

## Lecture 16 — March 7

Lecturer: Caramanis &amp; Sanghavi

Scribe: Tianyang Bai

## 16.1 Topics Covered

In this lecture, we introduced one method of matrix completion via SVD-based PCA. Specifically, we covered the following topics.

1. Introduction: problem formation and applications
2. One toy example: recovering a matrix of rank 1
3. SVD-based algorithm
4. Proof for the effectiveness of the algorithm

## 16.2 Introduction

We would like to motivate the matrix completion problem through the example of collaborative filtering. Imagine that each of  $m$  customers watches and rates a subset of the  $n$  movies available through a website. This yields a dataset of customer-movie pairs  $(i, j) \in E \subseteq [m] \times [n]$ . For each such pair, a rating  $M_{ij} \in \mathbb{R}$  is known. The objective of collaborative filtering is to predict the rating for the missing pairs as to provide suggestions, based on the previous rating results. In addition, another application of matrix completion is localization from distance measurements, a topic we discussed previously.

### 16.2.1 Mathematical Model

A mathematical model is made to solve the collaborative filtering problem as follows. We denote by  $M$  the matrix whose entry  $(i, j) \in [m] \times [n]$  corresponds to the rating from user  $i$  to movie  $j$ . Out of the  $m \times n$  entries of  $M$ , only a subset  $E \subseteq [m] \times [n]$  is known. We let  $M^E$  be the  $m \times n$  matrix defined as

$$M_{i,j}^E = \begin{cases} M_{i,j} & \text{if } (i, j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

The set  $E$  is uniformly chosen from  $[m] \times [n]$  given its size  $|E|$ . Furthermore, we also assume  $M$  has low rank  $r \ll m, n$  for the following reasons.

1. Empirical datasets show that  $M$  has low rank.

**Table 16.1.** Matrix Completion Algorithm

- 
- 1: Trimming: Trim  $M^E$ , and let  $\bar{M}^E$  be the output.
  - 2: Compute the SVD of  $\bar{M}^E$ . Let  $\bar{M}^E = \sum_{i=1}^{\min(m,n)} \sigma_i x_i y_i^T$ , where  $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$ .
  - 3: Projection: Compute the recovered matrix  $M$  as  $M = \mathbf{P}_r(\bar{M}^E) = \frac{mn}{|E|} \sum_{i=1}^k \sigma_i x_i y_i^T$ .
- 

2. The clustered nature of the matrix  $M$  suggests low rank. Namely, users of similar interests are more likely to give similar ratings to the films of similar features.

### 16.2.2 A Toy Problem: Recovering a matrix of rank 1

In the class, we first discussed the problem of how to complete a matrix of rank 1 to illustrate the key ideas of matrix completion. The toy problem is formatted as follows.

Suppose  $M = UV$ , where  $U$  is an  $m \times 1$  unknown vector, and  $V$  is an  $n \times 1$  unknown vector. Hence  $M$  has rank 1. We want to recover  $M$ , i.e. to find appropriate vector  $U$  and  $V$  from  $M^E$ , which is the matrix containing all observed entries as defined in the previous section.

Firstly, note that the solution of  $U$  and  $V$  is not unique. The reason is that if  $U^*$  and  $V^*$  satisfy the constraints, then for any constant  $c \neq 0$ ,  $cU^*$  and  $\frac{1}{c}V^*$  are also a proper pair of solutions. Hence without loss of generosity, we assume  $U_1 = 1$ .

Next,  $\forall j \in [n]$ , if  $(1, j) \in E$ , we can recover  $V_j = \frac{M_{1,j}^E}{U_1}$ . In the similar way, having known  $V_j$ , we can further recover  $U_i = \frac{M_{i,j}^E}{V_j}$  as long as  $(i, j) \in E$ . Hence, if the matrix  $M_E$  is "connected", we can iteratively recover all the entries of  $U$  and  $V$ . We say a matrix is connected in this problem if the bipartite constructed in the following way is a connected graph.

Consider a bipartite graph formed by two disjoint subsets  $X$  and  $Y$ .  $X$  has  $m$  vertices, while  $Y$  has  $n$  vertices. The vertex  $X_i$  is connected to  $Y_j$  by an edge if  $M_{i,j}^E > 0$ . If the bipartite is connected, then we can recover  $M$  through the iterative method introduced above.

Lastly, note that  $M^E$  is connected with high probability if  $O(n \log n)$  number of its entries are nonzero (revealed).

## 16.3 SVD-based Matrix Completion Method

The algorithm is shown in the Table 16.1. Details of the steps are discussed in the following paragraphs.

### 16.3.1 Trimming

The operation of trimming is defined as follows.

**Trimming:** Set to zero all columns in  $M^E$  with more than  $\frac{2|E|}{m}$  nonzero entries; Set to zero all rows in  $M^E$  with more than  $\frac{2|E|}{n}$  nonzero entries.

Note that the average number of nonzero entries in each column (row) of  $M^E$  is  $\frac{|E|}{m}$  ( $\frac{|E|}{n}$ ). So we just discard all the columns and rows that have more than twice the average number of nonzero entries through trimming.

In fact, if the size of revealed entries  $|E| = O(nr \log n)$ , trimming is not necessary. However, if  $|E| = O(nr)$ , the performance without the step of trimming is poor. The reason is that when  $|E| = O(nr)$ , the maximum number of revealed (nonzero) entries in a row has the order of  $O(\frac{\log n}{\log \log n})$ , while the average number is constant. These over-represented rows (columns) will alter the spectrum of  $M^E$  artificially. Thus we drop these rows (columns) to avoid poor performance.

### 16.3.2 Projection

If ignore the scaling constant  $\frac{|E|}{mn}$ ,  $\mathbf{P}_r(M^E)$  is the orthogonal projection of  $M^E$  onto the set of rank- $r$  matrices. The scaling constant compensates the smaller average size of the nonzero entries of  $M^E$  with respect to  $M$ . Namely, the scaling constant is obtained from the following equations. For  $\forall \mathbf{x} \in \mathbb{R}^{m \times 1}, \mathbf{y} \in \mathbb{R}^{n \times 1}$ ,

$$\begin{aligned} \mathbb{E} [\mathbf{x}^T M^E \mathbf{y}] &= \mathbb{E} \left[ \sum_{i,j} x_i M_{i,j} y_j \mathbf{1}_{(i,j) \in E} \right] \\ &= \frac{|E|}{mn} \sum_{i,j} x_i M_{i,j} y_j \\ &= \frac{|E|}{mn} \mathbf{x}^T M \mathbf{y}. \end{aligned}$$

## 16.4 Proof of Effectiveness

The effectiveness of the proposed algorithm is provided in the following theorem.

**Theorem 16.1.** Assume  $M$  to be a rank  $r$  matrix of dimension  $m \times n$  which satisfies  $|M_{i,j}| \leq M_{\max}$  for all  $i, j$ . Then with probability larger than  $1 - 1/n^3$ ,

$$\frac{1}{n^2 M_{\max}^2} \left\| M - \mathbf{P}_r(\bar{M}^E) \right\|_F^2 \leq C \frac{rn}{|E|},$$

for some constant  $C$ .

Theorem 16.1 can be proved using the following lemmas.

**Lemma 16.2.** *There exists a constant  $C > 0$  such that, with probability larger than  $1 - 1/n^3$ ,*

$$\left| \frac{\sqrt{mn}}{|E|} \sigma_q - \Sigma_q \right| \leq C M_{\max} \sqrt{\frac{n}{|E|}},$$

where  $\sigma_q$  is the  $q^{\text{th}}$  largest singular value of  $\bar{M}^E$ , and  $\Sigma_q$  the  $q^{\text{th}}$  largest singular value of  $M/\sqrt{mn}$ .

Note that for  $q > r$ ,  $\Sigma_q = 0$ . Hence for  $q > r$ ,  $\sigma_q \leq C' M_{\max} \sqrt{|E|/n}$ .

**Lemma 16.3.** *There exists a constant  $C > 0$  such that, with probability larger than  $1 - 1/n^3$*

$$\left\| \frac{|E|}{mn} M - \bar{M}^E \right\|_2 \leq C M_{\max} \sqrt{\frac{|E|}{n}}.$$

We prove Theorem 16.1 as follows.

**Proof:** (Theorem 16.1) Note that for any matrix  $A$  of rank at most  $r$ ,  $\|A\|_F \leq \sqrt{2r} \|A\|_2$ . Hence it follows that

$$\|M - \mathbf{P}_r(\bar{M}^E)\|_F \leq \sqrt{2r} \|M - \mathbf{P}_r(\bar{M}^E)\|_2. \quad (16.1)$$

Next, by triangle inequality

$$\begin{aligned} \|M - \mathbf{P}_r(\bar{M}^E)\|_2 &\leq \|M - \frac{mn}{|E|} \bar{M}^E\|_2 + \left\| \frac{mn}{|E|} \bar{M}^E - \mathbf{P}_r(\bar{M}^E) \right\|_2 \\ &\leq \frac{mn}{|E|} C_1 M_{\max} \sqrt{\frac{|E|}{n}} + \left\| \frac{mn}{|E|} \bar{M}^E - \mathbf{P}_r(\bar{M}^E) \right\|_2 \end{aligned} \quad (16.2)$$

$$\begin{aligned} &\leq C_2 M_{\max} n^{3/2} \frac{1}{\sqrt{|E|}} + \frac{mn}{|E|} \sigma_{r+1} \\ &\leq C_2 M_{\max} n^{3/2} \frac{1}{\sqrt{|E|}} + C_3 \frac{mn}{|E|} M_{\max} \sqrt{|E|/n} \end{aligned} \quad (16.3)$$

$$\leq C_4 M_{\max} n^{3/2} \frac{1}{\sqrt{|E|}}, \quad (16.4)$$

where (16.2) is from Lemma 16.3, (16.2) is from Lemma 16.2, and  $C_1 - C_4$  are some positive constants. Finally, we complete the proof by substituting (16.4) for (16.1).  $\square$

Lastly, we provide the proof for Lemma 16.2 as follows.

**Proof:** (Lemma 16.2) Observe that

$$\sigma_q = \min_{H, \dim(H)=n-q+1} \max_{y \in H, \|y\|=1} \|\bar{M}^E y\| \quad (16.5)$$

$$= \max_{H, \dim(H)=q} \min_{y \in H, \|y\|=1} \|\bar{M}^E y\|. \quad (16.6)$$

By (16.5), if denote  $H^*$  as the orthogonal complement space of  $\mathbf{span}(v_1, \dots, v_{q-1})$ , it follows

$$\begin{aligned}
\sigma_q &\leq \max_{y \in H^*, \|y\|=1} \|\bar{M}^E y\| \\
&= \max_{y \in H^*, \|y\|=1} \left\| \bar{M}^E y - \frac{|E|}{mn} M + \frac{|E|}{mn} M \right\| \\
&\leq \frac{|E|}{\sqrt{mn}} \max_{y \in H^*, \|y\|=1} \left\| \frac{1}{\sqrt{mn}} M y \right\| + \max_{y \in H^*, \|y\|=1} \left\| \left( \frac{|E|}{mn} M - \bar{M}^E \right) y \right\| \\
&\leq \frac{|E|}{\sqrt{mn}} \Sigma_q + C M_{\max} \sqrt{\frac{|E|}{n}},
\end{aligned}$$

where the last step is from Lemma 16.3.

Similarly using (16.6) and letting  $H^* = \mathbf{span}(v_1, \dots, v_q)$ , the lower bound follows

$$\begin{aligned}
\sigma_q &\geq \min_{y \in H^*, \|y\|=1} \|\bar{M}^E y\| \\
&\geq \frac{|E|}{\sqrt{mn}} \min_{y \in H^*, \|y\|=1} \left\| \frac{M}{\sqrt{mn}} \right\| - \max_{y \in H^*, \|y\|=1} \left\| \left( \frac{|E|}{mn} M - \bar{M}^E \right) y \right\| \\
&\geq \frac{|E|}{\sqrt{mn}} \Sigma_q - C M_{\max} \sqrt{\frac{|E|}{n}}.
\end{aligned}$$

□

## Reference

- [1] R. H. Keshavan, A. Montanari, and O. Sewoong, “Matrix Completion From a Few Entries”, *IEEE Journal on Information Theory*, vol. 56, no. 6, pp. 2980–2998, Jun. 2010.