A Supplementary Material

In the proofs below, for $\alpha, \beta \geq 0$, we let $Y_t(A,\beta)$ and $Z_t(A,\beta)$ be matrix polynomials such that

$$Y_t(A,\beta) = 2AY_{t-1}(A,\beta) - \beta Y_{t-2}(A,\beta),\ t \geq 2,\ Y_1(A,\beta) = A,\ Y_0(A,\beta) = I, \tag{20}$$
$$Z_t(A,\beta) = 2AZ_{t-1}(A,\beta) - \beta Z_{t-2}(A,\beta),\ t \geq 2,\ Z_1(A,\beta) = 2A,\ Z_0(A,\beta) = I. \tag{21}$$

and let $y_t(\alpha,\beta)$ and $z_t(\alpha,\beta)$ be recurrence polynomials such that

$$y_t(\alpha,\beta) = \sqrt{\alpha} y_{t-1}(\alpha,\beta) - \beta y_{t-2}(\alpha,\beta),\ t \geq 2,\ y_1(\alpha,\beta) = \frac{\sqrt{\alpha}}{2},\ y_0(\alpha,\beta) = 1, \tag{22}$$
$$z_t(\alpha,\beta) = \sqrt{\alpha} z_{t-1}(\alpha,\beta) - \beta z_{t-2}(\alpha,\beta),\ t \geq 2,\ z_1(\alpha,\beta) = \sqrt{\alpha},\ z_0(\alpha,\beta) = 1. \tag{23}$$

For a sequence of matrices $B_0, B_1, B_2, \cdots$, let

$$\prod_{i=j}^{k} B_i = \begin{cases} B_j B_{j-1} \cdots B_k & \text{if } j \geq k \\ I, & \text{otherwise} \end{cases}.$$

Since the eigenvectors $u_1, u_2, \ldots, u_d$ form an orthogonal basis, we frequently use the fact that for $w \in \mathbb{R}^d$, we have $\|w\|^2 = \sum_{k=1}^d (u_k^T w)^2$.

A.1 Main Results

Lemma A.1. *For $w_0 \in \mathbb{R}^d$, let $w = w_0/\|w_0\|$. Then, for $t \geq 0$, we have*

$$\|P\left[(1-\eta)I + \eta C\right]^t w\|^2 \leq 2(1-\eta+\eta\lambda_1)^{2t}(1-(u_1^T w)^2), \tag{24a}$$
$$\|PY_t((1-\eta)I + \eta C, \beta(\eta))w\|^2 \leq 4(1-(u_1^T w)^2) p_t(\alpha_1(\eta), \beta(\eta)), \tag{24b}$$
$$\|Z_t((1-\eta)I + \eta C, \beta(\eta))\|^2 \leq q_t(\alpha_1(\eta), \beta(\eta)). \tag{24c}$$

Proof. Since $u_1, u_2, \cdots, u_d$ forms an orthogonal basis in $\mathbb{R}^d$, we can write $w = \sum_{k=1}^d (u_k^T w) u_k$. From that (λ_k, u_k) are eigenpairs of C, we have

$$\left[(1-\eta)I + \eta C\right]^t w = \sum_{k=1}^d (u_k^T w)(1-\eta+\eta\lambda_k)^t u_k. \tag{25}$$

From the definition of w and P in (6), we have $P = I - ww^T$. Since

$$\begin{aligned}
\|P\left[(1-\eta)I + \eta C\right]^t w\|^2 &= w^T \left[(1-\eta)I + \eta C\right]^t P^2 \left[(1-\eta)I + \eta C\right]^t w \\
&= w^T \left[(1-\eta)I + \eta C\right]^t P \left[(1-\eta)I + \eta C\right]^t w \\
&= w^T \left[(1-\eta)I + \eta C\right]^t (I - ww^T) \left[(1-\eta)I + \eta C\right]^t w \\
&= \|\left[(1-\eta)I + \eta C\right]^t w\|^2 - \left(w^T \left[(1-\eta)I + \eta C\right]^t w\right)^2,
\end{aligned}$$

using (25), we have

$$\begin{aligned}
\|P\left[(1-\eta)I + \eta C\right]^t w\|^2 &= \sum_{k=1}^d (u_k^T w)^2 (1-\eta+\eta\lambda_k)^{2t} - \left(\sum_{k=1}^d (u_k^T w)^2 (1-\eta+\eta\lambda_k)^t\right)^2 \\
&\leq (1-\eta+\eta\lambda_1)^{2t} - (u_1^T w)^4 (1-\eta+\eta\lambda_1)^{2t} \\
&\leq 2(1-(u_1^T w)^2)(1-\eta+\eta\lambda_1)^{2t}
\end{aligned}$$

where the last inequality follows from

$$1-(u_1^T w)^4 = \left(1+(u_1^T w)^2\right)\left(1-(u_1^T w)^2\right) \leq 2\left(1-(u_1^T w)^2\right). \tag{26}$$

To prove (24b), we first show that

$$Y_t((1-\eta)I+\eta C,\beta(\eta))u_k = y_t(\alpha_k(\eta),\beta(\eta))u_k. \tag{27}$$

First, consider the cases when $t=0$ and $t=1$. For $t=0$, we have $Y_0((1-\eta)I+\eta C,\beta(\eta))u_k = y_0(\alpha_k(\eta),\beta(\eta))u_k$. For $t=1$, it follows that

$$Y_1((1-\eta)I+\eta C,\beta(\eta))u_k = ((1-\eta)I+\eta C)u_k = (1-\eta+\eta\lambda_k)u_k = \frac{\sqrt{\alpha_k(\eta)}}{2}u_k = y_1(\alpha_k(\eta),\beta(\eta))u_k.$$

Suppose that (27) holds for $t-1$ and $t-2$. Using the definition of Y_t in (20), we have

$$\begin{aligned}
Y_t((1-\eta)I+\eta C,\beta(\eta))u_k &= \left[2((1-\eta)I+\eta C)Y_{t-1}((1-\eta)I+\eta C,\beta(\eta)) - \beta(\eta)Y_{t-2}((1-\eta)I+\eta C,\beta(\eta))\right]u_k\\
&= \left[2(1-\eta+\eta\lambda_k)y_{t-1}(\alpha_k(\eta),\beta(\eta)) - \beta(\eta)y_{t-2}(\alpha_k(\eta),\beta(\eta))\right]u_k\\
&= \left[\sqrt{\alpha_k(\eta)}y_{t-1}(\alpha_k(\eta),\beta(\eta)) - \beta(\eta)y_{t-2}(\alpha_k(\eta),\beta(\eta))\right]u_k\\
&= y_t(\alpha_k(\eta),\beta(\eta))u_k.
\end{aligned}$$

This completes the proof of (27).

Next, we show that

$$(y_t(\alpha_k(\eta),\beta(\eta))^2 = p_t(\alpha_k(\eta),\beta(\eta)). \tag{28}$$

For the base cases, we have

$$(y_0(\alpha_k(\eta),\beta(\eta))^2 = 1 = p_0(\alpha_k(\eta),\beta(\eta)), \quad (y_1(\alpha_k(\eta),\beta(\eta))^2 = \frac{\alpha_k}{4} = p_1(\alpha_k(\eta),\beta(\eta))$$

and

$$(y_2(\alpha_k(\eta),\beta(\eta))^2 = \left(\sqrt{\alpha_k(\eta)}y_1(\alpha_k(\eta),\beta(\eta)) - \beta(\eta)y_0(\alpha_k(\eta),\beta(\eta))\right)^2 = \left(\frac{\alpha(\eta)}{2}-\beta(\eta)\right)^2 = p_2(\alpha_k(\eta),\beta(\eta)).$$

Using the definition of y_t in (22) for t and $t-1$, we have

$$\begin{aligned}
(y_t(\alpha_k(\eta),\beta(\eta)))^2 &= (\sqrt{\alpha_k(\eta)}y_{t-1}(\alpha_k(\eta),\beta(\eta)) - \beta(\eta)y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2\\
&= \alpha_k(\eta)(y_{t-1}(\alpha_k(\eta),\beta(\eta)))^2 - 2\sqrt{\alpha_k(\eta)}\beta(\eta)y_{t-1}(\alpha_k(\eta),\beta(\eta))y_{t-2}(\alpha_k(\eta),\beta(\eta))\\
&\quad + \beta(\eta)^2(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2
\end{aligned}$$

and

$$\begin{aligned}
(y_{t-1}(\alpha_k(\eta),\beta(\eta)))^2 &= \alpha_k(\eta)(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2 - 2\sqrt{\alpha_k(\eta)}\beta(\eta)y_{t-2}(\alpha_k(\eta),\beta(\eta))y_{t-3}(\alpha_k(\eta),\beta(\eta))\\
&\quad + \beta(\eta)^2(y_{t-3}(\alpha_k(\eta),\beta(\eta)))^2.
\end{aligned}$$

Moreover, since

$$y_{t-1}(\alpha_k(\eta),\beta(\eta))y_{t-2}(\alpha_k(\eta),\beta(\eta)) = \sqrt{\alpha_k(\eta)}(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2 - \beta(\eta)y_{t-2}(\alpha_k(\eta),\beta(\eta))y_{t-3}(\alpha_k(\eta),\beta(\eta)),$$

we have

$$\begin{aligned}
(y_t(\alpha_k(\eta),\beta(\eta)))^2 &= \alpha_k(\eta)(y_{t-1}(\alpha_k(\eta),\beta(\eta)))^2 - 2\alpha_k(\eta)\beta(\eta)(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2 + \beta(\eta)^2(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2\\
&\quad + 2\sqrt{\alpha_k(\eta)}\beta(\eta)^2 y_{t-2}(\alpha_k(\eta),\beta(\eta))y_{t-3}(\alpha_k(\eta),\beta(\eta))\\
&= \alpha_k(\eta)(y_{t-1}(\alpha_k(\eta),\beta(\eta)))^2 - 2\alpha_k(\eta)\beta(\eta)(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2 + \beta(\eta)^2(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2\\
&\quad + \beta(\eta)\left(\alpha_k(\eta)(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2 + \beta(\eta)^2(y_{t-3}(\alpha_k(\eta),\beta(\eta)))^2 - (y_{t-1}(\alpha_k(\eta),\beta(\eta)))^2\right)\\
&= (\alpha_k(\eta)-\beta(\eta))(y_{t-1}(\alpha_k(\eta),\beta(\eta)))^2 - \beta(\eta)(\alpha_k(\eta)-\beta(\eta))(y_{t-2}(\alpha_k(\eta),\beta(\eta)))^2\\
&\quad + \beta(\eta)^3(y_{t-3}(\alpha_k(\eta),\beta(\eta)))^2.
\end{aligned}$$

This proves (28).

Now, using (27), we have

$$Y_t((1-\eta)I+\eta C,\beta(\eta))w=\sum_{k=1}^{d}y_t(\alpha_k(\eta),\beta(\eta))(u_k^Tw)u_k. \tag{29}$$

Since $u_1, u_2, \cdots, u_d$ form an orthogonal basis in $\mathbb{R}^d$, we have

$$\|Y_t((1-\eta)I+\eta C,\beta(\eta))w\|^2=\sum_{k=1}^{d}(y_t(\alpha_k(\eta),\beta(\eta)))^2(u_k^Tw)^2=\sum_{k=1}^{d}p_t(\alpha_k(\eta),\beta(\eta))(u_k^Tw)^2.$$

Using (90) and (92) in Lemma A.4, for $k \geq 2$, we have

$$p_t(\alpha_k(\eta),\beta(\eta))\leq p_t(\alpha_1(\eta),\beta(\eta)) \tag{30}$$

Since $\sum_{k=1}^{d}(u_k^Tw)^2=1$, we have

$$\|Y_t((1-\eta)I+\eta C,\beta(\eta))w\|^2\leq p_t(\alpha_1(\eta),\beta(\eta)).$$

Moreover, using $(u_1^Tw)^2 \leq 1$ and (29), we obtain

$$\begin{aligned}\left(w^TY_t((1-\eta)I+\eta C,\beta(\eta))w\right)^2&=\Big(y_t(\alpha_1(\eta),\beta(\eta))(u_1^Tw)^2+\sum_{k=2}^{d}y_t(\alpha_k(\eta),\beta(\eta))(u_k^Tw)^2\Big)^2\\&\geq(y_t(\alpha_1(\eta),\beta(\eta)))^2(u_1^Tw)^4-2y_t(\alpha_1(\eta),\beta(\eta))\sum_{k=2}^{d}|y_t(\alpha_k(\eta),\beta(\eta))|(u_k^Tw)^2\\&\geq(y_t(\alpha_1(\eta),\beta(\eta)))^2(u_1^Tw)^4-2(y_t(\alpha_1(\eta),\beta(\eta)))^2(1-(u_k^Tw)^2)\end{aligned}$$

Therefore,

$$\begin{aligned}\|PY_t((1-\eta)I+\eta C,\beta(\eta))w\|^2&=\|Y_t((1-\eta)I+\eta C,\beta(\eta))w\|^2-\left(w^TY_t((1-\eta)I+\eta C,\beta(\eta))w\right)^2\\&\leq(y_t(\alpha_1(\eta),\beta(\eta)))^2(1-(u_1^Tw)^4)+2(y_t(\alpha_1(\eta),\beta(\eta)))^2(1-(u_k^Tw)^2)\\&\leq 4(y_t(\alpha_1(\eta),\beta(\eta)))^2(1-(u_k^Tw)^2)\end{aligned}$$

where the last inequality follows from (26).

Lastly, we prove (24c). In the same way we prove (27) and (28), we can show that

$$Z_t((1-\eta)I+\eta C,\beta(\eta))u_k=z_t(\alpha_k(\eta),\beta(\eta))u_k,\quad(z_t(\alpha_k(\eta),\beta(\eta)))^2=q_t(\alpha_k(\eta),\beta(\eta)). \tag{31}$$

Using (91) and (92) in Lemma A.4, for $k \geq 2$, we have

$$q_t(\alpha_k(\eta),\beta(\eta))\leq q_t(\alpha_1(\eta),\beta(\eta)). \tag{32}$$

Using (31), we have

$$w^TZ_t((1-\eta)I+\eta C,\beta(\eta))w=\sum_{k=1}^{d}z_t(\alpha_k(\eta),\beta(\eta))(u_k^Tw)^2\leq\sum_{k=1}^{d}|z_t(\alpha_k(\eta),\beta(\eta))|(u_k^Tw)^2.$$

Moreover, using (32) and the fact that $\sum_{k=1}^{d}(u_k^Tw)^2=1$, we have

$$\sum_{k=1}^{d}|z_t(\alpha_k(\eta),\beta(\eta))|(u_k^Tw)^2\leq|z_t(\alpha_1(\eta),\beta(\eta))|\sum_{k=1}^{d}(u_k^Tw)^2=|z_t(\alpha_1(\eta),\beta(\eta))|.$$

This results in

$$w^TZ_t((1-\eta)I+\eta C,\beta(\eta))w\leq|z_t(\alpha_1(\eta),\beta(\eta))|,$$

leading to

$$\|Z_t((1-\eta)I+\eta C,\beta(\eta))\|^2\leq|z_t(\alpha_1(\eta),\beta(\eta))|^2=q_t(\alpha_1(\eta),\beta(\eta)).$$

This complets the proof. □

A.1.1 VR Power

Proof of Lemma 3.1. Since $Pw_0 = \left(I - w_0 w_0^T\right) w_0 = 0$, we have

$$u_k^T w_1 = (1-\eta) u_k^T w_0 + \eta u_k^T C w_0 + \eta u_k^T (C_0 - C) P w_0 = (1 - \eta + \eta\lambda_k) u_k^T w_0. \tag{33}$$

Taking the expectation of the square of (33), we obtain

$$E[(u_k^T w_1)^2] = (1 - \eta + \eta\lambda_k)^2 E[(u_k^T w_0)^2]. \tag{34}$$

For $t \geq 2$, we have

$$u_k^T w_t = (1 - \eta + \eta\lambda_k) u_k^T w_{t-1} + \eta u_k^T (C_{t-1} - C) P w_{t-1}. \tag{35}$$

Since S_t is sampled uniformly at random, C_t is independent of $S_1, \ldots, S_{t-1}$ and w_0 with $E[C_t] = C$, leading to

$$\begin{aligned} E[u_k^T w_{t-1} u_k^T (C_{t-1} - C) P w_t] &= E[E[u_k^T w_{t-1} u_k^T (C_{t-1} - C) P w_t | w_0, S_1, \ldots, S_{t-2}]] \\ &= E[u_k^T w_{t-1} u_k^T E[C_{t-1} - C] P w_t] = 0. \end{aligned}$$

Therefore, taking the expectation of the square of (35), we have

$$\begin{aligned} E[(u_k^T w_t)^2] &= (1 - \eta + \eta\lambda_k)^2 E[(u_k^T w_{t-1})^2] + \eta^2 E[w_{t-1}^T P (C_{t-1} - C) u_k u_k^T (C_{t-1} - C) P w_{t-1}] \\ &= (1 - \eta + \eta\lambda_k)^2 E[(u_k^T w_{t-1})^2] + \eta^2 E[w_{t-1}^T P M_k P w_{t-1}] \end{aligned} \tag{36}$$

where the last equality follows from

$$\begin{aligned} E[w_{t-1}^T P (C_{t-1} - C) u_k u_k^T (C_{t-1} - C) P w_{t-1}] &= E[E[w_{t-1}^T P (C_{t-1} - C) u_k u_k^T (C_{t-1} - C) P w_{t-1} | w_0, S_1, \ldots, S_{t-2}]] \\ &= E[w_{t-1}^T P E[(C_{t-1} - C) u_k u_k^T (C_{t-1} - C)] P w_{t-1}] \\ &= E[w_{t-1}^T P M_k P w_{t-1}]. \end{aligned}$$

Repeatedly applying (36) and using (34), we obtain

$$E[(u_k^T w_t)^2] = (1 - \eta + \eta\lambda_k)^{2t} E[(u_k^T w_0)^2] + \eta^2 \sum_{i=1}^{t-1} (1 - \eta + \eta\lambda_k)^{2(t-i-1)} E[w_i^T P M_k P w_i].$$

□

Proof of Lemma 3.2. By Lemma A.2, we have

$$\sum_{k=2}^{d} E[w_t^T P M_k P w_t] = \sum_{k=2}^{d} E[w_t^T P M_k P w_t] = E[w_t^T P \sum_{k=2}^{d} M_k P w_t] \leq \| \sum_{k=2}^{d} M_k \| \cdot E[\|P w_t\|^2]. \tag{37}$$

Using the Jensen's inequality and the fact that $\| \sum_{k=2}^{d} u_k u_k^T \| = 1$, we have

$$\| \sum_{k=2}^{d} M_k \| = \| \sum_{k=2}^{d} E[(C_t - C) u_k u_k^T (C_t - C)] \| \leq E[\|C_t - C\|^2] = E[\|(C_t - C)^2\|] = K,$$

resulting in

$$\sum_{k=2}^{d} E[w_t^T P M_k P w_t] \leq K E[\|P w_t\|^2]. \tag{38}$$

Let

$$B_i = (1 - \eta) I + \eta C + \eta (C_i - C) P.$$

Since $Pw_0 = 0$ and

$$\begin{aligned} \prod_{i=t-1}^{0} B_i &= \prod_{i=t-1}^{1} B_i \eta (C_0 - C) P + \prod_{i=t-1}^{1} B_i \left((1 - \eta) I + \eta C\right) \\ &= \prod_{i=t-1}^{1} B_i \eta (C_0 - C) P + \sum_{j=1}^{t-1} \prod_{i=t-1}^{j+1} B_i \eta (C_j - C) P \left[(1 - \eta) I + \eta C\right]^j + \left[(1 - \eta) I + \eta C\right]^t, \end{aligned}$$

which can be seen by elementary manipulation, we have

$$w_t = \prod_{i=t-1}^{0} B_i w_0 = \left[\sum_{j=1}^{t-1} \prod_{i=t-1}^{j+1} B_i \eta (C_j - C) P \left[(1-\eta)I + \eta C\right]^j + \left[(1-\eta)I + \eta C\right]^t \right] w_0,$$

resulting in

$$Pw_t = P \prod_{i=t-1}^{0} B_i w_0 = \left[\sum_{j=1}^{t-1} P \prod_{i=t-1}^{j+1} B_i \eta (C_j - C) P \left[(1-\eta)I + \eta C\right]^j + P\left[(1-\eta)I + \eta C\right]^t \right] w_0. \tag{39}$$

Since $C_0, \cdots, C_{t-1}$ are independent with $E[C_i] = C$ for all $1 \leq i \leq t-1$, we obtain

$$E\Big[w_0^T \left[(1-\eta)I + \eta C\right]^t P^2 \prod_{i=t-1}^{j+1} B_i \eta (C_j - C) P \left[(1-\eta)I + \eta C\right]^j w_0\Big] = 0 \tag{40}$$

$$E\Big[w_0^T \left[(1-\eta)I + \eta C\right]^{j_1} P(C_{j_1} - C)\eta \prod_{i=j_1+1}^{t-1} B_i P^2 \prod_{i=t-1}^{j_2+1} B_i \eta (C_{j_2} - C) P \left[(1-\eta)I + \eta C\right]^{j_2} w_0\Big] = 0 \tag{41}$$

where $1 \leq j, j_1, j_2 \leq t-1$ and $j_1 \neq j_2$. Therefore, we have

$$E[\|Pw_t\|^2] = \sum_{j=1}^{t-1} E\big[\big\|P \prod_{i=t-1}^{j+1} B_i \eta (C_j - C) P \left[(1-\eta)I + \eta C\right]^j w_0\big\|^2\big] + E[\|P\left[(1-\eta)I + \eta C\right]^t w_0\|^2] \tag{42}$$

due to cross-terms being 0 from (40) and (41) when "squaring" (39). Using Lemma A.1 with $w = w_0/\|w_0\|$ and the fact that $\|w_0\|^2(1 - (u_1^T w_0)^2/\|w_0\|^2) = \sum_{k=2}^d (u_k^T w_0)^2$, we have

$$E[\|P[(1-\eta)I + \eta C]^t w_0\|^2] \leq 2(1 - \eta + \eta\lambda_1)^{2t} \sum_{k=2}^{d} E[(u_k^T w_0)^2]. \tag{43}$$

By Lemma A.2 and $\|P\| = 1$, we have

$$\big\|P \prod_{i=t-1}^{j+1} B_i \eta (C_j - C) P \left[(1-\eta)I + \eta C\right]^j w_0\big\|^2 \leq \eta^2 \big\| \prod_{i=t-1}^{j+1} B_i (C_j - C) P \left[(1-\eta)I + \eta C\right]^j w_0\big\|^2. \tag{44}$$

Moreover, by repeatedly using first the property that B_i is independent of $w_0, C_j, B_{j+1}, \cdots, B_{i-1}$ and Lemma A.2, we have

$$\begin{aligned}
&E\big[\big\| \prod_{i=t-1}^{j+1} B_i (C_j - C) P \big[(1-\eta)I + \eta C\big]^j w_0\big\|^2\big] \\
&= E[w_0^T[(1-\eta)I + \eta C]^j P(C_j - C)\Big(\prod_{i=t-2}^{j+1} B_i\Big)^T B_{t-1}^T B_{t-1} \prod_{i=t-2}^{j+1} B_i P(C_j - C)[(1-\eta)I + \eta C]^j w_0] \\
&= E[w_0^T[(1-\eta)I + \eta C]^j P(C_j - C)\Big(\prod_{i=t-2}^{j+1} B_i\Big)^T E[B_{t-1}^T B_{t-1}] \prod_{i=t-2}^{j+1} B_i P(C_j - C)[(1-\eta)I + \eta C]^j w_0] \\
&\leq \|E[B_{t-1}^T B_{t-1}]\| \cdot E\big[\big\| \prod_{i=t-2}^{j+1} B_i (C_j - C) P \big[(1-\eta)I + \eta C\big]^j w_0\big\|^2\big] \\
&\leq \prod_{i=t-1}^{j+1} \big\|E[B_i^T B_i]\big\| \cdot E[\|(C_j - C) P \left[(1-\eta)I + \eta C\right]^j w_0\|^2].
\end{aligned}$$

In the same way, using the fact that C_j is independent of w_0 and Lemma A.2, we have

$$E[\|(C_j - C) P \left[(1-\eta)I + \eta C\right]^j w_0\|^2] \leq \|E[(C_j - C)^2]\| \cdot E[\|P \left[(1-\eta)I + \eta C\right]^j w_0\|^2],$$

resulting in

$$E\big[\big\|\prod_{i=t-1}^{j+1} B_i(C_j - C)P[(1-\eta)I + \eta C]^j w_0\big\|^2\big] \leq \prod_{i=t-1}^{j+1} \|E[B_i^T B_i]\| \cdot \|E[(C_j - C)^2]\| \cdot E[\|P\,[(1-\eta)I + \eta C]^j\, w_0\|^2]. \tag{45}$$

Since C_i is independent of w_0 and $E[C_i] = C$, we have

$$\|E[B_i^T B_i]\| \leq \|\,[(1-\eta)I + \eta C]^2\,\| + \eta^2 \|E[P(C_i - C)^2 P]\|.$$

Since all induced norms are convex, using the Jensen's inequality, we have

$$\|E[P(C_i - C)^2 P]\|] \leq E[\|P(C_i - C)^2 P\|] \leq E[\|(C_i - C)^2\|] = K,$$

leading to

$$\|E[B_i^T B_i]\| \leq \|\,[(1-\eta)I + \eta C]^2\,\| + \eta^2 \|E[P(C_i - C)^2 P]\| \leq (1 - \eta + \eta\lambda_1)^2 + \eta^2 K. \tag{46}$$

In the same way, we obtain

$$\|E[(C_j - C)^2]\| \leq E[\|(C_j - C)^2\|] = K. \tag{47}$$

Using (46), (47) and (43) for (45), we have

$$E\big[\big\|\prod_{i=t-1}^{j+1} B_i(C_j - C)P[(1-\eta)I + \eta C]^j w_0\big\|^2\big] \leq K\left[(1 - \eta + \eta\lambda_1)^2 + \eta^2 K\right]^{t-j-1} (1 - \eta + \eta\lambda_1)^{2j} \sum_{k=2}^{d} E[(u_k^T w_0)^2]. \tag{48}$$

From (42), (43), (44) and (48), we finally have

$$\begin{aligned} E[\|Pw_t\|^2] &\leq 2\left[\sum_{j=1}^{t-1} \eta^2 K\left[(1 - \eta + \eta\lambda_1)^2 + \eta^2 K\right]^{t-j-1} (1 - \eta + \eta\lambda_1)^{2j} + (1 - \eta + \eta\lambda_1)^{2t}\right] \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2] \\ &\leq 2\left[(1 - \eta + \eta\lambda_1)^2 + \eta^2 K\right]^t \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2], \end{aligned}$$

where the last inequality can be checked by elementary manipulation. This results in

$$\sum_{k=2}^{d} E[w_t^T P M_k P w_t] \leq 2K\left[(1 - \eta + \eta\lambda_1)^2 + \eta^2 K\right]^t \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2]. \tag{49}$$

This proves the first part of the proof.

Next, we have

$$\sum_{k=2}^{d}\sum_{i=1}^{t-1} (1 - \eta + \eta\lambda_k)^{2(t-i-1)} E[w_i^T P M_k P w_i] \leq (1 - \eta + \eta\lambda_1)^{2t} \cdot \sum_{i=1}^{t-1} (1 - \eta + \eta\lambda_1)^{-2(i+1)} \sum_{k=2}^{d} E[w_i^T P M_k P w_i]$$

and

$$\begin{aligned} \sum_{i=1}^{t-1} (1 - \eta + \eta\lambda_1)^{-2(i+1)} \left[(1 - \eta + \eta\lambda_1)^2 + \eta^2 K\right]^i &\leq \frac{1}{(1 - \eta + \eta\lambda_1)^2} \sum_{i=1}^{t-1} \left(\frac{(1 - \eta + \eta\lambda_1)^2 + \eta^2 K}{(1 - \eta + \eta\lambda_1)^2}\right)^i \\ &\leq \frac{1}{\eta^2 K}\left[\left(1 + \frac{\eta^2 K}{(1 - \eta + \eta\lambda_1)^2}\right)^{t-1} - 1\right]\left(1 + \frac{\eta^2 K}{(1 - \eta + \eta\lambda_1)^2}\right) \\ &\leq \frac{1}{\eta^2 K}\left[\exp\left(\frac{\eta^2 K t}{(1 - \eta + \eta\lambda_1)^2}\right) - 1\right]. \end{aligned}$$

Using the condition that

$$0 < \frac{\eta^2 K m}{(1-\eta+\eta\lambda_1)^2} < 1$$

and the fact $\exp(x) - 1 \leq 2x$ for all $x \in (0,1)$, we further obtain

$$\sum_{i=1}^{t-1} (1-\eta+\eta\lambda_1)^{-2(i+1)} \left[(1-\eta+\eta\lambda_1)^2 + \eta^2 K\right]^i \leq \frac{2t}{(1-\eta+\eta\lambda_1)^2}.$$

Combined with (49), this results in

$$\begin{aligned}\eta^2 \sum_{k=2}^{d} \sum_{i=1}^{m-1} (1-\eta+\eta\lambda_k)^{2(m-i-1)} E[w_i^T P M_k P w_i] &\leq \eta^2 \sum_{i=1}^{m-1} (1-\eta+\eta\lambda_k)^{2(m-i-1)} \sum_{k=2}^{d} E[w_i^T P M_k P w_i] \\ &\leq 4\eta^2 K m (1-\eta+\eta\lambda_1)^{2(m-1)} \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2].\end{aligned}$$

Using Lemma 3.1 for $t = m$ and the fact that $(1-\eta+\eta\lambda_k)^{2m} \leq (1-\eta+\eta\lambda_2)^{2m}$ for $k \geq 2$, we finally have

$$\begin{aligned}\sum_{k=2}^{d} E[(u_k^T w_m)^2] &= \sum_{k=2}^{d} (1-\eta+\eta\lambda_k)^{2m} E[(u_k^T w_0)^2] + \eta^2 \sum_{k=2}^{d} \sum_{i=1}^{m-1} (1-\eta+\eta\lambda_k)^{2(m-i-1)} E[w_i^T P M_k P w_i] \\ &\leq \left((1-\eta+\eta\lambda_2)^{2m} + 4\eta^2 K m (1-\eta+\eta\lambda_1)^{2(m-1)}\right) \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2]. \end{aligned} \tag{50}$$

On the other hand, by Lemma 3.1 and the fact that PM_kP is positive semi-definite, we have

$$(1-\eta+\eta\lambda_1)^{2m} E[(u_1^T w_0)^2] \leq E[(u_1^T w_m)^2]. \tag{51}$$

Combining (51) with (50), we obtain

$$\frac{\sum_{k=2}^{d} E[(u_k^T w_m)^2]}{E[(u_1^T w_m)^2]} \leq \left[\left(\frac{1-\eta+\eta\lambda_2}{1-\eta+\eta\lambda_1}\right)^{2m} + \frac{4\eta^2 K m}{(1-\eta+\eta\lambda_1)^2}\right] \cdot \frac{\sum_{k=2}^{d} E[(u_k^T w_0)^2]}{E[(u_1^T w_0)^2]}.$$

□

Proof of Lemma 3.3. From the conditions on η, m and $|S|$, we have

$$0 < \frac{\eta^2 K m}{(1-\eta+\eta\lambda_1)^2} < \frac{1}{16}.$$

Therefore, using Lemma 3.2, we have

$$\frac{\sum_{k=2}^{d} E[(u_k^T w_m)^2]}{E[(u_1^T w_m)^2]} \leq \left[\left(\frac{1-\eta+\eta\lambda_2}{1-\eta+\eta\lambda_1}\right)^{2m} + \frac{4\eta^2 K m}{(1-\eta+\eta\lambda_1)^2}\right] \cdot \frac{\sum_{k=2}^{d} E[(u_k^T w_0)^2]}{E[(u_1^T w_0)^2]}.$$

By the choice of η and m, we have

$$\left(\frac{1-\eta+\eta\lambda_2}{1-\eta+\eta\lambda_1}\right)^{2m} = \left(1 - \frac{\eta(\lambda_1-\lambda_2)}{1-\eta+\eta\lambda_1}\right)^{2m} \leq \exp\left(-\frac{2\eta(\lambda_1-\lambda_2)m}{1-\eta+\eta\lambda_1}\right) \leq \exp(-\log 2) = \frac{1}{2}.$$

Also, by the choice of η, m and $|S|$, we have

$$\frac{4\eta^2 K m}{(1-\eta+\eta\lambda_1)^2} = \frac{4\sigma^2 \eta^2 m}{|S|(1-\eta+\eta\lambda_1)^2} \leq \frac{1}{4}.$$

Therefore, we have

$$\frac{\sum_{k=2}^{d} E[(u_k^T w_m)^2]}{E[(u_1^T w_m)^2]} \leq \frac{3}{4} \cdot \frac{\sum_{k=2}^{d} E[(u_k^T w_0)^2]}{E[(u_1^T w_0)^2]}.$$

□

Proof of Theorem 3.4. By repeatedly applying Lemma 3.3, we have

$$\frac{\sum_{k=2}^{d} E[(u_k^T \tilde{w}_\tau)^2]}{E[(u_1^T \tilde{w}_\tau)^2]} \le \left(\frac{3}{4}\right)^\tau \frac{\sum_{k=2}^{d} E[(u_k^T \tilde{w}_0)^2]}{E[(u_1^T \tilde{w}_0)^2]} = \left(\frac{3}{4}\right)^\tau \tilde{\theta}_0.$$

Since $\tau = \lceil \log(\tilde{\theta}_0/\epsilon)/\log(4/3) \rceil$, we have

$$\tau \log\left(\frac{3}{4}\right) \le \log\left(\frac{\epsilon}{\tilde{\theta}_0}\right),$$

resulting in

$$\frac{\sum_{k=2}^{d} E[(u_k^T \tilde{w}_\tau)^2]}{E[(u_1^T \tilde{w}_\tau)^2]} \le \epsilon.$$

□

A.1.2 VR HB Power

Proof of Lemma 3.5. From

$$\begin{aligned} w_1 &= (1-\eta)w_0 + \eta\tilde{g} \\ &= (1-\eta)w_0 + \eta C w_0, \end{aligned}$$

we have

$$\begin{aligned} u_k^T w_1 &= (1-\eta)u_k^T w_0 + \eta u_k^T C w_0 \\ &= (1-\eta)u_k^T w_0 + \eta\lambda_k u_k^T w_0 \\ &= (1-\eta+\eta\lambda_k)u_k^T w_0. \end{aligned} \tag{52}$$

Taking the expectation of the square of (52), we obtain

$$E[(u_k^T w_1)^2] = (1-\eta+\eta\lambda_k)^2 E[(u_k^T w_0)^2] = \frac{\alpha_k(\eta)}{4} E[(u_k^T w_0)^2]. \tag{53}$$

Next, from (5), we have

$$\begin{aligned} w_{t+1} &= 2\bigg((1-\eta)w_t + \eta\frac{1}{|S_t|}\sum_{i_t\in S_t} a_{i_t}a_{i_t}^T\Big(w_t - \frac{(w_t^T w_0)}{\|w_0\|^2}w_0\Big) + \frac{(w_t^T w_0)}{\|w_0\|^2}\tilde{g}\bigg) - \beta(\eta)w_{t-1} \\ &= 2\bigg((1-\eta)w_t + \eta\frac{1}{|S_t|}\sum_{i_t\in S_t} a_{i_t}a_{i_t}^T\Big(I - \frac{w_0 w_0^T}{\|w_0\|^2}\Big)w_t + C\frac{w_0 w_0^T}{\|w_0\|^2}w_t\bigg) - \beta(\eta)w_{t-1} \\ &= 2\bigg((1-\eta)w_t + \eta C w_t + \eta\frac{1}{|S_t|}\sum_{i_t\in S_t}(a_{i_t}a_{i_t}^T - C)\Big(I - \frac{w_0 w_0^T}{\|w_0\|^2}\Big)w_t\bigg) - \beta(\eta)w_{t-1} \\ &= 2\big((1-\eta)w_t + \eta C w_t + \eta(C_t - C)Pw_t\big) - \beta(\eta)w_{t-1}, \end{aligned} \tag{54}$$

leading to

$$u_k^T w_{t+1} = 2\big((1-\eta+\eta\lambda_k)u_k^T w_t + \eta u_k^T(C_t - C)Pw_t\big) - \beta(\eta)u_k^T w_{t-1}. \tag{55}$$

Taking the square of (55), we have

$$\begin{aligned} (u_k^T w_{t+1})^2 &= 4(1-\eta+\eta\lambda_k)^2(u_k^T w_t)^2 + 4\eta^2 w_t^T P(C_t - C)u_k u_k^T(C_t - C)Pw_t + (\beta(\eta))^2(u_k^T w_{t-1})^2 \\ &\quad + 8\eta(1-\eta+\eta\lambda_k)u_k^T w_t u_k^T(C_t - C)Pw_t - 4(1-\eta+\eta\lambda_k)\beta(\eta)u_k^T w_t u_k^T w_{t-1} \\ &\quad - 4\eta\beta(\eta)u_k^T(C_t - C)Pw_t u_k^T w_{t-1}. \end{aligned} \tag{56}$$

Since S_t is sampled uniformly at random, C_t is independent of $S_1, \ldots, S_{t-1}$ and identically distributed with $E[C_t] = C$. Therefore,

$$E[u_k^T w_t u_k^T(C_t - C)Pw_t] = E[E[u_k^T w_t u_k^T(C_t - C)Pw_t | w_0, S_1, \ldots, S_{t-1}]] = E[u_k^T w_t u_k^T E[C_t - C]Pw_t] = 0.$$

Similarly, we have

$$E[u_k^T(C_t - C)Pw_t u_k^T w_{t-1}] = 0. \tag{57}$$

As a result, we obtain

$$\begin{aligned} E[(u_k^T w_{t+1})^2] &= \alpha_k(\eta)E[(u_k^T w_t)^2] - 2\sqrt{\alpha_k(\eta)}\beta(\eta)E[(u_k^T w_t)(u_k^T w_{t-1})] + (\beta(\eta))^2 E[(u_k^T w_{t-1})^2] \\ &\quad + 4\eta^2 E[w_t^T P M_k P w_t]. \end{aligned} \tag{58}$$

Using (52) and (53) in (58) for $t = 1$, we have

$$E[(u_k^T w_2)^2] = \Big(\frac{\alpha_k(\eta)}{2} - \beta(\eta)\Big)^2 E[(u_k^T w_0)^2] + 4\eta^2 E[w_1^T P M_k P w_1]. \tag{59}$$

Moreover, by using (55) with $t-1$, multiplying it with $u_k^T w_{t-1}$, taking expectation and using (57) with w_t being w_{t-1} (which can be derived in the same way as (57)) , we have

$$E[(u_k^T w_t)(u_k^T w_{t-1})] = \sqrt{\alpha_k(\eta)}E[(u_k^T w_{t-1})^2] - \beta(\eta)E[(u_k^T w_{t-1})(u_k^T w_{t-2})]. \tag{60}$$

Using (60), we can further write (58) as

$$\begin{aligned} E[(u_k^T w_{t+1})^2] = \alpha_k(\eta)E[(u_k w_t)^2] - \beta(\eta)(2\alpha_k(\eta) - \beta(\eta))E[(u_k^T w_{t-1})^2] \\ + 2\sqrt{\alpha_k(\eta)}(\beta(\eta))^2 E[(u_k^T w_{t-1})(u_k^T w_{t-2})] + 4\eta^2 E[w_t^T PM_k P w_t]. \end{aligned} \tag{61}$$

With $t-1$ in (58), we have

$$\begin{aligned} E[(u_k^T w_t)^2] = \alpha_k(\eta)E[(u_k^T w_{t-1})^2] - 2\sqrt{\alpha_k(\eta)}\beta(\eta)E[(u_k^T w_{t-1})(u_k^T w_{t-2})] + (\beta(\eta))^2 E[(u_k^T w_{t-2})^2] \\ + 4\eta^2 E[w_{t-1}^T PM_k P w_{t-1}]. \end{aligned} \tag{62}$$

Adding (62) multiplied by $\beta(\eta)$ to (61), we obtain

$$\begin{aligned} E[(u_k^T w_{t+1})^2] = (\alpha_k(\eta) - \beta(\eta))E[(u_k^T w_t)^2] - \beta(\eta)(\alpha_k(\eta) - \beta(\eta))E[(u_k^T w_{t-1})^2] + (\beta(\eta))^3 E[(u_k^T w_{t-2})^2] \\ + 4\eta^2 E[w_t^T PM_k P w_t] + 4\eta^2\beta(\eta)E[w_{t-1}^T PM_k P w_{t-1}]. \end{aligned} \tag{63}$$

With $t-1$ in (63), we finally have

$$\begin{aligned} E[(u_k^T w_t)^2] = (\alpha_k(\eta) - \beta(\eta))E[(u_k^T w_{t-1})^2] - \beta(\eta)(\alpha_k(\eta) - \beta(\eta))E[(u_k^T w_{t-2})^2] + (\beta(\eta))^3 E[(u_k^T w_{t-3})^2] \\ + 4\eta^2 E[w_{t-1}^T PM_k P w_{t-1}] + 4\eta^2\beta(\eta)E[w_{t-2}^T PM_k P w_{t-2}] \end{aligned} \tag{64}$$

for $t \geq 3$.

Using Lemma A.4 for $E[(u_k^T w_t)^2]$ defined by (53), (59), and (64) with

$$\alpha = \alpha_k(\eta), \quad \beta = \beta(\eta), \quad L_0 = E[(u_k^T w_0)^2], \quad L_t = 4\eta^2 E[w_t^T PM_k P w_t],$$

we have

$$E[(u_k^T w_t)^2] = p_t(\alpha_k(\eta), \beta(\eta))E[(u_k^T w_0)^2] + 4\eta^2 \sum_{r=1}^{t-1} q_{t-r-1}(\alpha_k(\eta), \beta(\eta))E[w_r^T PM_k P w_r].$$

□

Proof of Lemma 3.6. Since $\| \sum_{k=2}^d u_k u_k^T \| \leq 1$, we have

$$\|\sum_{k=2}^d M_k\| = \|\sum_{k=2}^d E[(C_t - C)u_k u_k^T (C_t - C)]\| \leq E[\|C_t - C\|^2] = E[\|(C_t - C)^2\|] = K.$$

By Lemma A.2, this leads to

$$\sum_{k=2}^d E[w_t^T PM_k P w_t] = E[w_t^T P \sum_{k=2}^d M_k P w_t] \leq \|\sum_{k=2}^d M_k\| E[\|Pw_t\|^2] \leq KE[\|Pw_t\|^2]. \tag{65}$$

Let

$$F = \begin{bmatrix} I \\ 0 \end{bmatrix}, \; G = \begin{bmatrix} 2\left[(1-\eta)I + \eta C\right] & -\beta(\eta)I \\ I & 0 \end{bmatrix}, \; G_0 = \begin{bmatrix} (1-\eta)I + \eta C & -\beta(\eta)I \\ I & 0 \end{bmatrix}, \; H_t = 2\eta \begin{bmatrix} (C_t - C)P & 0 \\ 0 & 0 \end{bmatrix}.$$

From the update rule in Algorithm 2 expressed in (54), we can write

$$w_t = F^T (G + H_{t-1})(G + H_{t-2}) \cdots (G + H_1)(G_0 + H_0)Fw_0.$$

Using Lemma A.3 for the expansion of $(G+H_{t-1})(G+H_{t-2})\cdots(G+H_1)(G_0+H_0)$, we have

$$Pw_t = PF^T\left(G^{t-1}G_0 + \sum_{i=1}^{t-1}\left[\prod_{j=t-1}^{i+1}(G+H_j)H_iG^{i-1}G_0\right] + \prod_{j=t-1}^{1}(G+H_j)H_0\right)Fw_0. \tag{66}$$

Since $C_0, C_1, \cdots, C_{t-1}$ are independent and identically distributed with mean C, so are $H_0, H_1, \cdots, H_{t-1}$ with mean 0. Therefore, the expectation of all cross-terms in the "square" of (66) are zero. Using the fact that $H_0Fw_0 = 0$, we have

$$E[\|Pw_t\|^2] = E[\|PF^TG^{t-1}G_0Fw_0\|^2] + \sum_{i=1}^{t-1}E\Big[\Big\|PF^T\prod_{j=t-1}^{i+1}(G+H_j)H_iG^{i-1}G_0Fw_0\Big\|^2\Big]. \tag{67}$$

Note that this result is analogous to (42) in the analysis of VR Power. From $F^TG^{t-1}G_0F = Y_t((1-\eta)I+\eta C, \beta(\eta))$ (see (20) for the definition of Y_t) and (24b) in Lemma A.1 with $w = w_0/\|w_0\|$ and the fact that $\|w_0\|^2(1-(u_1^Tw_0)^2/\|w_0\|^2) = \sum_{k=2}^d(u_k^Tw_0)^2$, we have

$$E\big[\|PF^TG^{t-1}G_0Fw_0\|^2\big] = 4p_t(\alpha_1(\eta),\beta(\eta))\cdot\sum_{k=2}^d E[(u_k^Tw_0)^2]. \tag{68}$$

Using Lemma A.2, $\|P\| = 1$, $H_t = 2\eta F(C_t - C)PF^T$, we have

$$\begin{aligned}E\big[\Big\|PF^T\prod_{j=t-1}^{i+1}(G+H_j)H_iG^{i-1}G_0Fw_0\Big\|^2\big] &\le 4\eta^2\|P\|^2\cdot E\big[\Big\|F^T\prod_{j=t-1}^{i+1}(G+H_j)F(C_i-C)PF^TG^{i-1}G_0Fw_0\Big\|^2\big]\\ &\le \Big\|E\big[F^T\big[\prod_{j=t-1}^{i+1}(G+H_j)\big]^TFF^T\prod_{j=t-1}^{i+1}(G+H_j)F\big]\Big\|\\ &\quad\cdot 4\eta^2E\big[\big\|(C_i-C)PF^TG^{i-1}G_0Fw_0\big\|^2\big].\end{aligned} \tag{69}$$

Using mathematical induction on i, we prove that

$$E\big[\big[\prod_{j=t-1}^{i+1}(G+H_j)\big]^TFF^T\prod_{j=t-1}^{i+1}(G+H_j)\big] = \sum_{\substack{(v_{i+1},\cdots,v_{t-1})\\ \in\{0,1\}^{t-i-1}}}E\big[\big[\prod_{j=t-1}^{i+1}H_j^{1-v_j}G^{v_j}\big]^TFF^T\prod_{j=t-1}^{i+1}H_j^{1-v_j}G^{v_j}\big] \tag{70}$$

for any $i \le t-2$ and fixed $t \ge 2$. Since $E[H_{t-1}] = 0$, we have

$$E[(G^T+H_{t-1}^T)FF^T(G+H_{t-1})] = G^TFF^TG + E[H_{t-1}^TFF^TH_{t-1}].$$

This proves the base case for $i = t-2$.

Suppose that (70) holds for $i = k$. Then, since H_k is independent from $H_{k+1}, \cdots, H_{t-1}$ and $E[H_k] = 0$, we have

$$\begin{aligned}E\big[\big[\prod_{j=t-1}^{k}(G+H_j)\big]^TFF^T\prod_{j=t-1}^{k}(G+H_j)\big] &= G^TE\big[\big[\prod_{j=t-1}^{k+1}(G+H_j)\big]^TFF^T\prod_{j=t-1}^{k+1}(G+H_j)\big]G\\ &\quad + E\big[H_k^T\big[\prod_{j=t-1}^{k+1}(G+H_j)\big]^TFF^T\prod_{j=t-1}^{k+1}(G+H_j)H_k\big].\end{aligned}$$

From (70), we have

$$\begin{aligned}&G^TE\big[\big[\prod_{j=t-1}^{k+1}(G+H_j)\big]^TFF^T\prod_{j=t-1}^{k+1}(G+H_j)\big]G\\ &= \sum_{\substack{(v_{k+1},\cdots,v_{t-1})\\ \in\{0,1\}^{t-k-1}}}E\big[\big[\big(\prod_{j=t-1}^{k+1}H_j^{1-v_j}G^{v_j}\big)G\big]^TFF^T\big(\prod_{j=t-1}^{k+1}H_j^{1-v_j}G^{v_j}\big)G\big].\end{aligned}$$

Also, by the independence of H_k from $H_{k+1}, \cdots, H_{t-1}$ and (70), we have

$$\begin{aligned}
&E\big[H_k^T\big[\prod_{j=t-1}^{k+1}(G+H_j)\big]^T FF^T \prod_{j=t-1}^{k+1}(G+H_j)H_k\big] \\
&= E\big[H_k^T E\big[\big[\prod_{j=t-1}^{k+1}(G+H_j)\big]^T FF^T \prod_{j=t-1}^{k+1}(G+H_j)\big]H_k\big] \\
&= E\big[H_k^T \sum_{\substack{(v_{k+1},\cdots,v_{t-1})\\ \in\{0,1\}^{t-i-1}}} E\big[\big[\prod_{j=t-1}^{k+1} H_j^{1-v_j}G^{v_j}\big]^T FF^T \prod_{j=t-1}^{k+1} H_j^{1-v_j}G^{v_j}\big]H_k\big] \\
&= \sum_{\substack{(v_{k+1},\cdots,v_{t-1})\\ \in\{0,1\}^{t-k-1}}} E\big[\big[\big(\prod_{j=t-1}^{k+1} H_j^{1-v_j}G^{v_j}\big)H_k\big]^T FF^T \big(\prod_{j=t-1}^{k+1} H_j^{1-v_j}G^{v_j}\big)H_k\big].
\end{aligned}$$

Therefore, we have

$$E\big[\big[\prod_{j=t-1}^{k}(G+H_j)\big]^T FF^T \prod_{j=t-1}^{k}(G+H_j)\big] = \sum_{\substack{(v_k,\cdots,v_{t-1})\\ \in\{0,1\}^{t-k}}} E\big[\big[\prod_{j=t-1}^{k} H_j^{1-v_j}G^{v_j}\big]^T FF^T \prod_{j=t-1}^{k} H_j^{1-v_j}G^{v_j}\big],$$

which completes the proof of (70).

Using the Jensen's inequality and the norm property of a symmetric matrix, we have

$$\|E\big[F^T\big[\prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]\big]^T FF^T \prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]F\big]\| \le E\big[\|F^T \prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]F\|^2\big]. \tag{71}$$

For $(v_{i+1}, \cdots, v_{t-1}) \in \{0,1\}^{t-i-1}$, let $J = \{j_1, j_2, \cdots, j_{\bar{k}}\}$ be a set of indices such that $j_1 < j_2 < \cdots < j_{\bar{k}}$ and $v_j = 0$ if $j \in J$ and $v_j = 1$ otherwise. Also, let $j_0 = i$. Using that $H_j = FF^T H_j FF^T$, we have

$$\begin{aligned}
E\big[\|F^T \prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]F\|^2\big] &= E\big[\|F^T G^{t-j_{\bar{k}}-1}F \prod_{l=\bar{k}}^{1}\big(F^T H_{j_l} FF^T G^{j_l-j_{l-1}-1}F\big)\|^2\big] \\
&\le E\big[\|F^T G^{t-j_{\bar{k}}-1}F\|^2 \prod_{l=\bar{k}}^{1}\|F^T H_{j_l}F\|^2 \|F^T G^{j_l-j_{l-1}-1}F\|^2\big].
\end{aligned} \tag{72}$$

Since $F^T G^t F = Z_t((1-\eta)I + \eta C, \beta(\eta))$, using (24c) in Lemma A.1, we have

$$\|F^T G^t F\|^2 \le q_t(\alpha_1(\eta), \beta(\eta)). \tag{73}$$

Also, from that $F^T H_t F = 2\eta(C_t - C)P$, we have

$$E\big[\|F^T H_t F\|^2\big] \le 4\eta^2 E\big[\|(C_t - C)P\|^2\big] \le 4\eta^2 E\big[\|(C_t - C)\|^2\big] = 4\eta^2 E\big[\|(C_t - C)^2\|\big] = 4\eta^2 K. \tag{74}$$

where the last inequality follows from $\|P\| = 1$ and the second last equality follows from the symmetry of $C_t - C$. Using (73) and Lemma A.5, we have

$$\|F^T G^{t-j_{\bar{k}}-1}F\|^2 \prod_{l=\bar{k}}^{1}\|F^T G^{j_l-j_{l-1}-1}F\|^2 \le \left(\frac{1}{\alpha_1(\eta) - 4\beta(\eta)}\right)^{\bar{k}} q_{t-i-1}(\alpha_1(\eta), \beta(\eta)). \tag{75}$$

Note that there are $\bar{k}+1$ terms of the form $\|F^T G^t F\|^2$ for some $t \ge 0$ on the left-hand side of the above inequality and we use Lemma A.5 $\bar{k}$ times to obtain the term on the right-hand side.

Using (71), (72), (75), and the independence of $C_0, C_1, \cdots, C_{t-1}$, we obtain

$$\|E\big[F^T\big[\prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]\big]^T FF^T \prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]F\big]\| \le \left(\frac{4\eta^2 K}{\alpha_1(\eta) - 4\beta(\eta)}\right)^{\bar{k}} q_{t-i-1}(\alpha_1(\eta), \beta(\eta)).$$

Combined with (70), this results in

$$\begin{aligned}
&\|E\big[F^T\big[\prod_{j=t-1}^{i+1}(G+H_j)\big]^T FF^T \prod_{j=t-1}^{i+1}(G+H_j)F\big]\| \\
&= \|\sum_{\substack{(v_{i+1},\cdots,v_{t-1})\\ \in\{0,1\}^{t-i-1}}} E\big[F^T\big[\prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]\big]^T FF^T \prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]F\big]\| \\
&\le \sum_{\substack{(v_{i+1},\cdots,v_{t-1})\\ \in\{0,1\}^{t-i-1}}} \|E\big[F^T\big[\prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]\big]^T FF^T \prod_{j=t-1}^{i+1}[H_j^{1-v_j}G^{v_j}]F\big]\| \\
&\le \sum_{\bar{k}=0}^{t-i-1}\binom{t-i-1}{\bar{k}}\left(\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{\bar{k}} q_{t-i-1}(\alpha_1(\eta),\beta(\eta)) \\
&= q_{t-i-1}(\alpha_1(\eta),\beta(\eta))\left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{t-i-1}.
\end{aligned} \tag{76}$$

On the other hand, using Lemma A.2 and (68) for $t=i$, we have

$$\begin{aligned}
\eta^2 E\big[\big\|(C_i-C)PF^TG^{i-1}G_0Fw_0\big\|^2\big] &= \eta^2 E[w_0F^TG_0^T(G^{i-1})^TFP^TE[(C_i-C)^2]PF^TG^{i-1}G_0Fw_0] \\
&\le \eta^2\|E[(C_i-C)^2]\|E[\|PF^TG^{i-1}G_0Fw_0\|^2] \\
&\le 4\eta^2K\cdot p_i(\alpha_1(\eta),\beta(\eta))\cdot\sum_{k=2}^{d}E[(u_k^Tw_0)^2].
\end{aligned} \tag{77}$$

Using (76) and (77) to bound (69), we have

$$\begin{aligned}
&E\big[\big\|PF^T\prod_{j=t-1}^{i+1}(G+H_j)H_iG^{i-1}G_0Fw_0\big\|^2\big] \\
&\quad\le 16\eta^2K\cdot p_i(\alpha_1(\eta),\beta(\eta))\cdot q_{t-i-1}(\alpha_1(\eta),\beta(\eta))\left(1+\frac{4\eta^2K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{t-i-1}\cdot\sum_{k=2}^{d}E[(u_k^Tw_0)^2]
\end{aligned} \tag{78}$$

Using (68) and (78) for (67), we finally have

$$\begin{aligned}
E[\|Pw_t\|^2] \le &\left[4p_t(\alpha_1(\eta),\beta(\eta))+16\eta^2K\sum_{i=1}^{t-1}p_i(\alpha_1(\eta),\beta(\eta))\cdot q_{t-i-1}(\alpha_1(\eta),\beta(\eta))\left(1+\frac{4\eta^2K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{t-i-1}\right] \\
&\cdot\sum_{k=2}^{d}E[(u_k^Tw_0)^2].
\end{aligned}$$

By (90) and (91) in Lemma A.4, we have

$$\begin{aligned}
p_t(\alpha_1(\eta),\beta(\eta)) &\le \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2t}, \\
q_t(\alpha_1(\eta),\beta(\eta)) &\le \left(\frac{1}{\alpha_1(\eta)-\beta(\eta)}\right)\left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2(t+1)}.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
&\left[4p_t(\alpha_1(\eta),\beta(\eta)) + 16\eta^2 K \sum_{i=1}^{t-1} p_i(\alpha_1(\eta),\beta(\eta)) \cdot q_{t-i-1}(\alpha_1(\eta),\beta(\eta)) \left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{t-i-1}\right] \\
&\le 4\left[1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)} \sum_{i=1}^{t-1}\left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{t-i-1}\right] \cdot \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2t} \\
&= 4\left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{t-1} \cdot \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2t},
\end{aligned}$$

which results in

$$E[\|Pw_t\|^2] \le 4\left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{t-1} \cdot \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2t} \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2].$$

Finally, from (65), we have

$$\sum_{k=2}^{d} E[w_t^T P M_k P w_t] \le 4K\left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{t-1} \cdot \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2t} \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2]. \quad (79)$$

This completes the proof of the first statement.

Next, from $\alpha_2(\eta) = 4\beta(\eta) \ge \alpha_k(\eta)$ for $k \ge 2$ and (92) in Lemma A.4,

$$\sum_{k=2}^{d} p_m(\alpha_k(\eta),\beta(\eta)) E[(u_k^T w_0)^2] \le p_m(\alpha_2(\eta),\beta(\eta)) \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2]. \quad (80)$$

Also, using (91) and (92) in Lemma A.4 and (79), we have

$$\begin{aligned}
&4\eta^2 \sum_{k=2}^{d}\sum_{r=1}^{m-1} q_{m-r-1}(\alpha_k(\eta),\beta(\eta)) E[w_r^T P M_k P w_r] \\
&\le \frac{16\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)} \sum_{r=1}^{m-1}\left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{r-1} \cdot \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2m} \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2] \\
&\le 4\left[\left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{m-1} - 1\right] \cdot \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2m} \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2].
\end{aligned}$$

Since $0 < \dfrac{4\eta^2 Km}{\alpha_1(\eta)-\beta(\eta)} < 1$, using that $\exp(x) \le 1+2x$ for $x \in [0,1]$ we have

$$\left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{m-1} - 1 \le \left(1+\frac{4\eta^2 K}{\alpha_1(\eta)-4\beta(\eta)}\right)^{m} - 1 \le \exp\left(\frac{4\eta^2 Km}{\alpha_1(\eta)-4\beta(\eta)}\right) - 1 \le \frac{8\eta^2 Km}{\alpha_1(\eta)-4\beta(\eta)},$$

leading to

$$\begin{aligned}
4\eta^2 \sum_{k=2}^{d}\sum_{r=1}^{m-1} q_{m-r-1}(\alpha_k(\eta),\beta(\eta)) E[w_r^T P M_k P w_r] \le \frac{32\eta^2 Km}{\alpha_1(\eta)-4\beta(\eta)} \cdot \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2m} \\
\cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2]. \quad (81)
\end{aligned}$$

Using (80), (81) for Lemma 3.5, we finally have

$$\sum_{k=2}^{d} E[(u_k^T w_m)^2] \le \left[p_m(\alpha_2(\eta),\beta(\eta)) + \frac{32\eta^2 Km}{\alpha_1(\eta)-4\beta(\eta)} \cdot \left(\frac{\sqrt{\alpha_1(\eta)}}{2}+\frac{\sqrt{\alpha_1(\eta)-4\beta(\eta)}}{2}\right)^{2m}\right] \cdot \sum_{k=2}^{d} E[(u_k^T w_0)^2]. \quad (82)$$

Lastly, using Lemma 3.5 for $k = 1$, we have

$$E[(u_1^T w_m)^2] = p_m(\alpha_1(\eta), \beta(\eta))E[(u_1^T w_0)^2] + 4\eta^2 \sum_{r=1}^{m-1} q_{m-r-1}(\alpha_1(\eta), \beta(\eta))E[w_r^T PM_1 Pw_r].$$

Since PM_kP is positive semi-definite and $q_t(\alpha_1(\eta), \beta(\eta)) \geq 0$ for $1 \leq t < m$ by (91) in Lemma A.4, we have

$$E[(u_1^T w_m)^2] \geq p_m(\alpha_1(\eta), \beta(\eta))E[(u_1^T w_0)^2]. \tag{83}$$

Also, from $\alpha_1(\eta) > \alpha_2(\eta) = 4\beta(\eta)$ and (90) in Lemma A.4, we have

$$p_m(\alpha_1(\eta), \beta(\eta)) \geq \frac{1}{4}\left(\frac{\sqrt{\alpha_1(\eta)}}{2} + \frac{\sqrt{\alpha_1(\eta) - 4\beta(\eta)}}{2}\right)^{2m}. \tag{84}$$

Using (82), (83) and (84), we eventually obtain

$$\frac{\sum_{k=2}^d E[(u_k^T w_m)^2]}{E[(u_1^T w_m)^2} \leq \left[\frac{p_m(\alpha_2(\eta), \beta(\eta))}{p_m(\alpha_1(\eta), \beta(\eta))} + \frac{128\eta^2 Km}{\alpha_1(\eta) - 4\beta(\eta)}\right] \cdot \frac{\sum_{k=2}^d E[(u_k^T w_0)^2]}{E[(u_1^T w_0)^2},$$

which completes the proof. □

Proof of Lemma 3.7. Using the conditions on m and $|S|$, we have

$$0 \leq \frac{4\eta^2 Km}{\alpha_1(\eta) - 4\beta(\eta)} \leq \frac{1}{128}. \tag{85}$$

Also, from

$$p_m(\alpha_2(\eta), \beta(\eta)) = (\beta(\eta))^m, \quad p_m(\alpha_1(\eta), \beta(\eta)) \geq \frac{1}{4}\left(\frac{\sqrt{\alpha_1(\eta)}}{2} + \frac{\sqrt{\alpha_1(\eta) - 4\beta(\eta)}}{2}\right)^{2m}.$$

and the choice of and m, we have

$$\begin{aligned}
\frac{p_m(\alpha_2(\eta), \beta(\eta))}{p_m(\alpha_1(\eta), \beta(\eta))} &\leq 4 \cdot \left(\frac{\sqrt{4\beta(\eta)}}{\sqrt{\alpha_1(\eta)} + \sqrt{\alpha_1(\eta) - 4\beta(\eta)}}\right)^{2m} \\
&= 4 \cdot \left(1 - \frac{\sqrt{\alpha_1(\eta)} - \sqrt{4\beta(\eta)} + \sqrt{\alpha_1(\eta) - 4\beta(\eta)}}{\sqrt{\alpha_1(\eta)} + \sqrt{\alpha_1(\eta) - 4\beta(\eta)}}\right)^{2m} \\
&= 4 \cdot \left(1 - \frac{\eta\lambda_1\Delta + \sqrt{\eta\lambda_1\Delta(2(1-\eta) + \eta(\lambda_1 + \lambda_2))}}{1 - \eta + \eta\lambda_1 + \sqrt{\eta\lambda_1\Delta(2(1-\eta) + \eta(\lambda_1 + \lambda_2))}}\right)^{2m} \\
&\leq 4 \cdot \exp\left(-2\frac{\eta\lambda_1\Delta + \sqrt{\eta\lambda_1\Delta(2(1-\eta) + \eta(\lambda_1 + \lambda_2))}}{1 - \eta + \eta\lambda_1 + \sqrt{\eta\lambda_1\Delta(2(1-\eta) + \eta(\lambda_1 + \lambda_2))}} m\right) \\
&\leq \frac{1}{2}.
\end{aligned} \tag{86}$$

Therefore, using (85) and (86) in Lemma 3.6, we finally have

$$\frac{\sum_{k=2}^d E[(u_k^T w_m)^2]}{E[(u_1^T w_m)^2]} \leq \left(\frac{p_m(\alpha_2(\eta), \beta(\eta))}{p_m(\alpha_1(\eta), \beta(\eta))} + \frac{128\eta^2 Km}{\alpha_1(\eta) - 4\beta(\eta)}\right)\left(\frac{\sum_{k=2}^d E[(u_k^T w_0)^2]}{E[(u_1^T w_0)^2]}\right) \leq \frac{3}{4}\left(\frac{\sum_{k=2}^d E[(u_k^T w_0)^2]}{E[(u_1^T w_0)^2]}\right),$$

which completes the proof. □

Proof of Theorem 3.8. By repeatedly applying Lemma 3.7, we have

$$\frac{\sum_{k=2}^{d} E[(u_k^T \tilde{w}_\tau)^2]}{E[(u_1^T \tilde{w}_\tau)^2]} \leq \left(\frac{3}{4}\right)^\tau \frac{\sum_{k=2}^{d} E[(u_k^T \tilde{w}_0)^2]}{E[(u_1^T \tilde{w}_0)^2]} = \left(\frac{3}{4}\right)^\tau \tilde{\theta}_0.$$

Since $\tau = \lceil \log(\tilde{\theta}_0/\epsilon)/\log(4/3) \rceil$, we have

$$\tau \log\left(\frac{3}{4}\right) \leq \log\left(\frac{\epsilon}{\tilde{\theta}_0}\right),$$

resulting in

$$\frac{\sum_{k=2}^{d} E[(u_k^T \tilde{w}_\tau)^2]}{E[(u_1^T \tilde{w}_\tau)^2]} \leq \epsilon.$$

□

A.2 Technical Lemmas

Lemma A.2. *Let w be a vector in $\mathbb{R}^d$ and let M be a $d \times d$ symmetric matrix. Then, we have*

$$w^T M w \leq \|M\| \|w\|^2.$$

Proof. By the cyclic property of the trace, we have

$$w^T M w = \mathrm{Tr}[w^T M w] = \mathrm{Tr}[M w w^T].$$

Since ww^T is positive semi-definite, we have

$$\mathrm{Tr}[M w w^T] \leq \|M\| \mathrm{Tr}[w w^T].$$

Again, by the cyclic property of the trace, we finally have

$$w^T M w \leq \|M\| \mathrm{Tr}[w w^T] = \|M\| \mathrm{Tr}[w^T w] = \|M\| \|w\|^2.$$

□

Lemma A.3. *Let A_i and B_i be $d \times d$ matrices for $i = 0, \cdots, t-1$. Then, we have*

$$\prod_{i=t-1}^{0} (A_i + B_i) = (A_{t-1} + B_{t-1})(A_{t-2} + B_{t-2}) \cdots (A_0 + B_0) = \prod_{i=t-1}^{0} A_i + \sum_{i=0}^{t-1} \left[\prod_{j=t-1}^{i+1} (A_j + B_j) B_i \prod_{k=i-1}^{0} A_k \right]. \quad (87)$$

Proof. We prove the statement by induction. For $t = 1$, we have

$$\prod_{i=0}^{0} A_i + \sum_{i=0}^{0} \left[\prod_{j=0}^{i+1} (A_j + B_j) B_i \prod_{k=i-1}^{0} A_k \right] = A_0 + \left[\prod_{j=0}^{1} (A_j + B_j) B_0 \prod_{k=-1}^{0} A_k \right] = A_0 + B_0,$$

which proves the base case. Next, suppose that we have (87) for $t-2$. Then, we have

$$\begin{aligned}
\prod_{i=t-1}^{0} (A_i + B_i) &= (A_{t-1} + B_{t-1}) \prod_{i=t-2}^{0} (A_i + B_i) \\
&= (A_{t-1} + B_{t-1}) \left(\prod_{i=t-2}^{0} A_i + \sum_{i=0}^{t-2} \left[\prod_{j=t-2}^{i+1} (A_j + B_j) B_i \prod_{k=i-1}^{0} A_k \right] \right) \\
&= \prod_{i=t-1}^{0} A_i + B_{t-1} \prod_{i=t-2}^{0} A_i + \left(\sum_{i=0}^{t-2} \left[\prod_{j=t-1}^{i+1} (A_j + B_j) B_i \prod_{k=i-1}^{0} A_k \right] \right) \\
&= \prod_{i=t-1}^{0} A_i + \sum_{i=0}^{t-1} \left[\prod_{j=t-1}^{i+1} (A_j + B_j) B_i \prod_{k=i-1}^{0} A_k \right].
\end{aligned}$$

This completes the proof. □

Lemma A.4. *Let x_t be a sequence of real numbers such that*

$$x_t = (\alpha - \beta) x_{t-1} - \beta(\alpha - \beta) x_{t-2} + \beta^3 x_{t-3} + L_{t-1} + \beta L_{t-2}$$

for $t \geq 3$ and $x_0 = L_0$, $x_1 = \frac{\alpha}{4} L_0$, $x_2 = \left(\frac{\alpha}{2} - \beta\right)^2 L_0 + L_1$. Then, we have

$$x_t = p_t(\alpha, \beta) L_0 + \sum_{r=1}^{t-1} q_{t-r-1}(\alpha, \beta) L_r. \quad (88)$$

Moreover, for $t \geq 0$, we have

- *if* $0 \le \alpha = 4\beta$,

$$p_t(4\beta,\beta) = \beta^t \ge 0, \quad q_t(4\beta,\beta) = (t+1)^2\beta^t \ge 0, \tag{89}$$

- *if* $0 \le 4\beta < \alpha$,

$$p_t(\alpha,\beta) = \left[\frac{1}{2}\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha-4\beta}}{2}\right)^t + \frac{1}{2}\left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha-4\beta}}{2}\right)^t\right]^2 > p_t(4\beta,\beta) \ge 0, \tag{90}$$

$$q_t(\alpha,\beta) = \frac{1}{\alpha-4\beta}\left[\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha-4\beta}}{2}\right)^{t+1} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha-4\beta}}{2}\right)^{t+1}\right]^2 > q_t(4\beta,\beta) \ge 0, \tag{91}$$

- *if* $0 \le \alpha < 4\beta$,

$$p_t(\alpha,\beta) \le p_t(4\beta,\beta), \quad q_t(\alpha,\beta) \le q_t(4\beta,\beta). \tag{92}$$

Proof. It is easy to check that x_0, x_1, and x_2 satisfy (88). Suppose that (88) holds for $t-1, t-2, t-3$. Then, we have

$$\begin{aligned} x_t &= (\alpha-\beta)x_{t-1} - \beta(\alpha-\beta)x_{t-2} + \beta^3 x_{t-3} + L_{t-1} + \beta L_{t-2} \\ &= p_t(\alpha,\beta)L_0 + L_{t-1} + \alpha L_{t-2} + (\alpha-\beta)^2 L_{t-3} + \sum_{r=1}^{t-4} q_{t-r-1}(\alpha,\beta)L_r \\ &= p_t(\alpha,\beta)L_0 + \sum_{r=1}^{t-1} q_{t-r-1}(\alpha,\beta)L_r. \end{aligned}$$

Therefore, (88) holds by induction.

Next, we prove (89), (90), (91) and (92). The characteristic equation of (9) is

$$r^3 - (\alpha-\beta)r^2 + \beta(\alpha-\beta)r - \beta^3 = 0. \tag{93}$$

If $0 \le \alpha = 4\beta$, (93) has a cube root of $r = \beta$. From initial conditions (11) and (12), we obtain

$$p_t(4\beta,\beta) = \beta^t \ge 0, \quad q_t(4\beta,\beta) = (t+1)^2\beta^t \ge 0. \tag{94}$$

If $0 \le 4\beta < \alpha$, the roots of (93) are

$$r = \beta, \frac{\alpha-2\beta}{2} + \frac{\sqrt{\alpha^2-4\alpha\beta}}{2}, \frac{\alpha-2\beta}{2} - \frac{\sqrt{\alpha^2-4\alpha\beta}}{2}.$$

With initial conditions (11), we obtain

$$p_t(\alpha,\beta) = \frac{1}{4}\left(\frac{\alpha-2\beta}{2} + \frac{\sqrt{\alpha^2-4\alpha\beta}}{2}\right)^t + \frac{1}{4}\left(\frac{\alpha-2\beta}{2} - \frac{\sqrt{\alpha^2-4\alpha\beta}}{2}\right)^t + \frac{1}{2}\beta^t$$

Using the fact that $\alpha > 4\beta$ and the arithmetic-geometric mean inequality, we have

$$p_t(\alpha,\beta) > \beta^t \ge 0.$$

Moreover, we can further write $p_t(\alpha,\beta)$ as

$$p_t(\alpha,\beta) = \left[\frac{1}{2}\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha-4\beta}}{2}\right)^t + \frac{1}{2}\left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha-4\beta}}{2}\right)^t\right]^2$$

by expanding this expression.

On the other hand, using (12), we have

$$q_t(\alpha,\beta) = \frac{1}{\alpha-4\beta}\left[\left(\frac{\alpha-2\beta}{2}+\frac{\sqrt{\alpha^2-4\alpha\beta}}{2}\right)^{t+1}+\left(\frac{\alpha-2\beta}{2}-\frac{\sqrt{\alpha^2-4\alpha\beta}}{2}\right)^{t+1}-2\beta^{t+1}\right]$$
$$= \frac{1}{\alpha-4\beta}\left[\left(\frac{\sqrt{\alpha}}{2}+\frac{\sqrt{\alpha-4\beta}}{2}\right)^{t+1}-\left(\frac{\sqrt{\alpha}}{2}-\frac{\sqrt{\alpha-4\beta}}{2}\right)^{t+1}\right]^2 \geq 0.$$

Using the fact that $A^{t+1}-B^{t+1}=(A-B)(A^t+A^{t-1}B+\cdots+B^t)$ for any $A,B\in\mathbb{R}$, we have

$$q_t(\alpha,\beta) = \left[\sum_{i=0}^{t}\left(\frac{\sqrt{\alpha}}{2}+\frac{\sqrt{\alpha-4\beta}}{2}\right)^{i}\left(\frac{\sqrt{\alpha}}{2}-\frac{\sqrt{\alpha-4\beta}}{2}\right)^{t-i}\right]^2.$$

Again, using the arithmetic-geometric mean inequality and the fact that $\alpha > 4\beta$, we have

$$q_t(\alpha,\beta) \geq \left[(t+1)\left(\frac{\sqrt{\alpha}}{2}+\frac{\sqrt{\alpha-4\beta}}{2}\right)^{t/2}\left(\frac{\sqrt{\alpha}}{2}-\frac{\sqrt{\alpha-4\beta}}{2}\right)^{t/2}\right]^2 = (t+1)^2\beta^t = q_t(4\beta,\beta).$$

If $0 \leq \alpha < 4\beta$, the roots of (93) are

$$r = \beta, \frac{\alpha-2\beta}{2}+\frac{\sqrt{4\alpha\beta-\alpha^2}}{2}i, \frac{\alpha-2\beta}{2}-\frac{\sqrt{4\alpha\beta-\alpha^2}}{2}i.$$

Setting

$$\cos\,\theta_p = \frac{\alpha-2\beta}{2\beta}, \quad \sin\,\theta_p = \frac{\sqrt{4\alpha\beta-\alpha^2}}{2\beta}$$

it is easy to verify that

$$\begin{aligned}
p_t(\alpha,\beta) &= \frac{1}{4}\beta^t\left[\cos\,\theta_p + i\,\sin\,\theta_p\right]^t + \frac{1}{4}\beta^t\left[\cos\,\theta_p - i\,\sin\,\theta_p\right]^t + \frac{1}{2}\beta^t \\
&= \frac{1}{4}(e^{i\theta t}+e^{-i\theta t})\beta^t + \frac{1}{2}\beta^t \\
&= \frac{1}{4}|e^{i\theta t}+e^{-i\theta t}|\beta^t + \frac{1}{2}\beta^t \\
&\leq \frac{1}{4}(|e^{i\theta t}|+|e^{-i\theta t}|)\beta^t + \frac{1}{2}\beta^t \\
&= \beta^t.
\end{aligned}$$

Moreover, with

$$\cos\,\theta_q = \frac{\alpha-2\beta}{2\beta}, \quad \sin\,\theta_q = \frac{\sqrt{4\alpha\beta-\alpha^2}}{2\beta}, \quad \cos\,\phi_q = 1-\frac{\alpha}{2\beta}, \quad \sin\,\phi_q = -\frac{\sqrt{4\alpha\beta-\alpha^2}}{2\beta},$$

it can be seen by using elementary calculus that

$$q_t(\alpha,\beta) = \left[\frac{2\beta}{4\beta-\alpha}+\frac{2\beta}{4\beta-\alpha}\cos(\phi_q+t\theta_q)\right]\beta^t. \tag{95}$$

Let

$$Q(t) = \frac{q_t(4\beta,\beta)-q_t(\alpha,\beta)}{\beta^t}.$$

Then, from (9) and (11), we have

$$Q(0)=0, \quad Q(1)=\frac{4\beta-\alpha}{\beta}, \quad Q(2)=\frac{(4\beta-\alpha)(2\beta+\alpha)}{\beta^2}, \quad Q(3)=\frac{(\alpha^2+4\beta^2)(4\beta-\alpha)}{\beta^3} \tag{96}$$

resulting in

$$Q(2) - Q(0) = \frac{(4\beta - \alpha)(2\beta + \alpha)}{\beta^2} \geq 0, \quad Q(3) - Q(1) = \frac{(\alpha^2 + 3\beta^2)(4\beta - \alpha)}{\beta^3} \geq 0. \tag{97}$$

In order to show $Q(t) \geq 0$ for $t \geq 0$, we prove $Q(t+2) - Q(t) \geq 0$ for $t \geq 0$. Using (94), (95) and standard trigonometric equalities, it follows that

$$Q(t+2) - 2Q(t) + Q(t-2) = 8 + \frac{2\alpha}{\beta}\cos(\phi_q + t\theta_q).$$

In turn, we have

$$\begin{aligned} Q(t+2) - Q(t) &= Q(t) - Q(t-2) + 8 + \frac{2\alpha}{\beta}\cos(\phi_q + t\theta_q) \\ &\geq Q(t) - Q(t-2) + 8 - \frac{2\alpha}{\beta} \\ &= Q(t) - Q(t-2) + \frac{2(4\beta - \alpha)}{\beta} \\ &\geq Q(t) - Q(t-2). \end{aligned} \tag{98}$$

From (96), (97), and (98), for $t \geq 0$, we obtain $Q(t) \geq 0$ implying

$$q_t(\alpha, \beta) \leq q_t(4\beta, \beta).$$

□

Lemma A.5. *If $\alpha > 4\beta \geq 0$, then for $0 \leq t_1 < t_2$, we have*

$$q_{t_1}(\alpha,\beta) \cdot q_{t_2}(\alpha,\beta) \leq \left(\frac{1}{\alpha - 4\beta}\right) q_{t_1+t_2+1}(\alpha,\beta).$$

Proof. From (91) in Lemma A.4, we have

$$\begin{aligned} q_{t_1}(\alpha,\beta) \cdot q_{t_2}(\alpha,\beta) = &\left(\frac{1}{\alpha - 4\beta}\right)^2 \left[\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+1} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+1}\right]^2 \\ &\cdot \left[\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_2+1} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_2+1}\right]^2. \end{aligned}$$

Since

$$0 \leq \frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2} < \frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2},$$

we have

$$\begin{aligned} &\left[\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+1} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+1}\right]\left[\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_2+1} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_2+1}\right] \\ &= \left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+t_2+2} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+1}\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_2+1} \\ &\quad - \left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+1}\left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_2+1} + \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+t_2+2} \\ &\leq \left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+t_2+2} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+t_2+2}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} q_{t_1}(\alpha,\beta) \cdot q_{t_2}(\alpha,\beta) &\leq \left(\frac{1}{\alpha - 4\beta}\right)^2 \left[\left(\frac{\sqrt{\alpha}}{2} + \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+t_2+2} - \left(\frac{\sqrt{\alpha}}{2} - \frac{\sqrt{\alpha - 4\beta}}{2}\right)^{t_1+t_2+2}\right]^2 \\ &= \left(\frac{1}{\alpha - 4\beta}\right) q_{t_1+t_2+1}(\alpha,\beta). \end{aligned}$$

This completes the proof. □