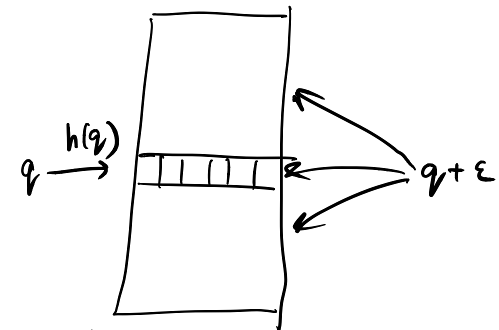
LSH recap:

n^ρ hash tables

n^ρ query time

Storage $n^{1+\rho}$

$\ell_1 : \rho = \frac{1}{c}$ $\quad \ell_2 : \rho = \frac{1}{c^2}$

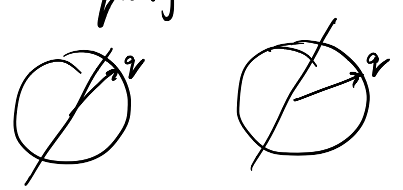


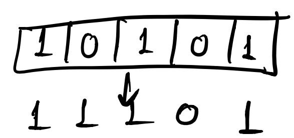
[Panigrahy '06] [Lv et al '07] : multi-probe

perturb query & map to hash bucket

polylog(n) hash tables

ℓ_1: $n^{2.06/c}$ query time





[Kapralov '14] smooth time-space tradeoffs

Sketching for Fast Numerical Linear Algebra

ℓ_2-regression

Given $n \times d$ matrix A vector $b \in \mathbb{R}^n$ $\quad n \gg d$

$$\operatorname{argmin}_x \|Ax - b\|_2^2 = \sum_{i=1}^{n} \left(b_i - \langle A_{i,*}, x\rangle\right)^2$$

ith row of A

Closed form solution:

(Normal eq^n) $\quad A^T A x^* = A^T b$

$x^* = (A^T A)^{-1} A^T b \quad$ if $A^T A$ full rank $d \times d$ matrix

Solving takes $O(nd^2)$ time using naïve matrix mult^n

[Sarlós] Sketching techniques can do better

randomized approximation

relaxed problem $\|Ax-b\|_2 \le (1+\varepsilon)\, \|Ax^* - b\|_2$

optimum solⁿ

Failure prob. δ

① Sample $r \times n$ matrix S $\quad r << n$

② Compute SA, Sb

③ Output exact solⁿ to $\underbrace{\min_x \|(SA)x - (Sb)\|_2}_{\text{smaller regression problem}}$

black box redⁿ to smaller regression problem

Which family of random matrices?

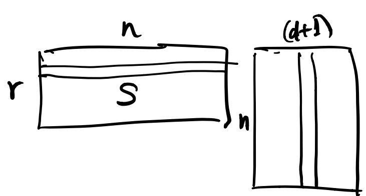
$r = \theta\left(\frac{d}{\varepsilon^2}\right)$ $\quad S_{ij} \sim N\left(0, \frac{1}{r}\right)$ $\quad$ i.i.d. Normal r.v's.

U : $n \times (d+1)$ matrix : orthonormal basis for $[A, b]$

orthonormal columns

col. space (U) = col space $([A, b])$

$S \cdot U$ is also $r \times (d+1)$ matrix of i.i.d. $N(0, \frac{1}{r})$ random variables



w. prob. $1 - \exp(-d)$ $\quad \forall x \quad \|SUx\|_2^2 = (1 \pm \varepsilon)\|x\|_2^2$

For fixed x, $\quad \|SUx\|_2^2 = (1 \pm \varepsilon)\|x\|_2^2$

w. prob. $(1 - \exp(-d))$

Place a sufficiently fine net on unit sphere

Apply union bound to all points on net. (will see this later)

$\forall y \quad \|SUy\|_2 = (1 \pm \varepsilon)\|Uy\|_2 = (1\pm\varepsilon)\|y\|_2$

$\|Uy\|_2 = \|y\|_2$

Consider regression:

$\min_x \|(SA)x - (Sb)\|_2 = \min_x \|S(Ax-b)\|_2$

$Ax-b$ is in col. space of U: $Ax-b = Uy$ for some y

$\|S(Ax-b)\|_2 = (1\pm\varepsilon)\|Ax-b\|_2$

by solving $\min_x \|(SA)x-(Sb)\|_2$ we get $(1+\varepsilon)$ approx to original regression.

$r = \frac{d}{\varepsilon^2}$

Smaller regression can be solved in time $O(rd^2)$
indpt. of n!

Bottleneck: computing SA takes $\Theta(nrd)$ time $S: r\times n$, $A: n\times d$

too expensive!

S is dense

Can choose S from family of structured matrices

s.t. computing SA is faster: (fast JL-transform)

$O(nd\log d) + \mathrm{poly}(d/\varepsilon)$

[Clarkson, Woodruff]: $O(\mathrm{nnz}(A)) + \mathrm{poly}(d/\varepsilon)$

ℓ_2-subspace embedding:

$(1\pm\varepsilon)$ ℓ_2 subspace embedding for col. space of $n\times d$ matrix A is a matrix S for which

$\forall x \in \mathbb{R}^d \quad \|SAx\|_2^2 = (1\pm\varepsilon)\|Ax\|_2^2$

Let U be orthormal basis for columns of A

$$\{Ax \mid x \in \mathbb{R}^d\} = \{Uy \mid y \in \mathbb{R}^t\} \quad t = \operatorname{rank}(A)$$

S is a $(1\pm\varepsilon)$ ℓ_2-subspace embedding for A

$\Leftrightarrow$ $\underline{\quad\quad}\|$ $\underline{\quad\quad}$ for U

$$\|SUy\|_2^2 = (1\pm\varepsilon)\|Uy\|_2^2 = (1\pm\varepsilon)\|y\|_2^2$$

$$y^TU^TS^TSUy = (1\pm\varepsilon)\,y^Ty$$

$$|y^T(I_d - U^TS^TSU)y| \le \varepsilon$$

operator norm $\quad \|I_d - U^TS^TSU\|_2 \le \varepsilon$

Find S w. small # of rows
& be able to compute $S.A$ quickly

oblivious ℓ_2-subspace embedding:

- distrib^n on $r\times n$ matrices S, $r = f(n, d, \varepsilon, \delta)$
  w. prob $1-\delta$ for any fixed $n\times d$ matrix A
  $S \sim \Pi$ is a $(1\pm\varepsilon)$ ℓ_2 subspace embedding for A

<u>Defn</u>: Random matrix $S \in \mathbb{R}^{k\times n}$ is a JL-transform with parameters ε, δ, f $\quad JLT(\varepsilon,\delta,f)$ if

w. prob $\geq 1-\delta$ for any f-element subset $V \subset \mathbb{R}^n$
for all $v, v' \in V$

$$|\langle Sv, Sv'\rangle - \langle v, v'\rangle| \le \varepsilon\|v\|_2\|v'\|_2$$

If we scale vectors in V to unit vectors

alt. phrasing: $\|Sv\|_2^2 = (1\pm\varepsilon)\|v\|_2^2$

$$\|S(v+v')\|_2^2 = (1\pm\varepsilon)\|v+v'\|_2^2$$

$$\forall\, v, v' \in V$$

$$\langle Sv, Sv' \rangle = \left(\|S(v+v')\|_2^2 - \|Sv\|_2^2 - \|Sv'\|_2^2 \right)/2$$
$$= \left((1\pm\varepsilon)\|v+v'\|_2^2 - (1\pm\varepsilon)\|v\|_2^2 - (1\pm\varepsilon)\|v'\|_2^2 \right)/2$$
$$= \langle v, v' \rangle \pm O(\varepsilon)$$

<u>Thm</u>: $S = \frac{1}{\sqrt{k}} R \in \mathbb{R}^{k\times n}$ R_{ij} indpt. standard normal r.v.s.

if $k = \Omega\left(\frac{\log(f/\delta)}{\varepsilon^2}\right)$ then S is a $JLT(\varepsilon, \delta, f)$

How to get subspace embedding property

Need ε-net

$T = \{ y \in \mathbb{R}^n \mid y = Ax$ for some $x \in \mathbb{R}^d, \|y\|_2 = 1 \}$

Find finite subset $N \subset T$ so that

if $\langle Sw, Sw' \rangle = \langle w, w' \rangle \pm \varepsilon$ for all $w, w' \in N$

then $\|Sy\|_2 = (1\pm\varepsilon)\|y\|_2 \quad \forall\, y \in T$

Suffices to choose N so that $\forall y \in T$

$\exists\, w \in N$ for which $\|y - w\|_2 \le 1/2$

N: $\frac{1}{2}$-net

y : unit vector $\quad y = y^0 + y^1 + y^2 + \dots\dots$

y^i : scalar multiple of vector in N

$\|y^i\|_2 \le \frac{1}{2^i}$

$y = y^0 + (y - y^0) \qquad y^0 \in N \quad \|y - y^0\| \le 1/2$

$y - y^0 = y^1 + ((y - y^0) - y^1) \qquad \|y - y^0 - y^1\| \le \frac{1}{4}$

y^1 : scalar multiple of vector in N

$$
\begin{aligned}
\|Sy\|_2^2 &= \|S(y^0 + y^1 + \cdots)\|_2^2 \\
&= \sum_i \|Sy^i\|_2^2 + 2\sum_{i<j} \langle Sy^i, Sy^j \rangle \\
&= \underbrace{\sum_i \|y^i\|_2^2 + 2\sum_{i<j} \langle y^i, y^j \rangle} \pm 2\varepsilon \Big(\underbrace{\sum_{i<j} \|y^i\|_2 \|y^j\|_2}_{\leq \left(\sum \|y^i\|_2\right)^2} \Big) \\
&= 1 \pm O(\varepsilon)
\end{aligned}
$$

Why is there a small $\frac{1}{2}$-net

<u>Lemma:</u> $\exists$ r-net N of T $|N| \leq \left(1 + \frac{4}{r}\right)$