Mixing Time Bounds for Asymmetric Random Walks on Hypercubes, Cycles, and Tori Using Coupling

Mrudula A Mahindrakar, Hrushikesh A Kant, and Avhishek Chatterjee
Department of Electrical Engineering, Indian Institute of Technology Madras, India
Email: ee25s011@smail.iitm.ac.in, ee22b108@smail.iitm.ac.in, avhishek@ee.iitm.ac.in

***Abstract*—Mixing time bounds for Markov chains play a central role in characterizing the sample complexity of learning and inference from correlated data. While the mixing behavior of symmetric random walks on standard graph structures such as cycles, tori, and hypercubes is well understood, the impact of transition asymmetry remains less explored. In this work, we study the mixing times of lazy, asymmetric random walks on cycles, tori, and hypercubes, motivated by their relevance in practical applications.**

For the n-cycle, we develop a novel coupling construction that yields an order-wise tight upper bound $O\left(\frac{n^2}{p+q}\right)$, explicitly capturing the dependence on asymmetric transition probabilities p and q. Numerical results indicate that this dependence is highly accurate. Building on this result, we derive corresponding bounds for d-dimensional tori.

For asymmetric random walks on n-dimensional hypercubes, motivated by applications, we consider the problem of estimating expectations of functions that depend only on a subset $\Delta \ll n$ of coordinates. We show that the effective sample complexity improves to $O(n \log \Delta)$, compared to $O(n \log n)$ for the full chain. In all cases, our bounds recover the tightest known results for symmetric walks as special cases.

***Index Terms*—Asymmetric Random Walks; Sample Complexity; Mixing Time.**

I. Introduction

Learning from correlated samples, particularly Markov samples, is central to numerous applications, including large language models (LLMs) [1], [2] and learning in social and biological networks [3]–[5]. Sample-complexity analyses in these settings rely on Markov concentration inequalities, which also arise in quantum fault tolerance [6].

A widely used Markov concentration inequality [7, Cor. 2.11] states that for a c-Lipschitz function f (with respect to the Hamming distance [7], [8]) defined on the sample path of a Markov chain $X = X_k : 1 \leq k \leq N$ over a discrete state space,

$$\mathbb{P}(|f(X) - \mathbb{E}[f(X)]| \geq \lambda) \leq 2 \exp\left(-\frac{2\lambda^2}{9c^2 N t_{mix;}}\right), \quad (1)$$

where t_{mix} is the $\frac{1}{4}$-mixing time of the chain. This highlights the importance of mixing-time estimates for obtaining meaningful concentration guarantees. Accordingly, we establish tight mixing-time bounds for several useful Markov chains.

In the context of LLMs, the latent and unobservable state evolution can be well approximated by random walks on graphs, wherein the observables correspond to the tokens emitted during state transitions [1], [2]. Such random walks are inherently asymmetric due to the structure of natural languages.

Similarly, in social networks and biological networks, although the underlying Markov chains are typically more intricate than simple random walks [3], asymmetry remains an intrinsic characteristic of these dynamics. The same holds for relevant Markov processes in quantum fault-tolerance [6].

Motivated by these considerations, we investigate the mixing times of asymmetric random walks on cycles, tori, and hypercubes. Although these graph structures do not constitute exact models for large language models or graph learning applications, they provide valuable insights into the impact of asymmetry on mixing times and the associated concentration bounds. Nevertheless, hypercubes often serve as good approximations for regular and sparse graphs [9], [10], and hypertori frequently arise in distributed computation and networks-on-chip architectures [11], [12]. There mixing time and concentration bounds are important for obtaining performance guarantees for distributed algorithms run on them. Understanding the mixing times of asymmetric walks on cycles is key to understanding the same on tori.

The mixing times of lazy symmetric random walks on hypercubes, cycles, and tori are well established [13] for the case in which the laziness probability is equal to $\frac{1}{2}$. In this work, we study asymmetric random walks on these graphs. Although not formally published, upper bounds on mixing time for these asymmetric walks can be obtained using the standard spectral methods [13]. However, those upper bounds are not tight (have additional $\log n$ factors) and do not correctly capture the dependence on asymmetric parameters.

A. Main Contribution

Our principal technical contributions consist of deriving upper bounds on mixing times for these asymmetric random walks using coupling arguments. To the best of our knowledge, the coupling construction developed for asymmetric random walks on cycles is novel. It results in a mixing time bound that is order-wise tight and accurately captures the effect of asymmetry. In particular, we observe that for sufficiently large cycles, the ratio between the derived theoretical bound and the numerically computed mixing time remains constant across the choices of asymmetry parameters. Building on the bound obtained for asymmetric cycles, we further derive a corresponding bound for asymmetric random walks on tori, which is also tight in an order sense. When evaluated with

symmetric parameters, these mixing time bounds derived for asymmetric walks on cycles and tori match the best bounds for the corresponding symmetric walks. In this setting, we similarly observe that the ratio between the theoretical bound and the numerical mixing time exhibits insensitivity to the asymmetry parameters, thereby indicating that the proposed bound effectively captures the influence of asymmetry.

There are many practical situations in which one is interested in estimating functions of a n-dimensional discrete-time Markov chain (DTMC) that depend only on a small subset $\Delta \ll n$ of the state variables, such as conditional distributions of a few coordinates [3], [4]. We study functions of lazy asymmetric random walks on n-hypercubes that depend on a Δ-subset of state variables and show that sample complexity improves by a factor $\frac{\log \Delta}{\log n}$.

II. Mixing time analysis on the n-cycle

We consider a lazy asymmetric random walk on a cycle with n nodes, where n is even. At each step, the walk stays in the same place with probability $1 - \frac{p+q}{2}$ and takes one clockwise (CW) or anticlockwise (ACW) step with (asymmetric) probabilities $\frac{q}{2}$ and $\frac{p}{2}$, respectively, for $p + q < 1$.

For any Markov chain with a unique stationary distribution, the mixing time, denoted $t_{\text{mix}}(\varepsilon)$, quantifies the time required for the chain distribution to be within the total variation distance ε of the stationary distribution π, $t_{\text{mix}}(\varepsilon) := \min\left\{t : \max_{x\in S} \|P^t_{x,\cdot} - \pi\|_{\text{TV}} < \varepsilon\right\}$. We denote $t_{\text{mix}}(\frac{1}{4})$ by t_{mix}.

As discussed above, a mixing time bound directly gives a concentration bound. Hence, the goal is to bound the mixing time. The symmetric case is well studied [13] and builds on the well known coupling inequality [13]. For these asymmetric dynamics we use the coupling inequality while constructing a coupling that is different from the symmetric case, and to the best of our knowledge, is a new contribution of this work.

Let (X_t, Y_t) be two coupled processes, where each of them evolves according to the asymmetric random walk on the cycle. We refer to the two marginals as random walks of two balls, Ball 1 and 2, respectively. This new coupling is implemented using a partition of the unit interval, as shown in Figure 3. At each step, a uniform random variable (on $[0, 1]$) is sampled independent of the past, and its value determines the movements of both balls. This ensures that at most one ball moves at any given time and leads to a tight bound on the coupling time $\tau = \inf\{t \geq 0 : X_t = Y_t\}$.

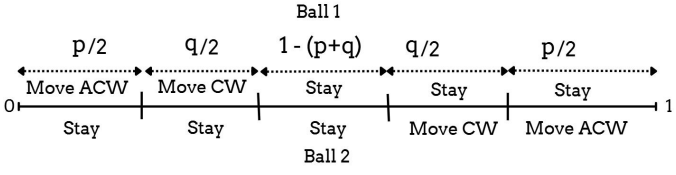

Fig. 1. Probability partition for Ball 1 and Ball 2

We define the cyclic distance between the two walks as $d_t = \min\{|X_t - Y_t|, n - |X_t - Y_t|\}$. Since n is even, the state-space of (d_t) is $\{0, 1, 2, \ldots, n/2\}$. The state $d = 0$ corresponds to the two balls occupying the same position, while $d = n/2$ represents the maximum possible distance, with the balls located at opposite ends of the cycle.

By construction of the coupling, at most one ball moves in each step, and the relative movements of the two balls are symmetric. Consequently, the distance between them can change by at most one at each step. It follows that (d_t) evolves as a birth–death Markov chain, as illustrated in Fig. 2.

Under this formulation, the coupling time τ for initial conditions x and y of the coupled chains coincides with the hitting time of state 0 for the birth–death chain (d_t) starting from $d_0 = |x - y|$.

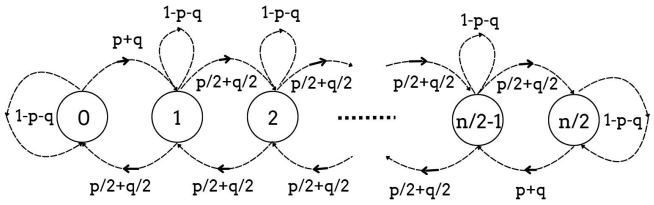

Fig. 2. Markov chain formed by the relative distance between the two balls in coupling

We define $h(k)$ as the expected hitting time to reach state $d = 0$ starting from distance $d = k$. The expected hitting time satisfies the standard recurrence [14],

$$h(k) = 1 + \sum_{j\notin\{0\}} p_{kj} h(j) \tag{2}$$

$$= 1 + (\frac{p}{2} + \frac{q}{2})h(k-1) + (1-p-q)h(k) + (\frac{p}{2} + \frac{q}{2})h(k+1), \text{ for } 1 \leq k \leq \frac{n}{2} - 1. \tag{3}$$

With boundary conditions, $h(0) = 0$, and

$$h\left(\frac{n}{2}\right) = 1 + (p+q)h\left(\frac{n}{2} - 1\right) + (1-p-q)h\left(\frac{n}{2}\right). \tag{4}$$

The recurrence relation in (2) has a general solution of the form $h(k) = ak^2 + bk + c$. Solving for the constants gives $h(k) = \frac{k(n-k)}{p+q}$. The maximum expected coupling time occurs at $k = n/2$ giving $\max_{0\leq k\leq n/2} h(k) = \frac{n^2}{4(p+q)}$. Then, by first invoking the coupling inequality for Markov chains [13] and then using Markov's inequality, we obtain

$$d(t) \leq \max_{x,y\in\mathbb{Z}_n} \mathbb{P}_{x,y}(\tau > t) \leq \max_{x,y} \frac{\mathbb{E}_{x,y}[\tau]}{t}. \tag{5}$$

Thus, $d(t) \leq \frac{n^2}{4(p+q)t}$. As discussed before, $\max_{x,y} \mathbb{E}_{x,y}[\tau]$ is upper bounded by $\max_{0\leq k\leq n/2} h(k)$. Thus, for $t = \frac{n^2}{(p+q)}$, we have $d(t) \leq \frac{1}{4}$, resulting in

$$t_{\text{mix}} \leq \frac{n^2}{p+q}. \tag{6}$$

As discussed later, this bound captures the dependence on p and q accurately. Also, the symmetric bound in [13, Ch. 5] is recovered when $p = q = \frac{1}{2}$.

A. Comparison with lower bounds

The mixing time bound can be evaluated using spectral analysis to obtain a lower bound. A standard spectral bound from Theorem 12.5 in [13] gives

$$t_{\text{mix}}(\varepsilon) \geq \left(\frac{1}{1-|\lambda_1|}-1\right)\log\left(\frac{1}{2\varepsilon}\right). \tag{7}$$

The transition probability matrix, on the cycle is circulant, and hence using Fourier analysis on the cycle, the eigenvalues are

$$\lambda_k = 1-\frac{p+q}{2}+\frac{p}{2}e^{2\pi ik/n}+\frac{q}{2}e^{-2\pi ik/n},\ k=0,1,\ldots,n-1. \tag{8}$$

For $k=1$ and large n, a Taylor expansion with $\theta=\frac{2\pi}{n}$ yields,

$$|\lambda_1| = \left|1-\left(\frac{p+q}{2}\right)\frac{\theta^2}{2}+i\left(\frac{p-q}{2}\right)\theta\right|. \tag{9}$$

We use the approximation $1-|\lambda_1| \approx \frac{1-|\lambda_1|^2}{2}$ to show that the spectral gap for the biased n-cycle is

$$\gamma = 1-|\lambda_1| \approx \frac{2\pi^2}{n^2}\left((p+q)-\frac{(p-q)^2}{2}\right). \tag{10}$$

Thus, $t_{mix} \geq \frac{\ln(2)}{\gamma} = \frac{n^2\ \ln 4}{\pi^2\left[(p+q)-\frac{(p-q)^2}{2}\right]}$ for $\epsilon=\frac{1}{4}$, where the approximation uses $\gamma \ll 1$ for large n. Whenever $|p-q|$, i.e., the asymmetry in the parameters is small compared to their average $\frac{p+q}{2}$, the lower bound is almost $\frac{n^2\ln 4}{(p+q)\pi^2}$.

B. Numerical results on the cycle

To compare our upper bound with the true mixing time, we numerically evaluate the mixing time of walks for $n \in \{10,20,40,60,80,100,120\}$ and many different (p,q) values including symmetric ($p=q$) and asymmetric cases. We follow the usual method of tracking the TV distance between the stationary distribution (uniform) and the distribution obtained using the transition kernel and noting the time it goes below $\frac{1}{4}$. Figure 3 shows that the ratio of the numerically calculated t_{mix} to our upper bound converges to a constant 0.10 for all p and q satisfying $(p+q) \gg \frac{(p-q)^2}{2}$.

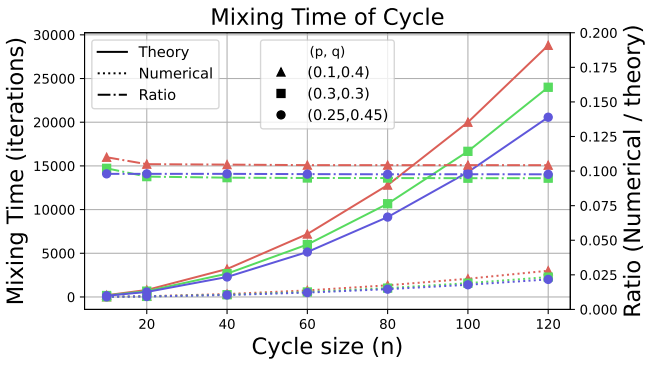

Fig. 3. Comparison of numerical t_{mix} and our upper bound for different n and p, q. The vertical axis on the right shows the ratio of the numerical to the theoretical upper bound.

Comparison of the numerically obtained mixing times and the spectral lower bound shows that their ratio is a constant irrespective of p and q. This indicates that the lower bound captures the dependence on p and q accurately up to a constant factor and the dependence of our upper bound on p and q is tight except for the second order term $\frac{(p-q)^2}{2}$.

III. Mixing Time Analysis on the d-Dimensional Torus $\mathbb{Z}_n^d$

The d-dimensional discrete torus $\mathbb{Z}_n^d$ is the graph whose vertex set is $V=\{(x_1,\ldots,x_d) \mid x_i \in \mathbb{Z}_n\ \forall\ i=1,2,\ldots,d\}$. Two vertices are adjacent if they differ in exactly one coordinate by $\pm 1 \mod n$. Each coordinate $x_i, 1 \leq i \leq d$, lives essentially in a cycle of length n.

We consider the well-known random walk on a torus [13] while adding asymmetry. At each step, a coordinate $i \in \{1,\ldots,d\}$ is chosen uniformly at random and the transition happens on this ith cycle according to the asymmetric lazy random walk on a cycle discussed in the previous section. We want to bound the mixing time of this asymmetric version of a random walk on torus. Towards that construct a coupling strategy building on the coupling for asymmetric random walk on a cycle.

Two random walks are started from two different initial points on the torus. At each step, the same coordinate is chosen in both the random walks, i.e., the same cycle is chosen. Then, within that cycle, the coupling described in Fig. 3 is used.

Let τ_i denote the coupling time of the i-th coordinate, i.e., the time by which the ith coordinates of the two processes become the same. The coupling time τ_{couple} is given by $\max_{1\leq i\leq d}\tau_i$. First, we obtain tail probabilities for τ_i, $1 \leq i \leq d$ using the results from previous section. Then, we introduce another coupling whose coupling time is stochastically larger and bound that using the tail probabilities for τ_i.

Following the same steps as in the case of the cycle, the expected coupling time for a single coordinate satisfies $\mathbb{E}[\tau_i] = dn^2/4(p+q)$, where the additional d factor arises because coordinate i is selected with probability $1/d$ at each step. Hence, there is a geometric waiting time with expectation d between two consecutive movements in the ith cycle. By Markov's inequality, the probability that the coordinate has not yet been coupled by time $t=dn^2/(p+q)$ is

$$\mathbb{P}\left(\tau_i > \frac{dn^2}{(p+q)}\right) \leq \frac{\mathbb{E}[\tau_i]}{\frac{dn^2}{(p+q)}} = \frac{\frac{dn^2}{4(p+q)}}{\frac{dn^2}{p+q}} = \frac{1}{4}. \tag{11}$$

Let us consider the following modification to the coupling. If by $t=\frac{dn^2}{p+q}$ all the coordinates are not coupled, we repeat the two processes from their initial condition. Clearly, the coupling time in this process stochastically dominates the original coupling time.

Building on the bound for $\mathbb{P}\left(\tau_i > \frac{dn^2}{(p+q)}\right)$ in the original coupling, for this new coupling with re-initialization we have,

$$\mathbb{P}\left(\tau_i > \frac{kdn^2}{(p+q)}\right) \leq \left(\frac{1}{4}\right)^k. \tag{12}$$

This is because at every $\frac{dn^2}{p+q}$th time the process is renewed independent of the past. By the union bound,

$$\mathbb{P}(\tau_{\text{couple}} > \frac{kdn^2}{p+q}) \leq \sum_{i=1}^{d} \mathbb{P}\left(\tau_i > \frac{kdn^2}{p+q}\right) \leq d\left(\frac{1}{4}\right)^k \quad (13)$$

Choosing $k \geq \log\left(\frac{d}{\varepsilon}\right)$ ensures that this probability is at most ε. By the coupling inequality, we obtain the mixing time of a torus as

$$t_{\text{mix}} \leq \frac{dn^2}{p+q} \log_4(4d), \quad (14)$$

which highlights the dependence of mixing time on the varying transition probabilities p and q. The symmetric bound in [13, Ch. 5] follows when $p = q = \frac{1}{2}$.

A. Numerical results on the torus

We numerically evaluate the mixing time for d-dimensional torus for $d \in \{1, \ldots, 6\}$ and $4 \leq n \leq 10$, and multiple pairs of (p, q), with $\epsilon = 0.25$. We observe that the ratio of the numerically calculated t_{mix} to the upper bound is approximately in the range of 0.08 to 0.09 for all choices of d, and (p, q) when $n \geq 5$. Thus, the upper bound accurately captures the dependence on p and q up to a constant factor.

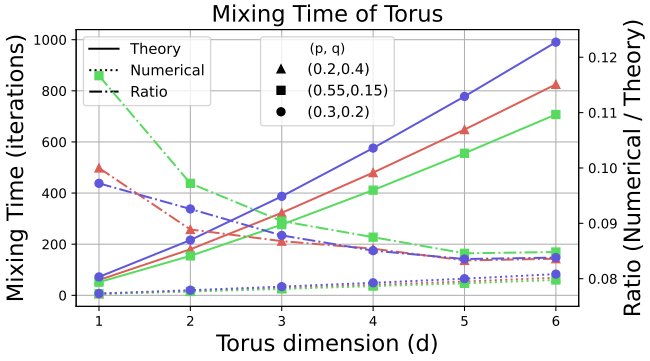

Fig. 4. Comparison of numerical t_{mix} and our upper bound for $n = 6$ for different d and p, q. The vertical axis on the right shows the ratio of the numerical to the theoretical upper bound.

IV. Mixing time analysis on the Hypercube

We start with a symmetric random walk on an n-dimensional hypercube having state-space $\{0, 1\}^n$, where each vertex is a binary n-tuple and differs from its neighbors in exactly one coordinate. Let $(X_t)_{t\geq 0}$ be the DTMC with state space $\{0, 1\}^n$, where $X_t = (X_{t,1}, \ldots, X_{t,n}) \in \{0, 1\}^n$ denotes the state at time t. At each step of the walk, the Markov chain stays in its current state with probability $1 - p$, and with probability p choosing to move any of the n neighbors uniformly at random [13].

A. Functions of a Few Coordinates of the Random Walk

As mentioned above, in many learning problems, including graph and causal learning [3]–[5], it is important to estimate the expectation of a function of a subset of coordinates of a high-dimensional Markov process. In general, such problems would involve estimating $\sum_{t=1}^{T} f(\mathbf{X}_t)$, where $f(\mathbf{X}_t) = f_\Delta(X_{t,i} : i \in C)$, for $C \subset [n]$ and $|C| = \Delta$.

A common approach to this problem involves a concentration inequality for the whole Markov chain $\mathbf{X}_t$ [3]–[5]. Here, we show that by focusing on the evolution of the relevant coordinates, one can tighten the bounds significantly. In particular, using [7, Cor. 2.11] for the whole process $\mathbf{X}_t$, the sample complexity for having ϵ-accurate estimate would scale $\sim n \log n$, whereas by focusing on the relevant coordinates, this can be tightened to $\sim n \log \Delta$.

We define $\mathbf{X}_t^{(\Delta)} = (X_{j,t}, X_{j+1,t}, \ldots, X_{j+\Delta,t})$, a subprocess of $\mathbf{X}_t$. For obtaining the above concentration bound we study $\mathbf{X}_t^{(\Delta)}$. First, note that $\mathbf{X}_{t\geq 0}^{(\Delta)}$ is a DTMC, which can be proved using the Markov property and the specific structure of $\{\mathbf{X}_t\}_{t\geq 0}$.

Fix $\mathbf{y}, \mathbf{y}' \in \{0, 1\}^\Delta$. Using the law of total probability for disjoint events $\{\mathbf{X}_t = \mathbf{x} : \mathbf{x}^{(\Delta)} = \mathbf{y}\}$, $\mathbb{P}\big(\mathbf{X}_{t+1}^{(\Delta)} = \mathbf{y}' \mid \mathbf{X}_t^{(\Delta)} = \mathbf{y}, \mathbf{X}_{t-1}^{(\Delta)}, \ldots\big)$ becomes

$$\sum_{\mathbf{x}:\mathbf{x}^{(\Delta)}=\mathbf{y}} \sum_{\mathbf{x}':\mathbf{x}'^{(\Delta)}=\mathbf{y}'} \mathbb{P}\big(\mathbf{X}_{t+1} = \mathbf{x}' \mid \mathbf{X}_t = \mathbf{x}, \mathbf{X}_t^{(\Delta)} = \mathbf{y}, \ldots\big) \mathbb{P}\big(\mathbf{X}_t = \mathbf{x} \mid \mathbf{X}_t^{(\Delta)} = \mathbf{y}, \ldots\big). \quad (15)$$

The term $\mathbb{P}\big(\mathbf{X}_{t+1} = \mathbf{x}' \mid \mathbf{X}_t = \mathbf{x}, \mathbf{X}_t^{(\Delta)} = \mathbf{y}, \ldots\big)$ is 0 if $\mathbf{x}'$ and $\mathbf{x}$ differ in more than one coordinate, $\frac{p}{n}$ if $\mathbf{x}'$ and $\mathbf{x}$ differ in exactly one coordinate, and $1-p$ otherwise. Since the second term, $\mathbb{P}\big(\mathbf{X}_t = \mathbf{x} \mid \mathbf{X}_t^{(\Delta)} = \mathbf{y}, \ldots\big)$, is not affected by $\mathbf{x}'$ and $\mathbf{y}'$, after taking the inner sum over $\mathbf{x}'$, Eq.(15) becomes

$$p(\mathbf{y}, \mathbf{y}') \cdot \sum_{\mathbf{x}:\mathbf{x}^{(\Delta)}=\mathbf{y}} \mathbb{P}\big(\mathbf{X}_t = \mathbf{x} \mid \mathbf{X}_t^{(\Delta)} = \mathbf{y}, \ldots\big) = p(\mathbf{y}, \mathbf{y}'),$$

where the last step follows by the law of total probability. Here, $p(\mathbf{y}, \mathbf{y}')$ is $1 - \frac{\Delta p}{n}$ if $\mathbf{y} = \mathbf{y}'$, it is 0 if $\mathbf{y}$ and $\mathbf{y}'$ differ in more than one coordinate, and it is $\frac{\Delta p}{n}$ when $\mathbf{y}$ and $\mathbf{y}'$ differ in exactly one coordinate.

We obtain a tighter mixing time upper bound $\frac{2n \log \Delta}{p(1-p)}$ for $\mathbf{X}_t^{(\Delta)}$ compared to $\frac{2n \log n}{p(1-p)}$ for the full process $\mathbf{X}_t$. We skip the derivation for this symmetric case in the interest of the more general asymmetric case discussed next.

B. Mixing Time for Asymmetric RW on Hypercube

The symmetric walk on the hypercube can be equivalently seen as the following. The walk chooses a dimension (coordinate) uniformly at random and moves one (or zero) step in that dimension if a coin toss is head with probability p. For the asymmetric version of the walk we consider the coin toss for dimension (coordinate) i to have probability of head p_i.

We couple two walks with the same kernel by having the same uniform random choice of the dimension. The coin tosses in the two processes are independent in those dimensions where the two processes have different values. When the two processes have the same value in a dimension (coordinate) they toss the same coin.

For each of the i^{th} coordinate, define $k_i = \frac{1}{2p_i(1-p_i)}$, and τ_i be the first time when coordinate i couples. Then τ_i is a geometric random variable with success probability $\frac{1}{k_i\ n}$.

Let $\tau := \max_{i\in[\Delta]} \tau_i$ be the time by which all Δ coordinates have been coupled. Then, using union bound,

$$\mathbb{P}(\tau > t) = \mathbb{P}\left(\max_i \tau_i > t\right) \leq \sum_i \mathbb{P}(\tau_i > t). \tag{16}$$

Now, $\mathbb{P}(\tau_i > t) = \left(1 - \frac{1}{n\ k_i}\right)^t$ and

$$\mathbb{P}(\tau > t) \leq \sum_{i\in[\Delta]} \left(1 - \frac{1}{n\ k_i}\right)^t \leq \sum_{i\in[\Delta]} \exp\left(-\frac{t}{n\ k_i}\right). \tag{17}$$

For $\mathbb{P}(\tau > t)$ to be less than $\frac{1}{4}$, it suffices to have $t \geq t_{\text{mix}}^{(\Delta)} = n \max_{i\in[\Delta]} k_i \log_4 \Delta$.

Also, since the maximum is less than the sum for non-negative random variables, $\mathbb{E}[\tau] \leq \sum_{i\in[\Delta]} \tau_i \leq n \sum_{i\in[\Delta]} k_i$. This implies another bound on $t_{\text{mix}}^{(\Delta)}$ since $t_{\text{mix}}^{(\Delta)} \leq \frac{1}{4}\mathbb{E}[\tau]$. Thus, we get,

$$t_{\text{mix}}^{(\Delta)} \leq n \cdot \min\left(\max_{i\in[\Delta]} k_i \log_4 \Delta, \frac{1}{4}\sum_{i\in[\Delta]} k_i\right). \tag{18}$$

Although it may be tempting to expect that, for sufficiently large Δ, the above bound simplifies to $n \max_{i\in[\Delta]} k_i \log_4 \Delta$, the situation is more subtle. For example, suppose $\{p_i\}$ are drawn randomly such that the induced distribution of $\{k_i\}$ is heavy-tailed with infinite mean. In this case, with high probability, $\max_{i\in[\Delta]} k_i \log_4 \Delta$ grows asymptotically faster than $\sum_{i\in[\Delta]} k_i$. In particular, this occurs when p_i are i.i.d. uniform on $[0, 1]$ or follow i.i.d. Beta distributions with at least one parameter less than 1, both of which yield heavy-tailed k_i with infinite mean.

For $p_i = \frac{1}{2}$, the symmetric bound [13, Ch. 5] is recovered and $t_{mix}^{(\Delta)}$ is $\frac{\log\Delta}{\log n}$ times better than the mixing time bound for the full chain. This results in $\frac{\log\Delta}{\log n}$ times improvement in sample complexity for problems involving estimation of the expectation of functions of Δ coordinates [7, Cor. 2.11].

We numerically evaluated mixing times of walks on n-dimensional hypercube and that of the processes restricted to Δ-coordinates, and observing that they consistently validate their corresponding theoretical upper bounds. When $\{p_i\}$ are chosen uniformly at random in $[0, 1]$, we observe that for most choices of $\{p_i\}$, $\max_{i\in[\Delta]} k_i \log_4 \Delta \geq \sum_{i\in[\Delta]} k_i$, reaffirming the theoretical predictions above.

V. Conclusion

Learning from Markov samples is fundamental in a wide range of applications, including LLMs and social and biological networks. Sample complexity in these settings often depends on upper bounds on the mixing times of asymmetric, high-dimensional Markov processes on regular graphs. Motivated by this, we analyze the mixing times of lazy, asymmetric random walks on cycles, tori, and hypercubes, emphasizing the role of transition asymmetry. Using coupling-based arguments, we derive order-wise tight upper bounds that capture asymmetry accurately and recover the best-known symmetric results as special cases.

For the cycle, to the best of our knowledge, the proposed coupling is novel and yields an order-wise tight bound that almost precisely reflects the dependence on asymmetric transition probabilities. Spectral techniques are not directly applicable due to non-reversibility and, even when used, introduce an additional logarithmic factor. The construction extends naturally to d-dimensional tori, yielding an order-wise tight bound.

For random walks on n-dimensional hypercubes, we show that when only Δ coordinates are observed, the effective process mixes faster. In the symmetric case, the improvement is by a factor of $\frac{\log\Delta}{\log n}$. Consequently, the sample complexity of estimating expectations of functions depending on Δ-coordinates improves by the same factor, with implications for practical graph learning problems [3]–[5].

References

[1] M. E. Ildiz, Y. Huang, Y. Li, A. S. Rawat, and S. Oymak, "From self-attention to Markov models: Unveiling the dynamics of generative transformers," in *Proceedings of the 41st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, and F. Berkenkamp, Eds., vol. 235. PMLR, Jul 2024, pp. 20 955–20 982.

[2] A. V. Makkuva, M. Bondaschi, A. Girish, A. Nagle, M. Jaggi, H. Kim, and M. Gastpar, "Attention with Markov: A curious case of single-layer transformers," in *International Conference on Learning Representations (ICLR)*, 2025.

[3] D. Katselis, C. L. Beck, and R. Srikant, "Mixing times and structural inference for bernoulli autoregressive processes," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 364–378, 2018.

[4] A. Chatterjee, A. S. Rawat, S. Vishwanath, and S. Sanghavi, "Learning the causal graph of markov time series," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2013, pp. 107–114.

[5] N. A. Barker, C. J. Myers, and H. Kuwahara, "Learning genetic regulatory network connectivity from time series data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 1, pp. 152–165, 2011.

[6] S. Bagewadi and A. Chatterjee, "Effect of correlated errors on quantum memory," *Phys. Rev. A*, vol. 112, p. 022422, Aug 2025.

[7] D. Paulin, "Concentration inequalities for Markov chains by Marton couplings and spectral methods," *Electronic Journal of Probability*, vol. 20, pp. 1 – 32, 2015.

[8] L. Kontorovich and K. Ramanan, "Concentration inequalities for dependent random variables via the martingale method," *Annals of Probability*, vol. 36, no. 6, pp. 2126–2158, 2008.

[9] S. Petti and S. S. Vempala, "Approximating sparse graphs: The random overlapping communities model," *Random Structures & Algorithms*, vol. 61, no. 4, pp. 844–908, 2022.

[10] S. Diskin, J. Erde, M. Kang, and M. Krivelevich, "Isoperimetric inequalities and supercritical percolation on high-dimensional graphs," *Combinatorica*, vol. 44, no. 4, pp. 741–784, 2024.

[11] Y. Ajima, S. Sumimoto, and T. Shimizu, "Tofu: A 6d mesh/torus interconnect for exascale computers," *IEEE Computer*, vol. 42, no. 11, pp. 36–40, 2009.

[12] Y. Li, S. Peng, W. Chu, and H.-M. Wang, "Some properties and algorithms for the hyper-torus network," *The Journal of Supercomputing*, vol. 68, no. 1, pp. 72–97, 2014.

[13] D. A. Levin and Y. Peres, *Markov chains and Mixing times*. American Mathematical Soc., 2017, vol. 107.

[14] J. J. Hunter, "The computation of the mean first passage times for Markov chains," *Linear Algebra and its Applications*, vol. 549, pp. 100–122, 2018.