Fun with ℓ_1 and ℓ_2

$\|x\|_1 \leq \sqrt{n}\|x\|_2$.

$\|x\|_1 = x \cdot sgn(x) \leq \|x\|_2 \|sgn(x)\|_2 \leq \sqrt{\|x\|_2}$

$\|x\|_1 = x \cdot sgn(x) \leq \|x\|_2 \|sgn(x)\|_2 \leq \sqrt{|supp(x)|}\|x\|_2$

$supp(x)$ is non-zero indices of x.

If concentrated mass, $\|x\|_1 = \|x\|_2$.
$x = (1, 0, 0, \ldots, 0)$.

If spreadout, $\sqrt{n}\|x\|_2 \leq \|x\|_1$.
$x = (1, 1, 1, \ldots, 1)$.

If kind of spread out, $\|x\|_2 \leq \frac{1}{\sqrt{k}}\|x\|_1$.
x has k 1's.

Fixing $\|v\|_2$, sparse vectors have small $\|v\|_1$ norm, dense ones have big $\|v\|_1$ norm.

Compressed Sensing.

Find x with small number of non-zeros using linear measurements.

$Ax = b$.

Application: MRI.

Find x with k-sparse x, i.e., $supp(x) \leq k$.

ℓ_0-minimization.

Extremely "non-convex".

Find solution to $\min \|w\|_1, Ax = b$.

Linear Program! Exercise.

Restricted Isometry Property (RIP) matrices.

Definition: A matrix A is RIP for δ_k if any k-sparse vector x
$(1-\delta_k)\|x\|_2 \leq \|Ax\|_2 \leq (1+\delta_k)\|x\|_2$.

Theorem [Candes-Tao]: For any matrix RIP matrix A with $\delta_{2k} + \delta_{3k} < 1$, for $Ax = b$ with a k-sparse solution, then the solution to $\min \|y\|_1, Ay = b$, has $y = x$.

Fun with ℓ_1 and ℓ_2

$\|x\|_1 \leq \sqrt{n}\|x\|_2$.

$\|x\|_1 = x \cdot sgn(x) \leq \|x\|_2 \|sgn(x)\|_2 \leq \sqrt{\|x\|_2}$

$\|x\|_1 = x \cdot sgn(x) \leq \|x\|_2 \|sgn(x)\|_2 \leq \sqrt{|supp(x)|}\|x\|_2$

$supp(x)$ is non-zero indices of x.

If concentrated mass, $\|x\|_1 = \|x\|_2$.
$x = (1, 0, 0, \ldots, 0)$.

If spreadout, $\sqrt{n}\|x\|_2 \leq \|x\|_1$.
$x = (1, 1, 1, \ldots, 1)$.

If kind of spread out, $\|x\|_2 \leq \frac{1}{\sqrt{k}}\|x\|_1$.
x has k 1's.

Almost Euclidean Nullspace.

Theorem: For a random ± 1, $d \times n$ matrix A, and for any x in $ker(A)$ some $d = \Omega(k \log \frac{n}{k})$ rows, has for any $T \subset [n]$ that
$\|x\|_2 < \frac{\sqrt{1}}{\sqrt{16k}}\|x\|_1$. $(*)$

Intuition: "Mass in x is spread out over k entries."

The nullspace of A, is almost euclidean.
Typical vectors are spread out: every vector is kind of spread out.

The ℓ_1 ball is closer to scaling of ℓ_2 ball for vectors in the null-space.

Idea: Consider random $r \times n$ matrix A over $GF(2)$.
For a vector x in $GF(2)$.
$A \cdot x = 0$, with probability $(1/2)^r$ if r rows.
There are $< X = 2\binom{n}{k}$ vectors x with fewer than k zeros.
If $r > \log(2\binom{n}{k}) = \Theta(k \log \frac{n}{k})$, plus union bound.
$\implies Ax \neq 0$ for all vectors that are k-sparse.

That is, random A has no sparse vectors in null-space.

Note: Parity check matrix of linear code!

Small projection onto small set of coordinates.

Consider A with property, $x \in ker(A)$, has $\|x\|_2 < \frac{1}{16\sqrt{k}}\|x\|_1$.

Lemma: For $v \in ker(A)$, $T \subset [n]$, $|T| < k$,
$\|v_T\|_1 < \frac{\|v\|_1}{4}$.

Proof:
$\|v_T\|_1 \leq \sqrt{|T|}\|v_T\|_2 \leq \sqrt{|T|}\|v\|_2 \leq \sqrt{|T|}\frac{1}{\sqrt{16k}}\|v\|_1 < \frac{1}{4}\|v\|_1$ $\square$

Intuition:
For any $v \in ker(A)$, the amount of mass in any small, k, set of coordinates is small, $\frac{1}{4}v_1$.

Mass is spread out over more than k coordinates.

Optimum is correct!

Want to find: k-sparse solution to $Ax = b$.

Recall: minimize $\|w\|_1$ with $Aw = b$.

Lemma: For $v \in ker(A)$, $T \subset [n]$, $|T| \leq k$,
$\|v_T\|_1 < \frac{\|v\|_1}{4}$.

Idea: any nonzero vector, $v \in ker(A)$ has small projection onto any k coordinates.

Consider solution w. $w = x + v$ where $v \in ker(A)$.

Will prove: $v = 0$ or $w = x$. Contradiction ? Hmmm.
Let T be non-zero coordinates of x.

$$
\begin{aligned}
\|w\|_1 &= \|x+v\|_1 \\
&= \|x_T + v_T\|_1 + \|v_{\overline{T}}\|_1 \quad \|v\| \geq \|v_T\| - \|v_{\overline{T}}\| \implies \\
&\geq \|x_T\|_1 - \|v_T\|_1 + \|v_{\overline{T}}\| \\
&\geq \|x_T\|_1 - \|v_T\|_1 - \|v_T\|_1 + \|v\|_1 \\
&\geq \|x\|_1 - 2\|v_T\|_1 + \|v\|_1 > \|x\|_1.
\end{aligned}
$$

If v is nonzero.

Imperfect Case.

What if x is mostly sparse?

$$\sigma_k(x) = \min_{supp(z) \leq k} \|x - z\|_1$$

"Amount of x outside of k coordinates."

Theorem: If $v \in ker(A) \implies \|v\|_2 \leq \frac{1}{16k}\|v\|_1$, then solution to $\min \|w\|_1, Ax = b$, has $\|x - w\|_1 \leq 4\sigma_k(x)$.

Still have.

Lemma: For $v \in ker(A)$, $T \subset [n]$, $|T| \leq \frac{k}{16}$,
$\|v_T\|_1 < \frac{\|v\|_1}{4}$.

Proof of $\|w - x\| \leq 4\sigma(x)$.

Again: $\sigma_k(x) = \min_{supp(z) \leq k} |x - z|_1$.

Lemma: For $v \in ker(A)$, $T \subset [n]$, $|T| \leq \frac{k}{16}$, $\|v_T\|_1 < \frac{\|v\|_1}{4}$.
Proof of Theorem: T be k largest in magnitude coordinates of x.

$$
\begin{aligned}
\|x - w\|_1 &= \|(x-w)_T\|_1 + \|(x-w)_{\overline{T}}\|_1 \\
&\leq \|(x-w)_T\|_1 + \|x_{\overline{T}}\|_1 + \|w_{\overline{T}}\|_1 \\
&\leq \|(x-w)_T\|_1 + \|x_{\overline{T}}\|_1 + \|w\|_1 - \|w_T\|_1 \quad \text{triangle inequality on } \|w\|
\end{aligned}
$$

$\|w_{\overline{T}}\|_1 = \|w\|_1 - \|w_T\|_1 \leq \|x\|_1$.

$\|x - w\|_1 \leq \|(x-w)_T\|_1 + \|x_{\overline{T}}\|_1 + \|x\|_1 - \|w_{\overline{T}}\|_1$.

$(*) = 2\|x_T\|_1 + \|x_{\overline{T}}\| - \|w_T\|_1 \leq 2\|x_T\|_1 + \|x_{\overline{T}} - w_T\|_1$

$$
\begin{aligned}
\|x - w\|_1 &\leq 2\|(x-w)_T\|_1 + 2\|x_{\overline{T}}\|_1 \\
&\leq 2\frac{\|(x-w)\|}{4} + 2\sigma(x)
\end{aligned}
$$

$\implies \|x - w\|_1 \leq 4\sigma(x)$.

$\square$

Almost Euclidean Matrices Proof.

Theorem:

For a random ± 1, $d \times n$ matrix, and for any x in with Ax some $d = \Omega(k \log \frac{n}{k})$ rows, has for any $T \subset [n]$ that
$\|x_T\|_2 < \frac{\sqrt{1}}{\sqrt{16k}}\|x_T\|_1$. $(*)$

Idea in GF(2):
Random dot product is 0 with probability $1/2$. All r rows 0: $(1/2)^r$.
Union bound over $\binom{n}{k}$ vectors. $\implies \log \binom{n}{k}$ vectors are enough.

Too many vectors. Real proof is fancy.
Discusses distribution of $X \cdot v$ for a vector v
and random ± 1 vector X
Poor Man's proof:
Group coordinates of v until groups of same size.
n_i in each group.
Deviation in group $\leq \sqrt{n_i}/2$ in each group is less than $1/2$.
Probability groups cancel is small.
Lots of rows. So, norm is good on average for each group.
"Few" vectors with most of mass in small set of coordinates.
Union bound over those.

Credits

Moitra, MIT,6.854. Roughgarden, CS168, Stanford.

See Jame Lee, TCS Blog, May 2008 for proof of Almost Euclidean Nature of random subspaces.

Possible Topics.

TODO: Long tailed distributions.

Interior Point Algorithms.

Matrix Concentration/Matrix Experts/Semidefinite Programs.

Coding Theory: Low Density Parity Check Codes or Expander codes.

Auctions. Mechanism Design.