Smoothed analysis of random matrices

Roman Vershynin

University of Michigan

Asymptotic Geometric Analsysis II

Euler International Mathematical Institute, Saint-Petersburg, Russia, June 2013

Smoothed analysis. [Spielman-Teng '01]

In theoretical computer science:

"An object should become better under a random perturbation."

Better = non-degenerate (hence algorithms are faster, more accurate).

Objects: polytopes, convex sets (?), polynomials, etc.
In this talk, an $n \times n$ **matrix** D.

Random perturbation = adding to D a random matrix R:

$$A = D + R.$$

"An $n \times n$ matrix D should become non-degenerate when replaced by $D + R$, where R is a random matrix."

Non-degeneracy

Qualitatively: A has full rank, invertible.

Quantitatively: control of $\|A^{-1}\|$.

Equivalently, the smallest singular value (smallest eigenvalue of $\sqrt{A^*A}$),

$$s_n(A) = \frac{1}{\|A^{-1}\|} = \min_x \frac{\|Ax\|_2}{\|x\|_2}$$

$$= \text{dist}_{\|.\|}(A, \text{non-invertible matrices}).$$

Problem (Smoothed analysis of matrices)

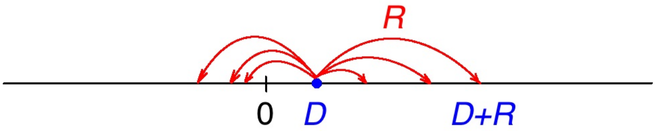
Let D be a $n \times n$ deterministic matrix,
R be an $n \times n$ random matrix (some natural distribution, or "ensemble").
Does the smallest singular value satisfy

$$s_n(D+R) \geq \text{something nice}$$

with high probability?

Intuition in 1D: if R has a continuous distribution, bounded density, then

$$|D+R| \gtrsim 1 \quad \text{w.h.p.}$$



The bound does not depend on D. Worst case: $D = 0$.

Gaussian random matrices R with iid entries ("Ginibre")

Matrix case: D, R are $n \times n$ matrices.

Theorem [Sankar-Spielman-Teng '06]

Let D be arbitrary, R be a **Gaussian** random matrix (entries iid $N(0,1)$). Then

$$\mathbb{P}\left\{ s_n(D+R) < \varepsilon n^{-1/2} \right\} \le \varepsilon, \quad \varepsilon > 0.$$

Hence

$$s_n(D+R) \gtrsim n^{-1/2} \quad \text{with high probability.}$$

The bound is **independent of** D.

"Worst case" is $D = 0$, since $s_n(R) \sim n^{-1/2}$ [Edelman '88, Szarek '90].

General random matrices with iid entries (general Ginibre)

Theorem [Rudelson-Vershynin '08]

Let $\|D\| = O(\sqrt{n})$ and R be a random matrix with iid **sub-gaussian** entries, zero means, unit variances. Then

$$\mathbb{P}\{s_n(D+R) < \varepsilon n^{-1/2}\} \leq C\varepsilon + c^n, \quad \varepsilon > 0.$$

Hence:

if $\|D\| \lesssim \sqrt{n}$, the result **does not depend on** D,
the "worst case" is $D = 0$.

If $\|D\| \gg n$, the result is generally false:

Example (Rudelson), see also [Tao-Vu '08]

$D = M \cdot \mathrm{diag}(0, 1, \ldots, 1)$,
R = Bernoulli random matrix (entries iid ± 1). Then

$$s_n(D + R) \leq \frac{C\sqrt{n}}{M} \quad \text{with probability } \frac{1}{2}.$$

Hence $D = 0$ is **not** the worst case!
$D + R$ can become **more degenerate** for D **large**.

Open question: How large?

When does $s_n(D + R)$ start to feel the deterministic part D?

What we know:
Does not feel for $\|D\| \lesssim \sqrt{n}$, feels for $\|D\| \gg n$. **Where is the threshold?**

Polynomiality

In any case:

If $\|D\|$ is polynomial in n, then $s_n(A+B)$ **is polynomial**, too.

Theorem. [Tao-Vu '08]

For any $B > 0$ there exists $A = A(\alpha, B)$ so that if $\|D\| \leq n^{\alpha}$, then

$$\mathbb{P}\left\{s_n(D+R) < C n^{-A}\right\} \leq n^{-B}.$$

Symmetric random matrices

R has iid sub-gaussian entries modulo **symmetry**: $R_{ij} = R_{ji}$.
("general Wigner")

Similar results, more difficult:

Theorem [Vershynin '11]

$$\mathbb{P}\big\{s_n(R) < \varepsilon n^{-1/2}\big\} \le C\varepsilon^{1/9} + \exp(-n^c), \quad \varepsilon > 0.$$

Same for $D + R$ where D is **any diagonal** matrix.
Thus Rudelson's example is not a problem for symmetric matrices.

Theorem [Nguyen '11]

For any $B > 0$ there exists $A = A(\alpha, B)$ so that if $\|D\| \le n^\alpha$, then

$$\mathbb{P}\big\{s_n(D + R) < Cn^{-A}\big\} \le n^{-B}.$$

When entries have **continuous** distributions

Conjecture

Suppose the entries of R have uniformly bounded densities.
Then $s_n(D+R)$ should **not** feel the deterministic part D.
The worst case should be $D=0$.

What we know: Polynomial bounds **independent** of D, but non-optimal.

Theorem (simple for indep. entries; [Farrell-Vershynin '12] for symmetric)

$$\mathbb{P}\left\{s_n(D+R) < \varepsilon n^{-p}\right\} \le C\varepsilon, \quad \varepsilon > 0.$$

$p = 3/2$ for indep. entries (maybe better), and $p = 2$ for symmetric.
C depends only on the maximal density of the entries of R.

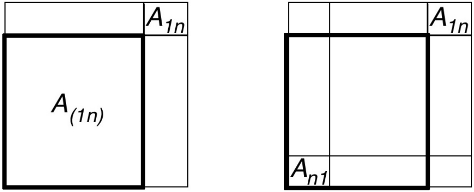
Question. Is $p = 1/2$, i.e. $s_n(D+R) \gtrsim \varepsilon n^{-1/2}$, like in the Gaussian case?

Proof for symmetric matrices [Farrell, Vershynin '12]

Enough to show that

$$(A^{-1})_{ij} = O(1) \quad \text{with high probability.}$$

Influence of A_{1n} on $(A^{-1})_{1n}$? Cramer's rule: $(A^{-1})_{1n} = \frac{\det A_{(1n)}}{\det A}$



$$|A| = aA_{11}^2 + 2b\,A_{11} + c, \qquad |A_{(11)}| = a\,A_{11} + b.$$

Divide, use that A_{1n} fluctuates continuously by $\gtrsim$ const w.h.p. □

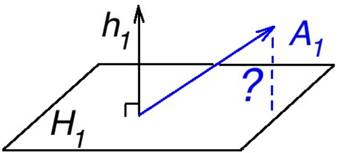
Proof for non-symmetric matrices: distance argument

$A := D + R$. $\qquad s_n(D+R) = 1/\|A^{-1}\| \geq ?$

Negative second moment identity (noticed by [Tao-Vu '08]):

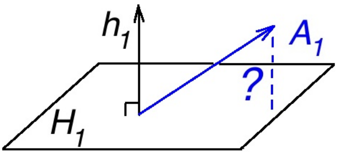
$$\|A^{-1}\|^2 \leq \|A^{-1}\|_{\mathrm{HS}}^2 = \sum_{i=1}^{n} d(A_i, H_i)^{-2}$$

where A_i = columns of A and $H_i = \mathrm{span}(A_i)_{i \neq j}$.



Remains to estimate each $d(A_i, H_i)$; finish by union bound.

Proof for non-symmetric matrices: distance argument



$$d(A_1, H_1) = |\langle A_1, h_1 \rangle| = \Big| \sum_{j=1}^{n} h_{1j} A_{1j} \Big|$$

where h_1 = unit normal for H_1. Condition on h_1; A_1 is independent.

Hence we have a **sum of independent random variables**.

A_{1j} are continuous, densities bounded by $M \Rightarrow$ same for their sum [Rogozin] + [Ball]. Hence

$$\mathbb{P}\{d(A_1, H_1) < \varepsilon\} \le CM\varepsilon. \qquad \square$$

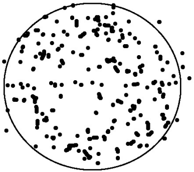
Remark. Discrete distributions - combinatorial arguments [Rudelson-V '08].

Theoretical applications: limit laws in RMT

Polynomial estimates of $s_n(A)$ are essential for validating **limit laws** of random matrix theory.

Two examples:

Circular law [Girko, Bai, Götze-Tikhomirov, Pan-Zhou, Tao-Vu]
Spectrum of $n^{-1/2}R$ converges to the uniform distribution on the unit disc:



Uses $s_n(R) \geq n^{-c}$ w.h.p.

Random unitary and orthogonal matrices

Conjecture (O. Zeitouni).

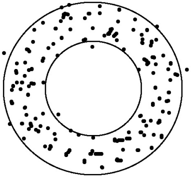
Let D be a deterministic matrix, U be a random matrix uniformly distributed in $U(n)$ or $O(n)$. Show that

$$s_n(D+R) \geq n^{-c} \quad \text{w.h.p.} \quad (1-n^{-10}).$$

This is needed to validate the Single ring theorem:

Single ring theorem [Guionnet, Krishnapur, Zeitouni '11]

Distribution of spectrum of UDV is supported in a single ring, where $U, V \in U(n)$ or $O(n)$ random uniform.



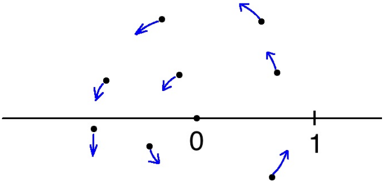
Naïve approach:

Instead of using the full power of $U \in U(n)$
just multiply by a random complex number r, $|r| = 1$.

$$s_n(D + U) \equiv s_n(D + U^{-1}) = s_n(D + r^{-1}U^{-1}) = s_n(rUD - I).$$

Condition on U.

Multiplication by r $\Leftrightarrow$ random rotation of spectrum $\sigma(UD)$ in $\mathbb{C}$.



$\sigma(UD) = \{n \text{ points}\}$. Rotation separates it from $\sigma(I) = \{1\}$ w.h.p.

$$\Rightarrow \quad \sigma(rUD - I) \quad \text{is bounded away from } 0.$$

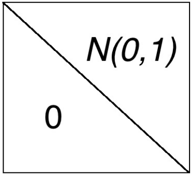
Q.E.D.?

Not Q.E.D. Fault:

Spectrum bounded away from 0 $\not\Rightarrow$ matrix well invertible.

In other words, No eigenvalues near 0 $\not\Rightarrow$ no singular values near 0.

Example [Trefethen, Viswanath '98] Triangular random Gaussian matrix A:



$$\sigma(A) = \mathrm{diag}(A) \gtrsim \frac{1}{n} \quad \text{while} \quad s_n(A) \sim e^{-cn}.$$

Random unitary matrices

Theorem (Unitary perturbations) [Rudelson, Vershynin '12]

Let D be any fixed matrix, and $U \in U(n)$ be random uniform. Then

$$\mathbb{P}\big\{s_n(D+U) \le t n^{-C}\big\} \le t^c, \quad t > 0.$$

Here $C, c > 0$ are absolute constants (independent of D).

Hence

$$s_n(D+U) \ge t n^{-C} \quad \text{w.h.p.}$$

Random orthogonal matrices

The result fails over $\mathbb{R}$, for $U \in O(n)$!

Example. If n is odd, *every* rotation $U \in SO(n)$ has eigenvalue 1.
$\Rightarrow -I + U$ is singular with probability $1/2$.

Moreover: by rotation invariance,
every orthogonal matrix D is a counterexample:
$D + U$ is singular with probability $1/2$.

Main result: These are *the only* counterexamples.
If D is not approximately orthogonal, then

$$s_n(D + U) \geq t n^{-C} \quad \text{w.h.p. :}$$

Random orthogonal matrices

Theorem (Orthogonal perturbations) [Rudelson, Vershynin '12]

Let D be a fixed matrix, and $U \in U(n)$ be random uniform. Suppose

$$\inf_{V \in O(n)} \|D - V\| \geq \delta, \qquad \|D\| \leq K.$$

Then

$$\mathbb{P}\left\{ s_n(D+U) \leq t(\delta/Kn)^C \right\} \leq t^c, \quad t > 0.$$

Here $C, c > 0$ are absolute constants (independent of D).

Remarks.

Orthogonal case is harder than unitary.

Nontrivial even in low dimensions $n = 3, 4$.

The bound $\|D\| \leq K$ may not be needed.

Optimal exponents C, c are unknown.

Approach: local perturbations

Difficulty: entries of $U \in U(n)$ are dependent.

Fixing it: like in the naïve approach, do not use the full strength of U. Instead, replace U by *infinitesimal* perturbations of identity $=$ skew-Hermitian matrices, $S^* = -S$.

Advantage: skew-Hermitian matrices can be forced to have independent entries.

Algebraically:

Local structure of Lie group $U(n)$ is given by the associated Lie algebra ($=$ tangent space at I) $=$ space of skew-Hermitian matrices.

Approach: local perturbations

Problem: skew-symmetric matrices themselves are singular (for odd n)!

Indeed, one one hand

$$\det(S) = \det(S^{\mathsf{T}}) = \det(-S) = (-1)^n \det(S).$$

So $\det(S) = 0$.

Approach: complementing by global perturbations

Global perturbation: rotation in one coordinate (say, first) in $\mathbb{C}^n$.
Multiply that coordinate by a random complex number r, $|r| = 1$.

Summary of the approach:
Use both local and global structures of $U(n)$.
Local: skew-symmetric matrices (Lie algebra).
Global: random uniform rotation in one coordinate.

Formalizing local and global perturbations

$$s_n(D+U) \gtrsim ?$$

Local:
$S :=$ skew-symmetric real Gaussian random matrix, $\varepsilon > 0$ small (n^{-10}).
Then $I + \varepsilon S$ is approximately unitary. $\Rightarrow$ Replace U by $I + \varepsilon S$.

Global:
$R := \mathrm{diag}(r, 1, \ldots, 1)$, where r random uniform, $|r| = 1$.
Replace further $I + \varepsilon S$ by $R^{-1}(I + \varepsilon S)$.

$$s_n(D+U) \cong s_n(D + R^{-1}(I + \varepsilon S)) \cong s_n(RD' + I + \varepsilon S).$$

Formalizing local and global perturbations

$$s_n(D+U) \cong s_n(RD' + I + \varepsilon S) \ \geq ?$$

Condition on V.

Summary: two layers of randomness,
local S (Gaussian skew-symmetric); global R (rotation in first coordinate).

Advantages: S has *independent entries* (modulo skew-symmetry);
R is very simple (determined by one random variable r).

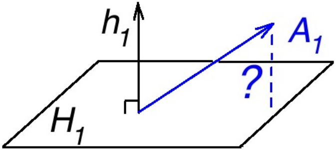
Challenges: skew-symmetry $\Rightarrow$ dependences in half of the entries.
Otherwise we would finish by the distance argument like before.

Distance argument revisited

Distance argument: estimating $s_n(A)$ reduces to estimating

$$d(A_1, H_1) = |h_1^\mathsf{T} A_1| \geq \cdots \quad \text{w.h.p.}$$

where A_1 = first column of A and $H_1 = \mathrm{span}(A_j)_{i>1}$, and $h_1 = H_1^\perp$.



Challenge of skew-symmetry: In our matrix $A = RVD + I + \varepsilon S$, the first column A_1 is correlated with H_1 through the first row.

How to express $h_1^\mathsf{T} A_1$?

Distance argument revisited

$$A = RD' + I + \varepsilon S = \begin{bmatrix} A_{11} & Y^{\mathsf{T}} \\ X & B^{\mathsf{T}} \end{bmatrix}$$

Lemma (distance via quadratic forms)

$$|h_1^{\mathsf{T}} A_1| = \frac{|A_{11} - X^{\mathsf{T}} M Y|}{\sqrt{1 + \|MY\|_2^2}}, \quad \text{where } M = B^{-1}.$$

Our situation: $Z \in \mathbb{R}^{n-1}$ random Gaussian vector,

$$S = \begin{bmatrix} 0 & -Z^{\mathsf{T}} \\ Z & 0 \end{bmatrix}, \quad D' = \begin{bmatrix} p & v^{\mathsf{T}} \\ u & Q \end{bmatrix} \quad \Rightarrow \quad A = \begin{bmatrix} rp + 1 & (rv - \varepsilon Z)^{\mathsf{T}} \\ u + \varepsilon Z & I + Q \end{bmatrix}$$

Good: $h_1^{\mathsf{T}} A_1$ is a self-normalized quadratic form in Gaussian random variables (Z). Essentially a *linear form* (ε^2 = second order term).

Bad: bound it below without knowing much about $M = (I + Q)^{-1}$.

Idea (local/global): Use r or Z (or both) depending on $\|M\|$.

Orthogonal perturbations

Same approach (local/global, via quadratic forms), with one difference:

Global perturbation: instead of random rotation in *one* coordinate, rotate in two coordinates.

Argument is more challenging.
Seems to differentiate odd and even n; reduces the problem to $n = 3$.

Entries of the inverse matrix

Question.

For A a random matrix, what is the magnitude of the entries of A^{-1}?

$$\text{Is } \max_{ij} |(A^{-1})_{ij}| \lesssim n^{-1/2} \quad \text{w.h.p. (up to log-factors)?}$$

This would imply $\|A^{-1}\| \leq \|A^{-1}\|_{\mathrm{HS}} \lesssim n^{1/2}$, so

$$s_n(A) \gtrsim n^{-1/2} \quad \text{w.h.p., as before.}$$

Work by [L. Erdös-Schlein-Yau+Yin '12], [Tao-Vu '12].

Entries of the inverse matrix and delocalization

Question.

Is $\max_{ij} |(A^{-1})_{ij}| \lesssim n^{-1/2}$ w.h.p. (up to log-factors) ?

Related to **delocalization** of eigenvectors of A.

Heuristics. Say, A is symmetric, iid entries. Spectral decomposition:

$$A = \sum \lambda_i u_i u_i^\mathsf{T} \quad \Rightarrow \quad A^{-1} = \sum \lambda_i^{-1} u_i u_i^\mathsf{T} \approx \lambda_n^{-1} u_n u_n^\mathsf{T}$$

where λ_n is the smallest eigenvalue in magnitude.

$$\max_{ij} |(A^{-1})_{ij}| \approx |\lambda_n^{-1} u_n(i) u_n(j)|.$$

Invertibility as before $\Rightarrow \lambda_n \gtrsim n^{-1/2}$. **Delocalization:** all $|u_n(i)| \lesssim n^{-1/2}$.

$$\Rightarrow \quad \max_{ij} |(A^{-1})_{ij}| \lesssim n^{-1/2}. \qquad \square$$