CS 15-859: Algorithms for Big Data Fall 2020

Lecture 1.2 — 9/10/2020

Prof. David Woodruff *Scribe: Bailey Miller*

1 Sketch and Solve Paradigm

Solving least squares involves computing $A^{-}b$ which can naively have $O(nd^2)$ time complexity when A is an $n \times d$ matrix. Even fast matrix multiplication only improves this time complexity to $O(nd^{1.376})$. We can improve this time complexity by solving least squares for an approximate solution x'.

Claim 1. The Sketch and Solve method (outlined below) produces an approximate solution x' that with high probability satisfies $\|Ax' - b\|_2 \leq (1+\epsilon)\min_x \|Ax - b\|_2$

- Draw S from a $k \times n$ random family of matrices, for a $k \ll n$
- Compute SA and Sb
- Output solution x' to $\min_{x'} \|(SA)x - (Sb)\|_2$

Remark 1. Lots of families of random matrices will produce an S that works in the sketch and solve paradigm. We choose S to be a $\frac{d}{\epsilon^2} \times n$ matrix of i.i.d. Normal random variables, $S_{i,j} \sim N(0, \frac{1}{k})$.

Remark 2. If SA is given, the time complexity of computing $(SA)^{-}$ is $O(kd^2)$ where $k << n$. Notice though that computing SA naively will take $O(nkd)$ time. Later in the course, we'll see that S can be structured so that SA can be computed much more efficiently than if S is an arbitrary matrix.

2 Subspace Embeddings

Definition (Subspace Embedding)**.** Let S be a $k \times n$ matrix. We say S is a subspace embedding for the column space of A if with high probability

$$\forall x \in \mathbb{R}^d \quad \|SAx\|_2 = (1 \pm \epsilon)\|Ax\|_2. \tag{1}$$

Remark 3. A subspace embedding is simultaneously valid for **all** vectors x in a fixed $d-$dimensional subspace. This means that the length all of vectors x in the column space of A are preserved by a subspace embedding.

We can use subspace embeddings to show that the sketch and solve paradigm produces approximate solutions that are close to the optimal solution with high probability.

Claim 2 (S is a subspace embedding)**.** The $k \times n$ matrix S of i.i.d. normal random variables $N(0, \frac{1}{k})$ is a subspace embedding for the column space of A for $k = O(d/\epsilon^2)$.

Lemma 1. *If $X \sim N(0, a^2)$ and $Y \sim N(0, b^2)$ are independent random variables, then $Z = X + Y$ where $Z \sim N(0, a^2 + b^2)$*

Proof. The pdf of Z, say $f_Z(x)$, is the convolution of the pdf of X and Y, say $f_X(x)$ and $f_Y(x)$ respectively

$$f_Z(z) = \int f_X(z-y) f_Y(y)\, dy \tag{2}$$

$$= \int \frac{1}{a\sqrt{2\pi}} \exp\left(\frac{-(z-y)^2}{2a^2}\right) \frac{1}{b\sqrt{2\pi}} \exp\left(\frac{-y^2}{2b^2}\right) dy \tag{3}$$

$$= \int \frac{1}{ab2\pi} \exp\left(\frac{-(z-y)^2}{2a^2} - \frac{y^2}{2b^2}\right) dy \tag{4}$$

$$= \frac{1}{\sqrt{a^2+b^2}} \int \frac{\sqrt{a^2+b^2}}{ab\sqrt{2\pi}} \exp\left(\frac{-(z-y)^2}{2a^2} - \frac{y^2}{2b^2}\right) dy. \tag{5}$$

With some calculation, it can be shown that:

$$\frac{-(z-y)^2}{2a^2} - \frac{y^2}{2b^2} = \frac{-z^2}{2(a^2+b^2)} - \frac{\left(y - \frac{b^2 z}{a^2+b^2}\right)^2}{2\frac{(ab)^2}{(a^2+b^2)}}. \tag{6}$$

This allows us to split the exponential and obtain two Normal density functions

$$f_Z(z) = \underbrace{\frac{1}{\sqrt{a^2+b^2}\sqrt{2\pi}} \exp\left(\frac{-z^2}{2(a^2+b^2)}\right)}_{\text{pdf for } N(0, a^2+b^2)} \int \underbrace{\frac{\sqrt{a^2+b^2}}{ab\sqrt{2\pi}} \exp\left(-\frac{\left(y - \frac{b^2 z}{a^2+b^2}\right)^2}{2\frac{(ab)^2}{a^2+b^2}}\right)}_{\text{pdf for } N(\frac{b^2 z}{a^2+b^2}, \frac{(ab)^2}{a^2+b^2})} dy \tag{7}$$

and so integrating over the Normal density on the domain $(-\infty, \infty)$ gives us our solution

$$f_Z(z) = \frac{1}{\sqrt{a^2+b^2}\sqrt{2\pi}} \exp\left(\frac{-z^2}{2(a^2+b^2)}\right) \cdot 1. \tag{8}$$

■

Fact 1. If g is an n-dimensional vector of i.i.d. $N(0,1)$ random variables and R is a fixed matrix, then the probability density function of Rg is

$$f(x) = \frac{1}{\det(RR^T)(2\pi)^{n/2}} \exp\left(-\frac{x^T (RR^T)^{-1} x}{2}\right) \tag{9}$$

where RR^T is the covariance matrix.

Remark 4. If R is a rotation matrix, then $RR^T = I$ and so the distribution of Rg and g are the same

$$\frac{1}{\det(RR^T)(2\pi)^{n/2}} \exp\left(-\frac{x^T (RR^T)^{-1} x}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{x^T x}{2}\right). \tag{10}$$

Lemma 2. *If u, v are vectors with $\langle u, v\rangle = 0$, then $\langle g, u\rangle$ and $\langle g, v\rangle$ are independent, where g is an n-dimensional vector of i.i.d. $N(0, \frac{1}{k})$ random variables.*

Proof. Since $\langle u, v\rangle = 0$ we can choose a rotation R such that

$$Ru = \alpha e_1 = \alpha(1, 0, \ldots, 0) \tag{11}$$
$$Rv = \beta e_2 = \beta(0, 1, \ldots, 0). \tag{12}$$

Then, we can apply the rotation R to the dot products $\langle g, u\rangle$ and $\langle g, v\rangle$ since rotations have no impact on dot products

$$\langle g, u\rangle = \langle Rg, Ru\rangle = \langle h, \alpha e_1\rangle = \alpha h_1 \tag{13}$$
$$\langle g, v\rangle = \langle Rg, Rv\rangle = \langle h, \beta e_2\rangle = \beta h_2. \tag{14}$$

Recall that h will also be a vector of i.i.d. random variables $N(0, \frac{1}{k})$ since Rg and g have the same distribution. By definition then, h_1 and h_2 are independent. ■

Lemma 3. *Let S be a $k \times n$ matrix of i.i.d. random variables $N(0, \frac{1}{k})$ and let A be an $n \times d$ orthonormal matrix, then SA is a $k \times d$ matrix of i.i.d. $N(0, \frac{1}{k})$ random variables*

Proof. Since the rows of S are independent, the rows of A are also independent. We can write each row of SA as follows

$$\langle g, A_1\rangle, \ldots, \langle g, A_d\rangle \tag{15}$$

where A_i is the i^{th} column of A and g is a row of S.

Each entry $\langle g, A_i\rangle$ is a sum of i.i.d. Normal random variables $N(0, \frac{1}{k})$ scaled by entry $A_{j,i}$, so by lemma 1 each entry is itself a Normal distribution with mean 0 and $N(0, \frac{\|A_i\|_2^2}{k})$ where $\|A_i\|_2^2 = 1$ since A is orthonormal.

Finally since the columns of A are independent, lemma 2 implies any two entries $\langle g, A_i\rangle$ and $\langle g, A_k\rangle$ must be independent. Therefore, all the entries of SA are independent. ■

Lemma 4. *If S be a $k \times n$ matrix of i.i.d. random variables $N(0, \frac{1}{k})$ and A is an $n \times d$ matrix with orthonormal columns, then for any fixed unit vector $x \in \mathbb{R}^d$*

$$\mathbb{E}[\|SAx\|_2^2] = 1 \tag{16}$$

Proof. Assume that A is orthonormal and consider any fixed unit vector $x \in \mathbb{R}^d$.

$$\|SAx\|_2^2 = \sum_{i\in[k]} \langle g_i, x\rangle^2 \tag{17}$$

where g_i is the i^{th} row of SA. Then each $\langle g_i, x\rangle^2$ is distributed as $N(0, \frac{\|x\|_2^2}{k})^2 = N(0, \frac{1}{k})^2$.

$$\mathbb{E}[\langle g_i, x\rangle^2] = \frac{1}{k} \tag{18}$$

and finally using linearity of expectation we have

$$\mathbb{E}[\sum_{i\in[k]} \langle g_i, x\rangle^2] = \sum_{i\in[k]} \mathbb{E}[\langle g_i, x\rangle^2] = \sum_{i\in[k]} \frac{1}{k} = 1.$$ ■

Remark 5. The last lemma shows that we're headed in the right direction for proving that S is a subspace embedding. Recall that we want to show that with high probability

$$\|SAx\|_2 = (1 \pm \epsilon)\|Ax\|_2. \tag{19}$$

If we assume A is orthonormal and x is a unity vector, then $\|Ax\|_2 = \|x\|_2 = 1$. So we want to show that

$$\|SAx\|_2 = (1 \pm \epsilon). \tag{20}$$

Squaring both sides, we see

$$\|SAx\|_2^2 = (1 \pm \epsilon)^2. \tag{21}$$

Since $\mathbb{E}[\|SAx\|_2^2] = 1$, we can see that S is in the right ballpark for being a subspace embedding. Now, we turn our attention to finding the concentration of $\|SAx\|_2^2$ about its expectation.

Theorem 1 (Johnson-Lindenstrauss Theorem)**.** Let $h_1, \ldots, h_k$ be i.i.d. $N(0,1)$ random variables, then $G = \sum_i h_i^2$ is a χ^2 random variable which means we can apply known tail bounds to G

$$\text{(Upper Bound)} \qquad \mathbf{Pr}[G \geq k + 2\sqrt{kx} + 2x] \leq \exp(-x) \tag{22}$$

$$\text{(Lower Bound)} \qquad \mathbf{Pr}[G \leq k - 2\sqrt{kx}] \leq \exp(-x) \tag{23}$$

If $x = \frac{\epsilon^2 k}{16}$ then the tail bounds become

$$\mathbf{Pr}[G \geq k + \frac{k\epsilon}{2} + \frac{k\epsilon^2}{8}] \leq \exp\Big(-\frac{\epsilon^2 k}{16}\Big) \tag{24}$$

$$\mathbf{Pr}[G \leq k - \frac{k\epsilon}{2}] \leq \exp\Big(-\frac{\epsilon^2 k}{16}\Big). \tag{25}$$

from which we can use these bounds to construct a concentration inequality around k

$$\mathbf{Pr}[G \in k(1 \pm \epsilon)] \geq 1 - 2\exp\Big(-\frac{\epsilon^2 k}{16}\Big). \tag{26}$$

We can see that if $k = \Theta(\epsilon^{-2}\log(\frac{1}{\delta}))$, then the probability is $1 - \delta$

$$\mathbf{Pr}[G \in k(1 \pm \epsilon)] \geq 1 - \delta. \tag{27}$$

Recall that $\|SAx\|_2^2$ is also the sum of random variables, in particular each $\langle g_i, x\rangle^2 \sim N(0, \frac{1}{k})^2$

$$\|SAx\|_2^2 = \sum_i \langle g_i, x\rangle^2. \tag{28}$$

Therefore, if we divide G by k we get an identical distribution

$$\mathbf{Pr}[\frac{G}{k} \in \frac{k}{k}(1 \pm \epsilon)] \geq 1 - \delta \tag{29}$$

$$\mathbf{Pr}[\|SAx\|_2^2 \in (1 \pm \epsilon)] \geq 1 - \delta. \tag{30}$$

Now, if we choose k such that $k = \Theta(\frac{d}{\epsilon^2})$ where d is the number of columns in A then our bound becomes

$$\mathbf{Pr}[\|SAx\|_2^2 \in (1 \pm \epsilon)] \geq 1 - 2^{-\Theta(d)}. \tag{31}$$

Remark 6. The Johnson-Lindenstrauss Theorem tells us that $\mathbf{Pr}[\|SAx\|_2^2 \in (1 \pm \epsilon)] \geq 1 - 2^{-\Theta(d)}$ for a fixed vector $x \in \mathbb{R}^d$. This means that we do not yet have a sufficient argument for why S is a subspace embedding. We could try to take a union bound over all possible $x \in \mathbb{R}^d$ but since there are infinitely many $x \in \mathbb{R}^d$ the probability of our union bound would vanish. Instead, we can try to use a *net argument* which will provide us with a finite set of vectors to reason about.

Definition (γ-Net)**.** Consider the sphere S^{d-1}. A subset N is a γ-net for all $x \in S^{d-1}$, there is a $y \in N$ such that $\|x - y\|_2 \leq \gamma$.

Definition (Greedy Construction of N)**.** To construct a net N we can use the following greedy process

- While there is a point $x \in S^{d-1}$ such that $\|x - y\|_2 > \gamma$ for all $y \in N$ then include x in N.

Claim 3. The Greedy Construction of N will terminate eventually.

Proof. Every net point is distance 1 from the center of the sphere 0^d. Let each net point $x \in N$ be the center of a ball with radius $\gamma/2$, say $B(x, \gamma/2)$. Then by the triangle inequality every net point ball is contained within a ball of $1 + \gamma/2$ centered at 0^d, say $B(0, 1 + \gamma/2)$. Each pair of net point balls $B(x, \gamma/2)$ and $B(y, \gamma/2)$ must be disjoint, if not then that would imply the second ball to be added during construction was within distance γ of the first point. ■

Claim 4. A net N has at most $\frac{(1+\gamma/2)^d}{(\gamma/2)^d}$ net points for $\gamma > 0$.

Proof. The ratio of the volume of a d-dimensional ball of radius $1 + \gamma/2$ to the d-dimensional ball of radius $\gamma/2$ is the ratio of the radii raised to the d^{th} power $\frac{(1+\gamma/2)^d}{(\gamma/2)^d}$.

Our Greedy Construction will continue to add net points surrounded by balls of radius $\gamma/2$ until it terminates. Therefore, a net N has at most $|N| \leq \frac{(1+\gamma/2)^d}{(\gamma/2)^d}$ net points. ■

Definition (Induced Net on $\mathbb{R}^d$)**.** We can use the γ-net for a sphere S^{d-1} to induce a net M on $\mathbb{R}^d$.

$$M = \{Ax \mid x \in N\} \tag{32}$$

where $|M| \leq \frac{(1+\gamma/2)^d}{(\gamma/2)^d}$.

Claim 5. For every $x \in S^{d-1}$, there is a $y \in M$ for which $\|Ax - y\|_2 \leq \gamma$.

Proof. Let $x' \in S^{d-1}$ be such that $\|x - x'\|_2 \leq \gamma$, then since A is an orthonormal matrix

$$\|Ax - Ax'\|_2 = \|x - x'\|_2 \leq \gamma \tag{33}$$

So we can choose $y = Ax'$. ■

Remark 7. At this point, We've shown that for a fixed unit vector x the quantity $\|SAx\|_2$ is preserved up to a factor of $1 \pm \epsilon$ with high probability

$$\mathbf{Pr}[\|SAx\|_2 = (1 \pm \epsilon)] \geq 1 - 2^{-\Theta(d)}. \tag{34}$$

So if we fix vectors x, x' we can show through a union bound that the quantities $\|SAx\|_2$, $\|SAx'\|_2$, and $\|SA(x-x')\|_2$ are also preserved up to a factor of $1 \pm \epsilon$.

Note that although $(x - x')$ is not a unit vector, we can fix the unit vector $\frac{(x-x')}{\|x-x'\|_2}$ and then the preservation of this unit vector up to $1 \pm \epsilon$ tells us that $\|SA(x-x')\|_2$ will be preserved up to a factor of $\|x - x'\|_2(1 \pm \epsilon)$ where $\|x - x'\|_2$ is bounded by 2. The bound of 2 comes from the fact that both x and x' are unit vectors.

$$\mathbf{Pr}[\|\frac{SA(x-x')}{\|x-x'\|_2}\|_2 = (1 \pm \epsilon)] \geq 1 - 2^{-\Theta(d)} \tag{35}$$

$$\mathbf{Pr}[\|SA(x-x')\|_2 = \|x-x'\|_2(1 \pm \epsilon)] \geq 1 - 2^{-\Theta(d)} \tag{36}$$

Claim 6. If $x, x' \in \mathbb{R}^d$ are fixed unit vectors, then $\mathbf{Pr}[\langle Ax, Ax' \rangle = \langle SAx, SAx' \rangle \pm O(\epsilon)] = 1 - 2^{-\Theta(d)}$.

Proof. We can expand the dot product in terms of various norms

$$2\langle SAx, SAx' \rangle = \|SAx\|_2^2 + \|SAx'\|_2^2 - \|SA(x-x')\|_2^2 \tag{37}$$

$$2\langle Ax, Ax' \rangle = \|Ax\|_2^2 + \|Ax'\|_2^2 - \|A(x-x')\|_2^2 \tag{38}$$

Subtracting the second expression from the first expression we have

$$2\langle SAx, SAx' \rangle - \langle Ax, Ax' \rangle = \|SAx\|_2^2 - \|Ax\|_2^2 + \|SAx'\|_2^2 - \|Ax'\|_2^2 + \|A(x-x')\|_2^2 - \|SA(x-x')\|_2^2$$

with probability $1 - 2^{-\Theta(d)}$ the residual of the norms is bounded

$$2(\langle SAx, SAx' \rangle - \langle Ax, Ax' \rangle) = \pm\epsilon \pm \epsilon \pm \epsilon\|x - x'\|_2^2. \tag{39}$$

Since x and x' are unit vectors, $\|x - x'\|_2$ is bounded by 2

$$\langle SAx, SAx' \rangle - \langle Ax, Ax' \rangle = \pm O(\epsilon). \tag{40}$$

Therefore, with high probability the dot products are within order ϵ of each other

$$\mathbf{Pr}[\langle Ax, Ax' \rangle = \langle SAx, SAx' \rangle \pm O(\epsilon)] = 1 - 2^{-\Theta(e)}. \qquad \blacksquare$$

Remark 8. We can now apply a union bound to the points of a net $M = \{Ax \mid x \in N\}$. If M is a $\frac{1}{2}$-net then there are at most $\left(\frac{1+\gamma/2}{\gamma/2}\right)^d = 5^d$ net points. This means there are $(5^d)^2 = 25^d$ pairs of net points, and so $\Theta(d)$ can be adjusted by a constant factor to account for the union bound. For instance, we can choose $\Theta(d)$ such that $1 - 2^{-\Theta(d)} = 1 - \frac{1}{50^d}$. Therefore,

$$\mathbf{Pr}[\forall y, y' \in M \ \langle y, y' \rangle = \langle Sy, Sy' \rangle \pm O(\epsilon)] = 1 - 2^{-\Theta(d)}. \tag{41}$$

For the remainder of the notes, we'll condition on the event that $y, y' \in M$ have their dot product preserved under S.

Remark 9. By linearity, if for $y, y' \in M$ and $\langle y, y' \rangle = \langle Sy, Sy' \rangle \pm O(\epsilon)$ then for $\alpha y, \beta y'$ we have

$$\langle \alpha y, \beta y' \rangle = \alpha\beta\langle y, y' \rangle = \alpha\beta\langle Sy, Sy' \rangle \pm O(\epsilon\alpha\beta). \tag{42}$$

Now let $y = Ax$ for any $x \in S^{d-1}$. This means y is a unit vector in the column space of A, but it is not necessarily a net point. Our goal is to show that S preserves the length of y within $\pm O(\epsilon)$. We do this by first representing y using a linear combination of net points and then by evaluating $\|Sy\|_2^2$ in terms of the linear combination of net points. To construct such a linear combination, we use the process outlined below.

Let $y = Ax$ for $x \in S^{d-1}$

- Choose $y_1 \in M$ such that $\|y - y_1\|_2 \leq \gamma$
- Select α such that $\|\alpha(y - y_1)\|_2 = 1$
- Let $y_2' \in M$ be such that $\|\alpha(y - y_1) - y_2'\|_2 \leq \gamma$
- Then $\|y - y_1 - \frac{y_2'}{\alpha}\|_2 \leq \frac{\gamma}{\alpha} \leq \gamma^2$
- Set $y_2 = \frac{y_2'}{\alpha}$
- Repeat process to obtain $y_1, y_2, y_3, \ldots$

when the process is complete, that is there are no more net points remaining, then we have

$$\|y - y_1 - y_2 - \ldots - y_i\|_2 \leq \gamma^i. \tag{43}$$

We can use this expression and the triangle inequality to determine the length of y_i

$$\begin{aligned} \|y_i\|_2 &= \|(y - y_1 - y_2 - \ldots - y_{i-1}) - (y - y_1 - y_2 - \ldots - y_i)\|_2 && (44)\\ &\leq \|y - y_1 - y_2 - \ldots - y_{i-1}\|_2 + \|y - y_1 - y_2 - \ldots - y_i\|_2 && (45)\\ &\leq \gamma^{i-1} + \gamma^i && (46)\\ &\leq 2\gamma^{i-1}. && (47) \end{aligned}$$

Therefore, we have that y is a linear combination of net points $y = \sum_i^\infty y_i$ and since $\|y_i\|_2 \leq 2\gamma^{i-1}$ the series will converge.

Proposition 1. For any $x \in R^d$, we have $\|SAx\|_2^2 = (1 \pm \epsilon)\|Ax\|_2^2$. In other words, S is a subspace embedding

Proof. Let $x \in S^{d-1}$ and then $SAx = Sy$ where $y = Ax$. We can construct a linear combination of net points $y_1, y_2, \ldots$ as outlined in the process above. Now, we use this linear combination to

determine $\|Sy\|_2^2$

$$\|Sy\|_2^2 = \|S\sum_i y_i\|_2^2 \tag{48}$$
$$= \sum_i \|Sy_i\|_2^2 + 2\sum_{i\neq j}\langle Sy_i, Sy_j\rangle \tag{49}$$
$$= \sum_i \|y_i\|_2^2 + 2\sum_{i\neq j}\langle y_i, y_j\rangle \pm O(\epsilon)\sum_{i,j}\|y_i\|_2\|y_j\|_2 \tag{50}$$
$$= \sum_i \|y_i\|_2^2 + 2\sum_{i\neq j}\langle y_i, y_j\rangle \pm O(\epsilon) \tag{51}$$
$$= \|\sum_i y_i\|_2^2 \pm O(\epsilon) \tag{52}$$
$$= \|y\|_2^2 \tag{53}$$
$$= 1 \pm O(\epsilon). \tag{54}$$

Note that we were able to reduce $\sum_{i,j}\|y_i\|_2\|y_j\|_2$ to $O(1)$ since we can form geometric series

$$\sum_{i,j}\|y_i\|_2\|y_j\|_2 = \sum_i \|y_i\|_2 \sum_j \|y_j\|_2 \tag{55}$$
$$\leq \sum_i \|y_i\|_2 \sum_j 2\gamma^{j-1} \tag{56}$$
$$\leq \sum_i \|y_i\|_2 O(1) \tag{57}$$
$$\leq \sum_i O(1)2\gamma^{i-1} \tag{58}$$
$$\leq O(1). \tag{59}$$

From remark 9, we know that we can scale x and the results will still hold. Therefore for all x we have that

$$\|SAx\|_2 = (1\pm\epsilon)\|Ax\|_2. \qquad \blacksquare$$

3 Back to Regression

Recall that we originally set out to find an approximate x' using the sketch and solve paradigm such that the following bound was true with high probability

$$\|Ax - b\|_2 \leq (1+\epsilon)\min_y \|Ay - b\|_2 \tag{60}$$

Theorem 2. *If S is a $k\times n$ matrix of i.i.d. random variables $N(0, \frac{1}{k})$ where $k = O(d/\epsilon^2)$, then if*

$$\tilde{x} = \arg\min_x \|S(Ax-b)\|_2, \tag{61}$$

then

$$\|A\tilde{x} - b\|_2 \leq (1+O(\epsilon))\|Ax - b\|_2.$$

Proof. Consider the $d+1$ dimensional subspace L spanned by the columns of A joined with b. For any $y \in L$ we have by proposition 1 that S is a subspace embedding and so

$$\|Sy\|_2 = (1 \pm \epsilon)\|y\|_2. \tag{62}$$

Let $x^* = \arg\min_x \|Ax - b\|_2$ and $\tilde{x} = \arg\min_x \|(SA)x - (Sb)\|_2$. Now

$$\begin{aligned}
\|A\tilde{x} - b\|_2 &\leq \frac{1}{1-\epsilon}\|SA\tilde{x} - Sb\|_2 \\
&\leq \frac{1}{1-\epsilon}\|SAx^* - Sb\|_2 \\
&\leq \frac{1+\epsilon}{1-\epsilon}\|Ax^* - b\|_2 \\
&\leq (1 + O(\epsilon))\|Ax^* - b\|_2.
\end{aligned}$$

■

As a result, we can compute a good approximation of least squares with high probability using the sketch and solve paradigm

$$\arg\min_x \|(SA)x - (Sb)\|_2 \tag{63}$$

which given SA and Sb can be solved in $\text{poly}(d/\epsilon)$ time where $d \ll n$.