Cutoff for a Stratified Random Walk on the Hypercube

Anna Ben-Hamou[1] and Yuval Peres[2]

1   IMPA, Rio de Janeiro, Brasil
`benhamou@impa.br`

2   Microsoft Research, Redmond, WA, USA
`peres@microsoft.com`

───── **Abstract** ─────

We consider the random walk on the hypercube which moves by picking an ordered pair (i, j) of distinct coordinates uniformly at random and adding the bit at location i to the bit at location j, modulo 2. We show that this Markov chain has cutoff at time $\frac{3}{2} n \log n$ with window of size n, solving a question posed by Chung and Graham (1997).

1998 ACM Subject Classification G.3 Probability

Keywords and phrases Mixing times, cutoff, hypercube

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2017.29

1   Introduction

Let $\mathrm{SL}_n(\mathbb{Z}_2)$ be the set of invertible matrices with coefficients in $\mathbb{Z}_2$, and consider the Markov chain on $\mathrm{SL}_n(\mathbb{Z}_2)$ which moves by picking two distinct rows at random and adding the first one to the other. This walk has received significant attention, both from group theoreticians and cryptologists. Diaconis and Saloff-Coste [4] showed that the ℓ_2-mixing time was $O(n^4)$, and the powerful results of Kassabov [5] yield the upper-bound $O(n^3)$. One may observe that if $Z_t \in \{0, 1\}^n \backslash \{\mathbf{0}\}$ denotes the first column of the matrix at time t, then the process $\{Z_t\}_{t \geq 0}$ is also a Markov chain (defined more precisely below). Diaconis and Saloff-Coste [4] showed that the log-Sobolev constant of this chain is $O(n^2)$, which yields an upper-bound of order $n^2 \log n$ on the ℓ_2-mixing time. They however conjectured that the right order for the total-variation mixing was $n \log n$. Chung and Graham [3] confirmed this conjecture. They showed that the relaxation time of $\{Z_t\}$ was of order n (which yields a tight upper-bound of order n^2 for ℓ_2-mixing) and that the total-variation mixing time $t_{\mathrm{mix}}(\varepsilon)$ was smaller than $c_\varepsilon n \log n$ for some constant c_ε. They asked whether one could make this bound more precise and replace c_ε by a universal constant which would not depend on ε. We answer this question positively by proving that the chain $\{Z_t\}$ has cutoff at time $\frac{3}{2} n \log n$, with window of order n.

The matrix walk problem was brought to our attention by Ron Rivest, who was mostly interested in computational mixing aspects [7]. The question of determining the total-variation mixing time of this walk is still largely open. By a diameter bound, it can be lower bounded by $\Omega\left(\frac{n^2}{\log n}\right)$ (see Andrén et al. [1], Christofides [2]). The best known upper-bound is $O(n^3)$ as established by Kassabov [5].

Main result

Let $\mathcal{X} = \{0, 1\}^n \backslash \{\mathbf{0}\}$ and consider the Markov chain $\{Z_t\}_{t \geq 0}$ on $\mathcal{X}$ defined as follows: if the current state is x and if $x(i)$ denotes the bit at the ith coordinate of x, then the walk

proceeds by choosing uniformly at random an ordered pair (i, j) of distinct coordinates, and replacing $x(j)$ by $x(j) + x(i)$ (mod 2).

The transition matrix P of this chain is symmetric, irreducible and aperiodic. Its stationary distribution π is the uniform distribution over $\mathcal{X}$, i.e. for all $x \in \mathcal{X}$, $\pi(x) = \frac{1}{2^n - 1}$. We are interested in the total-variation mixing time, defined as

$$t_{\mathrm{mix}}(\varepsilon) = \min \{ t \geq 0, \, d(t) \leq \varepsilon \} \,,$$

where $d(t) = \max_{x \in \mathcal{X}} d_x(t)$ and $d_x(t)$ is the total-variation distance between $P^t(x, \cdot)$ and π:

$$d_x(t) = \sup_{A \subset \mathcal{X}} \left(\pi(A) - P^t(x, A) \right) \; = \; \sum_{y \in \mathcal{X}} \left(P^t(x, y) - \pi(y) \right)_+ \,.$$

▶ **Theorem 1.** *The chain $\{Z_t\}$ has total-variation cutoff at time $\frac{3}{2} n \log n$ with window n, i.e.*

$$\lim_{\alpha \to +\infty} \liminf_{n \to +\infty} d \left(\frac{3}{2} n \log n - \alpha n \right) = 1 \,,$$

and

$$\lim_{\alpha \to +\infty} \limsup_{n \to +\infty} d \left(\frac{3}{2} n \log n + \alpha n \right) = 0 \,.$$

Before proving Theorem 1, we first state some useful properties of the birth-and-death chain given by the Hamming weight of Z_t. In particular, we show that this projected chain also has cutoff at $\frac{3}{2} n \log n$ (Section 2). Section 3 is then devoted to the proof of Theorem 1.

2   The Hamming weight

For a vertex $x \in \mathcal{X}$, we denote by $H(x)$ the Hamming weight of x, *i.e.*

$$H(x) = \sum_{i=1}^{n} x(i) \,.$$

Consider the birth-and-death chain $H_t := H(Z_t)$, and denote by P_H, π_H, and $d_H(\cdot)$ its transition matrix, stationary distribution, and total-variation distance to equilibrium. For $1 \leq k \leq n$, we have

$$P_H(k, k+1) = \frac{k(n-k)}{n(n-1)} \,,$$

$$P_H(k, k-1) = \frac{k(k-1)}{n(n-1)} \,,$$

$$P_H(k, k) = \frac{n-k}{n} \,,$$

and

$$\pi_H(k) = \frac{\binom{n}{k}}{2^n - 1} \,.$$

The hitting time of state k is defined as

$$T_k = \min \{ t \geq 0, \, H_t = k \} \,.$$

One standard result in birth-and-death chains is that, for $2 \leq \ell \leq n$,

$$\mathbb{E}_{\ell-1}(T_\ell) = \frac{1}{P_H(\ell, \ell-1)} \sum_{i=1}^{\ell-1} \frac{\pi_H(i)}{\pi_H(\ell)}, \tag{1}$$

(see for instance [6, Section 2.5]). The following lemma will be useful.

▶ **Lemma 2.** *Let* $0 < \beta < 1$ *and* $K = (1-\beta)\frac{n}{2}$. *Then there exist constants* $c_\beta, C_\beta \in \mathbb{R}$ *depending on* β *only such that*

$$\mathbb{E}_1(T_K) \leq n \log n + c_\beta n,$$

and

$$\mathrm{Var}_1 T_K \leq C_\beta n^2.$$

Proof of Lemma 2. For $2 \leq k \leq K$, let $\mu_k = \mathbb{E}_{k-1} T_k$ and $v_k = \mathrm{Var}_{k-1}(T_k)$. Resorting to (1), we have

$$\mu_k = \frac{\binom{n}{k-1}}{\binom{n-2}{k-2}} \sum_{i=1}^{k-1} \frac{\binom{n}{i}}{\binom{n}{k-1}} \leq \frac{\binom{n}{k-1}}{\binom{n-2}{k-2}} \sum_{i=1}^{k-1} \left(\frac{k-1}{n-k+2}\right)^{k-i-1} \leq \frac{n^2}{k(n-2k)} \tag{2}$$

Summing from 2 to K yields the desired bound on $\mathbb{E}_1 T_K$. Moving on to the variance, by independence of the successive hitting times, we have

$$\mathrm{Var}_1 T_K = \sum_{k=1}^{K-1} v_{k+1}.$$

Hence, it is sufficient to show that there exists a constant $a_\beta > 0$ such that $v_{k+1} \leq \frac{a_\beta n^2}{k^2}$ for all $k \leq K$. To do so, we consider the following distributional identity for the hitting time T_{k+1} starting from k:

$$T_{k+1} = 1 + (1-I)\widetilde{T}_{k+1} + IJ(\widehat{T}_k + \widehat{T}_{k+1}),$$

where I is the indicator that the chain moves (*i.e.* that a one is picked as updating coordinate), J is the indicator that the chain decreases given that it moves (*i.e.* that the chosen one is added to another one), $\widetilde{T}_{k+1}$ and $\widehat{T}_{k+1}$ are copies of T_{k+1}, and $\widehat{T}_k$ is the hitting time of k starting from $k-1$. All those variables may be assumed to be independent. After computation we obtain the following induction relation:

$$\begin{aligned} v_{k+1} &= \frac{k-1}{n-1}(v_k + v_{k+1}) + \left(1 - \frac{k}{n}\right)\mu_{k+1}^2 + \frac{k-1}{n-1}\left(1 - \frac{k(k-1)}{n(n-1)}\right)(\mu_k + \mu_{k+1})^2 \\ &\leq \frac{k}{n}(v_k + v_{k+1}) + \mu_{k+1}^2 + \frac{k}{n}(\mu_k + \mu_{k+1})^2. \end{aligned}$$

Using the fact that for all $k \leq K$, we have $\mu_k \leq \frac{n}{\beta k}$ (which can be seen by inequality (2)), and after some simplification, we get

$$v_{k+1} \leq \frac{k}{n-k} v_k + \frac{3n^3}{\beta^2 k^2(n-k)} \leq \frac{k}{n-k} v_k + \frac{6n^2}{\beta^2 k^2}.$$

By induction and using that $v_2 \leq n^2$, we obtain that $v_{k+1} \leq \frac{a_\beta n^2}{k^2}$ for all $k \leq K$. ◀

The following proposition establishes cutoff for the chain $\{H_t\}$ and will be used in the next section to prove cutoff for the chain $\{Z_t\}$.

▶ **Proposition 3.** *The chain H_t exhibits cutoff at time $\frac{3}{2}n\log n$ with window n.*

Proof. For the lower bound, we want to show that for $t = \frac{3}{2}n\log n - 2\alpha n$

$$d_H(t) \geq 1 - \varepsilon(\alpha),$$

where $\varepsilon(\alpha) \to 0$ as $\alpha \to +\infty$. Consider the chain started at $H_0 = 1$ and let $k = \frac{n}{2} - \alpha\sqrt{n}$ and $A = \{k, k+1, \ldots, n\}$. By definition of total-variation distance,

$$d_H(t) \geq \pi_H(A) - P_H^t(1, A) \geq \pi_H(A) - \mathbb{P}_1(T_k \leq t).$$

By the Central Limit Theorem, $\lim_{\alpha\to\infty} \lim_{n\to\infty} \pi_H(A) = 1$. Moving on to $\mathbb{P}_1(T_k \leq t)$, let us write

$$\mathbb{P}_1(T_k \leq t) = \mathbb{P}_1\left(T_{n/3} \leq n\log n - \alpha n\right) + \mathbb{P}_{n/3}\left(T_k \leq \frac{n\log n}{2} - \alpha n\right).$$

Note that $T_{n/3}$ is stochastically larger than $\sum_{i=1}^{n/3} G_i$, where $(G_i)_{i=1}^{n/3}$ are independent Geometric random variables with respective parameter i/n (this is because at each step, we need at least to pick a one to just move from the current position). By Chebyshev's Inequality,

$$\mathbb{P}_1\left(T_{n/3} \leq n\log n - \alpha n\right) = O\left(\frac{1}{\alpha^2}\right).$$

Now, starting from Hamming weight $n/3$ and up to time T_k, we may couple H_t with $\widetilde{H}_t$, the Hamming weight of the standard lazy random walk on the hypercube (at each step, pick a coordinate uniformly at random and randomize the bit at this coordinate), in such a way $T_k \geq S_k$, where $S_k = \inf\{t \geq 0, \widetilde{H}_t = k\}$. It is known that S_k satisfies

$$\mathbb{P}_{n/3}\left(S_k \leq \frac{n\log n}{2} - \alpha n\right) \leq \varepsilon(\alpha),$$

with $\varepsilon(\alpha) \to 0$ as $\alpha \to +\infty$ (see for instance the proof of [6, Proposition 7.13]), which concludes the proof of the lower bound.

For the upper bound, letting $t = \frac{3}{2}n\log n + 2\alpha n$, we have

$$d_H(t) \leq \mathbb{P}_1\left(T_{n/3} > n\log n + \alpha n\right) + \max_{k \geq n/3} d_H^{(k)}\left(\frac{n\log n}{2} + \alpha n\right). \tag{3}$$

Lemma 2 entails that $T_{n/3}$ concentrates well: $\mathbb{E}_1(T_{n/3}) = n\log n + cn$ for some absolute constant c, and $\mathrm{Var}_1(T_{n/3}) = O(n^2)$. By Chebyshev's Inequality,

$$\mathbb{P}_1\left(T_{n/3} > n\log n + \alpha n\right) = O\left(\frac{1}{\alpha^2}\right). \tag{4}$$

To control the second term in the right-hand side of (3), we use the coupling method (see Levin et al. [6, Chapter 5]). For all starting point $k \geq n/3$, we consider the following coupling between a chain H_t started at k and a chain H_t^π started from stationarity: at each step t, if H_t makes an actual move (a one is picked as updating bit in the underlying chain Z_t), we try "as much as possible" not to move H_t^π (picking a zero as updating bit). Conversely, when H_t does not move, we try "as much as possible" to move H_t^π, the goal being to increase the chance that the two chains do not cross each other. The chains stay

together once they have met for the first time. We claim that the study of the coupling time can be reduced to the study of the first time when the chain started at $n/3$ reaches $n/2$. Indeed, as $\pi_H([2n/3, n]) = o(1)$, with high probability, $H_0^\pi \leq 2n/3$, and as starting from a larger Hamming weight can only speed up the chain, $\mathbb{P}_{2n/3}(T_{n/2} > t) \leq \mathbb{P}_{n/3}(T_{n/2} > t)$. Now, when both chains have reached $n/2$, either they have met, or they have crossed each other. In this last situation, we know however that the expected time of their first return to $n/2$ is $O(\sqrt{n})$, so that $\mathbb{P}_{n/2}\left(T_{n/2}^+ > \sqrt{\alpha n}\right) = O(1/\sqrt{\alpha})$. Moreover, thanks to our coupling, during each of those excursions, the chains have positive probability to meet, so that after an additional time of order $\alpha\sqrt{n}$ we can guarantee that they have met with large probability. We are thus left to prove that $\mathbb{P}_{n/3}\left(T_{n/2} > \frac{n \log n}{2} + \alpha n\right) \leq \varepsilon(\alpha)$, for a function ε tending to 0 at $+\infty$.

Starting from $H_0 = n/3$, we first argue that H_t will remain above $2n/7$ for a very long time. Namely, defining $\mathcal{G}_t = \{T_{2n/7} > t\}$, we have

$$\mathbb{P}_{n/3}\left(\mathcal{G}_{n^2}\right) = 1 - o(1). \tag{5}$$

This can easily be seen by considering $T_k^+ = \min\{t \geq 1,\, H_t = k\}$ and taking a union bound over the excursions around $k = n/3$ which visit $m = 2n/7$:

$$\mathbb{P}_k(T_m \leq n^2) \leq n^2 \mathbb{P}_k(T_m \leq T_k^+),$$

and

$$\mathbb{P}_k(T_m \leq T_k^+) = \frac{\mathbb{E}_k(T_k^+)}{\mathbb{E}_m(T_k) + \mathbb{E}_k(T_m)} \leq \frac{\mathbb{E}_k(T_k^+)}{\mathbb{E}_m(T_m^+)} = \frac{\pi_H(m)}{\pi_H(k)},$$

which decreases exponentially fast in n.

Our goal now will be to analyse the tail of $\tau = \inf\{t \geq 0,\, D_t \leq 0\}$, where

$$D_t = \frac{n}{2} - H_t.$$

Observe that

$$D_{t+1} - D_t = \begin{cases} 1 & \text{with probability } \frac{H_t(H_t-1)}{n(n-1)} \\ -1 & \text{with probability } \frac{H_t(n-H_t)}{n(n-1)} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

We get

$$\mathbb{E}\left[D_{t+1} - D_t \,\middle|\, D_t\right] = -\frac{2\left(\frac{n}{2} - D_t\right)(D_t + 1)}{n(n-1)} \leq -\frac{D_t}{n} + \frac{2D_t^2}{n(n-1)}. \tag{7}$$

Writing a similar recursion for the second moment of D_t gives

$$\mathbb{E}\left[D_{t+1}^2 - D_t^2 \,\middle|\, D_t\right] = -\frac{4H_t D_t(D_t + 1/2)}{n(n-1)} + \frac{H_t}{n} \leq -\frac{4H_t D_t^2}{n^2} + 2.$$

On the event $\mathcal{G}_t$,

$$\mathbb{E}\left[D_{t+1}^2 - D_t^2 \,\middle|\, D_t\right] \leq -\frac{8D_t^2}{7n} + 2.$$

By induction, letting $\mathcal{D}_t = \mathbb{1}_{\mathcal{G}_t} D_t$ (and noticing that $\mathcal{G}_{t+1} \subset \mathcal{G}_t$), we get

$$\mathbb{E}\left[\mathcal{D}_t^2\right] \le \mathbb{E}[D_0^2]\left(1 - \frac{8}{7n}\right)^t + \frac{7n}{4} \le \frac{n^2}{4}\,\mathrm{e}^{-8t/7n} + 2n\,.$$

Plugging this back in (7),

$$\mathbb{E}\left[\mathcal{D}_{t+1}\right] \le \left(1 - \frac{1}{n}\right)\mathbb{E}\left[\mathcal{D}_t\right] + \mathrm{e}^{-8t/7n} + 4/n\,,$$

and by induction,

$$\mathbb{E}\left[\mathcal{D}_t\right] \le an\,\mathrm{e}^{-t/n} + b\,, \tag{8}$$

for absolute constants $a, b \ge 0$. Also, letting $\tau_\star = \inf\{t \ge 0,\, \mathcal{D}_t = 0\}$, we see by (6) that, provided $\tau_\star > t$, the process $\{\mathcal{D}_t\}$ is at least as likely to move downwards than to move upwards and that there exists a constant $\sigma^2 > 0$ such that $\mathrm{Var}\left(\mathcal{D}_{t+1} \mid \mathcal{D}_t\right) \ge \sigma^2$ (this is because, on $\mathcal{G}_t$ the probability to make a move at time t in larger than some positive absolute constant). By Levin et al. [6, Proposition 17.20], we know that for all $u > 0$ and $k \ge 0$,

$$\mathbb{P}_k(\tau_\star > u) \le \frac{4k}{\sigma\sqrt{u}}\,. \tag{9}$$

Now take $H_0 = n/3$, $D_0 = n/6$, $s = \frac{1}{2}n\log n$ and $u = \alpha n$. We have

$$\mathbb{P}_{D_0}\left(\tau > s + u\right) \le \mathbb{P}_{D_0}\left(\tau_\star > s + u\right) + \mathbb{P}_{H_0}\left(\mathcal{G}_{n^2}^c\right)\,.$$

By equation (5), $\mathbb{P}_{H_0}\left(\mathcal{G}_{n^2}^c\right) = o(1)$, and, combining (9) and (8), we have

$$\mathbb{P}_{D_0}\left(\tau_\star > s + u\right) = \mathbb{E}_{D_0}\left[\mathbb{P}_{\mathcal{D}_s}\left(\tau_\star > u\right)\right] \le \mathbb{E}_{D_0}\left[\frac{4\mathcal{D}_s}{\sigma\sqrt{u}}\right] = O\left(\frac{1}{\sqrt{\alpha}}\right)\,,$$

which implies

$$\max_{k \ge n/3} d_H^{(k)}\left(s + u\right) = O\left(\frac{1}{\sqrt{\alpha}}\right)\,, \tag{10}$$

and concludes the proof of the upper bound.   ◀

3 Proof of Theorem 1

First note that, as projections of chains can not increase total-variation distance, the lower bound on $d(t)$ readily follows from the lower bound on $d_H(t)$, as established in Proposition 3. Therefore, we only have to prove the upper bound.

Let $\mathcal{E} = \{x \in \mathcal{X},\, H(x) \ge n/3\}$ and $\tau_{\mathcal{E}}$ be the hitting time of set $\mathcal{E}$. For all $t, s > 0$, we have

$$d(t + s) \le \max_{x_0 \in \mathcal{X}} \mathbb{P}_{x_0}\left(\tau_{\mathcal{E}} > s\right) + \max_{x \in \mathcal{E}} d_x(t)\,.$$

By (4), taking $s = n\log n + \alpha n$, we have $\max_{x_0 \in \mathcal{X}} \mathbb{P}_{x_0}(\tau_{\mathcal{E}} > s) = O(1/\alpha^2)$, so that our task comes down to showing that for all $x \in \mathcal{E}$,

$$d_x\left(\frac{n\log n}{2} + \alpha n\right) \le \varepsilon(\alpha)\,,$$

with $\varepsilon(\alpha) \to 0$ as $\alpha \to +\infty$. Let us fix $x \in \mathcal{E}$. Without loss of generality, we may assume that x is the vertex with $\bar{x} \geq n/3$ ones on the first $\bar{x}$ coordinates, and $n - \bar{x}$ zeros on the last $n - \bar{x}$ coordinates. We denote by $\{Z_t\}$ the random walk started at $Z_0 = x$ and for a vertex $z \in \mathcal{X}$, we define a two-dimensional object $\mathbf{W}(z)$, keeping track of the number of ones within the first $\bar{x}$ and last $n - \bar{x}$ coordinates of z, that is

$$\mathbf{W}(z) = \left(\sum_{i=1}^{\bar{x}} z(i), \sum_{i=\bar{x}+1}^{n} z(i) \right).$$

The projection of $\{Z_t\}_{t\geq 0}$ induced by $\mathbf{W}$ will be denoted $\mathbf{W}_t = \mathbf{W}(Z_t) = (X_t, Y_t)$. We argue that the study of $\{Z_t\}_{t\geq 0}$ can be reduced to the study of $\{\mathbf{W}_t\}_{t\geq 0}$, and that, when coupling two chains distributed as $\mathbf{W}_t$, we can restrict ourselves to initial states with the same total Hamming weight. Indeed, letting $\nu_{\bar{x}}$ be the uniform distribution over $\{z \in \mathcal{X}, H(z) = \bar{x}\}$. By the triangle inequality,

$$d_x(t) \leq \left\| \mathbb{P}_x \left(Z_t \in \cdot \right) - \mathbb{P}_{\nu_{\bar{x}}} \left(Z_t \in \cdot \right) \right\|_{\mathrm{TV}} + \left\| \mathbb{P}_{\nu_{\bar{x}}} \left(Z_t \in \cdot \right) - \pi(\cdot) \right\|_{\mathrm{TV}} \tag{11}$$

Starting from $\nu_{\bar{x}}$, the conditional distribution of Z_t given $\{H(Z_t) = h\}$ is uniform over $\{y \in \mathcal{X}, H(y) = h\}$. This entails

$$\left\| \mathbb{P}_{\nu_{\bar{x}}} \left(Z_t \in \cdot \right) - \pi(\cdot) \right\|_{\mathrm{TV}} = \left\| \mathbb{P}_{\bar{x}} \left(H_t \in \cdot \right) - \pi_H(\cdot) \right\|_{\mathrm{TV}}.$$

For $t = \frac{n \log n}{2} + \alpha n$, we know by (10) in the proof of Proposition 3 that $\left\| \mathbb{P}_{\bar{x}} \left(H_t \in \cdot \right) - \pi_H(\cdot) \right\|_{\mathrm{TV}} = O(1/\sqrt{\alpha})$. As for the first term in the right-hand side of (11), note that if z and z' are two vertices such that $\mathbf{W}(z) = \mathbf{W}(z')$, then for all $t \geq 0$, $\mathbb{P}_x(Z_t = z) = \mathbb{P}_x(Z_t = z')$, and that for all $y \in \mathcal{X}$ such that $\mathbf{W}(y) = (k, \ell)$

$$\begin{aligned}
\mathbb{P}_{\nu_{\bar{x}}} \left(Z_t = y \right) &= \sum_{\substack{i,j \\ i+j=\bar{x}}} \sum_{z, \mathbf{W}(z)=(i,j)} \frac{1}{\binom{n}{\bar{x}}} \mathbb{P}_z \left(Z_t = y \right) \\
&= \sum_{\substack{i,j \\ i+j=\bar{x}}} \frac{\binom{\bar{x}}{i} \binom{n-\bar{x}}{j}}{\binom{n}{\bar{x}}} \sum_{z, \mathbf{W}(z)=(i,j)} \frac{\mathbb{P}_z \left(Z_t = y \right)}{\binom{\bar{x}}{i} \binom{n-\bar{x}}{j}} \\
&= \sum_{\substack{i,j \\ i+j=\bar{x}}} \frac{\binom{\bar{x}}{i} \binom{n-\bar{x}}{j}}{\binom{n}{\bar{x}}} \frac{\mathbb{P}_{(i,j)} \left(\mathbf{W}_t = (k, \ell) \right)}{\binom{\bar{x}}{k} \binom{n-\bar{x}}{\ell}}.
\end{aligned}$$

Hence,

$$\left\| \mathbb{P}_x \left(Z_t \in \cdot \right) - \mathbb{P}_{\nu_{\bar{x}}} \left(Z_t \in \cdot \right) \right\|_{\mathrm{TV}} \leq \max_{\substack{i,j \\ i+j=\bar{x}}} \left\| \mathbb{P}_{(\bar{x},0)} \left(\mathbf{W}_t \in \cdot \right) - \mathbb{P}_{(i,j)} \left(\mathbf{W}_t \in \cdot \right) \right\|_{\mathrm{TV}}.$$

Now let $y \in \mathcal{E}$ such that $H(y) = \bar{x}$, and consider the chains $Z_t, \widetilde{Z}_t$ started at x and y respectively. Let $\mathbf{W}(Z_t) = (X_t, Y_t)$ and $\mathbf{W}(\widetilde{Z}_t) = (\widetilde{X}_t, \widetilde{Y}_t)$. We couple Z_t and $\widetilde{Z}_t$ as follows: at each step t, provided $H(Z_t) = H(\widetilde{Z}_t)$ and $\mathbf{W}(Z_t) \neq \mathbf{W}(\widetilde{Z}_t)$, we consider a random permutation π_t which is such that $Z_t(i) = \widetilde{Z}_t(\pi_t(i))$ for all $1 \leq i \leq n$, that is, we pair uniformly at random the ones (resp. the zeros) of Z_t with the ones (resp. the zeros) of $\widetilde{Z}_t$. If Z_t moves to Z_{t+1} by choosing the pair (i_t, j_t) and updating $Z_t(j_t)$ to $Z_t(j_t) + Z_t(i_t)$, then we move from $\widetilde{Z}_t$ to $\widetilde{Z}_{t+1}$ by updating $\widetilde{Z}_t(\pi_t(j_t))$ to $\widetilde{Z}_t(\pi_t(j_t)) + \widetilde{Z}_t(\pi_t(i_t))$. Once $\mathbf{W}(Z_t) = \mathbf{W}(\widetilde{Z}_t)$, the permutation π_t is chosen in such a way that the ones in the top (resp. in the bottom) in Z_t are matched with the ones in the top (resp. in the bottom) in $\widetilde{Z}_t$,

guaranteeing that from that time $\mathbf{W}(Z_t)$ and $\mathbf{W}(\widetilde{Z}_t)$ remain equal. Note that this coupling ensures that for all $t \geq 0$, the Hamming weight of Z_t is equal to that of $\widetilde{Z}_t$, and we may unequivocally denote it by H_t. In particular, coupling of the chains $\mathbf{W}(Z_t)$ and $\mathbf{W}(\widetilde{Z}_t)$ occurs when X_t and $\widetilde{X}_t$ are matched. As $X_t \geq \widetilde{X}_t$ for all $t \geq 0$, we may consider

$$\tau = \inf\{t \geq 0, \mathbf{D}_t = 0\},$$

where $\mathbf{D}_t = X_t - \widetilde{X}_t$.

Before analysing the behaviour of $\{\mathbf{D}_t\}$, we first notice that the worst possible y for the coupling time satisfies $\mathbf{W}(y) = (\max\{0, 2\bar{x} - n\}, \min\{\bar{x}, n - \bar{x}\})$. We now fix y to be such a vertex, and show that, starting from x, y, the variables $\mathbf{W}(Z_t), \mathbf{W}(\widetilde{Z}_t)$ remain "nice" for a very long time. More precisely, defining

$$\mathcal{B}_t = \bigcap_{s=0}^{t} \left\{ H_s \geq 2n/7, \; X_s \geq \frac{\bar{x}}{p}, \; \widetilde{Y}_s \geq \frac{\min\{\bar{x}, n - \bar{x}\}}{p} \right\},$$

we claim that we can choose $p \geq 1$ fixed such that

$$\mathbb{P}_{x,y}\left(\mathcal{B}_{n^2}\right) = 1 - o(1). \tag{12}$$

Indeed, the fact that $\mathbb{P}_{n/3}(T_{2n/7} \leq n^2) = o(1)$ has already been established in the proof of Proposition 3 (equation (5)), and with the same kind of arguments, we show that $\mathbb{P}_{(\bar{x},0)}\left(\cup_{s=0}^{n^2}\{X_s < \bar{x}/p\}\right) = o(1)$. Letting $A = \{(\bar{x}/p, \ell), \ell = 0, \ldots, n - \bar{x}\}$, $\pi_{\mathbf{W}}$ be the stationary distribution of $\mathbf{W}_t$, and $k_{\bar{x}} = \min\left\{\frac{\bar{x}}{2}, \frac{n-\bar{x}}{2}\right\}$, we have

$$\mathbb{P}_{(\bar{x},0)}(T_A \leq n^2) \leq \mathbb{P}_{(\bar{x}/2,k_{\bar{x}})}(T_A \leq n^2) \; \leq \; n^2 \sum_{\ell=0}^{n-\bar{x}} \mathbb{P}_{(\bar{x}/2,k_{\bar{x}})}\left(T_{(\bar{x}/p,\ell)} \leq T^+_{(\bar{x}/2,k_{\bar{x}})}\right)$$

$$\leq n^2 \sum_{\ell=0}^{n-\bar{x}} \frac{\pi_{\mathbf{W}}(\bar{x}/p, \ell)}{\pi_{\mathbf{W}}(\bar{x}/2, k_{\bar{x}})} \; = \; \frac{n^2 2^{n-\bar{x}} \binom{\bar{x}}{\bar{x}/p}}{\binom{\bar{x}}{\bar{x}/2}\binom{n-\bar{x}}{k_{\bar{x}}}},$$

and we can choose p large enough such that this quantity decreases exponentially fast in n. Similarly, starting from y, the value of $\widetilde{Y}_s$ will remain at a high level for a very long time, establishing (12).

Let us now turn to the analysis of $\{\mathbf{D}_t\}$. On the event $\{t < \tau\}$,

$$\mathbf{D}_{t+1} - \mathbf{D}_t = \begin{cases} 1 & \text{with probability } p_1^t \\ -1 & \text{with probability } p_{-1}^t \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

where

$$p_1^t = \frac{H_t}{n} \cdot \frac{n - H_t}{n - 1} \cdot \frac{\bar{x} - X_t}{n - H_t} \cdot \frac{n - \bar{x} - \widetilde{Y}_t}{n - H_t} + \frac{H_t}{n} \cdot \frac{H_t - 1}{n - 1} \cdot \frac{Y_t}{H_t} \cdot \frac{\widetilde{X}_t}{H_t},$$

and

$$p_{-1}^t = \frac{H_t}{n} \cdot \frac{n - H_t}{n - 1} \cdot \frac{\bar{x} - \widetilde{X}_t}{n - H_t} \cdot \frac{n - \bar{x} - Y_t}{n - H_t} + \frac{H_t}{n} \cdot \frac{H_t - 1}{n - 1} \cdot \frac{X_t}{H_t} \cdot \frac{\widetilde{Y}_t}{H_t}.$$

After computation, we get, on $\{t < \tau\}$,

$$\mathbb{E}\left[\mathbf{D}_{t+1} - \mathbf{D}_t \mid Z_t, \widetilde{Z}_t\right] = -\frac{H_t \mathbf{D}_t}{n(n-1)}\left(1 + \frac{H_t - 1}{H_t}\right) \; \leq \; -\frac{\mathbf{D}_t}{n^2}(2H_t - 1) \tag{14}$$

From (14), it is not hard to see that the variable

$$M_t = \mathbb{1}_{\{\tau > t\}} \mathbf{D}_t \exp \left(\sum_{s=0}^{t-1} \frac{(2H_s - 1)}{n^2} \right)$$

is a super-martingale, which implies $\mathbb{E}_{x,y}[M_t] \leq \mathbb{E}_{x,y}[\mathbf{D}_0] \leq n$.

Now let $\tau_\star = \inf\{t \geq 0, \mathbb{1}_{\mathcal{B}_t} \mathbf{D}_t = 0\}$. By (13), we see that, provided $\{\tau_\star > t\}$, the process $\{\mathbb{1}_{\mathcal{B}_t} \mathbf{D}_t\}$ is a supermartingale ($p_{-1}^t \geq p_1^t$) and that there exists a constant $\sigma^2 > 0$ such that the conditional variance of its increments is larger than σ^2 (because on $\mathcal{B}_t$, the probability to make a move $p_{-1}^t + p_1^t$ is larger than some absolute constant). By Levin et al. [6, Proposition 17.20], for all $u > 0$ and $k \geq 0$,

$$\mathbb{P}_k(\tau_\star > u) \leq \frac{4k}{\sigma \sqrt{u}} \cdot \tag{15}$$

Now take $t = \frac{n \log n}{2}$ and $u = \alpha n$. We have

$$\mathbb{P}_{x,y}(\tau > t + u) \leq \mathbb{P}_{x,y}(\mathcal{B}_{n^2}^c) + \mathbb{P}_{x,y}(\tau_\star > t + u).$$

By (12), we know that $\mathbb{P}_{x,y}(\mathcal{B}_{n^2}^c) = o(1)$. Also, considering the event

$$\mathcal{A}_{t-1} = \left\{ \sum_{s=0}^{t-1} H_s \geq \frac{n^2 \log n}{4} - \beta n^2 \right\},$$

and resorting to (15), we get

$$\mathbb{P}_{x,y}(\tau_\star > t + u) \leq \mathbb{E}_{x,y}\left[\mathbb{1}_{\{\tau_\star > t\}} \mathbb{P}_{Z_t, \widetilde{Z}_t}(\tau_\star > u) \right]$$

$$\leq \mathbb{P}_{x,y}\left(\{\tau_\star > t\} \cap \mathcal{A}_{t-1}^c \right) + \mathbb{E}_{x,y}\left[\mathbb{1}_{\mathcal{A}_{t-1}} \mathbb{1}_{\{\tau_\star > t\}} \frac{4\mathbf{D}_t}{\sigma \sqrt{u}} \right].$$

On the one hand, recalling the notation and results of Section 2 (in particular equation (8)), and applying Markov's Inequality,

$$\mathbb{P}_{x,y}\left(\{\tau_\star > t\} \cap \mathcal{A}_{t-1}^c \right) \leq \mathbb{P}_{x,y}\left(\sum_{s=0}^{t-1} \mathcal{D}_s > \beta n^2 \right)$$

$$\leq \frac{1}{\beta n^2} \sum_{s=0}^{t-1} \left(a n\, e^{-s/n} + b \right) = O\left(\frac{1}{\beta} \right).$$

On the other hand,

$$\mathbb{E}_{x,y}\left[\mathbb{1}_{\{\tau_\star > t\}} \mathbb{1}_{\mathcal{A}_{t-1}} \mathbf{D}_t \right] \leq \exp\left(-\frac{\log n}{2} + \frac{t}{n^2} + 2\beta \right) \mathbb{E}_{x,y}[M_t] = O\left(e^{2\beta} \sqrt{n} \right).$$

In the end, we get

$$\mathbb{P}_{x,y}(\tau > t + u) = O\left(\frac{1}{\beta} + \frac{e^{2\beta}}{\sqrt{\alpha}} \right).$$

Taking for instance $\beta = \frac{1}{5} \log \alpha$ concludes the proof of Theorem 1.

References

1 Daniel Andrén, Lars Hellström, and Klas Markström. On the complexity of matrix reduction over finite fields. *Advances in applied mathematics*, 39(4):428–452, 2007.

2 Demetres Christofides. The asymptotic complexity of matrix reduction over finite fields. *arXiv preprint arXiv:1406.5826*, 2014.

3 Fan R. K. Chung and Ronald L. Graham. Stratified random walks on the n-cube. *Random Structures and Algorithms*, 11(3):199–222, 1997.

4 Persi Diaconis and Laurent Saloff-Coste. Walks on generating sets of abelian groups. *Probability theory and related fields*, 105(3):393–421, 1996.

5 Martin Kassabov. Kazhdan constants for $SL_n(\mathbb{Z})$. *International Journal of Algebra and Computation*, 15(05n06):971–995, 2005.

6 D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. AMS, 2009.

7 Aikaterini Sotiraki. *Authentication protocol using trapdoored matrices*. PhD thesis, Massachusetts Institute of Technology, 2016.