Supplementary Material of A Matrix Chernoff Bound for Markov Chains
and Its Application to Co-occurrence Matrices

A Convergence Rate of Co-occurrence Matrices

A.1 Proof of Claim 1

Claim 1 (Properties of Q). If P is a regular Markov chain, then Q satisfies:

1. Q is a regular Markov chain with stationary distribution \(\sigma_{(u_0, \ldots, u_T)} = \pi_{u_0} P_{u_0, u_1} \cdots P_{u_{T-1}, u_T}\);
2. The sequence \((X_1, \ldots, X_{L-T})\) is a random walk on Q starting from a distribution \(\rho\) such that \(\rho_{(u_0, \ldots, u_T)} = \phi_{u_0} P_{u_0, u_1} \cdots P_{u_{T-1}, u_T}\) and \(\|\rho\|_\sigma = \|\phi\|_\pi\);
3. \(\forall \delta > 0\), the \(\delta\)-mixing time of P and Q satisfies \(\tau(Q) < \tau(P) + T\);
4. \(\exists P\) with \(\lambda(P) < 1\) s.t. the induced Q has \(\lambda(Q) = 1\), i.e. Q may have zero spectral gap.

Proof. We prove the four parts of this Claim one by one.

Part 1 To prove Q is regular, it is sufficient to show that \(\exists N_1, \forall n_1 > N_1\), \((v_0, \ldots, v_T)\) can reach \((u_0, \ldots, u_T)\) at \(n_1\) steps. We know P is a regular Markov chain, so there exists \(N_2 \geq T\) s.t., for any \(n_2 \geq N_2, v_T\) can reach \(u_0\) at exact \(n_2\) step, i.e., there is a \(n_2\)-step walk s.t. \((v_T, w_{T-1}, \ldots, w_{n_2-1}, u_0)\) on P. This induces an \(n_2\)-step walk from \((v_0, \ldots, v_T)\) to \((w_{n_2-T+1}, \ldots, w_{n_2-1}, u_0)\). Take further \(T\) step, we can reach \((u_0, \ldots, u_T)\), so we construct a \(n_1 = n_2 + T\) step walk from \((v_0, \ldots, v_T)\) to \((u_0, \ldots, u_T)\). Since this is true for any \(n_2 \geq N_2\), we then claim that any state can be reached from any other state in any number of steps greater than or equal to a number \(N_1 = N_2 + T\).

Next to verify \(\sigma\) such that \(\sigma_{(u_0, \ldots, u_T)} = \pi_{u_0} P_{u_0, u_1} \cdots P_{u_{T-1}, u_T}\) is the stationary distribution of Markov chain Q,

\[
\sum_{(u_0, \ldots, u_T) \in S} \sigma_{(u_0, \ldots, u_T)} Q_{(u_0, \ldots, u_T), (u_0, \ldots, u_T)} = \sum_{u_0: (u_0, u_0, \ldots, u_{T-1}) \in S} \pi_{u_0} P_{u_0, u_0} P_{u_0, u_1} \cdots P_{u_{T-2}, u_{T-1}} P_{u_{T-1}, u_T} = \sigma_{u_0, u_0, \ldots, u_T}
\]

Part 2 Recall \((v_1, \ldots, v_L)\) is a random walk on P starting from distribution \(\phi\), so the probability we observe \(X_1 = (v_1, \ldots, v_{T+1})\) is \(\phi_{v_1} P_{v_1, v_2} \cdots P_{v_{T+1}, v_T} = \rho_{(v_1, \ldots, v_{T+1})}\), i.e., \(X_1\) is sampled from the distribution \(\rho\). Then we study the transition probability from \(X_i = (v_i, \ldots, v_{i+T})\) to \(X_{i+1} = (v_{i+1}, \ldots, v_{i+T+1})\), which is \(P_{v_i, v_{i+T+1}} = Q_{X_i, X_{i+1}}\). Consequently, we can claim \((X_1, \ldots, X_{L-T})\) is a random walk on Q. Moreover,

\[
\|\rho\|_\sigma^2 = \sum_{(u_0, \ldots, u_T) \in S} \rho_{(u_0, \ldots, u_T)}^2 = \sum_{(u_0, \ldots, u_T) \in S} (\phi_{u_0} P_{u_0, u_1} \cdots P_{u_{T-1}, u_T})^2
\]

\[
= \sum_{u_0} \phi_{u_0}^2 \sum_{(u_0, u_1, \ldots, u_T) \in S} P_{u_0, u_1} \cdots P_{u_{T-1}, u_T} = \sum_{u_0} \frac{\phi_{u_0}^2}{\pi_{u_0}} \|\phi\|_\pi^2,
\]

which implies \(\|\rho\|_\sigma = \|\phi\|_\pi\).

Part 3 For any distribution \(y\) on \(S\), define \(x \in \mathbb{R}^n\) such that \(x_i = \sum_{(v_1, \ldots, v_{T-1}, i) \in S} y_{v_1, \ldots, v_{T-1}, i}\).
Easy to see \(x\) is a probability vector, since \(x\) is the marginal probability of \(y\). For convenience, we
assume for a moment the \( x, y, \sigma, \pi \) are row vectors. We can see that:

\[
\|yQ^{(P)+T-1} - \sigma\|_{TV} = \frac{1}{2} \|yQ^{(P)+T-1} - \sigma\|_1 \\
= \frac{1}{2} \sum_{(u_1, \ldots, u_T) \in S} \left| yQ^{(P)+T-1} - \sigma \right|_{u_1, \ldots, u_T} \\
= \frac{1}{2} \sum_{(u_1, \ldots, u_T) \in S} \left| xP^{(P)} v_1 P_{v_1,v_2} \cdots P_{v_{T-1},v_T} - \pi v_1 P_{v_1,v_2} \cdots P_{v_{T-1},v_T} \right| \\
= \frac{1}{2} \sum_{v_1} \left| xP^{(P)} v_1 - \pi v_1 \right| \sum_{(u_1, \ldots, u_T) \in S} P_{v_1,v_2} \cdots P_{v_{T-1},v_T} \\
= \frac{1}{2} \sum_{v_1} \left| xP^{(P)} v_1 - \pi v_1 \right| = \frac{1}{2} \|xP^{(P)} - \pi\|_1 = \|xP^{(P)} - \pi\|_{TV} \leq \delta.
\]

which indicates \( \tau(Q) \leq \tau(P) + T - 1 < \tau(P) + T \).

**Part 4** This is an example showing that \( \lambda(Q) \) cannot be bounded by \( \lambda(P) \) — even though \( P \) has \( \lambda(P) < 1 \), the induced \( Q \) may have \( \lambda(Q) = 1 \). We consider random walk on the unweighted undirected graph \( S \) and \( T = 1 \). The transition probability matrix \( P \) is:

\[
P = \begin{bmatrix}
0 & 1/3 & 1/3 & 1/3 \\
1/2 & 0 & 1/2 & 0 \\
1/3 & 1/3 & 0 & 1/3 \\
1/2 & 0 & 1/2 & 0
\end{bmatrix}
\]

with stationary distribution \( \pi = [0.3 \quad 0.2 \quad 0.3 \quad 0.2]^{T} \) and \( \lambda(P) = \frac{3}{4} \). When \( T = 1 \), the induced Markov chain \( Q \) has stationary distribution \( \sigma_{u,v} = \pi_u P_{u,v} = \frac{d_u}{2m} \frac{1}{2} = \frac{1}{2m} \) where \( m = 5 \) is the number of edges in the graph. Construct \( y \in \mathbb{R}^{|S|} \) such that

\[
y(u,v) = \begin{cases} 
1 & (u,v) = (0,1), \\
-1 & (u,v) = (0,3), \\
0 & \text{otherwise}.
\end{cases}
\]

The constructed vector \( y \) has norm

\[
\|y\|_2 = \sqrt{(y,y)}_\sigma = \sqrt{\sum_{(u,v) \in S} y(u,v)y(u,v)} = \sqrt{\frac{y(0,1)y(0,1)}{\sigma(0,1)} + \frac{y(0,3)y(0,3)}{\sigma(0,3)}} = 2\sqrt{m}.
\]

And it is easy to check \( y \perp \sigma \), since \( \langle y, \sigma \rangle = \sum_{(u,v) \in S} \sigma_{u,v} y(u,v) = y(0,1) + y(0,3) = 0 \). Let \( x = (y^*Q)^* \), we have for \( (u,v) \in S \):

\[
x(u,v) = \begin{cases} 
1 & (u,v) = (1,2), \\
-1 & (u,v) = (3,2), \\
0 & \text{otherwise}.
\end{cases}
\]

This vector has norm:

\[
\|x\|_2 = \sqrt{(x,x)}_\sigma = \sqrt{\sum_{(u,v) \in S} x(u,v)x(u,v)} = \sqrt{\frac{y(1,2)y(1,2)}{\sigma(1,2)} + \frac{y(3,2)y(3,2)}{\sigma(3,2)}} = 2\sqrt{m}.
\]

Thus we have \( \frac{\|y^*Q\|_2}{\|y\|_2} = 1 \). Taking maximum over all possible \( y \) gives \( \lambda(Q) \geq 1 \). Also note that fact that \( \lambda(Q) \leq 1 \), so \( \lambda(Q) = 1 \).

**A.2 Proof of Claim 2**

**Claim 2** (Properties of \( f \). The function \( f \) in Equation \( 2 \) satisfies (1) \( \sum_{X \in S} \sigma_X f(X) = 0; \) (2) \( f(X) \) is symmetric and \( \|f(X)\|_2 \leq 1, \forall X \in S. \)
Proof. Note that Equation 2 is indeed a random value minus its expectation, so naturally Equation 2 has zero mean, i.e., \( \sum_{x \in \mathcal{X}} \sigma_x f(X) = 0 \). Moreover, \( \|f(X)\|_2 \leq 1 \) because
\[
\|f(X)\|_2 \leq \frac{1}{2} \left( \sum_{r=1}^T \frac{|\alpha_r|}{2} \left( \|e_v \circ e_v^\top\|_2 + \|e_v \circ e_v^\top\Pi\|_2 \right) + \sum_{r=1}^T \frac{|\alpha_r|}{2} \left( \|\Pi\|_2 \|P\|_2 + \|P\|_2 \|\Pi\|_2 \right) \right)
\leq \frac{1}{2} \left( \sum_{r=1}^T |\alpha_r| + \sum_{r=1}^T |\alpha_r| \right) = 1.
\]
where the first step follows triangle inequality and submultiplicativity of 2-norm, and the third step follows by (1) \( \|e_v \circ e_v^\top\|_2 = 1 \); (2) \( \|\Pi\|_2 = \|\text{diag}(\pi)\|_2 \leq 1 \) for distribution \( \pi \); (3) \( \|P\|_2 = \|P\Pi\|_2 = 1 \). \(\square\)

A.3 Proof of Corollary 1

Corollary 1 (Co-occurrence Matrices of HMMs). For a HMM with observable states \( y_t \in \mathcal{Y} \) and hidden states \( x_t \in \mathcal{X} \), let \( P(y_t|x_t) \) be the emission probability and \( P(x_{t+1}|x_t) \) be the hidden state transition probability. Given an \( L \)-step trajectory observations from the HMM, \( (y_1, \ldots, y_L) \), one needs a trajectory of length \( L = O(\tau(\log |\mathcal{Y}| + \log \tau)/\epsilon^2) \) to achieve a co-occurrence matrix within error bound \( \epsilon \) with high probability, where \( \tau \) is the mixing time of the Markov chain on hidden states.

Proof. A HMM can be model by a Markov chain \( P \) on \( \mathcal{Y} \times \mathcal{X} \) such that \( P(y_{t+1}, x_{t+1}|y_t, x_t) = P(y_{t+1}|x_{t+1})P(x_{t+1}|x_t) \). For the co-occurrence matrix of observable states, applying a similar proof like our Theorem 2 shows that one needs a trajectory of length \( O(\tau(P)(\log |\mathcal{Y}| + \log \tau(P))/\epsilon^2) \) to achieve error bound \( \epsilon \) with high probability. Moreover, the mixing time \( \tau(P) \) is bounded by the mixing time of the Markov chain on the hidden state space (i.e., \( P(x_{t+1}|x_t) \)). \(\square\)

B Matrix Chernoff Bounds for Markov Chains

B.1 Preliminaries

Kronecker Products If \( A \) is an \( M_1 \times N_1 \) matrix and \( B \) is a \( M_2 \times N_2 \) matrix, then the Kronecker product \( A \otimes B \) is the \( M_2M_1 \times N_2N_1 \) block matrix such that
\[
A \otimes B = \begin{bmatrix}
A_{1,1}B & \cdots & A_{1,N_1}B \\
\vdots & \ddots & \vdots \\
A_{M_1,1}B & \cdots & A_{M_1,N_1}B
\end{bmatrix}.
\]
Kronecker product has the mixed-product property. If \( A, B, C, D \) are matrices of such size that one can from the matrix products \( AC \) and \( BD \), then \( (A \otimes B)(C \otimes D) = (AC) \otimes (BD) \).

Vectorization For a matrix \( X \in \mathbb{C}^{d \times d} \), vec(\( X \) \( ) \in \mathbb{C}^{d^2} \) denote the vectorization of the matrix \( X \), s.t. vec(\( X \) \( ) = \sum_{i \in [d]} \sum_{j \in [d]} X_{i,j} \circ e_i \otimes e_j \), which is the stack of rows of \( X \). And there is a relationship between matrix multiplication and Kronecker product s.t. vec(\( AXB \) \( ) = (A \otimes B^\top) \text{vec}(X) \).

Matrices and Norms For a matrix \( A \in \mathbb{C}^{N \times N} \), we use \( A^\top \) to denote matrix transpose, use \( \overline{A} \) to denote entry-wise matrix conjugation, use \( A^* \) to denote matrix conjugate transpose \( (A^* = \overline{A}^\top) \). The vector 2-norm is defined to be \( \|x\|_2 = \sqrt{x^\top x} \), and the matrix 2-norm is defined to be \( \|A\|_2 = \max_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \).

We then recall the definition of inner-product under \( \pi \)-kernel in Section 2. The inner-product under \( \pi \)-kernel for \( \mathbb{C}^N \) is \( \langle x, y \rangle_{\pi} = y^\top \Pi^{-1} x \) where \( \Pi = \text{diag}(\pi) \), and its induced \( \pi \)-norm \( \|x\|_{\pi} = \sqrt{\langle x, x \rangle_{\pi}} \). The above definition allow us to define a inner product under \( \pi \)-kernel on \( \mathbb{C}^{Nd^2} \):

Definition 1. Define inner product on \( \mathbb{C}^{Nd^2} \) under \( \pi \)-kernel to be \( \langle x, y \rangle_{\pi} = y^\top (\Pi^{-1} \otimes I_{d^2}) x \).

Remark 1. For \( x, y \in \mathbb{C}^N \) and \( p, q \in \mathbb{C}^2 \), then inner product (under \( \pi \)-kernel) between \( x \otimes p \) and \( y \otimes q \) can be simplified as
\[
\langle x \otimes p, y \otimes q \rangle_{\pi} = (y \otimes q)^\top (\Pi^{-1} \otimes I_{d^2}) (x \otimes p) = (y^\top \Pi^{-1} x) \otimes (q^\top p) = \langle x, y \rangle_{\pi}(p, q).
\]
Remark 2. The induced $\pi$-norm is $\|x\|_\pi = \sqrt{\langle x, x \rangle_\pi}$. When $x = y \otimes w$, the $\pi$-norm can be simplified to be: $\|x\|_\pi = \sqrt{\langle y \otimes w, y \otimes w \rangle_\pi} = \sqrt{\langle y, y \rangle_\pi \langle w, w \rangle_2} = \|y\|_\pi \|w\|_2$.

Matrix Exponential The matrix exponential of a matrix $A \in \mathbb{C}^{d \times d}$ is defined by Taylor expansion $\exp(A) = \sum_{j=0}^{+\infty} \frac{A^j}{j!}$. And we will use the fact that $\exp(A) \otimes \exp(B) = \exp(A \otimes I + I \otimes B)$.

Golden-Thompson Inequality We need the following multi-matrix Golden-Thompson inequality from from Garg et al. [10].

**Theorem 4** (Multi-matrix Golden-Thompson Inequality, Theorem 1.5 in [10]). Let $H_1, \ldots, H_k$ be $k$ Hermitian matrices, then for some probability distribution $\mu$ on $[-\frac{\pi}{2}, \frac{\pi}{2}]$, \[
\log \left( \text{Tr} \left[ \exp \left( \sum_{j=1}^k H_j \right) \right] \right) \leq \frac{4}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \log \left( \text{Tr} \left[ \prod_{j=1}^k \exp \left( \frac{e^{i\phi}}{2} H_j \right) \prod_{j=k}^1 \exp \left( \frac{e^{-i\phi}}{2} H_j \right) \right] \right) d\mu(\phi).
\]

**B.2 Proof of Theorem 3**

**Theorem 3** (A Real-Valued Version of Theorem 1). Let $P$ be a regular Markov chain with state space $[N]$, stationary distribution $\pi$ and spectral expansion $\lambda$. Let $f : [N] \rightarrow \mathbb{R}^{d \times d}$ be a function such that (1) $\forall v \in [N], f(v)$ is symmetric and $\|f(v)\|_2 \leq 1$; (2) $\sum_{v \in [N]} \pi_v f(v) = 0$. Let $(v_1, \ldots, v_k)$ denote a $k$-step random walk on $P$ starting from a distribution $\phi$ on $[N]$. Then given $\epsilon \in (0, 1)$,

$$
\mathbb{P} \left[ \lambda_{\max} \left( \frac{1}{k} \sum_{j=1}^k f(v_j) \right) \geq \epsilon \right] \leq \|\phi\|_\pi d^2 \exp \left( -(c^2 (1 - \lambda) k/2) \right)
$$

$$
\mathbb{P} \left[ \lambda_{\min} \left( \frac{1}{k} \sum_{j=1}^k f(v_j) \right) \leq -\epsilon \right] \leq \|\phi\|_\pi d^2 \exp \left( -(c^2 (1 - \lambda) k/2) \right).
$$

**Proof.** Due to symmetry, it suffices to prove one of the statements. Let $t > 0$ be a parameter to be chosen later. Then

$$
\mathbb{P} \left[ \lambda_{\max} \left( \frac{1}{k} \sum_{j=1}^k f(v_j) \right) \geq \epsilon \right] = \mathbb{P} \left[ \lambda_{\max} \left( \sum_{j=1}^k f(v_j) \right) \geq k\epsilon \right]
$$

$$
\leq \mathbb{P} \left[ \text{Tr} \left[ \exp \left( t \sum_{j=1}^k f(v_j) \right) \right] \geq \exp(t k\epsilon) \right] \quad (3)
$$

The second inequality follows Markov inequality.

Next to bound $\mathbb{E}_{v_1, \ldots, v_k} \left[ \text{Tr} \left[ \exp \left( t \sum_{j=1}^k f(v_j) \right) \right] \right]$. Using Theorem 4, we have:

$$
\log \left( \text{Tr} \left[ \exp \left( t \sum_{j=1}^k f(v_j) \right) \right] \right) \leq \frac{4}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \log \left( \text{Tr} \left[ \prod_{j=1}^k \exp \left( \frac{e^{i\phi}}{2} f(v_j) \right) \prod_{j=k}^1 \exp \left( \frac{e^{-i\phi}}{2} f(v_j) \right) \right] \right) d\mu(\phi)
$$

$$
\leq \frac{4}{\pi} \log \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \text{Tr} \left[ \prod_{j=1}^k \exp \left( \frac{e^{i\phi}}{2} f(v_j) \right) \prod_{j=k}^1 \exp \left( \frac{e^{-i\phi}}{2} f(v_j) \right) \right] d\mu(\phi),
$$

where the second step follows by concavity of log function and the fact that $\mu(\phi)$ is a probability distribution on $[-\frac{\pi}{2}, \frac{\pi}{2}]$. This implies

$$
\text{Tr} \left[ \exp \left( t \sum_{j=1}^k f(v_j) \right) \right] \leq \left( \frac{e^{\frac{\pi}{4}}}{\pi} \right)^k \text{Tr} \left[ \prod_{j=1}^k \exp \left( \frac{e^{i\phi}}{2} f(v_j) \right) \prod_{j=k}^1 \exp \left( \frac{e^{-i\phi}}{2} f(v_j) \right) \right] d\mu(\phi) \frac{4}{\pi}.
$$

Note that $\|x\|_p \leq d^{1/p-1} \|x\|_1$ for $p \in (0, 1)$, choosing $p = \pi/4$ we have

$$
\left( \text{Tr} \left[ \exp \left( \frac{\pi}{4} t \sum_{j=1}^k f(v_j) \right) \right] \right)^{\frac{4}{\pi}} \leq d^{\frac{4}{\pi}-1} \text{Tr} \left[ \exp \left( t \sum_{j=1}^k f(v_j) \right) \right].
$$
Combining the above two equations together, we have
\[
\text{Tr} \left[ \exp \left( \frac{\pi}{4} t \sum_{j=1}^{k} f(v_j) \right) \right] \leq \left( d - \frac{2}{3} \right) \text{Tr} \left[ \prod_{j=1}^{k} \exp \left( \frac{e^{i\phi}}{2} t f(v_j) \right) \prod_{j=k}^{1} \exp \left( \frac{e^{-i\phi}}{2} t f(v_j) \right) \right] d\mu(\phi). \quad (4)
\]

Write \( e^{i\phi} = \gamma + ib \) with \( \gamma^2 + b^2 = |\gamma + ib|^2 = |e^{i\phi}|^2 = 1: 

**Lemma 1** (Analogous to Lemma 4.3 in [10]). Let \( P \) be a regular Markov chain with state space \([N]\) with spectral expansion \( \lambda \). Let \( f \) be a function \( f : [N] \rightarrow \mathbb{R}^{d \times d} \) such that (1) \( \sum_{v \in [N]} \pi(v) f(v) = 0; \) (2) \( \|f(v)\|_2 \leq 1 \) and \( f(v) \) is symmetric, \( v \in [N]. \) Let \( (v_1, \ldots, v_k) \) denote a \( k \)-step random walk on \( P \) starting from a distribution \( \phi \) on \([N]\). Then for any \( t > 0, \gamma \geq 0, b > 0 \) such that \( t^2(\gamma^2 + b^2) \leq 1 \) and \( t \sqrt{\gamma^2 + b^2} \leq \frac{1 - \lambda}{4\lambda}, \) we have
\[
E \left[ \text{Tr} \left[ \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \prod_{j=k}^{1} \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) \right] \right] \leq \|\phi\|_w d \exp \left( k t^2(\gamma^2 + b^2) \left( 1 + \frac{8}{1 - \lambda} \right) \right).
\]

Assuming the above lemma, we can complete the proof of the theorem as:
\[
\begin{align*}
E_{v_1, \ldots, v_k} \left[ \text{Tr} \left[ \exp \left( \frac{\pi}{4} t \sum_{j=1}^{k} f(v_j) \right) \right] \right] \\
\leq d^{1 - \frac{2}{3}} \text{Tr} \left[ \prod_{j=1}^{k} \exp \left( \frac{e^{i\phi}}{2} t f(v_j) \right) \prod_{j=k}^{1} \exp \left( \frac{e^{-i\phi}}{2} t f(v_j) \right) \right] d\mu(\phi) \\
= d^{1 - \frac{2}{3}} \int_{\mathbb{R}^2} \|\phi\|_w d \exp \left( k t^2 |e^{i\phi}|^2 \left( 1 + \frac{8}{1 - \lambda} \right) \right) d\mu(\phi) \\
= \|\phi\|_w d^{2 - \frac{2}{3}} \exp \left( k t^2 \left( 1 + \frac{8}{1 - \lambda} \right) \right) \int_{\mathbb{R}^2} d\mu(\phi) \\
= \|\phi\|_w d^{2 - \frac{2}{3}} \exp \left( k t^2 \left( 1 + \frac{8}{1 - \lambda} \right) \right)
\end{align*}
\]
where the first step follows Equation 4, the second step follows by swapping \( E \) and \( \int \), the third step follows by Lemma 1, the fourth step follows by \( |e^{i\phi}| = 1 \), and the last step follows by \( \mu \) is a probability distribution on \([\frac{-\lambda}{4}, \frac{\lambda}{4}]\) so \( \int_{\frac{-\lambda}{4}}^{\frac{\lambda}{4}} d\mu(\phi) = 1 \).

Finally, putting it all together:
\[
P \left[ \lambda_{\text{max}} \left( \frac{1}{k} \sum_{j=1}^{k} f(v_j) \right) \geq \epsilon \right] \leq \frac{E \left[ \text{Tr} \left[ \exp \left( t \sum_{j=1}^{k} f(v_j) \right) \right] \right]}{\exp(tk)} \\
\leq \frac{E \left[ \text{Tr} \left[ \exp \left( \frac{\pi}{4} t \left( \frac{1}{k} \sum_{j=1}^{k} f(v_j) \right) \right) \right] \right]}{\exp(tk)} \\
\leq \frac{\|\phi\|_w d^{2 - \frac{2}{3}} \exp \left( k \left( \frac{1}{k} \right)^2 \left( 1 + \frac{8}{1 - \lambda} \right) \right)}}{\exp(tk)} \\
\leq \frac{\|\phi\|_w d^{2 - \frac{2}{3}} \exp \left( \frac{k}{\pi} (1 - \lambda) \epsilon^2 \frac{1}{36} \frac{9}{1 - \lambda} - k \frac{1 - \lambda}{36} \right)}{\exp(tk)} \\
\leq \frac{\|\phi\|_w d^{2} \exp \left( -k \epsilon^2 (1 - \lambda)/72 \right)}{\exp(tk)}.
\]
where the first step follows by Equation 3, the second step follows by Equation 4, the third step follows by choosing \( t = (1 - \lambda)\epsilon^2/36 \). The only thing to be check is that \( t = (1 - \lambda)\epsilon^2/36 \) satisfies \( t \sqrt{\gamma^2 + b^2} = t \leq \frac{1 - \lambda}{4\lambda} \). Recall that \( \epsilon < 1 \) and \( \lambda \leq 1 \), we have \( t = \frac{(1 - \lambda)\epsilon^2}{36} \leq \frac{1 - \lambda}{4} \leq \frac{1 - \lambda}{4\lambda} \). \( \square \)
B.3 Proof of Lemma 1

Lemma 1 (Analogous to Lemma 4.3 in [10]). Let $P$ be a regular Markov chain with state space $[N]$ with spectral expansion $\lambda$. Let $f$ be a function $f : [N] \rightarrow \mathbb{R}^{d \times d}$ such that (1) $\sum_{v \in [N]} \pi_v f(v) = 0$; (2) $\|f(v)\|_2 \leq 1$ and $f(v)$ is symmetric, $v \in [N]$. Let $(v_1, \cdots, v_k)$ denote a $k$-step random walk on $P$ starting from a distribution $\phi$ on $[N]$. Then for any $t > 0, \gamma \geq 0, b > 0$ such that $t^2(\gamma^2 + b^2) \leq 1$ and $t\sqrt{\gamma^2 + b^2} \leq \frac{1-\lambda}{4\lambda}$, we have

$$E \left[ \text{Tr} \left[ \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) \right] \right] \leq \left\| \phi \right\|_d \exp \left( kt^2(\gamma^2 + b^2) \left( 1 + \frac{8}{1-\lambda} \right) \right).$$

Proof. Note that for $A, B \in \mathbb{C}^{d \times d}$, $\langle (A \otimes B) \text{vec}(I_d), \text{vec}(I_d) \rangle = \text{Tr} \left[ AB^\top \right]$. By letting $A = \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right)$ and $B = \left( \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) \right)^\top$, the trace term in LHS of Lemma 1 becomes

$$\text{Tr} \left[ \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) \right] = \left\langle \left( \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \otimes \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) \right) \text{vec}(I_d), \text{vec}(I_d) \right\rangle. \quad (6)$$

By iteratively applying $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, we have

$$\prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \otimes \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) = \prod_{j=1}^{k} \left( \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \otimes \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) \right) \triangleq \prod_{j=1}^{k} M_{v_j},$$

where we define

$$M_{v_j} \triangleq \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \otimes \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right). \quad (7)$$

Plug it to the trace term, we have

$$\text{Tr} \left[ \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) \right] = \left\langle \left( \prod_{j=1}^{k} M_{v_j} \right) \text{vec}(I_d), \text{vec}(I_d) \right\rangle.$$

Next, taking expectation on Equation 6 gives

$$E_{v_1, \cdots, v_k} \left[ \text{Tr} \left[ \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma + ib)}{2} \right) \prod_{j=1}^{k} \exp \left( \frac{tf(v_j)(\gamma - ib)}{2} \right) \right] \right] = \left\langle \left( \prod_{j=1}^{k} M_{v_j} \right) \text{vec}(I_d), \text{vec}(I_d) \right\rangle. \quad (8)$$

We turn to study $E_{v_1, \cdots, v_k} \left[ \prod_{j=1}^{k} M_{v_j} \right]$, which is characterized by the following lemma:

Lemma 2. Let $E \triangleq \text{diag}(M_1, M_2, \cdots, M_N) \in \mathbb{C}^{N^2 \times N^2}$ and $\tilde{P} \triangleq P \otimes I_{d^2} \in \mathbb{R}^{N^2 \times N^2}$. For a random walk $(v_1, \cdots, v_k)$ such that $v_j$ is sampled from an arbitrary probability distribution $\phi$ on $[N]$, $E_{v_1, \cdots, v_k} \left[ \prod_{j=1}^{k} M_{v_j} \right] = (\phi \otimes I_{d^2})^\top \left( (EP)^{k-1} E \right) (1 \otimes I_{d^2})$, where $1$ is the all-ones vector.

Proof. (of Lemma 2) We always treat $EP$ as a block matrix, s.t.,

$$EP = \begin{bmatrix}
M_1 & \cdots & M_1M_1 \\
\vdots & \ddots & \vdots \\
M_N & \cdots & M_NM_N
\end{bmatrix} = \begin{bmatrix}
P_{1,1,I_{d^2}} & \cdots & P_{1,N,I_{d^2}} \\
\vdots & \ddots & \vdots \\
P_{N,1,I_{d^2}} & \cdots & P_{N,N,I_{d^2}}
\end{bmatrix} = \begin{bmatrix}
P_{1,1}M_1 & \cdots & P_{1,N}M_1 \\
\vdots & \ddots & \vdots \\
P_{N,1}M_N & \cdots & P_{N,N}M_N
\end{bmatrix}.$$
I.e., the \((u, v)\)-th block of \(\mathbf{E} \mathbf{P}\), denoted by \((\mathbf{E} \mathbf{P})_{u,v}\), is \(\mathbf{P}_{u,v} \mathbf{M}_{u}\).

\[
\mathbb{E}_{v_1, \ldots, v_k} \left[ \prod_{j=1}^{k} \mathbf{M}_{v_j} \right] = \sum_{v_1, v_2, \ldots, v_k} \mathbf{\Phi}_{v_1} \mathbf{P}_{v_1, v_2} \cdots \mathbf{P}_{v_{k-1}, v_k} \prod_{j=1}^{k} \mathbf{M}_{v_j}
\]

\[
= \sum_{v_1} \mathbf{\Phi}_{v_1} \sum_{v_2} (\mathbf{P}_{v_1, v_2} \mathbf{M}_{v_1}) \cdots \sum_{v_k} (\mathbf{P}_{v_{k-1}, v_k} \mathbf{M}_{v_{k-1}}) \mathbf{M}_{v_k}
\]

\[
= \sum_{v_1} \mathbf{\Phi}_{v_1} \sum_{v_2} (\mathbf{E} \mathbf{P})_{v_1, v_2} \cdots \sum_{v_k} (\mathbf{E} \mathbf{P})_{v_2, v_3} \cdots \sum_{v_k} (\mathbf{E} \mathbf{P})_{v_k-1, v_k}
\]

\[
= \sum_{v_1} \mathbf{\Phi}_{v_1} \sum_{v_2} (\mathbf{E} \mathbf{P})^{k-1} \mathbf{E} \left[ \prod_{j=1}^{k} \mathbf{M}_{v_j} \right] = (\mathbf{\Phi} \otimes \mathbf{I}_d^2)^\top \left( (\mathbf{E} \mathbf{P})^{k-1} \mathbf{E} \right) (1 \otimes \mathbf{I}_d^2)
\]

Given Lemma\(^2\) Equation\(^8\) becomes:

\[
\mathbb{E}_{v_1, \ldots, v_k} \left[ \mathbf{\text{Tr}} \left[ \prod_{j=1}^{k} \mathbf{M}_{v_j} \right] \mathbf{\text{vec}}(\mathbf{I}_d), \mathbf{\text{vec}}(\mathbf{I}_d) \right]
\]

\[
= \langle (\mathbf{\Phi} \otimes \mathbf{I}_d^2)^\top \left( (\mathbf{E} \mathbf{P})^{k-1} \mathbf{E} \right) (1 \otimes \mathbf{I}_d^2), \mathbf{\text{vec}}(\mathbf{I}_d) \rangle
\]

\[
= \langle (\mathbf{E} \mathbf{P})^{k-1} \mathbf{E} (1 \otimes \mathbf{\text{vec}}(\mathbf{I}_d)), (\mathbf{\Phi} \otimes \mathbf{I}_d^2) \mathbf{\text{vec}}(\mathbf{I}_d) \rangle
\]

The third equality is due to \(\langle x, \mathbf{A} y \rangle = \langle \mathbf{A}^\top x, y \rangle\). The forth equality is by setting \(\mathbf{C} = \mathbf{1}\) (scalar) in \((\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})\). Then

\[
\mathbb{E}_{v_1, \ldots, v_k} \left[ \mathbf{\text{Tr}} \left[ \prod_{j=1}^{k} \mathbf{M}_{v_j} \right] \mathbf{\text{vec}}(\mathbf{I}_d), \mathbf{\text{vec}}(\mathbf{I}_d) \right]
\]

\[
= \langle (\mathbf{E} \mathbf{P})^{k-1} \mathbf{E} (1 \otimes \mathbf{\text{vec}}(\mathbf{I}_d)), (\mathbf{\Phi} \otimes \mathbf{I}_d^2) \mathbf{\text{vec}}(\mathbf{I}_d) \rangle
\]

\[
= \langle (\mathbf{\Phi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d))^\ast \left( (\mathbf{E} \mathbf{P})^{k-1} \mathbf{E} \right) (1 \otimes \mathbf{\text{vec}}(\mathbf{I}_d)) \rangle
\]

\[
= \langle (\mathbf{\Phi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d))^\ast \left( (\mathbf{E} \mathbf{P})^{k-1} \mathbf{E} \right) (1 \otimes \mathbf{\text{vec}}(\mathbf{I}_d)), (\mathbf{\Phi} \otimes \mathbf{I}_d^2) \mathbf{\text{vec}}(\mathbf{I}_d) \rangle
\]

\[
= \langle (\mathbf{\Phi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d))^\ast \left( (\mathbf{E} \mathbf{P})^{k-1} \mathbf{E} \right) (1 \otimes \mathbf{\text{vec}}(\mathbf{I}_d)), (\mathbf{\Phi} \otimes \mathbf{I}_d^2) \mathbf{\text{vec}}(\mathbf{I}_d) \rangle
\]

where we define \(z_0 = \mathbf{\Phi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d)\) and \(z_k = \left( z_k^* \left( \mathbf{E} \mathbf{P} \right)^k \right)^\ast \). Moreover, by Remark\(^2\) we have \(\| \mathbf{\pi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d) \|_\pi = \| \mathbf{\pi} \|_\pi \| \mathbf{\text{vec}}(\mathbf{I}_d) \|_2 = \sqrt{d} \) and \(\| z_0 \|_\pi = \| \mathbf{\Phi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d) \|_\pi = \| \mathbf{\Phi} \|_\pi \| \mathbf{\text{vec}}(\mathbf{I}_d) \|_2 = \| \mathbf{\Phi} \|_\pi \sqrt{d} \).

**Definition 2.** Define linear subspace \(\mathcal{U} = \{ \mathbf{\pi} \otimes \mathbf{w}, \mathbf{w} \in \mathbb{C}^{d^2} \}\).

**Remark 3.** \(\{ \mathbf{\pi} \otimes \mathbf{e}_i, i \in [d^2] \}\) is an orthonormal basis of \(\mathcal{U}\). This is because \(\langle \mathbf{\pi} \otimes \mathbf{e}_i, \mathbf{\pi} \otimes \mathbf{e}_j \rangle_\pi = \langle \mathbf{\pi}, \mathbf{\pi} \rangle_\pi \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}\) by Remark\(^2\) where \(\delta_{ij}\) is the Kronecker delta.

**Remark 4.** Given \(\mathbf{x} = \mathbf{y} \otimes \mathbf{w}\). The projection of \(\mathbf{x}\) on \(\mathcal{U}\) is \(\mathbf{x}^\perp = (1\ast \mathbf{y})(\mathbf{\pi} \otimes \mathbf{w})\). This is because

\[
\mathbf{x}^\perp = \sum_{i=1}^{d^2} (\mathbf{y} \otimes \mathbf{w}, \mathbf{\pi} \otimes \mathbf{e}_i)_\pi \mathbf{\pi} \otimes \mathbf{e}_i = \sum_{i=1}^{d^2} (\mathbf{y}, \mathbf{\pi})_\pi (\mathbf{w}, \mathbf{e}_i)(\mathbf{\pi} \otimes \mathbf{e}_i) = (1\ast \mathbf{y})(\mathbf{\pi} \otimes \mathbf{w})
\]

We want to bound

\[
\langle \mathbf{\pi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d), z_k \rangle_\pi = \langle \mathbf{\pi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d), z_k^\top + z_k^\perp \rangle_\pi = \langle \mathbf{\pi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d), z_k^\perp \rangle_\pi
\]

\[
\leq \| \mathbf{\pi} \otimes \mathbf{\text{vec}}(\mathbf{I}_d) \|_\pi \| z_k^\perp \|_\pi = \sqrt{d} \| z_k^\perp \|_\pi.
\]
As \( z_k \) can be expressed as recursively applying operator \( E \) and \( \bar{P} \) on \( z_0 \), we turn to analyze the effects of \( E \) and \( \bar{P} \) operators.

**Definition 3.** The spectral expansion of \( \bar{P} \) is defined as \( \lambda(\bar{P}) \triangleq \max_{x \perp \mathcal{U}, x \neq 0} \frac{\| (x^* \bar{P})^* \|_\pi}{\| x \|_\pi} \)

**Lemma 3.** \( \lambda(P) = \lambda(\bar{P}) \).

**Proof.** First show \( \lambda(\bar{P}) \geq \lambda(P) \). Suppose the maximizer of \( \lambda(P) \triangleq \max_{y \perp \pi, y \neq 0} \frac{\| (y^* P)^* \|_\pi}{\| y \|_\pi} \) is \( y \in \mathbb{C}^n \), i.e., \( \| (y^* P)^* \|_\pi = \lambda(P) \| y \|_\pi \). Construct \( x = y \otimes o \) for arbitrary non-zero \( o \in \mathbb{C}^{d^2} \). Easy to check that \( x \perp \mathcal{U} \), because \( \langle x, \pi \otimes w \rangle_\pi = \langle y, \pi \rangle_\pi \langle o, w \rangle = 0 \), where the last equality is due to \( y \perp \pi \). Then we can bound \( \| (x^* \bar{P})^* \|_\pi \) such that

\[
\| (x^* \bar{P})^* \|_\pi = \| \bar{P}^* x \|_\pi = \| (P^* \otimes I_{d^2}) (y \otimes o) \|_\pi = \| (P^* y) \otimes o \|_\pi = \| (y^* P)^* \|_\pi \| o \|_2 = \lambda(P) \| y \|_\pi \| o \|_2 = \lambda(P) \| x \|_\pi,
\]

which indicate for \( x = y \otimes o \), \( \| (x^* \bar{P})^* \|_\pi = \lambda(P) \). Taking maximum over all \( x \) gives \( \lambda(\bar{P}) \geq \lambda(P) \).

Next to show \( \lambda(P) \geq \lambda(\bar{P}) \). For \( \forall x \in \mathbb{C}^{Nd^2} \) such that \( x \perp \mathcal{U} \) and \( x \neq 0 \), we can decompose it to be

\[
x = \begin{bmatrix} x_1 \\
\vdots \\
x_{Nd^2}
\end{bmatrix} = \begin{bmatrix} x_1 \\
\vdots \\
x_{Nd^2}
\end{bmatrix} \otimes e_1 + \cdots + \begin{bmatrix} x_{d^2} \\
\vdots \\
x_{(N-1)d^2}
\end{bmatrix} \otimes e_{d^2} \triangleq \sum_{i=1}^{d^2} x_i \otimes e_i,
\]

where we define \( x_i \triangleq [x_i \cdots x_{(N-1)d^2+i}]^\top \) for \( i \in [d^2] \). We can observe that \( x_i \perp \pi, i \in [d^2] \), because for \( \forall j \in [d^2] \), we have

\[
0 = \langle x_i, \pi \otimes e_j \rangle_\pi = \sum_{i=1}^{d^2} \langle x_i \otimes e_i, \pi \otimes e_j \rangle_\pi = \sum_{i=1}^{d^2} \langle x_i, \pi \rangle_\pi \langle e_i, e_j \rangle_\pi = \langle x_j, \pi \rangle_\pi \delta_{ij},
\]

which indicates \( x_j \perp \pi, j \in [d^2] \). Furthermore, we can also observe that \( x_i \otimes e_i, i \in [d^2] \) is pairwise orthogonal. This is because for \( \forall i, j \in [d^2] \), \( \langle x_i \otimes e_i, x_j \otimes e_j \rangle_\pi = \langle x_i, x_j \rangle_\pi \langle e_i, e_j \rangle_\pi = \delta_{ij} \), which suggests us to use Pythagorean theorem such that \( \| x \|_\pi^2 = \sum_{i=1}^{d^2} \| x_i \otimes e_i \|_\pi^2 = \sum_{i=1}^{d^2} \| x_i \|_\pi^2 \| e_i \|_2^2 \).

We can use similar way to decompose and analyze \( (x^* \bar{P})^* \) :

\[
(x^* \bar{P})^* = \bar{P}^* x = \sum_{i=1}^{d^2} (P^* \otimes I_{d^2}) (x_i \otimes e_i) = \sum_{i=1}^{d^2} (P^* x_i) \otimes e_i.
\]

where we can observe that \( (P^* x_i) \otimes e_i, i \in [d^2] \) is pairwise orthogonal. This is because for \( \forall i, j \in [d^2] \), we have \( \langle (P^* x_i) \otimes e_i, (P^* x_j) \otimes e_j \rangle_\pi = \langle P^* x_i, P^* x_j \rangle_\pi \langle e_i, e_j \rangle_\pi = \delta_{ij} \). Again, applying Pythagorean theorem gives:

\[
\| (x^* \bar{P})^* \|_\pi^2 = \sum_{i=1}^{d^2} \| (P^* x_i) \otimes e_i \|_\pi^2 = \sum_{i=1}^{d^2} \| (x_i^* P)^* \|_\pi^2 \| e_i \|_2^2 \\
\leq \sum_{i=1}^{d^2} \lambda(P)^2 \| x_i \|_\pi^2 \| e_i \|_2^2 = \lambda(P)^2 \left( \sum_{i=1}^{d^2} \| x_i \|_\pi^2 \| e_i \|_2^2 \right) = \lambda(P)^2 \| x \|_\pi^2,
\]

which indicate that for \( \forall x \) such that \( x \perp \mathcal{U} \) and \( x \neq 0 \), we have \( \| (x^* \bar{P})^* \|_\pi \leq \lambda(P) \), or equivalently \( \lambda(\bar{P}) \leq \lambda(P) \).

Overall, we have shown both \( \lambda(\bar{P}) \geq \lambda(P) \) and \( \lambda(\bar{P}) \leq \lambda(P) \). We conclude \( \lambda(\bar{P}) = \lambda(P) \).
Lemma 4. (The effect of $\bar{P}$ operator) This lemma is a generalization of lemma 3.3 in [6].

1. \( \forall y \in \mathcal{U}, \text{then } (y^* \bar{P})^* = y. \)
2. \( \forall y \perp \mathcal{U}, \text{then } (y^* \bar{P})^* \perp \mathcal{U}, \text{and } \left\| (y^* \bar{P})^* \right\|_\pi \leq \lambda \left\| y \right\|_\pi. \)

Proof. First prove the Part 1 of lemma\(^4\) \( \forall y = \pi \otimes w \in \mathcal{U}: \)

\[
y^* \bar{P} = (\pi^* \otimes w^*) (P \otimes I_{d^2}) = (\pi^* P) \otimes (w^* I_{d^2}) = \pi^* \otimes w^* = y^*,
\]

where third equality is because $\pi$ is the stationary distribution. Next to prove Part 2 of lemma\(^4\) Given $y \perp \mathcal{U}$, want to show $(y^* \bar{P})^* \perp \pi \otimes w$, for every $w \in \mathbb{C}^{d^2}$. It is true because

\[
\left\langle \pi \otimes w, (y^* \bar{P})^* \right\rangle_\pi = y^* \bar{P} (\Pi^{-1} \otimes I_{d^2}) (\pi \otimes w) = y^* ((P \Pi^{-1} \pi) \otimes w) = y^* ((\Pi^{-1} \pi) \otimes w) = 0.
\]

The third equality is due to $P \Pi^{-1} \pi = \Pi^{-1} \pi$, for every $\pi$ and $w \in \mathbb{C}^{d^2}$. Moreover, $\left\| (y^* \bar{P})^* \right\|_\pi \leq \lambda \left\| y \right\|_\pi$ is simply a re-statement of definition\(^3\).

Remark 5. Lemma\(^4\) implies that \( \forall y \in \mathbb{C}^{nd^2} \)

1. \( \left\| (y^* \bar{P})^* \right\|_\pi \leq \alpha_1 \left\| y \right\|_\pi, \text{ where } \alpha_1 = \exp (t \ell) - t \ell. \)
2. \( \left\| (y^* \bar{P})^* \right\|_\pi \leq \alpha_2 \left\| y \right\|_\pi, \text{ where } \alpha_2 = \lambda (\exp (t \ell) - 1). \)
3. \( \left\| (y^* \bar{P})^* \right\|_\pi \leq \alpha_3 \left\| y \right\|_\pi, \text{ where } \alpha_3 = \exp (t \ell) - 1. \)
4. \( \left\| (y^* \bar{P})^* \right\|_\pi \leq \alpha_4 \left\| y \right\|_\pi, \text{ where } \alpha_4 = \lambda \exp (t \ell). \)

Lemma 5. (The effect of $E$ operator) Given three parameters \( \lambda \in [0,1], t \geq 0 \) and \( t > 0 \). Let \( P \) be a regular Markov chain on state space \( [N] \), with stationary distribution \( \pi \) and spectral expansion \( \lambda \).

Suppose each state \( i \in [N] \) is assigned a matrix \( H_i \in \mathbb{C}^{d \times d} \) s.t. \( \left\| H_i \right\|_2 \leq \ell \) and \( \sum_{i \in [N]} \pi_i H_i = 0. \)

Let \( \bar{P} = P \otimes I_{d^2} \) and \( E \) denotes the \( Nd^2 \times Nd^2 \) block matrix where the $i$-th diagonal block is the matrix \( \exp (t H_i), i.e., E = \text{diag}(\exp (t H_1), \cdots, \exp (t H_N)). \) Then for any \( \forall z \in \mathbb{C}^{Nd^2} \), we have:

1. \( \left\| (z^* E \bar{P})^* \right\|_\pi \leq \alpha_1 \left\| z \right\|_\pi, \text{ where } \alpha_1 = \exp (t \ell) - \ell \ell. \)
2. \( \left\| (z^* E \bar{P})^* \right\|_\pi \leq \alpha_2 \left\| z \right\|_\pi, \text{ where } \alpha_2 = \lambda (\exp (t \ell) - 1). \)
3. \( \left\| (z^* E \bar{P})^* \right\|_\pi \leq \alpha_3 \left\| z \right\|_\pi, \text{ where } \alpha_3 = \exp (t \ell) - 1. \)
4. \( \left\| (z^* E \bar{P})^* \right\|_\pi \leq \alpha_4 \left\| z \right\|_\pi, \text{ where } \alpha_4 = \lambda \exp (t \ell). \)

Proof. (of Lemma\(^5\)) We first show that, for \( z = y \otimes w, \)

\[
(z^* E)^* = E^* z = \begin{bmatrix}
\exp(t H_1^T) & \cdots & \exp(t H_N^T)
\end{bmatrix}
\begin{bmatrix}
y_1 w \\
\vdots \\
y_N w
\end{bmatrix}
= \begin{bmatrix}
y_1 \exp(t H_1^T) w \\
\vdots \\
y_N \exp(t H_N^T) w
\end{bmatrix}
= \sum_{i=1}^{N} y_i (\lambda_i \otimes (\exp(t H_i^T) w)).
\]
Due to the linearity of projection,
\[
((z^* E)^*)^\| = \sum_{i=1}^{N} y_i (e_i \otimes (\exp(tH_i^*)w))^\| = \sum_{i=1}^{N} y_i (1^* e_i) (\pi \otimes (\exp(tH_i^*)w)) = \pi \otimes \left(\sum_{i=1}^{N} y_i \exp(tH_i^*)w\right),
\]
where the second inequality follows by Remark 4.

**Proof of Lemma 5, Part 1**

Firstly we can bound \(\left\|\sum_{i=1}^{N} \pi_i \exp(tH_i^*)\right\|_2\) by
\[
\left\|\sum_{i=1}^{N} \pi_i \exp(tH_i^*)\right\|_2 = \left\|\sum_{i=1}^{N} \pi_i \exp(tH_i)\right\|_2 = \left\|\sum_{i=1}^{N} \pi_i \sum_{k=0}^{+\infty} \frac{t^k H_i^j}{k!}\right\|_2 = \left\|I + \sum_{i=1}^{N} \pi_i \sum_{j=2}^{+\infty} \frac{(t t_j)^j}{j!}\right\|_2
\]
\[
\leq 1 + \sum_{i=1}^{N} \pi_i \sum_{j=2}^{+\infty} \frac{\|H_i\|_2^j}{j!} \leq 1 + \sum_{i=1}^{N} \pi_i \sum_{j=2}^{+\infty} \frac{(t t_j)^j}{j!} = \exp(tt) - tt,
\]

where the first step follows by \(\|A\|_2 = \|A^*\|_2\), the second step follows by matrix exponential, the third step follows by \(\sum_{i \in [N]} \pi_i H_i = 0\), and the forth step follows by triangle inequality. Given the above bound, for any \(z\) which can be written as \(z = \pi \otimes w\) for some \(w \in \mathbb{C}^{d^2}\), we have
\[
\left\|\left((z^* E \bar{P})^*\right)^\|\right\|_\pi = \left\|\left((z^* E)^*\right)^\|\right\|_\pi = \left\|\pi \otimes \left(\sum_{i=1}^{N} \pi_i \exp(tH_i^*)w\right)\right\|_\pi = \|\pi\|_\pi \left\|\sum_{i=1}^{N} \pi_i \exp(tH_i^*)w\right\|_2
\]
\[
\leq \|\pi\|_\pi \left\|\sum_{i=1}^{N} \pi_i \exp(tH_i^*)\right\|_2 \|w\|_2 = \left\|\sum_{i=1}^{N} \pi_i \exp(tH_i^*)\right\|_2 \|z\|_\pi
\]
\[
\leq (\exp(tt) - tt) \|z\|_\pi,
\]

where step one follows by Part 1 of Remark 5 and step two follows by Equation 9.

**Proof of Lemma 5, Part 2**

For \(\forall z \in \mathbb{C}^{N d^2}\), we can write it as block matrix such that:
\[
z = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \sum_{i=1}^{N} e_i \otimes z_i,
\]

where each \(z_i \in \mathbb{C}^{d^2}\). Please note that above decomposition is pairwise orthogonal. Applying Pythagorean theorem gives \(\|z\|_\pi^2 = \sum_{i=1}^{N} \|e_i \otimes z_i\|_\pi^2 = \sum_{i=1}^{N} \|e_i\|_\pi^2 \|z_i\|_\pi^2\). Similarly, we can decompose \((E^* - I_{N d^2})z\) such that
\[
(E^* - I_{N d^2})z = \begin{bmatrix} (\exp(tH_1^*) - I_{d^2})z_1 \\ \vdots \\ (\exp(tH_N^*) - I_{d^2})z_N \end{bmatrix} = \begin{bmatrix} (\exp(tH_1^*) - I_{d^2})z_1 \\ \vdots \\ (\exp(tH_N^*) - I_{d^2})z_N \end{bmatrix} = \sum_{i=1}^{N} e_i \otimes ((\exp(tH_i^*) - I_{d^2})z_i)
\]

Note that above decomposition is pairwise orthogonal, too. Applying Pythagorean theorem gives
\[
\|E^* - I_{N d^2}\|_\pi^2 = \sum_{i=1}^{N} \|e_i \otimes ((\exp(tH_i^*) - I_{d^2})z_i)\|_\pi^2 = \sum_{i=1}^{N} \|e_i\|_\pi^2 (\|\exp(tH_i^*) - I_{d^2}\|_2 \|z_i\|_\pi^2)
\]
\[
\leq \sum_{i=1}^{N} \|e_i\|_\pi^2 \|\exp(tH_i^*) - I_{d^2}\|_2 \|z_i\|_\pi^2 \leq \max_{i \in [N]} \|\exp(tH_i^*) - I_{d^2}\|_2 \sum_{i=1}^{N} \|e_i\|_\pi^2 \|z_i\|_\pi^2
\]
\[
= \max_{i \in [N]} \|\exp(tH_i^*) - I_{d^2}\|_2 \|z\|_\pi^2 = \max_{i \in [N]} \|\exp(tH_i) - I_{d^2}\|_2 \|z\|_\pi^2,
\]

22
which indicates
\[ \| (E^* - I_{N,d^2}) z \|_\pi = \max_{i \in [N]} \| \exp(tH_i) - I_{d^2} \|_2 \| z \|_\pi = \max_{i \in [N]} \left\| \sum_{j=1}^{t \beta J} \frac{j}{J!} \right\|_2 \| z \|_\pi \leq \left( \sum_{j=1}^{\infty} \frac{t \beta J}{J!} \right) \| z \|_\pi = (\exp (t \beta) - 1) \| z \|_\pi. \]

Now we can formally prove Part 2 of Lemma 5 by:
\[
\left\| \left( z^* E \hat{P} \right)^* \right\|_\pi = \left\| \left( E^* z^* \right)^\dagger \right\|_\pi = \left\| E^* z^* - z^* + z^* \right\|_\pi = \left\| E^* - I_{N,d^2} \right\|_\pi \leq \left( \exp (t \beta) - 1 \right) \| z^\dagger \|_\pi.
\]

The first step follows by Part 2 of Remark 5, the second step follows by Part 1 on Lemma 4, and the forth step is due to \((z^\dagger)^2 = 0\).

**Proof of Lemma 5, Part 3**

Note that
\[
\left\| \left( z^* E \hat{P} \right)^* \right\|_\pi = \left\| \left( E^* z^* \right)^\dagger \right\|_\pi = \left\| E^* z^* - z^* + z^* \right\|_\pi = \left\| E^* - I_{N,d^2} \right\|_\pi \leq \left( \exp (t \beta) - 1 \right) \| z^\dagger \|_\pi,
\]

where the first step follows by Part 1 of Remark 5, the second step follows by \((z^\dagger)^2 = 0\), and the last step follows by Part 2 of Lemma 4.

**Proof of Lemma 5, Part 4**

Similar to Equation 10 for all \(z \in \mathbb{C}^{N \times d^2}\), we can decompose \(E^* z\) as \(E^* z = \sum_{i=1}^{N} e_i \otimes (\exp(tH_i^*) z_i)\). This decomposition is pairwise orthogonal, too. Applying Pythagorean theorem gives
\[
\| E^* z \|_\pi^2 = \sum_{i=1}^{N} \| e_i \otimes (\exp(tH_i^*) z_i) \|_\pi^2 = \sum_{i=1}^{N} \| e_i \|_\pi^2 \| \exp(tH_i^*) z_i \|_\pi^2 \leq \sum_{i=1}^{N} \| e_i \|_\pi^2 \| \exp(tH_i^*) \|_\pi^2 \| z_i \|_\pi^2 \leq \max_{i \in [N]} \| \exp(tH_i^*) \|_\pi^2 \sum_{i=1}^{N} \| e_i \|_\pi^2 \| z_i \|_\pi^2 \leq \max \exp (\| tH_i^* \|_\pi^2 \| z \|_\pi^2 \leq \exp (t \beta) \| z \|_\pi^2) \]
\]

which indicates \(\| E^* z \|_\pi \leq \exp (t \beta) \| z \|_\pi\). Now we can prove Part 4 of Lemma 5.

Note that
\[
\left\| \left( z^* E \hat{P} \right)^* \right\|_\pi = \left\| \left( E^* z^* \right)^\dagger \right\|_\pi = \left\| E^* z^* - z^* + z^* \right\|_\pi = \left\| E^* - I_{N,d^2} \right\|_\pi \leq \| E^* z \|_\pi \leq \lambda \| E^* z \|_\pi \leq \lambda \| E^* z \|_\pi.
\]

\(\square\)

**Recursive Analysis**

We now use Lemma 5 to analyze the evolution of \(z^\dagger\) and \(z^\dagger\). Let \(H_v \triangleq \frac{f(v)(\gamma + ib)}{2} \otimes I_{d^2} + I_{d^2} \otimes \frac{f(v)(\gamma - ib)}{2}\) in Lemma 5. We can see verify the following three facts: (1) \(\exp(tH_v) = M_v\); (2) \(\| H_v \|_2\) is bounded (3) \(\sum_{v \in [N]} \tau_v H_v = 0\).

Firstly, easy to see that
\[
\exp (tH_v) = \exp \left( \frac{tf(v)(\gamma + ib)}{2} \otimes I_{d^2} + I_{d^2} \otimes \frac{tf(v)(\gamma - ib)}{2} \right) = \exp \left( \frac{tf(v)(\gamma + ib)}{2} \otimes \exp \left( \frac{tf(v)(\gamma - ib)}{2} \right) = M_v,
\]

where the first step follows by definition of \(H_v\) and the second step follows by the fact that \(\exp(A \otimes I_d + I_d \otimes B) = \exp(A) \otimes \exp(B)\), and the last step follows by Equation 7.

Secondly, we can bound \(\| H_v \|_2\) by:
\[
\| H_v \|_2 \leq \frac{f(v)(\gamma + ib)}{2} \| I_{d^2} \|_2 + \| I_{d^2} \| \| I_{d^2} \|_2 \leq \frac{\sqrt{\gamma^2 + b^2}}{2}.
\]
where the first step follows by triangle inequality, the second step follows by the fact that \( \|A \otimes B\|_2 = \|A\|_2 \|B\|_2 \), the third step follows by \( \|I_d\|_2 = 1 \) and \( f(v)\|_2 \leq 1 \). We set \( \ell = \sqrt{\gamma^2 + \beta^2} \) to satisfy the assumption in Lemma 5 that \( \|H_v\|_2 \leq \ell \). According to the conditions in Lemma 1, we know that \( t\ell \leq 1 \) and \( t\ell \leq \frac{1-\lambda}{4\xi} \).

Finally, we show that \( \sum_{v \in [N]} \pi_v H_v = 0 \), because

\[
\sum_{v \in [N]} \pi_v H_v = \sum_{v \in [N]} \left( \frac{f(v)(\gamma + ib)}{2} \otimes I_d + I_d \otimes \frac{f(v)(\gamma - ib)}{2} \right)
= \gamma + ib \left( \sum_{v \in [N]} \pi_v f(v) \right) \otimes I_d + \gamma - ib I_d \otimes \left( \sum_{v \in [N]} \pi_v f(v) \right) = 0,
\]

where the last step follows by \( \sum_{v \in [N]} \pi_v f(v) = 0 \).

**Claim 4.** \( \|z^+_i\|_\pi \leq \frac{\alpha_2}{1-\alpha_4} \max_{0 \leq j < i} \|z^+_j\|_\pi \).

**Proof.** Using Part 2 and Part 4 of Lemma 5 we have

\[
\|z^+_i\|_\pi = \left\| \left( \left( z^+_i E \tilde{\phi} \right)^\delta \right) \right\|_\pi
\leq \left\| \left( \left( z^+_i E \tilde{\phi} \right)^\delta \right) + \left( \left( z^+_i E \tilde{\phi} \right)^\delta \right)^\delta \right\|_\pi
\leq \alpha_2 \|z^+_i \|_\pi + \alpha_4 \|z^+_i \|_\pi
\leq (\alpha_2 + \alpha_2 \alpha_4 + \alpha_2 \alpha_4^2 + \cdots) \max_{0 \leq j < i} \|z^+_j\|_\pi
\leq \frac{\alpha_2}{1-\alpha_4} \max_{0 \leq j < i} \|z^+_j\|_\pi.
\]

**Claim 5.** \( \|z^+_i\|_\pi \leq \left( \alpha_1 + \frac{\alpha_2 \alpha_3}{1-\alpha_4} \right) \max_{0 \leq j < i} \|z^+_j\|_\pi \).

**Proof.** Using Part 1 and Part 3 of Lemma 5 as well as Claim 4 we have

\[
\|z^+_i\|_\pi = \left\| \left( \left( z^+_i E \tilde{\phi} \right)^\delta \right) \right\|_\pi
\leq \left\| \left( \left( z^+_i E \tilde{\phi} \right)^\delta \right) \right\|_\pi + \left\| \left( \left( z^+_i E \tilde{\phi} \right)^\delta \right)^\delta \right\|_\pi
\leq \alpha_1 \|z^+_i \|_\pi + \alpha_3 \|z^+_i \|_\pi
\leq \alpha_1 \|z^+_i \|_\pi + \alpha_3 \frac{\alpha_2}{1-\alpha_4} \max_{0 \leq j < i-1} \|z^+_j\|_\pi
\leq \left( \alpha_1 + \frac{\alpha_2 \alpha_3}{1-\alpha_4} \right) \max_{0 \leq j < i} \|z^+_j\|_\pi.
\]

Combining Claim 4 and Claim 5 gives

\[
\|z^+_i\|_\pi \leq \left( \alpha_1 + \frac{\alpha_2 \alpha_3}{1-\alpha_4} \right) \max_{0 \leq j < k} \|z^+_j\|_\pi
\]

(because \( \alpha_1 + \alpha_2 \alpha_3/(1-\alpha_4) \geq \alpha_1 \geq 1 \)) \leq \left( \alpha_1 + \frac{\alpha_2 \alpha_3}{1-\alpha_4} \right)^k \|z^+_0\|_\pi
\leq \|\phi\|_\pi \sqrt{d} \left( \alpha_1 + \frac{\alpha_2 \alpha_3}{1-\alpha_4} \right)^k.
\]

24
where the second step is because \\

\[ 1 \]

According to the complexification technique, we know that (1) \( g \) we construct a real-valued matrix function \( \alpha_1 = \exp(t \ell) - t \ell \leq 1 + t^2 \ell^2 = 1 + t^2(\gamma^2 + b^2) \), and\\n
\[ \alpha_2 \alpha_3 = \lambda(\exp(t \ell) - 1)^2 \leq \lambda(2t \ell)^2 = 4\lambda t^2(\gamma^2 + b^2) \]

where the second step is because \( \exp(x) \leq 1 + 2x, \forall x \in [0,1] \) and \( t \ell < 1 \),

\[ \alpha_4 = \lambda \exp(t \ell) \leq \lambda (1 + 2t \ell) \leq \frac{1}{2} + \frac{1}{2} \lambda \]

where the second step is because \( t \ell < 1 \), and the third step follows by \( t \ell \leq \frac{1-\lambda}{4\lambda} \).

Overall, we have

\[
\left( \alpha_1 + \frac{\alpha_2 \alpha_3}{1 - \alpha_4} \right)^k \leq \left( 1 + t^2(\gamma^2 + b^2) + \frac{4\lambda t^2(\gamma^2 + b^2)}{\frac{1}{2} - \frac{1}{2} \lambda} \right)^k \\
\leq \exp \left( kt^2(\gamma^2 + b^2) \left( 1 + \frac{8}{1 - \lambda} \right) \right).
\]

This completes our proof of Lemma[1].

\[ \square \]

**B.4 Proof of Theorem[1]**

**Theorem 1** (Markov Chain Matrix Chernoff Bound). Let \( P \) be a regular Markov chain with state space \( [N] \), stationary distribution \( \pi \) and spectral expansion \( \lambda \). Let \( f : [N] \to \mathbb{C}^{d \times d} \) be a function such that (1) \( \forall v \in [N], f(v) \) is Hermitian and \( \|f(v)\|_2 \leq 1 \); (2) \( \sum_{v \in [N]} \pi_v f(v) = 0 \). Let \( \{v_1, \ldots, v_k\} \) denote a \( k \)-step random walk on \( P \) starting from a distribution \( \phi \). Given \( \epsilon \in (0,1) \),

\[
\mathbb{P} \left[ \lambda_{\max} \left( \frac{1}{k} \sum_{j=1}^k f(v_j) \right) \geq \epsilon \right] \leq 4 \|\phi\|_w d^2 \exp \left( -\epsilon^2 (1-\lambda)k/72 \right)
\]

\[
\mathbb{P} \left[ \lambda_{\min} \left( \frac{1}{k} \sum_{j=1}^k f(v_j) \right) \leq -\epsilon \right] \leq 4 \|\phi\|_w d^2 \exp \left( -\epsilon^2 (1-\lambda)k/72 \right).
\]

**Proof.** (of Theorem[1]) Our strategy is to adopt complexification technique [8]. For any \( d \times d \) complex Hermitian matrix \( X \), we may write \( X = Y + iZ \), where \( Y \) and \( iZ \) are the real and imaginary parts of \( X \), respectively. Moreover, the Hermitian property of \( X \) (i.e., \( X^* = X \)) implies that (1) \( Y \) is real and symmetric (i.e., \( Y^\top = Y \)); (2) \( Z \) is real and skew-symmetric (i.e., \( Z = -Z^\top \)). The eigenvalues of \( X \) can be found via a \( 2d \times 2d \) real symmetric matrix \( H \triangleq \begin{bmatrix} Y & Z \\ -Z & Y \end{bmatrix} \), where the symmetry of \( H \) follows by the symmetry of \( Y \) and skew-symmetry of \( Z \). Note the fact that, if the eigenvalues (real) of \( X \) are \( \lambda_1, \lambda_2, \ldots, \lambda_d \), then those of \( H \) are \( \lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_d, \lambda_d \). I.e., \( X \) and \( H \) have the same eigenvalues, but with different multiplicity.

Using the above technique, we can formally prove Theorem[1]. For any complex matrix function \( f : [N] \to \mathbb{C}^{d \times d} \) in Theorem[1] we can separate its real and imaginary parts by \( f = f_1 + if_2 \). Then we construct a real-valued matrix function \( g : [N] \to \mathbb{R}^{2d \times 2d} \) s.t. \( \forall v \in [N], g(v) = \begin{bmatrix} f_1(v) & f_2(v) \\ -f_2(v) & f_1(v) \end{bmatrix} \).

According to the complexification technique, we know that (1) \( \forall v \in [N], g(v) \) is real symmetric and \( \|g(v)\|_2 = \|f(v)\|_2 \leq 1 \); (2) \( \sum_{v \in [N]} \pi_v g(v) = 0 \). Then

\[
\mathbb{P} \left[ \lambda_{\max} \left( \frac{1}{k} \sum_{j=1}^k f(v_j) \right) \geq \epsilon \right] = \mathbb{P} \left[ \lambda_{\max} \left( \frac{1}{k} \sum_{j=1}^k g(v_j) \right) \geq \epsilon \right] \leq 4 \|\phi\|_w d^2 \exp \left( -\epsilon^2 (1-\lambda)k/72 \right),
\]

where the first step follows by the fact that \( \frac{1}{k} \sum_{j=1}^k f(v_j) \) and \( \frac{1}{k} \sum_{j=1}^k g(v_j) \) have the same eigenvalues (with different multiplicity), and the second step follows by Theorem[8]. The bound on \( \lambda_{\min} \) also follows similarly.

\[ \square \]

---

The additional factor 4 is because the constructed \( g(v) \) has shape \( 2d \times 2d \).