Universität Rostock Traditio et Innovatio

Institut für Nachrichtentechnik

Sampling and Reconstruction of Sparse Signals

Guest Lecture in Madrid, 26.09. - 27.09.2016

Volker Kühn

Universität Rostock
Institut für Nachrichtentechnik

Content Description - 1

In the last two decades the pervasion of our daily life by communication and multimedia devices like smart phones, digital cameras or MP3 players grows at increasing speed. The basic foundation of all devices is digital signal processing, especially the representation of analog signals by bits and bytes. In order to keep storage and data rate requirements at a moderate level, compression is an inevitable part of digital systems. It removes redundant parts from the signal and represents these signals with as few bits as possible. Thereby, lossy and lossless compression are distinguished. We call compressible signals as being sparse.

Conventional strategies first sample the analog signal at high rate according to Shannon's famous sampling theorem and compress it afterwards. This provokes the question why sampling itself cannot be directly perform the compression in order to avoid costly sampling at high rate. In order to answer this question, the lectures will give an overview of state-of-the-art sampling techniques and will focus on two different approaches.

Content Description - 2

First, a technique called 'Compressed Sensing' will be introduced allowing to compress sparse signals in a very efficient way. After discussing some toy examples to illustrate the underlying problem, the compression and its fundamental properties are introduced. Next, the reconstruction step is explained for one simple exemplary algorithm, the Orthogonal Matching Pursuit (OM) algorithm. In the second part, Shannon's famous sampling theorem will be revisited first. Next, the class of analog Finite Rate of Innovation (FRI) signals will be introduced which can be sampled at rates much lower than stated by Shannon. For reconstruction, the annihilating filter as one example of spectral estimation algorithms will be presented. Comparing these approaches with classical compression techniques like MP3, MPEG or H.265 shows that the complexity is shifted from the encoder which becomes very simple to the decoder.

Lectures

What is this course about?

1. Compressed Sensing
   - Sparse signals, single pixel camera, Magnetic Resonance Imaging (MRI)
   - Toy example with sparse polynomials
   - Norms (ℓ_0, ℓ_1, ℓ_2)
   - Compressed Sensing (CS)
   - Optimization problem
   - Greedy reconstruction algorithms

Lectures

What is this course about?

2 Sampling analog FRI signals

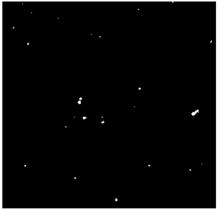
- Conventional sampling according to Shannon/Kotelnikov
- Sampling FRI signals using parametric description
- Spectral estimation by annihilating filter method
- Application to Particle Image Velocimetry
- Summary and Comparison of FRI and CS

What we need

- Linear algebra
- Fourier and z-transformation
- Sampling theorem of Shannon
- Curiosity and interest

Motivation

- Sparse signals allow exact description by few components (in certain domain / subspace) $\longrightarrow$ lossless compression (Huffman coding, run-length coding)
- Approximately sparse signals dominated by few large components $\longrightarrow$ lossy compression (rate-distortion theory)
- Rich area of applications: radar, sonar, medical (ultrasound), communications, multimedia signal processing, ...



Sampling Sparse Signals

- Conventional approaches: sample at high rate and compress signal afterwards
  - Audio signal processing (mp3, ...)
  - Video and image compression (MPEG, H.265, JPEG, ...)
- Sampling according to Shannon requires sampling rate twice as high as signal bandwidth (sufficient criterion, no necessary condition!)
- Questions:
  - Why high rate sampling and successive compression?
  - Can the sampling rate be reduced to perform compression directly?
  - Is (perfect) reconstruction possible even for non-bandlimited signals?
- Yes, but at the expense of increased reconstruction complexity!

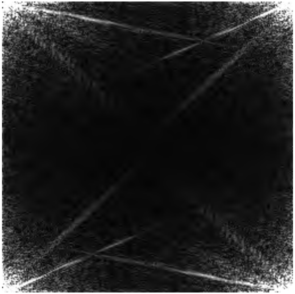
Image Compression

- Natural images typically are not sparse ...

Image Compression

- ... but they are approximately sparse in Wavelet-/DCT-/Fourierbasis[1]

Waveletbasis representation



FFT-basis (clipped)

[1] many coefficients are almost zero $\longrightarrow$ lossy compression

Image Compression

- Reconstruction results

Original 256×256

Reconstruction $M = 16384 c = 1/4$

Reconstruction $M = 8192 c = 1/8$

- Many sophisticated algorithms for
  - image compression (JPEG, JPEG 2000, ...)
  - video compression (MPEG, H26x, ...)
  - audio compression (AAC, MP3, ...)

Magnetic Resonance Imaging (MRI)

Principle

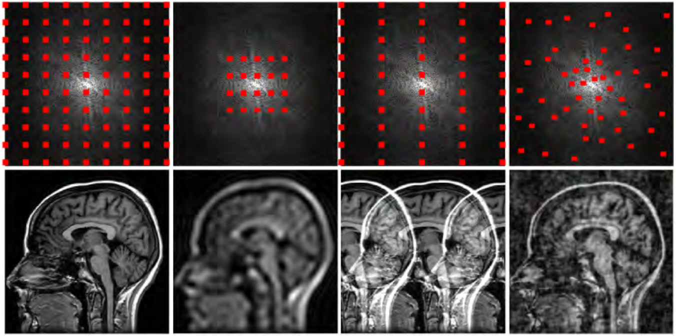
- Medical imaging tool with inherently slow data acquisition process
- Temporal MRI signal directly samples the spatial frequency domain of the image

$$s(t) = \int_R m(\vec{r}) e^{-i2\pi \vec{k}(t)\cdot \vec{r}} dr$$

- Received signal $s(t)$ equals Fourier transform of object $m(\vec{r})$ sampled at the spatial frequency $\vec{k}(t)$
- Incoherent detection (energy detection similar to camera sensors)
- Measurement time per coefficient limited by physical constraints (max. amplitude and max. slew-rate to avoid nerve stimulation)

Magnetic Resonance Imaging (MRI)

Example 1: MRI of human head[1]



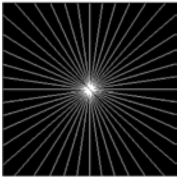
- Fewer measurements result in blurs, aliasing or random artifacts
- Using CS-techniques with random sampling yields better performance

[1] Lustig, Donoho, Santos, Pauly: "Compressed Sensing MRI", IEEE Signal Processing Magazine, vol.25, no.2, pp.72-82, March 2008

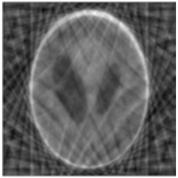
Magnetic Resonance Imaging (MRI)

Example 2: Logan Shepp test image[2]

original



Fourier coefficients sampled along radial lines



minimum-energy reconstruction

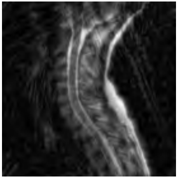
TV-norm minimization

[2] Candes, Romberg, Tao: "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information", IEEE Transactions on Information Theory, vol. 52, no. 2, pp.489-509, Feb. 2006

Magnetic Resonance Imaging (MRI)

Example 3: Real MRI Images[3]

fully sampled



$6\times$ undersampled

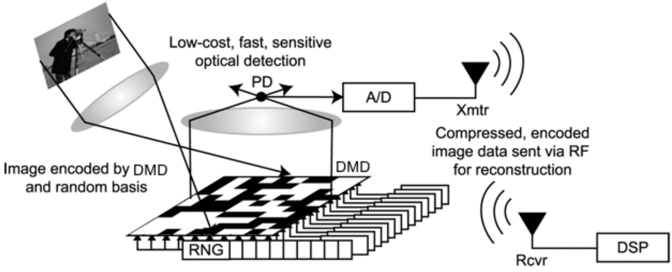
$6\times$ undersampled with CS reconstruction

- CS allows for less measurements required for approximately same resolution
- $\rightarrow$ decreased measurement time
- $\rightarrow$ higher convenience for patients

[3] Trzasko, Manduca: "Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic ℓ_0 minimization", IEEE Transactions on Medical Imaging, vol. 28, no. 1, pp.106-121, Jan. 2009

Single Pixel Camera[4]

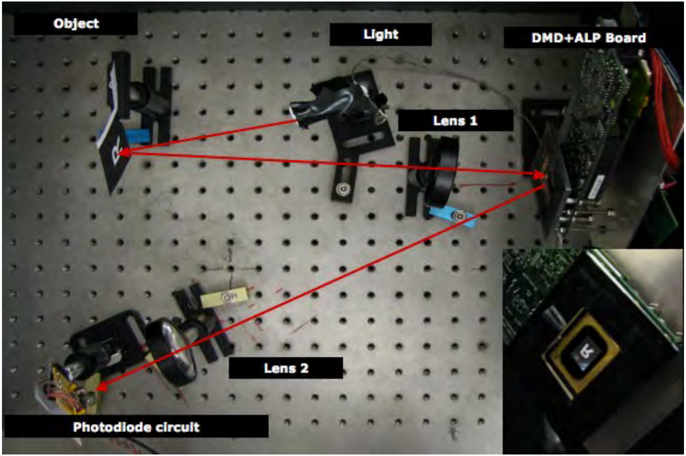
- Image acquisition with **only one** photo-detector
- Detector acquires multiple pseudorandom linear projections via DMD per scene
- Beneficial if detectors are expensive



[4] http://dsp.rice.edu/cscamera, DMD: Digital Micromirror Device, RNG: Random Number Generator

Single Pixel Camera

Hardware-setup [http://dsp.rice.edu/cscamera]



Single Pixel Camera

Some reconstruction results[5]

original, 16384 pixels

1300 samples (2%)

1600 samples (10%)

3300 samples (20%)

original, 4096 pixels

800 samples (20%)

1600 samples (40%)

[5] http://dsp.rice.edu/cscamera

Compressed Sensing

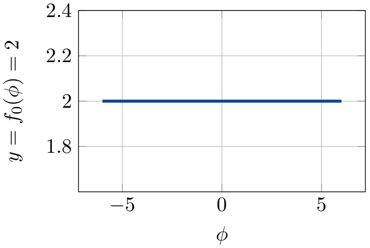
Toy Example with Polynomials

How many samples are needed to identify polynomial of rank $N-1$?

- Polynomial of rank $N-1$

$$y = f_{N-1}(\phi) = x_0 + x_1\phi + x_2\phi^2 + \cdots + x_{N-1}\phi^{N-1}$$

- rank $N-1=0$: $y = f_0(\phi) = x_0 \quad \forall \phi \in \mathbb{R}$

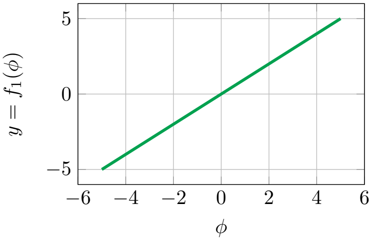


- 1 sample sufficient to identify x_0

Toy Example with Polynomials

How many samples are needed to identify polynomial of rank $N-1$?

- rank $N-1=1$: $y = f_1(\phi) = x_0 + x_1\phi$



- 2 samples with $\phi_1 \neq \phi_2$ required for identifying straight line

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 1 \end{bmatrix}}_{\boldsymbol{\Phi}} \cdot \underbrace{\begin{bmatrix} x_1 \\ x_0 \end{bmatrix}}_{\boldsymbol{x}} \longrightarrow \boldsymbol{x} = \boldsymbol{\Phi}^{-1} \cdot \boldsymbol{y}$$

Toy Example with Polynomials

How many samples are needed to identify polynomial of rank $N-1$?

- rank $N-1=2$: $y=f_2(\phi)=x_0+x_1\phi+x_2\phi^2$

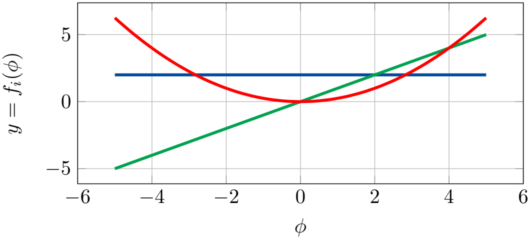
$$\Longrightarrow \quad \boldsymbol{y}=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}=\underbrace{\begin{bmatrix} \phi_1^2 & \phi_1 & 1 \\ \phi_2^2 & \phi_2 & 1 \\ \phi_3^2 & \phi_3 & 1 \end{bmatrix}}_{\boldsymbol{\Phi}} \cdot \underbrace{\begin{bmatrix} x_2 \\ x_1 \\ x_0 \end{bmatrix}}_{\boldsymbol{x}} \longrightarrow \boldsymbol{x}=\boldsymbol{\Phi}^{-1}\cdot\boldsymbol{y}$$

- Linear equation system can be uniquely solved as long as matrix $\boldsymbol{\Phi}$ is regular!
- Generally, N samples are needed to identify polynomial $f(\phi)$ of rank $N-1$
- What happens, if polynomial has maximal rank $N-1$, but only few coefficients are non-zero?
- Can polynomial be determined with less than N samples?

Toy Example with Sparse Polynomials

Polynomial of maximal rank $N-1=2$ and single non-zero coefficient

- $\binom{3}{1}=3$ possible polynomials: $f_0(\phi)=x_0$, $f_1(\phi)=x_1\phi$, $f_2(\phi)=x_2\phi^2$



- 2 unknown parameters (position and value of nonzero coefficient)
  $\rightarrow$ 2 instead of 3 samples suffice to identify $f_i(\phi)$

Toy Example with Sparse Polynomials

Polynomial of maximal rank $N-1=2$ and single non-zero coefficient

- 2 samples lead to underdetermined equation system

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \phi_1^2 & \phi_1 & 1 \\ \phi_2^2 & \phi_2 & 1 \end{bmatrix}}_{\boldsymbol{\Phi}} \cdot \underbrace{\begin{bmatrix} x_2 \\ x_1 \\ x_0 \end{bmatrix}}_{\boldsymbol{x}}$$

- Solution cannot be found by simple matrix inversion as $\boldsymbol{\Phi}$ is not quadratic
- Problem becomes combinatorial (ϕ_1 and ϕ_2 must have different magnitudes and signs for unique identification)
- More sophisticated (nonlinear) detection algorithms are required!

Compressed Sensing (CS) invented by mathematicians

David Donoho, Stanford

Emmanuel Candès, Stanford

Terence Tao, UCLA

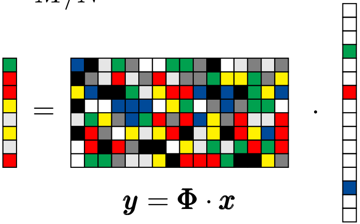
many others

- Meanwhile, CS found its way into various applications like
  - Source coding of sparse signals (sampling rates far below the classical theorem)
  - De-noising of images, audio, speech
  - Inpainting for images, audio
  - Radar signal processing (target detection)
  - Channel estimation (multipath propagation with few dominant paths)

Basic Principle of Compressed Sensing

Compressing sparse vector by linear projection

- Sparse vector $\boldsymbol{x} \in \mathbb{R}^N$ with at most $K \ll N$ nonzero components
- Compression by multiplying with $M \times N$ sensing matrix $\boldsymbol{\Phi}$, $M \ll N$
- Compression factor: $c = M/N$



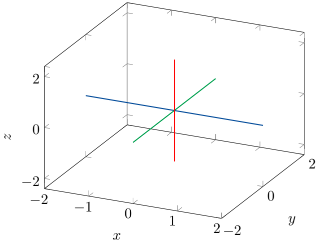
$$\boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{x}$$

- Equation system is underdetermined (infinitely many solutions) and can only be uniquely solved with prior knowledge (sparsity of $\boldsymbol{x}$)

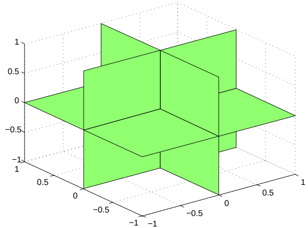
Sparsity

- Vector $\boldsymbol{x}$ is denoted as K-sparse if it contains at most K nonzero elements
- Common notation with ℓ_0 norm[6]: $\|\boldsymbol{x}\|_0 \leq K$
- Example: Set of all 2-sparse signals in $\mathbb{R}^3$

1D subspaces



2D subspaces



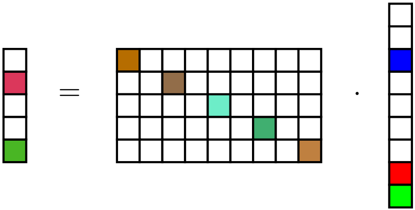
[6] ℓ_0 norm is no real norm in mathematical sense.

Sensing Matrix

How is matrix $\mathbf{\Phi}$ to be designed such that M is as small as possible?

- Sparse sensing matrix will not work

$$\boldsymbol{y} = \mathbf{\Phi} \cdot \boldsymbol{x}$$



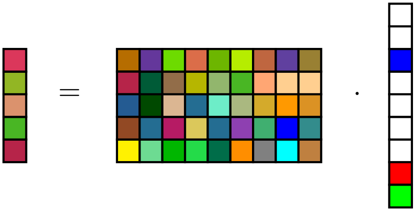
- As positions of nonzero components are not known in advance, catching them with sparse sensing matrices is totally random (we have to be very lucky).
- Some information might not be captured in $\boldsymbol{y}$ with high probability.

Sensing Matrix

How is matrix $\mathbf{\Phi}$ to be designed such that M is as small as possible?

- Dense sensing matrix: information has to be spread over all coefficients

$$\boldsymbol{y} = \mathbf{\Phi} \cdot \boldsymbol{x}$$



- Compressed vector $\boldsymbol{y}$ contains linear superposition of K columns of $\mathbf{\Phi}$
- Compressed vector $\boldsymbol{y}$ should be unique for each sparse vector $\boldsymbol{x}$

Properties of Sensing Matrix $\boldsymbol{\Phi}$

Spark of Sensing Matrix $\boldsymbol{\Phi}$

Definition

The spark of a given matrix $\boldsymbol{\Phi}$ is the smallest number of columns of $\boldsymbol{\Phi}$ that are linear dependent.

- If all K–sparse vectors $\boldsymbol{x}$ should be recoverable, two different vectors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ must not be mapped onto the same $\boldsymbol{y}$
- Recovery is not possible if $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}_1 = \boldsymbol{\Phi}\boldsymbol{x}_2 \Rightarrow \boldsymbol{\Phi}(\boldsymbol{x}_1 - \boldsymbol{x}_2) = \mathbf{0}$
- For any vector $\boldsymbol{y}$, there exists at most one K sparse signal $\boldsymbol{x}$ with $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ if and only if $\mathrm{spark}(\boldsymbol{\Phi}) > 2K$ (at least $2K+1$ columns can be linear dependent.).
- Nullspace of $\boldsymbol{\Phi}$ must not contain any $2K$-sparse vectors

Spark of Sensing Matrix $\mathbf{\Phi}$

Example for sparse polynomial of maximal rank $N-1=2$

- Polynomial with max. rank $N-1=2$ and single non-zero coefficient ($K=1$)
- Possible polynomials: $f_0(\phi)=x_0$, $f_1(\phi)=x_1\phi$, $f_2(\phi)=x_2\phi^2$
- 2 appropriate samples suffice to identify $f_i(\phi)$
- $\mathrm{spark}(\mathbf{\Phi}) \geq 2K+1=3$ required for unique identification
  - $\mathbf{\Phi} = \begin{bmatrix} \phi_1^2 & \phi_1 & 1 \\ \phi_2^2 & \phi_2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \longrightarrow \mathrm{spark}(\mathbf{\Phi}) = 2 < 2K+1$

    (straight line and parabola cannot be distinguished for $\phi_1=0$ and $\phi_2=1$)
  - $\mathbf{\Phi} = \begin{bmatrix} \phi_1^2 & \phi_1 & 1 \\ \phi_2^2 & \phi_2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \longrightarrow \mathrm{spark}(\mathbf{\Phi}) = 3 = 2K+1$

    ($\phi_1=1$ and $\phi_2=2$ ensure distinction of all three polynomials)

Properties of Sensing Matrix $\boldsymbol{\Phi}$

Restricted Isometry Property (RIP)

- Spark does not provide robustness against noise
- Euclidean distances should be preserved by projection

Definition

A matrix $\boldsymbol{\Phi}$ with unit-norm columns satisfies the restricted isometry property of order K if there exists a $\delta_K \in (0,1)$ such that

$$(1-\delta_K) \leq \frac{\|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \leq (1+\delta_K)$$

holds for all K-sparse $\boldsymbol{x}$ and matrices $\boldsymbol{\Phi}$.

- Replace $\boldsymbol{x}$ by $\boldsymbol{x}_1 - \boldsymbol{x}_2$: RIP of order $2K$ preserves Euclidean distance between vectors!

Restricted Isometry Property (RIP)

Example 1 for sparse polynomial of maximal rank $N-1=2$

- Choosing $\phi_1 = 0$ and $\phi_2 = 1$ delivers

$$\mathbf{\Phi}_1 = \begin{bmatrix} \phi_1^2 & \phi_1 & 1 \\ \phi_2^2 & \phi_2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \rightarrow \quad \mathbf{\Phi}_{1,\mathrm{norm}} = \begin{bmatrix} 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

- $\begin{bmatrix} 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ x_0 \end{bmatrix} \quad \rightarrow \quad \|\mathbf{\Phi}_{2,\mathrm{norm}}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2 = x_0^2$

- $\begin{bmatrix} 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ x_1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ x_1 \end{bmatrix} \quad \rightarrow \quad \|\mathbf{\Phi}_{2,\mathrm{norm}}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2 = x_1^2$

- $\begin{bmatrix} 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{x_2}{\sqrt{2}} \\ \frac{x_2}{\sqrt{2}} \end{bmatrix} \quad \rightarrow \quad \|\mathbf{\Phi}_{2,\mathrm{norm}}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2 = x_2^2$

Restricted Isometry Property (RIP)

Example 2 for sparse polynomial of maximal rank $N-1=2$

- Choosing $\phi_1 = 1$ and $\phi_2 = 2$ delivers

$$\boldsymbol{\Phi}_2 = \begin{bmatrix} \phi_1^2 & \phi_1 & 1 \\ \phi_2^2 & \phi_2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \quad \rightarrow \quad \boldsymbol{\Phi}_{2,\mathrm{norm}} = \begin{bmatrix} \frac{1}{\sqrt{17}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{2}} \\ \frac{4}{\sqrt{17}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

- $$\begin{bmatrix} \frac{1}{\sqrt{17}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{2}} \\ \frac{4}{\sqrt{17}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{17}} \\ \frac{4}{\sqrt{17}} \end{bmatrix} x_0 \quad \rightarrow \quad \left\| \boldsymbol{\Phi}_{2,\mathrm{norm}} \boldsymbol{x} \right\|_2^2 = \left\| \boldsymbol{x} \right\|_2^2 = x_0^2$$

- $$\begin{bmatrix} \frac{1}{\sqrt{17}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{2}} \\ \frac{4}{\sqrt{17}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ x_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} x_1 \quad \rightarrow \quad \left\| \boldsymbol{\Phi}_{2,\mathrm{norm}} \boldsymbol{x} \right\|_2^2 = \left\| \boldsymbol{x} \right\|_2^2 = x_1^2$$

- $$\begin{bmatrix} \frac{1}{\sqrt{17}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{2}} \\ \frac{4}{\sqrt{17}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} x_2 \quad \rightarrow \quad \left\| \boldsymbol{\Phi}_{2,\mathrm{norm}} \boldsymbol{x} \right\|_2^2 = \left\| \boldsymbol{x} \right\|_2^2 = x_2^2$$

Properties of Sensing Matrix $\mathbf{\Phi}$

Limited practical value of RIP and Spark

- RIP holds asympotically for Gaussian i.i.d. matrices with overwhelming probability provided that $M \geq \mathcal{O}(K \log \frac{N}{K})$
- Random Bernoulli matrices may work with overwhelming probability as well
- M random rows of an $N \times N$ Fourier matrix may also work appropriately
- However, proving spark and RIP property of $\mathbf{\Phi}$ are NP-hard problems
  $\Rightarrow$ infeasible in practice!
  - All $\binom{N}{2K}$ sets must be checked for linear independence
  - For small problem with $N = 300$ and $K = 30$:
    $\binom{300}{60} = 9 \cdot 10^{63}$ combinations to be checked!

Properties of Sensing Matrix $\mathbf{\Phi}$

Coherence

Definition

The coherence $\mu(\mathbf{\Phi})$ of a matrix $\mathbf{\Phi}$ is the largest absolute inner product between any two columns $\boldsymbol{\phi}_i, \boldsymbol{\phi}_j$ of $\mathbf{\Phi}$, i.e.

$$\mu(\mathbf{\Phi}) = \max_{1<i<j<N} \frac{|\langle \boldsymbol{\phi}_i, \boldsymbol{\phi}_j \rangle|}{\|\boldsymbol{\phi}_i\|_2 \, \|\boldsymbol{\phi}_j\|_2}$$

- Can be checked in reasonable time (only projections between columns of sensing matrix are considered, not all possible signal vectors)
- Valid interval of coherence with Welch bound: $\mu(\mathbf{\Phi}) \in \left[\sqrt{\frac{N-M}{M(N-1)}},\; 1 \right]$

Properties of Sensing Matrix $\mathbf{\Phi}$

Coherence

- Coherence represents lower bound for the spark:

$$\text{spark}(\mathbf{\Phi}) \geq 1 + \frac{1}{\mu(\mathbf{\Phi})}$$

- Suited sensing matrices have low coherence (columns shall be as "orthogonal as possible")
  - Random Gaussian or Bernoulli matrices deliver good results
  - Deterministic matrices can be constructed and analyzed as well
- Coding equivalence: distance between code words should be as large as possible

Coherence

Example for sparse polynomial of maximal rank $N-1=2$

- Sensing matrix $\mathbf{\Phi}_{1,\text{norm}} = \begin{bmatrix} 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix}$

$$\mu(\mathbf{\Phi}_{1,\text{norm}}) = \max\left\{1\,,\ \frac{1}{\sqrt{2}}\,,\ \frac{1}{\sqrt{2}}\right\} = 1$$

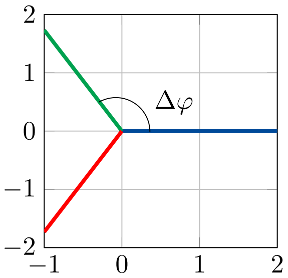
- Sensing matrix $\mathbf{\Phi}_{2,\text{norm}} = \begin{bmatrix} \frac{1}{\sqrt{17}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{2}} \\ \frac{4}{\sqrt{17}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

$$\mu(\mathbf{\Phi}_{2,\text{norm}}) = \max\left\{\frac{9}{\sqrt{85}}\,,\ \frac{5}{\sqrt{34}}\,,\ \frac{3}{\sqrt{10}}\right\} = \frac{9}{\sqrt{85}} \approx 0.9762$$

Coherence

Free Example for minimal coherence

- In our example, three column vectors in 2-dimensional space
- Angle between any pair of vectors shall be as large as possible $\rightarrow \quad \Delta\varphi = \frac{2\pi}{3}$
- $\mathbf{\Phi} = \begin{bmatrix} 1 & \cos(2\pi/3) & \cos(4\pi/3) \\ 0 & \sin(2\pi/3) & \sin(4\pi/3) \end{bmatrix}$
- $\mu(\mathbf{\Phi}) = \max\{0.5\,,\ 0.5,\ 0.5\} = 0.5$
  (Welch lower bound)
- Spark equals coherence:
  $\text{spark}(\mathbf{\Phi}) = 1 + \frac{1}{0.5} = 3$



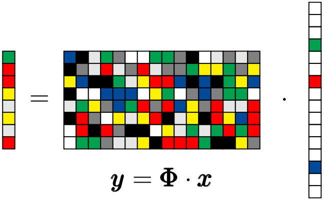
Lectures

What is this course about?

1. Compressed Sensing
   - Sparse signals, single pixel camera, Magnetic Resonance Imaging (MRI)
   - Toy example with sparse polynomials
   - Norms (ℓ_0, ℓ_1, ℓ_2)
   - Compressed Sensing (CS)
   - Optimization problem
   - Greedy reconstruction algorithms

Reconstruction of Sparse Signals

- Remember: Compression of K-sparse vector $\boldsymbol{x} \in \mathbb{R}^N$ by linear projection with sensing matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$



$$\boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{x}$$

- Equation system is underdetermined (infinitely many solutions) and can only be uniquely solved with prior knowledge (sparsity of $\boldsymbol{x}$)

- Optimization problem: Find the sparsest solution $\hat{\boldsymbol{x}}$ that solves $\boldsymbol{y} = \boldsymbol{\Phi}\hat{\boldsymbol{x}}$

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} \tag{1}$$

Reconstruction of Sparse Signals

Optimal Solution

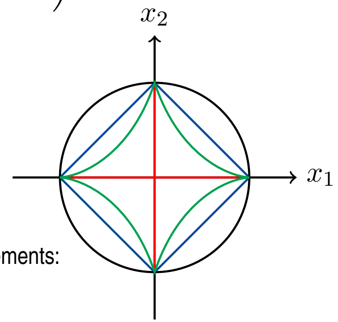
- Original problem is non-convex and NP-hard
- Optimal solution requires exhaustive search
  - Checking all $\binom{N}{K}$ K-sparse hypotheses, i.e. combination of K columns of $\mathbf{\Phi}$
  - For each hypothesis, amplitudes have to be estimated such that $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x}$, e.g. by least squares approach
  - Requires each time pseudo inversion of $M \times K$ submatrix of $\mathbf{\Phi}$
  - Even for small problem with $N = 300$ and $K = 30$:
    $\binom{300}{30} = 1.7 \cdot 10^{41}$ combinations to be checked $\Longrightarrow$ infeasible!
- Alternative solutions required

Illustration of Different Norms for 2D Case

- General definition of ℓ_p norm

$$\ell_p(\boldsymbol{x}) = \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p}$$

- ℓ_0 norm: number of nonzero elements
- ℓ_1 norm: sum of magnitudes
- ℓ_2 norm: sum of squares magnitudes
- $\ell_{0.5}$ norm
- ℓ_∞ norm delivers largest magnitude of all elements: $\|\boldsymbol{x}\|_\infty = \max_j |x_j|$



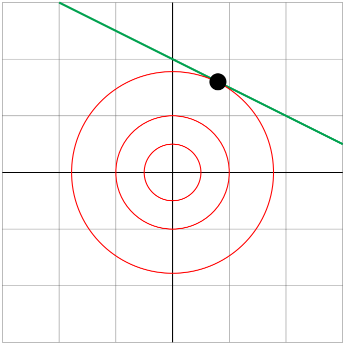
example for $\boldsymbol{x} = [x_1 \; x_2]^{\mathrm{T}}$

Graphical Illustration of Detection with ℓ_2 and ℓ_1 Norms

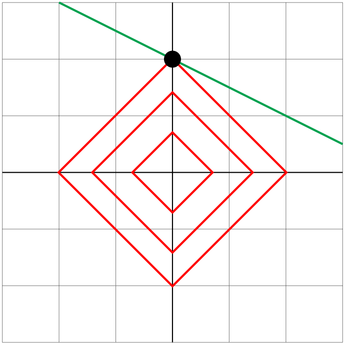
Estimate tuple (x_1, x_2) from observation $y = \mathbf{\Phi}\boldsymbol{x} = \phi_1 x_1 + \phi_2 x_2$

$$x_2 = \frac{1}{\phi_2}(y - \phi_1 x_1)$$

ℓ_2 norm solution is not sparse!



ℓ_1 norm solution is sparse!

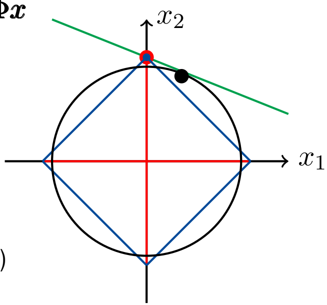


Reconstruction with Different Norms for 2D Case

- Estimate tuple (x_1, x_2) from observation $y = \mathbf{\Phi x} = \phi_1 x_1 + \phi_2 x_2$

$$x_2 = \frac{1}{\phi_2}(y - \phi_1 x_1)$$

- Infinitely many solutions as equation is underdetermined!
- Constraint: $\hat{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_p$ s.t. $y = \mathbf{\Phi}\boldsymbol{x}$
  - Choose solution with minimal ℓ_0 norm
    $\hat{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_0$ s.t. $y = \mathbf{\Phi}\boldsymbol{x}$
  - Choose solution with minimal ℓ_1 norm
    $\hat{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_1$ s.t. $y = \mathbf{\Phi}\boldsymbol{x}$
  - Choose solution with minimal ℓ_2 norm
    $\hat{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_2$ s.t. $y = \mathbf{\Phi}\boldsymbol{x}$ (not sparse!)

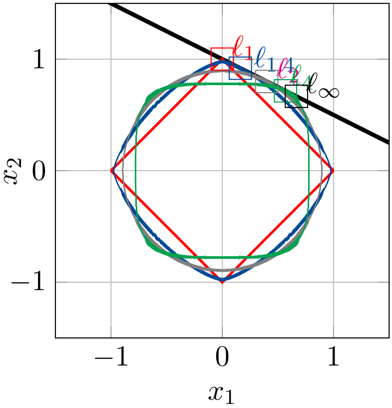


Reconstruction Results for Different Norms for 2D Case

- Reults obtained with CVX optimization toolbox from Boyd with Matlab[a]

```
N = 2;
  Phi = [0.5 1];
  y = 1;
  p = 1;

   cvx_begin
       variable x(N)
       minimize(norm(x,p)));
       subject to
       y == Phi*x;
  cvx_end
```



[a] http://cvxr.com/cvx/download, http://github.com/cvxr/cvx

Reconstrution of Sarse Signals

Near-optimal solution by convex relaxation

- Minimizing ℓ_1 norm provides optimal solution for almost all cases

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_1 \qquad \text{s.t.} \qquad \boldsymbol{y} = \boldsymbol{\Phi} \cdot \boldsymbol{x}$$

  - ℓ_1 norm leads to convex problem $\rightarrow$ efficient solutions by convex optimization tools

- Modified problem formulations possible in the presence of noise
  - Basis Pursuit Inequality Constraints (BPIC):

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_1 \qquad \text{s.t.} \qquad \|\boldsymbol{y} - \boldsymbol{\Phi} \cdot \boldsymbol{x}\|_2 \leq \epsilon$$

  - Lagrangian relaxation leads to Basis Pursuit DeNoising (BPDN), also known as LASSO (Least Absolute Shrinkage and Selection Operator):

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{x}\|_1 + \lambda \|\boldsymbol{y} - \boldsymbol{\Phi} \cdot \boldsymbol{x}\|_2$$

Reconstruction of Sparse Signals

Greedy algorithms

- Optimal reconstruction with ℓ_0 norm was NP-hard
- Relaxation to ℓ_1 norm enables application of convex optimization, but becomes infeasible for large problems as well
- Probabilistic graph based models (AMP: Approximate Message Passing) very attractive (but out of scope of this lecture)
- Greedy algorithms solving (1) approximately
  - Iterative reconstruction tries to minimize ℓ_0 norm
  - Low computational complexity, but suboptimal performance

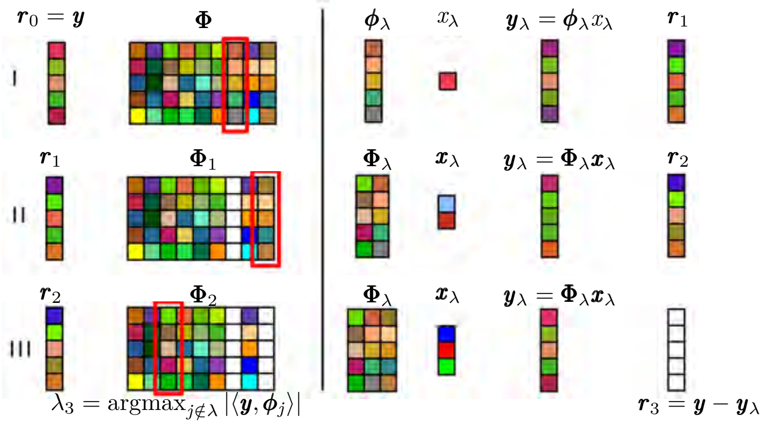
Reconstruction of Sparse Signals

Orthogonal Matching Pursuit (OMP)

- Greedy algorithm approximates optimal solution iteratively
- Basic principle:
  1. Set residuum $\boldsymbol{r}$ to compressed vector $\boldsymbol{y}$
  2. Choose column $\boldsymbol{\phi}_i^*$ of sensing matrix $\boldsymbol{\Phi}$ mostly correlated to residuum $\boldsymbol{r}$
  3. Determine optimal weight $\widehat{x}_i$, e.g. least squares solution
  4. Update residuum by subtracting contribution of $\widehat{x}_i$ from current residuum and continue with step 2 until
     - sparsity K of $\boldsymbol{x}$ has been reached (K has to be known a priori)
     - norm of residuum $\boldsymbol{r}$ falls below defined threshold
- Further pursuit methods: Stage-wise OMP (StOMP), CoSaMP, …

Reconstruction of Sparse Signals

Illustration of Orthogonal Matching Pursuit (OMP) for $K = 3$

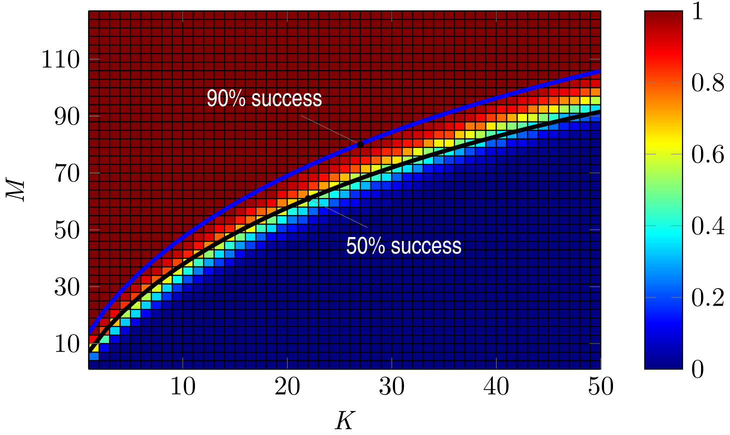


Phase Transition

- Phase transition illustrates regions with reliable and unreliable reconstruction
- Reconstruction succeeds with overwhelming probability if there are enough measurements M for given sparsity K
- If M is not sufficiently large, reconstruction via OMP is nearly impossible
- Transition area:
  - recovery success depends heavily on realization of $\boldsymbol{\Phi}$ and $\boldsymbol{x}$
  - sharpens for larger problem dimension N

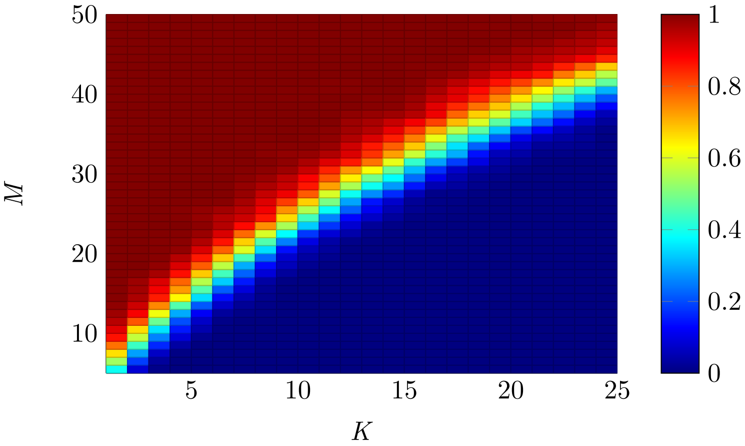
Example for Phase Transition with $N = 128$

- Basis Pursuit: $M \geq c \cdot K \log_2(1 + N/K)$ (black curve for $c = 1$)



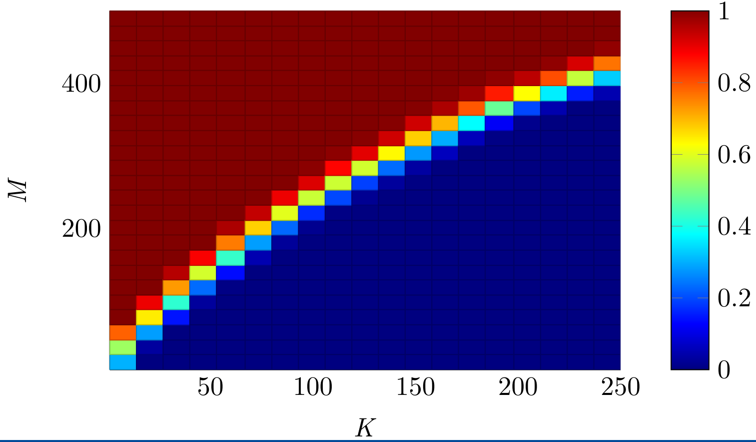
Phase Transition for OMP with $N = 50$

- Relative broad transition area



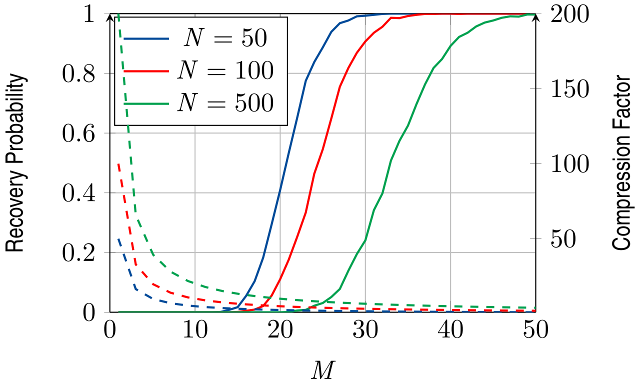
Phase Transition for OMP with $N = 500$

- Relative transition area significantly smaller for larger N



Phase Transitions for OMP with $K = 8$ and different N

- Required number of measurements M increases with N for constant K
- Nevertheless, compression ratio increases with N as M grows less than linear



Intermediate Summary

- Compressed Sensing allows compression of signals being sparse in some domain
- Compared to traditional compression techniques, complexity is shifted from encoder to decoder
- Sensing matrix has to fulfil certain constraints to ensure reconstruction
- Underdetermined linear equation system to be solved
  - Exploiting sparsity necessary to obtain unique solution
  - Due to sparsity constraint, optimization problem becomes NP hard
  - Relaxation to ℓ_1 norm turns optimization problem to become convex
  - We introduced greedy OMP algorithms for solving original problem