Matrix Completion Dream Team

Sergey Makarychev, Aleksandr Rozhnov

Mentors: Ivan Nazarov, Evgeny Frolov

Matrix completion is the task of filling in the missing entries of a partially observed matrix.

Matrix completion is the task of filling in the missing entries of a partially observed matrix.

- Collaborative filtering

Matrix completion is the task of filling in the missing entries of a partially observed matrix.

- Collaborative filtering
- Repairing damaged files (images)

Matrix completion is the task of filling in the missing entries of a partially observed matrix.

- Collaborative filtering
- Repairing damaged files (images)

Without any restrictions on the number of degrees of freedom in the completed matrix this problem is underdetermined since the hidden entries could be assigned arbitrary values. Thus matrix completion often seeks to find the lowest rank matrix

Define the following projection operator:

$$P_\Omega(X) : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}, X_{i,j} = \begin{cases} X_{i,j}, & \text{if } (i,j) \in \Omega \\ 0, & \text{otherwise} \end{cases}$$

We have the following optimization problem:

$$\begin{aligned} \text{minimize } & \text{rank}(Z) \\ \text{s.t } & P_\Omega(X) = P_\Omega(Z) \end{aligned}$$

The last problem is NP-hard. Convex relaxation:

$$\begin{aligned} &\text{minimize } \|Z\|_* \\ &\text{s.t } P_\Omega(X) = P_\Omega(Z), \end{aligned}$$

where $\|\|_*$ is a nuclear norm.
Due to the presence of noise, change the condition on more robust:

$$\|P_\Omega(X) - P_\Omega(Z)\|_F \leq \epsilon$$

Implemented approaches

- Alternating Least Squares
- Fast Alternating Least Squares [Mazumder, Hastie, Lee, Zadeh 2015]
- Riemannian optimization [Vandereycken, 2012]

Alternating Least Squares (ALS)

- Find Z in the form $Z = U^T V$, $U \in \mathbb{R}^{K \times M}$, $V \in \mathbb{R}^{K \times N}$
- Update U and V independently until convergence
- At each step optimal U and V can be found analytically

Alternating Least Squares (ALS)

- Find Z in the form $Z = U^T V$, $U \in \mathbb{R}^{K \times M}$, $V \in \mathbb{R}^{K \times N}$
- Update U and V independently until convergence
- At each step optimal U and V can be found analytically

Avoiding overfitting: add regularization term $\lambda(\|U\|_F^2 + \|V\|_F^2)$, still explicit formulas for optimal U and V^T at each step.

Alternating Least Squares (ALS)

- Find Z in the form $Z = U^T V$, $U \in \mathbb{R}^{K \times M}$, $V \in \mathbb{R}^{K \times N}$
- Update U and V independently until convergence
- At each step optimal U and V can be found analytically

Avoiding overfitting: add regularization term $\lambda(\|U\|_F^2 + \|V\|_F^2)$, still explicit formulas for optimal U and V^T at each step.

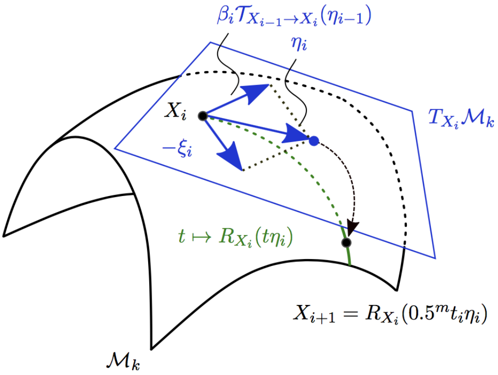
Functional to minimize:

$$\frac{1}{2}\|P_\Omega(X) - P_\Omega(U^T V)\|_F^2 + \lambda(\|U\|_F^2 + \|V\|_F^2) \to \min_{U,V},$$

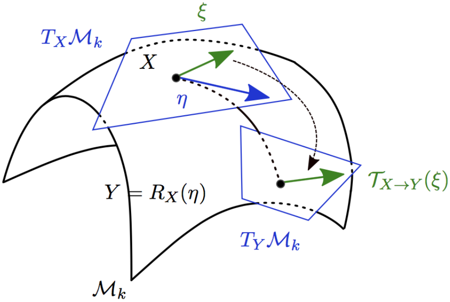
where Ω - indices of observed elements.

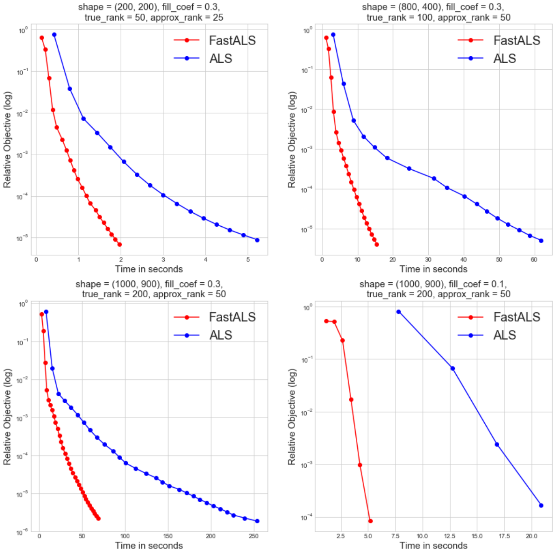
- Matrices of fixed-rank k forms a smooth manifold of dimensionality $(m+n-k)k$.
- Tangent space of the same dimensionality.
- Algorithm closely resembles a typical non-linear CG algorithm with Armijo line-search for unconstrained optimization.

Visualization of non-linear CG on a Riemannian manifold



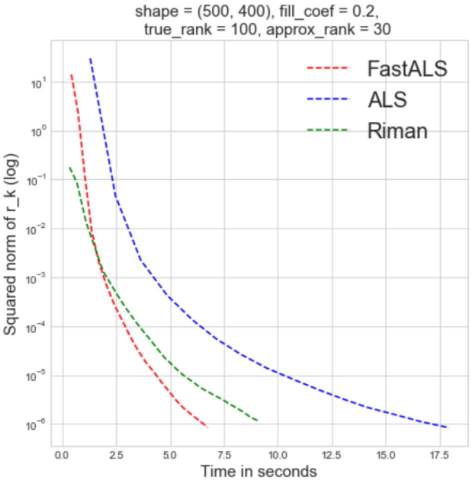
Vector transport on a Riemannian manifold



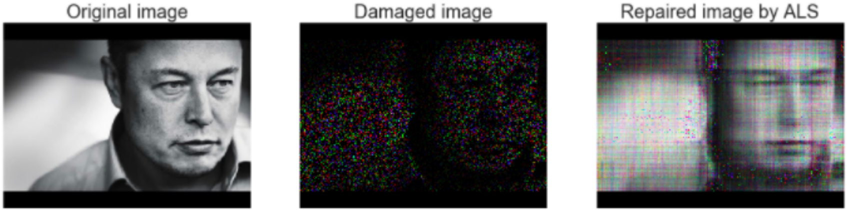

shape = (200, 200), fill_coef = 0.3,
true_rank = 50, approx_rank = 25
FastALS
ALS
Relative Objective (log)
Time in seconds
shape = (800, 400), fill_coef = 0.3,
true_rank = 100, approx_rank = 50
FastALS
ALS
Relative Objective (log)
Time in seconds
shape = (1000, 900), fill_coef = 0.3,
true_rank = 200, approx_rank = 50
FastALS
ALS
Relative Objective (log)
Time in seconds
shape = (1000, 900), fill_coef = 0.1,
true_rank = 200, approx_rank = 50
FastALS
ALS
Relative Objective (log)
Time in seconds

- ALS solves different regression problem for every row/column, because of their different amount of missingness. This can be costly.
- FastALS solves a single regresiion problem once and simultaneously for all rows/columns, because it operates on a filled-in matrix which is complete.

2-nd Experiment



We can recover image using ALS. Removing 90% of data we obtain damaged image. After that apply ALS and get Repaired image.



We can recover image using ALS. Removing 80% of data we obtain damaged image. After that apply ALS and get Repaired image.

Skoltech

Evaluating on Movielens1M dataset.

FastALS

	HR_10	HR_20	RMSE
mean	0.175	0.27	0.97
std	0.030	0.04	0.05

Riemannian Optimization

	HR_10	HR_20	RMSE
mean	0.120	0.17	0.98
std	0.034	0.05	0.06

All implemented algorithms have very nice and user-friendly interface, that is similar too sklearn ML algorithms:

- method fit(), predict()
- generator of random matrix completion problem
- generator of recovering image problem
- Metrics: Hit Rate, RMSE
- Exhaustive Grid Search for parameter tuning
- Cross-Validation for estimating

- Implementation of several competitive completion algorithms
- Comparative analysis of implemented algorithms
    - speed of convergence;
    - RMSE, HT metrics of quality;
    - application to images.
- The first step for creation Matrix Completion Library.