

Foundations of Compressed Sensing

Ronald DeVore

Usual Paradigm for Signal Processing

- Model signals as bandlimited functions $x(t)$
- Support of \hat{x} contained in $[-\Omega\pi, \Omega\pi]$
- Shannon-Nyquist
- Uniform time samples with spacing $h \leq \frac{1}{\Omega}$ allows for exact reconstruction
- A/D Converters: sample and quantize
- **Problem:** If Ω is too large we cant build circuitry to sample faithfully at the desired rate

Compressed Sensing

- **Compressed Sensing** seeks a way out of this dilemma
- Built on two new ingredients
- New model classes for signals
- Signals are **sparse** in some representation system: basis (frame)
- New meaning of samples
- Sample is a linear functional applied to the signal
- We shall consider only discrete signals
- Real world signal are typically **analog**

Discrete Compressed Sensing

- $x \in \mathbb{R}^N$ with N large

Discrete Compressed Sensing

- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x

Discrete Compressed Sensing

- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question (sample) is an inner product $v \cdot x$, $v \in \mathbb{R}^N$

Discrete Compressed Sensing

- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question (sample) is an inner product $v \cdot x$, $v \in \mathbb{R}^N$
- Any such sampling is given by is an $n \times N$ matrix Φ : the entries in $y = \Phi x$ are the answers to our questions

Discrete Compressed Sensing

- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question (sample) is an inner product $v \cdot x$, $v \in \mathbb{R}^N$
- Any such sampling is given by is an $n \times N$ matrix Φ : the entries in $y = \Phi x$ are the answers to our questions
- We are interested in the **good / best** matrices Φ , i.e. **what are the best questions to ask??**

Discrete Compressed Sensing

- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question (sample) is an inner product $v \cdot x$, $v \in \mathbb{R}^N$
- Any such sampling is given by is an $n \times N$ matrix Φ : the entries in $y = \Phi x$ are the answers to our questions
- We are interested in the **good / best** matrices Φ , i.e. **what are the best questions to ask??**
- Two issues: (i) Enough information in y to determine x ;
(ii) How to extract the information y holds about x :
Decoder

What is going on?

- Since $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^n$ many x give the same measurements y

What is going on?

- Since $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^n$ many x give the same measurements y
- $\mathcal{N} := \{\eta : \Phi\eta = 0\}$ the null space of Φ $\dim(\mathcal{N}) \geq N - n$

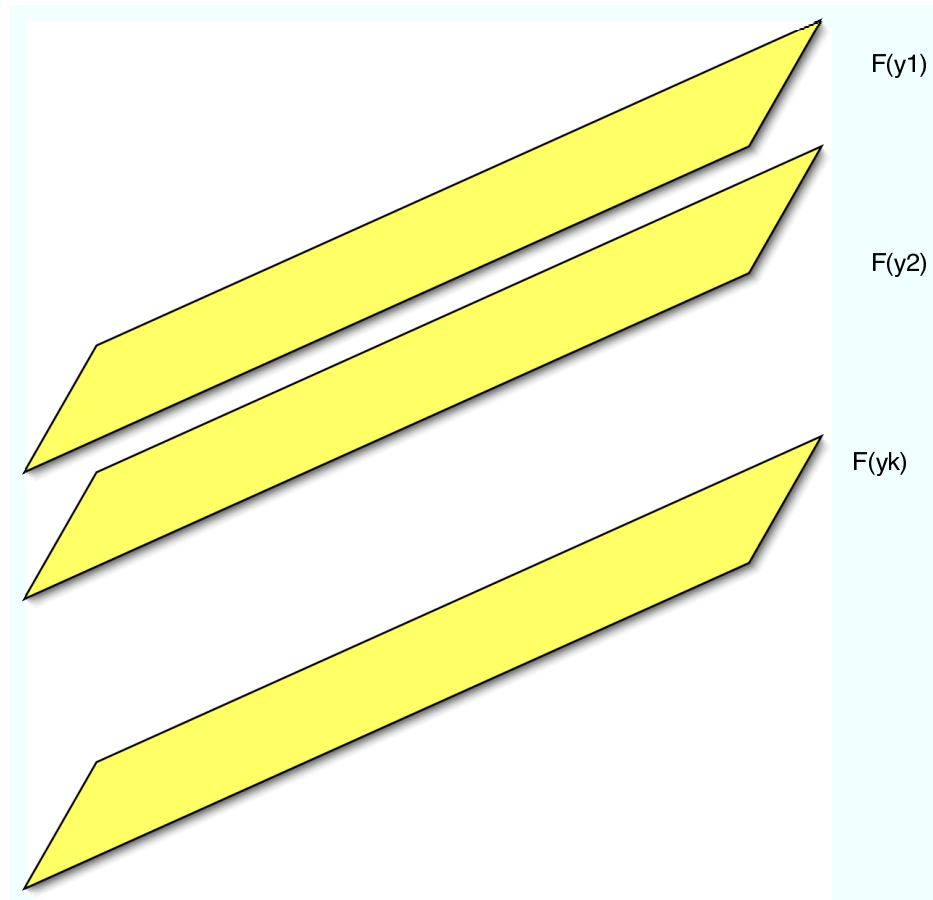
What is going on?

- Since $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^n$ many x give the same measurements y
- $\mathcal{N} := \{\eta : \Phi\eta = 0\}$ the null space of Φ $\dim(\mathcal{N}) \geq N - n$
- $\mathcal{F}(y) := \{x : \Phi x = y\} = x_0 + \mathcal{N}$ for any $x_0 \in \mathcal{F}(y)$

What is going on?

- Since $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^n$ many x give the same measurements y
- $\mathcal{N} := \{\eta : \Phi\eta = 0\}$ the null space of Φ $\dim(\mathcal{N}) \geq N - n$
- $\mathcal{F}(y) := \{x : \Phi x = y\} = x_0 + \mathcal{N}$ for any $x_0 \in \mathcal{F}(y)$
- The hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^n$ stratify \mathbb{R}^N

The sets $\mathcal{F}(y)$



What is going on?

- Since $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^n$ many x are encoded with same y
- $\mathcal{N} := \{\eta : \Phi\eta = 0\}$ the null space of Φ
- $\mathcal{F}(y) := \{x : \Phi x = y\} = x_0 + \mathcal{N}$ for any $x_0 \in \mathcal{F}(y)$
- The hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^n$ stratify \mathbb{R}^N
- **Decoder** is any (possibly nonlinear) mapping Δ from $\mathbb{R}^n \rightarrow \mathbb{R}^N$

What is going on?

- Since $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^n$ many x are encoded with same y
- $\mathcal{N} := \{\eta : \Phi\eta = 0\}$ the null space of Φ
- $\mathcal{F}(y) := \{x : \Phi x = y\} = x_0 + \mathcal{N}$ for any $x_0 \in \mathcal{F}(y)$
- The hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^n$ stratify \mathbb{R}^N
- **Decoder** is any (possibly nonlinear) mapping Δ from $\mathbb{R}^n \rightarrow \mathbb{R}^N$
- $\bar{x} := \Delta(\Phi(x))$ is our approximation to x from the information extracted

What is going on?

- Since $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^n$ many x are encoded with same y
- $\mathcal{N} := \{\eta : \Phi\eta = 0\}$ the null space of Φ
- $\mathcal{F}(y) := \{x : \Phi x = y\} = x_0 + \mathcal{N}$ for any $x_0 \in \mathcal{F}(y)$
- The hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^n$ stratify \mathbb{R}^N
- **Decoder** is any (possibly nonlinear) mapping Δ from $\mathbb{R}^n \rightarrow \mathbb{R}^N$
- $\bar{x} := \Delta(\Phi(x))$ is our approximation to x from the information extracted
- Note that all $x \in \mathcal{F}(y)$ are approximated by the same \bar{x}

Measuring Sparsity

- Compressed Sensing models signals as sparse in some basis
- By linear transformation, we can assume **WOLOG** that x is sparse with respect to the canonical basis on \mathbb{R}^N
- The support of x is $\text{supp}(x) := \{i : x_i \neq 0\}$
- $\Sigma_k := \{x : \#\text{supp}(x) \leq k\}$
- Note that Σ_k is a union of k dimensional subspaces:
 $\Sigma_k = \cup_{\#(T)=k} X_T$ where $X_T = \{x : \text{supp}(x) \subset T\}$
- **First Question:** Given k, N what is the smallest n for which there is (Φ, Δ) such each vector in Σ_k is captured exactly $\Delta(\Phi(x)) = x, \quad x \in \Sigma_k$

Measuring Sparsity

- Compressed Sensing models signals as sparse in some basis
- By linear transformation, we can assume **WOLOG** that x is sparse with respect to the canonical basis on \mathbb{R}^N
- The support of x is $\text{supp}(x) := \{i : x_i \neq 0\}$
- $\Sigma_k := \{x : \#\text{supp}(x) \leq k\}$
- Note that Σ_k is a union of k dimensional subspaces:
 $\Sigma_k = \cup_{\#(T)=k} X_T$ where $X_T = \{x : \text{supp}(x) \subset T\}$
- **First Question:** Given k, N what is the smallest n for which there is (Φ, Δ) such each vector in Σ_k is captured exactly $\Delta(\Phi(x)) = x, \quad x \in \Sigma_k$
- Answer $n = 2k$

What matrices do the job?

- $\Phi = [v_1, \dots, v_N]$, v_1, \dots, v_N columns of Φ
- We say Φ has the independence property (IP) of order k if all choices of k column vectors are independent
- If $T = \{i_1, \dots, i_m\}$ is a set of column indices
- $\Phi_T = [v_{i_1}, \dots, v_{i_m}]$ is the $n \times \#(T)$ submatrix of Φ formed from these columns
- IP means $\Phi_T^* \Phi_T := (\langle v_i, v_j \rangle)_{i,j \in T}$ is invertible (positive eigenvalues) whenever $\#(T) = k$

What matrices do the job?

- $\Phi = [v_1, \dots, v_n]$, v_1, \dots, v_N columns of Φ
- We say Φ has the independence property (IP) of order k if all choices of k column vectors are independent
- If $T = \{i_1, \dots, i_m\}$ is a set of column indices
- $\Phi_T = [v_{i_1}, \dots, v_{i_m}]$ is the $n \times \#(T)$ submatrix of Φ formed from these columns
- IP means $\Phi_T^* \Phi_T := (\langle v_i, v_j \rangle)_{i,j \in T}$ is invertible (positive eigenvalues) whenever $\#(T) = k$

Theorem: If Φ is any $n \times N$ matrix and $2k \leq n$, then the following are equivalent:

- There is a Δ such that $\Delta(\Phi(x)) = x$, for all $x \in \Sigma_k$,
- $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{0\}$,
- the matrix Φ_T has the independence property of order $2k$.

Optimal Matrices

- Given k can we construct matrices Φ of size $2k \times N$ with the properties of the theorem?

Optimal Matrices

- Given k can we construct matrices Φ of size $2k \times N$ with the properties of the theorem?
- We need N vectors in \mathbb{R}^{2k} such that any $2k$ of them are linearly independent

Optimal Matrices

- Given k can we construct matrices Φ of size $2k \times N$ with the properties of the theorem?
- We need N vectors in \mathbb{R}^{2k} such that any $2k$ of them are linearly independent
- Vandermonde matrix. Choose $x_1 < x_2 < \dots < x_N$

$$\Phi := \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_1^{2k-1} & x_2^{2k-1} & \dots & x_N^{2k-1} \end{pmatrix}$$

Naive Decoding

$$\Delta(y) := \underset{z \in \Sigma_k}{\text{Argmin}} \|y - \Phi(z)\|_{\ell_2^n}$$

- $X_T := \{z : \text{supp}(z) \subset T\}$
- $x_T := \underset{z \in X_T}{\text{Argmin}} \|y - \Phi z\|_{\ell_2^n} \rightarrow x_T = [\Phi_T^* \Phi_T]^{-1} \Phi_T y$
- $T^* := \underset{\#(T)=k}{\text{Argmin}} \|y - \Phi(x_T)\|_{\ell_2^n}$
- $\Delta(y) := x_{T^*}$

We Have a Problem

- Have we solved our first problem?

We Have a Problem

- Have we solved our first problem?
- None of us will be alive when the decoding is finished

We Have a Problem

- Have we solved our first problem?
- None of us will be alive when the decoding is finished
- Moreover, the decoding is also unstable

We Have a Problem

- Have we solved our first problem?
- None of us will be alive when the decoding is finished
- Moreover, the decoding is also unstable
- The first problem has an easy fix. We can take the first $2k$ rows of the discrete Fourier matrix and build a decoder which uses only $O(N + k^3)$ operations:

We Have a Problem

- Have we solved our first problem?
- None of us will be alive when the decoding is finished
- Moreover, the decoding is also unstable
- The first problem has an easy fix. We can take the first $2k$ rows of the discrete Fourier matrix and build a decoder which uses only $O(N + k^3)$ operations:
- However the stability problem is more substantial - no quick fix

We Have a Problem

- Have we solved our first problem?
- None of us will be alive when the decoding is finished
- Moreover, the decoding is also unstable
- The first problem has an easy fix. We can take the first $2k$ rows of the discrete Fourier matrix and build a decoder which uses only $O(N + k^3)$ operations:
- However the stability problem is more substantial - no quick fix
- Suppose we had any matrix Φ and we knew the support T of x then $x = x_T = [\Phi_T^* \Phi_T]^{-1} \Phi_T y$

We Have a Problem

- Have we solved our first problem?
- None of us will be alive when the decoding is finished
- Moreover, the decoding is also unstable
- The first problem has an easy fix. We can take the first $2k$ rows of the discrete Fourier matrix and build a decoder which uses only $O(N + k^3)$ operations:
- However the stability problem is more substantial - no quick fix
- Suppose we had any matrix Φ and we knew the support T of x then $x = x_T = [\Phi_T^* \Phi_T]^{-1} \Phi_T y$
- Need $\|[\Phi_T^* \Phi_T]^{-1}\|$ controlled

We Have a Problem

- Have we solved our first problem?
- None of us will be alive when the decoding is finished
- Moreover, the decoding is also unstable
- The first problem has an easy fix. We can take the first $2k$ rows of the discrete Fourier matrix and build a decoder which uses only $O(N + k^3)$ operations:
- However the stability problem is more substantial - no quick fix
- Suppose we had any matrix Φ and we knew the support T of x then $x = x_T = [\Phi_T^* \Phi_T]^{-1} \Phi_T y$
- Need $\|[\Phi_T^* \Phi_T]^{-1}\|$ controlled
- We would need this norm controlled for any T of size k

Optimal Stable Systems

- Candes-Romberg-Tao; Donoho: Compressed Sensing
- Two important discoveries: good matrices, decoding
- **Restricted Isometry Property (RIP)** of order k : There exists $0 < \delta = \delta_k < 1$ such that

$$(1 - \delta)\|x\|_{\ell_2^N}^2 \leq \|\Phi(x)\|_{\ell_2^n}^2 \leq (1 + \delta)\|x\|_{\ell_2^N}^2, \quad x \in \Sigma_k$$

- Equivalently the eigenvalues of $\Phi_T^* \Phi_T$ are in $[1 - \delta, 1 + \delta]$
- Decode by ℓ_1 minimization

$$\Delta(y) := \inf_{x \in \mathcal{F}(y)} \|x\|_{\ell_1^N}$$

- **Candes-Tao:** If Φ satisfies the **RIP** of order $3k$ then given any $x \in \Sigma_k$ we have $\Delta(\Phi(x)) = x$ for the ℓ_1 minimization decoder. Moreover, the decoding is stable

Building matrices

- How can we build matrices that satisfy RIP for the largest value of k
- Given n, N we can construct such matrices for any $k \leq c_0 n / \log(N/n)$
- The additional $\log(N/n)$ is the price we pay for stability
- This is the largest possible range of k
- How can we construct such Φ ?
- We want to create a lot of vectors v_1, \dots, v_N in \mathbb{R}^n such that any choice of k of them are far from being linearly dependent

Three constructions

- We choose at random N vectors from the unit sphere in \mathbb{R}^n and use these as the columns of Φ
- We choose each entry of Φ independently from the Gaussian distribution with mean 0 and variance n^{-1}
- We use Bernoulli process and create a matrix with entries $\frac{\pm 1}{\sqrt{n}}$, with equal probability
- With high probability each of these random constructions yields a matrix Φ with RIP of order k for the (largest) range $k \leq c_0 n / \log(N/n)$

Three constructions

- We choose at random N vectors from the unit sphere in \mathbb{R}^n and use these as the columns of Φ
- We choose each entry of Φ independently from the Gaussian distribution with mean 0 and variance n^{-1}
- We use Bernoulli process and create a matrix with entries $\frac{\pm 1}{\sqrt{n}}$, with equal probability
- With high probability each of these random constructions yields a matrix Φ with RIP of order k for the (largest) range $k \leq c_0 n / \log(N/n)$
- Probability is only used to prove existence of Φ . The sensing algorithm is constructive (not probabilistic) once we find a Φ .

Three constructions

- We choose at random N vectors from the unit sphere in \mathbb{R}^n and use these as the columns of Φ
- We choose each entry of Φ independently from the Gaussian distribution with mean 0 and variance n^{-1}
- We use Bernoulli process and create a matrix with entries $\frac{\pm 1}{\sqrt{n}}$, with equal probability
- With high probability each of these random constructions yields a matrix Φ with RIP of order k for the (largest) range $k \leq c_0 n / \log(N/n)$
- Probability is only used to prove existence of Φ . The sensing algorithm is constructive (not probabilistic) once we find a Φ .
- Geometry of Banach spaces: Kashin, Johnson-Lindenstrauss

Verification of RIP

- It is difficult (NP hard) to check whether a matrix satisfies **RIP**
- It is much easier to show that a random family $\Phi(\omega)$, $\omega \in \Omega$, of matrices has RIP with high probability
- Do not know specifically which matrix is favorable
- There is a simple strategy for verification of **RIP** for random families (**Baraniuk-Davenport-DeVore-Wakin**)
- Concentration of Measure Property **CMP**

$$\text{Prob}(|\|\Phi(\omega)x\|_{\ell_2^n}^2 - \|x\|_{\ell_2^N}^2| \geq \delta \|x\|_{\ell_2^N}^2) \leq C e^{-c(\delta)n}$$

- **CMP** is like RIP but only for one vector $x \in \mathbb{R}^N$
- Use **CMP** with ϵ nets for Σ_n

General Signals

- So far the results about performance apply only to sparse signals
- Can these results be extended to general signals: **How should such results be formulated?**
- We say (Φ, Δ) is **Instance-Optimal** of order k for X if for an absolute constant $C > 0$ (independent of k, n, N)

$$\|x - \Delta(\Phi(x))\|_X \leq C\sigma_k(x)_X$$

- Problem: for a given X and size $n \times N$ find the largest values of k for which we have instance-optimality and the encoder-decoder pairs (Φ, Δ) which admit these values of k
- **Cohen-Dahmen-DeVore** solve the instance-optimal problem for measuring error in $X = \ell_q^N$ for all $1 \leq q \leq 2$

Good News

- Let $X = \ell_1^N$ and let Φ satisfy **RIP** for $3k$, i.e. $\delta_{3k} < 1$ then there is a decoder such that (Φ, Δ) is instance optimal for k :

$$\|x - \Delta(\Phi(x))\|_{\ell_1^N} \leq C_0 \sigma_k(x)_{\ell_1^N}$$

- Given n we can have instance optimality if $k \leq c_0 n / \log(N/n)$
- This is provably the largest range of k
- Bonus: Decoding can be done by ℓ_1 **Minimization**

Bad News

- Let $X = \ell_2^N$, then in order to have instance optimality for $k = 1$ we need $n \geq c_0 N$

Bad News

- Let $X = \ell_2^N$, then in order to have instance optimality for $k = 1$ we need $n \geq c_0 N$
- Here c_0 depends on the instance optimality constant C

Bad News

- Let $X = \ell_2^N$, then in order to have instance optimality for $k = 1$ we need $n \geq c_0 N$
- Here c_0 depends on the instance optimality constant C
- OOPS: Instance Optimality is not a Viable Concept in ℓ_2^N

Bad News

- Let $X = \ell_2^N$, then in order to have instance optimality for $k = 1$ we need $n \geq c_0 N$
- Here c_0 depends on the instance optimality constant C
- OOPS: Instance Optimality is not a Viable Concept in ℓ_2^N
- Never give up hope

Bad News

- Let $X = \ell_2^N$, then in order to have instance optimality for $k = 1$ we need $n \geq c_0 N$
- Here c_0 depends on the instance optimality constant C
- OOPS: Instance Optimality is not a Viable Concept in ℓ_2^N
- Never give up hope
- Since we can never construct matrices satisfying RIP for the large without using probability then why not require the compressed sensing system to perform **only with high probability - not with certainty?**

Instance-Optimality in Probability

- Let $(\Phi(\omega), \Delta(\omega)), \omega \in \Omega$, be a random family of CS matrices and corresponding decoders

Instance-Optimality in Probability

- Let $(\Phi(\omega), \Delta(\omega)), \omega \in \Omega$, be a random family of CS matrices and corresponding decoders
- We say this family is **Instance Optimal in Probability (IOP)** of order k if for each $x \in \mathbb{R}^N$ with high probability on the draw $\omega \in \Omega$ we have

$$\|x - \Delta(\Phi x)\|_{\ell_2^N} \leq C_0 \sigma_k(x)_{\ell_2^N}$$

Instance-Optimality in Probability

- Let $(\Phi(\omega), \Delta(\omega)), \omega \in \Omega$, be a random family of CS matrices and corresponding decoders
- We say this family is **Instance Optimal in Probability (IOP)** of order k if for each $x \in \mathbb{R}^N$ with high probability on the draw $\omega \in \Omega$ we have

$$\|x - \Delta(\Phi x)\|_{\ell_2^N} \leq C_0 \sigma_k(x)_{\ell_2^N}$$

- This is the ℓ_2 accuracy we want

Instance-Optimality in Probability

- Let $(\Phi(\omega), \Delta(\omega)), \omega \in \Omega$, be a random family of CS matrices and corresponding decoders
- We say this family is **Instance Optimal in Probability (IOP)** of order k if for each $x \in \mathbb{R}^N$ with high probability on the draw $\omega \in \Omega$ we have

$$\|x - \Delta(\Phi x)\|_{\ell_2^N} \leq C_0 \sigma_k(x)_{\ell_2^N}$$

- This is the ℓ_2 accuracy we want
- However this is not guaranteed with certainty only with high probability

Instance-Optimality in Probability

- Let $(\Phi(\omega), \Delta(\omega)), \omega \in \Omega$, be a random family of CS matrices and corresponding decoders
- We say this family is **Instance Optimal in Probability (IOP)** of order k if for each $x \in \mathbb{R}^N$ with high probability on the draw $\omega \in \Omega$ we have

$$\|x - \Delta(\Phi x)\|_{\ell_2^N} \leq C_0 \sigma_k(x)_{\ell_2^N}$$

- This is the ℓ_2 accuracy we want
- However this is not guaranteed with certainty only with high probability
- Notice the probability is on the draw $\omega \in \Omega$ and not on x

Theorem: Cohen-Dahmen-DeVore

Theorem (CDD) Suppose $\Phi(\omega)$, $\omega \in \Omega$ has **CMP**. Then there are decoders $\Delta(\omega)$ such that $(\Phi(\omega), \Delta(\omega))$, $\omega \in \Omega$, is **Instance Optimal in Probability** for $k \leq c_0 n / \log(N/n)$

- General random families such as Gaussian and Bernoulli random matrices have **CMP** and will therefore satisfy the above theorem with probability of failure e^{-cn}
- Unfortunately the decoder is a naive impractical decoder

Decoding

- By far the most intriguing part of Compressed Sensing

Decoding

- By far the most intriguing part of Compressed Sensing
- There are continuing debates as to which decoding is numerically fastest

Decoding

- By far the most intriguing part of Compressed Sensing
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders

Decoding

- By far the most intriguing part of Compressed Sensing
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
 - ℓ_1 minimization: Long history in statistics (Donoho; Candes-Romberg in CS)

Decoding

- By far the most intriguing part of Compressed Sensing
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
 - ℓ_1 minimization: Long history in statistics (Donoho; Candes-Romberg in CS)
 - Greedy algorithms - find support of a good approximation vector and then decode using ℓ_2 minimization (Gilbert-Tropp; Needell-Vershynin; Needell-Tropp; Cohen-Dahmen-DeVore)

Decoding

- By far the most intriguing part of Compressed Sensing
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
 - ℓ_1 minimization: Long history in statistics (Donoho; Candes-Romberg in CS)
 - Greedy algorithms - find support of a good approximation vector and then decode using ℓ_2 minimization (Gilbert-Tropp; Needell-Vershynin; Needell-Tropp; Cohen-Dahmen-DeVore)
 - Iterative Reweighted Least Squares (Osborne, Lawson, Chartrand-Staneva, Daubechies-DeVore-Fornasier-Gunturk)

Decoding

- By far the most intriguing part of Compressed Sensing
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
 - ℓ_1 minimization: Long history in statistics (Donoho; Candes-Romberg in CS)
 - Greedy algorithms - find support of a good approximation vector and then decode using ℓ_2 minimization (Gilbert-Tropp; Needell-Vershynin; Needell-Tropp; Cohen-Dahmen-DeVore)
 - Iterative Reweighted Least Squares (Osborne, Lawson, Chartrand-Staneva, Daubechies-DeVore-Fornasier-Gunturk)
- More on decoding in the Special Session later today

Performance of ℓ_1 -minimization

- For the Gaussian random family, Przemek Wojtaszczyk proved ℓ_1 -minimization gives instance-optimality in probability for the large range $k \leq cn / \log(N/n)$

Performance of ℓ_1 -minimization

- For the Gaussian random family, Przemek Wojtaszczyk proved ℓ_1 -minimization gives instance-optimality in probability for the large range $k \leq cn / \log(N/n)$
- His proof depends on the following geometrical property of random Gaussian $n \times N$ matrices: with high probability the image of the unit ball in ℓ_1 contains an ℓ_2 ball of radius $\frac{c\sqrt{\log(N/n)}}{\sqrt{n}}$

Performance of ℓ_1 -minimization

- For the Gaussian random family, Przemek Wojtaszczyk proved ℓ_1 -minimization gives instance-optimality in probability for the large range $k \leq cn / \log(N/n)$
- His proof depends on the following geometrical property of random Gaussian $n \times N$ matrices: with high probability the image of the unit ball in ℓ_1 contains an ℓ_2 ball of radius $\frac{c\sqrt{\log(N/n)}}{\sqrt{n}}$
- This property does not hold for more general random constructions, e.g. Bernouli

Performance of ℓ_1 -minimization

- For the Gaussian random family, Przemek Wojtaszczyk proved ℓ_1 -minimization gives instance-optimality in probability for the large range $k \leq cn / \log(N/n)$
- His proof depends on the following geometrical property of random Gaussian $n \times N$ matrices: with high probability the image of the unit ball in ℓ_1 contains an ℓ_2 ball of radius $\frac{c\sqrt{\log(N/n)}}{\sqrt{n}}$
- This property does not hold for more general random constructions, e.g. Bernouli
- DeVore-Petrova-Wojtaszczyk prove instance-optimality for more general random matrices (including Bernouli) for the large range $k \leq cn / \log(N/n)$ by introducing new geometric properties

New Geometry

- Let $\Phi(\omega)$ be a family of random matrices whose entries are given by random draws of a real random variable ϕ with expectation 0 and variance $1/n$

New Geometry

- Let $\Phi(\omega)$ be a family of random matrices whose entries are given by random draws of a real random variable ϕ with expectation 0 and variance $1/n$
- Suppose $\Phi(\omega)$ satisfies the concentration of measure property for all n, N with constants depending only on δ , e.g. Gaussian or Bernoulli

New Geometry

- Let $\Phi(\omega)$ be a family of random matrices whose entries are given by random draws of a real random variable ϕ with expectation 0 and variance $1/n$
- Suppose $\Phi(\omega)$ satisfies the concentration of measure property for all n, N with constants depending only on δ , e.g. Gaussian or Bernoulli
- $\|y\|_J := \max\left\{ \sqrt{\log(N/n)} \|y\|_{\ell_\infty(\mathbb{R}^n)}, \|y\|_{\ell_2(\mathbb{R}^n)} \right\}$

New Geometry

- Let $\Phi(\omega)$ be a family of random matrices whose entries are given by random draws of a real random variable ϕ with expectation 0 and variance $1/n$
- Suppose $\Phi(\omega)$ satisfies the concentration of measure property for all n, N with constants depending only on δ , e.g. Gaussian or Bernoulli
- $\|y\|_J := \max\left\{ \sqrt{\log(N/n)} \|y\|_{\ell_\infty(\mathbb{R}^n)}, \|y\|_{\ell_2(\mathbb{R}^n)} \right\}$
- The unit ball U_J under this norm is a **clipped** ℓ_2 ball

New Geometry

- Let $\Phi(\omega)$ be a family of random matrices whose entries are given by random draws of a real random variable ϕ with expectation 0 and variance $1/n$
- Suppose $\Phi(\omega)$ satisfies the concentration of measure property for all n, N with constants depending only on δ , e.g. Gaussian or Bernoulli
- $\|y\|_J := \max\left(\left\{ \sqrt{\log(N/n)} \|y\|_{\ell_\infty(\mathbb{R}^n)}, \|y\|_{\ell_2(\mathbb{R}^n)} \right\}\right)$
- The unit ball U_J under this norm is a **clipped** ℓ_2 ball
- If $x \in U_J$ then $\|x\|_{\ell_2} \leq 1$ and $|x_j| \leq 1/\sqrt{\log(N/n)}$,
 $j = 1, \dots, N$

Random Mappings cover clipped balls

- If $x \in \mathbb{R}^N$ then with high probability on the draw of Φ :

$$\|\Phi x\|_J \leq C\|x\|_{\ell_2}$$

Random Mappings cover clipped balls

- If $x \in \mathbb{R}^N$ then with high probability on the draw of Φ :

$$\|\Phi x\|_J \leq C\|x\|_{\ell_2}$$

- With high probability on the draw $\Phi(\omega)$ we have

$$\Phi(U(\ell_1^N)) \supset \frac{c_0 \sqrt{\log(N/n)}}{\sqrt{n}} U_J$$

Random Mappings cover clipped balls

- If $x \in \mathbb{R}^N$ then with high probability on the draw of Φ :

$$\|\Phi x\|_J \leq C \|x\|_{\ell_2}$$

- With high probability on the draw $\Phi(\omega)$ we have

$$\Phi(U(\ell_1^N)) \supset \frac{c_0 \sqrt{\log(N/n)}}{\sqrt{n}} U_J$$

- Proved by Litvak, Pajor, Rudelson and Tomczak-Jaegermann 2005

Main Theorem of DPW

- Let $\Phi(\omega)$ be a family of random matrices whose entries are given by random draws of a real random variable ϕ with expectation 0 and variance $1/n$
- Suppose $\Phi(\omega)$ satisfies the concentration of measure property for all n, N with constants depending only on δ , e.g. Gaussian or Bernoulli
- Let $\Delta = \Delta(\omega)$ be the ℓ_1 -minimization decoder
- Given any $x \in \mathbb{R}^N$ then

$$\|x - \Delta(\Phi x)\|_{\ell_2^N} \leq C_0 \sigma_k(x)_{\ell_2^N}, \quad k \leq cn / \log(N/n)$$

with probability $\geq 1 - Cn e^{-c_0 n / \sqrt{\log(N/n)}}$