Lecture #22

Last Time: Planted Clique and Random Matrix Thy

(Find $c\sqrt{n}$-sized planted cliques in $G(n, \frac{1}{2})$, spectrally)

Today: Compressed Sensing

Our main focus is on sparse recovery:

$$\min \underset{\uparrow \text{ \# nonzeros}}{\|x\|_0} \quad \text{s.t.} \quad \overset{m\times n}{A}x = b \qquad (P_0)$$

when there are many solutions (e.g. $m \ll n$), find the one that is the sparsest.

This problem is NP-hard, but we'll solve it anyways in important special cases via:

$$\min \|x\|_1 \quad \text{s.t.} \quad Ax = b \qquad (P_1)$$

This is now a convex program, e.g. equivalent to:

$$\min \sum y_i \quad \text{s.t.} \quad Ax = b, \; x \le y, \; -x \le y$$

When does solving (P_1) give a good solution to (P_0)?

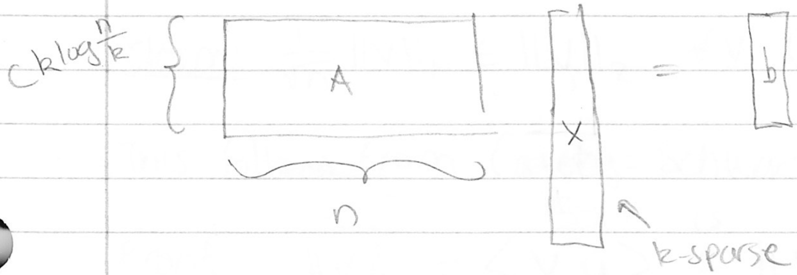
We say that A has the restricted isometry property (RIP) (k, δ_k) if $\forall \|x\|_0 \le k$

$$(1-\delta_k)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta_k)\|x\|_2^2$$

Let's give a sample theorem

Theorem 1 [Candes, Tao] If $Ax = b$ and $\|x\|_0 \le k$, and (*) $\delta_{2k} + \delta_{3k} < 1$ then x is the uniquely optimal solution to (P_1)

Fact: A random matrix A (independent, Gaussian entries) when scaled appropriately will satisfy (*) with $m = \Theta(k \log \frac{n}{k})$



Even though the system is highly underdetermined we can still find the true x, efficiently

Applications: MRI, single pixel camera (take many fewer measurements, reconstruct structured objects)

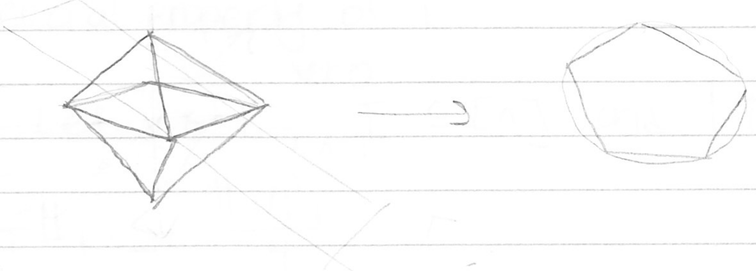
Even though finding the sparsest soln is NP-hard in the worst-case, there are important, practical cases where we can provably do it.

Restricted Isometry $\Longleftrightarrow$ Almost Euclidean Subspace

We say a subspace $\Gamma \subseteq \mathbb{R}^n$ is C-almost Euclidean if $\forall v \in \Gamma$

$$\frac{1}{\sqrt{n}} \|v\|_1 \le \|v\|_2 \le \frac{C}{\sqrt{n}} \|v\|_1$$

The following picture captures what's going on, and we'll unravel it:



We want $\Gamma \cap \{x \mid \|x\|_1 \leq 1\} \approx$ sphere

Claim: $\frac{1}{\sqrt{n}}\|v\|_1 \leq \|v\|_2 \quad \forall v$ (i.e. regardless of Γ)

This follows from Cauchy-Schwartz:

(first put in $\sqrt{n}$)

Proof: $\|v\|_1 = \langle v, \underset{\underset{u_i = \mathrm{sgn}(v_i)}{\uparrow}}{u} \rangle \overset{CS}{\leq} \|v\|_2 \|u\|_2 \leq \|v\|_2 \left(\underset{\underset{\#\text{non zeros}}{\uparrow}}{\sqrt{|\mathrm{supp}(v)|}} \right)$ ☒

Let's explore what properties $v \in \Gamma, v \neq 0$ must have:
Let $s = n/c^2$. Then

Lemma 1: If $v \in \Gamma$ and $v \neq 0$ then $|\mathrm{supp}(v)| \geq s$

Proof: From above

$$\|v\|_1 \leq \|v\|_2 \sqrt{|\mathrm{supp}(v)|} \overset{C\text{-}AE}{\leq} \frac{C}{\sqrt{n}} \|v\|_1 \sqrt{|\mathrm{supp}(v)|}$$

$$\Rightarrow \frac{n}{c^2} \leq |\mathrm{supp}(v)| \quad ☒$$

An almost Euclidean subspace is a real analogue of a linear error correcting code:

$$\mathcal{C} = \{ \overset{n \times k}{A} x \mid x \in \{0,1\}^k \} \quad \text{over } GF(2)$$

we need all non-zero Ax's to have many 1s

One more crucial property of Γ

Lemma 2: Let $v \in \Gamma, \overset{v \neq 0}{\wedge} T \subseteq [n]$ and $|T| \le \frac{s}{16}$. then

$$\|v_T\|_1 \le \frac{\|v\|_1}{4}$$

$\uparrow$

v restricted to coordinates in T

Not only does $v \neq 0$ have many non-zeros, no few coordinates contain majority of ℓ_1-mass

Proof: $\|v_T\|_1 \le \sqrt{|T|}\,\|v_T\|_2 \le \sqrt{|T|}\,\|v\|_2$ (*)

$$(*) \overset{C\text{-AE}}{\le} \frac{C\sqrt{|T|}}{\sqrt{n}}\|v\|_1 \le \frac{\cancel{C}\sqrt{\frac{\cancel{n}}{16\cancel{C^2}}}}{\cancel{\sqrt{n}}}\|v\|_1 = \frac{\|v\|_1}{4} \quad \boxtimes$$

Now we are ready to prove P_1 finds x:

Theorem 2: If $Ax = b$ and $\|x\|_0 \le \frac{\overset{= \frac{n}{c^2}}{s}}{16}$ and $\Gamma = \ker(A)$ is C-almost Euclidean then x is the uniquely optimal soln to (P_1)

Proof: Suppose not. Let w be the optimal soln. Then $w = x + v$, $v \in \Gamma$. Let $T = \text{supp}(x)$. Then

$$\|w\|_1 = \|w_T\|_1 + \|w_{\bar{T}}\|_1 \ge \|x_T\|_1 - \|v_T\|_1 + \|v_{\bar{T}}\|_1$$

$$= \|x\|_1 - 2\|v_T\|_1 + \|v\|_1$$

However since $v \in \Gamma$, $\|v\|_1 - 2\|v_T\|_1 \ge \frac{\|v\|_1}{2}$ Thus

$$\|w\|_1 \geq \|x\|_1 + \frac{\|v\|_1}{2} > \|x\|_1 \qquad \text{since } v \neq 0$$

Thus w is a strictly worse solution then x, and this holds for all $w = x + v$ with $v \in \Gamma$, $v \neq 0$ ☒

Hene almost Euclidean subsections yield algorithms for sparse recovery

But do good almost Euclidean subsections exist?

Theorem: [Kashin, Garnaev, Gluskin] A random subspace $\Gamma \subseteq \mathbb{R}^n$ of $\dim(\Gamma) = n - m$ is C-almost Eucliden with $C \leq \sqrt{\frac{n}{m} \log \frac{n}{m}}$ with high probability.

Thus we can obtain sparse recovery up to

$$\|x\|_0 = \frac{S}{16} = \frac{n}{16C^2} = \Omega\left(\frac{m}{\log \frac{n}{m}}\right)$$

with m measurements.

---

what happens when $Ax + \overset{\text{noise}}{\gamma} = b$ or x is not exactly k-sparse? In the latter case, can we find the k-largest coefficients

Let $\sigma_k(x) = \min_{\|w\|_0 \leq k} \|x - w\|_1$ "error of best k-sparse approximation"

Theorem 3: Let $Ax = b$ with $\Gamma = \ker(A)$ is C-almost Euclidean. Let $S = \frac{n}{C^2}$. If w is the optimal soln to (P_1) then

$$\|x - w\|_1 \leq 4\,\sigma_{\frac{S}{16}}(x)$$

Notice if $\|x\|_0 \le \frac{s}{16}$ this implies $w = x$ as a special case.

But more broadly, w is a good estimate for largest coordinates in x (can think of others as <u>noise</u>).

<u>Proof:</u> Let T be the $\frac{s}{16}$ largest magnitude coordinates of x

$$\|x-w\|_1 = \|(x-w)_T\|_1 + \|(x-w)_{\bar{T}}\|_1$$

$$\le \|(x-w)_T\|_1 + \|x_{\bar{T}}\|_1 + \|w_{\bar{T}}\|_1$$

But $\|w_{\bar{T}}\|_1 = \underbrace{\|w\|_1}_{\le \|x\|_1 \text{ b/c } w \text{ is optimal for } P_1} - \|w_T\|_1$

Thus we get $\|x-w\|_1 \le \|(x-w)_T\|_1 + \underbrace{\|x_{\bar{T}}\|_1 + \|x\|_1 - \|w_T\|_1}_{(*)}$

$$(*) = 2\|x_{\bar{T}}\|_1 + \underbrace{\|x_T\|_1 - \|w_T\|_1}_{\le \|(x-w)_T\|_1}$$

Putting it all together:

$$\|x-w\|_1 \le \underbrace{2\|(x-w)_T\|_1}_{\substack{\le \frac{\|x-w\|_1}{2} \\ \text{by lemma 2} \\ \text{since } x-w \in \Gamma}} + 2\underbrace{\|x_{\bar{T}}\|_1}_{\sigma_{\frac{s}{16}}(x)}$$

Rearranging finishes proof ☑

What about recovering other sorts of structured objects from few measurements?

Netflix problem matrix completion: Random observations of low rank matrix M

Many others, e.g phase retrieval, sparse PCA, tensor completion, etc