University of Wisconsin-Madison | Lecturer: Benjamin Recht

**CS838 Topics In Optimization:**
**Convex Geometry in High-Dimensional Data Analysis**

February 9, 2010 | Scribe: Jesse Holzer

---

# Lecture 7

## 1 Norms on $\mathbb{R}^n$

We will make use of all of the following norms

$$\|x\| = \|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

$$\|x\|_\infty = \max_{i=1\ldots n} |x_i|$$

**Fact 1.** *If $x$ is $s$-sparse then*

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq \sqrt{s}\|x\|_2 \leq s\|x\|_\infty.$$

These inequalities can all be verified immediately from the definitions of the norms.

## 2 Compressed sensing

Given $p, n, A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^m, \epsilon \geq 0$ the archetypal problem of compressed sensing is the following:

$$\begin{array}{ll} \min & \mathrm{card}(x) \\ \text{s.t.} & \|Ax - b\|_2 \leq \epsilon. \end{array} \tag{1}$$

As a relaxation, or, really, surrogate, of (1), we consider the following problem:

$$\begin{array}{ll} \min & \|x\|_1 \\ \text{s.t.} & \|Ax - b\|_2 \leq \epsilon. \end{array} \tag{2}$$

Now, why is (2) a sensible surrogate for (1)? In a sense, the $l^1$ norm is the best convex relaxation of the card function, which is not convex:

**Fact 2.** *If $f : [-1, 1]^n \to \mathbb{R}$ is convex and $f(x) \leq \mathrm{card}(x)$, then $f(x) \leq \|x\|_1$.*

*Proof (Sid Barman, next lecture).* The main idea is to express $|x|$ as a convex combination of the vertices of $[0, 1]^n$. □

Also, the $l^1$ unit ball is the convex hull of the length-one cardinality-one vectors:

**Fact 3.** $\{x \in \mathbb{R}^n : \|x\|_1 \leq 1\} = \mathrm{conv}\{e_1, -e_1, \ldots, e_n, -e_n\}$

*Proof.* Any nonzero $x$ in the $l^1$ unit ball satisfies

$$x = \sum_i \frac{|x_i|}{\|x\|_1} \operatorname{sgn}(x_i) e_i,$$

and, conversely, if

$$x = \sum_i (c_i - d_i) e_i$$

with $c, d \geq 0$ and $1^T(c + d) = 1$, then $\|x\|_1 = 1^T|c - d| \leq 1$. □

And, we can draw pictures that suggest that a minimum-cardinality solution to a random linear system is likely also to be a minimum $l^1$-norm solution. We now try to formalize this observation.

## 3 Restricted isometry property

**Definition 1.** *Let $A \in \mathbb{R}^{p \times n}, s \geq 0$ . The s-restricted isometry constant $\delta_s$ of $A$ is the smallest nonnegative number $\delta$ such that*

$$(1 - \delta)\|x\|_2 \leq \|Ax\|_2 \leq (1 + \delta)\|x\|_2$$

*for all $x$ with $\operatorname{card}(x) \leq s$. If $\delta_s < 1$, then $A$ is said to be an s-restricted isometry and to have the s-restricted isometry property (RIP).*

**Fact 4.** *If $s \leq s'$ then $\delta_s \leq \delta_{s'}$.*

How might we check $s$-RIP? For $I \subseteq \{1, \ldots, n\}$ let $A_I$ be the submatrix of $A$ consisting of those columns of $A$ with indices in $I$. Check the extreme singular values of all $A_I$, $|I| = s$. Then $\delta_s$ is the smallest number $\delta$ such that all these values lie in $[1 - \delta, 1 + \delta]$.

The point of RIP is that it sometimes ensures that the solution set to (1) is well-behaved. For example:

**Proposition 1.** *Suppose $A$ has $2s$-RIP constant $\delta_{2s} < 1$. If $Ax = b$ has a solution $x_0$ with $\operatorname{card}(x_0) \leq s$, then $x_0$ is the only such solution.*

*Proof.* Suppose there exists $z$ with $\operatorname{card}(z) \leq s$ and $Az = b$. Then $\operatorname{card}(z - x_0) \leq 2s$ so by RIP,

$$0 = \|Az - Ax_0\|_2 = \|A(z - x_0)\|_2 \geq (1 - \delta_{2s})\|z - x_0\|_2$$

so $z = x_0$. □

A subtler example of this principle is the following:

**Theorem 2** (Candes, Romberg, Tao [?])**.** *Suppose $A$ has $4s$-RIP constant $\delta_{4s} \leq \frac{1}{4}$. Suppose*

$$x_0 \in \operatorname{argmin}_x\{\operatorname{card}(x) : Ax = b\}$$

*and*

$$x_1 \in \operatorname{argmin}_x\{\|x\|_1 : Ax = b\}.$$

*If $\operatorname{card}(x_0) \leq s$ then $x_1 = x_0$.*

*Proof.* In the proof of the previous proposition, we used RIP on $z - x_0$. We would like to do so here on

$$r = x_1 - x_0,$$

but we can't since $r$ is not sparse. Instead the strategy will be to use RIP on sparse pieces of $r$, absorbing the smaller such pieces into the largest. So, for any vector $p \in \mathbb{R}^n$ and any index set $I \subseteq \{1, \dots, n\}$, define $p_I \in \mathbb{R}^n$ by

$$(p_I)_j = \begin{cases} p_j & j \in I \\ 0 & j \notin I \end{cases}.$$

Let $I = \operatorname{supp}(x_0)$. Then, since $x_0$ and $r_{I^c}$ have disjoint supports,

$$\|x_0\|_1 \geq \|x_1\|_1 = \|x_0 + r\|_1 \geq \|x_0 + r_{I^c}\|_1 - \|r_I\|_1 = \|x_0\|_1 + \|r_{I^c}\|_1 - \|r_I\|_1,$$

so

$$\|r_{I^c}\|_1 \leq \|r_I\|_1.$$

Now partition $I^c$ into sets $I_k, k = 1, 2, \dots$, of size $3s$ (except for the last one, which may be smaller), so that

$$|r_j| \geq |r_{j'}| \text{ if } j \in I_k \text{ and } j' \in I_{k'} \text{ with } k \leq k'.$$

Then for all $j' \in I_{k+1}$,

$$|r_{j'}| \leq \frac{1}{3s} \sum_{j \in I_k} |r_j|,$$

i.e.

$$\|r_{I_{k+1}}\|_\infty \leq \frac{1}{3s} \|r_{I_k}\|_1,$$

so

$$\|r_{I_{k+1}}\|_2 \leq \sqrt{3s} \|r_{I_{k+1}}\|_\infty \leq \frac{1}{\sqrt{3s}} \|r_{I_k}\|_1.$$

Then, using our earlier estimate of $r_{I^c}$ in terms of $r_I$, we are able to absorb the small pieces of $r$ into the main one, with, crucially, a factor less than 1:

$$\sum_{k \geq 2} \|r_{I_k}\|_2 \leq \frac{1}{\sqrt{3s}} \sum_{k \geq 1} \|r_{I_k}\|_1 = \frac{1}{\sqrt{3s}} \|r_{I^c}\|_1 \leq \frac{1}{\sqrt{3s}} \|r_I\|_1 \leq \frac{1}{\sqrt{3}} \|r_I\|_2.$$

Finally,

$$\begin{aligned} 0 = \|Ax_1 - Ax_0\|_2 &= \left\| A(r_I + r_{I_1}) + \textstyle\sum_{k \geq 2} A r_{I_k} \right\|_2 \\ &\geq \|A(r_I + r_{I_1})\|_2 - \textstyle\sum_{k \geq 2} \|A r_{I_k}\|_2 \\ &\geq (1 - \delta_{4s}) \|r_I + r_{I_1}\|_2 - (1 + \delta_{3s}) \textstyle\sum_{k \geq 2} \|r_{I_k}\|_2 \\ &\geq (1 - \delta_{4s}) \|r_I\|_2 - \tfrac{1}{\sqrt{3}} (1 + \delta_{4s}) \|r_I\|_2 \\ &\geq c(\delta_{4s}) \|r_I\|_2, \end{aligned}$$

where $c(\delta) > 0$ as long as $\delta < (1 - 1/\sqrt{3})/(1 + 1/\sqrt{3}) = 0.26795\dots$. Hence $r_I = 0$, so $r_{I^c} = 0$, so $r = 0$. ☐

## References


[1] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.