CS 15-851: Algorithms for Big Data	Spring 2025

Lecture 7 — Feb 27

Prof. David Woodruff | *Scribe: Yueqi Song*

1 $\ell_1 Regression$

Overview of the Algorithm Idea: We need to compute $poly(d)-$approximation and compute well-conditioned basis. Then we could sample rows from the well-conditioned basis and residual of the $poly(d)-$approximation. Then, we could solve the l_1 regression on the sample, obtaining vector x and output x.

1.1

First, we want to compute $poly(d)$- approximation.

To do so, we need to find x' such that

$$|Ax' - b|_1 \leq poly(d) \min_{x \in R^d} |Ax - b|_1$$

Let $b' = b - Ax'$ be the residual.

1.2

Note that

$$\min_{x \in R^d} |Ax - b|_1 = \min_{x \in R^d} |Ux - b'|_1$$

We want to sample $poly(d/\epsilon)$ rows of $U \circ b'$ proportional to their l_1 -norm.

1.3

Next, we want to compute well-conditioned basis by finding a basis $A = UW$ so that for all $x \in R^d$,

$$|x|_1/poly(d) \leq |Ux|_1 \leq |poly(d)x|_1$$

1.4

Then, we sample rows from the well-conditioned basis and the residual of $poly(d)$ approximation. Now, generic linear programming is efficient.

1.5

Lastly, we solve l_1- regression on the sample, obtaining vector x, and output x.

2 Sketching Theorem

2.1 Theorem

There is a probability space over $(d \log d) * n$ matrices R such that for any $n * d$ matrix A, with probability at least 99/100 we have for all x:

$$|Ax|_1 \leq |RAx|_1 \leq (d \log d) * |Ax|_1$$

A dense R that works: The entries of R are i.i.d. Cauchy random variables, scaled by $1/(d \log d)$

2.2 Embedding

- is linear
- is independent of A
- preserves lengths of an infinite number of vectors

2.3 Application of Sketching Theorem

2.3.1 Computing a $d(\log d)$-approximation

- Compute RA and Rb
- Solve $x' = argmin_x |RAx - Rb|_1$
- Main theorem applied to $A \circ b$ implies x' is a $d \log d$-approximation
- RA, Rb have $d \log d$ rows, so can solve l_1 -regression efficiently

2.3.2 Computing a well-conditioned basis

- Compute RA
- Compute W so that RAW is orthonormal (in the l_2 -sense)
- Output $U = AW$

Note: $U = AW$ **is well-conditioned** because

$$|AWx|_1 \leq |RAWx|_1 \leq (d\log d)^{\frac{1}{2}} * |RAWx|_2 = (d\log d)^{\frac{1}{2}} * |x|_2 \leq (d\log d)^{\frac{1}{2}} * |x|_1$$

and

$$|AWx|_1 \geq |RAWx|_1/(d\log d) \geq |RAWx|_2/(d\log d) = |x|_2/(d\log d) \geq |x|_1/(d^{\frac{3}{2}}\log d)$$

2.4 Cauchy Random Variables

2.4.1 Properties

- $pdf(z) = 1/(\pi(1+z^2))$ for z in $(-\infty, \infty)$
- Undefined expectation
- Infinite variance.
- 1-stable
- Can generate as the ratio of two standard normal random variables

1-stable: If $z_1, z_2, \cdots, z_n$ are i.i.d. Cauchy, then for $a \in R^n$,

$$a_1 * z_1 + a_2 * z_2 + \cdots + a_n * z_n = |a|_1 * z$$

where z is Cauchy.

2.5 Proof of Sketching Theorem

By 1-stability, For all rows r of R,

$$\langle r, Ax\rangle = |Ax|_1 * Z/(d\log d)$$

where Z is a Cauchy

$$RAx = (|Ax|_1 * Z_1, \cdots, |Ax|_1 * Z_{d\log d})/(d\log d)$$

where $Z_1, \cdots, Z_{d\log d}$ are i.i.d. Cauchy.

$$|RAx|_1 = |Ax|_1 \sum_j |Z_j|/(d\log d)$$

The $|Z_j|$s are half-Cauchy.

THen, we have

$$\sum_j |Z_j| = \Omega(d\log d) \text{ with probability } 1 - e^{-d\log d} \text{ by Chernoff}$$

But the $|Z_j|$s are heavy-tailed, so $\sum_j Z_j$ is heavy-tailed, so

$$|RAx|_1 = |Ax|_1 \sum_j |Z_j|/(d \log d)$$

may be large. Each $|Z_j|$ has c.d.f. asymptotic to $1 - \theta(1/z)$ for $z \in [0, \infty)$

Note that there exists a well-conditioned basis of A, Suppose w.l.o.g. the basis vectors are $A_{*1}, \cdots, A_{*d}$, then

$$|RA_{*i}|_1 = |A_{*i}|_1 * \sum_j |Z_{i,j}|/(d \log d)$$

Let $\mathbb{E}_{i,j}$ be the event that $|Z_{i,j}| \leq d^3$; Define $Z'_{i,j} = |Z_{i,j}|$ if $|Z_{i,j}| \leq d^3$, and $Z'_{i,j} = d^3$ otherwise. Then, we have

$$\mathbb{E}[Z_{i,j}|\mathbb{E}_{i,j}] = \mathbb{E}[Z'_{i,j}|\mathbb{E}_{i,j}] = O(\log d)$$

Let E be the event that for all i,j, $E_{i,j}$ occurs. Then,

$$\mathbb{P}(E) \geq 1 - \frac{\log d}{d}$$