Beyond Symmetry: Best Submatrix Selection for the Sparse SVD

Yongchun Li[1] and Weijun Xie[2]

[1] Department of Industrial & Systems Engineering, Virginia Tech, USA
liyc@vt.edu
[2] Department of Industrial & Systems Engineering, Virginia Tech, USA
wxie@vt.edu

Abstract. Truncated singular value decomposition (SVD), also known as the best low-rank matrix approximation with minimum error measured by a unitarily invariant norm, has been successfully applied to many domains such as biology, healthcare, and others, where high dimensional datasets are prevalent. To enhance the performance of dealing with high-dimensional data, sparse truncated SVD (SSVD) is used to select a few rows and columns of the original matrix along with the best low-rank approximation. Different from the literature on SSVD focusing on the top singular value or compromising the sparsity for the seek of computational efficiency, this paper presents a novel SSVD formulation that can select the best submatrix precisely up to a given size to maximize its truncated Ky Fan norm based on one or more singular values. The fact that the proposed SSVD problem is NP-hard motivates us to study effective algorithms with provable performance guarantees. To do so, we first reformulate SSVD as a mixed-integer semidefinite program, which can be solved exactly for small- or medium-sized instances within a branch and cut algorithm framework with closed-form cuts, and is extremely useful to evaluate the solution quality of approximation algorithms. We next develop three selection algorithms based on different selection criteria and two searching algorithms, greedy and local search. We prove the approximation ratios for all the approximation algorithms and show that all the ratios are tight when the selected submatrix admits the number of rows or columns no larger than that of the data matrix, i.e., our derived approximation ratios are unimprovable. Our numerical study demonstrates the high solution quality and computational efficiency of the proposed algorithms. Finally, all our analysis can be extended to row-sparse PCA.

Keywords: Sparse SVD, Branch and Cut, Approximation algorithms, and Row-sparse PCA

1 Introduction

Truncated singular value decomposition (SVD), also known as the best low-rank matrix approximation, has been successfully applied to many domains such as biology, healthcare, and others, where high-dimensional datasets are prevalent. To enhance the interpretability of the truncated SVD of a data matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, this paper proposes the Sparse SVD (SSVD) formulation

$$\text{(SSVD)} \quad w^* := \max_{S_1\subseteq[m],S_2\subseteq[n]} \left\{||\boldsymbol{A}_{S_1,S_2}||_{(k)} : |S_1| \le s_1, |S_2| \le s_2\right\}, \qquad (1)$$

where the Ky Fan k-norm $\|\cdot\|_{(k)}$ is defined as the sum of k largest singular values of a matrix, for any subsets S_1 and S_2, $\boldsymbol{A}_{S_1,S_2}$ denotes the submatrix of $\boldsymbol{A}$ with rows and columns indexed by S_1 and S_2, respectively.

The combinatorial formulation (1) leads to an intuitive explanation of SSVD from the perspective of submatrix selection, which is remarked below: for a given data matrix $\boldsymbol{A}$, the objective of SSVD (1) is to select the best submatrix of size at most $s_1 \times s_2$ with the maximum Ky Fan k-norm. The proposed SSVD (1) with submatrix selection and low-rank approximation can be applicable to many data-intensive problems such as biclustering [2, 3] and feature selection [1].

The SSVD (1) is versatile and generalizes many existing models. Particularly, we show that when matrix $\boldsymbol{A}$ is positive semidefinite and $s_1 = s_2$, SSVD (1) reduces to the well-known Sparse PCA (SPCA) problem [4]. Formally, SPCA can be defined as

$$\text{(SPCA)} \quad w^{spca} := \max_{\boldsymbol{U}\in\mathbb{R}^{n\times k}} \left\{ \operatorname{tr}(\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{U}) : \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_k, ||\boldsymbol{U}||_0 \le s\right\}, \qquad (2)$$

where $\boldsymbol{A} \in \S_+^n$ is a sample covariance matrix, $k \le s \le n$ are positive integers, and w^{spca} denotes the optimal value. SPCA (2) can be viewed as a special case of SSVD (1); thus, SSVD (1) is NP-hard with a reduction to the SPCA problem that has been notoriously known to be NP-hard and inapproximable.

1.1 Summary of Main Contributions

To solve SSVD (1), we derive an equivalent mixed integer semidefinite programming (MISDP) formulation and effective approximation algorithms using various selection and searching criteria. Below we list the major contributions in detail.

(i) Based on the MISDP, we derive a family of valid inequalities and a branch and cut algorithm for SSVD (1), which can solve the small- and medium-sized instances to optimality (e.g., $m = n = 500$, $s_1 = s_2 = 5$, $k = 2$) and help evaluate the solution quality of our proposed approximation algorithms;
(ii) Inspired by the combinatorial formulation (1) of SSVD, we consider three different selection criteria that are related to its objective function (i.e., the Ky Fan k-norm), but are much easier to compute, and then propose three selection algorithms;
(iii) We successfully tailor the greedy and local search algorithms to solve SSVD;
(iv) We derive the approximation ratios and time complexities of our proposed approximation algorithms for SSVD (1), which are displayed in Table 1. We prove that all the ratios are tight, i.e., un-improvable when the sparse parameters satisfy $s_1 \le m/2$ and $s_2 \le n/2$;

(v) We remark that the approximation ratio of the first selection algorithm is independent of k and matches the best-known one of rank-one SPCA;

(vi) All our analyses of exact and approximation algorithms can be extended to the general SPCA (2), where the results of approximation algorithms are displayed in Table 2; and

(vii) The numerical study on large-scale instances (e.g., $m = 16313, n = 2365$) shows the high solution quality and computational efficiency of our proposed approximation algorithms.

Table 1. Summary of Approximation Algorithms for SSVD (1)

	Selection Algorithms			Searching Algorithms	
Algorithm	Selection I	Selection II	Selection III	Greedy	Local Search
Ratio	$1/\sqrt{s'}$	$1/\sqrt{ks'}$	$\sqrt{s_1 s_2}/(k\sqrt{mn})$	$1/\sqrt{ks_1 s_2}$	$1/\sqrt{ks_1 s_2}$
Complexity	NP-hard	$O((m+n)(n' \log(n') + ks_1 s_2))$	$O(n' \log(n') + mn)$	$O(\max\{s_1, s_2\} (m+n)ks_1 s_2)$	$O(L/\delta ks_1 s_2 (ns_1 + ms_2))$

[1] $s' := \min\{s_1, s_2\}$, $n' = \max(m, n)$

[2] L = encoding length of $\boldsymbol{A}$, and $\delta > 0$ is the strict improvement factor

Table 2. Summary of Approximation Algorithms for SPCA

	Selection Algorithms			Searching Algorithms	
Algorithm	Selection I	Selection II	Selection III	Greedy	Local Search
Ratio	$1/\sqrt{s}$	$1/\sqrt{ks}$	$s/(kn)$	k/s	k/s
Complexity	NP-hard	$O(n(n\log(n) + ks^2))$	$O(n\log(n) + n^2)$	$O(nks^3)$	$O(L/\delta nks^3)$

[1] L = encoding length of $\boldsymbol{A}$, and $\delta > 0$ is the strict improvement factor

References

1. Pavan K Kumar, PESN Krishna Prasad, MV Ramakrishna, and BDCN Prasad. Feature extraction using sparse SVD for biometric fusion in multimodal authentication. *International Journal of Network Security & Its Applications*, 5(4):83, 2013.
2. Mihee Lee, Haipeng Shen, Jianhua Z Huang, and JS Marron. Biclustering via sparse singular value decomposition. *Biometrics*, 66(4):1087–1095, 2010.
3. Martin Sill, Sebastian Kaiser, Axel Benner, and Annette Kopp-Schneider. Robust biclustering by sparse singular value decomposition incorporating stability selection. *Bioinformatics*, 27(15):2089–2097, 2011.
4. Vincent Vu and Jing Lei. Minimax rates of estimation for sparse PCA in high dimensions. In *Artificial intelligence and statistics*, pages 1278–1286, 2012.