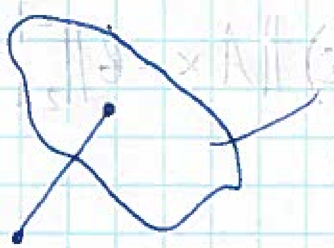
Today: fast least squares, application of (fast) random projections.

$\|Ax-b\|_2^2 \longrightarrow \min$

$A \in \mathbb{R}^{n\times d}$ $x \in \mathbb{R}^d$ $b \in \mathbb{R}^n$ $n \gg d$

"tall skinny" matrix.

geometrically



d-dimensional affine subspace $\{Ax-b\}_x$ of $\mathbb{R}^n$

exact algorithm:

$$x^* = (A^tA)^{-1}A^t b \quad (\text{if } A^tA \in \mathbb{R}^{d\times d} \text{ is invertible}).$$

bottleneck: computing A^tA, takes $O(nd^2)$ time.

Overall: $O(nd^2 + \text{poly}(d)) \approx O(nd^2)$.

Can we do faster?

Yes, if we allow approximation & randomization!!!

Idea: reduce $\|Ax-b\|_2^2 \to \min$ to

$$\|\tilde{A}x - \tilde{b}\|_2^2 \to \min \quad \text{s.t.} \quad \tilde{n} << n \text{ and}$$

$$\Pr\left[\forall x \quad \|\tilde{A}x-\tilde{b}\|_2^2 \overset{1\pm\varepsilon}{\approx} \|Ax-b\|_2^2\right] \geq 0.9 \quad (*).$$

Then optimal solution for the reduced problem is $(1\pm\varepsilon)$-approx. for the original problem.

$$\forall x \quad \Pr\left[\|\tilde{A}x-\tilde{b}\|_2^2 \in (1\pm\varepsilon)\|Ax-b\|_2^2\right] \geq 0.9$$

is not enough!!!

In spirit of JL:

$$\tilde{A} = SA \qquad \tilde{b} = Sb \qquad S \in \mathbb{R}^{\tilde{n}\times n}.$$

<u>Def</u>: $U \subset \mathbb{R}^n$ $\dim U = d$

S is a subspace embedding if

$$\Pr_S\left[\forall y \in U \quad \|Sy\|_2^2 \in (1\pm\varepsilon)\|y\|_2^2\right] \geq 0.9 \quad (**)$$

To achive $(*)$, it is enough to satisfy $(**)$ for $U = \mathrm{span}(\text{columns of } A, b)$.

$\dim U \leq d+1$

Comparison of JL and (**)

JL: any small set of points

(**): infinitely many points, but nicely organized in a subspace.

Two important parameters:

1) number of rows of S

2) Complexity $y \mapsto Sy$

Cl S must have $\geq d$ rows.

Th If S is (scaled) $N(0,1)$ matrix, yields (**) with $\approx \frac{d}{\varepsilon^2}$ rows.

(near-optimal).

We will see a weaker bound of

$$d\,\frac{\log(1/\varepsilon)}{\varepsilon^2}$$

Pf: plan:

1. $N_\varepsilon \subset \{ y \in U \mid \|y\|_2 = 1 \}$. $|N_\varepsilon| < +\infty$.
2. $\forall y \in N_\varepsilon \; \Pr\left[\|Sy\|_2^2 \in (1 \pm \varepsilon) \|y\|_2^2 \right] \geq 1 - \frac{1}{10 N_\varepsilon}$
3. $\Pr\left[\forall y \in N_\varepsilon \; \|Sy\|_2^2 \in (1 \pm \varepsilon) \|y\|_2^2 \right] \geq 0.9$
4. $\Pr\left[\forall y \in U \; \|Sy\|_2^2 \in (1 \pm O(\varepsilon)) \|y\|_2^2 \right] \geq 0.9$.

2) follow from JL with

$$O\left(\frac{\log|N_\varepsilon|}{\varepsilon^2}\right) \text{ rows}$$

3) union bound

Big idea: N_ε – ε-net of the unit sphere. $\{y \in U \mid \|y\|_2 = 1\} =: S_u$.

$$\forall y \in S_u \quad \exists y' \in N_\varepsilon \quad \|y' - y\|_2 \leq \varepsilon.$$

Cl 1 $\exists N_\varepsilon$ with $|N_\varepsilon| \leq O(1/\varepsilon)^d$

Cl 2 If $\forall y \in N_\varepsilon \quad \|Sy\|_2^2 \in (1 \pm \varepsilon)$, then $\forall y \in U \quad \|Sy\|_2^2 \in (1 \pm O(\varepsilon)) \|y\|_2^2$.

Get (**) with $\lesssim \dfrac{d \log(1/\varepsilon)}{\varepsilon^2}$ rows

Pf (Cl 1): can work in $S^{d-1} \subset \mathbb{R}^d$

Construct N_ε using a greedy algorithm: $z_1, \ldots, z_T \in S^{d-1}$.

$B(z_i, \frac{\varepsilon}{2})$ are disjoint and all lie in $B(0, 1 + \frac{\varepsilon}{2})$

Volume bound:

$$T \le \frac{\mathrm{Vol}\left(B(0, 1+\varepsilon/2)\right)}{\mathrm{Vol}\left(z_i, \frac{\varepsilon}{2}\right)} = \frac{\left(1+\frac{\varepsilon}{2}\right)^d}{\left(\frac{\varepsilon}{2}\right)^d} =$$

$$= \left(1+\frac{2}{\varepsilon}\right)^d .$$

Pf (Cl 2):

$$y \in U \qquad y = \alpha_0 v_0 + \alpha_1 v_1 + \alpha_2 v_2 + \dots$$

$$v_i \in N_\varepsilon \qquad |\alpha_i| \le \|y\|_2 \cdot \varepsilon^i \qquad |\alpha_0| = \|y\|_2 .$$

$$\|Sy\|_2 \le \sum_i |\alpha_i| \cdot \|Sv_i\|_2 \le$$

$$\le (1+O(\varepsilon)) \cdot (1+\varepsilon+\varepsilon^2+\dots) \cdot \|y\|_2 \le$$

$$\le (1+O(\varepsilon)) \cdot \|y\|_2$$

$$\|Sy\|_2 \ge |\alpha_0| \, \|Sv_0\|_2 - \sum_{i \ge 1} |\alpha_i| \cdot \|Sv_i\|_2 \ge$$

$$\ge (1-\varepsilon) \|y\|_2 - (1+O(\varepsilon)) (\varepsilon+\varepsilon^2+\dots) \|y\|_2 \ge$$

$$\ge (1-O(\varepsilon)) \cdot \|y\|_2 .$$

Summary:

constructed $S \in \mathbb{R}^{O\left(\frac{d\log(1/\varepsilon)}{\varepsilon^2}\right) \times n}$

that reduces

$$\|Ax - b\|_2^2 \to \min$$

to

$$\|SAx - Sb\|_2^2 \to \min \qquad (\mathrm{poly}(\tfrac{d}{\varepsilon})\ \text{time})$$

Did we gain anything?

No, takes time $O\left(nd^2 \frac{\log(1/\varepsilon)}{\varepsilon^2}\right)$ to compute SA. Worse than exact...

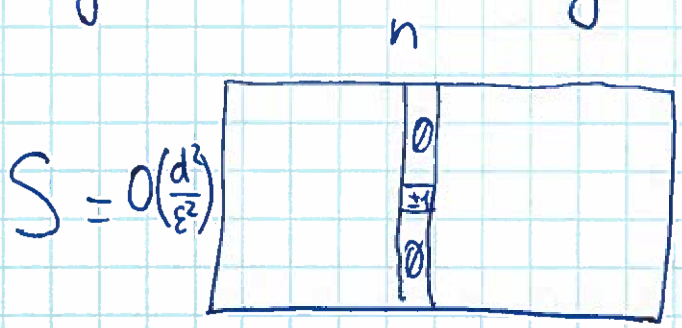
Fast JL to the rescue!!!

$$O\left(\frac{d\log(\frac{1}{\varepsilon})(d\log(\frac{1}{\varepsilon}) + \log n)}{\varepsilon^2}\right) \approx \text{rows}$$

$$\approx O\left(\frac{d^2\log^2(1/\varepsilon)}{\varepsilon^2}\right) \quad \text{but}$$

time is $O(nd\log n + \mathrm{poly}(\frac{d}{\varepsilon}))$
[Sarlos 2006]

Big breakthrough: [Clarkson, Woodruff 2013]



Count-Sketch matrix works!

Can't use our analysis, since

$\|Sy\|_2^2$ does not concentrate.

In particular, can't be used for JL.

Need to use the subspace structure.

Need delicate analysis.

Overall runtime: $O\left(nnz(A) + poly\left(\frac{d}{\varepsilon}\right)\right)$.

In practice:

- Care about $poly\left(\frac{d}{\varepsilon}\right)$
- $poly\left(\frac{1}{\varepsilon}\right) \longmapsto \log\left(\frac{1}{\varepsilon}\right)$.

See the book of Woodruff.