ELEG 867 - Compressive Sensing and Sparse Signal Representations

Gonzalo R. Arce
Depart. of Electrical and Computer Engineering
University of Delaware

Fall 2011

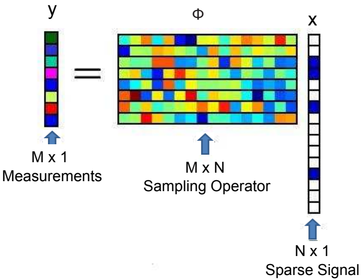
Outline

- Compressive Measurements
- Incoherent Orthobasis
- Restricted Isometry Property (RIP)
- Sparse Signal Recovery

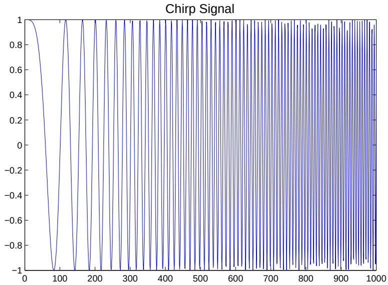
Compressive Measurements

- Measurements in CS are different than samples taken in traditional A/D converters.
- The compressed measurements are given by $y = \Phi x$.
- The signal x is acquired as a series of non-adaptive inner products of different waveforms $\{\phi_1, \phi_2, \ldots, \phi_M\}$

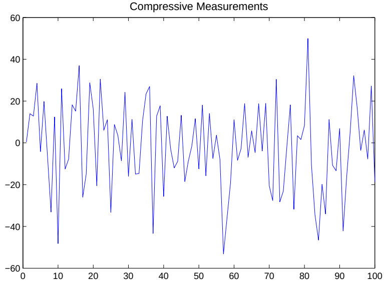
$$y_k = <\phi_k, x>; \quad k = 1, \ldots, M; \quad \text{with} \quad M \ll N$$



Example of Compressive Measurements:

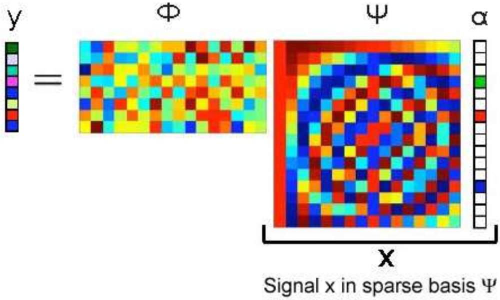


Sparse signal in the Time-Frequency basis.



Compressive Measurements.

- Random measurements can be used for signals sparse in any basis.

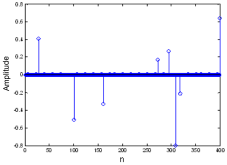


$$y_k =< \phi_k, x >; \quad k = 1, ..., M; \quad \text{with} \quad M \ll N$$

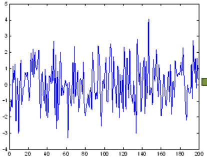
- Need to solve an under determined system of equations $y = \Phi x$.
- Infinitely solutions for the system since $M \ll N$.
- A sparse solution x is recovered from y by solving the following inverse problem

$$(P0): \quad \min_x \|x\|_{\ell_0} \quad s.t. \quad y = \Phi x \tag{1}$$

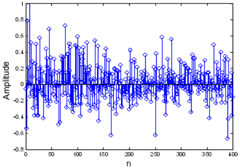
Example of the recovery of an under determined system of equations:



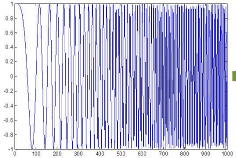
Original sparse signal



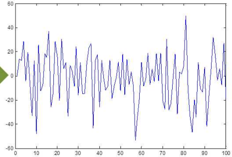
Compressed measurements



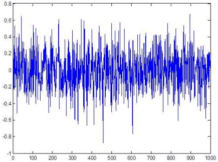
Reconstructed signal using least-squares.
Solution not sparse



Original sparse signal
in Time-Frequency basis



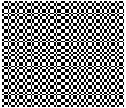
Compressed Measurements

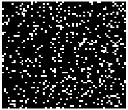


Reconstructed signal using
least-squares

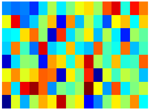
- Sparsity is what makes it possible to recover a signal from undersampled data.
- The number of measurements we need for successful reconstruction depends on the nature of the waveforms ϕ_k, and S

1. Incoherent Orthobasis


Ordered Hadamard Ensemble

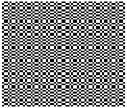

Scrambled Block Hadamard Ensemble

2. Random waveforms ϕ_k

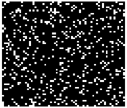

Gaussian Random Ensemble

- Sparsity is what makes it possible to recover a signal from undersampled data.
- The number of measurements we need for successful reconstruction depends on the nature of the waveforms ϕ_k, and S

1. Incoherent Orthobasis

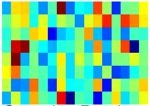


Ordered Hadamard Ensemble



Scrambled Block Hadamard Ensemble

2. Random waveforms ϕ_k

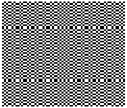


Gaussian Random Ensemble

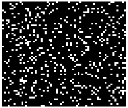
Recoverability

- Sparsity is what makes it possible to recover a signal from undersampled data.
- The number of measurements we need for successful reconstruction depends on the nature of the waveforms ϕ_k, and S

1. Incoherent Orthobasis

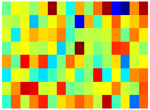


Ordered Hadamard Ensemble



Scrambled Block Hadamard Ensemble

2. Random waveforms ϕ_k



Gaussian Random Ensemble

Incoherent Orthobasis Example

Example of incoherent basis: the ”spike” basis (identity) and the Fourier basis.

Consider the case where the dictionary is the union of two orthobasis:

- I: the ”spike” basis (identity).
- F: the Fourier basis (sinusoids).

$$\Phi = [I; F]$$

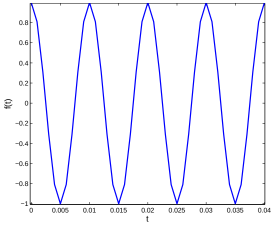
where I is a $N \times N$ matrix and F is a $N \times N$ matrix with

$$f_{m,\ell} = \frac{1}{\sqrt{(N)}} e^{j2\pi(m-1)(\ell-1)/N}$$

Incoherent Orthobasis Example

Note that:

- It takes N spikes to build up a single sinusoid.
- It takes N sinusoids to build a single spike.



Then, if f is a sinusoidal, there are two ways to decompose the signal with the given dictionary:

- $$f = \Phi\alpha = [I; F]\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline 1 \\ \vdots \\ 0 \end{bmatrix}$$
- $$f = \Phi\alpha = [I; F]\begin{bmatrix} x_1 \\ \vdots \\ x_N \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix}$$

But only one is sparse.

Incoherent Orthobasis (Uncertainty Principle)

Previous example of incoherent basis can be generalized using the **Uncertainty Principle**.

Theorem: Uncertainty Principle

Let $f \in \mathbb{R}^N$ be a discrete signal, and let $\hat{f} \in \mathbb{R}^N$ be its discrete Fourier Transform, then

$$|\text{supp}(f)| \cdot |\text{supp}(\hat{f})| \geq N \tag{2}$$

where $\text{supp}(f)$ is the support of f.

- Uncertainty Principle implies that f and $\hat{f}$ cannot both be highly concentrated or be sparse.

Proof of (2):

Let T be a subset of the time domain and let Ω be a subset of the frequency domain. Let $\Phi_{T\Omega} = [I_T; F_\Omega]$.

Let $x = \begin{bmatrix} f \\ -\hat{f} \end{bmatrix}$, where f is supported on T (Time domain), and $\hat{f} = F^* f$ is supported on Ω (frequency domain), then

$$\Phi x = \Phi \begin{bmatrix} f \\ -\hat{f} \end{bmatrix} = \Phi_{T\Omega} \begin{bmatrix} f_T \\ -\hat{f}_\Omega \end{bmatrix} = I_T f_T - F_\Omega \hat{f}_\Omega = I_T f_T - F_\Omega F_\Omega^* f_T = 0 \quad (3)$$

This is true, since $f_T \in \mathbb{R}^{|T|}$ and $\hat{f}_\Omega \in \mathbb{R}^{|\Omega|}$, throwing away all the columns of Φ that multiplies components in the vector that are zero.

If $(\Phi_{T\Omega}^H \Phi_{T\Omega})$ is positive definite, then for any vector $x = \begin{bmatrix} f \\ -\hat{f} \end{bmatrix} \neq 0$ the following is true:

$$x^H \Phi_{T\Omega}^H \Phi_{T\Omega} x = (\Phi_{T\Omega} x)^H (\Phi_{T\Omega} x) > 0.$$

This means that $\Phi_{T\Omega} x$ is either > 0, or < 0, but $\Phi_{T\Omega} x$ can not be $= 0$, for $x \neq 0$.

Remark

However, from (3), we know that there exists a matrix $\Phi_{T\Omega}$ such that $\Phi_{T\Omega} x = 0$, for $x \neq 0$. This means that $\Phi_{T\Omega}$ is a matrix such that $(\Phi_{T\Omega}^H \Phi_{T\Omega})$ is not positive definite.

We need to find the conditions such that $(\Phi_{T\Omega}^H \Phi_{T\Omega})$ is NOT positive definite.

Assume, first, that $(\Phi_{T\Omega}^H \Phi_{T\Omega})$ is positive definite, then all the eigenvalues of the matrix $\Phi_{T\Omega}^H \Phi_{T\Omega}$ are positive:

$$\lambda_{max}(\Phi_{T\Omega}^H \Phi_{T\Omega}) > \ldots > \lambda_{min}(\Phi_{T\Omega}^H \Phi_{T\Omega}) > 0$$

It is clear that all the eigenvalues are positive, if $\lambda_{min}(\Phi_{T\Omega}^H \Phi_{T\Omega}) > 0$.

The square matrix $(\Phi_{T\Omega}^H \Phi_{T\Omega})$ can be decomposed as follows:

$$\Phi_{T\Omega}^H \Phi_{T\Omega} = \begin{bmatrix} I_T^H \\ F_\Omega^H \end{bmatrix} \begin{bmatrix} I_T & F_\Omega \end{bmatrix} = \begin{bmatrix} I_T^H I_T & I_T^H F_\Omega \\ F_\Omega^H I_T & F_\Omega^H F_\Omega \end{bmatrix} = I + \begin{bmatrix} 0 & M \\ M^H & 0 \end{bmatrix}$$

$$\Phi_{T\Omega}^H \Phi_{T\Omega} = I + G.$$

By properties of the eigenvalues:

$$\lambda_i(I + A) = \lambda_i(I) + \lambda_i(A) = 1 + \lambda_i(A), \tag{4}$$

where I is the identity matrix with all the eigenvalues $= 1$. Thus,

$$\lambda_{min}(\Phi_{T\Omega}^H \Phi_{T\Omega}) = 1 + \lambda_{min}(G) > 0,$$

which means that $\lambda_{min}(\Phi_{T\Omega}^H \Phi_{T\Omega}) > 0$, if

$$\lambda_{min}(G) > -1 \tag{5}$$

Homework: Derive a proof of equation (4).

By eigen decomposition:

$$G = Q\Lambda Q^H, \text{ where } diag(\Lambda) = [\lambda_{max}(G), \ldots, \lambda_{min}(G)],$$

and

$$G^H G = Q\Lambda Q^H Q\Lambda Q^H = Q\Lambda^2 Q^H$$

where,

$$\begin{aligned} diag(\Lambda^2) &= [\lambda_{max}(G^H G), \ldots, \lambda_{min}(G^H G)] \\ &= [\lambda_1^2(G), \lambda_2^2(G), \ldots] \\ &\geq \underline{0} \end{aligned}$$

If $0 < \lambda_{max}(G^H G) < 1$, then all the eigenvalues of G satisfy, from (5):

$$-1 < \lambda_i(G) < 1, \Rightarrow \lambda_{min}(\Phi_{T\Omega}^H \Phi_{T\Omega}) > 0. \quad (6)$$

$$G^H G = \begin{bmatrix} 0 & M \\ M^H & 0 \end{bmatrix} \begin{bmatrix} 0 & M \\ M^H & 0 \end{bmatrix} = \begin{bmatrix} MM^H & 0 \\ 0 & M^H M \end{bmatrix},$$

where,

$$MM^H = I_T^H F_\Omega F_\Omega^H I_T; \quad \text{with size: } |T| \times |T|$$

and,

$$M^H M = F_\Omega^H I_T I_T^H F_\Omega \quad \text{with size: } |\Omega| \times |\Omega|.$$

The eigenvalues

$$\lambda_i(MM^H) = \lambda_i((MM^H)^H) = \lambda_i(M^H M),$$

therefore the eigenvalues of the block diagonal matrix $\lambda_i(G^H G) = \lambda_i(MM^H) = \lambda_i(M^H M)$.

Simple Example:

If $|T| = |\Omega| = N$, then

$$MM^H = I_T^H F_\Omega F_\Omega^H I_T = I_{\{N\times N\}} \text{ and, } M^H M = F_\Omega^H I_T I_T^H F_\Omega = I_{\{N\times N\}}.$$

Therefore,

$$G^H G = \begin{bmatrix} MM^H & 0 \\ 0 & M^H M \end{bmatrix} = I_{\{2N\times 2N\}}$$

Since the eigenvalues of $G^H G$, MM^H and $M^H M$ are equal, then

$$\lambda_{max}(G^H G) = \lambda_{max}(MM^H) = \lambda_{max}(M^H M). \quad (7)$$

We need to derive conditions such that $\lambda_{max}(M^H M) < 1$, and from (6) $\lambda_{max}(G^H G) < 1$, and $\lambda_{min}(\Phi_{T\Omega}^H \Phi_{T\Omega}) > 0$.

$$
\begin{aligned}
\lambda_{max}(M^H M) &\leq \text{Trace}(M^H M) & (8)\\
&= \text{Trace}(F_\Omega^H I_T I_T^H F_\Omega) & (9)\\
&= \frac{1}{N} \sum_{w \in \Omega} \sum_{t \in T} e^{-j\frac{2\pi wt}{N}} e^{j\frac{2\pi wt}{N}}.
\end{aligned}
$$

Therefore,

$$
\text{Trace}(M^H M) = \frac{|\Omega||T|}{N}, \text{ and } \lambda_{max}(M^H M) \leq \frac{|\Omega||T|}{N}.
$$

Thus, $(\Phi_{T\Omega}^H \Phi_{T\Omega})$ is PD when $|\Omega||T| < N$.

Hence, the condition such that $(\Phi_{T\Omega}^{H}\Phi_{T\Omega})$ is NOT positive definite is

$$|\mathrm{supp}(f)| \cdot |\mathrm{supp}(\hat{f})| \geq N \blacksquare \tag{10}$$

Consequence of the Uncertainty Principle (UP):

Since f and $\hat{f}$ cannot both be highly sparse, a sparse representation of f in Time has a *unique* image under the Fourier dictionary.

Proof:
If f is an unknown sparse signal in Time such that $\|f\|_{\ell_0} = S$, and we measure any $2S$ Fourier coefficients of f as:

$$y = F_{2S} f;$$

where, F_{2S} is the Fourier dictionary having only $2S$ rows.

Assume that there exist another S-sparse (in Time) signal f'. Take the same $2S$ Fourier coefficients of f' as:

$$y' = F_{2S} f'.$$

The signal $(f - f')$ is $2S$-sparse in Time. If $y = y'$, then the $2S$ Fourier coefficients of the signal $f - f'$ are given by:

$$F_{2S}(f - f') = 0.$$

and since $F_{2S}^H F_{2S}$ is PD, then $f = f'$.

The UP guarantees that we can recover a S-sparse (in Time) signal f, from $2S$ Fourier coefficients by solving

$$(P0): \quad \min_f \|f\|_{\ell_0} \;\; s.t. \;\; y = F_{2S}f \tag{11}$$

Random Waveforms

Randomness plays a major role in the measurement scenario.

Φ ⇒

Examples

- Each entry of Φ can be drawn from i.i.d. Gaussian distribution (i.e. $\phi_{i,j} \sim N(0,1)$).
- Each entry of Φ can be drawn from i.i.d. Bernoulli distribution (i.e. ± 1).

In random projections:

$$y = \Phi x$$

where, Φ follows a given random distribution, x can be recovered from M samples with high probability when M satisfies:

$$M \geq C \cdot S \cdot log(N/S), \quad C \geq 1$$

- Proved through the Restricted Isometry Property (RIP) as described shortly.

Remark

Note that when using incoherent orthobasis, the required number of measurements is $M = 2S$ and when using random projections, we require more measurements $M \geq C \cdot S \cdot log(N/S)$, to recover the signal.

Restricted Isometry Property (RIP)

1. Gives the probability that any s-sparse signal can be recovered from its random projections.
2. Uses probabilistic methods to prove [1].
3. Gives the minimum number of projections required to guarantee the recovery of any s-sparse signal from its random projections.

Restricted Isometry Property (RIP)

Theorem

A matrix $A \in R^{m \times n}$ satisfies the Restricted Isometry Property if there exists a constant $\delta > 0$ such that

$$(1-\delta)||x||_2^2 \leq ||Ax||_2^2 \leq (1+\delta)||x||_2^2$$

with high probability†.

Outline of the proof:

1. Show that for a fixed sparse vector x, $||Ax||_2^2 \approx ||x||_2^2$ with high probability.
2. Count up the "number" of sparse vectors, and show that $||Ax||_2^2 \approx ||x||_2^2$ for all of them with high probability.

†Baraniuk, R. *et al.*. A simple proof of the Restricted Isometry Property for Random Matrices. Springer science 2008.

RIP for Gaussian Random Matrices

Let $A \in R^{m\times n}$, with $m < n$, be a matrix with i.i.d. Gaussian random entries:

$$a_{i,j} \sim N\left(0, \frac{1}{m}\right)$$

Fix $x \in R^n$ and set $b = Ax$.

1. Show that for a fixed sparse vector x, $\|Ax\|_2^2 \approx \|x\|_2^2$ with high probability. The i^{th} element of b is

$$b_i = \sum_{j=1}^{n} a_{i,j}x_j \quad \sim \quad N(0, \sigma_i^2)$$

where $\sigma_i^2 = E\{b_i^2\} = \sum_{j=1}^{n} E\{a_{i,j}^2\}x_j^2 = \sum_{j=1}^{n} \frac{1}{m}x_j^2 = \frac{1}{m}\|x\|_2^2$. The ℓ_2-norm of b is

$$\|b\|_2^2 = \sum_{i=1}^{m} |b_i|^2 \text{ is a Chi-square r.v.}$$

and $E\{\|b\|_2^2\} = \sum_{i=1}^{m} E\{|b_i|^2\} = \sum_{i=1}^{m} \frac{1}{m}\|x\|_2^2 = \|x\|_2^2$.

To find the probability that $\|Ax\|_2^2 \approx \|x\|_2^2$, the Markov Inequality is used:

Markov Inequality

If y is a positive r.v.:

$$P(y > t) \leq \frac{E\{y\}}{t} \tag{12}$$

Proof:

$$\begin{aligned} E\{y\} &= \int_0^\infty y f(y) dy \\ &\geq \int_t^\infty y f(y) dy \\ &\geq t \int_t^\infty f(y) dy \\ &= tP(y > t) \end{aligned} \tag{13}$$

Let $y = ||b||_2^2$ (a positive r.v.), and without loss of generality, assume that $||x||_2^2 = 1$

$$
\begin{aligned}
P(y > (1+\delta)) &= P(e^{\lambda y} > e^{\lambda(1+\delta)});\ e^x \text{ is a monotonic function} \qquad (14)\\
&\leq \frac{E\{e^{\lambda y}\}}{e^{\lambda(1+\delta)}};\ \text{Markov Inequality}\\
&= \frac{E\{e^{\lambda(\sum_{i=1}^m b_i^2)}\}}{e^{\lambda(1+\delta)}}\\
&= \frac{E\{e^{\lambda b_1^2} e^{\lambda b_2^2} \ldots e^{\lambda b_m^2}\}}{e^{\lambda(1+\delta)}}\\
&= \frac{\prod_{i=1}^m E\{e^{\lambda b_i^2}\}}{e^{\lambda(1+\delta)}};\ \text{by independence of the } b_i's\\
&= \frac{E\{e^{\lambda b_1^2}\}^m}{e^{\lambda(1+\delta)}};\ b_i's \text{ are identically distributed}\\
P(y > (1+\delta)) &\leq \frac{E\{e^{\lambda b_1^2}\}^m}{e^{\lambda(1+\delta)}}
\end{aligned}
$$

Given $b_i \sim N(0, 1/m)$, then

$$
\begin{aligned}
E\{e^{\lambda b^2}\} &= \int_{-\infty}^{\infty} e^{\lambda b^2} f(b) db \\
&= \int_{-\infty}^{\infty} e^{\lambda b^2} \sqrt{\frac{m}{2\pi}} e^{-\frac{b^2 m}{2}} db \\
&= \sqrt{\frac{m}{m-2\lambda}} \int_{-\infty}^{\infty} \sqrt{\frac{m-2\lambda}{2\pi}} e^{-\frac{b^2(m-2\lambda)}{2}} db \\
E\{e^{\lambda b^2}\} &= \sqrt{\frac{m}{m-2\lambda}}; \text{ if } \lambda < m/2
\end{aligned} \tag{15}
$$

Replacing (15) in (14),

$$\begin{aligned} P(y > (1+\delta)) &\leq \frac{E\{e^{\lambda b_1^2}\}^m}{e^{\lambda(1+\delta)}} \qquad (16) \\ &= \frac{\left(\frac{m}{m-2\lambda}\right)^{m/2}}{e^{\lambda(1+\delta)}} \\ P(y > (1+\delta)) &\leq \left(\frac{e^{-2\lambda(1+\delta)/m}}{1-2\lambda/m}\right)^{m/2}; \forall\lambda < m/2 \end{aligned}$$

Choose $\lambda = \frac{m\delta}{2(1+\delta)}$;

$$\begin{aligned} P(y > (1+\delta)) &\leq \left(\frac{e^{-\delta}}{1-\delta/(1+\delta)}\right)^{m/2} \qquad (17) \\ &= \left((1+\delta)e^{-\delta}\right)^{m/2} \end{aligned}$$

By Taylor expansion:

$$\begin{aligned} \ln(1+\delta) &= \delta - \delta^2/2 + \delta^3/3 - \delta^4/4 + \ldots \\ \ln(1+\delta) &< \delta - \delta^2/2 + \delta^3/2 \\ (1+\delta) &< e^{\delta-\delta^2/2+\delta^3/2} \\ (1+\delta)e^{-\delta} &< e^{-(\delta^2/2-\delta^3/2)} \end{aligned} \quad (18)$$

(18) in (17):

$$P(y > (1+\delta)) \leq e^{-(\delta^2-\delta^3)m/4} \quad (19)$$

Thus, in general:

$$P(\|b\|_2^2 > (1+\delta)\|x\|_2^2) \leq e^{-(\delta^2-\delta^3)m/4} \quad (20)$$

Similarly, for the lower bound can be shown that

$$P(\|b\|_2^2 < (1-\delta)\|x\|_2^2) \leq e^{-(\delta^2-\delta^3)m/4} \tag{21}$$

proof:
Let $y = \|b\|_2^2$ (a positive r.v.), and without loss of generality, assume that $\|x\|_2^2 = 1$.

$$\begin{aligned} P(y < (1-\delta)) &= P(-y > -(1-\delta)) \\ &= P(e^{-\lambda y} > e^{-\lambda(1-\delta)});\ e^x \text{ is a monotonic function} \\ &\leq \frac{E\{e^{-\lambda y}\}}{e^{-\lambda(1-\delta)}};\ \text{Markov Inequality} \\ &= \frac{E\{e^{-\lambda(\sum_{i=1}^m b_i^2)}\}}{e^{-\lambda(1-\delta)}} \\ &= \frac{E\{e^{-\lambda b_1^2} e^{-\lambda b_2^2} \ldots e^{-\lambda b_m^2}\}}{e^{-\lambda(1-\delta)}} \end{aligned} \tag{22}$$

$$\begin{aligned}
P(y<(1-\delta)) &\leq \frac{E\{e^{-\lambda b_1^2}e^{-\lambda b_2^2}\ldots e^{-\lambda b_m^2}\}}{e^{-\lambda(1-\delta)}} \\
&= \frac{\prod_{i=1}^{m} E\{e^{-\lambda b_i^2}\}}{e^{-\lambda(1-\delta)}};\ \text{by independence of the } b_i's \\
&= \frac{E\{e^{-\lambda b_1^2}\}^m}{e^{-\lambda(1-\delta)}};\ b_i's \text{ are identically distributed} \\
P(y<(1-\delta)) &\leq \frac{E\{e^{-\lambda b_1^2}\}^m}{e^{-\lambda(1-\delta)}} \qquad (23)
\end{aligned}$$

Given $b \sim N(0, 1/m)$, then

$$
\begin{aligned}
E\{e^{-\lambda b^2}\} &= \int_{-\infty}^{\infty} e^{-\lambda b^2} f(b) db \\
&= \int_{-\infty}^{\infty} e^{-\lambda b^2} \sqrt{\frac{m}{2\pi}} e^{-\frac{b^2 m}{2}} db \\
&= \sqrt{\frac{m}{m+2\lambda}} \int_{-\infty}^{\infty} \sqrt{\frac{m+2\lambda}{2\pi}} e^{-\frac{b^2(m+2\lambda)}{2}} db \\
E\{e^{-\lambda b^2}\} &= \sqrt{\frac{m}{m+2\lambda}}; \text{ if } \lambda > -m/2 \qquad (24)
\end{aligned}
$$

Replacing (24) in (23),

$$\begin{aligned} P(y < (1-\delta)) &\leq \frac{E\{-e^{\lambda b_1^2}\}^m}{e^{-\lambda(1-\delta)}} \\ &= \frac{\left(\frac{m}{m+2\lambda}\right)^{m/2}}{e^{-\lambda(1-\delta)}} \\ P(y < (1-\delta)) &\leq \left(\frac{e^{2\lambda(1-\delta)/m}}{1+2\lambda/m}\right)^{m/2}; \forall\lambda < m/2 \end{aligned} \tag{25}$$

Choose $\lambda = \frac{m\delta}{2(1-\delta)}$;

$$\begin{aligned} P(y < (1-\delta)) &\leq \left(\frac{e^{\delta}}{1+\delta/(1-\delta)}\right)^{m/2} \\ &= \left((1-\delta)e^{-\delta}\right)^{m/2} \end{aligned} \tag{26}$$

By Taylor expansion:

$$\begin{aligned}
\ln(1-\delta) &= -\delta - \delta^2/2 - \delta^3/3 - \delta^4/4 + \ldots \\
\ln(1-\delta) &< -\delta - \delta^2/2 - \delta^3/3 \\
(1-\delta) &< e^{-\delta-\delta^2/2-\delta^3/3} \\
(1-\delta)e^{\delta} &< e^{-(\delta^2/2+\delta^3/3)} \qquad (27)
\end{aligned}$$

(27) in (26):

$$\begin{aligned}
P(y < (1-\delta)) &\leq e^{-(\delta^2/2+\delta^3/3)m/2} \\
P(y < (1-\delta)) &\leq e^{-(\delta^2-\delta^3)m/4};\ \text{since:} \delta^2/2 + \delta^3/3 > (\delta^2 - \delta^3)/2 \qquad (28)
\end{aligned}$$

Thus, in general:

$$P(\|b\|_2^2 < (1-\delta)\|x\|_2^2) \leq e^{-(\delta^2-\delta^3)m/4} \qquad (29)$$

From (20) and (29),

$$P((1-\delta)\|x\|_2^2 \leq \|b\|_2^2 \leq (1+\delta)\|x\|_2^2) > 1 - e^{-(\delta^2-\delta^3)m/4} - e^{-(\delta^2-\delta^3)m/4}$$

$$P((1-\delta)\|x\|_2^2 \leq \|b\|_2^2 \leq (1+\delta)\|x\|_2^2) > 1 - 2e^{-(\delta^2-\delta^3)m/4} \tag{30}$$

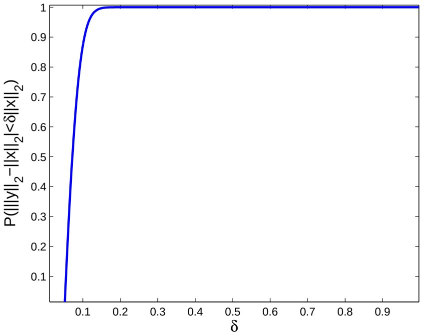
Example:

If $\delta = 1/2$ and $m = 1000$, then

$$P\left(\frac{1}{2}\|x\|_2^2 \leq \|b\|_2^2 \leq \frac{3}{2}\|x\|_2^2\right) > 1 - 2e^{-(\frac{1}{4}-\frac{1}{8})1000/4}$$

$$P\left(\frac{1}{2}\|x\|_2^2 \leq \|b\|_2^2 \leq \frac{3}{2}\|x\|_2^2\right) > 1 - 5.4 \times 10^{-14} \tag{31}$$

For $m = 1000$, the following plot shows the probability of satisfying the bound as a function of δ:



It has been proved that for a fixed sparse signal x, a matrix A with i.i.d. Gaussian entries satisfies:

$$P(|\|Ax\|_2^2 - \|x\|_2^2| > \delta\|x\|_2^2|) \leq 2e^{-C_0(\delta)m} \tag{32}$$

where $C_0(\delta) = e^{-(\delta^2-\delta^3)/4}$ is some constant that depends only on δ.

To show the RIP for all sparse signals x, it is necessary to find the probability for all possible support sets $\mathbf{T}$ with cardinality $|\mathbf{T}| \leq 2S$.

2. Count up the "number" of sparse vectors, and show that $\|Ax\|_2^2 \approx \|x\|_2^2$ for all of them with high probability.

To count the "number" of sparse vectors, it is necessary to find how many vectors x satisfy

$$\max_{|\mathbf{T}|\leq 2S} \sup_{x\in B_2^{\mathbf{T}}} \left|\|Ax\|_2^2 - \|x\|_2^2\right| \leq \delta \tag{33}$$

where,

- $B_2^{\mathbf{T}} = \{x \in \mathbb{R}^n : x \text{ is supported only in } \mathbf{T} \text{ and } \|x\|_2^2 = 1\}$.
- $\sup_{x\in B_2^{\mathbf{T}}}$ is the smallest upper bound of vectors $x \in B_2^{\mathbf{T}}$ satisfying $\left|\|Ax\|_2^2 - \|x\|_2^2\right| \leq \delta$.
- $\max_{|\mathbf{T}|\leq 2S}$ is the maximum over all support sets $\mathbf{T}$ of size $\leq 2S$.

Solution:

First, fix a set $\mathbf{T}$ of size $|\mathbf{T}| \leq 2S$ and find the $P\left(\sup_{x \in B_2^{\mathbf{T}}} |\|Ax\|_2^2 - \|x\|_2^2| > \delta \right)$.

Lemma 1

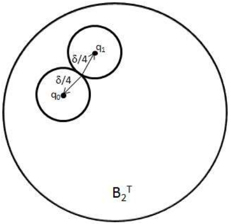
Let $A \in R^{n \times m}$ be a random matrix that satisfies (32). Let $\mathbf{T}$ a fixed set of size $|\mathbf{T}| \leq 2S$ and let δ be a fixed constant between 0 and 1, then

$$P\left(\sup_{x \in B_2^{\mathbf{T}}} |\|Ax\|_2^2 - \|x\|_2^2| > \delta \right) \leq 2\left(\frac{12}{\delta}\right)^{2S} e^{-C_0(\delta/2)m} \tag{34}$$

Proof: Approximate the set $B_2^{\mathbf{T}}$ by a finite set Q. The finite set Q, with elements $\{q_0, q_1, \ldots\}$, is such that every $x \in B_2^{\mathbf{T}}$ is within $\delta/4$ of an element in Q, *i.e.*

$$\min_{q \in Q} \|x - q\|_2 \leq \delta/4, \quad \forall x \in B_2^{\mathbf{T}}. \tag{35}$$

Essentially, Q is a set containing all the vectors x in $B_2^{\mathbf{T}}$ with a distortion $\leq \delta/4$.



The number of elements in the set Q is given by[†]:

$$|Q| \leq \left(\frac{12}{\delta}\right)^{2S}; \text{ where } |\mathbf{T}| \leq 2S. \tag{36}$$

[†] Lorentz, G., von Golitschek, M., Makovoz, Y. *Constructive Approximation: Advanced Problems*. vol. 304., Springer, Berlin. 1996

For any fixed $q_0 \in Q$, then, according to (32)

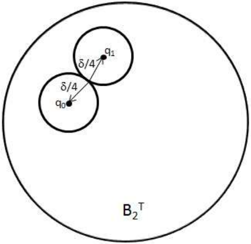
$$P(|\|Aq_0\|_2^2 - \|q_0\|_2^2| > \delta/2) \leq 2e^{-C_0(\delta/2)m} \tag{37}$$

Recall the union bound probability:

$$P(W_1 \cup W_2 \cup \ldots \cup W_k) \leq \sum_{i=1}^{k} P(W_i)$$

Applying the union bound probability along all the elements given in (36), then

$$P(\max_{q \in Q} |\|Aq\|_2^2 - \|q\|_2^2| > \delta/2) \leq 2\left(\frac{12}{\delta}\right)^{2S} e^{-C_0(\delta/2)m} \tag{38}$$



Note that, if it is true that all the $q \in Q$ are "well behaved" in that

$$\max_{q \in Q} \left| \|Aq\|_2^2 - \|q\|_2^2 \right| \leq \delta/2. \tag{39}$$

Then, it is also true that

$$\sup_{x \in B_2^{\mathbf{T}}} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \leq \delta. \tag{40}$$

This concludes that the probability of $\sup_{x \in B_2^{\mathbf{T}}} \left| \|Ax\|_2^2 - \|x\|_2^2 \right| > \delta$ is given by:

$$P(\left| \|Ax\|_2^2 - \|x\|_2^2 \right| > \delta) \leq 2 \left(\frac{12}{\delta} \right)^{2S} e^{-C_0(\delta/2)m} \tag{41}$$

Now, for all $2S-$sparse x signals simultaneously, $||Ax||_2^2 \approx ||x||_2^2$ has to be established.

Lemma 2

Let $A \in R^{n\times m}$ be a random matrix that satisfies (32). Then there exist a constant $C_1(\delta)$ depending only on δ, such that

$$P\left(max_{|\mathbf{T}|\leq 2S} \sup_{x\in B_2^{\mathbf{T}}} ||Ax||_2^2 - ||x||_2^2| \geq \delta\right) \text{ is small} \tag{42}$$

when

$$m \geq C_1(\delta) S log(n/S).$$

Proof: For a fixed $2S$-dimensional subspace $B_2^{\mathbf{T}}$

$$P\left(\max_{x \in B_2^{\mathbf{T}}} \|Ax\|_2^2 - \|x\|_2^2| > \delta\right) \leq 2(12/\delta)^{2S} e^{-C_0(\delta/2)m}$$

In $\mathbb{R}^n$, there are $\binom{n}{2S}$ such subspaces:

$$\binom{n}{2S} = \frac{n!}{(n-2S)!(2S)!} \leq \frac{n^{2S}}{(2S)!} \leq \left(\frac{ne}{2S}\right)^{2S}. \tag{43}$$

Homework: Provide a proof of Eq. (43).

Applying the union bound:

$$P\left(\max_{|\mathbf{T}| \leq 2S} \sup_{x \in B_2^{\mathbf{T}}} |\|Ax\|_2^2 - \|x\|_2^2| \geq \delta\right) \leq 2\left(\frac{ne}{2S}\right)^{2S} (12/\delta)^{2S} e^{-C_0(\delta/2)m}$$

$$
\begin{aligned}
P\left(\max_{|\mathbf{T}|\le 2S}\sup_{x\in B_2^{\mathbf{T}}}\left|\|Ax\|_2^2-\|x\|_2^2\right|\ge\delta\right) &\le 2\left(\frac{ne}{2S}\right)^{2S}(12/\delta)^{2S}e^{-C_0(\delta/2)m}\\
&= 2e^{-C_0(\delta/2)\left(m-\frac{2S}{C_0(\delta/2)}\log\left(\frac{12ne}{2\delta S}\right)\right)}.
\end{aligned}
$$

If

$$
m\ge\frac{2S}{C_0(\delta/2)}\log\left(\frac{12ne}{2\delta S}\right), \tag{44}
$$

then, the probability $P\left(\max_{|\mathbf{T}|\le 2S}\sup_{x\in B_2^{\mathbf{T}}}\left|\|Ax\|_2^2-\|x\|_2^2\right|\ge\delta\right)$ is small.

Homework:
Prove that Eq. (44) can be rewritten as $m > C_1(\delta)S\log(n/S)$, when $n\ge 6eS/\delta$.

This ends the proof of Lemma 2.

The Restricted Isometry Property (RIP) guarantees that we can recover a S-sparse signal x as a unique solution of the following problem:

$$\min_x \|x\|_0 \text{ s.t. } b = Ax. \text{ (P0)}$$

Because:
Assuming that there exists another signal x_1 having also minimum ℓ_0-norm (*i.e.* $\|x_1\|_0 \leq S$), then, if $x \neq x_1$

$$\|x - x_1\|_2^2 \neq 0 \tag{45}$$

and, by the RIP we know that $\|x - x_1\|_2^2 \approx \|Ax - Ax_1\|_2^2$, then

$$\|Ax - Ax_1\|_2^2 = \|b - b_1\|_2^2 \neq 0 \tag{46}$$

which means that any other S-sparse signal x_1 does not satisfy the constraint of the problem in (P0).

Recovery Via ℓ_1 Minimization

If the random matrix R obeys the RIP, then:

- Every S-sparse signal has a unique image under R; which means that $b = Rx$ is different for each S-sparse signal x.
- Given b, x can be recovered by solving: $\min_x \|x\|_0$ s.t. $b = Rx$.

MAIN PROBLEM:
The ℓ_0 minimization is NP-hard, then we want to solve a convex minimization problem, i.e.:

$$\min_x \|x\|_1 \text{ s.t. } b = Rx. \quad \text{(P1)}$$

If x_0 is the solution to (P0) and x^* is the solution to (P1), we need to

find the conditions under which $x_0 = x^*$.

Recovery Via ℓ_1 Minimization

Conditions for EXACT recovery using ℓ_1 Minimization.

Call x^* the solution to (P1) and x_0 the solution to (P0). Set $h = x^* - x_0$ as the recovery error. We need to show that $h = 0$ in order to show EXACT recovery.

1. Given the random projections b, then x^* and x_0 are both feasible solutions. But x^* is defined to be the feasible point with smallest ℓ_1-norm, i.e.,

$$\|x^*\|_1 \leq \|x_0\|_1, \quad \text{or} \quad \|x_0 + h\|_1 \leq \|x_0\|_1. \tag{47}$$

Proof: Let the set T be the support of h, this is:

$$h_T[i] = \begin{cases} h[i] & \text{if } i \in T \\ 0 & \text{if } i \notin T. \end{cases}$$

And, let the set T_0 be the support of x_0.

By the triangle inequality:

$$\begin{aligned} \|x_0 + h\|_1 &= \sum_{i \in T_0} |x_0[i] + h[i]| + \sum_{i \in T_0^C} |h[i]| \\ &\geq \sum_{i \in T_0} |x_0[i] + h[i]| + \sum_{i \in T_0^C} |h[i]| \\ &= \|x_0\|_1 - \|h_{T_0}\|_1 + \|h_{T_0^C}\|_1. \end{aligned} \quad (48)$$

Using (47) in (48), then:

$$\begin{aligned} \|x_0\|_1 &\geq \|x_0\|_1 - \|h_{T_0}\|_1 + \|h_{T_0^C}\|_1 \\ \|h_{T_0}\|_1 &\geq \|h_{T_0^C}\|_1 \end{aligned} \quad (49)$$

2. Since the matrix R obeys the RIP, then $\forall h \in \text{Null}(R)$:

$$\|h_{T_0}\|_1 \leq \rho \|h_{T_0^C}\|_1; \quad \rho < 1$$

for every set T_0 with $|T_0| \leq s$.

Proof: Let T_0^C divided into decreasing subsets $T_1, T_2, \ldots$ of size s'. We know that $h = x^* - x_0$ belongs to the Null space of R, (i.e. $Rh = 0$), then:

$$R(h_{T_0 \cup T_1} + h_{(T_0 \cup T_1)^C}) = 0; \quad \text{or} \quad R(h_{T_0 \cup T_1}) = -\sum_{j=2} Rh_{T_j}$$

and so

$$\|Rh_{T_0 \cup T_1}\|_2 = \|\sum_{j=2} Rh_{T_j}\|_2 \leq \sum_{j=2} \|Rh_{T_j}\|_2 \tag{50}$$

Since $h_{T_0 \cup T_1}$ is a $s+s'$ sparse vector, applying the $s+s'$-RIP

$$\sqrt{1-\delta_{s+s'}}\|h_{T_0 \cup T_1}\|_2 \leq \|Rh_{T_0 \cup T_1}\|_2. \tag{51}$$

Since each h_{T_j} is s'-sparse, applying the s'-RIP

$$\|h_{T_j}\|_2 \leq \sqrt{1+\delta_{s'}}\|h_{T_j}\|_2. \tag{52}$$

Replacing (51) and (52) in (50), then

$$\sqrt{1-\delta_{s+s'}}\|h_{T_0 \cup T_1}\|_2 \leq \sqrt{1+\delta_{s'}} \sum_{j \geq 2} \|h_{T_j}\|_2. \tag{53}$$

For each $j \geq 2$, all the magnitudes of the values in h_{T_j} are less than all the magnitudes of the $h_{T_{j-1}}$, since the set is organized in a decreasing way. Thus, the maximum value in h_{T_j} is smaller than the average of the magnitudes in $h_{T_{j-1}}$, i.e.

$$\|h_{T_j}\|_\infty \leq \frac{1}{s'}\|h_{T_{j-1}}\|_1$$

Thus,

$$\|h_{T_j}\|_2 \leq \sqrt{s'}\|h_{T_j}\|_\infty \leq \frac{1}{\sqrt{s'}}\|h_{T_{j-1}}\|_1$$

and

$$\sum_{j\geq 2}\|h_{T_j}\|_2 \leq \frac{1}{\sqrt{s'}}\sum_{j\geq 1}\|h_{T_j}\|_1 = \frac{1}{\sqrt{s'}}\|h_{T_0^C}\|_1 \tag{54}$$

Using (54) and (53), then

$$\|h_{T_0\cup T_1}\|_2 \leq \frac{\sqrt{1+\delta_{s'}}}{\sqrt{1-\delta_{s+s'}}}\frac{\|h_{T_0^C}\|_1}{\sqrt{s'}}. \tag{55}$$

Finally,

$$\|h_{T_0}\|_1 \leq \sqrt{s}\|h_{T_0}\|_2 \leq \sqrt{s}\|h_{T_0\cup T_1}\|_2 \tag{56}$$

Replacing (55) in (56), then

$$\begin{aligned} \|h_{T_0}\|_1 &\leq \frac{\sqrt{1+\delta_{s'}}}{\sqrt{1-\delta_{s+s'}}}\frac{\sqrt{s}\|h_{T_0^C}\|_1}{\sqrt{s'}} \\ &= \rho\|h_{T_0^C}\|_1 \end{aligned} \tag{57}$$

with $\rho = \frac{\sqrt{1+\delta_{2s}}}{\sqrt{1-\delta_{3s}}}\frac{1}{\sqrt{2}} \leq 1$ when $s' = 2s$ and $2\delta_{3s} + \delta_{2s} \leq 1$.

1. In (49), it has been proved that: $\|h_{T_0}\|_1 \geq \|h_{T_0^C}\|_1$

2. In (57), it has been proved that: $\|h_{T_0}\|_1 \leq \rho\|h_{T_0^C}\|_1$.

The only way h can obey [1] and [2], is that $h = 0$ which implies EXACT recovery.