Recitation #4:

TA: Rajesh Jayaram

October 11, 2019

Linear Dependence on ϵ in Regression

In class we saw that if $S \in \mathbb{R}^{k \times n}$ is a subspace embedding for $[A, b]$, where $A \in \mathbb{R}^{n \times d}$, then $\min_x \|S(Ax - b)\|_2^2 \leq (1 + \epsilon) \min_{x'} \|Ax - b\|_2^2$.

Linear Dependence on ϵ in Regression

In class we saw that if $S \in \mathbb{R}^{k \times n}$ is a subspace embedding for $[A, b]$, where $A \in \mathbb{R}^{n \times d}$, then $\min_x \|S(Ax - b)\|_2^2 \leq (1 + \epsilon) \min_{x'} \|Ax - b\|_2^2$.

Instantiations of S:

- i.i.d. Gaussian: $S_{i,j} \sim \mathcal{N}(0, 1/k)$
- $S = PHD$ Subsampled Randomized Hadamard Transform.
- $S :=$ Count-sketch.

Linear Dependence on ϵ in Regression

In class we saw that if $S \in \mathbb{R}^{k \times n}$ is a subspace embedding for $[A, b]$, where $A \in \mathbb{R}^{n \times d}$, then $\min_x \|S(Ax - b)\|_2^2 \leq (1 + \epsilon) \min_{x'} \|Ax - b\|_2^2$.

Instantiations of S:

- i.i.d. Gaussian: $S_{i,j} \sim \mathcal{N}(0, 1/k)$
- $S = PHD$ Subsampled Randomized Hadamard Transform.
- $S :=$ Count-sketch.

In each case, we required $k = \Omega(d/\epsilon^2)$ rows. We will now see that $k = O(d/\epsilon)$ is possible if S is i.i.d. Gaussian.

Linear Dependence on ϵ in Regression

First: let $A = U\Sigma V^T$, where $U \in \mathbb{R}^{n\times r}$ is an orthonormal basis for the column span of A, and $r =$rank(A).

- Suppose $x' = \arg\min_x \|SUx - Sb\|_2$ satisfies

$$\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$$

Now let $y' = \arg\min y\|SAy - Sb\|_2$. What can we say about $\|Ay' - b\|_2$?

Linear Dependence on ϵ in Regression

First: let $A = U\Sigma V^T$, where $U \in \mathbb{R}^{n\times r}$ is an orthonormal basis for the column span of A, and $r =$rank(A).

- Suppose $x' = \arg\min_x \|SUx - Sb\|_2$ satisfies

$$\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$$

Now let $y' = \arg\min y\|SAy - Sb\|_2$. What can we say about $\|Ay' - b\|_2$?

Column spans of U, A are the same! By change of variables:

$$\min y\|SAy - Sb\|_2 = \min y\|SU\Sigma V^T y - Sb\|_2 = \min y\|SUy - Sb\|_2$$

Linear Dependence on ϵ in Regression

First: let $A = U\Sigma V^T$, where $U \in \mathbb{R}^{n\times r}$ is an orthonormal basis for the column span of A, and $r =$rank(A).

- Suppose $x' = \arg\min_x \|SUx - Sb\|_2$ satisfies

$$\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$$

Now let $y' = \arg\min y\|SAy - Sb\|_2$. What can we say about $\|Ay' - b\|_2$?

Column spans of U, A are the same! By change of variables:

$$\min y\|SAy - Sb\|_2 = \min y\|SU\Sigma V^T y - Sb\|_2 = \min y\|SUy - Sb\|_2$$

So $Ay' = Ux'$.

Linear Dependence on ϵ in Regression

First: let $A = U\Sigma V^T$, where $U \in \mathbb{R}^{n\times r}$ is an orthonormal basis for the column span of A, and $r =$rank(A).

- Suppose $x' = \arg\min_x \|SUx - Sb\|_2$ satisfies

$$\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$$

Now let $y' = \arg\min y\|SAy - Sb\|_2$. What can we say about $\|Ay' - b\|_2$?

Column spans of U, A are the same! By change of variables:

$$\min y\|SAy - Sb\|_2 = \min y\|SU\Sigma V^T y - Sb\|_2 = \min y\|SUy - Sb\|_2$$

So $Ay' = Ux'$.

$$\|Ay' - b\|_2 = \|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2 = (1+\epsilon)\min_y \|Ay - b\|_2$$

Linear Dependence on ϵ in Regression

We have shown that if $x' = \arg\min_x \|SUx - Sb\|_2$ satisfies

$$\|Ux' - b\|_2 \leq (1+\epsilon) \min_x \|Ux - b\|_2$$

then $y' = \arg\min y \|SAy - Sb\|_2$ satisfies

$$\|Ay' - b\|_2 \leq (1+\epsilon) \min_y \|Ay - b\|_2$$

Thus we can focus on showing that $\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$ for x' as above, where S has $O(d/\epsilon)$ rows.

Linear Dependence on ϵ in Regression

W.T.S.: $\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$, where $x' = \arg\min_x \|S(Ux - b)\|_2$.

Linear Dependence on ϵ in Regression

W.T.S.: $\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$, where $x' = \arg\min_x \|S(Ux-b)\|_2$.

Let $x^* = \arg\min_x \|Ux - b\|_2$. We have:

$$\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$$

Why?

Linear Dependence on ϵ in Regression

W.T.S.: $\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$, where $x' = \arg\min_x \|S(Ux - b)\|_2$.

Let $x^* = \arg\min_x \|Ux - b\|_2$. We have:

$$\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$$

Why? Pythagorean Theorem: Ux^* is the projection of b onto the column span of U, so $Ux^* - b$ is orthogonal to the span of U.

Linear Dependence on ϵ in Regression

W.T.S.: $\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$, where $x' = \arg\min_x \|S(Ux - b)\|_2$.

Let $x^* = \arg\min_x \|Ux - b\|_2$. We have:

$$\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$$

Why? Pythagorean Theorem: Ux^* is the projection of b onto the column span of U, so $Ux^* - b$ is orthogonal to the span of U.

If $y \in V$ for a subspace V, then $\|y - b\|_2^2$ is equal to the cost of projecting b orthogonally onto V, and then the cost of moving from $P_V b$ to y. i.e. $\|y - b\|_2^2 = \|P_V b - b\|_2^2 + \|y - P_v b\|_2^2$.

Linear Dependence on ϵ in Regression

W.T.S.: $\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$, where $x' = \arg\min_x \|S(Ux - b)\|_2$.

We just showed that:

$$\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$$

where $x^* = \arg\min_x \|Ux - b\|_2$.

Linear Dependence on ϵ in Regression

W.T.S.: $\|Ux' - b\|_2 \leq (1+\epsilon)\min_x \|Ux - b\|_2$, where $x' = \arg\min_x \|S(Ux - b)\|_2$.

We just showed that:

$$\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$$

where $x^* = \arg\min_x \|Ux - b\|_2$.

To finish, we just need to show that

$$\|U(x' - x^*)\|_2^2 = O(\epsilon)\|Ux^* - b\|_2^2$$

Linear Dependence on ϵ in Regression

By definition: $x' = (SU)^{-}Sb$ and $x^* = (U)^{-}b = U^T b$. In class: if $S \in \mathbb{R}^{k \times n}$ has $k = O(d/\epsilon)$, then:

- $\|S[A, b]x\|_2 = (1 \pm c)\|x\|_2$ for all x for some constant $c > 0$.
- For fixed matrices C, D:

$$\Pr[\|CSS^T D - CD\|_F^2 \geq (\epsilon/d)\|A\|_F^2\|B\|_F^2] \leq 9/10$$

Linear Dependence on ϵ in Regression

By definition: $x' = (SU)^{-}Sb$ and $x^* = (U)^{-}b = U^Tb$. In class: if $S \in \mathbb{R}^{k \times n}$ has $k = O(d/\epsilon)$, then:

- $\|S[A,b]x\|_2 = (1 \pm c)\|x\|_2$ for all x for some constant $c > 0$.
- For fixed matrices C, D:

$$\Pr[\|CSS^TD - CD\|_F^2 \geq (\epsilon/d)\|A\|_F^2\|B\|_F^2] \leq 9/10$$

If $\|SUx\|_2 = (1 \pm c)\|x\|_2$ for all x, then SU has linearly independent columns (**Why?**).

Linear Dependence on ϵ in Regression

By definition: $x' = (SU)^- Sb$ and $x^* = (U)^- b = U^T b$. In class: if $S \in \mathbb{R}^{k \times n}$ has $k = O(d/\epsilon)$, then:

- $\|S[A, b]x\|_2 = (1 \pm c)\|x\|_2$ for all x for some constant $c > 0$.
- For fixed matrices C, D:

$$\Pr[\|CSS^T D - CD\|_F^2 \geq (\epsilon/d)\|A\|_F^2 \|B\|_F^2] \leq 9/10$$

If $\|SUx\|_2 = (1 \pm c)\|x\|_2$ for all x, then SU has linearly independent columns (**Why?**).
Fact: if C has full column rank, $C^- = (C^T C)^{-1} C^T$.
So $x' = (SU)^- Sb = (U^T S^T SU)^{-1} U^T S^T Sb$

Linear Dependence on ϵ in Regression

$$\begin{aligned}
\|U(x' - x^*)\|_2^2 &= \|U((SU)^- Sb - U^T b)\|_2^2 \\
&= \|U((U^T S^T SU)^{-1} U^T S^T Sb - U^T b)\|_2^2 \\
&= \|(U^T S^T SU)^{-1} U^T S^T Sb - U^T b\|_2^2
\end{aligned} \tag{1}$$

Linear Dependence on ϵ in Regression

$$\begin{aligned} \|U(x' - x^*)\|_2^2 &= \|U((SU)^- Sb - U^T b)\|_2^2 \\ &= \|U((U^T S^T SU)^{-1} U^T S^T Sb - U^T b)\|_2^2 \\ &= \|(U^T S^T SU)^{-1} U^T S^T Sb - U^T b\|_2^2 \end{aligned} \tag{2}$$

Now $\|SUx\|_2 = (1 \pm c)\|x\|_2$, so

$$\frac{1}{2} \leq \sigma_{\min}(U^T S^T SU) \leq \sigma_{\max}(U^T S^T SU) \leq 2$$

Linear Dependence on ϵ in Regression

$$\begin{aligned} \|U(x' - x^*)\|_2^2 &= \|U((SU)^{-}Sb - U^T b)\|_2^2 \\ &= \|U((U^T S^T SU)^{-1} U^T S^T Sb - U^T b)\|_2^2 \\ &= \|(U^T S^T SU)^{-1} U^T S^T Sb - U^T b\|_2^2 \end{aligned} \tag{3}$$

Now $\|SUx\|_2 = (1 \pm c)\|x\|_2$, so

$$\frac{1}{2} \leq \sigma_{\min}(U^T S^T SU) \leq \sigma_{\max}(U^T S^T SU) \leq 2$$

$$\|U(x' - x^*)\|_2^2 \leq O(1)\|(U^T S^T SU)((U^T S^T SU)^{-1} U^T S^T Sb - U^T b)\|_2^2 \tag{4}$$

Linear Dependence on ϵ in Regression

$$\begin{aligned} \|U(x' - x^*)\|_2^2 &= \|U((SU)^- Sb - U^T b)\|_2^2 \\ &= \|U((U^T S^T SU)^{-1} U^T S^T Sb - U^T b)\|_2^2 \\ &= \|(U^T S^T SU)^{-1} U^T S^T Sb - U^T b\|_2^2 \end{aligned} \quad (5)$$

Now $\|SUx\|_2 = (1 \pm c)\|x\|_2$, so

$$\frac{1}{2} \leq \sigma_{\min}(U^T S^T SU) \leq \sigma_{\max}(U^T S^T SU) \leq 2$$

$$\begin{aligned} \|U(x' - x^*)\|_2^2 &\leq O(1)\|(U^T S^T SU)((U^T S^T SU)^{-1} U^T S^T Sb - U^T b)\|_2^2 \\ &= O(1)\|U^T S^T Sb - U^T S^T SUU^T b\|_2^2 \\ &= O(1)\|U^T S^T S(b - UU^T b)\|_2^2 \end{aligned} \quad (6)$$

Linear Dependence on ϵ in Regression

We have:

$$\|U(x' - x^*)\|_2^2 \leq O(1)\|U^T S^T S(b - UU^T b)\|_2^2$$

Linear Dependence on ϵ in Regression

We have:

$$\|U(x' - x^*)\|_2^2 \leq O(1)\|U^T S^T S(b - UU^T b)\|_2^2$$

Observe $U^T(b - UU^T b) = U^T b - U^T U U^T b = U^T b - U^T b = 0$.

Look Familiar?

Linear Dependence on ϵ in Regression

We have:

$$\|U(x' - x^*)\|_2^2 \leq O(1)\|U^T S^T S(b - UU^T b)\|_2^2$$

Observe $U^T(b - UU^T b) = U^T b - U^T UU^T b = U^T b - U^T b = 0$.

Look Familiar? Approx matrix prod!

Linear Dependence on ϵ in Regression

We have:

$$\|U(x' - x^*)\|_2^2 \leq O(1)\|U^T S^T S(b - UU^T b)\|_2^2$$

Observe $U^T(b - UU^T b) = U^T b - U^T UU^T b = U^T b - U^T b = 0$.

Look Familiar? Approx matrix prod!

$$\begin{aligned} \|U^T S^T S(b - UU^T b)\|_2^2 &= \|U^T S^T S(b - UU^T b) - \mathbf{0}\|_2^2 \\ &\leq (\epsilon/d)\|U^T\|_F^2 \|b - UU^T b\|_2^2 \end{aligned} \tag{7}$$

Linear Dependence on ϵ in Regression

We have:

$$\|U(x' - x^*)\|_2^2 \leq O(1)\|U^T S^T S(b - UU^T b)\|_2^2$$

Observe $U^T(b - UU^T b) = U^T b - U^T U U^T b = U^T b - U^T b = 0.$

Look Familiar? Approx matrix prod!

$$\begin{aligned}
\|U^T S^T S(b - UU^T b)\|_2^2 &= \|U^T S^T S(b - UU^T b) - \mathbf{0}\|_2^2 \\
&\leq (\epsilon/d)\|U^T\|_F^2 \|b - UU^T b\|_2^2 \qquad (8) \\
&\leq \epsilon \|b - UU^T b\|_2^2
\end{aligned}$$

Linear Dependence on ϵ in Regression

We have:

$$\|U(x' - x^*)\|_2^2 \leq O(1)\|U^T S^T S(b - UU^T b)\|_2^2$$

Observe $U^T(b - UU^T b) = U^T b - U^T UU^T b = U^T b - U^T b = 0$.

Look Familiar? Approx matrix prod!

$$\begin{aligned} \|U^T S^T S(b - UU^T b)\|_2^2 &= \|U^T S^T S(b - UU^T b) - \mathbf{0}\|_2^2 \\ &\leq (\epsilon/d)\|U^T\|_F^2 \|b - UU^T b\|_2^2 \\ &\leq \epsilon\|b - UU^T b\|_2^2 \\ &\leq \epsilon\|b - Ux^*\|_2^2 \end{aligned} \qquad (9)$$

Linear Dependence on ϵ in Regression

We have:

$$\|U(x' - x^*)\|_2^2 \leq O(1)\|U^T S^T S(b - UU^T b)\|_2^2$$

Observe $U^T(b - UU^T b) = U^T b - U^T UU^T b = U^T b - U^T b = 0$.

Look Familiar? Approx matrix prod!

$$\begin{aligned} \|U^T S^T S(b - UU^T b)\|_2^2 &= \|U^T S^T S(b - UU^T b) - \mathbf{0}\|_2^2 \\ &\leq (\epsilon/d)\|U^T\|_F^2\|b - UU^T b\|_2^2 \\ &\leq \epsilon\|b - UU^T b\|_2^2 \\ &\leq \epsilon\|b - Ux^*\|_2^2 \end{aligned} \tag{10}$$

So

$$\|U(x' - x^*)\|_2^2 \leq O(\epsilon)\|b - Ux^*\|_2^2$$

Linear Dependence on ϵ in Regression

Summary: We showed that

$$\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$$

where $x' = \arg\min_x \|SUx - Sb\|_2^2$.

Linear Dependence on ϵ in Regression

Summary: We showed that

$$\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$$

where $x' = \arg\min_x \|SUx - Sb\|_2^2$.

We also have: $\|U(x' - x^*)\|_2^2 \leq O(\epsilon)\|Ux^* - b\|_2^2$. So

$$\|Ux' - b\|_2^2 = (1 + O(\epsilon))\|Ux^* - b\|_2^2$$

Linear Dependence on ϵ in Regression

Summary: We showed that

$$\|Ux' - b\|_2^2 = \|Ux^* - b\|_2^2 + \|U(x' - x^*)\|_2^2$$

where $x' = \arg\min_x \|SUx - Sb\|_2^2$.

We also have: $\|U(x' - x^*)\|_2^2 \leq O(\epsilon)\|Ux^* - b\|_2^2$. So

$$\|Ux' - b\|_2^2 = (1 + O(\epsilon))\|Ux^* - b\|_2^2$$

We only needed S to satisfy:

- S is a $(1 \pm 1/2)$ subspace embedding for the column span of $[U, b]$.
- S satisfies the approximate matrix product property with error $O(\epsilon/d)$.

Both of which we obtain with $O(d/\epsilon)$ rows for Gaussian S.