Transformed Schatten-1 Thresholding Algorithms for Low Rank Matrix Completion

Shuai Zhang
Joint with Penghang Yin and Jack Xin

Department of Mathematics
UC Irvine

May 25, 2016

Outline

Introduction

Transformed Schatten-1

TS1 Thresholding Algorithms

Numerical Experiments

Introduction

Matrix Completion

aims to recover a low-rank matrix from relatively few sampling of its entries. It is an application of matrix regularization.

TS1 regularization problem

$$\min_{X \in \Re^{m \times n}} \frac{1}{2} \|\mathscr{A}(X) - b\|_2^2 + \lambda T(X),$$

where $\mathscr{A}$ is the sampling operator and b is the given information.

TL1 and TS1

Transformed ℓ_1 penalty (TL1)

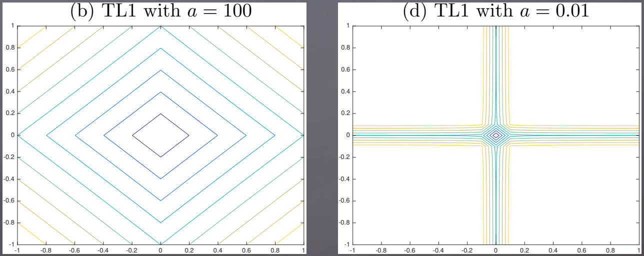
$$\rho_a(x) = \frac{(a+1)|x|}{a+|x|}, \text{ with } a \in (0, \infty).$$

Transformed Schatten-1 penalty (TS1)
is defined based on the singular values of matrix:

$$T(X) = \sum_{i=1}^{r} \rho_a(\sigma_i).$$

Properties of TL1

* TL1 bridges ℓ_0 ($a \to 0^+$) and ℓ_1 ($a \to \infty$);
* non-convex and increasing function;
* satisfies unbiasedness, continuity and sparsity properties;



TL1 Thresholding Function

Proximal point problem

$$y^* = arg\min_y \{ \frac{1}{2}(x-y)^2 + \lambda\rho_a(y) \},$$

has closed form solution

$$y^* = g_{\lambda,a}(x) = \begin{cases} 0, & |x| \leq t; \\ h_\lambda(x), & |x| > t. \end{cases}$$

Function $h_\lambda(x)$ is given by

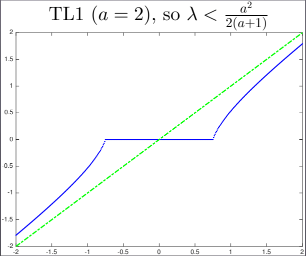
$$h_\lambda(x) = sgn(x)\left\{\frac{2}{3}(a+|x|)cos(\frac{\varphi(x)}{3}) - \frac{2a}{3} + \frac{|x|}{3}\right\},$$

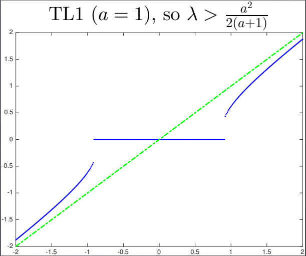
with $\varphi(x) = \arccos(1 - \frac{27\lambda a(a+1)}{2(a+|x|)^3})$.

Sub-critical and Super-critical Schemes

Threshold value t depends on regularization parameter λ,

- * if $\lambda \leq \frac{a^2}{2(a+1)}$ (sub-critical), $\quad t = t_2^* = \lambda \frac{a+1}{a}$;
- * if $\lambda > \frac{a^2}{2(a+1)}$ (super-critical), $t = t_3^* = \sqrt{2\lambda(a+1)} - \frac{a}{2}$.





TS1 Optimal Point Representation

Suppose X^* is a global minimizer of the TS1 regularization problem, and define $B_\mu(X^*) = X^* + \mu\mathscr{A}^*(b - \mathscr{A}(X^*))$ with singular values decomposition: $B_\mu(X^*) = U\,\mathrm{Diag}(\sigma_B^*)\,V^t$.

Theorem

If we choose $0 < \mu < \|\mathscr{A}\|^{-2}$, X^* *satisfies*

$$X^* = U\,\mathrm{Diag}(g_{\lambda\mu,a}(\sigma_B^*))\,V^t,$$

which means the singular values of X^* *satisfy*

$$\sigma_i^* = g_{\lambda\mu,a}(\sigma_{B,i}^*),\ \textit{for } i = 1, \cdots, m.$$

TS1 Thresholding Scheme

Fixed parameters iteration scheme

$$\begin{aligned} X_k &= G_{\lambda\mu,a}(B_\mu(X_{k-1})) \\ &= U_{k-1}\,\mathrm{Diag}\left(g_{\lambda\mu,a}(\sigma_{k-1})\right)\, V_{k-1}^t, \end{aligned}$$

where unitary matrices U_{k-1}, V_{k-1} and singular values $\{\sigma_{k-1,i}\}_i$ come from the SVD decomposition of matrix $B_\mu(X_{k-1})$.

Remark:
For this thresholding scheme, we have 3 tuning parameters: λ, μ and TL1 parameter a.

Relation of threshold value and rank

Suppose the rank of X^* is given or estimated as r,

$$\begin{aligned} \sigma_i^* > t \quad & \text{if } i \in \{1, 2, ..., r\}, \\ \sigma_j^* \le t \quad & \text{if } j \in \{r+1, r+2, ..., m\}, \end{aligned}$$

where σ^* is the singular values of $B_\mu(X^*)$.

t is the TL1 threshold value and equal to t_2^* or t_3^* depending on parameters values. It can be checked that $t_3^* \le t \le t_2^*$, so

$$\begin{aligned} &\sigma_r^* \ge t_3^* = \sqrt{2\lambda\mu(a+1)} - \tfrac{a}{2}; \\ &\sigma_{r+1}^* \le t_2^* = \lambda\mu\tfrac{a+1}{a}. \end{aligned}$$

Relation of threshold value and rank

From the previous inequalities, we get the bounds for λ,

$$\lambda_l^* \equiv \frac{a\sigma_{r+1}^*}{\mu(a+1)} \le \lambda \le \lambda_u^* \equiv \frac{(a+2\sigma_r^*)^2}{8(a+1)\mu}.$$

* λ_l^* comes from the formula of t_2^* (sub-critical scheme);
* λ_p^* comes from the formula of t_3^* (super-critical scheme).

Semi-adaptive TS1 Algorithm (TS1-s1)

Method: fix a and μ; change λ at each step.
At k-th iteration step, optimal parameter λ_k is

$$\lambda_k = \begin{cases} \lambda_l, & \text{if } \lambda_l \leq \frac{a^2}{2(a+1)\mu}, \\ \lambda_u, & \text{if } \lambda_l > \frac{a^2}{2(a+1)\mu}, \end{cases}$$

where λ_l uses the same formula of λ_l^* with σ^* approximated by $B_\mu(X_{k-1})$

Remark:
In the algorithm, it checks the value of λ_l to determine λ, which means TS1-s1 prefers sub-critical threshold scheme.

Adaptive TS1 Algorithm (TS1-s2)

Method: fix μ; change a and λ at each step.
At each iterative step, we choose a such that equality $\lambda = \frac{a^2}{2(a+1)\mu}$ holds, in which case

$$t = t_3^* = t_2^*.$$

By the formulas of threshold values, we have

$$\frac{2(\sigma_{r+1}^*)^2}{1 + 2\sigma_{r+1}^*} \leq \lambda \leq \frac{2(\sigma_r^*)^2}{1 + 2\sigma_r^*}.$$

In the algorithm, we evaluate λ first and then choose a.

Numerical Experiments

Random low rank matrices are
$M = M_L M_R^t \in \mathcal{R}_{m\times n}$,
where matrices $M_L \in \mathcal{R}_{m\times r}$ and $M_R \in \mathcal{R}_{n\times r}$ are generated with Gaussian distributions.

The difficulty of a recovery problem is quantified by

* Sampling ratio: $\mathrm{SR} = p/mn$.
* Freedom ratio: $\mathrm{FR} = r(m+n-r)/p$, which is the freedom of rank r matrix divided by the number of measurement.

Matrix Completion with Known Rank

Comparison of TS1-s1, TS1-s2, IRucL-q on recovery of uncorrelated multivariate Gaussian matrices at known rank, $m = n = 100$, SR = 0.4.

Problem		TS1-s1		TS1-s2		IRucL-q	
rank	FR	rel.err	time	rel.err	time	rel.err	time
10	0.4750	3.26e-05	0.33	1.11e-06	0.34	3.21e-04	2.49
14	0.6510	1.10e-05	0.53	1.03e-05	0.52	3.80e-05	7.25
15	0.6937	1.05e-05	0.66	9.88e-06	0.64	5.28e-05	9.29
16	0.7360	3.86e-05	0.91	1.79e-05	0.87	7.57e-05	12.34
17	0.7778	1.50e-04	1.03	7.10e-05	1.00	9.40e-05	15.31
18	0.8190	5.63e-04	1.00	4.15e-04	1.00	1.49e-04	22.27

Matrix Completion with Known Rank

Numerical experiments on multivariate Gaussian matrices with varying covariance at known rank, $m = n = 1000$, SR $= 0.4$.

Problem		TS1-s1		TS1-s2		IRucL-q	
rank	cor	rel.err	time	rel.err	time	rel.err	time
30	0.1	3.07e-06	9.71	3.07e-06	3.98	3.13e-06	222.90
30	0.2	2.90e-06	11.07	2.94e-06	3.92	3.16e-06	221.34
30	0.3	5.54e-03	26.64	3.02e-06	4.13	3.05e-06	218.57
30	0.4	1.19e-02	28.58	3.08e-06	4.31	3.29e-06	214.52
30	0.5	4.76e-02	34.25	2.89e-06	5.89	3.12e-06	209.05
30	0.6	6.89e-02	35.69	2.89e-06	10.28	3.30e-06	207.94
30	0.7	8.01e-02	33.92	6.99e-04	20.09	3.15e-06	210.06

Matrix Completion with Rank Estimation

Ground true matrices are generated by multivariate Gaussian with different covariance, $m = n = 100$, and SR $= 0.4$.

Problem		TS1-s1		TS1-s2		FPCA		IRucL-q	
rank	cor	rel.err	time	rel.err	time	rel.err	time	rel.err	time
5	0.5	5.49e-06	0.20	6.77e-02	0.86	1.61e-05	0.12	7.50e-06	2.07
5	0.6	5.45e-06	0.20	7.74e-02	0.91	1.69e-05	0.11	6.93e-06	1.76
5	0.7	5.25e-06	0.25	1.04e-01	1.33	1.53e-05	0.12	4.71e-04	2.06
10	0.5	1.10e-05	0.65	1.17e-01	1.14	1.21e-01	0.97	1.76e-05	3.35
10	0.6	1.61e-02	0.76	1.32e-01	1.04	1.02e-01	0.86	2.72e-05	4.26
10	0.7	9.14e-02	0.91	1.55e-01	0.93	9.11e-02	0.82	7.12e-04	4.59

Thanks for your attention!

Transformed Schatten-1 Thresholding Algorithms for Low Rank Matrix Completion

Shuai Zhang
Joint with Penghang Yin and Jack Xin

Department of Mathematics
UC Irvine

May 25, 2016