Chapter 4

Coupling

4.1 Background

To be written. See [dH, Section 2].

4.1.1 ▷ *Erdös-Rényi graphs: degree sequence*

To be written. See [vdH14, Section 5.3].

4.1.2 ▷ *Harmonic functions: lattices and trees*

To be written. See [Per, Section 6].*

4.2 Stochastic domination and correlation inequalities

It can be useful to compare the distributions of two random variables. For instance let $(X_i)_{i=1}^n$ be independent $\mathbb{Z}_+$-valued random variables with $\mathbb{P}[X_i \geq 1] \geq p$ and consider their sum $S = \sum_{i=1}^n X_i$. If $S_* \sim \mathrm{Bin}(n, p)$, then it is intuitively clear that one can extract information about S by studying S_* instead—which may be easier. Indeed, in some sense, S "dominates" S_* and one would expect that, say, $\mathbb{P}[S > x] \geq \mathbb{P}[S_* > x]$ among other relations. Coupling provides a way to make this intuition formal, as we detail in this section.

In particular we study an important special case known as positive association. In that case a measure "dominates itself" in the following sense: conditioning on certain events makes other events more likely. This concept, which is formalized in Section 4.2.4, has numerous applications in discrete probability.

*Requires: Section 3.3.1.

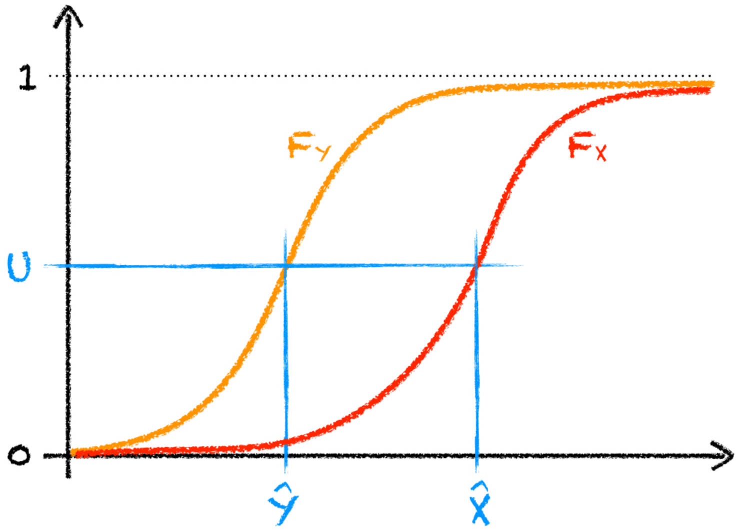

Figure 4.1: The law of X, represented here by its cumulative distribution function F_X in red, stochastically dominates the law of Y, in orange. The construction of a monotone coupling, $(\hat{X}, \hat{Y}) := (F_X^{-1}(U), F_Y^{-1}(U))$ where U is uniform in $[0, 1]$, is also depicted.

4.2.1 Definitions

We start with the case of real random variables.

Ordering of real random variables For real random variables, stochastic domination is defined as follows. See Figure 4.1 for an illustration.

Definition 4.1 (Stochastic domination)**.** *Let μ and ν be probability measures on $\mathbb{R}$. The measure μ is said to* stochastically dominate *ν, denoted $\mu \succeq \nu$, if for all $x \in \mathbb{R}$*

stochastic domination

$$\mu\big[(x, +\infty)\big] \geq \nu\big[(x, +\infty)\big].$$

A real random variable X stochastically dominates *Y, denoted by $X \succeq Y$, if the law of X dominates the law of Y.*

Example 4.2 (Bernoulli vs. Poisson)**.** Let $X \sim \mathrm{Poi}(\lambda)$ be Poisson with mean $\lambda > 0$ and let Y be a Bernoulli trial with success probability $p \in (0, 1)$, i.e., $\mathbb{P}[Y = 1] = 1 - \mathbb{P}[Y = 0] = p$. In order for X to stochastically dominate Y, it suffices to have

$$\mathbb{P}[X > \ell] \geq \mathbb{P}[Y > \ell], \qquad \forall \ell \geq 0.$$

This is always true for $\ell \geq 1$ since $\mathbb{P}[X > \ell] > 0$ but $\mathbb{P}[Y > \ell] = 0$. So it remains to consider the case $\ell = 0$. We have

$$1 - e^{-\lambda} = \mathbb{P}[X > 0] \geq \mathbb{P}[Y > 0] = p,$$

if and only if

$$\lambda \geq -\log(1-p).$$

◀

Note that stochastic domination does not require X and Y to be defined on the same probability space. The connection to coupling arises from the following characterization.

Theorem 4.3 (Coupling and stochastic domination)**.** *The real random variable X stochastically dominates Y if and only if there is a coupling $(\hat{X}, \hat{Y})$ of X and Y such that*

$$\mathbb{P}[\hat{X} \geq \hat{Y}] = 1. \tag{4.1}$$

We refer to $(\hat{X}, \hat{Y})$ as a monotone coupling *of X and Y.*

monotone coupling

Proof. One direction is clear. Suppose there is such a coupling. Then for all $x \in \mathbb{R}$

$$\mathbb{P}[Y > x] = \mathbb{P}[\hat{Y} > x] = \mathbb{P}[\hat{X} \geq \hat{Y} > x] \leq \mathbb{P}[\hat{X} > x] = \mathbb{P}[X > x].$$

For the other direction, define the cumulative distribution functions $F_X(x) = \mathbb{P}[X \leq x]$ and $F_Y(x) = \mathbb{P}[Y \leq x]$. Assume $X \succeq X'$. The idea of the proof is to use the following standard way of generating a real random variable. Recall (e.g. [Dur10, Section 1.2]) that

$$X \stackrel{\mathrm{d}}{=} F_X^{-1}(U), \tag{4.2}$$

where U is a $[0,1]$-valued uniform random variable and

$$F_X^{-1}(u) := \inf\{x \in \mathbb{R} \,:\, F_X(x) \geq u\},$$

is a generalized inverse. Indeed $F_X^{-1}(u) > x$ precisely when $u > F_X(x)$. It is natural to construct a coupling of X and Y by simply using the same uniform random variable U in this representation, i.e., we define $\hat{X} = F_X^{-1}(U)$ and $\hat{Y} = F_Y^{-1}(U)$. See Figure 4.1. By (4.2), this is a coupling of X and Y. It remains to check (4.1). Because $F_X(x) \leq F_Y(x)$ for all x by definition of stochastic domination, by the definition of the generalized inverse,

$$\mathbb{P}[\hat{X} \geq \hat{Y}] = \mathbb{P}[F_X^{-1}(U) \geq F_Y^{-1}(U)] = 1,$$

as required. ■

Example 4.4. Returning to the example in the first paragraph of Section 4.2, let $(X_i)_{i=1}^n$ be independent $\mathbb{Z}_+$-valued random variables with $\mathbb{P}[X_i \geq 1] \geq p$ and consider their sum $S := \sum_{i=1}^n X_i$. Further let $S_* \sim \mathrm{Bin}(n, p)$. Write S_* as the sum $\sum_{i=1}^n Y_i$ where (Y_i) are independent $\{0,1\}$-variables with $\mathbb{P}[Y_i = 1] = p$. To couple S and S_*, first set $(\hat{Y}_i) := (Y_i)$ and $\hat{S}_* := \sum_{i=1}^n \hat{Y}_i$. Let $\hat{X}_i$ be 0 whenever $\hat{Y}_i = 0$. Otherwise, i.e. if $\hat{Y}_i = 1$, generate $\hat{X}_i$ according to the distribution of X_i conditioned on $\{X_i \geq 1\}$, independently of everything else. By construction $\hat{X}_i \geq \hat{Y}_i$ a.s. for all i and as a result $\sum_{i=1}^n \hat{X}_i =: \hat{S} \geq \hat{S}_*$ a.s. or $S \succeq S_*$ by the previous theorem. That implies for instance that $\mathbb{P}[S > x] \geq \mathbb{P}[S_* > x]$ as we claimed earlier. A special case of this argument gives the following useful fact about binomials

$$n \geq m,\ q \geq p \implies \mathrm{Bin}(n, q) \succeq \mathrm{Bin}(m, p).$$

◀

Example 4.5 (Poisson distribution)**.** Let $X \sim \mathrm{Poi}(\mu)$ and $Y \sim \mathrm{Poi}(\nu)$ with $\mu > \nu$. Recall that a sum of independent Poisson is Poisson (use moment-generating functions or see e.g. [Dur10, Exercise 2.1.14]). This fact leads to a natural coupling: let $\hat{Y} \sim \mathrm{Poi}(\nu)$, $\hat{Z} \sim \mathrm{Poi}(\mu - \nu)$ independently of Y, and $\hat{X} = \hat{Y} + \hat{Z}$. Then $(\hat{X}, \hat{Y})$ is a coupling and $\hat{X} \geq \hat{Y}$ a.s. because $\hat{Z} \geq 0$. Hence $X \succeq Y$. ◀

We record two useful consequences of Theorem 4.3.

Corollary 4.6. *Let X and Y be real random variables with $X \succeq Y$ and let $f : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function. Then $f(X) \succeq f(Y)$ and furthermore, provided $\mathbb{E}|f(X)|, \mathbb{E}|f(Y)| < +\infty$, we have that*

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)].$$

Proof. Let $(\hat{X}, \hat{Y})$ be the monotone coupling of X and Y whose existence is guaranteed by Theorem 4.3. Then $f(\hat{X}) \geq f(\hat{Y})$ a.s. so that, provided the expectations exist,

$$\mathbb{E}[f(X)] = \mathbb{E}[f(\hat{X})] \geq \mathbb{E}[f(\hat{Y})] = \mathbb{E}[f(Y)],$$

and furthermore $(f(\hat{X}), f(\hat{Y}))$ is a monotone coupling of $f(X)$ and $f(Y)$. Hence $f(X) \succeq f(Y)$. ■

Corollary 4.7. *Let X_1, X_2 be independent random variables. Let Y_1, Y_2 be independent random variables such that $X_i \succeq Y_i$, $i = 1, 2$. Then*

$$X_1 + X_2 \succeq Y_1 + Y_2.$$

Proof. Let $(\hat{X}_1, \hat{Y}_1)$ and $(\hat{X}_2, \hat{Y}_2)$ be independent, monotone couplings of (X_1, Y_1) and (X_2, Y_2) (on the same probability space). Then

$$X_1 + X_2 \sim \hat{X}_1 + \hat{X}_2 \geq \hat{Y}_1 + \hat{Y}_2 \sim Y_1 + Y_2.$$

■

Example 4.8 (Binomial vs. Poisson)**.** A sum of n Poisson variables with mean λ is $\mathrm{Poi}(n\lambda)$. A sum of n Bernoulli trials with success probability p is $\mathrm{Bin}(n,p)$. Using Example 4.2 and Corollary 4.7, we get

$$\lambda \geq -\log(1-p) \quad \Longrightarrow \quad \mathrm{Poi}(n\lambda) \succeq \mathrm{Bin}(n,p). \tag{4.3}$$

The following special case will be useful later. Let $0 < \Lambda < 1$ and let m be an integer. Then

$$\frac{\Lambda}{m-1} \geq \frac{\Lambda}{m-\Lambda} = \frac{m}{m-\Lambda} - 1 \geq \log\left(\frac{m}{m-\Lambda}\right) = -\log\left(1 - \frac{\Lambda}{m}\right),$$

where we used that $\log x \leq x - 1$ for all $x \in \mathbb{R}$. So, setting $\lambda := \frac{\Lambda}{m-1}$, $p := \frac{\Lambda}{m}$ and $n := m-1$ in (4.3), we get

$$\Lambda \in (0,1) \quad \Longrightarrow \quad \mathrm{Poi}(\Lambda) \succeq \mathrm{Bin}\left(m-1, \frac{\Lambda}{m}\right). \tag{4.4}$$

◀

Ordering on partially ordered sets The definition of stochastic domination hinges on the totally ordered nature of $\mathbb{R}$. It also extends naturally to posets. Let $(\mathcal{X}, \leq)$ be a *poset*, i.e., for all $x, y, z \in \mathcal{X}$: *poset*

- *[Reflexivity]* $x \leq x$,
- *[Antisymmetry]* if $x \leq y$ and $y \leq x$ then $x = y$,
- *[Transitivity]* if $x \leq y$ and $y \leq z$ then $x \leq z$.

For instance the set $\{0,1\}^F$ is a poset when equipped with the relation $\boldsymbol{x} \leq \boldsymbol{y}$ if and only if $x_i \leq y_i$ for all $i \in F$. Equivalently the subsets of F, denoted by 2^F, form a poset with the inclusion relation. (A totally ordered set satisfies in addition that, for any x, y, we have either $x \leq y$ or $y \leq x$.)

Let $\mathcal{F}$ be a σ-field over the poset $\mathcal{X}$. An event $A \in \mathcal{F}$ is *increasing* if $x \in A$ implies that any $y \geq x$ is also in A. A function $f : \mathcal{X} \to \mathbb{R}$ is *increasing* if $x \leq y$ implies $f(x) \leq f(y)$. *increasing event, function*

Definition 4.9 (Stochastic domination for posets)**.** *Let $(\mathcal{X}, \leq)$ be a poset and let $\mathcal{F}$ be a σ-field on $\mathcal{X}$. Let μ and ν be probability measures on $(\mathcal{X}, \mathcal{F})$. The measure μ is said to* stochastically dominate *ν, denoted by $\mu \succeq \nu$, if for all increasing $A \in \mathcal{F}$*

stochastic domination for posets

$$\mu(A) \geq \nu(A).$$

An $\mathcal{X}$-valued random variable X stochastically dominates *Y, denoted by $X \succeq Y$, if the law of X dominates the law of Y.*

As before, a *monotone coupling* $(\hat{X}, \hat{Y})$ of X and Y is one which satisfies $\hat{X} \geq \hat{Y}$ a.s.

monotone coupling for posets

Example 4.10 (Monotonicity of the percolation function)**.** We have already seen an example of stochastic domination in Section 2.1.5. Consider bond percolation on the d-dimensional lattice $\mathbb{L}^d$. Here the poset is the collection of all subsets of edges, specifying the open edges, with the inclusion relation. Recall that the percolation function is given by

$$\theta(p) := \mathbb{P}_p[|\mathcal{C}_0| = +\infty],$$

where $\mathcal{C}_0$ is the open cluster of the origin. We argued in Section 2.1.5 that $\theta(p)$ is non-decreasing by considering the following alternative representation of the percolation process under $\mathbb{P}_p$: to each edge e, assign a uniform $[0,1]$-valued random variable U_e and declare the edge open if $U_e \leq p$. Using the same U_es for two different p-values, $p_1 < p_2$, gives a monotone coupling of the processes for p_1 and p_2. It follows immediately that $\theta(p_1) \leq \theta(p_2)$, where we used that the event $\{|\mathcal{C}_0| = +\infty\}$ is increasing. ◀

The existence of a monotone coupling is perhaps more surprising for posets. We prove the result in the finite case only, which will be enough for our purposes.

Theorem 4.11 (Strassen's theorem)**.** *Let X and Y be random variables taking values in a* finite *poset $(\mathcal{X}, \leq)$ with the σ-field $\mathcal{F} = 2^{\mathcal{X}}$. Then $X \succeq Y$ if and only if there exists a monotone coupling $(\hat{X}, \hat{Y})$ of X and Y.*

Proof. One direction is clear. Suppose there is such a coupling. Then for all increasing A

$$\mathbb{P}[Y \in A] = \mathbb{P}[\hat{Y} \in A] = \mathbb{P}[\hat{X} \geq \hat{Y} \in A] \leq \mathbb{P}[\hat{X} \in A] = \mathbb{P}[X \in A].$$

The proof in the other direction relies on the max-flow min-cut theorem. To see the connection with flows, let μ_X and μ_Y be the laws of X and Y respectively, and denote by ν their joint distribution under the desired coupling. Noting that we want $\nu(x, y) = 0$ if $x \leq y$, the marginal conditions on the coupling read

$$\sum_{y \leq x} \nu(x, y) = \mu_X(x), \qquad \forall x \in \mathcal{X},$$

and

$$\sum_{x \geq y} \nu(x, y) = \mu_Y(y), \qquad \forall y \in \mathcal{X}.$$

These equations can be interpreted as flow-conservation constraints. Consider the following directed graph. There are two vertices, $(w, 1)$ and $(w, 2)$, for each element w in $\mathcal{X}$ with edges connecting each $(x, 1)$ to those $(y, 2)$s with $x \geq y$. These edges have capacity $+\infty$. In addition there is a source a and a sink z. The source has a directed edge of capacity $\mu_X(x)$ to $(x, 1)$ for each $x \in \mathcal{X}$ and, similarly, each $(y, 2)$ has a directed edge of capacity $\mu_Y(y)$ to the sink. The existence of a monotone coupling will follow once we show that there is a flow of strength 1 between a and z. Indeed, in that case, all edges from the source and all edges to the sink are at capacity. If we let $\nu(x, y)$ be the flow on edge $\langle (x, 1), (y, 2) \rangle$, the constraints above then impose the conservation of the flow on the vertices $(\mathcal{X} \times \{1\}) \cup (\mathcal{X} \times \{2\})$. Hence the flow between $\mathcal{X} \times \{1\}$ and $\mathcal{X} \times \{2\}$ yields the desired coupling. See Figure 4.2.

By the max-flow min-cut theorem, it suffices to show that a minimum cut has capacity 1. Such a cut is of course obtained by choosing all edges out of the source. So it remains to show that no cut has capacity less than 1. This is where we use the fact that $\mu_X(A) \geq \mu_Y(A)$ for all increasing A. Because the edges between $\mathcal{X} \times \{1\}$ and $\mathcal{X} \times \{2\}$ have infinite capacity, they cannot be used in a minimum cut. So we can restrict our attention to those cuts containing edges from a to $A_* \times \{1\}$ and from $Z_* \times \{2\}$ to z for subsets $A_*, Z_* \subseteq \mathcal{X}$. We must have

$$A_* \supseteq \{x \in \mathcal{X} \,:\, \exists y \in Z_*^c,\ x \geq y\},$$

to block all paths of the form $a \sim (x, 1) \sim (y, 2) \sim z$ with x and y as above. In fact, for a minimum cut, we further have

$$A_* = \{x \in \mathcal{X} \,:\, \exists y \in Z_*^c,\ x \geq y\},$$

as adding an x not satisfying this property is redundant. See Figure 4.2. In particular A_* is increasing: if $x_1 \in A_*$ and $x_2 \geq x_1$, then $\exists y \in Z_*^c$ such that $x_1 \geq y$ and, since $x_2 \geq x_1 \geq y$, the same y works for x_2. Observe that, because $y \geq y$, the set A_* also includes Z_*^c. If it were the case that $A_* \neq Z_*^c$, then we could construct a cut with lower or equal capacity by fixing A_* and setting $Z_* := A_*^c$: because A_* is

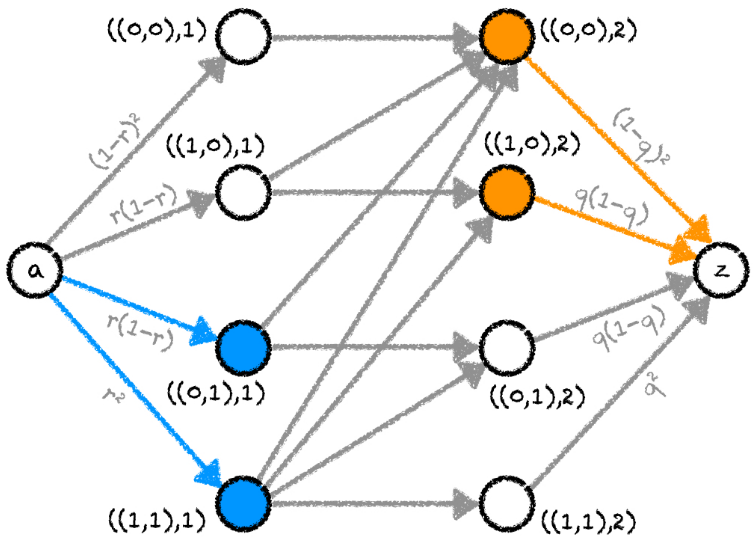

Figure 4.2: Construction of a monotone coupling through the max-flow representation for independent Bernoulli pairs with parameters r (on the left) and $q < r$ (on the right). Edge labels indicate capacity. Edges without labels have infinite capacity. The colored edges depict a suboptimal cut. The blue and orange vertices correspond respectively to the sets A_* and Z_* for this cut. The capacity of the cut is $r^2 + r(1-r) + (1-q)^2 + (1-q)q = r + (1-q) > r + (1-r) = 1$.

increasing, any $y \in A_* \cap Z_*$ is such that paths of the form $a \sim (x,1) \sim (y,2) \sim z$ with $x \geq y$ are cut by $x \in A_*$. Hence, for a minimum cut, we can assume that in fact $A_* = Z_*^c$. The capacity of the cut is

$$\mu_X(A_*) + \mu_Y(Z_*) = \mu_X(A_*) + 1 - \mu_Y(A_*) = 1 + (\mu_X(A_*) - \mu_Y(A_*)) \geq 1,$$

where the term in parenthesis is nonnegative by assumption and the fact that A_* is increasing. That concludes the proof. ■

Remark 4.12. *Strassen's theorem holds more generally on Polish spaces with a closed partial order. See e.g. [Lin02, Section IV.1.2] for the details.*

The proof of Corollary 4.6 immediately extends to:

Corollary 4.13. *Let X and Y be $\mathcal{X}$-valued random variables with $X \succeq Y$ and let $f : \mathcal{X} \to \mathbb{R}$ be an increasing function. Then $f(X) \succeq f(Y)$ and furthermore, provided $\mathbb{E}|f(X)|, \mathbb{E}|f(Y)| < +\infty$, we have that*

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)].$$

Ordering of Markov chains Stochastic domination also arises in the context of Markov chains. We begin with an example.

Example 4.14 (Lazier chain)**.** Consider a random walk (X_t) on the network $\mathcal{N} = ((V,E),c)$ where $V = \{0,1,\ldots,n\}$ and $i \sim j$ if and only $|i-j| \leq 1$ (including self-loops). Let $\mathcal{N}' = ((V,E),c')$ be a modified version of $\mathcal{N}$ on the same graph where for all i $c(i,i) \leq c'(i,i)$. That is, if (X'_t) is random walk on $\mathcal{N}'$, then (X'_t) is "lazier" than (X_t) in that it is more likely to stay put. Assume that both (X_t) and (X'_t) start at i_0 and define $M_s := \max_{t \leq s} X_t$ and $M'_s := \max_{t \leq s} X'_t$. Since (X'_t) "travels less" than (X_t) the following claim is intuitively obvious:

Claim 4.15.

$$M_s \succeq M'_s.$$

We prove this by producing a monotone coupling. First set $(\hat{X}_t) := (X_t)$. We then generate $(\hat{X}'_t)$ as a "sticky" version of $(\hat{X}_t)$. That is, $(\hat{X}'_t)$ follows exactly the same transitions as $(\hat{X}_t)$ (including the self-loops), but at each time it opts to stay where it currently is, say j, for an extra time step with probability

$$\frac{c'(j,j) - c(j,j)}{\sum_{i \sim j} c'(i,j)},$$

which is in $[0,1]$ by assumption. Marginally, $(\hat{X}'_t)$ is a random walk on $\mathcal{N}'$ because by construction

$$\frac{c'(j,j)}{\sum_{i\sim j} c'(i,j)} = \frac{c'(j,j)-c(j,j)}{\sum_{i\sim j} c'(i,j)} + \left(\frac{\sum_{i\sim j} c(i,j)}{\sum_{i\sim j} c'(i,j)}\right) \frac{c(j,j)}{\sum_{i\sim j} c(i,j)},$$

and for $i \neq j$ with $i \sim j$

$$\frac{c'(i,j)}{\sum_{i\sim j} c'(i,j)} = \left(\frac{\sum_{i\sim j} c(i,j)}{\sum_{i\sim j} c'(i,j)}\right) \frac{c(i,j)}{\sum_{i\sim j} c(i,j)},$$

since $c'(i,j) = c(i,j)$. This coupling satisfies

$$\widehat{M}_s := \max_{t\leq s} \hat{X}_t \geq \max_{t\leq s} \hat{X}'_t =: \widehat{M}'_s, \qquad \text{a.s.}$$

because $(\hat{X}'_t)_{t\leq s}$ visits a subset of the states visited by $(\hat{X}_t)_{t\leq s}$. In other words $(\widehat{M}_s, \widehat{M}'_s)$ is a monotone coupling of (M_s, M'_s) and this proves the claim. ◀

The previous example involved an asynchronous coupling of the chains. Often, a simpler step-by-step approach is possible.

Definition 4.16 (Stochastic domination for Markov kernels)**.** *Let P and Q be transition matrices on a finite or countable poset $(\mathcal{X}, \leq)$. The transition matrix Q is said to* stochastically dominate *the transition matrix P if*

stochastic domination for Markov chains

$$x \leq y \implies P(x,\cdot) \preceq Q(y,\cdot). \tag{4.5}$$

If the above condition is satisfied for $P = Q$, we say that P is stochastically monotone.

stochastic monotonicity

The equivalent of Strassen's theorem in this case is the following theorem, which we prove in the finite case only again.

Theorem 4.17 (Strassen's theorem for Markov kernels)**.** *Let (X_t) and (Y_t) be Markov chains on a finite poset $(\mathcal{X}, \leq)$ with transition matrices P and Q respectively. Assume that Q stochastically dominates P. Then for all $x_0 \leq y_0$ there is a coupling $(\hat{X}_t, \hat{Y}_t)$ of (X_t) started at x_0 and (Y_t) started at y_0 such that a.s.*

$$\hat{X}_t \leq \hat{Y}_t, \qquad \forall t.$$

Furthermore, if the chains are irreducible and have stationary distributions π and μ respectively, then $\pi \preceq \mu$.

Observe that, for a step-by-step monotone coupling to exist, it is not generally enough for the weaker condition $P(x,\cdot) \preceq Q(x,\cdot)$ to hold for all x, as should be clear from the proof. Also you should convince yourself that the chains in Example 4.14 do not in general satisfy (4.5). (Which pairs x, y cause problems?)

Proof of Theorem 4.17. Let

$$\mathcal{W} := \{(x,y) \in \mathcal{X} \times \mathcal{X} \, : \, x \leq y\}.$$

For all $(x,y) \in \mathcal{W}$, let $R((x,y),\cdot)$ be the joint law of a monotone coupling of $P(x,\cdot)$ and $Q(y,\cdot)$. Such a coupling exists by Strassen's theorem and Condition (4.5). Let $(\hat{X}_t, \hat{Y}_t)$ be a Markov chain on $\mathcal{W}$ with transition matrix R started at (x_0, y_0). By construction, $\hat{X}_t \leq \hat{Y}_t$ for all t a.s. That proves the first half of the theorem.

For the second half, let A be increasing in $\mathcal{X}$. Then, by the ergodic theorem for Markov chains (e.g. [Dur10, Exercise 6.6.4]),

$$\pi(A) \leftarrow \frac{1}{t}\sum_{s\leq t} \mathbb{1}_{\hat{X}_s \in A} \leq \frac{1}{t}\sum_{s\leq t} \mathbb{1}_{\hat{Y}_s \in A} \to \mu(A), \qquad a.s.$$

where we used that $\hat{X}_s \in A$ implies $\hat{Y}_s \in A$ because $\hat{X}_s \leq \hat{Y}_s$ and A is increasing. This proves the claim by definition of stochastic domination. ■

An example of application of this theorem is given in the next subsection.

4.2.2 ▷ *Ising model on $\mathbb{Z}^d$: extremal measures*

Consider the d-dimensional lattice $\mathbb{L}^d$. Let Λ be a finite subset of vertices in $\mathbb{L}^d$ and define $\mathcal{X} := \{-1,+1\}^\Lambda$, which is a poset when equipped with the relation $\sigma \leq \sigma'$ if and only if $\sigma_i \leq \sigma'_i$ for all $i \in \Lambda$. For shorthand, we occasionally write $+$ and $-$ instead of $+1$ and -1. For $\xi \in \{-1,+1\}^{\mathbb{L}^d}$, recall that the (ferromagnetic) Ising model on Λ with *boundary conditions* ξ and *inverse temperature* β is the probability distribution over *spin configurations* $\sigma \in \mathcal{X}$ given by

boundary conditions, inverse temperature, spins

$$\mu^\xi_{\beta,\Lambda}(\sigma) := \frac{1}{\mathcal{Z}_{\Lambda,\xi}(\beta)} e^{-\beta\mathcal{H}_{\Lambda,\xi}(\sigma)},$$

where

$$\mathcal{H}_{\Lambda,\xi}(\sigma) := -\sum_{\substack{i\sim j\\ i,j\in\Lambda}} \sigma_i\sigma_j - \sum_{\substack{i\sim j\\ i\in\Lambda, j\notin\Lambda}} \sigma_i\xi_j,$$

is the *Hamiltonian* and

Hamiltonian

$$\mathcal{Z}_{\Lambda,\xi}(\beta) := \sum_{\sigma\in\mathcal{X}} e^{-\beta\mathcal{H}_{\Lambda,\xi}(\sigma)},$$

is the *partition function*. (Warning: it is easy to get confused with the $-$ signs that cancel out in the exponent.) For the all-$(+1)$ and all-(-1) boundary conditions we write respectively $\mu^+_{\beta,\Lambda}(\sigma)$ and $\mu^-_{\beta,\Lambda}(\sigma)$. In this section, we show that these two measures are extremal in the following sense. For all boundary conditions $\xi \in \{-1,+1\}^{\mathbb{L}^d}$:

partition function

Claim 4.18.

$$\mu^+_{\beta,\Lambda} \succeq \mu^\xi_{\beta,\Lambda} \succeq \mu^-_{\beta,\Lambda}.$$

In words, because the ferromagnetic Ising model favors spin agreement, the all-$(+1)$ boundary condition tends to produce more $+1$s which in turn makes increasing events more likely.

The idea of the proof is to use Theorem 4.17 with a suitable Markov chain.

Stochastic domination In this context, vertices are often referred to as *sites*. Recall that the single-site Glauber dynamics of the Ising model is the Markov chain on $\mathcal{X}$ which, at each time, selects a site $i \in \Lambda$ uniformly at random and updates the spin σ_i according to $\mu^\xi_{\beta,\Lambda}(\sigma)$ conditioned on agreeing with σ at all sites in $\Lambda\backslash\{i\}$. Specifically, for $\gamma \in \{-1,+1\}$, $i \in \Lambda$, and $\sigma \in \mathcal{X}$, let $\sigma^{i,\gamma}$ be the configuration σ with the state at i being set to γ. Then, letting $n = |\Lambda|$, because the Ising measure factorizes, the transition matrix of the Glauber dynamics is simply

site

$$Q^\xi_{\beta,\Lambda}(\sigma,\sigma^{i,\gamma}) := \frac{1}{n}\cdot\frac{e^{\gamma\beta S^\xi_i(\sigma)}}{e^{-\beta S^\xi_i(\sigma)} + e^{\beta S^\xi_i(\sigma)}},$$

where

$$S^\xi_i(\sigma) := \sum_{\substack{j\sim i\\ j\in\Lambda}} \sigma_j + \sum_{\substack{j\sim i\\ j\notin\Lambda}} \xi_j.$$

All other transitions have probability 0.

This chain is clearly irreducible. It is also reversible with respect to $\mu^\xi_{\beta,\Lambda}$. Indeed, for all $\sigma \in \mathcal{X}$ and $i \in \Lambda$, letting

$$S^\xi_{\neq i}(\sigma) := \mathcal{H}_{\Lambda,\xi}(\sigma^{i,+}) + S^\xi_i(\sigma) = \mathcal{H}_{\Lambda,\xi}(\sigma^{i,-}) - S^\xi_i(\sigma),$$

we have

$$\begin{aligned}
\mu^{\xi}_{\beta,\Lambda}(\sigma^{i,-})\, Q^{\xi}_{\beta,\Lambda}(\sigma^{i,-},\sigma^{i,+}) &= \frac{e^{-\beta S^{\xi}_{\neq i}(\sigma)} e^{-\beta S^{\xi}_{i}(\sigma)}}{\mathcal{Z}_{\Lambda,\xi}(\beta)} \cdot \frac{e^{\beta S^{\xi}_{i}(\sigma)}}{n[e^{-\beta S^{\xi}_{i}(\sigma)} + e^{\beta S^{\xi}_{i}(\sigma)}]} \\
&= \frac{e^{-\beta S^{\xi}_{\neq i}(\sigma)}}{n\mathcal{Z}_{\Lambda,\xi}(\beta)[e^{-\beta S^{\xi}_{i}(\sigma)} + e^{\beta S^{\xi}_{i}(\sigma)}]} \\
&= \frac{e^{-\beta S^{\xi}_{\neq i}(\sigma)} e^{\beta S^{\xi}_{i}(\sigma)}}{\mathcal{Z}_{\Lambda,\xi}(\beta)} \cdot \frac{e^{-\beta S^{\xi}_{i}(\sigma)}}{n[e^{-\beta S^{\xi}_{i}(\sigma)} + e^{\beta S^{\xi}_{i}(\sigma)}]} \\
&= \mu^{\xi}_{\beta,\Lambda}(\sigma^{i,+})\, Q^{\xi}_{\beta,\Lambda}(\sigma^{i,+},\sigma^{i,-}).
\end{aligned}$$

In particular $\mu^{\xi}_{\beta,\Lambda}$ is the stationary distribution of $Q^{\xi}_{\beta,\Lambda}$.

Claim 4.19.

$$\xi' \geq \xi \implies Q^{\xi'}_{\beta,\Lambda} \textit{ stochastically dominates } Q^{\xi}_{\beta,\Lambda}. \tag{4.6}$$

Proof. Because the Glauber dynamics updates a single site at a time, establishing stochastic domination reduces to checking simple one-site inequalities:

Lemma 4.20. *To establish* (4.6)*, it suffices to show that, for all $\sigma \leq \tau$,*

$$Q^{\xi}_{\beta,\Lambda}(\sigma,\sigma^{i,+}) \leq Q^{\xi'}_{\beta,\Lambda}(\tau,\tau^{i,+}). \tag{4.7}$$

Proof. Assume (4.7) holds. Let A be increasing in $\mathcal{X}$ and let $\sigma \leq \tau$. Then, for the single-site Glauber dynamics, we have

$$Q^{\xi}_{\beta,\Lambda}(\sigma, A) = Q^{\xi}_{\beta,\Lambda}(\sigma, A \cap B_{\sigma}), \tag{4.8}$$

where

$$B_{\sigma} := \{\sigma^{i,\gamma} \,:\, i \in \Lambda,\ \gamma \in \{-1,+1\}\},$$

and similarly for τ, ξ'. Moreover, because A is increasing and $\tau \geq \sigma$,

$$\sigma^{i,\gamma} \in A \implies \tau^{i,\gamma} \in A, \tag{4.9}$$

and

$$\sigma^{i,-} \in A \implies \sigma^{i,+} \in A. \tag{4.10}$$

Letting

$$I^{\pm}_{\sigma,A} := \{i \in \Lambda \,:\, \sigma^{i,-} \in A\}, \qquad I^{+}_{\sigma,A} := \{i \in \Lambda \,:\, \sigma^{i,+} \in A\},$$

and similarly for τ, we have by (4.7), (4.8), (4.9), and (4.10),

$$\begin{aligned}
Q^{\xi}_{\beta,\Lambda}(\sigma, A) &= Q^{\xi}_{\beta,\Lambda}(\sigma, A \cap B_\sigma) \\
&= \sum_{i \in I^{+}_{\sigma,A}} Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,+}) + \sum_{i \in I^{\pm}_{\sigma,A}} \left[Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,-}) + Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,+}) \right] \\
&\leq \sum_{i \in I^{+}_{\sigma,A}} Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) + \sum_{i \in I^{\pm}_{\sigma,A}} \frac{1}{n} \\
&\leq \sum_{i \in I^{+}_{\tau,A}} Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) + \sum_{i \in I^{\pm}_{\tau,A}} \left[Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,-}) + Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) \right] \\
&= Q^{\xi'}_{\beta,\Lambda}(\tau, A),
\end{aligned}$$

as claimed. ■

Returning to the proof of Claim 4.19, observe that

$$Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,+}) = \frac{1}{n} \cdot \frac{e^{\beta S^{\xi}_i(\sigma)}}{e^{-\beta S^{\xi}_i(\sigma)} + e^{\beta S^{\xi}_i(\sigma)}} = \frac{1}{n} \cdot \frac{1}{e^{-2\beta S^{\xi}_i(\sigma)} + 1},$$

which is increasing in $S^{\xi}_i(\sigma)$. Now $\sigma \leq \tau$ and $\xi \leq \xi'$ imply that $S^{\xi}_i(\sigma) \leq S^{\xi'}_i(\tau)$. That proves the claim. ■

Finally:

Proof of Claim 4.18. Combining Theorem 4.17 and Claim 4.19 gives Claim 4.18. ■

Observe that we have not used any special property of the d-dimensional lattice. Indeed Claim 4.18 in fact holds for any countable, locally finite graph with positive coupling constants.

Thermodynamic limit To be written. See [RAS, Theorem 9.13].

4.2.3 ▷ *Random walk on trees: speed*

To be written. See [LP, Proposition 13.3 and Exercise 13.1].[†]

[†]Requires: Section 2.2.4.

4.2.4 FKG and Holley's inequalities

A special case of stochastic domination is positive associations. In this section, we restrict ourselves to posets of the form $\{0,1\}^F$ for F finite. We begin with an example.

Example 4.21 (Erdös-Rényi graphs: positive associations)**.** Consider an Erdös-Rényi graph $G \sim \mathbb{G}_{n,p}$. Let $\mathcal{E} = \{\{x,y\} : x,y \in [n],\ x \neq y\}$. Think of G as taking values in the poset $(\{0,1\}^{\mathcal{E}}, \leq)$ where a 1 indicates that the corresponding edge is present. In fact observe that the law of G, which we denote as usual by $\mathbb{P}_{n,p}$, is a product measure on $\{0,1\}^{\mathcal{E}}$. The event $\mathcal{A}$ that G is connected is increasing because adding edges cannot disconnect a connected graph. So is the event $\mathcal{B}$ of having a chromatic number larger than 4. Intuitively then, conditioning on $\mathcal{A}$ makes $\mathcal{B}$ more likely. Indeed the occurence of $\mathcal{A}$ tends to be accompanied with a larger number of edges which in turn makes $\mathcal{B}$ more probable. This is a more general phenomenon. That is, for any non-empty increasing events $\mathcal{A}$ and $\mathcal{B}$, we have:

Claim 4.22.

$$\mathbb{P}_{n,p}[\mathcal{B} \,|\, \mathcal{A}] \geq \mathbb{P}_{n,p}[\mathcal{B}]. \tag{4.11}$$

Or, put differently, the conditional measure $\mathbb{P}_{n,p}[\,\cdot\,|\,\mathcal{A}]$ stochastically dominates the unconditional measure $\mathbb{P}_{n,p}[\,\cdot\,]$. This is a special case of what is known as Harris' inequality (proved below). Note that (4.11) is equivalent to $\mathbb{P}_{n,p}[\mathcal{A} \cap \mathcal{B}] \geq \mathbb{P}_{n,p}[\mathcal{A}]\,\mathbb{P}_{n,p}[\mathcal{B}]$, i.e., to the fact that $\mathcal{A}$ and $\mathcal{B}$ are positively correlated. ◀

More generally:

Definition 4.23 (Positive associations)**.** *Let μ be a probability measure on $\{0,1\}^F$ where F is finite. Then μ is said to have* positive associations*, or is positively associated, if for all increasing functions $f, g : \{0,1\}^F \to \mathbb{R}$*

positive associations

$$\mu(fg) \geq \mu(f)\mu(g),$$

where

$$\mu(h) := \sum_{\omega \in \{0,1\}^F} \mu(\omega) h(\omega).$$

In particular, for any increasing events A and B it holds that

$$\mu(A \cap B) \geq \mu(A)\mu(B),$$

i.e., A and B are positively correlated*.*

positive correlation

Remark 4.24. *Note that positive associations is concerned only with* increasing *events. See Remark 4.40.*

Remark 4.25. *A notion of negative associations, which is a somewhat more delicate concept, was defined in Remark 3.59. See also [Pem00].*

Let μ be positively associated. Note that if A and B are *decreasing*, i.e. their complements are increasing, then

decreasing event

$$\begin{aligned}
\mu(A \cap B) &= 1 - \mu(A^c \cup B^c) \\
&= 1 - \mu(A^c) - \mu(B^c) + \mu(A^c \cap B^c) \\
&\geq 1 - \mu(A^c) - \mu(B^c) + \mu(A^c)\mu(B^c) \\
&= \mu(A)\mu(B).
\end{aligned}$$

Similarly, if A is increasing and B is decreasing, we have $\mu(A \cap B) \leq \mu(A)\mu(B)$.

Harris' inequality states that product measures on $\{0,1\}^F$ have positive associations. We prove a more general result known as the *FKG inequality*. For two configurations ω, ω' in $\{0,1\}^F$, we let $\omega \wedge \omega'$ and $\omega \vee \omega'$ be the coordinatewise minimum and maximum of ω and ω'.

Theorem 4.26 (FKG inequality)**.** *Let $\mathcal{X} = \{0,1\}^F$ where F is finite. Suppose μ is a positive probability measure on $\mathcal{X}$ satisfying the* FKG condition

FKG condition

$$\mu(\omega \vee \omega')\,\mu(\omega \wedge \omega') \geq \mu(\omega)\,\mu(\omega'), \qquad \forall \omega, \omega' \in \mathcal{X}. \tag{4.12}$$

This property is also known as log-convexity or log-supermodularity. We call such a measure an FKG measure*. Then μ has positive associations.*

FKG measure

Remark 4.27. *Strict positivity is not in fact needed [FKG71]. The FKG condition is equivalent to a strong form of positive associations. See Exercise 4.4.*

Note that product measures satisfy the FKG condition with equality. Indeed if $\mu(\omega)$ is of the form $\prod_{f \in F} \mu_f(\omega_f)$ then

$$\begin{aligned}
\mu(\omega \vee \omega')\,\mu(\omega \wedge \omega') &= \prod_f \mu_f(\omega_f \vee \omega'_f)\ \mu_f(\omega_f \wedge \omega'_f) \\
&= \prod_{f:\omega_f = \omega'_f} \mu_f(\omega_f)^2 \prod_{f:\omega_f \neq \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f) \\
&= \prod_{f:\omega_f = \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f) \prod_{f:\omega_f \neq \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f) \\
&= \mu(\omega)\,\mu(\omega').
\end{aligned}$$

So the FKG inequality applies, for instance, to bond percolation and Erdös-Rényi graphs. The pointwise nature of the FKG condition also makes it relatively easy to check for measures, such as that of the Ising model, which are defined explicitly up to a normalizing constant.

Example 4.28 (Ising model on $\mathbb{Z}^d$: checking FKG)**.** Consider again the setting of Section 4.2.2. Of course we work on the space $\mathcal{X} := \{-1,+1\}^\Lambda$ rather than $\{0,1\}^F$. Fix $\Lambda \subseteq \mathbb{L}^d$, $\xi \in \{-1,+1\}^{\mathbb{L}^d}$ and $\beta > 0$.

Claim 4.29. *The measure $\mu^\xi_{\beta,\Lambda}$ satisfies the FKG condition and therefore has positive associations.*

Intuitively, taking the maximum or minimum of two configurations tends to increase spin agreement and therefore leads to a higher likelihood. By taking logarithms in the FKG condition, one sees that proving the claim boils down to checking an inequality for each term in the Hamiltonian. For $\sigma, \sigma' \in \mathcal{X}$, let $\overline{\tau} = \sigma \vee \sigma'$ and $\underline{\tau} = \sigma \wedge \sigma'$. When $i \in \Lambda$ and $j \notin \Lambda$ such that $i \sim j$, we have

$$\overline{\tau}_i \xi_j + \underline{\tau}_i \xi_j = (\overline{\tau}_i + \underline{\tau}_i)\xi_j = (\sigma_i + \sigma'_i)\xi_j = \sigma_i \xi_j + \sigma'_i \xi_j. \tag{4.13}$$

For $i, j \in \Lambda$ with $i \sim j$, note first that the case $\sigma_j = \sigma'_j$ reduces to the previous calculation, so we assume $\sigma_i \neq \sigma'_i$ and $\sigma_j \neq \sigma'_j$. Then

$$\overline{\tau}_i \overline{\tau}_j + \underline{\tau}_i \underline{\tau}_j = (+1)(+1) + (-1)(-1) = 2 \geq \sigma_i \sigma_j + \sigma'_i \sigma'_j,$$

since 2 is the largest value the rightmost expression ever takes. We have shown that

$$\mathcal{H}_{\Lambda,\xi}(\overline{\tau}) + \mathcal{H}_{\Lambda,\xi}(\underline{\tau}) \leq \mathcal{H}_{\Lambda,\xi}(\sigma) + \mathcal{H}_{\Lambda,\xi}(\sigma'),$$

which implies the claim.

Again, we have not used any special property of the lattice and the same result holds for countable, locally finite graphs with positive coupling constants. Note however that in the anti-ferromagnetic case, i.e., if we multiply the Hamiltonian by -1, the above argument does not work. Indeed there is no reason to expect positive associations in that case. ◀

The FKG inequality in turn follows from a more general result known as Holley's inequality.

Theorem 4.30 (Holley's inequality)**.** *Let $\mathcal{X} = \{0,1\}^F$ where F is finite. Suppose μ_1 and μ_2 are positive probability measures on $\mathcal{X}$ satisfying*

$$\mu_2(\omega \vee \omega')\, \mu_1(\omega \wedge \omega') \geq \mu_2(\omega)\, \mu_1(\omega'), \qquad \forall \omega, \omega' \in \mathcal{X}. \tag{4.14}$$

Then $\mu_1 \preceq \mu_2$.

Before proving Holley's inequality, we check that it indeed implies the FKG inequality. See Exercise 4.1 for an elementary proof in the independent case, i.e., of Harris' inequality.

Proof of Theorem 4.26. Assume that μ satisfies the FKG condition and let f, g be increasing functions. Because of our restriction to positive measures in Holley's inequality, we will work with positive functions. This is done without loss of generality. Indeed, note that f and g are increasing if and only if $f' := f - f(0) + 1 > 0$ and $g' := g - g(0) + 1 > 0$ are increasing and that, moreover,

$$\begin{aligned}\mu(f'g') - \mu(f')\mu(g') &= \mu([f' - \mu(f')][g' - \mu(g')]) \\ &= \mu([f - \mu(f)][g - \mu(g)]) \\ &= \mu(fg) - \mu(f)\mu(g).\end{aligned}$$

In Holley's inequality, we let $\mu_1 := \mu$ and define the positive probability measure

$$\mu_2(\omega) := \frac{g(\omega)\mu(\omega)}{\mu(g)}.$$

We check that μ_1 and μ_2 satisfy the conditions of Holley's inequality. Note that $\omega' \leq \omega \vee \omega'$ for any ω so that, because g is increasing, we have $g(\omega') \leq g(\omega \vee \omega')$. Hence, for any ω, ω',

$$\begin{aligned}\mu_1(\omega)\mu_2(\omega') &= \mu(\omega)\frac{g(\omega')\mu(\omega')}{\mu(g)} \\ &= \mu(\omega)\mu(\omega')\frac{g(\omega')}{\mu(g)} \\ &\leq \mu(\omega \wedge \omega')\mu(\omega \vee \omega')\frac{g(\omega \vee \omega')}{\mu(g)} \\ &= \mu_1(\omega \wedge \omega')\mu_2(\omega \vee \omega'),\end{aligned}$$

where on the third line we used the FKG condition satisfied by μ.

So Holley's inequality implies that $\mu_2 \succeq \mu_1$. Hence, since f is increasing, by Corollary 4.13

$$\mu(f) = \mu_1(f) \leq \mu_2(f) = \frac{\mu(fg)}{\mu(g)},$$

and the theorem is proved. ∎

Proof of Theorem 4.30. We use Theorem 4.17. This is similar to what was done in Section 4.2.2. This time we couple Metropolis-like chains. For $x \in \mathcal{X}$ and $\gamma \in$

$\{0,1\}$, we let $x^{i,\gamma}$ be x with coordinate i set to γ. We write $x \sim y$ if $\|x-y\|_1 = 1$. Let $n = |F|$.

For $\alpha, \beta > 0$ small enough, the following transition matrix over $\mathcal{X}$ is irreducible and reversible w.r.t. its stationary distribution μ_2: for all $i \in F$, $y \in \mathcal{X}$,

$$Q(y^{i,0}, y^{i,1}) = \frac{\alpha}{n}\left\{\beta\right\},$$

$$Q(y^{i,1}, y^{i,0}) = \frac{\alpha}{n}\left\{\beta\,\frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})}\right\},$$

$$Q(y, y) = 1 - \sum_{z \sim y} Q(y, z).$$

Let P be similarly defined for μ_1 *with the same values of* α *and* β. For reasons that will be clear below, the value of β is chosen so that the *sum* of the two expressions in brackets above is smaller than 1 for all y, i. The value of α is then chosen so that $P(x,x), Q(y,y) \geq 0$ for all x, y. Reversibility follows immediately from the first two equations. We call the first transition above an *upward transition* and the second one a *downward transition*.

upward/downward transition

By Theorem 4.17, it remains to show that Q stochastically dominates P. That is, for any $x \leq y$, we want to show that $P(x,\cdot) \preceq Q(y,\cdot)$. We produce a monotone coupling $(\hat{X}, \hat{Y})$ of these two distributions. Our goal is never to perform an upward transition in x *simultaneously* with a downward transition in y. Observe that

$$\frac{\mu_1(x^{i,0})}{\mu_1(x^{i,1})} \geq \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})} \tag{4.15}$$

by taking $\omega = y^{i,0}$ and $\omega' = x^{i,1}$ in Condition (4.14).

The coupling works as follows. Fix $x \preceq y$. With probability $1-\alpha$, set $(\hat{X}, \hat{Y}) := (x, y)$. Otherwise, pick a coordinate $i \in F$ uniformly at random. There are several cases to consider depending on the values of x_i, y_i (with $x_i \leq y_i$ by assumption):

- $(x_i, y_i) = (0,0)$: With probability β, perform an upward transition in both, i.e., set $\hat{X} := x^{i,1}$ and $\hat{Y} := y^{i,1}$. With probability $1-\beta$, set $(\hat{X}, \hat{Y}) := (x, y)$ instead.

- $(x_i, y_i) = (1,1)$: With probability $\beta\frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})}$, perform a downward transition in both, i.e., set $\hat{X} := x^{i,0}$ and $\hat{Y} := y^{i,0}$. With probability

$$\beta\left(\frac{\mu_1(x^{i,0})}{\mu_1(x^{i,1})} - \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})}\right),$$

perform a downward transition in x only, i.e., set $\hat{X} := x^{i,0}$ and $\hat{Y} := y$. Note that (4.15) guarantees that the previous step is well-defined. With the remaining probability, set $(\hat{X}, \hat{Y}) := (x, y)$ instead.

- $(x_i, y_i) = (0, 1)$: With probability β, perform an upward transition in x only, i.e., set $\hat{X} := x^{i,1}$ and $\hat{Y} := y$. With probability $\beta \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})}$, perform a downward transition in y only, i.e., set $\hat{X} := x$ and $\hat{Y} := y^{i,0}$. With the remaining probability, set $(\hat{X}, \hat{Y}) := (x, y)$ instead. (This is where we use the odd choice of β.)

By construction, this coupling satisfies $\hat{X} \leq \hat{Y}$ a.s. An application of Theorem 4.17 concludes the proof. ■

Example 4.31 (Ising model: extremality revisited)**.** Holley's inequality gives another proof of Claim 4.18. To see this, just repeat the calculations of Example 4.28, where now (4.13) is replaced with an inequality. See Exercise 4.2. ◀

4.2.5 ▷ *Erdös-Rényi graphs: Janson's inequality, and application to the containment problem*

Let $G = (V, E) \sim \mathbb{G}_{n,p}$ be an Erdös-Rényi graph. Repeating the computations of Section 2.2.2 (or see Claim 2.22), we see that the property of being triangle-free has threshold n^{-1}. That is, the probability that G contains a triangle goes to 0 or 1 as $n \to +\infty$ depending on whether $p \ll n^{-1}$ or $p \gg n^{-1}$ respectively. In this section, we investigate what happens *at* the threshold. From now on, we assume that $p = \lambda/n$ for some $\lambda > 0$ not depending on n.

For any subset S of three distinct vertices of G, let B_S be the event that S forms a triangle in G. So

$$\varepsilon := \mathbb{P}_{n,p}[B_S] = p^3 \to 0. \tag{4.16}$$

Let $X_n = \sum_{S \in \binom{V}{3}} \mathbb{1}_{B_S}$ be the number of triangles in G. By the linearity of expectation, the expected number of triangles is

$$\mathbb{E}_{n,p} X_n = \binom{n}{3} p^3 = \frac{n(n-1)(n-2)}{6} \left(\frac{\lambda}{n}\right)^3 \to \frac{\lambda^3}{6},$$

as $n \to +\infty$. *If the events $\{B_S\}_S$ were mutually independent,* X_n would be binomially-distributed and the event that G is triangle-free would have probability

$$\prod_{S \in \binom{V}{3}} \mathbb{P}_{n,p}[B_S^c] = (1 - p^3)^{\binom{n}{3}} \to e^{-\lambda^3/6}. \tag{4.17}$$

In fact, by the Poisson approximation to the binomial (e.g. [Dur10, Theorem 3.6.1]), we would have that the number of triangles converges weakly to $\mathrm{Poi}(\lambda^3/6)$.

In reality, of course, the events $\{B_S\}$ are not *mutually* independent. Observe however that, for *most* pairs S, S', the events B_S and $B_{S'}$ are in fact *pairwise* independent. That is the case whenever $|S \cap S'| \leq 1$, i.e., whenever the edges connecting S are disjoint from those connecting S'. Write $S \sim S'$ if $S \neq S'$ are not independent, i.e. if $|S \cap S'| = 2$. The expected number of (unordered) pairs $S \sim S'$ both forming a triangle is

$$\Delta := \frac{1}{2} \sum_{\substack{S,S' \binom{V}{3} \\ S \sim S'}} \mathbb{P}_{n,p}[B_S \cap B_{S'}] = \frac{1}{2}\binom{n}{3}(n-3)p^5 = \Theta(n^4 p^5) \to 0. \quad (4.18)$$

Given that the events $\{B_S\}$ are "mostly" independent, it is natural to expect that X_n behaves asymptotically as it does in the independent case. Indeed we prove:

Claim 4.32.

$$\mathbb{P}_{n,p}[X_n = 0] \to e^{-\lambda^3/6}.$$

Remark 4.33. *In fact,* $X_n \xrightarrow{\mathrm{d}} \mathrm{Poi}(\lambda^3/6)$. *See Exercises 2.11 and 4.5.*

The FKG inequality immediately gives one direction. Recall that $\mathbb{P}_{n,p}$, as a product measure over edge sets, satisfies the FKG condition and therefore has positive associations by the FKG inequality. Moreover the events B_S^c are decreasing for all S. Hence

$$\mathbb{P}_{n,p}\left[\bigcap_{S \in \binom{V}{3}} B_S^c\right] \geq \prod_{S \in \binom{V}{3}} \mathbb{P}_{n,p}[B_S^c] \to e^{-\lambda^3/6},$$

by (4.17). As it turns out, the FKG inequality also gives a bound in the other direction. This is known as *Janson's inequality*, which we state in a more general context.

Janson's inequality Let $\mathcal{X} := \{0,1\}^F$ where F is finite. Let B_i, $i \in I$, be a finite collection of events of the form $B_i := \{\omega \in \mathcal{X} : \omega \geq \beta^{(i)}\}$ for some $\beta^{(i)} \in \mathcal{X}$. Think of these as "bad events" corresponding to a certain subset of coordinates being set to 1. By definition, the B_is are increasing. Assume $\mathbb{P}$ is a positive product measure on $\mathcal{X}$. Write $i \sim j$ if $\beta^{(i)}_r = \beta^{(j)}_r = 1$ for some r and note that B_i is independent of B_j if $i \nsim j$. Set

$$\Delta := \sum_{\substack{\{i,j\} \\ i \sim j}} \mathbb{P}[B_i \cap B_j].$$

Theorem 4.34 (Janson's inequality)**.** *Let $\mathcal{X}$, $\mathbb{P}$, $\{B_i\}_{i\in I}$ and Δ be as above. Assume further that there is $\varepsilon > 0$ such that $\mathbb{P}[B_i] \leq \varepsilon$ for all $i \in I$. Then*

$$\prod_{i\in I} \mathbb{P}[B_i^c] \leq \mathbb{P}[\cap_{i\in I} B_i^c] \leq e^{\frac{\Delta}{1-\varepsilon}} \prod_{i\in I} \mathbb{P}[B_i^c].$$

Before proving the theorem, we show that it implies Claim 4.32. We have already shown in (4.16) and (4.18) that $\varepsilon \to 0$ and $\Delta \to 0$. Janson's inequality immediately implies the claim by (4.17).

Proof of Theorem 4.34. The lower bound follows from the FKG inequality.

In the other direction, the first step is somewhat clear. We apply the chain rule to obtain

$$\mathbb{P}[\cap_{i\in I} B_i^c] \quad = \quad \prod_{i=1}^{m} \mathbb{P}[B_i^c \mid \cap_{j\in[i-1]} B_j^c].$$

The rest is clever manipulation. W.l.o.g. assume $I = [m]$. For $i \in [m]$, let $N(i) := \{\ell \in [m] \,:\, \ell \sim i\}$ and $N_<(i) := N(i) \cap [i-1]$. Note that B_i is independent of $\{B_\ell \,:\, \ell \in [i-1]\backslash N_<(i)\}$. Then

$$\begin{aligned}
\mathbb{P}[B_i \mid \cap_{j\in[i-1]} B_j^c] &= \frac{\mathbb{P}\left[B_i \cap \left(\cap_{j\in[i-1]} B_j^c\right)\right]}{\mathbb{P}[\cap_{j\in[i-1]} B_j^c]} \\
&\geq \frac{\mathbb{P}\left[B_i \cap \left(\cap_{j\in[i-1]} B_j^c\right)\right]}{\mathbb{P}[\cap_{j\in[i-1]\backslash N_<(i)} B_j^c]} \\
&= \mathbb{P}\left[B_i \cap \left(\cap_{j\in N_<(i)} B_j^c\right) \,\middle|\, \cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right] \\
&= \mathbb{P}\left[B_i \,\middle|\, \cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right] \\
&\qquad \times \mathbb{P}\left[\cap_{j\in N_<(i)} B_j^c \,\middle|\, B_i \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right] \\
&= \mathbb{P}\left[B_i\right] \\
&\qquad \times \mathbb{P}\left[\cap_{j\in N_<(i)} B_j^c \,\middle|\, B_i \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right],
\end{aligned}$$

By a union bound the second term on the last line is

$$\begin{aligned}
&\mathbb{P}\left[\cap_{j\in N_<(i)} B_j^c \,\middle|\, B_i \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right] \\
&\quad \geq 1 - \sum_{j\in N_<(i)} \mathbb{P}\left[B_j \,\middle|\, B_i \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right] \\
&\quad \geq 1 - \sum_{j\in N_<(i)} \mathbb{P}\left[B_j \,\middle|\, B_i\right],
\end{aligned}$$

where the last line follows from the FKG inequality applied to the product measure $\mathbb{P}[\,\cdot\,|\,B_i]$ (on $\{0,1\}^{F'}$ with $F' := \{\ell \in [m] \,:\, \beta_\ell^{(i)} = 0\}$). Combining the last three displays and using $1 + z \leq e^z$, we get

$$\begin{aligned}
\mathbb{P}[\cap_{i\in I} B_i^c] &\leq \prod_{i=1}^m \left[\mathbb{P}\left[B_i^c\right] + \sum_{j\in N_<(i)} \mathbb{P}\left[B_i \cap B_j\right]\right] \\
&\leq \prod_{i=1}^m \mathbb{P}\left[B_i^c\right]\left[1 + \frac{1}{1-\varepsilon}\sum_{j\in N_<(i)} \mathbb{P}\left[B_i \cap B_j\right]\right] \\
&\leq \prod_{i=1}^m \mathbb{P}\left[B_i^c\right] \exp\left(\frac{1}{1-\varepsilon}\sum_{j\in N_<(i)} \mathbb{P}\left[B_i \cap B_j\right]\right).
\end{aligned}$$

By the definition of Δ, we are done. ■

4.2.6 ▷ *Percolation on $\mathbb{Z}^2$: RSW theory, and a proof of Harris' theorem*

Consider bond percolation on the two-dimensional lattice $\mathbb{L}^2$. Recall that the percolation function is given by

$$\theta(p) := \mathbb{P}_p[|\mathcal{C}_0| = +\infty],$$

where $\mathcal{C}_0$ is the open cluster of the origin. We known from Example 4.10 that $\theta(p)$ is non-decreasing. Let

$$p_c(\mathbb{L}^2) := \sup\{p \geq 0 \,:\, \theta(p) = 0\},$$

be the critical value. We proved in Section 2.1.5 that there is a non-trivial transition, i.e., $p_c(\mathbb{L}^2) \in (0,1)$. In fact we showed that $p_c(\mathbb{L}^2) \in [1/3, 2/3]$ (see Exercise 2.2).

Our goal in this section is to use the FKG inequality to improve this further to:

Theorem 4.35 (Harris' theorem)**.**

$$\theta(1/2) = 0.$$

Or, put differently, $p_c(\mathbb{L}^2) \geq 1/2$.

Remark 4.36. *This bound is tight, i.e., in fact $p_c(\mathbb{L}^2) = 1/2$. The other direction, known as* Kesten's theorem*, is postponed to Section 8.3.4 where an additional ingredient is introduced, Russo's formula.*

Several proofs of Harris' theorem are known. A particularly elegant one is sketched in Exercise 6.1. Here we present a proof that uses an important tool in percolation theory, the *RSW theorem*, an application of the FKG inequality.

Harris' theorem To motivate the RSW theorem, we start with the proof of Harris' theorem.

Proof of Theorem 4.35. Fix $p = 1/2$. We use duality. Consider the $\mathbb{L}^2$ annulus

$$\mathrm{Ann}(\ell) := [-3\ell, 3\ell]^2 \backslash [-\ell, \ell].$$

The existence of a closed dual cycle inside $\mathrm{Ann}(\ell)$, which we denote by $\mathrm{O_d}(\ell)$, prevents the possibility of an infinite open self-avoiding path from the origin in the primal lattice $\mathbb{L}^2$. See Figure 4.3. That is,

$$\mathbb{P}_{1/2}[|\mathcal{C}_0| = +\infty] \leq \prod_{k=0}^{K} \{1 - \mathbb{P}_{1/2}[\mathrm{O_d}(3^k)]\}, \tag{4.19}$$

for all K, where we took powers of 3 to make the annuli disjoint and therefore independent.

To prove the theorem, it suffices to show that there is a constant $c^* > 0$ such that, for all ℓ, $\mathbb{P}_{1/2}[\mathrm{O_d}(\ell)] \geq c^*$. Then the r.h.s. of (4.19) tends to 0 as $K \to +\infty$. To simplify further, thinking of $\mathrm{Ann}(\ell)$ as a union of four rectangles $[-3\ell, -\ell) \times [-3\ell, 3\ell]$, $[-3\ell, 3\ell] \times (\ell, 3\ell]$, etc., it suffices to consider the event $\mathrm{O}_\mathrm{d}^{\#}(\ell)$ that each one of these rectangles contains a closed dual self-avoiding path connecting its two shorter sides. (More precisely, for the first rectangle above, the path connects $[-3\ell + 1/2, -\ell - 1/2] \times \{3\ell - 1/2\}$ to $[-3\ell + 1/2, -\ell - 1/2] \times \{-3\ell + 1/2\}$ and stays inside the rectangle, etc.) See Figure 4.3. By symmetry the probability that such a path exists is the same for all four rectangles. Denote it by ρ_ℓ. Moreover the event that such a path exists is increasing so, although the four events are not independent, we can apply the FKG inequality. Hence, because $\mathrm{O}_\mathrm{d}^{\#}(\ell) \subseteq \mathrm{O_d}(\ell)$, we finally get the bound

$$\mathbb{P}_{1/2}[\mathrm{O_d}(\ell)] \geq \rho_\ell^4.$$

The RSW theorem and some symmetry arguments, both of which are detailed below, imply that there is some $c > 0$ such that, for all ℓ:

Claim 4.37.

$$\rho_\ell \geq c.$$

That concludes the proof. ■

It remains to prove Claim 4.37. We first state the RSW theorem.

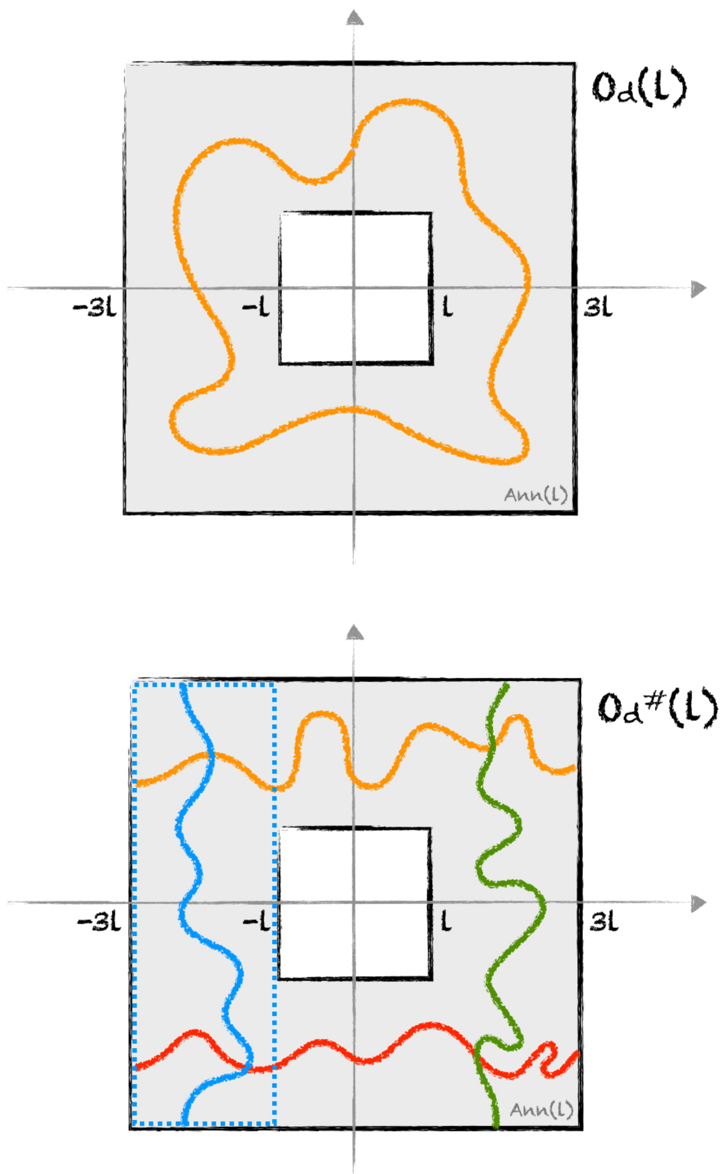

Figure 4.3: Top: the event $\mathrm{O}_{\mathrm{d}}(\ell)$. Bottom: the event $\mathrm{O}_{\mathrm{d}}^{\#}(\ell)$.

RSW theory We have reduced the proof of Harris' theorem to bounding the probability that certain closed paths exist in the dual lattice. To be consistent with the standard RSW notation, we switch to the primal lattice and consider open paths. We also let p take any value in $(0, 1)$. As we did in Section 2.1.5, we accept without proof some topological facts to be stated below.

Let $R_{n,\alpha}(p)$ be the probability that the rectangle

$$B(\alpha n, n) := [-n, (2\alpha - 1)n] \times [-n, n],$$

has an open self-avoiding path connecting its left and right sides with the path remaining inside the rectangle. Such a path is called an (open) *left-right crossing*. The event that a left-right crossing exists in a rectangle B is denoted by $\mathrm{LR}(B)$. We similarly define the event, $\mathrm{TB}(B)$, that a *top-bottom crossing* exists in B. In essence, the RSW theorem says this: if there is a significant probability that a left-right crossing exists in $B(n, n)$, then there is a significant probability that an open self-avoiding cycle exists in $\mathrm{Ann}(n)$. More precisely, here is a version of the theorem that will be enough for our purposes.

left-right/ top-bottom crossing

Theorem 4.38 (RSW theorem)**.** *For all $n \geq 2$ (divisible by 4) and $p \in (0, 1)$,*

$$R_{n,3}(p) \geq \frac{1}{4}\, R_{n,1}(p)^{11}\, R_{n/2,1}(p)^{12}. \tag{4.20}$$

Before presenting a proof, we finish the proof of Harris' theorem by proving Claim 4.37.

Proof of Claim 4.37. The point of (4.20) is that, if $R_{n,1}(p)$ is bounded away from 0 uniformly in n, so is the l.h.s. By the argument in the proof of Harris' theorem, this then implies that an open self-avoiding cycle exists in $\mathrm{Ann}(n)$ with a probability bounded away from 0 as well. Hence to prove Claim 4.37 it suffices to give a lower bound on $R_{n,1}(1/2)$. It is crucial that this bound not depend on the "scale," n. As it turns out, a simple duality-based symmetry argument does the trick. The following fact about $\mathbb{L}^2$ is a variant of the contour lemma, Lemma 2.12. Its proof is similar and Exercise 4.6 asks for the details (the "if" direction being the non-trivial implication).

Lemma 4.39. *There is an open left-right crossing in the primal rectangle $[0, n+1] \times [0, n]$ if and only if there is* no *closed top-bottom crossing in the dual rectangle $[1/2, n + 1/2] \times [-1/2, n + 1/2]$.*

By symmetry, when $p = 1/2$, the two events in Lemma 4.39 have equal probability. So they must have probability $1/2$ because they form a partition of the space of outcomes by the lemma. By monotonicity, that implies $R_{n,1}(1/2) \geq 1/2$ for all n. The RSW theorem then implies the required bound. ■

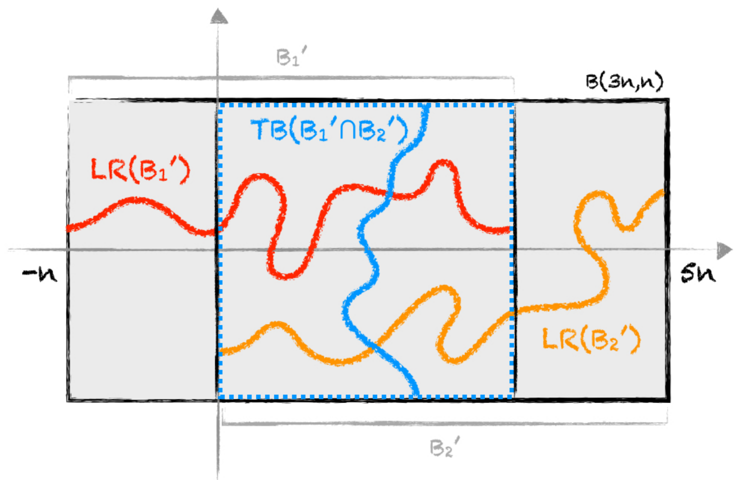

Figure 4.4: Illustration of the implication $\mathrm{LR}(B_1') \cap \mathrm{TB}(B_1' \cap B_2') \cap \mathrm{LR}(B_2') \subseteq \mathrm{LR}(B(3n, n))$.

The proof of the RSW theorem involves a clever choice of event that relates the existence of crossings in squares and rectangles.

Proof of Theorem 4.38. There are several steps in the proof.

Step 1: it suffices to bound $R_{n,3/2}(p)$ We first reduce the proof to finding a bound on $R_{n,3/2}(p)$. Let $B_1' := B(2n, n)$ and $B_2' := [n, 5n] \times [-n, n]$. Note that $B_1' \cup B_2' = B(3n, n)$ and $B_1' \cap B_2' = [n, 3n] \times [-n, n]$. Then we have the implication

$$\mathrm{LR}(B_1') \cap \mathrm{TB}(B_1' \cap B_2') \cap \mathrm{LR}(B_2') \subseteq \mathrm{LR}(B(3n, n)).$$

See Figure 4.4. Each event on the l.h.s. is increasing so the FKG inequality gives

$$R_{n,3}(p) \geq R_{n,2}(p)^2 R_{n,1}(p).$$

A similar argument over $B(2n, n)$ gives

$$R_{n,2}(p) \geq R_{n,3/2}(p)^2 R_{n,1}(p).$$

Combining the two, we have proved

$$R_{n,3}(p) \geq R_{n,3/2}(p)^4 R_{n,1}(p)^3. \tag{4.21}$$

Step 2: bounding $R_{n,3/2}(p)$ The heart of the proof is to bound $R_{n,3/2}(p)$ using an event involving crossings of squares. Let

$$\begin{aligned} B_1 &:= B(n,n) = [-n,n] \times [-n,n], \\ B_2 &:= [0,2n] \times [-n,n], \\ B_{12} &:= B_1 \cap B_2 = [0,n] \times [-n,n], \\ S &:= [0,n] \times [0,n]. \end{aligned}$$

Let Γ_1 be the event that there are paths P_1, P_2, where P_1 is a top-bottom crossing of S and P_2 is an open self-avoiding path connecting the left side of B_1 to P_1 and stays inside B_1. Similarly let Γ_2' be the event that there are paths P_1', P_2', where P_1' is a top-bottom crossing of S and P_2' is an open self-avoiding path connecting the right side of B_2 to P_1' and stays inside B_2. Then we have the implication

$$\Gamma_1 \cap \mathrm{LR}(S) \cap \Gamma_2' \subseteq \mathrm{LR}(B(3n/2, n)).$$

See Figure 4.5. By symmetry $\mathbb{P}_p[\Gamma_1] = \mathbb{P}_p[\Gamma_2']$. Moreover, the events on the l.h.s. are increasing so, by the FKG inequality,

$$R_{n,3/2}(p) \geq \mathbb{P}_p[\Gamma_1]^2 R_{n/2,1}(p), \tag{4.22}$$

and it remains to bound $\mathbb{P}_p[\Gamma_1]$. That requires several additional definitions.

Step 3: bounding $\mathbb{P}_p[\Gamma_1]$ Let P_1 and P_2 be top-bottom crossings of S. There is a natural partial order over such crossings. The path P_1 divides S into two subgraphs: $[P_1\}$ which includes the left side of S (including edges on the left incident with P_1 but not those edges on P_1 itself) and $\{P_1]$ which includes the right side of S (and P_1 itself). Then we write $P_1 \preceq P_2$ if $\{P_1] \subseteq \{P_2]$. Assuming $\mathrm{TB}(S)$ holds, one also gets the existence of a unique *rightmost crossing*: take the union of all top-bottom crossings of S as sets of edges; then the "right boundary" of this set is a top-bottom crossing P_S^* such that $P_S^* \preceq P$ for all top-bottom crossings P of S. (We accept as a fact the existence of a unique rightmost crossing. See Exercise 4.6 for a related construction.)

rightmost crossing

Let I_S be the set of self-avoiding (not necessarily open) paths connecting the top and bottom of S and stay inside S. For $P \in I_S$, we let P' be the reflection of P through the x-axis in $B_{12} \backslash S$ and we let $\frac{P}{P'}$ be the union of P and P'. Define $[\frac{P}{P'}\}$ to be the subgraph of B_1 to the left of $\frac{P}{P'}$ (including edges on the left incident with $\frac{P}{P'}$ but not those edges on $\frac{P}{P'}$ itself). Let $\mathrm{LR}^+\left([\frac{P}{P'}\}\right)$ be the event that there is a left-right crossing of $[\frac{P}{P'}\}$ ending on P, i.e., that there is an open self-avoiding path connecting the left side of B_1 and P that stays within $[\frac{P}{P'}\}$. See Figure 4.5.

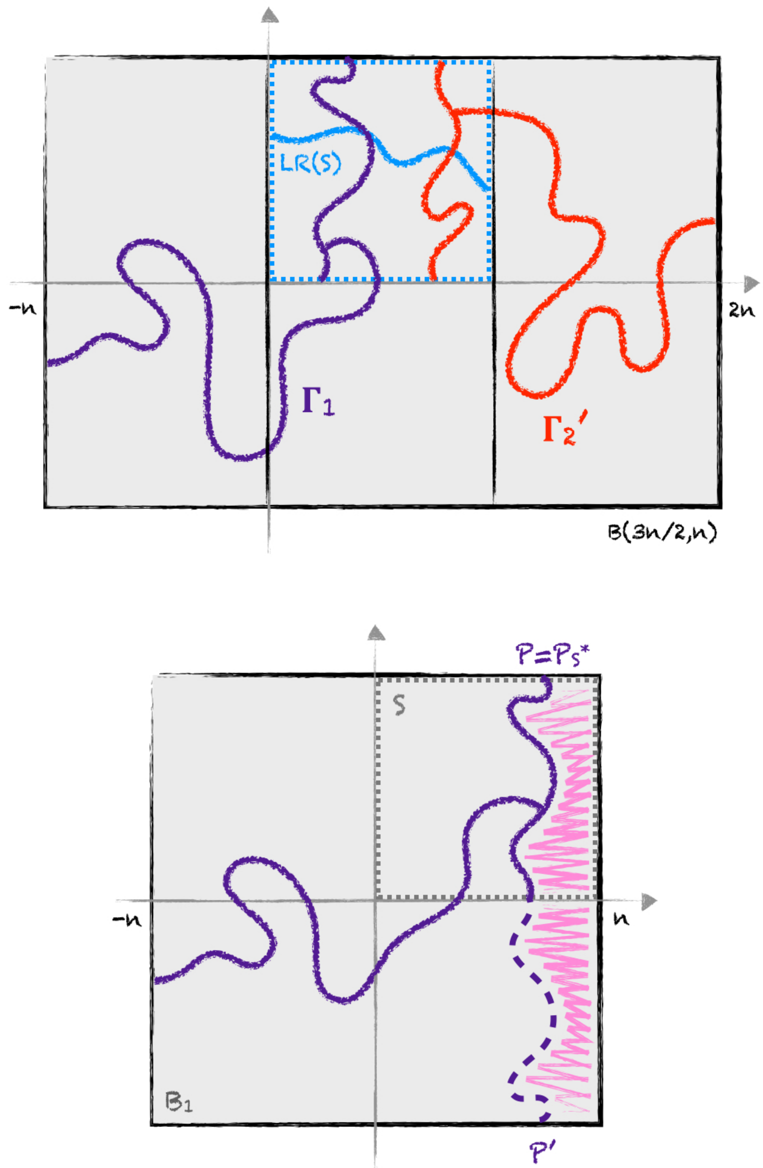

Figure 4.5: Top: illustration of the implication $\Gamma_1 \cap \mathrm{LR}(S) \cap \Gamma_2' \subseteq \mathrm{LR}(B(3n/2, n))1$. Bottom: the event $\mathrm{LR}^+\big([\frac{P}{P'}\}\big) \cap \{P = P_S^*\}$; the dashed path is the mirror image of the rightmost top-bottom crossing in S; the pink region is the complement in B_1 of the set $[\frac{P}{P'}\}$. Note that, because on the bottom figure the left-right path must stay within $[\frac{P}{P'}\}$, the configuration shown in the top figure where the purple left-right path "travels behind" the top-bottom crossing of S cannot occur.

Note that the existence of a left-right crossing of B_1 implies the existence of an open self-avoiding path connecting the left side of B_1 to $\frac{P}{P'}$. By symmetry we then get

$$\mathbb{P}_p\left[\mathrm{LR}^+\left([\tfrac{P}{P'}\}\right)\right] \geq \frac{1}{2}\mathbb{P}_p[\mathrm{LR}(B_1)] = \frac{1}{2}R_{n,1}(p). \tag{4.23}$$

Now comes a subtle point. We turn to the rightmost crossing of S—for two reasons. First, by uniqueness, $\{P = P_S^*\}_{P\in I_S}$ forms a partition of $\mathrm{TB}(S)$. Second, the rightmost crossing has a Markov-like property. Observe that, for $P \in I_S$, the event that $\{P = P_S^*\}$ *depends only the bonds in* $\{P]$. In particular it is *independent of the bonds in* $[\frac{P}{P'}\}$, e.g. of the event $\mathrm{LR}^+\left([\frac{P}{P'}\}\right)$. Hence

$$\mathbb{P}_p\left[\mathrm{LR}^+\left([\tfrac{P}{P'}\}\right) \mid P = P_S^*\right] = \mathbb{P}_p\left[\mathrm{LR}^+\left([\tfrac{P}{P'}\}\right)\right].$$

With (4.23), we get

$$\begin{aligned}
\mathbb{P}_p[\Gamma_1] &\geq \sum_{P\in I_S} \mathbb{P}_p[P = P_S^*]\,\mathbb{P}_p\left[\mathrm{LR}^+\left([\tfrac{P}{P'}\}\right) \mid P = P_S^*\right] \\
&\geq \frac{1}{2}R_{n,1}(p) \sum_{P\in I_S} \mathbb{P}_p[P = P_S^*] \\
&= \frac{1}{2}R_{n,1}(p)\,\mathbb{P}_p[\mathrm{TB}(S)] \\
&= \frac{1}{2}R_{n,1}(p)R_{n/2,1}(p).
\end{aligned} \tag{4.24}$$

Step 4: putting everything together Combining (4.21), (4.22) and (4.24) gives that

$$\begin{aligned}
R_{n,3}(p) &\geq R_{n,3/2}(p)^4 R_{n,1}(p)^3 \\
&\geq [\mathbb{P}_p[\Gamma_1]^2 R_{n/2,1}(p)]^4 R_{n,1}(p)^3 \\
&\geq \left[\left(\frac{1}{2}R_{n,1}(p)R_{n/2,1}(p)\right)^2 R_{n/2,1}(p)\right]^4 R_{n,1}(p)^3.
\end{aligned}$$

Rearranging concludes the proof of the RSW theorem. ■

Remark 4.40. *This argument is quite subtle. In that respect, it is instructive to read the remark after [Gri97, Theorem 9.3].*

4.3 Couplings of Markov chains

As we have seen, coupling is useful to bound total variation distance. A natural application is mixing.

4.3.1 Bounding the mixing time via coupling

Let P be an irreducible, aperiodic Markov transition matrix on the finite state space V with stationary distribution π. Recall that, for a fixed $0 < \varepsilon < 1/2$, the mixing time of P is

$$\mathrm{t_{mix}}(\varepsilon) := \min\{t \,:\, d(t) \leq \varepsilon\},$$

where

$$d(t) := \max_{x \in V} \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}}.$$

A *coupling of Markov chains* with transition matrix P is a Markov chain (X_t, Y_t) on $V \times V$ such that both (X_t) and (Y_t) are Markov chains with transition matrix P. For our purposes, the following special type of coupling will suffice.

Definition 4.41 (Markovian coupling)**.** *A* Markovian coupling *of P is a Markov chain $(X_t, Y_t)_t$ on $V \times V$ with transition matrix Q satisfying:*

Markovian coupling

- (Markovian coupling) *For all $x, y, x', y' \in V$,*

$$\sum_{z'} Q((x,y),(x',z')) = P(x,x'),$$

$$\sum_{z'} Q((x,y),(z',y')) = P(y,y').$$

We say that a Markovian coupling is coalescing *if further:*

coalescing

- (Coalescing) *For all $z \in V$,*

$$x' \neq y' \implies Q((z,z),(x',y')) = 0.$$

Let (X_t, Y_t) be a coalescing, Markovian coupling of P. By the coalescing condition, if $X_s = Y_s$ then $X_t = Y_t$ for all $t \geq s$. That is, once (X_t) and (Y_t) meet, they remain together. Let τ_{coal} be the first meeting time, i.e.,

meeting time

$$\tau_{\mathrm{coal}} := \inf\{t \geq 0 \,:\, X_t = Y_t\}.$$

The key to the coupling method is the following immediate consequence of the coupling inequality, Lemma **??**. For any starting point (x,y),

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{\mathrm{TV}} \leq \mathbb{P}_{(x,y)}[X_t \neq Y_t] = \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t]. \tag{4.25}$$

To relate this inequality to the mixing time, we define

$$\bar{d}(t) := \max_{x,y \in V} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{\mathrm{TV}}.$$

Lemma 4.42.

$$d(t) \leq \bar{d}(t) \leq 2d(t), \qquad \forall t.$$

Proof. The second inequality follows from an application of the triangle inequality.

For the first inequality, note that by definition of the total variation distance

$$\begin{aligned}
\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} &= \sup_{A \subseteq V} |P^t(x,A) - \pi(A)| \\
&= \sup_{A \subseteq V} \left| \sum_{y \in V} \pi(y)[P^t(x,A) - P^t(y,A)] \right| \\
&\leq \sup_{A \subseteq V} \sum_{y \in V} \pi(y) |P^t(x,A) - P^t(y,A)| \\
&\leq \sum_{y \in V} \pi(y) \left\{ \sup_{A \subseteq V} |P^t(x,A) - P^t(y,A)| \right\} \\
&\leq \sum_{y \in V} \pi(y) \|P^t(x,\cdot) - P^t(y,\cdot)\|_{\mathrm{TV}} \\
&\leq \max_{x,y \in V} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{\mathrm{TV}}.
\end{aligned}$$

■

Combining (4.25) and Lemma 4.42, we get the main result of this section.

Theorem 4.43 (Bounding the mixing time: coupling method)**.** *Let (X_t, Y_t) be a coalescing, Markovian coupling of an irreducible transition matrix P on a finite state space V with stationary distribution π. Then*

$$d(t) \leq \max_{x,y \in V} \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t].$$

In particular

$$\mathrm{t_{mix}}(\varepsilon) \leq \inf\left\{t \geq 0 \,:\, \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t] \leq \varepsilon,\ \forall x, y\right\}.$$

We give a few simple examples in the next subsection.

4.3.2 ▷ *Markov chains: mixing on cycles, hypercubes, and trees*

In this section, we consider *lazy simple random walk* on various graphs. By this we mean that the walk stays put with probability $1/2$ and otherwise picks an adjacent vertex uniformly at random. In each case, we construct a coupling to bound the mixing time. *lazy walk*

Cycle Let (Z_t) be lazy simple random walk on the cycle of size n, $\mathbb{Z}_n := \{0, 1 \ldots, n-1\}$, where $i \sim j$ if $|j - i| = 1 \pmod n$. For any starting points x, y, we construct a coalescing Markovian coupling (X_t, Y_t) of this chain. Set $(X_0, Y_0) := (x, y)$. At each time, flip a fair coin. On heads, Y_t stays put and X_t moves one step, the direction of which is uniform at random. On tails, proceed similarly with the roles of X_t and Y_t reversed. Let D_t be the clockwise distance between X_t and Y_t. Observe that, by construction, (D_t) is simple random walk on $\{0, \ldots, n\}$ and $\tau_{\mathrm{coal}} = \tau^D_{\{0,n\}}$, the first time (D_t) hits $\{0, n\}$.

We use Markov's inequality, Theorem 2.4, to bound $\mathbb{P}_{(x,y)}[\tau^D_{\{0,n\}} > t]$. Denote by $D_0 = d_{x,y}$ the starting distance. By Wald's second equation (e.g. [Dur10, Example 4.1.6]),

$$\mathbb{E}_{(x,y)}\left[\tau^D_{\{0,n\}}\right] = d_{x,y}(n - d_{x,y}).$$

Applying Theorem 4.43 and Markov's inequality we get

$$\begin{aligned} d(t) &\leq \max_{x,y \in V} \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t] \\ &\leq \max_{x,y \in V} \frac{\mathbb{E}_{(x,y)}\left[\tau^D_{\{0,n\}}\right]}{t} \\ &= \max_{x,y \in V} \frac{d_{x,y}(n - d_{x,y})}{t} \\ &\leq \frac{n^2}{4t}, \end{aligned}$$

or:

Claim 4.44.

$$t_{\mathrm{mix}}(\varepsilon) \leq \frac{n^2}{4\varepsilon}.$$

By our diameter-based lower bound on mixing in Section 2.3.4, this bound gives the correct order of magnitude in n up to logarithmic factors. Indeed, the diameter is $\Delta = n/2$ and $\pi_{\min} = 1/n$ so that Claim 2.48 gives

$$t_{\mathrm{mix}}(\varepsilon) \geq \frac{n^2}{64 \log n},$$

for n large enough. Exercise 4.8 sketches a tighter lower bound.

Hypercube Let (Z_t) be lazy simple random walk on the n-dimensional hypercube $\mathbb{Z}_2^n := \{0, 1\}^n$ where $i \sim j$ if $\|i - j\|_1 = 1$. We denote the coordinates of Z_t by $(Z_t^{(1)}, \ldots, Z_t^{(n)})$. The coupling (X_t, Y_t) started at (x, y) is the following.

At each time t, pick a coordinate i uniformly at random in $[n]$, pick a bit value b in $\{0,1\}$ uniformly at random independently of the coordinate choice. Set *both* i coordinates to b, i.e., $X_t^{(i)} = Y_t^{(i)} = b$. Because of the way the updating is done, the chains stay put with probability $1/2$ at each time as required. Clearly the chains meet when all coordinates have been updated at least once. The following standard bound on the coupon collector problem is what is needed to conclude.

Lemma 4.45 (Coupon collecting)**.** *Let $\tau_{\rm coll}$ be the time it takes to update each coordinate at least once. Then, for any $c > 0$,*

$$\mathbb{P}\left[\tau_{\rm coll} > \lceil n \log n + cn \rceil\right] \leq e^{-c}.$$

Proof. Let B_i be the event that the i-th coordinate has not been updated by time $\lceil n \log n + cn \rceil$. Then

$$\begin{aligned}
\mathbb{P}[\tau_{\rm coll} > \lceil n \log n + cn \rceil] &\leq \sum_i \mathbb{P}[B_i] \\
&= \sum_i \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} \\
&\leq n \exp\left(-\frac{n \log n + cn}{n}\right) \\
&= e^{-c}.
\end{aligned}$$

■

Applying Theorem 4.43, we get

$$\begin{aligned}
d(\lceil n \log n + cn \rceil) &\leq \max_{x,y \in V} \mathbb{P}_{(x,y)}[\tau_{\rm coal} > \lceil n \log n + cn \rceil] \\
&\leq \mathbb{P}[\tau_{\rm coll} > \lceil n \log n + cn \rceil] \\
&\leq e^{-c}.
\end{aligned}$$

Hence for $c := c_\varepsilon > 0$ large enough:

Claim 4.46.

$$t_{\rm mix}(\varepsilon) \leq \lceil n \log n + c_\varepsilon n \rceil.$$

Again we get a quick lower bound using our diameter-based result from Section 2.3.4. Here $\Delta = n$ and $\pi_{\min} = 1/2^n$ so that Claim 2.48 gives

$$t_{\rm mix}(\varepsilon) \geq \frac{n^2}{12 \log n + (4 \log 2) n} = \Omega(n),$$

for n large enough. So the upper bound we derived above is off at most by a logarithmic factor in n.

Remark 4.47. *In fact the upper bound is only off by a factor of* 2 *[LPW06, Proposition 7.13]. See also [LPW06, Theorem 18.3] for an improved upper bound and a discussion of the so-called* cutoff phenomenon*. The latter refers to the fact that for this chain*

$$\lim_n \frac{\mathrm{t_{mix}}(\varepsilon)}{\mathrm{t_{mix}}(1-\varepsilon)} = 1,$$

for all ε*, i.e., the total variation distance to stationarity drops from approximately* 1 *to approximately* 0 *in a short time window.*

b**-ary tree** Let (Z_t) be lazy simple random walk on the ℓ-level rooted b-ary tree, $\widehat{\mathbb{T}}_{b,\ell}$. The root, 0, is on level 0 and the leaves, L, are on level ℓ. All vertices have degree $b+1$, except for the root which has degree b and the leaves which have degree 1. Hence the stationary distribution is

$$\pi(x) := \frac{\delta(x)}{2(n-1)},$$

where n is the number of vertices and $\delta(x)$ is the degree of x. We construct a coupling (X_t, Y_t) of this chain started at (x, y). Assume w.l.o.g. that x is no further from the root than y. The coupling has two stages:

- In the first stage, at each time, flip a fair coin. On heads, Y_t stays put and X_t moves one step chosen uniformly at random among its neighbors. Similarly, on tails, reverse the roles of X_t and Y_t and proceed in the same manner. Do this until X_t and Y_t are on the same level.

- In the second stage, i.e., once the two chains are on the same level, at each time first let X_t move as a lazy random walk on $\widehat{\mathbb{T}}_{b,\ell}$. Then let Y_t move in the same direction as X_t, i.e., if X_t moves closer to the root, so does Y_t and so on.

The key observation is the following. Let τ^* be the first time (X_t) visits the root *after visiting the leaves*. Then $\tau_{\mathrm{coal}} \leq \tau^*$ because, by construction, the two chains have necessarily met by time τ^*.

We use Markov's inequality, Theorem 2.4, to estimate $\mathbb{P}_{(x,y)}[\tau^* > t]$. To bound $\mathbb{E}_{(x,y)}[\tau^*]$ we note that it is less than the mean time for the walk to go from the root to the leaves and back. Let L_t be the level of X_t and let $\mathcal{N}$ be the corresponding network (where the conductances are equal to the number of edges on each level of the tree). In terms of (L_t), the quantity we seek to bound is the mean of $\tau_{0,\ell}$, the commute time of (L_t) between 0 and ℓ. By the commute time identity, Theorem 3.53,

$$\mathbb{E}_0[\tau_{0,\ell}] = c_{\mathcal{N}}\, \mathscr{R}(0 \leftrightarrow \ell), \tag{4.26}$$

where

$$c_{\mathcal{N}} = 2 \sum_{e=\{x,y\}\in\mathcal{N}} c(e) = 4(n-1),$$

where we simply counted the number of edges in $\widehat{\mathbb{T}}_{b,\ell}$ and the factor of 4 accounts for self-loops. Using network reduction techniques, we computed the effective resistance $\mathscr{R}(0 \leftrightarrow \ell)$ in Examples 3.34 and 3.35—without self-loops. Of course adding self-loops does not affect the effective resistance as we can use the same voltage and current. So, ignoring them, we get

$$\mathscr{R}(0 \leftrightarrow \ell) = \sum_{j=0}^{\ell-1} r(j, j+1) = \sum_{j=0}^{\ell-1} b^{-(j+1)} = \frac{1}{b} \cdot \frac{b^{-\ell}-1}{b^{-1}-1} \leq 1.$$

Finally, applying Theorem 4.43 and Markov's inequality and using (4.26), we get

$$\begin{aligned} d(t) &\leq \max_{x,y\in V} \mathbb{P}_{(x,y)}[\tau^* > t] \\ &\leq \max_{x,y\in V} \frac{\mathbb{E}_{(x,y)}[\tau^*]}{t} \\ &\leq \frac{\mathbb{E}_0[\tau_{0,\ell}]}{t} \\ &\leq \frac{4n}{t}, \end{aligned}$$

or:

Claim 4.48.

$$\mathrm{t_{mix}}(\varepsilon) \leq \frac{4n}{\varepsilon}.$$

This time the diameter-based bound is far off. We have $\Delta = 2\ell = \Theta(\log n)$ and $\pi_{\min} = 1/2(n-1)$ so that Claim 2.48 gives

$$\mathrm{t_{mix}}(\varepsilon) \geq \frac{(2\ell)^2}{12\log n + 4\log(2(n-1))} = \Omega(\log n),$$

for n large enough. Here is a better lower bound. Intuitively the mixing time is significantly greater than the squared diameter because the chain tends to be pushed away from the root: "going from one side of the root the other" typically takes time linear in n. Formally let x_0 be a leaf of $\widehat{\mathbb{T}}_{b,\ell}$ and let A be the set of vertices "on the other side of root (inclusively)," i.e., vertices whose graph distance from x_0 is at least ℓ. Then $\pi(A) \geq 1/2$ by symmetry. We claim that, started at x_0, the walk typically takes time linear in n to reach A. See Figure 4.6. Consider again

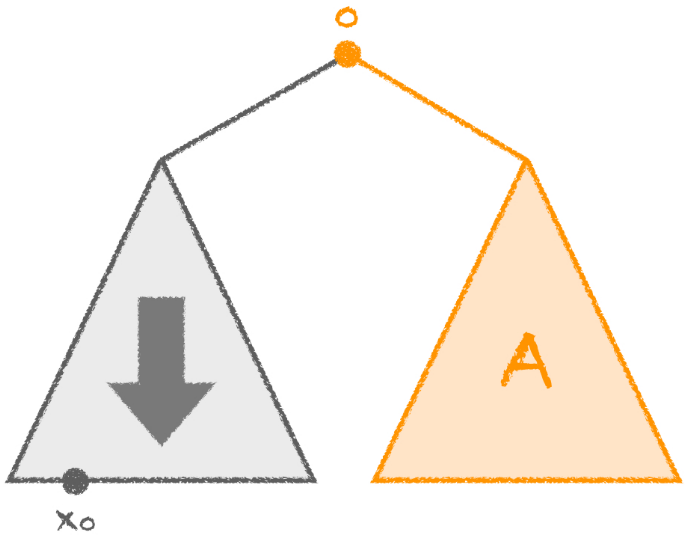

Figure 4.6: Setup for the lower bound on the mixing time on a b-ary tree. (Here $b = 2$.)

the level L_t of X_t. Using the expression for the effective resistance above, we have

$$\mathbb{P}_\ell[\tau_0 < \tau_\ell^+] = \frac{1}{c(\ell)\,\mathscr{R}(0 \leftrightarrow \ell)} = \frac{1}{2b^\ell} \cdot \frac{b-1}{1-b^{-\ell}} = \frac{b-1}{2b^\ell - 2} = O\left(\frac{1}{n}\right).$$

Hence, started from the leaves, the number of excursions back to the leaves needed to reach the root for the first time is geometric with success probability $O(n^{-1})$. Each such excursion takes time at least 2. So $P^t(x_0, A)$ is bounded above by the probability that at least one such excursion was successful among the first $t/2$ attempts. That is,

$$P^t(x_0, A) \leq 1 - \left(1 - O\left(n^{-1}\right)\right)^{t/2} < \frac{1}{2} - \varepsilon,$$

for all $t \leq \alpha_\varepsilon n$ with $\alpha_\varepsilon > 0$ small enough and

$$\|P^{\alpha_\varepsilon n}(x_0, \cdot) - \pi\|_{\mathrm{TV}} \geq |P^{\alpha_\varepsilon n}(x_0, A) - \pi(A)| > \varepsilon.$$

We have proved that $\mathrm{t_{mix}}(\varepsilon) \geq \alpha_\varepsilon n$.

4.3.3 Path coupling

Path coupling is a method for constructing Markovian couplings from "simpler" couplings. The building blocks are one-step couplings starting from pairs of initial states that are close in some dissimilarity graph.

Let (X_t) be an irreducible Markov chain on a finite state space V with transition matrix P and stationary distribution π. Assume that we are given a *dissimilarity graph* $H_0 = (V_0, E_0)$ on $V_0 := V$ with edge weights $w_0 : E_0 \to \mathbb{R}_+$. This graph need not have the same edges as the transition graph of (X_t). We extend w_0 to the *path metric*

dissimilarity graph, path metric

$$w_0(x, y) := \inf \left\{ \sum_{i=0}^{m-1} w_0(x_i, x_{i+1}) \,:\, x = x_0, \ldots, x_m = y \text{ is a path in } H_0 \right\},$$

where the infimum is over all paths connecting x and y in H_0. We call a path achieving the infimum a *minimum-weight path*. Let

minimum-weight path, weighted diameter

$$\Delta_0 := \max_{x,y} w_0(x, y),$$

be the *weighted diameter* of H_0.

Theorem 4.49 (Path coupling method)**.** *Assume that*

$$w_0(u, v) \geq 1,$$

for all $\{u, v\} \in E_0$. *Assume further that there exists* $\kappa \in (0, 1)$ *such that:*

- (Local couplings) *For all* x, y *with* $\{x, y\} \in E_0$*, there is a coupling* (X^*, Y^*) *of* $P(x, \cdot)$ *and* $P(y, \cdot)$ *satisfying the following* contraction *property*

$$\mathbb{E}[w_0(X^*, Y^*)] \leq \kappa \, w_0(x, y). \tag{4.27}$$

Then

$$d(t) \leq \Delta_0 \, \kappa^t,$$

or

$$\mathrm{t_{mix}}(\varepsilon) \leq \left\lceil \frac{\log \Delta_0 + \log \varepsilon^{-1}}{\log \kappa^{-1}} \right\rceil.$$

Proof. The crux of the proof is to extend (4.27) to arbitrary pairs of vertices.

Claim 4.50 (Global coupling)**.** *For all* $x, y \in V$*, there is a coupling* (X^*, Y^*) *of* $P(x, \cdot)$ *and* $P(y, \cdot)$ *such that* (4.27) *holds.*

Iterating the coupling in this claim immediately implies the existence of a coalescing Markovian coupling (X_t, Y_t) of P such that

$$\begin{aligned}
\mathbb{E}_{(x,y)}[w_0(X_t, Y_t)] &= \mathbb{E}_{(x,y)}\left[\mathbb{E}[w_0(X_t, Y_t) \,|\, X_{t-1}, Y_{t-1}]\right] \\
&\leq \mathbb{E}_{(x,y)}\left[\kappa\, w_0(X_{t-1}, Y_{t-1})\right] \\
&\leq \cdots \\
&\leq \kappa^t\, \mathbb{E}_{(x,y)}[w_0(X_0, Y_0)] \\
&= \kappa^t\, w_0(x, y) \\
&\leq \kappa^t\, \Delta_0.
\end{aligned}$$

By assumption, $\mathbb{1}_{\{x \neq y\}} \leq w_0(x, y)$ so that by the coupling inequality and Lemma 4.42, we have

$$d(t) \leq \bar{d}(t) \leq \max_{x,y} \mathbb{P}_{(x,y)}[X_t \neq Y_t] \leq \max_{x,y} \mathbb{E}_{(x,y)}[w_0(X_t, Y_t)] \leq \kappa^t\, \Delta_0,$$

which implies the theorem. It remains to prove Claim 4.50.

Proof of Claim 4.50. Fix $x', y' \in V$ such that $\{x', y'\}$ is *not* an edge in the dissimilarity graph H_0. The idea is to combine the local couplings on a minimum-weight path between x' and y' in H_0. Let $x' = x_0 \sim \cdots \sim x_m = y'$ be such a path. For all $i = 0, \ldots, m-1$, let $(Z^*_{i,0}, Z^*_{i,1})$ be a coupling of $P(x_i, \cdot)$ and $P(x_{i+1}, \cdot)$ satisfying the contraction property (4.27). Set $Z^{(0)} := Z^*_{0,0}$ and $Z^{(1)} := Z^*_{0,1}$. Then iteratively pick $Z^{(i+1)}$ according to the law $\mathbb{P}[Z^*_{i,1} \in \cdot \,|\, Z^*_{i,0} = Z^{(i)}]$. By induction on i, $(X^*, Y^*) := (Z^{(0)}, Z^{(m)})$ is then a coupling of $P(x', \cdot)$ and $P(y', \cdot)$. Formally, define the transition matrix

$$R_i(z^{(i)}, z^{(i+1)}) := \mathbb{P}[Z^*_{i,1} = z^{(i+1)} \,|\, Z^*_{i,0} = z^{(i)}],$$

and observe that

$$\sum_{z^{(i+1)}} R_i(z^{(i)}, z^{(i+1)}) = 1, \tag{4.28}$$

and

$$\sum_{z^{(i)}} P(x_i, z^{(i)}) R_i(z^{(i)}, z^{(i+1)}) = P(x_{i+1}, z^{(i+1)}), \tag{4.29}$$

by construction of the coupling $(Z^*_{i,0}, Z^*_{i,1})$. See Figure 4.7. The law of the full coupling $(Z^{(0)}, \ldots, Z^{(m)})$ is

$$\begin{aligned}
&\mathbb{P}[(Z^{(0)}, \ldots, Z^{(m)}) = (z^{(0)}, \ldots, z^{(m)})] \\
&\qquad = P(x_0, z^{(0)}) R_0(z^{(0)}, z^{(1)}) \cdots R_{m-1}(z^{(m-1)}, z^{(m)}).
\end{aligned}$$

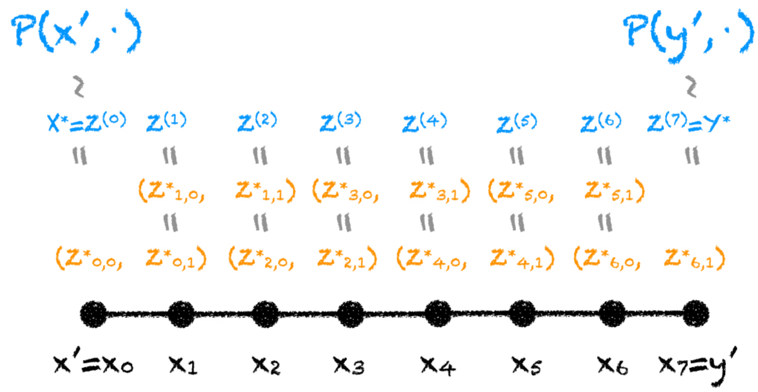

Figure 4.7: Coupling of $P(x',\cdot)$ and $P(y',\cdot)$ constructed from a sequence of local couplings $(Z^*_{0,0}, Z^*_{0,1}), \ldots, (Z^*_{0,m-1}, Z^*_{0,m-1})$.

Using (4.28) and (4.29) inductively gives

$$\mathbb{P}[X^* = z^{(0)}] = \mathbb{P}[Z^{(0)} = z^{(0)}] = P(x_0, z^{(0)}),$$
$$\mathbb{P}[Y^* = z^{(m)}] = \mathbb{P}[Z^{(m)} = z^{(m)}] = P(x_m, z^{(m)}),$$

as required.

By the triangle inequality for w_0, the coupling (X^*, Y^*) satisfies

$$\begin{aligned} \mathbb{E}[w_0(X^*, Y^*)] &= \mathbb{E}\left[w_0(X^{(0)}, X^{(m)})\right] \\ &\leq \sum_{i=0}^{m-1} \mathbb{E}\left[w_0(X^{(i)}, X^{(i+1)})\right] \\ &\leq \sum_{i=0}^{m-1} \kappa\, w_0(x_i, x_{i+1}) \\ &= \kappa\, w_0(x', y'), \end{aligned}$$

where, on the third line, we used (4.27) for adjacent pairs and the last line follows from the fact that we chose a minimum-weight path. ■

That concludes the proof of the theorem. ■

We illustrate the path coupling method in the next two subsections. See Exercise 4.9 for an optimal transport perspective on the path coupling method.

4.3.4 ▷ *Ising model: Glauber dynamics at high temperature*

Let $G = (V, E)$ be a finite, connected graph with maximal degree $\bar{\delta}$. Define $\mathcal{X} := \{-1, +1\}^V$. Recall that the (ferromagnetic) Ising model on V with *inverse temperature* β is the probability distribution over *spin configurations* $\sigma \in \mathcal{X}$ given by

$$\mu_\beta(\sigma) := \frac{1}{\mathcal{Z}(\beta)} e^{-\beta\mathcal{H}(\sigma)},$$

where

$$\mathcal{H}(\sigma) := -\sum_{i\sim j} \sigma_i\sigma_j,$$

is the *Hamiltonian* and

Hamiltonian

$$\mathcal{Z}(\beta) := \sum_{\sigma\in\mathcal{X}} e^{-\beta\mathcal{H}(\sigma)},$$

is the *partition function*. In this context, recall that vertices are often referred to as *sites*. The single-site Glauber dynamics of the Ising model is the Markov chain on $\mathcal{X}$ which, at each time, selects a site $i \in V$ uniformly at random and updates the spin σ_i according to $\mu_\beta(\sigma)$ conditioned on agreeing with σ at all sites in $V\backslash\{i\}$. Specifically, for $\gamma \in \{-1, +1\}$, $i \in V$, and $\sigma \in \mathcal{X}$, let $\sigma^{i,\gamma}$ be the configuration σ with the state at i being set to γ. Then, letting $n = |V|$, the transition matrix of the Glauber dynamics is

partition function

site

$$Q_\beta(\sigma, \sigma^{i,\gamma}) := \frac{1}{n}\cdot\frac{e^{\gamma\beta S_i(\sigma)}}{e^{-\beta S_i(\sigma)} + e^{\beta S_i(\sigma)}} = \frac{1}{n}\left\{\frac{1}{2} + \frac{1}{2}\tanh(\gamma\beta S_i(\sigma))\right\}, \quad (4.30)$$

where

$$S_i(\sigma) := \sum_{j\sim i} \sigma_j.$$

All other transitions have probability 0. Recall that this chain is irreducible and reversible with respect to μ_β. In particular μ_β is the stationary distribution of Q_β.

In this section we give an upper bound on the mixing time, $\mathrm{t_{mix}}(\varepsilon)$, of Q_β using path coupling. We first observe that:

Claim 4.51 (Glauber dynamics: lower bound on mixing)**.**

$$\mathrm{t_{mix}}(\varepsilon) = \Omega(n\log n), \qquad \forall\beta > 0.$$

Proof. Intuitively this is because, by a coupon collecting argument, it takes that long to update each site at least once. More formally let $\bar{\sigma}$ be the all-(-1) configuration and let A be the set of configurations where at least half of the sites are $+1$. Then, by symmetry, $\mu_\beta(A) = \mu_\beta(A^c) = 1/2$ where we assumed for simplicity that n is even. By definition of the total variation distance

$$d(t) \geq \|Q^t_\beta(\bar{\sigma}, \cdot) - \mu_\beta(\cdot)\|_{\mathrm{TV}} \geq |Q^t_\beta(\bar{\sigma}, A) - \mu_\beta(A)| = |Q^t_\beta(\bar{\sigma}, A) - 1/2|.$$

So it remains to show that by time $c\, n \log n$, for $c > 0$ small, the chain is unlikely to have reached A. That happens if, say, fewer than a third of the sites have been updated. By the computations in Example 2.15, for $t \leq c_\varepsilon\,(n/3)\log(n/3)$ for some $c_\varepsilon > 0$,

$$Q^t_\beta(\bar{\sigma}, A) \leq 1/2 - \varepsilon,$$

which proves the claim. ■

We say that the Glauber dynamics is *fast mixing* if $\mathrm{t_{mix}}(\varepsilon) = O(n \log n)$. In our main result, we show that this is the case when the inverse temperature β is small enough.

fast mixing

Claim 4.52 (Glauber dynamics: fast mixing at high temperature)**.**

$$\beta < \bar{\delta}^{-1} \implies \mathrm{t_{mix}}(\varepsilon) = O(n \log n).$$

Proof. We use path coupling. Let $H_0 = (V_0, E_0)$ where $V_0 := \mathcal{X}$ and $\{\sigma, \omega\} \in E_0$ if $\frac{1}{2}\|\sigma - \omega\|_1 = 1$ with unit w_0-weights on all edges. (To avoid confusion, we reserve the notation $\sim$ for adjacency in G.) Let $\{\sigma, \omega\} \in E_0$ differ at coordinate i. We construct a coupling (X^*, Y^*) of $Q_\beta(\sigma, \cdot)$ and $Q_\beta(\omega, \cdot)$. We first pick the same coordinate i_* to update. If i_* is such that all its neighbors in G have the same state in σ and ω, i.e., if $\sigma_j = \omega_j$ for all $j \sim i_*$, we update X^* from σ according to the Glauber rule and set $Y^* := X^*$. Note that this includes the case $i_* = i$. Otherwise, i.e. if $i_* \sim i$, we proceed as follows. From the state σ, the probability of updating site i_* to state $\gamma \in \{-1, +1\}$ is given by the expression in brackets in (4.30), and similarly for ω. Unlike the previous case, we cannot guarantee that the update is identical in both chains. However, in order to minimize the chances of increasing the distance between the two chains, we pick a uniform-$[-1, 1]$ variable U and set

$$X^*_{i_*} := \begin{cases} +1, & \text{if } U \leq \tanh(\beta S_{i_*}(\sigma)) \\ -1, & \text{o.w.} \end{cases}$$

and

$$Y^*_{i_*} := \begin{cases} +1, & \text{if } U \leq \tanh(\beta S_{i_*}(\omega)) \\ -1, & \text{o.w.} \end{cases}$$

We set $X_j^* := \sigma_j$ and $Y_j^* := \omega_j$ for all $j \neq i^*$. The expected distance between X^* and Y^* is then

$$\mathbb{E}[w_0(X^*, Y^*)] = 1 - \underbrace{\frac{1}{n}}_{\text{(a)}} + \underbrace{\frac{1}{n}\sum_{j\sim i}\frac{1}{2}\left|\tanh(\beta S_j(\sigma)) - \tanh(\beta S_j(\omega))\right|}_{\text{(b)}}, \quad (4.31)$$

where (a) corresponds to $i_* = i$ in which case $w_0(X^*, Y^*) = 0$ and (b) corresponds to $i_* \sim i$ in which case $w_0(X^*, Y^*) = 2$ with probability $\frac{1}{2}|\tanh(\beta S_{i_*}(\sigma)) - \tanh(\beta S_{i_*}(\omega))|$ by our coupling, and $w_0(X^*, Y^*) = 1$ otherwise. To bound (b), we note that for $j \sim i$

$$|\tanh(\beta S_j(\sigma)) - \tanh(\beta S_j(\omega))| = \tanh(\beta(s+2)) - \tanh(\beta s), \qquad (4.32)$$

where

$$s := S_j(\sigma) \wedge S_j(\omega).$$

The derivative of $\tanh$ is maximized at 0 where it is equal to 1. So the r.h.s. of (4.32) is $\leq 2\beta$. Plugging this back into (4.31), we get

$$\mathbb{E}[w_0(X^*, Y^*)] \leq 1 - \frac{1-\bar{\delta}\beta}{n} \leq \exp\left(-\frac{1-\bar{\delta}\beta}{n}\right) = \kappa\, w_0(\sigma, \omega),$$

where

$$\kappa := \exp\left(-\frac{1-\bar{\delta}\beta}{n}\right) < 1,$$

by assumption. The diameter of H_0 is $\Delta_0 = n$. By Theorem 4.49,

$$\mathrm{t_{mix}}(\varepsilon) \leq \left\lceil \frac{\log \Delta_0 + \log \varepsilon^{-1}}{\log \kappa^{-1}} \right\rceil = \left\lceil \frac{n(\log n + \log \varepsilon^{-1})}{1 - \bar{\delta}\beta} \right\rceil,$$

which implies the claim. ■

Remark 4.53. *A slighlty more careful analysis shows that the condition $\bar{\delta}\tanh(\beta) < 1$ is enough for the claim to hold. See [LPW06, Theorem 15.1].*

We will show in Section 6.2.6 that, at low temperatures, the Glauber dynamics on bounded-degree graphs may be *slow mixing*, i.e., for certain families of bounded-degree graphs $\mathrm{t_{mix}}(\varepsilon) = \Omega(n^\alpha)$ where $\alpha > 1$ depends on β and $\bar{\delta}$.

4.3.5 ▷ *Colorings: from approximate sampling to approximate counting*

To be written. See [MU05, Section 10.3] and [LPW06, Sections 14.3 and 14.4]. See also [JS97].

4.4 Duality

To be written. See [AF, Section 14.3] and [Ald13, Lig].

4.4.1 Graphical representations and coupling of initial configurations

4.4.2 ▹ *Interacting particles: voter model on the complete graph and on the line*

Exercises

Exercise 4.1 (Harris' inequality: alternative proof)**.** We say that $f : \mathbb{R}^n \to \mathbb{R}$ is *nondecreasing* if it is nondecreasing in each variable while keeping the other variables fixed.

- *(Chebyshev's association inequality)* Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be nondecreasing and let X be a real random variable. Show that

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

  [Hint: Consider the quantity $(f(X) - f(X'))(g(X) - g(X'))$ where X' is an independent copy of X.]

- *(Harris' inequality)* Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ be nondecreasing and let $X = (X_1, \ldots, X_n)$ be independent real random variables. Show by induction on n that

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

Exercise 4.2. Provide the details for Example 4.31.

Exercise 4.3 (FKG: sufficient conditions)**.** Let $\mathcal{X} := \{0,1\}^F$ where F is finite and let μ be a positive probability measure on $\mathcal{X}$. We use the notation introduced in the proof of Holley's inequality.

a) To check the FKG condition, show that it suffices to check that, for all $x \leq y \in \mathcal{X}$ and $i \in F$,

$$\frac{\mu(y^{i,1})}{\mu(y^{i,0})} \geq \frac{\mu(x^{i,1})}{\mu(x^{i,0})}.$$

   [Hint: Write $\mu(\omega \vee \omega')/\mu(\omega)$ as a telescoping product.]

b) To check the FKG condition, show that it suffices to check (4.12) only for those $\omega, \omega' \in \mathcal{X}$ such that $\|\omega - \omega'\|_1 = 2$ and neither $\omega \leq \omega'$ nor $\omega' \leq \omega$. [Hint: Use a).]

Exercise 4.4 (FKG and strong positive association)**.** Let $\mathcal{X} := \{0,1\}^F$ where F is finite and let μ be a positive probability measure on $\mathcal{X}$. For $\Lambda \subseteq F$ and $\xi \in \mathcal{X}$, let

$$\mathcal{X}^\xi_\Lambda := \{\omega_\Lambda \times \xi_{\Lambda^c} \,:\, \omega_\Lambda \in \{0,1\}^\Lambda\},$$

where $\omega_\Lambda \times \xi_{\Lambda^c}$ agrees with ω on coordinates in Λ and with ξ on coordinates in $F\backslash\Lambda$. Define the measure μ^ξ_Λ over $\{0,1\}^\Lambda$ as

$$\mu^\xi_\Lambda(\omega_\Lambda) := \frac{\mu(\omega_\Lambda \times \xi_{\Lambda^c})}{\mu(\mathcal{X}^\xi_\Lambda)}.$$

That is, μ^ξ_Λ is essentially μ conditioned on agreeing with ξ on $F\backslash\Lambda$. The measure μ is said to be *strongly positively associated* if $\mu^\xi_\Lambda(\omega_\Lambda)$ is positively associated for all Λ and ξ. Prove that the FKG condition is equivalent to strong positive association. [Hint: Use Exercise 4.3 as well as the FKG inequality.]

strong positive association

Exercise 4.5 (Triangle-freeness: a second proof)**.** Consider again the setting of Section 4.2.5.

a) Let e_t be the minimum number of edges in a t-vertex union of k not mutually vertex-disjoint triangles. Show that, for any $k \geq 2$ and $k \leq t < 3k$, it holds that $e_t > t$.

b) Use Exercise 2.11 to give a second proof of the fact that $\mathbb{P}[X_n = 0] \to e^{-\lambda^3/6}$.

Exercise 4.6 (Primal and dual crossings)**.** *Modify the proof of Lemma 2.12 to prove Lemma 4.39.*

Exercise 4.7 (Square-root trick)**.** *Let μ be an FKG measure on $\{0,1\}^F$ where F is finite. Let A_1 and A_2 be increasing events with $\mu(A_1) = \mu(A_2)$. Show that*

$$\mu(A_1) \geq 1 - \sqrt{1 - \mu(A_1 \cup A_2)}.$$

Exercise 4.8 (Mixing on cycles: lower bound)**.** Let (Z_t) be lazy, simple random walk on the cycle of size n, $\mathbb{Z}_n := \{0, 1 \ldots, n-1\}$, where $i \sim j$ if $|j - i| = 1 \pmod{n}$.

a) Let $A = \{n/2, \ldots, n-1\}$. By coupling (Z_t) with lazy, simple random walk on $\mathbb{Z}$, show that

$$P^{\alpha n^2}(n/4, A) < \frac{1}{2} - \varepsilon,$$

for $\alpha \geq \alpha_\varepsilon > 0$. [Hint: Use Kolmogorov's maximal inequality (e.g. [Dur10, Theorem 2.5.2]).]

b) Deduce that

$$t_{\mathrm{mix}}(\varepsilon) \geq \alpha_\varepsilon n^2.$$

Exercise 4.9 (Path coupling and optimal transport)**.** Let V be a finite state space and let P be an irreducible transition matrix on V with stationary distribution π. Let w_0 be a metric on V. For probability measures μ, ν on V, let

$$W_0(\mu, \nu) := \inf \left\{ \mathbb{E}[w_0(X, Y)] \,:\, (X, Y) \text{ is a coupling of } \mu \text{ and } \nu \right\},$$

be the so-called *transportation metric* or *Wasserstein distance* between μ and ν.

Wasserstein distance

a) Show that W_0 is a metric. [Hint: Proof of Claim 4.50.]

b) Assume that the conditions of Theorem 4.49 hold. Show that for any probability measures μ, ν

$$W_0(\mu P, \nu P) \leq \kappa\, W_0(\mu, \nu).$$

c) Use a) and b) to prove Theorem 4.49.

Notes

Bibliographic Remarks

Section 4.1 The coupling method is generally attributed to Doeblin [Doe38]. The standard reference on coupling is [Lin02]. See that reference for a history of coupling and a facsimile of Doeblin's paper. See also [dH] where much of the material of Section 4.1 is borrowed.

Section 4.2 Strassen's theorem is due to Strassen [Str65]. Harris' inequality is due to Harris [Har60]. The FKG inequality is due to Fortuin, Kasteleyn, and Ginibre [FKG71]. A "four-function" version of Holley's inequality, which also extends to distributive lattices, was proved by Ahlswede and Daykin [AD78]. See e.g. [AS11, Section 6.1]. An exposition of submodularity and its connections to convexity can be found in [Lov83]. For more on Markov random fields, see e.g. [RAS]. Section 4.2.5 follows [AS11, Sections 8.1, 8.2, 10.1]. Janson's inequality is due to Janson [Jan90]. Boppana and Spencer [BS89] gave the proof presented here. For more on Janson's inequality, see [JLR11, Section 2.2]. The presentation in Section 4.2.6 follows closely [BR06b, Sections 3 and 4]. See also [BR06a, Chapter 3]. Broadbent and Hammersley [BH57, Ham57] initiated the study of the critical value of percolation. Harris' theorem was proved by Harris [Har60] and

Kesten's theorem was proved two decades later by Kesten [Kes80], confirming non-rigourous work of Sykes and Essam [SE64]. The RSW theorem was obtained independently to Russo [Rus78] and Seymour and Welsh [SW78]. The proof we gave here is due to Bollobás and Riordan [BR06b]. Another short proof of a version of the RSW theorem for critical site percolation on the triangular lattice was given by Smirnov. See e.g. [Ste]. The type of "scale invariance" seen in the RSW theorem plays a key role in the contemporary theory of critical two-dimensional percolation and of two-dimensional lattice models more generally. See e.g. [Law05, Gri10a].

Section 4.3 The material in Section 4.3 borrows heavily from [LPW06, Chapters 5, 14, 15] and [AF, Chapter 12]. Aldous [Ald83] was the first author to make explicit use of coupling to bound total variation distance to stationarity of finite Markov chains. The link between couplings of Markov chains and total variation distance was also used by Griffeath [Gri75] and Pitman [Pit76]. Path coupling is due to Bubley and Dyer [BD97]. The optimal transport perspective on the path coupling method in Exercise 4.9 is from [LPW06, Chapter 14]. For more on optimal transport, see e.g. [Vil09]. The main result in Section 4.3.4 is taken from [LPW06, Theorem 15.1]. The connection between sampling and counting was first considered by Jerrum, L. Valiant and V. Vazirani [JVV86]. For more on this topic, see e.g. [Sin93].

Section 4.4 For much more on interacting particle systems, see [Lig85].