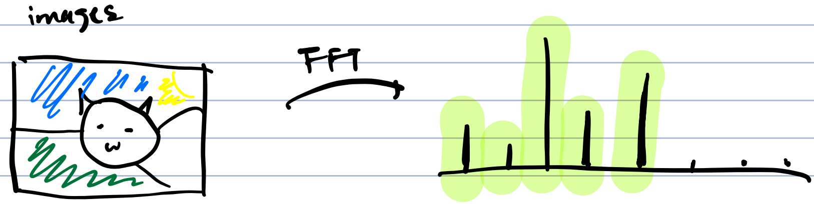
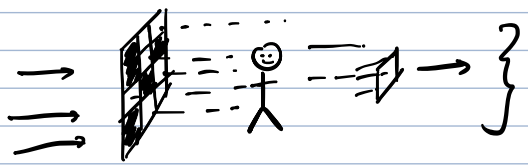
Lecture 13: Compressive sensing (aka compressed sensing) and linear programming.

How to find a sparse explanation of data.

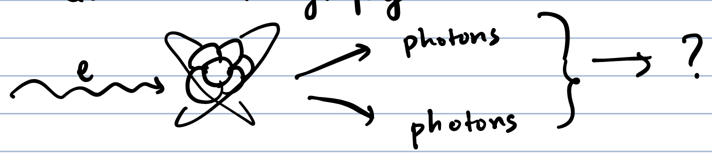
e.g. images



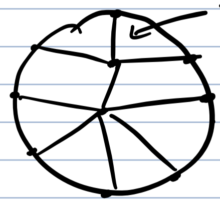
tend to be Fourier sparse.



- MRI → noisy measurements of Fourier signals through body, want to find underlying image.
- Quantum tomography



- Network tomography



individual delays.

send probes between boundary nodes.
total delay = Σ delays on path.

[Firooz, Roy '10]

UW EE

Framework: $x \in \mathbb{R}^N$ is an unknown signal. We are given n linear measurements $a_1, \dots, a_n \in \mathbb{R}^N$, $n \ll N$, i.e. we observe $b_1, \dots, b_n$, where $b_i = \langle a_i, x \rangle$.

Goal: given $\{(a_i, b_i)\}$, recover x.

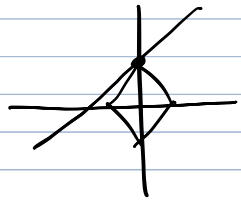
This is just linear regression! With no assumptions on x, this is impossible if $n \ll N$.

$$\underbrace{\begin{bmatrix} — a_1 — \\ — a_2 — \\ \vdots \\ — a_n — \end{bmatrix}}_{A} \begin{bmatrix} \\ x \\ \\ \end{bmatrix} = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}}_{b} \qquad (A)$$

Fact: there is a $(N-n)$-dim subspace of possible solutions.

to get around this, assume sparsity.

Recall: from previous lectures, we said to try ℓ_1-regularization.



This lecture: (somewhat) formalize this intuition.

Recall: $\|x\|_0 = \#\{i : x_i \neq 0\}$ is the sparsity of x. For this lecture, let's assume $\|x\|_0 = k \ll N$.

Thm: [Candes-Tao '04, Candes-Romberg-Tao '04]. Choose a random $n \times N$ matrix A with independent entries $N(0,1)$. If $n \geq \Omega(k \log(n/k))$, then whp over the choice of A, there is a unique k-sparse solution to $(*)$.

Moreover, it is also the solution to:

$$\min \|\hat{x}\|_1 \quad \text{s.t.} \quad A\hat{x} = b. \qquad \text{(basis pursuit)}$$

Remarks:

- very similar to LASSO!
$$\min \|A\hat{x} - b\|^2 + \lambda \|\hat{x}\|_1.$$
- Also works for other types of random matrices, e.g. $\{\pm 1, 0\}$, as long as the entries of A are independent
- For MRIs, the random matrix A is not independent, but has nice Fourier structures, so you can almost recover the same guarantees.
- many, many, many variants of the problem (> 20,000 papers!).
  - faster algos
  - different A
  - sparse A
  - robustness ← [KLLTS'24] :)

"Proof" of correctness

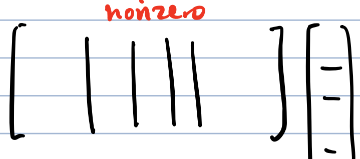
Consider first the following, non-convex problem:

$$\min \|\hat{x}\|_0 \quad \text{s.t.} \quad A\hat{x} = b. \qquad (**)$$

Why does this work?

not standard

Def: We say a matrix A satisfies the no-sparse-kernel property if there is no nonzero $2k$-sparse vector in $\ker(A)$.

 $\neq 0.$

Thm: If A satisfies NSK, then x is the unique solution to $(**)$.

pf: Suppose y is k-sparse and $Ay = b$. Then $A(x-y) = 0$, and

$x - y \in \ker(A)$

$x-y$ is $2k$-sparse.

$\Rightarrow x - y = 0$.

Thm: If A is random, $n = \Omega(k \log(n/k))$, then whp, it satisfies NSK property.

In fact, we can show something even stronger:

Def: We say a matrix A has the (p, ε)-restricted isometry property (RIP) if for all x p-sparse, we have that

$$(1-\varepsilon)\|x\|_2 \le \|Ax\|_2 \le (1+\varepsilon)\|x\|_2.$$

[Candes-Tao '04, Candes-Romberg-Tao '04].

Thm: If A is an appropriately scaled random matrix w/ $n = \Omega\left(\frac{p \log(n/p)}{\varepsilon^2}\right)$, then whp A satisfies the $(2p, \varepsilon)$-RIP.

Cor: A also satisfies NSK whp.

pf of Theorem: This is pretty much Johnson-Lindenstrauss!!

Recall: given m points $\{x_i\}_{i=1}^m$, a (suitably scaled) Gaussian matrix A w/ $n = \frac{\log m}{\varepsilon^2}$ rows satisfies

$$\|A(x_i - x_j)\|_2^2 \approx (1 \pm \varepsilon)\|x_i - x_j\|_2^2 \text{ whp.}$$

pretty much equivalently,

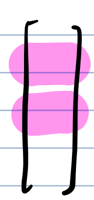
$$\|Ax_i\|_2^2 \approx (1 \pm \varepsilon)\|x_i\|_2^2 \quad \forall i.$$

How big is m? technically, infinite

But like in generalization, $m \approx$ # degrees of freedom.

To specify a vector w/ sparsity p:

1. Specify p coordinates $\longrightarrow \binom{n}{p}$ choices
2. specify values on coordinates $\longrightarrow 2^p$ choices.



$$\log m = \log\left(\binom{n}{p} + 2^p\right)$$
$$\le O\left(\log\binom{n}{p} + p\right)$$

$\binom{n}{p} \le n^p$, but slightly tighter: $\log\binom{n}{p} \le p\log(n/p)$

$\le O(p\log(n/p) + p)$.

What about for ℓ_1-minimization?

need something stronger than NSK property.

<u>Def:</u> We say A has the restricted nullspace property (RNP) if $\forall x \in \ker(A)$, and any set S of $\le k$ coordinates,

$$\sum_{i \in S} |x_i| < \sum_{i \notin S} |x_i|.$$

Fact: ℓ_1-minimization recovers the true support $\Leftrightarrow$ A has RNP.

<u>Fact:</u> If A satisfies $(2k, 1/3)$-RIP, then A satisfies RNP.

So random matrix also works for ℓ_1-minimization / basis pursuit.