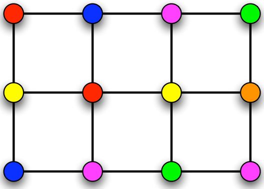
Tight bounds on mixing time of Markov chains

Mark Jerrum
School of Mathematical Sciences
Queen Mary, University of London

LMS Durham Symposium, 4th July 2007

Example 1: proper colourings (antiferro Potts)

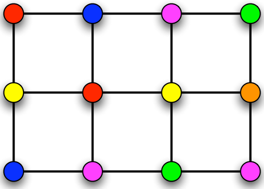
Instance: a graph $G = (V, E)$.



A (vertex) *colouring* of G is a an assignment $\sigma : V \to [q]$ of q "colours" $\{0, \dots, q-1\}$ to the vertices of G; it is *proper* if there are no monochromatic edges. Take "proper" as read.

Example 1: proper colourings (antiferro Potts)

Instance: a graph $G = (V, E)$.

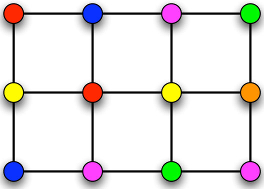


A (vertex) *colouring* of G is a an assignment $\sigma : V \to [q]$ of q "colours" $\{0, \ldots, q-1\}$ to the vertices of G; it is *proper* if there are no monochromatic edges. Take "proper" as read.

Problem

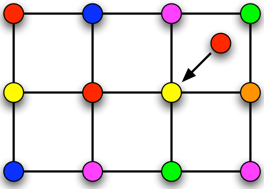
Sample a colouring of G uniformly at random (u.a.r.), *efficiently* (and certainly in time polynomial in $n = |V|$).

Markov chain Monte Carlo



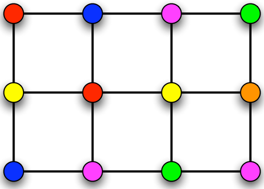
- Repeat:
  - Choose $v \in V$ and $c \in [q]$ u.a.r.
  - Let $\sigma' : V \to [q]$ be the colouring obtained by recolouring vertex v with colour c.
  - If σ' is a proper colouring then $\sigma := \sigma'$.

Markov chain Monte Carlo



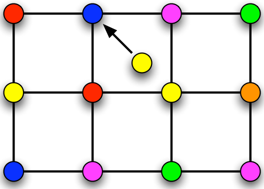
- Repeat:
  - Choose $v \in V$ and $c \in [q]$ u.a.r.
  - Let $\sigma' : V \to [q]$ be the colouring obtained by recolouring vertex v with colour c.
  - If σ' is a proper colouring then $\sigma := \sigma'$.

Markov chain Monte Carlo



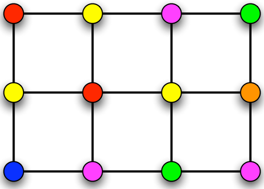
- Repeat:
  - Choose $v \in V$ and $c \in [q]$ u.a.r.
  - Let $\sigma' : V \to [q]$ be the colouring obtained by recolouring vertex v with colour c.
  - If σ' is a proper colouring then $\sigma := \sigma'$.

Markov chain Monte Carlo



- Repeat:
  - Choose $v \in V$ and $c \in [q]$ u.a.r.
  - Let $\sigma' : V \to [q]$ be the colouring obtained by recolouring vertex v with colour c.
  - If σ' is a proper colouring then $\sigma := \sigma'$.

Markov chain Monte Carlo



- Repeat:
  - Choose $v \in V$ and $c \in [q]$ u.a.r.
  - Let $\sigma' : V \to [q]$ be the colouring obtained by recolouring vertex v with colour c.
  - If σ' is a proper colouring then $\sigma := \sigma'$.

Mixing time

The trial just described defines the transition probabilities P of a Markov chain (X_t) on state space

$$\Omega = \{\text{All (proper) } q\text{-colourings of } G\}.$$

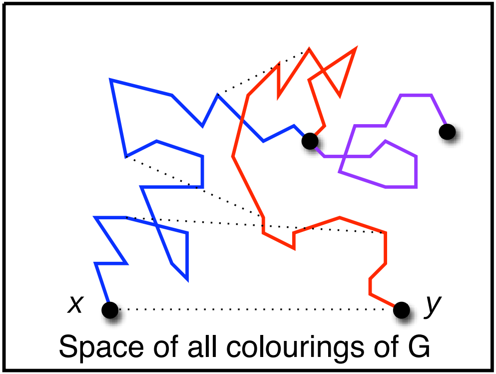
The Markov chain is irreducible and aperiodic (provided q is large enough) and its stationary distribution π is uniform.

We are interested in the *mixing time* τ of the Markov chain, i.e., the time to convergence to near stationarity:

$$\tau = \max_{x \in \Omega} \min \left\{ t : \| P^t(x, \cdot) - \pi \|_{\mathrm{TV}} \leq e^{-1} \right\},$$

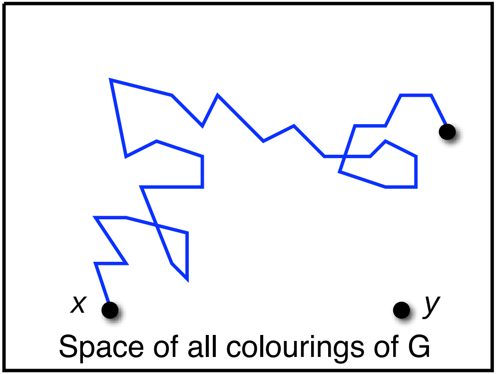
where $\|\varphi\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\varphi(x)|$.

Rough guide to coupling



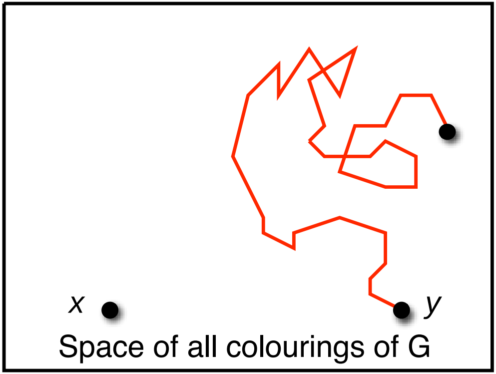
Two “coupled” evolutions of the Markov chain on the same sample space, but with different initial states.

Rough guide to coupling



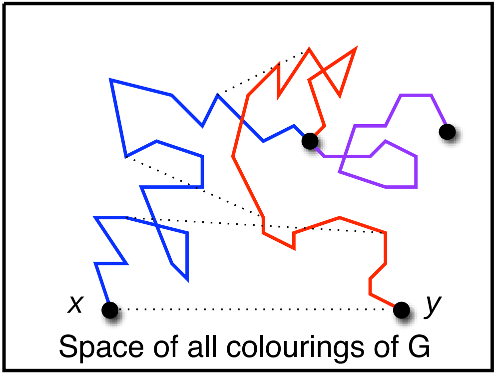
Projecting on the blue component we see a faithful copy. . .

Rough guide to coupling



Ditto projecting on red.

Rough guide to coupling



If the two can be made to coalesce rapidly, then the Markov chain must be rapidly mixing.

Basic coupling lemma

Consider a coupling $\big((X_t, Y_t) \in \Omega^2 : t \in \mathbb{N}\big)$.

Lemma

Suppose

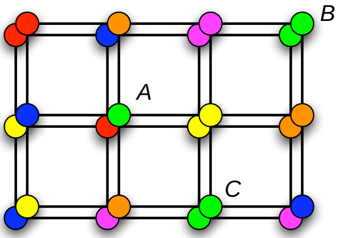
$$\Pr\left(X_t \neq Y_t \mid (X_0, Y_0) = (x_0, y_0)\right) \leq e^{-1}$$

for all choices of starting states (x_0, y_0). *Then* $\tau \leq t$.

[Doeblin 1938, Aldous 1983.]

Coupling: How it might be applied to colourings

Consider a pair of colourings $(X_t, Y_t) \in \Omega^2$.



The coupling:

- Choose the same vertex v in both copies.
- Choose the same colour c in both copies.
- Attempt to recolour vertex v in both X_t and Y_t with colour c; the result is (X_{t+1}, Y_{t+1}).

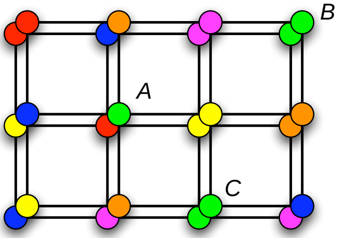
Analysis

Assume $q = 41$ colours and maximum degree 4.

Measure the progress of the coupling in terms of the Hamming distance $H(X_t, Y_t)$.

The are three basic cases, according to which vertex v is selected.

Analysis (continued)



- Type A. A vertex of disagreement. With probability at least $\frac{33}{41}$ the distance decreases by one.
- Type B. A vertex of agreement that is surrounded by other vertices of agreement. No change.
- Type C. A vertex of agreement adjacent to at least one vertex of disagreement. With probability at most $\frac{8}{41}$ the distance increases by one.

Analysis (continued)

Observe:

$$|\text{Type C}| \le 4 \times |\text{Type A}|.$$

Therefore

$$\begin{aligned} \mathrm{E}\, H(X_{t+1}, Y_{t+1} \mid X_t, Y_t) - H(X_t, Y_t) \\ \le \left(-\frac{33}{41} + 4 \times \frac{8}{41} \right) \frac{1}{n} H(X_t, Y_t), \end{aligned}$$

or

$$\mathrm{E}\, H(X_{t+1}, Y_{t+1} \mid X_t, Y_t) \le \left(1 - \frac{1}{41n} \right) H(X_t, Y_t).$$

Analysis (concluded)

$$
\begin{aligned}
\mathrm{E}\, H(X_t, Y_t \mid X_0, Y_0) &\le \left(1 - \frac{1}{41n}\right)^t H(X_0, Y_0) \\
&\le \left(1 - \frac{1}{41n}\right)^t n \\
&\le e^{-1},
\end{aligned}
$$

for $t = \lceil 41n(1 + \ln n) \rceil$.

So, by the Coupling Lemma, the mixing time is $O(n \log n)$.

Fewer colours

41 colours seems quite wasteful, and is!

Denote by Δ the maximum degree of the graph. It is not difficult to sharpen the coupling in order to establish $O(n \log n)$ mixing time whenever $q > 2\Delta$.

With much more effort one can weaken the condition to $q \geq (2 - \varepsilon)\Delta$, so 2Δ is not the barrier.

The Markov chain is irreducible when $q \geq \Delta + 2$, and there is no reason not to believe this is also the threshold for $O(n \log n)$ mixing. A lot of effort has gone into this, but current results are still quite a way short.

Matching lower bound

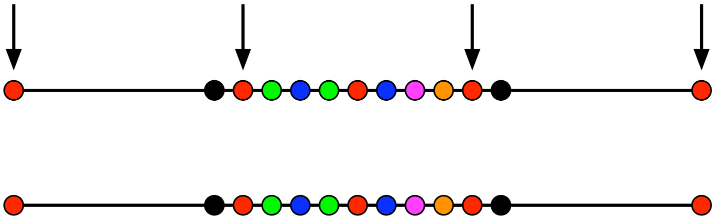
That $\Omega(n \log n)$ should be a lower bound seems obvious by a coupon-collector argument. But that is an illusion.

Dyer, Goldberg and Jerrum showed an $\Omega(n \log n)$ lower bound for a particular family of graphs.

Hayes and Sinclair showed that the same lower bound holds under quite general conditions, specifically for bounded degree graphs.

Sketch of lower bound proof

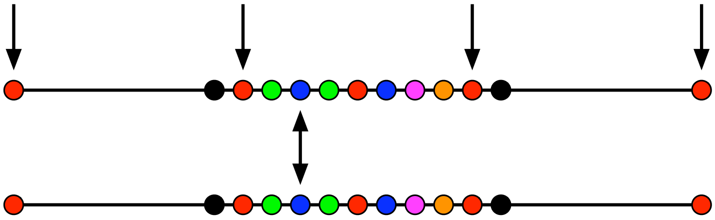
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof

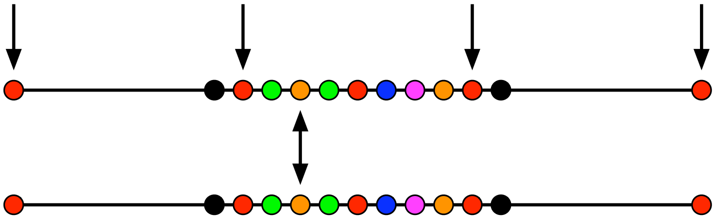
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is “free”.

Sketch of lower bound proof

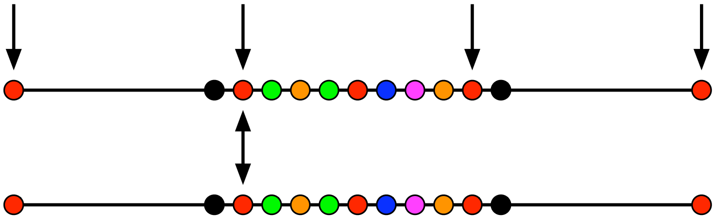
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is “free”.

Sketch of lower bound proof

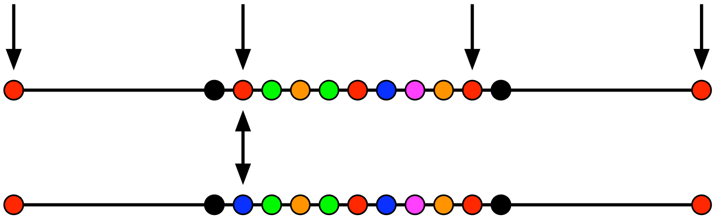
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof

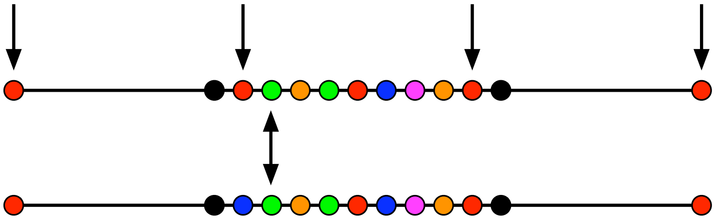
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof

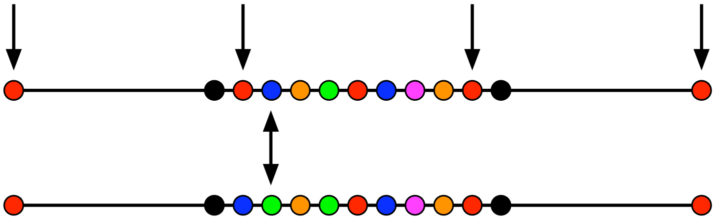
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof

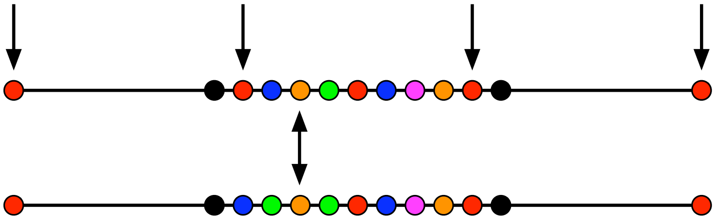
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof

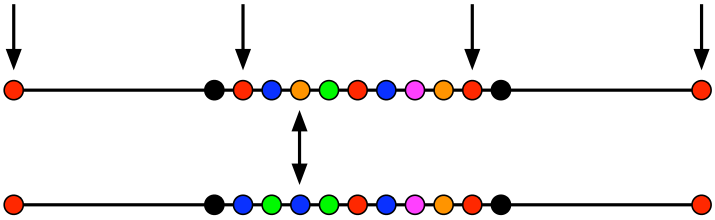
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof

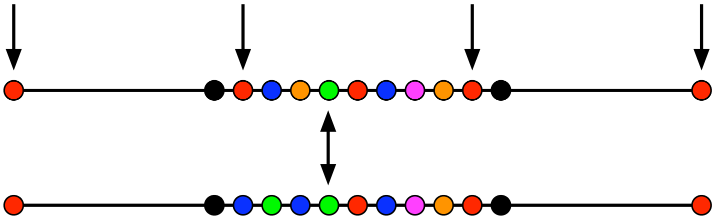
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof

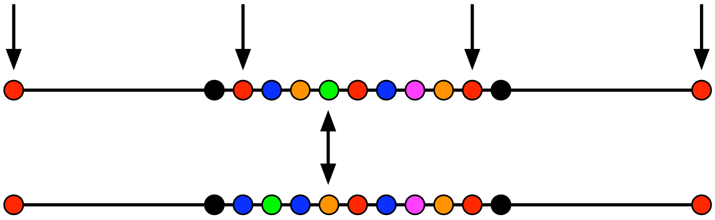
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof

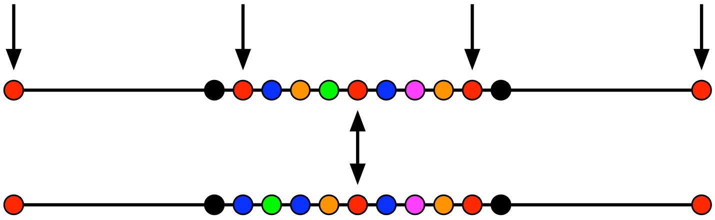
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is “free”.

Sketch of lower bound proof

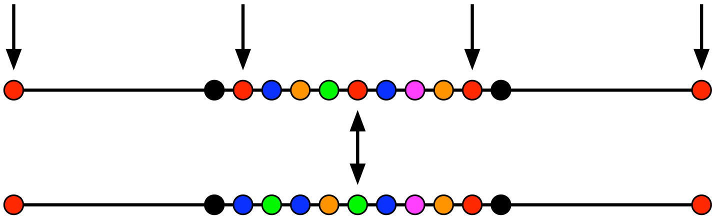
Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is “free”.

Sketch of lower bound proof

Specialise to the line (portion of $\mathbb{Z}$).
Top copy has certain vertices clamped to red.



Bottom version is "free".

Sketch of lower bound proof (concluded)

- Start with both copies in the “clamped” stationary distribution.

Sketch of lower bound proof (concluded)

- Start with both copies in the “clamped” stationary distribution.
- In the clamped stationary distribution, there is a slight preference for red in the centre of intervals.

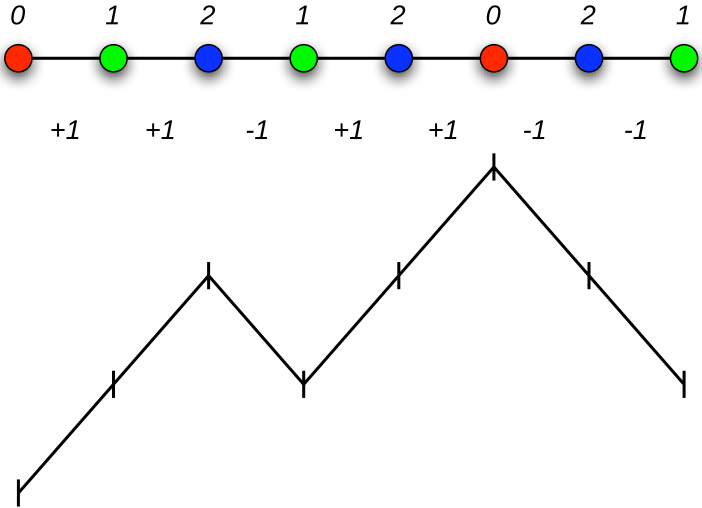
Sketch of lower bound proof (concluded)

- Start with both copies in the "clamped" stationary distribution.
- In the clamped stationary distribution, there is a slight preference for red in the centre of intervals.
- This preference for red persists also in the unclamped version until disagreements percolate from the ends of the intervals. This takes time $\Omega(n \log n)$.

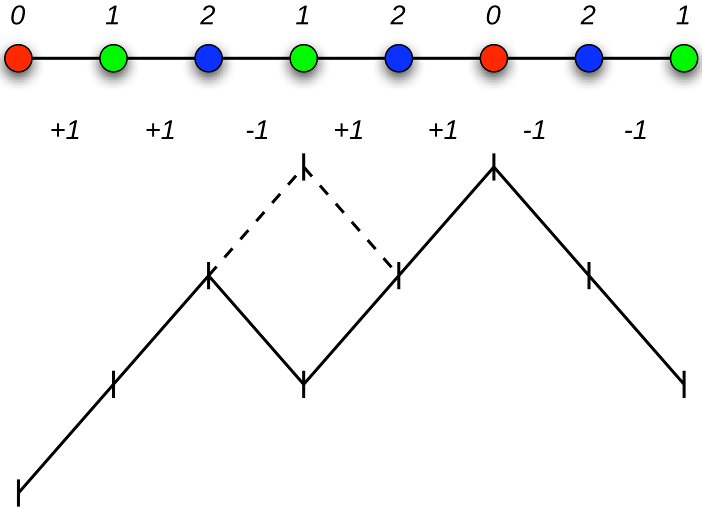
Sketch of lower bound proof (concluded)

- Start with both copies in the "clamped" stationary distribution.
- In the clamped stationary distribution, there is a slight preference for red in the centre of intervals.
- This preference for red persists also in the unclamped version until disagreements percolate from the ends of the intervals. This takes time $\Omega(n \log n)$.
- Until this happens, the unclamped version cannot be close to *its* stationary distribution.

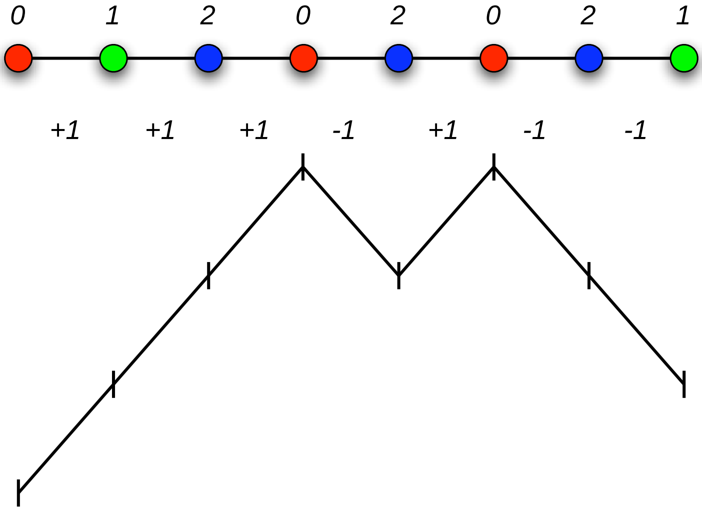
Example 2: 3-colourings of the line



Example 2: 3-colourings of the line



Example 2: 3-colourings of the line



Harmonic analysis

Introduced by Wilson (in the context of lozenge tilings).

General setting [Dyer, Goldberg & Jerrum]: states are vectors, and we require $\mathrm{E}[X_{t+1} \mid X_t] = A\,X_t$ for some matrix A.

Choose eigenvector w of A with eigenvalue λ. Define potential function $\Phi_t = w \cdot X_t$. To get an upper on mixing time we require $w > 0$ and "monotonicity". To get a lower bound we *require* a uniform bound $\mathrm{Var}(\Phi_t \mid \Phi_{t-1}) \le \varrho$ and we *want* λ to be as close as possible to 1.

Specialising to three-colourings

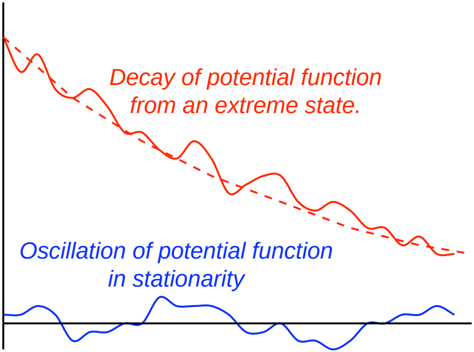
In our example, $A = I - \frac{1}{3n}B$, where

$$B = \begin{pmatrix} 3 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & & \ddots & & & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 3 \end{pmatrix},$$

and $w_i = \sin\left(\frac{(i-1/2)\pi}{n-1}\right) > 0$, for all i, and $1 - \lambda \sim 2\pi^2/3n^3$.

Lower bound: general considerations

Key facts: $\operatorname{E}\Phi_t = \lambda^t \Phi_0$ and $\operatorname{Var}\Phi_t \leq \varrho/(1-\lambda^2)$. Start at an initial state with Φ_0 large. We know that $\operatorname{E}\Phi_\infty = 0$.



Lower bound: three colourings

Starting at a state with $\Phi_0 = \Omega(n)$, we need $t = \Omega(n^3 \log n)$ steps for $\mathrm{E}\,\Phi_t$ to be consistent with near-stationarity. Thus we have a lower bound on mixing time of $\Omega(n^3 \log n)$.

This is for single-site updates. Using a similar but more involved calculation it is possible to prove a lower bound of $\Omega(n^2 \log n)$ for "systematic scan".

Upper bound: general considerations

Re-scale w so that $\min_i w_i = 1$. Define metric $d(x,y) = \sum_{i=1}^{n-1} w_i |x_i - y_i|$ for $x, y \in \{-1,+1\}^{n-1}$.

Consider a monotone coupling $(X(t), Y(t))$, with initial states $X(0) = (+1,+1,\dots,+1)'$ and $Y(0) = (-1,-1,\dots,-1)'$. Then, since $X(t) \geq Y(t)$,

$$\begin{aligned} \mathrm{E}[d(X(t),Y(t))] &= \mathrm{E}[w \cdot (X(t) - Y(t))] \\ &= w' A(X(t-1) - Y(t-1)) \\ &= \lambda\, w \cdot (X(t-1) - Y(t-1)) \\ &= \lambda\, d(X(t-1), Y(t-1)). \end{aligned}$$

Upper bound: 3 colourings

After $O(n^3 \log n)$ steps, $X(t) = Y(t)$ with high probability.

This is not quite the end of the story, since three 3-colourings map to one "height function". But a further coupling argument yield an upper bound of $O(n^3 \log n)$ for 3-colourings with single site updates.

With more work, one can show an upper bound of $O(n^2 \log n)$ for systematic scan.

A selection of open problems

- Show $O(n \log n)$ mixing of q-colourings of a graph of maximum degree Δ, where $q \geq \Delta + 2$.
- Mixing time of the bases-exchange walk for general matroids?
- Improved upper and lower bounds for existing Markov chains.
- Which functions arise as mixing times of spin systems with single-site dynamics: $n \log n$, $n^3 \log n, \ldots$. (Admittedly a vague question.)