Compressive Sensing with Random Matrices

Lucas Connell

University of Georgia

29 November 2017

Overview

We will discuss compressive sensing, and, in particular, the viability of using random matrices.
An outline:

1. Introduce some vector notation and terminology
2. Discuss Compressive Sensing
3. Review probability
4. Define random matrices and concentration of measure
5. Define an ϵ-covering number, and prove a short lemma
6. Use these tools to prove our main result

Vector Notation

The ℓ_p^N norm of a vector $\boldsymbol{x} \in \mathbb{R}^N$ is defined as

$$\|\boldsymbol{x}\|_{\ell_p^N} := \begin{cases} \left(\sum_{i=1}^N |x_i|^p\right)^{\frac{1}{p}} & \text{if } 0 < p < \infty \\ \max\limits_{i=1,\dots,N} |x_i| & \text{if } p = \infty \end{cases}$$

Vector Notation

The ℓ_p^N norm of a vector $\boldsymbol{x} \in \mathbb{R}^N$ is defined as

$$\|\boldsymbol{x}\|_{\ell_p^N} := \begin{cases} \left(\sum_{i=1}^N |x_i|^p\right)^{\frac{1}{p}} & \text{if } 0 < p < \infty \\ \max_{i=1,\ldots,N} |x_i| & \text{if } p = \infty \end{cases}$$

Definition (Sparsity)

We say a vector $\boldsymbol{x} \in \mathbb{R}^N$ is k-sparse if $\#\{i : x_i \neq 0\} \leq k$

Vector Notation

The ℓ_p^N norm of a vector $\boldsymbol{x} \in \mathbb{R}^N$ is defined as

$$\|\boldsymbol{x}\|_{\ell_p^N} := \begin{cases} \left(\sum_{i=1}^N |x_i|^p\right)^{\frac{1}{p}} & \text{if } 0 < p < \infty \\ \max\limits_{i=1,\ldots,N} |x_i| & \text{if } p = \infty \end{cases}$$

Definition (Sparsity)

We say a vector $\boldsymbol{x} \in \mathbb{R}^N$ *is k-sparse if* $\#\{i : x_i \neq 0\} \leq k$

In $\mathbb{R}^N$, with a fixed $k \leq N$ and given a set $T = \{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, N\}$ of indices, we define

$$X_T := \{[x_1, x_2, \ldots, x_N]^T : x_i = 0 \text{ for } i \notin T\}$$

which is a k-dimensional subspace of $\mathbb{R}^N$
We similarly define $\boldsymbol{x}_T$

Examples

Example (Sparse Vectors)

Let

$$\boldsymbol{x} = \begin{bmatrix} -3 \\ 4 \\ 0 \\ 2 \\ 0 \end{bmatrix} \quad \textit{and} \quad \boldsymbol{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$\boldsymbol{x}$ *is 3-sparse.*
$\boldsymbol{y}$ *is 1-sparse.*

Example (X_T Subspace)

Consider $\mathbb{R}^5$, *and set* $T = \{1, 2, 4\}$.
$X_T := \{[x_1, x_2, 0, x_4, 0]^T : x_1, x_2, x_4 \in \mathbb{R}\}$

Motivation

In many applications, one wants to solve

$$\Phi \boldsymbol{x} = \boldsymbol{b} \text{ where } \boldsymbol{x} \in \mathbb{R}^N \text{ is } k \text{ - sparse, } \Phi \in \mathbb{R}^{n \times N}, \text{ and } n << N$$

Motivation

In many applications, one wants to solve

$$\Phi \boldsymbol{x} = \boldsymbol{b} \text{ where } \boldsymbol{x} \in \mathbb{R}^N \text{ is } k \text{ - sparse, } \Phi \in \mathbb{R}^{n \times N}, \text{ and } n << N$$

This can be quite hard, and we instead solve:

$$\operatorname{argmin} \|\boldsymbol{x}\|_{\ell_1^N}, \text{ such that } \Phi \boldsymbol{x} = \boldsymbol{b} \tag{1}$$

Motivation

In many applications, one wants to solve

$$\Phi \boldsymbol{x} = \boldsymbol{b} \text{ where } \boldsymbol{x} \in \mathbb{R}^N \text{ is } k \text{ - sparse, } \Phi \in \mathbb{R}^{n \times N}, \text{ and } n << N$$

This can be quite hard, and we instead solve:

$$\operatorname{argmin} \|\boldsymbol{x}\|_{\ell_1^N}, \text{ such that } \Phi \boldsymbol{x} = \boldsymbol{b} \tag{1}$$

Theorem (Candes and Tao, "Decoding by linear programming", 2005)

Suppose $\boldsymbol{b} = \Phi \overline{\boldsymbol{x}}$, with $\overline{\boldsymbol{x}}$ k-sparse. If Φ has the Restricted Isometry Property of order 2k, then $\overline{\boldsymbol{x}}$ is the unique solution to (1)

Motivation

In many applications, one wants to solve

$$\Phi \boldsymbol{x} = \boldsymbol{b} \text{ where } \boldsymbol{x} \in \mathbb{R}^N \text{ is } k \text{ - sparse, } \Phi \in \mathbb{R}^{n \times N}, \text{ and } n << N$$

This can be quite hard, and we instead solve:

$$\operatorname{argmin} \|\boldsymbol{x}\|_{\ell_1^N}, \text{ such that } \Phi \boldsymbol{x} = \boldsymbol{b} \tag{1}$$

Theorem (Candes and Tao, "Decoding by linear programming", 2005)

Suppose $\boldsymbol{b} = \Phi \overline{\boldsymbol{x}}$, with $\overline{\boldsymbol{x}}$ k-sparse. If Φ has the Restricted Isometry Property of order $2k$, then $\overline{\boldsymbol{x}}$ is the unique solution to (1)

Theorem (Candes and Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?", 2006)

If $\Phi \in \mathbb{R}^{n \times N}$ is a subgaussian random matrix with $n \geq ck \log\left(\frac{N}{k}\right)$, then Φ satisfies the RIP of order k with overwhelming probability.

Probability Review

Definition (Random Variable)

A (real) random variable returns a real number with some probability determined by a probability distribution function.

Example: *Let* $X \sim \mathcal{N}(0,1)$ *(standard normal distribution)*

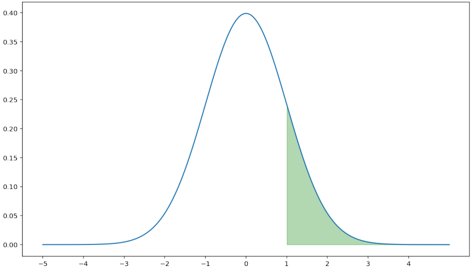

Figure: $\Pr[X \geq 1] = .15866$

Subgaussian Random Variable: A subgaussian random variable X is a random variable whose distribution function's tails decay exponentially fast. Or, more precisely, for any $t \in \mathbb{R}$:

$$\Pr\left[|X - \mu| \geq t\right] \leq 2e^{-\frac{t^2}{2\sigma^2}}$$

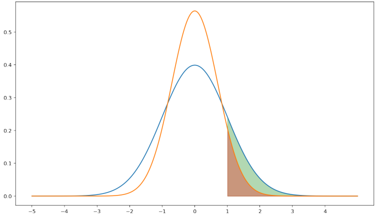

Figure: Blue is standard normal; orange is subgaussian. Notice the red area (subgaussian) is less than the green area (standard normal)

Probability Review

Union Bounds: Given random variables X_1, X_2, we know $\Pr[X_1 \geq t \text{ or } X_2 \geq t] \leq \Pr[X_1 \geq t] + \Pr[X_2 \geq t]$. The same inequality holds for any finite number of random variables.

Example

Let $X_1, X_2, ..., X_m \sim \mathcal{N}(0,1)$ be independent. Using a finite union bound:

$$Pr[X_i \geq t \text{ for at least one } i = 1, 2, ..., m] \leq \sum_{i=1}^{m} Pr[X_i \geq t] = mPr[X_1 \geq t]$$

Concentration of Measure

The sum of (independent) random variables concentrates around the mean.

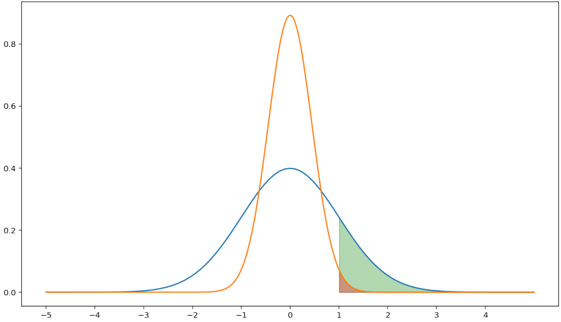

Figure: Blue is standard normal; orange is $\frac{\sum_{i=1}^{5} X_i}{5}$ where $X_1, ..., X_5 \sim \mathcal{N}(0, 1)$

Random Matrix

Definition (Random Matrix)

A random matrix, $\Phi \in \mathbb{R}^{n \times N}$, is a matrix whose entries, $\phi_{i,j}$, are each (independent) random variables.

Random Matrix

Definition (Random Matrix)

A random matrix, $\Phi \in \mathbb{R}^{n \times N}$, is a matrix whose entries, $\phi_{i,j}$, are each (independent) random variables.

Theorem (Subgaussian Random Matrix)

Letting $\boldsymbol{q} \in \mathbb{R}^N$, $\epsilon \in (0,1)$, and Φ be an $n \times N$ matrix with entries $\phi_{i,j} \sim \mathcal{N}(0, \frac{1}{n})$, this concentration results in $\exists c_0(\epsilon) > 0$ with:

$$Pr\left[(1-\epsilon)\left\|\boldsymbol{q}\right\|_{\ell_2^N}^2 \leq \left\|\Phi\boldsymbol{q}\right\|_{\ell_2^n}^2 \leq (1+\epsilon)\left\|\boldsymbol{q}\right\|_{\ell_2^N}^2 \ \ fails\right] \leq 2e^{-nc_0(\epsilon)}$$

Covering Numbers

Definition (ϵ-Covering Number ϵ-Packing Number)

1. ϵ-**covering number**: $N(U,\epsilon) = min\ n\ such\ that\ \exists\{q_1, q_2, ..., q_n\}$ *with the property*

$$U \subseteq \bigcup_{i=1}^{n} B_\epsilon(q_i)$$

2. ϵ-**packing number**: $P(U,\epsilon) = max\ m\ such\ that\ \exists\{q_1, q_2, ..., q_m\}$ *with the property* $d(q_i, q_j) > \epsilon,\ \forall i \neq j$

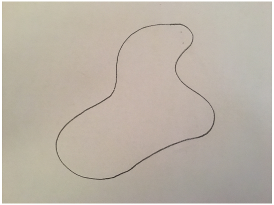

Figure: A space, U

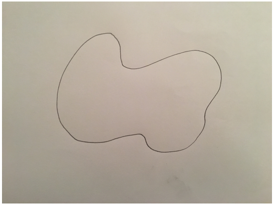

Figure: A space, V

Covering Numbers

Definition (ϵ-Covering Number ϵ-Packing Number)

1. ϵ-**covering number**: $N(U,\epsilon) = min\ n\ such\ that\ \exists\{q_1, q_2, ..., q_n\}$ *with the property*

$$U \subseteq \bigcup_{i=1}^{n} B_\epsilon(q_i)$$

2. ϵ-**packing number**: $P(U,\epsilon) = max\ m\ such\ that\ \exists\{q_1, q_2, ..., q_m\}$ *with the property* $d(q_i, q_j) > \epsilon,\ \forall i \neq j$

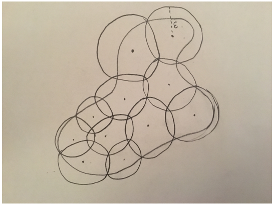

Figure: An ϵ-cover of our space U

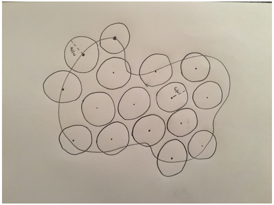

Figure: An ϵ-packing in our space V

An Upper Bound for the ϵ-Covering Number of a Unit Ball

Lemma (Upper Bound for Covering of a Unit Ball in $\mathbb{R}^k$)

If U is the unit ball in $\mathbb{R}^k$ and $\epsilon \in (0,1)$, $N(U,\epsilon) \leq (\frac{3}{\epsilon})^k$

An Upper Bound for the ϵ-Covering Number of a Unit Ball

Lemma (Upper Bound for Covering of a Unit Ball in $\mathbb{R}^k$)

If U is the unit ball in $\mathbb{R}^k$ and $\epsilon \in (0,1)$, $N(U,\epsilon) \leq (\frac{3}{\epsilon})^k$

It is known that $N(U,\epsilon) \leq P(U,\epsilon)$, so we bound the ϵ-packing number.

An Upper Bound for the ϵ-Covering Number of a Unit Ball

Lemma (Upper Bound for Covering of a Unit Ball in $\mathbb{R}^k$)

If U is the unit ball in $\mathbb{R}^k$ and $\epsilon \in (0,1)$, $N(U,\epsilon) \leq (\frac{3}{\epsilon})^k$

It is known that $N(U,\epsilon) \leq P(U,\epsilon)$, so we bound the ϵ-packing number. Let $\{q_1, q_2, ..., q_m\}$ be an optimal ϵ-packing. Hence

$$\bigcup_{i=1}^{m} B_{\frac{\epsilon}{2}}(q_i) \subseteq B_{1+\frac{\epsilon}{2}}(\mathbf{0})$$

An Upper Bound for the ϵ-Covering Number of a Unit Ball

Lemma (Upper Bound for Covering of a Unit Ball in $\mathbb{R}^k$)

If U is the unit ball in $\mathbb{R}^k$ and $\epsilon \in (0,1)$, $N(U,\epsilon) \leq (\frac{3}{\epsilon})^k$

It is known that $N(U,\epsilon) \leq P(U,\epsilon)$, so we bound the ϵ-packing number. Let $\{q_1, q_2, ..., q_m\}$ be an optimal ϵ-packing. Hence

$$\bigcup_{i=1}^{m} B_{\frac{\epsilon}{2}}(q_i) \subseteq B_{1+\frac{\epsilon}{2}}(\mathbf{0})$$

Since $d(q_i, q_j) > \epsilon$, $\forall i \neq j$, these balls are actually disjoint. Taking volume

$$Vol\left[\bigcup_{i=1}^{m} B_{\frac{\epsilon}{2}}(q_i)\right] = \sum_{i=1}^{m} Vol\left[B_{\frac{\epsilon}{2}}(q_i)\right] = mC_k\left(\frac{\epsilon}{2}\right)^k$$

$$\leq Vol\left[B_{1+\frac{\epsilon}{2}}(\mathbf{0})\right] = C_k\left(1+\frac{\epsilon}{2}\right)^k$$

An Upper Bound for the ϵ-Covering Number of a Unit Ball

Lemma (Upper Bound for Covering of a Unit Ball in $\mathbb{R}^k$)

If U is the unit ball in $\mathbb{R}^k$ and $\epsilon \in (0,1)$, $N(U,\epsilon) \leq (\frac{3}{\epsilon})^k$

It is known that $N(U,\epsilon) \leq P(U,\epsilon)$, so we bound the ϵ-packing number. Let $\{q_1, q_2, ..., q_m\}$ be an optimal ϵ-packing. Hence

$$\bigcup_{i=1}^{m} B_{\frac{\epsilon}{2}}(q_i) \subseteq B_{1+\frac{\epsilon}{2}}(\mathbf{0})$$

Since $d(q_i, q_j) > \epsilon, \ \forall i \neq j$, these balls are actually disjoint. Taking volume

$$Vol\left[\bigcup_{i=1}^{m} B_{\frac{\epsilon}{2}}(q_i)\right] = \sum_{i=1}^{m} Vol\left[B_{\frac{\epsilon}{2}}(q_i)\right] = mC_k\left(\frac{\epsilon}{2}\right)^k$$

$$\leq Vol\left[B_{1+\frac{\epsilon}{2}}(\mathbf{0})\right] = C_k\left(1+\frac{\epsilon}{2}\right)^k$$

Rearranging these results ultimately yields: $m \leq \left(\frac{2+\epsilon}{\epsilon}\right)^k \leq \left(\frac{3}{\epsilon}\right)^k$

The Restricted Isometry Property (RIP)

Definition

We say a matrix Φ *satisfies the RIP of order* k *if* $\exists \delta_k \in (0,1)$ *such that, for any* T *with* $\#(T) \leq k$

$$(1-\delta_k)\left\|\boldsymbol{x}_T\right\|_{\ell_2^N}^2 \leq \left\|\Phi \boldsymbol{x}_T\right\|_{\ell_2^n}^2 \leq (1+\delta_k)\left\|\boldsymbol{x}_T\right\|_{\ell_2^N}^2 \tag{2}$$

Our goal is now to show that a random matrix, Φ with entries from a subgaussian distribution, satisfies the RIP of order k with overwhelming probability.

Major Theorem (See also[1])

Theorem (Candes and Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?", 2006)

Given $\delta \in (0,1)$, $\exists c_1, c_2 \in \mathbb{R}^+$ *such that* Φ *satisfies the RIP of order* k *with probability greater than* $1 - 2e^{-c_2 n}$, *provided* $n \geq c_1 k \log\left(\frac{N}{k}\right)$

[1] Baraniuk et al., "A Simple Proof of the Restricted Isometry Property for Random Matrices".

Major Theorem (See also[1])

Theorem (Candes and Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?", 2006)

Given $\delta \in (0,1)$, $\exists c_1, c_2 \in \mathbb{R}^+$ *such that* Φ *satisfies the RIP of order* k *with probability greater than* $1 - 2e^{-c_2 n}$, *provided* $n \geq c_1 k \log\left(\frac{N}{k}\right)$

- Fix a k satisfying the hypothesis, let T be a set of indices with $\#(T) = k$. Note here that there are $\binom{N}{k}$ choices for T. Now consider $B_T := X_T \cap B_1(\mathbf{0})$

[1] Baraniuk et al., "A Simple Proof of the Restricted Isometry Property for Random Matrices".

Major Theorem (See also[1])

Theorem (Candes and Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?", 2006)

Given $\delta \in (0,1)$, $\exists c_1, c_2 \in \mathbb{R}^+$ such that Φ satisfies the RIP of order k with probability greater than $1 - 2e^{-c_2 n}$, provided $n \geq c_1 k \log\left(\frac{N}{k}\right)$

- Fix a k satisfying the hypothesis, let T be a set of indices with $\#(T) = k$. Note here that there are $\binom{N}{k}$ choices for T. Now consider $B_T := X_T \cap B_1(\mathbf{0})$
- By our Lemma on covering numbers, can pick $\frac{\delta}{4}$-covering

$$Q_T = \{\boldsymbol{q_1}, ..., \boldsymbol{q}_\alpha\} \text{ with } \alpha \leq \left(\frac{12}{\delta}\right)^k$$

[1] Baraniuk et al., "A Simple Proof of the Restricted Isometry Property for Random Matrices".

- Via concentration of measure, for a given $\boldsymbol{q}_i \in Q_T$:

$$\Pr\left[\left(1-\frac{\delta}{2}\right)\|\boldsymbol{q}_i\|_{\ell_2^N} \leq \|\Phi\boldsymbol{q}_i\|_{\ell_2^n} \leq \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}_i\|_{\ell_2^N} \text{ fails}\right] \leq 2e^{-nc_0\left(\frac{\delta}{2}\right)}$$

- Via concentration of measure, for a given $\boldsymbol{q}_i \in Q_T$:

$$\Pr\left[\left(1-\frac{\delta}{2}\right)\|\boldsymbol{q}_i\|_{\ell_2^N} \leq \|\Phi\boldsymbol{q}_i\|_{\ell_2^n} \leq \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}_i\|_{\ell_2^N} \text{ fails}\right] \leq 2e^{-nc_0\left(\frac{\delta}{2}\right)}$$

- Using a union bound on our finite number of points:

$$\Pr\left[\left(1-\frac{\delta}{2}\right)\|\boldsymbol{q}\|_{\ell_2^N} \leq \|\Phi\boldsymbol{q}\|_{\ell_2^n} \leq \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}\|_{\ell_2^N} \text{ fails for a } \boldsymbol{q} \in Q_T\right]$$
$$\leq 2\left(\frac{12}{\delta}\right)^k e^{-nc_0\left(\frac{\delta}{2}\right)}$$

- Via concentration of measure, for a given $\boldsymbol{q}_i \in Q_T$:

$$\Pr\left[\left(1-\frac{\delta}{2}\right)\|\boldsymbol{q}_i\|_{\ell_2^N} \le \|\Phi\boldsymbol{q}_i\|_{\ell_2^n} \le \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}_i\|_{\ell_2^N} \text{ fails}\right] \le 2e^{-nc_0\left(\frac{\delta}{2}\right)}$$

- Using a union bound on our finite number of points:

$$\Pr\left[\left(1-\frac{\delta}{2}\right)\|\boldsymbol{q}\|_{\ell_2^N} \le \|\Phi\boldsymbol{q}\|_{\ell_2^n} \le \left(1+\frac{\delta}{2}\right)\|\boldsymbol{q}\|_{\ell_2^N} \text{ fails for a } \boldsymbol{q} \in Q_T\right]$$
$$\le 2\left(\frac{12}{\delta}\right)^k e^{-nc_0\left(\frac{\delta}{2}\right)}$$

- From this, one can show:

$$\Pr\left[(1-\delta)\|\boldsymbol{x}\|_{\ell_2^N}^2 \le \|\Phi\boldsymbol{x}\|_{\ell_2^n}^2 \le (1+\delta)\|\boldsymbol{x}\|_{\ell_2^N}^2 \text{ fails for an } \boldsymbol{x} \in X_T\right]$$
$$\le 2\left(\frac{12}{\delta}\right)^k e^{-nc_0\left(\frac{\delta}{2}\right)}$$

- Since there are $\binom{N}{k} \leq \left(\frac{eN}{k}\right)^k$ such k-dimensional subspaces, use a union bound to see

$$\Pr\left[(1-\delta)\|\boldsymbol{x}\|_{\ell_2^N}^2 \leq \|\Phi\boldsymbol{x}\|_{\ell_2^n}^2 \leq (1+\delta)\|\boldsymbol{x}\|_{\ell_2^N}^2 \text{ fails for an } \boldsymbol{x} \in \bigcup_T X_T\right]$$

$$\leq 2\left(\frac{12}{\delta}\right)^k \left(\frac{eN}{k}\right)^k e^{-nc_0\left(\frac{\delta}{2}\right)} = 2e^{-nc_0\left(\frac{\delta}{2}\right)+k\left[\log\left(\frac{eN}{k}\right)+\log\left(\frac{12}{\delta}\right)\right]}$$

- Since there are $\binom{N}{k} \leq \left(\frac{eN}{k}\right)^k$ such k-dimensional subspaces, use a union bound to see

$$\Pr\left[(1-\delta)\,\|\boldsymbol{x}\|_{\ell_2^N}^2 \leq \|\Phi\boldsymbol{x}\|_{\ell_2^n}^2 \leq (1+\delta)\,\|\boldsymbol{x}\|_{\ell_2^N}^2 \text{ fails for an } \boldsymbol{x} \in \bigcup_T X_T\right]$$

$$\leq 2\left(\frac{12}{\delta}\right)^k \left(\frac{eN}{k}\right)^k e^{-nc_0\left(\frac{\delta}{2}\right)} = 2e^{-nc_0\left(\frac{\delta}{2}\right)+k\left[\log\left(\frac{eN}{k}\right)+\log\left(\frac{12}{\delta}\right)\right]}$$

- By picking some constant c_2 and taking the complement,

$$\Pr\left[\forall \boldsymbol{x} \in \bigcup_T X_T,\ (1-\delta)\,\|\boldsymbol{x}\|_{\ell_2^N}^2 \leq \|\Phi\boldsymbol{x}\|_{\ell_2^n}^2 \leq (1+\delta)\,\|\boldsymbol{x}\|_{\ell_2^N}^2\right]$$

$$\geq 1 - 2e^{-nc_2}$$

Which is to say: Φ satisfies the RIP of order k specified in the theorem with at least the above probability.

Summary

- Want to compress a high-sparse vector, in a way where we can recover it with little-to-no error.
- **How do we compress it?**
- Compressing with a subgaussian matrix Φ allows for unique ℓ_1-minimization; an accurate recovery
- **How small can we compress?**
- We can pick n logarithmic to N.

References

Baraniuk, Richard et al. "A Simple Proof of the Restricted Isometry Property for Random Matrices". In: *Constructive Approximation* 28.3 (Dec. 2008), pp. 253–263. ISSN: 1432-0940. DOI: 10.1007/s00365-007-9003-x. URL: https://doi.org/10.1007/s00365-007-9003-x.

Candes, Emmanuel J and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215.

– . "Near-optimal signal recovery from random projections: Universal encoding strategies?". In: *IEEE transactions on information theory* 52.12 (2006), pp. 5406–5425.