Low-Rank Matrix Completion project

Authors: Artem Bochkarev, Roman Isachenko
Ilya Zharikov, Anne-Laure Ducrocq

Skolkovo Institute of Science and Technology
Methods of Optimization course

December 16, 2016.

Problem Statement

Task

Given the amount of observed matrix entries to reconstruct low-rank matrix approximation.

Applications:

- recommender systems;
- image-processing;
- imputation of NAs for genomic data;
- rank estimation for SVD.

Problem Statement

Notations:

- M — $n \times m$ unknown matrix;
- $\Omega \in \{1, \ldots, n\} \times \{1, \ldots, m\}$ indices of observed elements;
- $$P_\Omega(M) = \begin{cases} M_{ij}, & \text{if } (i,j) \in \Omega; \\ 0, & \text{otherwise.} \end{cases}$$

Optimization Task $(NP - hard)$

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \operatorname{rank}(X) \\ \text{subject to} \quad & P_\Omega(X) = P_\Omega(M) \end{aligned}$$

Related Works

1. Candes E. J., Recht B. Exact matrix completion via convex optimization. 2009.
2. Cai J. F., Candes E. J., Shen Z. A singular value thresholding algorithm for matrix completion. 2010.
3. Mazumder R., Hastie T., Tibshirani R. Spectral regularization algorithms for learning large incomplete matrices. 2010.
4. Jain P., Meka R., Dhillon I. S. Guaranteed rank minimization via singular value projection. 2010.
5. Takacs G. et al. Scalable collaborative filtering approaches for large recommender systems. 2009.
6. Vandereycken B. Low-rank matrix completion by Riemannian optimization. 2013.

Convex Relaxation (Candes E. J., Recht B., 2009)

Original Task

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \operatorname{rank}(X) \\ \text{subject to} \quad & P_\Omega(X) = P_\Omega(M) \end{aligned}$$

Relaxation

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \|X\|_* \\ \text{subject to} \quad & P_\Omega(X) = P_\Omega(M) \end{aligned}$$

Motivation: (analogue for l_0 and l_1 regularizations)

$$\operatorname{rank}(X) = |\{i : \sigma_i(X) \neq 0\}| \qquad \|X\|_* = \sum_{i=1}^{k} \sigma_i(X)$$

Problems:

- expensive computations;
- only for small matrices.

SVP (Jain P., Meka R., Dhillon I. S., 2010)

Original Task

$$\underset{X\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad \text{rank}(X)$$
$$\text{subject to} \quad P_\Omega(X) = P_\Omega(M)$$

Relaxation

$$\underset{X\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad \|P_\Omega(X) - P_\Omega(M)\|_F^2$$
$$\text{subject to} \quad \text{rank}(X) \leq k$$

Steps

- gradient descent for convex function;
- *SVD* projection on the k-dimensional space;

Features:

- sensitive to given rank k;
- guaranteed rank minimization for *RIP*-matrices.

SVT (Cai J. F., Candes E. J., Shen Z., 2010)

Original Task

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \operatorname{rank}(X) \\ \text{subject to} \quad & P_\Omega(X) = P_\Omega(M) \end{aligned}$$

Relaxation

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \tau \|X\|_* + \|X\|_F^2 \\ \text{subject to} \quad & P_\Omega(X) = P_\Omega(M) \end{aligned}$$

Steps

- proximal gradient descent method;
- soft-thresholding for singular values.

Features:

- sensitive to noise;
- robust to matrix rank.

SoftImpute (Hastie T., Tibshirani R., 2010)

Original Task

$$\underset{X\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad \text{rank}(X)$$
$$\text{subject to} \quad P_\Omega(X) = P_\Omega(M)$$

Relaxation

$$\underset{X\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad \|X\|_*$$
$$\text{subject to} \; \|P_\Omega(X) - P_\Omega(M)\|_F \leq \delta$$

Motivation:

The method is the same as *SVT* with $\delta = 0$. When $\delta > 0$ the overfitting is less possible.

Steps:

- proximal gradient descent method;
- soft-thresholded *SVD*;
- warm starts.

RISMF

Original Task

$$\begin{aligned} \underset{X \in \mathbb{R}^{n \times m}}{\text{minimize}} \quad & \operatorname{rank}(X) \\ \text{subject to} \quad & P_\Omega(X) = P_\Omega(M) \end{aligned}$$

Relaxation

$$\begin{aligned} \underset{\substack{U \in \mathbb{R}^{n \times k} \\ V \in \mathbb{R}^{k \times m}}}{\text{minimize}} \quad & \|U\|_F^2 + \|V\|_F^2 \\ \text{subject to} \quad & P_\Omega(UV) = P_\Omega(M) \end{aligned}$$

Steps:

- incremental gradient descent;
- updating U and V simultaneously.

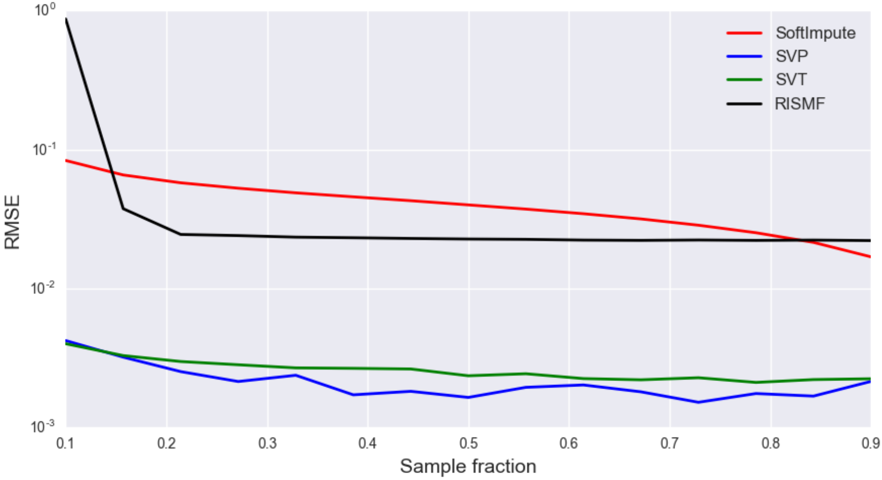
Features:

- need to choose learning rate;
- sensitive to regularization factor.

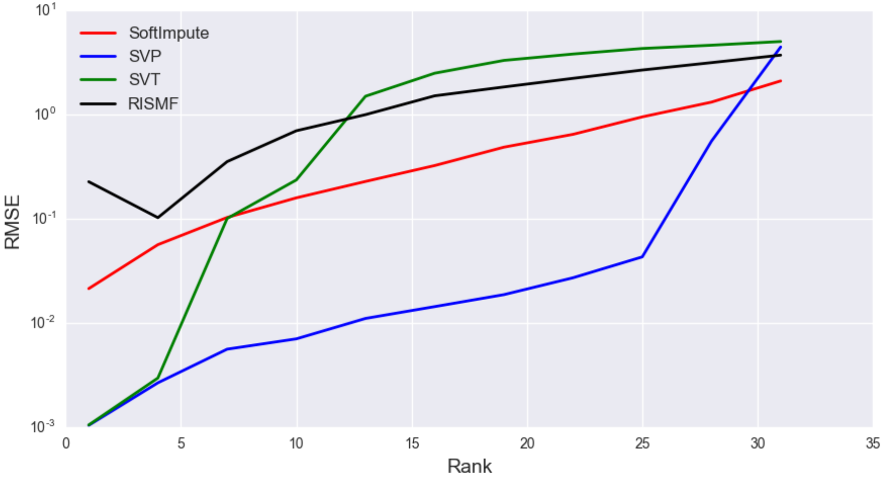
Numerical Experiments

- Synthetic data
  Low-rank matrices with random noise.
- Low-rank images
  Visual demonstration of algorithms.
- Assessment dataset
  Real dataset for collaborating filtering.

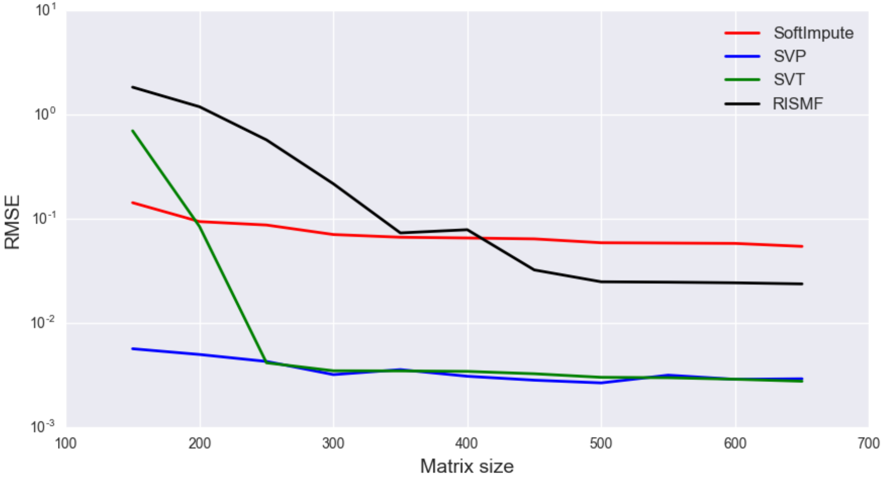
Synthetic Data



Synthetic Data

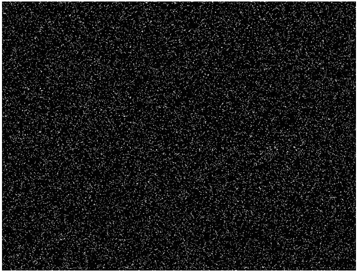


Synthetic Data

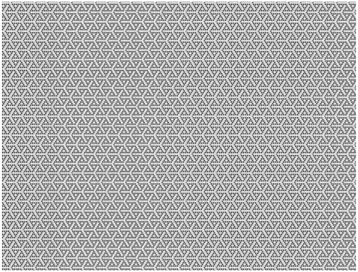


Images

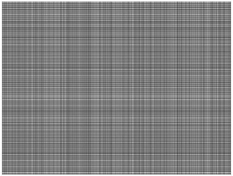
Sampled Image



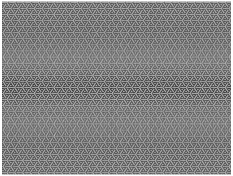
SoftImpute



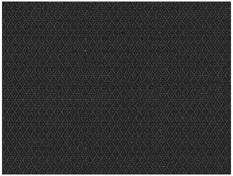
RISMF



SVT

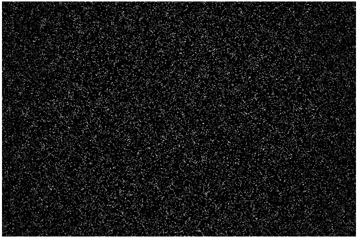


SVP

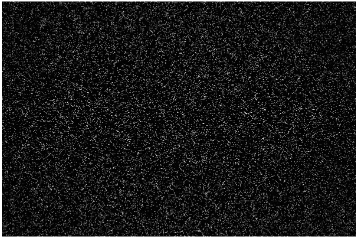


Images

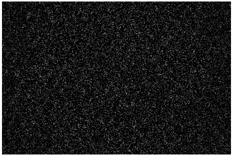
Sampled Image



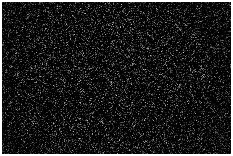
SVT



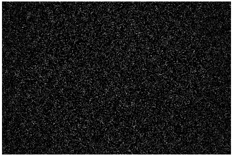
RISMF



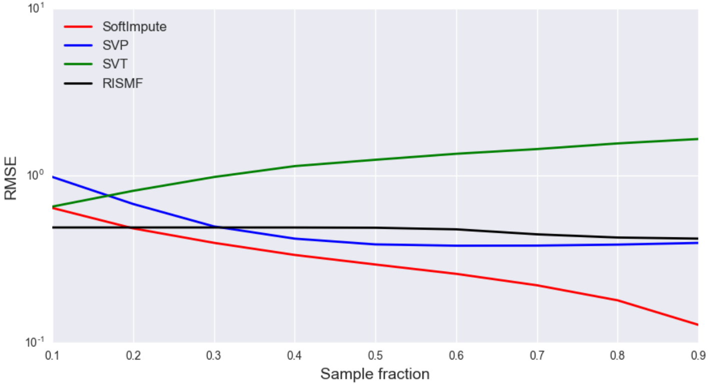
SVT



SVP



Assessment data



Assessment data