SVD-free Convex-Concave Approaches for Nuclear Norm Regularization

Yichi Xiao[1*], **Zhe Li**[2*], **Tianbao Yang**[2], **Lijun Zhang**[1]

[1]N[illegible] K[illegible] L[illegible] [illegible], N[illegible] U[illegible], N[illegible] 210023, C[illegible]

[2]D[illegible] C[illegible] S[illegible], U[illegible] I[illegible], I[illegible] C[illegible], IA 52242, USA

{[illegible]}@[illegible], {[illegible]}@[illegible]

Abstract

[illegible] (SVD) [illegible] SVD. [illegible] *first* SVD-[illegible] $O(1/\sqrt{T})$ [illegible]

1 Introduction

[illegible] [C[illegible], 2009; A[illegible] *et al.*, 2009] [illegible] [Ar[illegible] *et al.*, 2008; P[illegible] *et al.*, 2010] [illegible]

[illegible]:

$$\min_{A\in\mathbb{R}^{m\times n}} \quad (A) = \quad (A) + \quad \|A\|_* \tag{1}$$

[illegible] $(\cdot)$ [illegible] 0 [illegible] $\|A\|_* = \mathrm{trace}(\sqrt{A^\top A})$ [illegible] A [illegible] $\leq$ [illegible].

T[illegible] (1) [illegible] [N[illegible], 2004] [illegible] [D[illegible] S[illegible], 2009; N[illegible], 2013]. [illegible] (SVD) [illegible] SVD [illegible] SVD. H[illegible] $(\cdot)$ [illegible] [D[illegible] *et al.*, 2012; H[illegible] O[illegible], 2014] [illegible] [J[illegible] *et al.*, 2010; J[illegible], 2013].

[illegible] (1), [illegible] $(\cdot)$ [illegible] *non-smooth*, [illegible] SVD-[illegible]. F[illegible] (1) [illegible] SVD, [illegible] *first* SVD-[illegible] $O(1/\sqrt{T})$ [illegible] [N[illegible], 2004]. S[illegible] $(\cdot)$ [illegible] $O(1/\sqrt{T})$ [illegible] $_1$-[illegible].

---

*E[illegible] C[illegible]

Applications A [illegible] T [illegible]:

- **Robust Low-rank Matrix Approximation:** T [illegible] A [illegible] [B [illegible] *et al.*, 1996; Cr [illegible] F [illegible] r, 1998; K [illegible] K [illegible], 2005 .] F [illegible] $(A) = \sum_{ij} |A_{ij} - \; _{ij}|$ [illegible].
- **Sparse and Low-rank Link Prediction:** T [illegible] [R [illegible] *et al.*, 2012 .] T [illegible] $(A) = (A \quad) + \gamma \|A\|_1$, [illegible] $(A \quad)$ [illegible].

2 Related Work

I [illegible].

2.1 Nuclear-norm Regularized Problems

O [illegible] (1) [illegible] (GD):

$$A_{t+1} = A_t - \eta_t \left(\nabla \;(A_t) + \; \|A_t\|_* \right)$$

[illegible] $\|A_t\|_*$ [illegible] $\|\cdot\|_*$ [illegible] A_t, [illegible] η_t [illegible] 0 [illegible] GD [illegible] $(1/\sqrt{T})$ [illegible] T -138

Algorithm 1 SVD-[illegible] E CON[illegible] -C[illegible] E A[illegible] (SECONE)

1: **Initialize:** $A_1 = \quad_1 = 0 \in \mathbb{R}^{m\times n}$
2: **for** $t = 1$ **to** T **do**
3: U[illegible] A_{t+1}

$$A_{t+1} = A_t - \eta_t(\quad (A_t) + \quad_t)$$

4: U[illegible] $_{t+1}$

$$_{t+1} = \quad_t + \quad(\quad A_t - \quad [\| \quad_t\|_2 - 1]_+)$$

5: **end for**
6: **Output:** $\widehat{A}_T = \sum_{t=1}^{T} A_t/ T$

S[illegible] v[illegible] , [illegible] v[illegible] : H[illegible] v[illegible] r, [illegible] , v[illegible] , [illegible] SVD [illegible] r-[illegible].

T[illegible] , [illegible] $\| \quad \|_2 \leq 1$, [illegible] :

$$\min_{A\in\mathbb{R}^{m\times n}} \max_{U\in\mathbb{R}^{m\times n}} \quad (A) + \quad t \; (\quad ^\top A) - \quad [\| \quad \|_2 - 1]_+ \quad (3)$$

[illegible] 0 [illegible] v [illegible] [illegible] $[\;]_+ = \max\{ \quad 0\}$ [illegible] r. T [illegible] v [illegible] , [illegible] , [illegible] :

$$A_{t+1} = A_t - \eta_t(\quad (A_t) + \quad_t)$$
$$_{t+1} = \quad_t + \quad(\quad A_t - \quad [\| \quad_t\|_2 - 1]_+)$$

N[illegible] $[\| \quad \|_2 - 1]_+$ [illegible]

Algorithm 2 E (SECONE-S)

1: **Initialize:** $A_1 = \quad_1 = 0 \in \mathbb{R}^{m\times n}$
2: **for** $t = 1$ **to** T **do**
3: S $\quad_t$
4: U A_{t+1}

$$A_{t+1} = A_t - \eta_t(\quad (A_t; \quad_t) + \quad_t)$$

5: U $\quad_{t+1}$

$$\quad_{t+1} = \quad - \quad [\| \quad_t\|_2 - 1]_+)$$

6: **end for**
7: **Output:** $\widehat{A}_T =$

T v r ... (5)

r $(\cdot)$ ... r $\|A\|_1$ γ 0 ... N ... (5), ... T ... v ...

$$\min_{A\in\mathbb{R}^{m\times n}} \max_{U\in\mathbb{R}^{m}} \ldots t \; (\;^\top A) - \; [\| \; \|_2 - 1]_+$$

In ... , ... $\gamma \; (\cdot)$, ... fi , ... fi ... $(\cdot)$:

$$\mathrm{Pr}_{\eta,h} \ldots A) + \frac{1}{2\eta}\|A - \bar{A}\|_F^2$$

W $(A) =$ ... :

$$\bar{A}_{t+1} \ldots + \quad_t)$$
$$A_{t+1} \ldots)$$
$$\quad_{t+1}$$

L $\eta_t = /\sqrt{t}$,

$$\sum_{t=1}^{T}\frac{1}{2\eta_t}(\|A - A_t\|_F^2 - \|A - A_{t+1}\|_F^2)$$

$$\leq \frac{1}{2\eta_1}\|A - A_1\|_F^2 + \sum_{t=1}^{T-1}\left(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\right)\|A - A_t\|_F^2$$

$$\leq \frac{1}{2\eta_1}\ {}_1^2 + \left(\frac{1}{2\eta_T} - \frac{1}{2\eta_1}\right)\ {}_1^2 \leq \frac{1}{2\eta_T}\ {}_1^2 = \frac{\sqrt{T}}{2\ {}_1}\ {}_1^2$$

r $_1 = \|A\| + \ \geq \max_t \|A - A_t\|_F$.

W $_t = /\sqrt{t}$

$$\sum_{t=1}^{T}(A_t\ ) \leq \sum_{t=1}^{T}(A\ _t) + \frac{\sqrt{T}}{2\ _1}\ {}_1^2 + \frac{\sqrt{T}}{2\ _2}\ {}_2^2$$

$$+\ _1\sqrt{T}(\ +\ )^2 + \ _2\sqrt{T}(\ +\ )^2$$

r $\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq 1 + \int_{t=1}^{T}\frac{1}{\sqrt{t}}\mathrm{d}t \leq 2\sqrt{T}$ $_2 = \|\ \|_F + \ \geq \max_t \|\ -\ _t\|_F$.

D $\widehat{A}_T = \sum_{t=1}^{T} A_t/T$ $\widehat{\ }_T = \sum_{t=1}^{T}\ /T$. B $(A\ )$ A v ,

$$(\widehat{A}_T\ ) - (A\ \widehat{\ }_T) \leq \frac{1}{\sqrt{T}}\left(\frac{{}_1^2}{2\ _1} + \ _1(\ +\ )^2\right) + \frac{1}{\sqrt{T}}\left(\frac{{}_2^2}{2\ _2} + \ _2(\ +\ )^2\right)$$

T v v , $(\widehat{A}_t\ ) - (A\ \widehat{\ }_t)$ r $(1/\sqrt{T})$ r , $(\widehat{A}_T\ \widehat{\ }_T)$ r v r (3).

Part II: r r , - v (A_t) (A_*), r A_* r . T v

$$(\widehat{A}_T) - (A_*) \leq (\widehat{A}_T\ {}_T^*) - (A_*\ \widehat{\ }_T)$$

r ${}_T^* = \arg\max_{\|U\|_2\leq 1} t(\ ^\top \widehat{A}_T)$. F , v

$$(\widehat{A}_T\ {}_T^*) = (\widehat{A}_T) + \|\widehat{A}_T\|_* = (\widehat{A}_T)$$

T , r

$$\|A_*\|_* \geq t(\widehat{\ }_T^\top A_*) - [\|\widehat{\ }_T\|_2 - 1]_+ \qquad (7)$$

N $t(A^\top\ ) \leq \|A\|_2\|\ \|_*$ r A . W $\|\widehat{\ }_T\|_2 \leq 1$, v (7) . $\|\widehat{\ }_T\|_2 \geq 1$.

L $\widehat{\ }_T$ v SVD $\widehat{\ }_T = \Sigma^\top$, r $\Sigma = (\ _1\ _r\ _{r+1}\ _m)$. T $\{\ _i\}$ r $_r \geq 1$ $_{r+1}$. D $_T$ r $\widehat{\ }_T$. S , r $\widetilde{\ }_T = \widetilde{\Sigma}$, r $\widetilde{\Sigma} = (1\ 1\ _{r+1}\ _m)$.

$$t(\widetilde{\ }_T^\top A_*) \leq \|A_*\|_*$$

$$\|\widehat{\ }_T - \widetilde{\ }_T\|_2 \leq \ _1 - 1 = \|\widehat{\ }_T\|_2 - 1$$

W r v (7) . F A - 1, v $\|A_*\|_* \leq$. W v

$$t(\widehat{\ }_T^\top A_*) - \|A_*\|_* \leq t((\widehat{\ }_T - \widetilde{\ }_T)^\top A_*)$$

$$\leq \|\widehat{\ }_T - \widetilde{\ }_T\|_2\|A_*\|_* \leq (\|\widehat{\ }_T\|_2 - 1)$$

$$\leq (\|\widehat{\ }_T\|_2 - 1)$$

$\|A_*\|_* \leq$ r r $\geq$. T r T r 1.

4.2 Proof of Theorem 3

W r v . L $(A\ ) = (A) + t(\ ^\top A) + \gamma(A) - [\|\ \|_2 - 1]_+$. D

$$_t(A) = (A) + t(\ _t^\top A) \quad (A) = \gamma(A)$$

$$_t = \ _t(A_t) = (A_t) + \ _t$$

Fr $_t(A)$ (A), v

$$\eta_t(\ _t(A_t) + (A_{t+1}) - \ _t(A) - (A))$$
$$\leq\langle A_t - A\ \eta_t\ _t\rangle + \langle A_{t+1} - A\ \eta_t\ (A_{t+1})\rangle$$
$$=\langle A - A_{t+1}\ A_t - A_{t+1} - \eta_t\ _t - \eta_t\ (A_{t+1})\rangle$$
$$+\langle A - A_{t+1}\ A_{t+1} - A_t\rangle + \eta_t\langle A_t - A_{t+1}\ _t\rangle$$

Fr A_{t+1} A r 3, v

$$\langle A - A_{t+1}\ A_{t+1} - A_t + \eta_t\ _t + \eta_t\ (A_{t+1})\rangle \geq 0$$

B V $A = A_*$,

$$\eta_t(\ _t(A_t) + (A_{t+1}) - \ _t(A_*) - (A_*))$$
$$\leq\langle A_* - A_{t+1}\ A_{t+1} - A_t\rangle + \eta_t\langle A_t - A_{t+1}\ _t\rangle$$
$$\leq\frac{1}{2}\left(\|A_* - A_t\|_F^2 - \|A_* - A_{t+1}\|_F^2 - \|A_{t+1} - A_t\|_F^2\right)$$
$$+\frac{1}{2}\left(\|A_t - A_{t+1}\|_F^2 + \eta_t^2\|\ _t\|_F^2\right)$$
$$\leq\frac{1}{2}\left(\|A_* - A_t\|_F^2 - \|A_* - A_{t+1}\|_F^2\right) + \frac{\eta_t^2}{2}\|\ _t\|_F^2$$

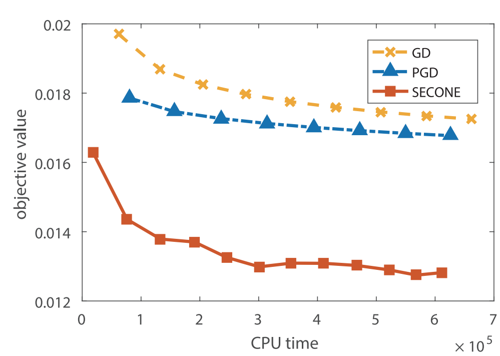
T fr C -S - . L ' r $(A_t\ _t) - (A_*\ _t)$,

$$(A_t\ _t) - (A_*\ _t)$$
$$= \ _t(A_t) + (A_t) - \ _t(A_*) - (A_*)$$
$$\leq\frac{1}{2\eta_t}(\|A_* - A_t\|_F^2 - \|A_* - A_{t+1}\|_F^2) + \frac{\eta_t}{2}\|\ _t\|_F^2$$
$$+\gamma(\ (A_t) - (A_{t+1}))$$

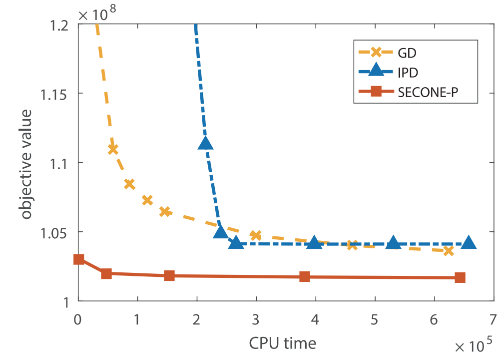
N r $\gamma(\ (A_t) - (A_{t+1}))$ - r $(A_1) = (0) = 0$. U r r T r 1, rr r r r .

5 Experiments

W r r r - fi fr r r , r r - r .



Fi r 1: R f r -r r r



Fi r 2: R f r -r r

T 1: S f r r

M	1	2	$\mathcal{T}$	T CPU
SECONE	1 7	1 4	8500	6 12 5
PGD	1 6		80	6 26 5
GD	1 6		90	6 62 5

T 2: S f r r

M	1 r	2	$\mathcal{T}$	T CPU
SECONE-P	1	1 −5	2000	6 43 5
IPD	0 1		50	6 57 5
GD	1		40	6 25 5

5.1 Robust Low-rank Matrix Approximation

W r r -r r [B et al., 1996; Cr F r, 1998; K K , 2005 :]

$$\min_{A\in\mathbb{R}^{m\times n}} \frac{1}{\ } \sum_{i=1}^{m}\sum_{j=1}^{n} | \ _{ij} - A_{ij}| + \ \|A\|_*$$

r . D fi , r r : r (GD) r r (PGD) [D S r, 2009]. W N 20^2 , = 11 269 , f = 20 302 f r (fi f r r 7). A r T r 1, A 1 $\eta_t = 1/\sqrt{t}$ $= 1/2\sqrt{t}$, r 1 2 r . T $\eta_t = 1/\sqrt{t}$ f r GD PGD. W v f 1 2 r f {1 −5 1 −4 1 10} r r v . In Fi . 1, v v r r f r = 1 −6. W v f r -r , r v r . A , SECONE r f r GD PGD. T SECONE SVD-fr - ffi , r T 1. A , f r f SECONE .

5.2 Sparse and Low-rank Link Prediction

G r f r 0 1 fi r -r r :

$$\min_{A\in\mathbb{R}^{m\times n}} \sum_{ij} \max(1-(2\ _{ij}-1)\cdot A_{ij}\ \ 0)+\gamma\|A\|_1+\ \ \|A\|_*$$

2 :// . / /20N r /

F [R r et al., 2012 ,] f r r F 100 r fr r . W v r f 41 554 fir = = 15 000 r r . W fl 15% f r r r -r fr . W r A 3 (SECONE-P) r (GD) In r Pr D (IPD), r f r v r f r [R r et al., 2012 .] T SECONE-P GD r S 5.1. T r r f IPD r r f {1 −2 1 −1 1 2}. In Fi . 2, v v r r = 10 γ = 0 4. A , SECONE-P r f r r . T f IPD r r - . T f ff r r T 2. A r r r f SECONE-P r .

Acknowledgments

T NSFC (61603177), J SF (BK20160658), C r In C r f Nv S f r T In N U r . L T. r N S F (IIS-1463988, IIS-1545995).

References

[A et al., 2009] J A , Fr B , T r E , J -P V . A - r fi : O r r f r . Journal of Machine Learning Research, 10:803 826, 2009.

[Ar... *et al.*, 2008] Ar... Ar..., T... E... M... P...: C... . *Machine Learning*, 73(3):243 272, 2008.

[B... *et al.*, 1996] A... B..., P. B..., A. F... A ...-... PCA ... . In *Proceedings of the International Conference on Ordinal and Symbolic Data Analysis*, 359 368, 1996.

[C... *et al.*, 2010] J...-F... C..., E... J C... S... . A ... . *SIAM Journal on Optimization*, 20(4):1956 1982, 2010.

[C... R..., 2009] E... J. C... B... R... . E... . *Foundations of Computational Mathematics*, 9(6):717 772, 2009.

[Cr... F..., 1998] C... Cr... P... F... R... . In *Proceedings in Computational Statistics*, 245 250, 1998.

[D... S..., 2009] J... D... S... . E... . *Journal of Machine Learning Research*, 10(D):2899 2934, 2009.

[D... *et al.*, 2012] M... D... H... J... M... L... . In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics*, 22, 327 336, 2012.

[H..., 2008] E... H... S... . In *Proceedings of the 8th Latin American Conference on Theoretical Informatics*, 306 316, 2008.

[H... O..., 2014] C... -J... H... P... A O... . N... . In *Proceedings of The 31st International Conference on Machine Learning*, 575 583, 2014.

[J... *et al.*, 2010] M... J... M... S... . A ... . In *Proceedings of the 27th International Conference on Machine Learning*, 471 478, 2010.

[J..., 2013] M... J... R... : Pr... . In *Proceedings of the 30th International Conference on Machine Learning*, 427 435, 2013.

[J..., 2009] S... J... J... . A... . In *Proceedings of the 26th Annual International Conference on Machine Learning*, 457 464, 2009.

[K... K..., 2005] Q... K... T... K... . R... . In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 739 746, 2005.

[L..., 2012] G... L... . A... . *Mathematical Programming*, 133:365 397, 2012.

[L... *et al.*, 2014] Q... L..., C..., J... P... . A ... . *Computational Optimization and Applications*, 58(2):455 482, 2014.

[M... *et al.*, 2012] M... M..., T..., R... J..., S..., J... S... . In *Advances in Neural Information Processing Systems (NIPS)*, 503 511, 2012.

[N... *et al.*, 1982] Ar... N..., D... B..., E-R D... . *Problem complexity and method efficiency in optimization*. W... -I... . W..., 1982.

[N..., 2004] ... N... . *Introductory lectures on convex optimization: a basic course*, 87 *Applied optimization*. K... A... P..., 2004.

[N..., 2013] ... N... . Gr... . *Mathematical Programming*, 140(1):125 161, 2013.

[P... *et al.*, 2010] T... K... P..., P... T..., S... J..., J... . Tr... : R... . *SIAM Journal on Optimization*, 20(6):3465 3489, 2010.

[R... S..., 2005] J... DM R... N... S... . F... . In *Proceedings of the 22nd International Conference on Machine Learning*, 713 719, 2005.

[R... *et al.*, 2012] E... R..., P... -A... S..., N... V... . E... . In *Proceedings of the 29th International Conference on Machine Learning*, 1351 1358, 2012.

[S... -S... *et al.*, 2011] S... S... -S..., A... G..., O... S... . L... . In *Proceedings of the 28th International Conference on Machine Learning*, 329 336, 2011.

[S... *et al.*, 2013] M... S..., V... C..., J A S... . A... . In *Proceedings of the IEEE Conference on Decision and Control*, 2013.

[Sr... *et al.*, 2005] N... Sr..., J... R..., T... S. J... . M... . In *Advances in Neural Information Processing Systems 17*, 1329 1336, 2005.

[T..., 2010] K... -C... T... S... . A... . *Pacific Journal of Optimization*, 6(615-640):15, 2010.