FASTER LEAST SQUARE APPROXIMATION

- Let $[n] = \{1, 2, 3, \dots n\}$

- Recall Hadamard transform: H_n defined recursively:

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad H_n = \begin{bmatrix} H_{n/2} & H_{n/2} \\ H_{n/2} & -H_{n/2} \end{bmatrix} \longrightarrow (1)$$

- $H = \frac{1}{\sqrt{n}} H_n$ - normalized version: $n = 2^k$

- Let $D = \text{diag}(d_{11}, d_{22}, \dots d_{nn}) \in \mathbb{R}^{n \times n}$
  $d_{ii} = \pm$ with probability $1/2$ $\Big\} \longrightarrow (2)$

- DH is known as the randomized Hadamard transform.

Two properties of (DH)

- When applied to a vector x, it spreads its energy in the sense that $\|x\|_\infty$ is bounded after application of (DH)
- The matrix vector product $(DH)x$ can be computed in $O(n \log n)$ time as in standard Fast Fourier transform

- In computing $(DH)x$, there is no multiplication involved.

Linear least squares problem :-

- Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ $m > n$. A - Full rank

- Our goal is to find x such that

$$Z = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 \longrightarrow (3)$$

- $\Rightarrow$ $x_{LS} = (A^T A)^{-1} A^T b = A^+ b$, $A^+ = (A^T A)^{-1} A^T \rightarrow (4)$

- Let B be preconditioning matrix. $B \in \mathbb{R}^{r \times m}$ $(r << m)$

- Define $\tilde{Z}$ such that

$$\tilde{Z} = \min_{x \in \mathbb{R}^n} \|B(Ax - b)\|_2^2 \longrightarrow (5)$$

$\Rightarrow$ $\tilde{x}_{LS} = (BA)^+ (Bb)$, $(BA)^+ = \left[(BA)^T (BA)\right]^{-1} (BA)^T$ $\rightarrow (6)$

- We can find $\tilde{x}_{LS}$ by using conjugate gradient algorithm.

- B also viewed as downscaling operator

Two choices for B

Random sampling	Random Projection
· $B = S^T H D \rightarrow (7)$	$B = T H D \rightarrow (8)$
· HD Randomized Hadamard Transform	· HD - Randomized Hadamard transform
· S - Sampling matrix.	· T is a random projection matrix

Choice of S: $S \in \mathbb{R}^{m \times n}$ - Initially empty

1) For $t = 1, 2, \ldots n$, select uniformly at random an integer from $[n]$ - i.i.d trails with replacement.

   · If i is that integer, add $\sqrt{\frac{m}{n}}\, e_i$ to S
   e_i = standard unit vector in $\mathbb{R}^m$

End For

Choice of T: $T = [T_{ij}] \in \mathbb{R}^{n \times m}$

1) $T_{ij} \sim$ i.i.d $\sim N(0, 1)$
2) $T_{ij} \sim \{-1, 1\}$ with probability $1/2$, $1/2$
3) $T_{ij} \sim \{\pm\sqrt{3}, 0\}$ $\quad T_{ij} = 0$ with prob $= 2/3$
   $\quad\quad = \pm\sqrt{3}$ " $= 1/6$

<u>Claim</u>: Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\varepsilon \in (0,1)$. Then

there exists a randomized algorithm that returns a vector $\tilde{x}_{opt} \in \mathbb{R}^n$ such that

(a) $\tilde{x}_{opt}$: $\|A\tilde{x}_{opt} - b\|_2^* \leq (1+\varepsilon) Z$ $\longrightarrow$ (9)

where Z is defined in (3), with a large probability.

(b) If $\mathcal{K}(A)$ is the condition number of A, $\gamma \in (0,1)$,

the fraction of the norm of b that lies in the column space of A, i.e.

$$\gamma = \frac{\|U_A U_A^T b\|_2}{\|b\|_2} \longrightarrow (10)$$

where U_A is the <u>orthonormal basis for the column space of A</u>, then

$$\|x_{opt} - \tilde{x}_{opt}\|_2 \leq \varepsilon \left[\mathcal{K}(A) \sqrt{\frac{1}{\gamma^2} - 1} \right] \|x_{opt}\|_2$$

$\longrightarrow$ (11)

When structural conditions are true:

$$\sigma^2_{min}(BU_A) \geq \frac{1}{\sqrt{2}} \longrightarrow (12)$$

$$\|U_A^T B^T B b^{\perp}\|_2 \leq \frac{\varepsilon}{2} Z^2 \longrightarrow (13)$$

Where Z is in (3).

<u>Recall</u>: $A = U_A \Sigma_A^{1/2} V_A^T$ - SVD of A.