A Proofs of Lemmas in Section 4

In this section, we restate the Lemmas in Section 4 which were used to prove Theorem 3.2, and present their proofs.

First, we prove lemmas which show a lower bound and a upper bound on the eigenvalues of the intermediate matrices $\mathbf{U_t}$ in Algorithm 1. This shows $\mathbf{U_t}$ always stay away from the surface where unwanted stationary point locate.

Lemma A.1 (Restatement of Lemma 4.2). *Suppose* $\eta < \frac{c\min\left(\sigma_{\min}(\mathbf{U_0}),\sqrt{\sigma_{\min}(\mathbf{M})}/10\right)}{\max\left(\|\mathbf{U_0}\|_2^3,\left(3\|\mathbf{M}\|_2\right)^{3/2}\right)}$*, where c is a small enough constant. Then, for every $t\in[T-1]$, we have $\mathbf{U_t}$ be a PD matrix with*

$$\lambda_{\min}\left(\mathbf{U_t}\right)\geq\min\left(\sigma_{\min}\left(\mathbf{U_0}\right),\frac{\sqrt{\sigma_{\min}\left(\mathbf{M}\right)}}{10}\right).$$

Proof. We will prove the lemma by induction. The base case $t=0$ holds trivially. Suppose the lemma holds for some t. We will now prove that it holds for $t+1$. We have

$$\begin{aligned}\lambda_{\min}\left(\mathbf{U_{t+1}}\right)=&\lambda_{\min}\left(\mathbf{U_t}-\eta\left(\mathbf{U_t}^2-\mathbf{M}\right)\mathbf{U_t}-\eta\mathbf{U_t}\left(\mathbf{U_t}^2-\mathbf{M}\right)\right)\\\geq&\lambda_{\min}\left(\frac{3}{4}\mathbf{U_t}-2\eta\mathbf{U_t}^3\right)+\lambda_{\min}\left(\frac{1}{4}\mathbf{U_t}+\eta(\mathbf{M}\mathbf{U_t}+\mathbf{U_t}\mathbf{M})\right)\\=&\lambda_{\min}\left(\frac{3}{4}\mathbf{U_t}-2\eta\mathbf{U_t}^3\right)+\lambda_{\min}\left(\left(\frac{1}{2}\mathbf{I}+\eta\mathbf{M}\right)\mathbf{U_t}\left(\frac{1}{2}\mathbf{I}+\eta\mathbf{M}\right)-\eta^2\mathbf{M}\mathbf{U_t}\mathbf{M}\right)\end{aligned}\tag{8}$$

When $\eta\leq\frac{1}{100\max\left(\|\mathbf{U_0}\|_2^2,3\|\mathbf{M}\|_2\right)}$, using Lemma 4.3 we can bound the first term as

$$\lambda_{\min}\left(\frac{3}{4}\mathbf{U_t}-2\eta\mathbf{U_t}^3\right)=\frac{3}{4}\sigma_{\min}\left(\mathbf{U_t}\right)-2\eta\sigma_{\min}\left(\mathbf{U_t}\right)^3.\tag{9}$$

To bound the second term, for any vector $\mathbf{w}\in\mathbb{R}^n$ with $\|\mathbf{w}\|_2=1$, let $\mathbf{w}=\sum_i^n\alpha_i\mathbf{v}_i$, where $\mathbf{v}_i$ is the i^{th} eigenvector of $\mathbf{M}$, and $\sum_{i=1}^n\alpha_i^2=1$. Then:

$$\begin{aligned}&\mathbf{w}^\top\left(\left(\frac{1}{2}\mathbf{I}+\eta\mathbf{M}\right)\mathbf{U_t}\left(\frac{1}{2}\mathbf{I}+\eta\mathbf{M}\right)-\eta^2\mathbf{M}\mathbf{U_t}\mathbf{M}\right)\mathbf{w}\\\geq&\lambda_{\min}\left(\mathbf{U_t}\right)\left\|\left(\frac{1}{2}\mathbf{I}+\eta\mathbf{M}\right)\mathbf{w}\right\|_2^2-\eta^2\left\|\mathbf{U_t}\right\|_2\left\|\mathbf{M}\mathbf{w}\right\|_2^2\\=&\lambda_{\min}\left(\mathbf{U_t}\right)\sum_{i=1}^n\left(\frac{1}{2}+\eta\lambda_i\left(\mathbf{M}\right)\right)^2\alpha_i^2-\eta^2\left\|\mathbf{U_t}\right\|_2\sum_{i=1}^n\lambda_i\left(\mathbf{M}\right)^2\alpha_i^2\\=&\lambda_{\min}\left(\mathbf{U_t}\right)\sum_{i=1}^n\alpha_i^2\left(\frac{1}{4}+\eta\lambda_i\left(\mathbf{M}\right)+\eta^2\lambda_i\left(\mathbf{M}\right)^2-\eta^2\kappa\left(\mathbf{U_t}\right)\lambda_i\left(\mathbf{M}\right)^2\right)\\\overset{(\zeta_1)}{=}&\sigma_{\min}\left(\mathbf{U_t}\right)\sum_{i=1}^n\alpha_i^2\left(\frac{1}{4}+\eta\sigma_i\left(\mathbf{M}\right)+\eta^2\sigma_i\left(\mathbf{M}\right)^2-\eta^2\kappa\left(\mathbf{U_t}\right)\sigma_i\left(\mathbf{M}\right)^2\right)\\\overset{(\zeta_2)}{\geq}&\sigma_{\min}\left(\mathbf{U_t}\right)\sum_{i=1}^n\alpha_i^2\left(\frac{1}{4}+\frac{1}{2}\eta\sigma_i\left(\mathbf{M}\right)\right)\geq\sigma_{\min}\left(\mathbf{U_t}\right)\left(\frac{1}{4}+\frac{1}{2}\eta\sigma_{\min}\left(\mathbf{M}\right)\right),\end{aligned}\tag{10}$$

where (ζ_1) is due to the fact that $\mathbf{U_t}$ is a PD matrix, so $\lambda_{\min}\left(\mathbf{U_t}\right)=\sigma_{\min}\left(\mathbf{U_t}\right)\geq0$, and (ζ_2) is because since $\eta\leq\frac{\min\left(\sigma_{\min}(\mathbf{U_0}),\sqrt{\sigma_{\min}(\mathbf{M})}/10\right)}{\max\left(\|\mathbf{U_0}\|_2^3,\left(3\|\mathbf{M}\|_2\right)^{3/2}\right)}\leq\frac{1}{2\kappa(\mathbf{U_t})\|\mathbf{M}\|_2}$, we have $\eta^2\kappa\left(\mathbf{U_t}\right)\sigma_i\left(\mathbf{M}\right)^2\leq\frac{1}{2}\eta\sigma_i\left(\mathbf{M}\right)$.

Plugging Eq.(9) and Eq.(10) into Eq.(8), we have:

$$\lambda_{\min}\left(\mathbf{U_{t+1}}\right)\geq\sigma_{\min}\left(\mathbf{U_t}\right)(1+\frac{1}{2}\eta\sigma_{\min}\left(\mathbf{M}\right)-2\eta\sigma_{\min}\left(\mathbf{U_t}\right)^2)$$

When $\sigma_{\min}(\mathbf{U_t}) \leq \sqrt{\sigma_{\min}(\mathbf{M})}/3$, we obtain:

$$\lambda_{\min}(\mathbf{U_{t+1}}) \geq \sigma_{\min}(\mathbf{U_t}) \geq \max\left(\sigma_{\min}(\mathbf{U_0}), \frac{\sqrt{\sigma_{\min}(\mathbf{M})}}{10}\right),$$

and when $\sigma_{\min}(\mathbf{U_t}) \geq \sqrt{\sigma_{\min}(\mathbf{M})}/3$, we have:

$$\begin{aligned}\lambda_{\min}(\mathbf{U_{t+1}}) &\geq \sigma_{\min}(\mathbf{U_t})\left(1 - 2\eta\sigma_{\min}(\mathbf{U_t})^2\right) \\ &\geq \sigma_{\min}(\mathbf{U_t})\left(1 - 2\eta\left\|\mathbf{U_t}\right\|_2^2\right) \geq \frac{9}{10}\sigma_{\min}(\mathbf{U_t}) \geq \min\left(\sigma_{\min}(\mathbf{U_0}), \frac{\sqrt{\sigma_{\min}(\mathbf{M})}}{10}\right).\end{aligned}$$

This concludes the proof. □

Lemma A.2 (Restatement of Lemma 4.3)**.** *Suppose* $\eta < \frac{1}{10\max\left(\|\mathbf{U_0}\|_2^2, 3\|\mathbf{M}\|_2\right)}$. *For every* $t \in [T-1]$, *we have:*

$$\left\|\mathbf{U_t}\right\|_2 \leq \max\left(\left\|\mathbf{U_0}\right\|_2, \sqrt{3\left\|\mathbf{M}\right\|_2}\right).$$

Proof. We will prove the lemma by induction. The base case $t = 0$ is trivially true. Supposing the statement is true for $\mathbf{U_t}$, we will prove it for $\mathbf{U_{t+1}}$.

Using the update equation of Algorithm 1, we have:

$$\begin{aligned}\left\|\mathbf{U_{t+1}}\right\|_2 &= \left\|\mathbf{U_t} - \eta\left(\mathbf{U_t}^2 - \mathbf{M}\right)\mathbf{U_t} - \eta\mathbf{U_t}\left(\mathbf{U_t}^2 - \mathbf{M}\right)\right\|_2 \\ &= \left\|\left(\mathbf{I} - 2\eta\mathbf{U_t}^2\right)\mathbf{U_t} + \eta\mathbf{M}\mathbf{U_t} + \eta\mathbf{U_t}\mathbf{M}\right\|_2 \\ &\leq \left\|\left(\mathbf{I} - 2\eta\mathbf{U_t}^2\right)\mathbf{U_t}\right\|_2 + 2\eta\left\|\mathbf{M}\right\|_2\left\|\mathbf{U_t}\right\|_2. \end{aligned} \tag{11}$$

The singular values of the matrix $\left(\mathbf{I} - 2\eta\mathbf{U_t}^2\right)\mathbf{U_t}$ are exactly $(1 - 2\eta\sigma^2)\cdot\sigma$ where σ is a singular value of $\mathbf{U_t}$. For $\sigma \leq \sqrt{2\left\|\mathbf{M}\right\|_2}$, we clearly have $(1 - 2\eta\sigma^2)\sigma \leq \sqrt{2\left\|\mathbf{M}\right\|_2}$. On the other hand, for $\sigma > \sqrt{2\left\|\mathbf{M}\right\|_2}$, we have $(1 - 2\eta\sigma^2)\sigma < (1 - 4\eta\left\|\mathbf{M}\right\|_2)\sigma$. Plugging this observation into Eq.(11), we obtain:

$$\begin{aligned}\left\|\mathbf{U_{t+1}}\right\|_2 &\leq \max\left(\sqrt{2\left\|\mathbf{M}\right\|_2}, \left(1 - 4\eta\left\|\mathbf{M}\right\|_2\right)\left\|\mathbf{U_t}\right\|_2\right) + 2\eta\left\|\mathbf{M}\right\|_2\left\|\mathbf{U_t}\right\|_2 \\ &\leq \max\left(\sqrt{2\left\|\mathbf{M}\right\|_2} + \frac{1}{15}\left\|\mathbf{U_t}\right\|_2, \left\|\mathbf{U_t}\right\|_2\right) \leq \max\left(\left\|\mathbf{U_0}\right\|_2, \sqrt{3\left\|\mathbf{M}\right\|_2}\right),\end{aligned}$$

where we used the inductive hypothesis in the last step. This proves the lemma. □

Finally, we prove the smoothness and gradient dominance in above regions.

Lemma A.3 (Restatement of Lemma 4.4)**.** *For any* $\mathbf{U_1}, \mathbf{U_2} \in \{\mathbf{U} | \left\|\mathbf{U}\right\|_2^2 \leq \Gamma\}$, *we have function* $f(\mathbf{U}) = \left\|\mathbf{M} - \mathbf{U}^2\right\|_F^2$ *satisfying:*

$$\|\nabla f(\mathbf{U_1}) - \nabla f(\mathbf{U_2})\|_F \leq 8\max\{\Gamma, \|\mathbf{M}\|_2\}\|\mathbf{U_1} - \mathbf{U_2}\|_F \tag{12}$$

Proof. By expanding gradient $\nabla f(\mathbf{U})$, and reordering terms, we have:

$$\begin{aligned}&\|\nabla f(\mathbf{U_1}) - \nabla f(\mathbf{U_2})\|_F \\ =&\|(2\mathbf{U_1}^3 - \mathbf{M}\mathbf{U_1} - \mathbf{U_1}\mathbf{M}) - (2\mathbf{U_2}^3 - \mathbf{M}\mathbf{U_2} - \mathbf{U_2}\mathbf{M})\|_F \\ =&\|2(\mathbf{U_1}^3 - \mathbf{U_2}^3) - \mathbf{M}(\mathbf{U_1} - \mathbf{U_2}) - (\mathbf{U_1} - \mathbf{U_2})\mathbf{M}\|_F \\ \leq&2\|\mathbf{M}\|_2\|\mathbf{U_1} - \mathbf{U_2}\|_F + 2\|\mathbf{U_1}^3 - \mathbf{U_2}^3\|_F \\ =&2\|\mathbf{M}\|_2\|\mathbf{U_1} - \mathbf{U_2}\|_F + 2\|\mathbf{U_1}^2(\mathbf{U_1} - \mathbf{U_2}) + \mathbf{U_1}(\mathbf{U_1} - \mathbf{U_2})\mathbf{U_2} + (\mathbf{U_1} - \mathbf{U_2})\mathbf{U_2}^2\|_F \\ \leq&2\|\mathbf{M}\|_2\|\mathbf{U_1} - \mathbf{U_2}\|_F + 6\Gamma\|\mathbf{U_1} - \mathbf{U_2}\|_F \\ \leq&8\max\{\Gamma, \|\mathbf{M}\|_2\}\|\mathbf{U_1} - \mathbf{U_2}\|_F\end{aligned}$$

□

Lemma A.4 (Restatement of Lemma 4.5)**.** *For any* $\mathbf{U} \in \{\mathbf{U} | \sigma_{\min}(\mathbf{U})^2 \geq \gamma\}$*, we have function* $f(\mathbf{U}) = \left\|\mathbf{M} - \mathbf{U}^2\right\|_F^2$ *satisfying:*

$$\|\nabla f(\mathbf{U})\|_F^2 \geq 4\gamma f(\mathbf{U}) \tag{13}$$

Proof. By expanding gradient $\nabla f(\mathbf{U})$, we have:

$$\begin{aligned}
&\|\nabla f(\mathbf{U})\|_F^2 = \|(\mathbf{U}^2 - \mathbf{M})\mathbf{U} + \mathbf{U}(\mathbf{U}^2 - \mathbf{M})\|_F^2 \\
=&\langle(\mathbf{U}^2 - \mathbf{M})\mathbf{U} + \mathbf{U}(\mathbf{U}^2 - \mathbf{M}), (\mathbf{U}^2 - \mathbf{M})\mathbf{U} + \mathbf{U}(\mathbf{U}^2 - \mathbf{M})\rangle \\
\geq&4\sigma_{\min}^2(\mathbf{U})\|\mathbf{U}^2 - \mathbf{M}\|_F^2 = 4\gamma f(\mathbf{U})
\end{aligned}$$

□

B Proof of Theorem 3.4

In this section, we will prove Theorem 3.4. We first state a useful lemma which is a stronger version of Lemma 4.2

Lemma B.1. *Suppose* $\mathbf{U_t}$ *is a PD matrix with* $\|\mathbf{U_t}\|_2 \leq \max(\|\mathbf{U_0}\|_2, \sqrt{3\|\mathbf{M}\|_2})$*, and* $\sigma_{\min}(\mathbf{U_t}) \geq \min(\sigma_{\min}(\mathbf{U_0}), \frac{1}{10}\sqrt{\sigma_{\min}(\mathbf{M})})$*. Suppose further that* $\eta < \frac{c\min(\sigma_{\min}(\mathbf{U_0}), \frac{1}{10}\sqrt{\sigma_{\min}(\mathbf{M})})}{\max(\|\mathbf{U_0}\|_2, \sqrt{3\|\mathbf{M}\|_2})^{3/2}}$*, where* c *is a small enough constant and denote* $\mathbf{U_{t+1}} \triangleq \mathbf{U_t} - \eta\left(\mathbf{U_t}^2 - \mathbf{M}\right)\mathbf{U_t} - \eta\mathbf{U_t}\left(\mathbf{U_t}^2 - \mathbf{M}\right)$*. Then,* $\mathbf{U_{t+1}}$ *is a PD matrix with:*

$$\lambda_{\min}(\mathbf{U_{t+1}}) \geq \left(1 + \frac{\eta\sigma_{\min}(\mathbf{M})}{15}\right)\min(\sigma_{\min}(\mathbf{U_0}), \frac{1}{10}\sqrt{\sigma_{\min}(\mathbf{M})}).$$

Indeed, our proof of Lemma 4.2 already proves this stronger result. Now we are ready to prove Theorem 3.4.

Proof of Theorem 3.4. The proof of the theorem is a fairly straight forward modification of the proof of Theorem 3.2. We will be terse since for most part we will use the arguments employed in the proofs of Theorem 3.2 and Lemmas 4.3 and 4.2.

We have the following two claims, which are robust versions of Lemmas 4.3 and 4.2, bounding the spectral norm and smallest eigenvalue of intermediate iterates. The proofs will be provided after the proof of the theorem.

Claim B.2. *For every* $t \in [T-1]$*, we have:*

$$\|\mathbf{U_t}\|_2 \leq \max(\|\mathbf{U_0}\|_2, \sqrt{3\|\mathbf{M}\|_2}).$$

Claim B.3. *For every* $t \in [T-1]$*, we have* $\mathbf{U_t}$ *be a PD matrix with*

$$\sigma_{\min}(\mathbf{U_t}) \geq \min(\sigma_{\min}(\mathbf{U_0}), \frac{\sqrt{\sigma_{\min}(\mathbf{M})}}{10}).$$

We prove the theorem by induction. The base case $t = 0$ holds trivially. Assuming the theorem is true for t, we will show it for $t+1$. Denoting $\widetilde{\mathbf{U}}_{\mathbf{t+1}} \triangleq \mathbf{U_t} - \eta\left(\mathbf{U_t}^2 - \mathbf{M}\right)\mathbf{U_t} - \eta\mathbf{U_t}\left(\mathbf{U_t}^2 - \mathbf{M}\right)$, we have

$$\begin{aligned}
\left\|\mathbf{M} - \mathbf{U_{t+1}}^2\right\|_F &= \left\|\mathbf{M} - \widetilde{\mathbf{U}}_{\mathbf{t+1}}^2 - \widetilde{\mathbf{U}}_{\mathbf{t+1}}\triangle_\mathbf{t} - \triangle_\mathbf{t}\widetilde{\mathbf{U}}_{\mathbf{t+1}} - \triangle_\mathbf{t}^2\right\|_F \\
&\leq \left\|\mathbf{M} - \widetilde{\mathbf{U}}_{\mathbf{t+1}}^2\right\|_F + 2\left\|\widetilde{\mathbf{U}}_{\mathbf{t+1}}\right\|_2\|\triangle_\mathbf{t}\|_F + \left\|\triangle_\mathbf{t}^2\right\|_F.
\end{aligned} \tag{14}$$

Using Claims B.2 and B.3, Lemma 4.3 tells us that

$$\left\|\widetilde{\mathbf{U}}_{\mathbf{t+1}}\right\|_2 \leq \max(\|\mathbf{U_0}\|_2, \sqrt{3\|\mathbf{M}\|_2}),$$

and Theorem 3.2 tells us that

$$\left\|\mathbf{M} - \widetilde{\mathbf{U}}_{\mathbf{t+1}}^2\right\|_F \leq \exp\left(-\widehat{c}\eta \min(\sigma_{\min}(\mathbf{U_0})^2, \sigma_{\min}(\mathbf{M}))\right) \left\|\mathbf{M} - \mathbf{U_t}^2\right\|_F.$$

Plugging the above two conclusions into (14), tells us that

$$\begin{aligned}
&\left\|\mathbf{M} - \mathbf{U_{t+1}}^2\right\|_F \\
&\leq \exp\left(-\widehat{c}\eta \min(\sigma_{\min}(\mathbf{U_0})^2, \sigma_{\min}(\mathbf{M}))\right) \left\|\mathbf{M} - \mathbf{U_t}^2\right\|_F + 2\max(\|\mathbf{U_0}\|_2, \sqrt{3\|\mathbf{M}\|_2}) \|\triangle_{\mathbf{t}}\|_F \\
&\quad + \frac{1}{30}\eta\sigma_{\min}(\mathbf{M}) \min(\sigma_{\min}(\mathbf{U_0}), \sqrt{\sigma_{\min}(\mathbf{M})}) \|\triangle_{\mathbf{t}}\|_F \\
&\leq \exp\left(-\widehat{c}\eta \min(\sigma_{\min}(\mathbf{U_0})^2, \sigma_{\min}(\mathbf{M}))\right) \left\|\mathbf{M} - \mathbf{U_0}^2\right\|_F \\
&\quad + 4\max(\|\mathbf{U_0}\|_2, \sqrt{3\|\mathbf{M}\|_2}) \sum_{s=0}^{t} \exp\left(-\widehat{c}\eta \min(\sigma_{\min}(\mathbf{U_0})^2, \sigma_{\min}(\mathbf{M}))(t-s)\right) \|\triangle_{\mathbf{s}}\|_F,
\end{aligned}$$

where we used the induction hypothesis in the last step. □

We now prove Claim B.2.

Proof of Claim B.2. Just as in the proof of Lemma 4.3, we will use induction. Assuming the claim is true for $\mathbf{U_t}$, by update equation

$$\mathbf{U_{t+1}} = \mathbf{U_t} - \eta\left(\mathbf{U_t}^2 - \mathbf{M}\right)\mathbf{U_t} - \eta\mathbf{U_t}\left(\mathbf{U_t}^2 - \mathbf{M}\right) + \triangle_{\mathbf{t}}$$

We can write out:

$$\|\mathbf{U_{t+1}}\|_2 \leq \left\|\left(\mathbf{I} - 2\eta\mathbf{U_t}^2\right)\mathbf{U_t}\right\|_2 + \eta\|\mathbf{M}\|_2\|\mathbf{U_t}\|_2 + \eta\|\mathbf{U_t}\|_2\|\mathbf{M}\|_2 + \|\triangle_{\mathbf{t}}\|_2. \tag{15}$$

Since $\eta < \frac{1}{10\max(\|\mathbf{U_0}\|_2^2, \|\mathbf{M}\|_2)}$ and $\|\mathbf{U_t}\|_2 \leq \max(\|\mathbf{U_0}\|_2, \sqrt{3\|\mathbf{M}\|_2})$, note that the singular values of the matrix $\left(\mathbf{I} - 2\eta\mathbf{U_t}^2\right)\mathbf{U_t}$ are exactly $(1-2\eta\sigma^2)\cdot\sigma$ where σ is a singular value of $\mathbf{U_t}$. For $\sigma \leq \sqrt{2\|\mathbf{M}\|_2}$, we clearly have $(1-2\eta\sigma^2)\sigma \leq \sqrt{2\|\mathbf{M}\|_2}$. On the other hand, for $\sigma > \sqrt{2\|\mathbf{M}\|_2}$, we have $(1-2\eta\sigma^2)\sigma < (1-4\eta\|\mathbf{M}\|_2)\sigma$. Plugging this observation into (15), we obtain:

$$\begin{aligned}
\|\mathbf{U_{t+1}}\|_2 &\leq \max\left(\sqrt{2\|\mathbf{M}\|_2}, (1-4\eta\|\mathbf{M}\|_2)\|\mathbf{U_t}\|_2\right) + 2\eta\|\mathbf{M}\|_2\|\mathbf{U_t}\|_2 + \|\triangle_{\mathbf{t}}\|_2 \\
&\leq \max(\|\mathbf{U_0}\|_2, \sqrt{3\|\mathbf{M}\|_2}),
\end{aligned}$$

proving the claim. □

We now prove Claim B.3.

Proof of Claim B.3. We will use induction, with the proof following fairly easily using Lemma B.1. Suppose $\sigma_{\min}(\mathbf{U_t}) \geq \min(\sigma_{\min}(\mathbf{U_0}), \frac{1}{10}\sqrt{\sigma_{\min}(\mathbf{M})})$. Denoting

$$\widetilde{\mathbf{U}}_{\mathbf{t+1}} \stackrel{\triangle}{=} \mathbf{U_t} - \eta\left(\mathbf{U_t}^2 - \mathbf{M}\right)\mathbf{U_t} - \eta\mathbf{U_t}\left(\mathbf{U_t}^2 - \mathbf{M}\right),$$

Lemma B.1 tells us that

$$\sigma_{\min}\left(\widetilde{\mathbf{U}}_{\mathbf{t+1}}\right) \geq \left(1 + \frac{\eta\sigma_{\min}(\mathbf{M})}{15}\right)\min(\sigma_{\min}(\mathbf{U_0}), \frac{1}{10}\sqrt{\sigma_{\min}(\mathbf{M})}),$$

which then implies the claim, since

$$\sigma_{\min}(\mathbf{U_{t+1}}) \geq \sigma_{\min}(\mathbf{U_{t+1}}) - \|\triangle_{\mathbf{t}}\|_2 \geq \min(\sigma_{\min}(\mathbf{U_0}), \frac{1}{10}\sqrt{\sigma_{\min}(\mathbf{M})}).$$

□

C Proof of Theorem 3.5

In this section, we will prove Theorem 3.5.

Proof. Consider two-dimensional case, where

$$\mathbf{M} = \begin{pmatrix} \|\mathbf{M}\|_2 & 0 \\ 0 & \sigma_{\min}(\mathbf{M}) \end{pmatrix}$$

We will prove Theorem 3.5 by considering two cases of step size (where $\eta \geq \frac{1}{4\|\mathbf{M}\|_2}$ or $\eta < \frac{1}{4\|\mathbf{M}\|_2}$) separately.

Case 1 : For step size $\eta \geq \frac{1}{4\|\mathbf{M}\|_2}$. Let $\beta = \frac{1}{2\eta\|\mathbf{M}\|_2} + 1$, and consider following initialization $\mathbf{U_0}$:

$$\mathbf{U_0} = \begin{pmatrix} \sqrt{\beta\|\mathbf{M}\|_2} & 0 \\ 0 & \sqrt{\sigma_{\min}(\mathbf{M})} \end{pmatrix}$$

Since $\eta \geq \frac{1}{4\|\mathbf{M}\|_2}$, we know $\beta \leq 3$, which satisfies our assumption about $\mathbf{U_0}$ and $\mathbf{M}$. By calculation, we have:

$$\mathbf{U_0}(\mathbf{U_0}^2 - \mathbf{M}) + (\mathbf{U_0}^2 - \mathbf{M})\mathbf{U_0} = \begin{pmatrix} 2(\beta-1)\sqrt{\beta\|\mathbf{M}\|_2^3} & 0 \\ 0 & 0 \end{pmatrix}$$

and since $2\eta(\beta - 1)\|\mathbf{M}\|_2 = 1$, we have:

$$\mathbf{U_1} = \mathbf{U_0} - \eta[\mathbf{U_0}(\mathbf{U_0}^2 - \mathbf{M}) + (\mathbf{U_0}^2 - \mathbf{M})\mathbf{U_0}] = \begin{pmatrix} \sqrt{\beta\|\mathbf{M}\|_2} - 2\eta(\beta-1)\sqrt{\beta\|\mathbf{M}\|_2^3} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Then, by induction we can easily show for all $t \geq 1$, $\mathbf{U_t} = \mathbf{U_1}$, thus $\left\|\mathbf{U_t}^2 - \mathbf{M}\right\|_F \geq \|\mathbf{M}\|_2 \geq \frac{1}{4}\sigma_{\min}(\mathbf{M})$.

Case 2 : For step size $\eta \leq \frac{1}{4\|\mathbf{M}\|_2}$, consider following initialization $\mathbf{U_0}$:

$$\mathbf{U_0} = \begin{pmatrix} \sqrt{\|\mathbf{M}\|_2} & 0 \\ 0 & \frac{1}{2}\sqrt{\sigma_{\min}(\mathbf{M})} \end{pmatrix}$$

According to the update rule in Algorithm 1, we can easily show by induction that: for any $t \geq 0$, $\mathbf{U_t}$ is of form:

$$\mathbf{U_t} = \begin{pmatrix} \sqrt{\|\mathbf{M}\|_2} & 0 \\ 0 & \alpha_t\sqrt{\sigma_{\min}(\mathbf{M})} \end{pmatrix}$$

where α_t is a factor that depends on t, satisfying $0 \leq \alpha_t \leq 1$ and:

$$\alpha_{t+1} = \alpha_t[1 + \eta\sigma_{\min}(\mathbf{M})(1 - \alpha_t^2)], \quad \alpha_0 = \frac{1}{2}$$

Since $\eta \leq \frac{1}{4\|\mathbf{M}\|_2}$, we know:

$$\alpha_{t+1} \leq \alpha_t[1 + \frac{1}{4\kappa}(1 - \alpha_t^2)] \leq \alpha_t + \frac{1}{4\kappa}$$

Therefore, for all $t \leq \kappa$, we have $\alpha_t \leq \frac{3}{4}$, and thus $\left\|\mathbf{U_t}^2 - \mathbf{M}\right\|_F \geq \frac{1}{4}\sigma_{\min}(\mathbf{M})$. □