Computation

- Most scientific problems do not have a closed form solution
- To solve them we must resort to computation and create an approximate solution
- To do this we create a numerical algorithm
- There are typically a myriad of algorithms
- Is there a way to tell whether algorithm is optimal or near optimal?
- How do we define optimal?
- Computations are not exact
- We cant even enter a real number into a computer without some error
- Numerical Stability

Discrete Sensing

- $x \in I\!R^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question means inner product $v \cdot x$ with $v \in I\!R^N$ - called sample
- 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

Optimal Matrices

- We have only built Optimal Matrices using probabilistic methods
- What can deterministic constructions give?

Optimal Matrices

- We have only built Optimal Matrices using probabilistic methods
- What can deterministic constructions give?
- Needs a precise formulation

Optimal Matrices

- We have only built Optimal Matrices using probabilistic methods
- What can deterministic constructions give?
- Needs a precise formulation
- We can not prove high performance for deterministic constructions
- Deterministic constructions based on coding/finite fields etc. give $k \leq C_0\sqrt{n}$

Computational Issues

- We have seen three competing issues: (i) optimality of the sparsity level k, (ii) number of computations to decode, (iii) stability

Computational Issues

- We have seen three competing issues: (i) optimality of the sparsity level k, (ii) number of computations to decode, (iii) stability

Computational Issues

- We have seen three competing issues: (i) optimality of the sparsity level k, (ii) number of computations to decode, (iii) stability
- It is not yet clear exactly how these demands play against one another
- Theoretical Computer Science: Gilbert, Muthukrishnan, Strauss, and many others

Analog Systems

- How to sense analog signals?

Analog Systems

- How to sense analog signals?
- Discrete to Analog

Analog Systems

- How to sense analog signals?
- Discrete to Analog
- Can we build it: circuit implementation?

Analog Systems

- How to sense analog signals?
- Discrete to Analog
- Can we build it: circuit implementation?
- In certain contexts there are already impressive results: Tomography (Candes)

Discrete Sensing

- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^N$ - called sample
- 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??

Analog Systems

- How to sense analog signals?
- Discrete to Analog
- Can we build it: circuit implementation?
- In certain contexts there are already impressive results: Tomography (Candes)
- Web site on Compressed Sensing: Rice ECE

Discrete Sensing

- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^N$ - called sample
- 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??
- Here good roughly means that the samples $y = \Phi x$ contain enough information to approximate x well

Discrete Sensing

- $x \in \mathbb{R}^N$ with N large
- We are able to ask n non-adaptive questions about x
- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^N$ - called sample
- 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??
- Here good roughly means that the samples $y = \Phi x$ contain enough information to approximate x well
- Two issues: (i) Enough information in y; (ii) How to extract this information: Decoder

What is going on?

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

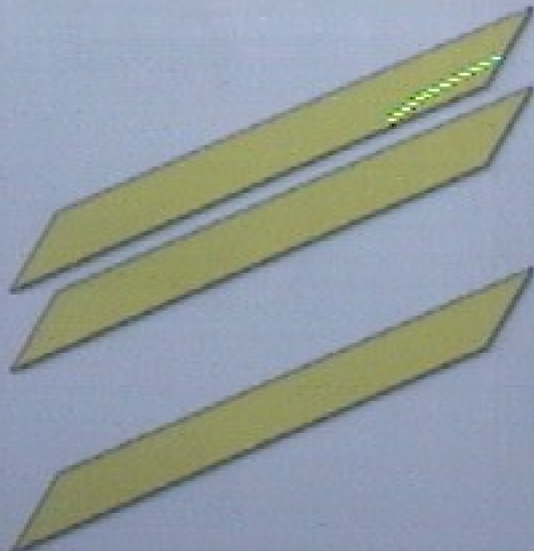
What is going on?

- Since $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ 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 $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ 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 sets $\mathcal{F}(y)$



What is going on?

- Since $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ 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 $\mathbb{R}^n \to \mathbb{R}^N$

What is going on?

- Since $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ 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 $\mathbb{R}^n \to \mathbb{R}^N$
- $\bar{x} := \Delta(\Phi(x))$ is our approximation to x from the information extracted

Many Settings

- Numerical PDEs
- Data Fitting
- Statistical Estimation
- Encoding/Compression
- Compressed Sensing

What is going on?

- Since $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ 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 $\mathbb{R}^n \to \mathbb{R}^N$
- $\bar{x} := \Delta(\Phi(x))$ 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 we can say anything

Structure

- With no additional information about x it is doubtful we can say anything
- Fortunately the x we are interested in have structure

Structure

- With no additional information about x it is doubtful we can say anything
- Fortunately the x we are interested in have structure
- Typically x can be well represented by sparse linear combination of certain building blocks - for our purposes these building blocks are a basis

Structure

- With no additional information about x it is doubtful we can say anything
- Fortunately the x we are interested in have structure
- Typically x can be well represented by 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

Structure

- With no additional information about x it is doubtful we can say anything
- Fortunately the x we are interested in have structure
- Typically x can be well represented by 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
- For the most part we shall assume the basis is known to us

First Measure of Optimality

- To begin with we shall assume x is sparse with respect to the canonical basis on $\mathbb{R}^N$

First Measure of Optimality

- 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)

First Measure of Optimality

- 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)

First Measure of Optimality

- 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 $\mathrm{supp}(x) := \{i : x_i \neq 0\}$

Many Settings

- Numerical PDEs
- Data Fitting
- Statistical Estimation
- Encoding/Compression
- Compressed Sensing

First Measure of Optimality

- 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 $\mathrm{supp}(x) := \{i : x_i \neq 0\}$
- $\Sigma_k := \{x : \#\mathrm{supp}(x) \leq k\}$

First Measure of Optimality

- 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 Measure of Optimality

- 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 Measure of Optimality

- 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 $\Delta(\Phi(x)) = x, \quad x \in \Sigma_k$
- Answer $n = 2k$

What matrices do the job?

- $\Phi = [v_1, \ldots, v_N]$, $v_1, \ldots, v_N$ columns of Φ

What matrices do the job?

- $\Phi = [v_1, \ldots, v_N]$, $\quad v_1, \ldots, v_N$ columns of Φ
- If T is a set of column indices

What matrices do the job?

- $\Phi = [v_1, \dots, v_N]$, $v_1, \dots, v_N$ columns of Φ
- If T 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 the columns with index $T = \{i_1, \dots, i_m\}$

What matrices do the job?

- $\Phi = [v_1, \dots, v_N]$, $v_1, \dots, v_N$ columns of Φ
- If T 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 the columns with index $T = \{i_1, \dots, i_m\}$
- $\Phi_T^* \Phi_T := (\langle v_i, v_j \rangle)_{i,j \in T}$

What matrices do the job?

- $\Phi = [v_1, \dots, v_n]$, $v_1, \dots, v_N$ columns of Φ
- If T is a set of column indices
- $\Phi_T = [v_{i_1}, \dots, v_{i_m}]$ is the $\#(T) \times n$ submatrix of Φ formed from the columns with index $T = \{i_1, \dots, i_m\}$
- $\Phi_T^* \Phi_T := (\langle v_i, v_j \rangle)_{i,j \in T}$ of size $\#(T) \times \#(T)$

Theorem: If Φ is any $n \times N$ matrix and $2k \le n$, 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) For any set T with $\#T = 2k$, the matrix Φ_T has rank $2k$.

(iv) For any set T with $\#T = 2k$, the $2k \times 2k$ matrix $\Phi_T^* \Phi_T$ is invertible: all its eigenvalues are positive.

Optimal Matrices

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

Many Settings

- Numerical PDEs
- Data Fitting
- Statistical Estimation
- Encoding/Compression
- Compressed Sensing
- Signal Processing: Searching for better ways to sample signals than Shannon-Nyquist Theory

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

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

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

Trouble in Madrid

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

Trouble in Madrid

- 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 Madrid

- 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 Madrid

- 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 Madrid

- 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 Madrid

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

Optimal Stable Systems

- Candes-Romberg-Tao; Donoho: Compressed Sensing

Many Settings

- Numerical PDEs
- Data Fitting
- Statistical Estimation
- Encoding/Compression
- Compressed Sensing
- Signal Processing: Searching for better ways to sample signals than Shannon-Nyquist Theory

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 $\#(T) = k$

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:

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

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

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

Optimal Stable Systems

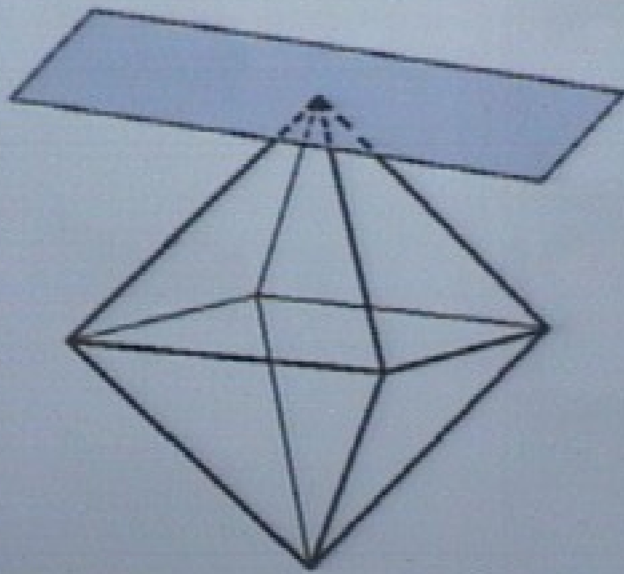
- Candes-Romberg-Tao; Donoho: Compressed Sensing
- Two important discoveries: good matrices, decoding
- Restricted Isometry Property (RIP) of order k: There exists $0 < \delta = \delta_k < 1$ such that the eigenvalues of $\Phi_T^*\Phi_T$ are in $[1-\delta, 1+\delta]$ whenever $\#(T) = k$
- Equivalently:

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

- 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

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

- 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

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

Many Settings

- Numerical PDEs
- Data Fitting
- Statistical Estimation
- Encoding/Compression
- Compressed Sensing
- Signal Processing: Searching for better ways to sample signals than Shannon-Nyquist Theory
- Audience Friendly: Easy to digest
- Interfaces many areas of mathematics: Functional Analysis, Approximation, Probability, Theoretical Computer Science ...

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 $k \le c_0 n / \log(N/n)$

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?

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

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 $v_1, \dots, v_N$ 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

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}

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}

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 $0, 1$)

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 $k \leq c_0 n / \log(N/n)$

Discrete Sensing

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

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 $k \leq c_0 n / \log(N/n)$
- Algorithm is constructive (not probabilistic) once we find a Φ. Probability is only used to prove existence of 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 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 $k \le c_0 n / \log(N/n)$
- Algorithm is constructive (not probabilistic) once we find a Φ. Probability is only used to prove existence of Φ

Lazy man verification of RIP

- Let $\Phi(\omega) = (\phi_{i,j}(\omega))$, $\omega \in \Omega$, be random matrices

Lazy man verification of RIP

- Let $\Phi(\omega) = (\phi_{i,j}(\omega))$, $\omega \in \Omega$, be random matrices
- Here each entry $\phi_{i,j}$ is an independent realization of some fixed random variable r with mean zero and variance $1/n$

Lazy man verification of RIP

- Let $\Phi(\omega) = (\phi_{i,j}(\omega))$, $\omega \in \Omega$, be random matrices
- Here each entry $\phi_{i,j}$ is an independent realization of some fixed random variable r with mean zero and variance $1/n$
- Trivially we have $E(\|\Phi(\omega)x\|_{\ell_2^n}^2) = \|x\|_{\ell_2^N}^2$

Lazy man verification of RIP

- Let $\Phi(\omega) = (\phi_{i,j}(\omega))$, $\omega \in \Omega$, be random matrices
- Here each entry $\phi_{i,j}$ is an independent realization of some fixed random variable r with mean zero and variance $1/n$
- Trivially we have $E(\|\Phi(\omega)x\|_{\ell_2^n}^2) = \|x\|_{\ell_2^N}^2$

Lazy man Proof

Theorem (Baraniuk, Davenport, DeVore, Wakin)
Given c_0 there is a $c_1 > 0$ such that with probability $1 - e^{-c_1 n}$, the matrix $\Phi(\omega)$ satisfies RIP of order k for all $k \leq c_0 n / \log(N/n)$

- We find a net of points $\mathcal{P}$ which cover the unit sphere in Σ_k to accuracy $\delta/4$
- Using the concentration inequality, we see that with high probability the draw of $\Phi = \Phi(\omega)$ satisfies

$$(1-\delta/2)\|q\|_{\ell_2^N} \leq \|\Phi(q)\|_{\ell_2^n} \leq (1+\delta/2)\|q\|_{\ell_2^N}, \quad q \in \mathcal{P}$$

- Extend this to all $x \in \Sigma_k$ by a boot strapping estimate: Given x with $\|x\|_{\ell_2^N} = 1$ find q such that $\|x - q\|_{\ell_2^N} \leq \delta/4$

$$\|\Phi(x)\|_{\ell_2^n} \leq \|\Phi(x-q)\|_{\ell_2^n} + \|\Phi(q)\|_{\ell_2^n} \leq M\delta/4 + 1 + \delta/2$$

Lazy man Proof

Theorem (Baraniuk, Davenport, DeVore, Wakin)
Given c_0 there is a $c_1 > 0$ such that with probability $1 - e^{-c_1 n}$, the matrix $\Phi(\omega)$ satisfies RIP of order k for all $k \le c_0 n / \log(N/n)$

- We find a net of points $\mathcal{P}$ which cover the unit sphere in Σ_k to accuracy $\delta/4$
- Using the concentration inequality, we see that with high probability the draw of $\Phi = \Phi(\omega)$ satisfies

$$(1-\delta/2)\|q\|_{\ell_2^N} \le \|\Phi(q)\|_{\ell_2^n} \le (1+\delta/2)\|q\|_{\ell_2^N}, \quad q \in \mathcal{P}$$

- Extend this to all $x \in \Sigma_k$ by a boot strapping estimate: Given x with $\|x\|_{\ell_2^N} = 1$ find q such that $\|x - q\|_{\ell_2^N} \le \delta/4$

$$\|\Phi(x)\|_{\ell_2^n} \le \|\Phi(x-q)\|_{\ell_2^n} + \|\Phi(q)\|_{\ell_2^n} \le M\delta/4 + 1 + \delta/2$$

Universality

- Suppose we have a collection of basis $\mathcal{B} := \{B\}$
- $\#(\mathcal{B}) \leq Ce^{cn}$
- These probability arguments give the existence of a matrix Φ which satisfies RIP of order k simultaneously for all $B \in \mathcal{B}$ and $k \leq cn/\log(N/n)$
- This means this one Φ will capture sparsity with respect to any and all of these bases
- However to decode we need to know the basis

Universality

- Suppose we have a collection of basis $\mathcal{B} := \{B\}$
- $\#(\mathcal{B}) \leq Ce^{cn}$
- These probability arguments give the existence of a matrix Φ which satisfies RIP of order k simultaneously for all $B \in \mathcal{B}$ and $k \leq c_0 n / \log(N/n)$
- This means this one Φ will capture sparsity with respect to any and all of these bases
- However to decode we need to know the basis

Discrete Sensing

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

Universality

- Suppose we have a collection of basis $\mathcal{B} := \{B\}$
- $\#(\mathcal{B}) \leq Ce^{cn}$
- These probability arguments give the existence of a matrix Φ which satisfies RIP of order k simultaneously for all $B \in \mathcal{B}$ and $k \leq c_0 n / \log(N/n)$
- This means this one Φ will capture sparsity with respect to any and all of these bases
- However to decode we need to know the basis

Model II: Compressible Signals

- 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 $\sigma_k(x)_{\ell_q^N} \le k^{1/q-1/p}, \quad p < q$
- Example ($q = 2$, $p = 1$): $\sigma_k(x)_{\ell_2^N} \le k^{-1/2}$

Model II: Compressible Signals

- 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 ($q = 2,\ p = 1$): $\sigma_k(x)_{\ell_2^N} \leq k^{-1/2}$

Optimality on Models II

- Let K be a compact set in $\mathbb{R}^N$
- Best performance on the class K in norm $\|\cdot\|_X$

$$E_n(K)_X := \inf_{(\Phi,\Delta)\in A_n} \sup_{x\in K} \|x - \Delta(\Phi(x))\|_X$$

- Here $A_n = \{(\Phi, \Delta) : \Phi \text{ is } n \times N\}$
- The pair (Φ, Δ) is near optimal if

$$\sup_{x\in K} \|x - \Delta(\Phi(x))\|_X \le C_0 E_n(K)_X$$

Optimality on Models II

- Let K be a compact set in $\mathbb{R}^N$
- Best performance on the class K in norm $\|\cdot\|_X$

$$E_n(K)_X := \inf_{(\Phi,\Delta)\in A_n} \sup_{x\in K} \|x - \Delta(\Phi(x))\|_X$$

- Here $A_n = \{(\Phi, \Delta) : \Phi \text{ is } n \times N\}$
- The pair (Φ, Δ) is near optimal if

$$\sup_{x\in K} \|x - \Delta(\Phi(x))\|_X \leq C_0 E_n(K)_X$$

Optimality on ℓ_p^N classes

- The asymptotic behavior of $E_n(K)_X$ is known for all $K = U(\ell_p^N)$ in all $X = \ell_q^N$
- Solved in 1970's and 1980's in Approximation Theory and Finite Dimensional Geometry
- Kashin, Gluskin main players
- 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 , Donoho show that RIP + ℓ_1 minimization give near optimal performance for $q = 2,\ p \leq 1$

Instance-Optimal

- 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 all $1 \leq q \leq 2$

Instance-Optimal

- 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 all $1 \leq q \leq 2$

Instance-Optimal

- 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 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

Discrete Sensing

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

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

Instance-Optimality in Probability

- We saw that Instance-Optimality for ℓ_2^N is not viable

Instance-Optimality in Probability

- We saw that Instance-Optimality for ℓ_2^N is not viable

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

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

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 $1-2\epsilon$

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

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 $1-2\epsilon$

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

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

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 $1-2\epsilon$

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

- Instance-optimality in probability
- Range of k is $k \leq c_0 n / \log(N/n)$
- Decoder is the least squares minimization

Optimal Matrices

- We have only built Optimal Matrices using probabilistic methods