Outline

1. An Example Communication Lower Bound – Randomized 1-way Communication Complexity of the INDEX problem

2. Projection onto Complicated Objects and Gaussian Mean Width

3. Compressed Sensing

Randomized 1-Way Communication Complexity

INDEX PROBLEM

$x \in \{0,1\}^n$

$j \in \{1, 2, 3, \ldots, n\}$

- Alice sends a single message M to Bob
- Bob, given M and j, should output x_j with probability at least 2/3
- Note: The probability is over the coin tosses, not inputs
- Prove that for some inputs and coin tosses, M must be $\Omega(n)$ bits long…

1-Way Communication Complexity of Index

- Consider a uniform distribution μ on X
- Alice sends a single message M to Bob
- We can think of Bob's output as a guess X_j' to X_j
- For all j, $\Pr\left[\mathrm{X}_j' = \mathrm{X}_j\right] \geq \frac{2}{3}$

- By Fano's inequality, for all j,

$$\mathrm{H}\left(\mathrm{X}_j \middle| \mathrm{M}\right) \leq \mathrm{H}\left(\frac{2}{3}\right) + \frac{1}{3}(\log_2 2 \ - 1) = \mathrm{H}(\frac{1}{3})$$

1-Way Communication of Index Continued

- Consider the mutual information $I(M ; X)$
- By the chain rule,

$$I(X ; M) = \Sigma_i \, I(X_i ; M \mid X_{<i})$$
$$= \Sigma_i \, H(X_i \mid X_{<i}) - H(X_i \mid M , X_{<i})$$

- Since the coordinates of X are independent bits, $H(X_i \mid X_{<i}) = H(X_i) = 1$.
- Since conditioning cannot increase entropy,

$$H(X_i \mid M , X_{<i}) \leq H(X_i \mid M)$$

So, $I(X ; M) \geq n - \sum_i H(X_i|M) \geq n - H\left(\frac{1}{3}\right) n$

So, $|M| \geq H(M) \geq I(X ; M) = \Omega(n)$

Typical Communication Reduction

$a \in \{0,1\}^n$
Create stream s(a)

$b \in \{0,1\}^n$
Create stream s(b)

Lower Bound Technique

1. Run Streaming Alg on s(a), transmit state of Alg(s(a)) to Bob

2. Bob computes Alg(s(a), s(b))

3. If Bob solves g(a,b), space complexity of Alg at least the 1-way communication complexity of g

Example: Distinct Elements

- Given $a_1, \ldots, a_m$ in $[n]$, how many *distinct* numbers are there?

- Index problem:
  - Alice has a bit string x in $\{0, 1\}^n$
  - Bob has an index i in $[n]$
  - Bob wants to know if $x_i = 1$

- Reduction:
  - $s(a) = i_1, \ldots, i_r$, where i_j appears if and only if $x_{i_j} = 1$
  - $s(b) = i$
  - If $\text{Alg}(s(a), s(b)) = \text{Alg}(s(a))+1$ then $x_i = 0$, otherwise $x_i = 1$

- Space complexity of Alg at least the 1-way communication complexity of Index

Strengthening Index: Augmented Indexing

- Augmented-Index problem:
  - Alice has $x \in \{0, 1\}^n$
  - Bob has $i \in [n]$, and $x_1, \ldots, x_{i-1}$
  - Bob wants to learn x_i

- Similar proof shows $\Omega(n)$ bound
- $I(M ; X) = \text{sum}_i\, I(M ; X_i \mid X_{< i})$

  $= n - \text{sum}_i\, H(X_i \mid M, X_{< i})$

- By Fano's inequality, $H(X_i \mid M, X_{< i}) \leq H(\delta)$ if Bob can predict X_i with probability $\geq 1- \delta$ from $M, X_{< i}$
- $CC_\delta(\text{Augmented-Index}) \geq I(M ; X) \geq n(1-H(\delta))$

Log n Bit Lower Bound for Estimating Norms

- Alice has $x \in \{0,1\}^{\log n}$ as an input to Augmented Index
- She creates a vector v with a single coordinate equal to $\sum_j 10^j x_j$
- Alice sends to Bob the state of the data stream algorithm after feeding in the input v
- Bob has i in [log n] and $x_{i+1}, x_{i+2}, \ldots, x_{\log n}$
- Bob creates vector w = $\sum_{j>i} 10^j x_j$
- Bob feeds –w into the state of the algorithm
- If the output of the streaming algorithm is at least $10^i/2$, guess $x_i = 1$, otherwise guess $x_i = 0$

$\frac{1}{\epsilon^2}$ Bit Lower Bound for Estimating Norms

$x \in \{0,1\}^n$ $y \in \{0,1\}^n$

- Gap Hamming Problem: Hamming distance $\Delta(x,y) > n/2 + 2\varepsilon n$ or $\Delta(x,y) < n/2 + \varepsilon n$
- Lower bound of $\Omega(\varepsilon^{-2})$ for randomized 1-way communication [Indyk, W], [W], [Jayram, Kumar, Sivakumar]
- Gives $\Omega(\varepsilon^{-2})$ bit lower bound for approximating any norm
- Same for 2-way communication [Chakrabarti, Regev]

Gap-Hamming From Index [JKS]

Public coin = $r^1, \ldots, r^t$, each in $\{0,1\}^t$

$t = \Theta(\epsilon^{-2})$

$x \in \{0,1\}^t$ → $a \in \{0,1\}^t$

$i \in [t]$ → $b \in \{0,1\}^t$

$a_k = \text{Majority}_{j \text{ such that } x_j = 1}\, r^k_j$

$b_k = r^k_i$

$E[\Delta(a,b)] = t/2 + x_i \cdot t^{1/2}$

Aspects of 1-Way Communication of Index

- Alice has $x \in \{0,1\}^n$
- Bob has $i \in [n]$
- Alice sends a (randomized) message M to Bob
- $I(M ; X \mid R) = \text{sum}_i\ I(M ; X_i \mid X_{< i}, R)$

  $\geq \text{sum}_i\ I(M; X_i \mid R)$

  $= n - \text{sum}_i\ H(X_i \mid M, R)$
- Fano: $H(X_i \mid M, R) \leq H(\delta)$ if Bob can guess X_i with probability $> 1- \delta$
- $CC_\delta(\text{Index}) \geq I(M ; X \mid R) \geq n(1-H(\delta))$

The same lower bound applies if the protocol is only correct on average over x and i drawn independently from a uniform distribution

Distributional Communication Complexity

f(X,Y)?

X Y

- $(X, Y) \sim \mu$

- μ-distributional complexity $D_{\mu}(f)$: the minimum communication cost of a protocol which outputs f(X,Y) with probability 2/3 for $(X, Y) \sim \mu$
  - Yao's minimax principle: $\mathrm{R(f)} = \max_{\mu} D_{\mu}\mathrm{(f)}$

- 1-way communication: Alice sends a single message M(X) to Bob

Indexing is Universal for Product Distributions [Kremer, Nisan, Ron]

- Communication matrix A_f of a Boolean function $f: X \times Y \to \{0,1\}$ has (x,y)-th entry equal to f(x,y)

- $\max_{\text{product } \mu} D_\mu(f) = \Theta(\text{VC} - \text{dimension})$ of A_f

- Implies a reduction from Index is optimal for product distributions

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Indexing with Low Error

- Index Problem with 1/3 error probability and 0 error probability both have $\Omega(n)$ communication

- Sometimes, want lower bounds in terms of error probability

- Indexing on Large Alphabets:
  - Alice has $x \in \{0,1\}^{n/\delta}$ with $wt(x) = n$, Bob has $i \in [n/\delta]$
  - Bob wants to decide if $x_i = 1$ with error probability δ
  - [Jayram, W] 1-way communication is $\Omega(n \log(1/\delta))$
  - Can be used to get an $\Omega(\log\left(\frac{1}{\delta}\right))$ bound for norm estimation
  - We've seen an $\Omega(\log n + \epsilon^{-2} + \log\left(\frac{1}{\delta}\right))$ lower bound for norm estimation
  - There is an $\Omega(\epsilon^{-2} \log\frac{1}{\delta} \log n)$ bit lower bound

Outline

1. An Example Communication Lower Bound – Randomized 1-way Communication Complexity of the INDEX problem

2. Projection onto Complicated Objects and Gaussian Mean Width

3. Compressed Sensing

Projection onto other Objects

- Least squares regression finds the closest point y in a subspace K to a given point b
- Given a (possibly infinite) set of points K, and a point b, compute $\min_{y \in K} |y - b|$
  - All norms are Euclidean norms
- Let S be a sketching matrix, we want that if y' = arg$\min_{y \in K} |Sy - Sb|$, then

$$|y' - b| \leq (1 + \epsilon) \min_{y \in K} |y - b|$$

- More generally, want to preserve distances of all vectors in a set K, that is, |S(y-y')| = $(1 \pm \epsilon)|y - y'|$ for all y, y' ∈ K

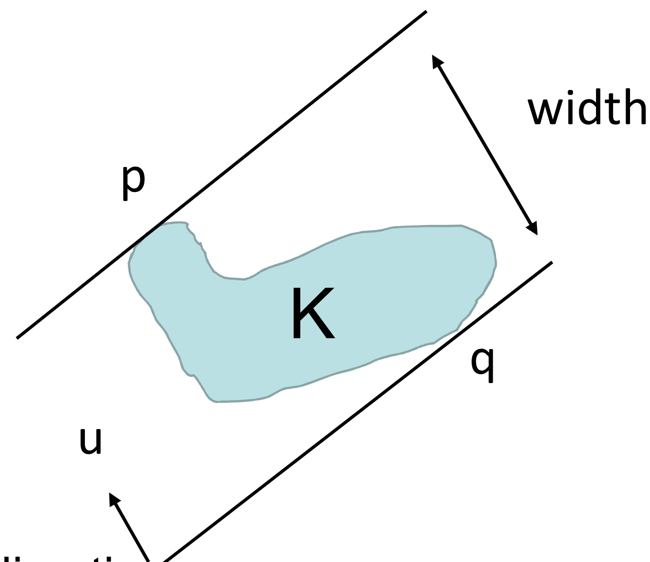
What properties of K determine the dimension and sparsity of S?

Example: Preserving Distances in a Set

- More generally, want to preserve distances of all vectors in a set K, that is, |S(y-y')| = $(1 \pm \epsilon)|y - y'|$ for all y, y' ∈ K
- What is the dimension of S needed if K is:
  - n arbitrary points in R^d?
  - n arbitrary points on a line in R^d?

Spherical Mean Width

- Let K be a bounded subset in R^n
- Consider the width in direction u for a unit vector u:



- Width in direction u = $\sup_{p,q \text{ in } K} < u, p - q >$
- Spherical mean width = $E_u[\sup_{p,q \text{ in } K} < u, p - q >]$

Gaussian Mean Width

- Let $g \sim N(0, I_n)$ be an i.i.d. Gaussian vector
- Gaussian mean width g(K) = $E_g[\sup_{p,q \text{ in } K} < g, p - q >]$
  $= \Theta(n^{.5}) \cdot$ spherical mean width
- Examples
  - K = S^{n-1}
    - $\Theta(n^{.5})$
  - K = set of unit vectors in a d-dimensional subspace of R^n
    - $\Theta(d^{.5})$
  - K = t arbitrary unit vectors in R^n
    - $\Theta(\log^{.5} t)$

Gaussian Mean Width of t Arbitrary Unit Vectors

- Let $u^1, \ldots, u^t$ be t arbitrary unit vectors in R^n
- Let g in R^n have iid N(0,1) entries
- Define random variables $Z_j = < u^j, g >$ which are N(0,1) random variables
- Want to bound $E_g[\max_j Z_j]$
- Fact: for an N(0,1) random variable W, $E\left[e^{\lambda W}\right] = e^{\lambda^2/2}$
- For any $\lambda > 0, E\left[e^{\lambda \max_j Z_j}\right] \leq \sum_j E\left[e^{\lambda Z_j}\right] \leq t\, e^{\lambda^2/2}$
- For all $\lambda > 0,\ E_g[\max_j Z_j] \leq \left(\frac{1}{\lambda}\right) \log E[e^{\lambda \max_j Z_j}] \leq (\frac{\log t}{\lambda} + \frac{\lambda}{2}) = 2\sqrt{\log t}$

Sketching Bounds

- [Gordon] Let K be a subset of S^{n-1}. A random Gaussian matrix S with $g(K)^2/\epsilon^2$ rows satisfies

$$|S(y-y')|^2 = (1 \pm \epsilon)|y-y'|^2 \text{ for all y, y' in K}$$

- What about sparse sketching matrices S?

- [Bourgain, Dirksen, Nelson] S can have m = $g(K)^2 \text{poly}(\log n)/\epsilon^2$ rows and $s = \text{poly}(\log n)/\epsilon^2$ non- zeros per column if m and s satisfy a condition related to higher moments of $\sup_{p,q} < g, p-q >$
  - Applied to finite and infinite unions of subspaces

Outline

1. An Example Communication Lower Bound – Randomized 1-way Communication Complexity of the INDEX problem

2. Projection onto Complicated Objects and Gaussian Mean Width

3. Compressed Sensing

Compressed Sensing

- We take random "linear measurements" of an n-dimensional vector x
- In our language, we choose a random r x n sketching matrix S and observe $S \cdot \mathrm{x}$
- Output a vector x' with $|\mathrm{x} - \mathrm{x}'|_\mathrm{p} = \mathrm{D} \cdot \min_{\mathrm{k-sparse\ z}} |\mathrm{x} - \mathrm{z}|_\mathrm{q}$, where D is the distortion (the $\ell_\mathrm{p}/\ell_\mathrm{q}$-guarantee)
- Let $\mathrm{x_k}$ be the best k-sparse approximation to x, i.e., the largest k coordinates in magnitude
- Randomized ("for-each") scheme versus deterministic ("for-all") scheme
- CountSketch is a randomized scheme achieving ℓ_2/ℓ_2 w.h.p.

$$|\mathrm{x} - \mathrm{x}'|_2 = O(1) \cdot |\mathrm{x} - \mathrm{x_k}|_2$$

CountSketch for Compressed Sensing

- CountSketch had O(log n) repetitions of hashing into O(k) buckets
- S is a random linear map S with O(k log n) rows
- For an n-dimensional vector x, estimate every x_i up to additive error $\frac{|x-x_k|_2}{\sqrt{k}}$
- Output a 2k-sparse x' consisting of the top 2k estimates given by CountSketch
- Say coordinate i is heavy if $|x_i| \geq |x - x_k|_2/\sqrt{k}$
  - How many heavy coordinates can there be?
- Say a coordinate i is super-heavy if $|x_i| \geq 3|x - x_k|_2/\sqrt{k}$
  - Claim: the set T of super-heavy coordinates is in the support of x'
- $|x - x'|_2 \leq |(x - x')_T|_2 + \left|(x - x')_{[n]\setminus T}\right|_2$

$$\leq \sqrt{2k} \cdot \frac{|x-x_k|_2}{\sqrt{k}} + |(x - x_k)_{[n]\setminus T}|_2 + \left|(x_k - x')_{[n]\setminus T}\right|_2 = O(|x - x_k|_2)$$

No Deterministic Algorithm Achieves ℓ_2/ℓ_2

- Recall ℓ_2/ℓ_2: output x' with $|x - x'|_2 = O(1) \cdot |x - x_k|_2$
- Consider k = 1
- Suppose S is a deterministic sketching matrix with r = o(n) rows
- Suffices to show there is a vector x in kernel(S) with $|x|_\infty \geq C|x - x_1|_2$ for any constant C > 0
- W.l.o.g., can assume S has orthonormal rows
- $\sum_i |Se_i|_2^2 = r$, so there exists an i with $|Se_i|_2^2 \leq \frac{r}{n}$
- Let $x = e_i - S^T S e_i$, so x is in kernel(S)
- But $|x|_\infty^2 \geq |x_i|^2 = \left(e_i^T e_i - e_i^T S^T S e_i\right)^2 \geq \left(1 - \frac{r}{n}\right)^2$, while
- $|x - x_1|_2 \leq |x - e_i|_2 = \left|S^T S e_i\right|_2 = |Se_i|_2 \leq \sqrt{\frac{r}{n}} = o(1)$

Deterministic Algorithms Achieve ℓ_2/ℓ_1

- ℓ_2/ℓ_1: output x' with $|x - x'|_2 = O(1/k^{.5}) \cdot |x - x_k|_1$
- S has the (ϵ, k)-restricted isometry property (RIP) if for all k-sparse vectors x,

$$(1-\epsilon)|x|_2^2 \leq |Sx|_2^2 \leq (1+\epsilon)|x|_2^2$$

- What are some matrices S with O(k log(n/k)) rows that have the (ϵ, k)-RIP property for constant ϵ?
- Deterministic, but not explicit!
- Major open question: explicit matrix with (ϵ, k)-RIP with $o(k^2)$ rows
- Bourgain et al.: can get $k^{2-\gamma}$ rows for a constant $\gamma > 0$ and $k \approx n^{.5}$

Deterministic Algorithms Achieve ℓ_2/ℓ_1

- If S has the (ϵ, k)-RIP then one can efficiently output an x' for which
$$|x - x'|_2 = O(1/k^{.5}) \cdot |x - x_k|_1$$

- In fact, can just solve a linear program!

$$\min_{z \in R^n} |z|_1$$
s.t. Sz = Sx

- If x' is the solution, then $|x - x'|_2 \leq O\left(\frac{1}{k^{.5}}\right) |x - x_k|_1$

- Proof uses (ϵ, k)-RIP and elementary norm manipulations