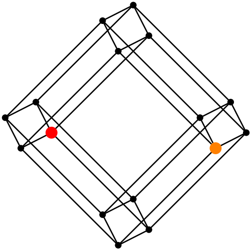
Coupling: co-adapted *vs.* maximal

(joint work with W.S. Kendall and S. Jacka, University of Warwick)

Stephen Connor
stephen.connor@york.ac.uk



Outline

Coupling

Let X be a Markov process with state space $\mathcal{S}$. We are interested in the situation where we have two copies of this process, X and Y, started from different states.

Definition

A *coupling* of X and Y is a process (X', Y') on $\mathcal{S} \times \mathcal{S}$, such that

$$X' \stackrel{\mathcal{D}}{=} X \quad \text{and} \quad Y' \stackrel{\mathcal{D}}{=} Y \,.$$

That is, viewed marginally, X' behaves as a version of X and Y' as a version of Y.

- The *coupling time* is defined by

$$\tau = \inf\left\{t \,:\, X'_s = Y'_s \quad \text{for all } s \geq t\right\}$$

- The coupling is *successful* if $\mathbb{P}\left(\tau < \infty\right) = 1$
  - τ is not, in general, a stopping time (for the marginal processes nor the joint process)
  - A 'good' coupling is usually one with a 'small' coupling time τ

- existence of a coupling is trivial: let X' and Y' be independent until they first meet, then stay together
  - this idea goes back to Doeblin (1938)

- a major use of coupling is to provide information about the convergence of X...

The coupling inequality

Definition

Let μ and ν be probability measures defined on S. The *total variation distance* between μ and ν is given by

$$\|\mu - \nu\|_{TV} = \sup_{A \subset \mathcal{S}} (\mu(A) - \nu(A))$$

The coupling inequality

Definition

Let μ and ν be probability measures defined on S. The *total variation distance* between μ and ν is given by

$$\|\mu - \nu\|_{TV} = \sup_{A \subset \mathcal{S}} (\mu(A) - \nu(A))$$

Lemma (The coupling inequality)

Let (X, Y) be a coupling as above. Then

$$\|\mathbb{P}(X_t \in \cdot) - \mathbb{P}(Y_t \in \cdot)\|_{TV} \leq \mathbb{P}(\tau > t).$$

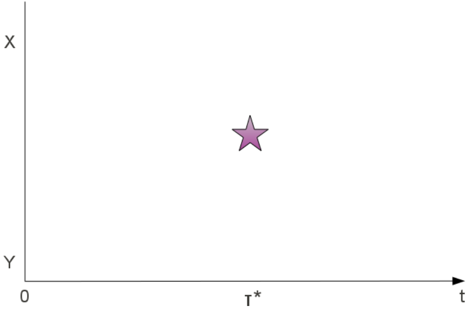
Maximal coupling

It is well known that there exists a **maximal coupling** of X and Y (Griffeath, 1975); that is, a joint process (X^*, Y^*) with coupling time τ^* satisfying

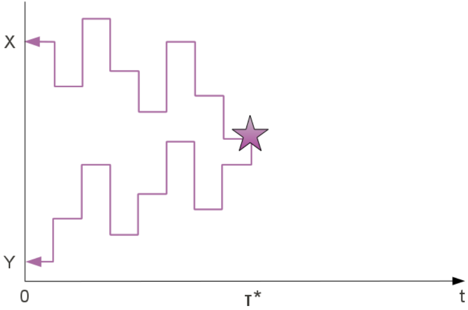
$$\|\mathbb{P}(X_t^* \in \cdot) - \mathbb{P}(Y_t^* \in \cdot)\|_{TV} = \mathbb{P}(\tau^* > t)\,.$$

- Thus there exists a successful coupling for X if and only if X is weakly ergodic

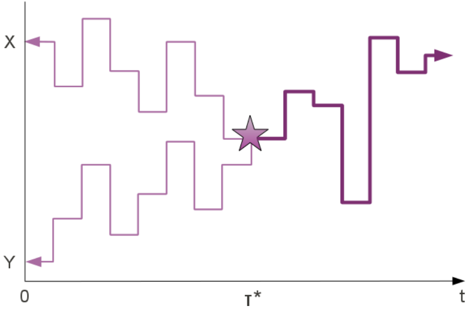
- Although a maximal coupling is known to exist, such a coupling is typically (at best) non-Markovian, unintuitive, and very difficult to compute explicitly – they are rarely used in practical applications



- Although a maximal coupling is known to exist, such a coupling is typically (at best) non-Markovian, unintuitive, and very difficult to compute explicitly – they are rarely used in practical applications



- Although a maximal coupling is known to exist, such a coupling is typically (at best) non-Markovian, unintuitive, and very difficult to compute explicitly – they are rarely used in practical applications



Co-adapted coupling

Definition

(X, Y) is called *co-adapted* if X and Y are both Markov with respect to a common filtration $(\mathcal{F}_t)$.

(We don't require that (X, Y) is Markov w.r.t. $(\mathcal{F}_t)$.)

Co-adapted coupling

Definition

(X, Y) is called *co-adapted* if X and Y are both Markov with respect to a common filtration $(\mathcal{F}_t)$.

(We don't require that (X, Y) is Markov w.r.t. $(\mathcal{F}_t)$.)

- It now suffices to study the first **collision time** of X and Y $\rightarrow$ X and Y can be made to agree from this time onwards
- co-adapted couplings are much more intuitive (neither process is allowed to 'cheat' by looking into the future)
- **maximal couplings are in general not co-adapted**

Co-adapted coupling

Definition

(X, Y) is called *co-adapted* if X and Y are both Markov with respect to a common filtration $(\mathcal{F}_t)$.

(We don't require that (X, Y) is Markov w.r.t. $(\mathcal{F}_t)$.)

- It now suffices to study the first **collision time** of X and Y $\rightarrow$ X and Y can be made to agree from this time onwards
- co-adapted couplings are much more intuitive (neither process is allowed to 'cheat' by looking into the future)
- **maximal couplings are in general not co-adapted**

But how good can a co-adapted coupling be?

Brownian motion: a maximal coupling

Consider two Brownian motions, X and Y, on $\mathbb{R}$, with $X_0 = x$ and $Y_0 = y$ (and $x \geq y$). Write

$$p_t(x, u) = \frac{e^{-(u-x)^2/2t}}{\sqrt{2\pi t}} .$$

It is simple to calculate the total variation distance between these two processes at any time t using the following result...

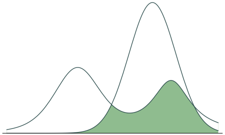
Lemma

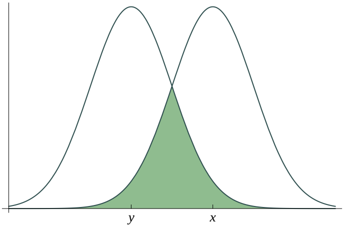
For probability measures μ *and* ν*, let* $\mu \wedge \nu$ *be their greatest common component, and let* λ *be a measure that dominates* μ *and* ν*. Write*

$$f = \frac{d\mu}{d\lambda}, \qquad f' = \frac{d\nu}{d\lambda}.$$

Then

$$\|\mu - \nu\|_{TV} = 1 - \int (f \wedge f')\, d\lambda.$$

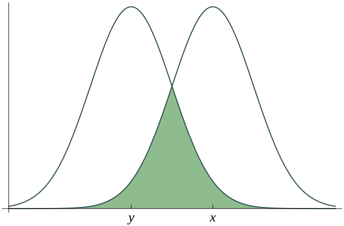




So...

$$\|\mathcal{L}(X_t) - \mathcal{L}(Y_t)\| = 1 - 2\int_{-\infty}^{(x+y)/2} p_t(x,z)dz$$
$$= \text{Erf}\left[\frac{(x-y)/2}{\sqrt{2t}}\right]$$
$$= \mathbb{P}_x\left(\tau_{(x+y)/2} > t\right) .$$

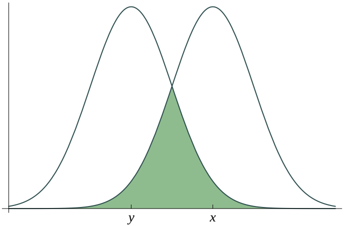
where $\tau_{(x+y)/2} = \inf\{t \geq 0 \,|\, X_t = (x+y)/2\}$.



So...

$$\|\mathcal{L}(X_t) - \mathcal{L}(Y_t)\| = 1 - 2\int_{-\infty}^{(x+y)/2} p_t(x,z)dz$$
$$= \operatorname{Erf}\left[\frac{(x-y)/2}{\sqrt{2t}}\right]$$
$$= \mathbb{P}_x\left(\tau_{(x+y)/2} > t\right) .$$

where $\tau_{(x+y)/2} = \inf\{t \geq 0 \mid X_t = (x+y)/2\}$.



So...

$$\|\mathcal{L}(X_t) - \mathcal{L}(Y_t)\| = 1 - 2\int_{-\infty}^{(x+y)/2} p_t(x,z)dz$$
$$= \text{Erf}\left[\frac{(x-y)/2}{\sqrt{2t}}\right]$$
$$= \mathbb{P}_x\left(\tau_{(x+y)/2} > t\right) .$$

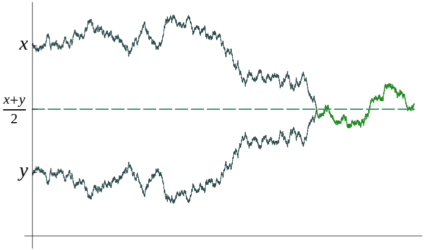
where $\tau_{(x+y)/2} = \inf\{t \geq 0 \,|\, X_t = (x+y)/2\}$.

Thanks to the symmetry of BM, this shows that *reflection coupling* is maximal for X and Y. In other words, if we define Y by

$$Y_t = \begin{cases} y - (X_t - x) & \text{for } t \le \tau_{(x+y)/2} \\ X_t & \text{for } t > \tau_{(x+y)/2}\,, \end{cases}$$

then

$$\|\mathcal{L}(X_t) - \mathcal{L}(Y_t)\| = \mathbb{P}_x\left(\tau_{(x+y)/2} > t\right) = \mathbb{P}(X_t \neq Y_t)\ .$$



Altering the initial conditions

- Note that this reflection coupling is co-adapted

Altering the initial conditions

- Note that this reflection coupling is co-adapted
- But what if we now alter the starting conditions, so that $X_0 = x$ is fixed, but $Y_0 \sim \mathrm{N}(0, \sigma^2)$?

Altering the initial conditions

- Note that this reflection coupling is co-adapted
- But what if we now alter the starting conditions, so that $X_0 = x$ is fixed, but $Y_0 \sim \mathrm{N}(0, \sigma^2)$?

Perhaps surprisingly...

Lemma

Under these initial conditions, reflection coupling for the pair of Brownian motions (X, Y) is not maximal

However, reflection is an *optimal co-adapted* coupling for X and Y when Y_0 is randomised using *any* distribution:

- any co-adapted coupling must be conditioned at time zero upon the σ-algebra $\mathcal{F}_0 = \sigma\{X_s, Y_s \ : \ s \leq 0\}$;
- in particular, the coupling scheme at time zero is conditioned on the event $\{Y_0 = y\}$;
- so the best that any co-adapted coupling can do is to match the coupling time of a maximal coupling between X and Y when $(X_0, Y_0) = (x, y)$, averaged over the distribution of Y_0;
- this bound is achieved **uniquely** by reflection coupling.

However, reflection is an *optimal co-adapted* coupling for X and Y when Y_0 is randomised using *any* distribution:

- any co-adapted coupling must be conditioned at time zero upon the σ-algebra $\mathcal{F}_0 = \sigma\{X_s, Y_s \;:\; s \leq 0\}$;
- in particular, the coupling scheme at time zero is conditioned on the event $\{Y_0 = y\}$;
- so the best that any co-adapted coupling can do is to match the coupling time of a maximal coupling between X and Y when $(X_0, Y_0) = (x, y)$, averaged over the distribution of Y_0;
- this bound is achieved **uniquely** by reflection coupling.

However, reflection is an *optimal co-adapted* coupling for X and Y when Y_0 is randomised using *any* distribution:

- any co-adapted coupling must be conditioned at time zero upon the σ-algebra $\mathcal{F}_0 = \sigma\{X_s, Y_s \ : \ s \leq 0\}$;
- in particular, the coupling scheme at time zero is conditioned on the event $\{Y_0 = y\}$;
- so the best that any co-adapted coupling can do is to match the coupling time of a maximal coupling between X and Y when $(X_0, Y_0) = (x, y)$, averaged over the distribution of Y_0;
- this bound is achieved **uniquely** by reflection coupling.

However, reflection is an *optimal co-adapted* coupling for X and Y when Y_0 is randomised using *any* distribution:

- any co-adapted coupling must be conditioned at time zero upon the σ-algebra $\mathcal{F}_0 = \sigma\{X_s, Y_s \;:\; s \leq 0\}$;
- in particular, the coupling scheme at time zero is conditioned on the event $\{Y_0 = y\}$;
- so the best that any co-adapted coupling can do is to match the coupling time of a maximal coupling between X and Y when $(X_0, Y_0) = (x, y)$, averaged over the distribution of Y_0;
- this bound is achieved **uniquely** by reflection coupling.

However, reflection is an *optimal co-adapted* coupling for X and Y when Y_0 is randomised using *any* distribution:

- any co-adapted coupling must be conditioned at time zero upon the σ-algebra $\mathcal{F}_0 = \sigma\{X_s, Y_s \ : \ s \leq 0\}$;
- in particular, the coupling scheme at time zero is conditioned on the event $\{Y_0 = y\}$;
- so the best that any co-adapted coupling can do is to match the coupling time of a maximal coupling between X and Y when $(X_0, Y_0) = (x, y)$, averaged over the distribution of Y_0;
- this bound is achieved **uniquely** by reflection coupling.

Question

Is there an intuitive description of a *maximal* coupling under these initial conditions?

Some extensions

- A similar set of results holds for couplings of two Ornstein-Uhlenbeck processes.
- What about couplings for more general diffusions?

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t$$

Question

Is *reflection* (reflecting B_t) maximal when X_0 and Y_0 are deterministic?

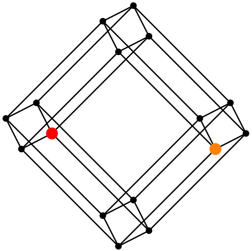
- it seems natural to conjecture that reflection is not maximal when $X_0 = x$ and Y_0 is randomised
- but is reflection still a (unique) optimal co-adapted coupling?

Random walk on the hypercube, $\mathbb{Z}_2^n$

Let $\mathbb{Z}_2^n$ be the group of binary n-tuples, under coordinate-wise addition modulo 2:

- this is the set of vertices of an n-dimensional cube
- we write $x \in \mathbb{Z}_2^n$ as $x = (x(1), \ldots, x(n)) \in \{0, 1\}^n$

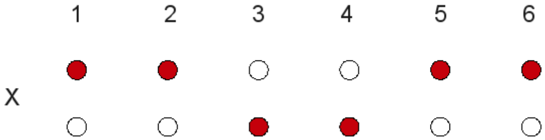
This is one of the simplest groups on which to study random walks, due to its high level of symmetry.



We define a simple, symmetric, continuous-time random walk X on $\mathbb{Z}_2^n$ as follows:

- let Λ_i, $1 \leq i \leq n$, be independent unit-rate Poisson processes
- at incident times of Λ_i, the i^{th} coordinate of X flips to its opposite value (0 or 1)
- unique equilibrium distribution is Uniform($\mathbb{Z}_2^n$)

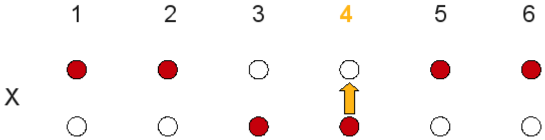
e.g., with $n = 6$:



We define a simple, symmetric, continuous-time random walk X on $\mathbb{Z}_2^n$ as follows:

- let Λ_i, $1 \leq i \leq n$, be independent unit-rate Poisson processes
- at incident times of Λ_i, the i^{th} coordinate of X flips to its opposite value (0 or 1)
- unique equilibrium distribution is Uniform($\mathbb{Z}_2^n$)

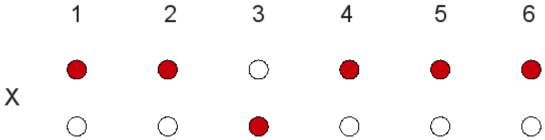
e.g., with $n = 6$:



We define a simple, symmetric, continuous-time random walk X on $\mathbb{Z}_2^n$ as follows:

- let Λ_i, $1 \leq i \leq n$, be independent unit-rate Poisson processes
- at incident times of Λ_i, the i^{th} coordinate of X flips to its opposite value (0 or 1)
- unique equilibrium distribution is Uniform($\mathbb{Z}_2^n$)

e.g., with $n = 6$:



Now suppose that we wish to couple two such random walks, X and Y, with $X_0 = (0, 0, \ldots, 0)$ and $Y_0 \sim \text{Uniform}(\mathbb{Z}_2^n)$

Maximal coupling

An almost-maximal coupling was described by Matthews (1987):

- not co-adapted, but still intuitive!
- expected coupling time $\sim (\log n)/4$

Now suppose that we wish to couple two such random walks, X and Y, with $X_0 = (0, 0, \ldots, 0)$ and $Y_0 \sim \text{Uniform}(\mathbb{Z}_2^n)$

Maximal coupling

An almost-maximal coupling was described by Matthews (1987):

- not co-adapted, but still intuitive!
- expected coupling time $\sim (\log n)/4$

But how good can a co-adapted coupling be for this process?

Optimal co-adapted coupling

Any co-adapted coupling must satisfy three constraints (all imposed by the marginal processes $X(i)$ being unit-rate Poisson processes):

- in any instant, no. of jumps of (X, Y) cannot exceed two
- all 'single' and 'double' jumps have rates bounded above by 1
- the *total* rate at which $X(i)$ jumps must equal 1

Optimal co-adapted coupling

Any co-adapted coupling must satisfy three constraints (all imposed by the marginal processes $X(i)$ being unit-rate Poisson processes):

- in any instant, no. of jumps of (X, Y) cannot exceed two
- all 'single' and 'double' jumps have rates bounded above by 1
- the *total* rate at which $X(i)$ jumps must equal 1

This allows us to describe *any* co-adapted coupling (X, Y) using

- **marked Poisson processes Λ_{ij} $(0 \leq i, j \leq n)$**
- **a $(n+1) \times (n+1)$ matrix-valued control process, Q**

Let $\mathcal{C}$ be the class of all co-adapted couplings.

Stochastic control problem

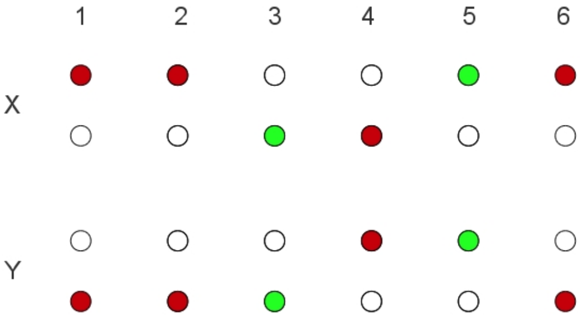
We now just need to find the best control process Q:

Method

1. Make a **guess** at a good process: call it $\hat{Q}$;
2. Choose a **cost function** v, that measures how good any control process Q is
   e.g. $v(Q) = \mathbb{E}[\tau_Q]$, or $v(Q,t) = \mathbb{P}(\tau_Q > t)$;
3. Use **Bellman's principle** to show that v is minimized by $\hat{Q}$.

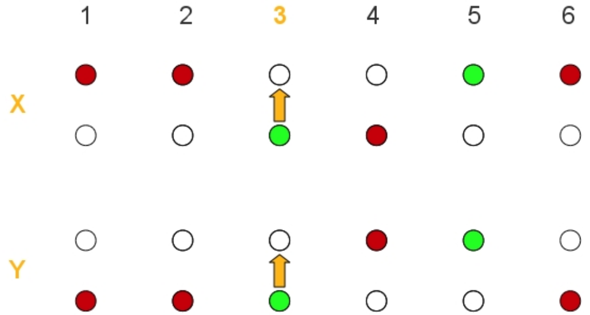
Making a good guess: intuition

- *Matching* coordinates should be made to move *synchronously*:



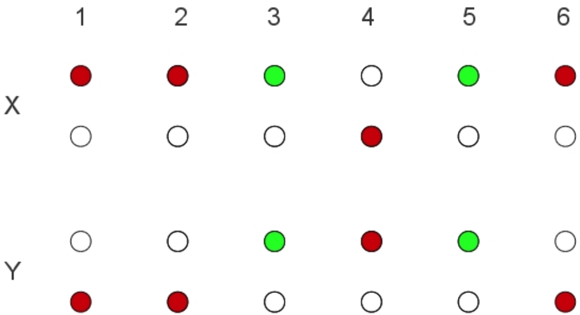
Making a good guess: intuition

- *Matching* coordinates should be made to move *synchronously*:



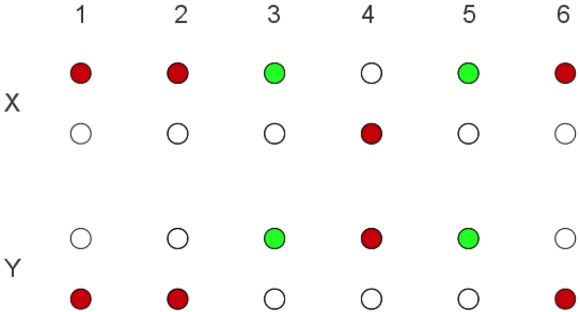
Making a good guess: intuition

- *Matching* coordinates should be made to move *synchronously*:



Making a good guess: intuition

- *Matching* coordinates should be made to move *synchronously*:

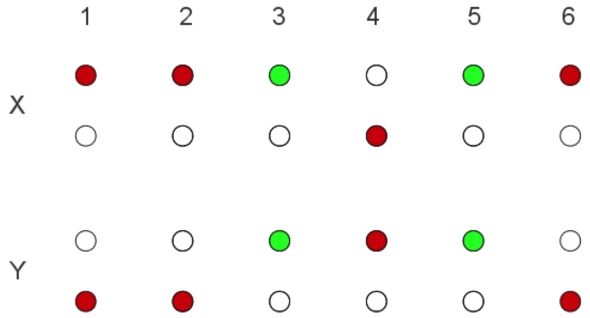


- Coupling strategy should depend only on

$$N_t = \text{no. of unmatched bits at time } t$$

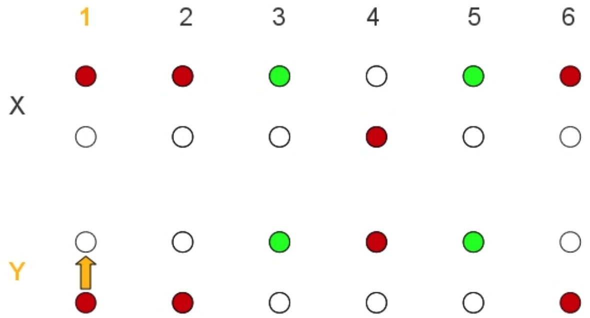
Two possible strategies (I)

- Allow unmatched bits to evolve **independently**:



Two possible strategies (I)

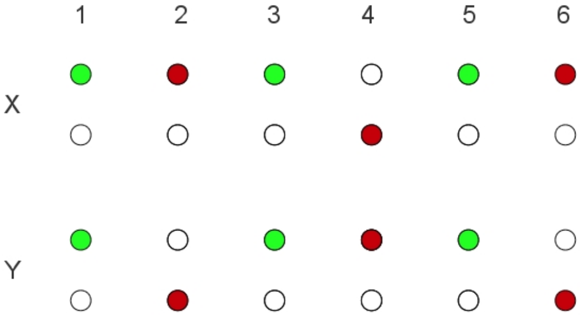
- Allow unmatched bits to evolve **independently**:



$\rightarrow$ **single** new matches are made at rate $\mathbf{2N_t}$

Two possible strategies (I)

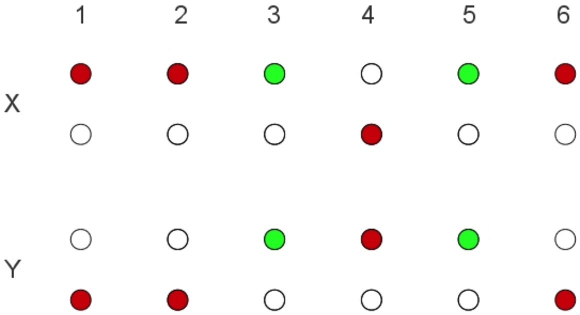
- Allow unmatched bits to evolve **independently**:



$\rightarrow$ **single** new matches are made at rate $\mathbf{2N_t}$

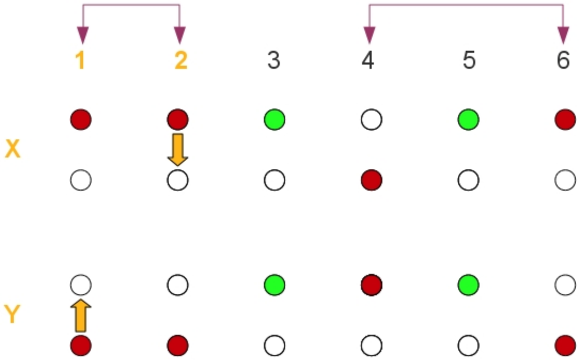
Two possible strategies (II)

- **Pair** unmatched bits:



Two possible strategies (II)

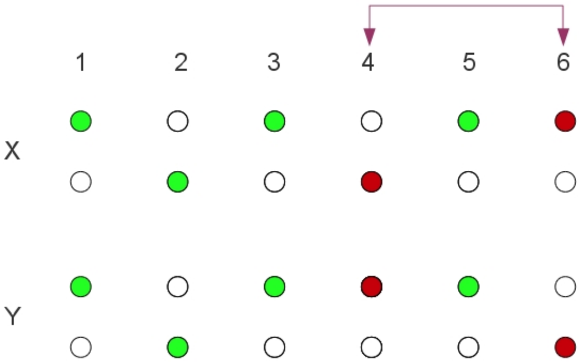
- **Pair** unmatched bits:



$\rightarrow$ **two** new matches are made at rate $\mathbf{N_t}$

Two possible strategies (II)

- **Pair** unmatched bits:



$\rightarrow$ **two** new matches are made at rate $\mathbf{N_t}$

Main result: the optimal co-adapted coupling

Let $\hat{Q}$ control (X, Y) as follows:

1. matched bits move synchronously;

Main result: the optimal co-adapted coupling

Let $\hat{Q}$ control (X, Y) as follows:

1. matched bits move synchronously;
2. if N_t is **even** then coupled unmatched bits in **pairs**;

Main result: the optimal co-adapted coupling

Let $\hat{Q}$ control (X, Y) as follows:

1. matched bits move synchronously;
2. if N_t is **even** then coupled unmatched bits in **pairs**;
3. if N_t is **odd** then let unmatched bits move **independently**.

Main result: the optimal co-adapted coupling

Let $\hat{Q}$ control (X, Y) as follows:

1. matched bits move synchronously;
2. if N_t is **even** then coupled unmatched bits in **pairs**;
3. if N_t is **odd** then let unmatched bits move **independently**.

Let $\hat{\tau}$ be the associated coupling time. What cost function should we use?

Main result: the optimal co-adapted coupling

Let $\hat{Q}$ control (X, Y) as follows:

1. matched bits move synchronously;
2. if N_t is **even** then coupled unmatched bits in **pairs**;
3. if N_t is **odd** then let unmatched bits move **independently**.

Let $\hat{\tau}$ be the associated coupling time. What cost function should we use?

$$\text{Define } \hat{v}(x, y, t) = \mathbb{P}\left(\hat{\tau} > t \,|\, X_0 = x, Y_0 = y\right) .$$

Main result: the *stochastically* optimal co-adapted coupling

Let $\hat{Q}$ control (X, Y) as follows:

1. matched bits move synchronously;
2. if N_t is **even** then coupled unmatched bits in **pairs**;
3. if N_t is **odd** then let unmatched bits move **independently**.

Let $\hat{\tau}$ be the associated coupling time. What cost function should we use?

$$\text{Define } \hat{v}(x, y, t) = \mathbb{P}\left(\hat{\tau} > t \,|\, X_0 = x, Y_0 = y\right) .$$

Theorem (Connor & Jacka, 2008)

For any states $x, y \in \mathbb{Z}_2^n$ and time $t \geq 0$,

$$\hat{v}(x, y, t) = \inf_{c \in \mathcal{C}} \mathbb{P}\left(\tau^c > t \,|\, X_0 = x, Y_0 = y\right) .$$

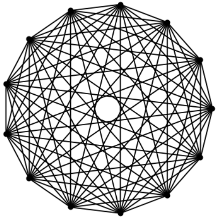
Remarks

- Proof uses notion of *totally monotone functions*;
- The optimal co-adapted coupling is **not** a maximal coupling:

$$\mathbb{E}\left[\hat{\tau}\right] \sim \frac{1}{2}\log n \quad \text{but} \quad \mathbb{E}\left[\tau^*\right] \sim \frac{1}{4}\log n\,;$$

- If the rate at which $X(i)$ flips is allowed to *vary* with i then there is, in general, no stochastically optimal coupling.

Analysis can be extended to simple symmetric random walk on G_d^n, where G_d is the *complete graph* on d vertices.



Theorem (Connor, 2009)

There exists a stochastically minimal co-adapted coupling for this random walk:

- *the coupling is not maximal for any fixed d;*
- *but as $d \to \infty$, the coupling tends to a maximal coupling.*

Concluding remarks

- For some processes (such as the transposition shuffle on S_n) co-adapted couplings perform *much* worse than maximal couplings, while for others (such as the examples above) there is not a big difference
- Not obvious *a priori* when a co-adapted maximal coupling will exist for a given process, nor how big the 'gap' between maximal and optimal co-adapted couplings will be
- Might lead to an interesting classification system for *e.g.* random walks on groups
- Possibly interesting consequences for perfect simulation algorithms?

References

Connor, S. B. (2009).
Optimal co-adapted coupling for a random walk on the hyper-complete-graph.

Connor, S. B. and S. Jacka (2008).
Optimal co-adapted coupling for the symmetric random walk on the hypercube.
Journal of Applied Probability 45, 703–713.

Griffeath, D. (1975).
A maximal coupling for Markov chains.
Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 31, 95–106.

Matthews, P. (1987).
Mixing rates for a random walk on the cube.
SIAM J. Algebraic Discrete Methods 8(4), 746–752.