On the role of sparsity in compressed sensing and random matrix theory

Roman Vershynin

University of Michigan

December 15, 2009
The concept of sparsity underlies the many recent developments in the area of compressed sensing.

Much earlier, sparsity has been used in a similar way (but often implicitly) in theoretical mathematics; most notably in geometric functional analysis.

Recently, the understanding of the role of sparsity led to formalizing some of the connections between the statements in those areas. This is leading to a new interplay between “pure” and “applied” mathematics.

Applications of “pure” mathematics to compressed sensing are expected and quite common. The reverse direction – applications of compressed sensing to mathematics – is rare.

We shall discuss one such application to the problem of invertibility of random matrices, going back to von Neumann and addressed recently by Tao-Vu and Rudelson-Vershynin.
Sparsity as entropy control

- Sparsity as a way to represent objects in an economical way. A sparse object can be accurately represented with a small number of basis objects.

- Our objects will be vectors in $\mathbb{R}^N$; our basis is the canonical basis of $\mathbb{R}^N$. We then say that a vector $x \in \mathbb{R}^N$ is $s$-sparse if it has few non-zero coordinates:

  $$|\text{supp}(x)| \leq s \ll N.$$ 

  The set of all such unit vectors in $\mathbb{R}^N$ is denoted $\text{Sparse}(N, s)$.

- Sparsity translates into the low metric entropy of the space $\text{Sparse}(N, s)$. Recall that, given a subset $S$ of a metric space and a number $\varepsilon > 0$, the covering number $\mathcal{N}(S, \varepsilon)$ is the smallest cardinality of an $\varepsilon$-net of $S$, i.e. the smallest number of $\varepsilon$-balls needed to cover $S$. 
Sparsity as entropy control

• A simple volume comparison leads to an exponential bound of the metric entropy of many natural subsets of $\mathbb{R}^N$. For example, the whole Euclidean sphere $S^{N-1}$ has entropy

$$\mathcal{N}(S^{N-1}, \varepsilon) \leq (3/\varepsilon)^N, \quad \varepsilon \in (0, 1).$$

• This bound improves significantly for the set of sparse vectors. Since there are $\binom{N}{s} \leq (eN/s)^s$ ways to choose the support of a sparse vector, we have

$$\mathcal{N}(\text{Sparse}(N, s), \varepsilon) \leq \binom{N}{s} \mathcal{N}(S^{N-1}, \varepsilon) \leq (CN/s\varepsilon)^s.$$ 

We see that the sparse vectors have smaller entropy than the whole sphere – essentially exponential in the sparsity $s$ rather than the dimension $N$. This advantage is crucially used in many arguments. Here is an example.
In compressed sensing, a basic quality of matrices is the Restricted Isometry Condition. An $n \times N$ matrix $A$ with $n \leq N$ is said to satisfy the Restricted Isometry Condition (RIC) if $A$ acts as an approximate isometry when restricted to the set of sparse vectors. Formally, for every integer $s \leq N$ the RIC constant $\delta_s$ of the matrix $A$ is the minimal number that satisfies the two-sided inequality

$$(1 - \delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta)\|x\|_2^2$$

for all $x \in \text{Sparse}(N, s)$.

Candes and Tao showed that, if a matrix $A$ satisfies RIC with $\delta_{2s} \leq \sqrt{2} - 1$, one can exactly recover every $s$-sparse vector $x$ from its “measurement vector” $y = Ax$ by solving the convex optimization problem

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad Ax = y.$$
Proposition (RIC for random Gaussian matrices [known])

Let $\tilde{A}$ be an $n \times N$ matrix whose entries are independent standard normal random variables. Let $1 \leq s \leq N$. If $n \gtrsim s \log(N/s)$ then, w.h.p., the matrix $A = \frac{1}{\sqrt{n}} \tilde{A}$ satisfies RIC with $\delta_s \leq 0.01$.

**Proof.** A simple approximation argument shows that it is enough to check RIC for all $x$ in a fixed $\delta$-net $\mathcal{N}$ of Sparse($N, s$), where $\delta \sim 0.001$. As we know, we can choose $|\mathcal{N}| \leq (CN/s)^s$.

For a fixed vector $x \in \mathcal{N}$, the random variable $\|Ax\|_2^2$ is distributed identically with $\chi^2$ with $n$ degrees of freedom. Deviation inequality:

$(1 - \delta)n \leq \chi^2 \leq (1 + \delta)n$ with probability at least $1 - e^{-cn}$.

So, with this probability, RIC holds for a fixed vector $x \in \mathcal{N}$.

Union bound over $\mathcal{N}$ implies that RIC holds for all vectors $x \in \mathcal{N}$ with probability $1 - |\mathcal{N}| e^{-cn} \geq 1 - (CN/s)^s e^{-cn} = 1 - o(1)$.
Sparsity in random matrix theory

- One can view RIC as the condition that all submatrices of $A$ of a given size are well conditioned. The question of how well conditioned random matrices are goes back to Von Neumann and his collaborators, in connection with their work on large matrix inversion.

- In 2007-2008, the original prediction of Von Neumann was verified for general random matrices:

Invertibility Theorem [Rudelson-Vershynin]

Let $A$ be an $N \times N$ matrix whose entries are i.i.d. random variables with mean zero, unit variance, and fourth moment $O(1)$. Then, with high probability, the smallest singular value and the condition number of $A$ satisfy

$$ s_N(A) \sim N^{-1/2}, \quad \kappa(A) \sim N. $$
We demonstrate the role of sparsity in the proof of the Invertibility Theorem. For simplicity, we do this for Gaussian matrices $A$. The smallest singular value can be defined as

$$s_N(A) = \min_{x \in S^{N-1}} \|Ax\|_2.$$

Our goal is then to bound $\|Ax\|_2$ below uniformly for all unit vectors $x$. We first prove this bound sparse vectors $x \in \text{Sparse}(N, cN)$, then for remaining, “spread” vectors.

Begin with sparse vectors. As we proved, Gaussian matrices satisfy RIC. The RIC’s lower bound is nothing else but

$$\min_{x \in \text{Sparse}(N, cN)} \|Ax\|_2 \gtrsim N^{1/2}.$$ 

We have proved a much better bound than the Invertibility Theorem claims: $N^{1/2}$ instead of $N^{-1/2}$!

It remains to handle the “spread” vectors (more difficult)...
- Sparse vectors have low entropy (their advantage).
- Spread vectors have high entropy. Most vectors on the sphere are spread. What is good about spread vectors $x \in S^{N-1}$?
- Their coefficients! Most of them have magnitude $|x_k| \sim N^{-1/2}$. (For simplicity, assume this for all coefficients).

This inspires an alternative, geometric argument to prove invertibility of $A$ on the spread vectors.
A qualitative argument (trivial rank comparison):
If $s_N(A) = 0$ then one of its columns $X_k$ lies in the span $H_k = \text{span}(A_i)_{i \neq k}$ of the others: $\text{dist}(X_k, H_k) = 0$.

Quantitative argument should look something like this:

$$\inf_{x \in S^{N-1}} \|Ax\|_2 \gtrsim \text{dist}(X_k, H_k).$$

This is indeed possible for spread vectors!
Quantitative argument for spread vectors $x \in S^{N-1}$:

$$\|Ax\|_2 \geq \text{dist}(Ax, H_k) = \text{dist}\left(\sum x_i X_i, H_k\right) = \text{dist}(x_k A_k, H_k)$$

$$= |x_k| \cdot \text{dist}(X_k, H_k) \gtrsim N^{-1/2} \text{dist}(X_k, H_k).$$

The right hand side does not depend on $x$; thus

$$\min_{\text{Spread } x} \|Ax\|_2 \gtrsim N^{-1/2} \text{dist}(X_1, H_1).$$
We have shown that $$\min_{\text{Spread } x} \|Ax\|_2 \gtrsim N^{-1/2} \text{ dist}(X_1, H_1).$$

To estimate this distance, note that $X_1$ is a standard Gaussian vector, $H_1$ is an independent hyperplane. Then

$$\text{dist}(X_1, H_1) = |N(0, 1)| = \Omega(1) \text{ with arbitrarily high probability.}$$

This argument fails badly for non-Gaussian vectors. Deeper methods are needed, Littlewood-Offord Theory.
We have thus shown that, with arbitrarily high constant probability,

\[ \min_{\text{Sparse}} \|Ax\|_2 \gtrsim N^{1/2}, \quad \min_{\text{Spread}} \|Ax\|_2 \gtrsim N^{-1/2}. \]

This proves the Invertibility Theorem as claimed:

\[ s_N(A) = \min_{x \in S^{N-1}} \|Ax\|_2 \gtrsim N^{-1/2}. \]
Invertibility Theorem for rectangular matrices [Rudelson-V.]

The random rectangular $n \times N$ matrices satisfy w.h.p.

$$s_N(A) \gtrsim \sqrt{N} - \sqrt{n - 1}.$$  

The following probability bounds hold, given $\varepsilon \in (0, 1)$:

- For square matrices:
  $$\mathbb{P}(s_N(A) < \varepsilon N^{-1/2}) \leq C\varepsilon + e^{-cN}.$$  

- For rectangular matrices:
  $$\mathbb{P}(s_n(A) \leq \varepsilon(\sqrt{N} - \sqrt{n - 1})) \leq (C\varepsilon)^{N-n+1} + e^{-cN}.$$