COMPRESSED SENSING, OR HOW TO GET MORE FROM LESS

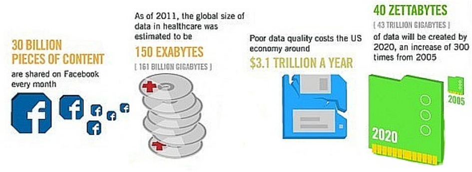
Irena Stoyanoska (Bojarovska)

Technische Universität Berlin
Airberlin

Algorithms & Data Chalenges Berlin
11 July 2016
Berlin

Berlin Mathematical School

The key idea of compressed sensing

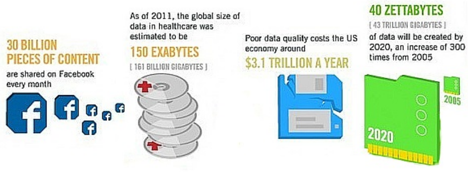


Source: IBM; 2008

"Measure what can be measured."

Galileo Galilei (1564 - 1642)

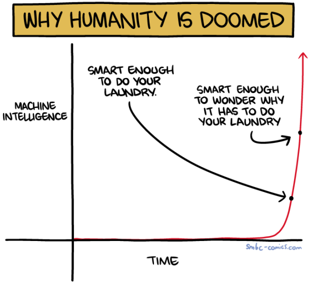
The key idea of compressed sensing



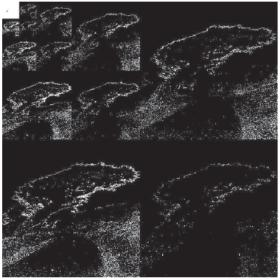
Source: IBM; 2008

"Measure what should be measured."

Thomas Strohmer, 2012



Why Compressed sensing?



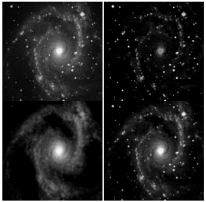
Being sparse is natural!

$\boxed{\text{First acquire}} \longrightarrow \boxed{\text{Then compress}}$

Can we directly acquire just the useful part of the signal?

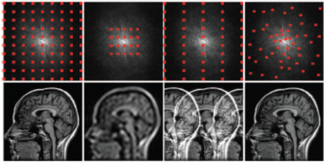
Why is this important?

- Hardware design: MRI, astronomy, imaging, radar and sonar signal processing
- Processing non-conventional signals: high-dimensional data, graph-based data, structured data



Morphological component analysis

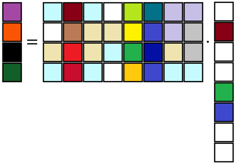
Source: Starck, Donoho, Candès; 2002



6 times faster MRI

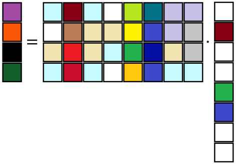
Source: Donoho, Lustig, Pauly; 2007

The mathematics behind it



$x \in \mathbb{R}^n-$ is the signal we are interested in,
$A \in \mathbb{R}^{m \times n} = \{a_i\}_{i=1}^m$ is *the measurement matrix*,
$y \in \mathbb{R}^m-$ are the *observations* we make, $y_i = \langle a_i, x \rangle$.

The mathematics behind it

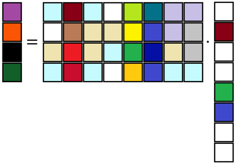


$x \in \mathbb{R}^n-$ is the signal we are interested in,
$A \in \mathbb{R}^{m\times n} = \{a_i\}_{i=1}^m$ is *the measurement matrix*,
$y \in \mathbb{R}^m-$ are the *observations* we make, $y_i = \langle a_i, x\rangle$.

Conventional Linear Algebra vs. Compressed Sensing

- $m \geq n$ in order to have a determined system.

The mathematics behind it



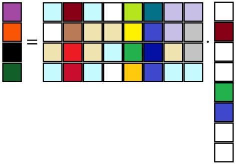
$x \in \mathbb{R}^n -$ is the signal we are interested in,
$A \in \mathbb{R}^{m \times n} = \{a_i\}_{i=1}^m$ is *the measurement matrix*,
$y \in \mathbb{R}^m -$ are the *observations* we make, $y_i = \langle a_i, x \rangle$.

Conventional Linear Algebra vs. Compressed Sensing

- $m \geq n$ in order to have a determined system.
- $m \geq k \log n$ for exact recovery via convex linear programming if x is sparse, i.e. has $k < m$ non-zero entries.

The mathematics behind it

$x \in \mathbb{R}^n-$ is the signal we are interested in,
$A \in \mathbb{R}^{m \times n} = \{a_i\}_{i=1}^m$ is *the measurement matrix*,
$y \in \mathbb{R}^m-$ are the *observations* we make, $y_i = \langle a_i, x \rangle$.



Conventional Linear Algebra vs. Compressed Sensing

- $m \geq n$ in order to have a determined system.
- $m \geq k \log n$ for exact recovery via convex linear programming if x is sparse, i.e. has $k < m$ non-zero entries.

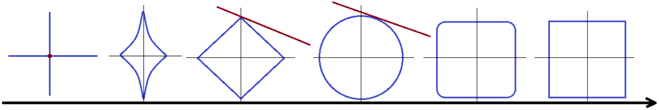
Sparsity is the crucial assumption!

"Strong" sparsity

$x \in \mathbb{R}^n$ is called k-sparse, if

$$\|x\|_0 := |supp(x)| = |\{i : x_i \neq 0\}| = k.$$

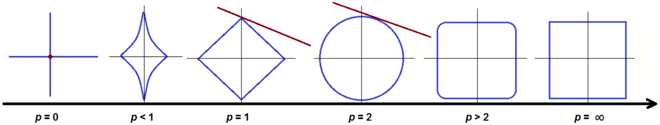
From Compressed Sensing to Machine Learning



ℓ_0- minimization

$$\min \|x\|_0 \quad \text{subject to} \quad Ax = y. \qquad (P_0)$$

From Compressed Sensing to Machine Learning



ℓ_0- minimization

$$\min \|x\|_0 \quad \text{subject to} \quad Ax = y. \tag{P_0}$$

This problem is NP-hard. We can relax this problem and take

ℓ_1- minimization

$$\min \|x\|_1 \quad \text{subject to} \quad Ax = y. \tag{P_1}$$

If we allow some error in the measurement process, we get to:

LASSO Regression!

$$\min \lambda\|x\|_1 + \|Ax - y\|_2^2. \tag{$P_{1,2}$}$$

Does this work?

Definition

Let $A \in \mathbb{R}^{m \times n}$. Then A has the restricted isometry property (RIP) of order k, if there exists a $\delta_k \in (0,1)$ such that

$$(1-\delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\|x\|_2^2 \quad \text{for all } k\text{-sparse } x \in \mathbb{R}^n$$

Does this work?

Definition

Let $A \in \mathbb{R}^{m \times n}$. Then A has the restricted isometry property (RIP) of order k, if there exists a $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k)\|x\|_2^2 \quad \text{for all } k\text{-sparse } x \in \mathbb{R}^n$$

Theorem (Cohen, Dahmen, DeVore; 2008, Candes; 2008)

Let $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2} - 1$. Let $x \in \mathbb{R}^n$, and let x^* be a solution of the associated ℓ_1 problem (P_1). Then

$$\|x - x^*\|_2 \leq C \cdot \frac{\sigma_k(x)_1}{\sqrt{k}},$$

for some constant C dependent on δ_{2k}.

Here $\sigma_k(x)_1 := \min_{\tilde{x} \in \Sigma_k} \|x - \tilde{x}\|_1$ is the error of the best k term approximation.

Research directions

Are there matrices that satisfy the RIP property and allow us to recover x with as few as possible measurements?

- Gaussian - entries are independent realization of $\mathcal{N}(0, \frac{1}{m})$.
- Block $m \times 2m$ *matrix with blocks Fourier and Dirac bases*

Research directions

Are there matrices that satisfy the RIP property and allow us to recover x with as few as possible measurements?

- Gaussian - entries are independent realization of $\mathcal{N}(0, \frac{1}{m})$.
- Block $m \times 2m$ *matrix with blocks Fourier and Dirac bases*

Are there efficient algorithm for solving the minimization problems (P_1) *and* $(P_{1,2})$*?*

- Basis pursuit (Lasso regression)
- Greedy algorithms (Orthogonal matching pursuit, thresholding-based methods)

Research directions

Are there matrices that satisfy the RIP property and allow us to recover x with as few as possible measurements?

- Gaussian - entries are independent realization of $\mathcal{N}(0, \frac{1}{m})$.
- Block $m \times 2m$ *matrix with blocks Fourier and Dirac bases*

Are there efficient algorithm for solving the minimization problems (P_1) and $(P_{1,2})$?

- Basis pursuit (Lasso regression)
- Greedy algorithms (Orthogonal matching pursuit, thresholding-based methods)

What to do if my signal/data is not really sparse?

- Find (learn) a sparse representation (in a dictionary)
- Discover some some structure (geometric sparsity, block sparsity)
- Go for extensions in higher dimensions: sparse (low rank) matrices [the Netflix problem]

Applications of Compressed Sensing: Data Separation

Morphological Component Analysis: image decomposition method which uses sparse representations of the components and the compressed sensing idea.

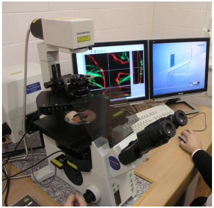
$$\min_{c_1, c_2} \|c_1\|_1 + \|c_2\|_1 \text{ subject to } x = \begin{bmatrix} \Phi_1 & \Phi_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

Applications of Compressed Sensing: Data Separation

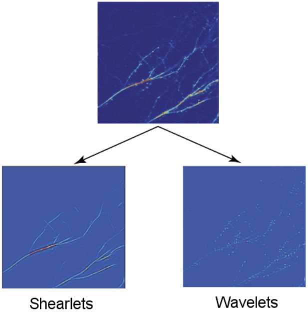
Morphological Component Analysis: image decomposition method which uses sparse representations of the components and the compressed sensing idea.

$$\min_{c_1, c_2} \|c_1\|_1 + \|c_2\|_1 \text{ subject to } x = \begin{bmatrix} \Phi_1 & \Phi_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

Example: Detection of characteristics of Alzheimer: separation of spines and dendrites



Confocal laser scanning microscopy



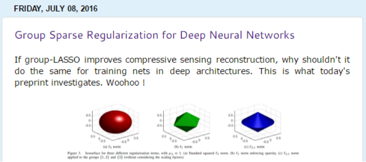
Kutyniok, Lim; 2010

Summary and take away

- "More from less" is possible
  Compressed sensing allows to recover the sparse signal from a small set of (linear) non-adaptive measurements in an efficient manner
- Sparsity is all around - search for it in your models.
  Many signals from various application fields are sparse or admit sparse representation.
- It is worth to learn the language of compressed sensing and sparse representations
  Many well-known machine learning problems can be seen from a different light and new ideas can be found.

THANK YOU!

- Nuit Blanche - great informative blog on compressed sensing
  A blog about Compressive Sensing, Computational Imaging, Machine Learning. Using priors to avoid the curse of dimensionality arising in Big Data. http://nuit-blanche.blogspot.com



- My contributions
  - Bojarovska I., Flinth A., Phase Retrieval from Gabor Measurements, J. Fourier. Anal. Appl. (2015)
  - Bojarovska I., Paternostro V., Gabor Fusion Frames Generated by Difference Sets, SPIE Proc., Wavelets and Sparsity XVI (2015)
  - Bojarovska I. Geometric Compressed Sensing and Structured Sparsity, PhD Thesis (TU Berlin, 2015)