Appendix of
Provable Inductive Robust PCA via Iterative Hard Thresholding

A PROOFS: NOISY CASE

A.1 Proof of Theorem 3.2

Proof. We prove this by induction over t. Note that Step 3 of Algorithm 1 initializes $\zeta_0 = 5\mu^2\sigma_{\max}^2(F)\frac{d}{n}c_W + \nu$ and sets $\zeta_t = \mu^2\sigma_{\max}^2(F)\frac{d}{n}\frac{c_W}{5^{t-1}} + \nu$ for all $t \geq 1$. Let $\nu = (3\mu^2 d\kappa^2 + 1)\|N^*\|_\infty$. For $t = 1$, since $L_0 = 0$ by our initialization, it is clear that $\|L^* - L_0\|_\infty \leq \mu^2\sigma_{\max}^2(F)\frac{d}{n}c_W$ and hence the base case holds. Next, for $t \geq 1$, by using Lemma 3.3, we have $\|S^* - S_t\|_\infty \leq 2\mu^2\sigma_{\max}^2(F)\frac{d}{n}\frac{c_W}{5^{t-1}} + 2(3\mu^2 d\kappa^2 + 1)\|N^*\|_\infty$ and further, by Lemma 3.4, we have $\|L^* - L_t\|_\infty \leq \mu^2\sigma_{\max}^2(F)\frac{d}{n}\frac{c_W}{5^t} + 3\mu^2 d\kappa^2\|N^*\|_\infty$. Moreover, setting $T > \lceil \log_5(2\mu^2\sigma_{\max}^2(F)\frac{d}{n}\frac{c_W}{\epsilon})\rceil + 1$, we obtain the result. □

A.2 Proof of Lemma 3.3

Proof. Recall that $S_t = \mathcal{P}_{\zeta_t}(M - L_{t-1}) = \mathcal{P}_{\zeta_t}(L^* - L_{t-1} + S^* + N^*)$. By the definition of our entry-wise hard thresholding operation, we have the following:

1. Term $e_i^\top S_t e_j = e_i^\top (M - L_{t-1})e_j = e_i^\top (L^* + S^* + N^* - L_{t-1})e_j$ when $\left|e_i^\top (M - L_{t-1})e_j\right| > \zeta_t$. Thus, $\left|e_i^\top (S^* - S_t)e_j\right| = \left|e_i^\top (L^* - L_{t-1})e_j\right| + \left|e_i^\top N^* e_j\right| \leq \mu^2\sigma_{\max}^2(F)\frac{d}{n}\frac{c_W}{5^{t-1}} + 3\mu^2 d\kappa^2\|N^*\|_\infty + \|N^*\|_\infty$.
2. Term $e_i^\top S_t e_j = 0$ when $\left|e_i^\top (M - L_{t-1})e_j\right| = \left|e_i^\top (L^* + S^* + N^* - L_{t-1})e_j\right| \leq \zeta_t$. Now, using the triangle inequality, we have $\left|e_i^\top (S^* - S_t)e^j\right| = \left|e_i^\top S^* e^j\right| \leq \zeta_t + \left|e_i^\top (L^* - L_{t-1})e_j\right| + \left|e_i^\top N^* e_j\right| \leq 2\left(\mu^2\sigma_{\max}^2(F)\frac{d}{n}\frac{c_W}{5^{t-1}} + 3\mu^2 d\kappa^2\|N^*\|_\infty + \|N^*\|_\infty\right)$.

Thus, the above two cases show the validity of the entry-wise hard thresholding operation. Next, we show that for any given (i,j), if $e_i^\top S^* e_j = 0$ then $e_i^\top S_t e_j$ is also zero for all t. Noting that $M = L^* + S^* + N^*$ and $e_i^\top S^* e_j = 0$, $e_i^\top S_t e_j = e_i^\top (M - L_{t-1})e_j = e_i^\top (L^* + N^* - L_{t-1})e_j \neq 0$ iff $\left|e_i^\top (L^* + N^* - L_{t-1})e_j\right| > \zeta_t$. But this is a contradiction since $\left|e_i^\top (L^* + N^* - L_{t-1})e_j\right| \leq \left|e_i^\top (L^* - L_{t-1})e_j\right| + \left|e_i^\top N^* e_j\right| \leq \mu^2\sigma_{\max}^2(F)\frac{d}{n}\frac{c_W}{5^{t-1}} + 3\mu^2 d\kappa^2\|N^*\|_\infty + \|N^*\|_\infty = \zeta_t$. □

A.3 Proof of Lemma 3.4

Proof. Using the fact that $F_1 = F_2$, $L^* = F^\top W^* F$ and $L_t = F^\top W_t F$, we have

$$\begin{aligned}
\|L^* - L_t\|_\infty &= \left\|F^\top (W^* - W_t)F\right\|_\infty \\
&= \max_{i,j}\left|e_i^\top F^\top (W^* - W_t)Fe_j\right| \\
&\overset{\xi_{11}}{=} \max_{i,j}\left|e_i^\top V_F\Sigma_F^\top U_F^\top (W^* - W_t)U_F\Sigma_F V_F^\top e_j\right| \\
&\overset{\xi_{12}}{\leq} \left(\max_i \left\|e_i^\top V_F\Sigma_F^\top\right\|_2\right)^2 \left\|U_F^\top (W^* - W_t)U_F\right\|_2
\end{aligned} \tag{10}$$

where ξ_{11} follows by substituting the SVD of $F = U_F\Sigma_F V_F^\top$ and ξ_{12} follows from the sub-multiplicative property of the spectral norm. Similar to the proof of Lemma 3.2, using Assumption 2 we have:

$$\max_i \left\|e_i^\top V_F\Sigma_F^\top\right\|_2 \leq \mu\sqrt{\frac{d}{n}}\sigma_{\max}(F). \tag{11}$$

Let the residual sparse perturbation be defined as $E_t := S - S_t$. Let $Q\Lambda Q^\top + Q_\perp\Lambda_\perp Q_\perp^\top$ be the full SVD of $W^* + G^\top (E_t + N^*)G$ where Q and $Q_\perp$ span orthogonal sub-spaces of dimensions r and $d - r$ respectively, and $G = F^\dagger$ is the pseudoinverse. Also, recall that from Step 7 of Algorithm 1 that W_t is computed as $\mathcal{P}_r\left((F_1^\top)^\dagger(M - S_t)(F_2)^\dagger\right)$

where $M = F_1^\top W^* F_2 + S^* + N^*$. Using these and the unitary invariance property of the spectral norm, we have

$$\begin{aligned}
&\left\|U_F^\top (W^* - W_t) U_F\right\|_2 \leq \left\|W^* - W_t\right\|_2 \\
&\leq \left\|W^* - \mathcal{P}_r(G^\top (F^\top W^* F + E_t + N^*)G)\right\|_2 \\
&\overset{\xi_{13}}{\leq} \left\|Q\Lambda Q^\top + Q_\perp \Lambda_\perp Q_\perp^\top - G^\top (E_t + N^*)G - Q\Lambda Q^\top\right\|_2 \\
&\overset{\xi_{14}}{\leq} \left\|G^\top (E_t + N^*)G\right\|_2 + \left\|Q_\perp \Lambda_\perp Q_\perp^\top\right\|_2 \\
&\overset{\xi_{15}}{\leq} 2\left\|G^\top (E_t + N^*)G\right\|_2 \leq 2\left\|G\right\|_2^2 \left\|E_t + N^*\right\|_2 \\
&\leq \frac{2\left\|E_t + N^*\right\|_2}{[\sigma_{\min}(F)]^2} \overset{\xi_{16}}{\leq} \frac{2z\left\|E_t\right\|_\infty}{[\sigma_{\min}(F)]^2} + \frac{2\left\|N^*\right\|_2}{[\sigma_{\min}(F)]^2}
\end{aligned} \tag{12}$$

where ξ_{13} is obtained by substituting $W^* = Q\Lambda Q^\top + Q_\perp \Lambda_\perp Q_\perp^\top - G^\top (E_t + N^*)G$, ξ_{14} by triangle inequality, ξ_{15} by using Weyl's eigenvalue perturbation lemma, ie,

$$\left\|Q_\perp \Lambda_\perp Q_\perp^\top\right\|_2 = \left\|\Lambda_\perp\right\|_\infty \leq \left\|G^\top (E_t + N^*)G\right\|_2$$

and ξ_{16} by using Lemma 4 of (Netrapalli *et al.* , 2014) along with triangle inequality. Now, combining Equations (10), (11) and (12), we have

$$\begin{aligned}
\left\|L^* - L_t\right\|_\infty &\leq 2\mu^2 \frac{d}{n} \kappa^2 \left(z\left\|E_t\right\|_\infty + \left\|N^*\right\|_2\right) \\
&\overset{\xi_{17}}{\leq} \frac{\left\|E_t\right\|_\infty}{10} + 2\mu^2 d\kappa^2 \left\|N^*\right\|_\infty
\end{aligned} \tag{13}$$

where ξ_{17} follows by using Assumption 3 and the inequality that $\left\|N^*\right\|_2 \leq n\left\|N^*\right\|_\infty$. Using the inequality $\left\|S^* - S_t\right\|_\infty \leq 2\left(\mu^2 \sigma_{\max}^2(F) \frac{d}{n} \frac{c_W}{5^{t-1}} + (3\mu^2 d\kappa^2 + 1)\left\|N^*\right\|_\infty\right)$ from Lemma 3.3 in Equation (13) completes the proof. □

B PROOFS: ASYMMETRIC CASE

B.1 Proof of Claim 3.1

Proof. Applying the symmetric embedding transformation to our data matrix, we get $\mathrm{Sym}(M) = \mathrm{Sym}(L^*) + \mathrm{Sym}(S^*)$. Now we characterize the properties of this symmetric embedding and show that it satisfies Assumptions 1, 2 and 3. First, we have

$$\begin{aligned}
\mathrm{Sym}(L^*) &= \begin{pmatrix} 0 & L^* \\ L^{*\top} & 0 \end{pmatrix} = \begin{pmatrix} 0 & F_1^\top W^* F_2 \\ F_2^\top W^{*\top} F_1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} F_1^\top & 0 \\ 0 & F_2^\top \end{pmatrix} \begin{pmatrix} 0 & W^* \\ W^{*\top} & 0 \end{pmatrix} \begin{pmatrix} F_1 & 0 \\ 0 & F_2 \end{pmatrix}.
\end{aligned}$$

Thus, $\mathrm{Sym}(L^*)$ is of the form $\widetilde{F}^\top \widetilde{W}^* \widetilde{F}$. If the SVD of W^* is $U_{W^*}\Sigma_{W^*}V_{W^*}^\top$, then the eigenvalue decomposition of $\widetilde{W}^*$ is given by

$$\begin{aligned}
\widetilde{W}^* &= \begin{pmatrix} 0 & W^* \\ W^{*\top} & 0 \end{pmatrix} = \begin{pmatrix} 0 & U_{W^*}\Sigma_{W^*}V_{W^*}^\top \\ V_{W^*}\Sigma_{W^*}^\top U_{W^*}^\top & 0 \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} U_{W^*} & U_{W^*} \\ V_{W^*} & -V_{W^*} \end{pmatrix} \begin{pmatrix} \Sigma_{W^*} & 0 \\ 0 & -\Sigma_{W^*} \end{pmatrix} \begin{pmatrix} U_{W^*} & U_{W^*} \\ V_{W^*} & -V_{W^*} \end{pmatrix}^\top,
\end{aligned}$$

implying that $\mathrm{rank}(\widetilde{W}^*) = 2 \cdot \mathrm{rank}(W^*)$. Next, let the SVDs of F_1 and F_2 be $U_{F_1}\Sigma_{F_1}V_{F_1}^\top$ and $U_{F_2}\Sigma_{F_2}V_{F_2}^\top$ respectively; also, without loss of generality, let $\sigma_{\min}(F_1) > \sigma_{\min}(F_2)$. Then, the SVD of $\widetilde{F} = U_{\widetilde{F}}\Sigma_{\widetilde{F}}V_{\widetilde{F}}^\top$ is given by

$$\widetilde{F} = \begin{pmatrix} F_1 & 0 \\ 0 & F_2 \end{pmatrix} = \begin{pmatrix} U_{F_1}\Sigma_{F_1}V_{F_1}^\top & 0 \\ 0 & U_{F_2}\Sigma_{F_2}V_{F_2}^\top \end{pmatrix}$$
$$= \begin{pmatrix} U_{F_1} & 0 \\ 0 & U_{F_2} \end{pmatrix} \begin{pmatrix} \Sigma_{F_1} & 0 \\ 0 & \Sigma_{F_2} \end{pmatrix} \begin{pmatrix} V_{F_1}^\top & 0 \\ 0 & V_{F_2}^\top \end{pmatrix}$$

Now, we verify that the right singular vectors of this new feature matrix $\widetilde{F}$ satisfies weak incoherence property. Specifically, we expect that the following holds:

$$\max_j \left\|V_{\widetilde{F}} e_j\right\|_2 \leq \mu_{\widetilde{F}} \sqrt{\frac{d_1 + d_2}{n_1 + n_2}} \tag{14}$$

On the other hand, we actually have

$$\max_j \|V e_j\|_2 \leq \max\left(\mu_{F_1}\sqrt{\frac{d_1}{n_1}}, \mu_{F_2}\sqrt{\frac{d_2}{n_2}}\right). \tag{15}$$

Wlog, let $\mu_{F_1}\sqrt{d_1/n_1} > \mu_{F_2}\sqrt{d_2/n_2}$. Then, combining Equations (14) and (15), we want $\frac{\mu_{\tilde{F}}}{\mu_{F_1}} \leq \sqrt{\frac{1+n_2/n_1}{1+d_2/d_1}}$. In particular, when $n_2/n_1 = d_2/d_1$, the incoherence constant for $\widetilde{F}$ satisfies $\mu_{\widetilde{F}} = \mu_{F_1}$.

Next, note that $\mathrm{Sym}(S^*)$ is also sparse; specifically, $\|S^*\|_{0,\infty} \leq z$ and $\|S^*\|_{\infty,0} \leq z$ where $z = \max(z_1, z_2)$.

Finally, our algorithm and guarantees hold for general matrices with noise, similar to noiseless case, due to the following observation: $\|\mathrm{Sym}(N^*)\|_\infty = \|N^*\|_\infty$. □