

CSE 392/CS 395T/M 397C: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin
Spring 2025

Lecture 9: Countsketch; sketch and solve

Outline

- 1 Countsketch
- 2 Sketch and solve

Types of sketching matrices

Gaussian sketching matrix:

- Performs well. Small sketch size.
- $S \in \mathbb{R}^{m \times n}$ requires generating $m \cdot n$ random i.i.d entries.
- Computing SA takes $O(mnd)$ time.

SHRT: Subsampled Randomized Hadamard Transform

- $S = PHD \in \mathbb{R}^{m \times n}$, fewer random bits.
- Faster to apply. SA in $O(mn \log(d))$ time.
- Sketch size needed is larger.
- A should be dense.
- Issues with parallel and distributed computing.

Faster Embeddings: Countsketch

- **Sparse Embeddings:** Adaptation of CountSketch from streaming algorithms.
- S is of the form:

$$\begin{bmatrix} 0 & -1 & 0 & 0 & \cdots & 0 \\ +1 & 0 & 0 & +1 & \cdots & 0 \\ 0 & 0 & -1 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

- One random ± 1 per column.
- Row A_{i*} of A contributes $\pm A_{i*}$ to one of the rows of SA .

Sparse Embeddings

- **Sparse sketching matrix:** For $i \in [n]$, pick uniformly and independently: $h_i \in [m]$, $s_i \in \{-1, +1\}$, and define $\mathbf{S} \in \mathbb{R}^{m \times n}$ as:

$$\mathbf{S}_{h_i, i} \rightarrow s_i \text{ for } i \in [n],$$

and $\mathbf{S}_{j, i} \rightarrow 0$ otherwise.

- \mathbf{s} is a sign (Radamacher) vector. The vector \mathbf{h} hashes to m “hash buckets”. That is,

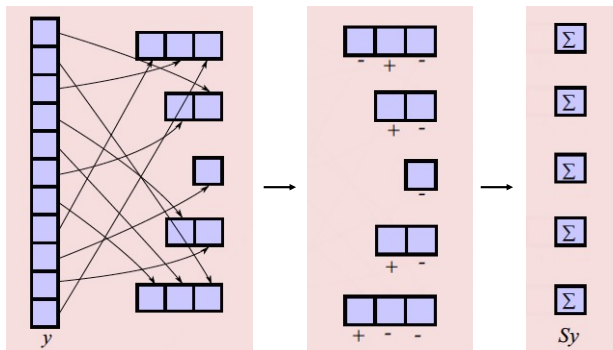
$$\mathbf{S}_{j*} = \sum_{i: h_i=j} s_i \mathbf{e}_i^\top,$$

and so

$$[\mathbf{SA}]_{j*} = \sum_{i: h_i=j} s_i \mathbf{e}_i^\top \mathbf{A} = \sum_{i: h_i=j} s_i \mathbf{A}_{i*}.$$

- **Fast sketching:** Can compute \mathbf{SA} in $O(nnz(\mathbf{A}))$ time.

- If \mathbf{s} is a sign (Radamacher) vector, then $\mathbb{E}[(\mathbf{s}^\top \mathbf{y})^2] = \|\mathbf{y}\|_2^2$.
- For $\mathbf{y} = \mathbf{A}\mathbf{x}$, each row of \mathbf{S} :
 - (a) collects a subset of entries y_i 's; (b) applies the signs, and (c) adds
- $\mathbb{E}[\|\mathbf{S}\mathbf{y}\|_2^2] = \|\mathbf{y}\|_2^2$.



Analysis of sparse embeddings

Variance of Countsketch

For $\mathbf{S} \in \mathbb{R}^{m \times n}$ a sparse sketching distribution, and $\mathbf{y} \in \mathbb{R}^n$ a unit vector,

$$\text{Var}[\|\mathbf{S}\mathbf{y}\|_2^2] \leq \frac{3}{m}.$$

Analysis of sparse embeddings

Variance of Countsketch

For $\mathbf{S} \in \mathbb{R}^{m \times n}$ a sparse sketching distribution, and $\mathbf{y} \in \mathbb{R}^n$ a unit vector,

$$\text{Var}[\|\mathbf{S}\mathbf{y}\|_2^2] \leq \frac{3}{m}.$$

Proof: Let $\mathbf{z} = \mathbf{S}\mathbf{y}$. We have $\mathbb{E}[\|\mathbf{z}\|_2^2] =$

$$\text{Var}[\|\mathbf{z}\|_2^2] =$$

Analysis of sparse embeddings

Variance of Countsketch

For $\mathbf{S} \in \mathbb{R}^{m \times n}$ a sparse sketching distribution, and $\mathbf{y} \in \mathbb{R}^n$ a unit vector,

$$\text{Var}[\|\mathbf{S}\mathbf{y}\|_2^2] \leq \frac{3}{m}.$$

Proof: Let $\mathbf{z} = \mathbf{S}\mathbf{y}$. We have $\mathbb{E}[\|\mathbf{z}\|_2^2] =$

$$\text{Var}[\|\mathbf{z}\|_2^2] =$$

$$\mathbb{E}[\|\mathbf{z}\|_2^4] =$$

Analysis of sparse embeddings

Variance of Countsketch

For $\mathbf{S} \in \mathbb{R}^{m \times n}$ a sparse sketching distribution, and $\mathbf{y} \in \mathbb{R}^n$ a unit vector,

$$\text{Var}[\|\mathbf{S}\mathbf{y}\|_2^2] \leq \frac{3}{m}.$$

Proof: Let $\mathbf{z} = \mathbf{S}\mathbf{y}$. We have $\mathbb{E}[\|\mathbf{z}\|_2^2] =$

$$\text{Var}[\|\mathbf{z}\|_2^2] =$$

$$\mathbb{E}[\|\mathbf{z}\|_2^4] =$$

$$\mathbb{E}_{s,h}[z_j^4] =$$

Countsketch Embedding

Countsketch - subspace embedding

For $\mathbf{S} \in \mathbb{R}^{m \times n}$ a countsketch matrix and $\mathbf{A} \in \mathbb{R}^{n \times d}$, if $m = O\left(\frac{d^2}{\delta \epsilon^2}\right)$, then with probability at least $1 - \delta$:

$$\|\mathbf{S}\mathbf{A}\mathbf{x}\|_2 = (1 \pm \epsilon)\|\mathbf{A}\mathbf{x}\|_2.$$

Countsketch Embedding

Countsketch - subspace embedding

For $\mathbf{S} \in \mathbb{R}^{m \times n}$ a countsketch matrix and $\mathbf{A} \in \mathbb{R}^{n \times d}$, if $m = O\left(\frac{d^2}{\delta \epsilon^2}\right)$, then with probability at least $1 - \delta$:

$$\|\mathbf{S}\mathbf{A}\mathbf{x}\|_2 = (1 \pm \epsilon)\|\mathbf{A}\mathbf{x}\|_2.$$

We use the AMM and JL moment result.

We have $\text{Var}[\|\mathbf{S}\mathbf{y}\|_2^2] \leq \frac{K}{m}$.

If $\frac{K}{m} \leq \epsilon^2 \delta$, we know \mathbf{S} is ϵd -embedding with probability at least $1 - \delta$.

Types of sketching matrices

Sketching matrix	Sketch size m	Cost to sketch \mathbf{SA}
JL - i.i.d subGaussians	$m = O\left(\frac{d \log(1/\delta)}{\epsilon^2}\right)$	$O(mnd)$
Fast JL -SRHT	$m = O\left(\frac{d \log(d) \log(1/\delta)}{\epsilon^2}\right)$	$O(mn \log(d))$
Countsketch	$m = O\left(\frac{d^2}{\delta \epsilon^2}\right)$	$O(nnz(\mathbf{A}))$

We have other sparse embeddings where nnz per column is > 1 , e.g, OSNAPs, sparse graphs.

Can improve $m = O\left(\frac{d \log(d) \log(1/\delta)}{\epsilon^2}\right)$ with $s = \Theta(\log(1/\delta))$ nonzero entries per column.

Further Reading

Countsketch was first introduced by:

- Clarkson, Kenneth L., and David P. Woodruff. “Low-rank approximation and regression in input sparsity time.” *Journal of the ACM (JACM)* 63.6 (2017): 1-45

Above analysis is from:

- Nelson, Jelani, and Huy L. Nguyen. “OSNAP: Faster numerical linear algebra algorithms via sparser subspace embeddings.” *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 2013.

Sketch and solve - least squares regression

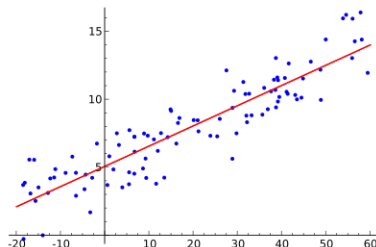
Least squares linear regression

Given a data matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ with n samples $\{\mathbf{a}_i\}_{i=1}^n \in \mathbb{R}^d$ of d -dimensional features, and a column vector $\mathbf{b} \in \mathbb{R}^n$ (targets):

- In the *least-squares* regression problem, assuming $d < n$, we solve:

$$\mathbf{x}^* = \min_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2.$$

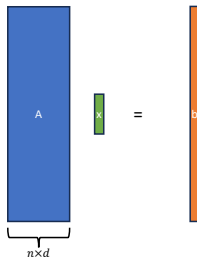
- A linear function and Euclidean- (ℓ_2) norm (squared) loss function.
- The observed targets, $b_i = \mathbf{a}_i^\top \mathbf{x} + \varepsilon_i$, for $i = 1, \dots, n$ and ε_i is noise..



Overdetermined problems

$$\mathbf{x}^* = \min_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2.$$

- We are interested in over-constrained least-squares problems, $n \gg d$.
- Typically, there is no \mathbf{x}^* such that $\mathbf{A}\mathbf{x}^* = \mathbf{b}$.
- Want to find the “best: \mathbf{x}^* such that $\mathbf{A}\mathbf{x}^* \approx \mathbf{b}$.



Exact solution and ϵ -approximation

- The solution is given by the psuedo-inverse $\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{b} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$.
- In terms of SVD, we have $\mathbf{A}^\dagger = \mathbf{V} \Sigma^{-1} \mathbf{U}^\top$, and
- QR factorization, we have $\mathbf{A}^\dagger = \mathbf{R}^{-1} \mathbf{Q}^\top$.

Complexity is $O(nd^2)$, but constant factors differ.

Exact solution and ϵ -approximation

- The solution is given by the psuedo-inverse $\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{b} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$.
- In terms of SVD, we have $\mathbf{A}^\dagger = \mathbf{V} \Sigma^{-1} \mathbf{U}^\top$, and
- QR factorization, we have $\mathbf{A}^\dagger = \mathbf{R}^{-1} \mathbf{Q}^\top$.

Complexity is $O(nd^2)$, but constant factors differ.

ϵ -approximation

For an error parameter ϵ , compute $\tilde{\mathbf{x}}$ such that

$$\|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}\|_2 \leq (1 + \epsilon) \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2$$

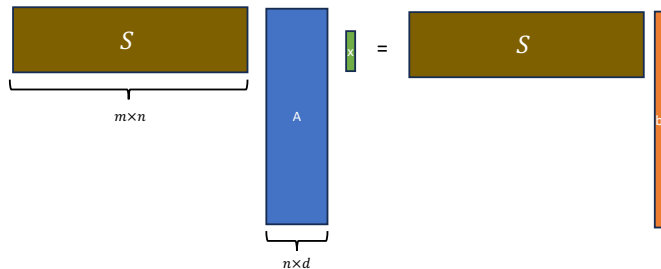
Sketch and solve

Use *Sketching*:

- Generate a sketching matrix $\mathbf{S} \in \mathbb{R}^{m \times n}$.
- Compute sketches \mathbf{SA} and \mathbf{Sb} .
- Solve:

$$\tilde{\mathbf{x}} = \min_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{SAx} - \mathbf{Sb}\|_2^2.$$

- Typically, $m = \text{poly}(d/\epsilon)$.



Recall: subspace embedding

Subspace embedding

For $\mathbf{A} \in \mathbb{R}^{n \times d}$, a matrix $\mathbf{S} \in \mathbb{R}^{m \times n}$ is a subspace ϵ -embedding for \mathbf{A} if \mathbf{S} is an ϵ -embedding for $\text{span}(\mathbf{A}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^d\}$. I.e., for all $\mathbf{x} \in \mathbb{R}^d$,

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

Sketching matrix	Sketch size m	Cost to sketch \mathbf{SA}
JL - i.i.d subGaussians	$m = O\left(\frac{d \log(1/\delta)}{\epsilon^2}\right)$	$O(mnd)$
Fast JL -SRHT	$m = O\left(\frac{d \log(d) \log(1/\delta)}{\epsilon^2}\right)$	$O(mn \log(d))$
Countsketch	$m = O\left(\frac{d^2}{\delta \epsilon^2}\right)$	$O(nnz(\mathbf{A}))$

Subspace embedding for sketch and solve

Sketch and solve

Suppose $\mathbf{S} \in \mathbb{R}^{m \times n}$ is a subspace ϵ -embedding for $\text{span}([\mathbf{A} \ \mathbf{b}])$.

Let,

$$\mathbf{x}^* = \min_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$$

$$\tilde{\mathbf{x}} = \min_{\mathbf{x} \in \mathbb{R}^d} \|\mathbf{S}(\mathbf{A}\mathbf{x} - \mathbf{b})\|_2,$$

for $\epsilon \leq 1/3$, we have

$$\|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}\|_2 \leq (1 + 3\epsilon)\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2$$

Proof:

For $\mathbf{y} = \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix}$, $\mathbf{x} \in \mathbb{R}^d$,

$$\|\mathbf{S}(\mathbf{A}\mathbf{x} - \mathbf{b})\|_2 =$$

Proof:

For $\mathbf{y} = \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix}$, $\mathbf{x} \in \mathbb{R}^d$,

$$\|\mathbf{S}(\mathbf{Ax} - \mathbf{b})\|_2 =$$

$$\|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}\|_2 \leq$$

Proof:

For $\mathbf{y} = \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix}$, $\mathbf{x} \in \mathbb{R}^d$,

$$\|\mathbf{S}(\mathbf{A}\mathbf{x} - \mathbf{b})\|_2 =$$

$$\|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}\|_2 \leq$$

and so for $\epsilon \leq 1/3$, $\|\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}\|_2 \leq (1 + 3\epsilon)\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2$.

Computational cost:

Matlab demo