Supplementary Material for "Streaming PCA for Markovian Data"

Syamantak Kumar[1] **Purnamrita Sarkar**[2]

[1]Department of Computer Science, UT Austin
[2]Department of Statistics and Data Sciences, UT Austin
syamantak@utexas.edu, purna.sarkar@austin.utexas.edu

Abstract

The supplementary material is organized as follows -

- Section S.1 introduces notation that will be useful for concise representation.
- Section S.2 provides the proof of Proposition 1.
- Section S.3 contains useful intermediate results which are used in subsequent proofs of our main results.
- Section S.4 proves bounds on $v_1 B_n B_n^T v_1^T$ and $V_\perp B_n B_n^T V_\perp^T$ (Theorems 2, 3, 4 and 5).
- Section S.5 puts everything together and provides proofs of our main result - Theorem 1, along with Corollary 1.
- Section S.6 provides additional experiments to further support our claims.

S.1 Notation and assumptions

For conciseness, we define the stochastic function $A : \Omega \to \mathbb{R}^{d\times d}$ which maps each state variable of the Markov chain to a $(d \times d)$ positive semi-definite symmetric matrix as

$$A\left(s_t\right) := X_t X_t^T$$

Where $X_t \sim D\left(s_t\right)$ is drawn from the distribution corresponding to the state at timestep s_t. All the theoretical results are derived under Assumptions 1, 2 and 3.

S.2 Offline PCA with Markovian Data

In this section, we prove Proposition 1. We note that [4] considers $F_j\left(s_j\right)$ to be random only with respect to the states. Therefore, we first show that their results generalize to our setting as well, using $F_j\left(s_j\right) := A\left(s_j\right) - \Sigma$. From Eq (5) in [4], we have

$$\begin{aligned}
\left\|\prod_{j=1}^{n}\exp\left(\frac{\theta}{2}\left(A\left(s_j\right)-\Sigma\right)\right)\right\|_F^2 &= \operatorname{Tr}\left(\prod_{j=1}^{n}\exp\left(\frac{\theta}{2}\left(A\left(s_j\right)-\Sigma\right)\right)\prod_{j=n}^{1}\exp\left(\frac{\theta}{2}\left(A\left(s_j\right)-\Sigma\right)\right)\right) \\
&= \operatorname{vec}\left(I_d\right)^T\left(\prod_{j=1}^{n}\exp\left(\theta H\left(s_j\right)\right)\right)\operatorname{vec}\left(I_d\right)
\end{aligned}$$

where $H\left(s_j\right) := \frac{1}{2}\left[\left(A\left(s_j\right)-\Sigma\right)\otimes I_d + I_d \otimes \left(A\left(s_j\right)-\Sigma\right)\right]$. Noting that conditioned on the state sequence, the matrices $A\left(s_i\right), i \in [n]$ are independent under our model, we can push in the expectation over the state-specific distributions inside. Let $\mathbb{E}_\pi$ denote the expectation over the stationary state-sequence of the Markov chain, and $\mathbb{E}_D$ denote the distribution over states. Therefore,

$$\mathbb{E}_\pi \mathbb{E}_D \left[\left\| \prod_{j=1}^{n} \exp\left(\frac{\theta}{2}\left(A\left(s_j\right) - \Sigma\right)\right) \right\|_F^2 \right] = \mathbb{E}_\pi \left[\operatorname{vec}\left(I_d\right)^T \left(\prod_{j=1}^{n} \mathbb{E}_{D(s_j)} \left[\exp\left(\theta H\left(s_j\right)\right)\right] \right) \operatorname{vec}\left(I_d\right) \right]$$

Defining the multiplication operator $\left(E_j^\theta \mathbf{h}\right)(x) = \mathbb{E}_{D(x)}\left[\exp\left(\theta H_j\left(x\right)\right)\right] \mathbf{h}\left(x\right)$ for any vector-valued function $\mathbf{h}$, we note that Eq (8) from [4] holds for our case as well.

Next, we adapt Proposition 5.3 from [4] for our setting. Specifically, we have the following lemma -

Lemma S.1. *Consider the operator* $H\left(x\right) := \frac{1}{2}\left[\left(A\left(x\right) - \Sigma\right) \otimes I_d + I_d \otimes \left(A\left(x\right) - \Sigma\right)\right]$. *Then, under assumptions 3 and 2 and the definition of* Σ*, we have,*

1. $\mathbb{E}_\pi \mathbb{E}_{D(x)}\left[H\left(x\right)\right] = 0$
2. $H\left(x\right) \preceq \mathcal{M} I$
3. $\left\| \mathbb{E}_\pi \mathbb{E}_{D(x)} \left[H\left(x\right)^2\right] \right\|_2 \leq \mathcal{V}$

Proof. The proof follows by using the same arguments as Proposition 5.3 from [4] and using the expectation $\mathbb{E}_\pi \mathbb{E}_{D(x)}$ over both the state sequence and the distribution over states, along with assumptions 3 and 2. □

Finally, to prove Bernstein's inequality, we prove that Lemma 6.7 from [4] holds for our case. To note this, we start with equation (57) in their work. We have, using Lemma S.1,

$$\begin{aligned}
\left|\left\langle v_2, \mathbb{E}_\pi \mathbb{E}_{D(x)}\left[\exp\left(\theta H\left(x\right)\right)\right] v_1 \right\rangle\right| &= \left|\left\langle v_2, \mathbb{E}_\pi \mathbb{E}_{D(x)}\left[\exp\left(\theta H\left(x\right)\right)\right] v_1 \right\rangle\right| \\
&= \left|\left\langle v_2, \left(I + \mathbb{E}_\pi \mathbb{E}_{D(x)}\left[H\left(x\right)\right] + \sum_{k=2}^{\infty} \frac{\theta^k}{k!} \mathbb{E}_\pi \mathbb{E}_{D(x)}\left[H\left(x\right)^k\right]\right) v_1 \right\rangle\right| \\
&= \left|\left\langle v_2, v_1 \right\rangle + \left\langle v_2 \left(\sum_{k=2}^{\infty} \frac{\theta^k}{k!} \mathbb{E}_\pi \mathbb{E}_{D(x)}\left[H\left(x\right)^k\right]\right) v_1 \right\rangle\right| \\
&\leq \left|\left\langle v_2, v_1 \right\rangle\right| \left(1 + \mathcal{V}\left(\sum_{k=2}^{\infty} \frac{\theta^k}{k!} \mathcal{M}^{k-2}\right)\right)
\end{aligned}$$

Therefore, Eq (60) from [4] follows. The other bounds in the proof of Lemma 6.7 from [4] follow similarly. Therefore, we have the following version of Theorem 2.2 from [4] -

Proposition S.1. *Under Assumptions 2 and 3, we have*

$$P\left(\left\|\sum_{j=1}^{n}\left(A\left(s_j\right) - \Sigma\right)\right\|_2 \geq t\right) \leq d^{2-\frac{\pi}{4}} \exp\left(\frac{-t^2 / \frac{32}{\pi^2}}{\frac{1+|\lambda_2(P)|}{1-|\lambda_2(P)|} n\mathcal{V} + \frac{8/\pi}{1-|\lambda_2(P)|}\mathcal{M}t}\right)$$

The proof of Proposition 1 now follows by converting the tail bound into a high probability bound and using Wedin's theorem [5]. See proof of Theorem 1.1 in [1] for details.

S.3 Useful Results

This section presents some useful lemmas and their proofs that are subsequently used in our proofs.

Proof of Lemma 1. Let the transition probabilities of the Markov chain be represented as $P(x|y) := P(Z_{t+1} = x|Z_t = y)$. Consider the time-reversed chain $Y_i := Z_{n-i+1}$ for $i = 1, 2, \dots n$. Then,

$$\begin{aligned}
&\mathbb{P}\left(Y_l = s_l|Y_{l-1} = s_{l-1}, Y_{l-2} = s_{l-2} \dots Y_1 = s_1\right) \\
&\quad = \mathbb{P}\left(Z_{n-l+1} = s_l|Z_{n-l+2} = s_{l-1}, Z_{n-l+3} = s_{l-2}, \dots Z_n = s_1\right) \\
&\quad = \mathbb{P}\left(Z_{n-l+1} = s_l|Z_{n-l+2} = s_{l-1}\right) \quad \text{using Lemma S.5} \\
&\quad = \frac{\mathbb{P}\left(Z_{n-l+1} = s_l, Z_{n-l+2} = s_{l-1}\right)}{\mathbb{P}\left(Z_{n-l+2} = s_{l-1}\right)} \\
&\quad = \frac{\pi\left(s_l\right) P(s_{l-1}|s_l)}{\pi\left(s_{l-1}\right)} \\
&\quad = P(s_l|s_{l-1}) \quad \text{using reversibility}
\end{aligned}$$

This proves that Y_n is an irreducible Markov chain with the same transition probabilities as the original Markov chain. The irreducibility of Y_n follows from the original Markov chain being irreducible. Therefore,

$$\mathbb{P}\left(Z_t = s_1|Z_{t+k} = s_2\right) = \mathbb{P}\left(Y_{n+1-t} = s_1|Y_{n+1-t-k} = s_2\right) \tag{S.1}$$

Then,

$$\frac{1}{2}\sup_{t\in\Omega}\sum_{s}\left|\mathbb{P}\left(Z_t = s|Z_{t+k} = t\right) - \pi\left(s\right)\right| = \frac{1}{2}\sup_{t\in\Omega}\sum_{s}\left|\mathbb{P}\left(Y_{n+1-t} = s|Y_{n+1-t-k} = t\right) - \pi\left(s\right)\right| = d_{\text{mix}}\left(k\right)$$

where the last inequality follows from the forward mixing properties of the Markov chain. □

Lemma S.2. *Let $C_{j,i} = \prod_{t=j}^{i}(I + Z_t)$ for $i \leq j \leq n$, where $Z_t \in \mathbb{R}^{d\times d}$ are symmetric PSD matrices. Let $U \in \mathbb{R}^{d\times d'}$. Then,*

$$\operatorname{Tr}\left(U^T C_{j,i+1} C_{j,i+1}^T U\right) \leq \operatorname{Tr}\left(U^T C_{j,i} C_{j,i}^T U\right)$$

Proof. Observe that,

$$\begin{aligned}
\operatorname{Tr}\left(U^T C_{j,i} C_{j,i}^T U\right) &= \operatorname{Tr}\left(U^T C_{j,i+1}(I + 2Z_i + Z_i^2) C_{j,i+1}^T U\right) \\
&= \operatorname{Tr}\left(U^T C_{j,i+1} C_{j,i+1}^T U\right) + \operatorname{Tr}\left(U^T C_{j,i+1}(2Z_i + Z_i^2) C_{j,i+1}^T U\right)
\end{aligned}$$

Since Z_i and Z_i^2 are both PSD, the second term on the RHS is always positive. This yields the proof. □

Lemma S.3. *Let $B_t = \prod_{i=t}^{1}(I + Z_i)$, where $Z_i \in \mathbb{R}^{d\times d}$ are symmetric PSD matrices.*

$$\operatorname{Tr}\left(B_{n-1} B_{n-1}^T\right) \leq \operatorname{Tr}\left(B_n B_n^T\right)$$

Proof. We have,

$$\begin{aligned}
\operatorname{Tr}\left(B_n B_n^T\right) &= \operatorname{Tr}\left((I + Z_n) B_{n-1} B_{n-1}^T (I + Z_n)\right) \\
&= \operatorname{Tr}\left(B_{n-1} B_{n-1}^T\right) + \operatorname{Tr}\left(Z_n B_{n-1} B_{n-1}^T\right) + \operatorname{Tr}\left(B_{n-1} B_{n-1}^T Z_n\right) + \operatorname{Tr}\left(Z_n B_{n-1} B_{n-1}^T Z_n\right) \\
&= \operatorname{Tr}\left(B_{n-1} B_{n-1}^T\right) + 2\operatorname{Tr}\left(B_{n-1}^T Z_n B_{n-1}\right) + \operatorname{Tr}\left(B_{n-1}^T Z_n^2 B_{n-1}\right)
\end{aligned}$$

Since Z_n and Z_n^2 are both PSD, the last two terms on the RHS are always positive. This yields the proof. □

Lemma S.4. *Consider matrices $X \in \mathbb{R}^{d\times d'}$ and $A \in \mathbb{R}^{d\times d}$. Then,*

$$\left|\operatorname{Tr}\left(X^T A X\right)\right| \leq \|A\|_2 \operatorname{Tr}\left(X^T X\right)$$

Proof. For a matrix $Z \in \mathbb{R}^{d\times d}$, let the singular values be denoted as :

$$\sigma_{max}\left(Z\right) = \sigma_1\left(Z\right) \geq \sigma_2\left(Z\right) \dots \geq \sigma_d\left(Z\right)$$

Using Von-Neumann's trace inequality, we have

$$\begin{aligned}\left|\text{Tr}\left(X^T A X\right)\right| = \left|\text{Tr}\left(A X X^T\right)\right| &\leq \sum_{i=1}^{d} \sigma_i\left(A\right)\sigma_i\left(XX^T\right)\\ &\leq \sigma_{max}\left(A\right)\sum_{i=1}^{d}\sigma_i\left(XX^T\right)\\ &= \|A\|_2 \text{Tr}\left(XX^T\right) = \|A\|_2 \text{Tr}\left(X^T X\right)\end{aligned}$$

□

Lemma S.5. *Given the Markov property in a Markov chain, the reverse Markov property holds, i.e*

$$P\left(Z_t = s | Z_{t+1} = w, Z_{t+2} = s_{t+2} \dots Z_n = s_n\right) = P\left(Z_t = s | Z_{t+1} = w\right)$$

Proof. We have,

$$\begin{aligned}&P\left(Z_t = s | Z_{t+1} = w, Z_{t+2} = s_{t+2} \dots Z_n = s_n\right)\\ &\quad = \frac{P\left(Z_t = s, Z_{t+1} = w, Z_{t+2} = s_{t+2} \dots Z_n = s_n\right)}{P\left(Z_{t+1} = t, Z_{t+2} = s_{t+2} \dots Z_n = s_n\right)}\\ &\quad = \frac{P\left(Z_t = s, Z_{t+1} = w\right) P\left(Z_{t+2} = s_{t+2} \dots Z_n = s_n | Z_t = s, Z_{t+1} = w\right)}{P\left(Z_{t+1} = w\right) P\left(Z_{t+2} = s_{t+2} \dots Z_n = s_n | Z_{t+1} = w\right)}\\ &\quad = \frac{P\left(Z_t = s, Z_{t+1} = w\right) P\left(Z_{t+2} = s_{t+2} \dots Z_n = s_n | Z_{t+1} = w\right)}{P\left(Z_{t+1} = w\right) P\left(Z_{t+2} = s_{t+2} \dots Z_n = s_n | Z_{t+1} = w\right)}\\ &\quad = \frac{P\left(Z_t = s, Z_{t+1} = w\right)}{P\left(Z_{t+1} = w\right)}\\ &\quad = P\left(Z_t = s | Z_{t+1} = w\right)\end{aligned}$$

□

S.3.1 Proof of Lemma 2

Now we are ready to provide a proof of Lemma 2.

Proof of Lemma 2. Without loss of generality, we prove the statement for $m = 1$. For convenience of notation, we denote $k := k_1$. Note that,

$$B_{k,1} = \sum_{r=0}^{k} \sum_{(i_1, i_2 \dots i_r) \in G_r} \prod_{j=1}^{r} \eta_{i_j} A(s_{i_j}),\ G_r = \left\{(i_1, \dots, i_r) \in \{1, \dots, N\}^r : i_1 < \dots < i_r\right\}$$

with the convention that $\prod_\phi = I$. Therefore, since η_i forms a non-increasing sequence and $|G_r| = \binom{k}{r}$, we have,

$$\begin{aligned}\|B_{k,1} - I\|_2 = \left\|\sum_{r=1}^{k} \sum_{(i_1, i_2 \dots i_r) \in G_r} \prod_{j=1}^{r} \eta_{i_j} A(s_{i_j})\right\|_2 &\leq \sum_{r=1}^{k} \sum_{(i_1, i_2 \dots i_r) \in G_r} \left\|\prod_{j=1}^{r} \eta_{i_j} A(s_{i_j})\right\|_2\\ &\leq \sum_{r=1}^{k} \binom{k}{r} \left(\prod_{i=1}^{r} \eta_i\right) \left(\mathcal{M} + \lambda_1\right)^r \leq \sum_{r=1}^{k} \frac{k^r}{r!} \left(\prod_{i=1}^{r} \eta_i\right) \left(\mathcal{M} + \lambda_1\right)^r\\ &\leq \sum_{r=1}^{k} \frac{k^r}{r!} \eta_1^r \left(\mathcal{M} + \lambda_1\right)^r\\ &\leq \exp\left(k\eta_1\left(\mathcal{M} + \lambda_1\right)\right) - 1\\ &\leq k\eta_1\left(\mathcal{M} + \lambda_1\right)\left(1 + k\eta_1\left(\mathcal{M} + \lambda_1\right)\right) \text{ using } S.3\\ &\leq \left(1 + \epsilon\right) k\eta_1\left(\mathcal{M} + \lambda_1\right)\end{aligned} \tag{S.2}$$

where we have used the assumptions that $\|A(s)\|_2 \leq \|A(s) - \Sigma\| + \|\Sigma\|_2 = (\mathcal{M} + \lambda_1)$, $k\eta_1 (\mathcal{M} + \lambda_1) < 1$ and the useful result that

$$e^x \leq 1 + x + x^2, x \in [0, 1.79] \tag{S.3}$$

This completes the proof for (a).

For part (b), we have

$$\begin{aligned}
\left\| B_{k,1} - I - \sum_{t=1}^{k} \eta_t A(s_t) \right\|_2 &= \left\| \sum_{r=2}^{k} \sum_{(i_1, i_2 \ldots i_r) \in G_r} \prod_{j=1}^{r} \eta_{i_j} A(s_{i_j}) \right\|_2 \\
&\leq \sum_{r=2}^{k} \sum_{(i_1, i_2 \ldots i_r) \in G_r} \left\| \prod_{j=2}^{r} \eta_{i_j} A(s_{i_j}) \right\|_2 \\
&\leq \sum_{r=2}^{k} \binom{k}{r} \left(\prod_{i=2}^{r} \eta_i \right) (\mathcal{M} + \lambda_1)^r \\
&\leq \sum_{r=2}^{k} \frac{k^r}{r!} \left(\prod_{i=2}^{r} \eta_i \right) (\mathcal{M} + \lambda_1)^r \\
&\leq \sum_{r=2}^{k} \frac{k^r}{r!} \eta_1^r (\mathcal{M} + \lambda_1)^r \\
&\leq \exp (k\eta_1 (\mathcal{M} + \lambda_1)) - 1 - k\eta_1 (\mathcal{M} + \lambda_1) \\
&\leq k^2 \eta_1^2 (\mathcal{M} + \lambda_1)^2 \text{ using } S.3 \text{ along with } k\eta_1 (\mathcal{M} + \lambda_1) < 1
\end{aligned} \tag{S.4}$$

which completes the proof. □

S.3.2 Proof of Lemma 3

Before proving Lemma 3, we will need the following lemma.

Lemma S.6. *For arbitrary matrices $M_i \in \mathbb{R}^{d \times d}, i \in [n]$ and $Q \in \mathbb{R}^{n \times n}$, we have*

$$\left\| \sum_{x,y \in [n]} Q(x,y) M_x M_y^T \right\|_2 \leq \|Q\|_2 \left\| \sum_{x \in [n]} M_x M_x^T \right\|_2$$

where $\|.\|_2$ denotes the spectral norm.

Proof. Define matrix $X \in \mathbb{R}^{d \times nd}$ as $X := [M_1 \quad M_2 \quad \ldots \quad M_n]$. We note that

$$\begin{aligned}
\|X\|_2 &= \sqrt{\lambda_{max} (XX^T)} \\
&= \sqrt{\lambda_{max} \left(\sum_{x \in [n]} M_x M_x^T \right)} \\
&= \sqrt{\left\| \sum_{x \in [n]} M_x M_x^T \right\|_2} \text{ since } \sum_{x \in [n]} M_x M_x^T \text{ is a symmetric matrix}
\end{aligned}$$

Then, we have,

$$\begin{aligned}
\sum_{x,y \in [n]} Q(x,y) M_x M_y^T &= X (Q \otimes I_{d \times d}) X^T, \text{ where } \otimes \text{ denotes the kronecker product} \\
&\leq \|X\|_2^2 \|Q \otimes I_{d \times d}\|_2 \text{ using submultiplicativity of the spectral norm} \\
&= \|X\|_2^2 \|Q\|_2 \text{ since } \|A \otimes B\|_2 = \|A\|_2 \|B\|_2
\end{aligned}$$

which completes our proof. □

Proof of Lemma 3. We denote $k_i := k$ for convenience of notation. By using reversibility (see S.1), we know that the time-reversed process is also a Markov chain with the same transition probabilities. Then, for $i < j \leq i + k$ and any m,

$$\begin{aligned} P(s_i = s, s_j = t | s_{i+k} = u) &= P(s_i = s | s_j = t) P(s_j = t | s_{i+k} = u) \\ &\overset{(i)}{=} P^{j-i}(t, s) P^{i+k-j}(u, t) \\ &= P(s_m = s | s_{m-j+i} = t) P(s_{m-j+i} = t | s_{m-k} = u) \\ &= P(s_m = s, s_{m-j+i} = t | s_{m-k} = u) \end{aligned} \tag{S.5}$$

Step (i) uses reversibility. Therefore,

$$\begin{aligned} \mathbb{E}\left[\left(A\left(s_i\right) - \Sigma\right) S A\left(s_j\right) | s_{i+k}, \ldots s_n\right] &= \sum_{s,t} \left(\Sigma_s + \mu_s \mu_s^T - \Sigma\right) S \left(\Sigma_t + \mu_t \mu_t^T\right) P(s_i = s, s_j = t | s_{i+k}, \ldots s_n) \\ \text{using Lemma S.5 } &= \sum_{s,t} \left(\Sigma_s + \mu_s \mu_s^T - \Sigma\right) S \left(\Sigma_t + \mu_t \mu_t^T\right) P(s_i = s, s_j = t | s_{i+k}) \\ \text{using Eq S.5 } &= \sum_{s,t} \left(\Sigma_s + \mu_s \mu_s^T - \Sigma\right) S \left(\Sigma_t + \mu_t \mu_t^T\right) P(s_m = s, s_{m-j+i} = t | s_{m-k} = u) \\ &= \mathbb{E}\left[\left(A\left(s_m\right) - \Sigma\right) S A\left(s_{m-j+i}\right) | s_{m-k}\right] \\ &= \mathbb{E}\left[\left(A\left(s_j\right) - \Sigma\right) S A\left(s_i\right) | s_{j-k}\right] \text{ setting } m := j \end{aligned}$$

Therefore, without loss of generality, we proceed with the second form.

$$\begin{aligned} &\left\|\mathbb{E}\left[\left(A\left(s_j\right) - \Sigma\right) S A\left(s_i\right) | s_{j-k} = x_0\right]\right\|_2 \\ &\leq \underbrace{\left\|\mathbb{E}\left[\left(A\left(s_j\right) - \Sigma\right) S \Sigma | s_{j-k} = x_0\right]\right\|_2}_{T_1} + \underbrace{\left\|\mathbb{E}\left[\left(A\left(s_j\right) - \Sigma\right) S \left(A\left(s_i\right) - \Sigma\right) | s_{j-k} = x_0\right]\right\|_2}_{T_2} \end{aligned}$$

$$\begin{aligned} T_1 &= \left\|\mathbb{E}\left[\left(A\left(s_j\right) - \Sigma\right) S \Sigma | s_{j-k} = x_0\right]\right\|_2 \\ &= \left\|\mathbb{E}\left[\mathbb{E}_{D(s_j)}\left[\left(A\left(s_j\right) - \Sigma\right)\right] | s_{j-k} = x_0\right] S \Sigma\right\|_2 \\ &= \left\|\mathbb{E}\left[\left(\Sigma_{s_j} + \mu_{s_j} \mu_{s_j}^T - \Sigma\right) | s_{j-k} = x_0\right] S \Sigma\right\|_2 \\ &= \left\|\sum_{s \in \Omega} P^k(s_{j-k}, s) \left(\Sigma_s + \mu_s \mu_s^T - \Sigma\right) S \Sigma\right\|_2 \\ &\leq \left\|\sum_{s \in \Omega} \left(P^k(s_{j-k}, s) - \pi\left(s\right)\right) \left(\Sigma_s + \mu_s \mu_s^T - \Sigma\right) + \underbrace{\mathbb{E}_\pi\left[\left(\Sigma_s + \mu_s \mu_s^T - \Sigma\right)\right]}_{=0}\right\|_2 \left\|S\right\|_2 \left\|\Sigma\right\|_2 \\ &= \lambda_1 \left\|S\right\|_2 \left(\left\|\sum_{s \in \Omega} \left(P^k(s_{j-k}, s) - \pi\left(s\right)\right) \left(\Sigma_s + \mu_s \mu_s^T - \Sigma\right)\right\|_2\right) \\ &\leq \lambda_1 \left\|S\right\|_2 \mathcal{M} \sum_{s \in \Omega} \left|P^k(s_{j-k}, s) - \pi\left(s\right)\right| \\ &\leq 2\lambda_1 \left\|S\right\|_2 \mathcal{M} d_{\text{mix}}\left(k_{i+1}\right) \\ &\leq 2\eta_i^2 \mathcal{M} \lambda_1 \left\|S\right\|_2 \end{aligned} \tag{S.6}$$

$$
\begin{aligned}
T_2 &= \left\|\mathbb{E}\left[\left(A\left(s_j\right)-\Sigma\right) S\left(A\left(s_i\right)-\Sigma\right) | s_{j-k}=x_0\right]\right\|_2 \\
&= \left\|\sum_{x,y\in\Omega} \mathbb{P}\left(s_j=x, s_i=y | s_{j-k}=x_0\right) \mathbb{E}_{D(x)}\left[A\left(x\right)-\Sigma\right] S \mathbb{E}_{D(y)}\left[A\left(y\right)-\Sigma\right]\right\|_2 \text{ using independence of} \\
&\quad D\left(x\right) \text{ and } D\left(y\right) \text{ conditioned on } x \text{ and } y \\
&= \left\|\sum_{x,y\in\Omega} \mathbb{P}\left(s_j=x, s_i=y | s_{j-k}=x_0\right) \underbrace{\left(\Sigma_x+\mu_x\mu_x^T-\Sigma\right) S^{\frac{1}{2}}}_{W_x} \underbrace{S^{\frac{1}{2}}\left(\Sigma_y+\mu_y\mu_y^T-\Sigma\right)}_{W_y^T}\right\|_2 \\
&= \left\|\sum_{x,y\in\Omega} \mathbb{P}\left(s_j=x | s_i=y\right) \mathbb{P}\left(s_i=y | s_{j-k}=x_0\right) W_x W_y^T\right\|_2 \quad \text{using the Markov property} \\
&= \left\|\sum_{x,y\in\Omega} P^{j-i}\left(y,x\right) P^{i-j+k}\left(x_0,y\right) W_x W_y^T\right\|_2 \\
&= \left\|\sum_{x,y\in\Omega} \left(P^{j-i}\left(y,x\right)-\pi\left(x\right)\right) P^{i-j+k}\left(x_0,y\right) W_x W_y^T + \sum_{x,y\in\Omega} \pi\left(x\right) P^{i-j+k}\left(x_0,y\right) W_x W_y^T\right\|_2 \\
&= \left\|\sum_{x,y\in\Omega} \left(P^{j-i}\left(y,x\right)-\pi\left(x\right)\right) P^{i-j+k}\left(x_0,y\right) W_x W_y^T + \underbrace{\sum_{x\in\Omega} \pi\left(x\right) W_x}_{=0} \sum_{y\in\Omega} P^{i-j+k}\left(x_0,y\right) W_y^T\right\|_2
\end{aligned}
$$

$$
\begin{aligned}
&= \left\|\sum_{x,y\in\Omega} \left(P^{j-i}\left(y,x\right)-\pi\left(x\right)\right) P^{i-j+k}\left(x_0,y\right) W_x W_y^T\right\|_2 \\
&\leq \underbrace{\left\|\sum_{x,y\in\Omega} \left(P^{j-i}\left(y,x\right)-\pi\left(x\right)\right) \left(P^{j-i+k}\left(x_0,y\right)-\pi\left(y\right)\right) W_x W_y^T\right\|_2}_{T_{21}} + \underbrace{\left\|\sum_{x,y\in\Omega} \left(P^{j-i}\left(y,x\right)-\pi\left(x\right)\right) \pi\left(y\right) W_x W_y^T\right\|_2}_{T_{22}}
\end{aligned} \tag{S.7}
$$

For T_{21}, we have,

$$
\begin{aligned}
T_{21} &\leq \sum_{x,y\in\Omega} \left|P^{j-i}\left(y,x\right)-\pi\left(x\right)\right| \left|P^{i-j+k}\left(x_0,y\right)-\pi\left(y\right)\right| \left\|W_x W_y^T\right\|_2 \\
&\leq \left\|S\right\|_2 \mathcal{M}^2 \sum_{y\in\Omega} \left|P^{i-j+k}\left(x_0,y\right)-\pi\left(y\right)\right| \sum_{x\in\Omega} \left|P^{j-i}\left(y,x\right)-\pi\left(x\right)\right| \\
&\leq 2\left\|S\right\|_2 \mathcal{M}^2 d_{\text{mix}}\left(j-i\right) \sum_{y\in\Omega} \left|P^{i-j+k}\left(x_0,y\right)-\pi\left(y\right)\right| \\
&\leq 4\left\|S\right\|_2 \mathcal{M}^2 d_{\text{mix}}\left(j-i\right) d_{\text{mix}}\left(i-j+k\right) \\
&\leq 4\left\|S\right\|_2 \mathcal{M}^2 2^{-\left\lfloor\frac{j-i}{\tau_{\text{mix}}}\right\rfloor} 2^{-\left\lfloor\frac{i-j+k}{\tau_{\text{mix}}}\right\rfloor} \\
&\leq 8\left\|S\right\|_2 \mathcal{M}^2 2^{-\left\lfloor\frac{j-i+i-j+k}{\tau_{\text{mix}}}\right\rfloor} \text{ since } \forall a,b \;\; \lfloor a\rfloor+\lfloor b\rfloor \geq \lfloor a+b\rfloor-1 \\
&\leq 8\left\|S\right\|_2 \mathcal{M}^2 2^{-\left\lfloor\frac{k}{\tau_{\text{mix}}}\right\rfloor} \leq 8\left\|S\right\|_2 \mathcal{M}^2 d_{\text{mix}}\left(k\right) \leq 8\eta_i^2 \mathcal{M}^2 \left\|S\right\|_2
\end{aligned} \tag{S.8}
$$

For T_{22}, we have,

$$
\begin{aligned}
T_{22} &= \left\| \sum_{x,y\in\Omega} \left(P^{j-i}\left(y,x\right) - \pi\left(x\right)\right)\pi\left(y\right)W_x W_y^T \right\|_2 \\
&= \left\| \sum_{x,y\in\Omega} \frac{\left(P^{j-i}\left(y,x\right) - \pi\left(x\right)\right)}{\sqrt{\pi\left(x\right)}}\sqrt{\pi\left(y\right)}\left(\sqrt{\pi\left(x\right)}W_x\right)\left(\sqrt{\pi\left(y\right)}W_y^T\right) \right\|_2 \\
&= \left\| \sum_{x,y\in\Omega} \frac{\left(P^{j-i}\left(y,x\right) - \pi\left(x\right)\right)}{\sqrt{\pi\left(x\right)}}\sqrt{\pi\left(y\right)}\left(\sqrt{\pi\left(x\right)}\left(\Sigma_x + \mu_x\mu_x^T - \Sigma\right)S^{\frac{1}{2}}\right)\left(\sqrt{\pi\left(y\right)}S^{\frac{1}{2}}\left(\Sigma_y + \mu_y\mu_y^T - \Sigma\right)\right) \right\|_2 \\
&\overset{(i)}{\leq} \left\|Q\right\|_2 \left\| \sum_{x\in\Omega} \pi\left(x\right)\left(\Sigma_x + \mu_x\mu_x^T - \Sigma\right)S\left(\Sigma_x + \mu_x\mu_x^T - \Sigma\right) \right\|_2 \\
&= \left\|Q\right\|_2 \left\|\mathbb{E}_\pi\left[\left(\Sigma_x + \mu_x\mu_x^T - \Sigma\right)S\left(\Sigma_x + \mu_x\mu_x^T - \Sigma\right)\right]\right\|_2 \\
&\leq \left\|Q\right\|_2 \left\|S\right\|_2 \left\|\mathbb{E}_\pi\left[\left(\Sigma_x + \mu_x\mu_x^T - \Sigma\right)^2\right]\right\|_2 \\
&\leq \mathcal{V}\left\|Q\right\|_2 \left\|S\right\|_2
\end{aligned} \tag{S.9}
$$

Step (i) uses Lemma S.6 with $Q(y,x) := \frac{\left(P^{i-j}(y,x)-\pi(x)\right)}{\sqrt{\pi(x)}}\sqrt{\pi\left(y\right)}$ and $M_x = \sqrt{\pi\left(x\right)}\left(\Sigma_x + \mu_x\mu_x^T - \Sigma\right)S^{\frac{1}{2}}$. Let's now bound $\left\|Q\right\|_2$. Let $\Pi := \operatorname{diag}\left(\pi\right) \in \mathbb{R}^{\Omega\times\Omega}$ and $t := j - i$. Then, we have

$$
\begin{aligned}
Q &= \Pi^{\frac{1}{2}}\left(P^t - \mathbb{1}\mathbb{1}^T\Pi\right)\Pi^{-\frac{1}{2}} \\
&= \Pi^{\frac{1}{2}}P^t\Pi^{-\frac{1}{2}} - \Pi^{\frac{1}{2}}\mathbb{1}\mathbb{1}^T\Pi^{\frac{1}{2}}
\end{aligned}
$$

Now, since we have a reversible Markov chain, $\Pi P = P^T\Pi$. Therefore,

$$
\begin{aligned}
\Pi^{\frac{1}{2}}P\Pi^{-\frac{1}{2}} &= \Pi^{\frac{1}{2}}\Pi^{-1}P^T\Pi\Pi^{-\frac{1}{2}} \\
&= \Pi^{-\frac{1}{2}}P^T\Pi^{\frac{1}{2}}
\end{aligned}
$$

Therefore, P is similar to the self-adjoint matrix $\Pi^{\frac{1}{2}}P\Pi^{-\frac{1}{2}}$ and their eigenvalues are real and the same. Further note that $\Pi^{\frac{1}{2}}\mathbb{1}$ is the leading eigenvector of $\Pi^{\frac{1}{2}}P\Pi^{-\frac{1}{2}}$ with eigenvalue 1 since

$$
\begin{aligned}
\Pi^{\frac{1}{2}}P\Pi^{-\frac{1}{2}}\Pi^{\frac{1}{2}}\mathbb{1} &= \Pi^{\frac{1}{2}}P\mathbb{1} \\
&= \Pi^{\frac{1}{2}}\mathbb{1} \text{ since } P \text{ is a stochastic matrix}
\end{aligned}
$$

Now,

$$
\begin{aligned}
\left\|Q\right\|_2 &= \left\|\Pi^{\frac{1}{2}}P^t\Pi^{-\frac{1}{2}} - \Pi^{\frac{1}{2}}\mathbb{1}\mathbb{1}^T\Pi^{\frac{1}{2}}\right\|_2 \\
&= \left\|\left(\Pi^{\frac{1}{2}}P\Pi^{-\frac{1}{2}}\right)^t - \Pi^{\frac{1}{2}}\mathbb{1}\mathbb{1}^T\Pi^{\frac{1}{2}}\right\|_2 \\
&\leq \left|\lambda_2\left(\Pi^{\frac{1}{2}}P\Pi^{-\frac{1}{2}}\right)\right|^t \\
&= \left|\lambda_2\left(P\right)\right|^t
\end{aligned}
$$

where $\left|\lambda_2\left(.\right)\right|$ denotes the second-largest eigenvalue in magnitude. Therefore, using S.6, S.8 and S.9, we have

$$
\begin{aligned}
\mathbb{E}\left[\left(A\left(s_i\right) - \Sigma\right)SA\left(s_j\right)|s_{i+k},...s_n\right] &\leq \left(\left|\lambda_2\left(P\right)\right|^{j-i}\mathcal{V} + 8\eta_i^2\mathcal{M}^2 + 2\eta_i^2\mathcal{M}\lambda_1\right)\left\|S\right\|_2 \\
&\leq \left(\left|\lambda_2\left(P\right)\right|^{j-i}\mathcal{V} + 8\eta_i^2\mathcal{M}\left(\mathcal{M} + \lambda_1\right)\right)\left\|S\right\|_2
\end{aligned}
$$

Hence proved. □

Lemma S.7. *Let $\forall i \in [n], \eta_i k_i (\mathcal{M}+\lambda_1) \leq \epsilon, \epsilon \in (0,1)$ and η_i forms a non-increasing sequence. Set $k_i := \tau_{mix}\left(\gamma\eta_i^2\right), \gamma \in (0,1]$. Then for constant matrix $U \in \mathbb{R}^{d\times d'}$, and constant positive semi-definite matrix $G \in \mathbb{R}^{d\times d}$, $i \leq j \leq n,\ j - i \geq k_i$, we have*

$$\begin{aligned}
&\left|\mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+1} G\left(A_i-\Sigma\right) B_{j,i+1}^T U\right)\right]\right| \\
&\quad \leq \eta_{i+1}\|G\|_2\left(\frac{2\mathcal{V}\left|\lambda_2(P)\right|}{1-\left|\lambda_2(P)\right|}+\eta_{i+1}\mathcal{M}\left(2\gamma(1+8\epsilon)+\left(2+(1+\epsilon)^2\right)k_{i+1}^2(\mathcal{M}+\lambda_1)^2\right)\right) \\
&\qquad \times \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k_{i+1}} B_{j,i+k_{i+1}}^T U\right)\right]
\end{aligned}$$

where $B_{j,i}$ is defined in 7.

Proof. For the convenience of notation, we denote $k_{i+1} := k$. Let $B_{j,i+1} = B_{j,i+k}(I+R)$, then

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+1} G\left(A_i-\Sigma\right) B_{j,i+1}^T U\right)\right] = \\
&\quad \mathbb{E}\left[\underbrace{\operatorname{Tr}\left(U^T B_{j,i+k} G\left(A_i-\Sigma\right) B_{j,i+k}^T U\right)}_{T_1}\right]+\mathbb{E}\left[\underbrace{\operatorname{Tr}\left(U^T B_{j,i+k} G\left(A_i-\Sigma\right) R^T B_{j,i+k}^T U\right)}_{T_2}\right]+ \\
&\quad \mathbb{E}\left[\underbrace{\operatorname{Tr}\left(U^T B_{j,i+k} R G\left(A_i-\Sigma\right) B_{j,i+k}^T U\right)}_{T_3}\right]+\mathbb{E}\left[\underbrace{\operatorname{Tr}\left(U^T B_{j,i+k} R G\left(A_i-\Sigma\right) R^T B_{j,i+k}^T U\right)}_{T_4}\right]
\end{aligned} \tag{S.10}$$

We will now bound each of the terms $\mathbb{E}[T_1], \mathbb{E}[T_2], \mathbb{E}[T_3]$ and $\mathbb{E}[T_4]$.

$$\begin{aligned}
\mathbb{E}[T_1] &= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G\left(A_i-\Sigma\right) B_{j,i+k}^T U\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G\left(A_i-\Sigma\right) B_{j,i+k}^T U\right)\middle| s_{i+k},\ldots s_{j-1}, s_j\right]\right] \\
&= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G\, \mathbb{E}\left[\left(A_i-\Sigma\right)\middle| s_{i+k},\ldots s_{j-1}, s_j\right] B_{j,i+k}^T U\right)\right] \\
&= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G\, \mathbb{E}\left[\left(A_i-\Sigma\right)\middle| s_{i+k}\right] B_{j,i+k}^T U\right)\right] \quad \text{using Lemma S.5}
\end{aligned}$$

Now, using Lemma 1, we have,

$$\begin{aligned}
\left\|\mathbb{E}\left[\left(A_i-\Sigma\right)\middle| s_{i+k}\right]\right\|_2 &= \left\|\sum_{s\in\Omega} P^k(s_{i+k}, s)\left(A_i-\Sigma\right)\right\|_2 \\
&= \left\|\sum_{s\in\Omega}\left(P^k(s_{i+k}, s)-\pi(s)\right)\left(A_i-\Sigma\right)+\underbrace{\mathbb{E}_\pi\left[\left(A_i-\Sigma\right)\right]}_{=0}\right\|_2 \\
&= \left\|\sum_{s\in\Omega}\left(P^k(s_{i+k}, s)-\pi(s)\right)\left(A_i-\Sigma\right)\right\|_2 \\
&\leq \mathcal{M}\sum_{s\in\Omega}\left|P^k(s_{i+k}, s)-\pi(s)\right| \\
&\leq 2\mathcal{M} d_{\text{mix}}(k_{i+1}) \\
&\leq 2\gamma\eta_{i+1}^2\mathcal{M}
\end{aligned} \tag{S.11}$$

where we have used Lemma S.4. Therefore,

$$\left|\mathbb{E}[T_1]\right| \leq \gamma\eta_{i+1}^2\mathcal{M}\|G\|_2\mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \tag{S.12}$$

We will now bound $\mathbb{E}[T_2]$. Let $R_0 := \sum_{\ell=i+1}^{i+k-1}\eta_\ell A_\ell$. Using Lemma 2 we have

$$\|R-R_0\|_2 \leq \eta_{i+1}^2 k_{i+1}^2(\mathcal{M}+\lambda_1)^2$$

Then,

$$\begin{aligned}
\mathbb{E}[T_2] &= \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} G\left(A_i - \Sigma\right) R^T B_{j,i+k}^T U\right)\right] \\
&= \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} G\left(A_i - \Sigma\right) R_0^T B_{j,i+k}^T U\right)\right] + \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} G\left(A_i - \Sigma\right)\left(R - R_0\right)^T B_{j,i+k}^T U\right)\right] \\
&= \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} G \mathbb{E}[\left(A_i - \Sigma\right) R_0^T | s_{i+k}, \ldots s_{j-1}, s_j] B_{j,i+k}^T U\right)\right] + \\
&\qquad \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} G\left(A_i - \Sigma\right)\left(R - R_0\right)^T B_{j,i+k}^T U\right)\right]
\end{aligned}$$

Using Lemma 3 with $S := I$ we have,

$$\begin{aligned}
\|\mathbb{E}\left[\left(A_i - \Sigma\right) R_0^T | s_{i+k}, \ldots, s_j\right]\|_2 &\leq \sum_{\ell=i+1}^{i+k-1} \eta_\ell \left(\left|\lambda_2\left(P\right)\right|^{\ell-i} \mathcal{V} + 8\gamma\eta_{i+1}^2 \mathcal{M}\left(\mathcal{M} + \lambda_1\right)\right) \\
&\leq \eta_{i+1} \mathcal{V} \frac{\left|\lambda_2\left(P\right)\right|}{1 - \left|\lambda_2\left(P\right)\right|} + 8\gamma\eta_{i+1}^3 k_{i+1} \mathcal{M}\left(\mathcal{M} + \lambda_1\right) \qquad \text{(S.13)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\mathbb{E}\left[T_2\right]\right| \\
&\leq \|G\|_2 \left(\eta_{i+1} \frac{\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1 - \left|\lambda_2\left(P\right)\right|} + 8\gamma\eta_{i+1}^3 k_{i+1} \mathcal{M}\left(\mathcal{M} + \lambda_1\right) + \eta_{i+1}^2 k_{i+1}^2 \mathcal{M}\left(\mathcal{M} + \lambda_1\right)^2\right) \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \\
&= \eta_{i+1}\|G\|_2 \left(\frac{\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1 - \left|\lambda_2\left(P\right)\right|} + 8\gamma\eta_{i+1}^2 k_{i+1} \mathcal{M}\left(\mathcal{M} + \lambda_1\right) + \eta_{i+1} k_{i+1}^2 \mathcal{M}\left(\mathcal{M} + \lambda_1\right)^2\right) \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right]
\end{aligned} \qquad \text{(S.14)}$$

Similarly using Lemma 3 with $S := G$,

$$\left|\mathbb{E}\left[T_3\right]\right| \leq \eta_{i+1}\|G\|_2 \left(\frac{\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1 - \left|\lambda_2\left(P\right)\right|} + 8\gamma\eta_{i+1}^2 k_{i+1} \mathcal{M}\left(\mathcal{M} + \lambda_1\right) + \eta_{i+1} k_{i+1}^2 \mathcal{M}\left(\mathcal{M} + \lambda_1\right)^2\right) \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \qquad \text{(S.15)}$$

Finally,

$$\begin{aligned}
\left|\mathbb{E}\left[T_4\right]\right| &\leq \mathcal{M}\|G\|_2\|R\|_2^2 \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \\
&\leq \left(1 + \epsilon\right)^2 \eta_{i+1}^2 k_{i+1}^2 \mathcal{M}\left(\mathcal{M} + \lambda_1\right)^2 \|G\|_2 \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \text{ using Lemma 2}
\end{aligned} \qquad \text{(S.16)}$$

Therefore, using Eqs S.12, S.14, S.15, S.16 along with S.10, we have

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+1} G\left(A_i - \Sigma\right) B_{j,i+1}^T U\right)\right]\right| \\
&\leq \eta_{i+1}\|G\|_2 \left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1 - \left|\lambda_2\left(P\right)\right|} + \eta_{i+1}\mathcal{M}\left(2\gamma + 16\gamma\eta_{i+1} k_{i+1}\left(\mathcal{M} + \lambda_1\right) + \left(2 + \left(1 + \epsilon\right)^2\right) k_{i+1}^2 \left(\mathcal{M} + \lambda_1\right)^2\right)\right) \\
&\qquad \times \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \\
&\leq \eta_{i+1}\|G\|_2 \left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1 - \left|\lambda_2\left(P\right)\right|} + \eta_{i+1}\mathcal{M}\left(2\gamma\left(1 + 8\epsilon\right) + \left(2 + \left(1 + \epsilon\right)^2\right) k_{i+1}^2 \left(\mathcal{M} + \lambda_1\right)^2\right)\right) \\
&\qquad \times \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right]
\end{aligned}$$

where in the last line we used $\eta_{i+1} k_{i+1}\left(\mathcal{M} + \lambda_1\right) \leq \epsilon$. Hence proved. □

Lemma S.8. *Let $\forall i \in [n], \eta_i k_i\left(\mathcal{M} + \lambda_1\right) \leq \epsilon, \epsilon \in (0,1)$ and η_i forms a non-increasing sequence. Set $k_i := \tau_{mix}\left(\gamma\eta_i^2\right), \gamma \in (0,1]$. Then for constant matrices $U \in \mathbb{R}^{d \times d'}$, $G \in \mathbb{R}^{d \times d}$, $i \leq j \leq n,\ j - i \geq k_i$, we have*

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+1} G\left(A_i - \Sigma\right)^2 B_{j,i+1}^T U\right)\right]\right| \\
&\qquad \leq \left(\mathcal{V} + \eta_{i+1}\mathcal{M}^2\left(2\gamma\eta_{i+1} + \left(1 + \epsilon\right)\left(2 + \epsilon\left(1 + \epsilon\right)\right) k_{i+1}\left(\mathcal{M} + \lambda_1\right)\right)\right) \|G\|_2 \mathbb{E}\left[\mathrm{Tr}\left(U^T B_{j,i+k_{i+1}} B_{j,i+k_{i+1}}^T U\right)\right]
\end{aligned}$$

where $B_{j,i}$ is defined in 7.

Proof. For convenience of notation, we denote $k_{i+1} := k$. Let $B_{j,i+1} = B_{j,i+k}(I+R)$, then

$$\mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+1} G (A_i - \Sigma)^2 B_{j,i+1}^T U\right)\right] =$$

$$\mathbb{E}\left[\underbrace{\operatorname{Tr}\left(U^T B_{j,i+k} G (A_i - \Sigma)^2 B_{j,i+k}^T U\right)}_{T_1}\right] + \mathbb{E}\left[\underbrace{\operatorname{Tr}\left(U^T B_{j,i+k} G (A_i - \Sigma)^2 R^T B_{j,i+k}^T U\right)}_{T_2}\right] +$$

$$\mathbb{E}\left[\underbrace{\operatorname{Tr}\left(U^T B_{j,i+k} R G (A_i - \Sigma)^2 B_{j,i+k}^T U\right)}_{T_3}\right] + \mathbb{E}\left[\underbrace{\operatorname{Tr}\left(U^T B_{j,i+k} R G (A_i - \Sigma)^2 R^T B_{j,i+k}^T U\right)}_{T_4}\right]$$

We will now bound each of the terms $\mathbb{E}[T_1], \mathbb{E}[T_2], \mathbb{E}[T_3]$ and $\mathbb{E}[T_4]$.

Since $\left\|\mathbb{E}_\pi\left[(A_t - \Sigma)^2\right]\right\|_2 \leq \mathcal{V}$, therefore

$$\begin{aligned}
\mathbb{E}[T_1] &= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G (A_i - \Sigma)^2 B_{j,i+k}^T U\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G (A_i - \Sigma)^2 B_{j,i+k}^T U\right) \middle| s_{i+k}, \ldots s_{j-1}, s_j\right]\right] \\
&= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G\, \mathbb{E}\left[(A_i - \Sigma)^2 \middle| s_{i+k}, \ldots s_{j-1}, s_j\right] B_{j,i+k}^T U\right)\right] \\
&= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G\, \mathbb{E}\left[(A_i - \Sigma)^2 \middle| s_{i+k}\right] B_{j,i+k}^T U\right)\right] \text{ using Lemma S.5} \\
&\overset{(i)}{\leq} \left(\mathcal{V} + 2 d_{\text{mix}}(k)\, \mathcal{M}^2\right) \|G\|_2\, \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k}\ B_{j,i+k}^T U\right)\right]
\end{aligned}$$

where in (i), we used similar steps as S.11 to get

$$\left\|\mathbb{E}\left[(A_i - \Sigma)^2 \middle| s_{i+k}\right]\right\|_2 \leq \left\|\mathbb{E}_\pi\left[(A_i - \Sigma)^2\right]\right\|_2 + 2 d_{\text{mix}}(k)\, \mathcal{M}^2 \tag{S.17}$$

Next, using Lemma 2 we have that

$$\|R\|_2 \leq (1+\epsilon)\, k_{i+1} \eta_{i+1} (\mathcal{M} + \lambda_1). \tag{S.18}$$

Therefore,

$$\begin{aligned}
\mathbb{E}[T_2] &= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} G (A_i - \Sigma)^2 R^T B_{j,i+k}^T U\right)\right] \\
&\leq (1+\epsilon)\, k_{i+1} \eta_{i+1} \mathcal{M}^2 (\mathcal{M} + \lambda_1) \|G\|_2\, \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right]
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}[T_3] &= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} R G (A_i - \Sigma)^2 B_{j,i+k}^T U\right)\right] \\
&\leq (1+\epsilon)\, k_{i+1} \eta_{i+1} \mathcal{M}^2 (\mathcal{M} + \lambda_1) \|G\|_2\, \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right]
\end{aligned}$$

Finally, using the bound on $\|R\|_2$ from Eq S.18, we have:

$$\begin{aligned}
\mathbb{E}[T_4] &= \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} R G (A_i - \Sigma)^2 R^T B_{j,i+k}^T U\right)\right] \\
&\leq (1+\epsilon)^2 k_{i+1}^2 \eta_{i+1}^2 \mathcal{M}^2 (\mathcal{M} + \lambda_1)^2 \|G\|_2\, \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \\
&\leq \epsilon (1+\epsilon)^2 k_{i+1} \eta_{i+1} \mathcal{M}^2 (\mathcal{M} + \lambda_1) \|G\|_2\, \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \text{ using } \forall i,\ \eta_i k_i (M + \lambda_1) \leq c
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+1} G (A_i - \Sigma)^2 B_{j,i+1}^T U\right)\right]\right| \\
&\overset{(i)}{\leq} \left(\mathcal{V} + \eta_{i+1}\left(2\gamma\eta_{i+1}\mathcal{M}^2 + (1+\epsilon)(2 + \epsilon(1+\epsilon))\, k_{i+1} \mathcal{M}^2 (\mathcal{M} + \lambda_1)\right)\right) \|G\|_2\, \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right] \\
&= \left(\mathcal{V} + \eta_{i+1}\mathcal{M}^2\left(2\gamma\eta_{i+1} + (1+\epsilon)(2 + \epsilon(1+\epsilon))\, k_{i+1} (\mathcal{M} + \lambda_1)\right)\right) \|G\|_2\, \mathbb{E}\left[\operatorname{Tr}\left(U^T B_{j,i+k} B_{j,i+k}^T U\right)\right]
\end{aligned}$$

where in (i), we used $d_{\text{mix}}(k) = d_{\text{mix}}(k_{i+1}) \leq \gamma\eta_{i+1}^2$. Hence proved. □

Lemma S.9. *Let $\forall i \in [n], \eta_i k_i (\mathcal{M} + \lambda_1) \le \epsilon, \epsilon \in (0,1)$ and step-sizes η_i forms a non-increasing sequence. Further, let the step-sizes follow a slow-decay property, i.e, $\forall i, \eta_i \le \eta_{i-k_i} \le 2\eta_i$. Set $k_i := \tau_{mix}\left(\gamma \eta_i^2\right), \gamma \in (0,1]$. Let $G \in \mathbb{R}^{d\times d}$ be a constant positive semi-definite matrix, and $P_t := \text{Tr}\left(B_{t-1}B_{t-1}^T G(A_t - \Sigma)\right)$, then,*

$$\mathbb{E}\left[P_t\right] \le \eta_{t-k_t}\left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|} + \eta_{t-k_t}\mathcal{M}\left(2\gamma\left(1+8\epsilon\right) + \left(2+\left(1+\epsilon\right)^2\right)k_t^2\left(\mathcal{M}+\lambda_1\right)^2\right)\right)\|G\|_2\,\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right]$$

where B_t is defined in 2.

Proof. Let $B_t = (I+R)\,B_{t-k_t}$ with $\|R\|_2 \le r$. Then,

$$\begin{aligned}\mathbb{E}\left[P_t\right] = {}& \mathbb{E}\left[\underbrace{\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T G(A_t-\Sigma)\right)}_{P_{t,1}}\right] + \mathbb{E}\left[\underbrace{\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T R^T G(A_t-\Sigma)\right)}_{P_{t,2}}\right] \\ &+ \mathbb{E}\left[\underbrace{\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T G(A_t-\Sigma)R\right)}_{P_{t,3}}\right] + \mathbb{E}\left[\underbrace{\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T R^T G(A_t-\Sigma)R\right)}_{P_{t,4}}\right]\end{aligned}$$

Let's consider each of the terms above. Using Von-Neumann's trace inequality and S.14, we have,

$$\begin{aligned}\mathbb{E}\left[P_{t,1}\right] &= \mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\mathbb{E}\left[G(A_t-\Sigma)|s_1,s_2,\ldots,s_{t-k_t}\right]\right)\right] \\ &\le \mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T G\mathbb{E}\left[(A_t-\Sigma)|s_{t-k_t}\right]\right)\right] \\ &\le \|G\mathbb{E}\left[(A_t-\Sigma)|s_{t-k_t}\right]\|_2\,\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\ &\le 2\mathcal{M}d_{\text{mix}}\left(k_t\right)\|G\|_2\,\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \text{ using } S.11 \\ &\le 2\gamma\eta_t^2\mathcal{M}\|G\|_2\,\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right]\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[P_{t,2}\right] &= \mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T, \mathbb{E}\left[R^T G(A_t-\Sigma)U|s_1,s_2,\ldots,s_{t-k_t}\right]\right)\right] \\ &\le \left\|\mathbb{E}\left[R^T G(A_t-\Sigma)|s_1,s_2,\ldots,s_{t-k_t}\right]\right\|_2\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\ &= \left\|\mathbb{E}\left[R^T G(A_t-\Sigma)|s_{t-k_t}\right]\right\|_2\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\ &\le \eta_{t-k_t}\|G\|_2\left(\frac{\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|} + 8\gamma\eta_{t-k_t}^2 k_t\mathcal{M}\left(\mathcal{M}+\lambda_1\right) + \eta_{t-k_t}k_t^2\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2\right)\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\ &\qquad \text{using } S.14\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[P_{t,3}\right] &\le \eta_{t-k_t}\|G\|_2\left(\frac{\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|} + 8\gamma\eta_{t-k_t}^2 k_t\mathcal{M}\left(\mathcal{M}+\lambda_1\right) + \eta_{t-k_t}k_t^2\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2\right)\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\ &\qquad \text{using similar steps as } \mathbb{E}\left[P_{t,2}\right]\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[P_{t,4}\right] &= \mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T R^T G(A_t-\Sigma)R\right)\right] \\ &\le r^2\mathcal{M}\|G\|_2\,\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\ &\le \left(1+\epsilon\right)^2\eta_{t-k_t+1}^2 k_t^2\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2\|G\|_2\,\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \text{ using Lemma 2} \\ &\le \left(1+\epsilon\right)^2\eta_{t-k_t}^2 k_t^2\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2\|G\|_2\,\mathbb{E}\left[\text{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right]\end{aligned}$$

Therefore we have,

$$\begin{aligned}
&\mathbb{E}\left[P_t\right] \\
&\leq \eta_{t-k_t}\left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|}+\mathcal{M}\left(2\gamma\eta_t+16\gamma\eta_{t-k_t}^2 k_t\left(\mathcal{M}+\lambda_1\right)+\left(2+\left(1+\epsilon\right)^2\right)\eta_{t-k_t}k_t^2\left(\mathcal{M}+\lambda_1\right)^2\right)\right)\left\|G\right\|_2 \\
&\qquad\qquad \times \mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\
&\overset{(i)}{\leq} \eta_{t-k_t}\left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|}+\eta_{t-k_t}\mathcal{M}\left(2\gamma+16\gamma\eta_t k_t\left(\mathcal{M}+\lambda_1\right)+\left(2+\left(1+\epsilon\right)^2\right)k_t^2\left(\mathcal{M}+\lambda_1\right)^2\right)\right)\left\|G\right\|_2 \\
&\qquad\qquad \times \mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\
&\overset{(ii)}{\leq} \eta_{t-k_t}\left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|}+\eta_{t-k_t}\mathcal{M}\left(2\gamma\left(1+8\epsilon\right)+\left(2+\left(1+\epsilon\right)^2\right)k_t^2\left(\mathcal{M}+\lambda_1\right)^2\right)\right)\left\|G\right\|_2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right]
\end{aligned}$$

where in (i) we used $2\eta_{t-k_t} \leq \eta_t \leq \eta_{t-k_t}$ along with $\eta_t k_t\left(\mathcal{M}+\lambda_1\right) \leq \epsilon$ in (ii). Hence proved. $\square$

Lemma S.10. *Let $\forall i \in [n], \eta_i k_i\left(\mathcal{M}+\lambda_1\right) \leq \epsilon, \epsilon \in (0,1)$ and η_i forms a non-increasing sequence. Set $k_i := \tau_{mix}\left(\gamma\eta_i^2\right), \gamma \in (0,1]$. Let $U \in \mathbb{R}^{d\times d}$ be a constant matrix and $Q_t := \mathrm{Tr}\left(B_{t-1}B_{t-1}^T(A_t-\Sigma)U(A_t-\Sigma)\right)$. Further, let the decay of the step-sizes be slow such that $\forall i,\ \eta_i \leq \eta_{i-k_i} \leq 2\eta_i$. Then*

$$\mathbb{E}\left[Q_t\right] \leq \left(\mathcal{V}+\eta_{t-k_t+1}\mathcal{M}^2\left(2\gamma\eta_t+2\left(1+\epsilon\right)\left(1+\epsilon\left(1+\epsilon\right)\right)k_t\left(\mathcal{M}+\lambda_1\right)\right)\right)\left\|U\right\|_2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right]$$

where B_t is defined in 2.

Proof. Let $B_t = \left(I+R\right)B_{t-k_t}$ with $\|R\|_2 \leq r$. Then,

$$\begin{aligned}
\mathbb{E}\left[Q_t\right] = &\ \mathbb{E}\Bigg[\underbrace{\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T(A_t-\Sigma)U(A_t-\Sigma)\right)}_{Q_{t,1}}\Bigg] + \mathbb{E}\Bigg[\underbrace{\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T R^T(A_t-\Sigma)U(A_t-\Sigma)\right)}_{Q_{t,2}}\Bigg] \\
&+ \mathbb{E}\Bigg[\underbrace{\mathrm{Tr}\left(RB_{t-k_t}B_{t-k_t}^T(A_t-\Sigma)U(A_t-\Sigma)\right)}_{Q_{t,3}}\Bigg] + \mathbb{E}\Bigg[\underbrace{\mathrm{Tr}\left(RB_{t-k_t}B_{t-k_t}^T R^T(A_t-\Sigma)U(A_t-\Sigma)\right)}_{Q_{t,4}}\Bigg]
\end{aligned}$$

Let's consider each of the terms above. Using Von-Neumann's trace inequality and noting that $\left\|\mathbb{E}_\pi\left[(A_t-\Sigma)^2\right]\right\|_2 \leq \mathcal{V}$, we have

$$\begin{aligned}
\mathbb{E}\left[Q_{t,1}\right] &= \mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\mathbb{E}\left[(A_t-\Sigma)U(A_t-\Sigma)|s_1,s_2,\ldots,s_{t-k_t}\right]\right)\right] \\
&= \mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\mathbb{E}\left[(A_t-\Sigma)U(A_t-\Sigma)|s_{t-k_t}\right]\right)\right] \\
&\leq \left\|\mathbb{E}\left[(A_t-\Sigma)U(A_t-\Sigma)|s_{t-k_t}\right]\right\|_2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\
&\leq \left\|U\right\|_2\left\|\mathbb{E}\left[(A_t-\Sigma)^2|s_{t-k_t}\right]\right\|_2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \text{ using } S.17 \\
&\leq \left\|U\right\|_2\left(\mathcal{V}+2d_{\text{mix}}\left(k_t\right)\mathcal{M}^2\right)\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\
&\leq \left\|U\right\|_2\left(\mathcal{V}+2\gamma\eta_t^2\mathcal{M}^2\right)\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[Q_{t,2}\right] &= \mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\mathbb{E}\left[R^T(A_t-\Sigma)U(A_t-\Sigma)|s_1,s_2,\ldots,s_{t-k_t}\right]\right)\right] \\
&\leq \left\|\mathbb{E}\left[R^T(A_t-\Sigma)U(A_t-\Sigma)|s_{t-k_t}\right]\right\|_2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\
&\leq \left(1+\epsilon\right)\eta_{t-k_t+1}k_t\mathcal{M}^2\left(\mathcal{M}+\lambda_1\right)\left\|U\right\|_2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \quad \text{using Lemma 2}
\end{aligned}$$

$$\mathbb{E}\left[Q_{t,3}\right] \leq \left(1+\epsilon\right)\eta_{t-k_t+1}k_t\mathcal{M}^2\left(\mathcal{M}+\lambda_1\right)\left\|U\right\|_2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \text{ using a similar argument as } Q_{t,2}$$

$$\begin{aligned}
\mathbb{E}\left[Q_{t,4}\right] &= \mathbb{E}\left[\mathrm{Tr}\left(RB_{t-k_t}B_{t-k_t}^T R^T(A_t-\Sigma)U(A_t-\Sigma)\right)\right] \\
&= \mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T R^T(A_t-\Sigma)U(A_t-\Sigma)R\right)\right] \\
&\leq r^2\left\|U\right\|_2\mathcal{M}^2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\
&\leq \left(1+\epsilon\right)^2\eta_{t-k_t+1}^2k_t^2\mathcal{M}^2\left(\mathcal{M}+\lambda_1\right)^2\left\|U\right\|_2\mathbb{E}\left[\mathrm{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \text{ using Lemma 2}
\end{aligned}$$

Therefore, we have

$$
\begin{aligned}
&\mathbb{E}\left[Q_t\right] \\
&\leq \left(\mathcal{V} + \eta_{t-k_t+1}\left(2\gamma\eta_t\mathcal{M}^2 + 2\left(1+\epsilon\right)k_t\mathcal{M}^2\left(\mathcal{M}+\lambda_1\right) + \left(1+\epsilon\right)^2\eta_{t-k_t+1}k_t^2\mathcal{M}^2\left(\mathcal{M}+\lambda_1\right)^2\right)\right)\left\|U\right\|_2 \times \\
&\qquad\qquad\qquad \mathbb{E}\left[\operatorname{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\
&\overset{(i)}{\leq} \left(\mathcal{V} + \eta_{t-k_t+1}\mathcal{M}^2\left(2\gamma\eta_t + 2\left(1+\epsilon\right)k_t\left(\mathcal{M}+\lambda_1\right) + 2\epsilon\left(1+\epsilon\right)^2k_t\left(\mathcal{M}+\lambda_1\right)\right)\right)\left\|U\right\|_2\mathbb{E}\left[\operatorname{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right] \\
&= \left(\mathcal{V} + \eta_{t-k_t+1}\mathcal{M}^2\left(2\gamma\eta_t + 2\left(1+\epsilon\right)\left(1+\epsilon\left(1+\epsilon\right)\right)k_t\left(\mathcal{M}+\lambda_1\right)\right)\right)\left\|U\right\|_2\mathbb{E}\left[\operatorname{Tr}\left(B_{t-k_t}B_{t-k_t}^T\right)\right]
\end{aligned}
$$

In (i), we used the slow-decay assumption on η_i mentioned in the lemma statement along with $\eta_i k_i\left(\mathcal{M}+\lambda_1\right) \leq \epsilon$. Hence proved. □

Lemma S.11. *(Learning Rate Schedule) Fix any $\delta \in (0,1)$. Set $k_i := \tau_{mix}\left(\eta_i^2\right)$. Suppose the step sizes are set such that*

$$
\eta_i = \frac{\alpha}{\left(\lambda_1-\lambda_2\right)\left(\beta+i\right)}
$$

Define the linear function

$$
\forall i \in [n],\ f\left(i\right) := \frac{1}{\eta_i} = \frac{\left(\lambda_1-\lambda_2\right)\left(\beta+i\right)}{\alpha},
$$

With $\epsilon := \frac{1}{100}$ and $\xi_{k,t}, \zeta_{k,t}, \mathcal{V}', \overline{\mathcal{V}_{k,t}}$ defined in S.38, set $\alpha > 2,\ f\left(0\right) \geq e,\ m := 200$ and

$$
\beta := 600\max\left\{\frac{\tau_{mix}\log\left(f\left(0\right)\right)\left(\mathcal{M}+\lambda_1\right)\alpha}{\lambda_1-\lambda_2},\ \frac{5\tau_{mix}\log\left(f\left(0\right)\right)\left(\mathcal{M}+\lambda_1\right)^2\alpha^2}{3\left(\lambda_1-\lambda_2\right)^2\log\left(1+\frac{\delta}{m}\right)}, \frac{\left(\mathcal{V}'+5\lambda_1^2\right)\alpha^2}{300\left(\lambda_1-\lambda_2\right)^2\log\left(1+\frac{\delta}{m}\right)}\right\}
$$

then we have

1. $\eta_i k_i\left(\mathcal{M}+\lambda_1\right) \leq \epsilon$
2. $\forall i,\ \eta_i \leq \eta_{i-k_i} \leq \left(1+2\epsilon\right)\eta_i \leq 2\eta_i$ *(slow-decay)*
3. $\sum\limits_{i=1}^{n}\left(\overline{\mathcal{V}_{k,i}} + \zeta_{k,i} + 4\lambda_1^2\right)\eta_i^2 \leq \log\left(1+\frac{\delta}{m}\right)$
4. $\sum\limits_{i=1}^{n}\left(\mathcal{V}'+\xi_{k,i}\right)\eta_{i-k_i}^2\exp\left(-\sum\limits_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right) \leq$

$$
\left(\frac{2\left(1+10\epsilon\right)\alpha^2}{2\alpha-1}\right)\frac{\mathcal{V}'}{\left(\lambda_1-\lambda_2\right)^2}\frac{1}{n} + \left(\frac{24\left(1+10\epsilon\right)\alpha^3}{\left(\alpha-1\right)}\right)\frac{\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2}{\left(\lambda_1-\lambda_2\right)^3}\frac{k_n^2}{n^2}
$$

Proof. We use the following inequalities -

$$
\sum_{j=i}^{t}\eta_j^2 \leq \frac{\alpha^2}{\left(\lambda_1-\lambda_2\right)^2\left(\beta+i-1\right)} \qquad \left(\text{Using } \frac{1}{x+1} \leq \sum_{i=1}^{\infty}\frac{1}{(x+i)^2} \leq \frac{1}{x}\right) \tag{S.19}
$$

$$
\sum_{j=i}^{t}\eta_j \geq \frac{\alpha}{\left(\lambda_1-\lambda_2\right)}\log\left(\frac{t+\beta+1}{i+\beta}\right) \tag{S.20}
$$

$$
\sum_{j=i}^{t}\eta_j \leq \frac{\alpha}{\left(\lambda_1-\lambda_2\right)}\log\left(\frac{t+\beta}{i+\beta-1}\right) \tag{S.21}
$$

$$
\sum_{j=i}^{t}(j+\beta)^{\ell} \leq \frac{\left(t+\beta+1\right)^{\ell+1}-\left(i+\beta\right)^{\ell+1}}{\ell+1} \leq \frac{\left(t+\beta+1\right)^{\ell+1}}{\ell+1}\ \forall\ \ell > 0 \tag{S.22}
$$

For the first result, we observe that $f(x) = \frac{\log(x)}{x}$ is a decreasing function of x for $x \geq e$. Using properties of the mixing time (see Section 2.1 in the manuscript), we have

$$k_i := \tau_{\text{mix}}\left(\eta_i^2\right) \leq \frac{2\tau_{\text{mix}}}{\log(2)} \log\left(\frac{1}{\eta_i^2}\right) = \frac{4\tau_{\text{mix}}}{\log(2)} \log\left(\frac{(\beta+i)(\lambda_1-\lambda_2)}{\alpha}\right) = \frac{4\tau_{\text{mix}}}{\log(2)} \log(f(i)) \tag{S.23}$$

for $\eta_i < 1$. For $i \geq 0$

$$f(i) \geq f(0) = \frac{\beta(\lambda_1-\lambda_2)}{\alpha} \geq e$$

Therefore,

$$\begin{aligned}
\eta_i k_i (\mathcal{M}+\lambda_1) &\leq \frac{4\tau_{\text{mix}}(\mathcal{M}+\lambda_1)}{\log(2)} \frac{\alpha}{(\beta+i)(\lambda_1-\lambda_2)} \log\left(\frac{(\beta+i)(\lambda_1-\lambda_2)}{\alpha}\right) \\
&= \frac{4\tau_{\text{mix}}(\mathcal{M}+\lambda_1)}{\log(2)} \frac{\log(f(i))}{f(i)} \\
&\leq \frac{4\tau_{\text{mix}}(\mathcal{M}+\lambda_1)}{\log(2)} \frac{\log(f(0))}{f(0)}
\end{aligned}$$

From the assumptions mentioned in the Lemma statement, we have

$$\frac{\log(f(0))}{f(0)} < \frac{\epsilon \log(2)}{4\tau_{\text{mix}}(\mathcal{M}+\lambda_1)} = \frac{\log(2)}{400\tau_{\text{mix}}(\mathcal{M}+\lambda_1)} \tag{S.24}$$

Therefore,

$$\forall\, i, \eta_i k_i (\mathcal{M}+\lambda_1) \leq \epsilon \tag{S.25}$$

For the second result, we note that $\forall i \in [n]$,

$$\begin{aligned}
\frac{\eta_{i-k_i}}{\eta_i} &= \frac{\beta+i}{\beta+i-k_i} \\
&= 1 + \frac{k_i}{\beta+i-k_i} \\
&= 1 + \frac{1}{\frac{\beta+i}{k_i}-1}
\end{aligned}$$

Consider the fraction $\frac{\beta+i}{k_i}$. We can simplify it as :

$$\begin{aligned}
\frac{\beta+i}{k_i} &\geq \frac{\log(2)}{4\tau_{\text{mix}}} \frac{\beta+i}{\log\left(\frac{(\beta+i)(\lambda_1-\lambda_2)}{\alpha}\right)} \\
&= \frac{\alpha\log(2)}{4\tau_{\text{mix}}(\lambda_1-\lambda_2)} \frac{f(i)}{\log(f(i))} \\
&\geq \frac{\alpha\log(2)}{4\tau_{\text{mix}}(\lambda_1-\lambda_2)} \frac{f(0)}{\log(f(0))} \\
&\geq \frac{1}{\epsilon} \text{ from } S.24
\end{aligned}$$

where we used the fact that $\frac{x}{\log(x)}$ is an increasing function for $x \geq e$. Therefore, we have that

$$\begin{aligned}
\frac{\eta_{i-k_i}}{\eta_i} &\leq 1 + \frac{1}{\frac{1}{\epsilon}-1} \\
&= \frac{1}{1-\epsilon} \\
&\leq 1 + 2\epsilon \text{ for } \epsilon \in (0, 0.1)
\end{aligned}$$

For the third result, we note that

$$
\begin{aligned}
\zeta_{k,t} &:= 40k_{t+1}\left(\mathcal{M}+\lambda_1\right)^2, \\
\xi_{k,t} &:= 2\eta_t\mathcal{M}\left[3+9k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^2\right] \\
&\leq 24\eta_t\mathcal{M}\left[k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^2\right] \text{ since } \left(\mathcal{M}+\lambda_1\right)\geq 1 \text{ WLOG} \\
&\leq 24\epsilon\left(1+\epsilon\right)k_{t+1}\left(\mathcal{M}+\lambda_1\right)^2 \text{ since } \eta_t \leq \left(1+2\epsilon\right)\eta_{t+1} \text{ and } \eta_{t+1}k_{t+1}\left(\mathcal{M}+\lambda_1\right)\leq\epsilon
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sum_{i=1}^{n}\left(\overline{\mathcal{V}_{k,i}}+\zeta_{k,i}\right)\eta_i^2 &= \left(\mathcal{V}'+5\lambda_1^2\right)\sum_{i=1}^{n}\eta_i^2+41\left(\mathcal{M}+\lambda_1\right)^2\sum_{i=1}^{n}\eta_i^2k_{i+1} \\
&\stackrel{(i)}{\leq} \left(\mathcal{V}'+5\lambda_1^2\right)\underbrace{\sum_{i=1}^{n}\eta_i^2}_{T_1}+45\left(\mathcal{M}+\lambda_1\right)^2\underbrace{\sum_{i=1}^{n}\eta_{i+1}^2k_{i+1}}_{T_2}
\end{aligned} \tag{S.26}
$$

where (i) follows from the slow decay property of η_i.

For T_1, using S.19 we have,

$$
T_1 \leq \frac{\alpha^2}{\left(\lambda_1-\lambda_2\right)^2\beta} \tag{S.27}
$$

For T_2, substituting the value of k_i from S.23 for $\eta_i < 1$ we have,

$$
T_2 := \sum_{i=1}^{n}\eta_{i+1}^2k_{i+1} \leq \frac{4\tau_{\text{mix}}}{\log\left(2\right)}\sum_{i=1}^{n}\left(\frac{\alpha}{\left(\lambda_1-\lambda_2\right)\left(\beta+i+1\right)}\right)^2\log\left(\frac{\left(\lambda_1-\lambda_2\right)\left(\beta+i+1\right)}{\alpha}\right) \tag{S.28}
$$

$$
= \frac{4\tau_{\text{mix}}}{\log\left(2\right)}\sum_{i=1}^{n}\frac{\log\left(f\left(i+1\right)\right)}{f\left(i+1\right)^2} \tag{S.29}
$$

Note that $f\left(i\right)$ is a linear function of i and $\forall i\ f\left(i+1\right)-f\left(i\right)=\frac{\lambda_1-\lambda_2}{\alpha}$. We observe that $g(x)=\frac{\log(x)}{x^2}$ is a decreasing function of x for $x\geq e^{\frac{1}{2}}\sim 1.65$. Therefore,

$$
\left(\frac{\lambda_1-\lambda_2}{\alpha}\right)\sum_{i=1}^{n}\frac{\log\left(f\left(i+1\right)\right)}{f\left(i+1\right)^2}\leq\int_{f(1)}^{f(n+1)}\frac{\log\left(x\right)}{x^2}\,dx
$$

Substituting in S.29 we have,

$$
\begin{aligned}
T_2 &\leq \frac{4\tau_{\text{mix}}}{\log\left(2\right)}\left(\frac{\alpha}{\lambda_1-\lambda_2}\right)\int_{f(1)}^{f(n+1)}\frac{\log\left(x\right)}{x^2}\,dx \\
&= \frac{4\tau_{\text{mix}}}{\log\left(2\right)}\left(\frac{\alpha}{\lambda_1-\lambda_2}\right)\left(-\left(\frac{\log\left(x\right)}{x}+\frac{1}{x}\right)\Bigg|_{f(1)}^{f(n)}\right) \\
&\leq \frac{4\tau_{\text{mix}}}{\log\left(2\right)}\left(\frac{\alpha}{\lambda_1-\lambda_2}\right)\left(\frac{\log\left(f\left(1\right)\right)}{f\left(1\right)}+\frac{1}{f\left(1\right)}\right) \\
&\leq \frac{8\tau_{\text{mix}}}{\log\left(2\right)}\left(\frac{\alpha}{\lambda_1-\lambda_2}\right)\left(\frac{\log\left(f\left(1\right)\right)}{f\left(1\right)}\right) \\
&\leq \frac{8\tau_{\text{mix}}}{\log\left(2\right)}\left(\frac{\alpha}{\lambda_1-\lambda_2}\right)\left(\frac{\log\left(f\left(0\right)\right)}{f\left(0\right)}\right) \text{ since } \frac{\log\left(x\right)}{x} \text{ is a decreasing function of } x \text{ for } x\geq e
\end{aligned}
$$

Putting everything together in S.26 and using the bounds on $\beta, f(0)$ mentioned in the lemma statement, we have,

$$\begin{aligned}
\sum_{i=1}^{n}\left(\overline{\mathcal{V}_{k,i}}+\zeta_{k,i}\right)\eta_i^2 &\leq 460\left(\mathcal{M}+\lambda_1\right)^2\tau_{\text{mix}}\left(\frac{\alpha}{\lambda_1-\lambda_2}\right)\frac{\log\left(f\left(0\right)\right)}{f\left(0\right)}+\frac{\alpha^2}{\left(\lambda_1-\lambda_2\right)^2\beta}\left(\mathcal{V}'+5\lambda_1^2\right)\\
&= 460\tau_{\text{mix}}\log\left(f\left(0\right)\right)\frac{\alpha^2}{\left(\lambda_1-\lambda_2\right)^2\beta}\left(\mathcal{M}+\lambda_1\right)^2+\frac{\alpha^2}{\left(\lambda_1-\lambda_2\right)^2\beta}\left(\mathcal{V}'+5\lambda_1^2\right)\\
&\leq \log\left(1+\frac{\delta}{m}\right)
\end{aligned}$$

Finally, for the last result we first note that

$$\begin{aligned}
\xi_{k,t} &:= 2\eta_t\mathcal{M}\left[3+9k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^2\right]\\
&\leq 24\eta_t\mathcal{M}\left[k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^2\right] \text{ since } \left(\mathcal{M}+\lambda_1\right)\geq 1 \text{ WLOG}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{i=1}^{n}\left(\mathcal{V}'+\xi_{k,i}\right)\eta_{i-k_i}^2\exp\left(-\sum_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right)\\
&\quad\leq\left(1+2\epsilon\right)^2\sum_{i=1}^{n}\left(\mathcal{V}'+\xi_{k,i}\right)\eta_i^2\exp\left(-\sum_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right)\\
&\quad\leq\left(1+5\epsilon\right)\sum_{i=1}^{n}\left(\mathcal{V}'+\xi_{k,i}\right)\eta_i^2\exp\left(-\sum_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right) \text{ since } \epsilon\in\left(0,0.1\right)\\
&\quad=\left(1+5\epsilon\right)\left[\sum_{i=1}^{n}\mathcal{V}'\eta_i^2\exp\left(-\sum_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right)+\sum_{i=1}^{n}\xi_{k,i}\eta_i^2\exp\left(-\sum_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right)\right]
\end{aligned} \tag{S.30}$$

Let's define

$$g\left(i\right):=\exp\left(-\sum_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right),\ \ T_3:=\sum_{i=1}^{n}\eta_i^2g\left(i\right),\ \ T_4:=\sum_{i=1}^{n}\eta_i^3g\left(i\right),\ \ T_5:=\sum_{i=1}^{n}\eta_i^3k_i^2g\left(i\right),$$

Note that since $k_n\geq k_i$,

$$T_5=\sum_{i=1}^{n}\eta_i^3k_i^2g\left(i\right)\leq k_n^2\sum_{i=1}^{n}\eta_i^3g\left(i\right)=k_n^2T_4$$

Then,

$$\begin{aligned}
\sum_{i=1}^{n}\left(\mathcal{V}'+\xi_{k,i}\right)\eta_{i-k_i}^2\exp\left(-\sum_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right) &\leq\left(1+5\epsilon\right)\left[\mathcal{V}'T_3+24\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2T_5\right]\\
&\leq\left(1+5\epsilon\right)\left[\mathcal{V}'T_3+24\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2k_n^2T_4\right]
\end{aligned} \tag{S.31}$$

Using S.20, $g(i) \le \left(\frac{i+\beta+1}{n+\beta+1}\right)^{2\alpha}$. Noting that $\left(\frac{\beta+1}{\beta}\right)^2 \le \left(\frac{\beta+1}{\beta}\right)^3 \le 2$, we have

$$\begin{aligned}
T_3 &:= \sum_{i=1}^{n} \eta_i^2 \exp\left(-2\sum_{j=i+1}^{n} \eta_j \left(\lambda_1 - \lambda_2\right)\right) \\
&= \left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \sum_{i=1}^{n} \frac{1}{\left(\beta+i\right)^2} \left(\frac{i+\beta+1}{n+\beta+1}\right)^{2\alpha} \\
&\le \left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \left(\frac{\beta+1}{\beta}\right)^2 \sum_{i=1}^{n} \frac{1}{\left(\beta+i+1\right)^2} \left(\frac{i+\beta+1}{n+\beta+1}\right)^{2\alpha} \\
&= \left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \left(\frac{\beta+1}{\beta}\right)^2 \sum_{i=1}^{n} \frac{1}{\left(\beta+i+1\right)^2} \left(\frac{i+\beta+1}{n+\beta+1}\right)^{2\alpha} \\
&\le 2\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \frac{1}{\left(n+\beta+1\right)^{2\alpha}} \sum_{i=1}^{n} \left(i+\beta+1\right)^{2\alpha-2} \\
&\le \frac{2}{2\alpha-1}\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \frac{1}{\left(n+\beta+2\right)} \left(\frac{n+\beta+2}{n+\beta+1}\right)^{2\alpha} \quad \text{using } S.22 \\
&= \frac{2}{2\alpha-1}\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \frac{1}{\left(n+\beta+2\right)} \left(1+\frac{1}{n+\beta+1}\right)^{2\alpha}
\end{aligned} \tag{S.32}$$

and similarly,

$$\begin{aligned}
T_4 &:= \sum_{i=1}^{n} \eta_i^3 \exp\left(-2\sum_{j=i+1}^{n} \eta_j \left(\lambda_1 - \lambda_2\right)\right) \\
&= \left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^3 \sum_{i=1}^{n} \frac{1}{\left(\beta+i\right)^3} \left(\frac{i+\beta+1}{n+\beta+1}\right)^{2\alpha} \\
&\le \left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^3 \left(\frac{\beta+1}{\beta}\right)^3 \sum_{i=1}^{n} \frac{1}{\left(\beta+i+1\right)^3} \left(\frac{i+\beta+1}{n+\beta+1}\right)^{2\alpha} \\
&= \left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^3 \left(\frac{\beta+1}{\beta}\right)^3 \sum_{i=1}^{n} \frac{1}{\left(\beta+i+1\right)^2} \left(\frac{i+\beta+1}{n+\beta+1}\right)^{2\alpha} \\
&\le 2\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^3 \frac{1}{\left(n+\beta+1\right)^{2\alpha}} \sum_{i=1}^{n} \left(i+\beta+1\right)^{2\alpha-3}
\end{aligned}$$

$$\begin{aligned}
&\le \frac{1}{\alpha-1}\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^3 \frac{1}{\left(n+\beta+2\right)^2} \left(\frac{n+\beta+2}{n+\beta+1}\right)^{2\alpha} \quad \text{using } S.22 \\
&= \frac{1}{\alpha-1}\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^3 \frac{1}{\left(n+\beta+2\right)^2} \left(1+\frac{1}{n+\beta+1}\right)^{2\alpha}
\end{aligned} \tag{S.33}$$

Using S.25, we have

$$\frac{\alpha}{n+\beta+1} = \eta_n\left(\lambda_1 - \lambda_2\right) \le \eta_n \lambda_1 \le \eta_n k_n \lambda_1 \le \epsilon \le 0.1 \tag{S.34}$$

Therefore, using [2]

$$\left(1+\frac{1}{n+\beta+1}\right)^{2\alpha} \overset{(i)}{\le} \frac{1}{1-\frac{2\alpha}{n+\beta+1}} \overset{(ii)}{\le} 1+\frac{4\alpha}{n+\beta+1} \le 1+4\epsilon \tag{S.35}$$

where (i) follows since $\frac{2\alpha}{n+\beta+1} < 1$ by S.34 and (ii) follows since $\frac{1}{1-x} \le 1+2x$ for $x \in [0, \frac{1}{2}]$.

Using S.35 with S.32, we have

$$
\begin{aligned}
T_3 &\leq \frac{2}{2\alpha - 1}\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \frac{1}{(n+\beta+2)}\left(1 + \frac{4\alpha}{n+\beta+1}\right) \\
&\leq \frac{2(1+4\epsilon)}{2\alpha-1}\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \frac{1}{(n+\beta+2)}
\end{aligned}
\tag{S.36}
$$

Using S.35 with S.33, we have

$$
T_4 \leq \frac{1+4\epsilon}{\alpha-1}\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^3 \frac{1}{(n+\beta+2)^2}
\tag{S.37}
$$

Let

$$
C_1 := \frac{2(1+10\epsilon)\alpha^2}{2\alpha-1}, C_2 := \frac{24(1+10\epsilon)\alpha^3}{(\alpha-1)},
$$

Putting together S.36, S.37 in S.31 and using the definition of k_i in S.23 we have

$$
\begin{aligned}
(1+5\epsilon)\mathcal{V}'T_3 &\leq \frac{2(1+5\epsilon)(1+4\epsilon)}{2\alpha-1}\left(\frac{\alpha}{\lambda_1 - \lambda_2}\right)^2 \frac{\mathcal{V}'}{(n+\beta+2)} \\
&\leq \frac{2(1+10\epsilon)\alpha^2}{2\alpha-1}\frac{\mathcal{V}'}{(\lambda_1-\lambda_2)^2}\frac{1}{n} \text{ since } \epsilon \leq 0.05
\end{aligned}
$$

and similarly,

$$
24(1+5\epsilon)\mathcal{M}(\mathcal{M}+\lambda_1)^2 k_n^2 T_4 \leq \frac{24(1+5\epsilon)(1+4\epsilon)\alpha^3}{\alpha-1}\frac{\mathcal{M}(\mathcal{M}+\lambda_1)^2}{(\lambda_1-\lambda_2)^3}\frac{k_n^2}{n^2}
$$

Therefore from S.31, we have

$$
\sum_{i=1}^{n}(\mathcal{V}'+\xi_{k,i})\eta_{i-k_i}^2 \exp\left(-\sum_{j=i+1}^{n} 2\eta_j(\lambda_1-\lambda_2)\right) \leq C_1\frac{\mathcal{V}'}{(\lambda_1-\lambda_2)^2}\frac{1}{n} + C_2\frac{\mathcal{M}(\mathcal{M}+\lambda_1)^2}{(\lambda_1-\lambda_2)^3}\frac{k_n^2}{n^2}
$$

Hence proved. □

S.4 Proofs : Convergence Analysis of Oja's Algorithm for Markovian Data

In this section, we present proofs of Theorems 2, 3, 4 and 5. We state versions of these theorems that are valid under more general conditions on the step sizes. Specifically, for the following, we only require a sequence of non-increasing step-sizes which satisfy, for $\epsilon := \frac{1}{100}, \forall i \in [n]$ -

C.1 $\eta_i k_i(\mathcal{M}+\lambda_1) \leq \epsilon$ **C.2** (Slow decay) $\eta_i \leq \eta_{i-k_i} \leq (1+2\epsilon)\eta_i \leq 2\eta_i$

The version of these theorems stated in the main manuscript are obtained by plugging in the step-sizes as $\eta_i := \frac{\alpha}{(\lambda_1-\lambda_2)(\beta+i)}$ for the values of α, β provided in Lemma S.11. Before starting with the proofs, we define the following scalar variables -

$$
\begin{aligned}
r &:= 2(1+\epsilon)k_n\eta_n(\mathcal{M}+\lambda_1), & \zeta_{k,t} &:= 40k_{t+1}(\mathcal{M}+\lambda_1)^2 \\
\psi_{k,t} &:= 6\mathcal{M}\left[1+3k_{t+1}^2(\mathcal{M}+\lambda_1)^2\right], & \xi_{k,t} &:= \eta_{t-k_t}\psi_{k,t} \\
\mathcal{V}' &:= \frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\mathcal{V}, & \overline{\mathcal{V}_{k,t}} &:= \mathcal{V}'+\lambda_1^2+\xi_{k,t}
\end{aligned}
\tag{S.38}
$$

The basic idea behind these proofs is illustrated in Figure 2 in the main paper, where we are trying to approximate the matrix product by conditioning back in time just the right amount, to balance the tradeoff between the advantage of the mixing decay and the norm of the product of matrices.

Theorem 2. *(General Version) Under Assumptions 1, 2 and 3, for all $n > k_n$, and any decaying step-size schedule η_i satisfying **C.1** and **C.2**, we have:*

$$\mathbb{E}\left[v_1^T B_{n,1} B_{n,1}^T v_1\right] \leq (1+r)^2 \exp\left(\sum_{t=1}^{n-k_n}\left(2\eta_t\lambda_1 + \eta_t^2\left(\mathcal{V}' + \lambda_1^2 + \xi_{k,t}\right)\right)\right)$$

where $B_{j,i}$ is defined in 7.

Proof. Define $\alpha_{n,t} := \mathbb{E}\left[\operatorname{Tr}\left(v_1^T B_{n,t} B_{n,t}^T v_1\right)\right] = \mathbb{E}\left[v_1^T B_{n,t} B_{n,t}^T v_1\right], i \leq t \leq n$. Then, we have

$$\begin{aligned}
v_1^T B_{n,t} B_{n,t}^T v_1 &= v_1^T B_{n,t+1}(I+\eta_t\Sigma)^2 B_{n,t+1}^T v_1 \;+\; 2\eta_t \underbrace{\left(v_1^T B_{n,t+1}(I+\eta_t\Sigma)(A_t-\Sigma)B_{n,t+1}^T v_1\right)}_{P_{n,t}} \\
&\quad + \;\eta_t^2 \underbrace{\left(v_1^T B_{n,t+1}(A_t-\Sigma)^2 B_{n,t+1}^T v_1\right)}_{Q_{n,t}} \qquad \text{(S.39)} \\
&\leq v_1^T B_{j,t+1} B_{j,t+1}^T v_1((1+\eta_t\lambda_1)^2) + \eta_t^2 Q_{n,t} + 2\eta_t P_{n,t}
\end{aligned}$$

Using Lemma S.7 with $U = v_1, G = (I+\eta_t\Sigma), \gamma = 1$ and noting that $\mathbb{E}_\pi[A_t - \Sigma] = 0$, along with observing that $\alpha_{n,t+k_{t+1}} \leq \alpha_{n,t+k_t}$ from Lemma S.2, we have

$$|\mathbb{E}[P_{n,t}]| \leq \eta_{t+1}(1+\eta_t\lambda_1)\left(\frac{2\mathcal{V}|\lambda_2(P)|}{1-|\lambda_2(P)|} + \eta_{t+1}\mathcal{M}\left(2+16\epsilon+\left(2+(1+\epsilon)^2\right)k_{t+1}^2(\mathcal{M}+\lambda_1)^2\right)\right)\alpha_{n,t+k_t}$$

We note that $\forall i,\ k_i \geq 1$, therefore, using the assumption in S.38, $1+\eta_t\lambda_1 \leq 1+\eta_t k_t(\mathcal{M}+\lambda_1) \leq 1+\epsilon$.

Next, using Lemma S.8 with $U = v_1, G = I, \gamma = 1$ and noting that $\left\|\mathbb{E}_\pi\left[(A_t-\Sigma)^2\right]\right\|_2 \leq \mathcal{V}$ along with observing that $\alpha_{n,t+k_{t+1}} \leq \alpha_{n,t+k_t}$ using Lemma S.2, we have

$$\begin{aligned}
|\mathbb{E}[Q_{n,t}]| &\leq \left(\mathcal{V} + \eta_{t+1}\mathcal{M}^2\left(2\eta_{t+1} + (1+\epsilon)(2+\epsilon(1+\epsilon))k_{t+1}(\mathcal{M}+\lambda_1)\right)\right)\alpha_{n,t+k_t} \\
&\leq \left(\mathcal{V} + 2\epsilon\eta_{t+1}\mathcal{M} + \eta_{t+1}\mathcal{M}^2\left((1+\epsilon)(2+\epsilon(1+\epsilon))k_{t+1}(\mathcal{M}+\lambda_1)\right)\right)\alpha_{n,t+k_t}
\end{aligned}$$

where in the last line, we used $\eta_{t+1}\mathcal{M} \leq \eta_{t+1}(\mathcal{M}+\lambda_1) \leq \eta_{t+1}k_{t+1}(\mathcal{M}+\lambda_1) \leq \epsilon$.

Then from S.39 for $n - k_t \geq t \geq 1$,

$$\alpha_{n,t} \leq (1+\eta_t\lambda_1)^2\alpha_{n,t+1} + \left(\frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\right)\mathcal{V}\eta_t^2\alpha_{n,t+k_t} + C_{k,t}\eta_t^3\alpha_{n,t+k_t} \qquad \text{(S.40)}$$

where $C_{k,t}$ is defined as

$$\begin{aligned}
C_{k,t} &:= \mathcal{M}\left[4(1+\epsilon)(1+8\epsilon) + 2\epsilon + k_{t+1}(\mathcal{M}+\lambda_1)\left((1+\epsilon)(2+\epsilon(1+\epsilon))\mathcal{M} + 2\left(2+(1+\epsilon)^2\right)k_{t+1}(\mathcal{M}+\lambda_1)\right)\right] \\
&\overset{(i)}{\leq} \mathcal{M}\left[4(1+\epsilon)(1+8\epsilon) + 2\epsilon + \left((1+\epsilon)(2+\epsilon(1+\epsilon)) + 2\left(2+(1+\epsilon)^2\right)\right)k_{t+1}^2(\mathcal{M}+\lambda_1)^2\right] \\
&= \mathcal{M}\left[4 + 38\epsilon + 32\epsilon^2 + \left(6 + 2\epsilon + (1+\epsilon)^2(1+2\epsilon)\right)k_{t+1}^2(\mathcal{M}+\lambda_1)^2\right]
\end{aligned}$$

where in (i) we used $\mathcal{M} \leq k_{t+1}(\mathcal{M}+\lambda_1)$.

Then recalling the definition of $\xi_{k,t}$ in S.38, and noting that $\alpha_{n,t+k_t} \leq \alpha_{n,t+1}$ using Lemma S.2 we have from S.40,

$$\begin{aligned}
\alpha_{n,t} &\leq (1+\eta_t\lambda_1)^2\alpha_{n,t+1} + \left(\left(\frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\right)\mathcal{V} + \xi_{k,t}\right)\eta_t^2\alpha_{n,t+k_t} \\
&= \left(1 + 2\eta_t\lambda_1 + \eta_t^2\left(\left(\frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\right)\mathcal{V} + \lambda_1^2 + \xi_{k,t}\right)\right)\alpha_{n,t+1}
\end{aligned}$$

Therefore using this recursion, we have,

$$\alpha_{n,1} \leq \alpha_{n,n-k_n+1} \exp\left(2\lambda_1 \sum_{t=1}^{n-k_n} \eta_t + \sum_{t=1}^{n-k_n} \eta_t^2 \left(\left(\frac{1+(3+4\epsilon)\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|}\right)\mathcal{V} + \lambda_1^2 + \xi_{k,t}\right)\right)$$

Let $B_{n,n-k_n+1} = I + R'$, where $\|R'\| \leq r$ a.s.

$$\begin{aligned}
\alpha_{n,n-k_n+1} &= \mathbb{E}\left[v_1^T B_{n,n-k_n+1} B_{n,n-k_n+1}^T v_1\right] \\
&= \mathbb{E}\left[v_1^T v_1\right] + \mathbb{E}\left[v_1^T (R' + R'^T) v_1\right] + \mathbb{E}\left[v_1^T R' R'^T v_1\right] \\
&\leq 1 + 2r + r^2
\end{aligned}$$

Using Lemma 2 we have

$$\begin{aligned}
r &\leq (1+\epsilon)\, k_n \eta_{n-k_n+1} \left(\mathcal{M} + \lambda_1\right) \\
&\leq (1+\epsilon)\, k_n \eta_{n-k_n} \left(\mathcal{M} + \lambda_1\right) \\
&\leq 2\,(1+\epsilon)\, k_n \eta_n \left(\mathcal{M} + \lambda_1\right) \quad \text{since } \eta_{n-k_n} \leq 2\eta_n
\end{aligned}$$

Therefore,

$$\alpha_{n,1} \leq \left(1 + 2r + r^2\right) \exp\left(2\lambda_1 \sum_{t=1}^{n-k_n} \eta_t + \sum_{t=1}^{n-k_n} \eta_t^2 \left(\left(\frac{1+(3+4\epsilon)\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|}\right)\mathcal{V} + \lambda_1^2 + \xi_{k,t}\right)\right)$$

Hence proved. $\square$

Theorem 3. *(General Version) Let $u := \min\{t : t \in [n], t - k_t \geq 0\}$. Under Assumptions 1, 2 and 3, for all $n > u$, and any decaying step-size η_i satisfying **C.1** and **C.2**, we have,*

$$\begin{aligned}
\mathbb{E}\left[\operatorname{Tr}\left(V_\perp^T B_n B_n^T V_\perp\right)\right] &\leq (1+5\epsilon) \exp\left(\sum_{i=u+1}^{n} 2\eta_i \lambda_2 + \left(\mathcal{V}' + \lambda_1^2 + \xi_{k,i}\right)\eta_{i-k_i}^2\right) \\
&\quad \times \left(d + \sum_{i=u+1}^{n} \left(\mathcal{V}' + \xi_{k,i}\right) C'_{k,i} \eta_{i-k_i}^2 \exp\left(\sum_{j=u+1}^{i} 2\eta_j \left(\lambda_1 - \lambda_2\right)\right)\right)
\end{aligned}$$

where $C'_{k,t} := \exp\left(2\lambda_1 \sum_{j=1}^{u} \left(\eta_j - \eta_{t-u+j}\right) + \sum_{j=1}^{t-u} \eta_j^2 \left(\overline{\mathcal{V}_{k,j}} - \overline{\mathcal{V}_{k,j+u}}\right)\right)$ and B_t is defined in 2.

Proof. For $t \leq n$, let

$$\begin{aligned}
\alpha_t &:= \alpha_{t,1} = \mathbb{E}\left[v_1^T B_t B_t^T v_1\right] = \mathbb{E}\left[\operatorname{Tr}\left(v_1^T B_t B_t^T v_1\right)\right], \text{ as defined in Theorem 2} \\
\beta_t &:= \mathbb{E}\left[\operatorname{Tr}\left(V_\perp^T B_t B_t^T V_\perp\right)\right]
\end{aligned}$$

Note that $\alpha_t + \beta_t = \operatorname{Tr}\left(B_t B_t^T\right)$ by definition. Then,

$$\begin{aligned}
\operatorname{Tr}\left(B_t B_t^T V_\perp V_\perp^T\right) &= \operatorname{Tr}\left(B_{t-1} B_{t-1}^T (I + \eta_t \Sigma) V_\perp V_\perp^T (I + \eta_t \Sigma)\right) + \eta_t \operatorname{Tr}\left(B_{t-1}^T (I + \eta_t \Sigma) V_\perp V_\perp^T (A_t - \Sigma) B_{t-1}\right) \\
&\quad + \eta_t \operatorname{Tr}\left(B_{t-1}^T (A_t - \Sigma) V_\perp V_\perp^T (I + \eta_t \Sigma) B_{t-1}\right) + \eta_t^2 \operatorname{Tr}\left(B_{t-1} B_{t-1}^T (A_t - \Sigma) V_\perp V_\perp^T (A_t - \Sigma)\right) \\
&\leq (1 + \eta_t \lambda_2)^2 \operatorname{Tr}\left(B_{t-1} B_{t-1}^T V_\perp V_\perp^T\right) + 2\eta_t \underbrace{\operatorname{Tr}\left(B_{t-1} B_{t-1}^T (I + \eta_t \Sigma) V_\perp V_\perp^T (A_t - \Sigma)\right)}_{P_t} \\
&\quad + \eta_t^2 \underbrace{\operatorname{Tr}\left(B_{t-1} B_{t-1}^T (A_t - \Sigma) V_\perp V_\perp^T (A_t - \Sigma)\right)}_{Q_t}
\end{aligned}$$

Let $B_{t-1} = (I + R)\, B_{t-k_t}$ with $\|R\|_2 \leq r$. Using Lemma S.9 with $G = (I + \eta_t \Sigma)\, V_\perp V_\perp^T = V_\perp (I + \eta_t \Lambda_\perp) V_\perp^T, \gamma = 1$, where $\Lambda_\perp$ is a $d-1 \times d-1$ diagonal matrix of eigenvalues $\lambda_2, \ldots, \lambda_d$ of Σ, and noting that $\left\|V_\perp V_\perp^T\right\|_2 = 1$,

$$\begin{aligned}
\mathbb{E}\left[P_t\right] &\leq (1 + \eta_t \lambda_1)\, \eta_{t-k_t} \left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|} + \eta_{t-k_t} \mathcal{M}\left(2\,(1+8\epsilon) + \left(2 + (1+\epsilon)^2\right) k_t^2 \left(\mathcal{M} + \lambda_1\right)^2\right)\right) \left(\alpha_{t-k_t} + \beta_{t-k_t}\right) \\
&\leq (1+\epsilon)\, \eta_{t-k_t} \left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1-\left|\lambda_2\left(P\right)\right|} + \eta_{t-k_t} \mathcal{M}\left(2\,(1+8\epsilon) + \left(2 + (1+\epsilon)^2\right) k_t^2 \left(\mathcal{M} + \lambda_1\right)^2\right)\right) \left(\alpha_{t-k_t} + \beta_{t-k_t}\right)
\end{aligned}$$

where in the last line, we used $\eta_t \lambda_1 \leq \eta_t k_t (\mathcal{M}+\lambda_1) \leq \epsilon$.

Using Lemma S.10 with $U = V_\perp V_\perp^T, \gamma = 1$,

$$\begin{aligned}
\mathbb{E}[Q_t] &\leq \left(\mathcal{V} + \eta_{t-k_t+1}\mathcal{M}^2 \left(2\eta_t + 2(1+\epsilon)(1+\epsilon(1+\epsilon))k_t(\mathcal{M}+\lambda_1)\right)\right)(\alpha_{t-k_t} + \beta_{t-k_t}) \\
&\overset{(i)}{\leq} \left(\mathcal{V} + 2\epsilon\eta_t\mathcal{M} + 2\eta_{t-k_t+1}\mathcal{M}^2 \left((1+\epsilon)(1+\epsilon(1+\epsilon))k_t(\mathcal{M}+\lambda_1)\right)\right)(\alpha_{t-k_t} + \beta_{t-k_t}) \\
&\overset{(ii)}{\leq} \left(\mathcal{V} + 2\epsilon\eta_t\mathcal{M} + 2\eta_{t-k_t+1}\mathcal{M}\left((1+\epsilon)(1+\epsilon(1+\epsilon))k_t^2(\mathcal{M}+\lambda_1)^2\right)\right)(\alpha_{t-k_t} + \beta_{t-k_t})
\end{aligned}$$

where in (i) we used $\forall i,\ \eta_i\mathcal{M} \leq \eta_i k_i(\mathcal{M}+\lambda_1) \leq \epsilon$ and in (ii) we used $\mathcal{M} \leq k_t(\mathcal{M}+\lambda_1)$. Putting everything together, we have,

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{Tr}\left(B_t B_t^T V_\perp V_\perp^T\right)\right] \\
&\quad\leq (1+\eta_t\lambda_2)^2 \beta_{t-1} \\
&\qquad + 2(1+\epsilon)\eta_t\eta_{t-k_t}\left(\frac{2\mathcal{V}|\lambda_2(P)|}{1-|\lambda_2(P)|} + \eta_{t-k_t}\mathcal{M}\left(2(1+8\epsilon) + \left(2+(1+\epsilon)^2\right)k_t^2(\mathcal{M}+\lambda_1)^2\right)\right)(\alpha_{t-k_t}+\beta_{t-k_t}) \\
&\qquad + \eta_t^2\left(\mathcal{V} + 2\epsilon\eta_t\mathcal{M} + 2\eta_{t-k_t+1}\mathcal{M}\left((1+\epsilon)(1+\epsilon(1+\epsilon))k_t^2(\mathcal{M}+\lambda_1)^2\right)\right)(\alpha_{t-k_t}+\beta_{t-k_t}) \\
&\quad\leq (1+\eta_t\lambda_2)^2 \beta_{t-1} \\
&\qquad + 2(1+\epsilon)\eta_{t-k_t}^2\left(\frac{2\mathcal{V}|\lambda_2(P)|}{1-|\lambda_2(P)|} + \eta_{t-k_t}\mathcal{M}\left(2(1+8\epsilon) + \left(2+(1+\epsilon)^2\right)k_t^2(\mathcal{M}+\lambda_1)^2\right)\right)(\alpha_{t-k_t}+\beta_{t-k_t}) \\
&\qquad + \eta_{t-k_t}^2\left(\mathcal{V} + 2\epsilon\eta_t\mathcal{M} + 2\eta_{t-k_t+1}\mathcal{M}\left((1+\epsilon)(1+\epsilon(1+\epsilon))k_t^2(\mathcal{M}+\lambda_1)^2\right)\right)(\alpha_{t-k_t}+\beta_{t-k_t}) \\
&\quad\leq (1+\eta_t\lambda_2)^2 \beta_{t-1} + \eta_{t-k_t}^2\left(\left(\frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\right)\mathcal{V} + \xi_{k,t}\right)(\alpha_{t-k_t}+\beta_{t-k_t})
\end{aligned}$$

where $\xi_{k,t}$ is as defined in S.38. Therefore using Lemma S.3,

$$\begin{aligned}
\mathbb{E}\left[\operatorname{Tr}\left(B_t B_t^T V_\perp V_\perp^T\right)\right] \leq &\left(1 + 2\eta_t\lambda_2 + \eta_{t-k_t}^2\left(\left(\frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\right)\mathcal{V} + \lambda_2^2 + \xi_{k,t}\right)\right)\beta_{t-1} \\
&+ \eta_{t-k_t}^2\left(\left(\frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\right)\mathcal{V} + \xi_{k,t}\right)\alpha_{t-1}
\end{aligned} \tag{S.41}$$

Let $\chi_\epsilon := 1 + 4\epsilon(1+\epsilon)\left(1+\epsilon+\epsilon^2\right) \leq 1.05$. From Theorem 2 denoting

$$r_{k,t} := 1 + 4(1+\epsilon)\eta_{t-1}k_{t-1}(\mathcal{M}+\lambda_1) + 4(1+c)^2\eta_{t-1}^2 k_{t-1}^2(\mathcal{M}+\lambda_1)^2 \leq 1 + 4\epsilon(1+\epsilon)\left(1+\epsilon+\epsilon^2\right) = \chi_\epsilon, \tag{S.42}$$

we have,

$$\alpha_{t-1} \leq r_{k,t}\exp\left(2\lambda_1\sum_{i=1}^{t-k_t-1}\eta_t + \sum_{i=1}^{t-k_t-1}\eta_i^2\left(\left(\frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\right)\mathcal{V} + \lambda_1^2 + \xi_{k,i}\right)\right)$$

Now, we note the definition of $\overline{\mathcal{V}_{k,t}}$ and $\mathcal{V}'$ as mentioned in S.38 -

$$\begin{aligned}
\overline{\mathcal{V}_{k,t}} &:= \left(\frac{1+(3+4\epsilon)|\lambda_2(P)|}{1-|\lambda_2(P)|}\right)\mathcal{V} + \lambda_1^2 + \xi_{k,t} \\
&= \mathcal{V}' + \lambda_1^2 + \xi_{k,t}
\end{aligned}$$

Therefore using S.41,

$$\beta_t \leq \left(1 + 2\eta_t\lambda_2 + \eta_{t-k_t}^2\overline{\mathcal{V}_{k,t}}\right)\beta_{t-1} + \eta_{t-k_t}^2 r_{k,t}(\mathcal{V}' + \xi_{k,t})\exp\left(2\lambda_1\sum_{i=1}^{t-k_t-1}\eta_i + \sum_{i=1}^{t-k_t-1}\eta_i^2\overline{\mathcal{V}_{k,i}}\right)$$

Recursing on the above inequality for $u < t \leq n$ where $u = \min\{i : i \in [n], i - k_i \geq 0\}$, we have,

$$\beta_n \leq \beta_u \exp\left(2\sum_{i=u+1}^{n} \eta_i \lambda_2 + \sum_{i=u+1}^{n} \overline{\mathcal{V}_{k,i}} \eta_{i-k_i}^2\right)$$
$$+ \sum_{i=u+1}^{n} r_{k,i}\left(\mathcal{V}' + \xi_{k,i}\right)\eta_{i-k_i}^2 \exp\left(\sum_{j=i+1}^{n}\left(2\eta_j\lambda_2 + \overline{\mathcal{V}_{k,j}}\eta_{j-k_j}^2\right)\right)\exp\left(\sum_{j=1}^{i-k_i} 2\eta_j\lambda_1 + \overline{\mathcal{V}_{k,j}}\eta_j^2\right)$$
$$\leq \exp\left(\sum_{i=u+1}^{n} 2\eta_i\lambda_2 + \overline{\mathcal{V}_{k,i}}\eta_{i-k_i}^2\right)$$
$$\times\left(\beta_u + \sum_{i=u+1}^{n} r_{k,i}\left(\mathcal{V}' + \xi_{k,i}\right)\eta_{i-k_i}^2 \exp\left(\sum_{j=1}^{i-k_i}(2\eta_j\lambda_1 + \overline{\mathcal{V}_{k,j}}\eta_j^2) - \sum_{j=u+1}^{i}\left(2\eta_j\lambda_2 + \overline{\mathcal{V}_{k,j}}\eta_{j-k_j}^2\right)\right)\right)$$

Now, since $k_i, k_j \geq k_u = u$, therefore, we have

$$\beta_n \leq \exp\left(\sum_{i=u+1}^{n} 2\eta_i\lambda_2 + \overline{\mathcal{V}_{k,i}}\eta_{i-k_i}^2\right)\times$$
$$\left(\beta_u + \sum_{i=u+1}^{n} r_{k,i}\left(\mathcal{V}' + \xi_{k,i}\right)\eta_{i-k_i}^2 \exp\left(\sum_{j=1}^{i-u}(2\eta_j\lambda_1 + \overline{\mathcal{V}_{k,j}}\eta_j^2) - \sum_{j=u+1}^{i}\left(2\eta_j\lambda_2 + \overline{\mathcal{V}_{k,j}}\eta_{j-u}^2\right)\right)\right)$$

Recall that $C'_{k,i} := \exp\left(2\lambda_1\sum_{j=1}^{u}\left(\eta_j - \eta_{i-u+j}\right) + \sum_{j=1}^{i-u}\eta_j^2\left(\overline{\mathcal{V}_{k,j}} - \overline{\mathcal{V}_{k,j+u}}\right)\right)$ as defined in S.38. Therefore,

$$\beta_n \leq \exp\left(\sum_{i=u+1}^{n} 2\eta_i\lambda_2 + \overline{\mathcal{V}_{k,i}}\eta_{i-k_i}^2\right)\times$$
$$\left(\beta_u + \sum_{i=u+1}^{n} r_{k,i}\left(\mathcal{V}' + \xi_{k,i}\right)C'_{k,i}\eta_{i-k_i}^2 \exp\left(\sum_{j=u+1}^{i} 2\eta_j\left(\lambda_1 - \lambda_2\right)\right)\right)$$

Let $B_u = I + R'$ with $\|R'\| \leq r'$ a.s. Using Lemma 2 we have

$$\begin{aligned}
r' &\leq (1+\epsilon)\, k_u\eta_1\left(\mathcal{M} + \lambda_1\right)\\
&\leq (1+\epsilon)\, k_u\eta_0\left(\mathcal{M} + \lambda_1\right)\\
&\leq 2(1+\epsilon)\, k_u\eta_u\left(\mathcal{M} + \lambda_1\right) \quad \text{since } \eta_0 = \eta_{u-k_u} \leq 2\eta_u\\
&< 2\epsilon(1+\epsilon)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\beta_u &= \mathbb{E}\left[\mathrm{Tr}\left(V_\perp^T B_u B_u^T V_\perp\right)\right]\\
&= \mathbb{E}\left[\mathrm{Tr}\left(V_\perp^T V_\perp\right)\right] + \mathbb{E}\left[\mathrm{Tr}\left(V_\perp^T (R' + R'^T) V_\perp\right)\right] + \mathbb{E}\left[\mathrm{Tr}\left(V_\perp^T R' R'^T V_\perp\right)\right]\\
&\leq d\left(1 + 2r' + r'^2\right)\\
&\leq d\left(1 + 4\epsilon(1+\epsilon) + 4\epsilon^2(1+\epsilon)^2\right)\\
&= d\left(1 + 4\epsilon(1+\epsilon)\left(1 + \epsilon + \epsilon^2\right)\right)\\
&= \chi_\epsilon d
\end{aligned}$$

The proof follows by noting that $r_{k,t} \leq \chi_\epsilon$ as shown in S.42. □

Theorem 4. *(General Version) Under Assumptions 1, 2 and 3, for all $n > k_n$, any decaying step-size η_i satisfying **C.1** and **C.2**, we have:*

$$\mathbb{E}\left[v_1^T B_{n,1} B_{n,1}^T v_1\right] \geq (1-t)\exp\left(\sum_{i=1}^{n-k_n} 2\eta_i\lambda_1 - \sum_{i=1}^{n-k_n} 4\eta_i^2\lambda_1^2\right)$$

where $t := 2r + s, s := 3\left(1+r\right)^2 \exp\left(2\lambda_1^2 \sum_{i=1}^{n} \eta_i^2\right) \sum_{t=1}^{n-k_n} W_{k,t}\eta_t^2 \exp\left(\sum_{i=t+1}^{n-k_n} \eta_i^2\right), W_{k,t} := \mathcal{V}' + \xi_{k,t}$ *and* $B_{j,i}$ *has been defined in 7.*

Proof. We will start will expanding the quantity of interest using Eq S.39.

$$\alpha_{n,t} = \mathbb{E}\left[v_1^T B_{n,t} B_{n,t}^T v_1\right] \geq \mathbb{E}\left[v_1^T B_{n,t+1}\left(I + \eta_t \Sigma\right)^2 B_{n,t+1}^T v_1 + 2\eta_t P_{n,t}\right] \tag{S.43}$$

where $P_{n,t}$ has been defined in Theorem 2. Let's define

$$\mathcal{S}_t := \prod_{i=t}^{1}\left(I + \eta_i \Sigma\right) \prod_{i=1}^{t}\left(I + \eta_i \Sigma\right),\ \mathcal{S}_0 = I \ \text{ and}$$
$$\delta_{n,t} := \mathbb{E}\left[v_1^T B_{n,t+1} \mathcal{S}_t B_{n,t+1}^T v_1\right]$$

Note that $\delta_{n,0} = \alpha_{n,1}$. First we bound $\delta_{n,n-k_n}$. Let $B_{n,n-k_n} = I + R'$. By Lemma 2 along with the slow-decay assumption on the step-sizes, we know that $\|R'\|_2 \leq r := 2\left(1+\epsilon\right)\eta_n k_n\left(\mathcal{M} + \lambda_1\right)$ a.s. Then,

$$\delta_{n,n-k_n} - \prod_{i=1}^{n-k_n}\left(1 + \eta_i \lambda_1\right)^2 \geq -2\left|E[v_1^T R' \mathcal{S}_{n-k_n} v_1]\right| \geq -2r \prod_{i=1}^{n-k_n}\left(1 + \eta_i \lambda_1\right)^2$$

Therefore,

$$\begin{aligned} \delta_{n,n-k_n} &\geq \prod_{i=1}^{n-k_n}\left(1 + \eta_i \lambda_1\right)^2\left(1 - 2r\right) \\ &= \left(1 - 2r\right)\|\mathcal{S}_{n-k_n}\|_2 \end{aligned} \tag{S.44}$$

Now using S.43, we have

$$\delta_{n,t-1} \geq \delta_{n,t} + 2\eta_t \mathbb{E}\left[\underbrace{v_1^T B_{n,t+1}\left(I + \eta_t \Sigma\right)\mathcal{S}_{t-1}\left(A_t - \Sigma\right) B_{n,t+1}^T v_1}_{U_t}\right]$$

First, observe that $\mathcal{S}_{t-1} = U \Lambda U^T$, where U denotes a matrix of eigenvectors of Σ, and Λ is a PSD diagonal matrix. Since $I + \eta_t \Sigma = U \Lambda' U^T$ for some other PSD diagonal matrix Λ', the product will also be PSD.

By using Lemma S.7 with $U = v_1, G = \left(I + \eta_t \Sigma\right)\mathcal{S}_{t-1}, \gamma = 1$ and noting that $\mathbb{E}_\pi\left[A_t - \Sigma\right] = 0$, we have

$$\begin{aligned} \left|\mathbb{E}\left[U_t\right]\right| &\leq \left(1 + \eta_t \lambda_1\right)\eta_{t+1}\|\mathcal{S}_{t-1}\|_2\left(\frac{2\mathcal{V}\left|\lambda_2\left(P\right)\right|}{1 - \left|\lambda_2\left(P\right)\right|}\right. \\ &\quad \left. + \eta_{t+1}\mathcal{M}\left(2\left(1 + 8\epsilon\right) + \left(2 + \left(1+\epsilon\right)^2\right) k_{t+1}^2\left(\mathcal{M} + \lambda_1\right)^2\right)\right)\alpha_{n,t+k_{t+1}} \\ &\leq \left(1+\epsilon\right)\eta_{t+1}\|\mathcal{S}_{t-1}\|_2 W_{k,t}\alpha_{n,t+1} \end{aligned}$$

where $W_{k,t} = \mathcal{V}' + \xi_{k,t}$. Therefore,

$$\delta_{n,t-1} \geq \delta_{n,t} - 2\left(1+\epsilon\right) W_{k,t}\eta_t^2 \alpha_{n,t+1}\|\mathcal{S}_{t-1}\|_2 \ \text{ for } t \leq n - k_n$$

Let

$$\mathcal{V}' := \left(\frac{1 + \left(3 + 4\epsilon\right)\left|\lambda_2\left(P\right)\right|}{1 - \left|\lambda_2\left(P\right)\right|}\right)\mathcal{V}$$

as defined in S.38. Unwinding the recursion for $t \leq n - k_n$, we have,

$$\begin{aligned} \delta_{n,0} &\geq \delta_{n,n-k_n} - 2\left(1+\epsilon\right)\sum_{t=1}^{n-k_n} W_{k,t}\eta_t^2 \alpha_{n,t+1}\|\mathcal{S}_{t-1}\|_2 \\ &\geq \left(1 - 2r\right)\|\mathcal{S}_{n-k_n}\|_2 \\ &\quad - 2\left(1+\epsilon\right)\left(1+r\right)^2 \sum_{t=1}^{n-k_n} W_{k,t}\eta_t^2 \exp\left(2\lambda_1 \sum_{i=t+1}^{n-k_n} \eta_i + \sum_{i=t+1}^{n-k_n} \eta_i^2\left(\mathcal{V}' + \lambda_1^2 + C_{k,i}\right)\right)\|\mathcal{S}_{t-1}\|_2 \end{aligned}$$

where second step followed from Theorem 2 and S.44.

Using the inequalities $\forall x \in \mathbb{R},\ 1+x \le e^x$ and $\forall x \in \mathbb{R},\ x \ge 0, 1+x \ge e^{x-x^2}, \quad \forall t$ we have,

$$\|\mathcal{S}_t\|_2 = \prod_{i=1}^{t}(1+\eta_i\lambda_1)^2 \le \exp\left(2\lambda_1\sum_{i=1}^{t}\eta_i\right), \text{ and}$$
$$\|\mathcal{S}_t\|_2 = \prod_{i=1}^{t}(1+\eta_i\lambda_1)^2 \ge \exp\left(2\lambda_1\sum_{i=1}^{t}\eta_i - 4\lambda_1^2\sum_{i=1}^{t}\eta_i^2\right)$$

Therefore denoting $\theta_\epsilon := 2(1+\epsilon)\exp\left(2\lambda_1^2\sum_{i=1}^{n}\eta_i^2\right)$, we have

$$\begin{aligned}
&\delta_{n,0}\\
&\ge \exp\left(2\lambda_1\sum_{i=1}^{n-k_n}\eta_i - 4\lambda_1^2\sum_{i=1}^{n-k_n}\eta_i^2\right)\left[(1-2r) - \theta_\epsilon(1+r)^2\sum_{t=1}^{n-k_n}W_{k,t}\eta_t^2\exp\left(\sum_{i=t+1}^{n-k_n}\eta_i^2\left(\mathcal{V}'+\lambda_1^2+C_{k,i}\right)\right)\right]\\
&\ge \exp\left(2\lambda_1\sum_{i=1}^{n-k_n}\eta_i - 4\lambda_1^2\sum_{i=1}^{n-k_n}\eta_i^2\right)\left[(1-2r) - \theta_\epsilon(1+r)^2\sum_{t=1}^{n-k_n}W_{k,t}\eta_t^2\exp\left(\sum_{i=t+1}^{n-k_n}\eta_i^2\left(\mathcal{V}'+\lambda_1^2+C_{k,i}\right)\right)\right]\\
&\ge \exp\left(2\lambda_1\sum_{i=1}^{n-k_n}\eta_i - 4\lambda_1^2\sum_{i=1}^{n-k_n}\eta_i^2\right)\left[1-\left(2r + \theta_\epsilon(1+r)^2\sum_{t=1}^{n-k_n}W_{k,t}\eta_t^2\exp\left(\sum_{i=t+1}^{n-k_n}\eta_i^2\overline{\mathcal{V}_{k,i}}\right)\right)\right]
\end{aligned}$$

where $\overline{\mathcal{V}_{k,i}}$ is defined in S.38. Hence proved. ☐

Theorem 5. *(General Version) Under Assumptions 1, 2 and 3, for all $n > k_n$, and decaying step-size η_i satisfying **C.1** and **C.2**, we have:*

$$\mathbb{E}\left[\left(v_1^T B_{n,1}B_{n,1}^T v_1\right)^2\right] \le (1+r)^4\exp\left(\sum_{i=1}^{n-k_n}4\eta_i\lambda_1 + \sum_{i=1}^{n-k_n}\eta_i^2\zeta_{k,i}\right)$$

where $B_{j,i}$ has been defined in 7.

Proof. Define $Q_{n,t} := v_1^T B_{n,t+1}(A_t-\Sigma)^2 B_{n,t+1}^T v_1$, and $P_{n,t} := v_1^T B_{n,t+1}(I+\eta_t\Sigma)(A_t-\Sigma)B_{n,t+1}^T v_1$. Using S.39, we have, for $n \ge t \ge 1$,

$$\begin{aligned}
0 \le v_1^T B_{n,t}B_{n,t}^T v_1 &= v_1^T B_{n,t+1}(I+\eta_t\Sigma)^2 B_{n,t+1}^T v_1 + \eta_t^2 Q_{n,t} + 2\eta_t P_{n,t}\\
&\le v_1^T B_{j,t+1}B_{j,t+1}^T v_1(1+\eta_t\lambda_1)^2 + \eta_t^2\mathcal{M}^2\left(v_1^T B_{n,t+1}B_{n,t+1}^T v_1\right) + 2\eta_t P_{n,t}\\
&\le v_1^T B_{j,t+1}B_{j,t+1}^T v_1\underbrace{\left((1+\eta_t\lambda_1)^2+\eta_t^2\mathcal{M}^2\right)}_{c_t} + 2\eta_t P_{n,t}
\end{aligned}$$

Thus, we have -

$$\begin{aligned}
\kappa_{n,t} := \mathbb{E}\left[(v_1^T B_{n,t}B_{n,t}^T v_1)^2\right] &\le \mathbb{E}\left[\left(c_t v_1^T B_{n,t+1}B_{n,t+1}^T v_1 + 2\eta_t P_{n,t}\right)^2\right]\\
&\le c_t^2\kappa_{n,t+1} + 4\eta_t^2\mathbb{E}\left[P_{n,t}^2\right] + 4c_t\eta_t\mathbb{E}\left[\left(v_1^T B_{n,t+1}B_{n,t+1}^T v_1\right)P_{n,t}\right]
\end{aligned} \tag{S.45}$$

Note that,

$$\begin{aligned}
\mathbb{E}\left[P_{n,t}^2\right] &\le \mathbb{E}\left[\left(v_1^T B_{n,t+1}(I+\eta_t\Sigma)(A_t-\Sigma)B_{n,t+1}^T v_1\right)^2\right]\\
&\le (1+\eta_t\lambda_1)^2\mathcal{M}^2\mathbb{E}\left[\left(v_1^T B_{n,t+1}B_{n,t+1}^T v_1\right)^2\right]\\
&= (1+\eta_t\lambda_1)^2\mathcal{M}^2\kappa_{n,t+1}
\end{aligned}$$

Now we work on the cross-term. For the convenience of notation, let's denote $k := k_{t+1}$ unless otherwise specified. Let $B_{n,t+1} = B_{n,t+k}(I+R)$ with,

$$\|R\|_2 \le (1+c)\eta_{t+1}k(\mathcal{M}+\lambda_1) =: r_t \le \epsilon(1+\epsilon)$$

Using Lemma 2, we have

$$\underbrace{|v_1^T B_{n,t+1} B_{n,t+1}^T v_1 - v_1^T B_{n,t+k} B_{n,t+k}^T v_1|}_{Y_1} = |v_1^T B_{n,t+k}(R + R^T + RR^T) B_{n,t+k}^T v_1|$$
$$\leq |v_1^T B_{n,t+k} B_{n,t+k}^T v_1| \left(2r_t + r_t^2\right) \quad \text{(S.46)}$$

We will also bound

$$\underbrace{|v_1^T B_{n,t+1}(I + \eta_t \Sigma)(A_t - \Sigma) B_{n,t+1}^T v_1 - v_1^T B_{n,t+k}(I + \eta_t \Sigma)(A_t - \Sigma) B_{n,t+k}^T v_1|}_{Y_2}$$
$$= |v_1^T B_{n,t+k} R(I + \eta_t \Sigma)(A_t - \Sigma)(I + R^T) B_{n,t+k}^T v_1 + v_1^T B_{n,t+k}(I + \eta_t \Sigma)(A_t - \Sigma) R^T B_{n,t+k}^T v_1|$$
$$\leq \left(2r_t + r_t^2\right)(1 + \eta_t \lambda_1) \mathcal{M} |v_1^T B_{n,t+k} B_{n,t+k}^T v_1| \quad \text{(S.47)}$$

So, now we have:

$$\begin{aligned}
&\mathbb{E}\left[\left(v_1^T B_{n,t+1} B_{n,t+1}^T v_1 P_{n,t}\right)\right] \\
&= \mathbb{E}\left[(v_1^T B_{n,t+1} B_{n,t+1}^T v_1)(v_1^T B_{n,t+1}(I + \eta_t \Sigma)(A_t - \Sigma) B_{n,t+1}^T v_1)\right] \\
&= \mathbb{E}\left[(Y_1 + v_1^T B_{n,t+k} B_{n,t+k}^T v_1)(Y_2 + v_1^T B_{n,t+k}(I + \eta_t \Sigma)(A_t - \Sigma) B_{n,t+k}^T v_1)\right] \\
&= \underbrace{\mathbb{E}\left[Y_1 Y_2\right]}_{T_1} + \underbrace{\mathbb{E}\left[Y_1 v_1^T B_{n,t+k}(I + \eta_t \Sigma)(A_t - \Sigma) B_{n,t+k}^T v_1\right]}_{T_2} + \underbrace{\mathbb{E}\left[Y_2 v_1^T B_{n,t+k} B_{n,t+k}^T v_1\right]}_{T_3} \\
&\quad + \underbrace{\mathbb{E}\left[(v_1^T B_{n,t+k} B_{n,t+k}^T v_1)(v_1^T B_{n,t+k}(I + \eta_t \Sigma)(A_t - \Sigma) B_{n,t+k}^T v_1)\right]}_{T_4}
\end{aligned}$$

Lets start with the last term, T_4. Using Lemma S.2 we have,

$$\begin{aligned}
|T_4| &\leq \left|\mathbb{E}\left[(v_1^T B_{n,t+k} B_{n,t+k}^T v_1)(v_1^T B_{n,t+k}(I + \eta_t \Sigma) \mathbb{E}\left[(A_t - \Sigma)|s_{t+k}\right] B_{n,t+k}^T v_1)\right]\right| \\
&\leq 2(1 + \eta_t \lambda_1) \mathcal{M} d_{\text{mix}}(k) \, \kappa_{n,t+k} \\
&\leq 2\eta_{t+1}^2 (1 + \eta_t \lambda_1) \mathcal{M} \kappa_{n,t+k} \\
&\leq 2\eta_{t+1}^2 (1 + \eta_t \lambda_1) \mathcal{M} \kappa_{n,t+1}
\end{aligned}$$

Using Eqs S.46 and S.47 the first three terms can be bounded as:

$$\begin{aligned}
|T_1| \leq \mathbb{E}\left[|Y_1 Y_2|\right] &\leq \left(2r_t + r_t^2\right)^2 (1 + \eta_t \lambda_1) \mathcal{M} \kappa_{n,t+k} \\
&\leq \left(2r_t + r_t^2\right)^2 (1 + \eta_t \lambda_1) \mathcal{M} \kappa_{n,t+1} \text{ using Lemma } S.2 \\
&= (2 + r_t)^2 r_t^2 (1 + \eta_t \lambda_1) \mathcal{M} \kappa_{n,t+1} \\
&\leq (1 + \epsilon)^2 (2 + \epsilon(1 + \epsilon))^2 (1 + \eta_t \lambda_1) \eta_{t+1}^2 k_{t+1}^2 \mathcal{M} (\mathcal{M} + \lambda_1)^2 \kappa_{n,t+1} \\
&\leq (1 + \epsilon)^3 \left(2 + \epsilon + \epsilon^2\right)^2 \eta_{t+1}^2 k_{t+1}^2 \mathcal{M} (\mathcal{M} + \lambda_1)^2 \kappa_{n,t+1} \text{ since } \eta_t \lambda_1 \leq \epsilon
\end{aligned}$$

$$\begin{aligned}
|T_2| &\leq \mathbb{E}\left[|Y_1 v_1^T B_{n,t+k}(I + \eta_t \Sigma)(A_t - \Sigma) B_{n,t+k}^T v_1|\right] \\
&\leq (2 + r_t) r_t (1 + \eta_t \lambda_1) \mathcal{M} \kappa_{n,t+k} \\
&\leq (2 + r_t) r_t (1 + \eta_t \lambda_1) \mathcal{M} \kappa_{n,t+1} \text{ using Lemma } S.2 \\
&\leq \left(2 + \epsilon + \epsilon^2\right)(1 + \epsilon)(1 + \eta_t \lambda_1) \eta_{t+1} k_{t+1} (\mathcal{M} + \lambda_1) \mathcal{M} \kappa_{n,t+1} \\
&\leq (1 + \epsilon)^2 \left(2 + \epsilon + \epsilon^2\right) \eta_{t+1} k_{t+1} \mathcal{M} (\mathcal{M} + \lambda_1) \kappa_{n,t+1}
\end{aligned}$$

and similarly,

$$\begin{aligned}
|T_3| &\leq \mathbb{E}\left[Y_2 v_1^T B_{n,t+k} B_{n,t+k}^T v_1\right] \\
&\leq r_t (2 + r_t)(1 + \eta_t \lambda_1) \mathcal{M} \kappa_{n,t+k} \\
&\leq (1 + \epsilon)\left(2 + \epsilon + \epsilon^2\right)(1 + \eta_t \lambda_1) \eta_{t+1} k_{t+1} \mathcal{M} (\mathcal{M} + \lambda_1) \kappa_{n,t+k} \\
&\leq (1 + \epsilon)^2 \left(2 + \epsilon + \epsilon^2\right) \eta_{t+1} k_{t+1} \mathcal{M} (\mathcal{M} + \lambda_1) \kappa_{n,t+k} \\
&\leq (1 + \epsilon)^2 \left(2 + \epsilon + \epsilon^2\right) \eta_{t+1} k_{t+1} \mathcal{M} (\mathcal{M} + \lambda_1) \kappa_{n,t+1} \text{ using Lemma } S.2
\end{aligned}$$

Note that

$$
\begin{aligned}
c_t &:= (1+\eta_t\lambda_1)^2 + \eta_t^2\mathcal{M}^2 \le 1+2\epsilon+2\epsilon^2, \text{ and} \\
c_t^2 &= \left(1+2\eta_t\lambda_1+\eta_t^2\left(\mathcal{M}^2+\lambda_1^2\right)\right)^2 \\
&= 1+4\eta_t^2\lambda_1^2+\eta_t^4\left(\mathcal{M}^2+\lambda_1^2\right)^2+4\eta_t\lambda_1+4\eta_t^3\lambda_1\left(\mathcal{M}^2+\lambda_1^2\right)+2\eta_t^2\left(\mathcal{M}^2+\lambda_1^2\right) \\
&\le 1+4\eta_t\lambda_1+\eta_t^2\left(2\mathcal{M}^2+6\lambda_1^2+\epsilon\mathcal{M}+\epsilon\lambda_1\right)+4\eta_t^3\lambda_1\left(\mathcal{M}^2+\lambda_1^2\right) \\
&\le 1+4\eta_t\lambda_1+6\eta_t^2\left(\mathcal{M}+\lambda_1\right)^2+4\eta_t^3\lambda_1\left(\mathcal{M}+\lambda_1\right)^2
\end{aligned}
$$

Define

$$
\begin{aligned}
\phi_\epsilon &:= (1+\epsilon)\left(2+\epsilon+\epsilon^2\right) \\
\omega_\epsilon &:= 1+2\epsilon+2\epsilon^2 \\
\zeta_{k,t} &:= \left(10+8\left(1+\epsilon\right)+4\left(1+2\epsilon\right)\phi_\epsilon\right)\phi_\epsilon c_t k_{t+1}\left(\mathcal{M}+\lambda_1\right)^2
\end{aligned}
$$

Putting everything together in Eq S.45, for $t \le n-k_{t+1}$ we have,

$$
\begin{aligned}
&\frac{\kappa_{n,t}}{\kappa_{n,t+1}} \\
&\le c_t^2+4\eta_t^2\left(1+\eta_t\lambda_1\right)^2\mathcal{M}^2+4\left(1+\epsilon\right)c_t\eta_t\mathcal{M}\left(2\phi_\epsilon\eta_{t+1}k_{t+1}\left(\mathcal{M}+\lambda_1\right)+\left(2+\phi_\epsilon^2k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^2\right)\eta_{t+1}^2\right) \\
&\le c_t^2+4\eta_t^2\left(1+\eta_t\lambda_1\right)^2\mathcal{M}^2+4\left(1+\epsilon\right)c_t\eta_t\mathcal{M}\left(2\phi_\epsilon\eta_t k_{t+1}\left(\mathcal{M}+\lambda_1\right)+\left(2+\phi_\epsilon^2k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^2\right)\eta_t^2\right) \\
&= c_t^2+4\eta_t^2\left[\mathcal{M}^2+2\phi_\epsilon\left(1+\epsilon\right)c_t\mathcal{M}\left(\mathcal{M}+\lambda_1\right)k_{t+1}\right]+4\eta_t^3\left[\left(1+2\epsilon\right)\lambda_1+\left(1+\epsilon\right)c_t\mathcal{M}\left(2+\phi_\epsilon^2k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^2\right)\right] \\
&\le c_t^2+4\eta_t^2\left[2+2\phi_\epsilon\left(1+\epsilon\right)c_tk_{t+1}\right]\mathcal{M}\left(\mathcal{M}+\lambda_1\right)+4\left(1+2\epsilon\right)\eta_t^3\left[\lambda_1+c_t\mathcal{M}\left(2+\phi_\epsilon^2k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^2\right)\right] \\
&\le 1+4\eta_t\lambda_1+\eta_t^2\left[10+8\phi_\epsilon\left(2+\epsilon\right)c_tk_{t+1}\right]\left(\mathcal{M}+\lambda_1\right)^2+4\left(1+2\epsilon\right)\eta_t^3\left[\lambda_1+2c_t\mathcal{M}+c_t\phi_\epsilon^2k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^3\right] \\
&\le \exp\left(4\eta_t\lambda_1+\eta_t^2\left(10+8\phi_\epsilon\left(1+\epsilon\right)c_tk_{t+1}\right)\left(\mathcal{M}+\lambda_1\right)^2+4\left(1+2\epsilon\right)\eta_t^3\left(\lambda_1+c_t\mathcal{M}+2c_t\phi_\epsilon^2k_{t+1}^2\left(\mathcal{M}+\lambda_1\right)^3\right)\right) \\
&\le \exp\left(4\eta_t\lambda_1+\eta_t^2\left(10+8\phi_\epsilon\left(1+\epsilon\right)c_tk_{t+1}\right)\left(\mathcal{M}+\lambda_1\right)^2+4\epsilon\left(1+2\epsilon\right)\eta_t^2\left(2c_t+c_t\phi_\epsilon^2k_{t+1}\left(\mathcal{M}+\lambda_1\right)^2\right)\right) \\
&\le \exp\left(4\eta_t\lambda_1+\eta_t^2\left(8\epsilon\left(1+2\epsilon\right)\omega_\epsilon+\left(10+\left(8\left(1+\epsilon\right)+4\epsilon\left(1+2\epsilon\right)\phi_\epsilon\right)\phi_\epsilon\omega_\epsilon k_{t+1}\right)\left(\mathcal{M}+\lambda_1\right)^2\right)\right) \\
&\le \exp\left(4\eta_t\lambda_1+\eta_t^2\left(1+\left(10+20k_{t+1}\right)\left(\mathcal{M}+\lambda_1\right)^2\right)\right) \\
&\le \exp\left(4\eta_t\lambda_1+\eta_t^2\left(1+\left(10+20k_{t+1}\right)\left(\mathcal{M}+\lambda_1\right)^2\right)\right) \\
&\le \exp\left(4\eta_t\lambda_1+40\eta_t^2k_{t+1}\left(\mathcal{M}+\lambda_1\right)^2\right) \text{ since } \left(\mathcal{M}+\lambda_1\right), k_{t+1}\ge 1
\end{aligned}
$$

Recall our definition of $k := k_{t+1}$. We can use the above recursion for $1 \le t \le n-k_{t+1}$. We note that $t = n-k_n$ satisfies the conditions. Therefore,

$$
\kappa_{n,1} \le \exp\left(\sum_{i=1}^{n-k_n}4\eta_i\lambda_1+\sum_{i=1}^{n-k_n}\eta_i^2\zeta_{k,i}\right)\kappa_{n,n-k_n+1}
$$

Let $B_{n,n-k_n+1} = I+R'$, with $\|R'\|_2 \le r$ a.s.

$$
\begin{aligned}
\kappa_{n,n-k_n+1} &= \mathbb{E}\left[\left(v_1^TB_{n,n-k_n+1}B_{n,n-k_n+1}^Tv_1\right)^2\right] \\
&= \mathbb{E}\left[\left(v_1^Tv_1+v_1^T(R'+R'^T)v_1+v_1^TR'R'^Tv_1\right)^2\right] \\
&\le \left(1+2r+r^2\right)^2\mathbb{E}\left[\left(v_1^Tv_1\right)^2\right]
\end{aligned}
$$

Using Lemma 2, we have

$$
\begin{aligned}
r &\le \left(1+\epsilon\right)k_{n+1}\eta_{n-k_{n+1}}\left(\mathcal{M}+\lambda_1\right) \\
&\le \left(1+\epsilon\right)k_n\eta_{n-k_n}\left(\mathcal{M}+\lambda_1\right) \\
&\le 2\left(1+\epsilon\right)k_n\eta_n\left(\mathcal{M}+\lambda_1\right) \quad \text{since } \eta_{n-k_n}\le 2\eta_n
\end{aligned}
$$

which completes our proof. □

S.5 Main Results : Details and Proofs

S.5.1 Proof of Theorem 1

Lemma S.12. *This lemma proves conditions required later in the proof. Let the step-sizes be set according to Lemma S.11 and $m := 200$. Define*

$$r := 2\left(1+\epsilon\right)\eta_n k_n\left(\mathcal{M}+\lambda_1\right),$$
$$s := 3(1+r)^2\sum_{t=1}^{n-k_n-1} W_{k,t}\eta_t^2\exp\left(\sum_{i=t+1}^{n-k_n-1}\overline{\mathcal{V}_{k,i}}\eta_i^2\right)$$

where $W_{k,t}$ is defined in Theorem 4, $\overline{\mathcal{V}_{k,i}}$ is defined in S.38 and $\alpha, \beta, f\left(.\right), \delta$ are defined in Lemma S.11. Then for sufficiently large number of samples n, such that

$$\frac{n}{\log\left(f\left(n\right)\right)} > \frac{\beta}{\log\left(f\left(0\right)\right)}$$

we have

1. $2r + s \leq \frac{1}{2}$ *(S.52)*
2. $r = 2\left(1+\epsilon\right)\eta_n k_n\left(\mathcal{M}+\lambda_1\right) < \frac{1}{50}\frac{\delta/m}{1+\delta/m}$ *(S.55)*

Proof. For (1), using Lemma S.11-(3), we note that

$$\begin{aligned}
s &\leq 3(1+r)^2\sum_{t=1}^{n-k_n-1} W_{k,t}\eta_t^2\exp\left(\sum_{i=t+1}^{n-k_n-1}\overline{\mathcal{V}_{k,i}}\eta_i^2\right)\\
&\leq 3(1+r)^2\sum_{t=1}^{n-k_n-1} W_{k,t}\eta_t^2\left(1+\frac{\delta}{m}\right)\\
&\leq \frac{3(1+r)^2}{100}\left(1+\frac{\delta}{m}\right)\log\left(1+\frac{\delta}{m}\right) \qquad \text{(S.48)}\\
&\leq \frac{3(1+r)^2\log\left(2\right)}{50} \text{ since } \frac{\delta}{m} < 1
\end{aligned}$$

Therefore,

$$\begin{aligned}
2r + s &\leq 2r + \frac{3\left(1+r\right)^2}{25}\\
&= \frac{3}{25} + \frac{56}{25}r + \frac{3}{25}r^2 \qquad \text{(S.49)}
\end{aligned}$$

Setting $\frac{3}{25} + \frac{56}{25}r + \frac{3}{25}r^2 \leq \frac{1}{2}$, we have,

$$\begin{aligned}
&\frac{3}{25} + \frac{56}{25}r + \frac{3}{25}r^2 \leq \frac{1}{2}\\
\implies &6r^2 + 112r - 19 \leq 0
\end{aligned}$$

which holds for $r \in \left[0, \frac{1}{10}\right]$.

For (2), using Lemma S.11 and substituting the value of $k_i := \tau_{\text{mix}}\left(\eta_i^2\right) \leq \frac{2\tau_{\text{mix}}}{\log(2)}\log\left(\frac{1}{\eta_i^2}\right)$ for $\eta_i < 1$, we note that

$$\begin{aligned}
r &\leq \frac{8\left(1+\epsilon\right)\tau_{\text{mix}}\left(\mathcal{M}+\lambda_1\right)}{\log\left(2\right)}\frac{\alpha}{\left(\lambda_1-\lambda_2\right)\left(\beta+n\right)}\log\left(\frac{\left(\lambda_1-\lambda_2\right)\left(\beta+n\right)}{\alpha}\right)\\
&= \frac{8\left(1+\epsilon\right)\tau_{\text{mix}}\left(\mathcal{M}+\lambda_1\right)}{\log\left(2\right)}\frac{\log\left(\frac{(\lambda_1-\lambda_2)(\beta+n)}{\alpha}\right)}{\frac{(\lambda_1-\lambda_2)(\beta+n)}{\alpha}}\\
&= \frac{8\left(1+\epsilon\right)\tau_{\text{mix}}\left(\mathcal{M}+\lambda_1\right)}{\log\left(2\right)}\frac{\log\left(f\left(n\right)\right)}{f\left(n\right)}
\end{aligned}$$

Therefore (2) holds for sufficiently large n, i.e,

$$\frac{f(n)}{\log(f(n))} \geq \frac{400\left(1+\frac{\delta}{m}\right)(1+\epsilon)\,\tau_{\text{mix}}\left(\mathcal{M}+\lambda_1\right)}{\log(2)\,\frac{\delta}{m}}$$

This is satisfied if

$$\frac{n}{\log(f(n))} \geq \frac{400\tau_{\text{mix}}\left(1+\frac{\delta}{m}\right)(1+\epsilon)}{\log(2)}\frac{\left(\mathcal{M}+\lambda_1\right)\alpha}{\left(\lambda_1-\lambda_2\right)\frac{\delta}{m}} \tag{S.50}$$

From Lemma S.11, we have

$$\frac{\beta}{\log(f(0))} \geq \frac{600\tau_{\text{mix}}(1+2\epsilon)^2\left(\mathcal{M}+\lambda_1\right)^2\alpha^2}{\left(\lambda_1-\lambda_2\right)^2\log\left(1+\frac{\delta}{m}\right)} \overset{(i)}{\geq} \frac{400\tau_{\text{mix}}\left(1+\frac{\delta}{m}\right)(1+\epsilon)}{\log(2)}\frac{\left(\mathcal{M}+\lambda_1\right)\alpha}{\left(\lambda_1-\lambda_2\right)\frac{\delta}{m}}$$

where (i) follows since $\frac{\mathcal{M}+\lambda_1}{\lambda_1-\lambda_2} > 1, \alpha > 2$ and $\log(1+x) \leq x\ \forall x$. Therefore, $\frac{n}{\log(f(n))} > \frac{\beta}{\log(f(0))}$ suffices. Further, we note that (2) implies (1) for m = 200, $\delta \leq 1$. Therefore, the condition on n is sufficient for both results. Hence proved. $\square$

Lemma S.13. *Let*

$$u := \min\left\{i : i \in [n], i-k_i \geq 0\right\}$$

where k_i is defined in Lemma S.11. Then,

$$u \leq \lfloor\beta\rfloor \leq \beta$$

Proof. Using the definition of k_i mentioned in Lemma S.11, we have

$$\begin{aligned} k_i := \tau_{\text{mix}}\left(\eta_i^2\right) &\leq \frac{2\tau_{\text{mix}}}{\log(2)}\log\left(\frac{1}{\eta_i^2}\right) \\ &= \frac{4\tau_{\text{mix}}}{\log(2)}\log\left(\frac{\left(\lambda_1-\lambda_2\right)(\beta+i)}{\alpha}\right) \end{aligned}$$

Therefore,

$$\begin{aligned} \lfloor\beta\rfloor - k_{\lfloor\beta\rfloor} &\geq \lfloor\beta\rfloor - \frac{4\tau_{\text{mix}}}{\log(2)}\log\left(\frac{\beta+\lfloor\beta\rfloor}{\frac{\alpha}{\lambda_1-\lambda_2}}\right) \\ &\geq \frac{\beta}{2} - \frac{4\tau_{\text{mix}}}{\log(2)}\log\left(\frac{2\beta}{\frac{\alpha}{\lambda_1-\lambda_2}}\right) \text{ since } \beta > 1 \\ &= \beta\left[\frac{1}{2} - \frac{4\tau_{\text{mix}}}{\log(2)}\frac{\log(2f(0))}{\beta}\right], \text{ where } f(.) \text{ is defined in Lemma } S.11 \end{aligned}$$

Now, from Lemma S.11, we know that $f(0) > e$. Therefore, $\log(2f(0)) \leq 2\log(f(0))$. Then,

$$\lfloor\beta\rfloor - k_{\lfloor\beta\rfloor} \geq \beta\left[\frac{1}{2} - \frac{8\tau_{\text{mix}}}{\log(2)}\frac{\log(f(0))}{\beta}\right]$$

Again, from the conditions in Lemma S.11, we know that

$$\frac{\log(f(0))}{\beta} \leq \frac{\epsilon}{6\tau_{\text{mix}}}\frac{\lambda_1-\lambda_2}{\left(\mathcal{M}+\lambda_1\right)\alpha} \leq \frac{1}{120\tau_{\text{mix}}} \text{ since } \alpha > 2, \frac{\lambda_1-\lambda_2}{\mathcal{M}+\lambda_1} \leq 1, \epsilon \leq \frac{1}{100}$$

Therefore,

$$\lfloor\beta\rfloor - k_{\lfloor\beta\rfloor} \geq \beta\left(\frac{1}{2} - \frac{8}{120\log(2)}\right) \geq 0$$

Hence proved. $\square$

S.5.1.1 Numerator

Using Theorem 3 and Markov's Inequality, we have with probability atleast $(1-\delta)$

$$\mathrm{Tr}\left(V_\perp^T B_n B_n^T V_\perp\right) \leq$$
$$1.05\frac{\exp\left(\sum_{i=u+1}^{n} 2\eta_i\lambda_2 + \overline{\mathcal{V}_{k,i}}\eta_{i-k_i}^2\right)}{\delta}\left(d + \sum_{i=u+1}^{n}\left(\mathcal{V}' + \xi_{k,i}\right)C'_{k,i}\eta_{i-k_i}^2 \exp\left(\sum_{j=u+1}^{i} 2\eta_j\left(\lambda_1-\lambda_2\right)\right)\right)$$

S.5.1.2 Denominator

Using Chebyshev's Inequality we have, with probability atleast $(1-\delta)$

$$v_1^T B_n B_n^T v_1 \geq \mathbb{E}\left[v_1^T B_n B_n^T v_1\right]\left(1 - \sqrt{\frac{1}{\delta}}\sqrt{\frac{\mathbb{E}\left[\left(v_1^T B_n B_n^T v_1\right)^2\right]}{\mathbb{E}\left[v_1^T B_n B_n^T v_1\right]^2} - 1}\right) \tag{S.51}$$

Let $r := 2\left(1+\epsilon\right)\eta_n k_n\left(\mathcal{M}+\lambda_1\right) \leq \frac{1}{10}$. Using Theorem 3, we have

$$\mathbb{E}\left[\left(v_1^T B_n B_n^T v_1\right)^2\right] \leq \left(1+r\right)^4 \exp\left(\sum_{i=1}^{n-k_n} 4\eta_i\lambda_1 + \sum_{i=1}^{n-k_n}\eta_i^2\zeta_{k,t}\right)$$

Using Theorem 4, we have

$$\mathbb{E}\left[v_1^T B_{n,1} B_{n,1}^T v_1\right] \geq \exp\left(2\lambda_1\sum_{i=1}^{n-k_n}\eta_i - 4\lambda_1^2\sum_{i=1}^{n-k_n}\eta_i^2\right)\left[1 - \left(2r + 3\left(1+r\right)^2\sum_{t=1}^{n-k_n} W_{k,t}\eta_t^2\exp\left(\sum_{i=t+1}^{n-k_n}\eta_i^2\overline{\mathcal{V}_{k,i}}\right)\right)\right]$$

Let

$$s := 3\left(1+r\right)^2\sum_{t=1}^{n-k_n} W_{k,t}\eta_t^2\exp\left(\sum_{i=t+1}^{n-k_n}\eta_i^2\overline{\mathcal{V}_{k,i}}\right)$$

Then,

$$\frac{\mathbb{E}\left[\left(v_1^T B_n B_n^T v_1\right)^2\right]}{\mathbb{E}\left[v_1^T B_n B_n^T v_1\right]^2} \leq \frac{\left(1+r\right)^4}{\left(1-2r-s\right)^2}\exp\left(\sum_{i=1}^{n-k_n}\eta_i^2\left(\zeta_{k,i}+4\lambda_1^2\right)\right)$$

By Lemma S.12, we have that

$$2r + s \leq \frac{1}{2}. \tag{S.52}$$

Then, using

$$\frac{1}{\left(1-x\right)^2} \leq 1 + 6x \text{ for } x \in \left[0, \frac{1}{2}\right] \text{ and, } \left(1+x\right)^4 \leq 1+5x \text{ for } x \in \left[0, \frac{1}{10}\right]$$

we have,

$$\begin{aligned}\frac{\mathbb{E}\left[\left(v_1^T B_n B_n^T v_1\right)^2\right]}{\mathbb{E}\left[v_1^T B_n B_n^T v_1\right]^2} &\leq \left(1+5r\right)\left(1+12r+6s\right)\exp\left(\sum_{i=1}^{n-k_n}\eta_i^2\left(\zeta_{k,i}+4\lambda_1^2\right)\right)\\ &\leq \left(1+17r+6s+60r^2+30rs\right)\exp\left(\sum_{i=1}^{n-k_n}\eta_i^2\left(\zeta_{k,i}+4\lambda_1^2\right)\right)\\ &\leq \left(1+22r+12s\right)\exp\left(\sum_{i=1}^{n-k_n}\eta_i^2\left(\zeta_{k,i}+4\lambda_1^2\right)\right) \text{ since } r \leq \frac{1}{10}\end{aligned}$$

By Lemma S.11-(3), we have that

$$\exp\left(\sum_{i=1}^{n-k_n}\eta_i^2\left(\zeta_{k,i}+4\lambda_1^2\right)\right)\le 1+\frac{\delta}{m} \tag{S.53}$$

By S.48, we have that

$$\begin{aligned}12s &\le \frac{48(1+r)^2}{100}\left(1+\frac{\delta}{m}\right)^2\log\left(1+\frac{\delta}{m}\right)\\ &\le \frac{3}{5}\left(1+\frac{\delta}{m}\right)^2\log\left(1+\frac{\delta}{m}\right)\ \text{ since } r\le\frac{1}{10}\end{aligned} \tag{S.54}$$

By Lemma S.12, we have that

$$r=2\left(1+\epsilon\right)\eta_n k_n\left(\mathcal{M}+\lambda_1\right)<\frac{1}{50}\frac{\delta/m}{1+\delta/m} \tag{S.55}$$

Then,

$$\begin{aligned}\frac{\mathbb{E}\left[\left(v_1^T B_n B_n^T v_1\right)^2\right]}{\mathbb{E}\left[v_1^T B_n B_n^T v_1\right]^2} &\le (1+22r+12s)\left(1+\frac{\delta}{m}\right)\\ &= 1+\frac{\delta}{m}+22r\left(1+\frac{\delta}{m}\right)+12s\left(1+\frac{\delta}{m}\right)\\ &\le 1+\frac{\delta}{m}+\frac{22}{50}\frac{\delta}{m}+\frac{3}{5}\left(1+\frac{\delta}{m}\right)^3\log\left(1+\frac{\delta}{m}\right)\\ &\le 1+\frac{\delta}{m}+\frac{22}{50}\frac{\delta}{m}+\frac{7}{10}\log\left(1+\frac{\delta}{m}\right)\ \text{ since } \delta\le 1, m=200\\ &\le 1+\frac{\delta}{m}+\frac{22}{50}\frac{\delta}{m}+\frac{7}{10}\frac{\delta}{m}\ \text{ since } \forall x,\ \log\left(1+x\right)\le x\\ &\le 1+3\frac{\delta}{m}\end{aligned}$$

Then setting $m=200$, from S.51 we have

$$\begin{aligned}v_1^T B_n B_n^T v_1 &\ge \exp\left(\sum_{i=1}^{n-k_n}2\eta_i\lambda_1-4\eta_i^2\lambda_1^2\right)(1-2r-s)\left(1-\sqrt{\frac{1}{\delta}}\sqrt{\frac{3\delta}{m}}\right)\\ &\ge \exp\left(\sum_{i=1}^{n-k_n}2\eta_i\lambda_1-4\eta_i^2\lambda_1^2\right)\left(1-\frac{1}{25}\frac{\delta/m}{1+\delta/m}-\frac{1}{20}\left(1+\frac{\delta}{m}\right)^2\log\left(1+\frac{\delta}{m}\right)\right)\left(1-\sqrt{\frac{3}{m}}\right)\\ &\ge \frac{5}{6}\exp\left(\sum_{i=1}^{n-k_n}2\eta_i\lambda_1-4\eta_i^2\lambda_1^2\right)\ \text{ since } \delta\le 1 \text{ and } m=200\end{aligned}$$

The second inequality uses Eqs S.54, S.55.

S.5.1.3 Fraction

Now that we have established this result let's calculate the fraction. Let the step-sizes be set according to Lemma S.11. Define

$$\begin{aligned}\mathcal{S} &:= \exp\left(\sum_{i=u+1}^{n}\overline{\mathcal{V}_{k,i}}\eta_{i-k_i}^2+\sum_{i=1}^{n-k_n}4\lambda_1^2\eta_i^2\right)\\ Q_u &:= \exp\left(2\lambda_1\left(\sum_{j=1}^{u}\eta_j-\sum_{j=n-k_n+1}^{n}\eta_j\right)\right)\\ \mathcal{R}_{k,t} &:= \frac{\exp\left(\sum_{j=1}^{t-u}\eta_j^2\left(\overline{\mathcal{V}_{k,j}}-\overline{\mathcal{V}_{k,j+u}}\right)\right)\exp\left(2\lambda_1\sum_{j=n-k_n+1}^{n}\eta_j\right)}{\exp\left(2\lambda_1\sum_{j=1}^{u}\eta_{t-u+j}\right)}\end{aligned}$$

Then, recall that

$$\begin{aligned}
u &:= \min\{i : i \in [n], i - k_i \geq 0\} \\
\xi_{k,t} &:= 6\eta_{t-k_t}\mathcal{M}\left[1 + 3k_{t+1}^2(\mathcal{M} + \lambda_1)^2\right] \\
\mathcal{V}' &:= \left(\frac{1 + (3 + 4\epsilon)|\lambda_2(P)|}{1 - |\lambda_2(P)|}\right)\mathcal{V} \\
\overline{\mathcal{V}_{k,t}} &:= \mathcal{V}' + \lambda_1^2 + \xi_{k,t} \\
C'_{k,t} &:= \exp\left(2\lambda_1 \sum_{j=1}^{u}(\eta_j - \eta_{t-u+j}) + \sum_{j=1}^{t-u}\eta_j^2\left(\overline{\mathcal{V}_{k,j}} - \overline{\mathcal{V}_{k,j+u}}\right)\right) = Q_u\mathcal{R}_{k,t}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{\text{Tr}\left(V_\perp^T B_n B_n^T V_\perp\right)}{v_1^T B_n B_n^T v_1} \\
&\leq \frac{1.3}{\delta}\frac{\exp\left(\sum_{i=u+1}^{n} 2\eta_i\lambda_2 + \overline{\mathcal{V}_{k,i}}\eta_{i-k_i}^2\right)}{\exp\left(\sum_{i=1}^{n-k_n} 2\eta_i\lambda_1 - 4\eta_i^2\lambda_1^2\right)}\left(d + \sum_{i=u+1}^{n}(\mathcal{V}' + \xi_{k,i})C'_{k,i}\eta_{i-k_i}^2 \exp\left(\sum_{j=u+1}^{i} 2\eta_j(\lambda_1 - \lambda_2)\right)\right) \\
&\leq \frac{1.3}{\delta}\frac{\mathcal{S}}{Q_u}\exp\left(\sum_{i=u+1}^{n} 2\eta_i(\lambda_2 - \lambda_1)\right)\left(d + \sum_{i=u+1}^{n}(\mathcal{V}' + \xi_{k,i})C'_{k,i}\eta_{i-k_i}^2 \exp\left(\sum_{j=u+1}^{i} 2\eta_j(\lambda_1 - \lambda_2)\right)\right) \\
&\leq \frac{1.3}{\delta}\mathcal{S}\left(\underbrace{\frac{d\exp\left(\sum_{i=u+1}^{n} 2\eta_i(\lambda_2 - \lambda_1)\right)}{Q_u}}_{X_1} + \underbrace{\sum_{i=u+1}^{n}(\mathcal{V}' + \xi_{k,i})\mathcal{R}_{k,i}\eta_{i-k_i}^2 \exp\left(-\sum_{j=i+1}^{n} 2\eta_j(\lambda_1 - \lambda_2)\right)}_{X_2}\right)
\end{aligned} \tag{S.56}$$

For X_1, we have

$$\begin{aligned}
X_1 &\leq \frac{d\exp\left(\sum_{i=u+1}^{n} 2\eta_i(\lambda_2 - \lambda_1)\right)}{Q_u} \\
&= \frac{d\exp\left(\sum_{i=u+1}^{n} 2\eta_i(\lambda_2 - \lambda_1)\right)}{\exp\left(2\lambda_1\left(\sum_{j=1}^{u}\eta_j - \sum_{j=n-k_n+1}^{n}\eta_j\right)\right)} \\
&\leq \frac{d\exp\left(\sum_{i=u+1}^{n} 2\eta_i(\lambda_2 - \lambda_1)\right)}{\exp\left(-2\lambda_1\left(\sum_{j=n-k_n+1}^{n}\eta_j\right)\right)} \\
&\leq d\exp\left(\sum_{i=u+1}^{n} 2\eta_i(\lambda_2 - \lambda_1)\right)\exp\left(2\lambda_1\left(\sum_{j=n-k_n+1}^{n}\eta_j\right)\right)
\end{aligned}$$

Note that

$$\begin{aligned}
\exp\left(2\lambda_1 \sum_{j=n-k_n+1}^{n}\eta_j\right) &\leq \exp\left(2(1+2\epsilon)\lambda_1 k_n \eta_{n-k_n+1}\right) \text{ using monotonicity of } \eta_i \\
&\leq \exp\left(4(1+2\epsilon)\lambda_1 k_n \eta_n\right) \text{ using slow-decay of } \eta_i \\
&\leq 1 + 2\frac{\delta}{m} \text{ using Lemma } S.12 \text{ along with } e^x \leq 1 + x + x^2 \text{ for } x \in (0,1)
\end{aligned}$$

Therefore, using S.20

$$X_1 \leq d\left(1 + \frac{2\delta}{m}\right)\left(\frac{\beta + u}{n}\right)^{2\alpha}$$

Next, for X_2, we first have

$$\mathcal{R}_{k,t} := \frac{\exp\left(\sum_{j=1}^{t-u}\eta_j^2\left(\overline{\mathcal{V}_{k,j}} - \overline{\mathcal{V}_{k,j+u}}\right)\right)\exp\left(2\lambda_1\sum_{j=n-k_n+1}^{n}\eta_j\right)}{\exp\left(2\lambda_1\sum_{j=1}^{u}\eta_{t-u+j}\right)}$$
$$\leq \exp\left(\sum_{j=1}^{t-u}\eta_j^2\overline{\mathcal{V}_{k,j}}\right)\exp\left(2\lambda_1\sum_{j=n-k_n+1}^{n}\eta_j\right)$$
$$\leq \left(1+\frac{2\delta}{m}\right)^2 \text{ using Lemmas } S.11-(3), S.12 \text{ and } e^x \leq 1+x+x^2 \text{ for } x \in (0,1)$$

Now, using S.11-(4) we have,

$$\sum_{i=1}^{n}\overline{\mathcal{V}_{k,i}}\eta_{i-k_i}^2\exp\left(-\sum_{j=i+1}^{n}2\eta_j\left(\lambda_1-\lambda_2\right)\right) \leq$$
$$\left(\frac{2\left(1+10\epsilon\right)\alpha^2}{2\alpha-1}\right)\frac{\mathcal{V}'}{\left(\lambda_1-\lambda_2\right)^2}\frac{1}{n} + \left(\frac{800\left(1+10\epsilon\right)\alpha^3}{\left(\alpha-1\right)}\right)\frac{\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2\tau_{\text{mix}}^2}{\left(\lambda_1-\lambda_2\right)^3}\frac{\log^2\left(\frac{(\beta+n)(\lambda_1-\lambda_2)}{\alpha}\right)}{n^2}$$

Then,

$$X_2 \leq \left(1+\frac{2\delta}{m}\right)^2\left[\underbrace{\left(\frac{2\left(1+10\epsilon\right)\alpha^2}{2\alpha-1}\right)}_{C_1}\frac{\mathcal{V}'}{\left(\lambda_1-\lambda_2\right)^2}\frac{1}{n} + \underbrace{\left(\frac{24\left(1+10\epsilon\right)\alpha^3}{\left(\alpha-1\right)}\right)}_{C_2}\frac{\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2}{\left(\lambda_1-\lambda_2\right)^3}\frac{k_n^2}{n^2}\right]$$

Therefore substituting in S.56,

$$\frac{\operatorname{Tr}\left(V_\perp^T B_n B_n^T V_\perp\right)}{v_1^T B_n B_n^T v_1} \leq \frac{1.3\mathcal{S}}{\delta}\left(1+\frac{2\delta}{m}\right)^2\left[d\left(\frac{\beta+u}{n}\right)^{2\alpha} + \frac{C_1\mathcal{V}'}{\left(\lambda_1-\lambda_2\right)^2}\frac{1}{n} + \frac{C_2\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2}{\left(\lambda_1-\lambda_2\right)^3}\frac{k_n^2}{n^2}\right] \tag{S.57}$$

Proof of Theorem 1. To complete our proof, we bound $\mathcal{S}$ to simplify S.57. We note that under the learning rate schedule presented in Lemma S.11-(3),

$$\mathcal{S} \leq \left(1+\frac{\delta}{m}\right)$$

Therefore,

$$\frac{\operatorname{Tr}\left(V_\perp^T B_n B_n^T V_\perp\right)}{v_1^T B_n B_n^T v_1} \leq \frac{1.3}{\delta}\left(1+\frac{2\delta}{m}\right)^3\left[d\left(\frac{\beta+u}{n}\right)^{2\alpha} + \frac{C_1\mathcal{V}'}{\left(\lambda_1-\lambda_2\right)^2}\frac{1}{n} + \frac{C_2\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2}{\left(\lambda_1-\lambda_2\right)^3}\frac{k_n^2}{n^2}\right]$$
$$\leq \frac{1.4}{\delta}\left[d\left(\frac{\beta+u}{n}\right)^{2\alpha} + \frac{C_1\mathcal{V}'}{\left(\lambda_1-\lambda_2\right)^2}\frac{1}{n} + \frac{C_2\mathcal{M}\left(\mathcal{M}+\lambda_1\right)^2}{\left(\lambda_1-\lambda_2\right)^3}\frac{k_n^2}{n^2}\right]$$

Using lemma S.13, we have that $u \leq \beta$. Then, using Lemma 3.1 from [1] completes our proof. □

S.5.2 Proof of Corollary 1

Proof of Corollary 1. We note that the downsampled data stream can be considered to be drawn from a Markov chain with transition kernel $P^k\left(.,.\right)$ since each data-point is k steps away from the previous one. We will denote the parameters of this transformed chain by $\tilde{y}$ when the corresponding parameter is y under the original chain. For example, $\tilde{\tau}_{\text{mix}}$ is the mixing time of the new chain.

Note that this modified transition matrix has the same stationary distribution π. It is also reversible. This can be seen by considering the diagonal matrix of stationary distribution probabilities Π, where $\Pi_{ii} = \pi_i$. For a reversible Markov Chain, we have $\Pi P = P\Pi$. However, that also implies

$\Pi P^2 = (\Pi P)P = (P\Pi)P = P(\Pi P) = P^2\Pi$. This same technique works for P^k yielding $\Pi P^k = P^k\Pi$.

Using standard results on Markov chains [3],

$$\frac{|\lambda_2(P)|}{1-|\lambda_2(P)|}\log\left(\frac{1}{2\epsilon}\right) \leq \tau_{\text{mix}}(\epsilon) \leq \frac{1}{1-|\lambda_2(P)|}\log\left(\frac{1}{\epsilon\pi_{\min}}\right), \tag{S.58}$$

where $\pi_{\min} := \min_i \pi_i$. Therefore, as noted in the theorem statement, we substitute the modified parameters in the bound we have proven for Theorem 1.

First, we will show that the mixing time for this new chain is $\Theta(1)$. We will use $k := \tau_{\text{mix}}\left(\eta_n^2\right)$. So by definition the $d_{\text{mix}}(k) \leq \eta_n^2$ using the definition of d_{mix} in Section 2.1. Hence $d_{\text{mix}}(k) \leq 1/4$ using conditions on the learning rate schedule imposed in Theorem 1. Therefore, in the transformed chain, the "new" $\tilde{\tau}_{\text{mix}}$ is $\Theta(1)$.

We also have:

$$k \leq \frac{2\tau_{\text{mix}}}{\log(2)}\log\left(\frac{1}{\eta_n^2}\right) \leq \underbrace{\frac{2\log(4/\pi_{\min})}{\log 2}}_{C}\frac{1}{1-|\lambda_2(P)|}\log(n)$$

We see that $C > 1$. Next, we note that for the transition kernel $P^k(.,.)$, the second-largest absolute eigenvalue is given as $|\lambda_2(P)|^k$. Consider the function $f(x) := x^{\frac{1}{1-x}}$ for $x \in (0,1)$. Then,

$$f'(x) = f(x)\left(\frac{1-x-x\log(x)}{x(1-x)^2}\right) > 0$$

Therefore, $f(x) < \lim_{x\to 1} f(x) = \frac{1}{e} < 1$. which implies $|\lambda_2(P)|^k \overset{(i)}{\leq} \left(\frac{1}{e}\right)^{C\log(n)} < \frac{1}{e}$. Here (i) follows if $C > 1, n > 3$, which is true. Therefore,

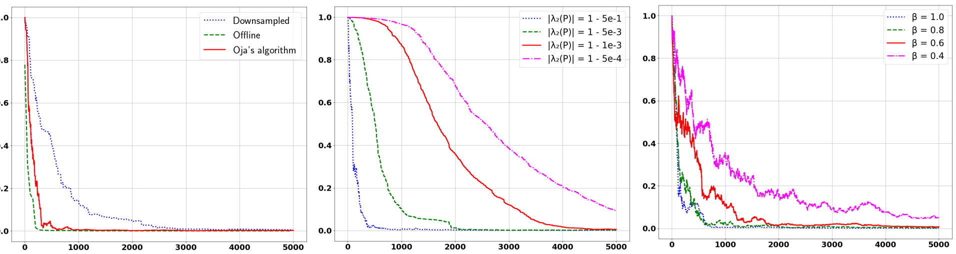
$$\widetilde{\mathcal{V}}' := \left(\frac{1+(3+4\epsilon)|\lambda_2(P)|^k}{1-|\lambda_2(P)|^k}\right)\mathcal{V} \leq 5\mathcal{V}$$

This also implies that the mixing time for the new Markov chain for sub-sampled data is $\Theta(1)$. The bound then follows by substituting n to be $\frac{n}{k} = n_k = \Theta\left(\frac{n}{C\tau_{\text{mix}}\log(n)}\right)$ and setting the τ_{mix} in the original expression of Theorem 1 to a constant. $\square$

S.6 Additional Experiments

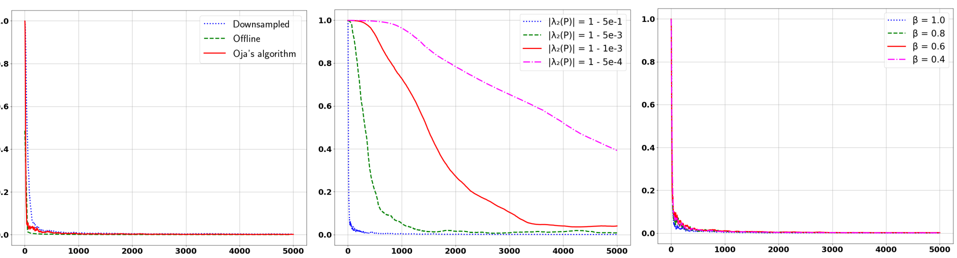
In this section, we first provide additional experiments to support the results established in Section 3 of the manuscript. We present experiments with distributions that have nonzero mean vectors at each state, but zero mean with respect to the stationary distribution. This means that the Z_i's are not necessarily zero-mean with respect to each state distribution $D(s)$. To normalize the data-points, we estimate the mean μ and covariance matrix Σ empirically from a much larger independently generated dataset.

We experiment with two different settings here - Figure S.1 contains the results for each state distribution being $D(s) :=$ Bernoulli(p_s) with $p_s \sim \mathcal{U}(0, 0.05)$ being fixed for each dataset. Figure S.2 provides results for each state distribution being $D(s) := \mathcal{U}(0, \ell_s)$ with $\ell_s \sim \mathcal{U}(0, 10)$ being selected at the start of each random run. We observe that these experiments depict similar trends to those shown in the main manuscript, which validates our results for the case of non-zero state means. Furthermore, the Bernoulli data, being sparse compared to the Uniform one, seems to exhibit a clearer difference between data downsampling and the traditional Oja's algorithm.



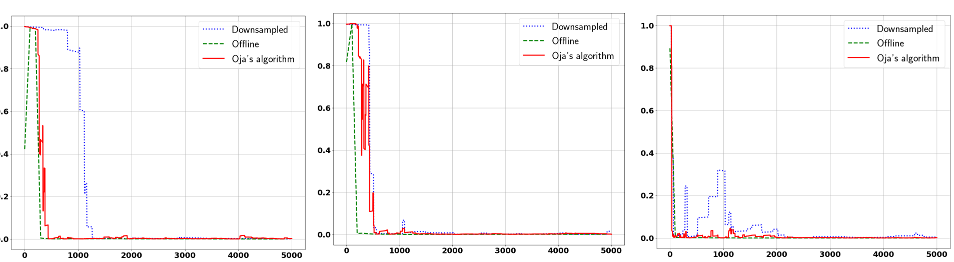
(a) (b) (c)

Figure S.1: Experiments with Bernoulli data. S.1a compares the three different algorithms, S.1b shows effect of changing the eigengap of the transition_matrix and S.1c records the variation in performance on changing the eigengap of the data covariance matrix.

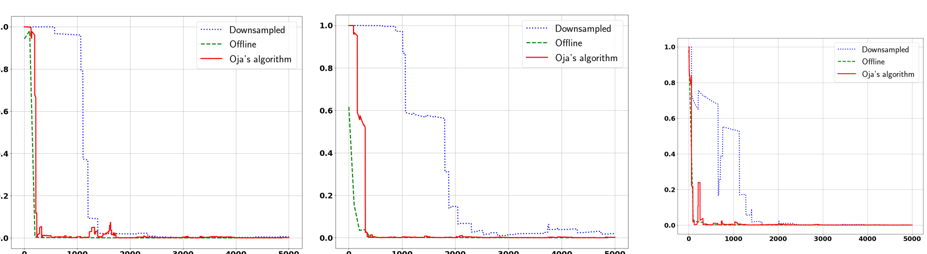


(a) (b) (c)

Figure S.2: Experiments with Uniform data. S.2a compares the three different algorithms, S.2b shows effect of changing the eigengap of the transition matrix and S.2c records the variation in performance on changing the eigengap of the data covariance matrix.



(a) (b) (c)



(d) (e) (f)

Figure S.3: Randomly chosen runs for the Bernoulli case

To provide clear plots demonstrating the relative behavior of the algorithms considered in this paper, we have shown the averaged $\sin^2$ errors in Figures S.1 and S.2. In Figure S.3 we show six random runs where we fixed the $p_s, s \in \Omega$ for each state for all runs. These figures clearly show that in general, Downsampled Oja has a worse performance than Oja's algorithm, which has a similar performance as the offline algorithm. It also shows that the Downsampled algorithm has the most variability, whereas Oja's algorithm on the whole dataset has much less variability, and finally, and not surprisingly, the offline algorithm has the least variability. Similar qualitative trends can be

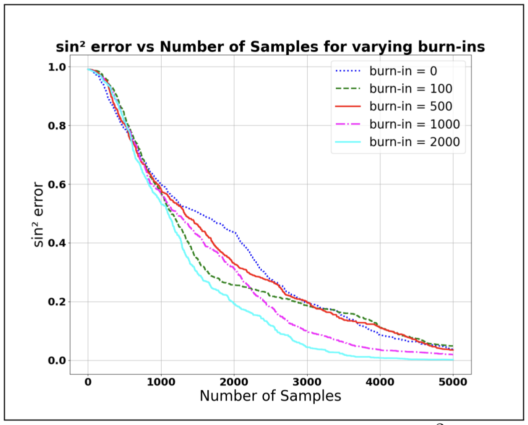


Figure S.4: Effect of burn-in period on $\sin^2$ error

observed for the other settings.

Next, we study the effect of different burn-in periods on the $\sin^2$ error in Figure S.4. We use the same experimental setup as described in Section 6 with $\rho = 0.001$, which implies that the mixing time τ_{mix} of the chain is ~ 1000. We start the chain with an initial distribution different from the stationary distribution and record the $\sin^2$ error after letting the chain run for burn-in timesteps. We observe that in general, the longer the burn-in period, the faster the error decay.

References

[1] Prateek Jain, Chi Jin, Sham Kakade, Praneeth Netrapalli, and Aaron Sidford. Streaming pca: Matching matrix bernstein and near-optimal finite sample guarantees for oja's algorithm. In *Proceedings of The 29th Conference on Learning Theory (COLT)*, June 2016.

[2] László Kozma. Inequalities cheat sheet, 2018. PDF file.

[3] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[4] Joe Neeman, Bobby Shi, and Rachel Ward. Concentration inequalities for sums of markov dependent random matrices, 2023.

[5] Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12:99–111, 1972.