Gaussian vector of the form $\varphi^t\eta+\theta$, where $\theta$ is a centered Gaussian vector whose law does not depend on $\eta$. Let us denote by $\hat{C}$ the covariance matrix of $\theta$, and $\hat{Q}=\hat{C}^{-1}$ —though for the time being it is not clear that $\hat{Q}$ exists.

Then, we can formally write the covariance matrix $\bar{C}$ of $(\eta,\eta')$ as $\bar{C}=\bar{Q}^{-1}$, with:

$$\bar{Q}(\eta,\eta')=\check{Q}(\eta)+\hat{Q}(\eta'-\varphi^t\eta). \tag{547}$$

(Note that $\bar{Q}$ is a quadratic form on $\mathbb{R}^{\mathbb{Z}\times\{p,q,p',q'\}}$, while $\check{Q}$ and $\hat{Q}$ were defined on $\mathbb{R}^{\mathbb{Z}\times\{p,q\}}$).

**5.4.4 Notation.** In the sequel, we shorthand "$\mathbb{Z}\times\{p,q\}$" into "$\mathbb{Z}^{\uplus 2}$", resp. "$\mathbb{Z}\times\{p,q,p',q'\}$" into "$\mathbb{Z}^{\uplus 4}$".

Now I claim that there exists constants $0<r\leqslant R<\infty$ such that $r\mathbf{I}\leqslant\bar{Q}\leqslant R\mathbf{I}$. Well, this is meaningless *stricto sensu*, because all the entries of $\bar{Q}$ do not have the same physical homogeneity, so we have to 'convert' momenta into positions by dividing them by some homogeneity parameter $\chi$, say $\chi=m\omega$ —but other choices may be more relevant.

First, I claim that $\bar{Q}\geqslant\frac{1}{2}(\chi^2m^{-1}\wedge m\omega^2)\mathbf{I}$. Let indeed $(\eta,\eta')=(\vec{p}_{\mathbb{Z}},\vec{q}_{\mathbb{Z}},\vec{p'}_{\mathbb{Z}},\vec{q'}_{\mathbb{Z}})\in\mathbb{R}^{\mathbb{Z}^{\uplus 4}}$ with finite support. We observe that

$$\|(\eta,\eta')\|^2=\sum_{i\in\mathbb{Z}}(\chi^{-2}p_i^2+q_i^2+\chi^{-2}p'^2_i+q'^2_i)=\|\eta\|^2+\|\eta'\|^2, \tag{548}$$

so that either $\|\eta\|^2\geqslant\frac{1}{2}\|(\eta,\eta')\|^2$ or $\|\eta'\|^2\geqslant\frac{1}{2}\|(\eta,\eta')\|^2$. Now, recalling the definition of $\check{Q}^\circ$ a few lines above, $\check{Q}(\eta)\geqslant\check{Q}^\circ(\eta)\geqslant(\chi^2m^{-1}\wedge m\omega^2)\|\eta\|^2$, so by (547), $\bar{Q}(\eta,\eta')\geqslant(\chi^2m^{-1}\wedge m\omega^2)\|\eta\|^2$. Since reversing the sense of time yields the same system with the sense of speed reversed, which does not change the norms of $\eta$ and $\eta'$, one has similarly $\bar{Q}(\eta,\eta')\geqslant(\chi^2m^{-1}\wedge m\omega^2)\|\eta'\|^2$. The claim follows.

The second point consists in proving that $\bar{Q}$ is bounded above. On the one hand, by (544)–(546),

$$\check{Q}(\eta)\leqslant(m^{-1}\chi^2\vee m(\omega^2+4c^2))\|\eta\|^2\leqslant(m^{-1}\chi^2\vee m(\omega^2+4c^2))\|(\eta,\eta')\|^2. \tag{549}$$

Next, the difficult point is to prove that $\hat{Q}(\eta'-\varphi^t\eta)$ (exists and) can be bounded above by a multiple of $\|(\eta,\eta')\|^2$. We begin with transforming the original problem of bounding a quadratic form on $\mathbb{R}^{\mathbb{Z}^{\uplus 4}}$ into a problem on $\mathbb{R}^{\mathbb{Z}^{\uplus 2}}$. Indeed, $\|\varphi^t\eta\|$ is bounded by a multiple of $\|\eta\|$, since the operator $\varphi^t$ dissipates the energy $H(\eta)$, energy which the previous work on $\check{Q}$ proved to be controlled below and above by $\|\eta\|^2$; therefore, it suffices to prove that the quadratic form $\hat{Q}(\eta)$ on $\mathbb{R}^{\mathbb{Z}^{\uplus 2}}$ is bounded by a multiple of $\|\eta\|^2$ to achieve our goal.

The natural quantity to be computed for $\theta$ (recall that $\theta$ denotes the total effect of noise between times 0 and $t$) is its covariance matrix $\hat{C}$. Its expression is the following (the notation is explained just below):

$$\hat{C}=2T\lambda m\int_0^t\varphi^{t-u}\mathbf{I}_p(\varphi^{t-u})^{\mathsf{T}}\,du, \tag{550}$$

where $\mathbf{I}_p$ is the diagonal matrix being 1 on diagonal entries indexed by some $p_i$ and 0 on diagonal entries indexed by some $q_i$, and $(\varphi^{t-u})^\top$ is the transpose of the linear operator $\varphi^{t-u}$ seen as a square matrix indexed by $\mathbb{Z}^{\uplus 2}$. This decomposition means that we are summing the contributions of all the elementary noises occuring at times $u \in [0,t]$, using that these elementary noises are independent.

Now we need an approximate expression for $\varphi^u$, $u \in [0,t]$. Here for the sake of legibility I will remain at a formal level, giving only limited expansions; it is essential nevertheless to keep in mind that all the "$O(*)$" can be made explicit by using Gronwall's lemma, and that these explicit values ensure that the $O(*)$ behave well provided $t$ is small enough. One finds that

$$\begin{aligned}
\varphi^u \delta_{p_i} \cdot p_j &= c^{2|j-i|} \frac{u^{2|j-i|}}{(2|j-i|)!} + O(u^{2|j-i|+2}); && (551)\\
\varphi^u \delta_{p_i} \cdot q_j &= m^{-1} c^{2|j-i|} \frac{u^{2|j-i|+1}}{(2|j-i|+1)!} + O(u^{2|j-i|+3}); && (552)\\
\varphi^u \delta_{q_i} \cdot p_i &= m\omega^2 u + O(u^3); && (553)\\
\varphi^u \delta_{q_i} \cdot p_{j\neq i} &= m c^{2|j-i|} \frac{u^{2|j-i|-1}}{(2|j-i|-1)!} + O(u^{2|j-i|+1}); && (554)\\
\varphi^u \delta_{q_i} \cdot q_j &= c^{2|j-i|} \frac{u^{2|j-i|}}{(2|j-i|)!} + O(u^{2|j-i|+2}). && (555)
\end{aligned}$$

Injecting Equations (551)–(555) into (550), one finds that:[¶]

$$\begin{aligned}
\hat{C}_{p_i p_i} &= 2T\lambda m t + O(t^3); && (556)\\
\hat{C}_{p_i q_i} &= T\lambda t^2 + O(t^4); && (557)\\
\hat{C}_{q_i q_i} &= \tfrac{2}{3} T\lambda m^{-1} t^3 + O(t^5); && (558)\\
\hat{C}_{p_i p_{j\neq i}} &= O(t^{2|j-i|+1}); && (559)\\
\hat{C}_{p_i q_{j\neq i}} &= O(t^{2|j-i|+2}); && (560)\\
\hat{C}_{q_i q_{j\neq i}} &= O(t^{2|j-i|+3}). && (561)
\end{aligned}$$

Consequently, the covariance matrix $\hat{C}$ can be seen as a perturbation of the matrix $\hat{C}^\circ$ which is defined by Equations (556)–(561), but with the "$O(*)$" terms replaced by 0. Since $\hat{C}^\circ$ is invertible, with an explicitly computable inverse, one finds that $\hat{C}$ is invertible too with:

$$\begin{aligned}
\hat{Q}_{p_i p_i} &= 2T^{-1}\lambda^{-1} m^{-1} t^{-1} + O(t); && (562)\\
\hat{Q}_{p_i q_i} &= -3T^{-1}\lambda^{-1} t^{-2} + O(1); && (563)\\
\hat{Q}_{q_i q_i} &= 6T^{-1}\lambda^{-1} m t^{-3} + O(t^{-1}); && (564)\\
\hat{Q}_{p_i p_{j\neq i}} &= O(t^{2|j-i|-1}); && (565)\\
\hat{Q}_{p_i q_{j\neq i}} &= O(t^{2|j-i|-2}); && (566)\\
\hat{Q}_{q_i q_{j\neq i}} &= O(t^{2|j-i|-3}). && (567)
\end{aligned}$$

In the end, provided that $t$ is small enough, we have proved that $\hat{Q}(\eta)/\|\eta\|^2 \leqslant 6T^{-1}\lambda m t^{-3} + O(t^{-1}) < \infty$.

[¶]Recall that $\hat{C}$, as a covariance matrix, is symmetric.

Actually we have proved more than that: not only we have a bound on the operator norm of $\bar{Q}$, but we have bounded it entry-wise. More precisely, expanding the $O(*)$, we find that provided $t$ is small enough, there exists constants $A < \infty$ and $\gamma > 0$ such that for all $i, j \in \mathbb{Z}$,

$$\underbrace{\hat{Q}_{p_i p_j}, \hat{Q}_{p_i q_j}, \ldots, \hat{Q}_{q'_i q'_j}}_{\text{all 16 possibilities}} \leqslant A e^{-\gamma |j-i|}. \tag{568}$$

**5.4.5 Notation.** From now on we denote the basic variables $p_i, q_i, p'_i, q'_i$ of our system by $X_i$, $i \in \mathbb{Z}^{\uplus 4}$.

Now the question is: for $i \neq j \in \mathbb{Z}^{\uplus 4}$, $K \subset \mathbb{Z}^{\uplus 4} \smallsetminus \{i, j\}$, what is the value of $\{X_i : X_j\}_{\vec{X}_K}$? By the properties of Gaussian variables [Theorem 1.2.6], the answer is the following. Let $\bar{Q}_{|_{\mathbb{Z}^{\uplus 4} \smallsetminus K}}$ be the restriction of $\bar{Q}$ to indexes in $(\mathbb{Z}^{\uplus 4} \smallsetminus K)$. Since $r\mathbf{I} \leqslant \bar{Q} \leqslant R\mathbf{I}$, the same holds for $\bar{Q}_{|_{\mathbb{Z}^{\uplus 4} \smallsetminus K}}$, so this matrix is invertible; denote by $\bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}}$ its inverse. This matrix is the covariance matrix of (the centered version of) $\vec{X}_{\mathbb{Z}^{\uplus 4} \smallsetminus K}$ under some fixed value for $\vec{X}_K$; thus:

$$\{X_i : X_j\}_{\vec{X}_K} = \frac{\left|\bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}}_{ij}\right|}{\sqrt{\bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}}_{ii} \bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}}_{jj}}}. \tag{569}$$

It remains to control the entries of $\bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}}$, uniformly in $K$. We need two types of control: first an exponential control when $i$ is far away from $j$, then a non-trivial control for the values of $i$ and $j$ corresponding to close (or even identical) atoms.

Let us start with the first one. $\bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}}_{ii}$ and $\bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}}_{jj}$ are bounded below by $R^{-1}$, so we just have to bound above $\bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}}_{ij}$. This is achieved by a direct use of Lemma 5.5.1 in appendix.

Concerning the uniform non-trivial control, since $r\mathbf{I} \leqslant \bar{Q} \leqslant R\mathbf{I}$ one has $r\mathbf{I} \leqslant \bar{Q}_{|_{\mathbb{Z}^{\uplus 4} \smallsetminus K}} \leqslant R\mathbf{I}$, hence $R^{-1}\mathbf{I} \leqslant \bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}} \leqslant r^{-1}\mathbf{I}$, hence $R^{-1}\mathbf{I} \leqslant (\bar{C}^{|^{\mathbb{Z}^{\uplus 4} \smallsetminus K}})_{|_{\{i,j\}^2}} \leqslant r^{-1}\mathbf{I}$; from this and (569),

$$\forall i \neq j \in I \qquad \{X_i : X_j\}_* \leqslant \frac{R - r}{R + r} < 1. \tag{570}$$

□

From Claim 5.4.3, we get the main result of this subsection:

**5.4.6 Theorem.** *For the model of Definition 5.4.1, for all $t > 0$, $\{\eta, \eta'\} < 1$.*

*Proof.* First, if $t$ is small enough so that Claim 5.4.3 holds, direct application of Lemma 3.6.8 proves the result, as the $\{X_i, X_j\}_*$ are summable (since they decrease exponentially) and they all are $< 1$.

Now for larger $t$, fix some $0 < t_1 < t$ so that Claim 5.4.3 holds for $t_1$. Then we notice that $\eta \to \eta(t_1) \to \eta'$ is a Markov chain (with "$\eta(t_1)$" standing for "$(\vec{p}(t_1), \vec{q}(t_1))$"), so by Proposition 1.1.13, $\{\eta, \eta'\} \leqslant \{\eta, \eta(t_1)\} < 1$. □

## 5.5 Appendix: Inverses of 'nearly diagonal' matrices

The goal of this appendix is to state and prove a few lemmas sharing the same spirit: "*if a matrix is 'nearly diagonal', then it shall be invertible and its inverse shall also be 'nearly diagonal' with the same type of decay*".

### 5.5.a Matrices with exponential decay

The goal of this subsection is to prove the following

**5.5.1 Lemma.** *Let $I \subset \mathbb{Z}$ and let $((M_{ij}))_{(i,j)\in I^2}$ be a matrix. Assume that, when seen as a quadratic form on $L^2(I)$, one has $r\mathbf{I} \leqslant M \leqslant R\mathbf{I}$ for $0 < r \leqslant R < \infty$ —in particular, $M$ is invertible. Assume moreover that there exists constants $A < \infty$ and $\gamma > 0$ such that for all $i, j \in I$, $|M_{ij}| \leqslant Ae^{-\gamma|j-i|}$.*

*Then there exist constants $A' < \infty$ and $\gamma' > 0$ which are explicit functions of $r, R, \gamma, A$ (so* they do not depend on *$I$), such that one has the following control on the entries of $M^{-1}$:*

$$\forall i, j \in I \qquad (M^{-1})_{ij} \leqslant A' e^{-\gamma'|j-i|}. \tag{571}$$

*Proof.* Up to multiplying by a scalar, one can assume that $R = 1$. Then $M$ writes $M = \mathbf{I} - H$, where $0 \leqslant H \leqslant (1-r)\mathbf{I}$; since $H$ is symmetric, that inequality means that $|||H||| \leqslant 1 - r < 1$. Therefore, for all $k \in \mathbb{N}$ one has $|||H^k||| \leqslant (1-r)^k$, which allows us to write $M^{-1}$ as a series expansion:

$$M^{-1} = \sum_{k=0}^{\infty} H^k. \tag{572}$$

Up to replacing $A$ by $A+1$, we have the same entry-wise control on $H$ as on $M$. Then one sees by induction that for all $k \in \mathbb{N}$,

$$\forall i, j \in I \qquad |(H^k)_{ij}| \leqslant A_1^k e^{-\gamma_1|j-i|}, \tag{573}$$

where $\gamma_1$ is an arbitrary parameter in $(0, \gamma)$ and

$$A_1 := \sum_{z\in\mathbb{Z}} Ae^{-\gamma|z|+\gamma_1 z} = \frac{(1-e^{-2\gamma})A}{(1-e^{-(\gamma-\gamma_1)})(1-e^{-(\gamma+\gamma_1)})} \tag{574}$$

—observe that $I$ does not appear in the expression of $A_1$. Since $A_1$ is greater than 1, (573) is not enough to get an entry-wise control on $M^{-1}$. But now observe that the bound $|||H^k||| \leqslant (1-r)^k$ implies that all the $(H^k)_{ij}$ are bounded by $(1-r)^k$ in absolute value; thus:

$$|(M^{-1})_{ij}| \leqslant \sum_{k=0}^{\infty} (e^{-\gamma_1|j-i|}A_1^k \wedge (1-r)^k) \leqslant \Big(\frac{A_1}{A_1-1} + \frac{1}{r}\Big) \exp\Big(-\frac{|\log(1-r)|\gamma_1}{|\log(1-r)| + \log A_1}|j-i|\Big), \tag{575}$$

from which you read suitable values for $A'$ and $\gamma'$. $\square$

### 5.5.b Convolution inverses of rapidly decreasing functions

☛ *In all this subsection, we work on $\mathbb{Z}^n$ for some $n \in \mathbb{N}^*$; $\mathbb{R}^n$ is endowed with some fixed norm $|\cdot|$.*

*5.5.2 Remark.* Here I will deal with fonctions on $\mathbb{Z}^n$, but the results of this subsection could also be tranposed for functions on $\mathbb{R}^n$.

**5.5.3 Definition.** If $a : \mathbb{Z}^n \to \mathbb{R}$ is some integrable function with $\|a\|_{l^1} < 1$, we define

$$B[a] = a + a * a + a * a * a + \cdots , \tag{576}$$

which is the sum of a convergent series in $l^1(\mathbb{Z}^n)$. $B[a]$ is the function $b \in l^1(\mathbb{Z}^n)$ characterized by:

$$(\delta_0 - a) * (\delta_0 + b) = \delta_0. \tag{577}$$

**5.5.4 Definition.** A function $a : \mathbb{Z}^n \to \mathbb{R}$ is said to have *exponential decay* if there exists some $\beta > 0$ such that, for all $\beta' < \beta$, $a(z) = O(e^{-\beta'|z|})$ when $|z| \to \infty$. The minimal $\beta$ satisfying that property is called the *(exponential) rate of decay* of $a$.

**5.5.5 Lemma.** *Let $a \in l^1(\mathbb{Z}^n)$ with $\|a\|_{l^1} < 1$. If $a$ has exponential decay, then so does $B[a]$.*

*Proof.* Denoting by $|a|$ the function defined by $|a|(z) = |a(z)|$, it is clear by (576) that

$$\forall z \in \mathbb{Z}^n \qquad \big|B[a](z)\big| \leqslant B\big[|a|\big](z), \tag{578}$$

therefore it suffices to prove the case where $a$ is nonnegative. In that case, $B[a]$ will also be nonnegative.

Let $(\mathbb{R}^n)^*$ denote the dual space of $\mathbb{R}^n$, endowed with the dual norm

$$\forall \lambda \in (\mathbb{R}^n)^* \qquad |\lambda|_* = \sup_{\substack{z \in \mathbb{R}^n \\ |z|=1}} |\langle \lambda, z \rangle|. \tag{579}$$

For a nonnegative function $a$, we define its Laplace transform $\mathscr{L}\{a\}$: $(\mathbb{R}^n)^* \to \mathbb{R}_+ \cup \{+\infty\}$ by

$$\mathscr{L}\{a\}(\lambda) = \sum_{z \in \mathbb{Z}^n} e^{\langle \lambda, z \rangle} a(z). \tag{580}$$

Then, saying that $a$ has exponential decay with rate $\gamma$ is equivalent to saying that, for all $\lambda \in (\mathbb{R}^n)^*$ with $|\lambda|_* < \gamma$, $\mathscr{L}\{a\}(\lambda)$ is finite.

Since Laplace transform is linear and turns convolution into ordinary product, (576) yields, for all $\lambda \in (\mathbb{R}^n)^*$:

$$\mathscr{L}\{B[a]\}(\lambda) = \mathscr{L}\{a\}(\lambda) + \mathscr{L}\{a\}(\lambda)^2 + \mathscr{L}\{a\}(\lambda)^3 + \cdots , \tag{581}$$

which converges if and only if $\mathscr{L}\{a\}(\lambda) < 1$.

Now, since $a$ is nonnegative, by (580) the function $\mathscr{L}\{a\}$ is convex, so it is continuous on the interior of the domain where it is finite. By the exponential decay hypothesis, that domain contains a neighbourhood of 0, so $\mathscr{L}\{a\}$ is continuous at 0. And since $\mathscr{L}\{a\}(0) = \sum_{z \in \mathbb{Z}^n} a(z) = \|a\|_{l^1} < 1$, there is a neighbourhood of 0 on which $\mathscr{L}\{a\} < 1$ and thus $\mathscr{L}\{B[a]\} < \infty$. This implies that $B[a]$ has exponential decay. □

*5.5.6 Remark.* This proof also shows that (for nonnegative $a$) the rate of decay of $B[a]$ will never be greater than the rate of decay of $a$. In general, it is even strictly smaller, since all the values of $\lambda$ for which $1 \leqslant \mathscr{L}\{a\}(\lambda) < \infty$ yield a finite Laplace tranform for $a$ but an infinite one for $B[a]$. For example, take $n = 1$ and $a = e^{-1}\delta_1$, which has exponential decay with infinite rate since it is compactly supported; then the $k$-th convolution power of $a$ is $a^{*k} = e^{-k}\delta_k$, so that $B[a]$ is the function

$$B[a](z) = \mathbf{1}_{z>0} e^{-z}, \tag{582}$$

which also has exponential decay, but with rate 1 only.

**5.5.7 Lemma.** *If $\|a\|_{l^1(\mathbb{Z}^n)} < 1$ and $a(z) = O(1/|z|^\alpha)$ when $|z| \to \infty$ for some $\alpha > n$, then $B[a](z) = O(1/|z|^\alpha)$ when $|z| \to \infty$.*

*Proof.* Let $a$ satisfy the assumptions of the lemma for some $\alpha$. Like in the proof of Lemma 5.5.7, we can assume that $a$ is nonnegative. For $d > 0$, we define the function $\varphi_d\colon\ \mathbb{Z}^n \to \mathbb{R}$ by:

$$\varphi_d(z) := 1/(|z| \wedge d)^\alpha, \tag{583}$$

which is in $l^1(\mathbb{Z}^n)$ since $\alpha > n$. Then the key claim is the following sub-lemma, whose proof is postponed:

**5.5.8 Lemma.** *Under the assumptions of Lemma 5.5.7, there exists some $\rho < 1$ and some $d \in (0,\infty)$ such that, pointwise,*

$$\varphi_d * a \leqslant \rho\varphi_d. \tag{584}$$

Admitting Lemma 5.5.8, take $\rho$ and $d$ such that (584) is satisfied. The assumption on $a$ implies that there exists some $C < \infty$ such that $a \leqslant C\varphi_d$; therefore by (584) one also has $a * a \leqslant C\varphi_d * a \leqslant \rho C\varphi_d$, whence by (584) again $a * a * a \leqslant \rho C\varphi_d * a \leqslant \rho^2 C\varphi_d$, etc.. In the end,

$$B[a] \leqslant C\varphi_d + \rho C\varphi_d + \rho^2 C\varphi_d + \cdots \leqslant \frac{C}{1-\rho}\varphi_d, \tag{585}$$

which implies that $B[a](z) = O(1/|z|^\alpha)$. □

*Proof of Lemma 5.5.8.* Denote $S := \|a\|_{l^1}$, which by hypothesis is $< 1$, and fix $\varepsilon \in (0,1/2)$ such that $(1-\varepsilon)^\alpha > S$. Let $d \in (0,\infty)$, devised to be quite large; our goal is to bound above $(\varphi_d * a)(z)$ for all $z \in \mathbb{Z}^n$. Since $\varphi_d$ is bounded above by $d^{-\alpha}$, one has obviously for all $z \in \mathbb{Z}^n$:

$$(\varphi_d * a)(z) \leqslant d^{-\alpha} \sum_{z \in \mathbb{Z}^n} a(z) = S d^{-\alpha}, \tag{586}$$

whence $(\varphi_d * a)(z) \leqslant S\varphi_d(z)$ for all $z$ with $|z| \leqslant d$. Since $S < 1$, the claim is therefore okay for $|z| \leqslant d$.

Now, let $z \in \mathbb{Z}^n$ with $|z| > d$. We have to bound above

$$(\varphi_d * a)(z) = \sum_{\substack{x,y \in \mathbb{Z}^n \\ x+y=z}} \varphi_d(x) a(y). \tag{587}$$

We decompose that sum into three pieces:

$$(\varphi_d * a)(z) = \sum_{|y| \leqslant \varepsilon|z|} \varphi_d(z-y) a_d(y) + \sum_{\substack{|x|>\varepsilon|z| \\ |z-x|>\varepsilon|z|}} \varphi_d(x) a_d(z-x) + \sum_{|x| \leqslant \varepsilon|z|} \varphi_d(x) a_d(z-x), \tag{588}$$

which we shorthand into "①+②+③".

We bound these three terms separately. For ①, we observe that for $|y| \leqslant \varepsilon|z|$, $|z-y| \geqslant (1-\varepsilon)|z|$ by the triangle inequality, thus $\varphi_d(z-y) \leqslant \big((1-\varepsilon)|z|\big)^{-\alpha} = (1-\varepsilon)^{-\alpha}\varphi_d(z)$, whence by summing:

$$① \leqslant (1-\varepsilon)^{-\alpha}\varphi_d(z) \sum_{|y|\leqslant\varepsilon|z|} a(y) \leqslant (1-\varepsilon)^{-\alpha} S\varphi_d(z). \tag{589}$$

Similarly, for $|x| \leqslant \varepsilon|z|$, $C$ denoting a constant such that $a \leqslant C\varphi_d$, one has $a(z-x) \leqslant C\big((1-\varepsilon)|z|\big)^{-\alpha}$, thus:

$$③ \leqslant (1-\varepsilon)^{-\alpha} C \|\varphi_d\|_{l^1}\, \varphi_d(z). \tag{590}$$

Of course, $\|\varphi_d\|_{l^1}$ depends on $d$; the important point is that, by dominated convergence, $\|\varphi_d\|_{l^1} \to 0$ when $d \to \infty$.

Finally, provided $d$ is large enough, Term ② will be well approximated by an integral:

$$② \leqslant \sum_{\substack{x\in\mathbb{Z}^n \\ |x|,|z-x|>\varepsilon|z|}} \frac{1}{|x|^\alpha} \times \frac{C}{|z-x|^\alpha} \simeq \int_{\substack{x\in\mathbb{R}^n \\ |x|,|z-x|>\varepsilon|z|}} \frac{C}{|x|^\alpha\,|z-x|^\alpha}\, dx, \tag{591}$$

where "$\simeq$" means that the ratio between the quantites at each side of that symbol can be made arbitrarily close to 1 when $d \to \infty$, uniformly in $z$. Indeed, the difference between the sum and the integral is due to two causes: first, approximating the integral on a unit square of $\mathbb{R}^n$ by the value of the integrand at the center of this square, second, summing (or not summing) terms of the discrete sum corresponding to squares that are not entirely in the domain of the integral. For the first cause, on the domain of the integral, $C/(|x|^\alpha|z-x|^\alpha)$ varies of at most $O(1/|z|)$ in relative value on all the unit squares. For the second cause, the border of the domain of the integral is made of two $(n-1)$-dimensional spheres of radius $\varepsilon|z|$, so it crosses $O(|z|^{n-1})$ unit squares. Since $C/(|x||z-x|)^\alpha$ is bounded by $C(\varepsilon(1-\varepsilon))^{-\alpha}|z|^{-2\alpha}$ on the domain of the integral, the (absolute) error due to boundary squares is at most $O(|z|^{n-1-2\alpha})$. As the integral itself is proportional to $|z|^{n-2\alpha}$ [cf. the change of variables below], the relative error due to boundary squares is at most $O(1/|z|)$ too, and $O(1/|z|) = o(1)$ since $|z| > \varepsilon d$.

Making the change of variables $x = |z|x'$, (591) becomes:

$$② \lesssim C|z|^{n-2\alpha} \int_{|x'|,|1-x'|>\varepsilon} \frac{1}{|x'|^\alpha|1-x'|^\alpha}\, dx', \tag{592}$$

which I shorthand into "$② \lesssim \mathscr{I}C|z|^{n-2\alpha}$". Since $|z| > d$ and $\alpha > n$, that bound implies:

$$② \lesssim \frac{\mathscr{I}C}{d^{\alpha-n}}\, \varphi_d(z). \tag{593}$$

Combining (589), (590) and (593), one finally gets that when $d \to \infty$, for all $|z| \geqslant d$,

$$(\varphi_d * a)(z) \leqslant \rho(d)\,\varphi_d(z), \tag{594}$$

with

$$\rho(d) = (1-\varepsilon)^{-\alpha}(S + C\|a\|_{l^1}) + (1+o(1))\mathscr{I}C/d^{\alpha-n}. \tag{595}$$

$\rho(d)$ tends to $(1-\varepsilon)^{-\alpha}S < 1$ when $d \to \infty$, so it is $< 1$ provided $d$ is large enough, what we wanted. □

# Bibliography


[1] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[2] M. Aizenmann, D. Barsky, and R. Fernández. The phase transition in a general class of Ising-type models is sharp. *J. Stat. Phys.*, 47:343–374, 1987.

[3] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.

[4] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[5] R. C. Bradley. Equivalent measures of dependence. *J. Multivariate Anal.*, 13(1):167–176, 1983.

[6] R. C. Bradley. On the spectral density and asymptotic normality of weakly dependent random fields. *J. Theoret. Probab.*, 5(2):355–373, 1992.

[7] R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144 (electronic), 2005. Update of, and a supplement to, the 1986 original.

[8] R. C. Bradley. *Introduction to strong mixing conditions. Vol. 1–3*. Kendrick Press, Heber City, UT, 2007.

[9] W. Bryc. Conditional moment representations for dependent random variables. *Electron. J. Probab.*, 1:no. 7, approx. 14 pp. (electronic), 1996.

[10] A. V. Bulinskiĭ. Measures of dependence that are close to the maximum correlation coefficient. *Soviet Math. Dokl.*, 30(1):249–252, 1984.

[11] T. Champion, L. De Pascale, and P. Juutinen. The $\infty$-Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.*, 40(1):1–20, 2008.

[12] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[13] P. Csáki and J. Fischer. On the general notion of maximal correlation. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8:27–51, 1963.

[14] J. Dedecker. A central limit theorem for stationary random fields. *Probab. Theory Related Fields*, 110(3):397–426, 1998.

[15] R. L. Dobrushin and S. Shlosman. Completely analytical interactions: constructive description. *J. Statist. Phys.*, 46(5-6):983–1014, 1987.

[16] R. 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.

[17] R. J. Gardner. The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, 39(3):355–405 (electronic), 2002.

[18] É. Ghys and P. de la Harpe. Infinite groups as geometric objects (after Gromov). In *Ergodic theory, symbolic dynamics, and hyperbolic spaces (Trieste, 1989)*, Oxford Sci. Publ., pages 299–314. Oxford Univ. Press, New York, 1991.

[19] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[20] R. J. Glauber. Time-dependent statistics of the Ising model. *J. Mathematical Phys.*, 4:294–307, 1963.

[21] G. Grimmett. *The random-cluster model*, volume 333 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2006.

[22] H. W. Guggenheimer. *Differential geometry*. Dover Publications Inc., New York, 1977. Corrected reprint of the 1963 edition, Dover Books on Advanced Mathematics.

[23] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[24] I. A. Ibragimov. A note on the central limit theorems for dependent random variables. *Theory of Probability and its Applications*, 20(1):135–141, 1975.

[25] A. Ioana. Cocycle superrigidity for profinite actions of property (T) groups. Submitted; preprint available online at `http://arxiv.org/abs/0805.2998`, 2008.

[26] A. N. Kolmogorov and Y. A. Rozanov. On strong mixing conditions for stationary gaussian processes. *Theory of Probability and its Applications*, 5(2):204–208, 1960.

[27] H. O. Lancaster. Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, 44:289–292, 1957.

[28] T. M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1985.

[29] F. 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, Berlin, 1999.

[30] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Comm. Math. Phys.*, 161(3):447–486, 1994.

[31] F. Martinelli, E. Olivieri, and R. H. Schonmann. For 2-D lattice spin systems weak mixing implies strong mixing. *Comm. Math. Phys.*, 165(1):33–47, 1994.

[32] R. Peierls. On Ising's model of ferromagnetism. *Proc. Cambridge Philos. Soc.*, 32:477–481, 1936.

[33] M. Peligrad. On the central limit theorem for $\rho$-mixing sequences of random variables. *Ann. Probab.*, 15(4):1387–1394, 1987.

[34] D. B. Percival and A. T. Walden. *Spectral analysis for physical applications*. Cambridge University Press, Cambridge, 1993. Multitaper and conventional univariate techniques.

[35] R. Peyre. Le théorème « $\mathbb{Z}$ contre $\mathbb{Z}$ » est optimal partout. Personal notes.

[36] R. Peyre. Pas de règle de chaîne pour les corrélations maximales, même dans un cas très gentil. Personal notes.

[37] R. Peyre. Hilbertian decorrelations. `arXiv:1004.1602v1`, 2010.

[38] R. A. Rankin. *The modular group and its subgroups*. The Ramanujan Institute, Madras, 1969.

[39] A. Robert. Functional analysis and NSA. In *Developments in nonstandard mathematics (Aveiro, 1994)*, volume 336 of *Pitman Res. Notes Math. Ser.*, pages 73–90. Longman, Harlow, 1995.

[40] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.*, 42:43–47, 1956.

[41] M. Rosenblatt. *Markov processes. Structure and asymptotic behavior*. Springer-Verlag, New York, 1971. Die Grundlehren der mathematischen Wissenschaften, Band 184.

[42] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[43] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[44] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

[45] J. Weidmann. *Linear operators in Hilbert spaces*, volume 68 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1980. Translated from the German by Joseph Szücs.

[46] C. S. Withers. Central limit theorems for dependent variables. I. *Z. Wahrsch. Verw. Gebiete*, 57(4):509–534, 1981.

[47] H. S. Witsenhausen. On sequences of pairs of dependent random variables. *SIAM J. Appl. Math.*, 28:100–113, 1975.