Supplementary Material of
Stability of Matrix Factorization for Collaborative Filtering

Appendices

A. Proof of Theorem 2: Partial Observation Theorem

In this appendix we prove Theorem 2. The proof involves a covering number argument and a concentration inequality for sampling without replacement. The two lemmas are stated below.

Lemma A.1 (Hoeffding Inequality for Sampling without Replacement (Serfling, 1974)). *Let $X = [X_1, ..., X_n]$ be a set of samples taken without replacement from a distribution $\{x_1, ...x_N\}$ of mean u and variance σ^2. Denote $a \triangleq \max_i x_i$ and $b \triangleq \min_i x_i$. Then we have:*

$$Pr(|\frac{1}{n}\sum_{i=1}^{n} X_i - u| \geq t) \leq 2\exp(-\frac{2nt^2}{(1-\frac{n-1}{N})(b-a)^2}). \tag{A.1}$$

Lemma A.2 (Covering number for low-rank matrices of bounded size). *Let $S_r = \{X \in \mathbb{R}^{n_1 \times n_2} : rank(X) \leq r, \|X\|_F \leq K\}$. Then there exists an ϵ-net $\bar{S}_r$ for the Frobenius norm obeying*

$$|\bar{S}_r(\epsilon)| \leq (9K/\epsilon)^{(n_1+n_2+1)r}.$$

This Lemma is essentially the same as Lemma 2.3 of (Candès & Plan, 2011), with the only difference being the range of $\|X\|_F$: instead of having $\|X\|_F = 1$, we have $\|X\|_F \leq K$. The proof is given in the next section of Appendix.

Proof of Theorem 2. Fix $X \in S_r$. Define the following to lighten notations

$$\hat{u}(X) = \frac{1}{|\Omega|}\|P_\Omega(X - \widehat{Y})\|_F^2 = (\hat{\mathcal{L}}(X))^2,$$
$$u(X) = \frac{1}{mn}\|X - \widehat{Y}\|_F^2 = (\mathcal{L}(X))^2.$$

Notice that $\left\{(X_{ij} - \widehat{Y}_{ij})^2\right\}_{ij}$ form a distribution of nm elements, u is its mean, and $\hat{u}$ is the mean of $|\Omega|$ random samples drawn without replacement. Hence, by Lemma A.1:

$$Pr(|\hat{u}(X) - u(X)| > t) \leq 2\exp\left(-\frac{2|\Omega|mnt^2}{(mn - |\Omega| + 1)M^2}\right), \tag{A.2}$$

where $M \triangleq \max_{ij}(X_{ij} - \widehat{Y}_{ij})^2 \leq 4k^2$. Apply union bound over all $X \in \bar{S}_r(\epsilon)$, we have

$$Pr(\sup_{\bar{X} \in \bar{S}_r(\epsilon)} |\hat{u}(\bar{X}) - u(\bar{X})| > t)$$
$$\leq 2|\bar{S}_r(\epsilon)|\exp\left(-\frac{2|\Omega|mnt^2}{(mn - |\Omega| + 1)M^2}\right)$$

Equivalently, with probability at least $1 - 2\exp(-n)$.

$$\sup_{\bar{X} \in \bar{S}_r(\epsilon)} |\hat{u}(\bar{X}) - u(\bar{X})|$$
$$\leq \sqrt{\frac{M^2}{2}(n + \log|\bar{S}_r(\epsilon)|)\left(\frac{1}{|\Omega|} - \frac{1}{mn} + \frac{1}{mn|\Omega|}\right)}.$$

Notice that $\|X\|_F \leq \sqrt{mn}k$. Hence substituting Lemma A.2 into the equation, we get:

$$\sup_{\bar{X} \in \bar{S}_r(\epsilon)} |\hat{u}(\bar{X}) - u(\bar{X})|$$
$$\leq \left[\frac{M^2}{2}\left(n + (m+n+1)r\log(9k\sqrt{mn}/\epsilon)\right)\right.$$
$$\left.\times\left(\frac{1}{|\Omega|} - \frac{1}{mn} + \frac{1}{mn|\Omega|}\right)\right]^{\frac{1}{2}}$$
$$:= \xi(\Omega),$$

where we define $\xi(\Omega)$ for convenience. Recall that $\hat{u}(\bar{X}) = (\hat{\mathcal{L}}(X))^2$ and $u(\bar{X}) = (\mathcal{L}(X))^2$. Notice that for any non-negative a and b, $a^2 + b^2 \leq (a+b)^2$. Hence the following inequalities hold for all $\bar{X} \in \bar{S}_r(\epsilon)$:

$$(\hat{\mathcal{L}}(\bar{X}))^2 \leq (\mathcal{L}(\bar{X}))^2 + \xi(\Omega) \leq (\mathcal{L}(\bar{X}) + \sqrt{\xi(\Omega)})^2,$$
$$(\mathcal{L}(\bar{X}))^2 \leq (\hat{\mathcal{L}}(\bar{X}))^2 + \xi(\Omega) \leq (\hat{\mathcal{L}}(\bar{X}) + \sqrt{\xi(\Omega)})^2,$$

which implies

$$\sup_{\bar{X} \in \bar{S}_r} |\hat{\mathcal{L}}(\bar{X}) - \mathcal{L}(\bar{X})| \leq \sqrt{\xi(\Omega)}.$$

To establish the theorem, we need to relate S_r and $\bar{S}_r(\epsilon)$. For any $X \in S_r$, there exists $c(X) \in \bar{S}_r(\epsilon)$ such

that:

$$\|X - c(X)\|_F \le \epsilon; \qquad \|P_\Omega(X - c(X))\|_F \le \epsilon;$$

which implies,

$$\begin{aligned}
&|\mathcal{L}(X) - \mathcal{L}(c(X))| \\
&= \frac{1}{\sqrt{mn}} \left| \|X - \widehat{Y}\|_F - \|c(X) - \widehat{Y}\|_F \right| \le \frac{\epsilon}{\sqrt{mn}}; \\
&|\hat{\mathcal{L}}(X) - \hat{\mathcal{L}}(c(X))| \\
&= \frac{1}{\sqrt{|\Omega|}} \left| \|P_\Omega(X - \widehat{Y})\|_F - \|P_\Omega(c(X) - \widehat{Y})\|_F \right| \le \frac{\epsilon}{\sqrt{|\Omega|}}.
\end{aligned}$$

Thus we have,

$$\begin{aligned}
&\sup_{X \in S_r} |\hat{\mathcal{L}}(X) - \mathcal{L}(X)| \\
\le &\sup_{X \in S_r} \Big\{ |\hat{\mathcal{L}}(X) - \hat{\mathcal{L}}(c(X))| + |\mathcal{L}(c(X)) - \mathcal{L}(X)| \\
&+ |\hat{\mathcal{L}}(c(X)) - \mathcal{L}(c(X))| \Big\} \\
\le &\frac{\epsilon}{\sqrt{|\Omega|}} + \frac{\epsilon}{\sqrt{mn}} + \sup_{X \in S_r} |\hat{\mathcal{L}}(c(X)) - \mathcal{L}(c(X))| \\
\le &\frac{\epsilon}{\sqrt{|\Omega|}} + \frac{\epsilon}{\sqrt{mn}} + \sup_{\bar{X} \in \bar{S}_r(\epsilon)} |\hat{\mathcal{L}}(\bar{X}) - \mathcal{L}(\bar{X})| \\
\le &\frac{\epsilon}{\sqrt{|\Omega|}} + \frac{\epsilon}{\sqrt{mn}} + \sqrt{\xi(\Omega)}.
\end{aligned}$$

Substitute in the expression of $\xi(\Omega)$ and take $\epsilon = 9k$, we have,

$$\begin{aligned}
&\sup_{X \in S_r} |\hat{\mathcal{L}}(X) - \mathcal{L}(X)| \\
\le &2\frac{\epsilon}{\sqrt{|\Omega|}} + \left(\frac{M^2}{2} \frac{2nr \log(9kn/\epsilon)}{|\Omega|} \right)^{\frac{1}{4}} \\
\le &\frac{18k}{\sqrt{|\Omega|}} + \sqrt{2}k \left(\frac{nr \log(n)}{|\Omega|} \right)^{\frac{1}{4}} \\
\le &Ck \left(\frac{nr \log(n)}{|\Omega|} \right)^{\frac{1}{4}},
\end{aligned}$$

for some universal constant C. This complete the proof. □

B. Proof of Lemma A.2: Covering number of low rank matrices

In this appendix, we prove the covering number lemma used in Appendix A. As explained in the main text of the paper, this is an extension of Lemma 2.1 in Candès & Plan (2011).

Proof of Lemma A.2. This is a two-step proof. First we prove for $\|X\|_F \le 1$, then we scale it to $\|X\|_F \le K$.

Step 1: The first part is almost identical to that in Page 14-15 of (Candès & Plan, 2011). We prove via SVD and bound the $\epsilon/3$-covering number of U, Σ and V individually. U and V are bounded the same way. So we only cover the part for of $r \times r$ diagonal singular value matrix Σ.

Now $\|\Sigma\| \le 1$ instead of $\|\Sigma\| = 1$. $diag(\Sigma)$ lying inside a unit r-sphere (denoted by A). We want to cover this r-sphere with smaller r-sphere of radius $\epsilon/3$ (denoted by B). Then there is a lower bound and an upper bound of the $(\epsilon/3)$-covering number $N(A, B)$.

$$\begin{aligned}
\frac{vol(A)}{vol(B)} \le N(A, B) \le \bar{N}(A, B) &= \bar{N}(A, \frac{B}{2} - \frac{B}{2}) \\
&\le M(A, \frac{B}{2}) \le \frac{vol(A + B/2)}{vol(B/2)}
\end{aligned}$$

where $\bar{N}(A, B)$ is the covering number from inside, and $M(A, B)$ is the number of separated points. Set $B = \frac{B}{2} - \frac{B}{2}$ because B is symmetrical (an n-sphere).

$$(\frac{1}{\epsilon/3})^r \le N(A, B) \le (\frac{1 + \epsilon/6}{\epsilon/6})^r$$

We are only interested in the upper bound of covering number:

$$N(A, B) \le (1 + 6/\epsilon)^r \le (6/\epsilon + \frac{1}{\epsilon/3}) = (9/\epsilon)^r$$

The inequality is due to the fact that $\epsilon/3 < 1$ (otherwise covering set $B > A$). In fact, we may further tighten the bound by using the fact that all singular values are positive, then A is further constrained in side the first orthant. This should reduce the covering number to its $\frac{1}{2^r}$.

Everything else follows exactly the same way as in Candès & Plan (2011) (Page 14-15).

Step 2: By definition, if $\|X\|_F = 1$, then a finite set of $(9/\epsilon)^{(n_1+n_2+1)r}$ elements are sufficient to ensure that, for every $X \in S_r$, it exists an $\bar{X} \in \bar{S}_r$, such that

$$\|\bar{X} - X\|_F \le \epsilon$$

Scale both side by K, we get:

$$\|K\bar{X} - KX\|_F \le K\epsilon$$

let $\beta = K\epsilon$, then the β-net covering number of the set of $\|X\|_F = K$ is:

$$|\bar{S}_r| \le (9/\epsilon)^{(n_1+n_2+1)r} = (9K/\beta)^{(n_1+n_2+1)r}$$

Revert the notation back to ϵ, the proof is complete. □

C. Proof of Proposition 1: σ_{min} bound

In this appendix, we develop proof for Proposition 1. As is explained in main text of the paper, σ_{min} can be arbitrarily small in general[1], unless we make assumptions about the structure of matrix. That is why we need strong incoherence property(Candes & Tao, 2010) for the proof of Proposition 1, which is stated below.

Strong incoherence property with parameter μ, implies that exist $\mu_1, \mu_2 \leq \mu$, such that:

A1 There exists $\mu_1 > 0$ such that for all pair of standard basis vector e_i and e_j (overloaded in both column space and row space of different dimension), there is:

$$\left|\langle e_i, P_U e_j\rangle - \frac{r}{m}1_{i=j}\right| \leq \mu_1\frac{\sqrt{r}}{m}$$
$$\left|\langle e_i, P_V e_j\rangle - \frac{r}{n}1_{i=j}\right| \leq \mu_1\frac{\sqrt{r}}{n}$$

A2 There exists $\mu_2 > 0$ such that for all i, j, the "sign matrix" E defined by $E = UV^T$ satisfies:

$$|E_{i,j}| = \mu_2\frac{\sqrt{r}}{\sqrt{mn}}$$

To interpret A1, again let singular subspace U be denoted by a orthonormal basis matrix N, $P_U = NN^T$. If $i = j$, we have

$$\frac{r-\mu\sqrt{r}}{m} \leq \|n_i\|^2 = \|n_j\|^2 \leq \frac{r+\mu\sqrt{r}}{m} \qquad \text{(C.1)}$$

When $i \neq j$, we have

$$-\frac{\mu\sqrt{r}}{m} \leq n_i^T n_j \leq \frac{\mu\sqrt{r}}{m}$$

Proof of Proposition 1. Instead of showing smallest singular value of N_1 directly, we find the $\sigma_{max}(N_2)$ or $\|N_2\|$, and then use the fact that all $\sigma_{min}(N) = 1$ to bound $\sigma_{min}(N_1)$ with their difference.

Let N_2 be of dimension $k \times r$. $\|N_2\| = \|N_2^T\|$, so the maximum singular value equals to $\max_u \|N_2^T u\|$ with u being a unit vector of dimension k. We may consider k a coefficient with $k = [c_1, c_2, ...c_k]^T$. It is easy to see that $c_1^2 + ... + c_k^2 = 1$.

$$\begin{aligned}
&\|N_2^T u\|^2 = u^T N_2 N_2^T u \\
=&(c_1 n_1^T + c_2 n_2^T + ... + c_k n_k^T)(c_1 n_1 + c_2 n_2 + ... + c_k n_k) \\
=&(c_1^2 n_1^T n_1 + ... + c_k^2 n_k^T n_k) + 2\sum_{i<j} c_i c_j n_i^T n_j^T \\
\leq&\sum_{i=1,...,k}(c_i^2)\max_i \|n_i\|^2 + \sum_{i<j} 2|c_i c_j|\max_{i,j} n_i^T n_j \\
\leq&\max_i \|n_i\|^2 + \sum_{i<j}(c_i^2 + c_j^2)\max_{i,j} n_i^T n_j \\
=&\max_i \|n_i\|^2 + (k-1)\sum_{i=1,...,k}(c_i^2)\max_{i,j} n_i^T n_j \\
=&\max_i \|n_i\|^2 + (k-1)\max_{i,j} n_i^T n_j \\
\leq&\frac{r+\mu_1\sqrt{r}}{m} + (k-1)\frac{\mu_1\sqrt{r}}{m} = \frac{r}{m} + k\frac{\mu_1\sqrt{r}}{m}
\end{aligned}$$

The second inequality is by $a^2 + b^2 \geq 2ab$ and last inequality is by the strong incoherence condition.

Similarly, using the $\min_i \|n_i\|^2$ and $\min_{i,j} \|n_i\|^2$ we have a lower bound of $\|N_2^T u\|^2 \geq \frac{r}{m} - k\frac{\mu_1\sqrt{r}}{m}$. But his bound is not useful/trivial because it decreases with the increase of k, which counters the intuition.

Now, we may express the bound of max singular value *sigma* in terms of sample rate p of N_1 (hence sample rate of N_2 is $(1-p)$)

$$\sigma_{max}(N_2) \leq \left(\frac{r}{m} + (1-p)\mu_1\sqrt{r}\right)^{\frac{1}{2}}$$

The desired bound on minimum singular value of N_1 is hence $1 - \sigma_{max}(N_2) = 1 - \left(\frac{r}{m} + (1-p)\mu_1\sqrt{r}\right)^{\frac{1}{2}}$.

□

D. Proof of Proposition 2:σ_{min} bound for random matrix

Proof. Without loss of generality we can assume $k > r$ (otherwise the theorem holds trivially), and normalize G such that $\mathbf{E}\|G\|_F^2 = r$. Indeed, random matrix theory (e.g., Rudelson & Vershynin, 2009; Silverstein, 1985; Davidson & Szarek, 2001) asserts that G is close to an orthonomal matrix, as the following lemma, adapted from Theorem II.13 of Davidson & Szarek (2001), shows:

Lemma D.1. *With probability of at least* $1 - 2\gamma$,

$$1 - \sqrt{\frac{r}{m}} - \sqrt{\frac{2\log(1/\gamma)}{m}} \leq \sigma_{min}(G)$$
$$\leq \sigma_{max}(G) \leq 1 + \sqrt{\frac{r}{m}} + \sqrt{\frac{2\log(1/\gamma)}{m}}.$$

[1] Consider a matrix N with first r rows identity matrix and the rest zero(verify that this is an orthonormal basis matrix). If no observations are taken from first r-rows of user y then all singular values of the N_1 will be zero and (4) is degenerate.

Now let $G = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}$ such that G_1 is of dimension $k \times r$. Notice that by Lemma D.1, we conclude that there exists an absolute constant such that with probability $1 - C'm^{-10}$,

$$\|G_1 - N_1\| \le \|G - N\| \le \sqrt{\frac{r}{m}} + C'\sqrt{\frac{\log m}{m}}.$$

To see this, take compact SVD of $G = USV^T$, U is $m \times r$, S and V are both $r \times r$. In particular, U is orthonormal and V is a rotation matrix. Let $N = UV^T$, then N is an orthonormal basis of G. Furthermore, $G - N = USV^T - UV^T = U(S - I_{r\times r})V^T$ implies $\|G - N\| = |\sigma_{max}(G) - 1|$.

Then using the fact that G_1 is again Gaussian random matrix, we apply Lemma D.1 on G_1 to obtain

$$Pr\left(\sigma_{min}(G_1) \le \sqrt{\frac{k}{m}} - \sqrt{\frac{r}{m}} - C'\sqrt{\frac{\log m}{m}}\right) \le C'm^{-10}.$$

This implies that with probability $1 - 2C'm^{-10}$

$$\begin{aligned}\sigma_{min}(N_1) &\ge \sigma_{min}(G_1) - \|G_1 - N_1\| \\ &\ge \sqrt{\frac{k}{m}} - 2\sqrt{\frac{r}{m}} - 2C'\sqrt{\frac{\log m}{m}}.\end{aligned}$$

□

E. Proof of Proposition 4: Weak Robustness for Mass Attack

Proof of Proposition 4. First observe $\|(E^{gnd^\perp})\|_F \le k\sqrt{mn_e}$. Note by assumption, sample rate in E block is capped at $3p/2$, thus $\|P_\Omega(E^{gnd^\perp})\|_F \le k\sqrt{\frac{3p}{2}mn_e}$. Apply Theorem 1, we obtained Frobenious norm error:

$$\begin{aligned}\|\Delta\|_F &\le \frac{1}{\sqrt{p}}\|P_\Omega(E^{gnd^\perp})\|_F + \|(E^{gnd^\perp})\|_F + |\tau(\Omega)| \\ &= (\sqrt{\frac{3}{2}} + 1)k\sqrt{mn_e} + Ck\sqrt{m(n+n_e)}\left(\frac{nr\log(n)}{|\Omega|}\right)^{\frac{1}{4}}.\end{aligned}$$

Simplify the equation by absorbing small terms into constant, we get:

$$\|\Delta\|_F \le Ck\left[\sqrt{mn_e} + \left(\frac{n^3 r\log(n)}{p}\right)^{\frac{1}{4}}\right]$$

By Theorem 3:

$$\|\mathbb{P}^{gnd} - \mathbb{P}^{N^*}\| \le \sqrt{2}\frac{\|\Delta\|}{\delta} = \rho.$$

and δ is greater than $\sigma_r - \sigma_1(E^{gnd^\perp})$.

With condition number κ:

$$\begin{aligned}\sigma_r &= \frac{\|Y\|_2}{\kappa} \ge \frac{\|Y\|_F}{\kappa\sqrt{r}} = \frac{\sqrt{mn\mathbf{E}|Y_{i,j}|^2}}{\kappa\sqrt{r}} \qquad \text{(E.1)} \\ &\ge \frac{\sqrt{mn\mathbf{E}|Y_{i,j}|^2}\sigma_1(E^{gnd^\perp})}{\kappa\sqrt{r}\|E^{gnd^\perp}\|_F} \ge \frac{\sqrt{n\mathbf{E}|Y_{i,j}|^2}\sigma_1(E^{gnd^\perp})}{k\kappa\sqrt{r}\sqrt{n_e}} \\ &\ge \sigma_1(E^{gnd^\perp})\sqrt{\frac{n\mathbf{E}|Y_{i,j}|^2}{k^2\kappa^2 r}/n_e}. \qquad \text{(E.2)}\end{aligned}$$

Substitute n_e into (E.2), we get $\sigma_r \ge n^{1/4}\sigma_1(E^{gnd^\perp})$, or rather $\sigma_1(E^{gnd^\perp}) \le \sigma_r/n^{1/4}$. Together with (E.1),

$$\delta \ge (1 - 1/n^{1/4})\sigma_r = \frac{(1 - 1/n^{1/4})\sqrt{mn\mathbf{E}|Y_{i,j}|^2}}{\kappa\sqrt{r}}$$

It follows that

$$\begin{aligned}\rho = \frac{\|\Delta\|_F}{\delta} &\le \frac{Ck\left[\sqrt{mn_e} + \left(\frac{n^3 r\log(n)}{p}\right)^{\frac{1}{4}}\right]\cdot\kappa\sqrt{r}}{(1 - 1/n^{1/4})\sqrt{mn\mathbf{E}|Y_{i,j}|^2}} \\ &\le C\left[\frac{1}{n^{1/4}} + \frac{k\kappa}{\sqrt{\mathbf{E}|Y_{i,j}|^2}}\left(\frac{r^3\log(n)}{pn}\right)^{\frac{1}{4}}\right] \qquad \text{(E.3)} \\ &\le \frac{Ck\kappa}{\sqrt{\mathbf{E}|Y_{i,j}|^2}}\left(\frac{r^3\log(n)}{pn}\right)^{1/4}. \qquad \text{(E.4)}\end{aligned}$$

To reach (E.3), we substitute n_e with its maximum value, which cancels out the $\mathbf{E}|Y_{i,j}|^2$, κ, $\sqrt{r}$ in δ and k as well. $(1 - 1/n^{1/4})$ is absorbed into the constant C. In the second term in the square brackets, $n^{3/4}$ is canceled out by $(mn)^{1/2}$ with the ratio $\sqrt{n/m}$ absorbed into constant term. Also note that $\frac{k}{\sqrt{\mathbf{E}|Y_{i,j}|^2}} > 1$, $\kappa > 1$, $\frac{r^3\log(n)}{p} > 1$, so the second term is larger than $\frac{1}{n^{1/4}}$ and we may reach (E.4).

Apply Theorem 4:

$$\begin{aligned}\|y^* - y^{gnd}\| &\le \frac{2Ck\kappa\|y\|}{\sigma_{min}\sqrt{\mathbf{E}|Y_{i,j}|^2}}\left(\frac{r^3\log(n)}{pn}\right)^{1/4} \qquad \text{(E.5)} \\ \|e^* - e^{gnd}\| &\le \frac{2Ck\kappa\|e^{gnd^\perp}\|}{\sigma_{min}\sqrt{\mathbf{E}|Y_{i,j}|^2}}\left(\frac{r^3\log(n)}{pn}\right)^{1/4} + \frac{\|e^{gnd^\perp}\|}{\sigma_{min}} \\ &= \frac{C\|e^{gnd^\perp}\|}{\sigma_{min}}. \qquad \text{(E.6)}\end{aligned}$$

Now let us deal with σ_{min}. By assumption, all user have sample rate of at least $\frac{p}{2}$. By Proposition 2 and union bound, we confirm that for some constant c, with probability greater than $1 - cn^{-10}$, $\sigma_{min} \ge \sqrt{\frac{p}{2}}$ (relaxed by another $\sqrt{2}$ to get rid of the small terms) for all users.

Summing (E.5) over all users, we get:

$$\begin{aligned}\|Y^* - Y\|_F &= \sqrt{\sum_{allusers} \|y* - y^{gnd}\|^2} \\ &= \frac{2Ck\kappa}{\sigma_{min}\sqrt{\mathbf{E}|Y_{i,j}|^2}} \left(\frac{r^3\log(n)}{pn}\right)^{1/4} \sqrt{\sum_{allusers} \|y\|^2} \\ &\leq \frac{2\sqrt{2}Ck\kappa}{\sqrt{p\mathbf{E}|Y_{i,j}|^2}} \left(\frac{r^3\log(n)}{pn}\right)^{1/4} \sqrt{mn\mathbf{E}|Y_{i,j}|^2} \\ &\leq C_1\kappa k\sqrt{mn}\left(\frac{r^3\log(n)}{p^3n}\right)^{1/4},\end{aligned}$$

so $RMSE_Y \leq C_1\kappa k\left(\frac{r^3\log(n)}{p^3n}\right)^{1/4}$ is proved.

Similarly from (E.6), $RMSE_E \leq C\frac{\|E^{gnd^\perp}\|_F}{\sqrt{mn_e}}\sqrt{\frac{2}{p}} \leq \frac{C_2k}{\sqrt{p}}$.

□

F. SVD Perturbation Theory

The following theorems in SVD Perturbation Theory (Stewart, 1998) are applied in our proof of the subspace stability bound (Theorem 3).

1. Weyl's Theorem gives a perturbation bound for singular values.

Lemma F.1 (Weyl)**.**

$$|\hat{\sigma}_i - \sigma_i| \leq \|E\|_2, i = 1, ..., n.$$

2. Wedin's Theorem provides a perturbation bound for singular subspace. To state the Lemma, we need to re-express the singular value decomposition of Y and $\widehat{Y}$ in block matrix form:

$$Y = \begin{pmatrix} L_1 & L_2 & L_3 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} \quad \text{(F.1)}$$

$$\widehat{Y} = \begin{pmatrix} \hat{L}_1 & \hat{L}_2 & \hat{L}_3 \end{pmatrix} \begin{pmatrix} \hat{\Sigma}_1 & 0 \\ 0 & \hat{\Sigma}_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{R}_1 \\ \hat{R}_2 \end{pmatrix} \quad \text{(F.2)}$$

Let Φ denotes the canonical angles between $span(L_1)$ and $span(\hat{L}_1)$; let Θ denotes the canonical angle matrix between $span(R_1)$ and $span(\hat{R}_1)$.

Also, define residuals:

$$\begin{aligned} Z &= Y\hat{R}_1^T - \hat{L}_1\hat{\Sigma}_1 \\ S &= Y^T\hat{L}_1 - \hat{R}_1^T\hat{\Sigma}_1 \end{aligned}$$

The Wedin's Theorem bounds Φ and Θ together using the Frobenious norm of Z and S.

Lemma F.2 (Wedin)**.** *If there is a $\delta > 0$ such that*

$$min|\sigma(\hat{\Sigma}_1) - \sigma(\Sigma_2)| \geq \delta \quad \text{(F.3)}$$

and

$$min\sigma(\hat{\Sigma}_1) \geq \delta \quad \text{(F.4)}$$

then

$$\sqrt{\|\sin\Phi\|_F^2 + \|\sin\Theta\|_F^2} \leq \frac{\sqrt{\|Z\|_F^2 + \|S\|_F^2}}{\delta} \quad \text{(F.5)}$$

Besides Frobenious norm, the same result goes for $\|\cdot\|_2$, the spectral norm of everything.

Lemma F.2(Wedin's Theorem) says that if the two separation conditions on singular value (F.3) and (F.4) are satisfied, we can bound the impact of perturbation on the left and right singular subspace simultaneously.

G. Discussion on Box Constraint in (1)

The box constraint is introduced due to the proof technique used in Section 3. We suspect that a more refined analysis may be possible to remove such a constraint. As for results of other sections, such constraint is not needed. Yet, it does not hurt to impose such constraint to (3), which will lead to similar results of subspace stability (though much more tedious in proof). Moreover, notice that for sufficiently large k, the solution will remain unchanged with or without the constraint.

On the other hand, we remark that such the box constraint is most natural for the application in collaborative filtering. Since user ratings are usually bounded in a pre-defined range. In real applications, either such box constraint or regularization will be needed to avoid over fitting to the noisy data. This is true regardless whether formulation (1) or (3) is used.

H. Additional Figures and Tables

References

Candès, E.J. and Plan, Y. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Info. Theory*, 57:2342–2359, 2011.

Candes, E.J. and Tao, T. The power of convex relaxation: Near-optimal matrix completion. *IEEE Info. Theory*, 56:2053–2080, 2010.

Davidson, K.R. and Szarek, S.J. Local operator theory, random matrices and banach spaces. *Handbook of the geometry of Banach spaces*, 1:317–366, 2001.

Rudelson, M. and Vershynin, R. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 62:1707–1739, 2009.

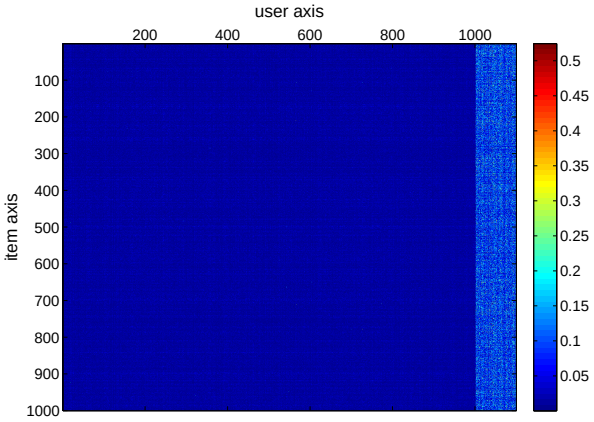

Figure 1. An illustration of error distribution for Random Attack, $n_e = 100$, $p = 0.3$. We can see a sharp transition in error level from honest user block on the left to the dummy user blocks on the right, which agrees with the prediction in Proposition 4 and the discussion in the beginning of Section 4.

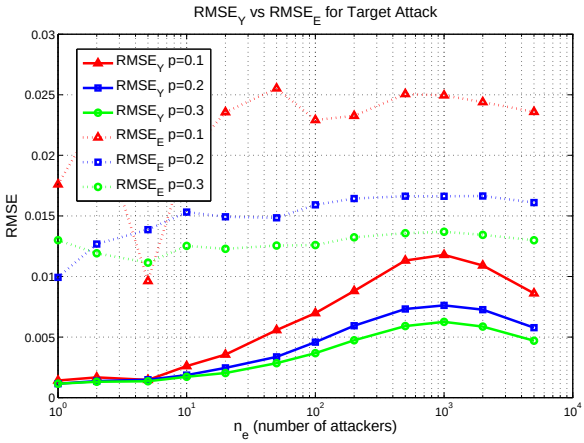

Figure 2. Comparison of $RMSE$ in Y-block and E-block for targeted attacks. This is the targeted attack version of Figure 2. From this figure, we can tell that while Proposition 3 bounds the total-$RMSE$, the gap between honest block and malicious block exists too. This leads to an even smaller manipulator impacts on honest users.

Serfling, R.J. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, 2:39–48, 1974.

Silverstein, J.W. The smallest eigenvalue of a large dimensional wishart matrix. *The Annals of Probability*, 13:1364–1368, 1985.

Stewart, G.W. Perturbation theory for the singular value decomposition. 1998.

Table 1. Table of Symbols and Notations

Y	$m \times n$ ground truth rating matrix.
E	$m \times n$ error matrix, in Section 6 dummy user matrix.
$\widehat{Y}$	Noisy observation matrix $\widehat{Y} = Y + E$.
Y^*, U^*, V^*	Optimal solution of (1) $Y^* = U^* V^{*T}$
$(\cdot)^*, (\hat{\cdot}), (\cdot)^{gnd}$	Refer to optimal solution, noisy observation, ground truth.
i, j	Item index and user index
r	Rank of ground truth matrix
Ω	The set of indices (i, j) of observed entries.
$\|\Omega\|$	Cardinality of set Ω.
P_Ω	The projection defined in (2).
k	$[-k, k]$ Valid range of user rating.
Δ	Frobenious norm error $\|Y^* - Y\|_F$
$\mathcal{N}, \mathcal{N}^\perp$	Denote subspace and complement subspace
$N, N^\perp$	Orthonormal basis matrix of $\mathcal{N}, \mathcal{N}^\perp$
N_i	Shortened N with only observed rows in column i
y_i	Observed subset of column i
$\mathbb{P}^{\mathcal{N}}$	Projection matrix to subspace $\mathcal{N}$
$\mathbb{P}_i$	Projection matrix to shortened subspace span(N_i)
τ	The gap of RMSE residual in the proof of Theorem 1.
$\mathcal{L}, \widehat{\mathcal{L}}$	Loss function in Theorem 2.
ρ	Bounded value of $\|\sin(\Theta)\|$ of Theorem 3.
δ	The r^{th} singular value of Y^* used in Theorem 3.
S_r	The collection of all rank-r $m \times n$ matrices.
μ	Coherence parameter in Proposition 1
s_{max}	Sparse parameter in Proposition 3
κ	Matrix condition number used in Proposition 4
p	Sample rate $\frac{\|\Omega\|}{m(n+n_e)}$ used in Proposition 4
C, c, C_1, C_2, C'	Numerical constants
$\sigma_i, \sigma_{min}, \sigma_{max}$	i^{th}, minimum, maximum singular value.
θ_i	i^{th} canonical angle.
Θ, Φ	Diagonal canonical angle matrix.
$\|\cdot\|$	Either absolute value or cardinality.
$\|\cdot\|_2$	2-norm of vector/spectral norm of matrix.
$\|\cdot\|_F$	Frobenious norm of a matrix.
$\|\cdot\|$	In Theorem 3 means both Frobenious norm and spectral norm, otherwise same as $\|\cdot\|_2$.