Analysis of Random Measurements

An introduction to the non-asymptotic theory of random matrices

Roman Vershynin

Department of Mathematics
University of California, Davis

IPAM Short Course, 2007

Goals:

- Short course in the *non-asymptotic theory of random matrices*.
- Methods of *geometric functional analysis*.
- Focus on techniques.

Sources:

- *Handouts*: use as a bibliography guide
- *My webpage*: copy of these slides; notes for my UCD course on Non-asymptotic theory of random matrices (Winter 2007)

Lecture 1. The sparse reconstruction problem and random matrices

1. The sparse reconstruction problem
2. Random matrices: asymptotic and non-asymptotic theories
3. The ε-net method

The sparse reconstruction problem

- Unknown signal: $x \in \mathbb{R}^d$ or $\mathbb{C}^d$.
  Sparse: $|\operatorname{supp}(x)| \le n$, and $n \ll d$.
- Want to reconstruct x from few (say, N) linear measurements.
- Measurements are given by a linear measurement operator

$$\Phi : \mathbb{R}^d \to \mathbb{R}^N; \qquad \text{Measurements} = \Phi x.$$

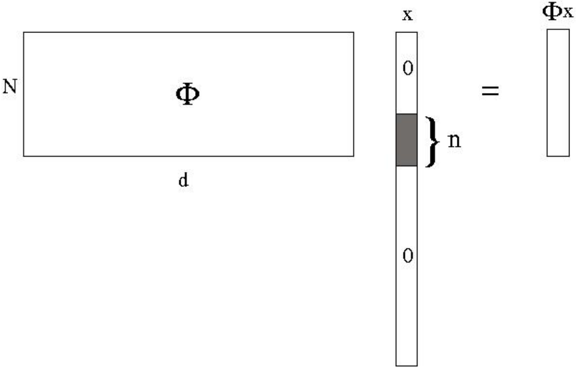
- Trivial to construct $N = d$ measurements (identity operator).
- We hope to achieve $N \ll d$, because x is sparse:
  its *effective* dimension $n \ll$ *nominal* dimension d.
- Lower bound: $N \ge n$. Hopefully, can make N closer to n than to d:

$$N \sim n \log d$$

  is everybody's dream (sometimes achieved).

Measurement matrices

- Signal $x \in \mathbb{R}^d$, which is sparse: $|\operatorname{supp}(x)| \leq n$, and $n \ll d$.
  Reconstruct x from linear measurements Φx, where $\Phi : \mathbb{R}^d \to \mathbb{R}^N$.
  Hope to take *few* measurements: $n \leq N \ll d$.
- Φ realizes dimension reduction (from d to N).
- We identify Φ with the $N \times d$ measurement matrix:



Restricted Isometries

- To reconstruct any n-sparse x from Φx, the measurement map Φ has to be one-to-one on n-sparse vectors.
- Candes-Tao (2004): a slightly stronger condition yields a reconstruction algorithm. A quantitative (rather than qualitative) condition: Φ has to be an almost isometry on the sparse vectors.

Restricted Isometries

Theorem (Candes-Tao, 2004)

Suppose the measurement map Φ is a restricted isometry:

$$0.8\|x\| \le \|\Phi x\| \le 1.2\|x\| \quad \text{for all } 3n\text{-sparse vectors } x.$$

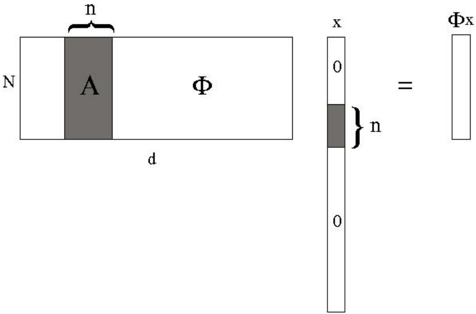
Then x can be reconstructed from the measurements $b = \Phi x$ as the solution to the convex optimization problem

$$\min \|x\|_1 \quad \text{subject to} \quad \Phi x = b.$$

- Meaning of 0.8 and 1.2: close to 1 (1 would be exact isometry).
- Proof: nontrivial, elementary
- Other applications of restricted isometries:
  - Vector quantization [Lyubarskii-V., 2006]
  - Invertibility conjectures on random matrices [Rudelson-V., 2007]. See Von Neumann Symposium talk.

Restricted Isometries

- Equivalently, all $N \times n$ minors of Φ are almost isometries:



- Verifying this is hard: there are $\binom{d}{n} \sim (\frac{n}{d})^n \sim \exp(n \log \frac{n}{d})$ minors.
- Probabilistic method. Draw the measurement matrix Φ at random. Verify the restricted isometry condition for *one* minor. Suppose holds with probability at least $1 - \exp(n \log \frac{n}{d})$. Then take union over all minors; with positive probability will satisfy all.
- The problem reduces to *one* minor = one $N \times n$ random matrix.

Random matrix ensembles

Generate a random $N \times d$ measurement matrix $\bar{\Phi} = \frac{1}{\sqrt{N}}\Phi$, where Φ:

- **Gaussian:** i.i.d. standard normal entries
- **Bernoulli:** i.i.d. symmetric ± 1 entries
- **Projections:** orthog. projection in $\mathbb{R}^d$ onto a random N-dim. subspace
- **Fourier:** random N rows of $d \times d$ Discrete Fourier Transform. ($\Phi x =$ random frequencies of x)

Unifying framework for Gaussian, Bernoulli, bdd.: subgaussian

Definition (Subgaussian random variables)

A random variable X is called subgaussian if its tail is dominated by the gaussian tail:

$$\mathbb{P}(|X| > t) \le 2\exp(-Ct^2) \qquad \text{for all } t > 0.$$

Will also assume variance 1.

From restricted isometries to singular values

- Restricted isometry condition $=$ all minors are almost isometries.
- Minors of subgaussian random matrices are also subgaussian.
- So the problem reduces to studying

  $N \times n$ subgaussian matrix A, normalized: $\bar{A} = \frac{1}{\sqrt{N}} A$.
- Main question: is $\bar{A}$ an approximate isometry?

  $$C_1 \|x\| \le \|\bar{A}x\| \le C_2 \|x\| \quad \text{for all vectors } x,$$

  with $C_1 \approx C_2 \approx 1$ (like 0.8 and 1.2).
- In terms of the singular values of $\bar{A}$ (eigenvalues of $|\bar{A}| = \sqrt{\bar{A}^* \bar{A}}$),
  the best constant C_2 is the largest singular value $\lambda_{max}(\bar{A})$;
  the best constant C_1 is the smallest singular value $\lambda_{min}(\bar{A})$.
- We want to bound $\lambda_{max}(\bar{A})$ above, $\lambda_{min}(\bar{A})$ below,
  with high probability.

Asymptotic theory of random matrices

- Studies limiting spectral properties of random $N \times n$ matrices A,
- as $n \to \infty$ and the aspect ratio $\frac{n}{N} \to y$.
- During 1980–1993, it was shown for subgaussian i.i.d. matrices:

$$\lambda_{max}(\bar{A}) \to 1 + \sqrt{y}; \qquad \lambda_{min}(\bar{A}) \to 1 - \sqrt{y} \qquad \text{a.s.}$$

  Thus close to 1 for small y (tall matrices are almost isometries).
- Does this solve our problem (proves R.I.C.?)
- No: the asymptotic theory holds in the limit, not for finite sizes n. Also, the probability is insufficient to take the union bound over all (exponentially many) minors A.
- We need a non-asymptotic theory of random matrices, valid for all finite sizes.

Non-asymptotic theory of random matrices

Theorem (Largest singular value)

Let A be an $N \times n$ subgaussian matrix. Then

$$\lambda_{max}(A) = \|A\| \le C(\sqrt{N} + \sqrt{n})$$

with exponentially large probability $1 - e^{-c(n+m)}$.

- Equivalently, for the normalized matrix $\bar{A} = \frac{1}{\sqrt{N}} A$,

$$\lambda_{max}(\bar{A}) \le C(1 + \sqrt{y}).$$

  Matches the asymptotic bound up to the constant C.
- Non-asymptotic: holds for every size N, n.
- Proof: the ε-net method.

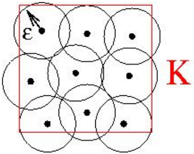
The ε-net method

- Need to bound $\|A\| = \max_{x \in S^{n-1}} \|Ax\|$
  (over the unit Euclidean sphere S^{n-1} in $\mathbb{R}^n$).
- Discretization: approximate the sphere with a finite set of points.
- Why possible? The sphere is *compact* $\Rightarrow$ has a finite ε-net.

The ε-net method

Definition (ε-net)

A subset $\mathcal{N}$ of a metric space T is called an ε-net if every point of T is within distance ε from some point in $\mathcal{N}$.
The minimum cardinality of an ε-net is the covering number $N(K, \varepsilon)$.



- $N(K, \varepsilon)$ is the minimal number of ε-balls needed to cover K.
- $N(K, \varepsilon)$ is the amount of information in K (complexity).
- $N(K, \varepsilon)$ is usually exponential in the dimension.

The ε-net method

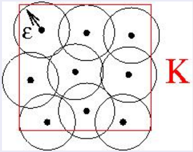
Proposition (Cardinality of an ε-net)

$$N(S^{n-1},\varepsilon) \leq \left(\frac{3}{\varepsilon}\right)^n.$$

Proof.

Construct an ε-net by the greedy algorithm:

- x_1: arbitrary;
- x_2: such that $\|x_2 - x_1\| > \varepsilon$;
- x_3: such that $\|x_3 - x_k\| > \varepsilon, \quad k = 1, 2$;
- ...until no such point can be found.

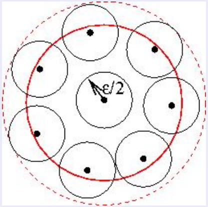


Then $\mathcal{N} = \{x_1, x_2, \dots\}$ is an ε-net.
(Or else there existed a point with distance $> \varepsilon$ from all x_k, which would contradict the stopping criterion).

$\mathcal{N} = \{x_1, x_2, \ldots\}$ is an ε-net,
which is ε-separated: $\|x_j - x_k\| > \varepsilon$.)

Want to bound above $|\mathcal{N}|$.

The $\frac{\varepsilon}{2}$-balls centered at x_k are disjoint
and are contained in a ball of radius $1 + \frac{\varepsilon}{2}$.



Count their volumes (B = unit ball):

$$|\mathcal{N}| \cdot \text{Vol}\left(\frac{\varepsilon}{2}B\right) \leq \text{Vol}\left(\left(1 + \frac{\varepsilon}{2}\right)B\right)$$

$$|\mathcal{N}| \cdot \left(\frac{\varepsilon}{2}\right)^n \leq \left(1 + \frac{\varepsilon}{2}\right)^n$$

$$|\mathcal{N}| \leq \left(\frac{2}{\varepsilon} + 1\right)^n$$

□

The ε-net method

We want to replace the sphere by its small ε-net in the operator norm

$$\|A\| = \max_{x \in S^{n-1}} \|Ax\| = \max_{x \in S^{n-1},\, y \in S^{N-1}} \langle Ax, y \rangle.$$

Lemma (Discretization of the norm)

Let $\mathcal{N}, \mathcal{M}$ be ε-nets of the unit spheres S^{n-1}, S^{N-1}. Then

$$\|A\| \le C(\varepsilon) \max_{x \in \mathcal{N}} \|Ax\| \le C'(\varepsilon) \max_{x \in \mathcal{N},\, y \in \mathcal{M}} \langle Ax, y \rangle.$$

Proof.

We write $x \in S^{n-1}$ as $x = y + z$, where $y \in \mathcal{N}$, $\|z\| \le \varepsilon$. Then

$$\|A\| \le \max_{x \in \mathcal{N}} \|Ax\|_2 + \max_{z : \|z\| \le \varepsilon} \|Az\|.$$

The last term is bounded by $\varepsilon\|A\|$. This proves the first inequality. □

The ε-net method

- Now we know how to compute the norm using $\frac{1}{2}$-nets $\mathcal{N}$ and $\mathcal{M}$:

$$\|A\| \lesssim \max_{x\in\mathcal{N},\, y\in\mathcal{M}} \langle Ax, y\rangle.$$

- And these nets are small small: $|\mathcal{N}| \le 6^n$, $|\mathcal{M}| \le 6^N$.
- For each x and y, the form $\langle Ax, y\rangle$ is a random variable.
- Entries of A are subgaussian $\Rightarrow$ $\langle Ax, y\rangle$ is subgaussian. (Exercise)
- This means: $\mathbb{P}(\langle Ax, y\rangle > t) \le 2e^{-Ct^2}$ for all t.
- Take the union bound over $x \in \mathcal{N}$, $y \in \mathcal{M}$,

$$\mathbb{P}\Big(\max_{x\in\mathcal{N},\, y\in\mathcal{M}} \langle Ax, y\rangle > t\Big) \le |\mathcal{N}||\mathcal{M}| 2e^{-Ct^2} \le 6^{n+N} 2e^{-Ct^2}.$$

- Choose $t \sim \sqrt{n} + \sqrt{N}$ so that the probability is $\le 2e^{-c(n+N)}$.
- This proves: $\|A\| \lesssim \sqrt{n} + \sqrt{N}$ with high probability. $\square$

Conclusion

Using an ε-net method, we proved an upper bound on subgaussian matrices:

Theorem (Largest singular value)

Let A be an $N \times n$ subgaussian matrix. Then

$$\lambda_{max}(A) = \|A\| \le C(\sqrt{N} + \sqrt{n})$$

with exponentially large probability $1 - e^{-c(n+m)}$.

For the lower bound, will need to refine the ε-net method. (Lecture 2).