

# Approximation Algorithms for $\ell_0$ -Low Rank Approximation

Karl Bringmann

Pavel Kolev

David P. Woodruff

## The $\ell_0$ -rank- $k$ Problem

**Input:** semiring/field  $S$ , matrix  $A \in S^{m \times n}$ , integer  $k \in [n]$

**Output:** rank- $k$  matrix  $A^* \in S^{m \times n}$  such that

$$A^* := U \cdot V = \operatorname{argmin}_{U \in S^{m \times k}, V \in S^{k \times n}} |A - U \cdot V|_0,$$

where  $|A|_0$  is the number of nonzero entries of  $A$ .

**Def:**  $\text{OPT}_S^k := |A - A^*|_0$  and  $0 \leq \text{OPT}_S^k \leq |A|_0$

## The Robust PCA Problem

**Input:**  $A = A^* + M$ ,  $\text{rank}(A^*) = k \ll n$ ,  $M$  sparse matrix

**Goal:** recover the low rank matrix  $A^*$

[3] relaxed the  $\ell_0$ -error measure to  $\ell_1$ -norm.

It is of fundamental importance for TCS to understand the theoretical guarantees for the original  $\ell_0$ -problem.

## Example: Reals $\ell_0$ -rank-1

**Hard instance** for algorithms that select a column:

$$A = \begin{bmatrix} 2 & 1 & \vdots & 1 \\ 1 & 2 & \vdots & 1 \\ \dots & \dots & \ddots & \dots \\ 1 & 1 & \vdots & 2 \end{bmatrix}_{n \times n} \quad u^* = v^* = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_n \quad \text{OPT}_F^1 = n$$

For any column  $A_{:,i}$  the best response vector is  $\mathbf{1}$ , so

$$|A_{:,i} \mathbf{1}^T - A|_0 = 2(n-1) = 2(1 - 1/n) \text{OPT}_F^1$$

**2-approximation scheme** [4] –  $O(|A|_0(m+n))$  time

For every column  $A_{:,i}$  compute *best response vector*  $v := \operatorname{argmin} |A_{:,i} v^T - A|_0$ . Return the best pair  $(A_{:,i}, v)$ .

## Reals $\ell_0$ -rank-1

**Theorem 1. (Sublinear)** Given  $A \in R^{m \times n}$  with column adjacency arrays,  $\epsilon \in (0, 1/10)$  and  $\text{OPT}_R^1 \geq 1$ , we can compute w.h.p. in time

$$O\left(\left(\frac{n \log m}{\epsilon^2} + \min\left\{|A|_0, n + \psi_R^{-1} \frac{\log n}{\epsilon^2}\right\}\right) \frac{\log^2 n}{\epsilon^2}\right)$$

a column  $A_{:,j}$  and a vector  $z$  such that

w.h.p.  $|A - A_{:,j} z^T|_0 \leq (2 + \epsilon) \text{OPT}_R^1$ .

**Def:**  $\psi_R := \text{OPT}_R^1 / |A|_0$ ,  $0 \leq \psi_R \leq 1$ , **sublinear**  $o(|A|_0)$ .

## Reals $\ell_0$ -rank- $k$

**Theorem 2. (Bicriteria Algorithm)** Given  $A \in R^{m \times n}$  and  $k \in [n]$  we can compute in expected time  $\text{poly}(m, n)$  a subset of columns  $A_{:,J}$  of size  $|J| = O(k \log(n/k))$  and a matrix  $Z \in R^{|J| \times n}$  such that  $|A - A_{:,J} Z|_0 \leq O(k^2 \log(n/k)) \text{OPT}_R^k$ .

**Lemma 1. (Structural)** For any  $A \in R^{m \times n}$  and  $k \in [n]$ , there is a subset of columns  $A_{:,J}$  of size  $k$  and a matrix  $Z \in R^{k \times n}$  such that  $|A - A_{:,J} Z|_0 \leq (k+1) \text{OPT}_R^k$ .

**Lemma 2. (Lower Bound)** Any algorithm that selects  $k$  columns of  $A$  incurs at least an  $\Omega(k)$ -approximation.

## Binary $\ell_0$ -rank-1

**Theorem 3. (Sublinear)** Given  $A \in \{0,1\}^{m \times n}$  with column adjacency arrays and with row and column sums, we can compute w.h.p. in time  $O(\min\{|A|_0 + m + n, \psi_B^{-1}(m+n)\} \log^3(mn))$  vectors  $u, v$  such that  $|A - uv^T|_0 \leq (1 + O(\psi_B)) \text{OPT}_B^1$ .

**Theorem 4. (Exact)** Given  $A \in \{0,1\}^{m \times n}$  with  $\text{OPT}_B^1 / |A|_0 \leq 1/300$ , we can solve exactly the Binary  $\ell_0$ -rank-1 problem in time  $2^{O(\text{OPT}_B^1 / \sqrt{|A|_0})} \text{poly}(mn)$ .

**Theorem 5. (Lower Bound on Sample Complexity)** Given  $A \in \{0,1\}^{n \times n}$  as in Theorem 3, and  $\sqrt{\log n / n} \ll \phi < 1/100$  such that  $\psi_B \leq \phi$ , computing a  $(1 + \phi)$ -approximation to  $\text{OPT}_B^1$  requires to read at least  $\Omega(n/\phi)$  entries of  $A$ .

## Proof Techniques (Theorem 2)

A subroutine from [1] that computes in time  $O(m^2 k^{\omega+1})$  a vector  $z$  such that  $|Az - b|_0 \leq k \cdot \min |Ax - b|_0$ .

We extend a result from [2] using **Lemma 1**, to show that w.p.  $1/3$  a set of  $2k$  columns  $A_{:,Q}$  selected u.a.r., yields a column set  $C \subset [n] \setminus Q$  of size  $\Omega(n)$  such that

$$\min_x |A_{:,Q} x - A_{:,i}|_0 \leq O(k+1) \text{OPT}_R^k / n, \quad \forall i \in C.$$

## References

- [\*] Karl Bringmann, Pavel Kolev, and David P. Woodruff, Approximation algorithms for  $\ell_0$ -low rank approximation, in Advances in Neural Information Processing Systems 30 (NIPS), 2017
- [1] Noga Alon, Rina Panigrahy, and Sergey Yekhanin. Deterministic approximation algorithms for the nearest codeword problem. In 12th International Workshop, APPROX-RANDOM 2009, USA
- [2] Flavio Chierichetti, Sreenivas Gollapudi, Ravi Kumar, Silvio Lattanzi, Rina Panigrahy, and David P. Woodruff. Algorithms for  $\ell_p$  low-rank approximation. In ICML 2017, Sydney, Australia
- [3] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal Component analysis? Journal of the ACM (JACM), 58(3):11, 2011.
- [4] Peng Jiang, Jiming Peng, Michael Heath, and Rui Yang. A clustering approach to constrained binary matrix factorization. In Data Mining and Knowledge Discovery for Big Data, Springer, 2014.