Electronic Journal of Probability

Vol. 12 (2007), Paper no. 10, pages 262–299.

Journal URL
http://www.math.washington.edu/~ejpecp/

# Robust Mixing

Murali K. Ganapathy
Google Inc.*

**Abstract**

In this paper, we develop a new "robust mixing" framework for reasoning about adversarially modified Markov Chains (AMMC). Let $\mathbb{P}$ be the transition matrix of an irreducible Markov Chain with stationary distribution $\pi$. An adversary announces a sequence of stochastic matrices $\{\mathbb{A}_t\}_{t>0}$ satisfying $\pi\mathbb{A}_t = \pi$. An AMMC process involves an application of $\mathbb{P}$ followed by $\mathbb{A}_t$ at time $t$. The robust mixing time of an ergodic Markov Chain $\mathbb{P}$ is the supremum over all adversarial strategies of the mixing time of the corresponding AMMC process. Applications include estimating the mixing times for certain non-Markovian processes and for reversible liftings of Markov Chains.

**Non-Markovian card shuffling processes:** The random-to-cyclic transposition process is a *non-Markovian* card shuffling process, which at time $t$, exchanges the card at position $L_t := t \pmod{n}$ with a random card. Mossel, Peres and Sinclair (2004) showed a lower bound of $(0.0345 + o(1))n \log n$ for the mixing time of the random-to-cyclic transposition process. They also considered a generalization of this process where the choice of $L_t$ is adversarial, and proved an upper bound of $Cn \log n + O(n)$ (with $C \approx 4 \times 10^5$) on the mixing time. We reduce the constant to 1 by showing that the random-to-top transposition chain (*a Markov Chain*) has robust mixing time $\leq n \log n + O(n)$ when the adversarial strategies are limited to holomorphic strategies, i.e. those strategies which preserve the symmetry of the underlying Markov Chain. We also show a $O(n \log^2 n)$ bound on the robust mixing time of the lazy random-to-top transposition chain when the adversary is not limited to holomorphic strategies.

**Reversible liftings:** Chen, Lovász and Pak showed that for a reversible ergodic Markov Chain $\mathbb{P}$, any reversible lifting $\mathbb{Q}$ of $\mathbb{P}$ must satisfy $\mathcal{T}(\mathbb{P}) \leq \mathcal{T}(\mathbb{Q}) \log(1/\pi_*)$ where $\pi_*$ is the minimum stationary probability. Looking at a specific adversarial strategy allows us to show that $\mathcal{T}(\mathbb{Q}) \geq r(\mathbb{P})$ where $r(\mathbb{P})$ is the relaxation time of $\mathbb{P}$. This gives an alternate proof of

*This work was done when the author was a student at the Univ. of Chicago

the reversible lifting result and helps identify cases where reversible liftings cannot improve the mixing time by more than a constant factor.

**Key words:** Markov Chains, Robust mixing time, Reversible lifting, random-to-cyclic transposition, non-Markovian processes.

**AMS 2000 Subject Classification:** Primary 60J10, 62M09, 62M15.

Submitted to EJP on October 3 2006, final version accepted February 14 2007.

# 1 Introduction

In this paper, we develop a "robust mixing" framework which allows us to reason about adversarially modified Markov Chains (AMMC). This framework can be used to bound mixing times of some *non-Markovian processes* in terms of the robust mixing time of related Markov Chains. Another type of application is to estimate mixing times of complex Markov Chains in terms of that of simpler Markov Chains. Finally, we also use this framework to give an alternate proof of a reversible lifting result due to Chen et al. (4).

## 1.1 Robust Mixing

All stochastic processes considered in this paper have finite state space and run in discrete time. Let $\mathcal{M}$ be an irreducible Markov chain on state space $\mathcal{X}$ with transition probability matrix $\mathbb{P}$ and stationary distribution $\pi$. When the context is clear, we use $\mathbb{P}$ to denote the Markov Chain as well as its transition probability matrix.

Markov Chains we consider here are not assumed to be reversible, unless otherwise specified. All logarithms are taken to the base e unless otherwise specified.

**Definition 1.1.** Let $\mathbb{P}$ be the transition matrix of a Markov Chain. Its mixing time and $L_2$-mixing time are defined by the equations

$$\mathcal{T}(\mathbb{P},\epsilon) = \max_{\mu}\min_{t}\{||\mu\mathbb{P}^t - \pi||_{TV} \le \epsilon\} \quad \text{and} \quad \mathcal{T}_2(\mathbb{P},\epsilon) = \max_{\mu}\min_{t}\{||\mu\mathbb{P}^t - \pi||_{2,\pi} \le \epsilon\}$$

respectively. Here $||\mu - \pi||_{TV} = \sum_x |\mu(x) - \pi(x)|/2$ is the total variation norm and $||\mu - \pi||_{2,\pi}^2 = \sum_x (\mu(x) - \pi(x))^2/\pi(x)$ is the $L_2(\pi)$ norm. When $\epsilon$ is not specified, we take it to be $1/4$ for $\mathcal{T}$ and $1/2$ for $\mathcal{T}_2$.

Note that for $\epsilon < 1/2$, the inequalities $2||\cdot||_{TV} \le ||\cdot||_{2,\pi} \le 2||\cdot||_{TV}\sqrt{\frac{1}{\pi_*}}$ together with submultiplicativity of $\mathcal{T}(\epsilon/2)$ give

$$\mathcal{T}(\mathbb{P},\epsilon) \le \mathcal{T}_2(\mathbb{P},2\epsilon) \le \mathcal{T}(\mathbb{P},\epsilon)\log_{1/2\epsilon}(1/\pi_*), \tag{1}$$

where $\pi_* = \min_x \pi(x)$ is the minimum stationary probability.

**Definition 1.2.** Let $\mathbb{P}$ be an irreducible Markov Chain with stationary distribution $\pi$. A stochastic matrix $\mathbb{A}$ (not necessarily irreducible) is said to be *compatible* with $\mathbb{P}$ if $\pi\mathbb{A} = \pi$.

Note that the notion of compatibility depends only on the stationary distribution of $\mathbb{P}$.

**Definition 1.3.** An *adversarially modified Markov Chain* (AMMC) $\mathcal{P}$ is a pair $(\mathbb{P}, \{\mathbb{A}_t\}_{t>0})$, where $\mathbb{P}$ is the transition matrix of an irreducible Markov Chain and $\mathbb{A}_t$ is a sequence of stochastic matrices compatible with $\mathbb{P}$. Given an AMMC and an initial distribution $\mu_0$, the AMMC process evolves as follows:

- At time $t = 0$, pick $X_0 \in \mathcal{X}$ according to $\mu_0$,
- Given $X_t$, pick $Y_t$ according to the distribution $\mathbb{P}(X_t, \cdot)$,

- Given $Y_t$, pick $X_{t+1}$ according to the distribution $\mathbb{A}_t(Y_t, \cdot)$

An application of $\mathbb{P}$ followed by $\mathbb{A}_t$ is called a *round*. Let $\mu_t$ and $\nu_t$ denote the distribution of $X_t$ and $Y_t$ respectively. Then $\mu_t$ is the distribution after $t$-rounds. Also

$$\nu_t = \mu_t \mathbb{P} \quad \text{and} \quad \mu_{t+1} = \nu_t \mathbb{A}_t \tag{2}$$

**Definition 1.4.** Let $\mathcal{P}$ be an AMMC. Its mixing time and $L_2$-mixing time are defined by the equations

$$\mathcal{T}(\mathcal{P}, \epsilon) = \max_{\mu_0} \min_t \{||\mu_t - \pi||_{TV} \leq \epsilon\} \quad \text{and} \quad \mathcal{T}_2(\mathcal{P}, \epsilon) = \max_{\mu_0} \min_t \{||\mu_t - \pi||_{2,\pi} \leq \epsilon\}$$

respectively. When $\epsilon$ is not specified, we take it to be $1/4$ for $\mathcal{T}$ and $1/2$ for $\mathcal{T}_2$.

The proof of (1) together with submultiplicativity of $\mathcal{T}(\mathcal{P})$ (Theorem 1.8) shows that for $\epsilon < 1/2$,

$$\mathcal{T}(\mathcal{P}, \epsilon) \leq \mathcal{T}_2(\mathcal{P}, 2\epsilon) \leq \mathcal{T}(\mathcal{P}, \epsilon) \log_{\frac{1}{2\epsilon}}(1/\pi_*) \tag{3}$$

**Definition 1.5.** Let $\mathbb{P}$ be an irreducible Markov Chain. An *adversarially modified version* of $\mathbb{P}$ is an AMMC $(\mathbb{P}, \{\mathbb{A}_t\}_{t>0})$.

**Definition 1.6.** Let $\mathbb{P}$ be an ergodic Markov Chain. The *robust mixing time* and *robust $L_2$-mixing time* of $\mathbb{P}$ are defined by the equations

$$R(\mathbb{P}, \epsilon) = \sup_{\mathcal{P}} \mathcal{T}(\mathcal{P}, \epsilon) \quad \text{and} \quad R_2(\mathbb{P}, \epsilon) = \sup_{\mathcal{P}} \mathcal{T}_2(\mathcal{P}, \epsilon)$$

respectively, where the suprema are taken over adversarially modified versions $\mathcal{P}$ of $\mathbb{P}$. When $\mathbb{P}$ is clear from context, we drop it and when $\epsilon$ is not specified we take it to be $1/4$ for $R$ and $1/2$ for $R_2$.

The set of stochastic matrices compatible with $\mathbb{P}$ is a bounded polytope and hence the convex hull of its vertices. Since the distances used to measure the mixing time are convex, it follows that the worst case for robust mixing time is achieved when each $\mathbb{A}_t$ is a vertex of the polytope. If the stationary distribution is uniform, the polytope is called the *assignment polytope* and its vertices are permutation matrices.

When we need to distinguish between the standard notion of mixing time and robust mixing time, we refer to the standard notion as "standard mixing time."

One can think of the standard mixing time of a Markov Chain as the number of (contiguous) applications of $\mathbb{P}$ required to get close to stationarity. In the same vein, the robust mixing time is the number of *not necessarily contiguous* applications of $\mathbb{P}$ required to get close to stationarity under reasonable assumptions on the intervening steps. Our adversary is *oblivious*. See Section 1.7 for related discussion. If $\pi$ is the uniform distribution then compatibility is equivalent to requiring that the $\mathbb{A}_t$ are doubly stochastic.

**Definition 1.7.** Let $\mathcal{R} = \{R_t\}_{t>0}$ be a stochastic process on a state space $\mathcal{X}$ and $\mathbb{P}$ an irreducible Markov Chain. We say that $\mathcal{R}$ can be *simulated by an adversarially modified* $\mathbb{P}$, if there is an adversarially modified version $\mathcal{P}$ of $\mathbb{P}$ such that for every initial distribution $\nu_0$ of $R_0$, there is an initial distribution $\mu_0$ of $\mathcal{P}$ such that the distribution $\nu_t$ of $R_t$ and $\mu_t$ are equal.

We give two examples in the card shuffling context of simulating a non-Markovian process by an adversarially modified Markov Chain in Section 1.5.

## 1.2 Properties of robust mixing time

Like standard mixing time, robust mixing time is also submultiplicative.

**Theorem 1.8.** ***(Submultiplicativity)*** *Let $\mathbb{P}$ be an ergodic Markov Chain. For $\epsilon, \delta > 0$,*

$$R(\mathbb{P}, \epsilon\delta/2) \leq R(\mathbb{P}, \epsilon/2) + R(\mathbb{P}, \delta/2) \quad \text{and} \quad R_2(\mathbb{P}, \epsilon\delta) \leq R_2(\mathbb{P}, \epsilon) + R_2(\mathbb{P}, \delta)$$

This will be proved in Section 2.1. A useful property enjoyed by robust mixing time not shared by the standard mixing time is the following convexity property.

**Theorem 1.9.** ***(Convexity)*** *Let $\mathbb{P}$ be an ergodic Markov Chain with stationary distribution $\pi$ and $\mathbb{Q}$ any Markov Chain compatible with $\mathbb{P}$. Let $0 < a = 1 - b < 1$ and $\pi_* = \min_x \pi(x)$. Then $R(\mathbb{P}\mathbb{Q}, 1/4) \leq R(\mathbb{P}, 1/4)$ and $R_2(\mathbb{P}\mathbb{Q}, 1/2) \leq R_2\mathbb{P}, 1/2)$. Also,*

$$R(a\mathbb{P} + b\mathbb{Q}, 1/4) \leq R(\mathbb{P}, 1/4) + R(\mathbb{Q}, 1/4) - 1$$
$$R_2(a\mathbb{P} + b\mathbb{Q}, 1/2) \leq R_2(\mathbb{P}, 1/2) + R_2(\mathbb{Q}, 1/2) - 1$$

*Moreover,*

- *if $R(\mathbb{P}, 1/4) \geq 11$, then $R(a\mathbb{P} + b\mathbb{Q}, 1/4) \leq 3R(\mathbb{P}, 1/4)/a$*
- *If $\pi_* \leq 1/16$ and $R_2(\mathbb{P}, 1/2) \geq \log(1/\pi_*)/2$ then $R_2(a\mathbb{P} + b\mathbb{Q}) \leq 7R(\mathbb{P})/a$*

Theorem 1.9 is proved in Section 2.3. Convex combinations of Markov Chains are considered in (3) to sample linear orderings. For reversible chains $\mathbb{P}$ and $\mathbb{Q}$, using standard results, one can show that a convex combination of $\mathbb{P}$ and $\mathbb{Q}$ mixes in time $O(\min(\mathcal{T}(\mathbb{P}), \mathcal{T}(\mathbb{Q})) \log(1/\pi_*))$. Our result allows us to eliminate the $\log(1/\pi_*)$ factor under some conditions.

## 1.3 Relation to classical parameters of Markov Chains

We now relate the robust mixing time of Markov chains to classical mixing parameters.

**Definition 1.10.** Let $\mathbb{P}$ be an ergodic chain with stationary distribution $\pi$.

- Denote by $\pi_*$ the smallest entry of $\pi$, i.e. $\pi_* = \min_x \pi(x)$,
- Let $\Pi$ denote a diagonal matrix with entries $\pi$, i.e. $\Pi(x, x) = \pi(x)$,
- $\mathbb{S}(\mathbb{P}) = \sqrt{\Pi^{-1}}\mathbb{P}\sqrt{\Pi}$,
- $\overleftarrow{\mathbb{P}} = \Pi^{-1}\mathbb{P}^T\Pi$ denotes the reverse of the Markov Chain $\mathbb{P}$, where $\mathbb{P}^T$ denotes the transpose of the matrix $\mathbb{P}$.

**Definition 1.11.** Let $\mathbb{A}$ be any $N \times N$ real matrix. By a *singular value decomposition of* $\mathbb{A}$, we mean two orthonormal bases $\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1}\}$ and $\{\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{N-1}\}$ and scalars $\sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{N-1} \geq 0$ which satisfy

$$\boldsymbol{x}_i\mathbb{A} = \sigma_i\boldsymbol{y}_i \qquad \boldsymbol{y_i}\mathbb{A}^T = \sigma_i\boldsymbol{x}_i \tag{4}$$

The $\sigma_i$ are called the *singular values* and the $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are called the *left and right singular vectors* respectively.

See Horn and Johnson (14, Chapter 3) for results about singular values. If $\mathbb{A}$ is the transition matrix of a reversible chain or a real symmetric matrix, let $\lambda_0(\mathbb{A}) \geq \lambda_1(\mathbb{A}) \geq \cdots \geq \lambda_{N-1}(\mathbb{A})$ denote its eigenvalues and put $\lambda_*(\mathbb{A}) = \max(|\lambda_1(\mathbb{A})|, |\lambda_{N-1}(\mathbb{A})|)$.

**Definition 1.12.** Let $\mathbb{P}$ be an ergodic reversible Markov Chain. Its *relaxation time* is defined by

$$r(\mathbb{P}) = \frac{-1}{\log \lambda_*(\mathbb{P})}$$

Note that for any ergodic reversible Markov Chain $\mathbb{P}$, $r(\mathbb{P}) \leq \mathcal{T}(\mathbb{P}) \leq r(\mathbb{P})(\log(1/\pi_*)/2 + 1)$.

Just like how the mixing time of a reversible chain is determined by the largest eigenvalue (in modulus) up to a $\log(1/\pi_*)$ factor, the robust mixing time of a Markov Chain (not necessarily reversible) is determined by its second largest singular value of $\mathbb{S}(\mathbb{P})$ up to a $\log(1/\pi_*)$ factor. More specifically, we have

**Theorem 1.13.** *Let $\mathbb{P}$ be an ergodic Markov Chain with stationary distribution $\pi$. Then*

$$r(\mathbb{P}\overleftarrow{\mathbb{P}}) \leq \max(\mathcal{T}(\mathbb{P}\overleftarrow{\mathbb{P}}), \mathcal{T}(\mathbb{P})) \leq R(\mathbb{P}) \leq 2r(\mathbb{P}\overleftarrow{\mathbb{P}})(\log(1/\pi_*)/2 + 1)$$

*In particular if $\mathbb{P}$ is reversible, we have*

$$r(\mathbb{P}) \leq \mathcal{T}(\mathbb{P}) \leq R(\mathbb{P}) \leq r(\mathbb{P}) \left( \frac{\log(1/\pi_*)}{2} + 1 \right)$$

In Section 4, we show that many of the techniques used to prove upper bounds on mixing time actually give bounds on the robust mixing time. These include eigenvalue estimation, conductance methods, log-Sobolev inequalities and most analytical methods. The most notable exception is coupling. Mixing time bounds established via coupling do not automatically lead to bounds on robust mixing time. However in certain cases, they lead to bounds on robust mixing time against certain types of restricted adversaries.

## 1.4 Cayley walks with restricted adversaries

We now turn to Markov Chains induced by walks on groups.

**Definition 1.14.** Let $G$ be a finite group and $P$ a probability distribution over $G$. By a *Cayley walk on $G$ induced by $P$* we mean a Markov Chain on $G$ with transition probability matrix $\mathbb{P}$ given by $\mathbb{P}(h, h \cdot s) = P(s)$ for all $h, s \in G$. By a *Cayley walk on $G$*, we mean a Cayley walk on $G$ induced by $P$ for some probability distribution $P$ over $G$.

In case of a *Cayley walk*, one can look at the robust mixing time when the adversary's strategies are limited to those preserving the symmetries of the group. We consider two natural restrictions.

**Definition 1.15.** Let $\mathbb{P}$ denote a Cayley walk on a group $G$. A *Cayley strategy* is a doubly stochastic matrix $\mathbb{A}$ such that it is the transition probability matrix of some Cayley walk (not necessarily irreducible) on $G$. Denote by $\mathcal{C}$ the set of all Cayley strategies for the group $G$ ($G$ will be clear from context). A *Cayley adversary* is an adversary whose strategies are limited to Cayley strategies.

Note that a Cayley adversary at time $t$, is only allowed to right multiply the current group element $g$ by a group element $s$ chosen from the distribution $P_t$ on $G$. See Section 5.2 for more discussion on the power of a Cayley adversary.

**Definition 1.16.** Let $G$ be a group. A permutation $J$ on $G$ is said to be a *holomorphism* if it can be written as the composition of

- right/left multiplication by elements of $G$, and
- automorphisms of $G$.

**Definition 1.17.** Let $\mathbb{P}$ be a Cayley walk on a group $G$. A *holomorphic strategy* is a doubly stochastic matrix $\mathbb{A}$ which can be written as a convex combination of holomorphisms of $G$. Denote by $\mathcal{H}$ the set of all holomorphic strategies of $G$ ($G$ will be clear from the context). A *holomorphic adversary* is one who is limited to holomorphic strategies.

We now turn to defining the robust mixing time against restricted adversaries.

**Definition 1.18.** Let $\mathbb{P}$ be an irreducible Markov Chain. A set $\mathcal{S}$ of stochastic matrices is said to be *a valid set of strategies against* $\mathbb{P}$ if it satisfies the following:

- $I \in \mathcal{S}$,
- $\mathbb{A} \in \mathcal{S} \implies \mathbb{A}$ is compatible with $\mathbb{P}$,
- $\mathcal{S}$ is closed under products and convex combinations.

**Definition 1.19.** Let $\mathbb{P}$ be an irreducible Markov Chain and $\mathcal{S}$ a valid set of strategies against $\mathbb{P}$. The *$\mathcal{S}$-robust mixing time* and *$\mathcal{S}$-robust $L_2$-mixing time* are defined by the equations

$$R^{\mathcal{S}}(\mathbb{P}, \epsilon) = \sup_{\mathcal{P}} \mathcal{T}(\mathcal{P}, \epsilon) \quad \text{and} \quad R_2^{\mathcal{S}}(\mathbb{P}, \epsilon) = \sup_{\mathcal{P}} \mathcal{T}_2(\mathcal{P}, \epsilon)$$

where $\mathcal{P} = (\mathbb{P}, \{\mathbb{A}_t\}_{t>0})$ ranges over adversarially modified versions of $\mathbb{P}$ where $\mathbb{A}_t \in \mathcal{S}$ for all $t$. In case $\mathbb{P}$ is a Cayley walk on a group $G$, define the *holomorphic robust mixing time* and *holomorphic robust $L_2$-mixing time* by taking $\mathcal{S} = \mathcal{H}$. Similarly taking $\mathcal{S} = \mathcal{C}$ define *Cayley robust mixing time* and *Cayley robust $L_2$-mixing time.*

Theorem 1.8 as well as Theorem 1.9 can be extended to work with any valid set of strategies against $\mathbb{P}$. Hence we also have the following

**Theorem 1.20.** ***(Submultiplicativity for Cayley walks)*** *Let $\mathbb{P}$ be an ergodic Cayley walk on a group $G$ and $\mathbb{Q}$ any Cayley walk on $G$. For $\epsilon, \delta > 0$ and $R' \in \{R^{\mathcal{C}}, R^{\mathcal{H}}\}$, we have*

$$R'(\mathbb{P}, \epsilon\delta/2) \le R'(\mathbb{P}, \epsilon/2) + R'(\mathbb{P}, \delta/2) \quad \textit{and} \quad R_2'(\mathbb{P}, \epsilon\delta) \le R_2'(\mathbb{P}, \epsilon) + R_2(\mathbb{P}, \delta)$$

**Theorem 1.21.** ***(Convexity for Cayley walks)*** *Let $\mathbb{P}$ be an ergodic Cayley walk on a group $G$ and $\mathbb{Q}$ any Cayley walk on $G$. Let $0 < a = 1 - b < 1$ and $R' \in \{R^{\mathcal{C}}, R^{\mathcal{H}}\}$. Then $R'(\mathbb{P}\mathbb{Q}, 1/4) \le R'(\mathbb{P}, 1/4)$ and $R_2'(\mathbb{P}\mathbb{Q}, 1/2) \le R_2'\mathbb{P}, 1/2)$. Also,*

$$R'(a\mathbb{P} + b\mathbb{Q}, 1/4) \le R'(\mathbb{P}, 1/4) + R'(\mathbb{Q}, 1/4) - 1$$
$$R_2'(a\mathbb{P} + b\mathbb{Q}, 1/2) \le R_2'(\mathbb{P}, 1/2) + R_2'(\mathbb{Q}, 1/2) - 1$$

*Moreover,*

- *if* $R'(\mathbb{P}, 1/4) \geq 11$*, then* $R'(a\mathbb{P} + b\mathbb{Q}, 1/4) \leq 3R'(\mathbb{P}, 1/4)/a$
- *If* $|G| \geq 16$ *and* $R_2'(\mathbb{P}, 1/2) \geq \log(|G|)/2$ *then* $R_2'(a\mathbb{P} + b\mathbb{Q}) \leq 7R'(\mathbb{P})/a$

Submultiplicativity and Convexity for Cayley walks are proved in Section 2.1 and Section 2.3 respectively.

**Example 1.** Suppose $G$ is a finite group generated by $S \subseteq G$. Let $T \subseteq G$ be arbitrary. Let $\mathbb{P}_S$ and $\mathbb{P}_{S\cup T}$ denote the Cayley walks on $G$ driven by the uniform distributions on $S$ and $S \cup T$, respectively. Assume also that for $R' \in \{R, R^{\mathcal{H}}, R^{\mathcal{C}}\}$ we have $R'(\mathbb{P}_S) \geq 11$. Then writing

$$\mathbb{P}_{S\cup T} = \frac{|S|}{|S| + |T|}\mathbb{P}_S + \frac{|T|}{|S| + |T|}\mathbb{P}_T$$

allows us to apply Theorem 1.21 to infer

$$R'(\mathbb{P}_{S\cup T}, 1/4) \leq 3\frac{|S| + |T|}{|S|}R'(\mathbb{P}_S, 1/4) = 3\left(1 + \frac{|T|}{|S|}\right)R'(\mathbb{P}_S, 1/4)$$

Thus we can remove some problematic generators while estimating the robust mixing time.

**Definition 1.22.** Let $X$ be a connected undirected graph and $s$ a special vertex, called the sink.

- A **configuration** $f$ consists of a non-negative integer $f(u)$ associated with each non-sink vertex $u$ of $X$. Vertex $u$ is said to have $f(u)$ grains.
- A configuration $f$ is said to be **stable** iff $f(u) < d(u)$ for all vertices $u \neq s$, where $d(u)$ is the degree of $u$.
- Suppose $f$ is an unstable configuration and $u \neq s$ is such that $f(u) \geq d(u)$. By **toppling** $f$ **at** $u$ we mean removing $d(u)$ grains from $u$ and adding one grain to each **non-sink neighbor** $v$ of $u$.

Note that the total number of grains in the system reduces exactly when a neighbor of the sink topples. (5) shows that by repeatedly toppling an unstable configuration one can reach a stable configuration (since $X$ is connected). More over the final stable configuration is independent of the order in which vertices were toppled.

- Given two stable configurations $f$ and $g$, define $f \circ g$ to be the stable configuration obtained after toppling the configuration with $f(u) + g(u)$ grains at vertex $u$.
- A stable configuration $f$ is said to be **recurrent** if for any configuration $g$ there exists a configuration $h$ for which $f = g \circ h$.
- The set of all **stable recurrent configurations** form an abelian group, called the **Sandpile group of** $X$.

For details and proof of these facts see (5).

**Example 2.** The natural Markov Chain on the Sandpile group $G$ of a graph $X$ with sink $s$, is the following: Given a stable recurrent configuration $f$, pick a vertex $u \neq s$ and move to the configuration $f \circ e_u$, where $e_u$ is the configuration with exactly one grain at $u$.

If we take $X_n = K_{n+2}$ (with one of the vertices being the sink), the Sandpile group $G_n = \oplus_n C_n$ is the product of $n$-copies of the cyclic group $C_n$.

The $n+1$ generators of the natural Markov Chain on $G_n$ are $e_1, \dots, e_n$ and $f = -(e_1 + e_2 + \cdots + e_n)$, where $e_i$ is 1 at the $i$'th coordinate and zero elsewhere. Here taking $S = \{e_1, \dots, e_n\}$ and $T = \{f\}$, allows us to estimate the mixing time of this Markov Chain by eliminating the generator $f$.

Theorem 1.13 shows that $R(\mathbb{P})$ and $R_2(\mathbb{P})$ are determined up to a $\log(1/\pi_*)$ factor by the singular values of $\mathbb{S}(\mathbb{P})$. However, it turns out that $R_2^{\mathcal{C}}$ and $R_2^{\mathcal{H}}$ are within a factor of 2 of $\mathcal{T}_2(\mathbb{P}\overleftarrow{\mathbb{P}})$. In fact we have,

**Theorem 1.23.** *Let $\mathbb{P}$ denote an irreducible Cayley walk on a group $G$. Then*

$$\max(\mathcal{T}_2(\mathbb{P}), \mathcal{T}_2(\mathbb{P}\overleftarrow{\mathbb{P}})) \leq R_2^{\mathcal{C}}(\mathbb{P}) \leq R_2^{\mathcal{H}}(\mathbb{P}) \leq 2\mathcal{T}_2(\mathbb{P}\overleftarrow{\mathbb{P}})$$

*In particular if $\mathbb{P}$ is a reversible ergodic Cayley walk on a group $G$, we have*

$$\mathcal{T}_2(\mathbb{P}) \leq R_2^{\mathcal{C}}(\mathbb{P}) \leq R_2^{\mathcal{H}}(\mathbb{P}) \leq 2\mathcal{T}_2(\mathbb{P}^2) \leq \mathcal{T}_2(\mathbb{P}) + 1$$

Thus for a reversible ergodic Cayley walk, a holomorphic adversary cannot change the $L_2$-mixing time. Theorem 1.23 will be proved in Section 5.

## 1.5 Applications: Card Shuffling

In this section we give some applications of the foregoing results in this paper.

**Definition 1.24.** By the chain $\mathbb{P}_{\alpha\beta;\theta}$ we mean the card shuffling process on $S_n$ where we choose two positions $i$ and $j$ according to the rules $\alpha$ and $\beta$ respectively and apply the operation $\theta$ to the cards at positions $i$ and $j$. Possible values for $\alpha$ and $\beta$ are $R$ (for random); $A$ (for adversarial); $C$ (for cyclic) i.e. $t \pmod{n}$; $k$ (for some fixed value), $T$ (for top, same as 1). A rule $(\gamma/\delta)$ implies that we choose according to $\gamma$ and $\delta$ with prespecified probabilities (or equal if not specified).

Possible values for the operation $\theta$ are $I$ (insertion) to move the card at position $i$ to position $j$; $T$ (transpose) to exchange the cards at positions $i$ and $j$. Similarly an operation $(\gamma/\delta)$ implies apply the two operations with prespecified probabilities (or equal if not specified).

Note that the $\mathbb{P}_{\alpha\beta;\theta}$ process is not necessarily Markovian as the rules and the operation may depend on the time $t$. We start with an example of simulating a *non-Markovian process* by an adversarially modified Markov Chain.

**Proposition 1.25.** *Let $\mathbb{P}_{RT;I}$ denote the random-to-top move chain i. e., we pick a position $r$ uniformly at random and move the card in position $r$ to the top. Also let $\mathbb{P}_{RC;I}$ denote the random-to-cyclic insertion process, where we move a random card to position $t \pmod{n}$. Then $\mathbb{P}_{RC;I}$ process can be simulated by an adversarially modified $\mathbb{P}_{RT;I}$ chain.*

*Proof.* Suppose $k$ and $r$ are arbitrary positions. Moving the card in position $r$ to position $k$ is equivalent to moving it to the top and then moving the top card to position $k$. Hence it follows that an adversarially modified $\mathbb{P}_{RT;I}$ chain can simulate a $\mathbb{P}_{RC;I}$ chain. Also note that the adversary involved in the simulation is a Cayley adversary. $\square$

The non-Markovian process we are mainly interested in this section is the random-to-cyclic transposition process $\mathbb{P}_{RC;T}$. The problem of estimating the mixing time of $\mathbb{P}_{RC;T}$ was raised by Aldous and Diaconis (1) in 1986. Recently Mironov (16) used this shuffle to analyze a cryptographic system known as RC4 and showed that $\mathbb{P}_{RC;T}$ mixes in time $O(n \log n)$ without an estimate on the hidden constant. Mossel et al. (19) showed that $\mathcal{T}(\mathbb{P}_{RC;T}) = \Theta(n \log n)$. They showed a lower bound of $(0.0345 + o(1))n \log n$. They also generalized $\mathbb{P}_{RC;T}$ to $\mathbb{P}_{RA;T}$ (random-to-adversarial transposition) and showed that $\mathcal{T}(\mathbb{P}_{RA;T}) \leq Cn \log n + O(n)$ giving the first explicit bound of $C = 32\vartheta^3 + \vartheta \approx 4 \times 10^5$ where $\vartheta = 2\mathrm{e}^3/(\mathrm{e} - 1)$. They also observe that since $\mathbb{P}_{RA;T}$ can simulate $\mathbb{P}_{RT;T}$ the constant $C \geq 1$. We are able to reduce the upper bound on the mixing time of $\mathbb{P}_{RA;T}$ to $C = 1$.

**Theorem 1.26.** *Let $\mathbb{P}_{RC;T}$ denote the random-to-cyclic transposition chain, i.e. at time $t$ we exchange the cards at positions $r$ and $t \pmod n$ where $r \in \{1, \dots, n\}$ is chosen uniformly at random. Then $\mathcal{T}_2(\mathbb{P}_{RC;T}) \leq \mathcal{T}_2(\mathbb{P}_{RT;T}) + 1 \leq n \log n + O(n)$*

*Proof.* In fact, we prove the following chain of inequalities: $\mathcal{T}_2(\mathbb{P}_{RC;T}) \leq \mathcal{T}_2(\mathbb{P}_{RA;T}) \leq R_2^{\mathcal{C}}(\mathbb{P}_{RT;T}) \leq R_2^{\mathcal{H}}(\mathbb{P}_{RT;T}) \leq \mathcal{T}_2(\mathbb{P}_{RT;T}) + 1 \leq n \log n + O(n)$.

For a particular choice of adversarial moves the $\mathbb{P}_{RA;T}$ process can simulate the $\mathbb{P}_{RC;T}$ process. Hence $\mathcal{T}_2(\mathbb{P}_{RC;T}) \leq \mathcal{T}_2(\mathbb{P}_{RA;T})$.

By convexity arguments, it is enough to consider the case that the adversary's choice is deterministic to estimate the mixing time of $\mathbb{P}_{RA;T}$. Let $\alpha_t \in \{1, \dots, n\}$ denote an adversarial choice for time $t$ (fixed before the process begins). We first observe that an adversarial version of $\mathbb{P}_{RT;T}$ can simulate $\mathbb{P}_{RA;T}$. For $k, r \in \{1, \dots, n\}, (k\, r) = (1\, k)(1\, r)(1\, k)$. Hence if we let $\mathbb{A}_t$ correspond to right multiplication by $(1\, \alpha_t)(1\, \alpha_{t+1})$, it follows that the given adversarial modification of $\mathbb{P}_{RT;T}$ simulates $\mathbb{P}_{RA;T}$. Since the simulation was done by a Cayley adversary, we have $\mathcal{T}_2(\mathbb{P}_{RA;T}) \leq R_2^{\mathcal{C}}(\mathbb{P}_{RT;T}) \leq R_2^{\mathcal{H}}(\mathbb{P}_{RT;T})$.

From Theorem 1.23 it follows that $R_2^{\mathcal{H}}(\mathbb{P}_{RT;T}) \leq \mathcal{T}_2(\mathbb{P}_{RT;T}) + 1$ since $\mathbb{P}_{RT;T}$ is reversible. But $\mathcal{T}_2(\mathbb{P}_{RT;T}) \leq n \log n + O(n)$ ((7; 20)). $\square$

Another application is in estimating the mixing time of a mixture of two reversible Cayley walks on a group $G$.

**Theorem 1.27.** *Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be two reversible ergodic Cayley walks on a group $G$ and put $\mathbb{Q} = a_1\mathbb{P}_1 + a_2\mathbb{P}_2$ where $0 < a_1 = 1 - a_2 < 1$. Then assuming $\mathcal{T}_2(\mathbb{P}_i) \geq \log(|G|)/2$ for $i = 1, 2$ and $|G| \geq 16$, we have*

$$\mathcal{T}_2(\mathbb{Q}) \leq 1 + \min\left(\frac{7\mathcal{T}_2(\mathbb{P}_1)}{a_1}, \frac{7\mathcal{T}_2(\mathbb{P}_2)}{a_2}, \mathcal{T}_2(\mathbb{P}_1) + \mathcal{T}_2(\mathbb{P}_2)\right)$$

*Proof.* Since the $\mathbb{P}_i$ are reversible, Theorem 1.23 implies $\mathcal{T}_2(\mathbb{P}_i) \leq R_2^{\mathcal{H}}(\mathbb{P}_i) \leq 2\mathcal{T}_2(\mathbb{P}_i^2) \leq \mathcal{T}_2(\mathbb{P}_i) + 1$. Similarly, we have $\mathcal{T}_2(\mathbb{Q}) \leq R_2^{\mathcal{H}}(\mathbb{Q}) \leq \mathcal{T}_2(\mathbb{Q}) + 1$. From Theorem 1.21, we have

$$R_2^{\mathcal{H}}(\mathbb{Q}) \leq \min\left(R_2^{\mathcal{H}}(\mathbb{P}_1) + R_2^{\mathcal{H}}(\mathbb{P}_2) - 1, 7R_2^{\mathcal{H}}(\mathbb{P}_1)/p, 7R_2^{\mathcal{H}}(\mathbb{P}_2)/q\right)$$ $\square$

Note that using standard results, one can only show $\mathcal{T}_2(\mathbb{Q}) \leq \min(\mathcal{T}_2(\mathbb{P}), \mathcal{T}_2(\mathbb{Q}))O(\log(|G|))$. Thus we have eliminated the pesky $\log|G|$ factor, which can be significant since usually $|G|$ is exponential.

We finish this section with a few examples of estimating the mixing time of complex card shuffling chains in terms of that of simpler ones. Given card positions $i$ and $j$, moving the card at position $j$ to position $i$ corresponds to right multiplying by the cycle $C_{ij} = (i, i+1, \ldots, j)$ if $i < j$ and $C_{ij} = C_{ji}^{-1}$ if $i > j$. If $i = j$, $C_{ij}$ is the identity permutation.

**Example 3.** "**Move a or b chain**" Fix $1 \leq a, b \leq n$ and $0 \leq q = 1 - p \leq 1$. Let $R' \in \{R, R_2, R^{\mathcal{H}}, R_2^{\mathcal{H}}, R^{\mathcal{C}}, R_2^{\mathcal{C}}\}$. Consider the following random-to-a-or-b move chain given by

$$\mathbb{P}_{R(a/b);I} = p\mathbb{P}_{Ra;I} + q\mathbb{P}_{Rb;I}$$

i.e. we either choose a random card and move it to either position $a$ with probability $p$ or position $b$ with the remaining probability.

Observe that $R'(\mathbb{P}_a) = R'(\mathbb{P}_b)$ because an adversary for one can simulate that for the other. Hence $R'(\mathbb{P}_a) = R'(\mathbb{P}_1) = R'(\mathbb{P}_{RT;I})$. Hence Theorem 1.9 implies $R'(\mathbb{P}_{R(a/b);I})$ is bounded by $2R'(\mathbb{P}_{RT;I}) - 1$. Also $R'(\mathbb{P}_{RT;I}) \leq R_2^{\mathcal{H}}(\mathbb{P}_{RT;I})$ and $\mathbb{P}_{RT;I}\overleftarrow{\mathbb{P}_{RT;I}} = \mathbb{P}_{RR;I}$. Hence by Theorem 1.23, we have $R_2^{\mathcal{H}}(\mathbb{P}_{R(a/b);I}) \leq 4\mathcal{T}_2(\mathbb{P}_{RR;I})$. Corollary A.4 shows $\mathcal{T}_2(\mathbb{P}_{RR;I}) \leq 1.5n\log n + O(n)$, giving a bound of $6n\log n + O(n)$ for $R_2^{\mathcal{H}}(\mathbb{P}_{R(a/b);I})$.

**Example 4. "Transpose or Move chain"** Let $\mathbb{P}_{RR;(T/M)}$ be the transpose or move chain, i.e. we pick two positions $i$ and $j$ at random. We transpose the selected cards with probability $1/2$ and move card at position $i$ to position $j$ with probability $1/2$. Note that $\mathcal{T}_2(\mathbb{P}_{RR;T}) \leq 0.5n\log n + O(n)$ (10), $\mathcal{T}_2(\mathbb{P}_{RR;I}) \leq 1.5n\log n + O(n)$ (Corollary A.4), and that both are reversible chains. Hence we have have $\mathcal{T}_2(\mathbb{P}_{RR;(T/M)}) \leq 2n\log n + O(n)$.

### 1.6 Applications: Reversible lifting

**Definition 1.28.** Let $\mathbb{P}$ and $\mathbb{Q}$ be Markov Chains on state spaces $\mathcal{X}$ and $\mathcal{Y}$ with stationary distributions $\pi$ and $\mu$ respectively. $\mathbb{P}$ is said to be a *collapsing* of $\mathbb{Q}$ if there exists a mapping $f : \mathcal{Y} \to \mathcal{X}$ such that the following hold:

- $\pi(x) = \mu(\mathcal{Y}_x)$ for all $x \in \mathcal{X}$ where $\mathcal{Y}_x = f^{-1}(x)$
- For all $x_1, x_2 \in \mathcal{X}$,
$$\mathbb{P}(x_1, x_2) = \sum_{y_1 \in \mathcal{Y}_{x_1}} \sum_{y_2 \in \mathcal{Y}_{x_2}} \mu^{x_1}(y_1)\mathbb{Q}(y_1, y_2) \tag{5}$$
where $\mu^x$ is the conditional distribution of $y \in \mathcal{Y}$ given $f(y) = x$, i.e. $\mu^x(y) = \mu(y)/\pi(x)$.

A *lifting* of $\mathbb{P}$ is a chain $\mathbb{Q}$ such that $\mathbb{P}$ is the collapsing of $\mathbb{Q}$.

Chen et al. (4) showed that if $\mathbb{Q}$ is a reversible lifting of a Markov chain $\mathbb{P}$, then $\mathcal{T}(\mathbb{Q}) \geq \mathcal{T}(\mathbb{P})/\log(1/\pi_*)$. We give an alternate proof of the same result which is motivated by adversarial strategies. The crucial observation is the following

$$\begin{pmatrix} 1-\upsilon & \upsilon & 0\dots0 \\ 1 & 0 & 0\dots0 \\ 0 & 0 & I \end{pmatrix} \tag{6}$$

The states are indexed starting with $y = \text{argmax}_x \pi(x)$ and the current state $z \neq y$. Here $\upsilon = \pi(z)/\pi(y)$.

Figure 1: An adaptive adversary is unreasonably powerful

**Theorem 1.29.** *Let $\mathbb{Q}$ be a lifting of $\mathbb{P}$. Then $R(\mathbb{Q}) \geq \mathcal{T}(\mathbb{P})$.*

If $\mathbb{Q}$ is reversible, Theorem 1.13 implies that $R(\mathbb{Q}) \leq \mathcal{T}(\mathbb{Q})(1 + \log(1/\mu_*)/2)$, where $\mu_* = \min_y \mu(y)$. This immediately gives

**Corollary 1.30.** *Let $\mathbb{Q}$ be a reversible Markov Chain with stationary distribution $\mu$ and $\mathbb{P}$ a collapsing of $\mathbb{Q}$ with stationary distribution $\pi$. Then $\mathcal{T}(\mathbb{Q})\log(1/\mu_*) \geq \mathcal{T}(\mathbb{P})$.*

When $\mu_*$ is only polynomially smaller than $\pi_*$, we have an alternate proof of the reversible lifting result. In order to fine tune the result, we look at the adversarial strategy used in the proof of Theorem 1.29 more closely and prove

**Theorem 1.31.** *Let $\mathbb{Q}$ be a reversible lifting of $\mathbb{P}$. Then $\mathcal{T}(\mathbb{Q}) \geq r(\mathbb{P})$*

This not only gives an alternate proof of the reversible lifting result of (4), it also shows that when $\mathcal{T}(\mathbb{P}) = O(r(\mathbb{P}))$ no reversible lifting $\mathbb{Q}$ of $\mathbb{P}$ can mix faster than $\mathbb{P}$ (ignoring constant factors).

### 1.7 Discussion: The power of the adversary

We discuss the necessity of our restrictions on the adversary.

Our definition of robust mixing time, requires that the adversary be *oblivious*, i. e., announce all his moves in advance. An *adaptive* adversary would be unreasonably powerful as shown below.

Let $y$ be the state with maximum stationary probability and suppose we allow an adaptive adversary. The adversary can ensure that the chain is always at state $y$ at the end of its turn as follows: Suppose the current state of the chain in $z$. If $z = y$ the adversary does not do any anything. Otherwise the adversary applies the stochastic matrix given in Figure 1. It is easily checked that this matrix is compatible with $\mathbb{P}$ and sends $z$ to $y$ with probability 1. For this reason we do not consider an adaptive adversary.

The requirement that the adversary's choices are compatible with $\mathbb{P}$ ensures that an AMMC process has a hope to have a limit.

## 2 Basic Properties

### 2.1 Submultiplicativity

In this section we establish some basic properties of $R$. We start by proving sub-multiplicativity of $R(\epsilon/2)$. Since this is a basic property we give two different proofs. Both proofs work for the standard mixing time as well. The second proof is new as far as we know, and simplifies the proof for standard mixing time as well. Apart from these two proofs, the argument based on coupling can also be extended to handle the robust case.

**Theorem 2.1.** *Let $\mathbb{P}$ be an ergodic Markov Chain and $\epsilon, \delta > 0$. Let $\mathcal{S}$ be a valid set of strategies against $\mathbb{P}$. Then $R^{\mathcal{S}}(\epsilon\delta/2) \leq R^{\mathcal{S}}(\epsilon/2) + R^{\mathcal{S}}(\delta/2)$.*

*Proof.* This proof is based on the standard proof using the triangle inequality. Let $S = R^{\mathcal{S}}(\epsilon/2)$ and $T = R^{\mathcal{S}}(\delta/2)$. Let $\mathbb{A}_{s,t}$ denote a sequence of $t - s + 1$ stochastic matrices in $\mathcal{S}$.

By convexity it is enough to consider initial distribution concentrated on a single point. For $x \in \mathcal{X}$, let $\delta_x$ denote the distribution concentrated on $x$. Define the following quantities:

$$\begin{aligned}
\Delta(x, t, \mathbb{A}_{1,t}) &= ||\mu_t - \pi||_{TV} \\
\Delta(x, t, \cdot) &= \max_{\mathbb{A}_{\leq t}} \Delta(x, t, \mathbb{A}_{1,t}) \\
\Delta(\cdot, t, \mathbb{A}_{1,t}) &= \max_{x \in \mathcal{X}} \Delta(x, t, \mathbb{A}_{1,t}) \\
\Delta(t) = \Delta(\cdot, t, \cdot) &= \max_{x} \Delta(x, t, \cdot)
\end{aligned}$$

where $\mu_t$ is the $t$-step distribution using the adversarial strategy $\mathbb{A}_{1,t}$ and initial distribution $\delta_x$. Now fix some adversarial strategy $\{\mathbb{A}_t\}_t$ from $\mathcal{S}$. Define $\mathbb{C} = \mathbb{P}\mathbb{A}_1\mathbb{P}\mathbb{A}_2 \ldots \mathbb{P}\mathbb{A}_S$ and $\mathbb{D} = \mathbb{P}\mathbb{A}_{S+1}\mathbb{P}\mathbb{A}_{S+2} \ldots \mathbb{P}\mathbb{A}_{S+T}$. Then

$$\begin{aligned}
2\Delta(x, S+T, \mathbb{A}_{1,S+T}) &= \sum_y |(\mathbb{C}\mathbb{D})(x, y) - \pi(y)| \\
&= \sum_y \left| \sum_z [\mathbb{C}(x, z)\mathbb{D}(z, y) - \pi(z)\mathbb{D}(z, y)] \right| \\
&= \sum_y \left| \sum_z [\mathbb{C}(x, z) - \pi(z)] [\mathbb{D}(z, y) - \pi(y)] \right| \\
&\leq \sum_z \sum_y |\mathbb{C}(x, z) - \pi(z)| \, |\mathbb{D}(z, y) - \pi(y)| \\
&= \sum_z |\mathbb{C}(x, z) - \pi(z)| \, 2\Delta(z, T, \mathbb{B}_{S+1,S+T}) \\
&\leq 2\Delta(\cdot, T, \mathbb{A}_{S+1,S+T}) \sum_z |\mathbb{C}(x, z) - \pi(z)| \\
&\leq 2\Delta(\cdot, T, \mathbb{A}_{S+1,S+T}) 2\Delta(x, S, \mathbb{A}_{1,S}) \\
&\leq 2\Delta(\cdot, T, \cdot) 2\Delta(x, S, \cdot)
\end{aligned}$$

Taking the maximum over all strategies $\{\mathbb{A}_t\}_t$ we have $\Delta(x, S+T, \cdot) \leq 2\Delta(x, S, \cdot)\Delta(\cdot, T, \cdot)$. Now taking maximum over all $x$ we have $2\Delta(S + T) \leq (2\Delta(S))(2\Delta(T))$. Using $R(\epsilon/2) \leq T \iff 2\Delta(T) \leq \epsilon$ gives the result. $\square$

Before we get to the new proof we need the following

**Lemma 2.2.** *Let $\mathbb{Q}$ be a stochastic matrix for which $\pi\mathbb{Q} = \pi$. Suppose we know that for all initial distributions $\mu$, $||\mu\mathbb{Q} - \pi||_{TV} < \epsilon$. Then for all initial distribution $\mu$,*

$$||\mu\mathbb{Q} - \pi||_{TV} \leq \epsilon \min(1, 2||\mu - \pi||_{TV}) \tag{7}$$

*Proof.* Let $\eta = ||\mu - \pi||_{TV}$. It follows that we can write $\mu - \pi = \eta(\nu_1 - \nu_2)$ for appropriate distributions $\nu_1, \nu_2$. Then

$$||\mu\mathbb{Q} - \pi||_{TV} = ||(\mu - \pi)\mathbb{Q}||_{TV} \leq \eta||\nu_1\mathbb{Q} - \pi||_{TV} + \eta||\nu_2\mathbb{Q} - \pi||_{TV} \leq 2\eta\epsilon = 2\epsilon||\mu - \pi||_{TV} \quad \square$$

The same proof using $|| \cdot ||_{2,\pi}$ instead of $|| \cdot ||_{TV}$ gives

**Corollary 2.3.** *Let $\mathbb{Q}$ be a stochastic matrix for which $\pi\mathbb{Q} = \pi$. Suppose we know that for all initial distributions $\mu$, $||\mu\mathbb{Q} - \pi||_{2,\pi} < \epsilon$. Then for all initial distributions $\mu$,*

$$||\mu\mathbb{Q} - \pi||_{2,\pi} \leq \epsilon \min(1, 2||\mu - \pi||_{TV}) \leq \epsilon \min(1, ||\mu - \pi||_{2,\pi}) \tag{8}$$

*Alternate proof of Theorem 2.1.* Let $S = R^{\mathcal{S}}(\epsilon/2)$ and $T = R^{\mathcal{S}}(\delta/2)$. Let $\mathbb{A}_1, \ldots, \mathbb{A}_S, \mathbb{A}_{S+1}, \ldots, \mathbb{A}_{S+T}$ be a sequence of stochastic matrices from $\mathcal{S}$. Let $\mathbb{C} = \mathbb{P}\mathbb{A}_1\mathbb{P}\mathbb{A}_2 \ldots \mathbb{P}\mathbb{A}_S$ and $\mathbb{D} = \mathbb{P}\mathbb{A}_{S+1}\mathbb{P}\mathbb{A}_{S+2} \ldots \mathbb{P}\mathbb{A}_{S+T}$.

From the lemma above, we have

$$||\mu\mathbb{C}\mathbb{D} - \pi||_{TV} \leq (\delta/2)2||\mu\mathbb{C} - \pi||_{TV} \leq \delta\epsilon/2 \tag{9}$$

Since $\mathbb{A}_i$ were arbitrary we have the result. $\square$

**Theorem 2.4.** *Let $\mathbb{P}$ be an ergodic Markov Chain and $\epsilon > 0$. Let $\mathcal{S}$ be a valid set of stochastic matrices compatible with $\mathbb{P}$. Then the following are sub-multiplicative: $R_2^{\mathcal{S}}(\epsilon), R^{\mathcal{S}}(\epsilon/2)$. Moreover for $k > 1$, we also have $R_2^{\mathcal{S}}(\epsilon^k) \leq R_2^{\mathcal{S}}(\epsilon) + (k-1)R^{\mathcal{S}}(\epsilon/2)$.*

*Proof.* Theorem 2.1 shows that $R^{\mathcal{S}}(\epsilon/2)$ is sub multiplicative. Replacing application of Lemma 2.2 with Corollary 2.3, in the proof of Theorem 2.1 shows that $R_2^{\mathcal{S}}(\epsilon)$ is sub multiplicative.

For the last part: Let $T_1 = R^{\mathcal{S}}(\epsilon/2)$ and $T_2 = R_2^{\mathcal{S}}(\epsilon)$. Then for any initial distribution $\mu_0$, we have $||\mu_{(k-1)T_1} - \pi||_{TV} \leq \epsilon^{k-1}/2$ (by submultiplicativity for total variation distance). Now using the tighter inequality in Corollary 2.3, we have $||\mu_{(k-1)T_1+T_2} - \pi||_{2,\pi} \leq 2\epsilon||\mu_{(k-1)T} - \pi||_{TV} \leq \epsilon^k$. $\square$

The second part can be useful in obtaining better non-asymptotic bounds for $L_2$ mixing when the total variation mixing time is a lot smaller than the $L_2$ mixing time.

## 2.2 Finiteness

Now we characterize chains with finite robust mixing time. First we observe that ergodicity is not enough to guarantee finiteness of $R$ as shown by the following examples.

**Example 5. "walk on directed edges":** Let $X$ be a connected $d$-regular undirected graph. Assume that the usual random walk on $X$ is ergodic. Consider the same walk except that this time we also keep track of the previous vertex. Thus the states of this walk are directed edges $(u, v)$ of $X$. If at time $t$ we are at $(u, v)$ we move to $(v, w)$ where $w$ is a uniformly chosen neighbor of $v$.

Since the transition rule doesn't care which vertex we came from, it follows that the new walk is ergodic exactly when the usual random walk on $X$ is. Also the mixing time of the new walk is bounded by one more than that of the usual walk. This is because picking a random neighbor of a random vertex of a regular graph is equivalent to picking a random directed edge.

Consider the following adversarial strategy $\mathbb{A}$: $\mathbb{A}((u, v)) = (v, u)$. The adversary simply reverses the direction of the edge. Let $v$ be any vertex of $X$ and let $\mu_0$ be the uniform distribution on all edges coming into $v$. Then $\nu_1$ is the uniform distribution of all edges going out of $v$. Applying the adversary's strategy we get $\mu_1 = \mu_0$. Thus $R(\mathbb{P}) = \infty$.

**Example 6. "Bottom $k$ to top shuffles":** Let $\mathbb{P}_k$ denote the following Markov Chain on $S_n$. Given a pack of $n$ cards, we pick a random card among the bottom $k$ cards (where $1 \le k \le n$) and move it to the top. Unless $k = n$, $R(\mathbb{P}_k) = \infty$. The adversarial strategy of exchanging the top two cards, ensures that the top card of the deck stays fixed (if $k < n$).

(12) defined "Bottom $k$ to top shuffles" and showed that it mixes in time between $O(n \log n)$ and $O(n^3 \log n)$ as $k$ varies from $n$ down to 2. Note that $k = n$ is the Random to Top transposition shuffle and $k = 2$ is related to the Rudvalis shuffle.

**Theorem 2.5.** *Let $\mathbb{P}$ be an ergodic Markov Chain with stationary distribution $\pi$.*

- *$R(\epsilon)$ is finite for all $\epsilon > 0$ iff $\sigma_1(\mathbb{S}(\mathbb{P})) < 1$.*
- *if $\mathbb{P}$ has all holding probabilities positive, then $\sigma_1(\mathbb{S}(\mathbb{P})) < 1$.*
- *If $\sigma_1(\mathbb{S}(\mathbb{P})) < 1$, then $2||\mu_t - \pi||_{TV} \le \sigma_1(\mathbb{S}(\mathbb{P}))^t \sqrt{\frac{1}{\pi_*}}$*

*Proof.* It is easy to see that $\overleftarrow{\mathbb{P}}$ is also a Markov Chain with same stationary distribution as $\mathbb{P}$. Hence if we put $\mathbb{A}_t = \overleftarrow{\mathbb{P}}$, we see that $R(\mathbb{P}, \epsilon) \ge \mathcal{T}(\mathbb{P}\overleftarrow{\mathbb{P}}, \epsilon)$.

Suppose $\sigma_1(\mathbb{S}(\mathbb{P})) = 1$. Observe that the singular values of $\mathbb{S}(\mathbb{P})$ are just the eigenvalues of $\mathbb{S}(\mathbb{P})\mathbb{S}(\mathbb{P})^T = \mathbb{S}(\mathbb{P}\overleftarrow{\mathbb{P}})$. But $\mathbb{S}(\mathbb{P}\overleftarrow{\mathbb{P}})$ and $\mathbb{P}\overleftarrow{\mathbb{P}}$ are similar matrices. Hence it follows that $\mathbb{P}\overleftarrow{\mathbb{P}}$ is not ergodic and hence $R(\mathbb{P}) \ge \mathcal{T}(\mathbb{P}\overleftarrow{\mathbb{P}}) = \infty$.

On the other hand if $\sigma_1(\mathbb{S}(\mathbb{P})) < 1$, it follows that

$$||\mu\mathbb{P}\mathbb{A} - \pi||_{2,\pi} = ||(\mu - \pi)\mathbb{P}\mathbb{A}||_{2,\pi} \le ||(\mu - \pi)\mathbb{P}||_{2,\pi} \le \sigma_1 ||\mu - \pi||_{2,\pi} \tag{10}$$

for an arbitrary distribution $\mu$ and an arbitrary $\mathbb{P}$ compatible $\mathbb{A}$ and $||\boldsymbol{x}||_{2,\pi} = \sum_i x_i^2/\pi_i$. Since the worst case initial distribution is a point distribution, one can check that $||\mu - \pi||_{2,\pi} \le$

$\sqrt{1/\pi_*}$ for any distribution $\mu$. Finally using the fact that $2||\mu - \pi||_{TV} \leq ||\mu - \pi||_{2,\pi}$, we have $2||\mu_t - \pi||_{TV} \leq \sigma_1(\mathbb{S}(\mathbb{P}))^t \sqrt{1/\pi_*}$.

If $\mathbb{P}$ has all holding probabilities positive, write $\mathbb{P} = aI + (1-a)\mathbb{Q}$ for some $a > 0$ and observe that

$$||\boldsymbol{x}\mathbb{S}(\mathbb{P})||_2 \leq a||\boldsymbol{x}||_2 + (1-a)||\boldsymbol{x}\mathbb{S}(\mathbb{Q})||_2 \leq ||x||_2 \tag{11}$$

Thus $||\boldsymbol{x}\mathbb{S}(\mathbb{P})||_2 = ||\boldsymbol{x}||_2 \implies \boldsymbol{x}\mathbb{S}(\mathbb{P}) = \boldsymbol{x}$ and hence ergodicity of $\mathbb{P}$ implies $\sqrt{\Pi}\boldsymbol{x}$ is a multiple of $\pi$. Hence $\sigma_1(\mathbb{S}(\mathbb{P})) < 1$. $\square$

In case $\mathbb{P}$ has uniform distribution (hence $S(\mathbb{P}) = \mathbb{P}$) and $\sigma_1(\mathbb{P}) = 1$, we can easily construct an adversarial strategy as follows. Let $\boldsymbol{\alpha}$ denote the left singular vector corresponding to $\sigma_1(\mathbb{P})$ and $\boldsymbol{\beta} = \boldsymbol{\alpha}\mathbb{P}$. Since $\mathbb{P}$ is a convex combination of permutation matrices and $||\boldsymbol{\beta}||_2 = ||\boldsymbol{\alpha}||_2$ it follows that $\boldsymbol{\beta}$ is a permutation of $\boldsymbol{\alpha}$. Let $I = \{x \in \mathcal{X} \colon \boldsymbol{\alpha}(x) \geq \boldsymbol{\alpha}(y)\, \forall y \in \mathcal{X}\}$, i. e., states $x$ where $\boldsymbol{\alpha}(x)$ attains its maximum. Similarly let $J \subseteq \mathcal{X}$ be the set of states $x$ where $\boldsymbol{\beta}(x)$ attains its maximum. Since $\boldsymbol{\beta}$ is a permutation of $\boldsymbol{\alpha}$ it follows that $\mathbb{P}$ maps the uniform distribution on $I$ to the uniform distribution of $J$ (consider the initial distribution $\pi + \epsilon\boldsymbol{\alpha}$ for a small enough $\epsilon$). Since $\boldsymbol{\alpha}$ is not the constant vector, $I$ and $J$ are non-trivial subsets of $\mathcal{X}$. Hence the adversary can choose a permutation on $\mathcal{X}$ which maps $J$ to $I$ and the initial distribution is taken to be uniform on $I$.

We now prove Theorem 1.13.

*Proof of Theorem 1.13.* Let $\sigma = \sigma_1(\mathbb{S}(\mathbb{P}))$. The case $\sigma = 1$ is easy, so we assume $\sigma < 1$. By considering the constant adversarial strategies $I$ and $\overleftarrow{\mathbb{P}}$, we have $\max(\mathcal{T}(\mathbb{P}\overleftarrow{\mathbb{P}}), \mathcal{T}(\mathbb{P})) \leq R(\mathbb{P})$. Consider the reversible Markov chain $\mathbb{Q} = \mathbb{P}\overleftarrow{\mathbb{P}}$ with second largest eigenvalue $\sigma^2$. Since $\mathbb{Q}$ is reversible, standard results imply $r(\mathbb{Q}) \leq \mathcal{T}(\mathbb{Q})$. By definition of relaxation time, we have

$$\sigma^{2r(\mathbb{Q})} = (\sigma^2)^{r(\mathbb{Q})} \leq 1/\mathrm{e}$$

Taking $t = 2kr(\mathbb{Q})$ in Theorem 2.5, gives

$$2||\mu_{2kr(\mathbb{Q})} - \pi||_{TV} \leq \frac{\mathrm{e}^{-k}}{\sqrt{\pi_*}}$$

Put $k = \log(1/\pi_*)/2 + 1$ to ensure $2||\mu_{kr(\mathbb{P})} - \pi||_{TV} \leq 1/\mathrm{e} < 1/2$. This gives the first half of the result.

If $\mathbb{P}$ is reversible then $\mathbb{S}(\mathbb{P})$ is symmetric and hence $\sigma_1(\mathbb{S}(\mathbb{P})) = \lambda_*(\mathbb{S}(\mathbb{P})) = \lambda_*(\mathbb{P})$. Clearly $R(\mathbb{P}) \geq \mathcal{T}(\mathbb{P}) \geq r(\mathbb{P})$. By definition of $r(\mathbb{P})$, we have $\lambda_*^{r(\mathbb{P})} \leq 1/\mathrm{e}$. Now substitute $t = r(\mathbb{P})(\log(1/\pi_*)/2 + 1)$ in Theorem 2.5 to conclude that $R(\mathbb{P}) \leq r(\mathbb{P})(\log(1/\pi_*)/2 + 1)$. $\square$

The tightness of this inequality is discussed at the end of Section 3.

## 2.3 Convexity

We now prove Theorem 1.9 and Theorem 1.21.

**Lemma 2.6.** *Let $\mathbb{P}$ be an irreducible Markov Chain and $\mathcal{S}$ a valid set of strategies against $\mathbb{P}$. If $\mathbb{Q} \in \mathcal{S}$, then $R'(\mathbb{P}\mathbb{Q}) \le R'(\mathbb{P})$ for $R' \in \{R^{\mathcal{S}}, R_2^{\mathcal{S}}\}$. In particular, if $\mathbb{P}$ is an irreducible Cayley walk on a group $G$ and $\mathbb{Q}$ any random walk on $G$ (not necessarily irreducible). Then $R'(\mathbb{P}\mathbb{Q}) \le R'(\mathbb{P})$ for $R' \in \{R^{\mathcal{C}}, R^{\mathcal{H}}, R_2^{\mathcal{C}}, R_2^{\mathcal{H}}\}$.*

*Proof.* Let $\mathcal{P} = (\mathbb{P}\mathbb{Q}, \{\mathbb{A}_t\}_{t>0})$ be any adversarially modified version of $\mathbb{P}\mathbb{Q}$ where $\mathbb{A}_t \in \mathcal{S}$. Then $\mathcal{P}' = (\mathbb{P}, \{\mathbb{Q}\mathbb{A}_t\}_{t>0})$ is an adversarially modified version of $\mathbb{P}$ where $\mathbb{Q}\mathbb{A}_t \in \mathcal{S}$ since $\mathbb{Q} \in \mathcal{S}$ and $\mathcal{S}$ is closed under products. Moreover the mixing times of $\mathcal{P}$ and $\mathcal{P}'$ are equal. Taking supremum over $\mathcal{P}$ we have the result.

For the case of Cayley walks, we just observe that the transition matrices of all Cayley walks belong to $\mathcal{C} \subseteq \mathcal{H}$. □

We now show that the robust mixing time of a convex combination of Markov Chains can be bounded in terms of that of the participating chains.

Let $\mathbb{P}$ and $\mathbb{Q}$ be two irreducible Markov Chains with same stationary distribution $\pi$. Suppose $\mathcal{S}$ is valid set of strategies against $\mathbb{P}$ and $\mathbb{Q}$. Also assume $\mathbb{P} \in \mathcal{S}$ and $\mathbb{Q} \in \mathcal{S}$. Fix $0 < a = 1 - b < 1$ and consider the chain $a\mathbb{P} + b\mathbb{Q}$. Let $\mathcal{P} = (a\mathbb{P} + b\mathbb{Q}, \{\mathbb{A}_t\}_{t>0})$ be any adversarially modified version of $a\mathbb{P} + b\mathbb{Q}$. Fix $S > 0$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_S)$ where $\epsilon_i \in \{0, 1\}$. Define the following quantities:

- $\mathbb{P}^{(0)} = \mathbb{Q}$, $\mathbb{P}^{(1)} = \mathbb{P}$
- $\xi(\boldsymbol{\epsilon}) = \prod_{i=1}^{S} \mathbb{P}^{(\epsilon_i)}\mathbb{A}_i$
- $H(\boldsymbol{\epsilon}) = \sum_{i=1}^{S} \epsilon_i$
- $w(\boldsymbol{\epsilon}) = \prod_{i=1}^{S} a^{\epsilon_i} b^{1-\epsilon_i} = a^{H(\boldsymbol{\epsilon})} b^{S-H(\boldsymbol{\epsilon})}$

If $\mu_0$ is any initial distribution, and $\mu_S$ is the distribution after $S$ rounds, we have

$$\mu_S - \pi = \sum_{\boldsymbol{\epsilon}} w(\boldsymbol{\epsilon}) \left(\mu_0 \xi(\boldsymbol{\epsilon}) - \pi\right)$$

where the sum ranges over all $2^S$ choices for $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_S)$.

**Lemma 2.7.** *Let $\mathbb{P}$ and $\mathbb{Q}$ be ergodic Markov Chains with the same stationary distribution. Let $\mathcal{S}$ be a valid set of strategies against both $\mathbb{P}$ and $\mathbb{Q}$ and assume that $\mathbb{P} \in \mathcal{S}$ and $\mathbb{Q} \in \mathcal{S}$. Let $0 < a = 1 - b < 1$. Then for $R' \in \{R^{\mathcal{S}}, R_2^{\mathcal{S}}\}$, $R'(a\mathbb{P} + b\mathbb{Q}) \le R'(\mathbb{P}) + R'(\mathbb{Q}) - 1$. In particular, if $\mathbb{P}$ and $\mathbb{Q}$ are ergodic Cayley walks on a group $G$, then we have $R'(a\mathbb{P} + b\mathbb{Q}) \le R'(\mathbb{P}) + R'(\mathbb{Q}) - 1$ for $R' \in \{R^{\mathcal{H}}, R^{\mathcal{C}}, R_2^{\mathcal{H}}, R_2^{\mathcal{C}}\}$.*

*Proof.* Choose $S = R^{\mathcal{S}}(\mathbb{P}) + R^{\mathcal{S}}(\mathbb{Q}) - 1$. Then we have

$$||\mu_S - \pi||_{TV} \le \sum_{\boldsymbol{\epsilon}} w(\boldsymbol{\epsilon}) ||\mu_0 \xi(\boldsymbol{\epsilon}) - \pi||_{TV}$$

Now for each $\boldsymbol{\epsilon}$, $\xi(\boldsymbol{\epsilon})$ either contains $\ge R^{\mathcal{S}}(\mathbb{P})$ occurrences of $\mathbb{P}$ or contains $\ge R^{\mathcal{S}}(\mathbb{Q})$ occurrences of $\mathbb{Q}$. The remaining matrices can be considered as an adversarial choice. Hence we have $||\mu_0 \xi(\boldsymbol{\epsilon}) - \pi||_{TV} \le 1/4$ for all $\boldsymbol{\epsilon}$. Hence $||\mu_S - \pi||_{TV} \le 1/4$.

Similarly, taking $S = R_2^{\mathcal{S}}(\mathbb{P}) + R_2^{\mathcal{S}}(\mathbb{Q}) - 1$, and looking at the $||\mu_S - \pi||_{2,\pi}$ we get $R_2^{\mathcal{S}}(a\mathbb{P} + b\mathbb{Q}) \leq R_2^{\mathcal{S}}(\mathbb{P}) + R_2^{\mathcal{S}}(\mathbb{Q}) - 1$.

In case $\mathbb{P}$ and $\mathbb{Q}$ are Cayley walks on $G$, we just observe that $\mathbb{P}$ and $\mathbb{Q}$ are valid choices for a Cayley adversary and hence also for a holomorphic adversary. □

Now we consider the case when $\mathbb{P}$ has finite robust mixing time and $\mathbb{Q}$ may not. We start with a concentration inequality.

**Lemma 2.8.** *Let $S = CT/p$ for $C > 1$ and $0 < p < 1$. Let $Z_1, \ldots, Z_S$ be independent Bernoulli random variables with $\Pr\{Z_i = 1\} = p$. Let $Z = \sum_i Z_i$. Then we have*

$$\log \Pr\{Z < T\} \leq -T\left((C-1) - \log C\right) \tag{12}$$

*Proof.* We use Hoeffding's inequality (13, Theorem 1), for $S - Z$ to conclude

$$\Pr\{Z < T\} \leq \left\{ \left(\frac{q}{q+a}\right)^{q+a} \left(\frac{p}{p-a}\right)^{p-a} \right\}^S$$

where $q = 1 - p, a = p - p/C$. After algebraic simplifications we get,

$$\Pr\{Z < T\} \leq \left\{ \left(\frac{q}{1 - \frac{p}{C}}\right)^{\frac{C}{p}-1} C \right\}^T$$

Taking logarithms, and using $\log(1-x) \leq -x$ for $0 < x < 1$ gives the result. □

**Lemma 2.9.** *Let $\mathbb{P}$ be an ergodic Markov Chain and $\mathcal{S}$ a valid set of strategies against $\mathbb{P}$. Fix $\mathbb{Q} \in \mathcal{S}$. Let $0 < a = 1 - b < 1$. Let $S = (1+\delta)R^{\mathcal{S}}(\mathbb{P}, \gamma)/a$, where $\gamma > 0$ and $\delta > 0$ are arbitrary. Then*

$$||\mu_S - \pi||_{TV} \leq \gamma + \exp\left(-R^{\mathcal{S}}(\mathbb{P}, \gamma) \cdot (\delta - \log(1+\delta))\right)$$

*Proof.* Let $S = (1+\delta)T/a$, where $T = R^{\mathcal{S}}(\mathbb{P}, \gamma)$. Write

$$\mu_S - \pi = \sum_{\boldsymbol{\epsilon}} w(\boldsymbol{\epsilon}) \left(\mu_0 \xi(\boldsymbol{\epsilon}) - \pi\right)$$

Put $\mathcal{D} = \{\boldsymbol{\epsilon} : H(\boldsymbol{\epsilon}) \geq T\}$, i.e. all choices of $\boldsymbol{\epsilon}$ which resulted in $\mathbb{P}$ being used at least $T$ times. For $\boldsymbol{\epsilon} \in \mathcal{D}$, we have $||\mu_0\xi(\boldsymbol{\epsilon}) - \pi||_{TV} \leq \gamma$. For $\boldsymbol{\epsilon} \notin \mathcal{D}$, $||\mu_0\xi(\boldsymbol{\epsilon}) - \pi||_{TV} \leq 1$.

We now estimate $\sum_{\boldsymbol{\epsilon} \notin \mathcal{A}} w(\boldsymbol{\epsilon})$. This is precisely $\Pr\{H(\boldsymbol{\epsilon}) < T\}$ where $\boldsymbol{\epsilon}$ is chosen at random with each coordinate taking a value 1 with probability $a$ and 0 with probability $b$. Note that $m = \mathbb{E}[H(\boldsymbol{\epsilon})] = (1+\delta)T$. Applying Lemma 2.8 we have

$$\sum_{\boldsymbol{\epsilon} \notin \mathcal{D}} w(\boldsymbol{\epsilon}) = \Pr\{H(\boldsymbol{\epsilon}) < m/(1+\delta)\} \leq \exp\left(-T(\delta - \log(1+\delta))\right)$$

Combining we have

$$\begin{aligned}
||\mu_S - \pi||_{TV} &\leq \sum_{\boldsymbol{\epsilon} \in \mathcal{D}} w(\boldsymbol{\epsilon}) ||\mu_0\xi(\boldsymbol{\epsilon}) - \pi||_{TV} + \sum_{\boldsymbol{\epsilon} \notin \mathcal{D}} w(\boldsymbol{\epsilon}) ||\mu_0\xi(\boldsymbol{\epsilon}) - \pi||_{TV} \\
&\leq \sum_{\boldsymbol{\epsilon} \in \mathcal{D}} w(\boldsymbol{\epsilon}) * \gamma + \sum_{\boldsymbol{\epsilon} \notin \mathcal{D}} w(\boldsymbol{\epsilon}) * 1 \\
&\leq \gamma + \exp\left(-T(\delta - \log(1+\delta))\right)
\end{aligned}$$ □

**Corollary 2.10.** *Let $\mathbb{P}$ be an ergodic Markov Chain and $\mathbb{Q}$ be compatible with $\mathbb{P}$. Let $\mathcal{S}$ be a valid set of strategies against $\mathbb{P}$. Assume $\mathbb{Q} \in \mathcal{S}$. Let $0 < a = 1 - b < 1$. Then $R^{\mathcal{S}}(a\mathbb{P} + b\mathbb{Q}) \leq 2(1+\delta)R^{\mathcal{S}}(\mathbb{P})/a$, as long as $2R^{\mathcal{S}}(\mathbb{P})(\delta - \log(1+\delta)) \geq \log 8$. If $R^{\mathcal{S}}(\mathbb{P}) \geq 11$, then $\delta$ may be taken to be $1/2$.*

*In particular, if $\mathbb{P}$ is an ergodic Cayley walk on the group $G$ and $\mathbb{Q}$ is any Cayley walk on $G$, the same conclusion holds with $R^{\mathcal{S}}$ replaced by $R^{\mathcal{H}}$ and $R^{\mathcal{C}}$.*

*Proof.* Let $T = R^{\mathcal{S}}(\mathbb{P})$ and $S = 2T(1+\delta)/a$. By submultiplicativity, we have $R^{\mathcal{S}}(\mathbb{P}, \gamma) \leq 2T$ for $\gamma = 1/8$. Lemma 2.9 now gives

$$||\mu_S - \pi||_{TV} \leq 1/8 + \exp\left(-2T(\delta - \log(1+\delta))\right)$$

If $2T(\delta - \log(1+\delta)) \geq \log 8$, we have $||\mu_S - \pi||_{TV} \leq 1/8 + 1/8 = 1/4$ as required. $\square$

Similarly for the $\mathcal{S}$-robust $L_2$-mixing time we get,

**Lemma 2.11.** *Let $\mathbb{P}$ be an ergodic Markov Chain and $\mathcal{S}$ a valid set of strategies against $\mathbb{P}$. Fix $\mathbb{Q} \in \mathcal{S}$. Let $0 < a = 1 - b < 1$. Let $S = (1+\delta)R_2^{\mathcal{S}}(\mathbb{P}, \gamma)/a$, where $\gamma > 0$ and $\delta > 0$ are arbitrary. Then*

$$||\mu_S - \pi||_{2,\pi} \leq \gamma + \exp\left(-R_2^{\mathcal{S}}(\mathbb{P}, \gamma) \cdot (\delta - \log(1+\delta))\right) \cdot \sqrt{\frac{1}{\pi_*}}$$

*Proof.* This proof is similar to that of Lemma 2.9 except for a slightly different choice of parameters. Put $T = R_2^{\mathcal{S}}(\mathbb{P}, \gamma)$ and $S = T(1+\delta)/a$. Put $\mathcal{D} = \{\boldsymbol{\epsilon} : H(\boldsymbol{\epsilon}) \geq T\}$. If $\boldsymbol{\epsilon} \in \mathcal{D}$, $||\mu_0\xi(\boldsymbol{\epsilon}) - \pi||_{2,\pi} \leq \gamma$. For $\boldsymbol{\epsilon} \notin \mathcal{D}$, $||\mu_0\xi(\boldsymbol{\epsilon}) - \pi||_{2,\pi} \leq \sqrt{1/\pi_*}$. Going along the same lines as Lemma 2.9, we have

$$||\mu_S - \pi||_{2,\pi} \leq \gamma + \exp\left(-T(\delta - \log(1+\delta))\right) \cdot \frac{1}{\sqrt{\pi_*}} \qquad \square$$

**Corollary 2.12.** *Let $\mathbb{P}$ be an ergodic Markov Chain and $\mathcal{S}$ a valid set of strategies against $\mathbb{P}$. Let $\mathbb{Q}$ be compatible with $\mathbb{P}$ and $\mathbb{Q} \in \mathcal{S}$. Let $0 < a = 1-b < 1$. Assume that $R_2^{\mathcal{S}}(\mathbb{P}) \geq \log(1/\pi_*)/2$ and $\pi_* \leq 1/16$. Then $R_2^{\mathcal{S}}(a\mathbb{P}+b\mathbb{Q}) \leq 2(1+\delta)R_2^{\mathcal{S}}(\mathbb{P})/a$, as long as $R_2^{\mathcal{S}}(\mathbb{P})(\delta - \log(1+\delta)) \geq \log(1/\pi_*)/2$. In particular $\delta$ may be taken to be $5/2$. In particular, if $\mathbb{P}$ and $\mathbb{Q}$ are Cayley walks on a group $G$, the conclusion holds with $\mathcal{S} = \mathcal{C}$ and $\mathcal{S} = \mathcal{H}$.*

*Proof.* Let $T = R_2^{\mathcal{S}}(\mathbb{P})$ and $S = 2T(1+\delta)/a$. By submultiplicativity, we have $R^{\mathcal{S}}(\mathbb{P}, \gamma) \leq 2T$ for $\gamma = 1/4$. Lemma 2.11 now gives

$$||\mu_S - \pi||_{2,\pi} \leq 1/4 + exp\left(-2T(\delta - \log(1+\delta))\right) \cdot \sqrt{\frac{1}{\pi_*}}$$

Now put $T = \alpha \log(1/\pi_*)/2$ for $\alpha > 1$. Then we have

$$||\mu_S - \pi||_{2,\pi} \leq 1/4 + \pi_*^{(2\alpha(\delta - \log(1+\delta)) - 1)}$$

The second term is bounded by $1/4$ if $\delta - \log(1+\delta) \geq 1/\alpha$. $\square$

*Proof of Theorem 1.9 and Theorem 1.21.* Follows from Lemma 2.6, Corollary 2.10 and Corollary 2.12. $\square$

# 3 Reversible liftings

In this section we reprove a result due to Chen et al. (4) on reversible liftings of Markov Chains. The proof is inspired by considering the Robust mixing time of a Markov Chain and looking at a particular adversarial strategy. We start with a proof of Theorem 1.29.

*Proof of Theorem 1.29.* We prove $R(\mathbb{Q}) \geq \mathcal{T}(\mathbb{P})$ by exhibiting an adversarial strategy which allows the adversary to simulate the evolution of $\mathbb{P}$.

Consider the following adversarial strategy $\mathbb{A}$: Given $y \in \mathcal{Y}$, the adversary picks a state $y' \in \mathcal{Y}$ according to the distribution $\mu^x$ where $x = f(y)$. Recall that $\mu^x$ is the conditional distribution of $\mu$ given that $f(y) = x$.

Since $\mu = \sum_{x\in\mathcal{X}} \pi(x)\mu^x$, it follows that this strategy fixes the stationary distribution $\mu$ of $\mathbb{Q}$. We now claim that with this strategy the adversary can simulate the evolution of $\mathbb{P}$ on $\mathcal{Y}$.

For a distribution $\nu$ on $\mathcal{X}$, consider the distribution $F(\nu) = \sum_{x\in\mathcal{X}} \nu(x)\mu^x$ on $\mathcal{Y}$. Then

$$\begin{aligned}
2||F(\nu) - \mu||_{TV} &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_x} |\nu(x)\mu^x(y) - \pi(x)\mu^x(y)| \\
&= \sum_{x\in\mathcal{X}} \left( |\nu(x) - \pi(x)| \sum_{y\in\mathcal{Y}_x} \mu^x(y) \right) \\
&= \sum_{x\in\mathcal{X}} |\nu(x) - \pi(x)| = 2||\nu - \pi||_{TV}
\end{aligned}$$

Hence for a distribution $\nu$ on $\mathcal{X}$ and $x_2 \in \mathcal{X}$ we have

$$\begin{aligned}
(F(\nu)\mathbb{Q})(\mathcal{Y}_{x_2}) &= \sum_{y_1\in\mathcal{Y}} \sum_{y_2\in\mathcal{Y}_{x_2}} \nu'(y_1)\mathbb{Q}(y_1, y_2) \\
&= \sum_{x_1\in\mathcal{X}} \sum_{y_1\in\mathcal{Y}_{x_1}} \sum_{y_2\in\mathcal{Y}_{x_2}} \nu(x_1)\mu^{x_1}(y_1)\mathbb{Q}(y_1, y_2) \\
&= \sum_{x_1\in\mathcal{X}} \nu(x_1)\mathbb{P}(x_1, x_2) \\
&= (\nu\mathbb{P})(x_2)
\end{aligned}$$

Hence we have $F(\nu)\mathbb{Q}\mathbb{A} = F(\nu\mathbb{P})$. This shows that alternating $\mathbb{Q}$ with the adversary's strategy ensures that this adversarially modified $\mathbb{Q}$ cannot mix faster than $\mathbb{P}$. Hence $R(\mathbb{Q}) \geq \mathcal{T}(\mathbb{P})$. ☐

Now if $\mathbb{Q}$ were reversible, Theorem 1.13 implies $R(\mathbb{Q}) \leq \mathcal{T}(\mathbb{Q})(1 + \log(1/\mu_*)/2)$. Hence

$$\mathcal{T}(\mathbb{Q}) \geq \frac{\mathcal{T}(\mathbb{P})}{\log(1/\mu_*)}$$

When $\mu_*$ is only polynomially smaller than $\pi_*$, this gives our result. We now improve the result by looking at the adversarial strategy in more detail and show $\mathcal{T}(\mathbb{Q}) \geq r(\mathbb{P})$.

*Proof of Theorem 1.31.* Let $\mathbb{A}$ denote the stochastic matrix representing the adversarial strategy. Note that $\mathbb{A}$ is reducible, but on each irreducible component, it reaches stationarity in one step. It follows that $\mathbb{A}$ is reversible.

Let $\boldsymbol{\alpha}$ denote the eigenvector (of length $|\mathcal{X}|$) corresponding to $\lambda_*(\mathbb{P})$ and define $\boldsymbol{\beta}$ (of length $|\mathcal{Y}|$) via $\boldsymbol{\beta}(y) = \boldsymbol{\alpha}(x)\mu^x(y)$ where $x = F(y)$. From our analysis before, it follows that for all $x \in \mathcal{X}$,

$$\sum_{y \in \mathcal{Y}_x} (\boldsymbol{\beta}\mathbb{Q})(y) = \lambda_*(\mathbb{P})\alpha(x)$$

Since $\mathbb{A}$ redistributes the probability in each $\mathcal{Y}_x$ according to $\mu^x$, we have for $y \in \mathcal{Y}$,

$$\boldsymbol{\beta}(\mathbb{Q}\mathbb{A})(y) = \lambda_*(\mathbb{P})\alpha(x)\mu^x(y) = \lambda_*(\mathbb{P})\boldsymbol{\beta}(y)$$

where $x = F(y)$. Thus $\lambda_*(\mathbb{Q}\mathbb{A}) \geq \lambda_*(\mathbb{P})$. Since $\mathbb{A}$ is a contraction (it is stochastic), we have

$$|\lambda_*(\mathbb{Q})| \geq |\lambda_*(\mathbb{Q}\mathbb{A})| \geq |\lambda_*(\mathbb{P})|$$

Hence $\mathcal{T}(\mathbb{Q}) \geq r(\mathbb{Q}) \geq r(\mathbb{P})$. $\square$

As a consequence, it follows that for many natural reversible chains where $\mathcal{T}(\mathbb{P}) = O(r(\mathbb{P}))$, one cannot gain more than a constant factor improvement by considering a reversible lifting.

(4) also gives an example $\mathbb{Q}$ of a reversible random walk on a tree (with $\pi_*$ exponentially small) and its collapsing $\mathbb{P}$ for which

$$\mathcal{T}(\mathbb{Q}) = \Theta\left(\mathcal{T}(\mathbb{P})\,\frac{\log\log(1/\pi_*)}{\log(1/\pi_*)}\right)$$

Since we know that $R(\mathbb{Q}) \geq \mathcal{T}(\mathbb{P})$ it shows that Theorem 1.13 is almost tight, even for reversible chains.

# 4 Upper bounds on Robust mixing time

In this section we observe that many proof techniques which establish bounds on mixing time in fact give us bounds on the Robust mixing time.

**Definition 4.1.** Let $\mathbb{P}$ be the transition matrix of an ergodic Markov Chain with stationary distribution $\pi$. For vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$, put $\langle \boldsymbol{\alpha}, \boldsymbol{\beta}\rangle_\pi = \sum_{x\in\mathcal{X}} \boldsymbol{\alpha}(x)\boldsymbol{\beta}(x)\pi(x)$. Associate the following bilinear forms with $\mathbb{P}$

$$\mathcal{E}_{\mathbb{P}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \langle (I - \mathbb{P})\boldsymbol{\alpha}, \boldsymbol{\beta}\rangle_\pi$$
$$\mathcal{F}_{\mathbb{P}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \langle (I + \mathbb{P})\boldsymbol{\alpha}, \boldsymbol{\beta}\rangle_\pi$$

$\mathcal{E}_{\mathbb{P}}$ is the Dirichlet form associated with $\mathbb{P}$ and $\mathcal{F}_{\mathbb{P}}$ is used to bound the negative eigenvalues of $\mathbb{P}$.

**Definition 4.2.** Let $\pi$ be a distributions on $\mathcal{X}$ and $\pi(x) > 0$ for all $x \in \mathcal{X}$. For any function $f : \mathcal{X} \to \mathbb{R}^+$, define the entropy of $f$ via

$$\mathrm{Ent}(f) = \mathbb{E}_\pi \left[ f \log f \right] \tag{13}$$

and for a distribution $\mu$ on $\mathcal{X}$, put $\mathrm{Ent}(\mu) = D(\mu||\pi) = \mathrm{Ent}(f)$, where $f(x) = \mu(x)/\pi(x)$ is the density function of $\mu$ w.r.t. $\pi$.

$\mathrm{Ent}(\mu)$ measures the entropy of $\mu$ relative to $\pi$. It is easy to see that $\mathrm{Ent}(\cdot)$ is convex non-negative and vanishes exactly when $\mu = \pi$. The log-Sobolev constant is defined as follows

**Definition 4.3.** Let $\mathbb{P}$ be a Markov Chain satisfying $\pi\mathbb{P} = \pi$. The log-Sobolev constant $\alpha(\mathbb{P})$ is defined as

$$\alpha(\mathbb{P}) = \min_f \frac{\mathcal{E}_{\mathbb{P}}(f, f)}{\mathrm{Ent}(f^2)} \tag{14}$$

where the minimum is taken over all real valued functions $f$ on $\mathcal{X}$ for which $\langle f, f \rangle_\pi = 1$.

**Lemma 4.4.** *Let $\mathbb{P}$ be an ergodic Markov Chain with stationary distribution $\pi$. Let $\mu$ be any distribution and $\mathbb{A}$ any stochastic matrix compatible with $\mathbb{P}$. Then for any $p \geq 1$,*

*(a)* $||\mu\mathbb{A} - \pi||_{p,\pi} \leq ||\mu - \pi||_{p,\pi}$

*(b)* $Var_\pi(\mu\mathbb{A}) \leq Var_\pi(\mu)$

*(c)* $||\mu\mathbb{A} - \pi||_{TV} \leq ||\mu - \pi||_{TV}$

*(d)* $D(\mu\mathbb{A}||\pi) \leq D(\mu||\pi)$

*Proof.* Since $\mathbb{A}$ is stochastic and compatible with $\mathbb{P}$, it follows that $\mathbb{A}$ is a contraction on $L^p(\pi)$ for $p \geq 1$. This proves $(a)$. $(b)$ and $(c)$ are special cases of $(a)$.
$\mathbb{A}$ is a contraction on $L^2(\pi)$ implies $\sigma_1(\mathbb{S}(\mathbb{A})) \leq 1$. Hence the log-Sobolev constant $\alpha(\mathbb{A}\overleftarrow{\mathbb{A}}) \geq 0$. (15, Prop 6) shows that $D(\mu\mathbb{A}||\pi) \leq D(\mu||\pi)$, when $\alpha(\mathbb{A}\overleftarrow{\mathbb{A}}) \geq 0$. ☐

As observed in (18, Chapter 3), many upper bounds on mixing time can be described via the following approach:

- Let $V(\eta)$ be a scalar associated with the distribution $\eta$ which satisfies the following
  - $V(\eta) \geq 0$ for all distributions $\eta$,
  - $V(\eta_t) \to 0 \iff \eta_t \to \pi$, for any sequence of distributions $\eta_t$ and
  - $V(\eta\mathbb{P}) \leq V(\eta)$.
- Let $I(t) = V(\mu_t)$ where $\mu_t$ is the $t$-step distribution. Note that $I(t)$ is a non-increasing function,
- Show that there is some non-decreasing function $G : \mathbb{R}^+ \to \mathbb{R}^+$ which satisfies $I(t) - I(t+1) \geq G(I(t))$

- Using $V(0) = I(0)$, solve for $T$ by which $V(T) \leq 1/4$ (or some appropriate constant $< 1$ depending on choice of $V$)

In case of an AMMC $\mathcal{P} = (\mathbb{P}, \{\mathbb{A}_t\})$ we can do the following

- *$V$ will usually also satisfy* $V(\mu\mathbb{A}) \leq V(\mu)$ for all $\mathbb{A}$ compatible with $\mathbb{P}$.
- Let $I(t) = V(\mu_t)$ and $J(t) = V(\nu_t)$ where $\nu_t = \mu_t\mathbb{P}$ and $\mu_{t+1} = \nu_t\mathbb{A}_t$.
- $I(t) - J(t) \geq G(I(t))$ and $I(t+1) \leq J(t)$ imply $I(t) - I(t+1) \geq G(I(t))$ as before.

Hence bounds on mixing time given by methods following the above approach also apply to Robust mixing time. This includes most analytical methods used to establish upper bounds on mixing time including log-Sobolev inequality based bounds, conductance based bounds and congestion based bounds (17). Bounds using Entropy constant (a.k.a Modified log-Sobolev constant) do not lead to a Robust mixing upper bound as they are not known to work in discrete time.

Unlike standard mixing time, establishing a bound on $R(\mathbb{P}^2)$ does not automatically imply any bound on $R(\mathbb{P})$.

**Example 7.** Let $\mathbb{P}$ be the **walk on directed edges** on a "good enough" undirected regular expander $X$. We have already seen that $R(\mathbb{P}) = \infty$ and $\mathcal{T}(\mathbb{P}) < \infty$. One can also show that $R(\mathbb{P}^2) < \infty$, thus showing that if the adversary is only allowed to intervene every alternate step, any adversarially modified version of $\mathbb{P}$ will still converge. A heuristic argument for that is that since the chain $\mathbb{P}$ remembers only the previous step of the usual random walk on $X$, when we run two steps of the chain in succession, the adversary does not gain any extra knowledge. A formal proof of $R(\mathbb{P}^2) < \infty$ is given in Theorem 4.11.

However, techniques similar to those used in Theorem 1.9 can be used to show

**Lemma 4.5.** *Let $\mathbb{P}, \mathbb{Q}$ be compatible Markov Chains and $k > 1$. Then for $0 < a = 1 - b < 1$, $\mathcal{T}(a\mathbb{P} + b\mathbb{Q}) \leq C \cdot kR(\mathbb{P}^k)/a^k$ for some absolute constant $C$.*

*Proof.* Consider $t = CkR(\mathbb{P}^k)/a^k$ for $C$ to be chosen later. Condition the run on the result of the initial coin toss (to decide if we run $\mathbb{P}$ or $\mathbb{Q}$). Break run of $a\mathbb{P} + b\mathbb{Q}$ into $t/k$ blocks of length $k$. The expected number of all $\mathbb{P}$ blocks is $CR(\mathbb{P}^k)$. For large enough $C$ we have at least $2R(\mathbb{P}^k)$ all $\mathbb{P}$ blocks. This together with an application of Chernoff type concentration inequality gives the result. $\square$

Coupling is one common approach to estimate mixing time of many Markov Chains. A coupling proof does not automatically give bounds on the robust mixing time. Many coupling proofs start by defining some notion of *distance* and show that it decreases in expectation after each step. Such coupling proofs lead to a bound on robust mixing time if one restricts to adversaries who cannot increase the *distance*.

**Example 8.** Consider the lazy random walk on the hypercube. The coupling proof showing a $n \log n + O(n)$ mixing time bound starts with two copies of the chain and couples them so that the Hamming distance between the two copies never increases and decreases in expectation.

Moves of a Cayley adversary in this case, amounts to flipping a subset of coordinates. Flipping the same subset in the other copy of the chain, ensures that the Hamming distance is not changed. Thus we have a coupling based proof of the $n \log n + O(n)$ mixing time bound for the lazy random walk against a Cayley adversary.

This also works against a holomorphic adversary, as automorphisms of the hypercube correspond to a combination of permutation of the coordinates and flipping a subset of the coordinates.

**Example 9.** Consider the random to top move chain, where we pick a random card and move it to the top position. The coupling proof starts with two copies of the chain and couples them so that after the move the two copies have the same top card. Thus we have created a pairing of cards in the two copies. The proof that we have not destroyed any existing pairing, depends on the fact that the paired cards are always continguous.

A Cayley adversary can easily destroy the continguous nature of the paired cards and thus in the next step, one may actually see a decrease in the number of paired cards.

Thus the naïve extension of the coupling based proof does not work in the robust setting.

The log-Sobolev constant as originally introduced by Diaconis works only in continuous time and needs to be adapted for the discrete time case. Miclo (15) adapted the result to discrete time and showed that $\mathrm{Ent}(\mu) - \mathrm{Ent}(\mu\mathbb{P}) \geq \alpha(\mathbb{P}\overleftarrow{\mathbb{P}})\,\mathrm{Ent}(\mu)$, where $\alpha(\mathbb{P}\overleftarrow{\mathbb{P}})$ is the log-Sobolev constant of the reversible chain $\mathbb{P}\overleftarrow{\mathbb{P}}$. This immediately translates to Robust mixing time bound as well.

**Lemma 4.6.** *Let $\mathbb{P}$ be an ergodic Markov Chain with stationary distribution $\pi$. Let $\alpha = \alpha(\mathbb{P}\overleftarrow{\mathbb{P}})$ denote the log-Sobolev constant for the chain $\mathbb{P}\overleftarrow{\mathbb{P}}$. Then $R(\mathbb{P}) = O(\log\log(1/\pi_*)/\alpha)$.*

*Proof.* (15, Proposition 6) shows that after each application of $\mathbb{P}$, the entropy of the resulting distribution falls by a factor $1 - \alpha(\mathbb{P}\overleftarrow{\mathbb{P}})$. Clearly adversarial moves cannot increase the entropy. Since the initial entropy is $\leq \log(1/\pi_*)$, the result follows. ☐

**Lemma 4.7.** *Let $\mathbb{P}$ be a reversible Markov Chain. Then*

$$\alpha\left(\left(\frac{I+\mathbb{P}}{2}\right)^2\right) \geq \frac{\alpha(\mathbb{P})}{2} \tag{15}$$

*If $\mathbb{P}$ has non-negative eigenvalues, then $\alpha(\mathbb{P}^2) \geq \alpha(\mathbb{P})$.*

*Proof.* Suppose $\mathbb{P}$ has non-negative eigenvalues. Let $1 = \lambda_0 > \lambda_1 \geq \cdots \geq \lambda_{N-1} \geq 0$ be the eigenvalues of $\mathbb{P}$ and let $\boldsymbol{x}_i$ be the corresponding eigenvectors which form an orthonormal basis of $L^2(\pi)$. Write $f = \sum_i c_i \boldsymbol{x}_i$ for appropriate scalars $c_i$. Then

$$\mathcal{E}_{\mathbb{P}^2}(f,f) = \langle (I-\mathbb{P}^2)f, f\rangle_\pi = \sum_i c_i^2(1-\lambda_i^2) \geq \sum_i c_i^2(1-\lambda_i) = \langle (I-\mathbb{P})f, f\rangle_\pi = \mathcal{E}_{\mathbb{P}}(f,f) \tag{16}$$

Since $\mathrm{Ent}(f)$ is independent of $\mathbb{P}$ it follows $\alpha(\mathbb{P}^2) \leq \alpha(\mathbb{P})$. In the general case, apply the previous result to $(I+\mathbb{P})/2$ and use the fact that $\alpha((I+\mathbb{P})/2) = \alpha(\mathbb{P})/2$ which follows from the definition of the log-Sobolev constant. ☐

**Example 10. Lazy Random walk on hypercube** Let $\mathbb{P}$ denote the (reversible) lazy random walk on the $n$-dimensional hypercube. All the eigenvalues of $\mathbb{P}$ are of the form $k/n$ for $k \in \{0, \ldots, n\}$. (6) shows that the log-Sobolev constant for the hypercube is $\Theta(1/n)$, and hence we have $R(\mathbb{P}) = O(n \log \log(2^n)) = O(n \log n) = \Theta(\mathcal{T}(\mathbb{P}))$.

**Theorem 4.8.** *Let $\mathbb{P}_{RT;T}$ denote the Random-Top transposition chain. Then $R((I+\mathbb{P}_{RT;T})/2) = O(n \log^2 n)$.*

*Proof.* We start by estimating the log-Sobolev constant of $\mathbb{P}_{RT;T}$. Lemma A.2 shows that $\mathcal{E}_{\mathbb{P}_{RR;T}} \leq 18\mathcal{E}_{\mathbb{P}_{RT;T}}$. (2) shows that $\alpha(\mathbb{P}_{RR;T}) \geq \frac{1}{2n \log n}$. Hence by the definition of the log-Sobolev constant, it follows that $\alpha(\mathbb{P}_{RT;T}) \geq \frac{1}{36n \log n}$. Lemma 4.7 now implies

$$\alpha\left(\left(\frac{I + \mathbb{P}_{RT;T}}{2}\right)^2\right) \geq \frac{1}{72n \log n} \tag{17}$$

The result now follows from Lemma 4.6. □

If we try to estimate the robust mixing time of $\mathbb{P}_{RT;T}$ via log-Sobolev methods directly via Lemma 4.6, one would need to estimate the log-Sobolev constant of $\mathbb{P}^2_{RT;T}$. The bound in Theorem 4.8 is against an adversary who is not required to respect the group structure. Later we show the optimal $O(n \log n)$ bound if we require the adversary to respect the group structure.

## 4.1 A spectral result

In this subsection, we prove a new rapid mixing result based on spectral machinery, which can be applied to chains whose top singular values are equal to or very close to 1. We use this to show that $R(\mathbb{P}^2) < \infty$ where $\mathbb{P}$ is the walk on directed edges considered before. For simplicity, we assume that $\mathbb{P}$ has uniform stationary distribution through out the rest of this section. The general case follows by considering $\mathcal{S}(\mathbb{P})$ instead of $\mathbb{P}$.

The approach of this spectral result is the following: Call a subspace $M$ "bad" if any vector in $M$ does not contract too much when $\mathbb{P}$ is applied to it. Similarly call a subspace "good" if every vector in $M$ contracts a lot when $\mathbb{P}$ is applied to it.

In case of a reversible Markov chain $\mathbb{P}$, the "bad" subspace is usually invariant under $\mathbb{P}$ (e.g. eigenvector corresponding to eigenvalue $\lambda_*$). However, if $\mathbb{P}$ is non-reversible, $\mathbb{P}$ may map "bad" subspaces to a "good" subspace. In such a case, two applications of $\mathbb{P}$ should contract a vector.

Let $\mathbb{P}$ be a Markov chain on $N$ states with uniform stationary distribution $\pi$. Consider the Singular Value Decomposition of $\mathbb{P}$ and let $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{N-1}$ and $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{N-1}$ denote the left and right singular vectors of $\mathbb{P}$. Since $\pi$ is uniform we may assume $\boldsymbol{x}_0 = \boldsymbol{y}_0 =$ constant vector. Let $X(\mathbb{P})$ (respy. $Y(\mathbb{P})$) denote the matrix whose columns are $\boldsymbol{x}_i$ (respy. $\boldsymbol{y}_i$) and $\Sigma(\mathbb{P})$ the diagonal matrix consisting of the singular values $\sigma_i$ of $\mathbb{P}$ so that the following equation holds

$$\mathbb{P} = X(\mathbb{P})\Sigma(\mathbb{P})Y(\mathbb{P})^T$$

Note that $X(\mathbb{P})$ and $Y(\mathbb{P})$ are both unitary.

**Definition 4.9.** Let $\mathbb{P}$ be a Markov chain with uniform stationary distribution and $X(\mathbb{P}), Y(\mathbb{P})$ be as above. Define the inner product matrix $IP(\mathbb{P})$ via

$$IP(\mathbb{P}) = \Sigma(\mathbb{P})Y(\mathbb{P})^T X(\mathbb{P})$$

Note that $IP(\mathbb{P})_{ij} = \langle \boldsymbol{x}_i\mathbb{P}, \boldsymbol{x}_j\rangle$ and that $IP(\mathbb{P})$ is just $\mathbb{P}$ written in the $\boldsymbol{x}_i$ basis. For $k = 1, \ldots, N-1$, denote by

- $E_k$, the linear span of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$,
- $E'_k$, the linear span of $\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_{N-1}$,
- $L_k$, the orthogonal projection (of the $N$ dimensional vector space) onto $E_k$,
- $c_k$, the euclidean norm of the $E_k \to E_k$ operator

$$\boldsymbol{z} \mapsto \boldsymbol{z}\mathbb{P}L_k$$

Note that $c_k$ is the norm of $(k-1)\times(k-1)$-minor of $IP(P)$ consisting of rows and columns $2, \ldots, k$.

Note that in terms of our earlier discussion the subspace $E_k$ is "bad" as it corresponds to high singular values and the subspace $E'_k$ is "good" as it corresponds to vectors of low singular values.

**Theorem 4.10.** *Let $\mathbb{P}$ be a doubly stochastic matrix with SVD $(\boldsymbol{x}_i, \boldsymbol{y}_i, \sigma_i)$. Then for any initial distribution $\mu$ and $t > 0$ and $1 \le k \le N$, we have*

$$||\mu\mathbb{P}^{2t} - \pi||_2 \le (\sigma_1(c_k + 2\sigma_k))^t ||\mu - \pi||_2$$

*where $\pi$ is the uniform stationary distribution. When $k = N$, we take $c_N = \sigma_1$ and $\sigma_N = 0$.*

*Proof.* Let $(\boldsymbol{x}_i, \boldsymbol{y}_i, \sigma_i)_{i=0}^{N-1}$ be the Singular Value Decomposition of $\mathbb{P}$. Let $E = E_k$ be the subspace spanned by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k-1}$ and $E' = E'_k$ be the subspace spanned by $\boldsymbol{x}_k, \ldots, \boldsymbol{x}_{N-1}$. Also, denote by $L$, the orthogonal projection onto $E$ and by $L'$ the orthogonal projection onto $E'$.

Since $\boldsymbol{x}_i\mathbb{P} = \sigma_i\boldsymbol{y}_i$ and we see that $\mathbb{P}$ maps $E$ and $E'$ to orthogonal subspaces. By definition $c_k$ is the Euclidean norm of the operator $L\mathbb{P}L$ on $E \oplus E'$. We now show that the Euclidean norm of $L'\mathbb{P}$ on $E \oplus E'$ is $\sigma_k$. Also note that the Euclidean norm of the operator $\mathbb{P}$ on $E \oplus E'$ is $\sigma_1$.

Let $\boldsymbol{z'} \in E'$ be arbitrary. Then

$$\boldsymbol{z'} = \sum_{i \ge k} \alpha_i \boldsymbol{x}_i$$

for some scalars $\alpha_i$, since $E'$ is spanned by $\boldsymbol{x}_k, \ldots, \boldsymbol{x}_{N-1}$. Since $\boldsymbol{x}_i\mathbb{P} = \sigma_i\boldsymbol{y}_i$, we have

$$||\boldsymbol{z'}\mathbb{P}||_2^2 = \sum_{i \ge k} \alpha_i^2 ||\sigma_i \boldsymbol{y}_i||_2 \le \sigma_k^2 \sum_{i \ge k} \alpha_i^2 = \sigma_k^2 ||\boldsymbol{z'}||_2^2 \tag{18}$$

since the $y_i$'s are orthonormal.

Let $\boldsymbol{z}$ be any vector orthogonal to $\pi$. Then

$$\begin{aligned} \boldsymbol{z}\mathbb{P}^2 &= \boldsymbol{z}L\mathbb{P}^2 + \boldsymbol{z}L'\mathbb{P}^2 \\ &= \boldsymbol{z}L\mathbb{P}L\mathbb{P} + \boldsymbol{z}L\mathbb{P}L'\mathbb{P} + \boldsymbol{z}L'\mathbb{P}^2 \end{aligned} \tag{19}$$

Estimating the individual norms, we get

$$||\boldsymbol{z}L\mathbb{P}L\mathbb{P}||_2 \quad \leq \sigma_1||\boldsymbol{z}L\mathbb{P}L||_2 \quad \leq \sigma_1 c_k||z||_2 \tag{20}$$

$$||\boldsymbol{z}L\mathbb{P}L'\mathbb{P}||_2 \quad \leq \sigma_k||\boldsymbol{z}L\mathbb{P}||_2 \quad \leq \sigma_k\sigma_1||z||_2 \tag{21}$$

$$||\boldsymbol{z}L'\mathbb{P}^2||_2 \quad \leq \sigma_1||\boldsymbol{z}L'\mathbb{P}||_2 \quad \leq \sigma_1\sigma_k||z||_2 \tag{22}$$

Combining we get

$$||\boldsymbol{z}\mathbb{P}^2||_2 \leq \sigma_1(c_k + 2\sigma_k)\boldsymbol{z}$$

If $\mu$ is any initial distribution, $\mu - \pi$ is orthogonal to $\pi$ and hence we have for $t > 0$,

$$||\mu\mathbb{P}^{2t} - \pi||_2 = ||(\mu - \pi)\mathbb{P}^{2t}||_2 \leq (\sigma_1(c_k + 2\sigma_k))^t||\mu - \pi||_2 \qquad \square$$

The assumption that $\pi$ is uniform is only for notational convenience. In the general case, we consider $\mathcal{S}(\mathbb{P})$ whose left and right singular vectors corresponding to $\sigma_0 = 1$ are equal.

**Theorem 4.11.** *Let $\mathbb{P}$ denote the* ***walk on directed edges*** *of a undirected $d$-regular graph $X$ considered before. Let $\mathbb{Q}$ denote the standard random walk on $X$ and $r(\mathbb{Q})$ its relaxation time. Then*

$$R_2(\mathbb{P}^2) \leq r(\mathbb{Q})(2 + \log(nd))$$

*where $n$ denotes the number of vertices of $X$ and $d \geq 2$.*

*Proof.* This is an application of Theorem 4.10. We calculate the Singular Value Decomposition of $\mathbb{P}$ and show that for a suitable $k$, $\sigma_1 = 1, c_k = \lambda_*, \sigma_k = 0$, where $\lambda_* = \lambda_*(\mathbb{Q})$. Thus after every two applications of $\mathbb{P}$ the Euclidean distance to stationarity contracts by $\lambda_*$. Adversarial moves cannot increase the distance.

Since the stationary distribution is uniform, the Euclidean distance is proportional to the $L_2(\pi)$ distance and the same conclusion holds for the $L_2(\pi)$ distance as well. Thus after $2r(\mathbb{Q})$ the distance to stationarity reduces by a factor of 1/e. Since the worst initial value is $\sqrt{nd}$, we have the result.

For each vertex $u$ of $X$, let $L_u^-$ denote the span of vectors corresponding to edges coming into $u$ and $L_u^+$ denote the span of vectors corresponding to edges leaving $u$. Also, let $\boldsymbol{x}_u^- \in L_u^-$ denote the vector which is 1 for edges coming into $u$ and 0 elsewhere. Similarly define $\boldsymbol{x}_u^+ \in L_u^+$.

Observe that the $nd$-dimensional space has the decomposition

$$\oplus_u L_u^- = \oplus_v L_v^+$$

and also that $\mathbb{P}$ maps $L_u^-$ to $L_u^+$. The matrix for $\mathbb{P}$ as a map from $L_u^-$ to $L_u^+$ is given by $J_d/d$, where $J_d$ is the $d \times d$ matrix with all entries equal to 1. $J_d/d$ is a symmetric matrix with eigenvalues 1 and 0 (multiplicity $d - 1$).

Since this holds for every vertex $u$, it follows that $\mathbb{P}$ has singular values 1 with multiplicity $n$, and 0 with multiplicity $n(d-1)$.

Let $E$ denote the span of the $n$ left singular vectors corresponding to the singular value 1. Note that $E$ has orthonormal basis $\{\frac{\boldsymbol{x}_u^-}{\sqrt{d}}\}_u$ and includes the stationary distribution. Note that image of $E$ under $\mathbb{P}$ has orthonormal basis $\{\frac{\boldsymbol{x}_u^+}{\sqrt{d}}\}_u$.

For arbitrary vertices $u, v$ of $X$, we have

$$\langle \frac{\boldsymbol{x}_u^-}{\sqrt{d}}\mathbb{P}, \frac{\boldsymbol{x}_v^-}{\sqrt{d}} \rangle = \frac{\langle \boldsymbol{x}_u^+, \boldsymbol{x}_v^- \rangle}{d} = \mathbb{Q}(u,v) = \begin{cases} 1/d & \text{if } (u,v) \text{ is an edge} \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

Thus $\mathbb{P}$ considered as an operator on $E$ behaves exactly like $\mathbb{Q}$. Now take $k = n$ in Theorem 4.10 and observe that $E = E_n \oplus < \pi >$, where $< \pi >$ denotes the linear span of the uniform stationary distribution of $\mathbb{Q}$. Since $< \pi >$ is invariant under $\mathbb{P}$, it follows that $c_n = \lambda_*(\mathbb{Q})$.

Thus Theorem 4.10 implies that for any initial distribution $\mu$, two successive applications of $\mathbb{P}$ reduces the euclidean norm of $\mu - \pi$ by a factor $\sigma_1(c_n + 2\sigma_n) = \lambda_*(\mathbb{Q})$. ☐

We conclude this section, with another example similar to the "walk on directed edges" example.

**Example 11. "walk without immediate reversal":** Let $X$ be a connected $d$-regular undirected graph. Assume that the usual random walk on $X$ is ergodic. Define a walk $\mathbb{P}$ as follows: If we are currently at vertex $u$, we pick a neighbor uniformly at random, except the vertex from which we came.

Like the "walk on directed edges" example, the states of this walk are directed edges $(u, v)$ of $X$. From $(u, v)$ we move to $(v, w)$ where $w$ is a uniformly chosen neighbor of $v$ except $u$.

The adversarial strategy of reversing the orientation of an edge, ensures that this walk has $R(\mathbb{P}) = \infty$, since as before $\mathbb{P}$ maps the uniform distribution on edges into a vertex $u$ to the uniform distribution on edges out of $u$.

The singular vectors of $\mathbb{P}$ are the same as that for the walk on directed edges, although the singular values are different. As in the "walk on directed edges" example, define subspaces $L_u^-$ and $L_v+$. The matrix for $\mathbb{P}$ as an map from $L_u^-$ to $L_u^+$ is the $d \times d$-matrix $(J_d - I)/(d-1)$, where $J_d$ is the all ones matrix. As before, $(J_d - I)/(d-1)$ is symmetric with eigenvalues 1 and $-1/(d-1)$ (multiplicity $d-1$).

Since this holds for all vertices $u$, $\mathbb{P}$ has singular values 1 with multiplicity $n$ and $1/(d-1)$ with multiplicity $nd - d$. A proof similar to Theorem 4.11 allows us to prove a decay factor of $\lambda_*(\mathbb{Q}) + 2/(d-1)$ where $\mathbb{Q}$ corresponds to the natural walk on $X$. Thus if $X$ is a sufficiently good expander and the degree is large enough, in the sense that $\lambda_*(\mathbb{Q}) + 2/(d-1) < 1$, the walk without immediate reversal is also rapidly mixing.

# 5 Cayley walks on groups

In this section, we specialize to Cayley walks driven by a probability measure $P$ over a group $G$. The chain is irreducible iff the support of $P$ generates $G$ and aperiodic if $P(id) > 0$ where $id$ is the identity element of $G$.

## 5.1 Walks against restricted adversaries

It is well known that the knowledge of all the singular values of the transition matrix $\mathbb{P}$ can give better bounds on the standard mixing time than those obtained just from the knowledge of $\sigma_1(\mathbb{P})$. In this section we show that the same conclusion holds for the robust mixing time against holomorphic adversaries.

**Definition 5.1.** For a group $G$, the *holomorph* of $G$, denoted $\mathrm{Hol}(G)$ is the semi-direct product of $G$ with $\mathrm{Aut}(G)$, where $\mathrm{Aut}(G)$ acts naturally on $G$.

$$\mathrm{Hol}(G) = G \rtimes \mathrm{Aut}(G) \tag{24}$$

Elements of $\mathrm{Hol}(G)$ are called *holomorphisms* of $G$.

Holomorphisms of $G$ can be identified with permutations of $G$ as follows: Elements of $G$ act on $G$ by right translation and those of $\mathrm{Aut}(G)$ act naturally, i.e.

$$(h, \tau) \cdot g = \tau(g)h \tag{25}$$

**Lemma 5.2.** *The permutation representation of* $\mathrm{Hol}(G)$ *is faithful and transitive.*

*Proof.* Transitivity is easily established by by setting $\tau$ to be the identity automorphism and choosing $h$ appropriately. To see that this representation is faithful, suppose that for some $(h, \tau) \in \mathrm{Hol}(G)$, we have

$$\tau(g)h = g \ \forall g \in G \tag{26}$$

Choosing $g = id$, we see that $h = id$ which in turn implies $\tau$ is the identity automorphism. ☐

Hence we can identify holomorphisms of $G$ by the permutations they induce on $G$ (which was how we defined them in the introduction).

**Definition 5.3.** A permutation $J : G \to G$ is said to be $G$*-respecting* if for some permutation $K : G \to G$, and all $g, h \in G$,

$$J(h)^{-1}J(g) = K(h^{-1}g)$$

Note that if $K = J$, then we get the definition of an automorphism of $G$.

**Lemma 5.4.** *A permutation $J$ of $G$ is $G$-respecting iff it is a holomorphism of $G$.*

*Proof.* Let $\mathcal{G}$ denote the set of all $G$-respecting permutations on $G$. $\mathrm{Hol}(G)$ is closed under composition. Observe that $\mathcal{G}$ is also closed under composition, since

$$J_1(J_2(h))^{-1}J_1(J_2(g)) = K_1(J_2(h)^{-1}J_2(g)) = K_1(K_2(h^{-1}g)$$

where $K_1, K_2$ correspond to $J_1$ and $J_2$.

All automorphisms and right translations are $G$-respecting. Since, $\mathrm{Hol}(G)$ is generated by right translations and automorphisms, it follows that $\mathrm{Hol}(G) \leq \mathcal{G}$.

For a $G$-respecting permutation $J$, define $J'(g) = g \cdot (J(id))^{-1}$. Then $J' \in \mathrm{Hol}(G)$ since it is a right translation. Since $J$ and $J'$ are $G$-respecting, $J''(g) = J'(J(g))$ is also a $G$-respecting permutation, as $G$-respecting permutations are closed under composition. But

$$J''(id) = J'(J(id)) = J(id) \cdot J(id)^{-1} = id \tag{27}$$

Hence $J''$ is a $G$-respecting permutation which fixes the identity. Hence from the definition of $G$-respecting, substituting $h = id$, there is a permutation $K''$ of $G$ for which

$$J''(id)^{-1}J''(g) = K''(id^{-1}g) \tag{28}$$

Since $J''(id) = id$, we have $K''(g) = J''(g)$. Substituting this back in the definition of $G$-respecting, we have for all $g, h \in G$,

$$J''(h)^{-1}J''(g) = J''(h^{-1}g) \tag{29}$$

i.e. $J''$ is an automorphism of $G$. For all $g \in G$, we now have

$$J(g) = J''(g) \cdot J(id) = (J(id), J'') \cdot g \tag{30}$$

Hence $J \in \mathrm{Hol}(G)$ and $\mathcal{G} \leq \mathrm{Hol}(G)$. □

Note that the holomorphic strategies $\mathcal{H}$ are precisely convex combinations of $\mathrm{Hol}(G)$ (viewed as permutations on $G$) and Cayley strategies $\mathcal{C}$ are precisely the convex combinations of the subgroup $G$ of $\mathrm{Hol}(G)$ viewed as permutations on $G$.

We now look at the holomorphic robust mixing time of the Cayley walk on $G$. We identify a permutation on $G$ with the $|G| \times |G|$ permutation matrix representing it.

Let $\mathbb{P}$ be the transition probability matrix of the Cayley walk on a group $G$. Fix a sequence $\{\mathbb{A}_t\}_{t>0}$ of holomorphic strategies and define

$$\mathbb{Q}_0 = I \quad \mathbb{Q}_{k+1} = \mathbb{Q}_k\mathbb{P}\mathbb{A}_{k+1}$$

If $\mu_t$ denotes the distribution after $t$ rounds we have $\mu_t = \mu_0\mathbb{Q}_t$, where $\mu_0$ is the initial distribution.

**Lemma 5.5.** *If $\mu_0$ is supported only at $g \in G$ then $||\mu_t - \pi||_2^2 = (\mathbb{Q}_t{\mathbb{Q}_t}^T)(g, g) - 1/N$, where $N = |G|$ and $||\cdot||_2$ denotes the Euclidean norm.*

*Proof.*

$$\begin{aligned}
||\mathbb{Q}_t(g,\cdot) - \pi||_2^2 &= \sum_h (\mathbb{Q}_t(g,h) - 1/N)^2 \\
&= \sum_h \left(\mathbb{Q}_t(g,h)\mathbb{Q}_t(g,h) - 2\mathbb{Q}_t(g,h)/N + 1/N^2\right) \\
&= \sum_h (\mathbb{Q}_t(g,h)\mathbb{Q}_t(g,h)) - 1/N \\
&= (\mathbb{Q}_t{\mathbb{Q}_t}^T)(g,g) - 1/N
\end{aligned}$$

□

**Definition 5.6.** A matrix $\mathbb{B}$ whose rows and columns are indexed by elements of $G$ is said to be a *$G$-circulant* if $\mathbb{B}(g,h) = P(g^{-1}h)$ for some function $P : G \to \mathbb{R}$.

**Lemma 5.7.** *The following are true:*

*(a) The transition matrix $\mathbb{P}$ of a Cayley walk on $G$ is $G$-circulant,*

*(b) $G$-circulant matrices are closed under multiplication, linear combinations and taking transpose,*

*(c) If $J$ is a holomorphism of $G$ and $\mathbb{B}$ is $G$-circulant, then so is $J^{-1}\mathbb{B}J$.*

*Proof of (c).* Let $\mathbb{B}(g,h) = P(g^{-1}h)$ for $P : G \to \mathbb{R}$. Let $Q : G \to G$ be such that $J(x)^{-1}J(y) = Q(x^{-1}y)$. Finally, put $\mathbb{C} = J^{-1}\mathbb{B}J$ Then

$$\mathbb{C}(g,h) = \mathbb{B}(J(g), J(h)) = P((J(g))^{-1}J(h)) = P(Q(g^{-1}h)) \qquad \square$$

For the standard $L_2$-mixing time the following result is now folklore.

**Theorem 5.8.** *Let $\mathbb{P}$ be the transition matrix of an ergodic Cayley walk on a finite group $G$ with $N = |G|$. Let $1 = \sigma_0 \geq \sigma_1 \geq \cdots \geq \sigma_{N-1} \geq 0$ denote the singular values of $\mathbb{P}$. If $v_t$ denotes the distribution of the Markov Chain after $t$-steps starting from any initial distribution $v_0$,*

$$||v_t - \pi||_{2,\pi}^2 \leq \sum_{i=1}^{N-1} \sigma_i^{2t}$$

*where $\pi$ is the (uniform) stationary distribution of $\mathbb{P}$.*

We now show that the same conclusion holds in the presence of holomorphic adversary as well.

**Theorem 5.9.** *Let $\mathbb{P}$ denote the transition matrix of a Cayley walk on a finite group $G$. Assume that $\mathbb{P}$ is ergodic and let $1 = \sigma_0 \geq \sigma_1 \geq \ldots \sigma_{N-1} \geq 0$ denote the singular values of $\mathbb{P}$. Also assume that the adversary is holomorphic. Then*

$$||\mu_t - \pi||_{2,\pi}^2 \leq \sum_{i=1}^{N-1} \sigma_i^{2t}$$

*where $\mu_t$ denotes the distribution after $t$-rounds and $\mu_0$ is any initial distribution.*

*Proof.* By convexity the worst case happens when all the $\mathbb{A}_i$ are holomorphisms of $G$. Having fixed such an adversarial strategy, it is enough to consider the case when the initial distribution $\mu_0$ is supported on one element of $G$.

Assume that $\mu_0$ puts all its weight on $g \in G$. Let $\mathbb{Q}_t = \mathbb{P}\mathbb{A}_1\mathbb{P}\mathbb{A}_2 \ldots \mathbb{P}\mathbb{A}_t$. By Lemma 5.5, and the relation $|| \cdot ||_{2,\pi}^2 = |G||| \cdot ||_2$, we have

$$||\mu_t - \pi||_{2,\pi}^2 \leq N(\mathbb{Q}_t{\mathbb{Q}_t}^T)(g,g) - 1$$

We first establish that $\mathbb{Q}_t{\mathbb{Q}_t}^T$ is $G$-circulant and hence has equal diagonal entries. In case of a Cayley adversary this follows from the fact that $\mathbb{Q}_t$ is a product of $G$-circulant matrices. In the holomorphic case, $\mathbb{Q}_t$ need not be $G$-circulant. However, $\mathbb{Q}_t{\mathbb{Q}_t}^T$ is $G$-circulant. To see this, consider evaluating $\mathbb{Q}_t{\mathbb{Q}_t}^T$ inside out, i.e. put

$$\mathbb{C}_{t+1} = I \quad \text{and for } k \leq t \quad \mathbb{C}_k = \mathbb{P}\mathbb{A}_k\mathbb{C}_{k+1}{\mathbb{A}_k}^T\mathbb{P}^T$$

Clearly $\mathbb{C}_{t+1}$ is $G$-circulant. If $\mathbb{C}_{k+1}$ is $G$-circulant, then Lemma 5.7 implies $\mathbb{A}_k\mathbb{C}_{k+1}{\mathbb{A}_k}^T = \mathbb{A}_k\mathbb{C}_{k+1}(\mathbb{A}_k)^{-1}$ is also $G$-circulant since $\mathbb{A}_k$ is a holomorphism. Since $G$-circulant matrices are closed under multiplication and taking transposes it follows that $\mathbb{C}_k$ is $G$-circulant. Hence $\mathbb{C}_1 = \mathbb{Q}_t{\mathbb{Q}_t}^T$ is $G$-circulant and has equal diagonal entries. Hence we have

$$||\mu_t - \pi||_{2,\pi}^2 \leq \operatorname{tr}(\mathbb{Q}_t{\mathbb{Q}_t}^T) - 1$$

Since the trace is just the sum of the eigenvalues and the eigenvalues of $\mathbb{D}\mathbb{D}^T$ are just the squares of the singular values of $\mathbb{D}$, we have

$$||\mu_t - \pi||_{2,\pi}^2 \leq \sum_{i=0}^{N-1} \sigma_i(\mathbb{Q}_t)^2 - 1$$

But $\sum_{i=0}^{N-1} \sigma_i^2(\mathbb{D}_1\mathbb{D}_2\dots\mathbb{D}_{2t}) \leq \sum_{i=0}^{N-1}\prod_{j=1}^{2t}\sigma_i^2(\mathbb{D}_j)$ (see (14, Chapter 3) for a proof). Using $\sigma_i(\mathbb{A}_k) = 1$ for all $i, k$ and $\sigma_0(\mathbb{P}) = \sigma_0(\mathbb{A}_k) = 1$, we have the result. □

Now we prove Theorem 1.23 and show that that holomorphic robust $L_2$-mixing time of $\mathbb{P}$ is within a factor 2 of the standard mixing time of $\mathbb{P}\overleftarrow{\mathbb{P}}$.

*Proof of Theorem 1.23.* Let $\mathbb{P}$ be the transition matrix of a random walk on $G$. Considering the adversarial strategy where $\mathbb{A}_t = I$ and the one where $\mathbb{A}_t = \overleftarrow{\mathbb{P}}$, we have that

$$\max(\mathcal{T}_2(\mathbb{P}), \mathcal{T}_2(\mathbb{P}\overleftarrow{\mathbb{P}})) \leq R_2^{\mathcal{C}}(\mathbb{P}) \leq R_2^{\mathcal{H}}(\mathbb{P})$$

Let $\sigma_i$ denote the singular values of $\mathbb{P}$. Let $v_t$ denote the $t$-step distribution without any adversary for the chain $\mathbb{Q} = \mathbb{P}\overleftarrow{\mathbb{P}}$ starting from $v_0$. Hence the eigenvalues of $\mathbb{Q}$ are $\{\sigma_i^2\}$. From Lemma 5.5, the fact that $||\cdot||_{2,\pi}^2 = |G|||\cdot||_2^2$ we have

$$||v_t - \pi||_{2,\pi}^2 = |G|\mathbb{Q}^{2t}(g, g) - 1$$

if the initial distribution $v_0$ is concentrated on $g$. Since $\mathbb{Q}^{2t}$ has equal diagonal entries (it is a $G$-circulant),

$$||v_t - \pi||_{2,\pi}^2 = \sum_{g\in G} \mathbb{Q}^{2t}(g, g) - 1$$

Note that the right hand side is independent of $v_0$. Since the trace equals the sum of the eigenvalues we have for $t = \mathcal{T}_2(\mathbb{Q})$,

$$\sum_{i>0} \sigma_i^{4t} = ||v_t - \pi||_{2,\pi}^2 \leq (1/2)^2$$

Now consider a run of an adversarially modified version of $\mathbb{P}$ for $s$-steps. Let $\mu_s$ be the distribution after $s$ rounds starting from $\mu_0$. Then from Theorem 5.9, we now have for $s = 2t$,

$$||\mu_s - \pi||_{2,\pi}^2 \leq \sum_{i>0} \sigma_i^{2s} \leq (1/2)^2$$

Hence $R_2^{\mathcal{H}}(\mathbb{P}) \leq 2\mathcal{T}_2(\mathbb{P}\overleftarrow{\mathbb{P}})$. □

### 5.2 Discussion: Adversarial strategies for Cayley walks

We now discuss the differences between permitted adversarial strategies for Cayley walks. Let $\mathbb{P}$ be the transition probability matrix of a Cayley walk on a group $G$. Let $\mathcal{D}$ denote the set of all doubly stochastic matrices. In the definition of $R(\mathbb{P})$ the adversary chooses each $\mathbb{A}_t$ from $\mathcal{D}$, while for $R^{\mathcal{C}}$ and $R^{\mathcal{H}}$ the adversary chooses $\mathbb{A}_t$ from $\mathcal{C}$ (Cayley strategies) and $\mathcal{H}$ (holomorphic strategies) respectively. Note that Cayley strategies only allowed the adversary to right multiply the current state in $G$ with a group element of the adversary's choosing. One could also consider the set of strategies $\mathcal{LR}$ where we allow the adversary to left and right multiply the current state with group elements of the adversary's choosing.

**Proposition 5.10.** *Let $\mathbb{P}$ be an ergodic Cayley walk on a group $G$. Then $R^{\mathcal{LR}} = R^{\mathcal{C}}$, $R_2^{\mathcal{LR}} = R_2^{\mathcal{C}}$.*

*Proof.* Let $\mathcal{P}_{LR}$ be an adversarially modified version of $\mathbb{P}$ using $\mathcal{LR}$-strategies. By convexity, we may assume that the adversaries choices are deterministic. Suppose the $\mathcal{LR}$ adversary's choices are $\{\ell_t, r_t\}_{t>0}$. Consider $\mathcal{P}_C$, the adversarially modified version of $\mathbb{P}$, where at time $t$ the adversary applies a right translation by $r_t$.
Then the distribution $\mu_t^{LR}$ of $\mathcal{P}_{LR}$ and $\mu_t^{C}$ of $\mathcal{P}_C$ are related. More specifically, we have

$$(\forall h \in G), \mu_t^{LR}(\ell(t)h) = \mu_t^{C}(h)$$

where $\ell(t) = \ell_t \ell_{t-1} \dots \ell_1$. In particular the two distributions are just permutations of each other and hence $\mathcal{P}_{LR}$ and $\mathcal{P}_C$ have the same mixing time under total variation as well as $L_2$-distance. □

**Corollary 5.11.** *If $G$ has no proper outer-automorphisms, i. e., all automorphisms are induced by conjugations, then for any Cayley walk $\mathbb{P}$ on $G$, $R^{\mathcal{C}}(\mathbb{P}) = R^{\mathcal{H}}(\mathbb{P})$ and $R_2^{\mathcal{C}}(\mathbb{P}) = R_2^{\mathcal{H}}(\mathbb{P})$.*
*In particular, this is the case for the symmetric groups $S_n$, $n \geq 5$, $n \neq 6$.*

Hence an $LR$-adversary is not more powerful than a Cayley adversary. Clearly a holomorphic adversary is at least as powerful as a Cayley adversary. However it is not clear if the holomorphic adversary is strictly more powerful.

## 6 Questions

(4, Example 3.5) constructs a reversible chain $\mathbb{P}$ and a lifting $\mathbb{Q}$ of $\mathbb{P}$ for which

$$\mathcal{T}(\mathbb{Q}) = \Theta\left(\mathcal{T}(\mathbb{P})\frac{\log\log(1/\pi_*)}{\log(1/\pi_*)}\right) \tag{31}$$

where $\pi$ is the stationary distribution of $\mathbb{P}$. Also, each state of $\mathbb{P}$ lifted to atmost 2 states in $\mathbb{Q}$. So, if $\mu$ denotes the stationary distribution of $\mathbb{Q}$, we have $\mu_* \geq \pi_*/2$.
Theorem 1.29 now implies that

$$R(\mathbb{Q}) \geq \mathcal{T}(\mathbb{P}) = \Theta\left(\mathcal{T}(\mathbb{Q})\frac{\log(1/\pi_*)}{\log\log(1/\pi_*)}\right) = \Theta\left(\mathcal{T}(\mathbb{Q})\frac{\log(1/\mu_*)}{\log\log(1/\mu_*)}\right) \tag{32}$$

Thus even for reversible $\mathbb{Q}$, $R(\mathbb{Q})$ and $\mathcal{T}(\mathbb{Q})$ can differ by more than a constant factor. However this example had $\pi_*$ exponentially small in the number of vertices of the graph under lying $\mathbb{Q}$.

**Question 1.** Is it true that for all natural examples of reversible chains $\mathbb{P}$, the robust mixing time differs from standard mixing time by no more than a constant factor? Specifically,

(a) If $\pi_*$ is polynomially small, is $R(\mathbb{P}) = O(\mathcal{T}(\mathbb{P}))$?

(b) If $\mathbb{P}$ is a reversible Cayley walk on a group $G$, then is $R(\mathbb{P}) = \Theta(\mathcal{T}(\mathbb{P}))$?

Theorem 1.23 shows that for reversible ergodic Cayley walks on a group $G$, $\mathcal{T}_2(\mathbb{P}) \leq R_2^{\mathcal{C}}(\mathbb{P}) \leq R_2^{\mathcal{H}}(\mathbb{P}) \leq \mathcal{T}_2(\mathbb{P}) + 1$. Does something similar hold of robust mixing in variation distance?

**Question 2.** Let $\mathbb{P}$ denote a reversible ergodic random walk on a group $G$. Is it true that $R^{\mathcal{H}}(\mathbb{P}) = \Theta(\mathcal{T}(\mathbb{P}))$? $R^{\mathcal{C}}(\mathbb{P}) = \Theta(\mathcal{T}(\mathbb{P}))$?

We were able to show that $R^{\mathcal{H}}(\mathbb{P}_{RT;T}) = O(n \log n)$ but only $R((\mathbb{P}_{RT;T} + I)/2) = O(n \log^2 n)$. One approach to proving $R(\mathbb{P}_{RT;T}) = O(n \log n)$ is via the modified log-Sobolev constant approach. Gao and Quastel (11) show that the modified log-sobolev constant (also called entropy constant) is $\Omega(1/n)$. Currently it is not known whether modified log-Sobolev constant inequalities imply bounds on mixing time in discrete time (without any adversary).

**Question 3.** Is is true that $R(\mathbb{P}_{RT;T}) = O(n \log n)$?

What is the difference in power between a holomorphic and a Cayley adversary? For robust $L_2$-mixing times, Theorem 1.23 shows the difference is at most a constant factor.

**Question 4.** Is it true that $R^{\mathcal{H}}(\mathbb{P}) = O(R^{\mathcal{C}}(\mathbb{P}))$ when $\mathbb{P}$ is a Cayley walk?

In all the examples we have seen, the adversarial strategy which achieves the robust mixing time can be taken to be homogenous. Is this always the case?

**Question 5.** Is it true that $R(\mathbb{P}) = \max_{\mathbb{A}} \mathcal{T}(\mathbb{P}\mathbb{A})$ where the maximum is taken over all $\mathbb{A}$ compatible with $\mathbb{P}$?

In Example 11, we saw that if $X$ is a sufficiently good expander and the degree is large enough, then the random walk on $X$ without immediate reversal is rapidly mixing. Can the degree requirement be removed?

**Question 6.** Let $X$ be an undirected $d$-regular expander on $n$ vertices with $d \geq 3$. Is it always the case, that the random walk on $X$ without immediate reversal mixes in $O(\log n)$ time?

# 7 Acknowledgements

I would like to thank László Babai for introducing me to the fascinating area of Markov Chains and helpful discussions.

# References


[1] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, May 1986. MR0841111

[2] Tzong-Yow Lee amd Horng-Tzer Yau. Logarithmic sobolev inequality for some models of random walks. *The Annals of Probability*, 26(4):1855–1873, 1998. MR1675008

[3] Ivona Bezáková and Daniel Štefankovič. Convex combinations of markov chains and sampling linear orderings. In preperation.

[4] Fang Chen, László Lovász, and Igor Pak. Lifting markov chains to speed up mixing. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 275–281, New York, NY, USA, 1999. ACM Press. MR1798046

[5] Deepak Dhar. The abelian sandpile and related models. *PHYSICA A*, 263:4, 1999.

[6] Perci Diaconis and Laurent Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996. MR1410112

[7] Persi Diaconis. *Group Representations in Probability and Statistics*, volume 11 of *Lecture Notes – Monograph Series*. Institute of Mathematical Statistics, 1998. MR0964069

[8] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Annals of Applied Probability*, 21(4):2131–2156, October 1993. MR1245303

[9] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible markov chains. *Annals of Applied Probability*, 3(3):696–730, Aug 1993. MR1233621

[10] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitstheorie*, 57:159–179, 1981. MR0626813

[11] Fuqing Gao and Jeremy Quastel. Exponential decay of entropy in random transposition and bernoulli-laplace models. *Annals of Applied Probability*, 13(4):1591–1600, 2003. MR2023890

[12] Sharad Goel. Analysis of top to bottom shuffles. *Annals of Applied Probability*, 16, February 2006. MR2209335

[13] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *American Statistical Association Journal*, pages 13–30, March 1963. MR0144363

[14] Roger Horn and Charles Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991. MR1091716

[15] Laurent Miclo. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de markov finies. *Séminaire de probabilités de Strasbourg*, 31:136–167, 1997. MR1478724

[16] Ilya Mironov. (Not so) random shuffles of RC4. In *Crypto '02*, pages 304–319, 2002. MR2054828

[17] Ravi Montenegro. Duality and evolving set bounds on mixing times.

[18] Ravi Montenegro and Prasad Tetali. *Mathematical Aspects of Mixing Times in Markov Chains*, volume 1 of *Foundations and Trends in Theoretical Computer Science*. NOW Publishers, Boston-Delft, 2006.

[19] Elchanan Mossel, Yuval Peres, and Alistair Sinclair. Shuffling by semi-random transpositions. In *Proceedings of FOCS 2004*, volume 00, pages 572–581, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[20] Laurent Saloff-Coste. Random walks on finite groups. In Harry Kesten, editor, *Probability on Discrete Structures*, volume 10 of *Encylopaedia Math. Sci.*, pages 263–346. Springer, Berlin, 2004. MR2023654

# A Comparing Dirichlet forms

Diaconis and Saloff-Coste (8) developed techniques to compare the Dirichlet forms of two reversible random walks on the same finite group $G$ and later extended it to reversible Markov Chains in (9). We restate their results.

**Theorem A.1** (Theorem 1,3 in (8))**.** *Let $G$ be a finite group and $|G| = N$. Let $p$ and $\tilde{p}$ denote two measures on $G$ with support $E$ and $\tilde{E}$ both of which generate $G$.*

*Suppose each $\tilde{\sigma} \in \tilde{E}$ is represented as a product of elements $\tau_1\tau_2 \dots \tau_k$ of $E$. Denote $|\tilde{\sigma}| = k$, $N(\tilde{\sigma}, \tau)$ as the number of times $\tau$ occurs in the representation of $\tilde{\sigma}$. Then*

$$\mathcal{E}_{\tilde{\mathbb{P}}}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \le A\mathcal{E}_{\mathbb{P}}(\boldsymbol{\alpha}, \boldsymbol{\alpha})$$

*where*

$$A = \max_{\tau \in E} A(\tau) \quad \textit{and} \quad A(\tau) = \frac{1}{p(\tau)} \sum_{\tilde{\sigma} \in \tilde{E}} |\tilde{\sigma}| N(\tilde{\sigma}, \tau) \tilde{p}(\tilde{\sigma}) \tag{33}$$

*In addition if all representations are chosen to have odd length, we also have*

$$\mathcal{F}_{\tilde{\mathbb{P}}}(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \le A\mathcal{F}_{\mathbb{P}}(\boldsymbol{\alpha}, \boldsymbol{\alpha})$$

**Lemma A.2.** *Let $\mathbb{P}_{RT;T}$ and $\mathbb{P}_{RR;T}$ denote the transition probability matrices for the random-top and random-random transposition chains. Then*

$$\mathcal{E}_{\mathbb{P}_{RR;T}} \le 18\mathcal{E}_{\mathbb{P}_{RT;T}} \quad \mathcal{F}_{\mathbb{P}_{RR;T}} \le 18\mathcal{F}_{\mathbb{P}_{RT;T}}$$

*Proof.* Let $G = S_n$ and $\tilde{p}$ to correspond to the random to random transposition chain, i.e. $\tilde{p}(\tilde{\sigma}) = 2/n^2$ for a non-trivial transposition and $1/n$ for the identity permutation. Let $p$ correspond to random to top transposition chain, i.e. $p(\tau) = 1/n$ if $\tau = (1k)$ for some $k = 1, \dots, n$. Then $\tilde{E} = \{(i\,j) | i \le j\} \cup \{e\}$ and $E = \{(1\,k) | 1 \le k \le n\}$. are the supports for $\tilde{p}$ and $p$ respectively. Every non-trivial transposition $(i\,j) \in \tilde{E}$ can be written as $(1\,i)(1\,j)(1\,i)$. The identity permutation can be written as itself. Thus all representations have odd length.

We now apply Theorem A.1. To compute $A$ via (33), note that $A(id) = 1$ and $A((1\,i)) = \frac{2}{n}\sum_{j=1}^{n} F(i,j)$ where

$$F(i,j) = |(i\,j)| N\left((i\,j), (1,i)\right) + |(j\,i)| N\left((j\,i), (1,i)\right) = 3 * 2 + 3 * 1 = 9$$

giving $A = 18$. $\square$

**Lemma A.3.** *Let $\mathbb{P}_{RR;I}$ denote the transition probability matrix of the random-to-random move chain, i.e. we pick i,j at random and move the card at position i to position j. Then*

$$\mathcal{E}_{\mathbb{P}_{RR;T}} \le 3\mathcal{E}_{\mathbb{P}_{RR;I}}$$

*Proof.* This is another application of Theorem A.1. For $i \neq j$, let $c_{i\to j}$ denote the permutation corresponding to moving the card at position $i$ to position $j$. Note that $c_{i\to j}c_{j\to i} = id$. Each $c_{i\to j}$ is chosen with probability $1/n^2$ if $|i-j| > 1$ $c_{i\to i+1} = c_{i+1\to i}$ is chosen with probability $2/n^2$ and $id$ is chosen with probability $1/n$.

Fix $1 \le i < j \le n$ and write $(i\,j)$ as a product of permutations of the form $c_{i\to j}$ as follows:

$$(i\,j) = \begin{cases} c_{i\to i+1} & \text{if } j = i+1 \\ c_{j\to i} \cdot c_{i+1\to j} & \text{if } j > i+1 \end{cases}$$

Note that each $c_{i\to j}$ appears in a representation at most once, i.e. $N(\cdot,\cdot) \le 1$. We now calculate $A(c_{i\to j})$ for various values of $i$ and $j$. Note that $c_{i\to j} = c_{j\to i}$ when $j = i+1$.

- $A(id) = 1$ as both chains have the same probability of choosing $id$ and $id$ is written as itself
- $(j = i+1)$ put $k = i+2$. $c_{i\to i+1}$ is used only in the representations of $(ij)$ and $(i\,k)$. Hence
$$A(c_{i\to j}) = \frac{2}{n^2}\left(|(i\,j)|\frac{n^2}{2} + |(i\,k)|\frac{n^2}{2}\right) = 3$$
- $(j > i+1)$ put $k = i-1$. Then $c_{i\to j}$ is only used for the representation of $(k\,j)$ and hence
$$A(c_{i\to j}) = \frac{1}{n^2}\cdot 2 \cdot \frac{n^2}{2} = 1$$
- $(j = i-1)$ $c_{i\to j} = c_{j\to i}$
- $(j < i-1)$ $c_{i\to j}$ is only used in the representation of $(j\,i)$. Hence
$$A(c_{i\to j}) = \frac{1}{n^2}\cdot 2 \cdot \frac{n^2}{2} = 1$$

Hence we have $A = 3$. □

**Corollary A.4.** *Let $\mathbb{P}_{RR;I}$ denote the random-to-random insertion chain. Then $\mathcal{T}_2(\mathbb{P}_{RR;I}) \le 1.5n\log n + O(n)$ and $R_2^{\mathcal{H}}(\mathbb{P}_{RT;I}) \le 3n\log n + O(n)$.*

*Proof.* Follows from Lemma A.3 and the fact that $\mathbb{P}_{RR;I}$ is reversible and that the $L_2$ mixing time of $\mathbb{P}_{RR;T}$ is $\le 1.5n\log n + O(n)$.

From Theorem 1.23, we have $R_2^{\mathcal{H}}(\mathbb{P}_{RT;I}) \le 2\mathcal{T}_2(\mathbb{Q})$ where $\mathbb{Q} = \mathbb{P}_{RT;I}\overleftarrow{\mathbb{P}_{RT;I}} = \mathbb{P}_{RT;I}\mathbb{P}_{TR;I} = \mathbb{P}_{RR;I}$. Hence the result. □