Robust Matrix Completion via Alternating Projection

Hing Cheung So

http://www.ee.cityu.edu.hk/~hcso

Department of Electronic Engineering, City University of Hong Kong

X. Jiang, Z. Zhong, X. Liu and H. C. So, "Robust matrix completion via alternating projection," *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 579-583, May 2017

Outline

- Introduction
- Matrix Completion as a Feasibility Problem
- Alternating Projection Algorithm
- Numerical Examples
- Concluding Remarks
- List of References

Introduction

What is Matrix Completion?

The aim is to recover a low-rank matrix given only a subset of its possibly noisy entries, e.g.,

$$\begin{pmatrix} 1 & ? & ? & 4 & ? \\ ? & 2 & 5 & ? & ? \\ ? & ? & 4 & 5 & ? \\ 5 & ? & ? & ? & 4 \end{pmatrix}$$

Denote the known entries of an incomplete matrix $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$ as $\boldsymbol{M}_\Omega$:

$$[\boldsymbol{M}_\Omega]_{i,j} = \begin{cases} \boldsymbol{M}_{i,j}, & \text{if } (i,j) \in \Omega \\ 0, & \text{otherwise.} \end{cases}$$

where Ω is a subset of the complete set of entries $[n_1] \times [n_2]$, with $[n]$ being the list $\{1, \cdots, n\}$ while the unknown entries are assumed zero.

Basically, matrix completion is to find a matrix $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$, which is an estimate of $\boldsymbol{M}$, given $\boldsymbol{M}_\Omega$ with the use of low-rank information of $\boldsymbol{M}$, which can be mathematically formulated as:

$$\min_{\boldsymbol{X}} \ \text{rank}(\boldsymbol{X}), \quad \text{s.t. } \boldsymbol{X}_\Omega = \boldsymbol{M}_\Omega.$$

Why Matrix Completion is Important?

It is a core problem in many applications including:

- Collaborative Filtering
- Image Inpainting and Restoration
- System Identification
- Node Localization
- Genotype Imputation

It is because many real-world signals can be approximated by a matrix whose rank is $r \ll \max\{n_1, n_2\}$.

Netflix Prize, whose goal was to accurately predict user preferences with the use of a database of over 100 million movie ratings made by 480,189 users in 17,770 films, which corresponds to the task of completing a matrix with around 99% missing entries.

Alice	1			4		
Bob		2	5			
Carol			4	5		
Dave	5				4	
⋮						

How to Recover an Incomplete Matrix?

Directly solving the noise-free version:

$$\min_{\boldsymbol{X}} \ \text{rank}(\boldsymbol{X}), \quad \text{s.t. } \boldsymbol{X}_\Omega = \boldsymbol{M}_\Omega$$

or noisy version:

$$\min_{\boldsymbol{X}} \ \text{rank}(\boldsymbol{X}), \quad \text{s.t. } \|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_F^2 \leq \epsilon_F$$

is difficult because the rank minimization problem is NP-hard.

A popular and practical solution is to replace the nonconvex rank by convex nuclear norm:

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_*, \quad \text{s.t. } \boldsymbol{X}_\Omega = \boldsymbol{M}_\Omega$$

or

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_*, \quad \text{s.t. } \|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_F^2 \leq \epsilon_2$$

where $\|\boldsymbol{X}\|_*$ equals the sum of singular values of $\boldsymbol{X}$. However, complexity of nuclear norm minimization is still high and this approach is not robust when $\boldsymbol{M}_\Omega$ contains outliers.

Another popular direction which is computationally simple is to apply low-rank matrix factorization:

$$\min_{\boldsymbol{U},\boldsymbol{V}} \ f_2(\boldsymbol{U},\boldsymbol{V}) := \|(\boldsymbol{U}\boldsymbol{V})_\Omega - \boldsymbol{M}_\Omega\|_F^2$$

where $\boldsymbol{U} \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{V} \in \mathbb{R}^{r \times n_2}$. However, generalization of the Frobenius norm to ℓ_p-norm for handling impulsive measurements is difficult.

Matrix Completion as a Feasibility Problem

We formulate matrix completion with noise-free entries as:

$$\text{find } \boldsymbol{X}, \quad \text{s.t. } \operatorname{rank}(\boldsymbol{X}) \leq r, \quad \boldsymbol{X}_\Omega = \boldsymbol{M}_\Omega$$

where an estimate or true value of r is needed.

It is called a feasibility problem because this optimization formulation has no objective function, but two constraints:

- Low-rank constraint: $\operatorname{rank}(\boldsymbol{X}) \leq r$
- Fidelity constraint: $\boldsymbol{X}_\Omega = \boldsymbol{M}_\Omega$

With Gaussian noise, the fidelity constraint is modified as:

$$\|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_F^2 \leq \epsilon_2.$$

To achieve robustness, the feasibility problem is:

$$\text{find } \boldsymbol{X}, \quad \text{s.t. } \operatorname{rank}(\boldsymbol{X}) \leq r, \quad \|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_p^p \leq \epsilon_p.$$

The rank constraint set is:

$$\mathcal{S}_r := \{\boldsymbol{X} | \operatorname{rank}(\boldsymbol{X}) \leq r\}$$

and the fidelity constraint set is:

$$\mathcal{S}_p := \left\{\boldsymbol{X} | \; \|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_p^p \leq \epsilon_p\right\}, \quad 0 < p \leq 2$$

where

$$\|\boldsymbol{X}_\Omega\|_p = \left(\sum_{(i,j)\in\Omega} |[\boldsymbol{X}]_{i,j}|^p \right)^{1/p}$$

is element-wise ℓ_p-norm which is robust to outliers if $p < 2$.

We may rewrite the robust feasibility problem as:

$$\text{find } \boldsymbol{X} \in \mathcal{S}_r \cap \mathcal{S}_p.$$

Remarks:

- $\mathcal{S}_p := \left\{\boldsymbol{X} | \ \|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_p^p \leq \epsilon_p\right\}$ is a generalization as $\epsilon_p = 0$ reduces to noise-free version while $p = 2$ reduces to conventional scenario of handling Gaussian noise.

- We restrict our study for $p = 1$ and $p = 2$ since their projections onto $\mathcal{S}_p$ have closed-form expressions and are not difficult to compute.

- $p < 1$ requires computing projection onto a nonconvex and nonsmooth ℓ_p-ball, which is difficult to compute.

Alternating Projection Algorithm

Define the projection of a point $\boldsymbol{Z} \notin \mathcal{S}$ onto any constraint set $\mathcal{S}$, denoted as $\Pi_{\mathcal{S}}(\boldsymbol{Z})$:

$$\Pi_{\mathcal{S}}(\boldsymbol{Z}) := \arg\min_{\boldsymbol{X} \in \mathcal{S}} \|\boldsymbol{X} - \boldsymbol{Z}\|_F^2.$$

That is, projection onto rank constraint set is:

$$\boldsymbol{X} = \Pi_{\mathcal{S}_r}(\boldsymbol{Z})$$

and projection onto fidelity constraint set is:

$$\boldsymbol{X} = \Pi_{\mathcal{S}_p}(\boldsymbol{Z})$$

High-Level Algorithm

The proposed alternating projection algorithm (APA) is outlined in Algorithm 1:

Algorithm 1 Alternating Projection for Matrix Completion

Input: $\boldsymbol{M}_\Omega$, Ω, and $\epsilon_p \geq 0$

Initialize: $\boldsymbol{X}^0 = \boldsymbol{M}_\Omega$

for $k = 0, 1, 2, \cdots$ **do**

$\boldsymbol{Y}^k = \Pi_{\mathcal{S}_r}\left(\boldsymbol{X}^k\right)$

$\boldsymbol{X}^{k+1} = \Pi_{\mathcal{S}_p}\left(\boldsymbol{Y}^k\right)$

Stop if a termination condition is satisfied.

end for

Output: $\boldsymbol{X}^{k+1}$

According to Eckart-Young theorem, the projection of $\boldsymbol{Z} \notin \mathcal{S}_r$ onto $\mathcal{S}_r$ can be computed via truncated singular value decomposition (SVD) of $\boldsymbol{Z}$:

$$\Pi_{\mathcal{S}_r}(\boldsymbol{Z}) = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

where $\{\sigma_i\}_{i=1}^r$, $\{\boldsymbol{u}_i\}_{i=1}^r \in \mathbb{R}^{n_1}$, and $\{\boldsymbol{v}_i\}_{i=1}^r \in \mathbb{R}^{n_2}$ are the r largest singular values and the corresponding left and right singular vectors of $\boldsymbol{Z}$, respectively.

Assuming $n_2 \leq n_1$, the complexity is $\mathcal{O}(n_1 n_2 r)$ which is much smaller than that of full SVD of $\mathcal{O}(n_1 n_2^2 + n_2^3)$ required in the nuclear norm minimization based methods, particularly when $r \ll \max\{n_1, n_2\}$.

Noting that projection onto $\mathcal{S}_p$ only affects the entries indexed by Ω, we first define $\boldsymbol{m}_\Omega \in \mathbb{R}^{|\Omega|}$, which is a vector that contains the observed entries of $\boldsymbol{M}$, e.g., if

$$\boldsymbol{M}_\Omega = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 0 & 0 \\ 0 & 6 & 0 \end{bmatrix}$$

then

$$\boldsymbol{m}_\Omega = [1\ 2\ 6\ 3]^T$$

Hence $\mathcal{S}_p := \left\{\boldsymbol{X} \middle| \ \|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_p^p \le \epsilon_p\right\}$ has the equivalent vector form:

$$\mathcal{B}_p := \left\{\boldsymbol{x}_\Omega \in \mathbb{R}^{|\Omega|} \middle| \ \|\boldsymbol{x}_\Omega - \boldsymbol{m}_\Omega\|_p^p \le \epsilon_p\right\}$$

which is an ℓ_p-ball with the observed vector $\boldsymbol{m}_\Omega$ being ball center.

We consider the following three cases with different values of p and ϵ_p:

- For $\epsilon_p = 0$, $\mathcal{B}_p$ reduces to equality constraint of $\boldsymbol{x}_\Omega = \boldsymbol{m}_\Omega$. For any vector $\boldsymbol{z} \in \mathbb{R}^{|\Omega|}$, the projection is simply calculated as $\Pi_{\mathcal{B}_p}(\boldsymbol{z}) = \boldsymbol{m}_\Omega$.

- For $p = 2$ and $\epsilon_2 > 0$, $\mathcal{B}_2$ is the conventional ℓ_2-ball in the Euclidean space. For any vector $\boldsymbol{z} \notin \mathcal{B}_2$, it is not difficult to derive the closed-form expression of the projection onto $\mathcal{B}_2$ as

$$\Pi_{\mathcal{B}_2}(\boldsymbol{z}) = \boldsymbol{m}_\Omega + \frac{\sqrt{\epsilon_2}(\boldsymbol{z} - \boldsymbol{m}_\Omega)}{\|\boldsymbol{z} - \boldsymbol{m}_\Omega\|_2}.$$

  With a proper value of ϵ_2, the robustness to Gaussian noise is enhanced.

- For $p=1$ and $\epsilon_1>0$, $\mathcal{B}_1$ is an ℓ_1-ball. For any vector $\boldsymbol{z}\notin\mathcal{B}_1$, the projection onto $\mathcal{B}_1$ is the solution of:

$$\min_{\boldsymbol{x}} \ \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2, \quad \text{s.t. } \|\boldsymbol{x}-\boldsymbol{m}_\Omega\|_1 \leq \epsilon_1.$$

Using the Lagrange multiplier method, we obtain:

$$[\Pi_{\mathcal{B}_1}(\boldsymbol{z})]_i = \operatorname{sgn}([\boldsymbol{z}-\boldsymbol{m}_\Omega]_i)\max(|[\boldsymbol{z}-\boldsymbol{m}_\Omega]_i|-\lambda^\star,0)$$

where $i=1,\cdots,|\Omega|$, and $\lambda^\star$ is the unique root of the nonlinear equation:

$$\sum_{i=1}^{|\Omega|}\max(|[\boldsymbol{z}-\boldsymbol{m}_\Omega]_i|-\lambda,0)=\epsilon_1, \quad \lambda^\star\in(0,\|\boldsymbol{z}-\boldsymbol{m}_\Omega\|_\infty)$$

The computational complexity of projection onto ℓ_1-ball is $\mathcal{O}(|\Omega|)$, which is much lower than that of projection onto $\mathcal{S}_r$.

Note that $1 < p < 2$ also involves the projection onto a convex ℓ_p-ball, which is not difficult to solve but requires an iterative procedure.

As $p = 1$ is more robust than $1 < p < 2$ in the presence of outliers, the latter case will not be considered.

Remarks:

- For the noise-free case, it is clear that $\epsilon_p = 0$ is the optimal value.

- Roughly speaking, larger noise requires a larger ϵ_p for $p \in [1,2]$. If we know the probability density function (PDF) of the noise, proper value of ϵ_p can be calculated.

- Note that the nuclear norm regularized problem:

$$\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_F^2 + \tau\|\boldsymbol{X}\|_*$$

also faces the issue of selecting the user-defined τ.

- Note also that our APA is different from the iterative hard thresholding (IHT) and its variants although they all use a rank-r projection.

More precisely, IHT solves the rank constrained Frobenius norm minimization:

$$\min_{\boldsymbol{X}} f(\boldsymbol{X}) := \frac{1}{2}\|\boldsymbol{X}_\Omega - \boldsymbol{M}_\Omega\|_F^2, \quad \text{s.t. } \operatorname{rank}(\boldsymbol{X}) \le r$$

using gradient projection with update step being

$$\boldsymbol{X}^{k+1} = \Pi_{\mathcal{S}_r}\left(\boldsymbol{X}^k - \mu \nabla f(\boldsymbol{X}^k)\right), \quad \mu > 0$$

where determining μ with a line search scheme requires computing projection $\Pi_{\mathcal{S}_r}(\cdot)$ for several times. Hence its computational cost is higher than APA per iteration.

- We prove that if initial point is close enough to $\mathcal{S}_r \cap \mathcal{S}_p$, then APA locally converges to $\boldsymbol{X} \in \mathcal{S}_r \cap \mathcal{S}_p$ at a linear rate.

Numerical Examples

Noise-free $\boldsymbol{M} \in \mathbb{R}^{n_1 \times n_2}$ of rank r is generated by the product of $\boldsymbol{M}_1 \in \mathbb{R}^{n_1 \times r}$ and $\boldsymbol{M}_2 \in \mathbb{R}^{r \times n_2}$ whose entries satisfy standard Gaussian distribution, where $n_1 = 150$, $n_2 = 300$, and $r = 10$.

45% of the entries of $\boldsymbol{M}$ are randomly selected as the known observations.

Impulsive noise is modelled as two-term Gaussian mixture model (GMM) whose PDF is

$$p_v(v) = \sum_{i=1}^{2} \frac{c_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{v^2}{2\sigma_i^2}\right), \quad c_1 = 0.9,\ c_2 = 0.1,\ \sigma_2^2 = 100\sigma_1^2$$

Signal-to-noise ratio (SNR) is defined as:

$$\frac{\|\boldsymbol{M}_\Omega\|_F^2}{|\Omega|\sigma_v^2}, \quad \sigma_v^2 = c_1\sigma_1^2 + c_2\sigma_2^2$$

Normalized root mean square error (RMSE) is defined as:

$$\mathrm{RMSE}(\boldsymbol{X}) = \sqrt{\mathrm{E}\left\{\frac{\|\boldsymbol{X} - \boldsymbol{M}\|_F^2}{\|\boldsymbol{M}\|_F^2}\right\}}$$

which is calculated based on 200 independent runs.

Comparison with singular value thresholding (SVT) and IHT with $\mu = 0.05$, 0.025 and 0.1, are included.

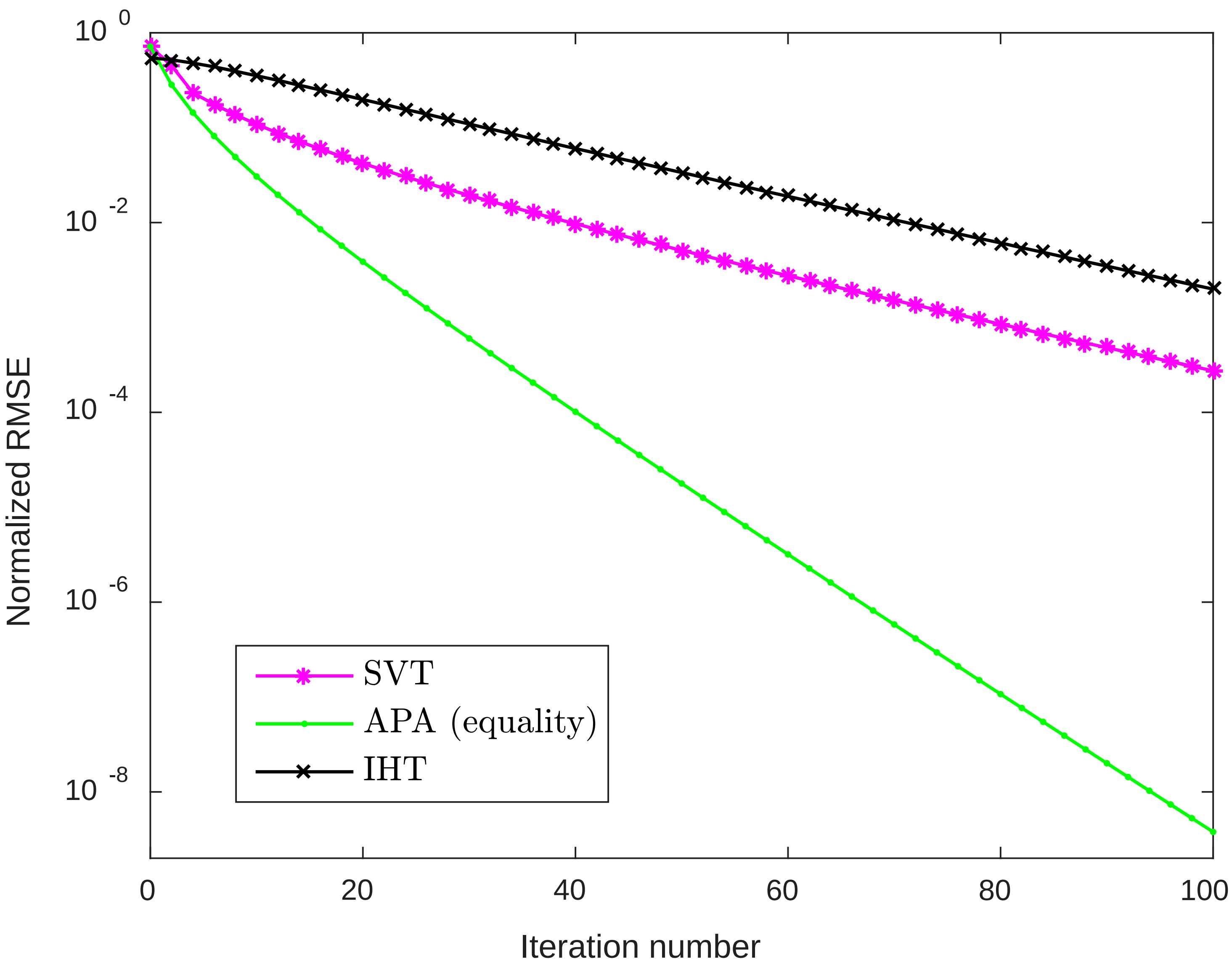

Figure 1: RMSE versus iteration number in noise-free case

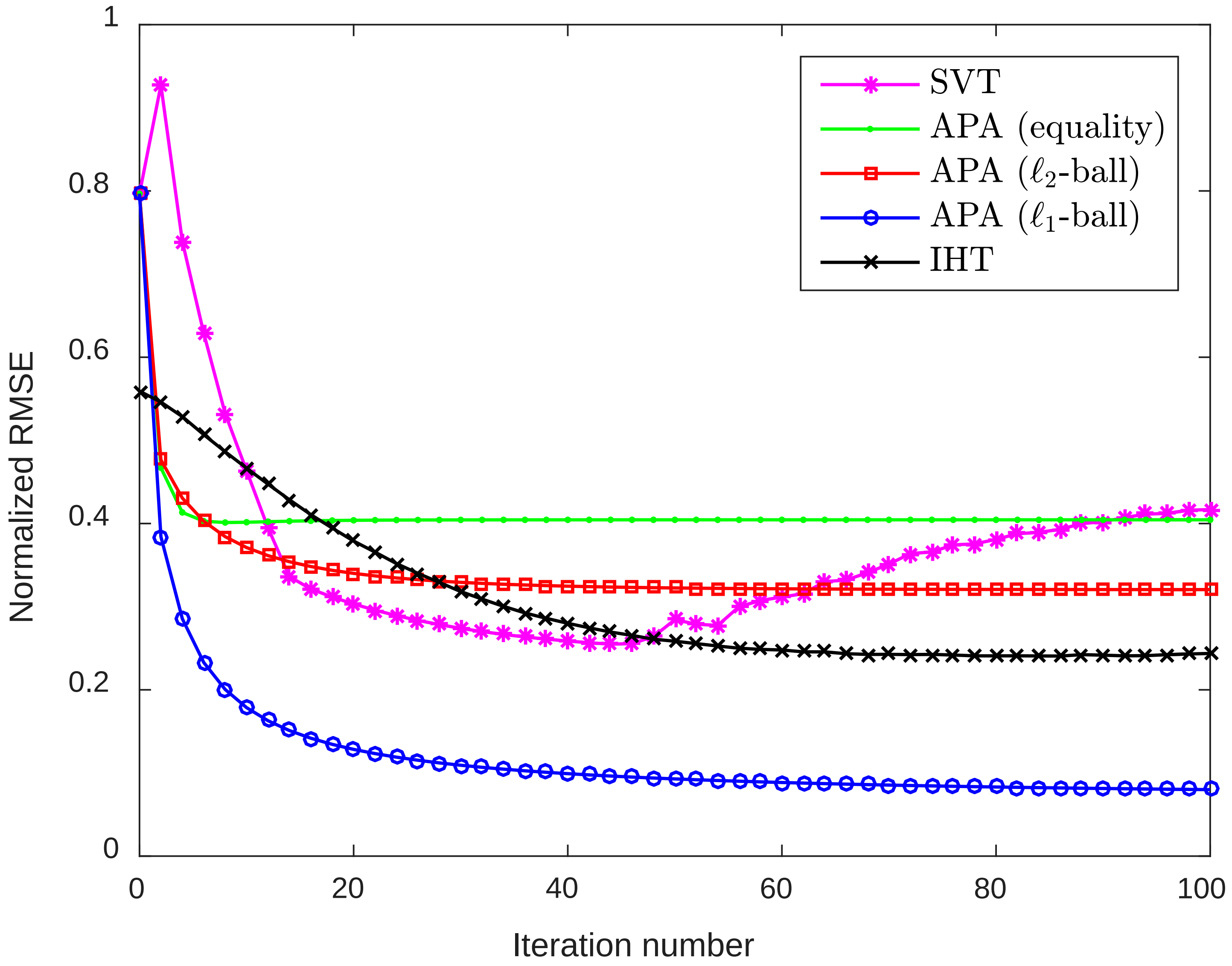

Figure 2: RMSE versus iteration number at SNR=6dB

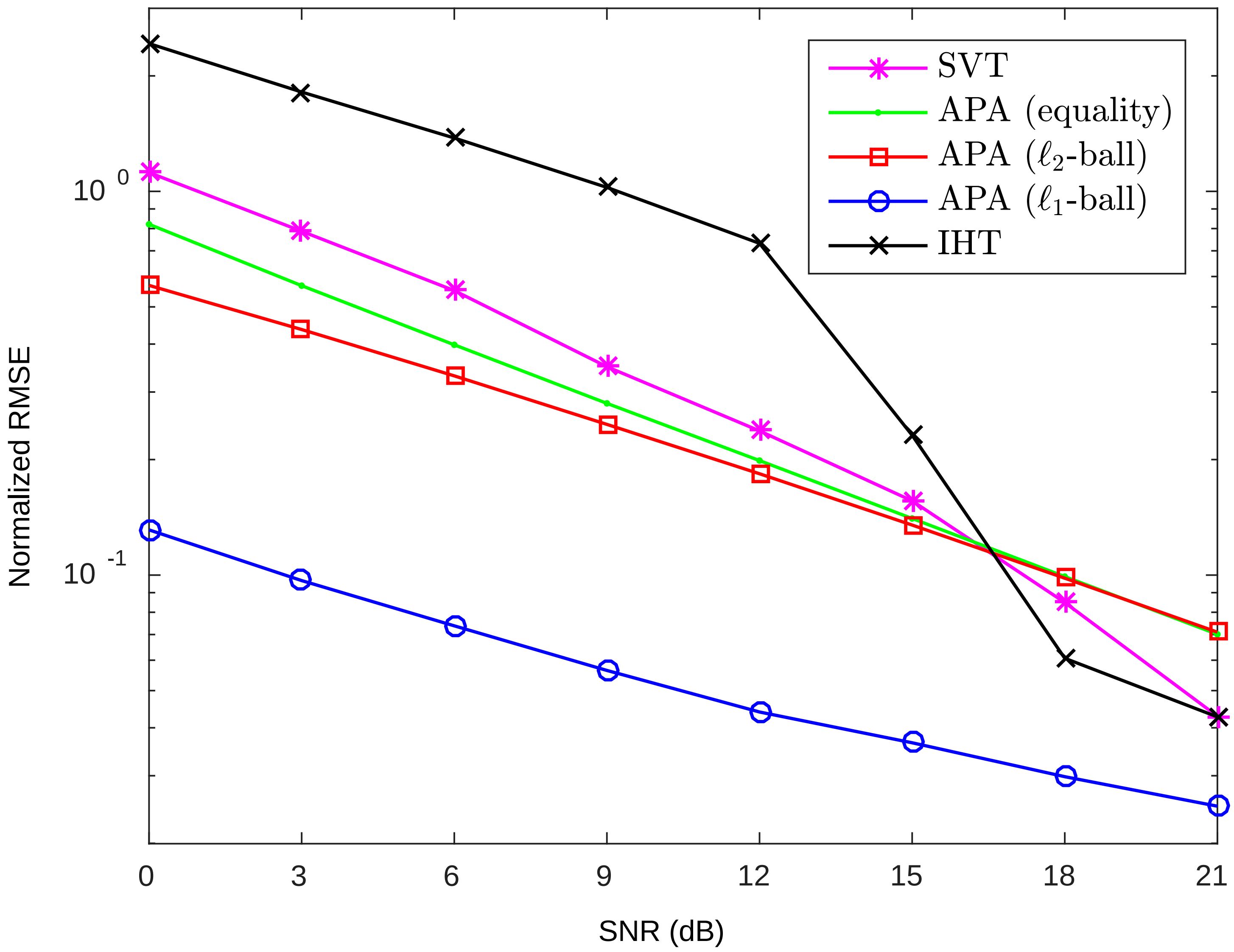

Figure 3: RMSE versus SNR in GMM noise

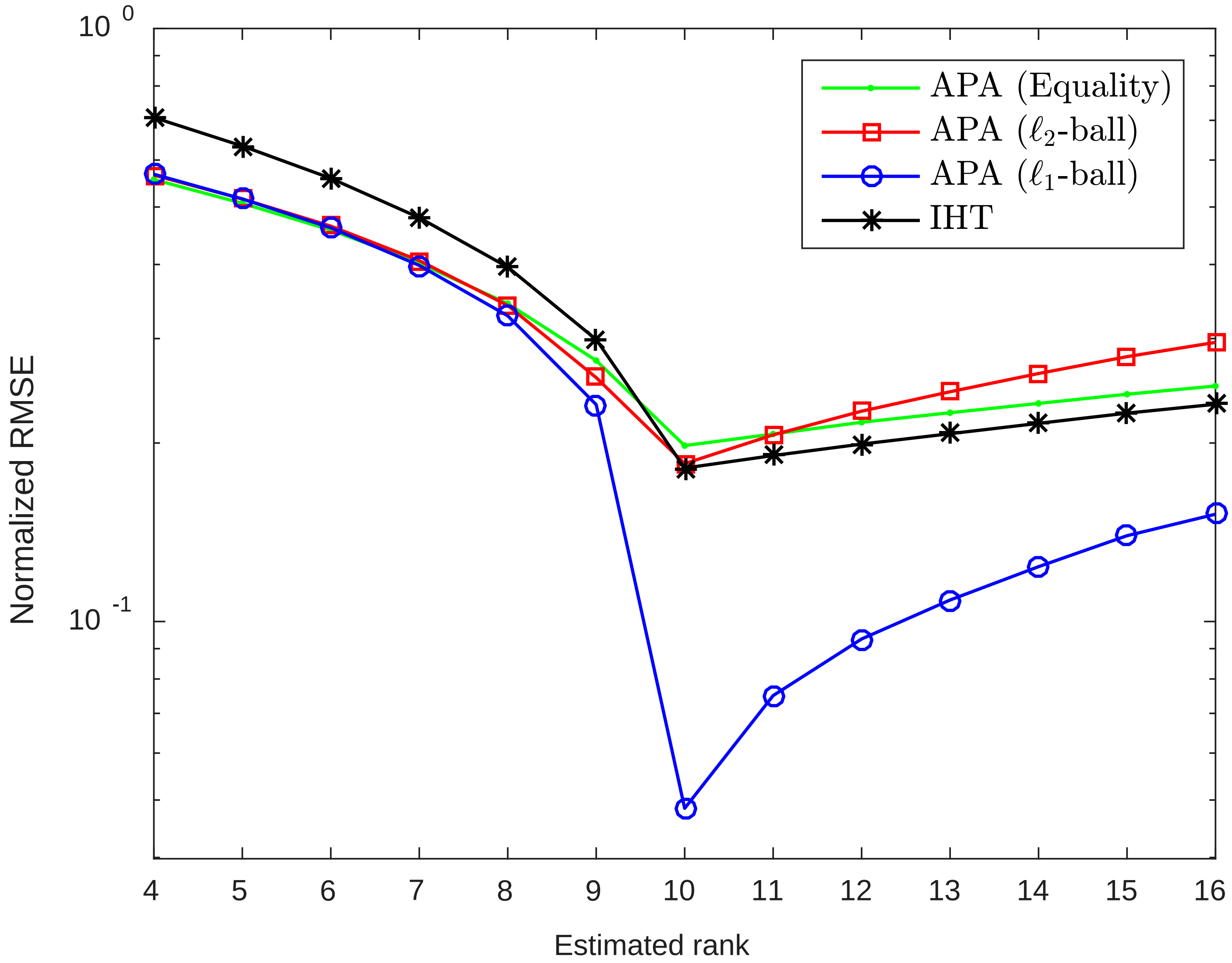

Figure 4: RMSE versus estimated rank at SNR=12dB

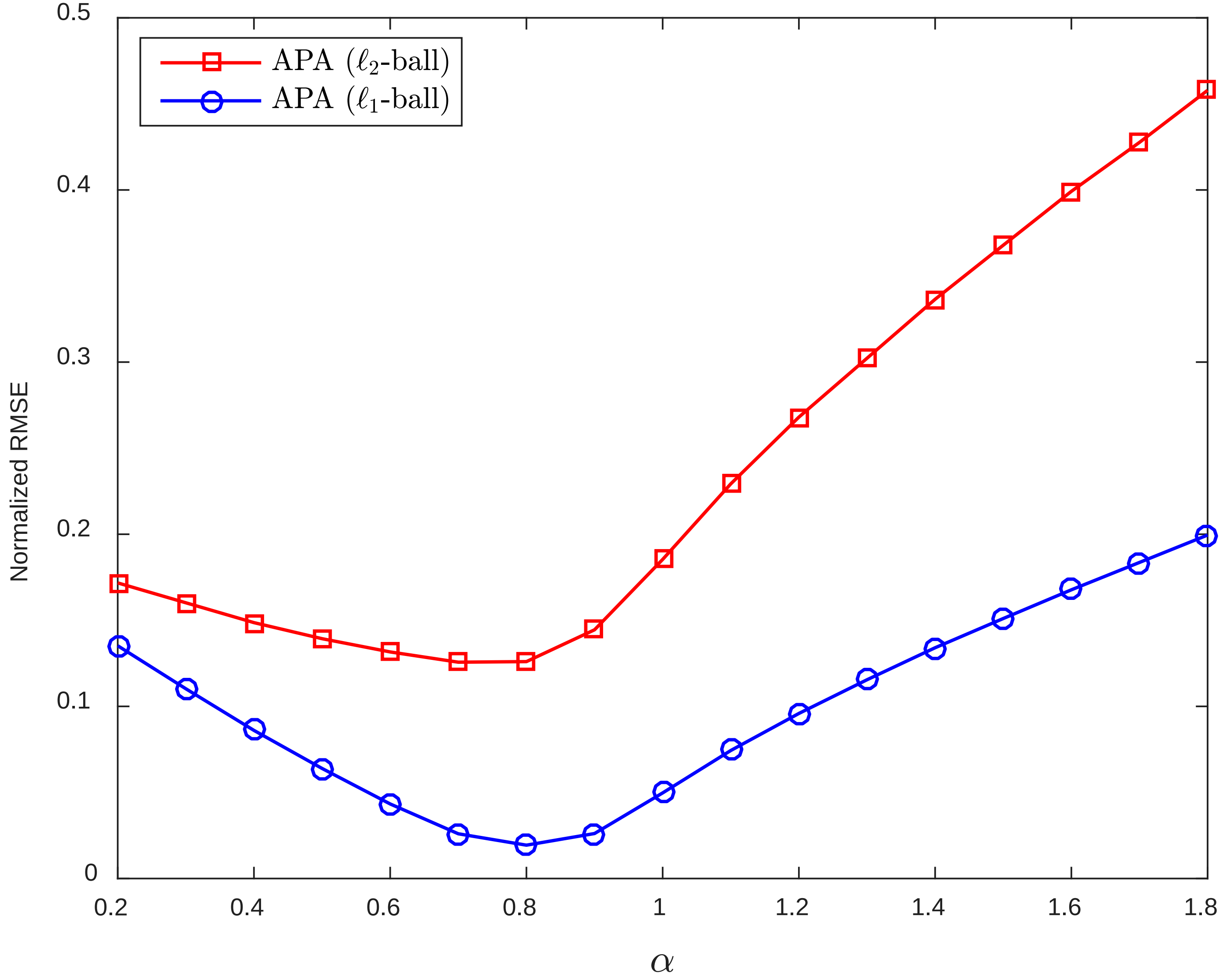

Figure 5: RMSE versus α with $\sqrt{\epsilon_p} = \alpha \|\boldsymbol{v}\|_p$ at SNR=12dB

Concluding Remarks

- The key idea is to formulate matrix completion as a feasibility problem, where a common point of the low-rank constraint set and fidelity constraint set is found by alternating projection.

- The fidelity constraint set is modelled as an ℓ_p-ball, where $p=1$ or $p=2$, which results in closed-form projection.

- The APA achieves robustness against Gaussian noise and outliers, with $p=2$ and $p=1$, respectively.

- The APA is conceptually simpler and computationally more efficient than the popular methods including the SVT and IHT.

List of References

[1] http://www.netflixprize.com/

[2] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925-936, Jun. 2010.

[3] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 4, pp. 608-622, Jun. 2016.

[4] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053-2080, May 2010.

[5] B. Recht, M. Fazel and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via

nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471-501, 2010.
[6] J.-F. Cai, E. J. Candès and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Opt.*, vol. 20, no. 4, pp. 1956-1982, 2010.
[7] P. Jain, R. Meka and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," in *Adv. Neural Inf. Process. Syst. (NIPS)*, pp. 937-945, 2010.
[8] J. Duchi, S. Shalev-Shwartz, Y. Singer and T. Chandra, "Efficient projections onto the L1-ball for learning in high dimensions," in *Proc. 25th Int. Conf. Machine Learning (ICML)*, pp.272-279, 2008.
[9] L. Condat, "Fast projection onto the simplex and the L1-ball," *Math. Program. Ser. A*, vol. 158, no. 1, pp. 575-585, Jul. 2016.

[10] A. S. Lewis, D. R. Luke and J. Malick, "Local linear convergence for alternating and averaged nonconvex projections," *Found. Comp. Math.*, vol. 9, no. 4, pp 485-513, Aug. 2009.

[11] D. R. Luke "Prox-regularity of rank constraint sets and implications for algorithms," *J. Math. Imaging and Vision*, vol. 47, no. 3, pp. 231-328, 2013.

[12] R. Hesse and D. R. Luke, "Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2397-2419, 2013.