Singular Value Thresholding for Matrix Completion

Sriram Ganesan*

Nitin Garg
gargn@umich.edu

Jeeheh Oh
jeeheh@umich.edu

Jonathan Stroud
stroud@umich.edu

Abstract

1 Introduction

2 Different Algorithms

2.1 Baseline Algorithm (Eigentaste)

2.2 SMC

2.3 Singular Value Thresholding

Singular Value Thresholding (SVT) [1] is an algorithm proposed for *nuclear norm minimization* of a matrix X from a few known entries $M_{ij}, (i, j) \in \Omega$. Formally, SVT addresses the optimization problem

$$\begin{array}{ll} \underset{X}{\text{minimize}} & \|X\|_* \\ \text{subject to} & \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M), \end{array}$$

where $\|\cdot\|_*$ is the *nuclear norm*, or the sum of the singular values and $\mathcal{P}_\Omega(\cdot)$ makes zero all entries $(i, j) \notin \Omega$. This can be thought of as a convex relaxation to the rank minimization problem, and the two are formally equivalent under some conditions. The rank minimization problem is, however, highly non-convex and therefore not a suitable candidate for black-box optimization algorithms.

Singular Value Thresholding works by iteratively constructing X using a low-rank, low-singular value approximation to an auxiliary sparse matrix Y. Y is then adjusted to ensure the resulting approximation in the subsequent step has matching entries $X_{ij} = M_{ij}$. Each iteration consists of the inductive steps

$$\begin{cases} X^k = \text{shrink}(Y^{k-1}, \tau) \\ Y^k = Y^{k-1} + \delta_k \mathcal{P}_\Omega(M - X^k), \end{cases}$$

where $\text{shrink}(\cdot, \cdot)$ is the *singular value shrinkage operator*. Given a singular value decomposition $X = U\Sigma V^T$, $\Sigma = \text{diag}(\{\sigma_i\}_{1 \le i \le r})$, we can write this as

$$\text{shrink}(X, \tau) = U\Sigma_\tau V^T, \;\; \Sigma_\tau = \text{diag}(\{(\sigma_i - \tau)_+\}).$$

*The names are printed in alphabetical order by last name.

These two operations, when repeated, approach a low-nuclear norm solution by repeatedly shrinking the singular values of X. This algorithm has shown success in recovering accurate low-rank solutions when the source of M is also low-rank, even though it does not optimize this objective directly. The original authors discuss its theoretical guarantees in detail, but we choose to omit them in this discussion.

In practice, this system has a number of hyperparameters that must be carefully tuned to guarantee convergence. The shrinkage value τ must be set fairly high in order for the algorithm to converge quickly, but not too high that it dwarfs the true singular values. The stepsizes δ_k are similarly sensitive. These can be set dynamically as well, though we choose to maintain a fixed stepsize throughout. We compute the decomposition of Y^K in batches, which introduces a new batch size parameter l. Also important is the initialization of Y^0, for which the authors provide helpful strategies. Finally, we use the relative error

$\|\mathcal{P}_\Omega(X^k - M)\|_F / \|P_\Omega(M)\|$ as a stopping criterion; we terminate when this drops below a small ϵ.

3 Results and Discussion

3.1 Synthetic Data

3.2 Jester Dataset

4 Conclusions

5 Future Work

References

[1] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

size ($n \times n$)	rank (r)	m/d_r	m/n^2	time (s)	#iters	relative error
1000	10	6	0.119	279.1	250.0	86.59246 x 10^{-4}
1000	50	4	0.390	800.4	220.4	0.99209 x 10^{-4}
1000	100	3	0.570	604.6	163.8	0.98189 x 10^{-4}
5000	10	6	0.024	12934.9	250.0	611.42446 x 10^{-4}
5000	50	5	0.100	-	-	-
5000	100	4	0.158	-	-	-
10000	10	6	0.012	-	-	-
10000	50	5	0.050	-	-	-
10000	100	4	0.080	-	-	-
20000	10	6	0.006	-	-	-
20000	50	5	0.025	-	-	-
30000	10	6	0.004	-	-	-

Table 1: Performance of Singular Value Thresholding on synthetic matrices of known rank. We generate two $n \times r$ matrices U and V whose entries are i.i.d. gaussian.We choose m random entries from $M = UV^T$ and measure convergence rates of SVT. m/d_r is the ratio of sampled entries m and the "true dimensionality" $d_r = r(2n - r)$.