SCIENCES
SORBONNE
UNIVERSITÉ

Gaussian measurements

Claire Boyer

The goal of today is to construct matrices verifying RIP(s) with an optimal number of measurements (i.e. $m \sim s \log(d/s)$).

+ RIP(s) $\Rightarrow$ ER($c_0 s$) (with an absolute constant c_0),
+ ER(s) requires at least $s \log(d/s)$ measurements,

$\rightsquigarrow$ Lower bound: one cannot hope to construct matrices satisfying RIP(s) with less than $s \log(d/s)$.

- Matrices with RIP & an optimal number m of rows:
  $\Rightarrow$ random matrices
- In the rest, results derived with high probability

Achtung!

Verifying if a given matrix satisfies RIP is NP-hard.

Random vs. deterministic?

1. The constructed random matrices will satisfy RIP with high probability, close to 1;
2. it is possible to construct deterministic matrices satisfying RIP, however they require a number of rows of the order

$$s^2 \gg s \log(d/s)$$

Remark: This is a very difficult open problem to construct deterministic matrices satisfying RIP with a number of rows of the order of $s \log(d/s)$.
$\Rightarrow$ could have important consequences!
(A realization of a random variable is not deterministic)

Summary

A first random matrix

We are going to focus on case where the entries of the matrix A are i.i.d. real-valued standard normal variables. As a result of this effort, the community now has a very precise understanding of the performance of ℓ^1-norm minimization in this Gaussian model.

Definition

A standard Gaussian random matrix is a matrix which entries are i.i.d. standard Gaussian random variables,

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1d} \\ \vdots & \ddots & \vdots \\ g_{m1} & \cdots & g_{md} \end{pmatrix},$$

with $(g_{ij})_{ij}$ i.i.d. $\mathcal{N}(0,1)$.

Isometry at least in expectation!

Let us have a look at the expectation of $\left(\frac{1}{\sqrt{m}}G\right)^* \frac{1}{\sqrt{m}}G$,

$$\mathbb{E}\frac{1}{m}G^*G = \mathrm{Id}. \tag{1}$$

- So $\frac{1}{\sqrt{m}}G$ is at least an isometry in expectation.
- The purpose of the rest of the lecture is to show that it does not deviate too much from its expectation with high probability for s-sparse vectors.

To show that Gaussian matrices satisfy RIP

We need some concentration tools.

Summary

Gaussian-type concentration

A typical result of concentration

If g is a random variable with law $\mathcal{N}(\mu, \sigma^2)$, then for all $t > 0$

$$\mathbb{P}\left(|g - \mu| \geq t\sigma\right) \leq 2\exp(-t^2/2). \tag{2}$$

The idea is that if g is distributed according to $\mathcal{N}(\mu, \sigma^2)$ then

- it will be in the interval centered in μ with length $2t\sigma$ with probability at least $1 - 2\exp(-t^2/2)$.
- The variable g is said to be "concentrated" around its expectation μ.

Beware! Any random variable is not concentrated around its expectation.Some does not admit expectation: see Cauchy variables that are absolutely continuous w.r.t. Lebesgue measure and with density $(\pi(1 + t^2))^{-1}$ for $t \in \mathbb{R}$.

Other types of concentrations

Concentration around expectation but slower than Gaussian var.

Let e be a r.v. with density $f(t) = (2\sigma)^{-1}\exp(-|t-\mu|/\sigma)$ for $t \in \mathbb{R}$, satisfies

$$\mathbb{P}\left(|e-\mu| \geq t\sqrt{2}\sigma\right) \leq \exp(-\sqrt{2}t).$$

We say that e concentrates in "$\exp(-t)$" around its expectation μ (while its variance is $2\sigma^2$).

There exists concentration of type

$$\exp(-t^\alpha), \quad \text{with } 0 < \alpha \leq 2 \quad \text{or of type} \quad 1/t^p \quad \text{for} p \geq 1$$

Concentration of type $\exp(-t^2)$, which is said to be sub-gaussian.
Concentration in $\exp(-t)$ will be said to be sub-exponential.

From moments to tails

How to characterize the type of concentration?
$\rightsquigarrow$ Use the speed of its moments growth.

Proposition (The tails-moments proposition)

Let X be a random variable and $\alpha \geq 1$. There is equivalence between the two following propositions:

1. *There exists an absolute constant c_0 such that*
$$\forall p \geq \alpha, \quad \|X\|_{L^p} \leq c_0 p^{1/\alpha}.$$
2. *There exist two absolute constants c_1 and c_2 such that for all $t \geq c_1$,*
$$\mathbb{P}(|X| \geq t) \leq \exp(-t^{\alpha}/(c_2)^{\alpha}).$$

A random variable is sub-Gaussian (resp. sub-exponential) if and only if its p-th moments grow slower than $\sqrt{p}$ (resp. p).

In the sequel we will focus on the square of random variables (see the RIP definition).

Proposition

Let X be a real sub-Gaussian variable. Then X^2 is sub-exponential.

Show it!

Proof: Using the tails-moments Proposition , X being sub-Gaussian, one has $\forall p \geq \alpha = 2$, $\|X\|_{L^p} \leq c_0\sqrt{p}$, and then $\|X^2\|_{L^p} = \|X\|_{L^{2p}}^2 \leq 2c_0^2 p$. Again, using the tails-moments Proposition, X^2 is sub-exponential.

A well-known subgaussian variable?

There is a certain type of random variables that one can expect to concentrate in a sub-Gaussian way. Which one?

The empirical means of i.i.d. variables!

If $(X_m)_m$ is a sequence of i.i.d. random variables admitting a 2nd-order moment then CLT provides that for $\bar{X}_m = \frac{1}{m}\sum_{i=1}^m X_i$, $\sqrt{m}(\bar{X}_m - \mathbb{E}X_1)$ is asymptotically Gaussian: for all $t > 0$,

$$\mathbb{P}\left(\sqrt{m}|\bar{X}_m - \mathbb{E}X_1| \geq t\sigma\right) \underset{m\to\infty}{\to} \mathbb{P}(|g| \geq t).$$

with σ^2 the variance of X_1.

Then, the deviations of $\sqrt{m}|\bar{X}_m - \mathbb{E}X_1|$ will be similar to Gaussian deviation when m is large.

CLT leads only to asymptotic results.
We're working with a fixed number $m \rightsquigarrow$ non-asymptotic results.

Non-asymptotic results

Proposition (Bernstein's result)

Let $Z_1, \ldots, Z_m$ be real i.i.d. random variables with the same law as Z such that for all $p \geq 0$,

$$\|Z\|_{L^p} \leq c_0 p,$$

with c_0 an absolute constant. Then, there exists an absolute constant $c_1 > 0$ such that for all $t > 0$,

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m} Z_i - \mathbb{E}Z\right| \geq tc_0\right) \leq 2\exp\left(-c_1 m \min(t^2, t)\right).$$

Without loss of generality, let us suppose that the (Z_i)'s are centered. Moreover,

$$\mathbb{P}\left(|\bar{Z}_m| \geq tC_0\right) = \mathbb{P}\left(\bar{Z}_m \geq tC_0\right) + \mathbb{P}\left(-\bar{Z}_m \geq -tC_0\right)$$

A classic way to show concentration inequality is to use the Laplace transform. By Markov inequality, one gets:

$$\mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} Z_i \geq t\right) \leq \inf_{\lambda>0} \frac{\mathbb{E}\exp(\lambda \bar{Z}_m)}{\exp(\lambda t)} = \inf_{\lambda>0}\left(e^{-\lambda t}\prod_{i=1}^{m}\mathbb{E}\exp\left(\frac{\lambda Z_i}{m}\right)\right). \tag{3}$$

One wants to bound above $\mathbb{E}\exp(\mu Z)$ when Z is a centered sub-exponential variable and $\mu > 0$. Using Taylor's expansion for the exponential function, one gets

$$\mathbb{E}\exp(\mu X) - 1 \leq \sum_{k\geq 2} \frac{\mu^k \mathbb{E}|Z|^k}{k!} \leq \sum_{k\geq 2} \frac{\mu^k C_0^k k^k}{k!},$$

Proof of Bernstein's proposition II

and since $k! \geq \sqrt{2\pi k}(k/e)^k$ and for $\mu > 0$ such that $e\mu C_0 < 1$,

$$\mathbb{E}\exp(\mu X) \leq 1+\sum_{k\geq 2}(e\mu C_0)^k = 1+\frac{(e\mu C_0)^2}{1-(e\mu C_0)} \leq \exp\left(\frac{(e\mu C_0)^2}{1-(e\mu C_0)}\right),$$

since $1+t \leq e^t$ for all $t \in \mathbb{R}$.
Therefore (3) gives

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} Z_i \geq t\right) &\leq \inf_{0<e\lambda C_0<m} e^{-\lambda t}\exp\left(\frac{m(e(\lambda/m)C_0)^2}{1-(e(\lambda/m)C_0)}\right) \\
&= \exp\left(\inf_{0<e\lambda C_0<m}\left\{-\lambda t + \frac{(e\lambda C_0)^2}{m-(e\lambda C_0)}\right\}\right) \\
&\leq \exp\left[\inf_{0<e\lambda C_0<m/2}\left\{-\lambda t + \frac{(e\lambda C_0)^2}{m-(e\lambda C_0)}\right\}\right]
\end{aligned}$$

Proof of Bernstein's proposition III

If $t \le 2eC_0$, take $\lambda = mt/(2eC_0)^2 \le m/(2eC_0)$, then

$$\mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} Z_i \ge t\right) \le \exp\left(-\frac{mt^2}{4e^2C_0^2}\right).$$

If $t > 2eC_0$, for $\lambda = m/(2eC_0)$,

$$\mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} Z_i \ge t\right) \le \exp\left(-\frac{mt}{4eC_0}\right).$$

Exercise (Bounded variables are sub-gaussian.)

Let X be a centered random variable such that $a \leq X \leq b$ a.s..

1. *Show that for all $\lambda > 0$,*

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

   Hint: use the convexity of exp on

$$\exp\left(\left(\frac{x-a}{b-a}\right)\lambda b + \left(\frac{b-x}{b-a}\right)\lambda a\right).$$

   Hint 2: Use a second order Taylor expansion for the majorizing function.

2. *Show that such a random variable is sub-gaussian.*
   Hint: Use Chernoff's inequality.

Exercise (Hoeffding's inequality)

Let $X_1, \dots, X_n$ be centered independent random variables such that for $i, \dots, n$ $a_i \le X_i \le b_i$ a.s.. Show that for all $t > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > t\right) \le \exp\left(-\frac{2nt^2}{\frac{1}{n}\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Summary

RIP for Gaussian matrices

The goal of this section is to show that Gaussian matrices satisfy RIP with an optimal number m of rows, $m \sim s \log(d/s)$.

Theorem

Let $G \in \mathbb{R}^{m \times d}$ be a Gaussian matrix. There exists two universal constants c_0 and c_1 such that for

$$m \geq c_1 s \log(d/s),$$

G satisfies for all $x \in \Sigma_s$

$$\frac{1}{2}\|x\|_2^2 \leq \frac{\|Gx\|_2^2}{m} \leq \frac{3}{2}\|x\|_2^2,$$

with probability larger than $1 - 2\exp(-c_0 m)$.

The normalization over m is natural: G should be an isometry at least in expectation.

Proof for a more general result

We are going to show a more general result, i.e. for a matrix A of the form

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} a_1^* \\ \vdots \\ a_m^* \end{pmatrix}$$

where the vectors (a_i) are random vectors distributed as a, under the following assumptions:

Assumptions

1. Isotropy condition: for $x \in \mathbb{R}^n$,

$$\mathbb{E} \langle a, x \rangle^2 = \|x\|_2^2$$

2. Subgaussian behaviour: for all $x \in \mathbb{R}^n$, for all $p \geq 1$,

$$\| \langle a, x \rangle \|_{L^p} \leq C_0 \sqrt{p} \|x\|_2.$$

Note that if g is a standard Gaussian vector, then the two previous assumptions are verified:

- the isotropy condition is satisfied:

$$\mathbb{E}\left\langle g, x\right\rangle^2 = \mathbb{E}\left[\left(\sum_{i=1}^{n} g_i x_i\right)^2\right] = \sum_{j=1}^{n} x_i^2 = \|x\|_2^2.$$

- the subgaussian condition is also satisfied: for all $x \in \mathbb{R}^n$, $\left\langle g, x\right\rangle$ is $\mathcal{N}(0, \|x\|_2^2)$,

$$\| \left\langle g, x\right\rangle \|_{L^p} = \|x\|_2 \|g\|_{L^p} \leq C_0 \sqrt{p} \|x\|_2.$$

ϵ-nets

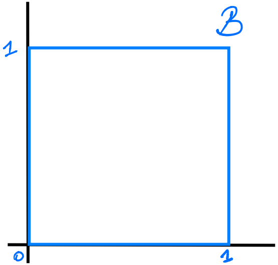
- Nets are convenient means to discretize compact sets.
- For RIP proof, we will mostly need to discretize the unit Euclidean sphere S^{d-1} or S^{s-1} in the definition of the spectral norm (cf. RIP).

Definition

Let (X, d) be a metric space and let $\epsilon > 0$.
A subset $\mathcal{N}_\epsilon$ of X is called an ϵ-net of X if every point $x \in X$ can be approximated to within ϵ by some point $y \in \mathcal{N}_\epsilon$, i.e. so that $d(x, y) \leq \epsilon$.
The minimal cardinality of an ϵ-net of X, if finite, is denoted $\mathcal{N}(X, \epsilon)$ and is called the covering number of X (at scale ϵ).

Equivalently, $\mathcal{N}(X, \epsilon)$ is the minimal number of balls with radii ϵ and with centers in X needed to cover X.

Example of ϵ-net



Construct an 1/4-net for the compact set B under the ℓ^∞ metrics.

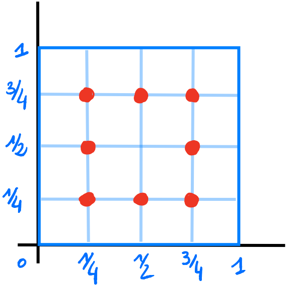


Figure: An example of a $1/4$-net for a compact set in the ℓ^∞ metrics.

Preliminary I

Here is the result that we already used but written differently:

Lemma

The unit Euclidean sphere S^{d-1} equipped with the Euclidean metric satisfies for every $\epsilon > 0$ that

$$\mathcal{N}(S^{d-1}, \epsilon) \leq \left(1 + \frac{2}{\epsilon}\right)^d .$$

Proof Blackboard time.

This is a simple volume argument. Let us fix $\epsilon > 0$ and choose $\mathcal{N}_\epsilon = \Lambda$ to be a maximal ϵ-separated subset of S^{d-1}. In other words, Λ is such that $d(x, y) \geq \epsilon$ for all $x, y \in \Lambda$, $x \neq y$, and no subset of S^{d-1} containing Λ has this property.

One can write

$$\cup_{x \in \Lambda}(x + \epsilon/2\overset{\circ}{B}) \subset B + \epsilon/2B = (1 + \epsilon/2)B.$$

Preliminary II

Note that the sets $\{x + \epsilon/2\overset{\circ}{B}\}_{x\in\Lambda}$ are disjoint. Applying the volume to the previous inequality, one gets

$$\sum_{x\in\Lambda} \mathrm{vol}(x + \epsilon/2\overset{\circ}{B}) \leq \mathrm{vol}((1+\epsilon/2)B).$$

The volume is invariant by translation, thus $\mathrm{vol}(x+\epsilon/2\overset{\circ}{B}) = \mathrm{vol}(\epsilon/2\overset{\circ}{B})$. By homothety, $\mathrm{vol}(\epsilon/2\overset{\circ}{B}) = (\epsilon/2)^d \mathrm{vol}(\overset{\circ}{B})$, and $\mathrm{vol}((1+\epsilon/2)B) = (1+\epsilon/2)^d \mathrm{vol}(B)$. The volume of the boundary ∂B is of zero measure, then $\mathrm{vol}(B) = \mathrm{vol}(\overset{\circ}{B})$. Finally, one has

$$|\Lambda|(\epsilon/2)^d \mathrm{vol}(B) \leq (1+\epsilon/2)^d \mathrm{vol}(B),$$

implying

$$|\Lambda| \leq \left(1 + \frac{2}{\epsilon}\right)^d.$$

A first lemma

The first lemma is just a reformulation of the previous volumic argument.

Lemma

Let s be an integer. Let B_2^s denote the euclidean unit-ball in $\mathbb{R}^s$. Let $0 < \epsilon < 1$, there exists $\Lambda_\epsilon \subset B_2^s$ such that

1. *for all $x \in B_2^s$, there exists $y \in \Lambda_\epsilon$ such that $\|x - y\|_2 \leq \epsilon$,*
2. $|\Lambda_\epsilon| \leq (5/\epsilon)^s$.

Proof.

One can use the previous volumic argument, and remark that $1 - 2/\epsilon \leq 5/\epsilon$ as soon as $\epsilon \leq 3$, which is true. □

A second lemma I

The second lemma allows to study the spectral norm of A restricted to the ϵ-net Λ_ϵ.

Lemma

Let $B \in \mathbb{R}^{s\times s}$ be a symmetric matrix and $0 < \epsilon < 1$. Let Λ_ϵ, be an ϵ-net of S_2^{s-1} (the euclidean unit sphere of $\mathbb{R}^s$). Then,

$$\|B\|_{2\to 2} = \sup_{x\in S_2^{s-1}} |\langle Bx, x\rangle| \leq (1-2\epsilon)^{-1} \sup_{y\in\Lambda_\epsilon} |\langle By, y\rangle|.$$

A second lemma II

(This part is quite obvious) Since B is symmetric there exists an orthogonal matrix $P \in \mathcal{O}(s)$ and a diagonal matrix $D = \mathrm{diag}(\lambda_1, \dots, \lambda_s)$ such that $B = PDP^*$. Hence,

$$\|B\|_{2\to 2} = \sup_{\|x\|_2=1} \|Bx\|_2 = \sup_{\|x\|_2=1} \|Dx\|_2 = \max(|\lambda_1|, \dots, |\lambda_s|).$$

One also has

$$\begin{aligned} \sup_{x \in S_2^{s-1}} |\langle Bx, x\rangle| &= \sup_{\|x\|_2=1} |\langle DP^*x, P^*x\rangle| = \sup_{\|x\|_2=1} |\langle Dx, x\rangle| \\ &= \max(|\lambda_1|, \dots, |\lambda_s|). \end{aligned}$$

We got $\|B\|_{2\to 2} = \sup_{x \in S_2^{s-1}} |\langle Bx, x\rangle|$, even if one can show it only by using the definition of $\|B\|_{2\to 2}$.

Rest of the proof

For the right hand side inequality, let $x \in S_2^{s-1}$ and $y \in \Lambda_\epsilon$ such that $\|x-y\|_2 \leq \epsilon$. Thus,

$$\begin{aligned} |\langle Bx, x\rangle| &= |\langle By, y\rangle + \langle Bx, (x-y)\rangle + \langle B(x-y), y\rangle| \\ &\leq |\langle By, y\rangle| + 2\|x-y\|_2\|B\|_{2\to 2}. \end{aligned}$$

One can conclude using $\|x-y\|_2 \leq \epsilon$.

A third lemma (more usual)

Lemma

$$\binom{n}{s} \leq \left(\frac{en}{s}\right)^s.$$

Proof: First,

$$\binom{d}{s} = \frac{d!}{s!(d-s)!} = \frac{d(d-1)\ldots(d-(s-1))}{s!} \leq \frac{d^s}{s!}. \qquad (4)$$

Using the expansion,

$$e^s = \sum_{i=0}^{+\infty} \frac{s^i}{i!},$$

and thus for $i = s$, $e^s \geq \frac{s^s}{s!} \Rightarrow \frac{1}{s!} \leq \left(\frac{e}{s}\right)^s$. Injecting in (4), one gets the desired result.

Proof of RIP for (sub)Gaussian matrices I

(Blackboard time) To show that A satisfies RIP(s), without loss of generality, it is sufficient to show that

$$\sup_{x \in \Sigma_s \cap S_2^{d-1}} \left| \|Ax\|_2^2 - 1 \right| \leq \frac{1}{2}. \tag{5}$$

One can write $\Sigma_s \cap S_2^{d-1} = \cup_{|J|=s} S^J$ where S^J is the set of all s-sparse vectors of S_2^{d-1} which support is included in J.
Using Lemma 10, for all $J \subset [|n|]$ such that $|J| = s$ there exists an ϵ-net Λ_ϵ^J of S^J with cardinality at most $(5/\epsilon)^s$.
The next step is to apply Lemma 11 to the symmetric matrix $B = A_J^* A_J - I_s$ where A_J is the matrix extracted from A with columns that are indexed by J. One gets

$$\sup_{x \in S^J} |\|Ax\|_2^2 - 1| \leq (1 - 2\epsilon)^{-1} \max_{x \in \Lambda_\epsilon^J} |\|Ax\|_2^2 - 1|,$$

Proof of RIP for (sub)Gaussian matrices II

leading to

$$\sup_{x \in \Sigma_s \cap S_2^{d-1}} \left| \|Ax\|_2^2 - 1 \right| \leq (1-2\epsilon)^{-1} \max_{x \in \Lambda_\epsilon} \left| \|Ax\|_2^2 - 1 \right|, \tag{6}$$

with $\Lambda_\epsilon = \cup_{|J|=s} \Lambda_\epsilon^J$.
In particular, the cardinal of Λ_ϵ is such that

$$|\Lambda_\epsilon| \leq \binom{d}{s} \left(\frac{5}{\epsilon}\right)^s \leq \left(\frac{5ed}{s\epsilon}\right)^s$$

where we used $\binom{d}{s} \leq (ed/s)^s$.
Applying Bernstein's inequality leads to

$$\mathbb{P}\left(\left| \|Ax\|_2^2 - 1 \right| \geq 1/8 \right) \leq 2\exp(-c_0 m),$$

Proof of RIP for (sub)Gaussian matrices III

for all $x \in \Lambda_\epsilon$. A union bound gives

$$\mathbb{P}\left(\forall x \in \Lambda_\epsilon, |\|Ax\|_2^2 - 1| \leq 1/8\right) \geq 1 - 2|\Lambda_\epsilon| \exp(-c_0 m)$$
$$= 1 - 2\left(\frac{5ed}{s\epsilon}\right)^s \exp(-c_0 m).$$

Indeed,

$$\mathbb{P}\left(\forall x \in \Lambda_\epsilon, |\|Ax\|_2^2 - 1| \leq 1/8\right) = \mathbb{P}\left(\cap_{x \in \Lambda_\epsilon} \left\{|\|Ax\|_2^2 - 1| \leq 1/8\right\}\right)$$
$$= 1 - \mathbb{P}\left(\cup_{x \in \Lambda_\epsilon} \left\{|\|Ax\|_2^2 - 1| \geq 1/8\right\}\right)$$

and

$$\mathbb{P}\left(\cup_{x\in\Lambda_\epsilon}\left\{|\|Ax\|_2^2-1|\geq 1/8\right\}\right) \leq \sum_{x\in\Lambda_\epsilon}\mathbb{P}\left(\left\{|\|Ax\|_2^2-1|\geq 1/8\right\}\right)$$
$$\leq 2|\Lambda_\epsilon|\exp(-c_0 m)$$
$$\leq 2\left(\frac{5ed}{s\epsilon}\right)^s\exp(-c_0 m).$$

Choosing $\epsilon = 3/8$ in (6) gives

$$\mathbb{P}\left(\sup_{x\in\Sigma_s\cap S_2^{d-1}}|\|Ax\|_2^2-1|\leq 1/2\right)\geq 1-2|\Lambda_\epsilon|\exp(-c_0 m) = 1-2\left(\frac{5ed}{s\epsilon}\right)^s$$

One wants this event to happen with high probability, let's say $1-\eta$ for $0<\eta<1$. This leads to

$$m \gtrsim s\log(ed/(s\eta)).$$

Why Gaussian matrices are theoretically great for CS? I

Universality of Gaussian matrices: If you fix the measurement basis to be Gaussian random, this measurement basis is compatible with *any* sparsity basis. Can you show it?

Why Gaussian matrices are theoretically great for CS? I

- Ψ orthogonal transform (change of basis for sparsity)
- $G \in \mathbb{R}^{n\times d} = (g_1|\ldots|g_m)^\top$ Gaussian measurements (w/ g_i standard Gaussian vectors)
  - satisfies RIP w/ constant $1/2$
- Sensing matrix

$$A = \frac{1}{\sqrt{m}} G\Psi,$$

  - has the same law as $\frac{1}{\sqrt{m}}G$
  - satisfies RIP!

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} a_1^* \\ \vdots \\ a_m^* \end{pmatrix} = \frac{1}{\sqrt{m}} G\Psi = \frac{1}{\sqrt{m}} \begin{pmatrix} (\Psi^* g_1)^* \\ \vdots \\ (\Psi^* g_m)^* \end{pmatrix},$$

where $(\Psi^* g_i) \sim \mathcal{N}(0, \mathrm{Id})$.

About the tightness of the proof

- Net-techniques are not always optimal! They are in particular sub-optimal for partial Fourier matrices defined as follows:

$$A_{k\ell} = e^{2i\pi x_k \ell/d}, \qquad 1 \leq \ell \leq d,$$

where $(x_k)_{1\leq k\leq m}$ are drawn uniformly in $[|d|]$.

- Gaussian is not for real-world applications: Even if Gaussian matrices are great for deriving recovery guarantees with an optimal number of measurements, they are unpractical:
    - They are full matrices: difficult to store or to make matrix-vector products
    - They are not often showing up in applications: it is very rare when you can model a physical acquisition model as a Gaussian matrix (except if you try to reconstruct images through a paint layer...)

Summary

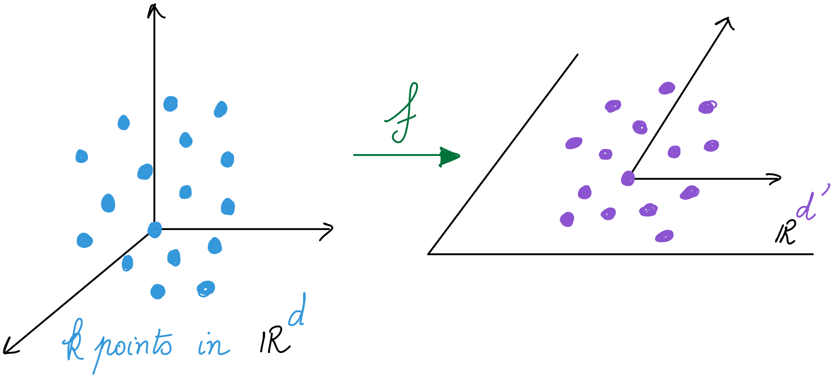
Suppose one has k points, $X = \{x_1, ..., x_k\}$, in $\mathbb{R}^d$ (with d large).

If $d > k$, since the points have to lie in a subspace of dimension k it is clear that one can consider the projection $f : \mathbb{R}^d \to \mathbb{R}^k$ of the points to that subspace without distorting the geometry of X. In particular, for every x_i and x_j, $\|f(x_i) - f(x_j)\|_2 = \|x_i - x_j\|_2$, meaning that f is an isometry in X.

Suppose now we allow a bit of distortion, and look for $f : \mathbb{R}^n \to \mathbb{R}^k$ that is an ϵ-isometry, meaning that

$$(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|f(x_i) - f(x_j)\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2$$

Motivations and context



Can we do better than $d' = k$?

In 1984, Johnson and Lindenstrauss showed a remarkable Lemma (below) that answers this question positively.

JL lemma

Theorem (Johnson-Lindenstrauss lemma)

For any $0 < \epsilon < 1$, *and for any integer* k, *let* d' *be such that*

$$d' \geq 4 \frac{1}{\epsilon^2/2 - \epsilon^3/3} \log(k).$$

then for any set X *of* k *points in* $\mathbb{R}^d$ *there is a linear map* $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ *that is an* ϵ*-isometry for* X*. This map can be found in randomized polynomial time.*

Proof of JL lemma

We need a few concentration of measure bounds, we will omit the proof of those but they are available in [1] and are essentially the same ideas as those used to show Hoeffding's inequality.

[1] Sanjoy Dasgupta and Anupam Gupta.
An elementary proof of a theorem of johnson and lindenstrauss.
Random Structures & Algorithms, 22(1):60–65, 2003.

Lemma (Dasgupta and Gupta[1])

Let $y_1, \ldots, y_d$ be i.i.d.standard Gaussian random variables and $Y = (y_1, \ldots, y_d)$. Let $\Pi_{d'} : \mathbb{R}^d \to \mathbb{R}^{d'}$ be the projection into the first d' coordinates and

$$Z := \Pi_{d'}\left(\frac{Y}{\|Y\|_2}\right) = \frac{1}{\|Y\|_2}(y_1, \ldots, y_{d'}) \qquad \text{and} \qquad L = \|Z\|_2^2.$$

It is clear that $\mathbb{E}L = \frac{d'}{d}$. In fact, L is very concentrated around its expectation,

- *if $\beta < 1$,*

$$\mathbb{P}\left(L \leq \beta \frac{d'}{d}\right) \leq \exp\left(\frac{d'}{2}(1 - \beta + \log(\beta)\right),$$

- *if $\beta > 1$,*

$$\mathbb{P}\left(L \geq \beta \frac{d'}{d}\right) \leq \exp\left(\frac{d'}{2}(1 - \beta + \log(\beta)\right).$$

(Blackboard time)

We will start by showing that, given a pair x_i, x_j a projection onto a random subspace of dimension d' will satisfy (after appropriate scaling) the property

$$(1-\epsilon)\|x_i - x_j\|_2^2 \leq \|f(x_i) - f(x_j)\|_2^2 \leq (1+\epsilon)\|x_i - x_j\|_2^2$$

with high probability. WLOG, we can assume that $u = x_i - x_j$ has unit norm. Understanding what is the norm of the projection of u on a random subspace of dimension d' is the same as understanding the norm of the projection of a (uniformly) random point on $\S^{d-1}$ the unit sphere in $\mathbb{R}^d$ on a specific d'-dimensional subspace, let's say the one generated by the first d' canonical basis vectors.
This means that we are interested in the distribution of the norm of the first d' entries of a random vector drawn from the uniform distribution over S^{d-1} – this distribution is the same as taking a standard Gaussian vector in $\mathbb{R}^d$ and normalizing it to the unit sphere.

Let $\Pi_{d'} : \mathbb{R}^d \to \mathbb{R}^{d'}$ be the projection on a random k-dimensional subspace and let $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ defined as $f = \sqrt{\frac{d'}{d}}\Pi_{d'}$. Then (by the above discussion), given a pair of distinct x_i and x_j, $\frac{\|f(x_i)-f(x_j)\|_2^2}{\|x_i-x_j\|_2^2}$ has the same distribution as $\frac{d}{d'}L$, as defined in Lemma 14. Using Lemma 14, we have, given a pair x_i, x_j,

$$\mathbb{P}\left(\frac{\|f(x_i)-f(x_j)\|_2^2}{\|x_i-x_j\|_2^2} \leq (1-\epsilon)\right) \leq \exp\left(\frac{d'}{2}(1-(1-\epsilon)+\log(1-\epsilon))\right),$$

since for $\epsilon \geq 0$, $\log(\epsilon \leq -\epsilon - \epsilon^2/2$, we have

$$\begin{aligned}\mathbb{P}\left(\frac{\|f(x_i)-f(x_j)\|_2^2}{\|x_i-x_j\|_2^2} \leq (1-\epsilon)\right) &\leq \exp\left(-\frac{d'\epsilon^2}{4}\right), \\ &\leq \exp(-2\log(k)) = \frac{1}{k^2}.\end{aligned}$$

On the other hand,

$$\mathbb{P}\left(\frac{\|f(x_i)-f(x_j)\|_2^2}{\|x_i-x_j\|_2^2} \geq (1+\epsilon)\right) \leq \exp\left(\frac{k}{2}(1-(1+\epsilon)+\log(1+\epsilon))\right),$$

since for $\epsilon \geq 0$, $\log(1+\epsilon) \leq \epsilon - \epsilon^2/2 + \epsilon^3/3$, one has

$$\mathbb{P}\left(\frac{\|f(x_i)-f(x_j)\|_2^2}{\|x_i-x_j\|_2^2} \geq (1+\epsilon)\right) \leq \exp\left(-\frac{d'(\epsilon^2-2\epsilon^3/3)}{4}\right),$$

$$\leq \exp(-2\log(k)) = \frac{1}{k^2}.$$

By union bound, one gets

$$\mathbb{P}\left(\frac{\|f(x_i)-f(x_j)\|_2^2}{\|x_i-x_j\|_2^2} \notin [1-\epsilon;1+\epsilon]\right) \leq \frac{2}{k^2}.$$

Since there exists $\binom{k}{2}$ such pairs, a simple union bound gives

$$\mathbb{P}\left(\exists(i,j), \quad \frac{\|f(x_i)-f(x_j)\|_2^2}{\|x_i-x_j\|_2^2} \notin [1-\epsilon;1+\epsilon]\right) \leq \frac{2}{k^2}\frac{k(k-1)}{2} = 1-\frac{1}{k}.$$

Therefore choosing f as a properly scaled projection onto a random d'-dimensional subspace is an ϵ-isometry on X with probability at least $1/k$. We can achieve any desirable constant probability of success by trying $O(k)$ such random projections, meaning we can find an ϵ-isometry in randomized polynomial time.

Note that by considering d' slightly larger one can get a good projection on the first random attempt with very good confidence. In fact, it's trivial to adapt the proof above to obtain the following Lemma.

Lemma (Improved JL lemma)

For any $0 < \epsilon < 1$, $\tau > 0$, and for any integer k, let d' be such that

$$d' \geq (2+\tau)\frac{2}{\epsilon^2/2 - \epsilon^3/3}\log(k).$$

Then, for any set X of k points in $\mathbb{R}^d$, take $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ to be a suitably scaled projection on a random subspace of dimension d', then f is an ϵ-isometry for X with probability at least $1 - \frac{1}{k^\tau}$.

This lemma is quite remarkable

- One can run dimension reduction in a streaming fashion
- $\neq$ PCA

JL and Gaussian matrices

The Johnson-Lindenstrauss lemma is not connected with sparsity features, but it can be closely related to the RIP property derived above.

Lemma (Lemma 9.35 in Foucart & Rauhut)

Let $x_1, \dots, x_k$ of $\mathbb{R}^d$ be an arbitrary set of points and $\eta > 0$. If

$$d' > c \cdot \eta^{-2} \ln(k),$$

with c a universal constant, then there exists a matrix $B \in \mathbb{R}^{d' \times d}$ such that

$$(1-\eta)\|x_j - x_\ell\|_2^2 \leq \|B(x_j - x_\ell)\|_2^2 \leq (1+\eta)\|x_j - x_\ell\|_2^2,$$

for all $1 \leq j, \ell \leq k$.

Hint: use a Gaussian matrix for B.

Proof of JL for Gaussian I

- Consider the set

$$E = \{x_j - x_\ell, 1 \le j, \ell \le k\}.$$

of cardinality $|E| \le k(k-1)/2$ it is enough to show the existence of B such that

$$(1-\eta)\|x\|_2^2 \le \|Bx\|_2^2 \le (1+\eta)\|x\|_2^2, \quad \forall x \in E. \tag{7}$$

- Take $A = \frac{1}{\sqrt{k}} G$ with G a Gaussian matrix of size $d' \times d$. Using the same tools as in the RIP proof for Gaussian measurements, one can show that for all $x \in E$ and an appropriate constant $\tilde{c}$,

$$\mathbb{P}\left(\left|\|Bx\|_2^2 - \|x\|_2^2\right| \ge \eta\|x\|_2^2\right) \le 2\exp(-\tilde{c}k\eta^2).$$

Proof of JL for Gaussian II

- By a union bound, (7) holds simultaneously for all $x \in E$ with probability at least
$$1 - k^2 e^{-\tilde{c} d' \eta^2}.$$
- Choosing $d' \geq \frac{1}{\tilde{c}\eta^2} \ln(k^2/\varepsilon)$ leads (7) to hold with probability greater than $1 - \varepsilon$, so the existence of such a map is established when $\varepsilon < 1$. Considering the limit $\varepsilon \to 1$ gives the claim with $c = 2\tilde{c}^{-1}$.

Connection JL and RIP

- The concentration inequality for Gaussian matrices

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - \|x\|_2^2\right| \geq \eta\|x\|_2^2\right) \leq 2\exp(-\tilde{c}(k \text{ or } m)\eta^2).$$

  is a cornerstone
  - to derive the Johnson-Lindenstrauss lemma;
  - to establish the RIP property for Gaussian matrices.

Now the question is: when I have a matrix satisfying RIP, can I use it for efficient embedding of k points (that are non-sparse)?

The answer is yes...

... but you need to randomize the columns!

Theorem

Let $E \subset \mathbb{R}^d$ be a finite point set of cardinality $|E| = k$. For $\eta, \varepsilon \in (0,1)$, let $A \in \mathbb{R}^{d' \times d}$ with restricted isometry constant satisfying $\delta_{2s} \leq \eta/4$ for some $s \geq 16 \ln(4k/\varepsilon)$ and let $\epsilon_1, \ldots, \epsilon_d$ be a Rademacher sequence. Then with probability exceeding $1 - \varepsilon$

$$(1-\eta)\|x\|_2^2 \leq \|A\mathit{diag}(\epsilon)x\|_2^2 \leq (1+\eta)\|x\|_2^2, \quad \forall x \in E.$$

The answer is yes...

- Without a randomization of the column signs, the theorem does not hold.
- Some additional condition should be added on the point cloud.
- Suppose that some points of E are in the kernel of A (which is not assumed to be random in this theorem) then there is no chance for the lower bound to hold.
- Randomization of the column signs ensures that the probability of E intersecting the kernel of $A\mathrm{diag}(\epsilon)$ is small.

Summary

Exercise

Let A be an $m \times d$ random matrix which entries are i.i.d. Rademacher random variables, i.e.

$$\mathbb{P}(a_{ij} = +1) = 1/2 \qquad \mathbb{P}(a_{ij} = -1) = 1/2.$$

Show the following concentration result: for $x \in \mathbb{R}^n$ and $t \in]0,1[$,

$$\mathbb{P}\left(\left|m^{-1}\|Ax\|_2^2 - \|x\|_2^2\right| \geq t\|x\|_2^2\right) \leq 2\exp\left(-c_0 \min(t, t^2)\right).$$

NB: a better bound is

$$\mathbb{P}\left(\left|m^{-1}\|Ax\|_2^2 - \|x\|_2^2\right| \geq t\|x\|_2^2\right) \leq 2\exp\left(-m(t^2/4 - t^3/6)\right).$$

What can you deduce on such random matrices?