Streaming Algorithms for ℓ_p Flows and ℓ_p Regression

Amit Chakrabarti, Jeffrey Jiang

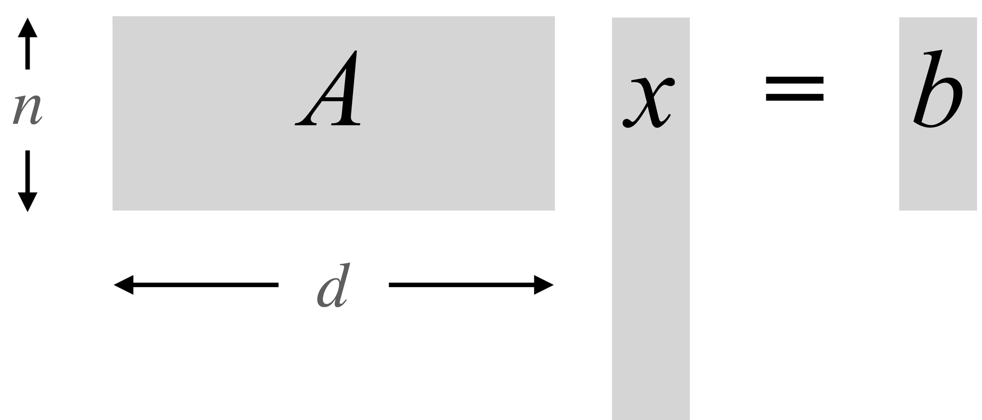
Dartmouth College

David P. Woodruff, Taisuke Yasuda

Carnegie Mellon University

Introduction and Problem Studied

Underdetermined linear systems. We wish to solve $\mathbf{Ax} = \mathbf{b}$ for an $n \times d$ matrix $\mathbf{A}$ and an n-dimensional vector $\mathbf{b}$ when $d \gg n$



- For unique solutions, consider the ℓ_p regression problem

$$\min \|\mathbf{x}\|_p^p \quad \text{s.t. } \mathbf{Ax} = \mathbf{b} \qquad \text{(P1)}$$

- Captures many well-studied problems
  - $p = 1$: basis pursuit, sparse recovery
  - $p = 2$: least squares solution to underdetermined systems
  - p-norm flow problem: $\mathbf{A}$ = incidence matrix of a graph and $\mathbf{b}$ = vector of demands (when $p = \infty$: max flow)
- We study two types of guarantees
  - Estimating the minimum cost of (P1) up to distortion κ
  - Outputting a good solution $\hat{\mathbf{x}}$ with distortion κ

Streaming algorithms

- Motivation: very large d, so working memory holds only one or a few columns of $\mathbf{A}$ at once
- **Column-arrival streaming model.** Algorithm receives the d columns of $\mathbf{A}$ and the vector $\mathbf{b}$ in an arbitrary order
  - Focus: algorithms that make one pass over data stream
  - If $\mathbf{A}$ = incidence matrix: edge insertion graph stream
  - Primary goal: minimizing the space used by the algorithm

Question. What is the space complexity of underdetermined ℓ_p linear regression in the one-pass column-arrival streaming model?

Estimating the Minimum Cost

Results. Approximating the cost of ℓ_p regression

Setting	Distortion	Space Complexity (bits)
$p = 2$	1	$\tilde{O}(n^2)$ (folklore)
$p \in (2,\infty]$	$(1+\varepsilon)$	$\tilde{O}(\varepsilon^{-2}n^2)$
$p \in (1,2)$	$(1+\varepsilon)$	$\tilde{O}(\varepsilon^{-2}n^{q/2+1})$
$p = 2$	$(1+\varepsilon)$	$\Omega(n^2)$
$p = 1$	$o(n^{1/2})$	$n^{\omega(1)}$
$p = 0$	2	$\Omega(d)$

Note: $q = p/(p-1)$, the Hölder conjugate exponent

Upper bound techniques

- We design streaming algorithms for constructing a **flow sparsifier**, i.e. a weighted subset of columns (edges) whose optimal value κ-approximates the optimal value of the original problem.
- First: reduce the problem to the construction of ℓ_q subspace embeddings via a duality lemma

Duality Lemma. If $\mathbf{Ax} = \mathbf{b}$ is feasible, then

$$\min_{\mathbf{Ax}=\mathbf{b}} \|\mathbf{x}\|_p = \max_{\|\mathbf{A}^\top\mathbf{y}\|_q \le 1} \mathbf{y}^\top\mathbf{b}$$

- Next: apply known streaming algorithms for constructing ℓ_q subspace embeddings

Theorem. There is a column-arrival streaming algorithm that computes a weighted subset of columns $\mathbf{AS}$ of $\mathbf{A}$ such that

$$\|\mathbf{S}^\top\mathbf{A}^\top\mathbf{y}\|_q = (1 \pm \varepsilon)\,\|\mathbf{A}^\top\mathbf{y}\|_q \text{ for all } \mathbf{y} \in \mathbb{R}^n,$$

using $\tilde{O}(\varepsilon^{-2}n^{\max\{1,q/2\}+1})$ bits of space.

- Key technique: online Lewis weight sampling [Cohen, Musco, Pachocki '16] [Woodruff, Yasuda '23]

Estimating the Minimum Cost (cont.)

Lower bound techniques (p = 1)

- Hard instance: the columns of $\mathbf{A}$ are n^D random vectors
- Two cases: $\mathbf{b}$ is a column of $\mathbf{A}$ or another random vector
  - Objective values for the two cases differ by factor of $\sqrt{n}$
  - Reduce from INDEX problem (communication complexity): distinguishing the two cases requires $\Omega(n^D)$ bits of space

Outputting a Good Solution

Results. Outputting $\hat{\mathbf{x}}$ to approximately minimize ℓ_p-norm

Setting	Distortion	Space Complexity (bits)
$p \in (1,\infty]$	$(1+\varepsilon)$	$\Omega(d)$
$p \in (1,\infty]$	β	$\tilde{\Omega}(d/\beta^{2q})$
$p \in (2,\infty]$	β	$n^2 \cdot \tilde{O}(d/\beta^q)$
$p \in (1,2)$	$n^{1/p-1/2}\beta$	$n^2 \cdot \tilde{O}(d/\beta^q)$
$p = 1$	$n^{1/2}$	$n^2 \cdot O(\operatorname{poly}\log d)$

Observations

- For $p > 1$, if we wish to output a $(1+\varepsilon)$-approximate solution, there are no algorithms using space sublinear in d.
- Reduce space by $\operatorname{poly}(d)$ factors, sacrificing $\operatorname{poly}(d)$ distortion
- For $p = 1$, there is an algorithm with $\sqrt{n}$ distortion using only $O(n^2 \cdot \operatorname{poly}\log d)$ bits of space

Upper bound techniques. Use streaming algorithms for constructing a well-conditioned subset of columns, which are essentially unweighted flow sparsifiers. Then a solution for the flow sparsifier is a valid solution to the original problem.

Lower bound techniques. Information theoretic argument: if $\mathbf{A}$ is random, then a good approximation $\hat{\mathbf{x}}$ must have a high mutual information with $\mathbf{A}$