A Taste of Compressed Sensing

Ronald DeVore

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images
  - In image compression we need much fewer pieces of information (bits) than the number of pixel values

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images
  - In image compression we need much fewer pieces of information (bits) than the number of pixel values
  - Can we bypass pixel sampling, transformation and thresholding

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images
  - In image compression we need much fewer pieces of information (bits) than the number of pixel values
  - Can we bypass pixel sampling, transformation and thresholding
  - Make much fewer measurements of image

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images
  - In image compression we need much fewer pieces of information (bits) than the number of pixel values
  - Can we bypass pixel sampling, transformation and thresholding
  - Make much fewer measurements of image
- Signal Processing (Shannon-Nyquist Theory)

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images
  - In image compression we need much fewer pieces of information (bits) than the number of pixel values
  - Can we bypass pixel sampling, transformation and thresholding
  - Make much fewer measurements of image
- Signal Processing (Shannon-Nyquist Theory)
  - Bandlimited signals- Sample at Nyquist rate

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images
  - In image compression we need much fewer pieces of information (bits) than the number of pixel values
  - Can we bypass pixel sampling, transformation and thresholding
  - Make much fewer measurements of image
- Signal Processing (Shannon-Nyquist Theory)
  - Bandlimited signals- Sample at Nyquist rate
  - Not possible for Broadbanded signals

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images
  - In image compression we need much fewer pieces of information (bits) than the number of pixel values
  - Can we bypass pixel sampling, transformation and thresholding
  - Make much fewer measurements of image
- Signal Processing (Shannon-Nyquist Theory)
  - Bandlimited signals- Sample at Nyquist rate
  - Not possible for Broadbanded signals
  - Real world signals frequently have low information content

Goal of Compressed Sensing

- Acquire signal/image with fewest measurements
- Images
  - In image compression we need much fewer pieces of information (bits) than the number of pixel values
  - Can we bypass pixel sampling, transformation and thresholding
  - Make much fewer measurements of image
- Signal Processing (Shannon-Nyquist Theory)
  - Bandlimited signals- Sample at Nyquist rate
  - Not possible for Broadbanded signals
  - Real world signals frequently have low information content
  - Can we sample at close to information rate??

The Setting

- Compressed Sensing assumes the signal/image has a sparse representation in some basis

The Setting

- Compressed Sensing assumes the signal/image has a sparse representation in some basis
- Discrete Setting

The Setting

- Compressed Sensing assumes the signal/image has a sparse representation in some basis
- Discrete Setting
- $x \in I\!R^N$ with N large

The Setting

- Compressed Sensing assumes the signal/image has a sparse representation in some basis
- Discrete Setting
- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x

The Setting

- Compressed Sensing assumes the signal/image has a sparse representation in some basis
- Discrete Setting
- $x \in I\!R^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question (sample) is an inner product $v \cdot x$, $v \in I\!R^N$

The Setting

- Compressed Sensing assumes the signal/image has a sparse representation in some basis
- Discrete Setting
- $x \in \mathbb{R}^N$ 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

The Setting

- Compressed Sensing assumes the signal/image has a sparse representation in some basis
- Discrete Setting
- $x \in I\!R^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question (sample) is an inner product $v \cdot x$, $v \in I\!R^N$
- 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??

The Setting

- Compressed Sensing assumes the signal/image has a sparse representation in some basis
- Discrete Setting
- $x \in I\!R^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question (sample) is an inner product $v \cdot x$, $v \in I\!R^N$
- 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 this information: Decoder

What is going on?

- Since $\Phi : I\!R^N \to I\!R^n$ many x give the same measurements y

What is going on?

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

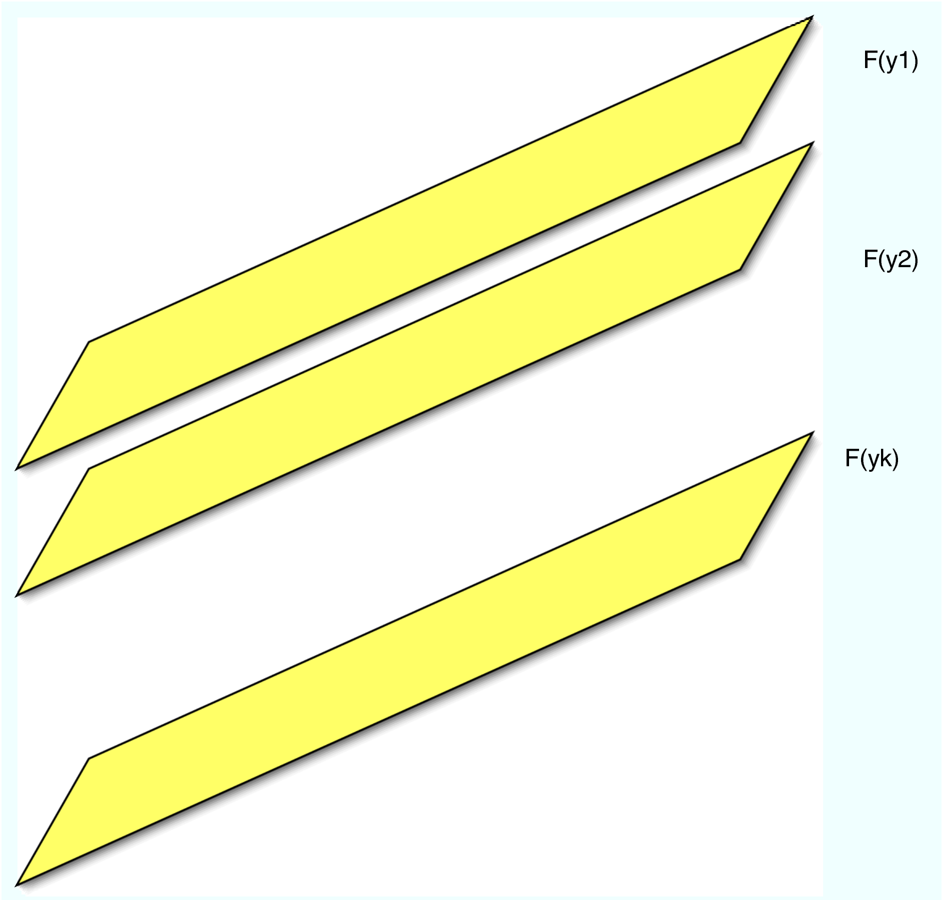
What is going on?

- Since $\Phi : I\!R^N \to I\!R^n$ 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 $\Phi : I\!R^N \to I\!R^n$ 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 $y \in I\!R^n$ stratify $I\!R^N$

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



What is going on?

- Since $\Phi : I\!R^N \to I\!R^n$ 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 $y \in I\!R^n$ stratify $I\!R^N$
- Decoder is any (possibly nonlinear) mapping Δ from $I\!R^n \to I\!R^N$

What is going on?

- Since $\Phi : I\!R^N \to I\!R^n$ 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 $y \in I\!R^n$ stratify $I\!R^N$
- Decoder is any (possibly nonlinear) mapping Δ from $I\!R^n \to I\!R^N$
- $\bar{x} := \Delta(\Phi(x))$ is our approximation to x from the information extracted

What is going on?

- Since $\Phi : I\!R^N \to I\!R^n$ 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 $y \in I\!R^n$ stratify $I\!R^N$
- Decoder is any (possibly nonlinear) mapping Δ from $I\!R^n \to I\!R^N$
- $\bar{x} := \Delta(\Phi(x))$ is our approximation to x from the information extracted
- Pessimism: all $x \in \mathcal{F}(y)$ are approximated by the same $\bar{x}$

Structure

- With no additional information about x it is doubtful what we can say
- Fortunately the x we are interested in have structure
- Namely we are assuming that x can be well represented by a sparse linear combination of certain building blocks - for our purposes these building blocks are a basis
- In some(many) problems we do not necessarily know the right basis. However the basis needs to be known for the decoding
- For the most part we shall assume the basis is known to us
- We will discuss three models for sparsity - progressively more demanding

First Model for Signals: Sparse

- To begin with we shall assume x is sparse with respect to the canonical basis on $\mathbb{R}^N$
- Any other basis could be handled by transformation (if the basis is known)
- The support of x is $\mathrm{supp(x)} := \{\mathrm{i} : \ \mathrm{x_i} \neq 0\}$
- $\Sigma_k := \{x : \ \#\mathrm{supp(x)} \leq \mathrm{k}\}$
- Note that Σ_k is a union of k dimensional subspaces: $\Sigma_k = \cup_{\#(T)=k} X_T$ where $X_T = \{x : \mathrm{supp(x)} \subset \mathrm{T}\}$
- 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 $\Delta(\Phi(x)) = x, \quad x \in \Sigma_k$

First Model for Signals: Sparse

- To begin with we shall assume x is sparse with respect to the canonical basis on $\mathbb{R}^N$
- Any other basis could be handled by transformation (if the basis is known)
- The support of x is $\mathrm{supp(x) := \{i : \ x_i \neq 0\}}$
- $\Sigma_k := \{x : \ \#\mathrm{supp(x) \leq k}\}$
- Note that Σ_k is a union of k dimensional subspaces: $\Sigma_k = \cup_{\#(T)=k} X_T$ where $X_T = \{x : \mathrm{supp(x) \subset T}\}$
- 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?

- $\Phi = [v_1, \ldots, v_N], \quad v_1, \ldots, v_N$ columns of Φ
- We say Φ has the independence property (IP) of order k if all choices of k column vectors are independent
- If $T = \{i_1, \ldots, i_m\}$ is a set of column indices
- $\Phi_T = [v_{i_1}, \ldots, v_{i_m}]$ 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?

- $\Phi = [v_1, \dots, v_n], \quad v_1, \dots, v_N$ 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:

(i) There is a Δ such that $\Delta(\Phi(x)) = x$, for all $x \in \Sigma_k$,
(ii) $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{0\}$,
(iii) 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 $x_1 < x_2 < \cdots < x_N$

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

Naive Decoding

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

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

Trouble in LA

- Have we solved our first problem?

Trouble in LA

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

Trouble in LA

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

Trouble in LA

- 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:

Trouble in LA

- 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

Trouble in LA

- 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 $x = x_T = [\Phi_T^* \Phi_T]^{-1} \Phi_T y$

Trouble in LA

- 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

Trouble in LA

- 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

Optimal Stable Systems

- Candes-Romberg-Tao; Donoho: Compressed Sensing
- Two important discoveries: good matrices, decoding

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 the eigenvalues of $\Phi_T^* \Phi_T$ are in $[1-\delta, 1+\delta]$ whenever $\#(T) = k$

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 the eigenvalues of $\Phi_T^*\Phi_T$ are in $[1-\delta, 1+\delta]$ whenever $\#(T) = k$
- Equivalently:

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

Optimal Stable Systems

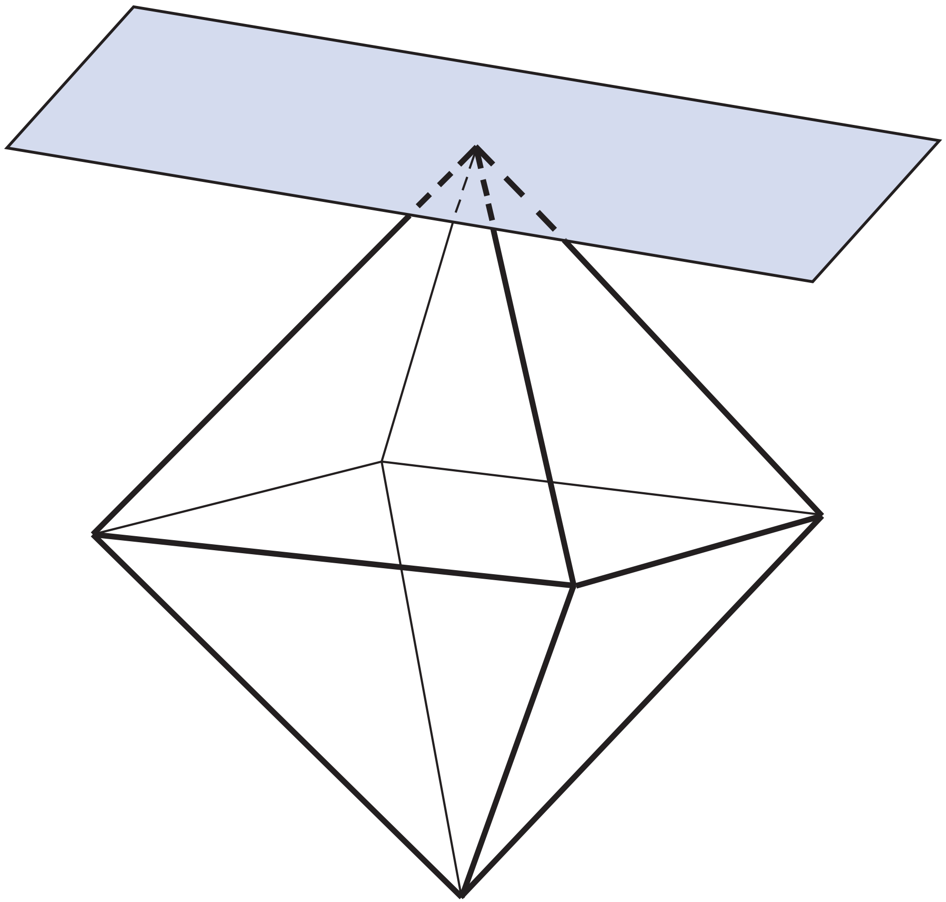
- Candes-Romberg-Tao; Donoho: Compressed Sensing
- Two important discoveries: good matrices, decoding
- Restricted Isometry Property (RIP) of order k: There exists $0 < \delta = \delta_k < 1$ such that the eigenvalues of $\Phi_T^*\Phi_T$ are in $[1-\delta, 1+\delta]$ whenever $\#(T) = k$
- Equivalently:

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

- Decode by ℓ_1 minimization

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

ℓ_1 ball meets the set $\mathcal{F}(y)$



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

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

- 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
- How can we construct such Φ?
- We want to create a lot of vectors $v_1, \ldots, v_N$ 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 and at random from the Gaussian distribution with mean 0 and variance n^{-1}
- We use Bernouli process and create a matrix with entries $-1, 1$ (or $0, 1$)
- With high probability each of these random constructions yields a matrix Φ with RIP of order k for any $k \leq c_0 n / \log(N/n)$

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 and at random from the Gaussian distribution with mean 0 and variance n^{-1}
- We use Bernouli process and create a matrix with entries $-1, 1$ (or $0, 1$)
- With high probability each of these random constructions yields a matrix Φ with RIP of order k for any $k \leq c_0 n / \log(N/n)$
- Probability is only used to prove existence of Φ. The sensing algorithm is constructive (not probabilistic) once we find a Φ.

Verification of RIP

- In general it is difficult to check whether a given matrix satisfies RIP
- It is much easier to show that a random family $\Phi(\omega) = (\phi_{i,j}(\omega))$, $\omega \in \Omega$ of matrices has RIP with high probability
- However, one then does not know specifically which of these matrices is favorable
- The primitive property is the concentration inequality

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

- From this one can use operator ideas to verify RIP (Baraniuk,Davenport, DeVore, Wakin)

Second Model for Signals: Compressible

- The sparse signal classes do not represent real signals: signals will typically have all entries nonzero but most will be small
- A compressible signal x is one that can be approximated well by elements from Σ_k:
  $$\sigma_k(x)_X := \inf_{z \in \Sigma_k} \|x - z\|_X$$
- Typical signal classes are $U(\ell_p^N)$ and typical $X = \ell_q^N$
  $$\|x\|_{\ell_p^N} := \left(\sum_{i=1}^N |x_i|^p \right)^{1/p}$$
- If $x \in U(\ell_p^N)$ then $\sigma_k(x)_{\ell_q^N} \leq k^{1/q - 1/p}, \quad p < q$
- Example ($q = 2,\ p = 1$): $\sigma_k(x)_{\ell_2^N} \leq k^{-1/2}$
- What is the best performance we can obtain on compressible signal classes?

Optimality on Models II

- These problems were solved in the late 1970's by Functional Analysts
- Deepest results are due to Kashin and Gluskin
- They say that with n measurements we can almost achieve the same performance as n term approximation
- Example ($p = 1, q = 2$)

$$C_0\sqrt{\frac{\log(N/n)}{n}} \leq E_n(U(\ell_1^N))_{\ell_2^N} \leq C_1\sqrt{\frac{\log(N/n)}{n}}$$

- These results do not provide practical encoding/decoding schemes
- Candes-Tao: RIP + ℓ_1 minimization give near optimal performance for $q = 2,\ p \leq 1$

Third Model for Signals: Arbitrary

- We say (Φ, Δ) is Instance-Optimal of order k for X if for an absolute constant $C > 0$ (independent of k, n, N)

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

- We will be interested in $X = \ell_q^N$
- 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 $k \leq c_0 n / \log(N/n)$
- Bonus: Decoding can be done by ℓ_1 Minimization
- Although not explicitly stated there this result is easily derived from the work of Candes-Tao

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

Instance-Optimality in Probability

- We saw that Instance-Optimality for ℓ_2^N is not viable
- Suppose $\Phi(\omega)$ is a collection of random matrices
- We say this family satisfies RIP of order k with probability $1-\epsilon$ if a random draw $\{\Phi(\omega)\}$ will satisfy RIP of order k with probability $1-\epsilon$
- We say $\{\Phi(\omega)\}$ is bounded with probability $1-\epsilon$ if given any $x \in \mathbb{R}^N$ with probability $1-\epsilon$ a random draw $\{\Phi(\omega)\}$ will satisfy

$$\|\Phi(\omega)(x)\|_{\ell_2^N} \leq C_0 \|x\|_{\ell_2^N}$$

  with C_0 an absolute constant
- Our earlier analysis showed that Gaussian and Bernouli random matrices have these properties with $\epsilon = e^{-cn}$

Theorem: Cohen-Dahmen-DeVore

- If $\{\Phi(\omega)\}$ satisfies RIP of order $3k$ and boundedness each with probability $1-\epsilon$ then there are decoders $\Delta(\omega)$ such that given any $x \in \ell_2^N$ we have with probability $1-2\epsilon$

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

- Instance-optimality in probability
- Range of k is $k \le c_0 n / \log(N/n)$
- Decoder is impractical

Decoding

- By far the most intriguing part of Compressed Sensing is the decoding

Decoding

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

Decoding

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

Decoding

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

Decoding

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

Decoding

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

Decoding

- By far the most intriguing part of Compressed Sensing is the decoding
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
  - ℓ_1 minimization: Long history (Donoho; Candes-Romberg)
  - Greedy algorithms - find support of a good approximation vector and then decode using ℓ_2 minimization (Gilbert-Tropp; Needel-Vershynin)
  - Iterative Reweighted Least Squares (Osborne, Daubechies-DeVore-Fornasier-Gunturk)
- Current Numerical winnner: ℓ_1 minimization via Bregman iteration: Yin-Osher-Goldfarb-Darbon

Decoding

- By far the most intriguing part of Compressed Sensing is the decoding
- There are continuing debates as to which decoding is numerically fastest
- Some common decoders
  - ℓ_1 minimization: Long history (Donoho; Candes-Romberg)
  - Greedy algorithms - find support of a good approximation vector and then decode using ℓ_2 minimization (Gilbert-Tropp; Needel-Vershynin)
  - Iterative Reweighted Least Squares (Osborne, Daubechies-DeVore-Fornasier-Gunturk)
- Current Numerical winnner: ℓ_1 minimization via Bregman iteration: Yin-Osher-Goldfarb-Darbon

Two outstanding Problems

- Show decoders are instance optimal in probability

Two outstanding Problems

- Show decoders are instance optimal in probability
  - For Gaussian random variables, Wojtaszek recently showed that ℓ_1 minimization decoding gives instance optimality of best order

Two outstanding Problems

- Show decoders are instance optimal in probability
  - For Gaussian random variables, Wojtaszek recently showed that ℓ_1 minimization decoding gives instance optimality of best order
  - For other processes, Cohen-Dahmen-DeVore have shown instance optimality modulo an arbitrary small error for ℓ_1 minimization and greedy

Two outstanding Problems

- Show decoders are instance optimal in probability
  - For Gaussian random variables, Wojtaszek recently showed that ℓ_1 minimization decoding gives instance optimality of best order
  - For other processes, Cohen-Dahmen-DeVore have shown instance optimality modulo an arbitrary small error for ℓ_1 minimization and greedy
- Deterministic construction of matrices

Two outstanding Problems

- Show decoders are instance optimal in probability
  - For Gaussian random variables, Wojtaszek recently showed that ℓ_1 minimization decoding gives instance optimality of best order
  - For other processes, Cohen-Dahmen-DeVore have shown instance optimality modulo an arbitrary small error for ℓ_1 minimization and greedy
- Deterministic construction of matrices
  - Constructions that satisfy RIP are know only for $k \leq C\sqrt{n}/\log(N/n)$

Two outstanding Problems

- Show decoders are instance optimal in probability
  - For Gaussian random variables, Wojtaszek recently showed that ℓ_1 minimization decoding gives instance optimality of best order
  - For other processes, Cohen-Dahmen-DeVore have shown instance optimality modulo an arbitrary small error for ℓ_1 minimization and greedy
- Deterministic construction of matrices
  - Constructions that satisfy RIP are know only for $k \leq C\sqrt{n}/\log(N/n)$
  - These constructions are similar to coding theory

Two outstanding Problems

- Show decoders are instance optimal in probability
  - For Gaussian random variables, Wojtaszek recently showed that ℓ_1 minimization decoding gives instance optimality of best order
  - For other processes, Cohen-Dahmen-DeVore have shown instance optimality modulo an arbitrary small error for ℓ_1 minimization and greedy
- Deterministic construction of matrices
  - Constructions that satisfy RIP are know only for $k \leq C\sqrt{n}/\log(N/n)$
  - These constructions are similar to coding theory
  - Can we find deterministic constructions which break the $\sqrt{n}$ limitation?