

OPTIMAL COADAPTED COUPLING FOR A RANDOM WALK ON THE HYPER-COMplete GRAPH

STEPHEN CONNOR,* *University of York*

Abstract

The problem of constructing an optimal coadapted coupling for a pair of symmetric random walks on \mathbb{Z}_2^d was considered by Connor and Jacka (2008), and the existence of a coupling which is stochastically fastest in the class of all such coadapted couplings was demonstrated. In this paper we show how to generalise this construction to an optimal coadapted coupling for the continuous-time symmetric random walk on K_n^d , where K_n is the complete graph with n vertices. Moreover, we show that although this coupling is not maximal for any n (i.e. it does not achieve equality in the coupling inequality), it does tend to a maximal coupling as $n \rightarrow \infty$.

Keywords: Optimal coupling; coadapted; stochastic control; random walk on a group; cutoff phenomenon

2010 Mathematics Subject Classification: Primary 93E20
Secondary 60J27

1. Introduction

The concept of coupling two copies of a Markov chain in order to prove ergodicity statements dates back to [7], and is now a well-used and elegant technique. Suppose that we have a Markov process X on some state space \mathcal{S} , and let Y be a copy of X started from a different initial distribution: a *coupling* of these two processes is defined as follows.

Definition 1.1. (*Coupling.*) A *coupling* of X and Y is a process (X^c, Y^c) on $\mathcal{S} \times \mathcal{S}$ such that

$$X^c \stackrel{D}{=} X \quad \text{and} \quad Y^c \stackrel{D}{=} Y,$$

where ' $\stackrel{D}{=}$ ' denotes equality in distribution.

That is, viewed marginally, X^c behaves as a version of X , and Y^c behaves as a version of Y . The *coupling time* τ^c is defined by

$$\tau^c = \inf\{t : X_s^c = Y_s^c \text{ for all } s \geq t\},$$

and the coupling is called *successful* if $\mathbb{P}(\tau^c < \infty) = 1$.

Recall that the coupling inequality bounds the tail distribution of *any* coupling of X and Y by the total variation distance between the two processes (see, e.g. [14]):

$$\|\mathcal{L}(X_t) - \mathcal{L}(Y_t)\|_{\text{TV}} \leq \mathbb{P}(\tau^c > t). \tag{1.1}$$

Received 23 August 2011; revision received 2 November 2012.

* Postal address: Department of Mathematics, University of York, York, YO10 5DD, UK.

Email address: stephen.connor@york.ac.uk

Here $\mathcal{L}(X_t)$ is the law of X_t . Moreover, there always exists a *maximal* coupling of X and Y , that is, one which achieves equality in (1.1) [9], [17]. Thus, in order to obtain a good estimate of the rate at which the distributions of X_t and Y_t converge, it suffices to find a ‘good’ coupling—one which has a small coupling time τ^c —and preferably one which is maximal. However, maximal couplings are, in most cases, unintuitive non-Markovian affairs, and extremely difficult to work with. It is therefore natural to consider the class of *coadapted* couplings. These may not be maximal, but are generally more intuitive; nearly all couplings used in practice fall into this class.

Definition 1.2. (*Coadapted coupling.*) A coupling (X^c, Y^c) is called *coadapted* if there exists a filtration $(\mathcal{F}_t)_{t \geq 0}$ such that

1. X^c and Y^c are both adapted to $(\mathcal{F}_t)_{t \geq 0}$;
2. for any $0 \leq s \leq t$,

$$\mathcal{L}(X_t^c \mid \mathcal{F}_s) = \mathcal{L}(X_t^c \mid X_s^c) \quad \text{and} \quad \mathcal{L}(Y_t^c \mid \mathcal{F}_s) = \mathcal{L}(Y_t^c \mid Y_s^c).$$

In other words, (X^c, Y^c) is coadapted if X^c and Y^c are both Markov with respect to a common filtration: for coadapted couplings, it is always possible to make X^c and Y^c coalesce at the first collision time of the two chains. (Note that some authors (e.g. Hsu and Sturm [10] and Kuwada [12], [13]) use the term ‘Markovian’ in place of ‘coadapted’, but this seems confusing when the *joint* process (X^c, Y^c) is not required to be Markov; we therefore prefer to reserve ‘Markovian’ for couplings where this stronger condition is satisfied.)

With some families of chains, it is possible to produce coadapted couplings with coupling times of the same order of magnitude as the mixing time. For example, for simple random walk on \mathbb{Z}_2^d , the mixing time is $\log d/4$ and the coupling of [3] takes time $\log d/2$. However, as pointed out by Griffeath [9], in general, a Markovian maximal coupling need not exist, and there are well-known chains for which coadapted couplings perform significantly worse than maximal. For example, the transposition shuffle on the symmetric group S_n and the Gibbs sampler on the n -simplex both have mixing times of order $O(n \log n)$, but it is impossible for coadapted couplings to do better than $O(n^2)$ and $O(n^2 \log n)$ respectively. It is therefore necessary in these instances to turn to noncoadapted couplings in order to obtain a good bound on the mixing time; see the preprints [2] and [16] for details.

Owing to the fact that maximal couplings are usually unintuitive and impractical to work with, whereas coadapted couplings are used extensively both in theoretical problems (e.g. bounding mixing times) and in practical applications (e.g. perfect simulation techniques), it is clearly of interest to obtain a better understanding of how good coadapted couplings can be for various processes. This is difficult to do in general, but an explicit understanding can be obtained in relatively simple cases: below we introduce one such process for which exact calculations are possible.

1.1. Random walk on K_n^d

Let K_n be the complete graph with n vertices (labelled $0, 1, \dots, n - 1$), and let K_n^d be the set of d -tuples of the form $(x(1), \dots, x(d))$ with $x(i) \in K_n, 1 \leq i \leq d$. The set K_n^d forms a group under coordinatewise addition modulo n , and $K_2^d \cong \mathbb{Z}_2^d$. A simple symmetric continuous-time random walk X on K_n^d may be defined by moving the i th coordinate of X to a different, uniformly chosen value at incident times of a unit-rate Poisson process, independently of all other coordinates. The unique equilibrium distribution of X is the uniform distribution on K_n^d .

Suppose now that we have a coupled pair (X^c, Y^c) , and let

$$U_t^c = \{1 \leq i \leq d: X_i^c(t) \neq Y_i^c(t)\} \quad \text{and} \quad M_t^c = \{1 \leq i \leq d: X_i^c(t) = Y_i^c(t)\}$$

respectively denote the sets of unmatched and matched coordinates at time $t \geq 0$. The coupling time τ^c clearly satisfies

$$\tau^c = \inf\{t \geq 0: X_s^c = Y_s^c \text{ for all } s \geq t\} = \inf\{t \geq 0: U_s^c = \emptyset \text{ for all } s \geq t\}.$$

In [3], an explicit, intuitive coupling strategy is described when $n = 2$, and is shown to yield the stochastically minimal coupling time of all coadapted couplings. This coupling strategy at time t depends only on the parity of $N_t = |U_t|$, and may be summarised as follows:

- matched coordinates are always made to move synchronously;
- if N is odd, all unmatched coordinates of X and Y are made to evolve *independently* until N becomes even;
- if N is even, unmatched coordinates are coupled in *pairs*—when an unmatched coordinate on X flips (thereby making a new match), a different unmatched coordinate on Y is flipped at the same instant (making a total of two new matches).

This motivates the following question: what is the optimal coadapted coupling when $n > 2$? Intuitively, we expect the optimal strategy of [3] to become inefficient as n gets large, since the rate at which unmatched coordinates can be made to agree using either ‘independent’ or ‘pairwise’ coupling (as described above) is proportional to N/n . In Section 2 we show how to describe the problem of finding an optimal coadapted coupling as an exercise in stochastic control, and solve this problem to once again obtain a stochastically minimal coupling time (in which new matches are made asymptotically at rate N as $n \rightarrow \infty$); the proof is deferred to Appendix A. In Section 3 we study the behaviour of this coupling as $d \rightarrow \infty$ for fixed n : we show that the time to stationarity exhibits a cutoff phenomenon which occurs strictly earlier than the coupling time, showing that our coupling is not maximal. As $n \rightarrow \infty$ however (with d fixed), the optimal coadapted coupling does tend to a maximal coupling. These are the results that we find most interesting since, to the best of the author’s knowledge, this is the first time that such behaviour (a natural sequence of optimal coadapted but submaximal couplings tending to a maximal coupling) has been observed. We conclude with some comments in Section 4 about a faster coupling for the random walks on K_n^d , and about optimal couplings for random walks on \mathbb{Z}_n^d .

2. Optimal coadapted coupling on K_n^d

In order to find the optimal coadapted coupling of X and Y , it is first necessary to be able to describe a general coupling strategy $c \in \mathcal{C}$; \mathcal{C} is the set of all coadapted couplings. To begin, let $\Lambda_{(i,k)(j,l)}$ ($1 \leq i, j \leq d$ and $0 \leq k, l \leq n - 1$) be independent Poisson processes on $[0, \infty)$, each of rate $(n - 1)^{-1}$. Now let $\{W_{(i,k)(j,l)}\}$ be a collection of piecewise-constant processes on $[0, 1]$, where $W_{(i,k)(j,l)}$ jumps at event times of $\Lambda_{(i,k)(j,l)}$ to values which are independent and identically distributed Uniform $[0, 1]$ (independently of all other Λ and W processes). We let $(\mathcal{F}_t)_{t \geq 0}$ be any filtration to which all of the processes $\{\Lambda_{(i,k)(j,l)}, W_{(i,k)(j,l)}\}$ are adapted.

The transitions of X^c and Y^c will be driven by the W processes, and controlled by a process $\{Q^c(t)\}_{t \geq 0}$ which is adapted to $(\mathcal{F}_t)_{t \geq 0}$, where

$$Q^c(t) = \{q_{(r,s)}^c(t): 0 \leq r, s \leq nd - 1\}$$

is an $(nd) \times (nd)$ doubly stochastic matrix.

A similar argument to that in [3] shows that a general coadapted coupling for X and Y may be defined as follows: if there is a jump in the process $W_{(i,k)(j,l)}$ at time $t \geq 0$, and the value of $W_{(i,k)(j,l)}(t)$ satisfies $W_{(i,k)(j,l)}(t) \leq q_{([i-1]n+k, [j-1]n+l)}^c(t)$, then set $X_t^c(i) = k$ and $Y_t^c(j) = l$. To ease notation, in the sequel we shall write $q_{(i,k)(j,l)}^c(t)$ instead of $q_{([i-1]n+k, [j-1]n+l)}^c(t)$, thus, $q_{(i,k)(j,l)}^c(t)$ is proportional to the instantaneous rate at which $(X_t^c(i), Y_t^c(j))$ jumps to (k, l) . (Note that this construction allows for the possibility of only one of X^c and Y^c actually changing its value at any given instant.)

Using the above construction, the rate at which $X^c(i)$ jumps from r to $s \neq r$ is equal to

$$\frac{1}{n-1} \sum_{j=1}^d \sum_{l=0}^{n-1} q_{(i,s)(j,l)}^c(t) = \frac{1}{n-1},$$

since the double sum is simply the sum of the $([i-1]n + s)$ th row of $Q^c(t)$, and, hence, equal to 1. From this, it follows directly that X^c and Y^c both have the correct marginal transition rates to be continuous-time simple random walks on K_n^d as described in Section 1.1, and are coadapted.

2.1. Stochastically optimal coupling

Our proposed optimal coupling $\hat{c} = \hat{c}_{n,d}$ once again depends upon the parity of $\hat{N}_t = |\hat{U}_t|$, the number of unmatched coordinates of $(\hat{X}, \hat{Y}) = (X^{\hat{c}}, Y^{\hat{c}})$ at time t . It now also depends upon how this number relates to the parameter n .

Definition 2.1. The matrix process \hat{Q} corresponding to the coupling \hat{c} has nonzero entries given by the following rules.

(C1) $\hat{q}_{(i,k)(i,k)}(t) = 1$ for all $i \in \hat{M}_{t-}$ and all $k = 0, \dots, n-1$.

(C2) If \hat{N}_{t-} is even, or $\hat{N}_{t-} \geq 2(n-1)/(n-2)$, then, for $i, j \in \hat{U}_{t-}$, $i \neq j$, and all $k, l \in \{0, \dots, n-1\}$,

$$(i) \hat{q}_{(i,k)(j,l)}(t) = \left(\frac{1}{\hat{N}_{t-} - 1} \right) \mathbf{1}_{[k=Y_{t-}(i), l=X_{t-}(j)]};$$

$$(ii) \hat{q}_{(i,k)(i,k)}(t) = \mathbf{1}_{[k \neq X_{t-}(i), k \neq Y_{t-}(i)]}.$$

(C3) If \hat{N}_{t-} is odd and $\hat{N}_{t-} < 2(n-1)/(n-2)$, then $\hat{q}_{(i,k)(i,k)}(t) = 1$ for all $i \in \hat{U}_{t-}$ and all $k = 0, \dots, n-1$.

Part (C1) of Definition 2.1 ensures that no matches are ever broken under \hat{c} . The final two parts define the strategy for making new matches. If \hat{N}_{t-} is even, or else sufficiently large, we will see that (C2)(i) implies that the rate at which two new matches are made is maximised; (C2)(ii) then maximises the rate at which a single new match is made, subject to the constraint imposed by (C2)(i). Finally, (C3) implies that if \hat{N}_{t-} is odd, with $\hat{N}_{t-} < 2(n-1)/(n-2)$, the coupling maximises the rate at which single matches are made. (Note that if $n = 2$, (C3) applies whenever \hat{N}_{t-} is odd; if $n = 3$ then it applies when $\hat{N}_{t-} \in \{1, 3\}$; while if $n \geq 4$, (C3) applies only when $\hat{N}_{t-} = 1$.)

Informally, when $n \geq 4$ and $\hat{N}_{t-} \geq 2$, \hat{c} couples \hat{X} and \hat{Y} as follows (see Figure 1). If an unmatched coordinate $\hat{X}(i)$ jumps to a different state k at time t (i.e. $k \neq \hat{X}_{t-}(i)$), then the following statements hold.

1. If (with probability $1/(n-1)$) we are lucky and $\hat{Y}_{t-}(i) = k$, choose another unmatched coordinate j uniformly at random, and set $\hat{Y}_t(j) = \hat{X}_{t-}(j)$. This decreases \hat{N} by 2.

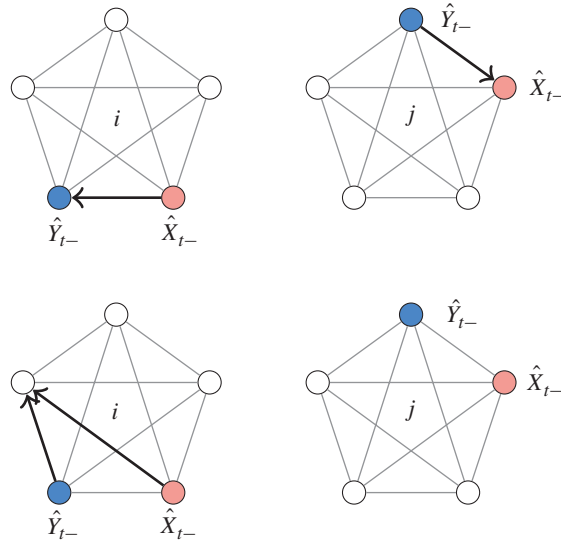


FIGURE 1: Optimal coadapted coupling strategy \hat{c}_n when $n = 5$: the values of \hat{X}_{t-} and \hat{Y}_{t-} on coordinates $i, j \in \hat{U}_{t-}$ are highlighted. Suppose that $\hat{X}(i)$ jumps to a new value at time t . *Top row*: the new value at time t happens to equal $\hat{Y}_{t-}(i)$ —in this case we select another unmatched coordinate (j) and move $\hat{Y}(j)$ to the value of $\hat{X}_{t-}(j)$, making a total of *two* new matches. *Bottom row*: the value of $\hat{X}_t(i)$ does not agree with that of $\hat{Y}_{t-}(i)$ —here $\hat{Y}(i)$ is also made to jump so that $\hat{X}_t(i) = \hat{Y}_t(i)$, making *one* new match.

2. If $\hat{Y}_{t-}(i) \neq k$, set $\hat{Y}_t(i) = k$. This decreases \hat{N} by 1.

Now define

$$\hat{v}(x, y, t) = \mathbb{P}(\hat{\tau} > t \mid \hat{X}_0 = x, \hat{Y}_0 = y)$$

to be the tail probability of the coupling time $\hat{\tau} = \hat{\tau}_{n,d}$ under \hat{c} . The main result of this paper is the following generalisation of Theorem 3.1 of [3].

Theorem 2.1. For any states $x, y \in K_n^d$ and time $t \geq 0$,

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

In other words, $\hat{\tau}$ is the stochastic minimum of all coadapted coupling times for the pair (X, Y) . The proof of this theorem can be found in Appendix A.

3. Limiting behaviour

Now that we have established the existence of an optimal coadapted coupling \hat{c} , a natural question to ask is whether or not this coupling is also maximal. (This was answered in the negative when $n = 2$ in [3].) Denote by π_n^d the uniform distribution on K_n^d (recall that π_n^d is the equilibrium distribution of X), and by $\tau_{n,d}^*$ the maximal coupling time for the pair (X, Y) , where $X_0 = \mathbf{0}$ and $Y_0 \sim \pi_n^d$. The following result is a simple generalisation of Proposition 1 of [6].

Lemma 3.1. Let

$$T_{n,d} = \frac{1}{2} \left(\frac{n-1}{n} \right) \log d.$$

Then, as $d \rightarrow \infty$, for all $\theta \in \mathbb{R}$,

$$\|\mathcal{L}(X_{T_{n,d}+\theta}) - \pi_n^d\|_{TV} = 2\Phi\left(\frac{\sqrt{n-1}}{2}e^{-n\theta/(n-1)}\right) - 1 + o(1), \tag{3.1}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

This shows that the distance between $\mathcal{L}(X)$ (the law of X) and π_n^d exhibits a *cutoff phenomenon* [1], [4], [5] at time $T_{n,d}$, the window size being $O(1)$. Thus, $\mathbb{E}[\tau_{n,d}^*] \sim T_{n,d}$.

On the other hand, we can bound $\mathbb{E}[\hat{\tau}_{n,d}]$ as follows. As in Appendix A, we write $\hat{\lambda}_t(k, k + s)$ for the rate (according to $\hat{Q}(t)$) at which \hat{N}_{t-} jumps from k to $k + s$. Under \hat{c} , \hat{N} is a decreasing process, with jumps being of size -1 or -2 ; the total rate at which \hat{N} jumps is equal (by (C2)) to

$$\hat{\lambda}_t(k, k - 2) + \hat{\lambda}_t(k, k - 1) = \frac{k}{n - 1} + \frac{k(n - 2)}{n - 1} = k.$$

Now let M be a process that takes only steps of size $k \rightarrow k - 2$ at rate k , and let τ_M be the time taken for M to be absorbed at 0. If $\hat{N}_0 = M_0 = 2m$ then $\mathbb{E}[\tau_M \mid M_0 = 2m] \leq \mathbb{E}[\hat{\tau}_{n,d} \mid \hat{N} = 2m]$, thanks to Lemma A.1 of Appendix A. Furthermore,

$$\mathbb{E}[\tau_M \mid M_0 = 2m] = \sum_{i=1}^m (2i)^{-1} \sim \frac{1}{2} \log m.$$

Since $\hat{N}_0 = |\hat{X}_0 - \hat{Y}_0| \sim \text{Bin}(d, (n - 1)/n)$, Chebychev’s inequality implies that \hat{N}_0 (and, thus, M_0) is concentrated around $d(n - 1)/n$, and so

$$\mathbb{E}[\tau_M] \sim \frac{1}{2} \log d \quad \text{as } d \rightarrow \infty.$$

Therefore,

$$\frac{\mathbb{E}[\hat{\tau}_{n,d}]}{\mathbb{E}[\tau_{n,d}^*]} \geq \frac{\mathbb{E}[\tau_M]}{\mathbb{E}[\tau_{n,d}^*]} \sim \frac{n}{n - 1},$$

and so the optimal coadapted coupling is not maximal for any fixed n .

Let us now consider what happens if we let $n \rightarrow \infty$ while keeping d fixed. Suppose that the n points of K_n are equally spaced on the unit interval $[0, 1)$, at locations $\{0, 1/n, 2/n, \dots\}$. As $n \rightarrow \infty$, the random walk X on K_n^d , with $X_0 = \mathbf{0}$, converges in distribution to the random walk \tilde{X} on $[0, 1)^d$ for which each coordinate jumps, at incident times of an independent unit-rate Poisson process, to a new location distributed uniformly on $[0, 1)$. The equilibrium distribution of \tilde{X} is of course $\pi_\infty^d = \text{Uniform}[0, 1)^{\otimes d}$.

Lemma 3.2. *For d fixed, as $n \rightarrow \infty$, the optimal coadapted coupling $\hat{c}_{n,d}$ of Section 2 tends to a maximal coupling.*

Proof. Let A_0 be the set of points in $[0, 1)^d$ which have at least one coordinate equal to 0. Then, by the definition of the total variation distance,

$$\begin{aligned} \|\mathcal{L}(\tilde{X}_t) - \pi_\infty^d\|_{TV} &= \sup_{A \subset [0,1)^d} (\mathbb{P}(\tilde{X}_t \in A) - \pi_\infty^d(A)) \\ &= \mathbb{P}(\tilde{X}_t \in A_0) \\ &= 1 - \mathbb{P}(\text{all coordinates of } \tilde{X} \text{ have jumped by time } t) \\ &= 1 - (1 - e^{-t})^d. \end{aligned}$$

Now consider the optimal coadapted coupling strategy $\hat{c}_{n,d}$ as $n \rightarrow \infty$. Let $\tilde{Y}_0 \sim \pi_\infty^d$, and note that $|X_0 - Y_0| = d$ almost surely. From (A.6) we see that $\lambda_t^{\hat{c}_{n,d}}(m, m + s)$, the rate under $\hat{c}_{n,d}$ at which \hat{N}_{t-} jumps from m to $m + s$, satisfies (when $n \geq 4$)

$$\lambda_t^{\hat{c}_{n,d}}(m, m + s) = \begin{cases} \frac{m(n-2)}{n-1} & \text{if } s = -1, \\ \frac{m}{n-1} & \text{if } s = -2, \end{cases} \quad \text{when } m \geq 2,$$

$$\lambda_t^{\hat{c}_{n,d}}(1, 0) = \frac{n}{n-1},$$

with all other rates equal to 0. Let B be a process that takes steps of size $m \rightarrow m - 1$ at rate $m(n - 2)/(n - 1)$. Since $\hat{N}_0 = d$, the time taken for \hat{N} to be absorbed at 0 (equal to $\hat{\tau}_{n,d}$) is bounded above by the time taken for B to hit 0 from d (again thanks to Lemma A.1). But this time, τ_B is simply equal to the maximum of d independent $\text{Exp}((n - 2)/(n - 1))$ random variables, and so

$$\mathbb{P}(\tau_B > t \mid B_0 = d) = 1 - (1 - e^{-(n-2)t/(n-1)})^d.$$

Finally, using the coupling inequality to lower bound the tail distribution of $\hat{\tau}_{n,d}$, we see that

$$1 - (1 - e^{-t})^d \leq \mathbb{P}(\hat{\tau}_{n,d} > t) \leq 1 - (1 - e^{-(n-2)t/(n-1)})^d.$$

As $n \rightarrow \infty$, it follows that the coupling time $\hat{\tau}_{\infty,d}$ achieves equality in the coupling inequality, showing $\hat{c}_{\infty,d}$ to be a maximal coupling.

Furthermore, if we now let $d \rightarrow \infty$, the distance between $\mathcal{L}(\tilde{X}_t)$ and π_∞^d again obeys a cutoff phenomenon, with cutoff time equal to $\log d$. (This may appear surprising, since $T_{n,d} \rightarrow \frac{1}{2} \log d$ as $n \rightarrow \infty$ in Lemma 3.1. However, note that the expression on the right-hand side of (3.1) tends to 1 for all $\theta \in \mathbb{R}$ as $n \rightarrow \infty$, showing that $\frac{1}{2} \log d$ is not the cutoff time for the limiting process.)

4. Final comments

4.1. A faster coupling for random walks on K_n^d

We have demonstrated the existence of (and explicitly described) a maximal coadapted coupling \hat{c} for a pair of continuous-time symmetric random walks on K_n^d ; this has been seen to be submaximal for any fixed n , but to converge naturally to a maximal coupling as $n \rightarrow \infty$. The reason that \hat{c} is submaximal can be found in (C2)(i), where the strategy for making two new matches is described. Recall that if $\hat{X}(i)$ jumps to state k at time t , where $\hat{Y}_{t-} = k$, then we choose a different coordinate $j \in U_t^{\hat{c}}$ and set $\hat{Y}_t(j) = \hat{X}_{t-}(j)$. In order for the coupling to be coadapted, j must be chosen in a way that depends only on the information in \mathcal{F}_t . Matthews [15] showed how to use information about the future of one process to more efficiently pair coordinates when $n = 2$, producing a noncoadapted (and near-maximal) coupling. This, combined with the result of Lemma 3.2, motivates the following suggestion for a faster coupling on K_n^d .

Suppose that $X_0 = \mathbf{0}$ and $Y_0 = \mathbf{1}$. (Vectors of 0s and 1s respectively—since K_n is a complete graph, the coupling time is unaffected by coordinatewise relabelling of the nonzero states of Y_0 .) Run X until time $\tau = \min\{\tau_1, \tau_2\}$, where

$$\tau_1 = \inf\{t : \text{all coordinates have visited the set } \{2, 3, \dots, n - 1\} \text{ by time } t\},$$

$$\tau_2 = \inf\{t : \#\{i : X_t(i) = 0\} = \#\{i : X_t(i) = 1\}\},$$

and where ‘#’ denotes cardinality. Now, if $\tau = \tau_1$ then each coordinate has visited a state not equal to 0 or 1 by time τ , thus, Y can be evolved over the period $[0, \tau]$ with every coordinate being independently coupled using the idea depicted in the bottom row of Figure 1. If $\tau = \tau_2$ however then let

$$A_\tau^0 = \{i: X_\tau(i) = 0\}, \quad A_\tau^1 = \{i: X_\tau(i) = 1\}, \quad \text{and} \quad A_\tau^+ = \{i: X_\tau(i) \geq 2\}.$$

On the event $\{\tau = \tau_2\}$, $\#A_\tau^0 = \#A_\tau^1$, and so we may choose some arbitrary bijection $\rho: A_\tau^0 \rightarrow A_\tau^1$. Now run Y over $[0, \tau]$ as follows: if $i \in A_\tau^0$ then let $Y(i)$ copy the path of $X(\rho(i))$, but with jumps $k \rightarrow 0$ replaced by $k \rightarrow 1$ and vice versa; if $i \in A_\tau^1$ then let $Y(i)$ similarly copy the path of $X(\rho^{-1}(i))$; coordinates $i \in A_\tau^+$ are coupled independently, as when $\tau = \tau_1$.

It is simple to check that this produces a coupling of X and Y , with coupling time τ . Note that, when $n = 2$, $\tau = \tau_2$ and we obtain the coupling of [15]. Moreover, $\mathbb{P}(\tau = \tau_1) \rightarrow 1$ as $n \rightarrow \infty$, and in the limit we find the maximal coupling \hat{c}_∞ from the proof of Lemma 3.2.

4.2. Random walk on \mathbb{Z}_n^d

The other natural generalisation of the random walk on \mathbb{Z}_2^d is of course to the random walk on \mathbb{Z}_n^d for $n \geq 2$. This is messier to analyse using the stochastic control arguments of this paper, since $|X_t(i) - Y_t(i)|$ can now be greater than 1; although the optimal coadapted coupling strategy (if one exists) will again be invariant under coordinate permutation, the value function \hat{v} will now depend not only upon N_t , but upon the entire vector $(\zeta_t(0), \zeta_t(1), \dots, \zeta_t(n/2))$, where

$$\zeta_t(k) = \#\{i: |X_t(i) - Y_t(i)| = k\}.$$

It seems obvious that if $d = 1$, the best coadapted coupling policy is for Y to jump at the same time as X , but in the opposite direction around the circle \mathbb{Z}_d , with the exception that if $|X_t - Y_t| = 1$ then both processes should evolve independently until this equality is broken. When $d \geq 2$, heuristics point towards the best strategy being to couple each coordinate independently (using reflection around the circle), *unless* $\zeta_t(1)$ happens to be even; in this case, we should *pair* those coordinates which are 1 away from coupling, and try to make two new matches at once (in the style depicted in the top row of Figure 1).

Finally, note that if we appropriately rescale these processes then the coordinates of X converge as $n \rightarrow \infty$ to a set of d independent Brownian motions on the unit circle. Coupling each coordinate independently by reflection—the limit of the suggested optimal coadapted coupling—should once again be *maximal* for the limiting process.

Appendix A

Proof of Theorem 2.1. From Definition 2.1, it is evident that \hat{c} is invariant under coordinate permutation, and that $\hat{v}(x, y, t)$ depends only on (x, y) through $|x - y|$ (the Hamming distance between x and y), and so we shall write

$$\hat{v}(m, t) = \mathbb{P}(\hat{\tau} > t \mid \hat{N}_0 = m),$$

with the convention that $\hat{v}(m, t) = 0$ for $m \leq 0$.

Following [3], we shall write $\lambda_t^c(m, m + s)$ for the rate (according to $Q^c(t)$) at which N_t^c jumps from m to $m + s$ for $s \in \{-2, \dots, 2\}$. For example,

$$(n - 1)\lambda_t^c(m, m - 2) = \sum_{\substack{i, j \in U_{t-}^c \\ i \neq j}} q_{(i, Y_{t-}^c(i))(j, X_{t-}^c(j))}^c(t) \tag{A.1}$$

and

$$\begin{aligned}
 & (n - 1)\lambda_t^c(m, m - 1) \\
 &= \sum_{i \in U_{t-}^c} \sum_{0 \leq k \leq n-1} q_{(i,k)(i,k)}^c(t) \\
 &+ \sum_{\substack{i,j \in U_{t-}^c \\ i \neq j}} \sum_{0 \leq l, k \leq n-1} q_{(i,k)(j,l)}^c(t) (\mathbf{1}_{[k=Y_{t-}^c(i), l \neq X_{t-}^c(j)]} + \mathbf{1}_{[k \neq Y_{t-}^c(i), l = X_{t-}^c(j)]}) \\
 &+ \sum_{\substack{i \in U_{t-}^c \\ j \in M_{t-}^c}} q_{(i, Y_{t-}^c(i))(j, Y_{t-}^c(j))}^c(t) + \sum_{\substack{i \in M_{t-}^c \\ j \in U_{t-}^c}} q_{(i, X_{t-}^c(i))(j, X_{t-}^c(j))}^c(t). \tag{A.2}
 \end{aligned}$$

Similar decompositions may be written down for $\lambda_t^c(m, m + 1)$ and $\lambda_t^c(m, m + 2)$, but we will have no need of them in the sequel.

The expression for $\lambda_t^c(m, m - 2)$ in (A.1) is easy to understand: N^c decreases by 2 if and only if different unmatched coordinates on X^c and Y^c flip at the same instant, with each flip making one new match.

The expression for $\lambda_t^c(m, m - 1)$ comprises four sums however, and so requires a little more explanation. The first sum in (A.2) gives the rate at which the same unmatched coordinate flips on both X^c and Y^c to the same value, making one new match. The second term is the rate at which an unmatched coordinate on one process flips to make a new match, while a different unmatched coordinate on the other process jumps (or possibly stays at its current value) *without* making another new match. Finally, the third and fourth sums in (A.2) give the rate at which an unmatched coordinate on one process flips and makes a new match, while on the other process a matched coordinate is selected and made to stay at its current value.

Using the constraints on the row sums of Q_t^c , it is possible to bound the terms in (A.1) and (A.2) as

$$(n - 1)\lambda_t^c(m, m - 2) = \sum_{\substack{i,j \in U_{t-}^c \\ i \neq j}} q_{(i, Y_{t-}^c(i))(j, X_{t-}^c(j))}^c(t) \leq |U_{t-}^c| = m \tag{A.3}$$

and, similarly,

$$(n - 1)\lambda_t^c(m, m - 1) \leq nm.$$

Moreover, since Q_t^c is doubly stochastic,

$$\begin{aligned}
 & (n - 1)\lambda_t^c(m, m - 1) + 2(n - 1)\lambda_t^c(m, m - 2) \\
 &\leq \sum_{i \in U_{t-}^c} \left(\sum_{j=1}^d \sum_{0 \leq k, l < n} q_{(i,k)(j,l)}^c(t) \mathbf{1}_{[k \neq X_{t-}^c(i), k \neq Y_{t-}^c(i)]} \right) \\
 &+ \sum_{i \in U_{t-}^c} \left(\sum_{j=1}^d \sum_{l=0}^{n-1} q_{(i,k)(j,l)}^c(t) (\mathbf{1}_{[j \neq i]} + \mathbf{1}_{[j=i, l = Y_{t-}^c(i)]}) \right) \\
 &+ \sum_{j \in U_{t-}^c} \left(\sum_{i=1}^d \sum_{k=0}^{n-1} q_{(i,k)(j,l)}^c(t) (\mathbf{1}_{[i \neq j]} + \mathbf{1}_{[i=j, k = X_{t-}^c(j)]}) \right) \\
 &\leq m(n - 2) + m + m \\
 &= nm.
 \end{aligned}$$

Denote by L_n the set of nonnegative λ satisfying the key linear constraint

$$(n - 1)\lambda(m, m - 1) + 2(n - 1)\lambda(m, m - 2) \leq nm. \tag{A.4}$$

When $n = 2$, this reduces to the constraint of [3]: $\lambda(m, m - 1) + 2\lambda(m, m - 2) \leq 2m$.

Proposition A.1. *Under \hat{c} , the following set of equations hold:*

$$\hat{\lambda}_t(m, m + 1) = \hat{\lambda}_t(m, m + 2) = 0; \tag{A.5}$$

if m is even, or if $m \geq 2(n - 1)/(n - 2)$, then

$$(n - 1)\hat{\lambda}_t(m, m - 2) = m \quad \text{and} \quad (n - 1)\hat{\lambda}_t(m, m - 1) = (n - 2)m; \tag{A.6}$$

while if m is odd and $m < 2(n - 1)/(n - 2)$ then

$$\hat{\lambda}_t(m, m - 2) = 0 \quad \text{and} \quad (n - 1)\hat{\lambda}_t(m, m - 1) = nm. \tag{A.7}$$

Proof. Equation (A.5) is an immediate consequence of (C1), which implies that no matches are ever broken under \hat{c} . When $\hat{N}_{t-} = m$ is even or satisfies $m \geq 2(n - 1)/(n - 2)$, it follows from (C2)(i) and (A.1) that

$$(n - 1)\hat{\lambda}_t(m, m - 2) = \sum_{\substack{i, j \in \hat{U}_{t-} \\ i \neq j}} \hat{q}_{(i, \hat{Y}_{t-(i)})(j, \hat{X}_{t-(j)}}(t) = \sum_{\substack{i, j \in \hat{U}_{t-} \\ i \neq j}} \frac{1}{m - 1} = m.$$

Finally, (C1) and (C2) imply that, under \hat{c} , the only nonzero term in (A.2) is the first sum, and so

$$(n - 1)\hat{\lambda}_t(m, m - 1) = \sum_{i \in \hat{U}_{t-}} \sum_{k=0}^{n-1} \hat{q}_{(i, k)(i, k)}(t).$$

Substituting the values of $\hat{q}_{(i, k)(i, k)}(t)$ from (C2) and (C3) completes the proof.

It follows that the upper bound of (A.4) is always attained under \hat{c} . Although the framework laid out in Section 2 for describing a general coupling $c \in \mathcal{C}$ differs from the setup in [3], we can immediately obtain the result of that paper.

Corollary A.1. ([3].) *Theorem 2.1 holds when $n = 2$.*

Proof. When $n = 2$, Proposition A.1 shows that, for all $m \in \mathbb{N}$,

$$\begin{aligned} \hat{\lambda}_t(m, m + 1) &= \hat{\lambda}_t(m, m + 2) = 0, \\ \hat{\lambda}_t(2m, 2m - 2) &= 2m, \quad \text{and} \quad \hat{\lambda}_t(2m - 1, 2m - 2) = 2(2m - 1). \end{aligned}$$

The optimality of $\hat{c} = \hat{c}_{2,d}$ now follows from the proof of Theorem 3.1 of [3].

For a strategy $c \in \mathcal{C}$, define the process S_t^c by

$$S_t^c = \hat{v}(X_t^c, Y_t^c, T - t),$$

where $T > 0$ is some fixed time. This is the conditional probability of X and Y not having coupled by time T , when strategy c has been followed over the interval $[0, t]$ and \hat{c} has then

been used from time t onwards. As in [3], the optimality of \hat{c} will follow by Bellman’s principle [11] if it can be shown that $S_{t \wedge \tau}^c$ is a submartingale for all $c \in \mathcal{C}$ (where $s \wedge t = \min\{s, t\}$).

Once again, S^c satisfies

$$dS_t^c = dZ_t^c + \left(\mathcal{A}_t^c \hat{v} - \frac{\partial \hat{v}}{\partial t} \right) dt,$$

where Z_t^c is a martingale, and \mathcal{A}_t^c is the ‘generator’ corresponding to the matrix $Q^c(t)$. Since \hat{v} is invariant under coordinate permutation, and the Poisson processes $\Lambda_{(i,k)(j,l)}$ are independent,

$$\mathcal{A}_t^c \hat{v}(m, t) = \sum_{s=-2}^2 \lambda_t^c(m, m + s) [\hat{v}(m + s, t) - \hat{v}(m, t)].$$

To prove that S^c is a submartingale, it suffices to show that $\mathcal{A}_t^c \hat{v}$ is minimised by setting $c = \hat{c}$. Thus we seek to maximise over $\lambda \in L_n$, for all $m \geq 0$ and all $t \geq 0$,

$$\sum_{s=-2}^2 \lambda(m, m + s) [\hat{v}(m, t) - \hat{v}(m + s, t)]. \tag{A.8}$$

Our first step is to simplify this maximisation problem by showing that $\hat{v}(m, t)$ is strictly increasing in m . When $n = 2$, this result follows trivially from the explicit representation of $\hat{\tau}$ given in [3]. For $n > 2$ however, the result is less obvious and requires a formal proof.

Lemma A.1. *The tail probability $\hat{v}(m, t)$ is strictly increasing in m .*

Proof. We detail here the proof for the case when $n \geq 4$ (for which case (C2) of Definition 2.1 applies for all $m > 1$): the proof when $n = 3$ is similar, using the remark following (A.10) below whenever (C3) applies (i.e. when $m \in \{1, 3\}$).

We begin by considering $\hat{v}(1, t)$. By (A.5) and (A.7), it directly follows that, for all values of n ,

$$\hat{v}(1, t) = \exp\left(-\frac{nt}{n-1}\right). \tag{A.9}$$

Now consider (for $m > 1$) that part of the coupling \hat{c} described in (C2). From (A.6), the total rate at which \hat{N}_{t-} can change under (C2) is given by

$$\hat{\lambda}_t(m, m - 2) + \hat{\lambda}_t(m, m - 1) = \frac{m}{n-1} + \frac{(n-2)m}{n-1} = m.$$

Using this, along with (A.5) and (A.6), we obtain, for $m > 1$,

$$\begin{aligned} \hat{v}(m, t) &= e^{-mt} \\ &+ \int_0^t m e^{-mu} [\hat{\lambda}_t(m, m - 2) \hat{v}(m - 2, t - u) + \hat{\lambda}_t(m, m - 1) \hat{v}(m - 1, t - u)] du \\ &= e^{-mt} + \int_0^t \frac{m e^{-mu}}{n-1} [m \hat{v}(m - 2, t - u) + m(n-2) \hat{v}(m - 1, t - u)] du. \end{aligned} \tag{A.10}$$

(A similar expression can be obtained for $\hat{v}(m, t)$ under (C3), noting that the total rate at which \hat{N}_{t-} can change in this case is $nm/(n-1)$.)

Define $\hat{V}^m(\alpha)$ to be the Laplace transform of $\hat{v}(m, \cdot)$:

$$\hat{V}^m(\alpha) = \int_0^\infty e^{-\alpha t} \hat{v}(m, t) dt.$$

It then follows from (A.10) that, for $m > 1$,

$$\hat{V}^m(\alpha) = \frac{1}{m + \alpha} + \frac{1}{(n - 1)(m + \alpha)} (m \hat{V}^{m-2}(\alpha) + m(n - 2) \hat{V}^{m-1}(\alpha)), \tag{A.11}$$

and so (rearranging)

$$(m + \alpha)(n - 1) \hat{V}^m(\alpha) = (n - 1) + m \hat{V}^{m-2}(\alpha) + m(n - 2) \hat{V}^{m-1}(\alpha). \tag{A.12}$$

We need to show that $r(m, t) \geq 0$ for all $m \geq 1$ and $t \geq 0$, where

$$r(m, t) = \hat{v}(m, t) - \hat{v}(m - 1, t).$$

By the Bernstein–Widder theorem (see [8, Theorem 1a, Chapter XIII.4]), this is equivalent to showing that $R^m(\alpha)$ is totally monotone (TM), where $R^m(\alpha)$ is defined for $\alpha \geq 0$ by

$$R^m(\alpha) = \int_0^\infty e^{-\alpha t} r(m, t) dt.$$

We begin by showing that this is true when $m = 1$ and $m = 2$, and then use induction. From (A.9) we see that

$$R^1(\alpha) = \hat{V}^1(\alpha) = \frac{n - 1}{n + \alpha(n - 1)}; \tag{A.13}$$

therefore, $R^1(\alpha)$ is TM.

Furthermore, using (A.11), we obtain

$$R^2(\alpha) = \frac{1}{2 + \alpha} + \frac{2(n - 2)}{(2 + \alpha)(n - 1)} \hat{V}^1(\alpha) - \hat{V}^1(\alpha) = \frac{n - 2}{(2 + \alpha)(n + \alpha(n - 1))}. \tag{A.14}$$

Since the product of two TM functions is itself TM [8], it follows that $R^2(\alpha)$ is also TM, as desired.

Now suppose that we have already shown $R^m(\alpha)$ and $R^{m-1}(\alpha)$ are TM for some $m \geq 2$. Subtracting $(m + \alpha)(n - 1) \hat{V}^{m-1}(\alpha)$ from both sides of (A.12) yields

$$(m + \alpha)(n - 1) R^m(\alpha) = (n - 1) [1 - \alpha \hat{V}^{m-1}(\alpha)] - m R^{m-1}(\alpha). \tag{A.15}$$

Substituting $m + 1$ for m in this expression we obtain

$$(m + 1 + \alpha)(n - 1) R^{m+1}(\alpha) = (n - 1) [1 - \alpha \hat{V}^m(\alpha)] - (m + 1) R^m(\alpha), \tag{A.16}$$

and then subtracting (A.15) from (A.16) yields

$$\begin{aligned} (m + 1 + \alpha)(n - 1) R^{m+1}(\alpha) &= (n - 1) \alpha [\hat{V}^{m-1}(\alpha) - \hat{V}^m(\alpha)] + m R^{m-1}(\alpha) \\ &\quad - [(m + 1) - (m + \alpha)(n - 1)] R^m(\alpha) \\ &= m R^{m-1}(\alpha) + [m(n - 2) - 1] R^m(\alpha). \end{aligned} \tag{A.17}$$

Since $m \geq 2$ and $n \geq 4$, we see that $m(n - 2) - 1 > 0$. Hence, by our induction hypothesis, $R^{m+1}(\alpha)$ can be expressed as the sum of two TM functions, and so is itself TM. This completes the proof.

Thus, $\hat{v}(m, t) - \hat{v}(m - 1, t) \geq 0$ for all $m \geq 0$ and $t \geq 0$. It follows that the terms appearing on the right-hand-side of (A.8) are nonpositive if and only if s is nonnegative. Hence, we must set

$$\lambda(m, m + 1) = \lambda(m, m + 2) = 0$$

in order to achieve the maximum in (A.8). It therefore now suffices to maximise

$$\begin{aligned} &\lambda(m, m - 2)[\hat{v}(m, t) - \hat{v}(m - 2, t)] + \lambda(m, m - 1)[\hat{v}(m, t) - \hat{v}(m - 1, t)] \\ &= \lambda(m, m - 2)[r(m, t) + r(m - 1, t)] + \lambda(m, m - 1)r(m, t) \end{aligned}$$

subject to the constraint from (A.4):

$$(n - 1)\lambda(m, m - 1) + 2(n - 1)\lambda(m, m - 2) \leq nm.$$

Putting these together we see that we need to maximise (for $m \geq 2$)

$$\lambda(m, m - 2)[r(m - 1, t) - r(m, t)].$$

Theorem A.1. For all $n \geq 4$, $m \geq 2$, and $t \geq 0$, $r(m - 1, t) \geq r(m, t)$.

Proof. In a similar fashion to the proof of Lemma A.1, we show positivity of $r(m - 1, t) - r(m, t)$ by showing that $R^{m-1}(\alpha) - R^m(\alpha)$ is TM, again using induction in m . From (A.13) and (A.14), we see that

$$R^1(\alpha) - R^2(\alpha) = \frac{1}{2 + \alpha}$$

and so is TM. Using (A.17), it can be deduced that

$$R^2(\alpha) - R^3(\alpha) = \frac{n - 4}{(2 + \alpha)(3 + \alpha)(n - 1)}.$$

Since $n \geq 4$, this difference is also TM.

Now assume that $R^{m-1}(\alpha) - R^m(\alpha)$ is TM for some $m \geq 3$. Substituting $m - 1$ for m into (A.17) yields

$$(m + \alpha)(n - 1)R^m(\alpha) = (m - 1)R^{m-2}(\alpha) + [(m - 1)(n - 2) - 1]R^{m-1}(\alpha), \quad (\text{A.18})$$

and subtracting (A.17) from (A.18) shows that

$$\begin{aligned} (m + 1 + \alpha)(n - 1)[R^m(\alpha) - R^{m+1}(\alpha)] &= ((m - 1)(n - 2) - 2)[R^{m-1}(\alpha) - R^m(\alpha)] \\ &\quad + (m - 1)[R^{m-2}(\alpha) - R^{m-1}(\alpha)]. \end{aligned}$$

Finally, since $m \geq 3$ and $n \geq 4$, $(m - 1)(n - 2) - 2 \geq 2$ and so it follows from our induction hypothesis that $R^m(\alpha) - R^{m+1}(\alpha)$ is the sum of two TM functions, and, hence, is itself TM, as claimed.

This result allows us to finally complete the proof of Theorem 2.1 when $n \geq 4$. (The proof for $n = 3$ follows a similar line of argument, using an amended version of Theorem A.1.) When $n \geq 4$, Theorem A.1 and the preceding discussion show that the optimal strategy must maximise $\lambda(m, m - 2)$ for $m \geq 2$, and $\lambda(1, 0)$. Using the bounds in (A.3) and (A.4), it follows that this is equivalent to requiring $\lambda(1, 0) = n/(n - 1)$ and, for $m \geq 2$,

$$(n - 1)\lambda(m, m - 2) = m \quad \text{and} \quad (n - 1)\lambda(m, m - 1) = (n - 2)m.$$

But, by Proposition A.1, this is in complete agreement with the rates $\hat{\lambda}_t$ arising from using our candidate optimal strategy, \hat{c} . Thus, \hat{c} is truly an optimal coadapted coupling, as claimed.

References

- [1] ALDOUS, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on Probability XVII* (Lecture Notes Math. **986**), Springer, Berlin, pp. 243–297.
- [2] BURTON, R. AND KOVCHegov, Y. (2011). Mixing times via super-fast coupling. *Tech. Rep.*, Oregon State University.
- [3] CONNOR, S. B. AND JACKA, S. (2008). Optimal co-adapted coupling for the symmetric random walk on the hypercube. *J. Appl. Prob.* **45**, 703–713.
- [4] DIACONIS, P. (1988). *Group Representations in Probability and Statistics* (Inst. Math. Statist. Lecture Notes Monogr. Ser. **11**). Institute of Mathematical Statistics, Hayward, CA.
- [5] DIACONIS, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. USA* **93**, 1659–1664.
- [6] DIACONIS, P., GRAHAM, R. L. AND MORRISON, J. A. (1990). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms* **1**, 51–72.
- [7] DOEBLIN, W. (1938). Exposé de la theorie des chaines constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkanique* **2**, 77–105.
- [8] FELLER, V. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn. John Wiley, New York.
- [9] GRIFFEATH, D. (1975). A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitsth.* **31**, 95–106.
- [10] HSU, E. P. AND STURM, K.-T. (2003). Maximal coupling of Euclidean Brownian motions. *Commun. Math. Statist.* **1**, 93–104.
- [11] KRYLOV, N. V. (1980). *Controlled Diffusion Processes* (Appl. Math. **14**). Springer, New York.
- [12] KUWADA, K. (2007). On uniqueness of maximal coupling for diffusion processes with a reflection. *J. Theoret. Prob.* **20**, 935–957.
- [13] KUWADA, K. (2009). Characterization of maximal Markovian couplings for diffusion processes. *Electron. J. Prob.* **14**, 633–662.
- [14] LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover Publications, Mineola, NY.
- [15] MATTHEWS, P. (1987). Mixing rates for a random walk on the cube. *SIAM J. Algebraic Discrete Methods* **8**, 746–752.
- [16] SMITH, A. (2011). A Gibbs sampler on the N-simplex. To appear in *Ann. Appl. Prob.*
- [17] SVERCHKOV, M. Y. AND SMIRNOV, S. N. (1990). Maximal coupling of D-valued processes. *Soviet Math. Dokl.* **41**, 352–354.