A Proof of Lemma 3

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

- $\beta_n = 5\gamma_{n}^2 + 2\gamma_{n}^3$,
- $Z_n = 2\gamma_n(\hat{V}U^TU^TA_nv - \|U^TV\|^2\hat{V}A_nv)$, and
- $E[Z_n|F_{n-1}] \geq 2\gamma_n(\lambda - \hat{\lambda})\Phi_{n-1}^{(v)}(1 - \Phi_{n-1}^{(v)}) \geq 0$.

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

$$|Z_n| \leq 2\gamma_n \|\hat{V}U^TU - \|U^TV\|^2\hat{V}\| : A_n\hat{V},$$

where $\|A_n\hat{V}\| \leq 1$ and

$$\|\hat{V}U^TU - \|U^TV\|^2\hat{V}\|^2 = \|U^TV\|^2 - 2\|U^TV\|^4 + \|U^TV\|^4 = \|U^TV\|^2(1 - \|U^TV\|^2).$$

As $\|U^TV\|^2 \leq 1$ and $(1 - \|U^TV\|^2) = \Phi_{n-1}^{(v)}$, we have

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

B Proof of Lemma 4

Assume that the event $\Gamma_0$ 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 $v \in \mathbb{R}^k$,

$$\frac{\|U^TY_nv\|}{\|Y_n\|} \geq \left(\frac{m}{n}\right)^{3c} \frac{\|U^TY_mv\|}{\|Y_m\|}.$$

From Proposition 1, we know that for any $v \in S$,

$$\frac{\|U^TY_nv\|}{\|Y_n\|} \geq (n_0/n)^{3c} \frac{\|U^TY_0v\|}{\|Y_0\|},$$

where $(n_0/n)^{3c} \geq (n_0/n_1)^{3c} \geq (1/c_1)^{3c}$ for the constant $c_1$ given in Remark 1. As $Y_0 = Q_0$ and $\|Q_0v\| = 1 = \|Q_0v\|$, we obtain

$$\frac{\|U^TY_nv\|}{\|Y_nv\|} \geq \frac{\|U^TY_0v\|}{\|Q_0v\|} \geq \sqrt{1 - \rho_i} = \sqrt{c_{1i}^2kd}$$

Therefore, assuming $\Gamma_0$, we always have

$$\Phi_n = \max_v \left(1 - \frac{\|U^TY_nv\|^2}{\|Y_nv\|^2}\right) \leq 1 - \frac{c_{1i}^2kd}{c_{1i}^2kd} = \rho_i.$$

B.1 Proof of Proposition 1

Recall that for any $n$, $Y_n = Y_{n-1} + \gamma_nX_nx_n^TY_{n-1}$ and $\|X_nx_n\| \leq 1$. Then for any $v \in \mathbb{R}^k$,

$$\frac{\|U^TY_nv\|}{\|Y_n\|} \geq \frac{\|U^TY_{n-1}v - \gamma_n\|U^TY_{n-1}v\|}{\|Y_{n-1}\| + \gamma_n\|Y_{n-1}\|},$$

which is

$$1 - \gamma_n \frac{\|U^TY_{n-1}v\|}{\|Y_{n-1}\|} \geq e^{-3\gamma_n} \frac{\|U^TY_{n-1}v\|}{\|Y_{n-1}\|},$$

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

$$\frac{\|U^TY_mv\|}{\|Y_m\|} \geq e^{-3\sum_{i=m}^{n} \gamma_i} \frac{\|U^TY_nv\|}{\|Y_n\|}.$$

The Proposition follows as

$$e^{-3\sum_{i=m}^{n} \gamma_i} = e^{-3c\sum_{i=m}^{n} \frac{i}{n}} \geq \left(\frac{m}{n}\right)^{3c}$$

using the fact that $\sum_{i=1}^{n} \frac{1}{i} \leq \int_{1}^{n} \frac{1}{x} dx = \ln(n)$.

C Proof of Lemma 5

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

**Lemma 9.** Let $\rho_i = \rho_i[\epsilon^{5/(3\alpha)}(1-\rho_i)]$. Then for any $v \in S$ and $\alpha_i \geq 12c_{1i}^{-2}/n_{i-1}$,

$$\Pr \left[ \sup_{n \geq n_i} \Phi_{n}^{(u)} \geq \rho_i + \alpha_i \mid \Gamma_i \right] \leq e^{-\Omega((\alpha_i/(c_{1i}^2\rho_i))n_{i-1})}.$$

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

$$\Pr \left[ \exists v \in S : \sup_{n_i \leq n < n_{i+1}} \Phi_{n}^{(v)} \geq \rho_i + 1 \mid \Gamma_i \right].$$

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

**Lemma 10.** Suppose $\Gamma_i$ happens. Then for any $n \in [n_i,n_{i+1})$, any $\epsilon \leq \sqrt{1 - \rho_i}/(2c_{1i}^2\rho_i)$, and any $u,v \in S$ with $|u - v| \leq \epsilon$, we have

$$|\Phi_{n}^{(v)} - \Phi_{n}^{(u)}| \leq 16c_{1i}^5c_{1i}^2\epsilon/\sqrt{1 - \rho_i}.$$
According to this, we can choose $\alpha_i = (\rho_{i+1} - \rho_i)/2$ and $\epsilon = \alpha_i\sqrt{1 - \rho_i}/(16c_i^3) \geq \epsilon$, so that $\|u - v\| \leq \epsilon$, we have $|\Phi_n^{(v)} - \Phi_n^{(u)}| \leq \alpha_i$. This means that given any $v \in S$ with $\Phi_n^{(v)} \geq \rho_{i+1}$, there exists some $u \in D_i$ with $\Phi_n^{(u)} \geq \rho_{i+1} - \alpha_i = \rho_i + \alpha_i$. As a result, we can now apply a union bound over $D_i$ and have

$$\Pr[-\Gamma_{i+1} | \Gamma_i] \leq \sum_{u \in D_i} \Pr\left[ \sup_{n \geq n_i} \Phi_n^{(u)} \geq \rho_i + \alpha_i | \Gamma_i \right].$$

(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 $\eta_1 = 1 - \rho_i \leq 1/4$, so that

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

Then $\alpha_i \geq (1 - 2\eta_i) / (1 - 3\eta_i) / 2 = \eta_i / 2$, which is at least $12c_i^2/n_{i-1}$, as $\eta_i \geq \eta_1 \geq 1/\epsilon(\epsilon_i^3kd)$ and $n_{i-1} \geq n_0 = \bar{c}_i k_3 d^2 \log d$ for a large enough constant $\bar{c}_i$. Therefore, we can apply Lemma 3 and the bound in (7) becomes

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

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

$$\bar{\rho}_i \leq \rho_i/[\epsilon^5/c_0] / 4 \leq \rho_i/[\epsilon^5/c_0]^3,$$

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

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

(8)

This completes the proof of Lemma 3.

C.1 Proof of Lemma 9

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

First, when given $\Gamma_i$, we have $|Z_n| \leq 2\gamma n \sqrt{\bar{\rho}_i}$ 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 $t \geq 0$,

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

by noting that $(n_i + 1)/(n_{i-1} + 1) = [\epsilon^5 / c_0]$ and $n \geq n_0 = \bar{c}_i k_3 d^2 \log d$ according to our choice of parameters.

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

$$\Pr \left[ \sup_{n \geq n_i} \Phi_n^{(v)} \geq \rho_i + \alpha_i | \Gamma_i \right] \leq \mathbb{E}\left[ e^{(\Phi_n^{(v)}) | \Gamma_i } \exp \left( -t(\bar{\rho}_i + \alpha_i) + \frac{c_2}{n_i} (6t + 2t^2 \rho_i) \right) \right] \leq \exp \left( -t(\bar{\rho}_i + \frac{c_2}{n_i} (6t + 2t^2 \rho_i) \right) \leq \exp \left( -\frac{t\alpha_i}{2} + \frac{2c_2^2 t^2 \rho_i}{n_{i-1}} \right),$$

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

C.2 Proof of Lemma 10

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

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

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

$$\frac{\|U^\top Y_n v\|^2}{\|Y_n v\|^2} \leq \frac{\|U^\top Y_n u\|^2 + c\|U^\top Y_n u\|}{\|Y_n u\|^2} - \epsilon \|Y_n u\|^2.$$  

(9)

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

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

(10)

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

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

given $\Gamma_i$. What we have obtained so far is a lower bound for both $\|U^\top Y_n u\| / \|Y_n u\|$ and $\|Y_n u\| / \|Y_n u\|$. Plugging this into (9), with $\epsilon = \epsilon c_1^3 / \sqrt{1 - \rho_i}$, we get

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

As a result, we have

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

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

In particular, their proofs for Lemma 2.9 and Lemma 2.10.
D Proof of Lemma \[7\]
As \(\cos(U, Q_i) - 1 = \frac{1}{1 + \tan(U, Q_i)} \geq \beta_i^2\), we have \(|G_i| \leq \Delta \beta_1 \leq \Delta \cos(U, Q_{i-1})\). Thus, we can apply Lemma \[5\] 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 \(\rho = \Delta \beta_i\) 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
\[
\Pr[|G_i| > \rho] \leq \Pr[|A - F_i| > \rho] \leq d e^{-\Omega(\rho^2|I_i|)} \leq \delta_i,
\]
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
\[
L \leq O\left(\frac{\lambda}{\lambda - \bar{\lambda}} \log \frac{\varepsilon_0}{\varepsilon}\right) \leq O\left(\frac{\lambda}{\lambda - \bar{\lambda}} \log d\right).
\]
As the number of samples in iteration \(i\) is
\[
|I_i| = O\left(\frac{\log(d) / \gamma_i}{(\lambda - \bar{\lambda})^2 \beta_i^2}\right) \leq O\left(\frac{\log(d) / \gamma_i}{(\lambda - \bar{\lambda})^2 \beta_i^2}\right),
\]
the total number of samples needed is
\[
\sum_{i=1}^{L} |I_i| \leq O\left(\frac{\log(dL)}{(\lambda - \bar{\lambda})^2}\right) \cdot \sum_{i=1}^{L} \frac{1}{\beta_i^2}.
\]
With \(\beta_i = \min(\gamma/\sqrt{1 + \varepsilon_{i-1}^2}, \gamma \varepsilon_{i-1})\), one sees that for some \(i_0 \leq O(\log d)\), \(\beta_i = \gamma/\sqrt{1 + \varepsilon_{i-1}^2}\) 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
\[
\frac{i_0}{\gamma^2} + \sum_{i=1}^{i_0} \varepsilon_{0}^2 \gamma^{2i-4} \leq O(\log d) \frac{\varepsilon_0^2}{\gamma^2} + \frac{\varepsilon_0^2}{\gamma^2(1 - \gamma^2)},
\]
while the second sum is
\[
\sum_{i=i_0+1}^{L} \gamma^{2(L-i)} \frac{1}{\varepsilon_i^2} \leq \frac{1}{(1 - \gamma^2)\varepsilon_L^2} \leq \frac{1}{\gamma^2(1 - \gamma^2)\varepsilon^2},
\]
using the fact that \(\varepsilon_L = \gamma \varepsilon_{L-1} \geq \gamma \varepsilon\). Since \(\gamma^2 = (1 - \frac{\lambda - \bar{\lambda}}{\lambda})^{1/2} \leq 1 - \frac{\lambda - \bar{\lambda}}{2\lambda}\), we have \(\frac{1}{1 - \gamma^2} \leq \frac{2}{\lambda - \bar{\lambda}}\), and since \(\lambda \leq O(\bar{\lambda})\), we also have \(\frac{1}{\gamma^2} \leq O(1)\). Moreover, as we assume that \(\varepsilon \leq 1/\sqrt{kd}\), we can conclude that the total number of samples needed is at most
\[
\sum_{i=1}^{L} |I_i| \leq O\left(\frac{\log(dL)}{(\lambda - \bar{\lambda})^2}\right) O\left(\frac{\lambda}{(\lambda - \bar{\lambda})^2 \varepsilon^2}\right) \leq O\left(\frac{\lambda \log(dL)}{\varepsilon^2(\lambda - \bar{\lambda})^3}\right).
\]

References