Sparse recovery under weak moment assumptions

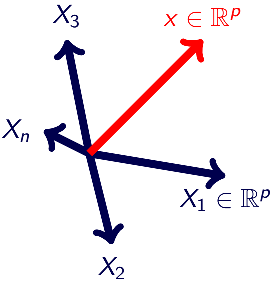
Guillaume Lecué

CNRS, CREST, ENSAE

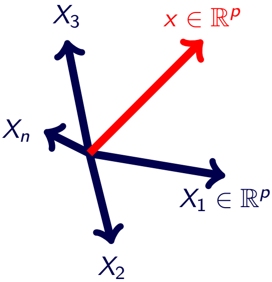
August 2016 - journée MAS Grenoble

joint works with Sjoerd Dirksen, Shahar Mendelson and Holger Rauhut

Exact reconstruction from few linear measurements (Compressed sensing)

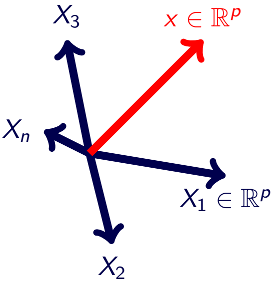


Exact reconstruction from few linear measurements (Compressed sensing)



$X_1, \ldots, X_n$: n measurements vectors

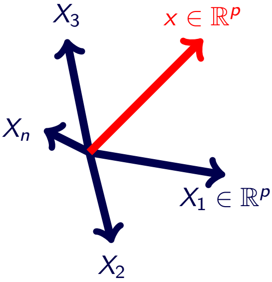
$\|x\|_0 = |\{j : x_j \neq 0\}| \leq s$: s-sparse



$X_1, \ldots, X_n$: n measurements vectors

$\|x\|_0 = |\{j : x_j \neq 0\}| \leq s$: s-sparse

aim : exact reconstruction of any s-sparse vector x from $\big(\langle X_i, x\rangle\big)_{i=1}^n$



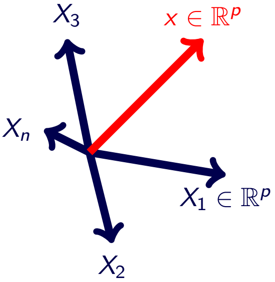
$X_1, \ldots, X_n$: n measurements vectors

$\|x\|_0 = |\{j : x_j \neq 0\}| \leq s$: s-sparse

aim : exact reconstruction of any s-sparse vector x from $\big(\langle X_i, x\rangle\big)_{i=1}^n$

Questions :

Exact reconstruction from few linear measurements (Compressed sensing)



$X_1, \dots, X_n$: n measurements vectors

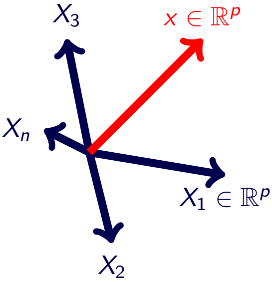
$\|x\|_0 = |\{j : x_j \neq 0\}| \leq s$: s-sparse

aim : exact reconstruction of any s-sparse vector x from $\big(\langle X_i, x\rangle\big)_{i=1}^n$

Questions :

- what is the minimal number of measurements n ?

Exact reconstruction from few linear measurements (Compressed sensing)



$X_1, \dots, X_n$: n measurements vectors

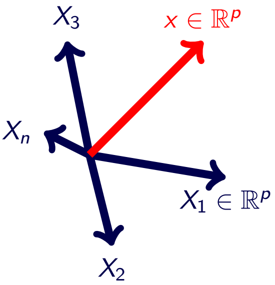
$\|x\|_0 = |\{j : x_j \neq 0\}| \leq s$: s-sparse

aim : exact reconstruction of any s-sparse vector x from $\big(\langle X_i, x\rangle\big)_{i=1}^n$

Questions :

- what is the minimal number of measurements n ?
- how to choose the measurement vectors $X_1, \dots, X_n$?

Exact reconstruction from few linear measurements (Compressed sensing)



$X_1, \ldots, X_n$: n measurements vectors

$\|x\|_0 = |\{j : x_j \neq 0\}| \leq s$: s-sparse

aim : exact reconstruction of any s-sparse vector x from $\big(\langle X_i, x\rangle\big)_{i=1}^n$

Questions :

- what is the minimal number of measurements n ?
- how to choose the measurement vectors $X_1, \ldots, X_n$?

p : space dimension, n : number of measurements, s : sparsity parameter.

ℓ_0-minimization is NP-hard

ℓ_0-minimization procedure

$$\text{minimize } \|t\|_0 \text{ subject to } \langle X_i, t\rangle = \langle X_i, x\rangle, i = 1, \ldots, n.$$

ℓ_0-minimization is NP-hard

ℓ_0-minimization procedure

$$\text{minimize } \|t\|_0 \text{ subject to } \langle X_i, t\rangle = \langle X_i, x\rangle, i = 1, \ldots, n.$$

$$\operatorname{argmin}\left(\|t\|_0 : \Gamma t = \Gamma x\right) = \{x\} \text{ for any } \|x\|_0 \le s.$$

ℓ_0-minimization procedure

$$\text{minimize } \|t\|_0 \text{ subject to } \left\langle X_i, t \right\rangle = \left\langle X_i, x \right\rangle, i = 1, \ldots, n.$$

$$\text{argmin} \left(\|t\|_0 : \Gamma t = \Gamma x \right) = \{x\} \text{ for any } \|x\|_0 \leq s.$$

$\Leftrightarrow$

$$\Gamma = \left[\begin{array}{c|c|c} X_1^\top / \sqrt{n} & & \\ \cdots & \overset{|I| = 2s}{\Gamma_I} & \\ X_n^\top / \sqrt{n} & & \end{array}\right] \qquad \ker \Gamma_I = \{0\}, \forall |I| = 2s$$

1. $n \geq 2s$ is the minimal number of measurements.

ℓ_0-minimization is NP-hard

ℓ_0-minimization procedure

$$\text{minimize } \|t\|_0 \text{ subject to } \left\langle X_i, t\right\rangle = \left\langle X_i, x\right\rangle, i = 1, \ldots, n.$$

$$\text{argmin}\left(\|t\|_0 : \Gamma t = \Gamma x\right) = \{x\} \text{ for any } \|x\|_0 \leq s.$$

$\Leftrightarrow$

$$\begin{array}{cc} & |I| = 2s \\ \Gamma = \begin{bmatrix} X_1^\top/\sqrt{n} \\ \cdots \\ X_n^\top/\sqrt{n} \end{bmatrix} & \boxed{\Gamma_I} \end{array} \qquad \ker\Gamma_I = \{0\}, \forall |I| = 2s$$

1. $n \geq 2s$ is the minimal number of measurements.
2. $\Gamma =$ the $2s$ first Fourier basis vectors

ℓ_0-minimization procedure

$$\text{minimize } \|t\|_0 \text{ subject to } \langle X_i, t\rangle = \langle X_i, x\rangle, i = 1, \dots, n.$$

$$\text{argmin}\left(\|t\|_0 : \Gamma t = \Gamma x\right) = \{x\} \text{ for any } \|x\|_0 \le s.$$

$\Leftrightarrow$

$|I| = 2s$

$$\Gamma = \begin{bmatrix} X_1^\top/\sqrt{n} \\ \cdots \\ X_n^\top/\sqrt{n} \end{bmatrix} \qquad \Gamma_I \qquad \ker\Gamma_I = \{0\}, \forall |I| = 2s$$

1. $n \ge 2s$ is the minimal number of measurements.
2. $\Gamma =$ the $2s$ first Fourier basis vectors
3. Natarajan, 1995 : ℓ_0-minimization is NP-hard (solves the "exact cover by 3-sets problem").

Basis pursuit – [Logan, 1965], [Donoho, Logan, 1992], [...]

$$\text{minimize } \|t\|_1 \text{ subject to } \langle X_i, t\rangle = \langle X_i, x\rangle, i = 1, \dots, n.$$

Basis pursuit – [Logan, 1965], [Donoho, Logan, 1992], [...]

$$\text{minimize } \|t\|_1 \text{ subject to } \langle X_i, t\rangle = \langle X_i, x\rangle, i = 1, \ldots, n.$$

BP can be recasted to a linear programming

Basis pursuit – [Logan, 1965], [Donoho, Logan, 1992], [...]

$$\text{minimize } \|t\|_1 \text{ subject to } \langle X_i, t\rangle = \langle X_i, x\rangle, i = 1, \ldots, n.$$

BP can be recasted to a linear programming

Definition

We say that $\Gamma = n^{-1/2}\sum_{i=1}^n \langle X_i, \cdot\rangle e_i$ satisfies the exact reconstruction property of order s when :

$$\text{for any } \|x\|_0 \leq s, \quad \operatorname{argmin}\Big(\|t\|_1 : \Gamma t = \Gamma x\Big) = \{x\} \qquad (ER(s))$$

Basis pursuit – [Logan, 1965], [Donoho, Logan, 1992], [...]

$$\text{minimize } \|t\|_1 \text{ subject to } \langle X_i, t\rangle = \langle X_i, x\rangle, i = 1, \ldots, n.$$

BP can be recasted to a linear programming

Definition

We say that $\Gamma = n^{-1/2}\sum_{i=1}^n \langle X_i, \cdot\rangle e_i$ satisfies the exact reconstruction property of order s when :

$$\text{for any } \|x\|_0 \leq s, \quad \operatorname{argmin}\Big(\|t\|_1 : \Gamma t = \Gamma x\Big) = \{x\} \qquad (ER(s))$$

Proposition : Γ satisfies ER(s) $\Rightarrow$ $n \gtrsim s\log(ep/s)$

Basis pursuit – [Logan, 1965], [Donoho, Logan, 1992], [...]

$$\text{minimize } \|t\|_1 \text{ subject to } \big\langle X_i, t\big\rangle = \big\langle X_i, x\big\rangle, i = 1, \ldots, n.$$

BP can be recasted to a linear programming

Definition

We say that $\Gamma = n^{-1/2}\sum_{i=1}^n \langle X_i, \cdot\rangle e_i$ satisfies the exact reconstruction property of order s when :

$$\text{for any } \|x\|_0 \leq s, \quad \text{argmin}\Big(\|t\|_1 : \Gamma t = \Gamma x\Big) = \{x\} \qquad (ER(s))$$

Proposition : Γ satisfies ER(s) $\Rightarrow$ $n \gtrsim s\log(ep/s)$

Question : construction of Γ satisfying ER(s) with $n \sim s\log(ep/s)$.

RIP($c_0 s$) : for any $\|x\|_0 \leq c_0 s, \quad \frac{1}{2}\|x\|_2 \leq \|\Gamma x\|_2 \leq \frac{3}{2}\|x\|_2.$

[Candès & Romberg & Tao , 05, 06]

$$\mathrm{RIP}(c_0 s): \qquad \text{for any } \|x\|_0 \le c_0 s, \qquad \tfrac{1}{2}\|x\|_2 \le \|\Gamma x\|_2 \le \tfrac{3}{2}\|x\|_2.$$

[Candès & Romberg & Tao , 05, 06]

$$\Downarrow$$

$$\text{for any } x \in \sqrt{c_1 s} B_1^p \cap S_2^{p-1}, \qquad \|\Gamma x\|_2 > 0$$

[Kashin & Temlyakov , 07]

RIP($c_0 s$) : for any $\|x\|_0 \leq c_0 s, \quad \frac{1}{2}\|x\|_2 \leq \|\Gamma x\|_2 \leq \frac{3}{2}\|x\|_2.$

[Candès & Romberg & Tao , 05, 06]

$$\Downarrow$$

for any $x \in \sqrt{c_1 s} B_1^p \cap S_2^{p-1}, \quad \|\Gamma x\|_2 > 0$

[Kashin & Temlyakov , 07]

$$\Downarrow$$

NSP(s) : for any $h \in \ker \Gamma \backslash \{0\}, |I| \leq s, \|h_I\|_1 < \|h_{I^c}\|_1$

[Donoho , Elad , Huo , 01, 03]

RIP($c_0 s$) : for any $\|x\|_0 \le c_0 s$, $\frac{1}{2}\|x\|_2 \le \|\Gamma x\|_2 \le \frac{3}{2}\|x\|_2$.

[Candès & Romberg & Tao , 05, 06]

$$\Downarrow$$

for any $x \in \sqrt{c_1 s} B_1^p \cap S_2^{p-1}$, $\|\Gamma x\|_2 > 0$

[Kashin & Temlyakov , 07]

$$\Downarrow$$

NSP(s) : for any $h \in \ker\Gamma \backslash \{0\}, |I| \le s, \|h_I\|_1 < \|h_{I^c}\|_1$

[Donoho , Elad , Huo , 01, 03]

$$\Updownarrow$$

ER(s) for any $\|x\|_0 \le s$, $\operatorname{argmin}\Big(\|t\|_1 : \Gamma t = \Gamma x\Big) = \{x\}$

$$\text{RIP}(c_0 s): \quad \text{for any } \|x\|_0 \le c_0 s, \quad \tfrac{1}{2}\|x\|_2 \le \|\Gamma x\|_2 \le \tfrac{3}{2}\|x\|_2.$$

[Candès & Romberg & Tao , 05, 06]

$$\Downarrow$$

$$\text{for any } x \in \sqrt{c_1 s} B_1^p \cap S_2^{p-1}, \quad \|\Gamma x\|_2 > 0$$

[Kashin & Temlyakov , 07]

$$\Downarrow$$

$$\text{NSP}(s): \quad \text{for any } h \in \ker \Gamma \backslash \{0\}, |I| \le s, \|h_I\|_1 < \|h_{I^c}\|_1$$

[Donoho , Elad , Huo , 01, 03]

$$\Updownarrow$$

$$\text{ER}(s) \quad \text{for any } \|x\|_0 \le s, \quad \text{argmin}\Big(\|t\|_1 : \Gamma t = \Gamma x\Big) = \{x\}$$

$(\Leftrightarrow \Gamma B_1^p$ is $s -$ neighborly, [Donoho, 05])

$(\Leftarrow$ Incoherency conditions)

Random matrix

So far, only random matrices satisfy ER(s) with $n \sim s \log\left(ep/s\right)$ (with large probability).

Random matrix

So far, only random matrices satisfy ER(s) with $n \sim s \log\big(ep/s\big)$ (with large probability).

Examples :

1. Independent, isotropic (i.e. $\mathbb{E}\big\langle X, t\big\rangle^2 = \|t\|_2^2$) and subgaussian (i.e. $\mathbb{P}[|\big\langle X, t\big\rangle| \geq u\|t\|_2] \leq 2\exp(-c_0 u^2)$) rows : RIP($s$) is satisfied when $n \sim s\log(ep/s)$ [Candès, Tao, Vershynin, Rudelson, Mendelson, Pajor, Tomjack-Jaegermann].

Random matrix

So far, only random matrices satisfy ER(s) with $n \sim s \log\left(ep/s\right)$ (with large probability).

Examples :

1. Independent, isotropic (i.e. $\mathbb{E}\langle X, t\rangle^2 = \|t\|_2^2$) and subgaussian (i.e. $\mathbb{P}[|\langle X, t\rangle| \geq u\|t\|_2] \leq 2\exp(-c_0 u^2)$) rows : RIP($s$) is satisfied when $n \sim s\log(ep/s)$ [Candès, Tao, Vershynin, Rudelson, Mendelson, Pajor, Tomjack-Jaegermann].
2. independent log-concave rows or independent sub-exponential columns satisfy RIP(s) when $n \sim s\log^2(ep/s)$ [Adamczack, Latala, Litvak, Pajor, Tomjack-Jaegermann].

Random matrix

So far, only random matrices satisfy ER(s) with $n \sim s \log\left(ep/s\right)$ (with large probability).

Examples :

1. Independent, isotropic (i.e. $\mathbb{E}\langle X, t\rangle^2 = \|t\|_2^2$) and subgaussian (i.e. $\mathbb{P}[|\langle X, t\rangle| \geq u\|t\|_2] \leq 2\exp(-c_0 u^2)$) rows : RIP($s$) is satisfied when $n \sim s\log(ep/s)$ [Candès, Tao, Vershynin, Rudelson, Mendelson, Pajor, Tomjack-Jaegermann].
2. independent log-concave rows or independent sub-exponential columns satisfy RIP(s) when $n \sim s\log^2(ep/s)$ [Adamczack, Latala, Litvak, Pajor, Tomjack-Jaegermann].
3. Structured matrices : partial Fourier matrices satisfy RIP(s) [Rudelson, Vershynin, Candès, Tao, Bourgain] when $n \gtrsim s\log^3(p)$.

What property of randomness is used for exact reconstruction ? concentration ?

What property of randomness is used for exact reconstruction ? concentration ?

Can we take "Cauchy measurements" (density $\propto (1 + x^2)^{-1}$) ?

$$X = (x_1, \dots, x_p) \text{ where } x_i \text{ are ind. Cauchy variables}$$

and still get the same properties as for "Gaussian measurements"?

Can we take "Cauchy measurements" (density $\propto (1+x^2)^{-1}$) ?

$$X = (x_1, \ldots, x_p) \text{ where } x_i \text{ are ind. Cauchy variables}$$

and still get the same properties as for "Gaussian measurements"?

Here : we will need $\log p$ moments for the coordinates x_j's.

What property of randomness is used for exact reconstruction ? concentration ?

1. Let $\Gamma = n^{-1/2}(e_{ij}) \in \mathbb{R}^{n\times p}$ where e_{ij} are iid symmetric exponential.

1. Let $\Gamma = n^{-1/2}(e_{ij}) \in \mathbb{R}^{n\times p}$ where e_{ij} are iid symmetric exponential.
   If Γ satisfies RIP(s) with probability at least 1/2 then

$$n \gtrsim s \log^2 \left(ep/s\right).$$

[Adamczack, Latala, Litvak, Pajor, Tomjack-Jaegermann]

1. Let $\Gamma = n^{-1/2}(e_{ij}) \in \mathbb{R}^{n\times p}$ where e_{ij} are iid symmetric exponential.

   If Γ satisfies RIP(s) with probability at least $1/2$ then

   $$n \gtrsim s \log^{2}\left(ep/s\right).$$

   [Adamczack, Latala, Litvak, Pajor, Tomjack-Jaegermann]

2. If Γ has independent isotropic sub-exponential rows then it satisfies ER(s) with large probability when

   $$n \gtrsim s \log\left(ep/s\right).$$

   [Koltchinksii] and [Foucart and Lai]

What property of randomness is used for exact reconstruction ? concentration ?

1. Let $\Gamma = n^{-1/2}(e_{ij}) \in \mathbb{R}^{n\times p}$ where e_{ij} are iid symmetric exponential. If Γ satisfies RIP(s) with probability at least $1/2$ then

$$n \gtrsim s \log^{2}\left(ep/s\right).$$

[Adamczack, Latala, Litvak, Pajor, Tomjack-Jaegermann]

2. If Γ has independent isotropic sub-exponential rows then it satisfies ER(s) with large probability when

$$n \gtrsim s \log\left(ep/s\right).$$

[Koltchinksii] and [Foucart and Lai]

Exact reconstruction under weak concentration property cannot be studied via RIP

A weaker condition than RIP

Proposition (L. and Mendelson)

Let $\Gamma : \mathbb{R}^p \mapsto \mathbb{R}^n$ such that

1. *for any $\|t\|_0 \leq s$: $\|\Gamma t\|_2 \geq \kappa_0 \|t\|_2$,*

Proposition (L. and Mendelson)

Let $\Gamma : \mathbb{R}^p \mapsto \mathbb{R}^n$ *such that*

1. *for any* $\|t\|_0 \leq s$ *:* $\|\Gamma t\|_2 \geq \kappa_0 \|t\|_2$,
2. $\|\Gamma e_j\|_2 \leq c_0, \forall 1 \leq j \leq p$ *(where* $(e_1, \ldots, e_p)$ *is the canonical basis)*

A weaker condition than RIP

Proposition (L. and Mendelson)

Let $\Gamma : \mathbb{R}^p \mapsto \mathbb{R}^n$ *such that*

1. *for any* $\|t\|_0 \le s$: $\|\Gamma t\|_2 \ge \kappa_0 \|t\|_2$,
2. $\|\Gamma e_j\|_2 \le c_0, \forall 1 \le j \le p$ *(where* $(e_1, \dots, e_p)$ *is the canonical basis)*

Then, $\forall t \in \sqrt{c_1 s} B_1^p \cap S_2^{p-1}$, $\|\Gamma t\|_2 \ge c_2 > 0$ *and so* *ER(s)* *holds.*

Comparison with RIP

$$\forall \|t\|_0 \leq s, \frac{1}{2}\|t\|_2 \leq \|\Gamma t\|_2 \leq \frac{3}{2}\|t\|_2$$

Comparison with RIP

$$\forall \|t\|_0 \le s, \frac{1}{2}\|t\|_2 \le \|\Gamma t\|_2 \le \frac{3}{2}\|t\|_2$$

1. $\forall \|t\|_0 \le s \; \|\Gamma t\|_2 \ge \kappa_0 \|t\|_2$,
2. $\max_{1 \le j \le p} \|\Gamma e_j\|_2 \le c_0$.

Comparison with RIP

$$\forall \|t\|_0 \le s, \frac{1}{2}\|t\|_2 \le \|\Gamma t\|_2 \le \frac{3}{2}\|t\|_2$$

1. $\forall \|t\|_0 \le s \ \|\Gamma t\|_2 \ge \kappa_0 \|t\|_2$,
2. $\max_{1 \le j \le p} \|\Gamma e_j\|_2 \le c_0$.

Both implies Exact reconstruction ER(s).

$$\forall \|t\|_0 \leq s, \frac{1}{2}\|t\|_2 \leq \|\Gamma t\|_2 \leq \frac{3}{2}\|t\|_2$$

1. $\forall \|t\|_0 \leq s \; \|\Gamma t\|_2 \geq \kappa_0 \|t\|_2$,
2. $\max_{1 \leq j \leq p} \|\Gamma e_j\|_2 \leq c_0$.

Both implies Exact reconstruction ER(s).

- LHS of RIP is implied by the small ball property : “no cost”

$$\forall \|t\|_0 \leq s, \frac{1}{2}\|t\|_2 \leq \|\Gamma t\|_2 \leq \frac{3}{2}\|t\|_2$$

1. $\forall \|t\|_0 \leq s \; \|\Gamma t\|_2 \geq \kappa_0 \|t\|_2$,
2. $\max_{1 \leq j \leq p} \|\Gamma e_j\|_2 \leq c_0$.

Both implies Exact reconstruction ER(s).

- LHS of RIP is implied by the small ball property : "no cost"
- RHS of RIP requires deviation (ψ_2).

1. Z is subgaussian $(\psi_2) \Leftrightarrow \|Z\|_{L_q} \lesssim \sqrt{q}$ for all q's

1. Z is subgaussian $(\psi_2) \Leftrightarrow \|Z\|_{L_q} \lesssim \sqrt{q}$ for all q's
2. Z is subexponential $(\psi_1) \Leftrightarrow \|Z\|_{L_q} \lesssim q$ for all q's

1. Z is subgaussian $(\psi_2) \Leftrightarrow \|Z\|_{L_q} \lesssim \sqrt{q}$ for all q's
2. Z is subexponential $(\psi_1) \Leftrightarrow \|Z\|_{L_q} \lesssim q$ for all q's
3. Z is $\psi_{\alpha} \Leftrightarrow \|Z\|_{L_q} \lesssim q^{1/\alpha}$ for all q's

1. Z is subgaussian $(\psi_2) \Leftrightarrow \|Z\|_{L_q} \lesssim \sqrt{q}$ for all q's
2. Z is subexponential $(\psi_1) \Leftrightarrow \|Z\|_{L_q} \lesssim q$ for all q's
3. Z is $\psi_\alpha \Leftrightarrow \|Z\|_{L_q} \lesssim q^{1/\alpha}$ for all q's

Here, we assume that the measurement vector $X = (x_1, \ldots, x_p) \in \mathbb{R}^p$ is such that : $\|x_j\|_{L_2} = 1$ and

$$\|x_j\|_{L_q} \leq \kappa_0 q^\eta, \text{ for } q \sim \log(p)$$

for some $\eta \geq 1/2$.

Small ball property - the second assumption. [Mendelson], [Koltchinskii, Mendelson]

Small ball property - the second assumption. [Mendelson], [Koltchinskii, Mendelson]

There exists two constants u_0, β_0 such that : $\forall \|t\|_0 \leq s$,

$$P\left[|\langle X, t\rangle| \geq u_0\|t\|_2\right] \geq \beta_0$$

Small ball property - the second assumption. [Mendelson], [Koltchinskii, Mendelson]

There exists two constants u_0, β_0 such that : $\forall \|t\|_0 \leq s$,

$$P\big[|\langle X, t\rangle| \geq u_0 \|t\|_2\big] \geq \beta_0$$

Examples :

1) if X is isotropic ($\mathbb{E}\langle X, t\rangle^2 = \|t\|_2^2$) and for all $\|t\|_0 \leq s$,

$$\|\langle X, t\rangle\|_{L_{2+\epsilon}} \leq \kappa_0 \|\langle X, t\rangle\|_{L_2}$$

for some $\epsilon > 0$.

Small ball property - the second assumption. [Mendelson], [Koltchinskii, Mendelson]

There exists two constants u_0, β_0 such that : $\forall \|t\|_0 \leq s$,

$$P\left[|\langle X, t\rangle| \geq u_0 \|t\|_2\right] \geq \beta_0$$

Examples :

1) if X is isotropic ($\mathbb{E}\langle X, t\rangle^2 = \|t\|_2^2$) and for all $\|t\|_0 \leq s$,

$$\|\langle X, t\rangle\|_{L_{2+\epsilon}} \leq \kappa_0 \|\langle X, t\rangle\|_{L_2}$$

for some $\epsilon > 0$.

2) if X is isotropic and for all $\|t\|_0 \leq s$,

$$\|\langle X, t\rangle\|_{L_2} \leq \kappa_0 \|\langle X, t\rangle\|_{L_1}.$$

[RV13] : "Small ball probabilities for linear images of high dimensional distributions" M. Rudelson and R. Vershynin.

Small ball property - other examples from [RV13]

[RV13] : "Small ball probabilities for linear images of high dimensional distributions" M. Rudelson and R. Vershynin.

3) $X = (x_1, \dots, x_p)$ with independent absolutly continous coordinates with density bounded by K a.s. then for any $t \in \mathbb{R}^p$,

$$P\left[\left|\langle X, t\rangle\right| \geq (4\sqrt{2}K)^{-1}\|t\|_2\right] \geq \frac{1}{2}.$$

(For example, a Cauchy measurement vector satisfies the small ball property).

Small ball property - other examples from [RV13]

[RV13] : "Small ball probabilities for linear images of high dimensional distributions" M. Rudelson and R. Vershynin.

3) $X = (x_1, \dots, x_p)$ with independent absolutly continous coordinates with density bounded by K a.s. then for any $t \in \mathbb{R}^p$,

$$P\Big[\big|\langle X, t\rangle\big| \geq (4\sqrt{2}K)^{-1}\|t\|_2\Big] \geq \frac{1}{2}.$$

(For example, a Cauchy measurement vector satisfies the small ball property).

4) $X = (x_1, \dots, x_p)$ with independent coordinates such that

$$\mathcal{L}(x_i, t_0) := \sup_{u\in\mathbb{R}} P[|x_i - u| \leq t_0] \leq p_0$$

for some $t_0, p_0 > 0$.

[RV13] : "Small ball probabilities for linear images of high dimensional distributions" M. Rudelson and R. Vershynin.

3) $X = (x_1, \dots, x_p)$ with independent absolutly continous coordinates with density bounded by K a.s. then for any $t \in \mathbb{R}^p$,

$$P\Big[\big|\langle X, t\rangle\big| \geq (4\sqrt{2}K)^{-1}\|t\|_2\Big] \geq \frac{1}{2}.$$

(For example, a Cauchy measurement vector satisfies the small ball property).

4) $X = (x_1, \dots, x_p)$ with independent coordinates such that

$$\mathcal{L}(x_i, t_0) := \sup_{u \in \mathbb{R}} P[|x_i - u| \leq t_0] \leq p_0$$

for some $t_0, p_0 > 0$. then for any $t \in \mathbb{R}^p$,

$$P\Big[\big|\langle X, t\rangle\big| \geq t_0\|t\|_2\Big] \geq 1 - c_0 p_0.$$

Theorem (L. & Mendelson)

Let $X_1, \ldots, X_n$ be n iid $\sim X = (x_1, \ldots, x_p)^\top$ random variables in $\mathbb{R}^p$ s.t. :

Theorem (L. & Mendelson)

Let $X_1, \dots, X_n$ *be* n *iid* $\sim X = (x_1, \dots, x_p)^\top$ *random variables in* $\mathbb{R}^p$ *s.t. :*

1. $\|x_j\|_{L_2} = 1$ *and for some* $\eta \geq 1/2$ *and* $q = \kappa_1 \log(wp)$ *:*

$$\|x_j\|_{L_q} \leq \kappa_0 q^\eta,$$

Theorem (L. & Mendelson)

Let $X_1, \dots, X_n$ *be n iid* $\sim X = (x_1, \dots, x_p)^\top$ *random variables in* $\mathbb{R}^p$ *s.t. :*

1. $\|x_j\|_{L_2} = 1$ *and for some* $\eta \geq 1/2$ *and* $q = \kappa_1 \log(wp)$ *:*

$$\|x_j\|_{L_q} \leq \kappa_0 q^\eta,$$

2. *there exists* u_0, β_0 *such that :* $\forall t, \|t\|_0 \leq s$*,*

$$P\big[|\langle X, t\rangle| \geq u_0 \|t\|_2\big] \geq \beta_0$$

Theorem (L. & Mendelson)

Let $X_1, \dots, X_n$ *be* n *iid* $\sim X = (x_1, \dots, x_p)^\top$ *random variables in* $\mathbb{R}^p$ *s.t. :*

1. $\|x_j\|_{L_2} = 1$ *and for some* $\eta \geq 1/2$ *and* $q = \kappa_1 \log(wp)$ *:*

$$\|x_j\|_{L_q} \leq \kappa_0 q^\eta,$$

2. *there exists* u_0, β_0 *such that :* $\forall t, \|t\|_0 \leq s$,

$$P\left[|\langle X, t\rangle| \geq u_0\|t\|_2\right] \geq \beta_0$$

3. $n \gtrsim s \log(ep/s)$.

Theorem (L. & Mendelson)

Let $X_1, \ldots, X_n$ *be n iid* $\sim X = (x_1, \ldots, x_p)^\top$ *random variables in* $\mathbb{R}^p$ *s.t. :*

1. $\|x_j\|_{L_2} = 1$ *and for some* $\eta \geq 1/2$ *and* $q = \kappa_1 \log(wp)$ *:*

$$\|x_j\|_{L_q} \leq \kappa_0 q^\eta,$$

2. *there exists* u_0, β_0 *such that :* $\forall t, \|t\|_0 \leq s$,

$$P\big[|\langle X, t\rangle| \geq u_0\|t\|_2\big] \geq \beta_0$$

3. $n \gtrsim s\log(ep/s)$.

Then, with probability at least $1 - 2\exp(-c_1 n\beta_0^2) - 1/(w^{\kappa_1}p^{\kappa_1 - 1})$,

$$\Gamma = \frac{1}{\sqrt{n}}\begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \text{ satisfies } ER(c_2 u_0^2 \beta_0 s).$$

Theorem (L. & Mendelson)

There exists a real-valued random variable x such that

1. $\mathbb{E}x = 0$, $\mathbb{E}x^2 = 1$, $\mathbb{E}x^4 \lesssim 1$

Theorem (L. & Mendelson)

There exists a real-valued random variable x such that

1. $\mathbb{E}x = 0$, $\mathbb{E}x^2 = 1$, $\mathbb{E}x^4 \lesssim 1$
2. $\|x\|_{L_q} \lesssim \sqrt{q}$ *for* $q \sim (\log p)/\log\log p$

$\log p$ moments are almost necessary

Theorem (L. & Mendelson)

There exists a real-valued random variable x such that

1. $\mathbb{E}x = 0$, $\mathbb{E}x^2 = 1$, $\mathbb{E}x^4 \lesssim 1$
2. $\|x\|_{L_q} \lesssim \sqrt{q}$ *for* $q \sim (\log p)/\log\log p$

for which if x_{ij} *are iid*$\sim x$*, for* $n \sim \log p$ *then*

$$\Gamma = n^{-1/2}\big(x_{ij} : 1 \le i \le n, 1 \le j \le p\big),$$

Γ *does not satisfy the ER(1) with probability at least* $1/2$*.*

$\log p$ moments are almost necessary

Theorem (L. & Mendelson)

There exists a real-valued random variable x such that

1. $\mathbb{E}x = 0$, $\mathbb{E}x^2 = 1$, $\mathbb{E}x^4 \lesssim 1$
2. $\|x\|_{L_q} \lesssim \sqrt{q}$ *for* $q \sim (\log p)/\log\log p$

for which if x_{ij} are iid$\sim x$, for $n \sim \log p$ then

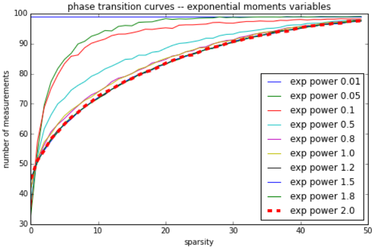
$$\Gamma = n^{-1/2}\big(x_{ij} : 1 \leq i \leq n, 1 \leq j \leq p\big),$$

Γ does not satisfy the ER(1) with probability at least $1/2$.

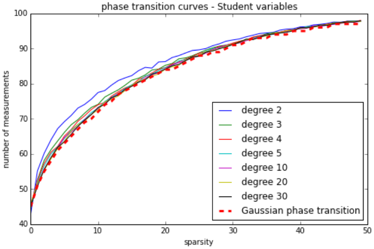
$\Rightarrow$ We need at least $\log p / \log\log p$ moments for exact reconstruction via basis pursuit.

phase transition diagram for Exponential and Student variables.

ψ_γ variables $\mathrm{sign}(g)|g|^{2/\gamma}$ where $g \sim \mathcal{N}(0,1)$



Student variables of degree k density $\sim (1+t^2)^{-(k+1)/2}$



A price to pay for convex relaxation

Theorem (L. & Mendelson)

Let $X_1, \ldots, X_n$ be n iid $\sim X$ random variables in $\mathbb{R}^p$ s.t. :

A price to pay for convex relaxation

Theorem (L. & Mendelson)

Let $X_1, \ldots, X_n$ be n iid $\sim X$ random variables in $\mathbb{R}^p$ s.t. :

1. *there exists u_0, β_0 such that : $\forall t, \|t\|_0 \leq s$,*

$$P\big[|\langle X, t\rangle| \geq u_0\|t\|_2\big] \geq \beta_0$$

Theorem (L. & Mendelson)

Let $X_1, \dots, X_n$ be n iid $\sim X$ random variables in $\mathbb{R}^p$ s.t. :

1. *there exists u_0, β_0 such that : $\forall t, \|t\|_0 \leq s$,*

$$P\big[|\langle X, t\rangle| \geq u_0 \|t\|_2\big] \geq \beta_0$$

2. *$n \gtrsim s \log(ep/s)$.*

Theorem (L. & Mendelson)

Let $X_1, \ldots, X_n$ be n iid $\sim X$ random variables in $\mathbb{R}^p$ s.t. :

1. *there exists u_0, β_0 such that : $\forall t, \|t\|_0 \leq s$,*

$$P\big[|\langle X, t\rangle| \geq u_0 \|t\|_2\big] \geq \beta_0$$

2. $n \gtrsim s\log(ep/s)$.

Then, with probability at least $1 - 2\exp(-c_1 n \beta_0^2)$,

$$\Gamma = \frac{1}{\sqrt{n}} \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix}.$$

is such that for any $\|x\|_0 \leq s$, $\operatorname{argmin}\big(\|t\|_0 : \Gamma t = \Gamma x\big) = \{x\}$.

Theorem (L. & Mendelson)

Let $X_1, \dots, X_n$ be n iid $\sim X$ random variables in $\mathbb{R}^p$ s.t. :

1. *there exists u_0, β_0 such that : $\forall t, \|t\|_0 \leq s$,*
$$P\left[|\langle X, t\rangle| \geq u_0\|t\|_2\right] \geq \beta_0$$
2. $n \gtrsim s\log(ep/s)$.

Then, with probability at least $1 - 2\exp(-c_1 n\beta_0^2)$,

$$\Gamma = \frac{1}{\sqrt{n}}\begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix}.$$

is such that for any $\|x\|_0 \leq s$, $\mathrm{argmin}\left(\|t\|_0 : \Gamma t = \Gamma x\right) = \{x\}$.

$\Rightarrow$ We don't need moment assumption for $\ell_0 -$ *minimization*. This proves that there is a *price to pay* in terms of concentration for convex relaxation.

1. Exact reconstruction via Basis Pursuit for random linear measurements under $\log p$ moments is possible with the same number of measurements as in the Gaussian case.

1. Exact reconstruction via Basis Pursuit for random linear measurements under $\log p$ moments is possible with the same number of measurements as in the Gaussian case.
2. RIP needs ψ_2-concentration (and thus may not be the "optimal" way to prove exact reconstruction)

1. Exact reconstruction via Basis Pursuit for random linear measurements under $\log p$ moments is possible with the same number of measurements as in the Gaussian case.
2. RIP needs ψ_2-concentration (and thus may not be the "optimal" way to prove exact reconstruction)
3. the property of randomness that looks "important" for exact reconstruction of s-sparse vectors : $\forall \|t\|_0 \leq s$,

$$P\left[|\langle X, t\rangle| \geq u_0\|t\|_2\right] \geq \beta_0.$$

1. Exact reconstruction via Basis Pursuit for random linear measurements under $\log p$ moments is possible with the same number of measurements as in the Gaussian case.
2. RIP needs ψ_2-concentration (and thus may not be the "optimal" way to prove exact reconstruction)
3. the property of randomness that looks "important" for exact reconstruction of s-sparse vectors : $\forall \|t\|_0 \leq s$,

$$P\left[|\langle X, t\rangle| \geq u_0\|t\|_2\right] \geq \beta_0.$$

4. $\log p / \log\log p$ moments is a necessary price to pay for convex relaxation.

Data : $y = Q_\theta(\Gamma x)$ where $Q_\theta : \mathbb{R}^m \to (\theta\mathbb{Z} + \theta/2)^m$.

Data : $y = Q_\theta(\Gamma x)$ where $Q_\theta : \mathbb{R}^m \to (\theta\mathbb{Z} + \theta/2)^m$.
"Model" : $y = \Gamma x + e$ where $\|e\|_\infty \le \theta/2$

Data : $y = Q_\theta(\Gamma x)$ where $Q_\theta : \mathbb{R}^m \to (\theta\mathbb{Z} + \theta/2)^m$.
"Model" : $y = \Gamma x + e$ where $\|e\|_\infty \leq \theta/2$
Procedure : $BPDN_\infty$

$$\min_{t\in\mathbb{R}^p} \|t\|_1 \text{ s.t.} \|y - \Gamma t\|_\infty \leq \theta/2$$

Data : $y = Q_\theta(\Gamma x)$ where $Q_\theta : \mathbb{R}^m \to (\theta\mathbb{Z} + \theta/2)^m$.
"Model" : $y = \Gamma x + e$ where $\|e\|_\infty \leq \theta/2$
Procedure : $BPDN_\infty$

$$\min_{t \in \mathbb{R}^p} \|t\|_1 \text{ s.t.} \|y - \Gamma t\|_\infty \leq \theta/2$$

Results for BPDN_q from Jacques, Hammond, and Fadili, *Dequantizing compressed sensing : when oversampling and non-Gaussian constraints combine*, when $q \to \infty$ shows that if

$$n \gtrsim \Big(s \log(ep/s)\Big)^{q/2}$$

then $(RIP)_{q,2}$ holds with large probability for Gaussian measurements matrices Γ and then for any x

$$\|\hat{x}_{BPDN_q} - x\|_2 \lesssim \frac{\sigma_s(x)_1}{\sqrt{s}} + \frac{\theta}{\sqrt{q+1}}$$

Application to quantized CS

Theorem (Dirksen, L. and Rauhut)

For Gaussian measurements, when $n \gtrsim s\log(ep/s)$ *then*

$$\|\hat{x}_{BPDN_\infty} - x\|_2 \lesssim \frac{\sigma_s(x)_1}{\sqrt{s}} + \theta.$$

Moreover, $\hat{x}_{BPDN_\infty}$ *is quantized consistent :* $y = Q_\theta(\Gamma \hat{x}_{BPDN_\infty})$.

Theorem (Dirksen, L. and Rauhut)

For Gaussian measurements, when $n \gtrsim s \log(ep/s)$ then

$$\|\hat{x}_{BPDN_\infty} - x\|_2 \lesssim \frac{\sigma_s(x)_1}{\sqrt{s}} + \theta.$$

Moreover, $\hat{x}_{BPDN_\infty}$ is quantized consistent : $y = Q_\theta(\Gamma \hat{x}_{BPDN_\infty})$.

For the quantization problem $RIP_{q,2}$ requires more measurements.

Theorem (Dirksen, L. and Rauhut)

For Gaussian measurements, when $n \gtrsim s\log(ep/s)$ then

$$\|\hat{x}_{BPDN_\infty} - x\|_2 \lesssim \frac{\sigma_s(x)_1}{\sqrt{s}} + \theta.$$

Moreover, $\hat{x}_{BPDN_\infty}$ is quantized consistent : $y = Q_\theta(\Gamma\hat{x}_{BPDN_\infty})$.

For the quantization problem $RIP_{q,2}$ requires more measurements.
For analysis based on $RIP_{q,r}$: for any $x \in \Sigma_s$

$$c\|x\|_r \leq \|\Gamma x\|_q \leq C\|x\|_r$$

two phenomena occur :

1. more measurements than $s\log(ep/s)$
2. other type of matrices than Gaussian (adjacency matrices, stable processes).

Bypassing the RIP based approach show that none of these two phenomena actually occur : one can use $s\log(ep/s)$ Gaussian measures.

Thanks for your attention

G. Lecué and S. Mendelson, *Sparse recovery under weak moment assumption*. To appear in *Journal of the European Mathematical society*, Jan. 2014.

S. Dirksen, G. Lecué and H. Rauhut, *On the gap between restricted isometry properties and sparse recovery conditions*. To appear in *IEEE Transactions on Information Theory*, March 2015.

Sparse Linear Regression

Noisy data - LASSO

Data : $y = \mathcal{X}\beta^* + \sigma W$ where $W \sim \mathcal{N}_p(0, I_n)$ and

$$\mathcal{X} = \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \left(= \sqrt{n}\Gamma \right)$$

Noisy data - LASSO

Data : $y = \mathcal{X}\beta^* + \sigma W$ where $W \sim \mathcal{N}_p(0, I_n)$ and

$$\mathcal{X} = \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \Big(= \sqrt{n}\Gamma \Big)$$

Aims : Estimation of β^* / denoising of $\mathcal{X}\beta^*$ / prediction of outputs / support recovery.

Noisy data - LASSO

Data : $y = \mathcal{X}\beta^* + \sigma W$ where $W \sim \mathcal{N}_p(0, I_n)$ and

$$\mathcal{X} = \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \left(= \sqrt{n}\Gamma \right)$$

Aims : Estimation of β^* / denoising of $\mathcal{X}\beta^*$ / prediction of outputs / support recovery.

LASSO :

$$\hat{\beta} \in \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \left(\frac{1}{n} \|y - \mathcal{X}\beta\|_2^2 + \lambda \|\beta\|_{n,1} \right) \text{ for } \lambda \sim \sigma \sqrt{\frac{\log p}{n}}$$

Data : $y = \mathcal{X}\beta^* + \sigma W$ where $W \sim \mathcal{N}_p(0, I_n)$ and

$$\mathcal{X} = \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \Big(= \sqrt{n}\Gamma \Big)$$

Aims : Estimation of β^* / denoising of $\mathcal{X}\beta^*$ / prediction of outputs / support recovery.

LASSO :

$$\hat{\beta} \in \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \left(\frac{1}{n} \|y - \mathcal{X}\beta\|_2^2 + \lambda \|\beta\|_{n,1} \right) \text{ for } \lambda \sim \sigma \sqrt{\frac{\log p}{n}}$$

where

$$\|\beta\|_{n,1} = \sum_{j=1}^{p} r_{n,j} |\beta_j| \text{ and } r_{n,j} = \left(\frac{\mathcal{X}^\top \mathcal{X}}{n} \right)_{jj}.$$

Restricted eigenvalue condition : For any $I \subset [p]$ s.t. $|I| \leq s$, $v \in \mathbb{R}^p$

$$\|v_{I^c}\|_1 \leq 3\|v_I\|_1 \quad \Rightarrow \quad \|\Gamma v\|_2 \geq \kappa(s)\|v_I\|_2 \qquad (REC(s))$$

Restricted eigenvalue condition : For any $I \subset [p]$ s.t. $|I| \leq s$, $v \in \mathbb{R}^p$

$$\|v_{I^c}\|_1 \leq 3\|v_I\|_1 \quad \Rightarrow \quad \|\Gamma v\|_2 \geq \kappa(s)\|v_I\|_2 \qquad (REC(s))$$

Remark : (Null space property) $\forall I \subset [p]$ s.t. $|I| \leq s$, $v \in \mathbb{R}^p$

$$\|v_{I^c}\|_1 < \|v_I\|_1 \quad \Rightarrow \quad \|\Gamma v\|_2 > 0 \qquad (NSP(s))$$

If REC(s) holds and $\|\beta^*\|_0 = s$

If REC(s) holds and $\|\beta^*\|_0 = s$ then with probability larger than $1 - 1/p^{\Box}$,

$$\|\hat{\beta} - \beta^*\|_1 \lesssim \frac{s\lambda}{\kappa(s)}.$$

Estimation under REC via the LASSO - [BRT,07]

If REC(s) holds and $\|\beta^*\|_0 = s$ then with probability larger than $1 - 1/p^{\square}$,

$$\|\hat{\beta} - \beta^*\|_1 \lesssim \frac{s\lambda}{\kappa(s)}.$$

and

$$\|\mathcal{X}(\hat{\beta} - \beta^*)\|_2 \lesssim \frac{\sigma^2 s \log p}{\kappa^2(s)}.$$

L. & Mendelson

$X_1, \ldots, X_n$ be n iid $\sim X = (x_1, \ldots, x_p)^\top$

1. $\|x_j\|_{L_2} = 1$ and $\|x_j\|_{L_q} \le \kappa_0 q^\eta$ for some $q = \kappa_1 \log(wp)$.

REC under weak moment assumption

L. & Mendelson

$X_1, \ldots, X_n$ be n iid $\sim X = (x_1, \ldots, x_p)^\top$

1. $\|x_j\|_{L_2} = 1$ and $\|x_j\|_{L_q} \leq \kappa_0 q^\eta$ for some $q = \kappa_1 \log(wp)$.
2. $\exists u_0, \beta_0$ such that : $\forall \|t\|_0 \leq s, P\big[|\langle X, t\rangle| \geq u_0\|t\|_2\big] \geq \beta_0$

L. & Mendelson

$X_1, \dots, X_n$ be n iid $\sim X = (x_1, \dots, x_p)^\top$

1. $\|x_j\|_{L_2} = 1$ and $\|x_j\|_{L_q} \leq \kappa_0 q^\eta$ for some $q = \kappa_1 \log(wp)$.
2. $\exists u_0, \beta_0$ such that : $\forall \|t\|_0 \leq s, P\left[|\langle X, t\rangle| \geq u_0 \|t\|_2\right] \geq \beta_0$
3. $n \gtrsim \max\left(s \log(ep/s), \log^{(2\eta-1)\vee 1}(wp)\right)$.

REC under weak moment assumption

L. & Mendelson

$X_1, \ldots, X_n$ be n iid $\sim X = (x_1, \ldots, x_p)^\top$

1. $\|x_j\|_{L_2} = 1$ and $\|x_j\|_{L_q} \leq \kappa_0 q^\eta$ for some $q = \kappa_1 \log(wp)$.
2. $\exists u_0, \beta_0$ such that : $\forall \|t\|_0 \leq s, P\big[|\langle X, t\rangle| \geq u_0\|t\|_2\big] \geq \beta_0$
3. $n \gtrsim \max\big(s\log(ep/s), \log^{(2\eta-1)\vee 1}(wp)\big)$.

Then, with probability at least $1 - 2\exp(-c_1 n\beta_0^2) - 1/(w^{\kappa_1} p^{\kappa_1 - 1})$,

$$\Gamma = \frac{1}{\sqrt{n}} \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \text{ satisfies } REC(c_1 s).$$

L. & Mendelson

$X_1, \dots, X_n$ be n iid $\sim X = (x_1, \dots, x_p)^\top$

1. $\|x_j\|_{L_2} = 1$ and $\|x_j\|_{L_q} \le \kappa_0 q^\eta$ for some $q = \kappa_1 \log(wp)$.
2. $\exists u_0, \beta_0$ such that : $\forall \|t\|_0 \le s, P\left[|\langle X, t\rangle| \ge u_0\|t\|_2\right] \ge \beta_0$
3. $n \gtrsim \max\left(s\log(ep/s), \log^{(2\eta-1)\vee 1}(wp)\right)$.

Then, with probability at least $1 - 2\exp(-c_1 n\beta_0^2) - 1/(w^{\kappa_1}p^{\kappa_1 - 1})$,

$$\Gamma = \frac{1}{\sqrt{n}}\begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \text{ satisfies } REC(c_1 s).$$

1. $\log p$ moments is almost necessary

L. & Mendelson

$X_1, \dots, X_n$ be n iid $\sim X = (x_1, \dots, x_p)^\top$

1. $\|x_j\|_{L_2} = 1$ and $\|x_j\|_{L_q} \leq \kappa_0 q^\eta$ for some $q = \kappa_1 \log(wp)$.
2. $\exists u_0, \beta_0$ such that : $\forall \|t\|_0 \leq s, P\big[|\langle X, t\rangle| \geq u_0\|t\|_2\big] \geq \beta_0$
3. $n \gtrsim \max\big(s\log(ep/s), \log^{(2\eta-1)\vee 1}(wp)\big)$.

Then, with probability at least $1 - 2\exp(-c_1 n\beta_0^2) - 1/(w^{\kappa_1}p^{\kappa_1 - 1})$,

$$\Gamma = \frac{1}{\sqrt{n}}\begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \text{ satisfies } REC(c_1 s).$$

1. $\log p$ moments is almost necessary
2. the same is true for the Compatibility Condition of S. van de Geer

L. & Mendelson

$X_1, \dots, X_n$ be n iid $\sim X = (x_1, \dots, x_p)^\top$

1. $\|x_j\|_{L_2} = 1$ and $\|x_j\|_{L_q} \leq \kappa_0 q^\eta$ for some $q = \kappa_1 \log(wp)$.
2. $\exists u_0, \beta_0$ such that : $\forall \|t\|_0 \leq s, P\left[|\langle X, t\rangle| \geq u_0 \|t\|_2\right] \geq \beta_0$
3. $n \gtrsim \max\left(s \log(ep/s), \log^{(2\eta-1)\vee 1}(wp)\right)$.

Then, with probability at least $1 - 2\exp(-c_1 n\beta_0^2) - 1/(w^{\kappa_1} p^{\kappa_1 - 1})$,

$$\Gamma = \frac{1}{\sqrt{n}} \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \text{ satisfies } REC(c_1 s).$$

1. $\log p$ moments is almost necessary
2. the same is true for the Compatibility Condition of S. van de Geer
3. the same is true for normalized measurement matrices.

Problem

Given a collection $\{C_j : j \in [p]\}$ of 3-element subsets of $[n]$, does there exists a partition of $[n]$ by elements C_j ?

(This problem is NP-complete = NP and NP-hard)

recasting basis pursuit to a linear program

Basis pursuit

$$x^\star \hookrightarrow \text{minimize }_{t\in\mathbb{R}^p} \|t\|_1 \text{ subject to } \Gamma t = \Gamma x.$$

is equivalent to the

linear program

$$\left((\mathbf{z}^+)^\star, (\mathbf{z}^-)^\star\right) \hookrightarrow \text{minimize }_{\mathbf{z}^+,\mathbf{z}^-\in\mathbb{R}^p} \sum_{j=1}^{N} (z_j^+ + z_j^-)$$

$$\text{subject to } \left[\Gamma \middle| -\Gamma\right] \begin{bmatrix} \mathbf{z}^+ \\ \mathbf{z}^- \end{bmatrix} = \Gamma x, \begin{bmatrix} \mathbf{z}^+ \\ \mathbf{z}^- \end{bmatrix} \geq 0.$$

$$x^\star = (\mathbf{z}^+)^\star - (\mathbf{z}^-)^\star$$

Definition

A centrally symmetric polytope $P \subset \mathbb{R}^n$ is said s-neighborly if every set of s vertices, containing no antipodal pair, is the set of all vertices of some faces of P.

Example : ΓB_1^p is s-neighborly when : $\forall I \subset [p], |I| \leq s, (\epsilon_i)_{i \in I} \in \{\pm 1\}^I$,

$$\mathrm{aff}(\{\epsilon_i X_i : i \in I\}) \cap \mathrm{conv}(\{\theta_j X_j, j \notin I, \theta_j \in \{\pm 1\}\}) = \emptyset$$

1. Paley-Zygmund : if $\|Z\|_{2+\epsilon} \leq \kappa \|Z\|_2$,

$$P\big[|Z| \geq (1/2)\|Z\|_2\big] \geq \left[\frac{3\|Z\|_2^2}{4\|Z\|_{2+\epsilon}^2}\right]^{\frac{2+\epsilon}{\epsilon}} \geq \left[\frac{3}{4\kappa}\right]^{\frac{2+\epsilon}{\epsilon}}.$$

2. Einmahl-Mason : if $Z \geq 0$ then for $t > 0$,

$$P[Z \leq \mathbb{E}Z - t\|Z\|_2] \leq \exp(-ct^2).$$

So if $\|Z\|_2 \leq \kappa \|Z\|_1$,

$$P[|Z| \geq (1-t)\|Z\|_2] \geq 1 - \exp(-ct^2).$$

Study the small ball probability function

$$\phi(\epsilon) = P[\|X\|_2 \leq \epsilon] \text{ when } \epsilon \to 0.$$

[Kashin, 1975], [Alon, Goldreich, Hastad, Peralta, 1992], [Devore, 2007], [Nelson, Temlyakov, 2010] $n \gtrsim s^2$

[Kashin, 1975], [Alon, Goldreich, Hastad, Peralta, 1992], [Devore, 2007], [Nelson, Temlyakov, 2010] $n \gtrsim s^2$

[Bourgain, Dilworth, Ford, Konyagin, Kutzarova, 2011] $n \gtrsim s^{2-\epsilon_0}$

[Kashin, 1975], [Alon, Goldreich, Hastad, Peralta, 1992], [Devore, 2007], [Nelson, Temlyakov, 2010] $n \gtrsim s^2$

[Bourgain, Dilworth, Ford, Konyagin, Kutzarova, 2011] $n \gtrsim s^{2-\epsilon_0}$

Still far from the number of mesurements that can be obtained by random matrices : $n \gtrsim s \log(ep/s)$.

ℓ_0-minimization is NP-hard and BP is solved by linear programing but...

ℓ_0-minimization is NP-hard and BP is solved by linear programing but...

- more measurements : $n \geq 2s$ (for ℓ_0) and $n \gtrsim s\log(ep/s)$ (for ℓ_1).

ℓ_0-minimization is NP-hard and BP is solved by linear programing but...

- more measurements : $n \geq 2s$ (for ℓ_0) and $n \gtrsim s \log(ep/s)$ (for ℓ_1).

- deterministic measurements for ℓ_0 (the first $2s$ discrete Fourier measurements) to random measurements for ℓ_1.

Proof 1/4

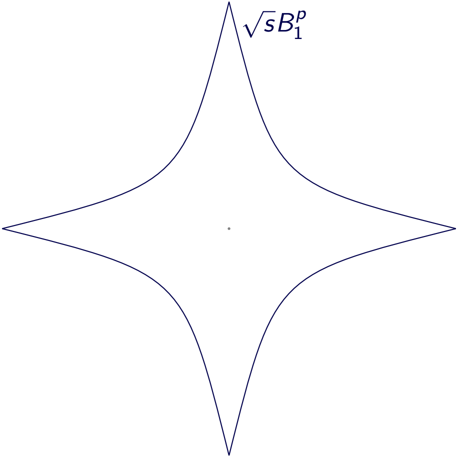
We prove : for any $t \in \sqrt{s}B_1^p \cap S_2^{p-1}$,

$$\|\Gamma t\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\left\langle X_i, t\right\rangle^2 \geq c_0 > 0.$$

Proof 1/4

We prove : for any $t \in \sqrt{s}B_1^p \cap S_2^{p-1}$,

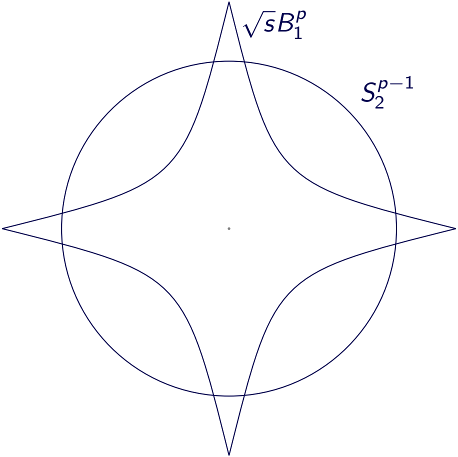
$$\|\Gamma t\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\left\langle X_i, t\right\rangle^2 \geq c_0 > 0.$$



Proof 1/4

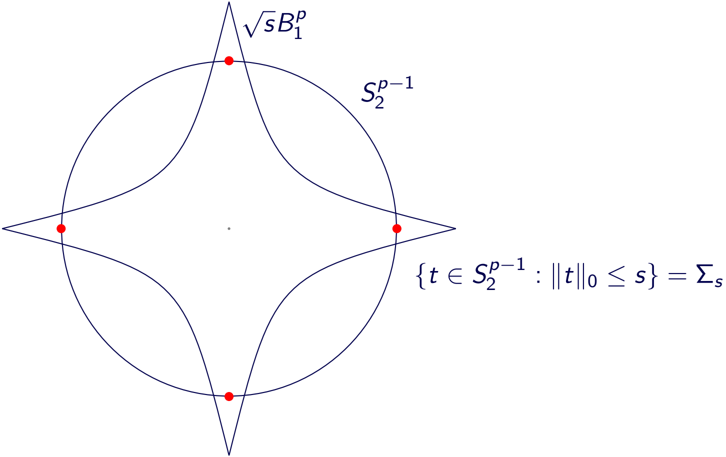
We prove : for any $t \in \sqrt{s}B_1^p \cap S_2^{p-1}$,

$$\|\Gamma t\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\big\langle X_i, t\big\rangle^2 \geq c_0 > 0.$$



We prove : for any $t \in \sqrt{s}B_1^p \cap S_2^{p-1}$,

$$\|\Gamma t\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\big\langle X_i, t\big\rangle^2 \geq c_0 > 0.$$



1. for any $t \in \Sigma_s$:

$$\frac{1}{n}\sum_{i=1}^{n}\langle X_i, t\rangle^2 \geq c_0 > 0.$$

1. for any $t \in \Sigma_s$:
$$\frac{1}{n}\sum_{i=1}^{n}\langle X_i, t\rangle^2 \geq c_0 > 0.$$
(small ball property)

1. for any $t \in \Sigma_s$:
$$\frac{1}{n}\sum_{i=1}^{n}\left\langle X_i, t\right\rangle^2 \geq c_0 > 0.$$
(small ball property)
2. s-sparse vectors to $\sqrt{s}B_1^p \cap S^{p-1}$ via Maurey's representation :
write $x \in \sqrt{s}B_1^p \cap S_2^{p-1}$ as a mean of s-sparse vectors

1. for any $t \in \Sigma_s$:
$$\frac{1}{n}\sum_{i=1}^{n}\left\langle X_i, t\right\rangle^2 \geq c_0 > 0.$$
(small ball property)
2. s-sparse vectors to $\sqrt{s}B_1^p \cap S^{p-1}$ via Maurey's representation :
write $x \in \sqrt{s}B_1^p \cap S_2^{p-1}$ as a mean of s-sparse vectors
(log(p) moments)

small ball assumption : $\forall t \in \Sigma_s, P\Big[|\big\langle X, t\big\rangle| \geq u_0\|t\|_2\Big] \geq \beta_0$

small ball assumption : $\forall t \in \Sigma_s, P\left[|\langle X, t\rangle| \geq u_0\|t\|_2\right] \geq \beta_0$

$$\Downarrow$$

empirical small ball property : w.h.p. $\forall t \in \Sigma_s$,

$$\left|\{i \in \{1, \ldots, n\}, |\langle X_i, t\rangle| \geq u_0\|t\|_2\}\right| \geq \frac{\beta_0 n}{2}.$$

small ball assumption $:\forall t \in \Sigma_s, P\left[|\left\langle X, t\right\rangle| \geq u_0\|t\|_2\right] \geq \beta_0$

$$\Downarrow$$

empirical small ball property : w.h.p.$\forall t \in \Sigma_s$,

$$\left|\{i \in \{1, \dots, n\}, |\left\langle X_i, t\right\rangle| \geq u_0\|t\|_2\}\right| \geq \frac{\beta_0 n}{2}.$$

when $n \gtrsim s \log(ep/s)$.

small ball assumption : $\forall t \in \Sigma_s, P\left[|\left\langle X, t\right\rangle| \geq u_0\|t\|_2\right] \geq \beta_0$

$$\Downarrow$$

empirical small ball property : w.h.p. $\forall t \in \Sigma_s$,

$$\left|\{i \in \{1, \dots, n\}, |\left\langle X_i, t\right\rangle| \geq u_0\|t\|_2\}\right| \geq \frac{\beta_0 n}{2}.$$

when $n \gtrsim s\log(ep/s)$. For any $t \in \Sigma_s$:

$$\frac{1}{n}\sum_{i=1}^{n}\left\langle X_i, t\right\rangle^2 \geq \frac{1}{n}\sum_{i=1}^{n} u_0^2\|t\|_2^2 I\left(|\left\langle X_i, t\right\rangle| \geq u_0\|t\|_2\}\right) \geq u_0^2\|t\|_2^2\frac{\beta_0}{2}.$$

$$\text{small ball assumption} : \forall t \in \Sigma_s, P\left[|\langle X, t\rangle| \geq u_0\|t\|_2\right] \geq \beta_0$$

$$\Downarrow$$

$$\text{empirical small ball property : w.h.p.} \forall t \in \Sigma_s,$$
$$\left|\{i \in \{1, \dots, n\}, |\langle X_i, t\rangle| \geq u_0\|t\|_2\}\right| \geq \frac{\beta_0 n}{2}.$$

when $n \gtrsim s \log(ep/s)$. For any $t \in \Sigma_s$:

$$\frac{1}{n}\sum_{i=1}^{n}\langle X_i, t\rangle^2 \geq \frac{1}{n}\sum_{i=1}^{n} u_0^2\|t\|_2^2 I\left(|\langle X_i, t\rangle| \geq u_0\|t\|_2\}\right) \geq u_0^2\|t\|_2^2\frac{\beta_0}{2}.$$

[Mendelson, Koltchinskii] under moment assumptions.

Proof 3/4 – s-sparse vectors to $\sqrt{s}B_1^p \cap S_2^{p-1}$ via Maurey's method

Proposition

Let $\Gamma : \mathbb{R}^p \mapsto \mathbb{R}^n$ such that

1. for any $t \in \Sigma_s \cap S_2^{p-1}$: $\|\Gamma t\|_2 \geq \kappa_0$,

Proposition

Let $\Gamma : \mathbb{R}^p \mapsto \mathbb{R}^n$ such that

1. for any $t \in \Sigma_s \cap S_2^{p-1}$: $\|\Gamma t\|_2 \geq \kappa_0$,
2. $\|\Gamma e_j\|_2 \leq c_0, \forall 1 \leq j \leq p$ (where $(e_1, \ldots, e_p)$ is the canonical basis)

Proposition

Let $\Gamma : \mathbb{R}^p \mapsto \mathbb{R}^n$ such that

1. for any $t \in \Sigma_s \cap S_2^{p-1}$: $\|\Gamma t\|_2 \geq \kappa_0$,
2. $\|\Gamma e_j\|_2 \leq c_0, \forall 1 \leq j \leq p$ (where $(e_1, \ldots, e_p)$ is the canonical basis)

Then, for any $t \in \sqrt{c_1 s}B_1^p \cap S_2^{p-1}$, $\|\Gamma t\|_2 \geq c_2 > 0$.

Proposition

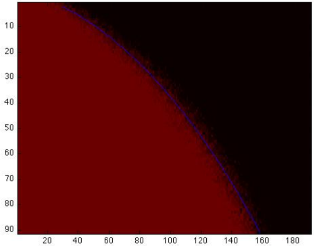
Let $\Gamma : \mathbb{R}^p \mapsto \mathbb{R}^n$ such that

1. for any $t \in \Sigma_s \cap S_2^{p-1}$: $\|\Gamma t\|_2 \geq \kappa_0$,
2. $\|\Gamma e_j\|_2 \leq c_0, \forall 1 \leq j \leq p$ (where $(e_1, \dots, e_p)$ is the canonical basis)

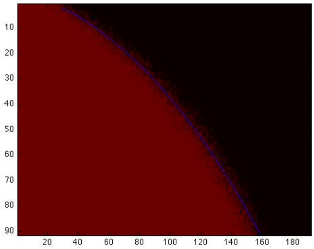
Then, for any $t \in \sqrt{c_1 s}B_1^p \cap S_2^{p-1}$, $\|\Gamma t\|_2 \geq c_2 > 0$.

The uniform control $\max_{1\leq j\leq p} \|\Gamma e_j\|_2 \leq c_0$ costs $\log(p)$ moments.

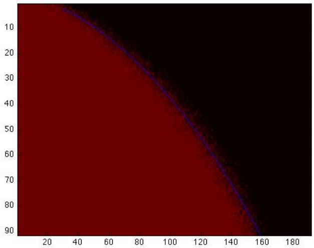
phase transition diagram for Gaussian measurements.



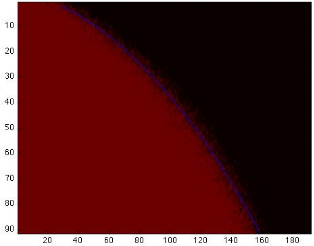
For every (n, s),
n : number of measurements
s : sparsity



For every (n, s),
n : number of measurements
s : sparsity
$\star$ Construct 20 s-sparse vectors $x \in \mathbb{R}^{200}$.



For every (n, s),
n : number of measurements
s : sparsity
$\star$ Construct 20 s-sparse vectors $x \in \mathbb{R}^{200}$.
$\star$ Run Basis Pursuit $\hat{x}_{BP}$ using $\langle X_i, x \rangle, i = 1, \ldots, n$



For every (n, s),

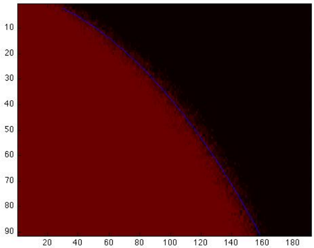
n : number of measurements

s : sparsity

$\star$ Construct 20 s-sparse vectors $x \in \mathbb{R}^{200}$.

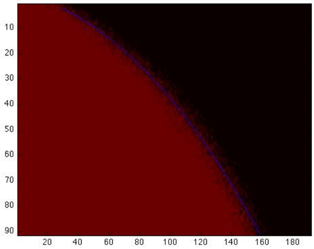
$\star$ Run Basis Pursuit $\hat{x}_{BP}$ using $\langle X_i, x \rangle, i = 1, \ldots, n$

$\star$ Check if $\|x - \hat{x}_{BP}\|_2 \leq 0.01$.



For every (n, s),
n : number of measurements
s : sparsity
$\star$ Construct 20 s-sparse vectors $x \in \mathbb{R}^{200}$.
$\star$ Run Basis Pursuit $\hat{x}_{BP}$ using $\langle X_i, x \rangle, i = 1, \ldots, n$
$\star$ Check if $\|x - \hat{x}_{BP}\|_2 \leq 0.01$.

1. black pixel = 20 "exact" recovery (0 mistakes)



For every (n, s),

n : number of measurements

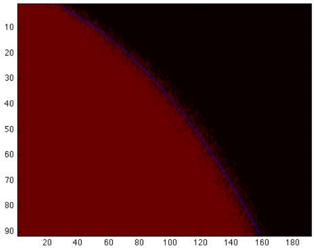
s : sparsity

$\star$ Construct 20 s-sparse vectors $x \in \mathbb{R}^{200}$.

$\star$ Run Basis Pursuit $\hat{x}_{BP}$ using $\langle X_i, x \rangle, i = 1, \ldots, n$

$\star$ Check if $\|x - \hat{x}_{BP}\|_2 \leq 0.01$.

1. black pixel = 20 “exact” recovery (0 mistakes)
2. red pixel = 0 exact recovery (20 mistakes).



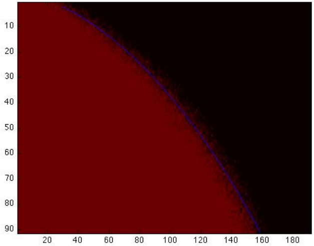
For every (n, s),
n : number of measurements
s : sparsity
$\star$ Construct 20 s-sparse vectors $x \in \mathbb{R}^{200}$.
$\star$ Run Basis Pursuit $\hat{x}_{BP}$ using $\langle X_i, x \rangle, i = 1, \ldots, n$
$\star$ Check if $\|x - \hat{x}_{BP}\|_2 \leq 0.01$.

1. black pixel = 20 "exact" recovery (0 mistakes)
2. red pixel = 0 exact recovery (20 mistakes).

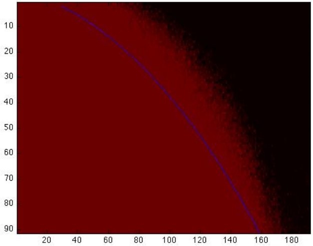
Theoretical phase transition $n \sim s \log(ep/s)$.

phase transition diagram for Gaussian and Cauchy measurements.

Gaussian measurements

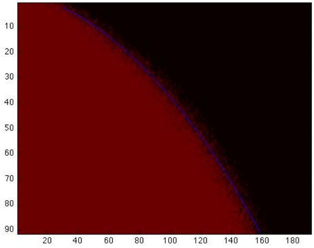


Cauchy measurements

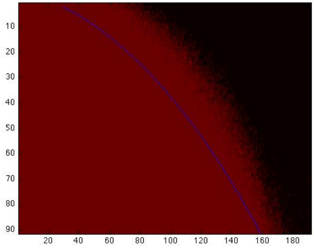


phase transition diagram for Gaussian and Cauchy measurements.

Gaussian measurements



Cauchy measurements



$\log(ep/s)$ moments may be necessary (?)

Smallest singular value of a random matrix

$X_1, \ldots, X_n$ iid vectors in $\mathbb{R}^p$ such that $n \gtrsim p$,

$X_1, \ldots, X_n$ iid vectors in $\mathbb{R}^p$ such that $n \gtrsim p$, we have

$$\inf_{\|t\|_2=1} \frac{1}{n} \sum_{i=1}^{n} \left\langle X_i, t \right\rangle^2 \geq c_0 > 0$$

proportional case ($s = n \gtrsim p$) = lower bound on the smallest singular value

$X_1, \dots, X_n$ iid vectors in $\mathbb{R}^p$ such that $n \gtrsim p$, we have

$$\inf_{\|t\|_2=1} \frac{1}{n}\sum_{i=1}^n \left\langle X_i, t\right\rangle^2 \geq c_0 > 0$$

1. in expectation : [Srivastava, Vershynin, 2012] X is isotropic and

$$\sup_{\|t\|_2=1} \mathbb{E}\left|\left\langle X, t\right\rangle\right|^{2+\epsilon} \leq c_1.$$

proportional case ($s = n \gtrsim p$) = lower bound on the smallest singular value

$X_1, \dots, X_n$ iid vectors in $\mathbb{R}^p$ such that $n \gtrsim p$, we have

$$\inf_{\|t\|_2=1} \frac{1}{n}\sum_{i=1}^n \left\langle X_i, t\right\rangle^2 \geq c_0 > 0$$

1. in expectation : [Srivastava, Vershynin, 2012] X is isotropic and
$$\sup_{\|t\|_2=1} \mathbb{E}\left|\left\langle X, t\right\rangle\right|^{2+\epsilon} \leq c_1.$$

2. with probability larger than $1-\exp(-c_2 n)$ in [Koltchinksii, Mendelson] when X is isotropic and for every $t \in \mathbb{R}^p$,
$$\|\left\langle t, X\right\rangle\|_{L_2} \leq c_3 \|\left\langle t, X\right\rangle\|_{L_1}.$$

proportional case ($s = n \gtrsim p$) = lower bound on the smallest singular value

$X_1, \ldots, X_n$ iid vectors in $\mathbb{R}^p$ such that $n \gtrsim p$, we have

$$\inf_{\|t\|_2=1} \frac{1}{n} \sum_{i=1}^{n} \big\langle X_i, t \big\rangle^2 \geq c_0 > 0$$

1. in expectation : [Srivastava, Vershynin, 2012] X is isotropic and

$$\sup_{\|t\|_2=1} \mathbb{E}\big|\big\langle X, t \big\rangle\big|^{2+\epsilon} \leq c_1.$$

2. with probability larger than $1 - \exp(-c_2 n)$ in [Koltchinksii, Mendelson] when X is isotropic and for every $t \in \mathbb{R}^p$,

$$\|\big\langle t, X \big\rangle\|_{L_2} \leq c_3 \|\big\langle t, X \big\rangle\|_{L_1}.$$

3. with probability larger than $1 - \exp(-c_2 n)$ in [L., Mendelson] when for every $t \in \mathbb{R}^p$,

$$P\big[|\big\langle t, X \big\rangle| \geq u_0 \|t\|_2\big] \geq \beta_0.$$

proportional case ($s = n \gtrsim p$) = lower bound on the smallest singular value

$X_1, \dots, X_n$ iid vectors in $\mathbb{R}^p$ such that $n \gtrsim p$, we have

$$\inf_{\|t\|_2=1} \frac{1}{n} \sum_{i=1}^{n} \left\langle X_i, t \right\rangle^2 \geq c_0 > 0$$

1. in expectation : [Srivastava, Vershynin, 2012] X is isotropic and

$$\sup_{\|t\|_2=1} \mathbb{E} \left| \left\langle X, t \right\rangle \right|^{2+\epsilon} \leq c_1.$$

2. with probability larger than $1 - \exp(-c_2 n)$ in [Koltchinksii, Mendelson] when X is isotropic and for every $t \in \mathbb{R}^p$,

$$\left\| \left\langle t, X \right\rangle \right\|_{L_2} \leq c_3 \left\| \left\langle t, X \right\rangle \right\|_{L_1}.$$

3. with probability larger than $1 - \exp(-c_2 n)$ in [L., Mendelson] when for every $t \in \mathbb{R}^p$,

$$P\left[\left| \left\langle t, X \right\rangle \right| \geq u_0 \|t\|_2 \right] \geq \beta_0.$$

$\Rightarrow$ Lower bound on the smallest singular value has nothing to do with concentration (true for Cauchy matrices).