CS 15-851: Algorithms for Big Data Spring 2025

Lecture Lecture 6, Part 2 — 2/20/2025

Prof. David Woodruff *Scribe: Yixuan Xu*

1 Distributed Problems Beyond Low Rank Approximation

In the preivous part of the lecture, we have covered the distributed low rank approximation problem. The problem is optimally solved by the BWZ protocol. However, in general, distributed problems with a similar flavor are not as well understood. We will list some of the open problems below.

Rank computation. There are s servers, each with a real matrix $A_i \in \mathbb{R}^{n\times n}$, and the goal is to compute the rank of the sum of the matrices $\sum_{i=1}^{s} A_i$, minimizing the communication cost.

Linear programming. There are s servers, each with a real matrix $A_i \in \mathbb{R}^{n\times d}$ and a vector $b_i \in \mathbb{R}^n$. A_i and b_i stand for a set of linear constraints. The goal is to solve the linear program $\min\{c^T x : A_i x \le b_i, i = 1, \ldots, s\}$, minimizing the communication cost.

Maximum matching. There are s servers, each with a set of edges E_i in a graph $G = (V, E)$. The goal is to compute the maximum matching when $E = \bigcup_{i=1}^{s} E_i$, minimizing the communication cost.

2 Robust Regression

In the following section, we will introduce the robust regression problem. That is, give a matrix A and a vector b, we want to find a vector x that minimizes the objective function

$$\|Ax - b\|_1 = \sum_{i=1}^{n} |\langle A_i, x\rangle - b_i|, \text{ where } A_i \text{ is the } i\text{-th row of } A.$$

2.1 The Linear Programming Solution

We do not have a good closed-form solution for the robust regression problem. However, we can still solve the problem exactly by formulating it as a linear program. The linear program is as follows:

$$\begin{array}{r|l} \textbf{Minimize} & (1,1,1,\ldots,1)\cdot(\alpha^+ + \alpha^-) \\ \textbf{Subject to} & Ax + \alpha^+ - \alpha^- = b \\ & \alpha^+, \alpha^- \in \mathbb{R}^d_{\ge 0} \\ & x \in \mathbb{R}^d \end{array}$$

One key observation is for each i, at least one of α_i^+ and α_i^- is zero. This is because if $\alpha_i^+ > 0$ and $\alpha_i^- > 0$, we can decrease both α_i^+ and α_i^- by the same amount and the objective function will decrease. Therefore, the optimal solution will have the objective function equal to $\|Ax - b\|_1$.

Thus, we can see that the problem can be solved by in poly(nd) time using linear programming.

2.2 Well-Conditioned Bases

The poly(nd) time complexity of the linear programming solution is not satisfactory in the context of big data. To develop a faster algorithm, we first introduce the concept of well-conditioned bases.

Recall that in the ℓ_2 regression problem, we can use the SVD of A to decompose $A = UW$, where U has orthonormal columns, so that $\|Ux\|_2 = \|x\|_2$ for all $x \in \mathbb{R}^d$. For the ℓ_1 regression problem, we would like to have a similar property. That is, we would like to have a matrix U such that

$$A = UW \text{ and } \|Ux\|_1 \approx \|x\|_1 \text{ for all } x \in \mathbb{R}^d.$$

To do this, we first write $A = QW$, where Q is a matrix with full column rank. Note that the number of columns k of Q is at most d. Below, we define the $(Q, 1)$-norm of a vector x.

Definition 2.1. *Let $Q \in \mathbb{R}^{n\times k}$ be a matrix with full column rank. For any $x \in \mathbb{R}^k$, we define*

$$\|x\|_{Q,1} = \|Qx\|_1 .$$

Lemma 2.2. *$\|\cdot\|_{Q,1}$ is a norm on $\mathbb{R}^k$.*

Proof. To show that $\|\cdot\|_{Q,1}$ is a norm, we need to show that it satisfies the three properties.

For any $x, y \in \mathbb{R}^k$, we have

$$\|x + y\|_{Q,1} = \|Q(x + y)\|_1 = \|Qx + Qy\|_1 \leq \|Qx\|_1 + \|Qy\|_1 = \|x\|_{Q,1} + \|y\|_{Q,1} .$$

For any $x \in \mathbb{R}^k$ and $c \in \mathbb{R}$, we have

$$\|cx\|_{Q,1} = \|Q(cx)\|_1 = \|cQx\|_1 = |c| \|Qx\|_1 = |c| \|x\|_{Q,1} .$$

Finally, for any $x \in \mathbb{R}^k$, we have $\|x\|_{Q,1} = \|Qx\|_1 = 0$ if and only if $Qx = 0$ if and only if $x = 0$.

Therefore, $\|\cdot\|_{Q,1}$ is a norm on $\mathbb{R}^k$. ■

For the ease of notation, we assume $k = d$ in the following discussion.

Let $C = \{z \in \mathbb{R}^d \mid \|z\|_{Q,1} \leq 1\}$ be the unit ball of the $(Q, 1)$-norm. By the triangle inequality and absolute homogeneity of the $(Q, 1)$-norm, we know that C is a convex set. Moreover, C is symmetric about the origin. For this symmetric convex body, we can apply the Lowner-John ellipsoid theorem.

Theorem 2.3. *Let C be a symmetric convex body in $\mathbb{R}^d$. Then there exists an ellipsoid E such that*

$$E \subseteq C \subseteq \sqrt{d}E.$$

Algebraically, there exists some $G \in \mathbb{R}^{d\times d}$ such that $E = \{z \in \mathbb{R}^d \mid z^\top G^\top Gz \leq 1\}$.

Corollary 2.4. *There exists a matrix $G \in \mathbb{R}^{d\times d}$ such that for all $z \in \mathbb{R}^d$, we have*

$$\sqrt{z^\top G^\top Gz} \leq \|z\|_{Q,1} \leq \sqrt{d} \cdot \sqrt{z^\top G^\top Gz}.$$

Define $U = QG^{-1}$. We will show that U is a well-conditioned basis such that $\|Ux\|_1 \approx \|x\|_1$ for all $x \in \mathbb{R}^d$. Let $z = G^{-1}x$. Then, we can compute the followings:

$$\|Ux\|_1 = \left\|QG^{-1}x\right\|_1 = \|Qz\|_1 = \|z\|_{Q,1}\,.$$
$$z^\top G^\top G z = x^\top (G^{-1})^\top G^\top G (G^{-1}) x = x^\top x = \|x\|_2^2\,.$$

Combining the above two equations with Corollary 2.4, we have

$$\|x\|_2 \le \|Ux\|_1 \le \sqrt{d} \cdot \|x\|_2\,.$$

To conclude the discussion, we will use the following relationship between the ℓ_1 and ℓ_2 norms.

Fact 2.5. *Let $x \in \mathbb{R}^d$. Then, $\|x\|_2 \le \|x\|_1 \le \sqrt{d} \cdot \|x\|_2$.*

Therefore, we can see that

$$\frac{\|x\|_1}{\sqrt{d}} \le \|x\|_2 \le \|Ux\|_1 \le \sqrt{d}\,\|x\|_2 \le \sqrt{d}\,\|x\|_1\,.$$

Here, we have shown that U perserves the ℓ_1 norm up to a factor of $\sqrt{d}$.

2.3 Net Argument

Previously, in the ℓ_2 regression problem, given a sketching matrix S, we used a net argument to show that the sketching matrix S preserves the ℓ_2 norm for a finite set of vectors, and in general, each vector is close to, and thus can be approximated by some vector in the net. We would like to obtain a similar result for the ℓ_1 regression problem in this section.

Net for the Ball. Let $B = \{x \in \mathbb{R}^d \mid \|x\|_1 = 1\}$ be the unit ball of the ℓ_1 norm. We would like to find a γ-net N for B with respect to the ℓ_1 norm. That is, for any $x \in B$, there exists $y \in N$ such that $\|x - y\|_1 \le \gamma$. To do this, we could construct the net N greedily:

While there is a point $x \in B$ of distance larger than γ from all points in N, add x to N.

Consider each point $x \in N$, if we construct a ℓ_1 ball of radius $\gamma/2$ around x, then the balls are all disjoint by the triangle inequality. Moreover, the balls are all contained in the ℓ_1 ball of radius $1 + \gamma/2$ around the origin. Therefore, we can apply a volume argument to bound the size of the net N. Note that the ratio of volume of d-dimensional ℓ_1 ball of radius $1 + \gamma/2$ to that of ℓ_1 ball of radius $\gamma/2$ is $(1 + \gamma/2)^d/(\gamma/2)^d$. Therefore, we can conclude that

$$|N| \le \left(\frac{1 + \gamma/2}{\gamma/2}\right)^d.$$

Net for the Subspace. After constructing the net for the ball, we would like to lift the net to a net for the subspace. Let U be a well-conditioned basis such that

$$\|x\|_1 \leq \|Ux\|_1 \leq d \|x\|_1 .$$

Unlike the ℓ_2 norm, the ℓ_1 norm is not preserved by the linear transformation U. Therefore, we need a denser net for the ball to ensure the performance of the net for the subspace. Let N be a (γ/d)-net for the ball B with respect to the ℓ_1 norm, and let $M = \{Ux \mid x \in N\}$. From the previous paragraph, we already know that $|M| \leq (1 + \gamma/(2d))^d/(\gamma/(2d))^d = \left(\frac{d}{\gamma}\right)^{O(d)}$.

We conclude with the following lemma.

Lemma 2.6. *For every $x \in B$, there exists $y \in M$ such that $\|Ux - y\|_1 \leq \gamma$.*

Proof. Let $x' \in N$ be the closest point to x, then $\|x - x'\|_1 \leq \gamma/d$. Therefore,

$$\|Ux - Ux'\|_1 \leq d \|x - x'\|_1 \leq \gamma.$$

Setting $y = Ux'$, we have $\|Ux - y\|_1 \leq \gamma$. ■