Compressed Sensing Basics and Beyond

EUROCAST 2015

Feb 12, 2015

S.M. Fosson E. Magli

POLITECNICO DI TORINO

Department of Electronics and Telecommunications

CRISP

Towards Compressive Information Processing Systems

erc

Outline

1. Mathematical problem
2. Applications
3. Recovery
4. Distributed compressed sensing

Outline

Mathematical problem

Compressed sensing (compressed sampling, compressive sensing... CS) deals with

Underdetermined linear systems ...

$$Ax = y$$

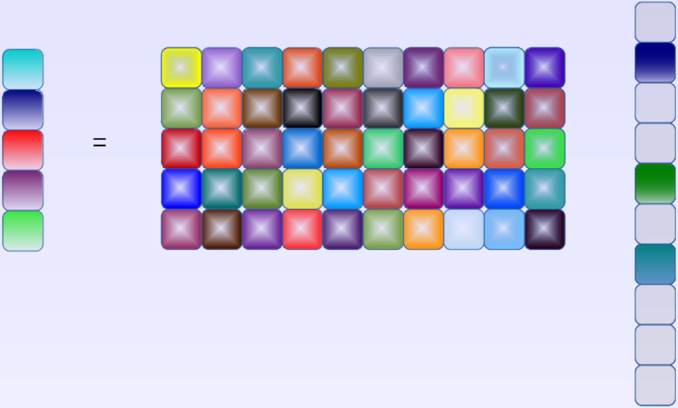
$x \in \mathbb{R}^n$ (unknown), $y \in \mathbb{R}^m$ (measurements), $A \in \mathbb{R}^{m\times n}$, $m < n$

Within the infinite set of solutions, CS looks for the sparsest one

... with sparsity assumptions

x is k-sparse, i.e., it has k non-zero components, where $k \ll n$

Mathematical problem



Questions

$$Ax = y, x \in \mathbb{R}^n (\text{sparse}), y \in \mathbb{R}^m, m < n$$

1. Is the problem well-posed (= is the solution unique)?
2. Are there feasible algorithms to find the solution?
3. Which applications motivate this study?

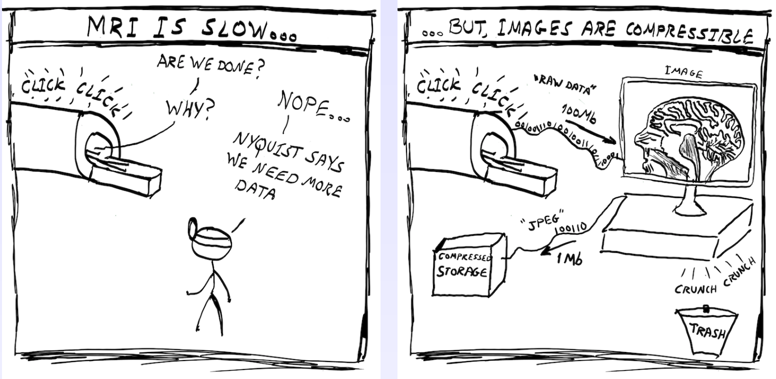
Answers

1. Yes, under some conditions
2. A number of recovery algorithms have been proposed
3.
   - Sparsity is ubiquitous: many signals are sparse in some basis ($y = A\phi x$ where ϕ is the sparsifying basis, e.g., DCT, wavelets, Fourier...)
   - Applications where data acquisition is difficult/expensive, and one aims to move the computational load to the receiver

Outline

1. Mathematical problem
2. Applications
3. Recovery
4. Distributed compressed sensing

Magnetic Resonance Imaging (MRI): acquisition is slow [Lustig (2012)]

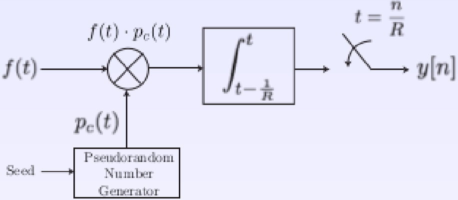


→ sense the compressed information directly

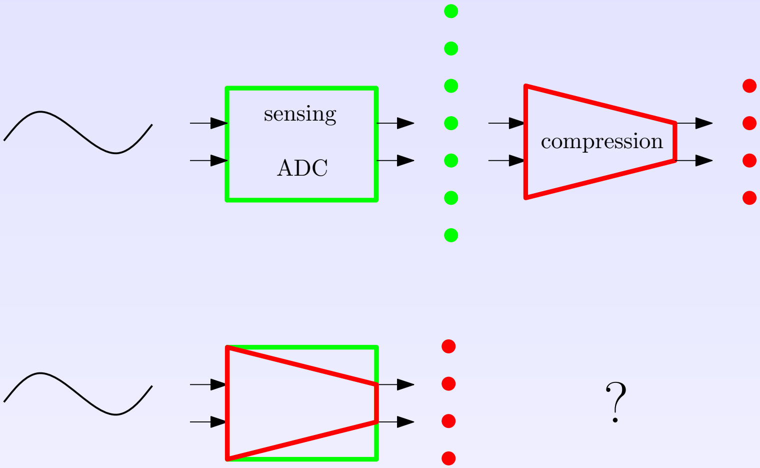
Compression and sampling

$$Ax = y, x \in \mathbb{R}^n(\text{sparse}), y \in \mathbb{R}^m, m < n$$

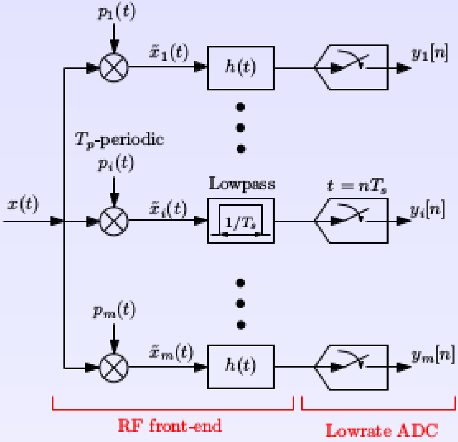
- Sampling: Nyquist-Shannon sampling theorem states given a signal bandlimited in (B, B), to represent it over a time interval T, we need at least $2BT$ samples
- CS indicates a way to merge compression and sampling, and sample at a sub-Nyquist rate [Tropp et al. (2009)]



Compression and sampling

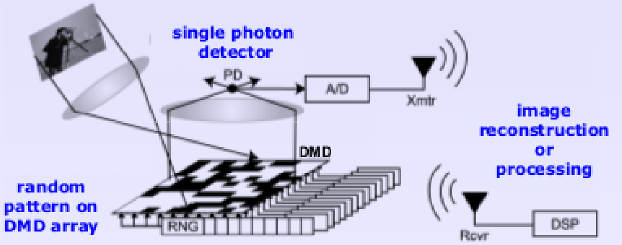


Modulated wideband converter (MWC) [Mishali and Eldar (2010)]



- Sub-Nyquist sampling for signals sparse in the frequency domain
- Realized in hardware (with commercial devices)

Single-pixel camera



Boufonos et al., ICASSP 2008

Key ingredient: a microarray consisting of a large number of small mirrors that can be turned on or off individually

Light from the image is reflected on this microarray and a lens combines all the reflected beams in one sensor, the single pixel of the camera

Outline

Mathematical formulation

ℓ_0-norm

$\|x\|_0 :=$ number of nonzeros entries of $x \in \mathbb{R}^n$

Natural formulation of the CS problem:

$$P_0 : \quad \min_{x \in \mathbb{R}^n} \|x\|_0 \text{ subject to } Ax = y$$

- Is the solution unique?
- P_0 is NP-hard!

Uniqueness of the solution

Spark

spark(A) := minimum number of columns of A that are linearly dependent

Theorem [D. Donoho, M. Elad (2003)]

For any vector $y \in \mathbb{R}^m$, there exists at most one k-sparse signal $x \in \mathbb{R}^n$ such that $y = Ax$ if and only if spark$(A) > 2k$.

Uniqueness of the solution

Coherence

$\mu(A) := \max_{i \neq j} \frac{|A_i^T A_j|}{\|A_i\|_2 \|A_j\|_2}$ ($A_i = i$th column of A)

Theorem [D. Donoho, M. Elad (2003)]

If

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right)$$

$y \in \mathbb{R}^m$, there exists at most one k-sparse signal $x \in \mathbb{R}^n$ such that $y = Ax$.

Possible solution: convex relation

Basis Pursuit

$$P_1: \quad \min_{x \in \mathbb{R}^n} \|x\|_1 \text{ subject to } Ax = y$$

- P_1 is convex; can be solved by linear programming
- Are P_0 and P_1 equivalent?

Coherence

Coherence

$\mu(A) := \max_{i\neq j} \frac{|A_i^T A_j|}{\|A_i\|_2 \|A_j\|_2}$ ($A_i = i$th column of A)

Theorem [Elad and Bruckstein (2002)]

If for the sparset solution $x^\star$ we have

$$\|x^\star\|_0 < \frac{\sqrt{2} - \frac{1}{2}}{\mu(A)}$$

then the solution of P_1 is equal to the solution of P_0.

Restricted Isometry Property (RIP)

RIP

Matrix A satisties the RIP of order k if there exists $\delta_k \in (0,1)$ such that the following relation holds for any k-sparse x:

$$(1-\delta_k)\,\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\,\|x\|_2^2$$

Theorem [Candès (2008)]

If $\delta_k < \sqrt{2}-1$, then for all k-sparse $x \in \mathbb{R}^n$ such that $Ax = y$, the solution of P_1 is equal to the solution of P_0.

Which matrices?

- Coherence, spark, RIP: not easy to compute
- Random matrices A with i.i.d. entries drawn from continuous distributions have $\text{spark}(A) = m + 1$ with probability one.
- Gaussian, Bernoulli matrices: given $\delta \in (0, 1)$ there exist c_1, c_2 depending on δ such that G. and B. matrices satisfy the RIP with constant δ and any $m \geq c_1 k \log(n/k)$ with probability $\geq 1 - 2e^{-c_2 m}$ [Baraniuk (2008)]
- Structured matrices: circulant matrices, partial Fourier matrices

Orthogonal Matching Pursuit (OMP)

- "When we talk about BP, we often say that the linear program can be solved in polynomial time with standard scientific software, and we cite books on convex programming [...]. This line of talk is misleading because it may take a long time to solve the linear program, even for signals of moderate length" [Tropp and Gilbert (2007)]
- Possible solution: greedy algorithm, fast, easy to implement $\rightarrow$ OMP

Orthogonal Matching Pursuit (OMP)

1. Initialize $r_0 = y$, $\Lambda_0 = \emptyset$
2. For $t = 1, \dots, T_{max}$
3. $\lambda_t = \underset{j=1,\dots,n}{\text{argmax}} |A_j^T r_{t-1}|$
4. $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$
5. $\widehat{x}_t = \underset{x \in \mathbb{R}^n}{\text{argmin}} \|y - A_{\Lambda_t} x\|_2$
6. $r_t = y - A_{\Lambda_t} \widehat{x}_t$

- $T_{max} \approx k$
- OMP requires the knowledge of k!

Variants of BP

Basis Pursuit Denoise (BPDN)

$$P_1: \quad \min_{x \in \mathbb{R}^n} \|x\|_1 \text{ subject to } \|Ax = y\|_2 \leq \varepsilon$$

Unconstrained version of BPDN

Lasso

$\min_{x \in \mathbb{R}^n} \left(\|Ax - y\|_2^2 + \lambda \|x\|_1 \right)$

For some $\lambda > 0$, Lasso and BPDN have the same solution (the choice of λ is tricky!)

Iterative soft thresholding (IST)

1. $\widehat{x}_0 = 0$
2. For $t = 1, \dots, T_{max}$
3. $\widehat{x}_t = \mathsf{S}_\lambda(\widehat{x}_{t-1} + \tau A^T(y - A * \widehat{x}_{t-1}))$

where the operator S_λ is defined entry by entry as
$\mathsf{S}_\lambda(x) = \mathrm{sgn}(|x| - \lambda)$ if $|x| > \lambda$, 0 otherwise

- IST achieves a minimum of the Lasso [Fornasier (2010)], and in many common situations such minimum is unique [Tibshirani (2012)]
- Faster method to get a minimum of the Lasso: alternating direction method of multipliers (ADMM)

Iterative hard thresholding

1. $\widehat{x}_0 = 0$
2. For $t = 1, \ldots, T_{max}$
3. $\widehat{x}_t = \mathrm{H}_k(\widehat{x}_{t-1} + A^T(y - A\widehat{x}_{t-1}))$

where the operator $\mathrm{H}_k(x)$ is the non-linear operator that sets all but the largest (in magnitude) k elements of x to zero [Blumensath (2008)]

Outline

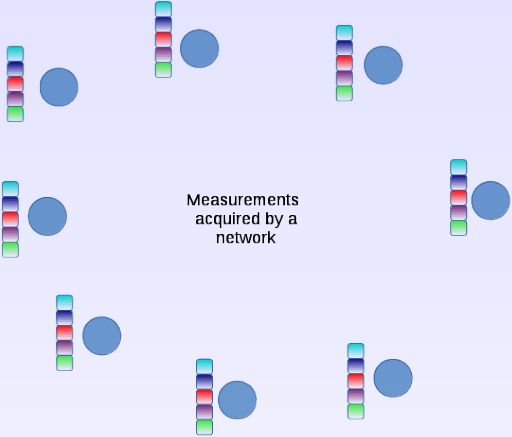
Distributed compressed sensing (DCS)

- Data acquisition is perfomed by a network of sensors

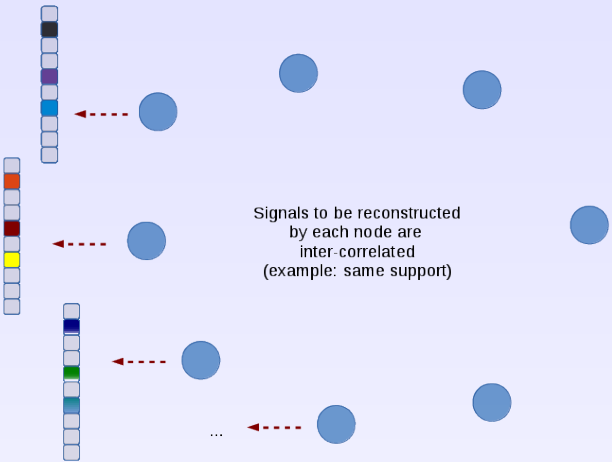
$$y_v = A_v x_v \qquad v \in \mathcal{V} = \{ \text{ sensors } \}$$

- First works: recovery is performed by a fusion center that gathers information from the network (sensing matrices, measurements)
- New: in-network recovery, exploiting local communication and consensus procedures
- We need iterative algorithms

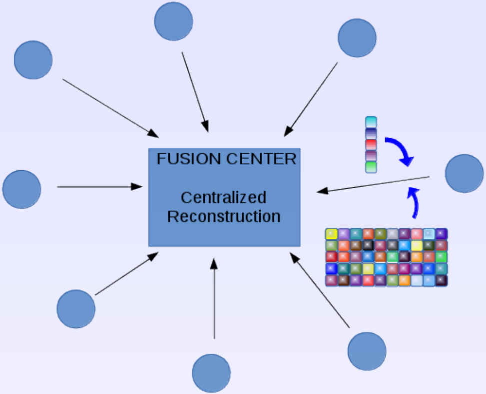
Distributed Compressed Sensing (DCS)



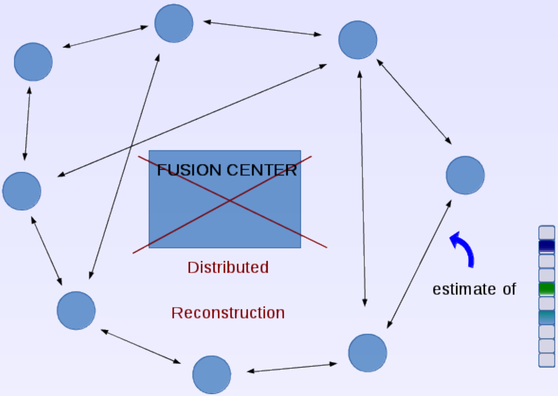
Distributed Compressed Sensing (DCS)



Distributed Compressed Sensing (DCS)



Distributed Compressed Sensing (DCS)



DCS: in-network recovery

1. C. Ravazzi, S.M. Fosson, E. Magli, E., Energy-saving Gossip Algorithm for Compressed Sensing in Multi-agent Systems, ICASSP, 2014
2. S.M. Fosson, J. Matamoros, C. Antón-Haro , E. Magli, Distributed Support Detection of Jointly Sparse Signals, ICASSP, 2014
3. J. Matamoros, S.M. Fosson, E. Magli, C. Antón-Haro, Distributed ADMM for in-network reconstruction of sparse signals with innovations, IEEE GlobalSIP, 2014
4. C. Ravazzi, S.M. Fosson, E. Magli, Distributed iterative thresholding for ℓ_0/ℓ_1-regularized linear inverse problems, IEEE Trans. Inf. Theory, 2015.

CS references

1. http://dsp.rice.edu/cs
2. E. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory, Feb. 2006
3. D. Donoho, Compressed sensing, IEEE Trans. Inf. Theory, Apr. 2006
4. A mathematical Introduction to Compressive Sensing, edited by S. Foucart and H. Rauhut, 2013
5. Compressed Sensing: Theory and Applications, edited by Y. C. Eldar and G. Kutyniok, 2012
6. Theoretical Foundations and Numerical Methods for Sparse Recovery, edited by M. Fornasier, 2010