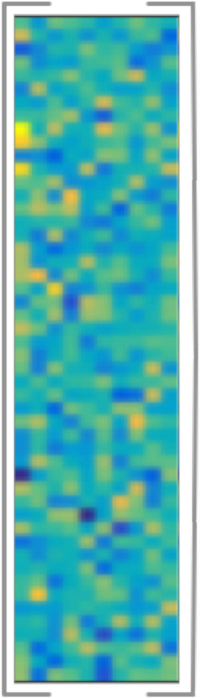
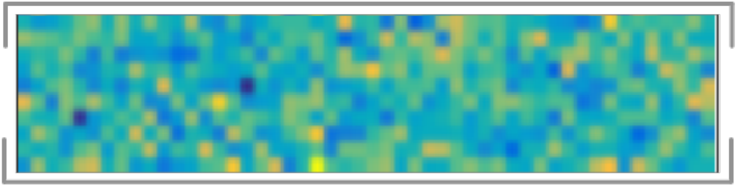




Mathematical Fundamentals of Compressive Sensing: Random matrices and ℓ_1-recovery

Justin Romberg, Georgia Tech ECE

NMI, IISc, Bangalore, India

February 20, 2015

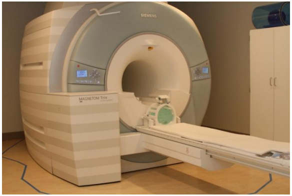
Agenda

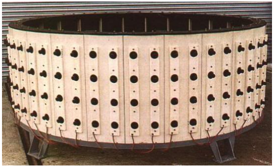
Today: Mathematical foundations of compressive sensing
Random embeddings and recovery using ℓ_1

Saturday: Low rank recovery and bilinear problems in signal processing

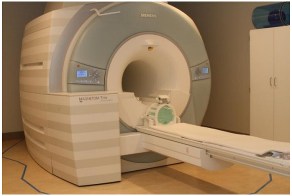
Sunday: Dynamic recovery, subspace matching and CS on the continuum

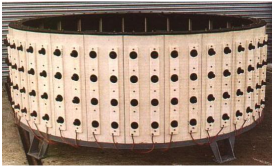
Linear systems of equations are ubiquitous





Linear systems of equations are ubiquitous





All of these can be abstracted to

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

Linear systems of equations are ubiquitous

Model:

$$\begin{bmatrix} \\ \boldsymbol{y} \\ \\ \\ \end{bmatrix} = \begin{bmatrix} & & \\ & \boldsymbol{A} & \\ & & \\ & & \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \\ \end{bmatrix}$$

$\boldsymbol{y}$: data coming off of sensor
$\boldsymbol{A}$: mathematical (linear) model for sensor
$\boldsymbol{x}$: signal/image to reconstruct

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix $\boldsymbol{A}$ with $M \geq N$ (MORE observations than unknowns), through which we observe

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \text{noise}$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix $\boldsymbol{A}$ with $M \geq N$ (MORE observations than unknowns), through which we observe

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \quad \Leftrightarrow \quad \hat{\boldsymbol{x}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix $\boldsymbol{A}$ with $M \geq N$ (MORE observations than unknowns), through which we observe

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \quad \Leftrightarrow \quad \hat{\boldsymbol{x}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix $\boldsymbol{A}$ with $M \geq N$ (MORE observations than unknowns), through which we observe

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \quad \Leftrightarrow \quad \hat{\boldsymbol{x}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$$

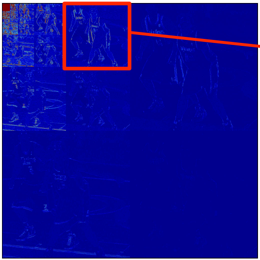
- Q: When is this recovery stable? That is, when is

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix $\boldsymbol{A}$ preserves *distances* ...

$$\|\boldsymbol{A}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2^2 \approx \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 \quad \text{for all } \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^N$$

Sparsity



wavelet transform

zoom

Wavelet approximation

Take 1% of *largest* coefficients, set the rest to zero (adaptive)

original

approximated

rel. error = 0.031

When can we stably recover an S-sparse vector?

$$\begin{bmatrix} \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} & \boldsymbol{\Phi} & \end{bmatrix} \begin{bmatrix} \boldsymbol{x_0} \\ \\ \\ \\ \end{bmatrix}$$

- Now we have an underdetermined $M \times N$ system $\boldsymbol{\Phi}$ (FEWER measurements than unknowns), and observe

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}_0 + \text{noise}$$

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system $\mathbf{\Phi}$ (FEWER measurements than unknowns), and observe

$$\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}_0 + \text{noise}$$

- We can recover $\boldsymbol{x}_0$ when $\mathbf{\Phi}$ *keeps sparse signals separated*

$$\|\mathbf{\Phi}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2^2 \;\approx\; \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2$$

  for all S-sparse $\boldsymbol{x}_1, \boldsymbol{x}_2$

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system $\mathbf{\Phi}$ (FEWER measurements than unknowns), and observe

$$\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}_0 + \text{noise}$$

- We can recover $\boldsymbol{x}_0$ when $\mathbf{\Phi}$ *keeps sparse signals separated*

$$\|\mathbf{\Phi}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2^2 \;\approx\; \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2$$

  for all S-sparse $\boldsymbol{x}_1, \boldsymbol{x}_2$

- To recover $\boldsymbol{x}_0$, we might solve

$$\min_{\boldsymbol{x}} \; \#\text{NonZeroTerms}(\boldsymbol{x}) \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{x} \approx \boldsymbol{y}$$

- This program is *computationally intractable*

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system $\mathbf{\Phi}$ (FEWER measurements than unknowns), and observe

$$\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}_0 + \text{noise}$$

- We can recover $\boldsymbol{x}_0$ when $\mathbf{\Phi}$ *keeps sparse signals separated*

$$\|\mathbf{\Phi}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2^2 \;\approx\; \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2$$

  for all S-sparse $\boldsymbol{x}_1, \boldsymbol{x}_2$

- A relaxed (convex) program

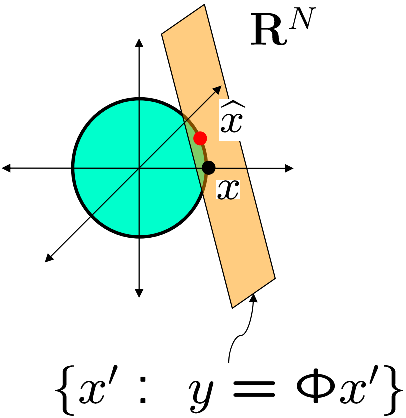
$$\min_{\boldsymbol{x}} \; \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{x} \approx \boldsymbol{y}$$

  $\|\boldsymbol{x}\|_1 = \sum_k |x_k|$

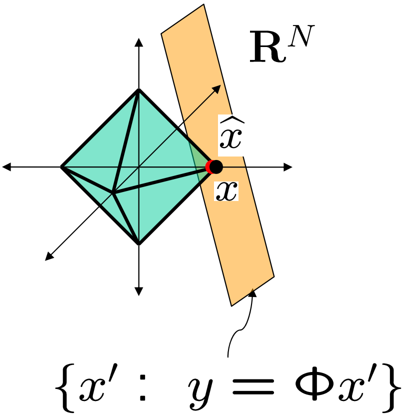
- This program is very tractable (linear program)
- The convex program can recover nearly all "identifiable" sparse vectors, and it is *robust*.

Intuition for ℓ_1

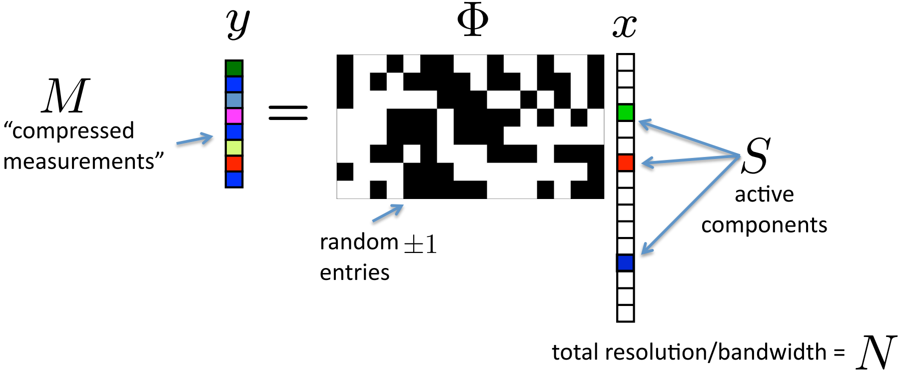
$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_2$ s.t. $\boldsymbol{\Phi x} = \boldsymbol{y}$



$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1$ s.t. $\boldsymbol{\Phi x} = \boldsymbol{y}$



What kind of matrices keep sparse signals separated?



- *Random matrices* are provably efficient
- We can recover S-sparse x from

$$M \gtrsim S \cdot \log(N/S)$$

measurements

Agenda

We will prove (almost from top to bottom) two things:

- That an $M \times N$ iid Gaussian random matrix satisfies

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \ \le \ \|\boldsymbol{\Phi x}\|_2^2 \ \le \ (1+\delta)\|\boldsymbol{x}\|_2^2 \quad \forall\ 2S\text{-sparse } \boldsymbol{x} \tag{1}$$

  with (extraordinarily) high probability when

$$M \ \ge \ \text{Const} \cdot S \log(N/S)$$

Agenda

We will prove (almost from top to bottom) two things:

- That an $M \times N$ iid Gaussian random matrix satisfies

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \;\leq\; \|\boldsymbol{\Phi x}\|_2^2 \;\leq\; (1+\delta)\|\boldsymbol{x}\|_2^2 \quad \forall\ 2S\text{-sparse } \boldsymbol{x} \tag{1}$$

  with (extraordinarily) high probability when

$$M \;\geq\; \mathrm{Const}\cdot S\log(N/S)$$

- Suppose an $M \times N$ matrix $\boldsymbol{\Phi}$ obeys (1). Let $\boldsymbol{x}_0$ be an S-sparse vector, and suppose we observe $\boldsymbol{y} = \boldsymbol{\Phi x}_0$. Given $\boldsymbol{y}$, the solution to

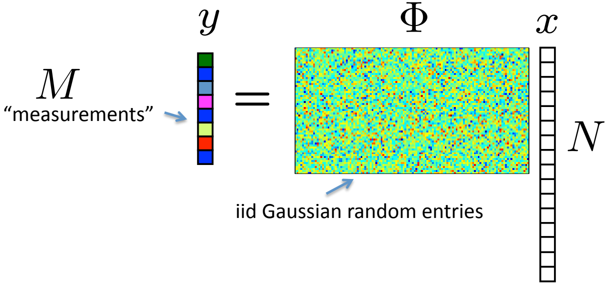
$$\min_{\boldsymbol{x}} \; \|\boldsymbol{x}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi x} = \boldsymbol{y}$$

  is *exactly* $\boldsymbol{x}_0$.

Restricted Isometries for Gaussian Matrices

Gaussian random matrices

- Each entry of $\mathbf{\Phi}$ is iid $\mathrm{Normal}(0, M^{-1})$

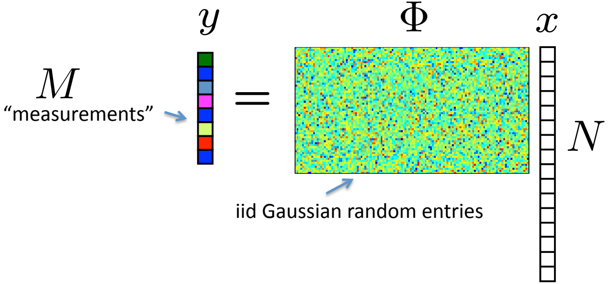


- For *any fixed* $\boldsymbol{x} \in \mathbb{R}^N$, each measurement is

$$y_m \sim \mathrm{Normal}(0, \|\boldsymbol{x}\|_2^2/M)$$

Gaussian random matrices

- Each entry of $\boldsymbol{\Phi}$ is iid $\mathrm{Normal}(0, M^{-1})$



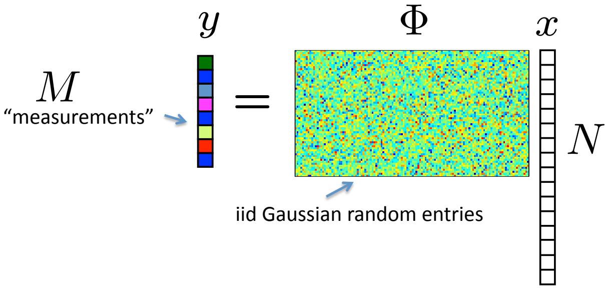
- For *any fixed* $\boldsymbol{x} \in \mathbb{R}^N$, we have

$$\mathrm{E}[\|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2] = \|\boldsymbol{x}\|_2^2$$

the mean of the measurement energy is exactly $\|\boldsymbol{x}\|_2^2$

Gaussian random matrices

- Each entry of $\mathbf{\Phi}$ is iid $\mathrm{Normal}(0, M^{-1})$

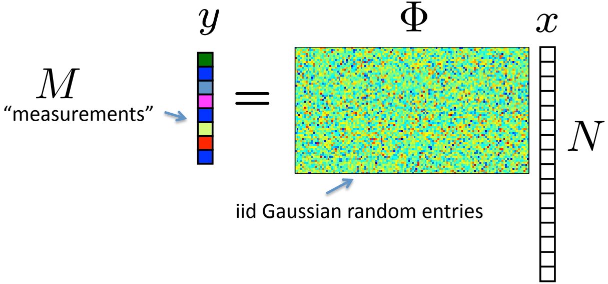


- For *any fixed* $\boldsymbol{x} \in \mathbb{R}^N$, we have

$$\mathrm{P}\left(\left|\|\mathbf{\Phi}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| < \delta\|\boldsymbol{x}\|_2^2\right) \;\geq\; 1 - 2e^{-M\delta^2/8}$$

Gaussian random matrices

- Each entry of $\mathbf{\Phi}$ is iid $\mathrm{Normal}(0, M^{-1})$



- For *all $2S$-sparse* $\boldsymbol{x} \in \mathbb{R}^N$, we have

$$\mathrm{P}\left(\max_{\boldsymbol{x}} \left| \|\mathbf{\Phi}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| < \delta \|\boldsymbol{x}\|_2^2 \right) \;\geq\; 1 - 2e^{c_1 S \log(N/S)} e^{-c_2 M \delta^2}$$

So we can make this probability close to 1 by taking

$$M \;\geq\; \mathrm{Const} \cdot S \log(N/S)$$

Random projection of a fixed vector

For Gaussian random $\mathbf{\Phi}$ operating on a *fixed* $\boldsymbol{x} \in \mathbb{R}^N$

$$\|\mathbf{\Phi}\boldsymbol{x}\|_2^2 \approx \|\boldsymbol{x}\|_2^2$$

Theorem: Let $\mathbf{\Phi}$ be an $M \times N$ matrix whose entries are iid Gaussian

$$\Phi_{i,j} \sim \mathrm{Normal}(0, 1/M).$$

Then

$$\mathrm{E}\,\|\Phi\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2,$$

as, with $\boldsymbol{v} = \mathbf{\Phi}\boldsymbol{x}$,

$$\mathrm{E}\left[\sum_{m=1}^{M} v_m^2\right] = \sum_{m=1}^{M} \mathrm{E}[v_m^2] = \sum_{m=1}^{M} \frac{1}{M}\|\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2,$$

since $v_m = \langle \boldsymbol{x}, \boldsymbol{\phi}_m \rangle \sim \mathrm{Normal}(0, M^{-1}\|\boldsymbol{x}\|_2^2)$

Random projection of a fixed vector

For Gaussian random $\mathbf{\Phi}$ operating on a *fixed* $\boldsymbol{x} \in \mathbb{R}^N$

$$\|\mathbf{\Phi}\boldsymbol{x}\|_2^2 \approx \|\boldsymbol{x}\|_2^2$$

Theorem: Let $\mathbf{\Phi}$ be an $M \times N$ matrix whose entries are iid Gaussian

$$\Phi_{i,j} \sim \mathrm{Normal}(0, 1/M).$$

Then

$$\mathrm{E}\,\|\mathbf{\Phi}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2,$$

and for any $0 < \delta \leq 1$

$$\begin{aligned}\mathrm{P}\left(\left|\|\mathbf{\Phi}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) &\leq 2\exp\left(-\frac{(\delta^2-\delta^3)M}{4}\right) \\ &\leq 2\exp\left(-\delta^2 M/8\right)\end{aligned}$$

for $\delta \leq 1/2$.

The Markov inequality

Let Y be a positive random variable. Then for all $t > 0$

$$\mathrm{P}\left(Y \geq t\right) \;\leq\; \frac{\mathrm{E}[Y]}{t}$$

The Markov inequality

Let Y be a positive random variable. Then for all $t > 0$

$$\mathrm{P}\left(Y \geq t\right) \leq \frac{\mathrm{E}[Y]}{t}$$

Proof:

$$\begin{aligned}
\mathrm{E}[Y] &= \int_0^\infty y\, f_Y(y)\ dy \\
&\geq \int_t^\infty y\, f_Y(y)\ dy \\
&\geq t \int_t^\infty f_Y(y)\ dy \\
&= t\,\mathrm{P}\left(Y \geq t\right).
\end{aligned}$$

The Markov inequality

Let Y be a positive random variable. Then for all $t > 0$

$$\mathrm{P}\left(Y \geq t\right) \;\leq\; \frac{\mathrm{E}[Y]}{t}$$

Also:

$$\mathrm{P}\left(Y^2 \geq t^2\right) \leq \frac{\mathrm{E}[Y^2]}{t^2}$$

$$\mathrm{P}\left(Y^3 \geq t^3\right) \leq \frac{\mathrm{E}[Y^3]}{t^3}$$

$$\mathrm{P}\left(e^{\lambda Y} \geq e^{\lambda t}\right) \leq \frac{\mathrm{E}[e^{\lambda Y}]}{e^{\lambda t}} \qquad \lambda > 0$$

$$\vdots$$

$$\mathrm{P}\left(\phi(Y) \geq \phi(t)\right) \leq \frac{\mathrm{E}[\phi(Y)]}{\phi(t)}$$

for any strictly monotonic $\phi(\cdot)$.

The Markov inequality

Let Y be a positive random variable. Then for all $t > 0$

$$\mathrm{P}\left(Y \geq t\right) \;\leq\; \frac{\mathrm{E}[Y]}{t}$$

Chernoff-type bound:

$$\mathrm{P}\left(Y \geq t\right) \;\leq\; \frac{\mathrm{E}[e^{\lambda Y}]}{e^{\lambda t}} \quad \text{for any} \quad \lambda > 0.$$

A first upper concentration bound ...

For $\boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{x}$, $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1 + \delta\right) \leq \frac{\mathrm{E}[e^{\lambda\|\boldsymbol{v}\|_2^2}]}{e^{\lambda(1+\delta)}}$$

A first upper concentration bound ...

For $\boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{x}$, $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\begin{aligned} \mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1 + \delta\right) &\leq \frac{\mathrm{E}[e^{\lambda\|\boldsymbol{v}\|_2^2}]}{e^{\lambda(1+\delta)}} \\ &= \frac{\mathrm{E}[e^{\lambda\sum_m v_m^2}]}{e^{\lambda(1+\delta)}} \end{aligned}$$

A first upper concentration bound ...

For $\boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{x}$, $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1+\delta\right) \leq \frac{\mathrm{E}[e^{\lambda\|\boldsymbol{v}\|_2^2}]}{e^{\lambda(1+\delta)}}$$

$$= \frac{\mathrm{E}[e^{\lambda\sum_m v_m^2}]}{e^{\lambda(1+\delta)}}$$

$$= \frac{\mathrm{E}[e^{\lambda v_1^2} e^{\lambda v_2^2} \cdots e^{\lambda v_M^2}]}{e^{\lambda(1+\delta)}}$$

A first upper concentration bound ...

For $\boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{x}$, $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\begin{aligned} \mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1 + \delta\right) &\leq \frac{\mathrm{E}[e^{\lambda \|\boldsymbol{v}\|_2^2}]}{e^{\lambda(1+\delta)}} \\ &= \frac{\mathrm{E}[e^{\lambda \sum_m v_m^2}]}{e^{\lambda(1+\delta)}} \\ &= \frac{\mathrm{E}[e^{\lambda v_1^2} e^{\lambda v_2^2} \cdots e^{\lambda v_M^2}]}{e^{\lambda(1+\delta)}} \\ &= \frac{\mathrm{E}[e^{\lambda v_1^2}]\,\mathrm{E}[e^{\lambda v_2^2}] \cdots \mathrm{E}[e^{\lambda v_M^2}]}{e^{\lambda(1+\delta)}} \end{aligned}$$

A first upper concentration bound ...

For $\boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{x}$, $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\begin{aligned}
\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1+\delta\right) &\le \frac{\mathrm{E}[e^{\lambda\|\boldsymbol{v}\|_2^2}]}{e^{\lambda(1+\delta)}} \\
&= \frac{\mathrm{E}[e^{\lambda\sum_m v_m^2}]}{e^{\lambda(1+\delta)}} \\
&= \frac{\mathrm{E}[e^{\lambda v_1^2} e^{\lambda v_2^2} \cdots e^{\lambda v_M^2}]}{e^{\lambda(1+\delta)}} \\
&= \frac{\mathrm{E}[e^{\lambda v_1^2}]\,\mathrm{E}[e^{\lambda v_2^2}] \cdots \mathrm{E}[e^{\lambda v_M^2}]}{e^{\lambda(1+\delta)}} \\
&= \frac{(\mathrm{E}[e^{\lambda v_1^2}])^M}{e^{\lambda(1+\delta)}} \quad \text{(since } v_m \text{ i.i.d.)}
\end{aligned}$$

A first upper concentration bound ...

For $\boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{x}$, $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1 + \delta\right) \;\leq\; \frac{(\mathrm{E}[e^{\lambda v_1^2}])^M}{e^{\lambda(1+\delta)}}, \qquad v_1 \sim \mathrm{Normal}(0, M^{-1})$$

A first upper concentration bound ...

For $\boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{x}$, $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1 + \delta\right) \;\leq\; \frac{(\mathrm{E}[e^{\lambda v_1^2}])^M}{e^{\lambda(1+\delta)}}, \qquad v_1 \sim \mathrm{Normal}(0, M^{-1})$$

It is known that

$$\mathrm{E}[e^{\lambda v_1^2}] \;=\; \frac{1}{\sqrt{1 - 2\lambda/M}} \qquad \text{for } \lambda < M/2.$$

A first upper concentration bound ...

For $\boldsymbol{v} = \boldsymbol{\Phi x}$, $\|\boldsymbol{x}\|_2 = 1$, we have that

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1 + \delta\right) \;\leq\; \frac{(\mathrm{E}[e^{\lambda v_1^2}])^M}{e^{\lambda(1+\delta)}}, \qquad v_1 \sim \mathrm{Normal}(0, M^{-1})$$

And so

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1 + \delta\right) \;\leq\; \left(\frac{e^{-2\lambda(1+\delta)/M}}{1 - 2\lambda/M}\right)^{M/2} \qquad \forall\ \lambda < M/2$$

A first upper concentration bound ...

We have

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1+\delta\right) \quad \leq \quad \left(\frac{e^{-2\lambda(1+\delta)/M}}{1-2\lambda/M}\right)^{M/2} \qquad \forall\ \lambda < M/2$$

Choose

$$\lambda = \frac{M\delta}{2(1+\delta)}$$

(easy to see that in this case $\lambda < M/2$).

A first upper concentration bound ...

We have

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1+\delta\right) \;\leq\; \left(\frac{e^{-2\lambda(1+\delta)/M}}{1-2\lambda/M}\right)^{M/2} \quad \forall\ \lambda < M/2$$

Choose

$$\lambda = \frac{M\delta}{2(1+\delta)}$$

(easy to see that in this case $\lambda < M/2$).

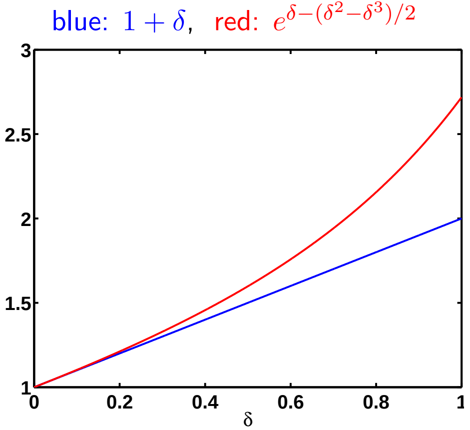
And so

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1+\delta\right) \;\leq\; \left((1+\delta)e^{-\delta}\right)^{M/2}.$$

The upper concentration bound

We have

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1+\delta\right) \;\leq\; \left((1+\delta)e^{-\delta}\right)^{M/2}.$$



The upper concentration bound

We have

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1+\delta\right) \;\; \leq \;\; \Big((1+\delta)e^{-\delta}\Big)^{M/2}.$$

and so

$$\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 > 1+\delta\right) \;\; \leq \;\; e^{-(\delta^2-\delta^3)M/4}$$

The lower concentration bound

The lower bound follows the exact same sequence of steps:

$$\begin{aligned}
\mathrm{P}\left(\|\boldsymbol{v}\|_2^2 < 1-\delta\right) &\leq \left(\frac{e^{2(1-\delta)\lambda/M}}{1+2\lambda/M}\right)^{M/2} \\
&\leq \left((1-\delta)e^{\delta}\right)^{M/2} \quad \text{by taking} \quad \lambda = \frac{M\delta}{2(1-\delta)} \\
&\leq e^{-(\delta^2-\delta^3)M/4}
\end{aligned}$$

The Johnson-Lindenstrauss Lemma

We have shown that for any *fixed* $\boldsymbol{x} \in \mathbb{R}^N$

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \;\le\; \|\boldsymbol{\Phi x}\|_2^2 \;\le\; (1+\delta)\|\boldsymbol{x}\|_2^2$$

with probability exceeding $1 - 2e^{-\delta^2 M/8}$.

A simple application of the union bound means that for any set of K vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_K$, the above holds with probability exceeding $1 - 2Ke^{-\delta^2 M/8}$...

The Johnson-Lindenstrauss Lemma

Theorem: (J&L, 1984): Let $\mathcal{Q}$ be a arbitrary set of Q vectors in $\mathbb{R}^N$, and let $\mathbf{\Phi}$ be an $M \times N$ random linear mapping. Then

$$(1-\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2 \;\le\; \|\Phi(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2^2 \;\le\; (1+\delta)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2$$

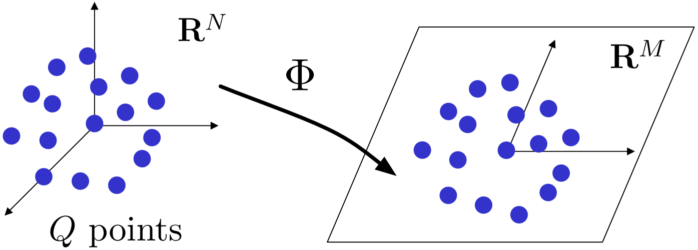
for *all* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{Q}$ with

$$\mathrm{P}\left(\text{Failure}\right) \;\le\; Q^2 e^{-\delta^2 M/8} \;\le\; \epsilon$$

when

$$M \;\ge\; \frac{8}{\delta^2}\left[\log(Q) + \log\left(\frac{1}{\epsilon}\right) + 0.7\right]$$

The Johnson-Lindenstrauss Lemma



$\mathbf{\Phi}$ embeds to precision δ with probability ϵ when

$$M \;\geq\; \frac{8}{\delta^2}\left[2\log(Q) + \log\left(\frac{1}{\epsilon}\right) + 0.7\right]$$

Concentration bound

We have: For any *fixed* $\boldsymbol{x} \in \mathbb{R}^N$

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \;\leq\; \|\boldsymbol{\Phi x}\|_2^2 \;\leq\; (1+\delta)\|\boldsymbol{x}\|_2^2$$

with probability exceeding $1 - 2e^{-\delta^2 M/8}$.

We want: this for *all* $2S$-sparse $\boldsymbol{x}$ simultaneously...

A single $2S$-dimensional subspace

Theorem: Let V be a $2S$-dimensional subspace of $\mathbb{R}^N$. Then

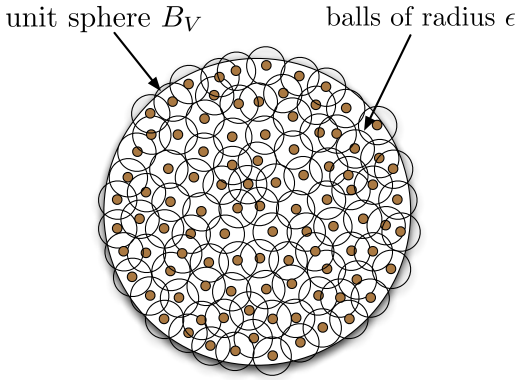
$$\mathrm{P}\left(\sup_{\boldsymbol{x}\in V} \left|\|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) \;\leq\; 2 \cdot 9^{2S} \cdot e^{-\delta^2 M/32}$$

As before, it is enough to prove this for

$$\boldsymbol{x} \in B_V = \{\boldsymbol{x} \in V : \|\boldsymbol{x}\|_2 = 1\}$$

Covering the sphere

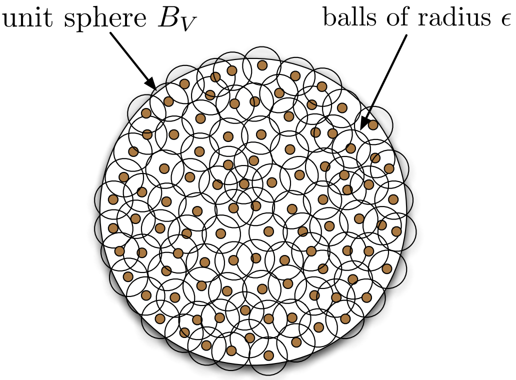
An ϵ-net for B_V:



for every $\boldsymbol{x} \in B_V$, there is a $\boldsymbol{y} \in \text{Net}$ such that $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \epsilon$

$N(B_V, \epsilon)$ is the *size of the smallest ϵ-net*

Covering the sphere



It is a fact that

$$N(B_V, \epsilon) \;\leq\; \left(1 + \frac{2}{\epsilon}\right)^{2S}$$

From discrete to continuous

Lemma: Fix $0 \leq \epsilon < 1/2$, and let $\mathcal{N}_\epsilon$ be the minimal ϵ-net for B_V. Then

$$\sup_{\boldsymbol{x} \in B_V} \left| \|\boldsymbol{\Phi x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| \;\leq\; \frac{1}{1-2\epsilon} \max_{\boldsymbol{y} \in \mathcal{N}_\epsilon} \left| \|\boldsymbol{\Phi x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right|$$

A single $2S$-dimensional subspace

Theorem: Let V be a $2S$-dimensional subspace of $\mathbb{R}^N$. Then

$$\mathrm{P}\left(\sup_{\boldsymbol{x}\in V}\left|\|\boldsymbol{\Phi x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) \;\leq\; 2\cdot 9^{2S}\cdot e^{-\delta^2 M/32}$$

where the constant $1/32 = 1/4\cdot 1/8$ ($1/8$ is from the previous theorem).

A single $2S$-dimensional subspace

Theorem: Let V be a $2S$-dimensional subspace of $\mathbb{R}^N$. Then

$$\mathrm{P}\left(\sup_{\boldsymbol{x}\in V}\left|\|\boldsymbol{\Phi x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) \;\leq\; 2\cdot 9^{2S}\cdot e^{-\delta^2 M/32}$$

where the constant $1/32 = 1/4 \cdot 1/8$ ($1/8$ is from the previous theorem).

So $\boldsymbol{\Phi}$ is "well-conditioned" on V when

$$M \;\geq\; \mathrm{Const}\cdot S$$

A single $2S$-dimensional subspace

Theorem: Let V be a $2S$-dimensional subspace of $\mathbb{R}^N$. Then

$$\mathrm{P}\left(\sup_{\boldsymbol{x}\in V}\left|\|\boldsymbol{\Phi x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) \;\leq\; 2\cdot 9^{2S}\cdot e^{-\delta^2 M/32}$$

where the constant $1/32 = 1/4\cdot 1/8$ ($1/8$ is from the previous theorem).

We want this for *all subspaces* in which $2S$-sparse signals live...

A single $2S$-dimensional subspace

Theorem: Let V be a $2S$-dimensional subspace of $\mathbb{R}^N$. Then

$$\mathrm{P}\left(\sup_{\boldsymbol{x}\in V}\left|\|\boldsymbol{\Phi x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) \;\leq\; 2\cdot 9^{2S}\cdot e^{-\delta^2 M/32}$$

where the constant $1/32 = 1/4\cdot 1/8$ ($1/8$ is from the previous theorem).

We want this for *all subspaces* in which $2S$-sparse signals live...

There are $\binom{N}{2S} \leq \left(\frac{Ne}{2S}\right)^{2S}$ such subspaces...

All $2S$-dimensional subspaces

For $\Gamma \subset \{1, \ldots, N\}$, let

$$B_\Gamma = \left\{ \boldsymbol{x} \in \mathbb{R}^N \ : \ x_\gamma = 0, \ \gamma \notin \Gamma, \ \|\boldsymbol{x}\|_2 = 1 \right\}.$$

Theorem:

$$\mathrm{P}\left(\max_{|\Gamma| \leq 2S} \sup_{\boldsymbol{x} \in B_\Gamma} \left| \|\boldsymbol{\Phi x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| > \delta \right) \ \leq \ 2 \left(\frac{Ne}{2S} \right)^{2S} 9^{2S} \, e^{-\delta^2 M/32}$$

All $2S$-dimensional subspaces

Theorem:

$$\mathrm{P}\left(\sup_{\text{all } 2S \text{ sparse } \boldsymbol{x}} \left|\|\boldsymbol{\Phi x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \delta\right) \leq 2\left(\frac{Ne}{2S}\right)^{2S} 9^{2S}\, e^{-\delta^2 M/32}$$

$$= e^{\log 2 + 2S\log(Ne/2S) + 2S\log 9 - \delta^2 M/32}$$

Which is to say

$$(1-\delta)\|\boldsymbol{x}\|_2^2 \;\leq\; \|\boldsymbol{\Phi x}\|_2^2 \;\leq\; (1+\delta)\|\boldsymbol{x}\|_2^2 \quad \forall\ 2S\text{-sparse } \boldsymbol{x}$$

with high probability when

$$M \;\geq\; \frac{\text{Const}}{\delta^2} \cdot S\log(N/S)$$

SUCCESS!!!

Sparse Recovery using ℓ_1 minimization

Sparse recovery

We will show the following:

Let $\mathbf{\Phi}$ be an $M \times N$ matrix that is an approximate isometry for $3S$-sparse vectors. Let $\boldsymbol{x}_0$ be an S-sparse vector, and suppose we observe $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}_0$. Given $\boldsymbol{y}$, the solution to

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{x} = \boldsymbol{y}$$

is *exactly* $\boldsymbol{x}_0$.

Moving to the solution

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \quad \text{such that} \quad \boldsymbol{\Phi x} = \boldsymbol{y}$$

Call the solution to this $\boldsymbol{x}^\sharp$. Set

$$\boldsymbol{h} = \boldsymbol{x}^\sharp - \boldsymbol{x}_0.$$

Moving to the solution

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \quad \text{such that} \quad \boldsymbol{\Phi x} = \boldsymbol{y}$$

Call the solution to this $\boldsymbol{x}^\sharp$. Set

$$\boldsymbol{h} = \boldsymbol{x}^\sharp - \boldsymbol{x}_0.$$

Two things must be true:

- $\boldsymbol{\Phi h} = \boldsymbol{0}$
  Simply because both $\boldsymbol{x}^\sharp$ and $\boldsymbol{x}_0$ are feasible: $\boldsymbol{\Phi x}^\sharp = \boldsymbol{y} = \boldsymbol{\Phi x}_0$
- $\|\boldsymbol{x}_0 + \boldsymbol{h}\|_1 \leq \|\boldsymbol{x}_0\|_1$
  Simply because $\boldsymbol{x}_0 + \boldsymbol{h} = \boldsymbol{x}^\sharp$, and $\|\boldsymbol{x}^\sharp\|_1 \leq \|\boldsymbol{x}_0\|_1$

Moving to the solution

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \quad \text{such that} \quad \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{y}$$

Call the solution to this $\boldsymbol{x}^\sharp$. Set

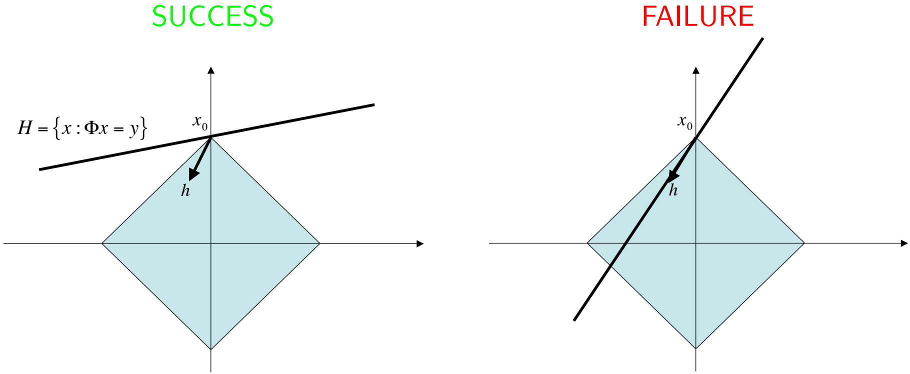
$$\boldsymbol{h} = \boldsymbol{x}^\sharp - \boldsymbol{x}_0.$$

Two things must be true:

- $\boldsymbol{\Phi}\boldsymbol{h} = \boldsymbol{0}$
  Simply because both $\boldsymbol{x}^\sharp$ and $\boldsymbol{x}_0$ are feasible: $\boldsymbol{\Phi}\boldsymbol{x}^\sharp = \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}_0$
- $\|\boldsymbol{x}_0 + \boldsymbol{h}\|_1 \leq \|\boldsymbol{x}_0\|_1$
  Simply because $\boldsymbol{x}_0 + \boldsymbol{h} = \boldsymbol{x}^\sharp$, and $\|\boldsymbol{x}^\sharp\|_1 \leq \|\boldsymbol{x}_0\|_1$

We'll show that if $\boldsymbol{\Phi}$ is $3S$-RIP, then these conditions are *incompatible* unless $\boldsymbol{h} = \boldsymbol{0}$

Geometry



Two things must be true:

- $\boldsymbol{\Phi h = 0}$
- $\|\boldsymbol{x}_0 + \boldsymbol{h}\|_1 \leq \|\boldsymbol{x}_0\|_1$

Cone condition

For $\Gamma \subset \{1, \ldots, N\}$, define $\boldsymbol{h}_\Gamma \in \mathbb{R}^N$ as

$$h_\Gamma(\gamma) = \begin{cases} h(\gamma) & \gamma \in \Gamma \\ 0 & \gamma \notin \Gamma \end{cases}$$

Let Γ_0 be the support of $\boldsymbol{x}_0$. For any "descent vector" $\boldsymbol{h}$, we have

$$\|\boldsymbol{h}_{\Gamma_0^c}\|_1 \;\leq\; \|\boldsymbol{h}_{\Gamma_0}\|_1$$

Cone condition

For $\Gamma \subset \{1, \ldots, N\}$, define $\boldsymbol{h}_\Gamma \in \mathbb{R}^N$ as

$$h_\Gamma(\gamma) = \begin{cases} h(\gamma) & \gamma \in \Gamma \\ 0 & \gamma \notin \Gamma \end{cases}$$

Let Γ_0 be the support of $\boldsymbol{x}_0$. For any "descent vector" $\boldsymbol{h}$, we have

$$\|\boldsymbol{h}_{\Gamma_0^c}\|_1 \;\leq\; \|\boldsymbol{h}_{\Gamma_0}\|_1$$

Why? The triangle inequality..

$$\begin{aligned} \|\boldsymbol{x}_0\|_1 \geq \|\boldsymbol{x}_0 + \boldsymbol{h}\|_1 &= \|\boldsymbol{x}_0 + \boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_0^c}\|_1 \\ &\geq \|\boldsymbol{x}_0 + \boldsymbol{h}_{\Gamma_0^c}\|_1 - \|\boldsymbol{h}_{\Gamma_0}\|_1 \\ &= \|\boldsymbol{x}_0\|_1 + \|\boldsymbol{h}_{\Gamma_0^c}\|_1 - \|\boldsymbol{h}_{\Gamma_0}\|_1 \end{aligned}$$

Cone condition

For $\Gamma \subset \{1, \ldots, N\}$, define $\boldsymbol{h}_\Gamma \in \mathbb{R}^N$ as

$$h_\Gamma(\gamma) = \begin{cases} h(\gamma) & \gamma \in \Gamma \\ 0 & \gamma \notin \Gamma \end{cases}$$

Let Γ_0 be the support of $\boldsymbol{x}_0$. For any "descent vector" $\boldsymbol{h}$, we have

$$\|\boldsymbol{h}_{\Gamma_0^c}\|_1 \;\leq\; \|\boldsymbol{h}_{\Gamma_0}\|_1$$

We will show that if $\boldsymbol{\Phi}$ is $3S$-RIP, then

$$\boldsymbol{\Phi h} = \boldsymbol{0} \quad \Rightarrow \quad \|\boldsymbol{h}_{\Gamma_0}\|_1 \leq \rho \|\boldsymbol{h}_{\Gamma_0^c}\|_1$$

for some $\rho < 1$, and so $\boldsymbol{h} = \boldsymbol{0}$.

Some basic facts about ℓ_p norms

- $\|\boldsymbol{h}_\Gamma\|_\infty \leq \|\boldsymbol{h}_\Gamma\|_2 \leq \|\boldsymbol{h}_\Gamma\|_1$

- $\|\boldsymbol{h}_\Gamma\|_1 \leq \sqrt{|\Gamma|} \cdot \|\boldsymbol{h}_\Gamma\|_2$

- $\|\boldsymbol{h}_\Gamma\|_2 \leq \sqrt{|\Gamma|} \cdot \|\boldsymbol{h}_\Gamma\|_\infty$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned} \Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ \Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ &\vdots \end{aligned}$$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned}
\Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\
\Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\
&\vdots
\end{aligned}$$

Then

$$0 = \|\boldsymbol{\Phi}\boldsymbol{h}\|_2 = \|\boldsymbol{\Phi}(\sum_{j\geq 1} \boldsymbol{h}_{\Gamma_j})\|_2 \geq \|\boldsymbol{\Phi}(\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1})\|_2 - \|\sum_{j\geq 2} \boldsymbol{\Phi}\boldsymbol{h}_{\Gamma_j}\|_2$$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned}\Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c},\\ \Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c},\\ &\vdots\end{aligned}$$

Then

$$\begin{aligned}0 = \|\boldsymbol{\Phi}\boldsymbol{h}\|_2 = \|\boldsymbol{\Phi}(\sum_{j\geq 1}\boldsymbol{h}_{\Gamma_j})\|_2 &\geq \|\boldsymbol{\Phi}(\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1})\|_2 - \|\sum_{j\geq 2}\boldsymbol{\Phi}\boldsymbol{h}_{\Gamma_j}\|_2\\ &\geq \|\boldsymbol{\Phi}(\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1})\|_2 - \sum_{j\geq 2}\|\boldsymbol{\Phi}\boldsymbol{h}_{\Gamma_j}\|_2\end{aligned}$$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned} \Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ \Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ &\vdots \end{aligned}$$

Then

$$\|\boldsymbol{\Phi}(\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1})\|_2 \;\leq\; \sum_{j \geq 2} \|\boldsymbol{\Phi}\boldsymbol{h}_{\Gamma_j}\|_2$$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned}
\Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\
\Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\
&\vdots
\end{aligned}$$

Applying the $3S$-RIP gives

$$\begin{aligned}
\sqrt{1-\delta_{3S}}\,\|\boldsymbol{h}_{\Gamma_0}+\boldsymbol{h}_{\Gamma_1}\|_2 &\le \|\boldsymbol{\Phi}(\boldsymbol{h}_{\Gamma_0}+\boldsymbol{h}_{\Gamma_1})\|_2 \\
&\le \sum_{j\ge 2}\|\boldsymbol{\Phi}\boldsymbol{h}_{\Gamma_j}\|_2 \;\le\; \sum_{j\ge 2}\sqrt{1+\delta_{2S}}\|\boldsymbol{h}_{\Gamma_j}\|_2
\end{aligned}$$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned} \Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ \Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ &\vdots \end{aligned}$$

Applying the $3S$-RIP gives

$$\|\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1}\|_2 \ \le \ \sqrt{\frac{1+\delta_{2S}}{1-\delta_{3S}}} \sum_{j\ge 2} \|\boldsymbol{h}_{\Gamma_j}\|_2$$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned}\Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c},\\ \Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c},\\ &\vdots\end{aligned}$$

Then

$$\|\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1}\|_2 \;\leq\; \sqrt{\frac{1+\delta_{2S}}{1-\delta_{3S}}} \sum_{j\geq 2} \sqrt{2S}\|\boldsymbol{h}_{\Gamma_j}\|_\infty$$

since $\|\boldsymbol{h}_{\Gamma_j}\|_2 \leq \sqrt{2S}\|\boldsymbol{h}_{\Gamma_j}\|_\infty$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned} \Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ \Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ &\vdots \end{aligned}$$

Then

$$\|\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1}\|_2 \ \le \ \sqrt{\frac{1+\delta_{2S}}{1-\delta_{3S}}} \sum_{j\ge 1} \frac{1}{\sqrt{2S}} \|\boldsymbol{h}_{\Gamma_j}\|_1$$

since $\|\boldsymbol{h}_{\Gamma_j}\|_\infty \le \frac{1}{2S}\|\boldsymbol{h}_{\Gamma_{j-1}}\|_1$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned} \Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ \Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ &\vdots \end{aligned}$$

Which means

$$\|\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1}\|_2 \;\le\; \sqrt{\frac{1+\delta_{2S}}{1-\delta_{3S}}}\,\frac{\|\boldsymbol{h}_{\Gamma_0^c}\|_1}{\sqrt{2S}}$$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned}
\Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\
\Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\
&\vdots
\end{aligned}$$

Working to the left

$$\|\boldsymbol{h}_{\Gamma_0}\|_2 \;\le\; \|\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1}\|_2 \;\le\; \sqrt{\frac{1+\delta_{2S}}{1-\delta_{3S}}}\,\frac{\|\boldsymbol{h}_{\Gamma_0^c}\|_1}{\sqrt{2S}}$$

Dividing up $\boldsymbol{h}_{\Gamma_0^c}$

Recall that Γ_0 is the support of $\boldsymbol{x}_0$

Fix $\boldsymbol{h} \in \mathrm{Null}(\boldsymbol{\Phi})$. Let

$$\begin{aligned} \Gamma_1 &= \text{locations of } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ \Gamma_2 &= \text{locations next } 2S \text{ largest terms in } \boldsymbol{h}_{\Gamma_0^c}, \\ &\vdots \end{aligned}$$

Working to the left

$$\frac{\|\boldsymbol{h}_{\Gamma_0}\|_1}{\sqrt{S}} \;\leq\; \|\boldsymbol{h}_{\Gamma_0}\|_2 \;\leq\; \|\boldsymbol{h}_{\Gamma_0} + \boldsymbol{h}_{\Gamma_1}\|_2 \;\leq\; \sqrt{\frac{1+\delta_{2S}}{1-\delta_{3S}}}\,\frac{\|\boldsymbol{h}_{\Gamma_0^c}\|_1}{\sqrt{2S}}$$

Wrapping it up

We have shown

$$\|\boldsymbol{h}_{\Gamma_0}\|_1 \leq \sqrt{\frac{1+\delta_{2S}}{1-\delta_{3S}}}\sqrt{\frac{S}{2S}}\|\boldsymbol{h}_{\Gamma_0^c}\|_1$$

$$= \rho\|\boldsymbol{h}_{\Gamma_0^c}\|_1$$

for

$$\rho = \sqrt{\frac{1+\delta_{2S}}{2(1-\delta_{3S})}}$$

Wrapping it up

We have shown

$$\|\boldsymbol{h}_{\Gamma_0}\|_1 \le \sqrt{\frac{1+\delta_{2S}}{1-\delta_{3S}}}\sqrt{\frac{S}{2S}}\|\boldsymbol{h}_{\Gamma_0^c}\|_1$$

$$= \rho\|\boldsymbol{h}_{\Gamma_0^c}\|_1$$

for

$$\rho = \sqrt{\frac{1+\delta_{2S}}{2(1-\delta_{3S})}}$$

Taking $\delta_{2S} \le \delta_{3S} < 1/3 \quad \Rightarrow \quad \rho < 1$.

SUCCESS!!

Theorem: Let $\mathbf{\Phi}$ be an $M \times N$ matrix that is an approximate isometry for $3S$-sparse vectors. Let $\boldsymbol{x}_0$ be an S-sparse vector, and suppose we observe $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}_0$. Given $\boldsymbol{y}$, the solution to

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{x} = \boldsymbol{y}$$

is *exactly* $\boldsymbol{x}_0$.

Other fundamental results

Iterative methods for sparse recovery

There are other iterative methods that have similar recovery guarantees:

- orthogonal matching pursuit (Tropp, Zhang, Foucart, and others)
- iterative hard thresholding (Blumensath, Davies)
- "iterative model selection" CoSAMP, etc. (Tropp, Needell, others)

Deterministic matrices

- Coherence bounds: can recover S-sparse vector from

$$S \lesssim \frac{1}{\mu}, \quad \mu = \text{max inner product between columns}$$

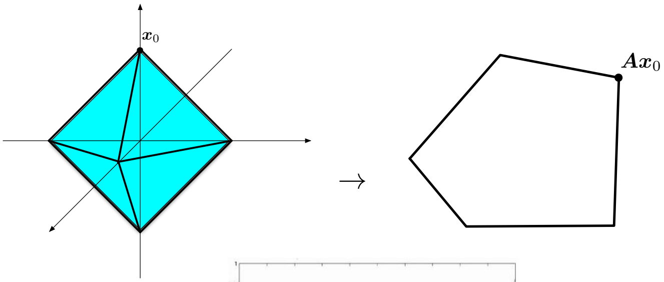
Donoho, Huo, Elad, Bruckstein, Nielson, Gribonval, ...

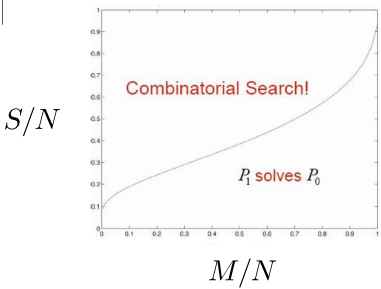
- Connections to channel coding:
  Specially constructed matrices coupled with specialized "decoding" algorithms can yield similar performance guarantees
  (Tarokh and collaborators on low-density frames)

- Other deterministic constructions based on Vandermonde and Fourier matrices yield weaker (but easily verifiable) bonds

Phase transitions for Gaussian + ℓ_1

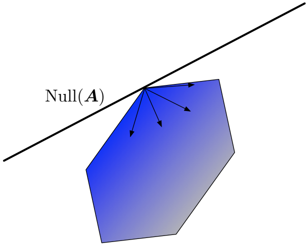
Donoho and Tanner get *sharp* results by looking at properties of projected polytopes:





Sharp upper bounds for Gaussian + ℓ_1

Chandrasekaran, Parrilo, Recht, and Wilsky get a sharp upper bound by estimating the *Gaussian width* of the descent cone



$$M \geq \omega(\mathcal{T}(\boldsymbol{x}))^2, \quad \mathcal{T}(\boldsymbol{x}) = \text{descent cone from } \boldsymbol{x}$$

$$\omega(\mathcal{X}) = \mathrm{E}[\sup_{\boldsymbol{v} \in \mathcal{X} \cap S^{N-1}} \langle \boldsymbol{g}, \boldsymbol{v} \rangle], \quad \boldsymbol{g} \sim \mathrm{Normal}(\mathbf{0}, \mathbf{I})$$

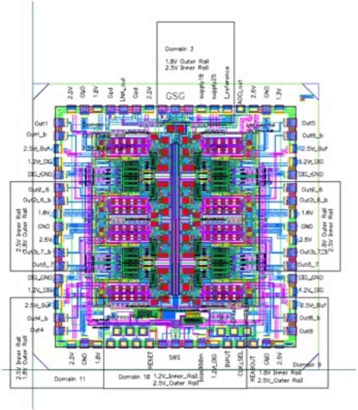
For ℓ_1 problem, $\boldsymbol{x}_0$ S-sparse,

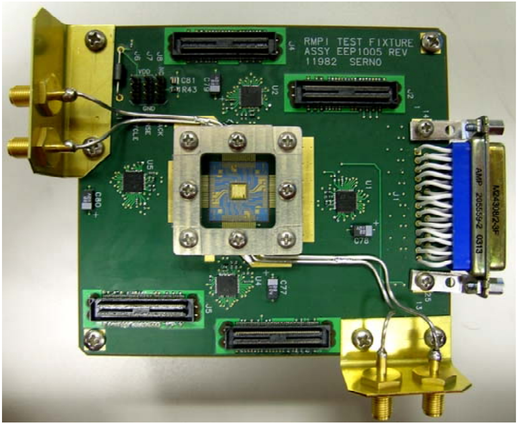
$$\omega(\mathcal{T}(\boldsymbol{x}_0))^2 \leq 2S \log((N-S)/S + 1)$$

Applications of random projections: Hyperspectral imaging

256 frequency bands, 10s of megapixels, 30 frames per second ...

Applications of random projections: Coded ADCs



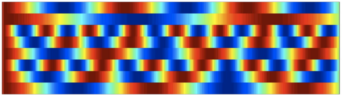


Multichannel ADC/receiver for identifying radar pulses

Covers ~ 3 GHz with ~ 400 MHz sampling rate

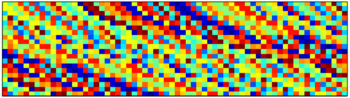
Matrices with structured randomness for sparse recovery

- Subsampled rows of “incoherent” orthogonal matrix



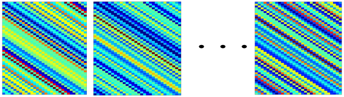
applications: MRI, channel estimation, radar, ...

- Random convolution + subsampling



applications: computed imaging, radar, sonar, ...

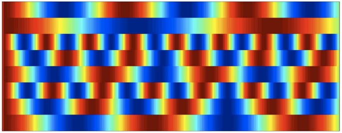
- Multi-toeplitz matrices



applications: MIMO channel estimation, fast forward modeling, ...

Compressive sensing with structured randomness

Subsampled rows of "incoherent" orthogonal matrix



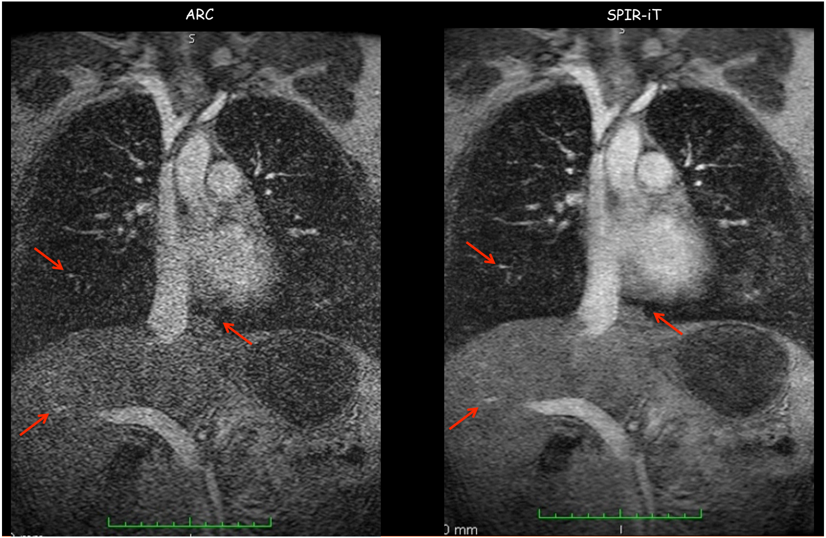
Can recover S-sparse $\boldsymbol{x}_0$ with

$$M \gtrsim S \log^a N$$

measurements

Candes, R, Tao, Rudelson, Vershynin, Tropp, ...

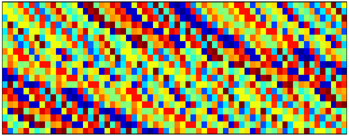
Accelerated MRI



(Lustig et al. '08)

Matrices for sparse recovery with structured randomness

Random convolution + subsampling



Universal; Can recover S-sparse $\boldsymbol{x}_0$ with

$$M \gtrsim S \log^a N$$

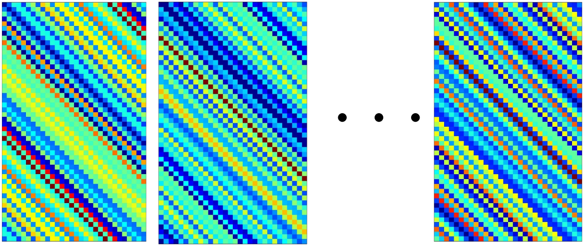
Applications include:

- radar imaging
- sonar imaging
- seismic exploration
- channel estimation for communications
- super-resolved imaging

R, Bajwa, Haupt, Tropp, Rauhut, ...

Matrices for sparse recovery with structured randomness

Multi-toeplitz:



Can recover S-sparse $\boldsymbol{x}_0$ with

$$M \gtrsim S \log^a N$$

R, Neelamani

Application: simultaneous activation

- Run a single simulation with all of the sources activated simultaneously with random waveforms
- The channel responses interfere with one another, but the randomness "codes" them in such a way that they can be separated later