Movie Recommendation: Aggregation of Collaborative Filtering and Low-rank Matrix Recovery

Yuan Chen*, Yanjun Han†, Yixin Wang‡

* Stanford Applied Physics, † Stanford EE, ‡ Stanford CS

Email: {ychen12, yjhan, wyixin}@stanford.edu

Matrix Completion Problem

Target (*Netflix Problem*): recover a large matrix from a small subset of its entries

Mathematical formulation: the original matrix $M \in \mathbb{R}^{m\times n}$ is corrupted by some (deterministic or random) noise Z and is partially revealed:

$$N = \mathcal{P}_\Lambda(M+Z).$$

A general model setup [1]:

- M: low-rank matrix
- Z: another low-dimensional structure (e.g., sparsity) and/or Gaussian noise
- $\mathcal{P}_\Lambda$: projection or a general linear operator
- $\Lambda \subset [m]\times[n]$: randomly revealed positions

Goal: reconstruct some $\hat{M}$ to minimize the root mean squared error (RMSE)

$$R(\hat{M}, M) = \sqrt{\mathbb{E}\left[\frac{1}{mn}\|\hat{M}-M\|_F^2\right]}.$$

Literature Review

Collaborative Filtering: similar users tend to give similar ratings to the same item

- Compute the similarity scores between users/items, and use weighted average;
- Widely used in practical recommender systems, where similarity scores can be trained.

Optimization-based methods: there are two categories:

Convex relaxation: under mild sample complexity and coherence assumptions [2],

$$\min_M \frac{1}{2}\|\mathcal{P}_\Lambda(M)-N\|_F^2 + \lambda\cdot\|M\|_* \Longleftrightarrow \min_M \frac{1}{2}\|\mathcal{P}_\Lambda(M)-N\|_F^2 + \lambda'\cdot \mathrm{rank}(M).$$

Gradient descent: the matrix completion problem shares a "nice" loss surface [3]:

1. simple gradient descent enjoys a good performance guarantee
2. requires good initialization and variance reduction: throw away some information!

Implemented Methods

Collaborative Filtering: based on Pearson similarity and trained similarity scores

Accelerated Gradient [4]: borrow Nesterov's idea and use proximal regularization:

$$\min_M \frac{1}{2\alpha^{(t)}}\left\|M-\left(M^{(t)}-\alpha^{(t)}\nabla f(M^{(t)})\right)\right\|_F^2 + \lambda\|M\|_*$$

admits a closed-form solution and can be accelerated.

Manifold Gradient [5]: use simple SVD after trimming as initialization, then minimize $F(U,V) = \min_{S\in\mathbb{R}^{r\times r}} \frac{1}{2}\|\mathcal{P}_\Lambda(N-USV^T)\|_F^2$ on Grassmann manifold $\mathrm{M} = \mathrm{G}(m,r)\times\mathrm{G}(n,r)$.

Performance on Simulated Data

Parameter for simulation:

- M: 600×1000 random matrix with rank $r=5$, adding random user-bias and item-bias;
- p: reveal probability of each entry;
- Noise: no noise or Gaussian noise with $\sigma = 0.1$.

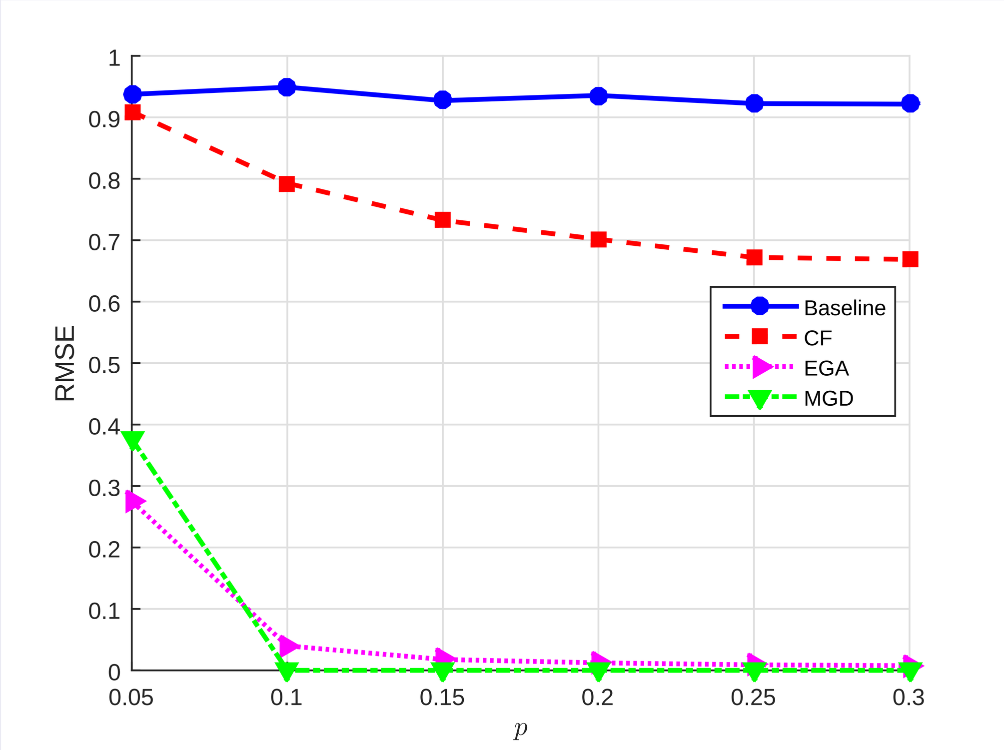


Figure 1: noise-free case.

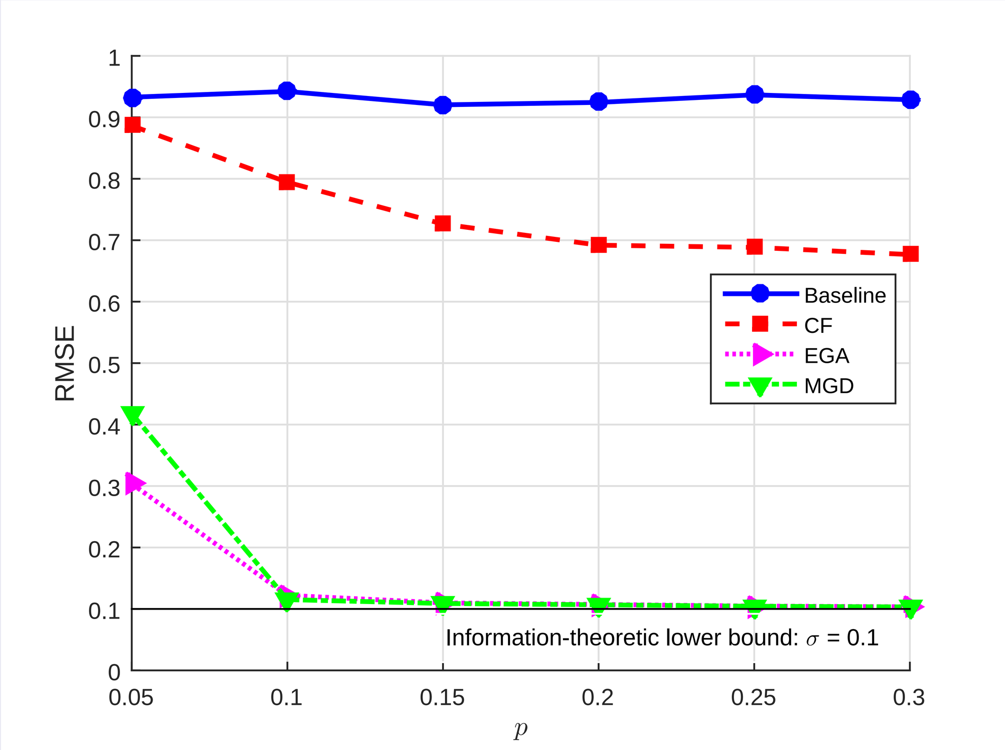
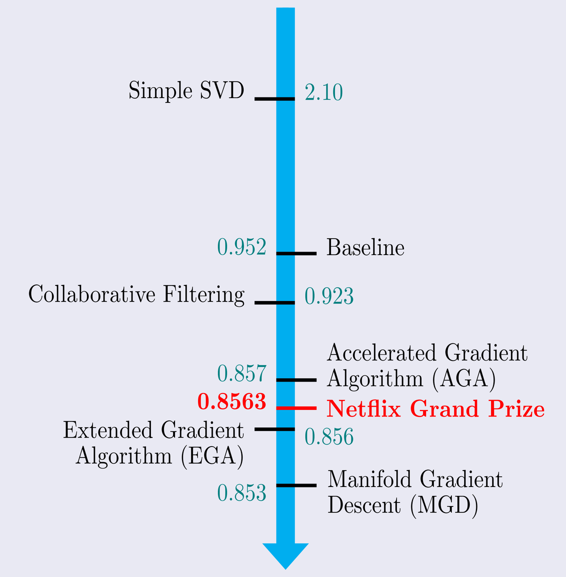
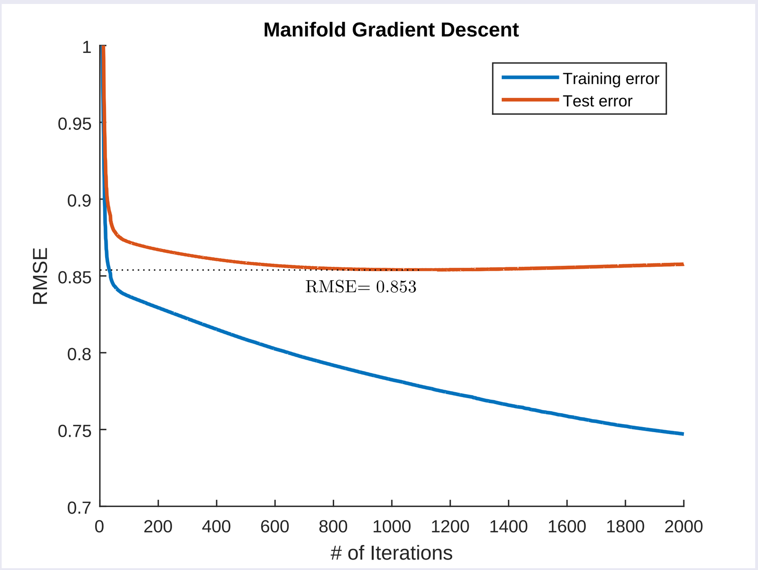


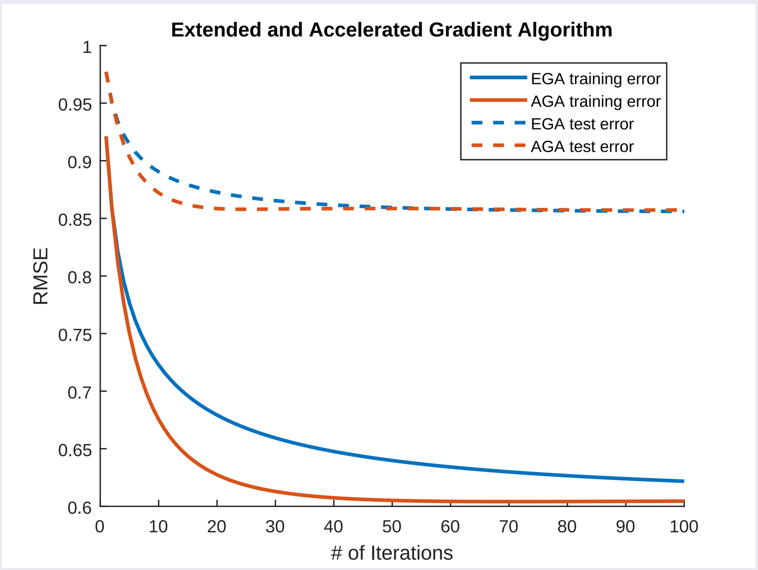
Figure 2: noisy case.

Performance on Real Data

MovieLens dataset: 138,493 users and 27,278 movies, 5-star ratings (fill rate 5%). Randomly select 90% as training, and 10% as testing.







References

[1] A. Agarwal, S. Negahban, and M. J. Wainwright, *Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions*, The Annals of Statistics, pp 1171-1197, 2015.

[2] E. J. Candès and T. Tao, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Transactions on Information Theory, Vol. 56, No. 5, pp 2053-2080, 2010.

[3] R. Ge, J. D. Lee, and T. Ma, *Matrix completion has no spurious local minimum*, arXiv preprint arXiv:1605.07272.

[4] S. Ji and J. Ye, *An accelerated gradient method for trace norm minimization*, ICML 2009.

[5] R. H. Keshavan, S. Oh and A. Montanari, *Matrix completion from a few entries*, IEEE Transactions on Information Theory, Vol. 56, No. 6, pp 2980-2998, 2010.