Lecture 9:

Fast Dimension Reduction Sketching

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Plan

- PS2 due tomorrow, 7pm
- My office hours after class

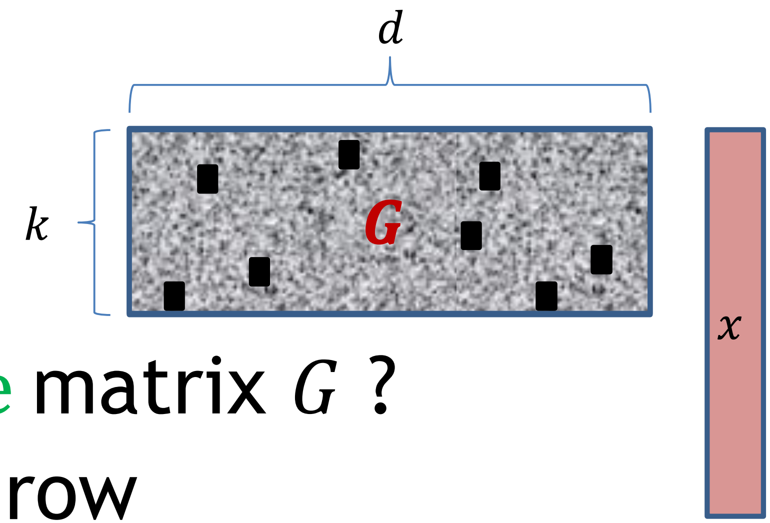
- Fast Dimension Reduction
- Sketching

- Scriber?
  - Due on Fri eve

Johnson Lindenstrauss Lemma

- $F(x) = \frac{1}{\sqrt{k}} Gx = (g_1 \cdot x, g_2 \cdot x, \ \ldots g_k \cdot x) \ / \sqrt{k}$
  - $||F(x)|| = \ (1 \pm \epsilon)||x||$ with probability $\geq 1 - \delta$
  - for $k = O\left(\frac{1}{\epsilon^2} \log \frac{1}{\delta}\right)$
- Time to compute Gx :
  - $O(kd)$
- Faster?
  - $O(d + k)$ time ?
  - Will show: $O(d \log d + k^3)$ time

Fast JL Transform

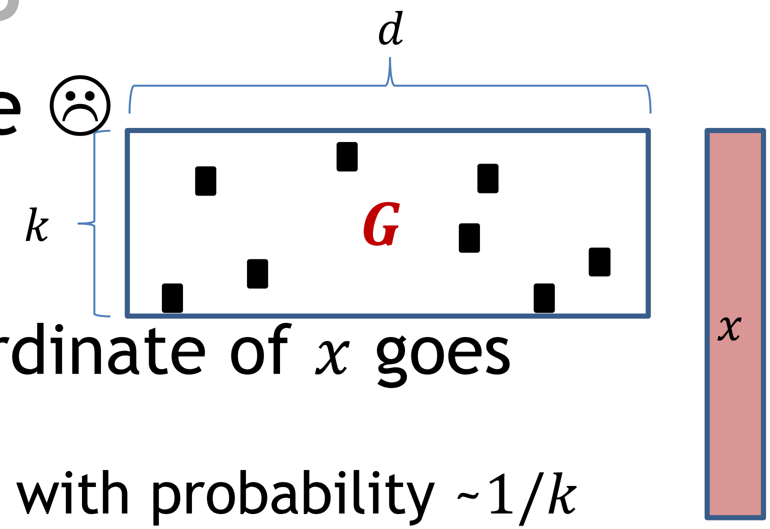


- $z = Gx$
- Costly because G is dense
- Meta-approach: use sparse matrix G ?
- Suppose sample s entries/row
- Analysis of one row:
  - $h: [d] \to \{0,1\}$ s.t. $h(i) = 1$ with probability s/d
  - $z_1 = \eta \cdot \sum_{i=1}^{d} h(i) \cdot g_i x_i$
  - Expectation of z_1^2:
  - $E[z_1^2] = \eta^2 \, E\left[\sum_i h(i) g_i^2 \, x_i^2\right] = \eta^2 \cdot \frac{s}{d} \cdot ||x||^2$
  - What about variance?

Set $\eta = \sqrt{d/s}$

normalization constant

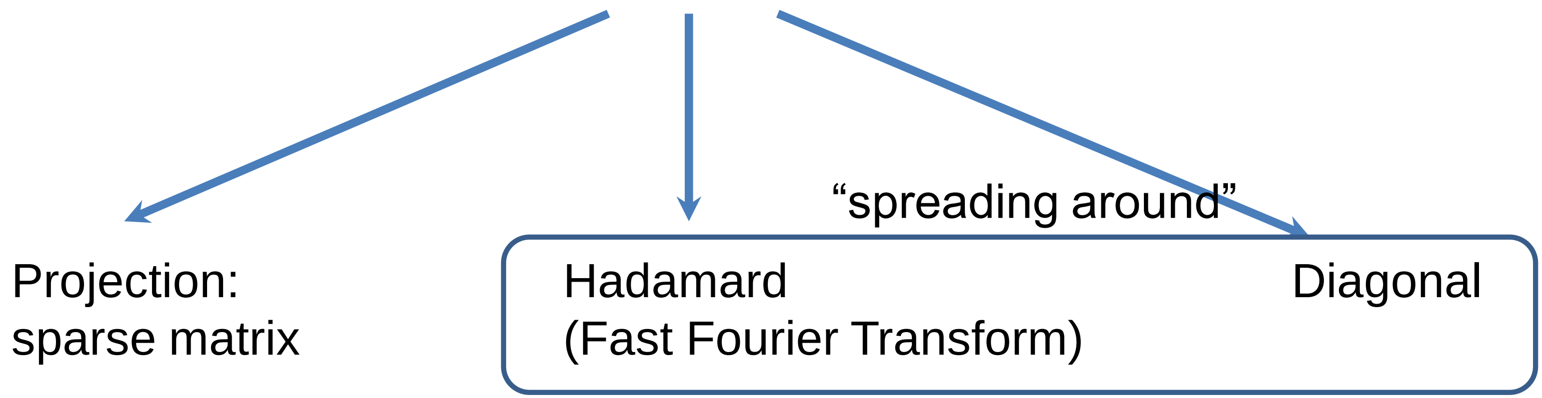
Fast JLT: sparse projection



- Variance of z_1 can be large ☹
  - Bad case: x is sparse
    - think: $x = e_1 - e_2$
  - Even for $s \approx d/k$ (each coordinate of x goes somewhere)
    - two coordinates collide (bad) with probability ~$1/k$
    - want exponential in k failure probability
    - really would need $s \approx d$
- But, take away: may work if x is "spread around"
- New plan:
  - "spread around" x
  - use sparse G

FJLT: construction

$$z = PHD \cdot x$$



- D = matrix with random ± 1 on diagonal
- H= Hadamard matrix (Fourier transform)
  - A non-trivial rotation
  - Hx can be computed in time $O(d \cdot \log d)$
- P = projection matrix: sparse matrix as before, with size $k' \times d$, with $k' \approx k^2$

Spreading around: intuition

$$z = PHD \cdot x$$

Projection:
sparse matrix

"spreading around"

Hadamard
(Fast Fourier Transform)

Diagonal

- $y = HDx$
- Idea for Hadamard/Fourier Transform:
  - "Uncertainty principle": if the original x is sparse, then the transform is dense!
  - Though can "break" x's that are already dense

$$H_1 = 1$$

$$H_{2^l} = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{2^{l-1}} & H_{2^{l-1}} \\ H_{2^{l-1}} & -H_{2^{l-1}} \end{pmatrix}$$

H_d composed of $\pm \frac{1}{\sqrt{d}}$

Spreading around: proof

- $y = HDx$
- Suppose $||x|| = 1$
  - Without loss of generality since the map is linear!
- Ideal spreading around:
  - would like $||y|| = 1$, and
  - $y_i^2 = \frac{1}{d}$ for all i
- Lemma: $y_i^2 \le \frac{1}{d} \cdot O\left(\log\frac{1}{\delta}\right)$ with probability at least $1 - \delta$, for each coordinate i
- Proof:
  - $y_i = H_i D x = rx$
    - where $r = H_i D$ is a random ± 1 vector, times $1/\sqrt{d}$!
    - as mentioned before, rx "behaves like" gx, for Gaussian g

      (needs proof: at the end of the lecture if time permits)
  - Hence $y_i^2 \le \frac{1}{d} \cdot O\left(\log\frac{1}{\delta}\right)$ with probability $\ge 1 - \delta$

Why projection P ?

$$z = PHDx$$

- Why aren't we done?
  - choose first few coordinates of $y = HDx$?
  - each has same distribution:
    - Roughly $||x|| \times$ gaussian
  - Issue:
    - $y_1, y_2, \dots$ are not independent!
- Nevertheless:
  - $||y|| = ||x||$ since HD is a change of basis (rotation in $\Re^d$)

Projection P

$$z = PHDx$$

- So far: $y = HDx$
  - $m = \max y_i^2 \leq \frac{1}{d} \cdot O\left(\log \frac{1}{\delta}\right)$ with probability $1 - d\delta$
  - Or: $m \leq \frac{1}{d} \cdot O\left(\log \frac{d}{\delta}\right)$ with probability $1 - \delta$
- P = projection onto just k' random coordinates!
  - $s = 1$
- Proof: standard concentration
  - $y_1^2 + y_2^2 + \cdots + y_d^2 = ||x||^2 = 1$
  - Chernoff: enough to sample $O\left(dm \cdot \frac{1}{\epsilon^2} \cdot \log \frac{1}{\delta}\right)$ terms for $1 + \epsilon$ approximation
  - Hence $k' = O\left(\log \frac{d}{\delta} \cdot \frac{1}{\epsilon^2} \log \frac{1}{\delta}\right)$ suffices

FJLT: wrap-up

$$z = PHDx$$

- Obtain:
  - $||z||^2 = (1 \pm \epsilon)||x||^2$ with probability $\geq 1 - 2\delta$
  - dimension of z is $k' = O\left(\log\frac{d}{\delta} \cdot \frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$
  - time: $O(d \log d + k')$
- Dimension k' not optimal:
  - apply regular (dense) JL on z
  - to reduce further to $k = O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$
- Final time: $O(d \log d + kk') = O(d \log d + k^3)$