A. Additional Proofs

A.1. Proof of Lemma 2

The result trivially holds for $k=0$, so we will assume $k>0$ from now. Let

$$f(s) = (1+\eta s)^k(1-\epsilon-s).$$

Differentiating f and setting to zero, we have

$$\begin{aligned}
& k\eta(1+\eta s)^{k-1}(1-\epsilon-s) - (1+\eta s)^k = 0 \\
\Leftrightarrow \;\; & k\eta(1-\epsilon-s) = 1+\eta s \\
\Leftrightarrow \;\; & \frac{k\eta(1-\epsilon)-1}{k\eta+\eta} = s \\
\Leftrightarrow \;\; & s = \frac{k(1-\epsilon)-1/\eta}{k+1}.
\end{aligned}$$

Let $s_c = \frac{k(1-\epsilon)-1/\eta}{k+1}$ denote this critical point, and consider two cases:

- $s_c \notin [0,1]$: In that case, f has no critical points in the domain, hence is maximized at one of the domain endpoints, with a value of at most
$$\max\{f(0), f(1)\} = \max\{1-\epsilon, -\epsilon(1+\eta)^k\} \leq 1.$$

- $s_c \in [0,1]$: In that case, we must have $k(1-\epsilon) - \frac{1}{\eta} \geq 0$, and the value of f at s_c is
$$\begin{aligned}
& \left(1+\frac{\eta k(1-\epsilon)-1}{k+1}\right)^k \left(1-\epsilon-\frac{k(1-\epsilon)-1/\eta}{k+1}\right) \\
& = \left(1+\frac{\eta k(1-\epsilon)-1}{k+1}\right)^k \left(\frac{1-\epsilon+\frac{1}{\eta}}{k+1}\right) \\
& \leq (1+\eta(1-\epsilon))^k \left(\frac{1+\frac{1}{\eta}}{k+1}\right) \\
& \leq \frac{2\,(1+\eta(1-\epsilon))^k}{\eta(k+1)}.
\end{aligned}$$

  The maximal value of f is either the value above, or the maximal value of f at the domain endpoints, which we already showed to be most 1. Overall, the maximal value f can attain is at most
$$\max\left\{1, \frac{2\,(1+\eta(1-\epsilon))^k}{\eta(k+1)}\right\} \leq 1 + \frac{2\,(1+\eta(1-\epsilon))^k}{\eta(k+1)}.$$

Combining the two cases, the result follows.

A.2. Proof of Lemma 3

To simplify notation, define for all $t=1,\ldots,T$ the matrices

$$C_0^t = I + \eta A \quad , \quad C_1^t = \eta(\tilde{A}_t - A).$$

Note that C_0^t is deterministic whereas C_1^t is random and zero-mean. Moreover, $\|C_0^t\| \leq 1+\eta$ and $\|C_1^t\| \leq \eta b$.

By definition of the algorithm, we have the following:

$$\begin{aligned} V_T &= \mathbf{w}_T^\top((1-\epsilon)I - A)\mathbf{w}_T \\ &= \mathbf{w}_0^\top\left(\prod_{t=1}^{T}\left(I+\eta\tilde{A}_t\right)\right)((1-\epsilon)I-A)\left(\prod_{t=T}^{1}\left(I+\eta\tilde{A}_t\right)\right)\mathbf{w}_0 \\ &= \mathbf{w}_0^\top\left(\prod_{t=1}^{T}\left(C_0^t+C_1^t\right)\right)((1-\epsilon)I-A)\left(\prod_{t=T}^{1}\left(C_0^t+C_1^t\right)\right)\mathbf{w}_0 \\ &= \sum_{(i_1,\ldots,i_T)\in\{0,1\}^T}\ \sum_{(j_1,\ldots,j_T)\in\{0,1\}^T}\mathbf{w}_0^\top\left(\prod_{t=1}^{T}C_{i_t}^t\right)((1-\epsilon)I-A)\left(\prod_{t=T}^{1}C_{j_t}^t\right)\mathbf{w}_0. \end{aligned}$$

Since $C_1^1,\ldots,C_1^T$ are independent and zero-mean, the expectation of each summand in the expression above is non-zero only if $i_t = j_t$ for all t. Therefore,

$$\mathbb{E}\left[\mathbf{w}_T^\top((1-\epsilon)I-A)\mathbf{w}_T\right] \;=\; \sum_{(i_1,\ldots,i_T)\in\{0,1\}^T}\mathbb{E}\left[\mathbf{w}_0^\top\left(\prod_{t=1}^{T}C_{i_t}^t\right)((1-\epsilon)I-A)\left(\prod_{t=T}^{1}C_{i_t}^t\right)\mathbf{w}_0\right].$$

We now decompose this sum according to what is the largest value of t for which $i_t = 1$ (hence $C_{i_t}^t = C_1^t$). The intuition for this, as will be seen shortly, is that Lemma 2 allows us to attain tighter bounds on the summands when t is much smaller than T. Formally, we can rewrite the expression above as

$$\begin{aligned} &\mathbb{E}\left[\mathbf{w}_0^\top\left(\prod_{t=1}^{T}C_0^t\right)((1-\epsilon)I-A)\left(\prod_{t=T}^{1}C_0^t\right)\mathbf{w}_0\right] \\ &\quad+\sum_{k=0}^{T-1}\ \sum_{(i_1,\ldots,i_k)\in\{0,1\}^k}\mathbb{E}\left[\mathbf{w}_0^\top\left(\prod_{t=1}^{k}C_{i_t}^t\right)C_1^{k+1}\left(\prod_{t=k+2}^{T}C_0^t\right)((1-\epsilon)I-A)\left(\prod_{t=T}^{k+2}C_0^t\right)C_1^{k+1}\left(\prod_{t=k}^{1}C_{i_t}^t\right)\mathbf{w}_0\right]. \end{aligned}$$

Since $C_0^t = I+\eta A$ is diagonal and the same for all t, and $((1-\epsilon)I-A)$ is diagonal as well, we can simplify the above to

$$\begin{aligned} &\mathbf{w}_0^\top(C_0^1)^{2T}((1-\epsilon)I-A)\mathbf{w}_0 \\ &\quad+\sum_{k=0}^{T-1}\ \sum_{(i_1,\ldots,i_k)\in\{0,1\}^k}\mathbb{E}\left[\mathbf{w}_0^\top\left(\prod_{t=1}^{k}C_{i_t}^t\right)C_1^{k+1}(C_0^1)^{2(T-k-1)}((1-\epsilon)I-A)C_1^{k+1}\left(\prod_{t=k}^{1}C_{i_t}^t\right)\mathbf{w}_0\right]. \end{aligned}$$

Using the fact that the spectral norm is sub-multiplicative, and that for any symmetric matrix B, $\mathbf{v}^\top B\mathbf{v} \leq \|\mathbf{v}^2\|\lambda_{\max}(B)$, where $\lambda_{\max}(B)$ denotes the largest eigenvalue of B, we can upper bound the above by

$$\begin{aligned} &\leq\ \mathbf{w}_0^\top(C_0^1)^{2T}((1-\epsilon)I-A)\mathbf{w}_0 \\ &\quad+\sum_{k=0}^{T-1}\ \sum_{(i_1,\ldots,i_k)\in\{0,1\}^k}\mathbb{E}\left[\|\mathbf{w}_0\|^2\left(\prod_{t=1}^{k}\|C_{i_t}^t\|^2\right)\|C_1^{k+1}\|^2\lambda_{\max}\left((C_0^1)^{2(T-k-1)}((1-\epsilon)I-A)\right)\right]. \end{aligned}$$

Since $\|\mathbf{w}_0\| = 1$, and $\|C_0^t\| \leq (1+\eta)$, $\|C_1^t\| \leq \eta b$, this is at most

$$\begin{aligned}
&\mathbf{w}_0^\top (C_0^1)^{2T}((1-\epsilon)I - A)\mathbf{w}_0 \\
&\quad + \sum_{k=0}^{T-1} \sum_{(i_1,\ldots,i_k)\in\{0,1\}^k} \left((1+\eta)^{2\left(k-\sum_{t=1}^k i_t\right)} (\eta b)^{2\sum_{t=1}^k i_t}\right) (\eta b)^2 \lambda_{\max}\left((C_0^1)^{2(T-k-1)}((1-\epsilon)I - A)\right) \\
&= \mathbf{w}_0^\top (C_0^1)^{2T}((1-\epsilon)I - A)\mathbf{w}_0 \\
&\quad + \sum_{k=0}^{T-1} \left((1+\eta)^2 + (\eta b)^2\right)^k (\eta b)^2 \lambda_{\max}\left((C_0^1)^{2(T-k-1)}((1-\epsilon)I - A)\right) \\
&= \mathbf{w}_0^\top (I+\eta A)^{2T}((1-\epsilon)I - A)\mathbf{w}_0 \\
&\quad + (\eta b)^2 \sum_{k=0}^{T-1} \left((1+\eta)^2 + (\eta b)^2\right)^k \lambda_{\max}\left((I+\eta A)^{2(T-k-1)}((1-\epsilon)I - A)\right)
\end{aligned} \tag{6}$$

Recalling that $A = \text{diag}(s_1,\ldots,s_d)$ with $s_1 = 1$, that $\|\mathbf{w}_0\|^2 = \sum_{j=1}^d w_{0,j}^2 = 1$, and that $w_{0,1}^2 \geq \frac{1}{p}$, the first term in Eq. (6) equals

$$\begin{aligned}
\mathbf{w}_0^\top (I+\eta A)^{2T}((1-\epsilon)I - A)\mathbf{w}_0 &= \sum_{j=1}^d (1+\eta s_j)^{2T}(1-\epsilon-s_j)w_{0,j}^2 \\
&= (1+\eta)(-\epsilon)w_{0,1}^2 + \sum_{j=2}^d (1+\eta s_j)^{2T}(1-\epsilon-s_j)w_{0,j}^2 \\
&\leq -(1+\eta)^{2T}\frac{\epsilon}{p} + \max_{s\in[0,1]}(1+\eta s)^{2T}(1-\epsilon-s).
\end{aligned}$$

Applying Lemma 2, and recalling that $\eta \leq 1$, we can upper bound the above by

$$\begin{aligned}
& -(1+\eta)^{2T}\frac{\epsilon}{p} + 1 + 2\frac{(1+\eta(1-\epsilon))^{2T}}{\eta(2T+1)} \\
&= \ (1+\eta)^{2T}\left(-\frac{\epsilon}{p} + (1+\eta)^{-2T} + 2\frac{\left(\frac{1+\eta(1-\epsilon)}{1+\eta}\right)^{2T}}{\eta(2T+1)}\right) \\
&\leq \ (1+\eta)^{2T}\left(-\frac{\epsilon}{p} + (1+\eta)^{-2T} + \frac{(1-\frac{1}{2}\eta\epsilon))^{2T}}{\eta T}\right).
\end{aligned} \tag{7}$$

As to the second term in Eq. (6), again using the fact that $A = \text{diag}(s_1,\ldots,s_d)$, we can upper bound it by

$$(\eta b)^2 \sum_{k=0}^{T-1} \left((1+\eta)^2 + (\eta b)^2\right)^k \max_{s\in[0,1]} (1+\eta s)^{2(T-k-1)}(1-\epsilon-s).$$

Applying Lemma 2, and recalling that $\eta \leq 1$, this is at most

$$\begin{aligned}
&(\eta b)^2 \sum_{k=0}^{T-1} \left((1+\eta)^2 + (\eta b)^2\right)^k \left(1 + 2\frac{(1+\eta(1-\epsilon))^{2(T-k-1)}}{\eta(2(T-k)-1)}\right) \\
&= \ (\eta b)^2(1+\eta)^{2T} \sum_{k=0}^{T-1} \left(1 + \left(\frac{\eta b}{1+\eta}\right)^2\right)^k \left((1+\eta)^{-2(T-k)} + 2\frac{\left(\frac{1+\eta(1-\epsilon)}{1+\eta}\right)^{2(T-k)}}{\eta(2(T-k)-1)}\right) \\
&\leq \ (\eta b)^2(1+\eta)^{2T} \sum_{k=0}^{T-1} \left(1 + (\eta b)^2\right)^k \left((1+\eta)^{-2(T-k)} + 2\frac{\left(1-\frac{1}{2}\eta\epsilon\right)^{2(T-k)}}{\eta(2(T-k)-1)}\right).
\end{aligned}$$

Upper bounding $\left(1+(\eta b)^2\right)^k$ by $\left(1+(\eta b)^2\right)^T$, and rewriting the sum in terms of k instead of $T-k$, we get

$$(\eta b)^2(1+\eta)^{2T}\left(1+(\eta b)^2\right)^T\sum_{k=1}^{T}\left((1+\eta)^{-2k}+2\frac{\left(1-\frac{1}{2}\eta\epsilon\right)^{2k}}{\eta(2k-1)}\right).$$

Since $k\geq 1$, we have $\frac{1}{2k-1}=\frac{2k}{2k-1}\frac{1}{2k}\leq 2\frac{1}{2k}$, so the above is at most

$$\begin{aligned}
&(\eta b)^2(1+\eta)^{2T}\left(1+(\eta b)^2\right)^T\sum_{k=1}^{T}\left((1+\eta)^{-2k}+\frac{4}{\eta}\frac{\left(1-\frac{1}{2}\eta\epsilon\right)^{2k}}{2k}\right)\\
&\leq\ (\eta b)^2(1+\eta)^{2T}\left(1+(\eta b)^2\right)^T\left(\sum_{k=1}^{\infty}(1+\eta)^{-2k}+\frac{4}{\eta}\sum_{k=1}^{\infty}\frac{\left(1-\frac{1}{2}\eta\epsilon\right)^{k}}{k}\right)\\
&=(\eta b)^2(1+\eta)^{2T}\left(1+(\eta b)^2\right)^T\left(\frac{1}{(1+\eta)^2-1}-\frac{4}{\eta}\log\left(\frac{1}{2}\eta\epsilon\right)\right)\\
&\leq(\eta b)^2(1+\eta)^{2T}\left(1+(\eta b)^2\right)^T\left(\frac{1}{2\eta}+\frac{4}{\eta}\log\left(\frac{2}{\eta\epsilon}\right)\right)\\
&=\eta b^2(1+\eta)^{2T}\left(1+(\eta b)^2\right)^T\left(\frac{1}{2}+4\log\left(\frac{2}{\eta\epsilon}\right)\right).
\end{aligned}$$

Recalling that this is an upper bound on the second term in Eq. (6), and combining with the upper bound in Eq. (7) on the first term, we get overall a bound of

$$(1+\eta)^{2T}\left(-\frac{\epsilon}{p}+(1+\eta)^{-2T}+\frac{\left(1-\frac{1}{2}\eta\epsilon\right)^{2T}}{\eta T}+\eta b^2\left(1+(\eta b)^2\right)^T\left(\frac{1}{2}+4\log\left(\frac{2}{\eta\epsilon}\right)\right)\right). \tag{8}$$

We now argue that under suitable choices of η,ϵ, the expression above is $-\Omega((1+\eta)^{2T}(\epsilon/p)$. For example, this is satisfied if $\eta=\frac{1}{b\sqrt{pT}}$, and we pick $\epsilon=\frac{c\log(T)b\sqrt{p}}{\sqrt{T}}$ for some sufficiently large constant c. Under these choices, the expression inside the main parentheses above becomes

$$-c\frac{\log(T)b}{\sqrt{pT}}+\left(1+\frac{1}{b\sqrt{pT}}\right)^{-2T}+b\sqrt{\frac{p}{T}}\left(1-\frac{c\log(T)}{2T}\right)^{2T}+\frac{b}{\sqrt{pT}}\left(1+\frac{1}{pT}\right)^{T}\left(\frac{1}{2}+4\log\left(\frac{2T}{c\log(T)}\right)\right).$$

Using the facts that $(1-a/t)^t\leq\exp(-a)$ for all positive t,a such that $a/t<1$, and that $c\log(T)/2T<1$ by the assumption that $\epsilon\leq 1$, the above is at most

$$\begin{aligned}
&-c\frac{\log(T)b}{\sqrt{pT}}+\frac{b}{\sqrt{pT}}\left(p\exp(-c\log(T))+\exp(1/p)\left(\frac{1}{2}+4\log\left(\frac{2T}{c\log(T)}\right)\right)\right)+\left(1+\frac{1}{b\sqrt{pT}}\right)^{-2T}\\
&=\ c\frac{\log(T)b}{\sqrt{pT}}\left(-1+\frac{p}{c\log(T)T^c}+\frac{\exp(1/p)}{c\log(T)}\left(\frac{1}{2}+4\log\left(\frac{2T}{c\log(T)}\right)\right)\right)+\left(1+\frac{1}{b\sqrt{pT}}\right)^{-2T}.
\end{aligned}$$

Note that $p,b\geq 1$ by assumption, and that we can assume $T\geq p$ (by the assumption that $\epsilon\leq 1$). Therefore, picking c sufficiently large ensures that the above is at most

$$c\frac{\log(T)b}{\sqrt{pT}}\left(-\frac{1}{2}\right)+\left(1+\frac{1}{b\sqrt{pT}}\right)^{-2T}.$$

The second term is exponentially small in T, and in particular can be verified to be less than $\frac{1}{4}c\frac{\log(T)b}{\sqrt{pT}}$ in the regime where $\epsilon=c\frac{\log(T)b\sqrt{p}}{\sqrt{T}}$ is at most 1 (assuming c is large enough). Overall, we get a bound of $-c\frac{\log(T)b}{\sqrt{pT}}\cdot\frac{1}{4}=-\frac{\epsilon}{4p}$. Plugging this back into Eq. (8), the result follows.

A.3. Proof of Lemma 4

Since $I - A$ is a positive semidefinite matrix, we have

$$V_T = \mathbf{w}_T^\top((1-\epsilon)I - A)\mathbf{w}_T \geq -\epsilon\|\mathbf{w}_T\|^2.$$

Thus, it is sufficient to prove that

$$\|\mathbf{w}_T\|^2 < \exp\left(\eta b\sqrt{T\log(1/\delta)} + (b^2+3)T\eta^2\right)(1+\eta)^{2T}. \tag{9}$$

The proof goes through a martingale argument. We have

$$\begin{aligned}\log(\|\mathbf{w}_T\|^2) &= \log\left(\prod_{t=0}^{T-1}\frac{\|\mathbf{w}_{t+1}\|^2}{\|\mathbf{w}_t\|^2}\right)\\ &= \sum_{t=0}^{T-1}\log\left(\frac{\|\mathbf{w}_{t+1}\|^2}{\|\mathbf{w}_t\|^2}\right)\\ &= \sum_{t=0}^{T-1}\log\left(\frac{\|(I+\eta\tilde{A}_t)\mathbf{w}_t\|^2}{\|\mathbf{w}_t\|^2}\right)\\ &= \sum_{t=0}^{T-1}\log\left(1+\left(\frac{\|(I+\eta\tilde{A}_t)\mathbf{w}_t\|^2}{\|\mathbf{w}_t\|^2}-1\right)\right).\end{aligned}$$

Note that since $\tilde{A}_t$ is positive semidefinite, we always have $(1+\eta b)\|\mathbf{w}_t\|^2 \geq \|(I+\eta\tilde{A}_t)\mathbf{w}_t\|^2 \geq \|\mathbf{w}_t\|^2$, and therefore each summand is of the form $\log(1+a_t)$ where $a_t \in [0,\eta b]$. Using the identity $\log(1+a) \leq a$ for any non-negative a, we can upper bound the above by

$$\sum_{t=0}^{T-1}\left(\frac{\|(I+\eta\tilde{A}_t)\mathbf{w}_t\|^2}{\|\mathbf{w}_t\|^2}-1\right). \tag{10}$$

Based on the preceding discussion, this is a sum of random variables bounded in $[0,\eta b]$, and the expectation of the t-th summand over $\tilde{A}_t$, conditioned on $\tilde{A}_1,\ldots,\tilde{A}_{t-1}$, equals

$$\begin{aligned}&\frac{\mathbf{w}_t^\top\mathbb{E}\left[(I+\eta\tilde{A}_t)^\top(I+\eta\tilde{A}_t)\right]\mathbf{w}_t}{\|\mathbf{w}_t\|^2}-1\\ =\ &\frac{\mathbf{w}_t^\top\left((I+\eta A)^2+\eta^2\left(\tilde{A}_t^\top\tilde{A}_t - A^2\right)\right)\mathbf{w}_t}{\|\mathbf{w}_t\|^2}-1\\ \leq\ &\frac{\mathbf{w}_t^\top(I+\eta A)^2\mathbf{w}_t}{\|\mathbf{w}_t\|^2}+\eta^2\frac{\mathbf{w}_t^\top\tilde{A}_t^\top\tilde{A}_t\mathbf{w}_t}{\|\mathbf{w}_t\|^2}-1\\ \leq\ &\|(I+\eta A)^2\|+\eta^2\|\tilde{A}_t^\top\tilde{A}_t\|-1\\ \leq\ &(1+\eta)^2+\eta^2\|\tilde{A}_t\|^2-1\\ \leq\ &2\eta+(b^2+1)\eta^2.\end{aligned}$$

Using Azuma's inequality, it follows that with probability at least $1-\delta$, Eq. (10) is at most

$$T\left(2\eta+(b^2+1)\eta^2\right)+\eta b\sqrt{T\log(1/\delta)}.$$

Combining the observations above, and the fact that $\log(1+z) \geq z - z^2$ for any $z \geq 0$, we get that with probability at least $1-\delta$,

$$\begin{aligned}\log(\|\mathbf{w}_T\|^2) &< 2T\eta+(b^2+1)T\eta^2+\eta b\sqrt{T\log(1/\delta)}\\ &= \eta b\sqrt{T\log(1/\delta)}+(b^2+3)T\eta^2+2T(\eta-\eta^2)\\ &\leq \eta b\sqrt{T\log(1/\delta)}+(b^2+3)T\eta^2+2T\log(1+\eta),\end{aligned}$$

and therefore

$$\|\mathbf{w}_T\|^2 < \exp\left(\eta b\sqrt{T\log(1/\delta)} + (b^2+3)T\eta^2\right)(1+\eta)^{2T},$$

which establishes Eq. (9) and proves the lemma.

A.4. Proof of Lemma 5

Inverting the bound in the lemma, we have that for any $z \in [1,\infty)$,

$$\Pr(X \geq z) \leq \exp(-(\log(z)/\beta)^2).$$

Now, let $r_2 > r_1 > 0$, be parameters to be chosen later. We have

$$\begin{aligned}\mathbb{E}[X] = \int_{z=0}^{\infty}\Pr(X>z)dz &= \int_{z=0}^{r_1}\Pr(X>z)dz + \int_{z=r_1}^{r_2}\Pr(X>z)dz + \int_{z=r_2}^{\infty}\Pr(X>z)dz \\ &\leq r_1 + (r_2-r_1)\Pr(X>r_1) + \int_{z=r_2}^{\infty}\exp(-(\log(z)/\beta)^2)dz \qquad (11)\end{aligned}$$

Performing the variable change $y = (\log(z)/\beta)^2$ (which implies $z = \exp(\beta\sqrt{y})$ and $dy = \frac{2\sqrt{y}}{\exp(\beta\sqrt{y})}dz$), we get

$$\begin{aligned}\int_{z=r_2}^{\infty}\exp(-(\log(z)/\beta)^2)dz &= \int_{y=\left(\frac{\log(r_2)}{\beta}\right)^2}^{\infty}\frac{1}{2\sqrt{y}}\exp(\beta\sqrt{y}-y)dy \\ &\leq \frac{\beta}{2\log(r_2)}\int_{y=\left(\frac{\log(r_2)}{\beta}\right)^2}^{\infty}\exp(\beta\sqrt{y}-y)dy.\end{aligned}$$

Suppose that we choose $r_2 \geq \exp(2\beta^2)$. Then $\frac{\log(r_2)}{2\beta} \geq \beta$, which implies that for any y in the integral above, $\frac{1}{2}\sqrt{y} \geq \beta$, and therefore $\beta\sqrt{y} - y \leq \frac{1}{2}y - y = -\frac{1}{2}y$. As a result, we can upper bound the above by

$$\frac{\beta}{2\log(r_2)}\int_{y=\left(\frac{\log(r_2)}{\beta}\right)^2}^{\infty}\exp\left(-\frac{1}{2}y\right)dy \;=\; \frac{\beta}{\log(r_2)}\exp\left(-\frac{\log^2(r_2)}{2\beta^2}\right).$$

Plugging this upper bound back into Eq. (11), extracting $\Pr(X > r_1)$, and using the assumption $\mathbb{E}[X] \geq \alpha$, we get that

$$\Pr(X > r_1) \;\geq\; \frac{\alpha - r_1 - \frac{\beta}{\log(r_2)}\exp\left(-\frac{\log^2(r_2)}{2\beta^2}\right)}{r_2 - r_1}.$$

Choosing $r_1 = \alpha/2$ and $r_2 = \exp(2)$ (which ensures $r_2 \geq \exp(2\beta^2)$ as assumed earlier, since $\beta \leq 1$), we get

$$\Pr\left(X > \frac{\alpha}{2}\right) \;\geq\; \frac{\alpha - \beta\exp\left(-\frac{2}{\beta^2}\right)}{2\exp(2) - \alpha}.$$

Since $\beta, \alpha \leq 1$, and $2\exp(2) < 15$, this can be simplified to

$$\Pr\left(X > \frac{\alpha}{2}\right) \;\geq\; \frac{\alpha - \exp\left(-\frac{2}{\beta^2}\right)}{15}.$$

A.5. Proof of Thm. 2

The proof is very similar to that of Thm. 1, using some of the same lemmas, and other lemmas having slight differences to take advantage of the eigengap assumption. Below, we focus on the differences, referring to parts of the proof of Thm. 1 where necessary.

First, as in the proof of Thm. 1, we assume that we work in a coordinate system where A is diagonal, $A = \text{diag}(s_1,\ldots,s_d)$, where $s_1 \geq s_2 \geq \ldots \geq s_d \geq 0$, and s_1 is the eigenvalue corresponding to $\mathbf{v}$. By the eigengap assumption, we can assume that $s_2,\ldots,s_d$ are all at most $1-\lambda$ for some strictly positive $\lambda \in (0,1]$. Under these assumptions, the theorem's conditions reduce to:

- $\frac{1}{w_{0,1}^2} \leq p$, for some $p \geq 8$
- $b \geq 1$ is an upper bound on $\|\tilde{A}_t\|, \|\tilde{A}_t - A\|$,

and as in the proof of Thm. 1, it is enough to lower bound $\Pr(V_T \leq 0)$ where

$$V_T = \mathbf{w}_T^\top((1-\epsilon)I - A)\mathbf{w}_T.$$

We begin by a technical lemma, which bounds a certain quantity appearing later in the proofs:

Lemma 6. *Under the conditions of Thm. 2,*

$$\frac{\log^2(T)b^2}{\lambda^2 T} \leq \frac{1}{p} \leq 1.$$

Proof. By the assumption $\frac{\log^2(T)b^2 p}{\lambda T} \leq \frac{\log(T)b\sqrt{p}}{\sqrt{T}}$, it follows that $\frac{\log(T)b}{\lambda\sqrt{T}} \leq \frac{1}{\sqrt{p}}$, and the result follows by squaring both sides. □

We now continue by presenting the following variant of Lemma 3:

Lemma 7. *Under the conditions of Thm. 2, if we pick $\eta = \frac{\log(T)}{\lambda T} \leq 1$ and $\epsilon = c\frac{\log^2(T)b^2 p}{\lambda T}$ for some sufficiently large numerical constant c, then*

$$\mathbb{E}[V_T] \;\leq\; -(1+\eta)^{2T}\frac{\epsilon}{4p}.$$

Proof. By the exact same proof as in Lemma 3 (up till Eq. (6)), we have

$$\begin{aligned}\mathbb{E}[V_T] &= \mathbb{E}[\mathbf{w}_T^\top((1-\epsilon)I - A)\mathbf{w}_T]\\ &\leq \mathbf{w}_0(I+\eta A)^{2T}((1-\epsilon)I - A)\mathbf{w}_0\\ &\quad + (\eta b)^2\sum_{k=0}^{T-1}\left((1+\eta)^2 + (\eta b)^2\right)^k \lambda_{\max}\left((I+\eta A)^{2(T-k-1)}((1-\epsilon)I - A)\right) \end{aligned} \tag{12}$$

Recalling that $A = \mathrm{diag}(s_1, \ldots, s_d)$ with $s_1 = 1$, that $\|\mathbf{w}_0\|^2 = \sum_{j=1}^d w_{0,j}^2 = 1$, and that $w_{0,1}^2 \geq \frac{1}{p}$, the first term in Eq. (6) equals

$$\begin{aligned}\mathbf{w}_0(I+\eta A)^{2T}((1-\epsilon)I - A)\mathbf{w}_0 &= \sum_{j=1}^d (1+\eta s_j)^{2T}(1-\epsilon-s_j)w_{0,j}^2\\ &= (1+\eta)(-\epsilon)w_{0,1}^2 + \sum_{j=2}^d (1+\eta s_j)^{2T}(1-\epsilon-s_j)w_{0,j}^2\\ &\leq -(1+\eta)^{2T}\frac{\epsilon}{p} + \max_{s\in[0,1-\lambda]}(1+\eta s)^{2T}(1-\epsilon-s)\\ &\leq -(1+\eta)^{2T}\frac{\epsilon}{p} + (1+\eta(1-\lambda))^{2T}\\ &\leq (1+\eta)^{2T}\left(-\frac{\epsilon}{p} + \left(1 - \frac{\eta\lambda}{1+\eta}\right)^{2T}\right)\\ &\leq (1+\eta)^{2T}\left(-\frac{\epsilon}{p} + \left(1 - \frac{\eta\lambda}{2}\right)^{2T}\right), \end{aligned} \tag{13}$$

where we used the assumption that $\eta \leq 1$. As to the second term in Eq. (12), upper bounding it in exactly the same way as in the proof of Lemma 3 (without using the eigengap assumption), we get an upper bound of

$$\eta b^2(1+\eta)^{2T}\left(1+(\eta b)^2\right)^T\left(\frac{1}{2} + 4\log\left(\frac{2}{\eta\epsilon}\right)\right).$$

Combining this with Eq. (13), and plugging back to Eq. (12), we get that

$$\mathbb{E}[V_T] \;\leq\; (1+\eta)^{2T}\left(-\frac{\epsilon}{p}+\left(1-\frac{\eta\lambda}{2}\right)^{2T}+\eta b^2\left(1+(\eta b)^2\right)^T\left(\frac{1}{2}+4\log\left(\frac{2}{\eta\epsilon}\right)\right)\right). \tag{14}$$

Picking $\eta=\frac{\log(T)}{\lambda T}$, and $\epsilon=\frac{c\log^2(T)b^2p}{\lambda T}$ for some constant $c\geq 2$, the above equals

$$(1+\eta)^{2T}\left(-\frac{c\log^2(T)b^2}{\lambda T}+\left(1-\frac{\log(T)}{2T}\right)^{2T}+\frac{b^2\log(T)}{\lambda T}\left(1+\frac{b^2\log^2(T)}{\lambda^2T^2}\right)^T\left(\frac{1}{2}+4\log\left(\frac{2\lambda^2T^2}{c\log^3(T)b^2p}\right)\right)\right).$$

Using the facts that $(1+a/t)^t\leq\exp(a)$ for all positive t,a, that $c\log^3(T)b^2p\geq 2$, and that $\lambda\leq 1$, the above is at most

$$(1+\eta)^{2T}\left(-\frac{c\log^2(T)b^2}{\lambda T}+\frac{1}{T}+\frac{b^2\log(T)}{\lambda T}\exp\left(\frac{b^2\log^2(T)}{\lambda^2T}\right)\left(\frac{1}{2}+4\log\left(T^2\right)\right)\right).$$

By Lemma 6, $\frac{b^2\log^2(T)}{\lambda^2T}\leq 1$, so the above is at most

$$\begin{aligned}&(1+\eta)^{2T}\left(-\frac{c\log^2(T)b^2}{\lambda T}+\frac{1}{T}+\frac{b^2\log(T)}{\lambda T}\exp(1)\left(\frac{1}{2}+8\log\left(T\right)\right)\right)\\&\leq(1+\eta)^{2T}\frac{b^2\log^2(T)}{\lambda T}\left(-c+\frac{\lambda}{b^2\log^2(T)}+\exp(1)\left(\frac{1}{2\log(T)}+8\right)\right).\end{aligned}$$

Clearly, for large enough c, the expression in the main parenthesis above is at most $-c/4$, so we get an upper bound of

$$-(1+\eta)^{2T}\frac{cb^2\log^2(T)}{4\lambda T}\;=\;-(1+\eta)^{2T}\frac{\epsilon}{4p},$$

from which the result follows. □

Rather similar to the proof of Thm. 1, we now define the non-negative random variable

$$R_T=\max\left\{0,-\frac{V_T}{\exp((b^2+3)T\eta^2)(1+\eta)^{2T}\epsilon}\right\}.$$

By Lemma 7,

$$\mathbb{E}[R_T]\;\geq\;\mathbb{E}\left[-\frac{V_T}{\exp((b^2+3)T\eta^2)(1+\eta)^{2T}\epsilon}\right]\;\geq\;\frac{1}{4p\exp((b^2+3)T\eta^2)},$$

and by Lemma 4,

$$\Pr\left(R_T\geq\exp\left(\eta b\sqrt{T\log(1/\delta)}\right)\right)\leq\delta.$$

Therefore, applying Lemma 5 on R_T, with $\alpha=\frac{1}{4p\exp((b^2+3)T\eta^2)}$ (which is in $[0,1]$) and with $\beta=\eta b\sqrt{T}$ (which can be verified to be in $[0,1]$ by the fact that $\eta=\frac{\log(T)}{\lambda T}$ and Lemma 6), we get that

$$\Pr\left(R_T>\frac{1}{8p\exp((b^2+3)T\eta^2)}\right)\;\geq\;\frac{1}{15}\left(\frac{1}{4p\exp((b^2+3)T\eta^2)}-\exp\left(-\frac{2}{\eta^2b^2T}\right)\right). \tag{15}$$

By definition of R_T, the left hand side of this inequality is at most

$$\begin{aligned}&=\;\Pr\left(\max\left\{0,-\frac{V_T}{\exp((b^2+3)T\eta^2)(1+\eta)^{2T}\epsilon}\right\}>\frac{1}{8p\exp((b^2+3)T\eta^2)}\right)\\&=\;\Pr\left(-\frac{V_T}{\exp((b^2+3)T\eta^2)(1+\eta)^{2T}\epsilon}>\frac{1}{8p\exp((b^2+3)T\eta^2)}\right)\\&=\;\Pr\left(V_T\leq-\frac{(1+\eta)^{2T}\epsilon}{8p}\right)\\&\leq\;\Pr\left(V_T\leq 0\right),\end{aligned}$$

and the right hand side of Eq. (15) (by definition of η, the assumption $b \geq 1$, and Lemma 6) equals

$$\begin{aligned}
&\frac{1}{15}\left(\frac{1}{4p\exp\left(\frac{(b^2+3)\log^2(T)}{\lambda^2 T}\right)} - \exp\left(-\frac{2\lambda^2 T}{b^2\log^2(T)}\right)\right)\\
\geq\ &\frac{1}{15}\left(\frac{1}{4p\exp\left(\frac{4b^2\log^2(T)}{\lambda^2 T}\right)} - \frac{1}{\exp\left(2\frac{\lambda^2 T}{b^2\log^2(T)}\right)}\right)\\
\geq\ &\frac{1}{15}\left(\frac{1}{4p\exp\left(\frac{4}{p}\right)} - \frac{1}{\exp\left(2p\right))}\right),
\end{aligned}$$

which can be verified to be at least $\frac{1}{100p}$ for any $p \geq 8$. Plugging these bounds back to Eq. (15), we obtained

$$\Pr(V_T \leq 0) \geq \frac{1}{100p}.$$

By definition of V_T, $V_T \leq 0$ implies that

$$\frac{\mathbf{w}_T(I-A)\mathbf{w}_T}{\|\mathbf{w}_T\|^2} \leq \epsilon,$$

where $\epsilon = c\frac{\log^2(T)b^2 p}{\lambda T}$ is the value chosen in Lemma 7, and the theorem is established.