Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization

Tae-Hyun Oh[†], Yasuyuki Matsushita[‡], Yu-Wing Tai[†], In So Kweon[†]

[†]KAIST, South Korea. [‡]Microsoft Research Asia, China / Osaka University, Japan.

In this work, we propose an accurate and fast approximation method for singular value thresholding (SVT), called *fast randomized SVT* (FRSVT), where we avoid direct computation of SVD and reduce the expensive computation caused from it. Our theoretical analysis provides a stepping stone between the approximation bound of SVD and its effect to NNM via SVT. Along with the analysis, our empirical results on both quantitative and qualitative studies show our approximation rarely harms the convergence behavior of the host algorithms. We apply it and validate the efficiency of our method on various vision problems, *e.g.* subspace clustering, weather artifact removal, simultaneous multi-image alignment and rectification.

Rank minimization is a crucial regularizer to derive a low-rank solution, and is required in many mathematical models in computer vision and machine learning. As rank minimization using rank($\cdot$) is an NP-hard problem [2], the problem is typically boiled down to Nuclear Norm Minimization (NNM), by relaxing using the nuclear norm (*i.e.*, $\|\cdot\|_*$, the sum of all the singular values), which is the convex envelope of the rank function. Also, weighted nuclear norm minimization (WNNM), which can be non-convex according to weights [3, 6], is used to better approximate the rank function.

As a simple example, a nuclear norm minimization (NNM) problem is expressed as:

$$\mathbf{X}^* = \operatorname{argmin}_{\mathbf{X}} f(\mathbf{X}) + \tau \|\mathbf{X}\|_*, \qquad (1)$$

where $\mathbf{X} \in \mathbb{R}^{m \times n}$, and $\tau > 0$ is a regularization parameter (the problem can have a more complex form with multiple variables, but it can be also analyzed in a similar manner). The function $f(\mathbf{X})$ can be defined according to different applications. By utilizing first-order optimization algorithms, most of the problems related to NNM (or WNNM) can be solved by iteratively solving the simple NNM subproblem defined as the following nuclear norm plus the proximity term:

Problem (**Nuclear norm minimization**). *For $\tau \geq 0$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$,*

$$\mathbf{X}^* = \operatorname{argmin}_{\mathbf{X}} \tau \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{X} - \mathbf{A}\|_F^2, \qquad (2)$$

where the optimal $\mathbf{X}^*$ can be obtained by the singular value thresholding operator defined as:

Definition 1 (**Singular value thresholding** [1]). *The problem (2) has a closed form solution by the singular value thresholding (SVT) operator $\mathbb{S}_\tau(\cdot)$ as*

$$\mathbf{X}^* = \mathbb{S}_\tau(\mathbf{A}) = \mathbf{U}_\mathbf{A} \mathcal{S}_\tau(\Sigma_\mathbf{A}) \mathbf{V}_\mathbf{A}^\top, \qquad (3)$$

where $\mathcal{S}_\tau(x) = \operatorname{sgn}(x) \cdot \max(|x| - \tau, 0)$ is the soft shrinkage operator [4], and $\mathbf{U}_\mathbf{A} \Sigma_\mathbf{A} \mathbf{V}_\mathbf{A}^\top$ is the SVD of $\mathbf{A}$.

A similar result for WNNM can be found in [3], called WSVT. They suffer from high computational cost to compute a SVD at each iteration. Hence, we propose an accurate and fast approximation method for SVT, called FRSVT, to accelerate general NNM and WNNM methods, where we avoid direct computation of SVD. Our method is motivated by the previous study of a randomized SVD proposed by Halko *et al.* [5], and we extend the original general method in several respects for better solving the NNM and WNNM problems that we focus in this work. The key idea is to extract an approximate basis for the range of a matrix from its compressed matrix. Given the basis, we compute the partial singular values of the original matrix from a small factored matrix. While the basis approximation is the bottleneck, our method is already severalfold faster than thin SVD. By adopting a range propagation technique, we can further avoid one of the bottleneck at each iteration. These lead to our algorithm with the following bound:
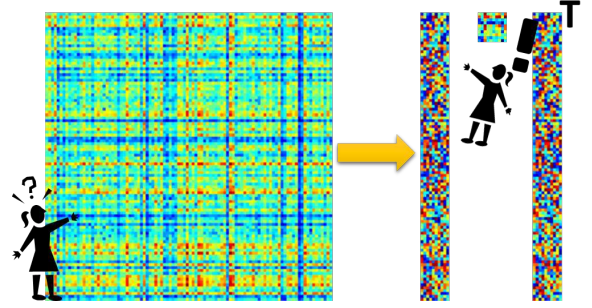
Figure 1: Illustration of the key idea. We use the approximately decomposed matrices, which are thin orthonormal matrices and a small very square core matrix. Instead of directly optimizing on the full dense matrix, we apply singular value thresholding to the very small core matrix based on the orthonormal invariant property with guarantee of Theorem 1 (see the details in the main paper).

Theorem 1 (**Average error bound of the approximate SVT**). *Let $\mathbb{S}_\tau(\cdot)$ be the SVT operator. Then, the average error satisfies the following inequality.*

$$\mathbb{E}\|\mathbb{S}_\tau(\mathbf{A}) - \mathbb{S}_\tau(\hat{\mathbf{A}}_k)\|_F^2 \leq (1 + k/(p-1)) \cdot \left(\sum_{j>k} \sigma_j^2(\mathbf{A}) \right) - G(\mathbf{A}),$$

where $G(\mathbf{A}) = \sum_{j>k} \min(\sigma_j(\mathbf{A}), \tau)^2 \geq 0$.

The proposed method is assessed using the problems of affine constrained NNM, non-convex NNM and NNM on tensor structure as well as the original NNM. The empirical evaluations showed the consistent result with the theoretical analysis, and our approach can reduce the computational time of low-rank applications without degrading accuracy and hurting convergence behavior.

References

[1] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 2010.

[2] E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 2011.

[3] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014.

[4] Elaine T. Hale, Wotao Yin, and Yin Zhang. Fixed-point continuation for l_1-minimization: Methodology and convergence. *SIAM Journal on Optimization*, 2008.

[5] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 2011.

[6] Tae-Hyun Oh, Yu-Wing Tai, Jean-Charles Bazin, Hyeongwoo Kim, and In So Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *arXiv preprint arXiv:1503.01444*, 2015.