A Proof of Lemma 3

Using the notation $\hat{\mathbf{v}} = Y_{n-1}\mathbf{v}/\|Y_{n-1}\mathbf{v}\|$ and $A_n = \mathbf{x}_n\mathbf{x}_n^\top$, one can follow the analysis in Balsubramani et al. (2013) to show that $\Phi_n^{(\mathbf{v})} \le \Phi_{n-1}^{(\mathbf{v})} + \beta_n - Z_n$, with

- $\beta_n = 5\gamma_n^2 + 2\gamma_n^3$,
- $Z_n = 2\gamma_n(\hat{\mathbf{v}}^\top UU^\top A_n\hat{\mathbf{v}} - \|U^\top\hat{\mathbf{v}}\|^2\hat{\mathbf{v}}^\top A_n\hat{\mathbf{v}})$, and
- $\mathbb{E}\left[Z_n|\mathcal{F}_{n-1}\right] \ge 2\gamma_n(\lambda-\hat{\lambda})\Phi_{n-1}^{(\mathbf{v})}(1-\Phi_{n-1}^{(\mathbf{v})}) \ge 0$.

We omit the proof here as the adaptation is straightforward. It remains to show our better bound on $|Z_n|$. For this, note that

$$|Z_n| \le 2\gamma_n \left\|\hat{\mathbf{v}}^\top UU^\top - \|U^\top\hat{\mathbf{v}}\|^2\hat{\mathbf{v}}^\top\right\| \cdot \|A_n\hat{\mathbf{v}}\|,$$

where $\|A_n\hat{\mathbf{v}}\| \le 1$ and

$$\begin{aligned}&\left\|\hat{\mathbf{v}}^\top UU^\top - \|U^\top\hat{\mathbf{v}}\|^2\hat{\mathbf{v}}^\top\right\|^2\\ &\quad= \|U^\top\hat{\mathbf{v}}\|^2 - 2\|U^\top\hat{\mathbf{v}}\|^4 + \|U^\top\hat{\mathbf{v}}\|^4\\ &\quad= \|U^\top\hat{\mathbf{v}}\|^2\left(1-\|U^\top\hat{\mathbf{v}}\|^2\right).\end{aligned}$$

As $\|U^\top\hat{\mathbf{v}}\|^2 \le 1$ and $\left(1-\|U^\top\hat{\mathbf{v}}\|^2\right) = \Phi_{n-1}^{(\mathbf{v})}$, we have

$$|Z_n| \le 2\gamma_n\sqrt{\Phi_{n-1}^{(\mathbf{v})}}.$$

B Proof of Lemma 4

Assume that the event Γ_0 holds and consider any $n \in [n_0, n_1)$. We need the following, which we prove in Appendix B.1.

Proposition 1. *For any $n > m$ and any $\mathbf{v} \in \mathbb{R}^k$,*

$$\frac{\|U^\top Y_n\mathbf{v}\|}{\|Y_n\|} \ge \left(\frac{m}{n}\right)^{3c} \cdot \frac{\|U^\top Y_m\mathbf{v}\|}{\|Y_m\|}.$$

From Proposition 1, we know that for any $\mathbf{v} \in \mathcal{S}$,

$$\frac{\|U^\top Y_n\mathbf{v}\|}{\|Y_n\mathbf{v}\|} \ge \frac{\|U^\top Y_n\mathbf{v}\|}{\|Y_n\|} \ge \left(\frac{n_0}{n}\right)^{3c}\frac{\|U^\top Y_0\mathbf{v}\|}{\|Y_0\|},$$

where $(n_0/n)^{3c} \ge (n_0/n_1)^{3c} \ge (1/c_1)^{3c}$ for the constant c_1 given in Remark 1. As $Y_0 = Q_0$ and $\|Q_0\| = 1 = \|Q_0\mathbf{v}\|$, we obtain

$$\frac{\|U^\top Y_n\mathbf{v}\|}{\|Y_n\mathbf{v}\|} \ge \frac{\|U^\top Q_0\mathbf{v}\|}{c_1^{3c}\|Q_0\mathbf{v}\|} \ge \frac{\sqrt{1-\rho_0}}{c_1^{3c}} = \sqrt{\frac{\bar{c}}{c_1^{6c}kd}}.$$

Therefore, assuming Γ_0, we always have

$$\Phi_n = \max_{\mathbf{v}}\left(1 - \frac{\|U^\top Y_n\mathbf{v}\|^2}{\|Y_n\mathbf{v}\|^2}\right) \le 1 - \frac{\bar{c}}{c_1^{6c}kd} = \rho_1.$$

B.1 Proof of Proposition 1

Recall that for any n, $Y_n = Y_{n-1} + \gamma_n\mathbf{x}_n\mathbf{x}_n^\top Y_{n-1}$ and $\|\mathbf{x}_n\mathbf{x}_n^\top\| \le 1$. Then for any $\mathbf{v} \in \mathbb{R}^k$,

$$\frac{\|U^\top Y_n\mathbf{v}\|}{\|Y_n\|} \ge \frac{\|U^\top Y_{n-1}\mathbf{v}\| - \gamma_n\|U^\top Y_{n-1}\mathbf{v}\|}{\|Y_{n-1}\| + \gamma_n\|Y_{n-1}\|},$$

which is

$$\frac{1-\gamma_n}{1+\gamma_n} \cdot \frac{\|U^\top Y_{n-1}\mathbf{v}\|}{\|Y_{n-1}\|} \ge e^{-3\gamma_n}\frac{\|U^\top Y_{n-1}\mathbf{v}\|}{\|Y_{n-1}\|},$$

using the fact that $1-x \ge e^{-2x}$ for $x \le 1/2$ and $\gamma_n \le 1/2$. Then by induction, we have

$$\frac{\|U^\top Y_n\mathbf{v}\|}{\|Y_n\|} \ge e^{-3\sum_{t>m}^n \gamma_i} \cdot \frac{\|U^\top Y_m\mathbf{v}\|}{\|Y_m\|}.$$

The Proposition follows as

$$e^{-3\sum_{t>m}^n \gamma_i} = e^{-3c\sum_{t>m}^n \frac{1}{t}} \ge \left(\frac{m}{n}\right)^{3c}$$

using the fact that $\sum_{t>m}^n \frac{1}{t} \le \int_m^n \frac{1}{x}dx = \ln(\frac{n}{m})$.

C Proof of Lemma 5

According to Lemma 3, our $\Phi_n^{(\mathbf{v})}$'s satisfy the same recurrence relation as the functions Ψ_n's of Balsubramani et al. (2013). We can therefore have the following, which we prove in Appendix C.1.

Lemma 9. *Let $\hat{\rho}_i = \rho_i/\lceil e^{5/c_0}\rceil^{c_0(1-\rho_i)}$. Then for any $\mathbf{u} \in \mathcal{S}$ and $\alpha_i \ge 12c^2/n_{i-1}$,*

$$\Pr\left[\sup_{n\ge n_i}\Phi_n^{(\mathbf{u})} \ge \hat{\rho}_i + \alpha_i \mid \Gamma_i\right] \le e^{-\Omega((\alpha_i^2/(c^2\rho_i))n_{i-1})}.$$

Our goal is to bound $\Pr\left[\neg\Gamma_{i+1}|\Gamma_i\right]$, which is

$$\Pr\left[\exists\mathbf{v}\in\mathcal{S} : \sup_{n_i\le n<n_{i+1}}\Phi_n^{(\mathbf{v})} \ge \rho_{i+1}|\Gamma_i\right].$$

As discussed before, we cannot directly apply a union bound on the bound in Lemma 9 as there are infinitely many $\mathbf{v}$'s in $\mathcal{S}$. Instead, we look for a small "ϵ-net" $\mathcal{D}_i$ of $\mathcal{S}$, with the property that any $\mathbf{v} \in \mathcal{S}$ has some $\mathbf{u} \in \mathcal{D}_i$ with $\|\mathbf{v}-\mathbf{u}\| \le \epsilon$. Such a $\mathcal{D}_i$ with $|\mathcal{D}_i| \le (1/\epsilon)^{\mathcal{O}(k)}$ is known to exist (see e.g. Milman and Schechtman (1986)). Then what we need is that when $\mathbf{v}$ and $\mathbf{u}$ are close, $\Phi_n^{(\mathbf{v})}$ and $\Phi_n^{(\mathbf{u})}$ are close as well. This is guaranteed by the following, which we prove in Appendix C.2.

Lemma 10. *Suppose Γ_i happens. Then for any $n \in [n_i, n_{i+1})$, any $\epsilon \le \sqrt{1-\rho_i}/(2c_1^{6c})$, and any $\mathbf{u},\mathbf{v} \in \mathcal{S}$ with $\|\mathbf{u}-\mathbf{v}\| \le \epsilon$, we have*

$$\left|\Phi_n^{(\mathbf{v})} - \Phi_n^{(\mathbf{u})}\right| \le 16c_1^{6c}\epsilon/\sqrt{1-\rho_i}.$$

According to this, we can choose $\alpha_i = (\rho_{i+1} - \hat{\rho}_i)/2$ and $\epsilon = \alpha_i\sqrt{1-\rho_i}/(16c_1^{6c})$ so that with $\|\mathbf{u}-\mathbf{v}\| \leq \epsilon$, we have $|\Phi_n^{(\mathbf{v})} - \Phi_n^{(\mathbf{u})}| \leq \alpha_i$. This means that given any $\mathbf{v} \in \mathcal{S}$ with $\Phi_n^{(\mathbf{v})} \geq \rho_{i+1}$, there exists some $\mathbf{u} \in \mathcal{D}_i$ with $\Phi_n^{(\mathbf{u})} \geq \rho_{i+1} - \alpha_i = \hat{\rho}_i + \alpha_i$. As a result, we can now apply a union bound over $\mathcal{D}_i$ and have

$$\Pr\left[\neg\Gamma_{i+1}|\Gamma_i\right] \leq \sum_{\mathbf{u}\in\mathcal{D}_i} \Pr\left[\sup_{n\geq n_i} \Phi_n^{(\mathbf{u})} \geq \hat{\rho}_i + \alpha_i \mid \Gamma_i\right]. \tag{7}$$

To bound this further, consider the following two cases.

First, for the case of $i < \pi_1$, we have $\rho_i \geq 3/4$ and $\eta_i = 1 - \rho_i \leq 1/4$, so that

$$\hat{\rho}_i \leq \rho_i e^{-5(1-\rho_i)} = (1-\eta_i)e^{-5\eta_i} \leq e^{-6\eta_i} \leq 1 - 3\eta_i.$$

Then $\alpha_i \geq ((1-2\eta_i) - (1-3\eta_i))/2 = \eta_i/2$, which is at least $12c^2/n_{i-1}$, as $\eta_i \geq \eta_1 \geq \bar{c}/(c_1^{6c}kd)$ and $n_{i-1} \geq n_0 = \hat{c}^c k^3 d^2 \log d$ for a large enough constant $\hat{c}$. Therefore, we can apply Lemma 9 and the bound in (7) becomes

$$(c_1^c/\eta_i)^{\mathcal{O}(k)}\, e^{-\Omega((\eta_i^2/c^2)n_{i-1})} \leq \frac{\delta_0}{2(i+1)^2}.$$

Next, for the case of $i \geq \pi_1$, we have $\rho_i \leq 3/4$ so that

$$\hat{\rho}_i \leq \rho_i/\lceil e^{5/c_0}\rceil^{c_0/4} \leq \rho_i/\lceil e^{5/c_0}\rceil^3,$$

as $c_0 \geq 12$ by assumption. Since $\rho_{i+1} \geq \rho_i/\lceil e^{5/c_0}\rceil^2$, this gives us $\alpha_i \geq \rho_i(\lceil e^{5/c_0}\rceil^{-2} - \lceil e^{5/c_0}\rceil^{-3})/2$, which is at least $12c^2/n_{i-1}$, as ρ_i, according to our choice, is about $c_2(c^3k\log n_{i-1})/(n_{i-1}+1)$ for a large enough constant c_2. Thus, we can apply Lemma 9 and the bound in (7) becomes

$$(c_1^c/\rho_i)^{\mathcal{O}(k)}\, e^{-\Omega((\rho_i/c^2)n_{i-1})} \leq \frac{\delta_0}{2(i+1)^2}. \tag{8}$$

This completes the proof of Lemma 5.

C.1 Proof of Lemma 9

By Lemma 3, the random variables $\Phi_n^{(\mathbf{v})}$'s satisfy the same recurrence relation of Balsubramani et al. (2013) for their random variables Φ_n's. Thus, we can follow their analysis[1], but use our better bound on $|Z_n|$, and have the following.

First, when given Γ_i, we have $|Z_n| \leq 2\gamma_n\sqrt{\rho_i}$ for $n_{i-1} \leq n < n_i$. Then one can easily modify the analysis in Balsubramani et al. (2013) to show that for any $t \geq 0$,

$$\mathbb{E}\left[e^{t\Phi_{n_i}^{(\mathbf{v})}}|\Gamma_i\right] \leq \exp\left(t\hat{\rho}_i + c^2(6t+2t^2\rho_i)\left(\frac{1}{n_{i-1}} - \frac{1}{n_i}\right)\right),$$

by noting that $(n_i+1)/(n_{i-1}+1) = \lceil e^{5/c_0}\rceil$ and $n \geq n_0 = \hat{c}^c k^3 d^2 \log d$ according to our choice of parameters.

[1] In particular, their proofs for Lemma 2.9 and Lemma 2.10.

Next, following Balsubramani et al. (2013) and applying Doob's martingale inequality, we obtain

$$\begin{aligned}
&\Pr\left[\sup_{n\geq n_i} \Phi_n^{(\mathbf{v})} \geq \hat{\rho}_i + \alpha_i|\Gamma_i\right] \\
&\leq \mathbb{E}\left[e^{t\Phi_{n_i}^{(\mathbf{v})}}|\Gamma_i\right]\exp\left(-t(\hat{\rho}_i+\alpha_i) + \frac{c^2}{n_i}(6t+2t^2\rho_i)\right) \\
&\leq \exp\left(-t\alpha_i + \frac{c^2}{n_{i-1}}(6t+2t^2\rho_i)\right) \\
&\leq \exp\left(-\frac{t\alpha_i}{2} + \frac{2c^2t^2\rho_i}{n_{i-1}}\right),
\end{aligned}$$

as $\alpha_i \geq \frac{12c^2}{n_{i-1}}$. Finally, by choosing $t = \frac{\alpha_i n_{i-1}}{8c^2\rho_i}$, we have the lemma.

C.2 Proof of Lemma 10

Assume without loss of generality that $\Phi_n^{(\mathbf{v})} \leq \Phi_n^{(\mathbf{u})}$ (otherwise, we switch $\mathbf{v}$ and $\mathbf{u}$), so that

$$\left|\Phi_n^{(\mathbf{v})} - \Phi_n^{(\mathbf{u})}\right| = \frac{\|U^\top Y_n\mathbf{v}\|^2}{\|Y_n\mathbf{v}\|^2} - \frac{\|U^\top Y_n\mathbf{u}\|^2}{\|Y_n\mathbf{u}\|^2}.$$

As $\|\mathbf{v}-\mathbf{u}\| \leq \epsilon$, we have

$$\frac{\|U^\top Y_n\mathbf{v}\|}{\|Y_n\mathbf{v}\|} \leq \frac{\|U^\top Y_n\mathbf{u}\| + \epsilon\|U^\top Y_n\|}{\|Y_n\mathbf{u}\| - \epsilon\|Y_n\|}. \tag{9}$$

To relate this to $\frac{\|U^\top Y_n\mathbf{u}\|^2}{\|Y_n\mathbf{u}\|^2}$, we would like to express $\|U^\top Y_n\|$ in terms of $\|U^\top Y_n\mathbf{u}\|$ and $\|Y_n\|$ in terms of $\|Y_n\mathbf{u}\|$. For this, note that both $\|U^\top Y_n\mathbf{u}\|/\|U^\top Y_n\|$ and $\|Y_n\mathbf{u}\|/\|Y_n\|$ are at least $\|U^\top Y_n\mathbf{u}\|/\|Y_n\|$, which by Proposition 1 is at least

$$\left(\frac{n_{i-1}}{n}\right)^{3c}\frac{\|U^\top Y_{n_{i-1}}\mathbf{u}\|}{\|Y_{n_{i-1}}\|} \geq c_1^{-6c}\frac{\|U^\top Y_{n_{i-1}}\mathbf{u}\|}{\|Y_{n_{i-1}}\|}, \tag{10}$$

using the fact that $n_{i-1}/n \geq n_{i-1}/n_{i+1} \geq 1/c_1^2$. Then as $Y_{n_{i-1}} = Q_{n_{i-1}}$ and $\|Q_{n_{i-1}}\| = \|Q_{n_{i-1}}\mathbf{u}\|$, the righthand side of (10) becomes

$$c_1^{-6c}\frac{\|U^\top Q_{n_{i-1}}\mathbf{u}\|}{\|Q_{n_{i-1}}\mathbf{u}\|} = c_1^{-6c}\sqrt{1-\Phi_{n_{i-1}}^{(\mathbf{u})}} \geq c_1^{-6c}\sqrt{1-\rho_i},$$

given Γ_i. What we have obtained so far is a lower bound for both $\|U^\top Y_n\mathbf{u}\|/\|U^\top Y_n\|$ and $\|Y_n\mathbf{u}\|/\|Y_n\|$. Plugging this into (9), with $\hat{\epsilon} = \epsilon c_1^{6c}/\sqrt{1-\rho_i}$, we get

$$\frac{\|U^\top Y_n\mathbf{v}\|}{\|Y_n\mathbf{v}\|} \leq \frac{\|U^\top Y_n\mathbf{u}\|(1+\hat{\epsilon})}{\|Y_n\mathbf{u}\|(1-\hat{\epsilon})}.$$

As a result, we have

$$\left|\Phi_n^{(\mathbf{v})} - \Phi_n^{(\mathbf{u})}\right| \leq \frac{\|U^\top Y_n\mathbf{u}\|^2}{\|Y_n\mathbf{u}\|^2}\left(\frac{(1+\hat{\epsilon})^2}{(1-\hat{\epsilon})^2} - 1\right) \leq 16\hat{\epsilon},$$

since $\frac{(1+\hat{\epsilon})^2}{(1-\hat{\epsilon})^2} - 1 \leq \frac{4\hat{\epsilon}}{(1-\hat{\epsilon})^2} \leq 16\hat{\epsilon}$ for $\hat{\epsilon} \leq 1/2$.

D Proof of Lemma 7

As $\cos(U, Q_{i-1})^2 = \frac{1}{1+\tan(U,Q_{i-1})^2} \geq \frac{1}{1+\varepsilon_{i-1}^2} \geq \beta_i^2$, we have $\|G_i\| \leq \triangle\beta_i \leq \triangle\cos(U, Q_{i-1})$. Thus, we can apply Lemma 6 and have

$$\tan(U, AQ_{i-1} + G_i) \leq \max(\beta_i, \max(\beta_i, \gamma)\varepsilon_{i-1}),$$

which is at most $\max(\beta_i, \gamma\varepsilon_{i-1}) \leq \gamma\varepsilon_{i-1} = \varepsilon_i$. The lemma follows as $\tan(U, Q_i) = \tan(U, AQ_{i-1} + G_i)$.

E Proof of Lemma 8

Let $\rho = \triangle\beta_i$ and note that $\|G_i\| \leq \|A - F_i\|$, where F_i is the average of $|I_i|$ i.i.d. random matrices, each with mean A. Recall that $\|A\| \leq 1$ by Assumption 1. Then from a matrix Chernoff bound, we have

$$\Pr\left[\|G_i\| > \rho\right] \leq \Pr\left[\|A - F_i\| > \rho\right] \leq de^{-\Omega(\rho^2|I_i|)} \leq \delta_i,$$

for $|I_i|$ given in (3).

F Proof of Lemma 9

Let L be the iteration number such that $\varepsilon_{L-1} > \varepsilon$ and $\varepsilon_L \leq \varepsilon$. Note that with $\varepsilon_L = \varepsilon_0\gamma^L = \varepsilon_0(1 - (\lambda - \bar{\lambda})/\lambda)^{L/4} \leq \varepsilon_0 e^{-L(\lambda-\bar{\lambda})/(4\lambda)}$, we can have

$$L \leq \mathcal{O}\left(\frac{\lambda}{\lambda - \bar{\lambda}}\log\frac{\varepsilon_0}{\varepsilon}\right) \leq \mathcal{O}\left(\frac{\lambda}{\lambda - \bar{\lambda}}\log\frac{d}{\varepsilon}\right).$$

As the number of samples in iteration i is

$$|I_i| = \mathcal{O}\left(\frac{\log(d/\delta_i)}{(\lambda - \bar{\lambda})^2\beta_i^2}\right) \leq \mathcal{O}\left(\frac{\log(di)}{(\lambda - \bar{\lambda})^2\beta_i^2}\right),$$

the total number of samples needed is

$$\sum_{i=1}^{L}|I_i| \leq \mathcal{O}\left(\frac{\log(dL)}{(\lambda - \bar{\lambda})^2}\right) \cdot \sum_{i=1}^{L}\frac{1}{\beta_i^2}.$$

With $\beta_i = \min(\gamma/\sqrt{1+\varepsilon_{i-1}^2}, \gamma\varepsilon_{i-1})$, one sees that for some $i_0 \leq \mathcal{O}(\log d)$, $\beta_i = \gamma/\sqrt{1+\varepsilon_{i-1}^2}$ when $i \leq i_0$ and $\beta_i = \gamma\varepsilon_{i-1} = \varepsilon_i$ when $i > i_0$. This implies that

$$\sum_{i=1}^{L}\frac{1}{\beta_i^2} = \sum_{i=1}^{i_0}\frac{1+\varepsilon_{i-1}^2}{\gamma^2} + \sum_{i=i_0+1}^{L}\frac{1}{\varepsilon_i^2}, \quad (11)$$

where the first sum in the righthand side of (11) is

$$\frac{i_0}{\gamma^2} + \sum_{i=1}^{i_0}\varepsilon_0^2\gamma^{2i-4} \leq \frac{\mathcal{O}(\log d)}{\gamma^2} + \frac{\varepsilon_0^2}{\gamma^2(1-\gamma^2)},$$

while the second sum is

$$\sum_{i=i_0+1}^{L}\frac{\gamma^{2(L-i)}}{\varepsilon_L^2} \leq \frac{1}{(1-\gamma^2)\varepsilon_L^2} \leq \frac{1}{\gamma^2(1-\gamma^2)\varepsilon^2}$$

using the fact that $\varepsilon_L = \gamma\varepsilon_{L-1} \geq \gamma\varepsilon$. Since $\gamma^2 = \left(1 - \frac{\lambda-\bar{\lambda}}{\lambda}\right)^{1/2} \leq 1 - \frac{\lambda-\bar{\lambda}}{2\lambda}$, we have $\frac{1}{1-\gamma^2} \leq \frac{2\lambda}{\lambda-\bar{\lambda}}$, and since $\lambda \leq \mathcal{O}(\bar{\lambda})$, we also have $\frac{1}{\gamma^2} \leq \mathcal{O}(1)$. Moreover, as we assume that $\varepsilon \leq 1/\sqrt{kd}$, we can conclude that the total number of samples needed is at most

$$\sum_{i=1}^{L}|I_i| \leq \mathcal{O}\left(\frac{\log(dL)}{(\lambda-\bar{\lambda})^2}\right)\cdot\mathcal{O}\left(\frac{\lambda}{(\lambda-\bar{\lambda})\varepsilon^2}\right) \leq \mathcal{O}\left(\frac{\lambda\log(dL)}{\varepsilon^2(\lambda-\bar{\lambda})^3}\right).$$

References

Balsubramani, A., Dasgupta, S., and Freund, Y. (2013). The fast convergence of incremental pca. In *Advances in Neural Information Processing Systems*.

Milman, V. D. and Schechtman, G. (1986). *Asymptotic theory of finite-dimensional normed spaces*. Lecture Notes in Mathematics. Springer.