A Non-asymptotic singular value bounds

Proposition A.1 (Theorem 1.1 of [RV09]). *Let* $\mathbf{A}$ *be an* $d \times r$ *matrix,* $d \geq r$*, whose entries are independently drawn from* $\mathcal{N}(0,1)$*. Then for every* $\tau \geq 0$*,*

$$Pr\left(\sigma_r(\mathbf{A}) \leq \tau(\sqrt{d} - \sqrt{r-1})\right) \leq (C_1\tau)^{d-r+1} + e^{-C_2 d}$$

where $C_1, C_2 > 0$ *are universal constants.*

Proposition A.2 ([Ver10]). *Let* $\mathbf{A}$ *be an* $d \times r$ *matrix whose entries are independently drawn from* $\mathcal{N}(0,1)$*. Then for every* $t \geq 0$*, with probability at least* $1 - \exp(-t^2/2)$*, we have*

$$\sigma_r(\mathbf{A}) \geq \sqrt{d} - \sqrt{r} - t$$

and for every $t \geq 0$*, with probability at least* $1 - \exp(-t^2/2)$*, we have*

$$\sigma_1(\mathbf{A}) \leq \sqrt{d} + \sqrt{r} + t$$

B Proof of Proposition 4.2

First, observe that by assumption, $\|\mathbf{X}_0\|^2, \|\mathbf{Y}_0\|^2 \leq \frac{9}{16\eta} \leq \frac{1}{\eta}$. Now, suppose that that $\|\mathbf{X}_0\|^2, \|\mathbf{Y}_0\|^2 \leq \frac{9}{16\eta}$ and $\|\mathbf{X}_t\|^2, \|\mathbf{Y}_t\|^2 \leq \frac{1}{\eta}$ for $t = 0, \ldots T-1$, and $1 \leq T \leq \lfloor \frac{1}{32\eta^2 f_0} \rfloor$. Then by Lemma 4.1,

$$\sum_{t=0}^{T-1} \left\| \nabla_{\mathrm{x}} f(\mathbf{X}_t, \mathbf{Y}_t) \right\|_{\mathrm{F}}^2 \leq \frac{2}{\eta} f_0. \tag{10}$$

Hence,

$$\begin{aligned}\|\mathbf{X}_T - \mathbf{X}_0\| &\leq \eta \left\| \sum_{t=0}^{T-1} \nabla_{\mathrm{x}} f(\mathbf{X}_t, \mathbf{Y}_t) \right\| \leq \eta \left\| \sum_{t=0}^{T-1} \nabla_{\mathrm{x}} f(\mathbf{X}_t, \mathbf{Y}_t) \right\|_{\mathrm{F}} \\ &\leq \eta \sqrt{\sum_{t=0}^{T-1} \|\nabla_{\mathrm{x}} f(\mathbf{X}_t, \mathbf{Y}_t)\|_F^2} \leq \eta \sqrt{\frac{2}{\eta} f_0} = \sqrt{2\eta f_0}. \end{aligned} \tag{11}$$

Then, for $T \leq T_*$, $\|\mathbf{X}_T\| \leq \|\mathbf{X}_0\| + \|\mathbf{X}_T - \mathbf{X}_0\| \leq \frac{3}{4\sqrt{\eta}} + \sqrt{2T\eta f_0} \leq \frac{1}{\sqrt{\eta}}$. It follows that $\|\mathbf{X}_t\|^2 \leq \frac{1}{\eta}$ for $t = 0, \ldots T$. Using Lemma 4.1 again, repeating the same argument,

$$\|\mathbf{Y}_t\| \leq \frac{1}{\sqrt{\eta}}, \quad t = 0, \ldots, T.$$

Iterate the induction until $T = T_* = \lfloor \frac{1}{32\eta^2 f_0} \rfloor$, to obtain $\|\mathbf{X}_t\|^2, \|\mathbf{Y}_t\|^2 \leq \frac{1}{\eta}$ for $t = 1, \ldots, T_*$.

Because $\|\mathbf{X}_T - \mathbf{X}_0\| \leq \sqrt{2\eta T f_0}$ for $T \leq T_* = \lfloor \frac{1}{32\eta^2 f_0} \rfloor$,

$$\sigma_r(\mathbf{X}_T) \geq \sigma_r(\mathbf{X}_0) - \|\mathbf{X}_T - \mathbf{X}_0\|; \qquad \sigma_1(\mathbf{X}_T) \leq \sigma_1(\mathbf{X}_0) + \|\mathbf{X}_T - \mathbf{X}_0\|.$$

A similar argument applies to achieve the stated bounds for $\sigma_r(\mathbf{Y}_T)$ and $\sigma_1(\mathbf{Y}_T)$.

C Proof of Proposition 4.3

Write the SVD $\mathbf{A} = \mathbf{U}_{m\times r}\mathbf{\Sigma}_{r\times r}\mathbf{V}_{r\times n}^{\mathsf{T}}$ so that $\mathbf{A\Phi}_1 = \mathbf{U}_{m\times r}\mathbf{\Sigma}_{r\times r}(\mathbf{V}^{\mathsf{T}}\Phi_1)$. Note that $\mathbf{V}^{\mathsf{T}}\mathbf{\Phi}_1 \in \mathbb{R}^{r\times d}$ has i.i.d. Gaussian entries $\mathcal{N}(0, \frac{1}{d})$. By Proposition A.1, with probability at least $1 - (C_1\epsilon)^{d-r+1} - e^{-C_2 d}$,

$$\sigma_r(\mathbf{V}^{\mathsf{T}}\mathbf{\Phi}_1) \geq \epsilon \left(1 - \frac{\sqrt{r-1}}{\sqrt{d}}\right)$$

On the other hand, Proposition A.2 implies that with probability at least $1 - e^{-r/2} - e^{-d/2}$,

$$\sigma_1(\mathbf{\Phi}_1) \le \left(1 + \frac{2\sqrt{r}}{\sqrt{d}}\right) \le 3, \quad \text{and} \quad \sigma_1(\mathbf{\Phi}_2) \le \left(1 + \frac{2\sqrt{d}}{\sqrt{m}}\right) \le 3.$$

If all aforementioned events hold, $\sigma_1(\mathbf{V}^\intercal \mathbf{\Phi}_1) \le \sigma_1(\mathbf{V})\sigma_1(\mathbf{\Phi}_1) \le 3$, and

$$\frac{\epsilon\left(1 - \frac{\sqrt{r-1}}{\sqrt{d}}\right)}{\sqrt{\eta} C \sigma_1(\mathbf{A})}\sigma_r(\mathbf{A}) \le \sigma_r(\mathbf{X}_0) \le \sigma_1(\mathbf{X}_0) \le \frac{3}{\sqrt{\eta} C}, \quad \sigma_1(\mathbf{Y}_0) \le 3\sqrt{\eta} D \sigma_1(\mathbf{A}) \le \frac{\sqrt{\eta} C \nu \sigma_1(\mathbf{A})}{3}.$$

where the last inequality uses $D \le \frac{C\nu}{9}$. Consequently,

$$1 - \nu \le 1 - \frac{D}{C}\sigma_1(\mathbf{\Phi}_1)\sigma_1(\mathbf{\Phi}_2) \le \left\| \mathbf{I} - \frac{D}{C}\mathbf{\Phi}_1 \mathbf{\Phi}_2^\intercal \right\| \le 1 + \frac{D}{C}\sigma_1(\mathbf{\Phi}_1)\sigma_1(\mathbf{\Phi}_2) \le 1 + \nu.$$

Hence,

$$2f(\mathbf{X}_0, \mathbf{Y}_0) = \left\| \mathbf{A}(\mathbf{I} - \frac{D}{C}\mathbf{\Phi}_1 \mathbf{\Phi}_2^\intercal) \right\|_{\mathrm{F}}^2 \le (1+\nu)^2 \|\mathbf{A}\|_{\mathrm{F}}^2$$

$$2f(\mathbf{X}_0, \mathbf{Y}_0) \ge \sigma_{\min}^2(\mathbf{I} - \frac{D}{C}\mathbf{\Phi}_1 \mathbf{\Phi}_2^\intercal)\|\mathbf{A}\|_F^2 \ge \frac{1}{(1-\nu)^2}\|\mathbf{A}\|_F^2 \tag{12}$$

D Proof of Corollary 5.3

Set $\beta_1 = \beta$ as in (A2c). Set $f_{0(1)} = f_0$.

By Corollary 5.2, iterating Assumption 1 for $T_1 = \lfloor \frac{\beta_1}{8\eta^2 f_{0(1)}} \rfloor$ iterations with step-size

$$\eta \le \frac{\beta_1}{\sqrt{32 f_{0(1)} \log(1/\epsilon)}}$$

guarantees that

$$\frac{1}{2}\|\mathbf{A} - \mathbf{X}_{T_1}\mathbf{Y}_{T_1}'\|_{\mathrm{F}}^2 \le f_{0(2)} := \epsilon f_{0(1)};$$

$$\|\sigma_r(\mathbf{X}_{T_1})\|^2 \ge \frac{1}{4}\frac{\beta_1}{\eta}.$$

This means that at time T_1, we can restart the analysis, and appeal again to Proposition 4.2 with modified parameters

- $f(\mathbf{X}_{T_1}, \mathbf{Y}_{T_1}) \le f_{02} := \epsilon f_{01}$,
- $\beta_2 := \frac{\beta_1}{4}$.

Corollary 5.2 again guarantees that provided

$$\eta \le \frac{\beta_2}{\sqrt{32 f_{0(2)} \log(1/\epsilon)}} = \frac{1}{4\sqrt{\epsilon}} \frac{\beta_1}{\sqrt{32 f_{0(1)} \log(1/\epsilon)}} \tag{13}$$

then $f(\mathbf{X}_{T_1+T_2}, \mathbf{Y}_{T_1+T_2}) \le \epsilon f(\mathbf{X}_{T_1}, \mathbf{Y}_{T_1}) \le \epsilon^2 f(\mathbf{X}_0, \mathbf{Y}_0)$ where

$$T_2 = \frac{T_1}{4\epsilon}. \tag{14}$$

We have that (13) is satisfied by assumption as we assume $\epsilon \le \frac{1}{16}$. Repeating this inductively, we find that after $T = T_1 + \cdots + T_k = T_1 \sum_{\ell=0}^{k-1} (\frac{1}{4\epsilon})^\ell \le T_1 (\frac{1}{4\epsilon})^k$ iterations, we are guaranteed that $f(\mathbf{X}_T, \mathbf{Y}_T) \le \epsilon^k f(\mathbf{X}_0, \mathbf{Y}_0)$. This is valid for any $T \in \mathbb{N}$ because we may always apply

Proposition 4.2 in light of summability and $C \geq 8$: for any t,

$$\begin{aligned}
\sigma_1(\mathbf{X}_t) &\leq \sigma_1(\mathbf{X}_0) + \sqrt{\eta}\sum_{j=1}^{k}\sqrt{2T_k f_{0(k)}} \\
&\leq \frac{3}{8\sqrt{\eta}} + \frac{1}{\sqrt{\eta}}\sum_{j=1}^{k}\sqrt{2(1/(4\epsilon))^j \frac{\beta_1}{8f_{0_1}}\epsilon^j f_{0(1)}} \\
&\leq \frac{3}{8\sqrt{\eta}} + \frac{\sqrt{\beta}}{2\sqrt{\eta}}\sum_{j=1}^{k}(1/2)^j \\
&\leq \frac{3+4\sqrt{\beta}}{8\sqrt{\eta}} \leq \frac{1}{2\sqrt{\eta}}.
\end{aligned}$$