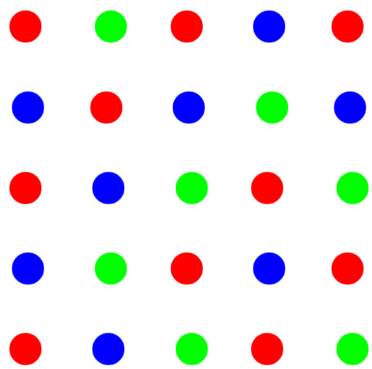
Efficient Algorithms for Finding a Random Needle in a Combinatorial Haystack

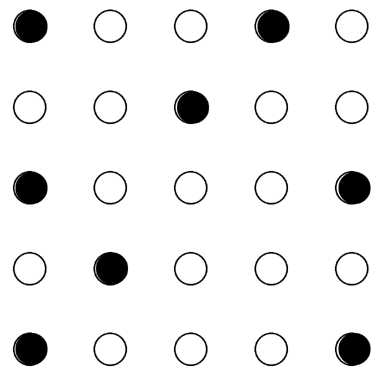
Dana Randall

Georgia Tech

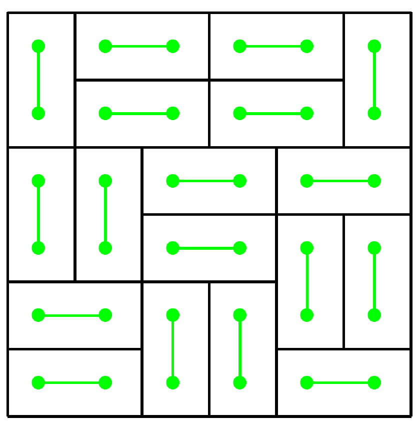
Models from Statistical Mechanics



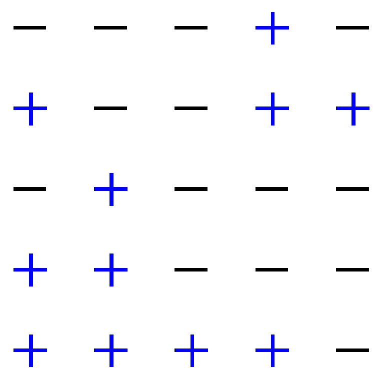
3-Colorings



Independent Sets



Dimer Model



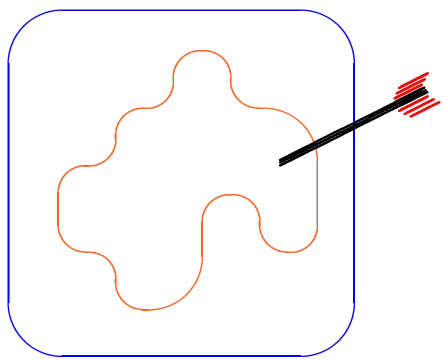
Ising Model

What does a “typical” element look like?

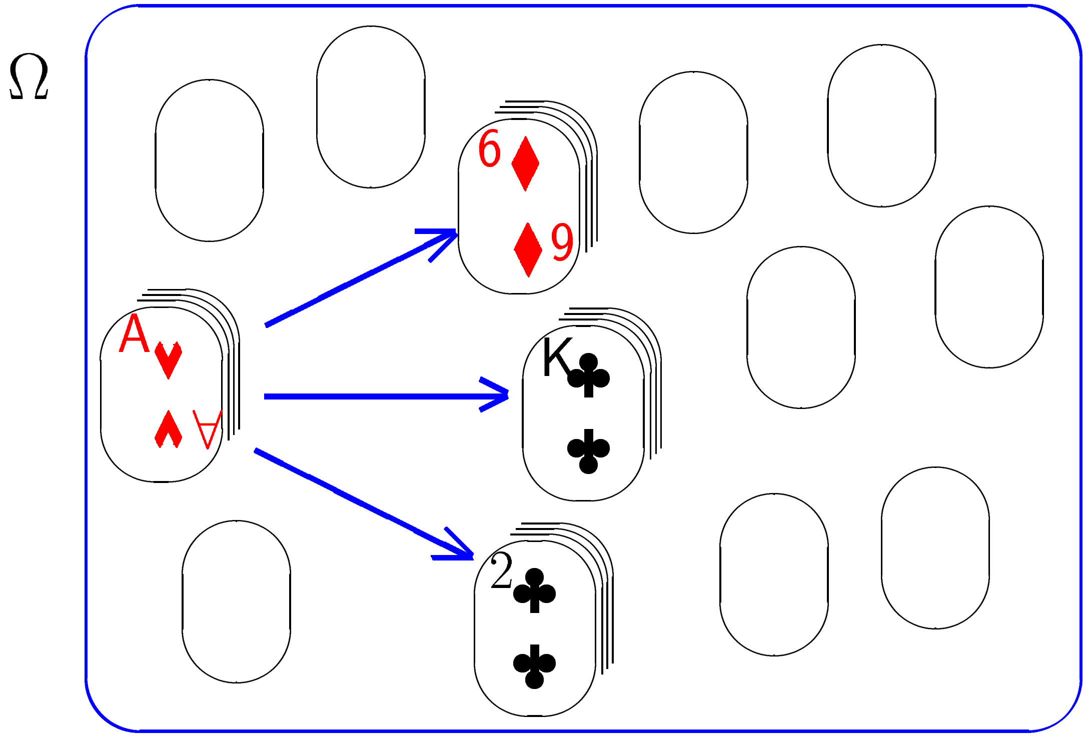
Sampling can be used to:

- Evaluate thermodynamic properties.
- Determine properties of “typical” elements.
- Estimate the cardinality of a set.

(“Markov chain Monte Carlo”)



Random Sampling Using Markov Chains



$(\Omega,\ P,\ \pi)$

$\pi(\cdot) = \frac{1}{52\,!}$

Thm: If a finite, reversible M.C. $\mathcal{M}$ with transition probs P is ergodic on Ω, then it converges to a unique stationary distribution.

The Metropolis Algorithm:

Thm: If P is defined so that

$$P(x,y) = \frac{1}{\Delta}\min\left(1, \frac{\pi(y)}{\pi(x)}\right),$$

then the stationary distribution will be π.

...But for how long must we run the chain?

Bounding Convergence Time

Def: The total variation distance

$$||\pi, \pi'||_{\mathrm{TV}} = \max_{A \subseteq \Omega} (\pi(A) - \pi'(A)).$$

Def: The mixing time is

$$\tau(\epsilon) = \max_{x} \min\{t : ||P^t(x, \cdot), \pi||_{\mathrm{TV}}\} \leq \epsilon.$$

Def: A Markov chain is rapidly mixing if

$$\tau(\epsilon) \leq \mathrm{poly}(n, \epsilon^{-1}).$$

Some methods:

- spectral gap $\longrightarrow$ $\mathrm{Gap}(P) = \lambda_1 - |\lambda_2|$,
- (path) coupling $\quad (\lambda_1 \geq |\lambda_2| \geq ... \geq |\lambda_{|\Omega|}|)$.
- canonical paths / flows
- conductance / isoperimetry
- comparison
- decomposition
- stopping rules

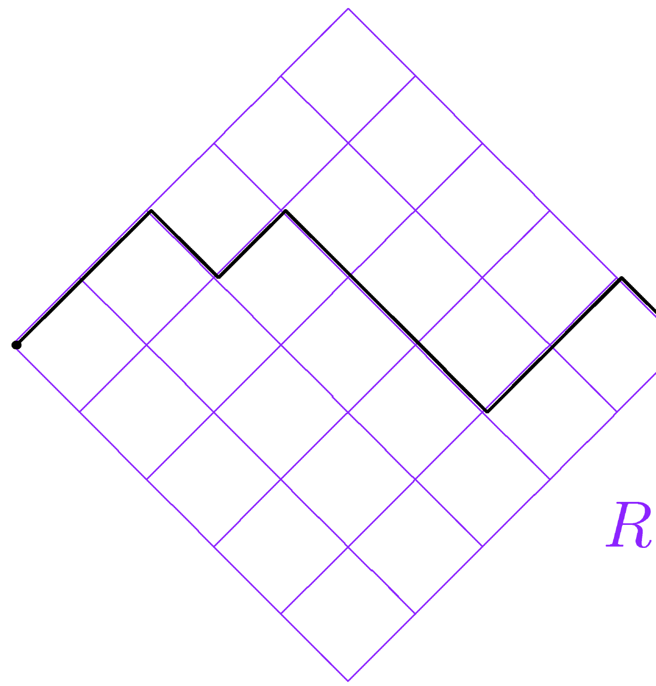
Examples:

The efficiency and limitations of various sampling algorithms for:

1. Lattice paths
2. Dimer models and 3-colorings
3. Ising / Potts models

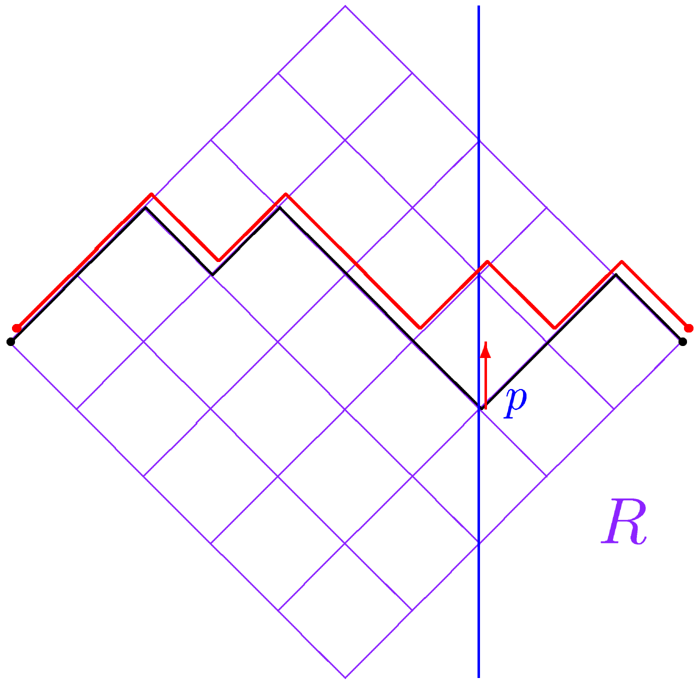
Example 1: Lattice paths

Find a shortest path in $R \subseteq \mathbb{Z}^2$.



Example 1: Lattice paths

Find a shortest path in $R \subseteq \mathbb{Z}^2$.



The Mountain/Valley Chain

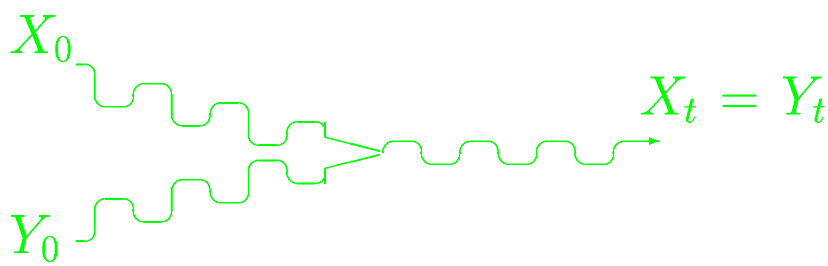
Repeat:

- Pick vertical line uniformly.
- If path intersect is "mountain" or "valley," invert it with probability $1/2$;
  else do nothing.

(This is ergodic for any simply connected R, and the stationary distribution is uniform.)

Coupling (To bound convergence rates)

For a Markov chain $\mathcal{M} = (\Omega, P, \pi)$:



Def: A coupling is a stochastic process $(X_t, Y_t)_{t=0}^{\infty}$ on $\Omega \times \Omega$ s.t.:

1. X_t and Y_t are each faithful copy of $\mathcal{M}$;
2. If $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$.

Def: The coupling time

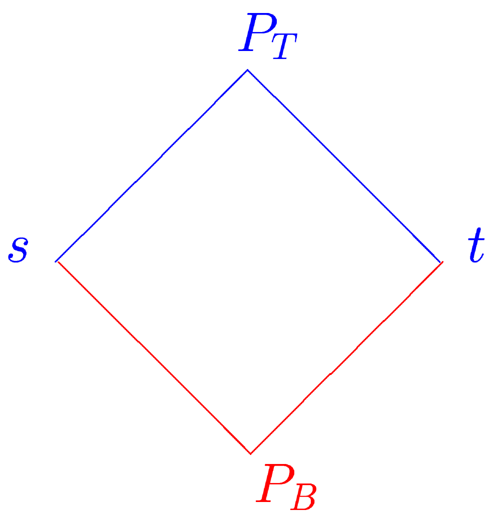
$$T = \max_{X_0, Y_0} \mathrm{E}\left[\min\{t : X_t = Y_t | X_0, Y_0\}\right].$$

Thm: [Aldous]

$$\tau(\epsilon) \leq c\, T \ln \epsilon^{-1}.$$

↑ mixing time ↑ coupling time

Proof of Fast Mixing



$\forall$ paths P,

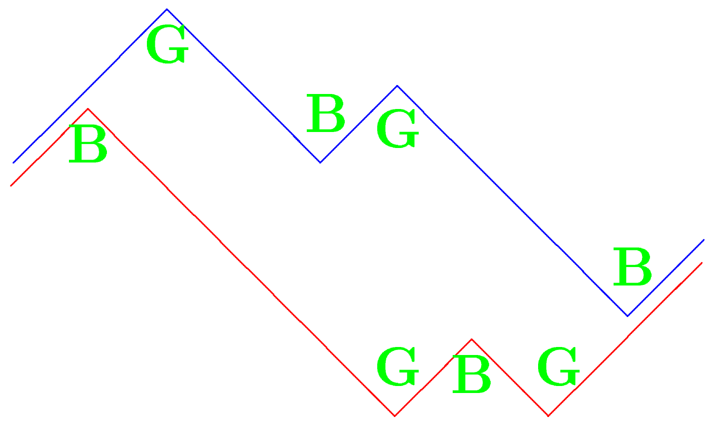
$P_T \geq P \geq P_B$

To couple, choose the same vertical line and the same direction for both processes.

1. Initially, vol $=$ vol $[P_T, P_B] = n$.
*2. $\mathrm{E}\left[\Delta\text{vol}\right] \leq 0$.
3. Pr $\left[\text{vol changes in 1 step}\right] \geq \frac{1}{n}$ (if vol $\neq 0$).
4. vol $= 0 \Leftrightarrow$ paths agree.

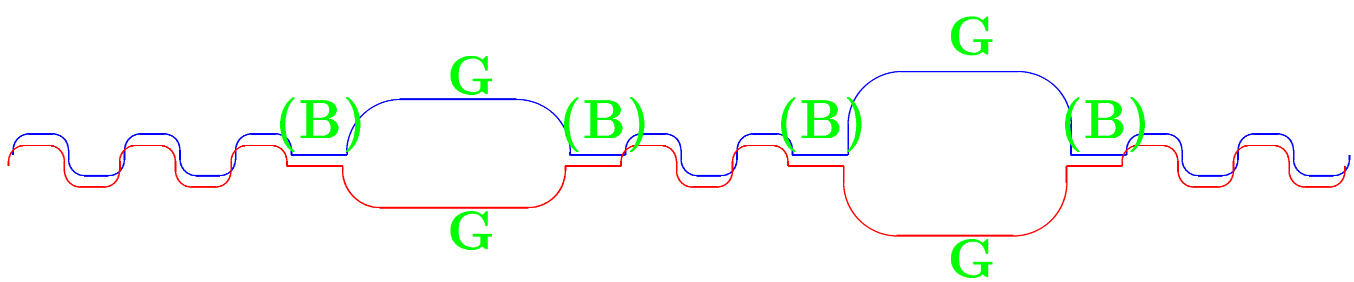
$\Rightarrow$ Mixing Time $\leq$ O(n^3).

Pf. cont.: (* $\mathrm{E}[\Delta \text{vol}] \leq 0$)



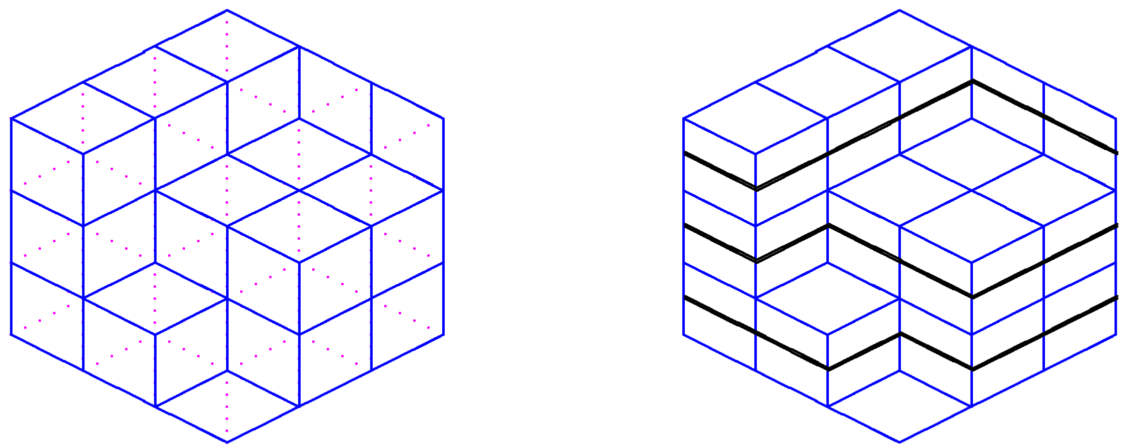
On P_1: Label Mtns w/ **G**; Valleys w/ **B**.
On P_2: Label Mtns w/ **B**; Valleys w/ **G**.

$$\mathrm{E}\left[\Delta \text{vol}\right] = \frac{1}{2m}\left[(\#B) - (\#G)\right] \overset{?}{\leq} 0.$$

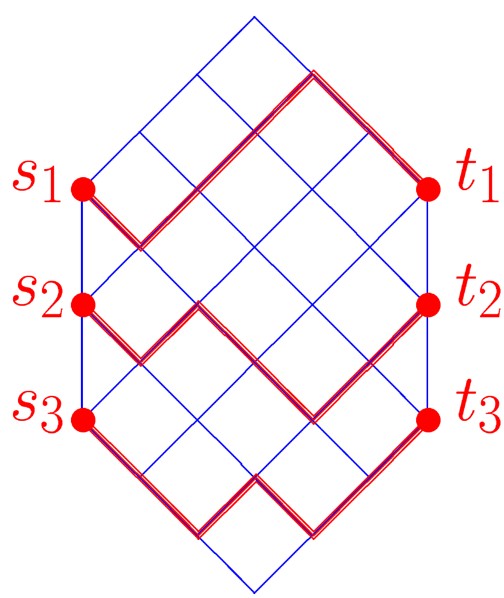


$$\Longrightarrow \quad (\#G) \geq (\#B) \qquad !$$

Ex. 2: Lozenge tilings

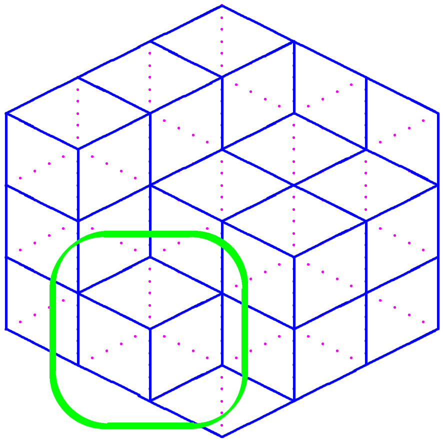


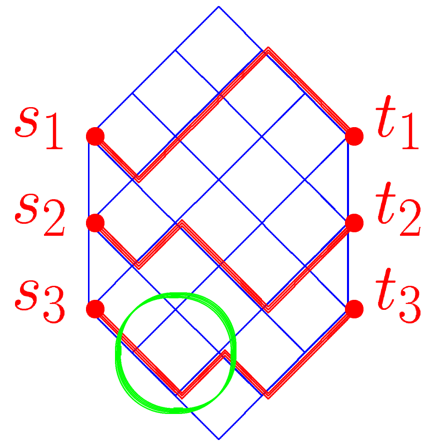
The dimer model on the triangular lattice



The “routing” interpretation

Two Markov chains on lozenge tilings





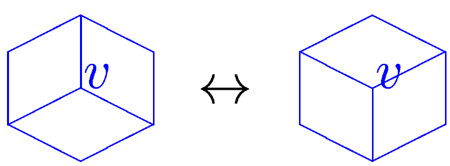
Markov chain 1: Glauber dynamics

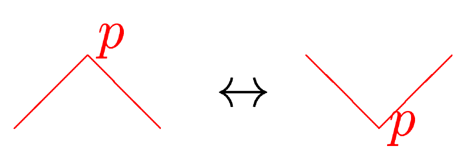
Repeat:

- Pick $v \in_u S$;
- Add/remove the "cube" at v w.p. $\frac{1}{2}$, if possible.

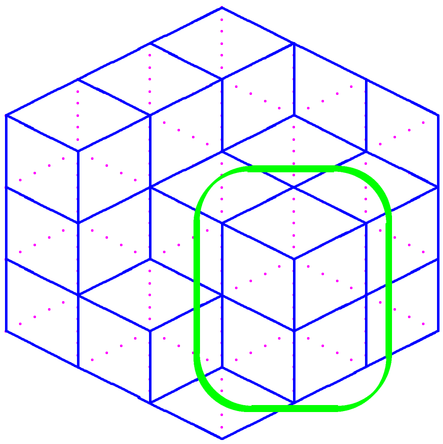
Repeat:

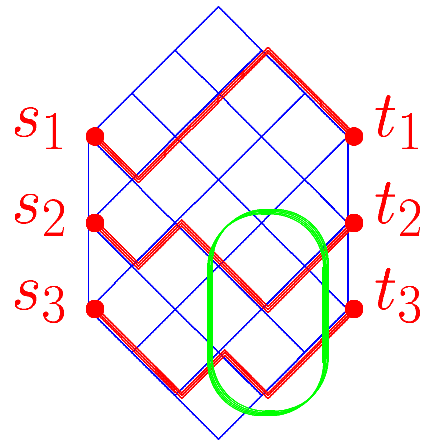
- Pick p on one of the paths uniformly;
- Invert Mountain/Valley, w.p. $\frac{1}{2}$, if possible.





Two Markov chains on lozenge tilings

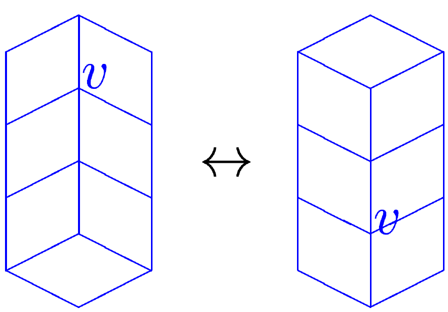




Markov chain 1: "Tower" moves

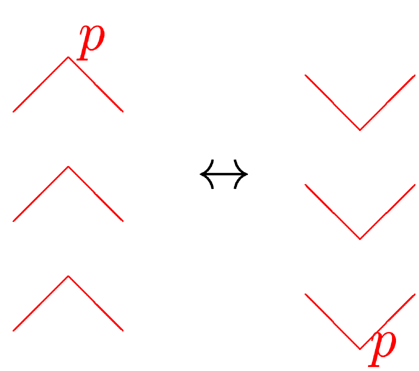
Repeat:

- Pick $v \in_u S$;
- Invert "tower" (ht h) at v w.p. $\frac{1}{2h}$, if pos.



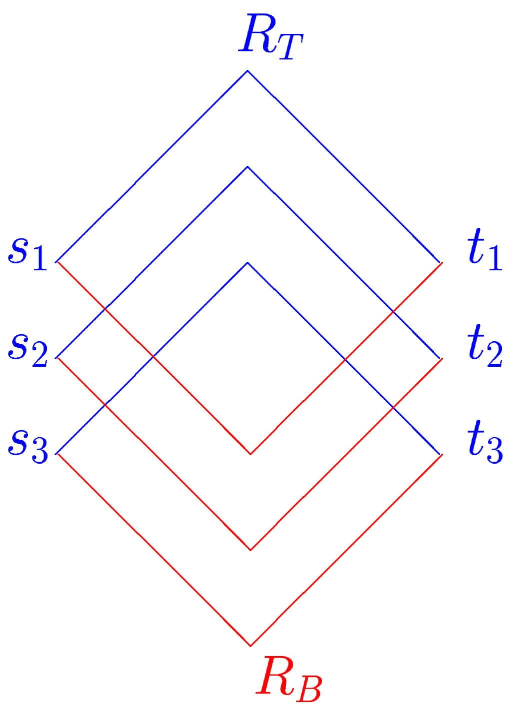
Repeat:

- Pick p on one of the paths uniformly;
- Invert the Mtn/val tower, w.p. $\frac{1}{2h}$, if possible.



Analysis of Markov chain 2 (Towers)

[Luby, R., Sinclair]



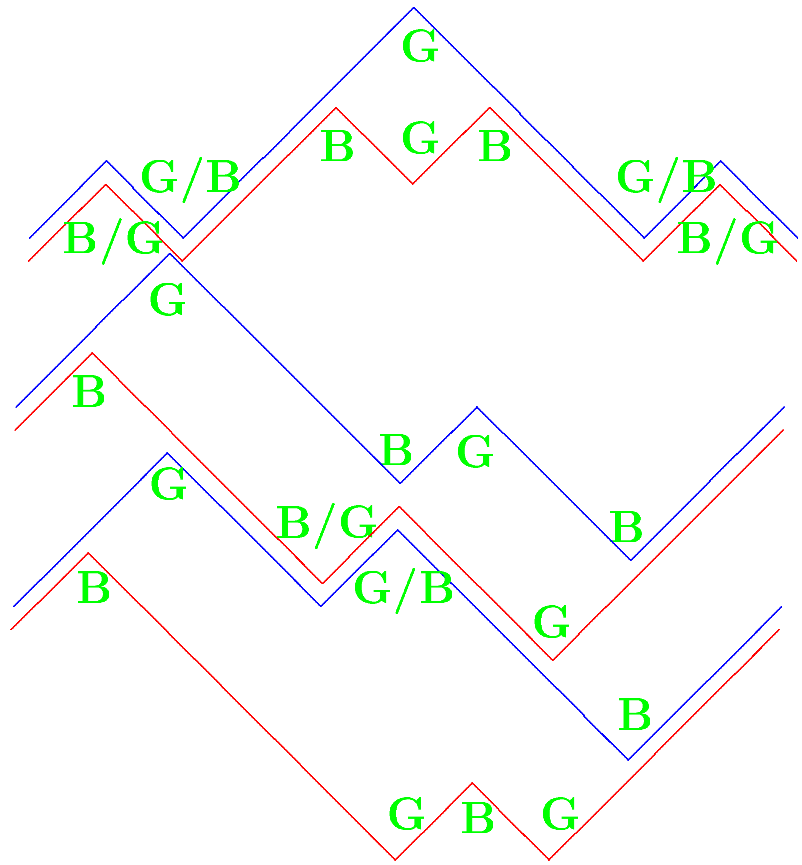
$\forall$ routings R,

$R_T \geq R \geq R_B$

1. Initially, vol $=$ vol $[R_T, R_B] = n^{1.5}$.
*2. E [Δvol] ≤ 0.
3. Pr [vol changes in 1 step] $\geq \frac{1}{n}$ (if vol $\neq 0$).
4. vol $= 0 \Leftrightarrow$ paths agree.

$\Rightarrow$ Mixing Time $\leq$ O(n^4).

Proof (cont.): (* $\mathrm{E}[\Delta\text{vol}] \leq 0$)



On P_1: Label Mtns w/ $\mathbf{G}$; Valleys w/ $\mathbf{B}$.

On P_2: Label Mtns w/ $\mathbf{B}$; Valleys w/ $\mathbf{G}$.

change due to inversion

prob of inversion

$$\mathrm{E}\left[\Delta\text{vol}\right] = \frac{1}{2m}\left(\sum_{\text{bad } t} h_t \left(\frac{1}{2h_t}\right) - \sum_{\text{good } t} h_t \left(\frac{1}{2h_t}\right)\right).$$

$$= \frac{1}{4m}\left((\#B) - (\#G)\right) \leq 0.$$

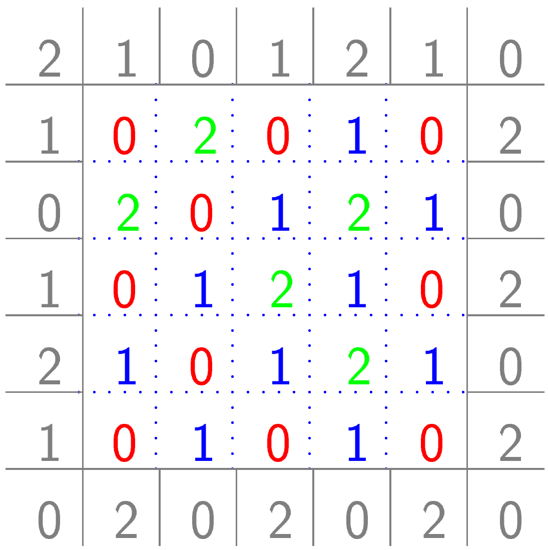
$\Longrightarrow$ The tower chain is rapidly mixing.

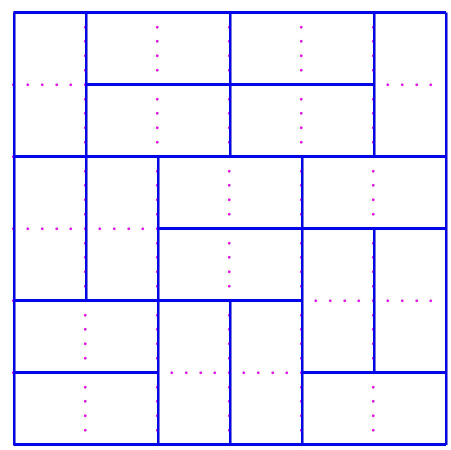
Extensions

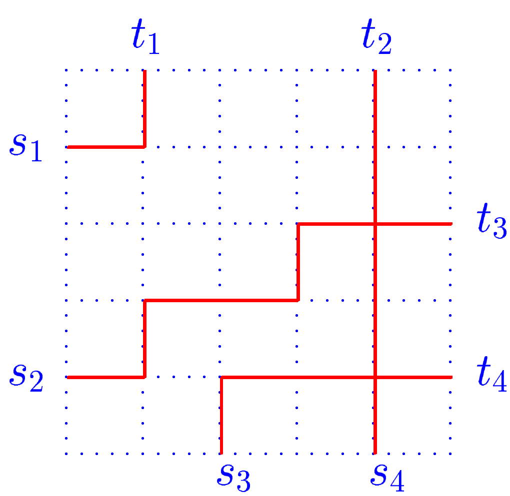
Thm: [R., Tetali] Glauber dynamics (MC 1) is also rapidly mixing on 3-colorings.

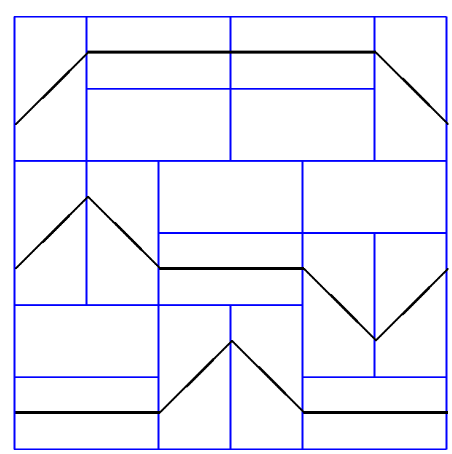
By the Comparison Thm of [Diaconis, Saloff-Coste] .

Same ideas also work for 3-colorings (with fixed boundary cond'ns) and domino tilings.









Example 3: Ising and Potts models

The Ising Model

− − − − − − − + + −
+ − − − − + + + + −
− − − − − + + + + +
− + − − − − + − − +
− − − + − − − − − −
− − − + − − − − − −
+ − − − − − − − − −
+ + + − − − − + − −
− + − − − − − − − −
+ + + + + + + + − −

$$\sigma \in \Omega = \{+,-\}^N$$

(Hamiltonian)

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

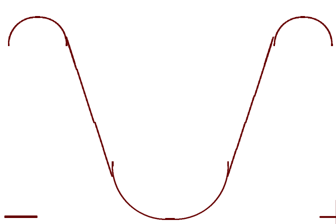
(Gibbs measure) $(\beta = 1/T)$

- $\pi_\beta(\sigma) = e^{\beta H(\sigma)}/Z_\beta$

(Partition function)

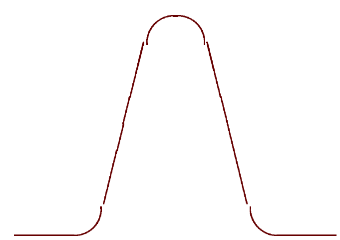
- $Z_\beta = \sum_{\sigma' \in \Omega} e^{\beta H(\sigma')}$

Energy landscapes:



− +

← Phase Transition →



− +

− + − − − − − −
− + − − − − − −
− − − − − − − −
− − − − + − − −
− − − − + + − −
− − − − − − − −
− − − − − − − −
+ + − − − − − −

Low temperature

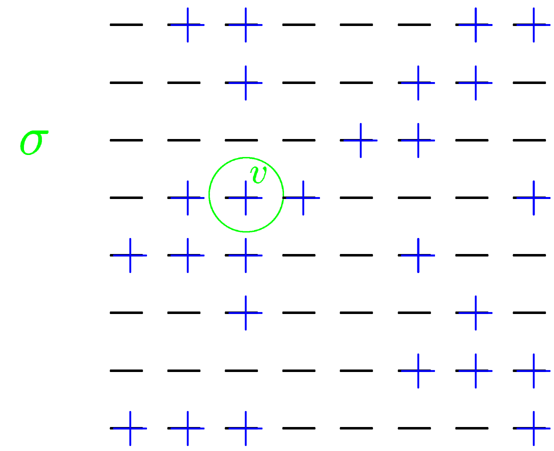
− + + − − − + +
− − + − − + + −
− − − − + + − −
− + + + − − − +
+ + + − − + − −
− − + − − − + −
− − − − − + + +
+ + + − − − − +

High temperature

Glauber dynamics on Ising configurations

Repeat:

- Pick v uniformly;
- Change the sign of v (going from σ to τ) w.p. $\min\left(1, \frac{\pi(\tau)}{\pi(\sigma)}\right)$.



Conductance:

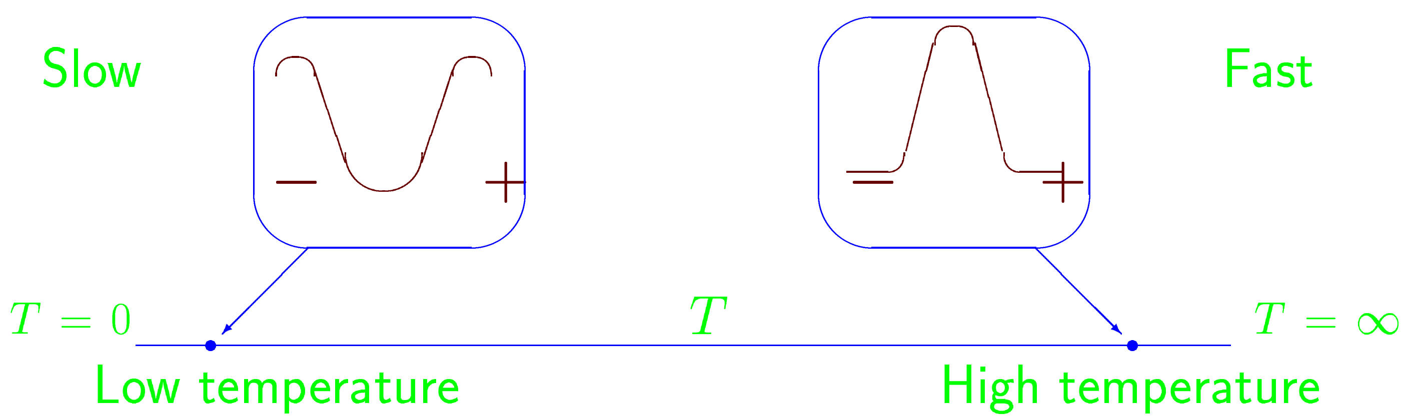
For $S \subseteq \Omega$, let

$$\Phi_S = \frac{\sum_{x \in S, y \in \overline{S}} \pi(x) P(x,y)}{\sum_{x \in S} \pi(x)}$$

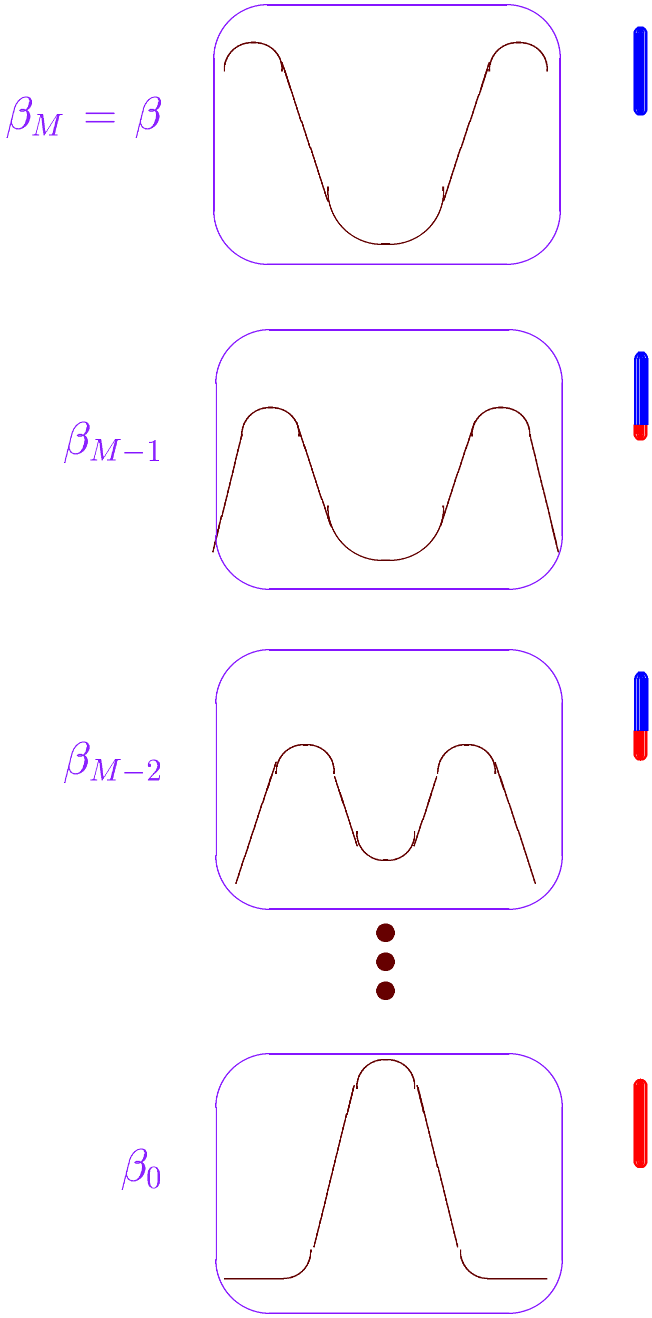
$$\Phi = \min_{S:\ \pi(S) \leq \frac{1}{2}} \Phi_S.$$

Thm: [Jerrum, Sinclair]

$$\frac{\Phi^2}{2} \leq Gap \leq 2\Phi.$$



Tempering / Swapping



$$\widehat{\Omega} = \Omega^{M+1}$$

$$\beta_i = \frac{i}{M}\beta$$

$$\pi_i(\sigma_i) = \pi_{\beta_i}(\sigma_i)$$

$$\widehat{\pi}(\sigma) = \prod_{i=0}^{M} \pi_i(\sigma_i)$$

The Swap Algorithm

Repeat:

w.p. 1/2: Do a LEVEL move:

Pick i; update σ_i

w.p. 1/2: Do a SWAP move:

Pick $(i, i+1)$;

"swap" σ_i and σ_{i+1}

$$\sigma = (\sigma_0, \ldots, \sigma_i, \sigma_{i+1}, \ldots, \sigma_M)$$

$$\sigma = (\sigma_0, \ldots, \sigma_{i+1}, \sigma_i, \ldots, \sigma_M)$$

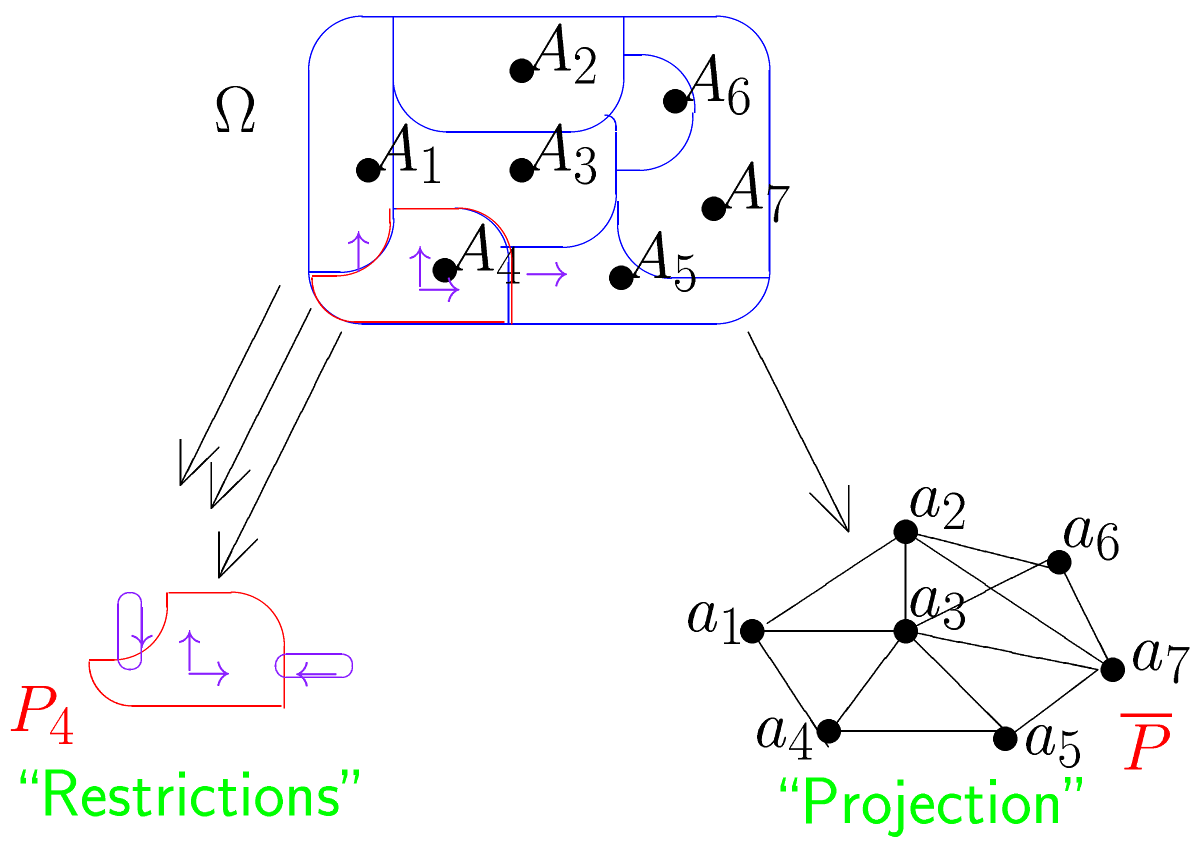
Swapping on the Mean-Field Ising Model

In the Mean-field model, the underlying graph is $G = K_n$.

Thm: [Madras, Zheng] Swap is fast for the mean-field Ising model for all β.

Disjoint Decomposition

[Caracciolo, Pelisetto, Sokal],
[Madras, R.],
[Martin, R.]

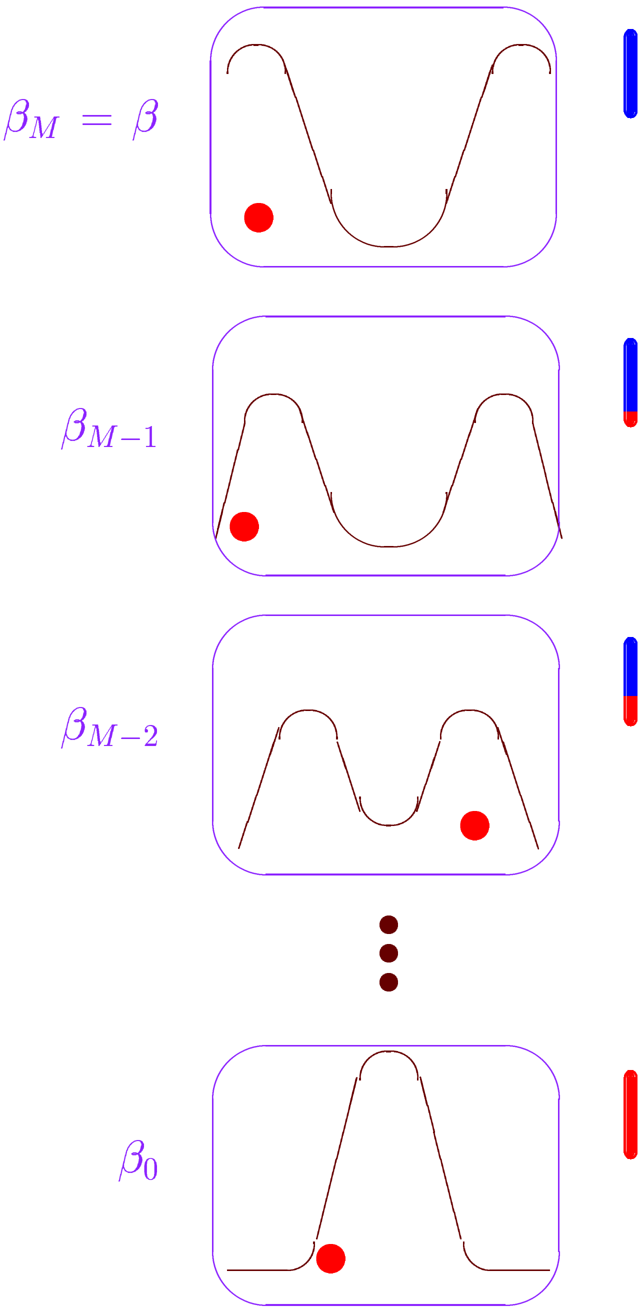


$$\overline{\pi}(a_i) = \pi(A_i)$$

$$\overline{P}(a_i, a_j) = \sum_{\substack{x \in A_i, \\ y \in A_j}} \frac{\pi(x)P(x,y)}{\pi(A_i)}$$

<u>Thm:</u> $Gap(\ P\ ) \geq \frac{1}{2}\, Gap(\ \overline{P}\ )(\min_i Gap(\ P_i\ ))$

Tempering / Swapping



$\mathrm{Tr}(\sigma) = (0, ..., 1, 0, 0)$

$$\widehat{\Omega} = \Omega^{M+1}$$

$$\beta_i = \frac{i}{M}\beta$$

$$\pi_i(\sigma_i) = \pi_{\beta_i}(\sigma_i)$$

$$\widehat{\pi}(\sigma) = \prod_{i=0}^{M} \pi_i(\sigma_i)$$

Def: The trace

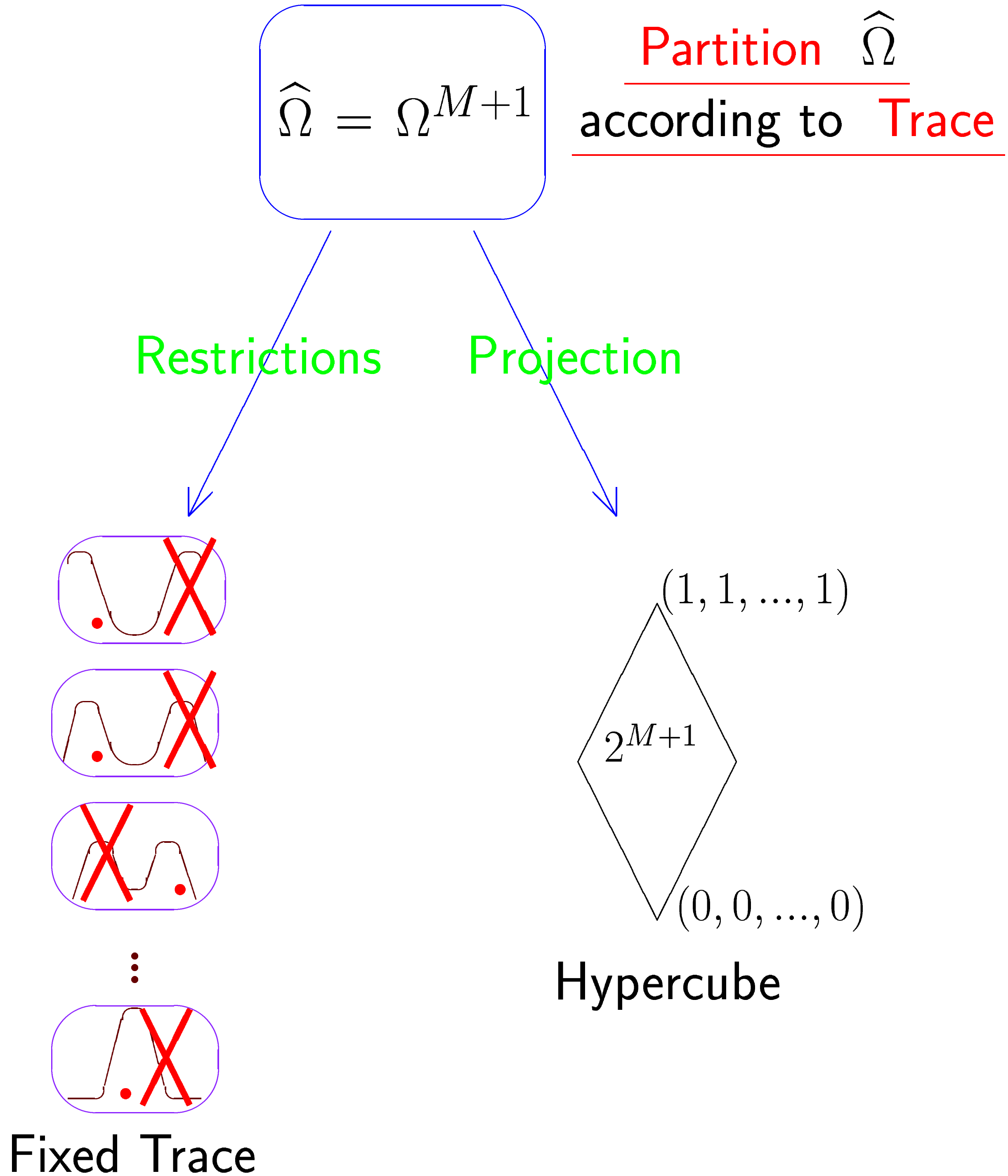
Tr: $\widehat{\Omega} \to \{0, 1\}^{M+1}$,

$\mathrm{Tr}(\sigma) = (b_0, ..., b_M)$,

$b_i = 0$ if σ_i is mostly $-$;
$b_i = 1$ if σ_i is mostly $+$.

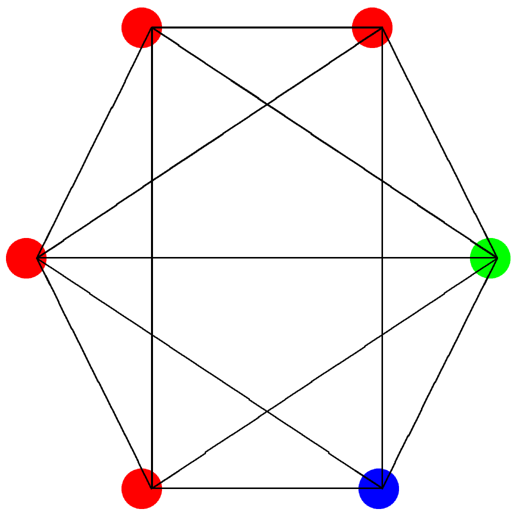
Thm: [Madras, Zheng] Swap is fast for the mean-field Ising model for all β.

Pf: (sketch) (using decomposition)



Swapping (Tempering): the Magic Answer?

The Potts Model



$\sigma \in \Omega = \{R, B, G\}^N$

(Hamiltonian)

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

(Gibbs measure) $(\beta = 1/T)$

- $\pi_\beta(\sigma) = e^{\beta H(\sigma)} / Z_\beta$

(Partition function)

- $Z_\beta = \sum_{\sigma' \in \Omega} e^{\beta H(\sigma')}$

Thm: [Bhatnagar, R.] The swap algorithm can be slow for the mean-field 3-state Potts model.

(Because of the first order (discontinuous) phase transition.)

Moreover, it can be exponentially slower than fixed-temperature Glauber dynamics !!!

Comments

1. There is a different way to sampling Ising states:

[Jerrum, Sinclair]: $\exists$ an efficient algorithm for estimating the partition function of an Ising system on any graph for any β.

[R., Wilson]: $\exists$ an efficient sampler for the Ising model for any graph, any β.

(But almost nothing is known for Potts!)

2. Phase trans'ns also exist for the uniform dist'n:

[Luby, Vigoda]: Glauber dynamics is fast for the Independent Set model on $R \subseteq \mathbb{Z}^2$.

[Galvin, Kahn] Above some sufficiently large dimension, Glauber dynamics is exponentially slow for the Independent Set model on $\mathbb{Z}^d$.

Future challenges ...

1. Characterize which sampling problems are computationally intractable.

2. Determine when specific Markov chain algorithms are inefficient.

3. Develop new techniques for anayzing potentially fast chains.

4. Design fast(er) Markov chains.