Approximation Algorithms for ℓ_0-Low Rank Approximation

Karl Bringmann   Pavel Kolev   David P. Woodruff

The ℓ_0-rank-k Problem

Input: a field F, a matrix $A \in F^{m\times n}$, an integer $k \in [n]$

Output: a matrix $A^\star \in F^{m \times n}$ such that

$$A^\star := \underset{\operatorname{rank}(A')=k}{\operatorname{argmin}} |A - A'|_0 = \underset{U\in F^{m\times k}, V\in F^{k\times n}}{\operatorname{argmin}} |A - UV|_0,$$

where $|A|_0 \coloneqq \operatorname{nnz}(A)$ is the number of nonzero entries.

Def: $\mathrm{OPT}_F^k \coloneqq |A - A^\star|_0$ and $0 \le \mathrm{OPT}_F^k \le |A|_0$

The Robust PCA Problem

Promise: $A = A^\star + S$, $\operatorname{rank}(A^\star) = k \ll n$, S is sparse.

Goal: recover the low rank matrix $A^\star$

[3] relaxed the ℓ_0-error measure to ℓ_1-norm.

It is of fundamental importance for TCS to understand the theoretical guarantees for the original ℓ_0-problem.

Example: Reals ℓ_0-rank-1

Hard instance for algorithms that select a column:

$$A = \begin{bmatrix} 2 & 1 & \vdots & 1 \\ 1 & 2 & \vdots & 1 \\ \cdots & \cdots & \ddots & \cdots \\ 1 & 1 & \vdots & 2 \end{bmatrix}_{n\times n} \qquad u^\star = v^\star = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_n \qquad \mathrm{OPT}_F^1 = n$$

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

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

Reals ℓ_0-rank-1

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

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

Theorem 1. (Sublinear) Given $A \in R^{m\times n}$ with column adjacency arrays, $\epsilon \in (0, 1/10)$ and $\mathrm{OPT}_R^1 \ge 1$, 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. $\left|A - A_{:,j}z^T\right|_0 \le (2+\epsilon)\,\mathrm{OPT}_R^1$.

Def: $\psi_F \coloneqq \mathrm{OPT}_F^1/|A|_0,\ 0 \le \psi_F \le 1,$ Sublinear $o(|A|_0)$.

Boolean ℓ_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 $|A - uv^T|_0 \le \left(1 + O(\psi_B)\right)\mathrm{OPT}_B^1$.

Theorem 4. (Exact) Given $A \in \{0,1\}^{m\times n}$ with $\mathrm{OPT}_B^1/|A|_0 \le 1/300$, we can solve exactly the Boolean ℓ_0-rank-1 problem in time $2^{O\left(\mathrm{OPT}_B^1/\sqrt{|A|_0}\right)}\operatorname{poly}(mn)$.

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 $\psi_B \le \phi$, computing a $(1+\phi)$-approximation to OPT_B^1 requires to read at least $\Omega(n/\phi)$ entries of A.

Reals ℓ_0-rank-k

Theorem 2. (Bicriteria Algorithm) Given $A \in R^{m\times n}$ and $k \in [n]$ we can compute in expected time $\operatorname{poly}(m,n)$ 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 $\left|A - A_{:,J}Z\right|_0 \le O(k^2\log(n/k))\,\mathrm{OPT}_R^k$.

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 $\left|A - A_{:,J}Z\right|_0 \le (k+1)\,\mathrm{OPT}_R^k$.

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

Proof Techniques (Theorem 2)

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

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 $C \subset [n]\backslash Q$ of size $\Omega(n)$ such that

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

References

[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 ℓ_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.

max planck institut informatik

Carnegie Mellon University
School of Computer Science