Advanced Topics in Signal Processing
Lectures Series on Compressed Sensing

January 2017, Ilmenau
Advanced Topics in Signal Processing
Lectures Series on Compressed Sensing

January 2017, Ilmenau

Dr.-Ing. Florian Römer
on behalf of...
Prof. Giovanni Del Galdo

With contributions from...
Anastasia Lavrenko
Alexandra Craciun
Magdalena Prus
Mohamed Gamal Ibrahim
Roman Alieiev

http://tu-ilmenau.de/dvt
Outline

- PART I
  - Fundamentals of Compressed Sensing

- PART II
  - Advanced aspects of Compressed Sensing
Outline Part I

- Introduction
  - Problem Statement
  - Data Model
  - Motivation

- Deriving the Compressive Sensing Approach
  - Designing the measurement kernel
  - Reconstruction strategy

- Compressed Sensing in Practice
  - Limitations
  - Exemplary Applications

- Conclusions
Goals of the lecture (Part I)

- Answer the following questions:
  - What is Compressed Sensing (CS)?
  - How and why does it work?
  - Which signals can be treated via CS?
  - How to design the measurement kernel without loss of information?
  - How to recover the desired signal?
Can we avoid the redundancy when taking the measurements?

David Brady: „One can regard the possibility of digital compression as a failure of sensor design. If it is possible to compress measured data, one might argue that too many measurements were taken.”
Compressed Sensing: The Big Picture

- Compressed sensing: a new paradigm in sampling theory
  - finds applications in almost all fields of science and engineering

- History
  - from 2004 important breakthroughs of Donoho, Candès, Tao, et al.
  - earlier important results in multidimensional geometry, information theory, and seismology

- Status
  - > 100k papers published
    - “Compressed Sensing” by Donoho: 20k citations in 10 y, 6th-most cited Paper in the entire IEEE library!
  - special sessions appearing at international conferences
  - substantial research grants
  - exponentially growing number of patents
Compressed Sensing: The Big Picture

Published items per year

source: google scholar

- searching for compressed sensing and sparsity publications
Compressed Sensing: The Big Picture

- Special Sessions and Paper Tracks at most major conferences on
  - signal processing: ICASSP, EUSIPCO, Asilomar SSC, …
  - communications: ICC, Globecom, …
  - information theory: ITW, ISIT, …
- Special Issues in Journals and Magazines
  - IEEE Journal of Selected Topics in Signal Processing: vol. 4, no. 2, Apr 2010
- Dedicated Workshops and Conferences
  - CSSIP Workshop on Compressed Sensing, Sparsity and Inverse Problems
  - ICNAAM Symposium on Sparse Approximation and Compressed Sensing
  - EURASIP Workshop on Sparsity and Compressive Sensing
  - International Conference of Compressive Sensing
  - International workshop on Compressed Sensing applied to RADAR
  - Local workshops at many major universities (Duke, UCLA, Edinburgh, Manchester….)
Traditional sampling and sparsity

- Traditional sampling: one value at one time
- At which rate?
  - Shannon-Nyquist: Bandlimited to $\pm B$, periodic with $T$: $2BT$ samples.
  - It can be hard!
  - It can be wasteful!

Figure from: A. Khilo et.al. “Photonic ADC: overcoming the bottleneck of electronic jitter”, Optics Express, Volume 20, Issue 4, pp. 4454-4469 (2012)
Traditional sampling and sparsity

- Traditional sampling: one value at one time
- At which rate?
  - Shannon-Nyquist: Bandlimited to $\pm B$, periodic with $T$: $2BT$ samples.
  - How would we know we sampled too high?
    - zeros in the spectrum ("sparsity").
    - reduce rate until $\#$samples $\approx \#$nonzeros
- Do nonzeros have to be around $f = 0$?
  - no, e.g., bandpass sampling
- But then, what if the pattern of (non)zeros is
  - even more irregular
  - unknown (but sparsity is known)
  - in another domain?

Unified framework: Compressive Sensing
Compressed Sensing

- Relevant advantages
  - Hardware complexity ↓, frame rate ↑, acquisition time ↓, accuracy ↑
  - Flexibility, adaptivity

- Two reasons for success: the sensing trick, the reco trick.
Assume a Sparse Analog Signal...

- \( s(t) \) uses only \( K \) degrees of freedom
- The signal \( s(t) \) can be expressed as a \( K \)-sparse vector \( x \) in the basis \( A \)

\[ s(t) \begin{bmatrix} A \in \mathbb{R}^{\alpha \times N} \end{bmatrix} \begin{bmatrix} x \in \mathbb{R}^N \end{bmatrix} \]

\[ K = \text{supp}\{x\} \]
Classical Approach to Measurements and Compression

- Classical approach: Assume band-limitation and measure at Nyquist-rate or more

- Problems:
  - N samples need to be measured, although we know that the signal possesses a much more compact representation
  - compression is achieved via post-measurement processing
The measurement kernel comprises \( N \) functions.

Assuming band limited functions, \( s \) contains all the information available.

\[ s \in \mathbb{R}^N \]

\[ \mathcal{A} \in \mathbb{R}^{\alpha \times N} \]

\[ x \in \mathbb{R}^N \]
Compressed Sensing Approach (analog signals)

- Compressive Sensing Solution: The signal $s(t)$ is measured by a fixed kernel which achieves compression
  - efficiently, i.e., achieving compression
  - without knowledge of the sparsifying basis $\mathcal{A}$ at the encoder
  - without loss of information, i.e., the reconstruction can retrieve the signal $s(t)$ perfectly as it delivers the correct sparse vector $x$
Compressive Sensing Approach

\[ d \in \mathbb{R}^M \]

\[ \mathcal{A} \in \mathbb{R}^{\alpha \times N} \]

\[ x \in \mathbb{R}^N \]

- \( M<<N \) measurements of \( s(t) \) are taken
- The functions in the kernel need to be chosen carefully
Compressed sensing is often studied in a purely discrete setting. To do so, we assume a Nyquist-rate sampling prior to any further processing.
Mathematical Model of the CS Measurement

\[ d \in \mathbb{R}^M \quad \Phi^T \in \mathbb{R}^{M \times N} \quad s \in \mathbb{R}^N \]

- Assumption: linearity
- We measure by applying the $N \times M$ matrix $\Phi$
The Sparsity Assumption

- $s \in \mathbb{R}^N$
- $A \in \mathbb{R}^{N \times N}$
- $x \in \mathbb{R}^N$

$s$ uses only $K$ degrees of freedom of the $N$ available.

The signal $s$ can be expressed as a $K$-sparse vector $x$ in the basis $A$. 

\[ K = \text{supp}\{x\} \]
Example: Sparsity in Frequency Domain: the iDCT

\[ s \in \mathbb{R}^N \quad A \in \mathbb{R}^{N \times N} \quad x \in \mathbb{R}^N \]

- \( x \) contains the coefficients in frequency domain
- \( s \) is the signal in time domain sampled at Nyquist-rate
Example: Sparsity in Frequency Domain: the iDCT

atomic functions $A(:, k)$

signal $s$
Sparsity in Time Domain

- The sparsifying matrix $A$ is an identity matrix.
- Trivial scenario; the signal $s$ is already sparse.
Overcomplete Basis

\[ s = A \cdot x \]
Overcomplete Basis
Complete Data Model of the Encoder

- Goal: be able to reconstruct $s$ from $d$ with $M$ as small as possible
- Ideally: $M \approx K = \text{supp}\{x\}$
Compressed Sensing Approach

\[ s \in \mathbb{R}^N \quad d \in \mathbb{R}^M \quad x \in \mathbb{R}^N \]

- Compressive Sensing Solution
  - The signal \( s \) can be measured by a fixed matrix \( \Phi \)
    - non-adaptively, i.e., without knowing the right \( A \) at the encoder
    - efficiently, i.e., with \( M \ll N \) and close to \( K \)
    - in a lossless manner, i.e., without losing information
  - Perfect reconstruction is possible only via non-linear methods
Outline Part I

- Introduction
  - Problem Statement
  - Data Model
  - Motivation

- Deriving the Compressive Sensing Approach
  - Designing the measurement kernel
  - Reconstruction strategy

- Compressed Sensing in Practice
  - Limitations
  - Exemplary Applications

- Conclusions
Deriving the Compressive Sensing Approach

- **Question 1**
  - How do we guarantee that the information reaches \( d \)? How do we make \( x \) observable by \( d \)?
  - Designing the measurement kernel \( \Phi \)

- **Question 2**
  - How do we use the information contained in \( d \) to retrieve \( x \)?
  - The reconstruction strategy
Deriving the Compressive Sensing Approach

- **Question 1**
  - How do we guarantee that the information reaches $d$? How do we make $x$ observable by $d$?

- **Designing the measurement kernel $\Phi$**

- **Question 2**
  - How do we use the information contained in $d$ to retrieve $x$?

- **The reconstruction strategy**

$$d \overset{\Phi}{=} x$$
Mutual Coherence

- **Definition**

\[ \mu(\Phi, A) = \sqrt{N} \cdot \max_{k,h} |\varphi_k^H \cdot a_h| \]

- **Assumptions**

\[ \|\varphi_k\|^2_{\ell_2} = \|a_k\|^2_{\ell_2} = 1 \quad \forall k \]
\[ A^H A = A A^H = I_N \]

- **Property**

\[ 1 \leq \mu(\Phi, A) \leq \sqrt{N} \]
Geometrical Interpretation of Incoherence

\[ 1 \leq \mu(\varphi, A) \leq \sqrt{N} \]

maximally incoherent

maximally coherent
Maximally Coherent $\Phi$ - The Worst Choice

Let us select $M$ specific atomic functions from $A$

$M << N$, therefore, we cannot pick all
Maximally Coherent $\Phi$ - The Worst Choice

Unless we know where the non-sparse elements are, this approach fails miserably
Maximal Coherence and Full Basis: Transform Coding

Revealing the structure underneath \( s \) (i.e., \( A \)) allows us to find the most compact description of the phenomenon.

Before reducing the description to \( K \) coefficients, all measurements \( M = N \) have to be computed!
Maximal Coherence and Full Basis: Transform Coding

Revealing the structure underneath $s$ (i.e., $A$) allows us to find the most compact description of the phenomenon.

Before reducing the description to $K$ coefficients, all measurements $M = N$ have to be computed!
Maximally Coherent $\Phi$ - The Worst Choice

- With a fully coherent measurement kernel, the vector $\mathbf{x}$ is observable by $\mathbf{d}$ only in special cases.

\[
\mathbf{x} = [1, \ 0]^T
\]

\[
A = I_2
\]

\[
\mathbf{x} = [0, \ 1]^T
\]
Maximally Incoherent $\Phi$ – The Best Choice

With an incoherent measurement kernel, the vector $x$ is observable by $d$ in all cases.

$$x = [1, \ 0]^T$$

$$x = [0, \ 1]^T$$

$$A = I_2$$
Random Matrices as a Measuring Kernels

- How to build incoherent bases without knowing $A$?
- "Randomness" is incoherent with any "structure"!

![Diagram showing random vector and PDF for different values of N]
Random Matrices as a Measuring Kernels

- Random matrices achieve very high incoherence wrt any basis $A$, especially for larger $N$

\[
\tilde{\varphi}_{k,h} \text{ i.i.d. } \mathcal{N}(0, 1)
\]

\[
\tilde{\varphi}_{k,h} \text{ i.i.d. } \pm 1
\]
Recap

- **Encoding Strategy**
  - Obtain $M << N$ measurements as linear functionals of the vector $s = Ax$
  - The measurement kernel $\Phi$ should be maximally incoherent with $A$
  - Without knowing $A$, this is achieved by introducing randomness

$$d = \Phi^T A x$$
Compressed Sensing

- Relevant advantages
  - Hardware complexity↓, frame rate↑, acquisition time↓, accuracy↑
  - Flexibility, adaptivity
- Two reasons for success: the sensing trick, the reco trick.
Deriving the Compressive Sensing Approach

- Question 1
  - How do we guarantee that the information reaches $d$? How do we make $x$ observable by $d$?
  - Designing the measurement kernel $\Phi$

- Question 2
  - How do we use the information contained in $d$ to retrieve $x$?
  - The reconstruction strategy
Reconstruction Strategy – Trivial Approach

- Basically, the LSE \( d = \Phi^T \cdot A \cdot x \) has infinitely many solutions.
- The sparse one is found by solving the problem

\[
\begin{align*}
x^* &= \arg \min_x \supp \{x\} \\
\text{s.t.} \quad d &= \Phi^T \cdot A \cdot x
\end{align*}
\]

- \( A \) must be known at the reconstruction!
- On the contrary, recall that \( A \) may be unknown at the sensor
- It’s an NP-hard problem
Complexity of the \( \ell_0 \) search

- For the \( \ell_0 \)-problem we need to test all possible support sets
  - \( K = 1 \): \( N \) possible choices
  - \( K = 2 \): \( N \) (\( N-1 \)) possible choices
  - …
  - \( k = 1, 2, \ldots, K \): \( O(N^K) \) possible choices
    - Prohibitive: doubling \( N \) increases complexity by \( 2^K \), \( N \) can be large.
Reconstruction Strategy – CS Approach

- The sparse vector $x$ is reconstructed by solving a convex L1 optimization problem

$$x^* = \arg \min_x \|x\|_{\ell_1}$$

subject to

$$d = \Phi^T \cdot A \cdot x$$

- $A$ must be known at the reconstruction!
- On the contrary, recall that $A$ may be unknown at the sensor
- Different $A$’s lead to different reconstructions
Why is an L1 Problem Efficiently Solvable?

- L1 can be recast as a linear problem
  \[ \mathbf{x}^* = \arg \min_{\mathbf{x}} \| \mathbf{x} \|_{\ell_1} \]
  \[ \text{s.t.} \quad d = \Phi^T \cdot A \cdot \mathbf{x} \]

- Every linear inequality constraint defines a half-space
  - search region becomes an intersection of half-spaces ("polytope")
  - optima are always at the intersection points
  - efficient search algorithms exist, e.g., Interior point method (polynomial-time)

\[ \mathbf{x}^* = \arg \min_{\mathbf{x}, \mathbf{z}} \mathbf{1}^T \cdot \mathbf{z} \]
\[ \text{s.t.} \quad \mathbf{x} \preceq \mathbf{z} \]
\[ \mathbf{x} \succeq -\mathbf{z} \]
\[ d = \Phi^T \cdot A \cdot \mathbf{x} \]
Complexity Comparison

- For the L0-problem we need to test all possible support sets
  - $K = 1$: $N$ possible choices
  - $K = 2$: $N(N-1)$ possible choices
  - ...
  - $k = 1, 2, \ldots, K$: $O(N^K)$ possible choices
    - Prohibitive: doubling $N$ increases complexity by $2^K$, $N$ can be large.

- For the L1-problem, polynomial-time algorithms are available
  - $O(N^3 \log(N))$ (Basis Pursuit)
    - For large $N$ the polynomial-time nature is crucial
  - Even more so, approximate L1 algorithms exist with $O(N^2)$ or even $O(NK)$
L0 – L1 Equivalence and Uniqueness (I)

- \( x \), solution to the L1 problem, is also the unique solution to the L0 problem iff
  - \( x \) is \( K \) sparse
  - \( \Phi^T \cdot A \) satisfies the null space property of order \( K \)

\[
\forall v : v \in \text{null}(\Phi^T \cdot A) \quad \| v_{[\text{supp}\{x\}]} \|_1 > \| v_{[\text{supp}\{x\}]} \|_1
\]

\[
x = \begin{pmatrix} \vdots \end{pmatrix} \quad v = \begin{pmatrix} \vdots \end{pmatrix} \quad > \begin{pmatrix} \vdots \end{pmatrix} \quad \ell_1 \quad \ell_1
\]
L0 – L1 Equivalence and Uniqueness (II)

- The null space property is guaranteed if the RIP holds with \( \delta_{2K} \leq \frac{1}{3} \)

\[
(1 - \delta_K) \| x \|_{\ell_2}^2 \leq \| \Phi^T A x \|_{\ell_2}^2 \leq (1 + \delta_K) \| x \|_{\ell_2}^2
\]

\( \forall \ K\text{-sparse } x \)

- Restricted Isometry Property: wrt sparse vectors the matrix \( \Phi^T A \) behaves almost as a unitary matrix, i.e., length preserving transformation.
L0 – L1 Equivalence and Uniqueness (III)

If:
- $\Phi$ is chosen randomly...
- $N$ is large
- $M \gtrsim \mu^2(\tilde{\Phi}, A) \cdot \text{supp}\{x\} \cdot \log_2(N)$

then:
- the RIP holds with overwhelming probability
- and therefore $x$ can be reconstructed by solving an L1 problem

Non-linear Sampling Theorem (Candès et.al 2006)
Identifiability: L0 vs. L1

- In general, for L0 we have

\[ K < \frac{1}{2}(K\text{-rank}(\Phi^T A) + 1) \leq \frac{M + 1}{2} \]

\[ M \geq 2K + 1 \]

NB: there are 2K degrees of freedom!
(support indices + amplitudes)

- whereas for L1 the bounds are of the form

\[ M \geq C \cdot K \cdot \log(N/K) \]

L1 “penalty”

Kruskal-rank ≥ r if all sets of r columns are linearly independent
Solution Set for the Equality Constraint

\[ d = \phi^T \cdot A \cdot x \]

\[ d = b^T \cdot x \]

\[ d = b_1 \cdot x_1 + b_2 \cdot x_2 \]

\( x \) must lie on a line
Lp-Norms and Lp-Balls

- Definition of Lp-norm: \[ \| \mathbf{x} \|_{L^p} = \left( \sum_{k=1}^{N} |x_k|^p \right)^{\frac{1}{p}} \]

Lp-Balls for \( N = 2 \)

- \( p = 0.5 \)
- \( p = 1 \)
- \( p = 2 \)
- \( p = 3 \)
$p = 0.3$
Minimization Problem: $p=2$
Minimization Problem: $p=2$
Minimization Problem: $p=2$

This is the LS solution

\[ x_{LS} = (\Phi^T A)^\dagger d \]
Minimization Problem: $p=0.5$
Minimization Problem: $p=0.5$
Minimization Problem: $p=0.5$

The “arms” of the $L_{0.5}$ ball reach out making sparse solutions favored wrt non-sparse
Minimization Problem: $p=1$

- Same effect as $L^{0.5}$, although not as prominent
Minimization Problem: $p=1$
Minimization Problem: $p=1$
Escape Velocities for the L1-ball

- Vertices move faster than edges, which move faster than sides
- They correspond to 1-sparse, 2-sparse, and 3-sparse respectively
Choice of p-Norm

- **p > 1** lead to non-sparse solutions
- **p ≤ 1** lead to sparse solutions
  - p = 0 = cardinality
    - logical choice, sparsest solution
    - combinatorial NP-hard problem
  - 0 ≤ p < 1
    - norms are non-convex
  - p = 1
    - Convex hull of p < 1 norm
    - identical to p = 0 under some conditions
    - linear and convex problem
Recap

- Reconstruction Strategy

*IF*

- the non-linear sampling theorem holds
  - i.e., $M$ is large enough
  - $A$ is known (at the decoder)

*THEN*

- $s$ can be perfectly reconstructed by solving a convex optimization problem in L1