Interactions of Matrix Shape, Coherence, and Rank

Alexander Osypov, Adam Shaker, David Terry

The University of Texas at Dallas

Abstract

We explore the problem of matrix completion that often arises in the recommender systems. Using synthetic data and real datasets such as Amazon and MovieLens, we draw the relations among matrix coherence, rank, and regularization models. We observe that the most-squared matrix gives the highest accuracy for matrix completion via nuclear norm minimization, which can be justified by matrix coherence. Singular value decomposition (SVD) plays an important role in matrix completion. We find an approximated scheme of SVD is not only faster but also better than the standard SVD. Lastly, we discover that the transform L_1 model achieves the best recovery results among some existing approaches such as nuclear norm and L_p Schatten norm.

Introduction

Matrix completion refers to recovering a data matrix from partial samples of its entries. For example, in a recommender rating system such as Netflix, one seeks to predict the "rating" or "preference" a user would give to an item, based on a user's past behaviors and similar decisions made by other users. Since users often have similar preferences that produce linearly dependent ratings, the data matrix shall be low rank. Since directly minimizing the rank is NP-hard, one popular method minimizes the sum of singular values, referred to as the nuclear norm [1]. Some nonconvex models are listed as follows,

Nonconvex Regularization Models

- L_p for $0 < p \leq 1$: $\|\mathbf{x}\|_p^p = \Sigma_j |x_j|^p$;
- Transformed L_1 (TL1): $\Sigma_j \dfrac{(a+1)|x_j|}{a+|x_j|}$;
- Smoothly clipped absolute deviation: $\Sigma_j \Phi^{\text{SCAD}}_{\lambda,\gamma}(x_j)$ with

$$\Phi^{\text{SCAD}}_{\lambda,\gamma}(x_j) = \begin{cases} \lambda|x_j|, & |x_j| \leq \lambda; \\ \dfrac{2\gamma\lambda|x_j| - |x_j|^2 - \lambda^2}{2(\gamma-1)}, & \lambda < |x_j| \leq \gamma\lambda; \\ \dfrac{(\gamma+1)\lambda^2}{2}, & |x_j| > \gamma\lambda; \end{cases}$$

- Minimax concave penalty: $\Sigma_j \Phi^{\text{MCP}}_{\lambda,\gamma}(x_j)$ with

$$\Phi^{\text{MCP}}_{\lambda,\gamma}(x_j) = \begin{cases} \lambda|x_j| - \dfrac{|x_j|^2}{2\gamma}, & |x_j| \leq \gamma\lambda; \\ \dfrac{1}{2}\gamma\lambda^2, & |x_j| > \gamma\lambda; \end{cases}$$

- L_1-L_2 : $\|\mathbf{x}\|_1 - \|\mathbf{x}\|_2 = \Sigma_j |x_j| - \sqrt{\Sigma_j x_j^2}$.

References

[1] Candés and Recht. *Exact matrix completion via convex optimization* (2009).

[2] Ma, Goldfarb, and Chen. *Fixed point and Bregman iterative methods for matrix rank minimization* (2011)

[3] Harper and Konstan. *The MovieLens datasets: History and context* (2015).

Research Questions

We rely on a fixed-point continuation (FPC) algorithm [2] to examine the influence of matrix shape, coherence, rank, and various regularization functionals on the performance of matrix completion. The FPC algorithm is summarized in Algorithm 1.

Algorithm 1 Fixed Point Continuation (FPC) [2]

Input: a sampling operator $\mathcal{A}$ that contains ones (zeros) where data are recorded (missing) and recorded data $\mathbf{b}$.
Select $\mu_1 > \mu_2 > \ldots > \mu_L = \mu > 0$ and $\tau > 0$.
Initialize X by zero filling.
for $\mu = \mu_1, \mu_2, \ldots, \mu_L$ **do**
  while NOT converged, **do**
    Compute $Y = X - \tau\mathcal{A} * (\mathcal{A}(X) - b)$
    Compute SVD of Y, i.e., $Y = U\Sigma V^\top$
    Compute $X = U\text{diag}[\text{shrink}(\sigma, \tau\mu)]V^\top$

Note that the shrink operator in the last step is applied to every singular value, which is defined by

$$\text{shrink}(\sigma, \rho) = \max(\sigma - \rho, 0). \quad (1)$$

As each iteration of FPC involves SVD, we explore an approximated SVD to combine with FPC, labelled as FPCA, to speed up the computational efficiency. As the matrix coherence [1] is often used to determine to what degree it is possible to complete a matrix, we will further investigate whether coherence has any influence over the FPC/A algorithms. In summary, we aim at the following three research questions:

1. Under the same sampling ratio, what type of matrix can be recovered with higher probability?
2. Which algorithm is better, FPC or FPCA?
3. Which model is the best for low-rank matrix completion?

Coherence Definition/Theory

Let $U \in \mathbb{R}^{n\times r}$ contain orthonormal columns with $r < n$ and define $\mathbf{P}_U = UU^\top$ as its associated orthogonal projection matrix. Then the **coherence** of **U** is:

$$\mu(U) = \frac{n}{r}\max_{1\leq i\leq n} \|\mathbf{P}_U\mathbf{e}_i\|^2 = \frac{n}{r}\max_{1\leq i\leq n}\|U_i\|^2. \quad (2)$$

Under the following two assumptions,

A0 The coherence obeys $\max(\mu(U), \mu(V)) \leq \mu_0$ for $\mu_0 > 0$;

A1 The $m \times n$ matrix $\Sigma_{1\leq k\leq r}\mathbf{u}_k\mathbf{v}_k^*$ has a maximum magnitude bounded by $\mu_1\sqrt{r/(mn)}$ for $\mu_1 > 0$;

Theorem Let X be an $m \times n$ matrix of rank r obeying **A0** and **A1** and put $N = \max(m, n)$. Suppose we observe M entries of X with locations sampled uniformly at random. Then there exist constants C, c such that if

$$M \leq C\max(\mu_1^2, \sqrt{\mu_0}\mu_1, \mu_0 N^{1/4})Nr(\beta\log N) \quad (3)$$

for some $\beta > 2$, then the solution to the nuclear minimization is unique and equal to X with probability at least $1 - cN^{-\beta}$.

This Theorem asserts that if the coherence is low, few samples are required to recover X, and a large value indicates that the recovery of the matrix is unlikely.

Matrix Shape and Rank

We artificially generate a low-rank matrix by multiplying two randomly generated matrices together of dimensions $m \times r$ and $r \times n$, respectively, with $r \ll \min(m, n)$. The resulting ground-truth matrix X_g is $m \times n$, and has a rank that is at most r. We investigate three matrix shapes: $10 \times 160, 20 \times 80$, and 40×40 with three levels of rank by setting $r = 2, 5$, and 10. With randomly sampling 50% of the entries, we perform matrix completion by FPCA, thus getting a recovered matrix, denoted by X_{comp}. We define the *percentage error* of X_g and X_{comp} by,

$$E(X_g, X_{comp}) = \frac{\|X_{comp} - X_g\|_F}{\|X_g\|_F} \times 100\%.$$

We repeat each combination of matrix shape and rank over 100 trials, and plot the mean percentage errors, the variance of said errors, and the coherence of X_g in Figure 1.

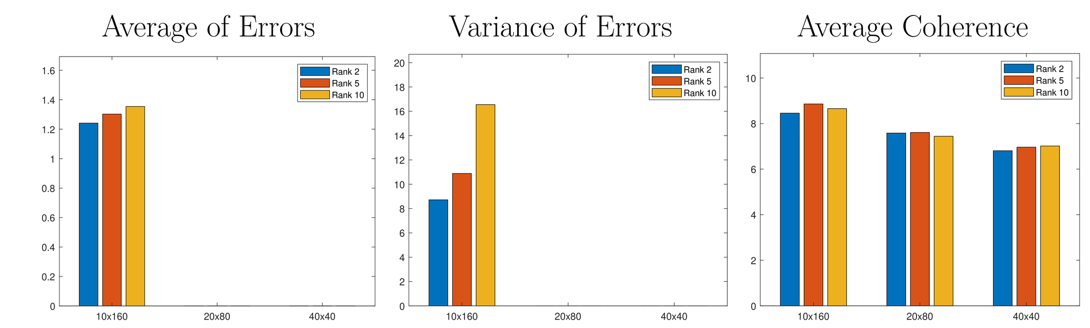

Figure 1:The influence of shape and rank on FPCA completion accuracy: the most-square ground-truth matrices have the highest recovery accuracy with the smallest coherence value.

FPC versus FPCA

Instead of random Gaussian matrices, we consider another type of matrix that is often used in sparse signal recovery, called over-sampled DCT matrix. It is defined by $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{m\times n}$ with

$$\mathbf{x}_j := \frac{1}{\sqrt{m}}\cos\left(\frac{2\pi j\mathbf{w}}{F}\right), \quad j = 1, \cdots, n, \quad (4)$$

where $\mathbf{w}$ is a random vector uniformly distributed in $[0, 1]^m$ and $F \in \mathbb{R}$ is a positive parameter. We multiply two over-sampled DCT matrices of size $m \times r$ and $r \times n$ together to generate the low-rank ground-truth matrix. By randomly taking out 50% of the entries, we attempt the matrix completion by FPC or FPCA. We repeat this experiment across F values of 1, 5, and 10, and the average percentage errors are shown in Figure 2.

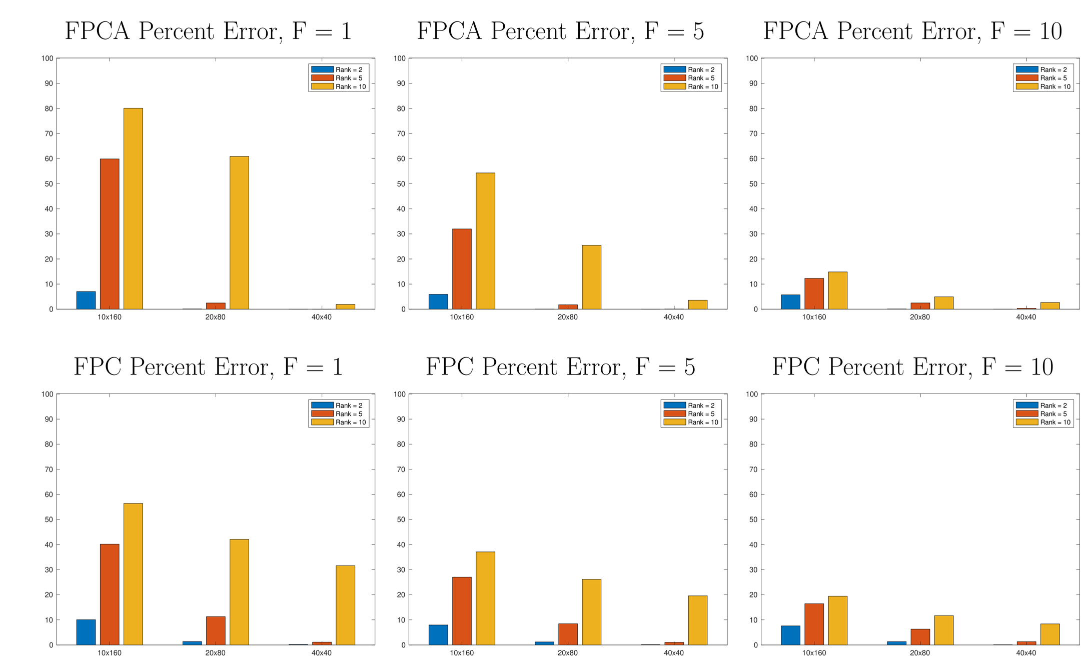

Figure 2:The comparison between FPC and FPCA of over-sampled DCT matrices across varying shapes, ranks, and F values.

Model Selection

We use two real datasets: Amazon Movies and MovieLens [3]. Both data matrices originally have 1000 rows and 1000 columns. We can progressively eliminate rows and columns in order to make the data set more dense. We summarize the data information in Table 1.

Data Set	Dimension	Density
MovieLens	1000 × 1000	34.80%
MovieLens	500 × 500	47.70%
MovieLens	308 × 366	80.14%
MovieLens	68 × 247	93.23%
Amazon Movie	1000 × 1000	3.71%
Amazon Movie	500 × 500	3.71%
Amazon Movie	38 × 91	41.90%

Table 1:Real data sets.

We can minimize any aforementioned non-convex regularization by replacing the shrinkage operator (1) by its own proximal operator. By doing so, we can examine these nonconvex regularizations for low-rank matrix completion. To guarantee the low-rankness of real data, we manually process the data matrix by keeping the largest r singular values for $r = 5, 10$, and 20.

Amazon Movies 38 × 91 | MovieLens 68 × 247 | MovieLens 303 × 366

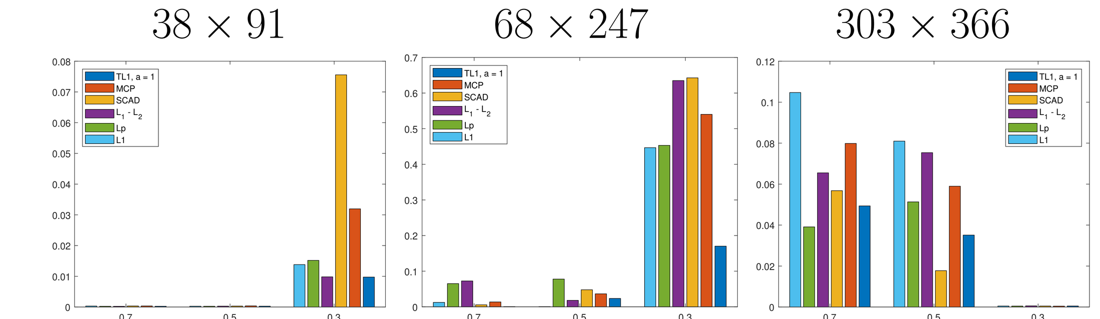

Figure 3:Comparison among various regularizations for matrix completion with ground-truth matrix rank of 20 in terms of relative errors.

Rank = 5 | Rank = 10 | Rank = 20

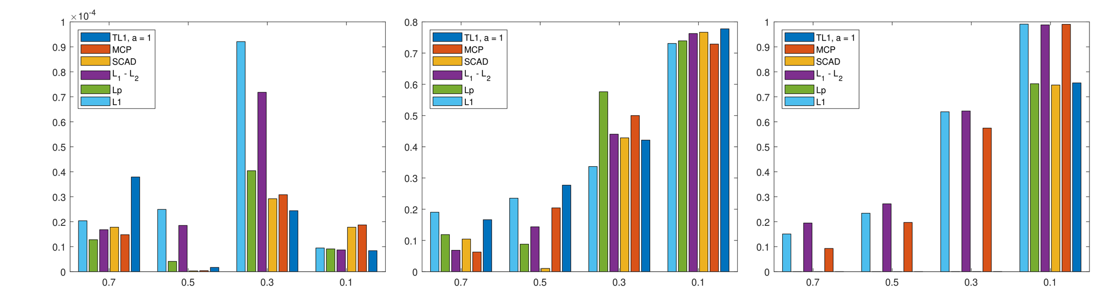

Figure 4:Comparison among various regularizations for matrix completion of 500 × 500 MovieLens with prescribed rank of 5, 10, and 20.

Conclusions

- Matrix completion prefers matrices that are low rank and square. This is justified by a decreased coherence for matrices with these desirable properties.
- FPCA *generally* achieves better recovery than FPC. But for matrices with undesirable properties, FPC with an exact SVD is better than FPCA.
- TL1 outperforms the other methods when the underlying matrices are skinny or tall, while TL1, SCAD, and L_p all perform good at a higher rank.