# Markov chain mixing: an introduction, and applications to statistical mechanics

*Fabio Toninelli – TU Wien*

FWF Austrian Science Fund

# Plan (tentative)



1. Lecture I
   - Motivations: MC, sampling, enumeration
   - Examples: Ising, (monomer)-dimer, graph coloring, matchings,…
   - Speed of convergence: $T_{\mathrm{mix}}$, spectral gap
2. Lecture II
   - Some classical tools: (path) coupling, canonical paths, bottleneck ratio,…
   - Examples: fast mixing for monomers-dimers with large monomer density
3. Lecture III
   - Monotone MC.
   - A case study: Glauber dynamics of lozenge tilings.

Goals:

- sample (exactly or approximately) from a *complicated* probability measure $\pi(\sigma) \propto w(\sigma), \sigma \in \Omega, |\Omega| \gg 1$

  Ex 1 : $\Omega = \{\text{proper colorings of graph } G\}, w \equiv 1$

  Ex 2: $\Omega = \{\text{matchings of } G\}, w = e^{\lambda|\{\text{edges of } \sigma\}|}$

  Ex 3: Ising. $G = (V, E), \Omega = \{-1, +1\}^V, w(\sigma) = e^{\beta \sum_{(x,y) \in E} \sigma_x \sigma_y}$

- Enumerate (exactly or approximately). More generally, compute/estimate $\sum_{\sigma \in \Omega} w(\sigma)$

  Ex 1: $\sum_\sigma w(\sigma) = \#\{\text{proper colorings of } G\}$

  Ex 3: $\sum_\sigma w(\sigma) = Z_{\beta,G}$

Other classical example: card shuffling. $\Omega = S_n$ (symmetric group)

# Introduction



Common features: many "degrees of freedom", $\Omega$ large.

Hard to sample directly. Better option: MC simulation

(e.g. graph matchings: simple-minded "rejection sampling" vs MCMC)

Basic idea:

- define ergodic MC $X = (X_n)_{n\geqslant 0}$ on $\Omega$ with stationary distribution $\pi$
- wait "long enough": $\mathbb{P}(X_n = \sigma) \xrightarrow{n\to\infty} \pi(\sigma)$
- how long is "long enough"?
- math goal: results of the type "if $n > f(|\Omega|, \varepsilon)$ then $\|\mathbb{P}(X_n = \cdot) - \pi(\cdot)\| < \varepsilon$. Useful for simulations if $f$ does not grow too fast with $|\Omega|, \varepsilon^{-1}$.

NB: efficient algorithm vs. "physically relevant" algorithm

**Meta-theorem:** fast approximate sampling algorithm implies fast approximate counting algorithm.

Example: $\mathcal{M}(G) = \{\text{matchings of } G = (V,E)\}, \pi = $ uniform

(For those interested: see [M. Jerrum's book: “Counting, sampling and integrating: Algorithms and complexity”])

- assume: MC samples from $\pi$ with error $\varepsilon$ in time $T(|V|,|E|,\varepsilon)$ (for every $G$)

**Meta-theorem:** fast approximate sampling algorithm implies fast approximate counting algorithm.

Example: $\mathcal{M}(G) = \{\text{matchings of } G = (V,E)\}, \pi = \text{uniform}$

(For those interested: see [M. Jerrum's book: "Counting, sampling and integrating: Algorithms and complexity"])

- assume: MC samples from $\pi$ with error $\varepsilon$ in time $T(|V|,|E|,\varepsilon)$ (for every $G$)
- then: we can approximate $|\mathcal{M}(G)|$ in time $\mathcal{T}(|V|,|E|,\varepsilon) \lesssim |E|^2 \times \varepsilon^{-2} \times T(|V|,|E|,\varepsilon/(6|E|))$

**Meta-theorem:** fast approximate sampling algorithm implies fast approximate counting algorithm.

Example: $\mathcal{M}(G) = \{\text{matchings of } G = (V,E)\}, \pi = $ uniform

(For those interested: see [M. Jerrum's book: "Counting, sampling and integrating: Algorithms and complexity"])

- assume: MC samples from $\pi$ with error $\varepsilon$ in time $T(|V|,|E|,\varepsilon)$ (for every $G$)
- then: we can approximate $|\mathcal{M}(G)|$ in time $\mathcal{T}(|V|,|E|,\varepsilon) \lesssim |E|^2 \times \varepsilon^{-2} \times T(|V|,|E|,\varepsilon/(6|E|))$
- Note: if $T(|V|,|E|,\varepsilon)$ is polynomial in $|V|,|E|,\varepsilon^{-1}$ then $\mathcal{T}(|V|,|E|,\varepsilon)$ is, too.

# Markov chain notations



- $P$: transition matrix. $P = (P(\sigma,\eta))_{\sigma,\eta\in\Omega}$

  $\mathbb{P}(X_n = \eta | X_{n-1} = \sigma) = P(\sigma,\eta)$
- if $\sigma \in \Omega, \mathbb{P}_\sigma$: law of process $X$ started from $X_0 = \sigma$. If $\nu$ is distribution on $\Omega$, then $\mathbb{P}_\nu$: law of process started from $X_0 \sim \nu$. $P_\nu^t$: law of $X_t$ started from $X_0 \sim \nu$.
- assume irreducibility (ergodicity) + aperiodicity. $\pi$ unique stationary distribution

  $$\pi P = \pi, \text{ that is, } \mathbb{P}_\pi(X_n = \sigma) = \pi(\sigma)$$

  General theory: $\mathbb{P}_\nu(X_n = \sigma) = (\nu P^n)(\sigma) \xrightarrow{n\to\infty} \pi(\sigma)$ exp. fast (not quantitative)
- We will deal with reversible MC, i.e., $\pi(\sigma)P(\sigma,\eta) = \pi(\eta)P(\eta,\sigma)$

  Implies stationarity and time-reversal symmetry

# Example 1: Ising



$G=(V,E)$ finite graph. $\Omega=\{-1,+1\}^V$. $\sigma=(\sigma_x)_{x\in V}\in\Omega$, $\beta>0$.

Boltzmann distribution: $\pi(\sigma)=\frac{e^{\beta\sum_{(x,y)\in E}\sigma_x\sigma_y}}{Z_{\beta,G}}$

Define MC $(\sigma(n))_{n\geqslant 0}$ on $\Omega$ as follows:

- start from $\sigma(0)\in\Omega$
- at step $n$, choose $x\in V$ uniformly at random
- define $\sigma(n)$ as $\sigma(n)_y=\sigma(n-1)_y$ if $y\neq x$ and sample $\sigma(n)_x$ from $\pi(\cdot|\sigma_{V\setminus\{x\}}=\sigma(n-1)_{V\setminus\{x\}})$. Explicitly, $\sigma(n)_x$ is chosen to be $\varepsilon=\pm1$ with probability

$$\frac{e^{\beta\varepsilon\sum_{y:(x,y)\in E}\sigma(n-1)_y}}{\sum_{\varepsilon=\pm1}e^{\beta\varepsilon\sum_{y:(x,y)\in E}\sigma(n-1)_y}}=\frac{e^{\beta\varepsilon\sum_{y:(x,y)\in E}\sigma(n-1)_y}}{2\cosh(\beta\sum_{y:(x,y)\in E}\sigma(n-1)_y)} \qquad (1)$$

# Example 1: Ising

**Exercise 1.** Write down the transition matrix $P$

**Exercise 2.** Prove that the MC is reversible: $\pi(\sigma)P(\sigma,\eta)=\pi(\eta)P(\eta,\sigma)$

**Exercise 3.** Prove that the MC is aperiodic irreducible: for every $\sigma,\sigma'\in W$ there exists a path $\sigma(n), 0\leqslant n\leqslant M$ with $\sigma(0)=\sigma, \sigma(M)=\sigma'$ and $P(\sigma_n,\sigma_{n+1})>0$.

**Remark.** The transition rate

$$\frac{e^{\beta\varepsilon\sum_{y:(x,y)\in E}\sigma(n-1)_y}}{2\cosh(\beta\sum_{y:(x,y)\in E}\sigma(n-1)_y)} \tag{2}$$

is a local function of $\sigma(n-1)$ around $x$. Does not require to compute $Z_{\beta,G}$!

# Example 2: monomer-dimer model



$G=(V,E)$ finite graph. $\lambda>0$ "fugacity parameter". $\Omega=\mathcal{M}(G)=\{\text{matchings of } G\}$.

NB: not just perfect matchings.

Each $M\in\Omega$ is a subset of $E$. $e\in M$: "dimer at $e$". $x\in V$ unmached in $M$: "monomer at $x$".

Call $m(M)$ the number of monomers of $M$. Boltzmann distribution:

$$\pi(M)=\frac{\lambda^{m(M)}}{\sum_{M'\in\Omega}\lambda^{m(M')}}=\frac{\lambda^{m(M)}}{Z_{\lambda,G}}. \tag{3}$$

Note: for $\lambda\to 0$, $\pi$ supported by maximal matchings.

# Example 2: monomer-dimer model



Define MC $(M(n))_{n\geqslant 0}$ on $\Omega$ as follows:

- start from $M(0)\in\Omega$
- at step $n$, choose $e=(x,y)\in E$ uniformly at random
- call $\tilde{M}(n-1)=M(n-1)\setminus\{(x,y)\}$. (NB: it may or may not coincide with $M(n-1)$)
- if either $x$ or $y$ belong to an edge of $\tilde{M}(n-1)$, do nothing
- otherwise, let $M(n)=\tilde{M}(n-1)\cup\{(x,y)\}$ with probability $\frac{1}{1+\lambda^2}$

  and $M(n)=\tilde{M}(n-1)$ with probability $\frac{\lambda^2}{1+\lambda^2}$

  **Exercise 4.** Check irreducibility, aperiodicity and reversibility.

**Remark.** If $\lambda=0$, irreducibility can fail. There are better algorithms

# Example 3: planar dimer model

$G=(V,E)$ finite, **planar** graph. $\Omega=\{\text{perfect matchings of } G\}$. Given $e\in E$, let $w_e>0$: edge weight. Assume that $G$ admits perfect matchings.

E.g.: $G=\{1,\dots,N\}\times\{1,\dots,2L\}$ with nearest-neighbor edges.

If $M\in\Omega$, $e\in M$ we say that "$e$ is occupied by a dimer in $M$"

Boltzmann distribution $\pi(M)=\frac{\prod_{e\in M} w_e}{\sum_{M'\in\Omega}\prod_{e\in M'} w_{e'}}$

Given face $f$ of $G$ and $M\in\Omega$, write $\partial f$ for the collection of edges around $f$.

Say that "a rotation at $f$ is possible in $M$" if every second edge of $\partial f$ belongs to $M$. In this case, call $M^f\in\Omega$ the configuration where occupied/empty edges of $\partial f$ switch roles, and nothing else changes.

NB: rotation possible only at faces $f$ with $|\partial f|$ even.

# Example 3: planar dimer model

Introduce MC $(M(n))_{n\geqslant 0}$ as follows:

- start from $M(0) \in \Omega$
- at each step, choose an (inner) face $f$ uniformly at random
- if a rotation at $f$ is possible in $M(n-1)$, then let $M(n) = M(n-1)^f$ with probability

$$p = \frac{\prod_{e\in\partial f: e\notin M(n-1)} w_e}{\prod_{e\in\partial f: e\notin M(n-1)} w_e + \prod_{e\in\partial f: e\in M(n-1)} w_e} \tag{4}$$

  and $M(n) = M(n-1)$ with probability $1-p$.
- if the rotation at $f$ is not possible in $M(n-1)$, do nothing.

**Exercise 5.** (Non trivial). If $G$ is bipartite, then the MC is irreducible.

# Speed of convergence



General theory: irreducible, aperiodic MC on finite state space:

$$|\mathbb{P}_\nu(X_n=\sigma)-\pi(\sigma)|\leqslant Ce^{-cn}. \tag{5}$$

$C, c^{-1}$ depend on $\Omega$, can diverge as $|\Omega|\to\infty$.

PB: quantify speed of convergence. Several classical ways, among which:

1. Total variation distance mixing time $(T_{\text{mix}})$
2. Mixing time w.r.t other distances ($L^p$, Hellinger distance, separation distance…)
3. spectral gap/relaxation time $T_{\text{rel}}$
4. log-Sobolev constant, relative entropy

In these lectures, we focus on 1 and 3. “Coupling arguments” adapt well to $T_{\text{mix}}$, spectral/variational arguments to $T_{\text{rel}}$.

# (Total variation) mixing time

(classical book: Levin and Peres, Markov chains and Mixing times)

**Definition (Total variation distance)** *Let $\mu,\nu$ be probability measures on finite set $\Omega$. Then,*

$$\|\nu-\mu\| = \frac{1}{2}\sum_{x\in\Omega} |\mu(x)-\nu(x)| = \max_{A\subset\Omega}|\mu(A)-\nu(A)| = \inf\{\mathbb{P}(X\neq Y) : X\sim\mu, Y\sim\nu\}. \tag{6}$$

**Exercise 6.** It is indeed a distance. Try to prove the equalities.

**Definition (Mixing time)** *Let $\varepsilon\in(0,1)$, $X$ be Markov Chain on $\Omega$. Then,*

$$T_{\text{mix}}(\varepsilon) = \inf\left\{n\geqslant 0 : \max_{x\in\Omega}\|P_x^n-\pi\| < \varepsilon\right\}. \tag{7}$$

NB: worst-case initial condition ($\max_{x\in\Omega}$).

# Spectral gap/Relaxation time

Let $X$ be a **reversible** MC on $\Omega$.

Define $\langle f, g\rangle_\pi := \sum_{x\in\Omega} f(x)g(x)\pi(x)$ for $f, g: \Omega \mapsto \mathbb{R}$.

Reversibility implies $\langle f, Pg\rangle_\pi = \langle Pf, g\rangle_\pi$: $P$ is self-adjoint

Easy/classical facts:

- Spectrum $(\lambda_i)_{i\leqslant|\Omega|}$ is real. $\lambda_1 \geqslant \lambda_2 \geqslant \cdots$
- $|\lambda_i| \leqslant 1$
- Up to redefining $P \to \frac{(I+P)}{2}$, we can assume that $0 \leqslant \lambda_i \leqslant 1$
- Irreducibility $\Rightarrow \lambda_1 = 1$ is simple.

Note: $P\mathbf{1} = \mathbf{1}$, $\pi P = P$

**Definition** **(Spectral gap)** *Let $X$ be reversible. We define the spectral gap equivalently as*

1. $\gamma = 1 - \lambda_2 > 0$. *Relaxation time* $T_{\mathrm{rel}} := \frac{1}{\gamma}$.
2. *(Variational principle)*

$$\gamma = \inf_{f:\Omega\mapsto\mathbb{R}, \mathrm{Var}_\pi(f)\neq 0} \frac{\mathcal{E}(f)}{\mathrm{Var}_\pi(f)} \tag{8}$$

*where* $\mathcal{E}(f) = \langle (I-P)f, f\rangle_\pi = \frac{1}{2}\sum_{x,y\in\Omega} \pi(x)P(x,y)|f(x)-f(y)|^2$ *(Dirichlet form)*

3. *($L^2$ relaxation)* $\gamma$ *is the best (i.e. largest) constant such that for all* $n\in\mathbb{N}, f:\Omega\mapsto\mathbb{R}$

$$\mathrm{Var}_\pi(P^n f) \leqslant \mathrm{Var}_\pi(f)(1-\gamma)^{2n} \tag{9}$$

# $T_{\mathrm{mix}}, T_{\mathrm{rel}}$: basic facts

- $T_{\mathrm{mix}}(\varepsilon)$ decreases with $\varepsilon$ (obvious) and

$$T_{\mathrm{mix}}(\varepsilon) \leqslant \lceil \log_2(\varepsilon^{-1}) \rceil T_{\mathrm{mix}}\left(\frac{1}{4}\right) \tag{10}$$

  that is, at time $m\, T_{\mathrm{mix}}\left(\frac{1}{4}\right)$ the variation distance from equilibrium is at most $2^{-m}$.

- General $T_{\mathrm{mix}} / T_{\mathrm{rel}}$ comparison:

$$(T_{\mathrm{rel}} - 1)\log\left(\frac{1}{2\varepsilon}\right) \leqslant T_{\mathrm{mix}}(\varepsilon) \leqslant T_{\mathrm{rel}}\log\left(\frac{1}{\varepsilon \min_{x\in\Omega}\pi(x)}\right).$$

- **Meta-statement:** upper bounds on $T_{\mathrm{mix}}, T_{\mathrm{rel}}$ are harder to get than lower bounds.

  In fact, $T_{\mathrm{mix}}(\varepsilon) \gtrsim_{\varepsilon} T_{\mathrm{rel}} \geqslant \frac{\mathrm{Var}_\pi(f)}{\mathcal{E}(f)}$ for any test function $f: \Omega \mapsto \mathbb{R}$.

Advantage of total variation distance in the definition of $T_{\text{mix}}$: variational characterization of $\|\mu-\nu\|$ in terms of optimal coupling.

Coupling methods powerful in estimating $T_{\text{mix}}$.

Warm-up result:

**Proposition** *For every $x, y \in \Omega$, let $(X_n, Y_n)_{n\geqslant 0}$ be a coupling of the processes started at $(x, y)$. Call $\tau_{\text{couple}} = \min\{n : X_n = Y_n\}$ and $\mathbb{P}_{x,y}$ the distribution of $(X_n, Y_n)_{n\geqslant 0}$.*

*If $\max_{x,y\in\Omega}\mathbb{P}_{x,y}(\tau_{\text{couple}} > n) < \varepsilon$, then $T_{\text{mix}}(\varepsilon) \leqslant n$.*

Example: lazy RW on $\{1,\ldots,n\}, T_{\text{mix}} \lesssim n^2$.

**Proof.** Assume wlog that $X_n = Y_n$ for $n \geqslant \tau_{\text{couple}}$. Then,

$$\|P_x^n - P_y^n\| = \inf\{\mathbb{P}(X \neq Y) : X \sim P_x^n, Y \sim P_y^n\} \leqslant \mathbb{P}(X_n \neq Y_n) = \mathbb{P}_{x,y}(\tau_{\text{couple}} > n). \tag{11}$$

We deduce $\max_{x,y\in\Omega} \|P_x^n - P_y^n\| < \varepsilon$.

Now note that

$$\begin{aligned} \|P_x^n - \pi\| &= \max_{A\subset\Omega} |P_x^n(A) - \pi(A)| = \max_{A\subset\Omega} \left| \sum_{y\in\Omega} \pi(y)[P_x^n(A) - P_y^n(A)] \right| \\ &\leqslant \sum_{y\in\Omega} \pi(y) \max_{A\subset\Omega} |P_x^n(A) - P_y^n(A)| = \sum_{y\in\Omega} \pi(y) \|P_x^n - P_y^n\| < \varepsilon. \end{aligned}$$

□

Drawback: in general, difficult to control $\tau_{\text{couple}}$. Better idea: “path coupling”.

Let $\mathcal{G}=(\mathcal{V},\mathcal{E})$ be connected a graph with $\mathcal{V}=\Omega$. Assign $\mathcal{E}\ni e\mapsto \ell_e\geqslant 1$ “length”.

$\rho$: induced graph distance on $\mathcal{G}$. $\rho_K$: Kantorovich transportation distance between distributions on $\Omega$:

$$\rho_K(\mu,\nu)=\inf\{\mathbb{E}(\rho(X,Y)): X\sim\mu, Y\sim\nu)\}.$$

Since $\rho(x,y)\geqslant \mathbf{1}_{x\neq y}$, it follows $\|\mu-\nu\|\leqslant\rho_K(\mu,\nu)$.

**Theorem** **(Bubley-Dyer '97)** *Assume that there exists* $\alpha < 1$ *and, for every* $x \sim^{\mathcal{G}} y$*, a coupling* $(X,Y)$ *of* $P(x,\cdot), P(y,\cdot)$ *such that* $\mathbb{E}(\rho(X,Y)) \leqslant \rho(x,y)e^{-\alpha}$*. Then,* $\rho_K(\mu P, \nu P) \leqslant e^{-\alpha}\rho_K(\mu,\nu)$.

Immediate consequence:

$$\begin{aligned} \|P^n_\nu - P^n_\mu\| \leqslant \rho_K(\nu P^n, \mu P^n) &= \rho_K(\nu P^{n-1}P, \mu P^{n-1}P) \\ &\leqslant e^{-\alpha}\rho_K(\nu P^{n-1}, \mu P^{n-1}) \\ \text{(iterate)} &\leqslant e^{-\alpha n}\rho_K(\mu,\nu) \leqslant e^{-\alpha n}\mathrm{diam}_\rho(\Omega) \end{aligned}$$

**Corollary** *The mixing time is* $T_{\mathrm{mix}}(\varepsilon) \leqslant \frac{\log\left(\frac{\mathrm{diam}_\rho(\Omega)}{\varepsilon}\right)}{\alpha}$ *(one can also deduce* $\gamma \geqslant 1 - e^{-\alpha}$*).*

Advantage: needs to check contraction only after 1 step, and for $x \sim^{\mathcal{G}} y$.

Drawback: if condition fails even for one pair $x \sim^{\mathcal{G}} y$, theorem does not apply.

Recall: $G=(V,E)$ finite graph. $\lambda>0$ "fugacity parameter". $\Omega=\mathcal{M}(G)=\{\text{matchings of } G\}$.

Boltzmann distribution: $\pi(M)\propto\lambda^{m(M)}$.

- at step $n$, choose $e=(x,y)\in E$ uniformly at random
- call $\tilde{M}(n-1)=M(n-1)\setminus\{(x,y)\}$.
- if either $x$ or $y$ belong to an edge of $\tilde{M}(n-1)$, do nothing
- otherwise, let $M(n)=\tilde{M}(n-1)\cup\{(x,y)\}$ with probability $\frac{1}{1+\lambda^2}$, $M(n)=\tilde{M}(n-1)$ else

Define $\mathcal{G}=(\Omega,\mathcal{E})$ with $(M,M')\in\mathcal{E}$ iff $M=M'\cup\{e\}$. Set $\ell_{(M,M')}=1$, i.e., $\rho$ is usual graph distance, i.e., $\rho(M,M')=\sum_e|\mathbf{1}_{e\in M}-\mathbf{1}_{e\in M'}|$. Let $\Delta_G$ be the maximal degree of $G$.

**Proposition** *Fix $\Delta>0$. There exists $\lambda_0(\Delta)<\infty$ such that path coupling works with $\alpha=1/(2|E|)$, for all $\lambda>\lambda_0$ and graphs with $\Delta_G\leqslant\Delta$.*

Proof

- let $M, M' \in \Omega, e = (x,y) \in E, e \notin M', M = M' \cup \{e\}$. Note $\rho(M, M') = 1$.

Proof

- let $M, M' \in \Omega, e = (x, y) \in E, e \notin M', M = M' \cup \{e\}$. Note $\rho(M, M') = 1$.
- Goal: couple $X \sim P(M, \cdot), Y \sim P(M', \cdot)$ so that $\mathbb{E}(\rho(X, Y)) \leqslant e^{-\alpha} < 1$ with $\alpha \sim \frac{1}{|E|}$

Proof

- let $M, M' \in \Omega, e = (x, y) \in E, e \notin M', M = M' \cup \{e\}$. Note $\rho(M, M') = 1$.
- Goal: couple $X \sim P(M, \cdot), Y \sim P(M', \cdot)$ so that $\mathbb{E}(\rho(X, Y)) \leqslant e^{-\alpha} < 1$ with $\alpha \sim \frac{1}{|E|}$
- Note: in $M'$, monomers at $x$ and $y$.

Proof

- let $M, M' \in \Omega, e = (x,y) \in E, e \notin M', M = M' \cup \{e\}$. Note $\rho(M, M') = 1$.
- Goal: couple $X \sim P(M,\cdot), Y \sim P(M',\cdot)$ so that $\mathbb{E}(\rho(X,Y)) \leqslant e^{-\alpha} < 1$ with $\alpha \sim \frac{1}{|E|}$
- Note: in $M'$, monomers at $x$ and $y$.
- Pick $e' \in E$ uniformly at random. If $x \notin e', y \notin e'$, do same move for both processes. Distance unchanged

# Path coupling method for the monomer-dimer model



Proof

- let $M, M' \in \Omega, e = (x,y) \in E, e \notin M', M = M' \cup \{e\}$. Note $\rho(M, M') = 1$.
- Goal: couple $X \sim P(M, \cdot), Y \sim P(M', \cdot)$ so that $\mathbb{E}(\rho(X,Y)) \leqslant e^{-\alpha} < 1$ with $\alpha \sim \frac{1}{|E|}$
- Note: in $M'$, monomers at $x$ and $y$.
- Pick $e' \in E$ uniformly at random. If $x \notin e', y \notin e'$, do same move for both processes. Distance unchanged
- If $e = e'$, do same move for both. Distance decreases by 1

# Path coupling method for the monomer-dimer model



Proof

- let $M, M' \in \Omega, e = (x, y) \in E, e \notin M', M = M' \cup \{e\}$. Note $\rho(M, M') = 1$.
- Goal: couple $X \sim P(M, \cdot), Y \sim P(M', \cdot)$ so that $\mathbb{E}(\rho(X, Y)) \leqslant e^{-\alpha} < 1$ with $\alpha \sim \frac{1}{|E|}$
- Note: in $M'$, monomers at $x$ and $y$.
- Pick $e' \in E$ uniformly at random. If $x \notin e', y \notin e'$, do same move for both processes. Distance unchanged
- If $e = e'$, do same move for both. Distance decreases by 1
- If $e' = (x, u), u \neq y$ (there are at most $\Delta_G$ such cases) then $e' \notin M, e' \notin M$.

Proof

- let $M, M' \in \Omega, e = (x,y) \in E, e \notin M', M = M' \cup \{e\}$. Note $\rho(M, M') = 1$.
- Goal: couple $X \sim P(M, \cdot), Y \sim P(M', \cdot)$ so that $\mathbb{E}(\rho(X,Y)) \leqslant e^{-\alpha} < 1$ with $\alpha \sim \frac{1}{|E|}$
- Note: in $M'$, monomers at $x$ and $y$.
- Pick $e' \in E$ uniformly at random. If $x \notin e', y \notin e'$, do same move for both processes. Distance unchanged
- If $e = e'$, do same move for both. Distance decreases by 1
- If $e' = (x,u), u \neq y$ (there are at most $\Delta_G$ such cases) then $e' \notin M, e' \notin M$.

  For process started at $M$, no move possible. For the other, put edge with probability $1/(1+\lambda^2)$. Distance increases by $1$ with probability $1/(1+\lambda^2)$.

# Path coupling method for the monomer-dimer model

Proof

- let $M, M' \in \Omega, e = (x,y) \in E, e \notin M', M = M' \cup \{e\}$. Note $\rho(M,M') = 1$.
- Goal: couple $X \sim P(M,\cdot), Y \sim P(M',\cdot)$ so that $\mathbb{E}(\rho(X,Y)) \leqslant e^{-\alpha} < 1$ with $\alpha \sim \frac{1}{|E|}$
- Note: in $M'$, monomers at $x$ and $y$.
- Pick $e' \in E$ uniformly at random. If $x \notin e', y \notin e'$, do same move for both processes. Distance unchanged
- If $e = e'$, do same move for both. Distance decreases by 1
- If $e' = (x,u), u \neq y$ (there are at most $\Delta_G$ such cases) then $e' \notin M, e' \notin M$.

  For process started at $M$, no move possible. For the other, put edge with probability $1/(1+\lambda^2)$. Distance increases by $1$ with probability $1/(1+\lambda^2)$.
- Altogether, new average distance is $\leqslant 1 + \frac{1}{|E|}\left(-1 + 2\frac{\Delta_G}{1+\lambda^2}\right) \leqslant e^{-1/(2|E|)}$ if $\lambda \gg \sqrt{\Delta_G}$.

# Path coupling method



Other examples of application of path coupling:

- Ising dynamics for $\beta \leqslant \beta_0(\Delta_G)$ (keyword: Dobrushin's uniqueness condition)

# Path coupling method



Other examples of application of path coupling:

- Ising dynamics for $\beta \leqslant \beta_0(\Delta_G)$ (keyword: Dobrushin's uniqueness condition)
- perfect colorings for $q > 2\Delta_G$

# Path coupling method



Other examples of application of path coupling:

- Ising dynamics for $\beta \leqslant \beta_0(\Delta_G)$ (keyword: Dobrushin's uniqueness condition)
- perfect colorings for $q > 2\Delta_G$
- hardcore model with small fugacity $\lambda$. $\pi(x) \propto \lambda^{\sum_{v \in V} x_v}$, $x \in \{0,1\}^V$, $x_v x_w = 0$ if $v \sim w$

Other examples of application of path coupling:

- Ising dynamics for $\beta \leqslant \beta_0(\Delta_G)$ (keyword: Dobrushin's uniqueness condition)
- perfect colorings for $q > 2\Delta_G$
- hardcore model with small fugacity $\lambda$. $\pi(x) \propto \lambda^{\sum_{v\in V} x_v}$, $x \in \{0,1\}^V$, $x_v x_w = 0$ if $v \sim w$
- biased RW on $\{1,\ldots,N\}$. Let $p \in (1/2,1)$ and

$$P(x,y) = \begin{cases} 1/2 & \text{if } \; x = y \notin \{1,N\} \\ p/2 & \text{if } \; y = x+1, 1 \leqslant x < N \\ (1-p)/2 & \text{if } \; y = x-1, 1 < x \leqslant N \\ 1-p/2 & \text{if } \; x = y = 1 \\ 1-(1-p)/2 & \text{if } \; x = y = N \end{cases}$$

# Path coupling method



Other examples of application of path coupling:

- Ising dynamics for $\beta \leqslant \beta_0(\Delta_G)$ (keyword: Dobrushin's uniqueness condition)
- perfect colorings for $q > 2\Delta_G$
- hardcore model with small fugacity $\lambda$. $\pi(x) \propto \lambda^{\sum_{v\in V} x_v}$, $x \in \{0,1\}^V$, $x_v x_w = 0$ if $v \sim w$
- biased RW on $\{1,\ldots,N\}$. Let $p \in (1/2,1)$ and

$$P(x,y) = \begin{cases} 1/2 & \text{if } x = y \notin \{1,N\} \\ p/2 & \text{if } y = x+1, 1 \leqslant x < N \\ (1-p)/2 & \text{if } y = x-1, 1 < x \leqslant N \\ 1-p/2 & \text{if } x = y = 1 \\ 1-(1-p)/2 & \text{if } x = y = N \end{cases}$$

- In latter example, “exponentially tilted metric”

For every $x,y \in \Omega$ fix a "canonical path" $\Gamma_{x\to y}$ of allowed transitions $e$ from $x$ to $y$.

Given $f:\Omega \mapsto \mathbb{R}$ write

$$
\begin{aligned}
2\mathrm{Var}_\pi(f) = \sum_{x,y} \pi(x)\pi(y)((f(x)-f(y))^2) &= \sum_{x,y} \pi(x)\pi(y)\left(\left(\sum_{e\in\Gamma_{x\to y}} \nabla_e f\right)^2\right) \\
&= \sum_{x,y} \pi(x)\pi(y)|\Gamma_{x\to y}|^2 \left(\left(\frac{1}{|\Gamma_{x\to y}|}\sum_{e\in\Gamma_{x\to y}} \nabla_e f\right)^2\right) \\
&\leqslant \sum_{x,y} \pi(x)\pi(y)|\Gamma_{x\to y}| \sum_{e\in\Gamma_{x\to y}} |\nabla_e f|^2. \\
&= \sum_{x,y} \pi(x)\pi(y)|\Gamma_{x\to y}| \sum_{(a,b)\in\Gamma_{x\to y}} |f(a)-f(b)|^2
\end{aligned}
$$

Summarizing: $\mathrm{Var}_\pi(f) \leqslant \frac{1}{2}\sum_{x,y} \pi(x)\pi(y)|\Gamma_{x\to y}|\sum_{(a,b)\in\Gamma_{x\to y}} \frac{\pi(a)P(a,b)}{\pi(a)P(a,b)}|f(a)-f(b)|^2$.

Now recall

$$\mathcal{E}(f) = \frac{1}{2}\sum_{x,y\in\Omega} \pi(x)P(x,y)|f(x)-f(y)|^2$$

Reorganizing the sum, deduce

$$\mathrm{Var}_\pi(f) \leqslant M\,\mathcal{E}(f), \quad M = \max_{e=(a,b)}\left(\sum_{x,y:\Gamma_{x\to y}\ni(a,b)} \frac{\pi(x)\pi(y)}{Q(e)}|\Gamma_{x\to y}|\right), \quad Q(e) = \pi(a)P(a,b).$$

($M$: "congestion ratio") and

$$\gamma = \frac{1}{T_{\mathrm{rel}}} = \inf_f \frac{\mathcal{E}(f)}{\mathrm{Var}_\pi(f)} \geqslant \frac{1}{M}.$$

# Applications of canonical paths 

- SRW on $\{1,\ldots,N\}^d$
- $T_{\mathrm{rel}} \lesssim N$ for two glued copies of $K_N$
- $T_{\mathrm{rel}} \leqslant e^{O(L^{d-1})}$ for Ising on $\{1,\ldots,L\}^d$
- $T_{\mathrm{rel}} \lesssim_\lambda |V|\,|E|$ for matchings of $(V,E)$. $\lambda$=monomer “fugacity”

  (see Jerrum's book, Sec. 5.3)

# Generalization of canonical paths: flows



Given $x \neq y \in \Omega$, replace single path $\Gamma_{x \to y}$ with a "multicommodity flow" $f$ (Sinclair '92)

$\mathcal{P}_{x \to y}$: directed simple paths from $x$ to $y$. $\mathcal{P} = \cup_{x \neq y} \mathcal{P}_{x \to y}$

- flow $f$ is map $f : \mathcal{P} \mapsto \mathbb{R}^+$, with

# Generalization of canonical paths: flows



Given $x \neq y \in \Omega$, replace single path $\Gamma_{x \to y}$ with a "multicommodity flow" $f$ (Sinclair '92)

$\mathcal{P}_{x \to y}$: directed simple paths from $x$ to $y$. $\mathcal{P} = \cup_{x \neq y} \mathcal{P}_{x \to y}$

- flow $f$ is map $f : \mathcal{P} \mapsto \mathbb{R}^+$, with

$$\sum_{p \in \mathcal{P}_{x \to y}} f(p) = \pi(x)\pi(y)$$

# Generalization of canonical paths: flows



Given $x \neq y \in \Omega$, replace single path $\Gamma_{x \to y}$ with a "multicommodity flow" $f$ (Sinclair '92)

$\mathcal{P}_{x \to y}$: directed simple paths from $x$ to $y$. $\mathcal{P} = \cup_{x \neq y} \mathcal{P}_{x \to y}$

- flow $f$ is map $f : \mathcal{P} \mapsto \mathbb{R}^+$, with

$$\sum_{p \in \mathcal{P}_{x \to y}} f(p) = \pi(x)\pi(y)$$

- define for $e = (a, b)$ with positive conductance $Q(e) = \pi(a)P(a, b)$

# Generalization of canonical paths: flows



Given $x \neq y \in \Omega$, replace single path $\Gamma_{x\to y}$ with a "multicommodity flow" $f$ (Sinclair '92)

$\mathcal{P}_{x\to y}$: directed simple paths from $x$ to $y$. $\mathcal{P} = \cup_{x\neq y}\mathcal{P}_{x\to y}$

- flow $f$ is map $f : \mathcal{P} \mapsto \mathbb{R}^+$, with

$$\sum_{p\in\mathcal{P}_{x\to y}} f(p) = \pi(x)\pi(y)$$

- define for $e = (a,b)$ with positive conductance $Q(e) = \pi(a)P(a,b)$

$$\bar{f}(e) = \sum_{p\ni e} f(p)|p|, \quad M(f) = \max_e \frac{\bar{f}(e)}{Q(e)}$$

# Generalization of canonical paths: flows



Given $x \neq y \in \Omega$, replace single path $\Gamma_{x\to y}$ with a "multicommodity flow" $f$ (Sinclair '92)

$\mathcal{P}_{x\to y}$: directed simple paths from $x$ to $y$. $\mathcal{P} = \cup_{x\neq y}\mathcal{P}_{x\to y}$

- flow $f$ is map $f : \mathcal{P} \mapsto \mathbb{R}^+$, with

$$\sum_{p\in\mathcal{P}_{x\to y}} f(p) = \pi(x)\pi(y)$$

- define for $e = (a,b)$ with positive conductance $Q(e) = \pi(a)P(a,b)$

$$\bar{f}(e) = \sum_{p\ni e} f(p)|p|, \quad M(f) = \max_e \frac{\bar{f}(e)}{Q(e)}$$

- With same proof as above: $\gamma = \frac{1}{T_{\text{rel}}} = \inf_f \frac{\mathcal{E}(f)}{\mathrm{Var}_\pi(f)} \geqslant \frac{1}{M(f)}$

# Generalization of canonical paths: flows

Idea:

- for every $x \neq y$, want to transport quantity $\pi(x)\pi(y)$ from $x$ to $y$.

# Generalization of canonical paths: flows



Idea:

- for every $x \neq y$, want to transport quantity $\pi(x)\pi(y)$ from $x$ to $y$.
- rather than deterministic path $\Gamma_{x \to y}$, random path with weight

# Generalization of canonical paths: flows



Idea:

- for every $x \neq y$, want to transport quantity $\pi(x)\pi(y)$ from $x$ to $y$.
- rather than deterministic path $\Gamma_{x\to y}$, random path with weight

$$w(p) = \frac{f(p)}{\pi(x)\pi(y)}\mathbf{1}_{p\in\mathcal{P}_{x\to y}}.$$

# Generalization of canonical paths: flows



Idea:

- for every $x \neq y$, want to transport quantity $\pi(x)\pi(y)$ from $x$ to $y$.
- rather than deterministic path $\Gamma_{x\to y}$, random path with weight

$$w(p) = \frac{f(p)}{\pi(x)\pi(y)}\mathbf{1}_{p\in\mathcal{P}_{x\to y}}.$$

- diversification decreases congestion rate.

# Generalization of canonical paths: flows



Idea:

- for every $x \neq y$, want to transport quantity $\pi(x)\pi(y)$ from $x$ to $y$.
- rather than deterministic path $\Gamma_{x\to y}$, random path with weight

$$w(p) = \frac{f(p)}{\pi(x)\pi(y)} \mathbf{1}_{p \in \mathcal{P}_{x\to y}}.$$

- diversification decreases congestion rate.
-

**Exercise 11.** For RW on complete bipartite graph $K_{2,N-2}$, prove $T_{\mathrm{rel}} \lesssim N$ with canonical paths and $T_{\mathrm{rel}} \lesssim 1$ with multicommodity flows.

Up to now: upper bounds on $T_{\mathrm{mix}}, T_{\mathrm{rel}}$. A useful lower bound tool: bottleneck ratio

Given $S \subset \Omega$,

$$\gamma \leqslant \frac{\mathcal{E}(\mathbf{1}_S)}{\mathrm{Var}_\pi(\mathbf{1}_S)} = \frac{1}{2}\frac{\sum_{x,y}\pi(x)P(x,y)(\mathbf{1}_S(x)-\mathbf{1}_S(y))^2}{\pi(S)-\pi(S)^2} = \frac{\sum_{x\in S, y\notin S}\pi(x)P(x,y)}{\pi(S)(1-\pi(S))}$$

Bottleneck ratio ("isoperimetric constant")

$$\Phi := \inf_{S\subset\Omega:\pi(S)\leqslant 1/2}\frac{\sum_{x\in S, y\notin S}\pi(x)P(x,y)}{\pi(S)} \Rightarrow \quad \gamma \leqslant 2\Phi, \quad T_{\mathrm{rel}} \geqslant \frac{1}{2\Phi}.$$

**Exercise 12.** $T_{\mathrm{rel}} \gtrsim N$ for RW on two copies of $K_N$ with one vertex in common

**Exercise 13.** $T_{\mathrm{rel}} \gtrsim e^{\Omega(N)}$ for size-$N$ Curie-Weiss Glauber dynamics, $\beta > \beta_c$.

Let $\preceq$ be a partial order on $\Omega$. Assume $\exists$ maximal/minimal configuration $\omega^+/\omega^-$.

Example: $\Omega = \{-1,+1\}^V, \sigma \preceq \eta$ iff $\sigma_x \leqslant \eta_x$ for every $x \in V$. $\omega^\pm \equiv \pm 1$.

Example: $\Omega = \{\text{perfect matchings of } G\}$, $G$ planar, bipartite graph.

$\sigma \preceq \eta$ iff $h_\sigma(f) \leqslant h_\eta(f)$ for every face $f$.

**Definition** *We say that a MC on $(\Omega, \preceq)$ is monotone (or "attractive") if $\forall x,y \in \Omega$ with $x \preceq y$ the processes $X,Y$ started from $x,y$ can be coupled so that $X_n \preceq Y_n$ almost surely for all $n \geqslant 0$.*

Most useful when coupling is Markovian and global.

Let $\tau_{+/-} = \inf\{n \geqslant 0 : X^+(n) = X^-(n)\}$ with $X^{+/-}$ the process started from maximal/minimal configuration $\omega^\pm$. Then:

**Proposition** *If* $\mathbb{P}(\tau_{+/-} > N) < \varepsilon$ *then* $T_{\mathrm{mix}}(\varepsilon) \leqslant N$.

**Proof.** Proof: Let $X^x$ be the process started from $x$ and $X^\pi$ the stationary process.

Write

$$\begin{aligned} \|P_x^N - \pi\| &\leqslant \mathbb{P}(X^x(N) \neq X^\pi(N)) && \text{(third definition of } \|\cdot\|\text{)} \\ &\leqslant \mathbb{P}(X^+(N) \neq X^-(N)) && \text{(by the well known sandwich principle)} \\ &= \mathbb{P}(\tau_{+/-} > N) && \text{(by definition of coalescence time)} \\ &< \varepsilon && \text{(by assumption)} \end{aligned}$$

so we deduce $T_{\mathrm{mix}}(\varepsilon) \leqslant N$. □

# Monotone dynamics: lazy random walk



Let $\Omega = \{1, \dots, N\}$, usual order. $\omega^+ = N, \omega^- = 1.$

$P(x, x \pm 1) = \frac{1}{4}$ if $x \pm 1 \in \Omega$. $P(x,x) = \frac{1}{2}$ if $1 < x < N$, $P(x,x) = \frac{3}{4}$ if $x \in \{1, N\}$.

Couple $(X^x)_{x \in \Omega}$ as follows:

- at step $n$ sample a uniform r.v. $U_n \sim \mathrm{Unif}(0,1)$

# Monotone dynamics: lazy random walk



Let $\Omega = \{1, \ldots, N\}$, usual order. $\omega^+ = N, \omega^- = 1.$

$P(x, x \pm 1) = \frac{1}{4}$ if $x \pm 1 \in \Omega$. $P(x,x) = \frac{1}{2}$ if $1 < x < N$, $P(x,x) = \frac{3}{4}$ if $x \in \{1, N\}$.

Couple $(X^x)_{x \in \Omega}$ as follows:

- at step $n$ sample a uniform r.v. $U_n \sim \text{Unif}(0,1)$
- If $U_n \leqslant 1/2,$ let $X^x(n) = X^x(n-1)$ for all $x$.

# Monotone dynamics: lazy random walk

Let $\Omega = \{1,\ldots,N\}$, usual order. $\omega^+ = N, \omega^- = 1.$

$P(x,x\pm 1) = \frac{1}{4}$ if $x\pm 1 \in \Omega$. $P(x,x) = \frac{1}{2}$ if $1 < x < N$, $P(x,x) = \frac{3}{4}$ if $x \in \{1,N\}$.

Couple $(X^x)_{x\in\Omega}$ as follows:

- at step $n$ sample a uniform r.v. $U_n \sim \text{Unif}(0,1)$
- If $U_n \leqslant 1/2,$ let $X^x(n) = X^x(n-1)$ for all $x$.
- If $1/2 < U_n \leqslant 3/4,$ let $X^x(n) = \max\,(X^x(n-1) - 1, 1)$

# Monotone dynamics: lazy random walk

Let $\Omega = \{1, \ldots, N\}$, usual order. $\omega^+ = N, \omega^- = 1.$

$P(x, x \pm 1) = \frac{1}{4}$ if $x \pm 1 \in \Omega$. $P(x,x) = \frac{1}{2}$ if $1 < x < N$, $P(x,x) = \frac{3}{4}$ if $x \in \{1, N\}$.

Couple $(X^x)_{x \in \Omega}$ as follows:

- at step $n$ sample a uniform r.v. $U_n \sim \mathrm{Unif}(0,1)$
- If $U_n \leqslant 1/2,$ let $X^x(n) = X^x(n-1)$ for all $x$.
- If $1/2 < U_n \leqslant 3/4,$ let $X^x(n) = \max\left(X^x(n-1) - 1, 1\right)$
- If $U_n > 3/4$, let $X^x(n) = \min\left(X^x(n-1) + 1, N\right)$

# Monotone dynamics: lazy random walk



Let $\Omega = \{1, \ldots, N\}$, usual order. $\omega^+ = N, \omega^- = 1.$

$P(x, x \pm 1) = \frac{1}{4}$ if $x \pm 1 \in \Omega$. $P(x,x) = \frac{1}{2}$ if $1 < x < N$, $P(x,x) = \frac{3}{4}$ if $x \in \{1, N\}$.

Couple $(X^x)_{x \in \Omega}$ as follows:

- at step $n$ sample a uniform r.v. $U_n \sim \mathrm{Unif}(0,1)$
- If $U_n \leqslant 1/2,$ let $X^x(n) = X^x(n-1)$ for all $x$.
- If $1/2 < U_n \leqslant 3/4,$ let $X^x(n) = \max\left(X^x(n-1) - 1, 1\right)$
- If $U_n > 3/4$, let $X^x(n) = \min\left(X^x(n-1) + 1, N\right)$

  NB: Once $X^x = X^y$, they coincide forever. The coupling is global, montone, Markovian.

# Monotone dynamics: ferromagnetic Ising



Global Markovian coupling for ferromagnetic ($\beta > 0$) Ising

- goal: find coupled processes $(X^\sigma(n))_{n\geqslant 0, \sigma\in\{-1,+1\}^V}$ that satisfy monotonicity

# Monotone dynamics: ferromagnetic Ising

Global Markovian coupling for ferromagnetic ($\beta > 0$) Ising

- goal: find coupled processes $(X^\sigma(n))_{n \geqslant 0, \sigma \in \{-1,+1\}^V}$ that satisfy monotonicity
- at step $n$, choose vertex $x_n \in V$ uniformly at random (the same for all processes)

# Monotone dynamics: ferromagnetic Ising



Global Markovian coupling for ferromagnetic ($\beta > 0$) Ising

- goal: find coupled processes $(X^{\sigma}(n))_{n\geqslant 0, \sigma\in\{-1,+1\}^V}$ that satisfy monotonicity
- at step $n$, choose vertex $x_n \in V$ uniformly at random (the same for all processes)
- sample a uniform variable $U_n \sim \mathrm{Unif}((0,1))$ (the same for all processes)

# Monotone dynamics: ferromagnetic Ising



Global Markovian coupling for ferromagnetic ($\beta > 0$) Ising

- goal: find coupled processes $(X^{\sigma}(n))_{n\geqslant 0, \sigma\in\{-1,+1\}^V}$ that satisfy monotonicity
- at step $n$, choose vertex $x_n \in V$ uniformly at random (the same for all processes)
- sample a uniform variable $U_n \sim \mathrm{Unif}((0,1))$ (the same for all processes)
- Set $X^{\sigma}_{x_n}(n) \to 1$ if $U_n < p^{\sigma}_{x_n}(n-1)$ and $X^{\sigma}_{x_n}(n) \to -1$ if $U_n \geqslant p^{\sigma}_{x_n}(n-1)$ where

$$p^{\sigma}_{x}(n-1) = \frac{e^{\beta \sum_{y\sim x} X^{\sigma}_{y}(n-1)}}{2\cosh(\beta \sum_{y\sim x} X^{\sigma}_{y}(n-1))} \in (0,1)$$

# Monotone dynamics: ferromagnetic Ising



Global Markovian coupling for ferromagnetic ($\beta > 0$) Ising

- goal: find coupled processes $(X^\sigma(n))_{n \geqslant 0, \sigma \in \{-1,+1\}^V}$ that satisfy monotonicity
- at step $n$, choose vertex $x_n \in V$ uniformly at random (the same for all processes)
- sample a uniform variable $U_n \sim \mathrm{Unif}((0,1))$ (the same for all processes)
- Set $X^\sigma_{x_n}(n) \to 1$ if $U_n < p^\sigma_{x_n}(n-1)$ and $X^\sigma_{x_n}(n) \to -1$ if $U_n \geqslant p^\sigma_{x_n}(n-1)$ where

$$p^\sigma_x(n-1) = \frac{e^{\beta \sum_{y \sim x} X^\sigma_y(n-1)}}{2\cosh(\beta \sum_{y \sim x} X^\sigma_y(n-1))} \in (0,1)$$

- Note: $u \mapsto \frac{e^{\beta u}}{2\cosh(\beta u)}$ is increasing if $\beta \geqslant 0$, hence coupling is monotone.

# Other examples



Other classical monotone dynamics:

- hardcore model on bipartite graph $G$. $\omega \preceq \eta$ iff $\omega_x \leqslant \eta_x$ for $x$ even and $\omega_x \geqslant \eta_x$ for $x$ odd. $\omega^{\pm}$: all even/odd vertices occupied
- heat-bath dynamics on perfect matchings of planar, bipartite $G$.
- heat-bath dynamics for height models

$$\pi(h) \propto e^{-\sum_{x\sim y} V(h_x - h_y)}, \quad h : V \mapsto \mathbb{R}, V \subset \mathbb{Z}^d, \quad h_{\partial V} \equiv 0$$

with $V$ symmetric and convex

# Naive (wrong) idea

Let $X$ be a MC on $(\Omega, \preceq)$ that admits a global monotone Markovian coupling.

Algorithmic description of process:

- randomness is iid sequence $(\xi_n)_{n\geqslant 1}$ (e.g. for Ising: $\xi_n = (x_n, U_n), x_n \sim \mathrm{Unif}(V), U_n \sim \mathrm{Unif}(0,1)$)
- state of chain $X(n)$ at time $n$ is **deterministic** function of $\xi_n$ and of $X(n-1)$. Write $X(n) = F_{\xi_n}(X(n-1))$.
- Iterating, $X(n) = F_{\xi_n} \circ F_{\xi_{n-1}} \circ \ldots \circ F_{\xi_1}(X(0))$

Appealing idea: stop algorithm at $\tau_{+/-}$ (coupling time of maximal/minimal configurations) and output $Z = X(\tau_{+/-}) = X^x(\tau_{+/-})$ for all initial conditions $x$

Problem: $\mathbb{P}(Z = \sigma) \neq \pi(\sigma)$ (e.g.: lazy RW)

# Better idea: coupling from the past (Propp-Wilson)



- View time as running from $-\infty$ to $0$: $\xi = (\ldots, \xi_{-2}, \xi_{-1}, \xi_0)$.

# Better idea: coupling from the past (Propp-Wilson)



- View time as running from $-\infty$ to $0$: $\xi = (\ldots, \xi_{-2}, \xi_{-1}, \xi_0)$.
- Find

$$n_{\min} := \inf\{n : F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^+) = F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^-)\}$$

(in most interesting cases, $n_{\min} < \mathrm{infty}$ a.s.)

# Better idea: coupling from the past (Propp-Wilson)

- View time as running from $-\infty$ to $0$: $\xi = (\ldots, \xi_{-2}, \xi_{-1}, \xi_0)$.
- Find

$$n_{\min} := \inf\{n : F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^+) = F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^-)\}$$

  (in most interesting cases, $n_{\min} < \mathrm{infty}$ a.s.)
- Output $Z := F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n_{\min}}}(\omega^\pm) \in \Omega$

# Better idea: coupling from the past (Propp-Wilson)



- View time as running from $-\infty$ to $0$: $\xi = (\ldots, \xi_{-2}, \xi_{-1}, \xi_0)$.
- Find

$$n_{\min} := \inf\{n : F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^+) = F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^-)\}$$

(in most interesting cases, $n_{\min} < \mathrm{infty}$ a.s.)

- Output $Z := F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n_{\min}}}(\omega^\pm) \in \Omega$
- Claim: $Z \sim \pi$ (no bias!)

# Better idea: coupling from the past (Propp-Wilson)



- View time as running from $-\infty$ to $0$: $\xi = (\ldots, \xi_{-2}, \xi_{-1}, \xi_0)$.
- Find

$$n_{\min} := \inf\{n : F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^+) = F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^-)\}$$

(in most interesting cases, $n_{\min} < \text{infty}$ a.s.)

- Output $Z := F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n_{\min}}}(\omega^\pm) \in \Omega$
- Claim: $Z \sim \pi$ (no bias!)
- Proof: with a picture

# Better idea: coupling from the past (Propp-Wilson)

- View time as running from $-\infty$ to $0$: $\xi = (\ldots, \xi_{-2}, \xi_{-1}, \xi_0)$.
- Find

$$n_{\min} := \inf\{n : F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^+) = F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^-)\}$$

  (in most interesting cases, $n_{\min} < \text{infty}$ a.s.)
- Output $Z := F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n_{\min}}}(\omega^\pm) \in \Omega$
- Claim: $Z \sim \pi$ (no bias!)
- Proof: with a picture
- Algorithmically better to replace $-n$ by $-2^n$.

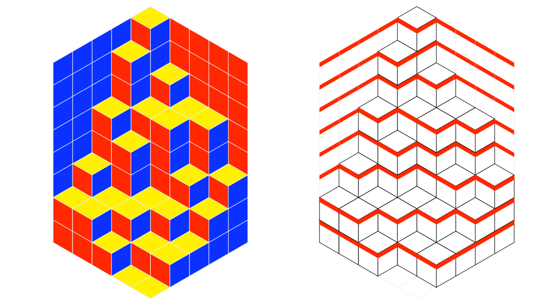
# Better idea: coupling from the past (Propp-Wilson)



- View time as running from $-\infty$ to $0$: $\xi = (\ldots, \xi_{-2}, \xi_{-1}, \xi_0)$.
- Find

$$n_{\min} := \inf\{n : F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^+) = F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n}}(\omega^-)\}$$

  (in most interesting cases, $n_{\min} < \text{infty}$ a.s.)
- Output $Z := F_{\xi_0} \circ F_{\xi_{-1}} \circ \ldots \circ F_{\xi_{-n_{\min}}}(\omega^\pm) \in \Omega$
- Claim: $Z \sim \pi$ (no bias!)
- Proof: with a picture
- Algorithmically better to replace $-n$ by $-2^n$.
- NB: “termination bias”, because $Z$ and $n_{\min}$ not independent. See J. Fill '98 for unbiased interruptible perfect simulation algorithm



$L$ non-intersecting paths $\varphi := (\varphi^i), i = 1, \ldots, L$, $\varphi^i(x) - \varphi^i(x-1) = \pm 1, \varphi^i(0) = \varphi^i(L) = i$

Partial order: $\varphi \preceq \psi \Leftrightarrow \varphi^i(x) \leqslant \psi^i(x)$ for all $i, x$

Note: concides with the partial order induced by height function

**"Tower-move"** sampling algorithm:

- Choose $0 < x < L$ uniformly at random

# Fast mixing of lozenge tiling dynamics



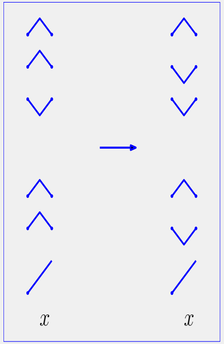
**"Tower-move"** sampling algorithm:

- Choose $0 < x < L$ uniformly at random
- resample $\varphi^i(x), i = 1, \ldots, L$ uniformly conditionally on $\varphi^i(x \pm 1), i = 1, \ldots, L$.

# Fast mixing of lozenge tiling dynamics

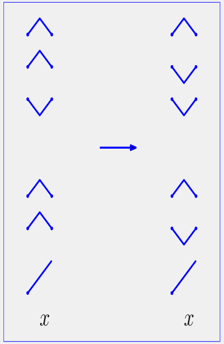
**“Tower-move”** sampling algorithm:

- Choose $0 < x < L$ uniformly at random
- resample $\varphi^i(x), i = 1, \ldots, L$ uniformly conditionally on $\varphi^i(x \pm 1), i = 1, \ldots, L$.



**"Tower-move"** sampling algorithm:

- Choose $0 < x < L$ uniformly at random
- resample $\varphi^i(x), i = 1, \dots, L$ uniformly conditionally on $\varphi^i(x \pm 1), i = 1, \dots, L$.

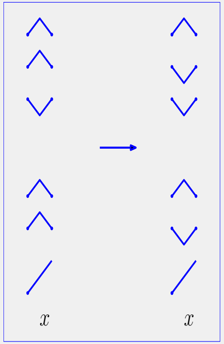



- Easy to check: (1) uniform measure $\pi$ is reversible (2) global monotone coupling

# Fast mixing of lozenge tiling dynamics

**"Tower-move"** sampling algorithm:

- Choose $0 < x < L$ uniformly at random
- resample $\varphi^i(x), i = 1, \ldots, L$ uniformly conditionally on $\varphi^i(x \pm 1), i = 1, \ldots, L$.




- Easy to check: (1) uniform measure $\pi$ is reversible (2) global monotone coupling
- Single step easy to simulate

Define function $\varphi \mapsto \Phi(\varphi) \in \mathbb{R}$ as $\Phi(\varphi) = \sum_{i=1}^{L} \sum_{x=0}^{L} \varphi^i(x) \sin\left(\frac{\pi x}{L}\right)$

Note:

- if $\varphi \preceq \psi$ then $\Phi(\varphi) \leqslant \Phi(\psi)$, and $\Phi(\psi) - \Phi(\varphi) \geqslant \frac{1}{L}$ if $\varphi \neq \psi$.

# Fast mixing of lozenge tiling dynamics



Define function $\varphi \mapsto \Phi(\varphi) \in \mathbb{R}$ as $\Phi(\varphi) = \sum_{i=1}^{L} \sum_{x=0}^{L} \varphi^i(x) \sin\left(\frac{\pi x}{L}\right)$

Note:

- if $\varphi \preceq \psi$ then $\Phi(\varphi) \leqslant \Phi(\psi)$, and $\Phi(\psi) - \Phi(\varphi) \geqslant \frac{1}{L}$ if $\varphi \neq \psi$.
- Deduce $\mathbb{P}(\varphi^+(n) \neq \varphi^-(n)) \leqslant L \times \mathbb{E}(\Phi(\varphi^+(n)) - \Phi(\varphi^-(n)))$

Define function $\varphi \mapsto \Phi(\varphi) \in \mathbb{R}$ as $\Phi(\varphi) = \sum_{i=1}^{L} \sum_{x=0}^{L} \varphi^i(x) \sin\left(\frac{\pi x}{L}\right)$

Note:

- if $\varphi \preceq \psi$ then $\Phi(\varphi) \leqslant \Phi(\psi)$, and $\Phi(\psi) - \Phi(\varphi) \geqslant \frac{1}{L}$ if $\varphi \neq \psi$.
- Deduce $\mathbb{P}(\varphi^+(n) \neq \varphi^-(n)) \leqslant L \times \mathbb{E}(\Phi(\varphi^+(n)) - \Phi(\varphi^-(n)))$
- Will prove (next slide):

$$\mathbb{E}(\Phi(\varphi^+(n)) - \Phi(\varphi^-(n))) \leqslant \frac{1}{L^2} \quad \text{if} \quad n \approx L^3 \log(L)$$

# Fast mixing of lozenge tiling dynamics



Define function $\varphi \mapsto \Phi(\varphi) \in \mathbb{R}$ as $\Phi(\varphi) = \sum_{i=1}^{L} \sum_{x=0}^{L} \varphi^i(x) \sin\left(\frac{\pi x}{L}\right)$

Note:

- if $\varphi \preceq \psi$ then $\Phi(\varphi) \leqslant \Phi(\psi)$, and $\Phi(\psi) - \Phi(\varphi) \geqslant \frac{1}{L}$ if $\varphi \neq \psi$.
- Deduce $\mathbb{P}(\varphi^+(n) \neq \varphi^-(n)) \leqslant L \times \mathbb{E}(\Phi(\varphi^+(n)) - \Phi(\varphi^-(n)))$
- Will prove (next slide):

$$\mathbb{E}(\Phi(\varphi^+(n)) - \Phi(\varphi^-(n))) \leqslant \frac{1}{L^2} \quad \text{if} \quad n \approx L^3 \log(L)$$

- Conclude: $T_{\text{mix}} \lesssim L^3 \log(L)$.

- Crucial observation: if at step $n$ we choose site $x$, $\sum_{i=0}^{L}\varphi^i(x)$ changes on average by

$$\frac{1}{2}\sum_{i=1}^{L-1}(\Delta\varphi^i)(x), \quad \Delta = \text{discrete Laplacian.}$$

- Crucial observation: if at step $n$ we choose site $x$, $\sum_{i=0}^{L} \varphi^i(x)$ changes on average by

$$\frac{1}{2}\sum_{i=1}^{L-1} (\Delta\varphi^i)(x), \quad \Delta = \text{discrete Laplacian.}$$

- Therefore, $\Phi(\varphi)$ changes on average by

$$\frac{1}{2L}\sum_{x=1}^{L-1}\sum_{i=1}^{L} (\Delta\varphi^i)(x)\sin\left(\frac{\pi x}{L}\right) = \frac{1}{2L}\sum_{x=1}^{L-1}\sum_{i=1}^{L} \varphi^i(x)\Delta\left(\sin\left(\frac{\pi x}{L}\right)\right)$$

- Crucial observation: if at step $n$ we choose site $x$, $\sum_{i=0}^{L} \varphi^i(x)$ changes on average by

$$\frac{1}{2}\sum_{i=1}^{L-1} (\Delta\varphi^i)(x), \quad \Delta = \text{discrete Laplacian.}$$

- Therefore, $\Phi(\varphi)$ changes on average by

$$\frac{1}{2L}\sum_{x=1}^{L-1}\sum_{i=1}^{L} (\Delta\varphi^i)(x)\sin\left(\frac{\pi x}{L}\right) = \frac{1}{2L}\sum_{x=1}^{L-1}\sum_{i=1}^{L} \varphi^i(x)\Delta\left(\sin\left(\frac{\pi x}{L}\right)\right)$$

- Note: $\Delta\left(\sin\left(\frac{\pi x}{L}\right)\right) = -\lambda_L \sin\left(\frac{\pi x}{L}\right), \quad \lambda_L \approx \frac{\pi^2}{L^2}$

# Fast mixing of lozenge tiling dynamics



- Crucial observation: if at step $n$ we choose site $x$, $\sum_{i=0}^{L} \varphi^i(x)$ changes on average by

$$\frac{1}{2}\sum_{i=1}^{L-1} (\Delta\varphi^i)(x), \quad \Delta = \text{discrete Laplacian.}$$

- Therefore, $\Phi(\varphi)$ changes on average by

$$\frac{1}{2L}\sum_{x=1}^{L-1}\sum_{i=1}^{L} (\Delta\varphi^i)(x)\sin\left(\frac{\pi x}{L}\right) = \frac{1}{2L}\sum_{x=1}^{L-1}\sum_{i=1}^{L} \varphi^i(x)\Delta\left(\sin\left(\frac{\pi x}{L}\right)\right)$$

- Note: $\Delta\left(\sin\left(\frac{\pi x}{L}\right)\right) = -\lambda_L \sin\left(\frac{\pi x}{L}\right), \quad \lambda_L \approx \frac{\pi^2}{L^2}$
- wrapping up: $\mathbb{E}(\Phi(\varphi(n))) - \mathbb{E}(\Phi(\varphi(n-1))) = -\frac{\lambda_L}{2L}\mathbb{E}(\Phi(\varphi(n-1)))$

Conclusion:

- Iterating,

$$
\begin{aligned}
\mathbb{E}(\Phi(\varphi^+(n)) - \Phi(\varphi^-(n))) &= \left(1 - \frac{\lambda_L}{2L}\right)^n \mathbb{E}(\Phi(\varphi^+(0)) - \Phi(\varphi^-(0))) \\
&\approx e^{-n\lambda_L/(2L)} \mathbb{E}(\Phi(\varphi^+(0)) - \Phi(\varphi^-(0))) \\
&\approx e^{-n\frac{\pi^2}{2L^3}} \mathbb{E}(\Phi(\varphi^+(0)) - \Phi(\varphi^-(0))) \\
&\approx L^3 e^{-n\frac{\pi^2}{2L^3}} \\
&\leqslant \frac{1}{L^2} \quad \text{if } n \gtrsim L^3 \log(L).
\end{aligned}
$$

- Argument (almost immediately) implies also $\gamma = \frac{\lambda_L}{L}$ exactly!

**Exercise 14.** Prove the following: any random walk on a (finite) tree $T$ is a reversible Markov chain. Find the stationary measure in terms of the transition matrix $P$.

**Exercise 15.** Prove that $\mathrm{Var}_\pi(P^n f) \leqslant \mathrm{Var}_\pi(f)(1-\gamma)^{2n}$ and that $\gamma$ is the largest constant such that this holds for every function $f$.

**Exercise 16.** Let $q \in \mathbb{N}, G = (V, E)$ a graph of maximal degree $\Delta$. Define a reversible MC on the set of perfect $q$-colorings of $G$ such that it is irreducible for $q$ large (for instance, $q > 2\Delta$).

**Exercise 17.** Prove that the elementary-rotation dynamics on a finite, planar, bipartite graph $G$ is irreducible.

**Exercise 18.** Let $X$ be a RW on $\mathbb{N}^2$ with transition probabilities $P((x,y),(x+1,y)) = P((x,y+1),(x,y)) = p$, $P((x,y),(x-1,y)) = P((x,y-1),(x,y)) = (1-p)/2$ assuming that the endpoint is in $\mathbb{N}^2$, otherwise $X$ stays put. Compute the stationary measure $\pi$.

# Exercises: session II

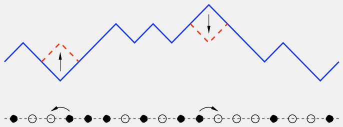
**Exercise 19.** Prove that the elementary-rotation dynamics on a finite, planar, bipartite graph $G$ is irreducible.

**Exercise 20.** Deduce that there exists a configuration with maximal height

**Exercise 21.** Prove $T_{\mathrm{rel}} \asymp N$, $T_{\mathrm{rel}} = O(1)$ for biased RW on $\{1, \ldots, N\}$ (jump probabilities $p \neq q$) using path coupling with exponential metric

**Exercise 22.** Prove that the Ising Glauber dynamic on $G = (V, E)$ has $T_{\mathrm{mix}} \lesssim |V| \log(|V|)$ and $\gamma \gtrsim |V|^{-1}$ for $\beta$ sufficiently small (depending on the maximal degree of $G$)

**Exercise 23.** Prove that the symmetric corner-flip dynamic admits a global monotone coupling



**Exercise 24.** Prove that the dimer dynamic on a planar, bipartite finite graph admits a global Markovian monotone coupling.