Mathematics in Image Processing

Ronald DeVore

University of South Carolina

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software
- Some of the primary tasks of image processing are

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software
- Some of the primary tasks of image processing are
  - Compression

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software
- Some of the primary tasks of image processing are
  - Compression
  - Denoising / Deblurring

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software
- Some of the primary tasks of image processing are
  - Compression
  - Denoising / Deblurring
  - Registration

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software
- Some of the primary tasks of image processing are
  - Compression
  - Denoising / Deblurring
  - Registration
  - Fast acquisition of images

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software
- Some of the primary tasks of image processing are
  - Compression
  - Denoising / Deblurring
  - Registration
  - Fast acquisition of images
- We will confine this lecture to the acquisition of signals and images

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software
- Some of the primary tasks of image processing are
  - Compression
  - Denoising / Deblurring
  - Registration
  - Fast acquisition of images
- We will confine this lecture to the acquisition of signals and images
- This summer's workshop in Park City will address the full spectrum of issues in image/signal processing

Overview

- Mathematics provides both the theory behind image processing and the foundation for image processing algorithms and software
- Some of the primary tasks of image processing are
  - Compression
  - Denoising / Deblurring
  - Registration
  - Fast acquisition of images
- We will confine this lecture to the acquisition of signals and images
- This summer's workshop in Park City will address the full spectrum of issues in image/signal processing

Signal and Image acquisition

- Signals and images are acquired by sensors (cameras)

Signal and Image acquisition

- Signals and images are acquired by sensors (cameras)
- We are interested in obtaining the signal with the fewest number of measurements

Signal and Image acquisition

- Signals and images are acquired by sensors (cameras)
- We are interested in obtaining the signal with the fewest number of measurements
- We shall discuss this only in the discrete setting

Discrete Compressed Sensing

- $x \in I\!R^N$ with N large

Discrete Compressed Sensing

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

Discrete Compressed Sensing

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

Discrete Compressed Sensing

- $x \in I\!R^N$ 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

Discrete Compressed Sensing

- $x \in I\!R^N$ 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??

Discrete Compressed Sensing

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

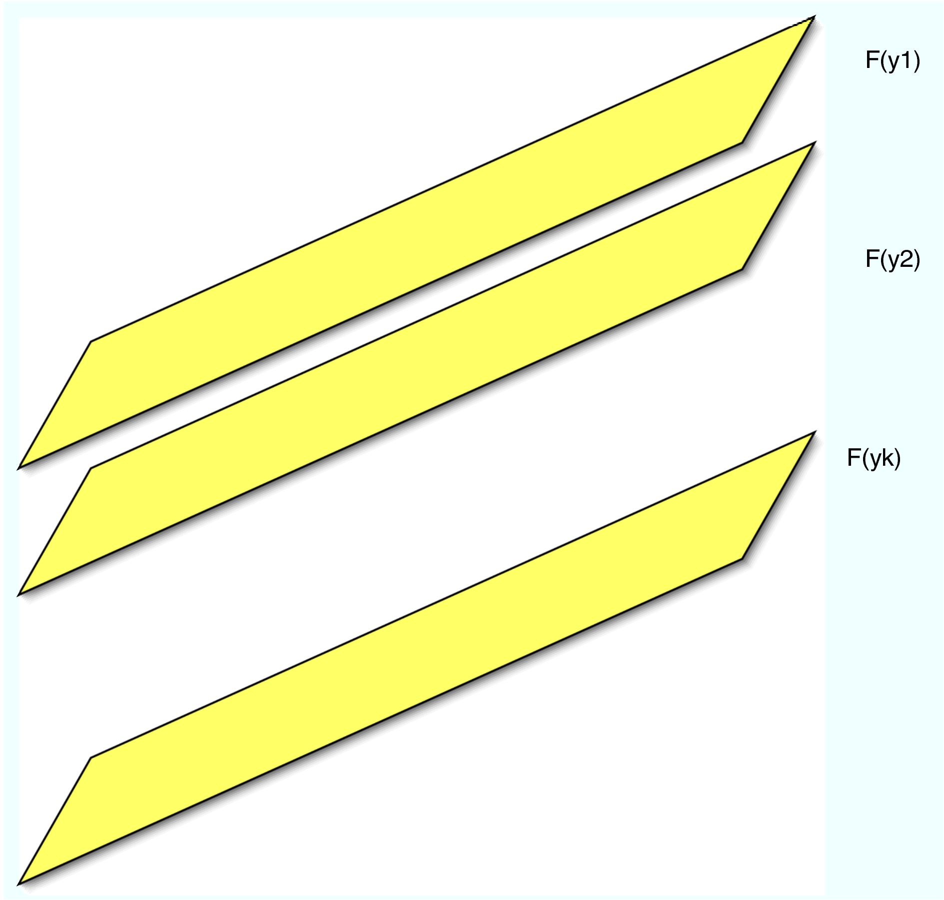
What is going on?

- 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 $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ 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 hyperplanes $\mathcal{F}(y)$ with $y \in \mathbb{R}^n$ stratify $\mathbb{R}^N$

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



What is going on?

- Since $\Phi : I\!R^N \to I\!R^n$ 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
- 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 $\mathrm{supp}(x) := \{i : \ x_i \neq 0\}$
- $\Sigma_k := \{x : \ \#\mathrm{supp}(x) \leq k\}$
- 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 $\Delta(\Phi(x)) = x, \quad x \in \Sigma_k$

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

- $\Phi = [v_1, \dots, v_N], \quad v_1, \dots, v_N$ 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 $\#(T) = k$

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 $x_1 < x_2 < \cdots < x_N$

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \\ \vdots & \vdots & \cdots & \vdots \\ x_1^{2k-1} & x_2^{2k-1} & \cdots & x_N^{2k-1} \end{pmatrix} \tag{3}$$

Naive Decoding

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

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

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 $v_1, \ldots, v_N$ 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 $I\!R^n$ 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 Bernouli 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 Bernouli 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.

Summary

- We see that the construction of good sensing systems utilizes the following areas of mathematics

Summary

- We see that the construction of good sensing systems utilizes the following areas of mathematics
  - Linear Algebra and properties of matrices

Summary

- We see that the construction of good sensing systems utilizes the following areas of mathematics
  - Linear Algebra and properties of matrices
  - Numerical Optimization (for decoding)

Summary

- We see that the construction of good sensing systems utilizes the following areas of mathematics
  - Linear Algebra and properties of matrices
  - Numerical Optimization (for decoding)
  - Probability (for construction of optimal matrices)

Summary

- We see that the construction of good sensing systems utilizes the following areas of mathematics
  - Linear Algebra and properties of matrices
  - Numerical Optimization (for decoding)
  - Probability (for construction of optimal matrices)

- Park City, Utah: July 2009