

# Adaptive Singular Value Thresholding

Nematollah Zarmehi and Farokh Marvasti  
Advanced Communication Research Institute  
Department of Electrical Engineering

Sharif University of Technology, Tehran, Iran

Email: <http://zarmehi.ir/contact.html>, [marvasti@sharif.edu](mailto:marvasti@sharif.edu)

**Abstract**—In this paper, we propose an Adaptive Singular Value Thresholding (ASVT) for low rank recovery under affine constraints. Unlike previous iterative methods that the threshold level is independent of the iteration number, in our proposed method, the threshold in adaptively decreases during iterations. The simulation results reveal that we get better performance with this thresholding strategy.

## I. INTRODUCTION

Finding a minimum rank matrix subject to affine constraints is the heart of many engineering applications such as minimal realization theory [1], minimum order controller design [2]–[4], collaborative filtering [5], and machine learning [6]–[8]. This problem is usually known as Affine Rank Minimization (ARM). Mathematically speaking, we want to solve the following minimization problem:

$$\begin{aligned} \min_{\mathbf{X}} \text{rank}(\mathbf{X}), \\ \text{s.t.} : \mathcal{A}(\mathbf{X}) = \mathbf{b}, \end{aligned} \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$  is the unknown matrix,  $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$  denotes the linear mapping, and  $\mathbf{b}$  is the measurement vector. The affine constraint  $\mathcal{A}(\mathbf{X}) = \mathbf{b}$  can be simplified as

$$\text{Vec}(\mathbf{X}) = \mathbf{v}$$

where  $\mathbf{A} \in \mathbb{R}^{m \times n_1 n_2}$  is the matrix representation of  $\mathcal{A}$  and  $\text{vec}(\cdot)$  is the standard vectorization operator.

Since rank is the number of non-zero Singular Values (SVs), the minimization problem (1) is the same as

$$\begin{aligned} \min_{\mathbf{X}} \|\sigma(\mathbf{X})\|_0, \\ \text{s.t.} : \mathcal{A}(\mathbf{X}) = \mathbf{b}, \end{aligned} \quad (2)$$

where  $(\sigma)(\mathbf{X})$  is the vector of SVs of  $\mathbf{X}$ .

An important special case of ARM problem introduced in (1) is the matrix completion problem:

$$\begin{aligned} \min_{\mathbf{X}} \text{rank}(\mathbf{X}), \\ \text{s.t.} : [\mathbf{X}]_{i,j} = [\mathbf{X}_*]_{i,j}, \quad \forall (i,j) \in \Omega, \end{aligned} \quad (3)$$

where  $\mathbf{X}$  is the unknown matrix,  $[\mathbf{X}]_{i,j}$  is the element of matrix  $\mathbf{X}$  at the intersection of  $i$ -th row and  $j$ -th column, and  $\mathbf{X}_*$  is the low rank matrix that we would like to recover but we only know a small subset of all entries of  $\mathbf{X}_*$ , that is,  $\Omega \subseteq \{1, 2, \dots, n_1\} \times \{1, 2, \dots, n_2\}$ .

There are about two main approaches to solve the ARM problem or better saying approximate the non-convex minimization problem (2).

In the first approach, the rank function is approximated by the nuclear norm. The nuclear norm of a matrix is defined as  $\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{X})$  where  $\sigma_i$  is the  $i$ -th largest SV of  $\mathbf{X}$ . Therefore, we can write the nuclear norm as  $\|\mathbf{X}\|_* = \|\sigma(\mathbf{X})\|_1$  and finally, the nuclear norm minimization will be equivalent to an  $\ell_1$ -norm minimization as follows [9], [10]:

$$\begin{aligned} \min_{\mathbf{X}} \|\sigma(\mathbf{X})\|_1, \\ \text{s.t.} : \mathcal{A}(\mathbf{X}) = \mathbf{b}. \end{aligned} \quad (4)$$

Above nuclear norm minimization can be efficiently solved using some convex optimization techniques such as a semi-definite program [11]. For the purpose of matrix completion, a Singular Value Thresholding (SVT) algorithm is proposed to solve the nuclear norm minimization [12]. It is an iterative algorithm that for a fixed value of  $\tau > 0$  and positive step-sizes  $\{\delta_k\}$ , starts with  $\mathbf{Y}_0 = \mathbf{0}$  and repeats

$$\begin{cases} \mathbf{X}_k = \mathcal{D}(\mathbf{Y}_{k-1}, \tau) \\ \mathbf{Y}_k = \mathbf{Y}_{k-1} + \delta_k \mathcal{A}(\mathbf{X}_* - \mathbf{X}_k) \end{cases} \quad (5)$$

until a stopping criterion is satisfied.  $\mathcal{D}(\cdot, \tau)$  is called the singular value shrinkage operator. For a matrix  $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2}$  with Singular Value Decomposition (SVD) of  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ , we have

$$\mathcal{D}(\mathbf{A}, \tau) = \mathbf{U} \text{diag}(\{\max(\sigma_i - \tau, 0)\}_{i=1}^{\min(n_1, n_2)}) \mathbf{V}^T. \quad (6)$$

In the second approach, the  $\ell_0$ -norm used in the rank function is approximated by families of smoothing  $\ell_0$ -norm functions [13]. For example, a class of Gaussian smoothing functions are defined as:

$$f_\delta(x) = \exp\left(-\frac{x^2}{2\delta^2}\right). \quad (7)$$

As  $\delta \rightarrow 0$ ,  $f_\delta(\cdot)$  gives better approximation of the Kronecker delta function. In this approach, the rank of  $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$  is approximated as follows:

$$\text{rank}(\mathbf{X}) \approx \min(n_1, n_2) - \sum_{i=1}^{\min(n_1, n_2)} f_\delta(\sigma_i(\mathbf{X})) \quad (8)$$

and the final optimization problem is solved using the Gradient Projection (GP) algorithm [14].

In this paper, we consider the ARM problem (1) and propose an adaptive singular value thresholding method. Here, the threshold level, unlike the threshold level in the SVT algorithm is not fixed.

## II. THE PROPOSED ALGORITHM

We aim to recover a low rank matrix under some linear constraints. As stated in Section I, this problem is known as ARM. It should be noted that we don't know the rank of the primary matrix. Since the low rank matrix has sparsity in SV domain, we propose to threshold the SVs. But the main difference between our algorithm and previous ones is that the threshold level is adaptively changed during iterations. The idea of adaptive thresholding is used in [15] for recovery of the original signal from its random samples. As suggested in [15], we choose threshold level of  $k$ -th iteration as follows:

$$\tau_k = b \exp(-ak), \quad (9)$$

where  $a$  and  $b$  are two constants. The proposed algorithm is as follows:

$$\begin{cases} \mathbf{X}_k = \mathcal{T}(\mathbf{Y}_{k-1}, \tau_k) \\ \mathbf{Y}_k = \mathbf{Y}_{k-1} + \delta_k \mathcal{A}^*(\mathbf{b} - \mathcal{A}\mathbf{X}_k) \end{cases} \quad (10)$$

where  $\mathcal{A}^*(\cdot)$  is the adjoint of  $\mathcal{A}(\cdot)$ . For a matrix  $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$  with SVD  $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ ,  $\mathcal{T}(\mathbf{X}, \tau_k)$  is defined as follows:

$$\mathcal{T}(\mathbf{X}, \tau_k) \triangleq \mathbf{U}[\text{if } \sigma_{i,j} < \tau_k : 0, \text{ else: } \sigma_{i,j}]_{i=1,j=1}^{n_1,n_2} \mathbf{V}^T. \quad (11)$$

We call our algorithm Adaptive Singular Value Thresholding (ASVT) which is shown in Algorithm 1.

## III. SIMULATION RESULTS

The numerical experiment results are presented in this section. All simulations are done by MATLAB R2015a on Intel(R) Core(TM) i7-5960X @ 3GHz with 32GB-RAM.

We consider matrix completion as a special case of linear constraints. The random matrices are generated according to the scheme proposed in [12]. An  $r$ -rank matrix  $\mathbf{X}$  of size  $n_1 \times n_2$  is generated as  $\mathbf{X} = \mathbf{A}\mathbf{B}$  where  $\mathbf{A} \in \mathbb{R}^{n_1 \times r}$  and  $\mathbf{B} \in \mathbb{R}^{r \times n_2}$  are sampled independently from a standard Gaussian distribution ( $\mathcal{N}(0,1)$ ). The measurements or the subset of observed entries  $\Omega$  is sampled uniformly at random of size  $m$ . As [12], we choose constant step-sizes as  $\delta = 1$ .

To evaluate the performance of the algorithms, we report the relative error of the reconstruction defined as follows:

$$RE(\mathbf{X}, \hat{\mathbf{X}}) = \frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_F}{\|\mathbf{X}\|_F} \quad (12)$$

where  $\mathbf{X}$  and  $\hat{\mathbf{X}}$  are original and reconstructed matrices, respectively.

### A. Comparison with the SVT Algorithm

We generate matrices of size  $n_1 = n_2 = 100, 300, 500, 1000, \text{ and } 2000$  with different ranks and different fraction of observations  $\left(\frac{m}{n_1 n_2}\right)$  and provide a comparison between our algorithm with the SVT algorithm proposed in [12]. We have downloaded the SVT MATLAB codes from <http://svt.stanford.edu/code.html>.

According to the results of Table I, our proposed method has better performance in terms of relative error even with

---

### Algorithm 1 Adaptive Singular Value Thresholding: ASVT

---

```

1: input:
2:   Measurement vector:  $\mathbf{b} \in \mathbb{R}^m$ 
3:   Linear mapping:  $\mathcal{A}$ 
4:   Maximum number of iterations:  $K$ 
5:   Positive step-sizes:  $\{\delta_k\}_{k=1}^K$ 
6:   Two constants for thresholding:  $a, b$ 
7:   Termination tolerance:  $\epsilon$ 
8: initialization:
9:    $k \leftarrow 1$ 
10:   $e \leftarrow \infty$ 
11:   $\mathbf{X}_0 \leftarrow \mathbf{0}$ 
12:   $\mathbf{Y} \leftarrow \mathbf{0}$ 
13: while  $e > \epsilon$  &  $k < K$  do
14:    $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{svd}(\mathbf{Y})$ 
15:    $\tau_k = b \exp(-ak)$ 
16:    $\mathbf{\Sigma}(\mathbf{\Sigma} < \tau_k) = 0$ 
17:    $\mathbf{X}_k = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ 
18:    $\mathbf{Y} = \mathbf{Y} + \delta_k \mathcal{A}^*(\mathbf{b} - \mathcal{A}\mathbf{X}_k)$ 
19:    $e = \|\mathbf{X}_k - \mathbf{X}_{k-1}\|_F$ 
20:    $k \leftarrow k + 1$ 
21: end while
22:    $\hat{\mathbf{X}} \leftarrow \mathbf{X}_{k-1}$ 
23: return  $\hat{\mathbf{X}}$ 

```

---

a fewer number of iterations. For example, when the size of the matrix is  $3000 \times 3000$  and its rank is 100 and only 0.3 of its entries are observed, our algorithm recovers it after 29 iterations while the SVT algorithm does this after 65 iterations. Moreover, the relative error of our algorithm and the SVT algorithm is  $9.29 \times 10^{-4}$  and  $1.53 \times 10^{-3}$ , respectively.

### B. Effect of Parameters

In this Subsection, we investigate the effect of damping factor in thresholding operator ( $\alpha$ ) and step size ( $\delta$ ). For this simulation, we generate  $300 \times 300$  random matrices of rank 10 where only 1/3 of their entries are observed. We test the proposed method with different values of  $\alpha$  and  $\delta$ . The simulation results are shown in Figs. 1 and 2. According to the results of Figs. 1 and 2, the run time of the algorithm decrease as  $\alpha$  and  $\delta$  increases.

### C. Phase Transition Plot

To better show the performance of the proposed method, we plot the phase transition between  $\frac{m}{n_1 n_2}$  and  $\frac{d_r}{m}$  where  $d_r = r(n_1 + n_2 - r)$  is the degree of freedom.

TABLE I  
COMPARISON OF THE PROPOSED ALGORITHM AND THE SVT ALGORITHM.

Size	Rank	$\frac{m}{n_1 n_2}$	ASVT		SVT [12]	
			#iters	RE	#iters	RE
$500 \times 500$	10	0.15	75	9.60e-4	200	5.02e-3
	50	0.4	64	9.62e-4	200	4.38e-2
	100	0.5	138	9.64e-4	200	3.28e-1
$1000 \times 1000$	10	0.15	38	9.22e-4	62	1.49e-3
	50	0.4	29	9.21e-4	70	1.60e-3
	100	0.5	37	9.88e-4	103	1.89e-3
$1500 \times 1500$	10	0.15	23	8.17e-4	49	1.28e-3
	50	0.4	22	8.69e-4	53	1.32e-3
	100	0.5	26	8.79e-4	68	1.51e-3
$2000 \times 2000$	10	0.1	49	9.24e-4	54	1.31e-3
	50	0.3	27	8.05e-4	56	1.35e-3
	100	0.4	28	9.98e-4	68	1.53e-3
$2500 \times 2500$	10	0.1	38	9.60e-4	47	1.65e-3
	50	0.25	33	8.36e-4	56	1.37e-3
	100	0.3	37	8.96e-4	77	1.61e-3
$3000 \times 3000$	10	0.05	29	9.29e-4	65	1.53e-3
	50	0.25	29	9.33e-4	50	1.33e-3
	100	0.3	32	8.37e-4	66	1.51e-3

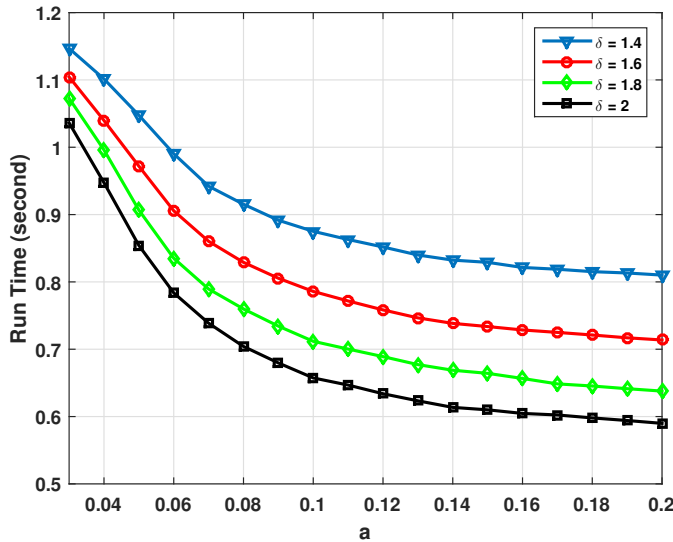


Fig. 1. Effect of damping factor  $\alpha$ .

In this simulation, we generate random matrix  $M$  of size  $80 \times 80$ . Each experiment is repeated 100 times and a matrix is declared to be successfully recovered if the relative error between the recovered matrix and the original one is less than  $10^{-3}$ . The phase transition is charted in Fig. 3. The gray color of each cell indicates the probability of success recovery during simulations. For example, the white color indicates that the proposed algorithm can successfully recover the original matrix while the black indicates that it fails.

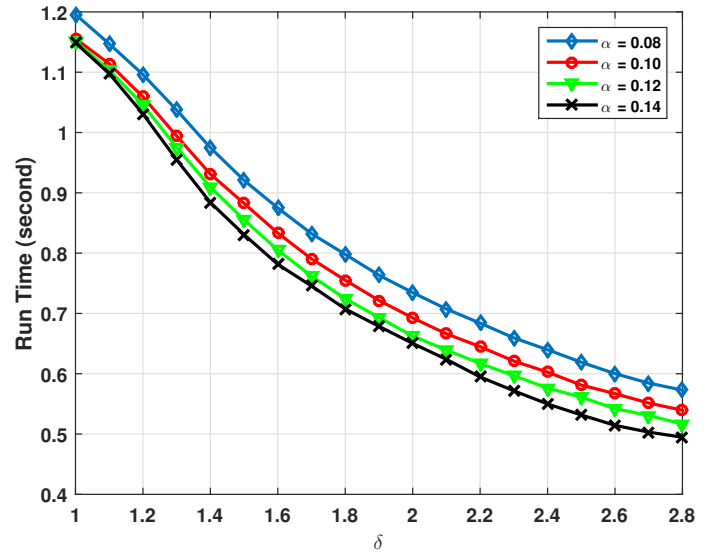


Fig. 2. Effect of step size  $\delta$ .

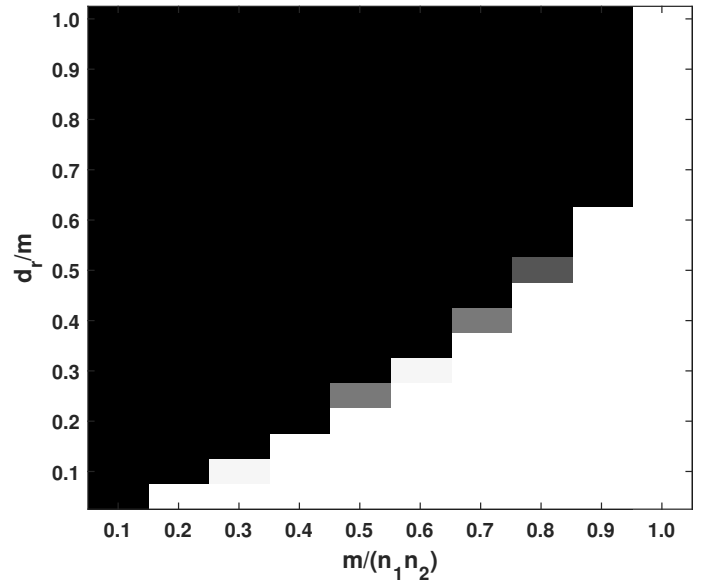


Fig. 3. Phase transition plot between  $d_r/m$  and  $m/(n_1 n_2)$ .

#### IV. CONCLUSION

In this paper, we proposed an adaptive singular value thresholding for recovery of low rank matrices under affine constraints. The word adaptive means that the threshold level is adaptively changed during the iterations. Recently, the SVT algorithm is proposed for the matrix completion which is an important case of low rank matrix recovery under affine constraints. We also compared our algorithm with the SVT algorithm and the simulation results show that we can achieve improvement with adaptive thresholding. We show that the relative error of the proposed method is less than the relative error of the SVT algorithm. Moreover, our algorithm does this task in a fewer number of iterations.

## REFERENCES

- [1] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the American Control Conference*, 2001.
- [2] M. Mesbahi, "On the semi-definite programming solution of the least order dynamic output feedback synthesis," in *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*, vol. 4, 1999, pp. 2355–2359.
- [3] M. Mesbahi and G. P. Papavassilopoulos, "On the rank minimization problem over a positive semidefinite linear matrix inequality," *IEEE Transactions on Automatic Control*, vol. 42, no. 2, pp. 239–243, February 1997.
- [4] L. E. Ghaoui and P. Gahinet, "Rank minimization under LMI constraints: A framework for output feedback problems," in *Proceedings of the European Control Conference*, 1993.
- [5] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, June 2010.
- [6] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the 22Nd International Conference on Machine Learning*. ACM, 2005, pp. 713–719.
- [7] Y. Amit, M. Fink, N. Srebro, and S. Ullman, "Uncovering shared structures in multiclass classification," in *Proceedings of the 24th International Conference on Machine Learning*. ACM, 2007, pp. 17–24.
- [8] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Machine Learning*, vol. 73, no. 3, pp. 243–272, 2008.
- [9] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the American Control Conference, IEEE*, September 2001, pp. 4734–4739.
- [10] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Electrical Engineering Department, Stanford University, Stanford, CA, USA, 2002.
- [11] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 55, pp. 471–501, 2010.
- [12] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, March 2010.
- [13] M. Malek-Mohammadi, M. Babaie-Zadeh, A. Amini, and C. Jutten, "Recovery of low-rank matrices under affine constraints via a smoothed rank function," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 981–992, February 2014.
- [14] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
- [15] F. Marvasti, A. Amini, F. Haddadi, M. Soltanolkotabi, B. Khalaj, A. Aldroubi, S. Sanei, and J. Chambers, "A unified approach to sparse signal processing," *EURASIP Journal on Advances in Signal Processing*, 2012.