A gentle introduction to row-rank matrix completion

Zhiling Gu

November 22, 2022

Abstract

In this report, we studied the Section 3.8 of Chen et al. [2021]. Overall, the objective is to estimate the unobserved entries of a not fully observed matrix of interest, $\boldsymbol{M}^\star$, of dimension $[n_1] \times [n_2]$. We considered the SVD of $\boldsymbol{M}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top}$, and constructed the rank-r decomposition $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ of a candidate approximation matrix $\mathbf{M}$. The approximation accuracy of $\mathbf{U}$, $\mathbf{V}$, and $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ for $\boldsymbol{U}^\star$, $\boldsymbol{V}^\star$, and $\boldsymbol{M}^\star$ were assessed with condition number κ and the incoherence parameter μ in Theorems 3.22–3.23 [Chen et al., 2021]. These two theorems and related lemmas were studied in detail. This report serves as EE623 Term Paper in Fall 2022.

1. Motivation

In the practice, it is extremely common to encounter missing data due to collection difficulty, erroneous data, and etc. And most of the data can be represented in the matrix. For example, if we consider each row of a matrix is the features/ measurements of a single subject, a matrix would represent the features of all the subjects/ population of interest. To tackle the missing data problem, one important tool is matrix completion. See Figure 1 for illustration.

2. Preliminaries

In this section, the framework of matrix completion is given. We refer the readers to Section 6 for the basic definitions and theorems used.

2.1. Problem formulation and assumption

Suppose the data matrix M^* is of dimension $n_1 \times n_2$ with rank r. Assume

$$n_1 \leq n_2.$$

We start with the single value decomposition of $\boldsymbol{M}^\star$ as follows

$$\boldsymbol{M}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top},$$

where $col(\boldsymbol{U}^\star) \in \mathbb{R}^{n_1 \times r}, col(\boldsymbol{V}^\star) \in \mathbb{R}^{n_2 \times r}$, and $\boldsymbol{\Sigma}^\star$ is a diagonal matrix with entries singular values, denoted as $\sigma_1(\boldsymbol{M}^\star), \ldots, \sigma_r(\boldsymbol{M}^\star)$ in descending order. And we introduce *condition number* of matrix $\boldsymbol{M}^\star$ to be the ratio of the largest singular value and the r-th largest singular value,

$$\kappa := \sigma_1(\boldsymbol{M}^\star)/\sigma_r(\boldsymbol{M}^\star),$$

and we define an index subset $\Omega \subset [n_1] \times [n_2]$ such that $(i, j) \in \Omega \iff {\boldsymbol{M}^\star}_{ij}$ is observed.

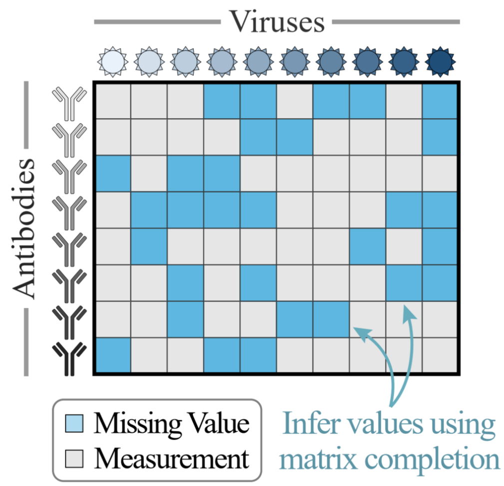

Figure 1: source: `https://www.fredhutch.org/en/news/spotlight/2022/08/bs-einav-cellsys.html`

Assumption 1 (Random sampling). In this report, we assume each entry of $\boldsymbol{M}^\star$ is observed independently with probability $0 < p < 1$. This corresponds to *missing at random* in statistics terminology.

Example 1 (Incoherence). Here we provide an example that satisfies random sampling but causes unfaithful recovery. Consider $\boldsymbol{M}^\star$ being a zero matrix except for 1 entry. If $p = o(1)$, then with high probability, the single nonzero entry would be missing, and any recovery method would be in vain to recover the rank 1 property.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

μ-incoherent. Motivated by the previous example, we define the *incoherence parameter* μ of $\boldsymbol{M}^\star$ as follows

$$\mu := \max\left\{\frac{n_1\|\boldsymbol{U}^\star\|_{2,\infty}^2}{r}, \frac{n_2\|\boldsymbol{V}^\star\|_{2,\infty}^2}{r}\right\}.$$

Recall that $\|\boldsymbol{U}^\star\|_{2,\infty} = \max_i \|\boldsymbol{U}^\star{}_{i,\cdot}\|_2$ is the largest ℓ_2 norm among rows of $\boldsymbol{U}^\star$. Also note by SVD, $\boldsymbol{U}^\star$ and $\boldsymbol{V}^\star$ are unitary matrices, and thus $\boldsymbol{U}^\star\boldsymbol{U}^{\star\top} = \mathbf{I}_r$ leading to $\|\boldsymbol{U}^\star\|_F = r$. The following inequality shows $\mu \geq 1$, and a smaller μ indicates the information of matrix is spread out across different rows and columns.

$$\frac{r}{n_1} = \frac{1}{n_1}\|\boldsymbol{U}^\star\|_F^2 \leq \|\boldsymbol{U}^\star\|_{2,\infty}^2 \leq \|\boldsymbol{U}^\star\|^2 = 1 \implies 1 \leq \mu \leq \max\{n_1, n_2\}/r = n_2/r.$$

Euclidean projection operator: $\mathcal{P}_\Omega : \mathbb{R}^{n_1\times n_2} \to \mathbb{R}^{n_1\times n_2}$. It is now natural to define a projection from original space $\mathbb{R}^{n_1\times n_2}$ where $\boldsymbol{M}^\star$ lies in a subspace of $\mathbb{R}^{n_1\times n_2}$ as follows:

$$[\mathcal{P}_\Omega(\boldsymbol{M}^\star)]_{ij} = \begin{cases} \boldsymbol{M}^\star{}_{ij}, & \text{if}(i,j) \in \Omega \\ 0, & \text{else.} \end{cases}$$

And our goal is to recover $\boldsymbol{M}^\star$ on the basis of $\mathcal{P}_\Omega(\boldsymbol{M}^\star)$.

2.2. Algorithm

Under the assumption of random sampling, we consider the recovery of $\boldsymbol{M}^\star$, $\boldsymbol{M}$, as the *inverse probability weighted average* of observed data matrix

$$\boldsymbol{M} := p^{-1}\mathcal{P}_\Omega(\boldsymbol{M}^\star). \tag{1}$$

Since the observed data is in the random subspace $\mathcal{P}_\Omega(\boldsymbol{M}^\star)$, $\boldsymbol{M}$ is in fact a random recovery due to the randomness of sampling over Ω. This construction leads to

$$\mathbb{E}_\Omega(\boldsymbol{M}) = \boldsymbol{M}^\star.$$

Then we compute the rank-r SVD of $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, following which $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, $\boldsymbol{U}$, and $\boldsymbol{V}$ are employed as the estimates of $\boldsymbol{M}^\star$, $\boldsymbol{U}^\star$, and $\boldsymbol{V}^\star$, respectively.

Example 2 (Inverse probability weighting). Here we provide an example of the process of inverse probability weighting.

$$\text{True matrix } \boldsymbol{M}^\star = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 1 & 3 & 1 \\ 4 & 1 & 1 & 3 \end{pmatrix}$$

$$\text{Observed matrix} = \begin{pmatrix} 1 & ? & ? & 2 \\ 2 & 1 & 3 & 1 \\ 4 & 1 & 1 & ? \end{pmatrix}$$

$$\mathcal{P}_\Omega(\boldsymbol{M}^\star) = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 2 & 1 & 3 & 1 \\ 4 & 1 & 1 & 0 \end{pmatrix}$$

$$\text{Approximation matrix } \boldsymbol{M} := p^{-1}\mathcal{P}_\Omega(\boldsymbol{M}^\star).$$

Assume p and r are known. $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}$ is the rank-r SVD of M. A natural question to ask is how close are $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}$ and $\boldsymbol{M}^\star$; $\boldsymbol{U}^\star$ and $\boldsymbol{U}$; $\boldsymbol{V}^\star$ and $\boldsymbol{V}$ from each other, respectively.

3. Main results

In this section, we present the main theoretical results.

Lemma 1 (Useful bounds of matrix norms, Lemma 3.20 of Chen et al. [2021]). Assume $\boldsymbol{M}^\star \in \mathbb{R}^{n_1 \times n_2}$ is μ-coherent. Then the following relations holds

$$\|\boldsymbol{M}^\star\|_{2,\infty} \le \sqrt{\mu r/n_1}\|\boldsymbol{M}^\star\| \tag{2}$$

$$\|\boldsymbol{M}^{\star\top}\|_{2,\infty} \le \sqrt{\mu r/n_2}\|\boldsymbol{M}^\star\| \tag{3}$$

$$\|\boldsymbol{M}^\star\|_\infty \le \mu r\sqrt{1/n_1 n_2}\|\boldsymbol{M}^\star\|. \tag{4}$$

Lemma 2 (Perturbation bound of $\boldsymbol{M}$, Lemma 3.21 of Chen et al. [2021]). Suppose $n_2 p \ge C\mu r \log n_2$ for some constant $C > 0$, then with probability at least $1 - O(n_2^{-10})$, one has

$$\|\boldsymbol{M} - \boldsymbol{M}^\star\| \lesssim \sqrt{\frac{\mu r \log n_2}{n_1 p}}\|\boldsymbol{M}^\star\|.$$

Theorem 1 (Recovery of $\boldsymbol{U}, \boldsymbol{V}$, Theorem 3.22 of Chen et al. [2021]). Suppose $n_1 p \geq C\kappa^2 \mu r \log n_2$ for some constant $C > 0$, then with probability at least $1 - O(n_2^{-10})$, one has

$$\max\left\{\text{dist}\left(\boldsymbol{U}, \boldsymbol{U}^\star\right), \text{dist}\left(\boldsymbol{V}, \boldsymbol{V}^\star\right)\right\} \lesssim \kappa\sqrt{\frac{\mu r \log n_2}{n_1 p}}.$$

Note that when the sample size $pn_1n_2 \gg \kappa^2 \mu r n_2 \log n_2$, the spectral estimate achieves consistent estimation $\max\left\{\text{dist}\left(\boldsymbol{U}, \boldsymbol{U}^\star\right), \text{dist}\left(\boldsymbol{V}, \boldsymbol{V}^\star\right)\right\} = o_p(1)$.

Theorem 2 (Recovery of $\boldsymbol{M}$, Theorem 3.23 of Chen et al. [2021]). Suppose $n_2 p \geq C\mu r \log n_2$ for some constant $C > 0$, then with probability at least $1 - O(n_2^{-10})$, one has

$$\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top - \boldsymbol{M}^\star\|_F \lesssim \sqrt{\frac{\mu r^2 \log n_2}{n_1 p}}\|\boldsymbol{M}^\star\|$$

The theorem above only requires Lemma 2 and characterizes the statistical accuracy of $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$.

4. Proofs of lemmas

We start with the lemma for basic inequalities of matrix norms.

Proof of Lemma 1. Sketch of proof: Prove the three basic inequalities of matrix norms, and make use of the definition of incoherence parameter.

We start with the following basic inequalities of matrix norms.

$$\text{(i) } \|\boldsymbol{A}\boldsymbol{B}\|_{2,\infty} \leq \|\boldsymbol{A}\|_{2,\infty}\|\boldsymbol{B}\|,$$
$$\text{(ii) } \|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|,$$
$$\text{(iii) } \|\boldsymbol{A}\boldsymbol{B}^\top\|_\infty \leq \|\boldsymbol{A}\|_{2,\infty}\|\boldsymbol{B}\|.$$

To see the three inequalities, we can proceed as below. Define $\boldsymbol{e}_j$ as the indicator vector, where the j-th entry is one, zero elsewhere. Consider SVD $\boldsymbol{A} = \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^\top$, and $\boldsymbol{B} = \boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}_2^\top$.

(i) $\|\boldsymbol{A}\boldsymbol{B}\|_{2,\infty} = \max_i \|\boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{B}\|_2 = \max_i \|\boldsymbol{e}_i^\top \boldsymbol{A}\|_2 \|\boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}_2^\top\|_2 \leq \|\boldsymbol{A}\|_{2,\infty}\|\boldsymbol{\Sigma}_2\|_2 \leq \|\boldsymbol{A}\|_{2,\infty}\|\boldsymbol{B}\|$.

(ii) $\|\boldsymbol{A}\boldsymbol{B}\| = \|\boldsymbol{A}\boldsymbol{B}\|_{op} = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{B}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq \left\{\max_{\boldsymbol{B}\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{B}\boldsymbol{x}\|_2}{\|\boldsymbol{B}\boldsymbol{x}\|_2}\right\} \cdot \left\{\max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{B}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}\right\} = \|\boldsymbol{A}\|_{op}\|\boldsymbol{B}\|_{op} = \|\boldsymbol{A}\|\|\boldsymbol{B}\|$.

(iii) $\|\boldsymbol{A}\boldsymbol{B}^\top\|_\infty = \max_{ij} |\boldsymbol{e}_i^\top \boldsymbol{A}\boldsymbol{B}^\top \boldsymbol{e}_j| \leq \max_i \|\boldsymbol{e}_i^\top \boldsymbol{A}\|_2 \|\boldsymbol{B}^\top \boldsymbol{e}_j\|_2 = \|\boldsymbol{A}\|_{2,\infty}\|\boldsymbol{B}\|_{2,\infty}$ by Cauchy-Schwartz inequality. In addition, since $\|\boldsymbol{B}\|_{2,\infty} \leq \|\boldsymbol{B}\|$, we proved the third inequality.

Equipped with the inequalities above, we consider

$$\begin{aligned}
\|\boldsymbol{M}^\star\|_{2,\infty} &= \|\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{V}^{\star\top}\|_{2,\infty} \\
&\leq \|\boldsymbol{U}^\star\|_{2,\infty}\|\boldsymbol{\Sigma}^\star\|\|\boldsymbol{V}^\star\| \text{ by (i)\& (ii)} \\
&\leq \frac{\sqrt{\mu r}}{\sqrt{n_1}}\|\boldsymbol{M}^\star\| \text{ by definition of incoherence parameter} \mu \geq n_1\|\boldsymbol{U}^\star\|_{2,\infty}^2/r,
\end{aligned}$$

In addition,

$$\begin{aligned}\|\boldsymbol{M}^\star\|_\infty &= \|\boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{V}^{\star\top}\|_\infty \\ &\le \|\boldsymbol{U}^\star\|_{2,\infty}\|\boldsymbol{\Sigma}^\star\|\|\boldsymbol{V}^\star\|_{2,\infty} \text{ by (i)\& (iii)} \\ &\le \frac{\sqrt{\mu r}}{\sqrt{n_1}}\|\boldsymbol{M}^\star\|\frac{\sqrt{\mu r}}{\sqrt{n_2}} = \frac{\mu r}{\sqrt{n_1 n_2}}\|\boldsymbol{M}^\star\|.\end{aligned}$$

We proved the first and third inequality of Lemma 1, and the second inequality follows exactly the same as the first one. □

Proof of Lemma 2. Sketch of proof: (i) Decompose elements of $\boldsymbol{E}$ as sum of independent random matrices $\boldsymbol{X}_{ij}$. (ii) Apply matrix Bernstein inequality to $\boldsymbol{E}$.

Step (i): Recall that $\boldsymbol{E} = p^{-1}\mathcal{P}_\Omega(\boldsymbol{M}^\star) - \boldsymbol{M}^\star$, and can be written as follows

$$\begin{aligned}p^{-1}\mathcal{P}_\Omega(\boldsymbol{M}^\star) - \boldsymbol{M}^\star &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\boldsymbol{X}_{ij} \\ \boldsymbol{X}_{ij} &= (p^{-1}\delta_{ij} - 1)M^*_{ij}\boldsymbol{e}_i\boldsymbol{e}_j^\top,\end{aligned}$$

where $\delta_{ij} \sim Ber(p)$ is indicator random variable for that (i,j)-th entry is observed; $\boldsymbol{e}_i$ is the i-th standard basis vector of appropriate dimension. It cann be seen that

$$\mathbb{E}(\boldsymbol{X}_{ij}) = \boldsymbol{0}, \quad \|\boldsymbol{X}_{ij}\| \le \frac{1}{p}\|\boldsymbol{M}^\star\|_\infty \le \frac{\mu r}{p\sqrt{n_1 n_2}}\|\boldsymbol{M}^\star\|,$$

by Lemma 1.

Step (ii): Apply matrix Bernstein inequality (Theorem 3), we have

$$\|\boldsymbol{E}\| \le \sqrt{2av\log n_2} + \frac{2a}{3}\frac{\mu r}{p\sqrt{n_1 n_2}}\|\boldsymbol{M}^\star\|\log n_2, \quad \forall a > 2,$$

where

$$v = \max\left\{\left\|\sum_{ij}\mathbb{E}\left[(\boldsymbol{X}_{ij})(\boldsymbol{X}_{ij})^\top\right]\right\|, \left\|\sum_{ij}\mathbb{E}\left[(\boldsymbol{X}_{ij})^\top(\boldsymbol{X}_{ij})\right]\right\|\right\}.$$

For the first term in v, we have

$$\begin{aligned}\sum_{ij}\mathbb{E}(\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^\top) &= \sum_{ij}\mathbb{E}\left\{(p^{-1}\delta_{ij}-1)^2(M^*_{ij})^2\boldsymbol{e}_i\boldsymbol{e}_j^\top\boldsymbol{e}_j\boldsymbol{e}_i^\top\right\} \\ &= \frac{1-p}{p}\sum_{ij}(M^*_{ij})^2\boldsymbol{e}_i\boldsymbol{e}_i^\top \text{ by random sampling } \delta_{ij} \sim Ber(p) \\ &= \frac{1-p}{p}\sum_{i=1}^{n_1}\|\boldsymbol{M}^\star{}_{i,\cdot}\|_2^2\boldsymbol{e}_i\boldsymbol{e}_i^\top \\ &\preceq \frac{1-p}{p}\|\boldsymbol{M}^\star\|_{2,\infty}^2\sum_{i=1}^{n_1}\boldsymbol{e}_i\boldsymbol{e}_i^\top \\ &\preceq \frac{\mu r}{n_1 p}\|\boldsymbol{M}^\star\|^2\boldsymbol{I}_{n_1} \text{ by Lemma 1}\end{aligned}$$

where $\boldsymbol{A} \preceq \boldsymbol{B} \iff \boldsymbol{B} - \boldsymbol{A}$ is positive semidefinite. Similarly, we can derive

$$\sum_{ij} \mathbb{E}(\boldsymbol{X}_{ij}^\top \boldsymbol{X}_{ij}) \preceq \frac{\mu r}{n_1 p} \|\boldsymbol{M}^\star\|^2 \boldsymbol{I}_{n_2}$$

Thus, we can bound v as follows

$$v \leq \frac{\mu r}{n_1 p} \|\boldsymbol{M}^\star\|^2$$

by noting that $n_1 \leq n_2$. Combine the bound of v and the result of Bernstein inequality, we have

$$\|\boldsymbol{E}\| \lesssim \sqrt{\frac{\mu r \|\boldsymbol{M}^\star\|^2 \log n_2}{n_1 p}} + \frac{\mu r \|\boldsymbol{M}^\star\| \log n_2}{p\sqrt{n_1 n_2}}$$

with probability at least $1 - O(n_2^{-10})$ by setting $a = 11$. Since $\log n_2 \ll \sqrt{n_2}$, the second term above diminishes as n_2 becomes large. In particular, when $n_2 \gtrsim \mu r \|\boldsymbol{M}^\star\|^2 \log n_2$, the first term dominates the second term, which leads to

$$\|\boldsymbol{E}\| \lesssim \sqrt{\frac{\mu r \|\boldsymbol{M}^\star\|^2 \log n_2}{n_1 p}}.$$

□

5. Proofs of theorems

Proof of Theorem 1. Sketch of the proof: (i) Prove $\boldsymbol{E} = \boldsymbol{M} - \boldsymbol{M}^\star$ satisfy $\|\boldsymbol{E}\| < \sigma_r(\boldsymbol{M}^\star) - \sigma_{r+1}(\boldsymbol{M}^\star)$, where $\sigma_r(\boldsymbol{M}^\star)$ is the r-th largest singular value of $\boldsymbol{M}^\star$. (ii) Apply Wedin's theorem to $\boldsymbol{E}$ and use lemma 2.

Step (i): recall $n_1 \geq n_2$, thus the condition of lemma 2 is satisfied. Then

$$\|\boldsymbol{E}\| = \|\boldsymbol{M} - \boldsymbol{M}^\star\| \lesssim \sqrt{\frac{\mu r \log n_2}{n_1 p}} \|\boldsymbol{M}^\star\|.$$

In addition, recall that $\sigma_1(\boldsymbol{M}^\star) = \|\boldsymbol{M}^\star\| = \kappa \sigma_r(\boldsymbol{M}^\star)$ by definition of singular value and κ. Therefore

$$\begin{aligned} \|\boldsymbol{E}\| \lesssim & \sqrt{\frac{\mu r \log n_2}{n_1 p}} \|\boldsymbol{M}^\star\| = \sqrt{\frac{\kappa^2 \mu r \log n_2}{n_1 p}} \sigma_r(\boldsymbol{M}^\star) \\ \leq & \frac{1}{C} \sigma_r(\boldsymbol{M}^\star) \text{ for some large enough } C > 0. \end{aligned}$$

Choose C such that $1/C < 1 - 1/\sqrt{2}$, we have

$$\|\boldsymbol{E}\| \lesssim (1 - \frac{1}{\sqrt{2}}) \sigma_r(\boldsymbol{M}^\star).$$

Note we can always choose a large enough C such that the condition of Wedin's theorem $\|\boldsymbol{E}\| < \sigma_r(\boldsymbol{M}^\star) - \sigma_{r+1}(\boldsymbol{M}^\star)$ holds.

Step (ii): Apply Wedin's theorem to $\boldsymbol{E}$, we have

$$\begin{aligned}
\max\left\{\operatorname{dist}\left(\boldsymbol{U},\boldsymbol{U}^{\star}\right),\operatorname{dist}\left(\boldsymbol{V},\boldsymbol{V}^{\star}\right)\right\} &\leq \frac{\sqrt{2}\max\left\{\left\|\boldsymbol{E}^{\top}\boldsymbol{U}^{\star}\right\|,\left\|\boldsymbol{E}\boldsymbol{V}^{\star}\right\|\right\}}{\sigma_r(\boldsymbol{M}^{\star})-\sigma_{r+1}(\boldsymbol{M}^{\star})-\|\boldsymbol{E}\|} \text{ by Wedin's theorem} \\
&\leq \frac{\sqrt{2}\|\boldsymbol{E}\|\max\left\{\|\boldsymbol{U}^{\star}\|,\|\boldsymbol{V}^{\star}\|\right\}}{\sigma_r(\boldsymbol{M}^{\star})-\|\boldsymbol{E}\|} \text{ by } \|AB\|\leq\|A\|\|B\| \\
&\leq \frac{\sqrt{2}\|\boldsymbol{E}\|}{\sigma_r^{\star}-(1-\frac{1}{\sqrt{2}})\sigma_r(\boldsymbol{M}^{\star})} \text{ by unitary matrix } \boldsymbol{U}^{\star},\boldsymbol{V}^{\star} \\
&= 2\|\boldsymbol{E}\|/\sigma_r(\boldsymbol{M}^{\star}) = 2\kappa\|\boldsymbol{E}\|/\sigma_1(\boldsymbol{M}^{\star}) \\
&\lesssim \kappa\sqrt{\frac{\mu r\log n_2}{n_1 p}} \text{ by Lemma 2.}
\end{aligned}$$

□

Proof of Theorem 2. By triangle inequality, we have

$$\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}-\boldsymbol{M}^{\star}\| \leq \|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}-\boldsymbol{M}\|+\|\boldsymbol{M}-\boldsymbol{M}^{\star}\|.$$

Note that $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}$ is the SVD of $\boldsymbol{M}$ and thus the best rank-r approximation to $\boldsymbol{M}$. Therefore $\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}-\boldsymbol{M}\|\leq\|\boldsymbol{M}-\boldsymbol{M}^{\star}\|$, where $\boldsymbol{M}^{\star}$ is an unknown rank-r matrix.

In addition, since both $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}$ and $\boldsymbol{M}^{\star}$ are of rank r, the difference between them would have rank at most $2r$. This leads to

$$\begin{aligned}
\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}-\boldsymbol{M}^{\star}\| &\leq \|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}-\boldsymbol{M}\|+\|\boldsymbol{M}-\boldsymbol{M}^{\star}\| \\
&\leq 2\|\boldsymbol{M}-\boldsymbol{M}^{\star}\| \\
\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}-\boldsymbol{M}^{\star}\|_F &\leq \sqrt{2r}\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}-\boldsymbol{M}^{\star}\| \text{ by Remark *} \\
&\leq 2\sqrt{2r}\|\boldsymbol{M}-\boldsymbol{M}^{\star}\| \\
&\lesssim \sqrt{\frac{\mu r^2\log n_2}{n_1 p}}\|\boldsymbol{M}^{\star}\| \text{ by Lemma 2.}
\end{aligned}$$

Remark *: $\|\boldsymbol{M}\|_F\leq\sqrt{r}\|\boldsymbol{M}\|$. This can be proved by recognizing $\|\boldsymbol{M}\|_F=\sqrt{\sum_i(\boldsymbol{e}_i^{\top}\boldsymbol{M}^{\top}\boldsymbol{M}\boldsymbol{e}_i)}=\sqrt{\sum_i(\boldsymbol{e}_i^{\top}\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\top}\boldsymbol{e}_i)}\leq\sqrt{\sum_{i=1}^{r}\|\boldsymbol{e}_i^{\top}\boldsymbol{U}\|_2\|\boldsymbol{M}\|^2\|\boldsymbol{V}^{\top}\boldsymbol{e}_i\|_2}=\sqrt{r}\|\boldsymbol{M}\|$. The second last inequality holds because $\boldsymbol{\Sigma}$ is only nonzero on the first r diagonal entries (sigular values of $\boldsymbol{M}$), which are at most $\|\boldsymbol{M}\|=\|\boldsymbol{M}\|_{op}$. The last equality holds because $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices. □

6. Miscellaneous

6.1. Norms

For any vector $\boldsymbol{v}$, we denote by $\|\boldsymbol{v}\|_2$, $\|\boldsymbol{v}\|_1$ and $\|\boldsymbol{v}\|_\infty$ its ℓ_2 norm, ℓ_1 norm and ℓ_∞ norm, respectively. For any matrix $\boldsymbol{A}=[A_{i,j}]_{1\leq i\leq m,1\leq j\leq n}$, we let $\|\boldsymbol{A}\|$, $\|\boldsymbol{A}\|_*$, $\|\boldsymbol{A}\|_{\mathrm{F}}$ and $\|\boldsymbol{A}\|_\infty$ represent respectively its spectral norm (i.e., the largest singular value of $\boldsymbol{A}$), its nuclear norm (i.e., the sum of singular values of $\boldsymbol{A}$), its Frobenius norm (i.e., $\|\boldsymbol{A}\|_{\mathrm{F}}:=\sqrt{\sum_{i,j}A_{i,j}^2}$), and its entrywise ℓ_∞ norm (i.e.,

$\|\boldsymbol{A}\|_\infty := \max_{i,j} |A_{i,j}|)$. In particular, singular values of $\boldsymbol{A}$ are square roots of the eigenvalues of $\boldsymbol{A}^H\boldsymbol{A}$. The largest singular value $\sigma_1(\boldsymbol{A})$= operator norm $\|\boldsymbol{A}\|_{op} := \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2$.

We also refer to $\|\boldsymbol{A}\|_{2,\infty}$ as the $\ell_{2,\infty}$ norm of $\boldsymbol{A}$, defined as $\|\boldsymbol{A}\|_{2,\infty} := \max_i \|\boldsymbol{A}_{i,\cdot}\|_2$. Similarly, we define the $\ell_{\infty,2}$ norm of $\boldsymbol{A}$ as $\|\boldsymbol{A}\|_{\infty,2} := \left\|\boldsymbol{A}^\top\right\|_{2,\infty}$. In addition, for any matrices $\boldsymbol{A} = [A_{i,j}]_{1\le i\le m,1\le j\le n}$ and $\boldsymbol{B} = [B_{i,j}]_{1\le i\le m,1\le j\le n}$, the inner product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as and denoted by $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \sum_{1\le i\le m,1\le j\le n} A_{i,j}B_{i,j} = \operatorname{Tr}\left(\boldsymbol{A}^\top\boldsymbol{B}\right)$.

6.2. *Singular Value Decomposition*

Consider $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E}$ and $\boldsymbol{M}^\star$ be two matrices of $\mathbb{R}^{n_1\times n_2}$, $n_1 \le n_2$. Let $\boldsymbol{M}^\star = \boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{V}^\star$, $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}$ as follows

$$\boldsymbol{M}^\star = \sum_{i=1}^{n_1} \sigma_i^\star u_i^\star \boldsymbol{v}_i^{\star\top} = \left[\begin{array}{cc} \boldsymbol{U}^\star & \boldsymbol{U^\star}_\perp \end{array}\right] \left[\begin{array}{ccc} \boldsymbol{\Sigma}^\star & 0 & 0 \\ 0 & \boldsymbol{\Sigma^\star}_\perp & 0 \end{array}\right] \left[\begin{array}{c} \boldsymbol{V}^{\star\top} \\ \boldsymbol{V}_\perp^{\star\top} \end{array}\right];$$
$$\boldsymbol{M} = \sum_{i=1}^{n_1} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \left[\begin{array}{cc} \boldsymbol{U} & \boldsymbol{U}_\perp \end{array}\right] \left[\begin{array}{ccc} \boldsymbol{\Sigma} & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_\perp & 0 \end{array}\right] \left[\begin{array}{c} \boldsymbol{V}^\top \\ \boldsymbol{V}_\perp^\top \end{array}\right].$$

Here, $\sigma_1 \ge \cdots \ge \sigma_{n_1}$ (resp. $\sigma_1^\star \ge \cdots \ge \sigma_{n_1}^\star$) stand for the singular values of M (resp. $M^\star$) arranged in descending order, $\boldsymbol{u}_i$ (resp. $\boldsymbol{u}_i^\star$) denotes the left singular vector associated with the singular value σ_i (resp. $\sigma_i^\star$), and $\boldsymbol{v}_i$ (resp. $\boldsymbol{v}_i^\star$) represents the right singular vector associated with σ_i (resp. $\sigma_i^\star$). In addition, we denote

$$\begin{aligned} &\boldsymbol{\Sigma} := \operatorname{diag}\left([\sigma_1, \cdots, \sigma_r]\right), && \boldsymbol{\Sigma}_\perp := \operatorname{diag}\left([\sigma_{r+1}, \cdots, \sigma_{n_1}]\right), \\ &\boldsymbol{U} := [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r] \in \mathbb{R}^{n_1\times r}, && \boldsymbol{U}_\perp := [\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_{n_1}] \in \mathbb{R}^{n_1\times(n_1-r)}, \\ &\boldsymbol{V} := [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r] \in \mathbb{R}^{n_2\times r}, && \boldsymbol{V}_\perp := [\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_{n_2}] \in \mathbb{R}^{n_2\times(n_2-r)} \end{aligned}$$

We define $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ as the rank-r SVD of $\boldsymbol{M}^\star$. The matrices $\boldsymbol{\Sigma}^\star, \boldsymbol{\Sigma}_\perp^\star, \boldsymbol{U}^\star, \boldsymbol{U}_\perp^\star, \boldsymbol{V}^\star, \boldsymbol{V}_\perp^\star$ are defined analogously.

In addition, we define the distance between two matrices as

$$\operatorname{dist}\left(\boldsymbol{U}, \boldsymbol{U}^\star\right) := \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\| \tag{5}$$

6.3. *Matrix Bernstein and Wedin Theorem*

Theorem 3 (Matrix Bernstein, Corollary 3.3 of Chen et al. [2021]). Let $\{\boldsymbol{X}_i\}_{1\le i\le m}$ be a set of independent real random matrices with dimension $n_1 \times n_2$. Suppose that

$$\mathbb{E}\left[\boldsymbol{X}_i\right] = \boldsymbol{0}, \quad \text{and} \quad \|\boldsymbol{X}_i\| \le L, \quad \text{for all } i.$$

Set $n := \max\{n_1, n_2\}$, and variance statistic

$$\begin{aligned} v := \max \Bigg\{ &\left\| \sum_{i=1}^m \mathbb{E}\left[\left(\boldsymbol{X}_i - \mathbb{E}\left[\boldsymbol{X}_i\right]\right)\left(\boldsymbol{X}_i - \mathbb{E}\left[\boldsymbol{X}_i\right]\right)^\top\right] \right\|, \\ &\left\| \sum_{i=1}^m \mathbb{E}\left[\left(\boldsymbol{X}_i - \mathbb{E}\left[\boldsymbol{X}_i\right]\right)^\top\left(\boldsymbol{X}_i - \mathbb{E}\left[\boldsymbol{X}_i\right]\right)\right] \right\| \Bigg\}. \end{aligned} \tag{6}$$

For any $a \geq 2$, with probability exceeding $1 - 2n^{-a+1}$ one has

$$\left\| \sum_{i=1}^{m} \boldsymbol{X}_i \right\| \leq \sqrt{2av \log n} + \frac{2a}{3} L \log n.$$

Theorem 4 (Wedin $\sin \Theta$ theorem for singular subspace perturbation, Theorem 3.22 of Chen et al. [2021]). If $\|\boldsymbol{E}\| < \sigma_r^\star - \sigma_{r+1}^\star$, then one has

$$\max \left\{ \operatorname{dist}\left(\boldsymbol{U}, \boldsymbol{U}^\star\right), \operatorname{dist}\left(\boldsymbol{V}, \boldsymbol{V}^\star\right) \right\} \leq \frac{\sqrt{2} \max\left\{ \left\|\boldsymbol{E}^\top \boldsymbol{U}^\star\right\|, \left\|\boldsymbol{E}\boldsymbol{V}^\star\right\| \right\}}{\sigma_r^\star - \sigma_{r+1}^\star - \|\boldsymbol{E}\|}.$$

7. Summary

In this report, we studied a simple setting of random sampling (i.e., missing at random), where the probability of an entry being missing is equal across the matrix, and the events of missing entries are independent. We introduced two key parameters: (i) condition number κ, and (ii) incoherence parameter μ of matrix $\boldsymbol{M}^\star$ to describe the property of matrix $\boldsymbol{M}^\star$. The former shows the gap between the information provided by the most and least informative singular sub-spaces. The latter describes how spread-out the information is across the singular sub-spaces.

We then adopted the inverse probability weighting to produce approximation of $\boldsymbol{U}^\star$, $\boldsymbol{V}^\star$ and $\boldsymbol{M}^\star$ by $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ is the rank-r SVD of $p^{-1}\mathcal{P}_\Omega(\boldsymbol{M}^\star)$, and $\mathcal{P}_\Omega(\boldsymbol{M}^\star)$ is the matrix with missing entries replaced by 0. The proofs of approximation bounds

$$\max \left\{ \operatorname{dist}\left(\boldsymbol{U}, \boldsymbol{U}^\star\right), \operatorname{dist}\left(\boldsymbol{V}, \boldsymbol{V}^\star\right) \right\} \lesssim \kappa \sqrt{\frac{\mu r \log n_2}{n_1 p}},$$

$$\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top - \boldsymbol{M}^\star\|_F \lesssim \sqrt{\frac{\mu r^2 \log n_2}{n_1 p}} \|\boldsymbol{M}^\star\|,$$

and related lemmas are then carefully studied.

References

Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806, 2021. doi: 10.1561/2200000079. URL https://doi.org/10.1561%2F2200000079.