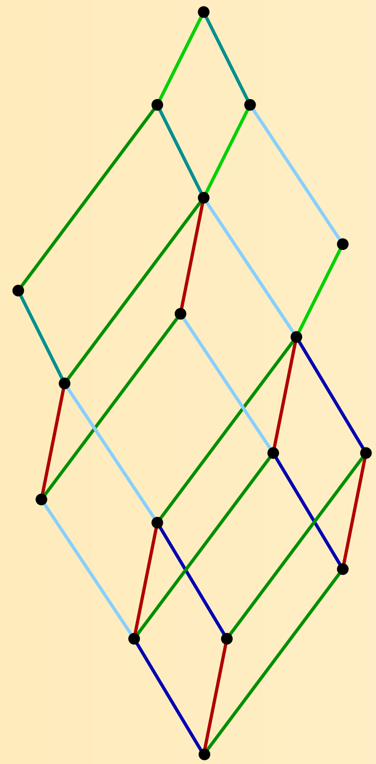
Sampling from distributive lattices – the Markov chain approach

Graduiertenkolleg MDS
TU Berlin
April 20., 2009



Stefan Felsner

Technische Universität Berlin
felsner@math.tu-berlin.de

Topics

Markov Chain Monte Carlo

Coupling and CFTP

Distributive Lattices

α-Orientations and Heights

Block Coupling for Heights

The Sampling Problem

- Ω a (large) finite set
- $\mu : \Omega \to [0, 1]$ a probability distribution

Problem. Sample from Ω according to μ.

$$\text{i.e., } \Pr(\mathsf{output} = \omega) = \mu(\omega).$$

The Sampling Problem

- Ω a (large) finite set
- $\mu : \Omega \to [0, 1]$ a probability distribution

Problem. Sample from Ω according to μ.

i.e., $\Pr(\mathsf{output} = \omega) = \mu(\omega)$.

There are many hard instances of the sampling problem.
Relaxation: Approximate sampling

i.e., $\Pr(\mathsf{output} = \omega) = \widetilde{\mu}(\omega)$ for some $\widetilde{\mu} \approx \mu$.

Applications of Sampling

- Get hand on typical examples from Ω.

- Approximate counting.

Preliminaries on Markov Chains

$\mathbf{M}$ transition matrix

- size $\Omega \times \Omega$
- entries $\in [0, 1]$
- row sums $= 1$ (stochastic)

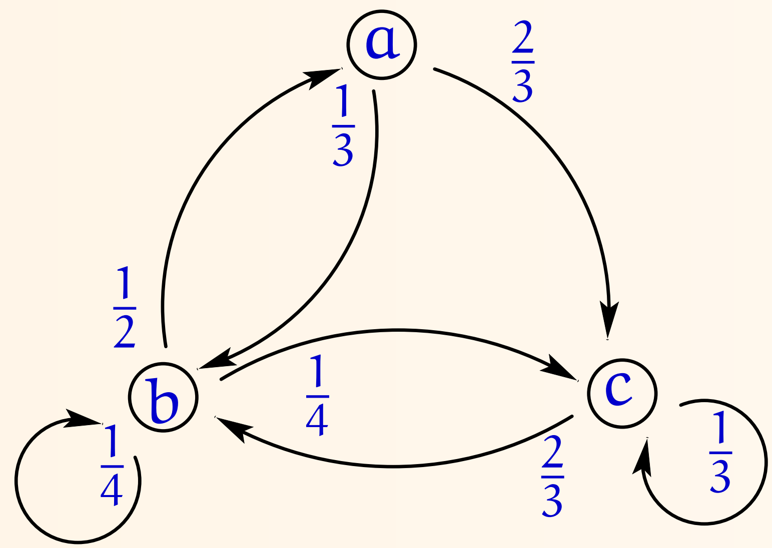
Preliminaries on Markov Chains

$\mathbf{M}$ transition matrix

- size $\Omega \times \Omega$
- entries $\in [0, 1]$
- row sums $= 1$ (stochastic)

Intuition:

$$\mathbf{M} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$



$\mathbf{M}$ specifies a random walk

Instance of a Markov Chains

$(X_0, X_1, X_2, \ldots X_r, \ldots)$ an instance of $\mathbf{M}$

- X_i random variable with values in Ω
- $\Pr(X_{i+1} = x \mid X_i = s) = \mathbf{M}(s, x)$

Proposition.
Probability distribution of X_t is μ_t with

$$\mu_t = \mu_0 \, \mathbf{M}^t$$

Ergodic Markov Chains

$\mathbf{M}$ is ergodic (i.e., irreducible and aperiodic)

$\Longrightarrow$ multiplicity of eigenvalue 1 is one

$\Longrightarrow$ unique π with $\pi = \pi\mathbf{M}$.

Fundamental Theorem.

$$\mathbf{M} \text{ ergodic} \implies \lim_{t\to\infty} \mu_0 \mathbf{M}^t = \pi.$$

Ergodic Markov Chains

$\mathbf{M}$ is ergodic (i.e., irreducible and aperiodic)

$\Longrightarrow$ multiplicity of eigenvalue 1 is one

$\Longrightarrow$ unique π with $\pi = \pi \mathbf{M}$.

Fundamental Theorem.

$$\mathbf{M} \text{ ergodic } \Longrightarrow \lim_{t \to \infty} \mu_0 \mathbf{M}^t = \pi.$$

$\mathbf{M}$ symmetric and ergodic

$\Longrightarrow$ $\mathbf{M}^\top \mathbb{1}^\top = \mathbf{M} \mathbb{1}^\top = \mathbb{1}^\top$, hence $\mathbb{1}\mathbf{M} = \mathbb{1}$

$\Longrightarrow$ π is the uniform distribution.

Example: Linear Extensions

A Markov chain for linear extensions

$L_t = x_1, x_2, \ldots, x_n$ the state at time t.

- Choose $i \in \{1, 2, \ldots, n-1\}$ uniformly.
- If x_i and x_{i+1} are incomparable, then $L_{t+1} = x_1, x_2, \ldots, x_{i-1}, x_{i+1}, x_i, x_{i+2}, \ldots, x_n$

Proposition. The chain is ergodic and symmetric.

Measuring Convergence

Variation distance

$$\|\mu - \mu'\|_{VD} := \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \mu'(x)|$$

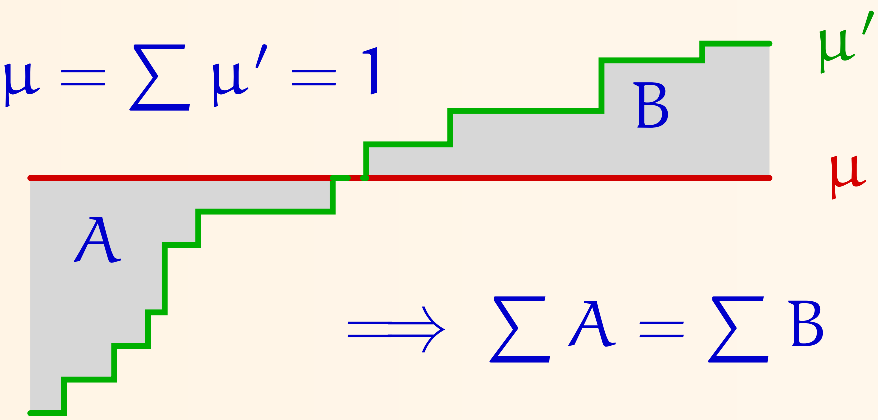
Measuring Convergence

Variation distance

$$\|\mu - \mu'\|_{VD} := \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \mu'(x)|$$

Lemma. $$\|\mu - \mu'\|_{VD} = \max_{A \subset \Omega} (\mu(A) - \mu'(A))$$

$$\sum \mu = \sum \mu' = 1$$



$$\implies \sum A = \sum B$$

Mixing Time

$\mu_x^t = \delta_x \mathbf{M}^t$ the distrib. after t steps starting in x

$\Delta(t) := \max(\|\mu_x^t - \pi\|_{VD} \ : \ x \in \Omega)$

$\tau(\varepsilon) = \min(t \ : \ \Delta(t) \leq \varepsilon)$

- $\tau(\varepsilon)$ is the mixing time.
- $\mathbf{M}$ is rapidly mixing $\iff$ $\tau(\varepsilon)$ is a polynomial function of the *problem size* and $\log(\varepsilon^{-1})$.

Mixing Time and Eigenvalues

- $\mathbf{M}$ stochastic $\implies |\lambda| \leq 1$ for all eigenvalues λ.
- $\mathbf{M}$ lazy (i.e., $m_{i,i} \geq 1/2$ for all i)
  $\implies \lambda \geq 0$ for all eigenvalues λ.
- $\mathbf{M}$ ergodic $\implies$ multiplicity of eigenvalue 1 is one.
- $\mathbf{M}$ symmetric $\implies$ ONB of eigenvectors.

Proposition. Mixing time, i.e., Convergence rate to π, depends on second largest eigenvalue.

Topics

Markov Chain Monte Carlo

Coupling and CFTP

Distributive Lattices

α-Orientations and Heights

Block Coupling for Heights

Coupling for Distributions

μ, ν distributions on Ω.

A distribution ω on $\Omega \times \Omega$ is a coupling of μ and ν
$\iff$ ω has μ and ν as marginals, i.e.,
$\sum_y \omega(x, y) = \mu(x)$ for all x and
$\sum_x \omega(x, y) = \nu(y)$ for all y.

Coupling Lemma.
ω a coupling of μ and ν and (X, Y) chosen from ω then

$$\|\mu - \nu\|_{VD} \leq \Pr(X \neq Y).$$

Coupling for Distributions

Lemma. $\|\mu - \nu\|_{VD} \le \Pr(X \neq Y)$.

Proof. We use $\mu(z) = \sum_y \omega(z, y) \ge \omega(z, z)$
$\nu(z) = \sum_x \omega(x, z) \ge \omega(z, z)$.

$$\begin{aligned}
\Pr(X \neq Y) &= 1 - \Pr(X = Y) \\
&= \sum_z \mu(z) - \sum_z \omega(z, z) \\
&\ge \sum_z \mu(z) - \sum_z \min(\mu(z), \nu(z)) \\
&= \sum_{z:\nu \le \mu} \mu(z) - \nu(z) \\
&= \max_{A \subset \Omega} \big(\mu(A) - \nu(A)\big) = \|\mu - \nu\|_{VD}
\end{aligned}$$

Coupling for Markov Chains

A coupling for $\mathbf{M}$ is a sequence $(Z_0, Z_1, Z_2, \ldots)$ with $Z_i = (X_i, Y_i)$ such that $(X_0, X_1, X_2, \ldots)$ and $(Y_0, Y_1, Y_2, \ldots)$ are instances for $\mathbf{M}$.

In particular

$$\begin{aligned} \Pr(X_{i+1} = x' \mid Z_i = (x, y)) &= \\ \Pr(X_{i+1} = x' \mid X_i = x) &= \mathbf{M}(x, x') \end{aligned}$$

Coupling and Mixing Times

$Z_i = (X_i, Y_i)$ a coupling for $\mathbf{M}$.

Theorem [Döblin 1938]**.**
If $\Pr\left(X_T \neq Y_T \mid Z_0 = (x_0, y_0)\right) < \varepsilon$ for every initial (x_0, y_0) and T steps

$$\Longrightarrow \tau(\varepsilon) \leq T$$

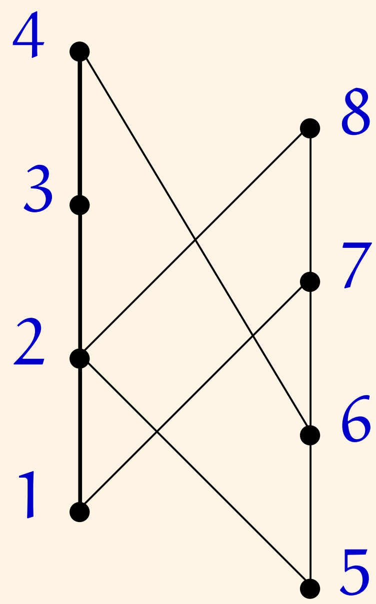
Proof. Choose y_0 from stationary distribution π
Y_t is in stationary distribution π for all t
X_t is in distribution $\mu_{x_0}^t$.

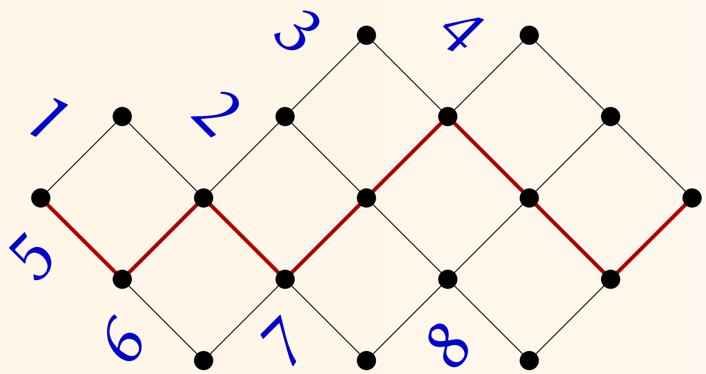
$$\Pr\left(X_T \neq Y_T \mid Z_0 = (x_0, y_0)\right) < \varepsilon$$

Coupling Lemma $\Longrightarrow$ $\max_x \|\mu_x^T - \pi\|_{VD} < \varepsilon$

definition of τ $\Longrightarrow$ $\tau(\varepsilon) \leq T$

Example :
Linear Extensions of Width 2 Orders





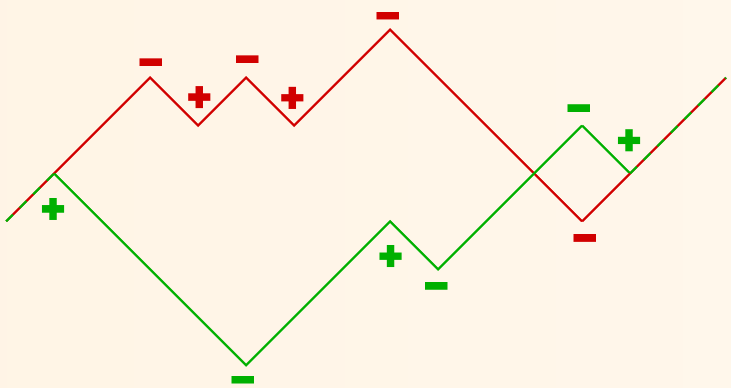
Linear extensions *are* paths.

The Markov chain and the coupling

- choose position k and $s \in \{\uparrow, \downarrow\}$
- Flip the path at position k in direction s (if possible)

Linear Extensions of Width 2 Orders
the Analysis

- $\mathsf{dist}(X, Y) =$ Area between paths $\leq n^2$
- $E(\mathsf{dist}(X_{i+1}, Y_{i+1})) \leq \mathsf{dist}(X_i, Y_i)$



The *distance* is a projection to a random walk on the line

$\Longrightarrow$ expected coupling time $O(n^4 \log n)$.

$\Longrightarrow$ $\tau(\varepsilon) \in O(n^4 \log n \log \varepsilon^{-1})$.

Coupling From the Past

$\mathbf{M}$ a Markov chain on Ω

$\mathcal{F}$ a family of maps $f : \Omega \to \Omega$ such that for random $f \in \mathcal{F}$:

$$\Pr(f(x) = x') = \mathbf{M}(x, x')$$

Coupling From the Past

$\mathbf{M}$ a Markov chain on Ω

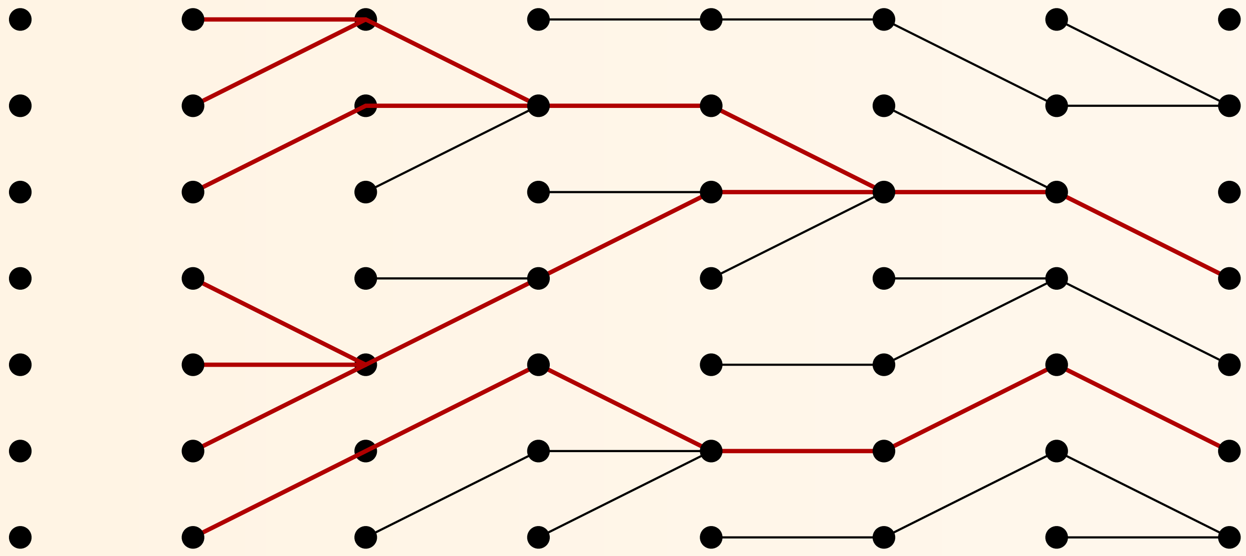
$\mathcal{F}$ a family of maps $\mathsf{f} : \Omega \to \Omega$ such that for random $\mathsf{f} \in \mathcal{F}$:

$$\Pr(\mathsf{f}(x) = x') = \mathbf{M}(x, x')$$

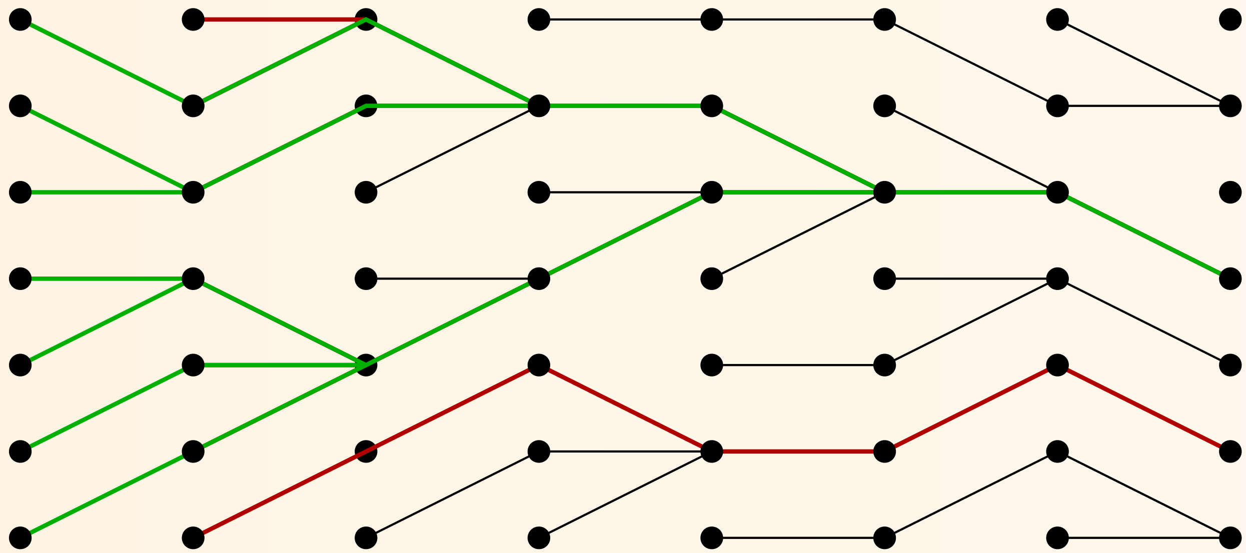
Coupling-FTP

```
F ← id_Ω
repeat
    choose f ∈ 𝓕 at random
    F ← F ∘ f
until F is a constant map
return F(x)
```

Coupling From the Past



Coupling From the Past



Theorem. The state returned by Coupling-FTP is exactly(!) in the stationary distribution.

Monotone Coupling From the Past: An Example

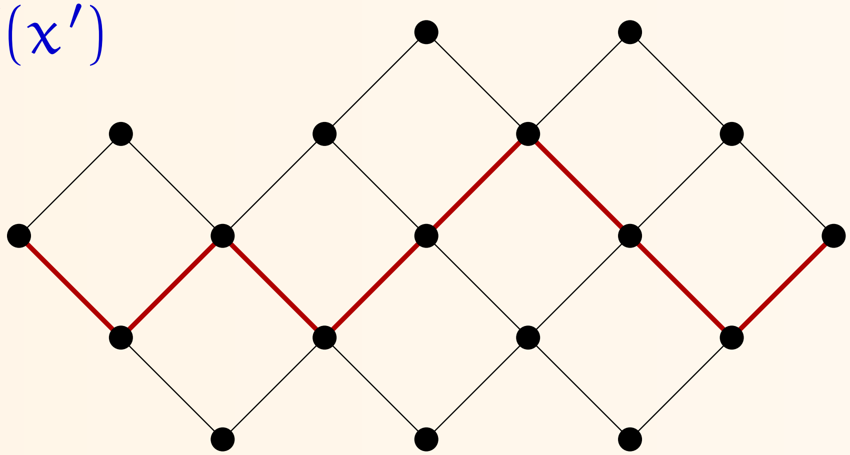
The problem with CFTP is the need of functions f on Ω.

Monotone Coupling From the Past: An Example

The problem with CFTP is the need of functions f on Ω.

Order relation $<_\Omega$ on Ω with $\hat{0}$ and $\hat{1}$

- $x <_\Omega x' \implies f(x) <_\Omega f(x')$
  for all $f \in \mathcal{F}$



Example:

Objects:
Lattice path in a grid

$\mathcal{F} = \{ f_{k,s}$: apply position k and direction s to all paths $\}$

This family is monotone!

Topics

Markov Chain Monte Carlo

Coupling and CFTP

Distributive Lattices

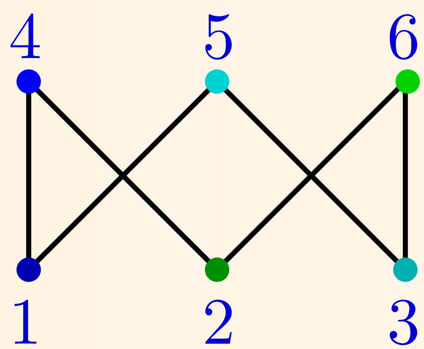
α-Orientations and Heights

Block Coupling for Heights

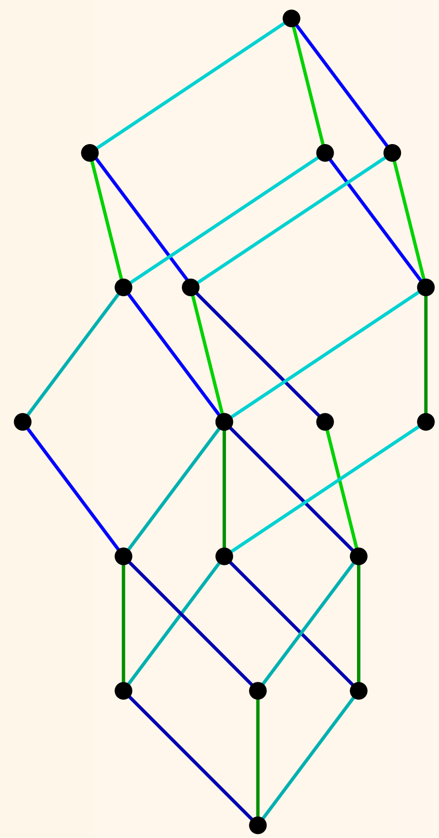
Distributive Lattices

Fact. $\mathcal{L}$ is a finite distributive lattice $\iff$ there is a poset P such that that $\mathcal{L}$ is isomorphic to the inclusion order on downsets of P.

P



$\mathcal{L}_{\mathsf{P}}$



Markov Chains on Distributive Lattices

A *natural* Markov chain on $\mathcal{L}_P$ (lattice walk):

Identify state with downset D

- choose $x \in P$
  choose $s \in \{\uparrow, \downarrow\}$
- depending on s move to $D + x$ or $D - x$
  (if possible)

Fact. The chain is ergodic and symmetric, i.e, π is uniform.

Monotone Coupling on Distributive Lattices

The coupling family $\mathcal{F}$:

$f_{x,s}$: Use element x and direction s for all D.
Is monotone!

$\implies$ uniform sampling from distributive lattices is easy.

Monotone Coupling on Distributive Lattices

The coupling family $\mathcal{F}$:

$f_{x,s}$: Use element x and direction s for all D.
Is monotone!

$\Longrightarrow$ uniform sampling from distributive lattices is easy.

Q: Is it fast (rapidly mixing)?

A: In most cases not.

Slow Mixing

- On distributive lattices based on Kleitman-Rothschild posets the mixing time of the lattice walk is exponential.
- The mixing time of the lattice walk is exponential for random bipartite graphs with degrees ≥ 6.
  (Dyer, Frieze and Jerrum)

Fast Mixing

- The mixing time of the lattice walk is polynomial for random bipartite graphs with max-degree ≤ 4. (Dyer and Greenhill)

In several situations where planarity plays a role rapid mixing could be proven:

- Monotone paths in the grid.
- Lozenge tilings of an $a \times b \times c$ hexagon.
- Domino tilings of a rectangle.

Topics

Markov Chain Monte Carlo

Coupling and CFTP

Distributive Lattices

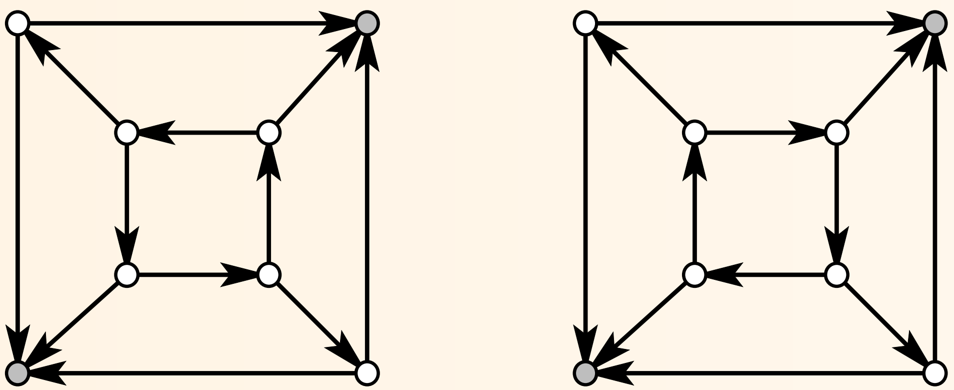
α-Orientations and Heights

Block Coupling for Heights

alpha-Orientations

Definition. Given $G = (V, E)$ and $\alpha : V \to \mathbb{N}$.
An α-orientation of G is an orientation with $\text{outdeg}(v) = \alpha(v)$ for all v.

Example.



Two orientations for the same α.

Potentials and Lattice Structure

Definition. An α-potential for G is a mapping $\wp:$ Faces $(G) \to \mathbb{Z}$ such that $\wp(\text{outer}) = 0$ and

- $|\wp(C) - \wp(C')| \leq 1$, if C and C' share an edge e.
- $\wp(C^{l(e)}) \leq \wp(C^{r(e)})$ for all e

  relative to some fixed α-orientation.

Lemma. There is a bijection between α-potentials and α-orientations.

Potentials and Lattice Structure

Definition. An α-potential for G is a mapping $\wp : \text{Faces}(G) \rightarrow \mathbb{Z}$ such that $\wp(\text{outer}) = 0$ and

- $|\wp(C) - \wp(C')| \leq 1$, if C and C' share an edge e.
- $\wp(C^{l(e)}) \leq \wp(C^{r(e)})$ for all e
  relative to some fixed α-orientation.

Lemma. There is a bijection between α-potentials and α-orientations.

Theorem. α-potentials are a distributive lattice with

$$(\wp_1 \vee \wp_2)(C) = \max\left\{\wp_1(C), \wp_2(C)\right\} \text{ and}$$
$$(\wp_1 \wedge \wp_2)(C) = \min\left\{\wp_1(C), \wp_2(C)\right\}.$$

Counting and Sampling

Proposition. Counting α-orientations is #P-complete for

- planar maps with $d(v) = 4$ and $\alpha(v) \in \{1, 2, 3\}$ and
- planar maps with $d(v) \in \{3, 4, 5\}$ and $\alpha(v) = 2$.

Problem.

- Is counting 3-orientations in triangulations #P-complete?
- Is counting 2-orientations in quadrangulations #P-complete?

Approximate Counting

Fact. The fully polynomial randomized approximation scheme for counting perfect matchings of bipartite graphs (Jerrum, Sinclair and Vigoda 2001) can be used for approximate counting of α-orientations.

Approximate Counting

Fact. The fully polynomial randomized approximation scheme for counting perfect matchings of bipartite graphs (Jerrum, Sinclair and Vigoda 2001) can be used for approximate counting of α-orientations.

- What about the lattice walk?

Lattice Walks for alpha-Orientations

Theorem [Fehrenbach 03].

- Sampling Eulerian orientations of simply connected patches of the quadrangular grid using the LW Markov chain is polynomial.

Theorem [Creed 05].

- Sampling Eulerian orientations of simply connected patches of the triangular grid using the LW Markov chain is polynomial.

- Sampling Eulerian orientations of patches of the triangular grid with holes using the LW Markov chain can be exponential.

alpha-Orientations and Heights

G planar

Definition. An α-potential for G is a mapping $\wp :$ Faces $(G) \to \mathbb{Z}$ such that $\wp(\text{outer}) = 0$ and

- $|\wp(C) - \wp(C')| \leq 1$, if C and C' share an edge e.
- $\wp(C^{l(e)}) \leq \wp(C^{r(e)})$ for all e relative to some fixed α-orientation.

Definition. A k-height for G is a mapping $H :$ Faces $(G) \to \{0, \ldots, k\}$ such that

- $|H(C) - H(C')| \leq 1$, if C and C' share an edge e.

Topics

Markov Chain Monte Carlo

Coupling and CFTP

Distributive Lattices

α-Orientations and Heights

Block Coupling for Heights

Height Lattices

Definition. A k-height for G is a mapping
$H: \text{Faces}(G) \to \{0, \ldots, k\}$ such that

- $|H(C) - H(C')| \leq 1$, if C and C' share an edge e.

Proposition. k-heights are a distributive lattice with

$$(H_1 \vee H_2)(C) = \max\left\{H_1(C), H_2(C)\right\} \text{ and}$$
$$(H_1 \wedge H_2)(C) = \min\left\{H_1(C), H_2(C)\right\}.$$

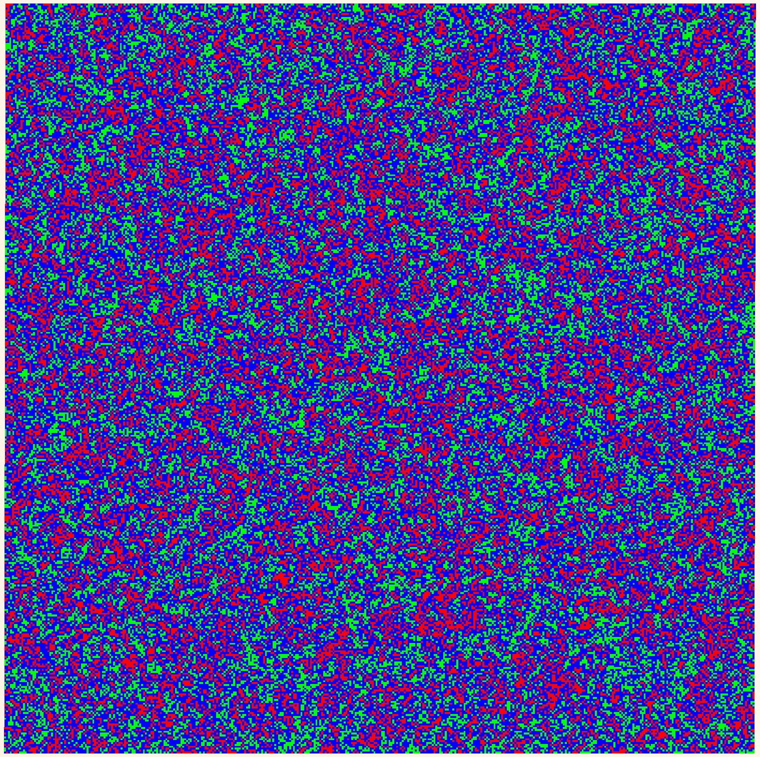
Sampling from Height Lattices

We can use monotone CFTP to sample uniformly from height lattices.

Sampling from Height Lattices

We can use monotone CFTP to sample uniformly from height lattices.

A random 2-height on the 400×400 square-grid. (38240593 steps)



Block Dynamics

- Experiments strongly suggest rapid mixing
  Our guess $c_k N^4 \log(N)$.

- A rigorous proof of rapid mixing for 2-heights on torus grids. We use block dynamics.

Block Dynamics

- Experiments strongly suggest rapid mixing
  Our guess $c_k N^4 \log(N)$.

- A rigorous proof of rapid mixing for 2-heights on torus grids. We use block dynamics.

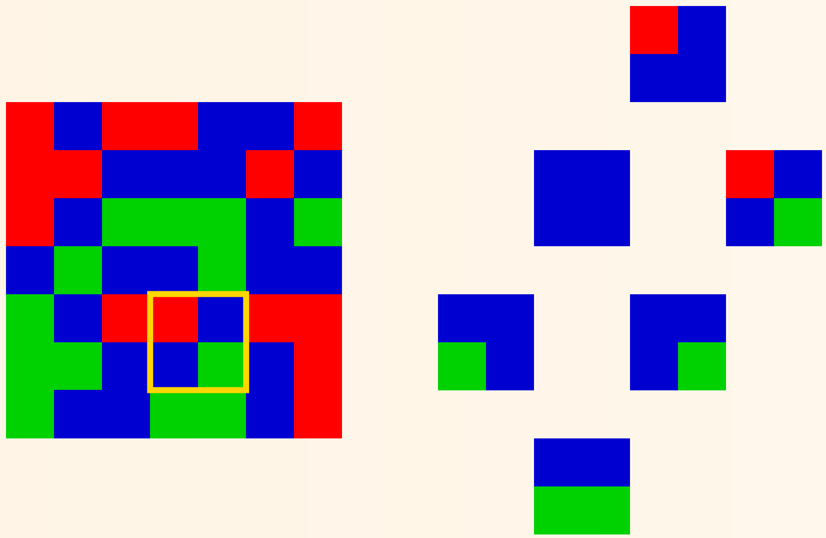
Block dynamics:

- choose a block $B \in \mathcal{B}$ such that $\Pr(f \in B) = \Pr(g \in B)$.

- choose heights for all faces in B respecting the heights on the border ∂B (uniform distribution).

Example

- choose heights for all faces in B respecting the heights on the border ∂B (uniform distribution).



Using Block Dynamics

Fact. The comparison technique yields:
If block dynamics is rapidly mixing then this also holds for the single step lattice walk.

Bound the mixing time via coupling

- Given instances H and H' choose the same block B for replacement in both.
- $\mathsf{dist}(\mathsf{H}, \mathsf{H}') := \sum_{\mathsf{f}} |\mathsf{H}(\mathsf{f}) - \mathsf{H}'(\mathsf{f})|$

Path Coupling

- With H and H' define $H = H_0, H_1, \ldots, H_d = H'$ such that $\mathsf{dist}(H_i, H_{i+1}) = 1$.
- Do the coupled block move on each H_i.

Goal: $E(\mathsf{dist}(H_i^+, H_{i+1}^+)) \leq 1$

- Consider f with $H_i(f) \neq H_{i+1}(f)$

$$f \in B \implies \mathsf{dist}(H_i^+, H_{i+1}^+) = 0$$

$$f \notin B \cup \partial B \implies \mathsf{dist}(H_i^+, H_{i+1}^+) = 1$$

$f \in \partial B$. (The hard case)
We sample from different distributions.

The Hard Case

Set up a monotone coupling

$$H_i \geq H_{i+1} \implies H_i^+ \geq H_{i+1}^+$$

(more about the existence later).

$$\begin{aligned} E(\mathrm{dist}(H_i^+, H_{i+1}^+)) &= E\Big(\sum_f |H_i^+(f) - H_{i+1}^+(f)|\Big) \\ &= E\Big(\sum_f H_i^+(f) - H_{i+1}^+(f)\Big) \\ &= E\Big(\sum_f H_i^+(f)\Big) - E\Big(\sum_f H_{i+1}^+(f)\Big) \end{aligned}$$

Combining the Cases

$$\delta := \max(E(H_h) - E(H_{h'}) \ : h, h' \text{ heights on } \partial B \text{ with } \mathsf{dist}(h, h') = 1)$$

For H_i, H_{i+1} with $\mathsf{dist}(H_i, H_{i+1}) = 1$ and a random block move on B with $|B| = k^2$ we get

$$E(\mathsf{dist}(H_i^+, H_{i+1}^+)) \leq 1 + \frac{4k\delta - k^2}{|\mathcal{B}|}$$

Hence we need: $4k\delta - k^2 \leq 0$

A Computer Proof

Blocks of size 6×6 suffice

- There are $3,3 \cdot 10^9$ possible h for the boundary.
- For a given h there are up to $3,7 \cdot 10^{12}$ compatible H for the block.

(work done by Daniel Heldt)

Stochastic Dominance and Strassen's

Definition. Stochastic dominance for distributions p_1 and p_2 on an ordered set $(A, \leq)$

$$p_1 \leq_{stoch} p_2 \iff \sum_{a \in F} p_1(a) \leq \sum_{a \in F} p_2(a) \text{ for all filter } F \subseteq A$$

Theorem [Strassen]**.** If $p_1 \leq_{stoch} p_2$ on $(A, \leq)$ then there is a distribution q on $A \times A$ with

- $q(x, y) > 0 \implies x \leq y$
- $\sum_y q(x, y) = p_1(x)$ and $\sum_x q(x, y) = p_2(y)$ (p_1 and p_2 are the marginals of q).

Existence of a Monotone Coupling

Strassen's Theorem implies the existence of the monotone block coupling if we can show that for $h_1 \leq h_2$ distributions on ∂B the induced distributions on B are in stochastic dominance.

Consider the intervals $A = \mathcal{D}_1$ and $B = \mathcal{D}_2$ of the height lattice over blocks.

We need that for every filter F of $\mathcal{D}$:

$$\frac{|A \cap F|}{|A|} \leq \frac{|B \cap F|}{|B|}$$

Existence of a Monotone Coupling

Goal: $|A \cap F||B| \leq |B \cap F||A|$

Restrict attention to the lattice L spanned by $\min A$ and $\max B$. L is distributive, A is an ideal, B a filter of L.

Define $f_1 = \chi_{A \cap F}$, $f_2 = \chi_B$, $f_3 = \chi_{B \cap F}$ and $f_4 = \chi_A$.

Lemma. $f_1(u)f_2(v) \leq f_3(u \vee v)f_4(u \wedge v)$

Ahlswede Daykin 4-Functions Theorem:

$$f_1(U)f_2(V) \leq f_3(U \vee V)f_4(U \wedge V)$$

We only need this for $U = V = L$.

Summary for Height Sampling

Theorem. The lattice walk for 2-heights on the square torus grid is rapidly mixing.

- Block dynamics and comparison method
- Monotone coupling from Strassen's via 4-FT.
- Valid blocks (6×6) from massive computations.

Summary for Height Sampling

Theorem. The lattice walk for 2-heights on the square torus grid is rapidly mixing.

- Block dynamics and comparison method
- Monotone coupling from Strassen's via 4-FT.
- Valid blocks (6×6) from massive computations.

Extension. The lattice walk for 2-heights on the planar traingulations is rapidly mixing.

Problems.

- $k > 2$.
- Other planar graphs.
- α-orientations.

The End

The End

Thank you.