Compressed sensing for distributed communications

INRIA, Oct. 16, 2012

Talk sponsored by EURASIP

Enrico Magli

Department of Electronics and Telecommunications
Politecnico di Torino (Italy)
enrico.magli@polito.it

Acknowledgments

This talk is part of the dissemination activities of ERC starting grant project 279848

CRISP

Towards compressive information processing systems

Credits

- Valerio Bioglio
- Giulio Coluccia
- Attilio Fiandrotti
- Sophie Fosson
- Chiara Ravazzi
- Aline Roumy
- Diego Valsesia

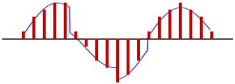
Outline

- Warm-up: distributed compressed sensing
- Distributed sensing: Joint reconstruction
- Distributed sensing: Distributed reconstruction
- Reconstruction of “big” images

Warm-up: distributed compressed sensing

Sampling + compression = no good

Theoretical foundations: Shannon's sampling theorem

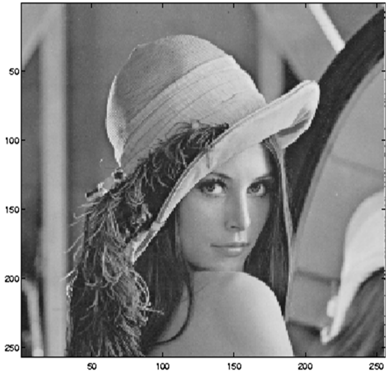


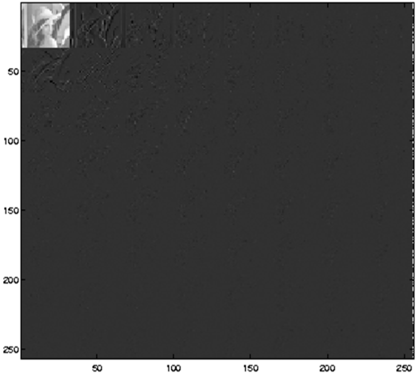
- Sampling the signal densely enough, exact reconstruction can be achieved
- Linear signal model: signal is bandlimited in $(-B, B)$
  - Nyquist criterion: to represent a signal over a time interval T, we need at least $2BT$ samples
  - This signal has $2BT$ degrees of freedom, i.e. different time or frequency components

Is that the "true" number of degrees of freedom?

Sparsity

Left: image; Right: discrete cosine transform of image





Sparsity

- $x \in \mathbb{R}^n$ is said to be k-sparse if it has at most k non-zero entries, i.e. $\|x\|_0 \leq k$
- In many practical cases, x has a sparse representation in some basis Φ, i.e. $x = \Phi c$, and $\|c\|_0 \leq k$.

Sparsity

- $x \in \mathbb{R}^n$ is said to be *k-sparse* if it has at most k non-zero entries, i.e. $\|x\|_0 \leq k$
- In many practical cases, x has a sparse representation in some basis Φ, i.e. $x = \Phi c$, and $\|c\|_0 \leq k$.

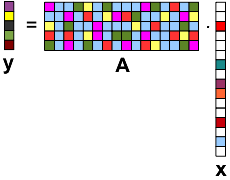
Sparsity in Italian: sparsità, parsimonia

Compressed sensing: linear acquisition model

Let $x \in \mathbb{R}^{n\times 1}$ be a signal.
Let $A \in \mathbb{R}^{m\times n}$, with $m \ll n$, be a sensing matrix.

We take linear projections of the signal as

$$y = Ax$$



with $y \in \mathbb{R}^{m\times 1}$.

- A is typically a random matrix
- If x is sparse in another domain: $y = Ax = A\Phi c \Rightarrow y = Bc$
- Underdetermined system $\Rightarrow$ infinitely many solutions

Quest for the sparsest solution

Out of all possible solutions, we would like to pick the one that is sparsest, i.e. solve the following problem:

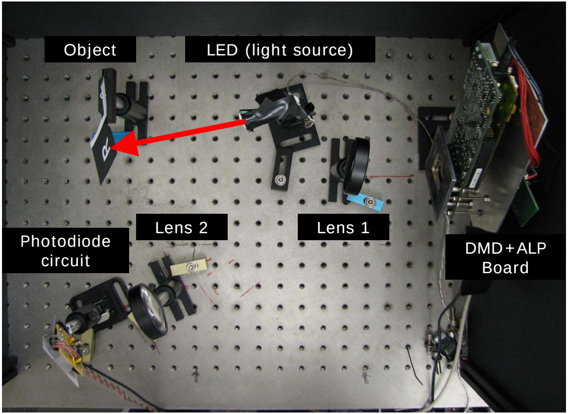
$$\min_{\widehat{x} \in \mathbb{R}^n} \|\widehat{x}\|_0 \text{ subject to } A\widehat{x} = y$$

Since this is NP-hard $\Rightarrow$ convexify (Basis Pursuit - BP):

$$\min_{\widehat{x} \in \mathbb{R}^n} \|\widehat{x}\|_1 \text{ subject to } A\widehat{x} = y$$

- Solutions to the BP problem still tend to be rather sparse, $O(n^3)$ complexity
- Model with noise: $y = Ax + n$, solve $\min_{\widehat{x} \in \mathbb{R}^n} \|\widehat{x}\|_1$ subject to $\|A\widehat{x} - y\|_2 \leq \epsilon$

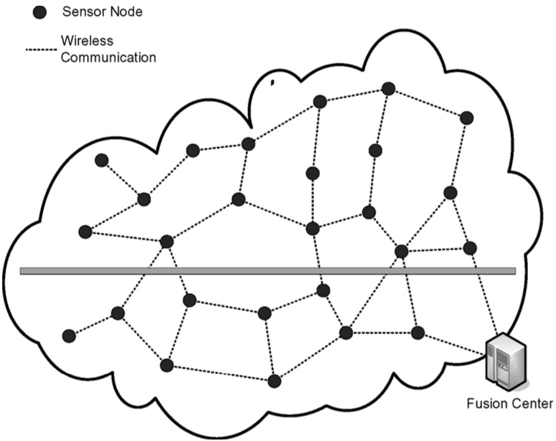
Single-pixel camera



Picture from Duarte *et al.*, "Single-pixel imaging via compressive sampling", *IEEE Sig. Proc. Mag.*, Mar. 2008

Distributed setting

In applications such as sensor networks, we have distributed acquisition of multiple correlated signals.



Signal models for distributed scenario

Joint sparsity model 1 (JSM-1): the signal acquired by the i-th sensor can be written as

$$x_i = x_C + \alpha x_i$$

with $i = 1, 2, \ldots, R$, $x_C = \Phi\theta_C$, $x_i = \Phi\theta_i$, $K_C = |\theta_C|$ and $K_i = |\theta_i|$.

- x_C is the "common" component, with K_C nonzero entries.
- x_i is the "innovation" component observed by sensor i, with K_i nonzero entries.
  - For a single x_i, we need around $c(K_C + K_i)$ measurements
  - Assuming $R = 2$, we expect something around $c(K_C + K_1 + K_2)$

D. Baron *et al.*, "Distributed compressed sensing", preprint available at www.arxiv.org

Problems in distributed compressed sensing

Challenge: taking advantage of intra- and inter-sensor correlations through proper design of sensing and reconstruction algorithms

- signal model (innovation, noise)
- coding model (quantization, entropy coding)
- availability of fusion center (joint vs. distributed reconstruction)
- application-specific aspects

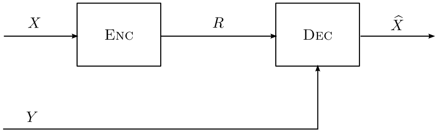
Distributed sensing: Joint reconstruction

Joint reconstruction with side information

- Scenario: signals acquired by a sensor network
  - correlated in time
  - correlated in space
- Compressed Sensing (CS)
  - captures intra-sensor correlation
  - allows a universal encoding scheme with low complexity
- Distributed (Slepian-Wolf) Source Coding (DSC)
  - captures inter-sensor correlation
  - allows efficient compression with no cooperation between the encoders

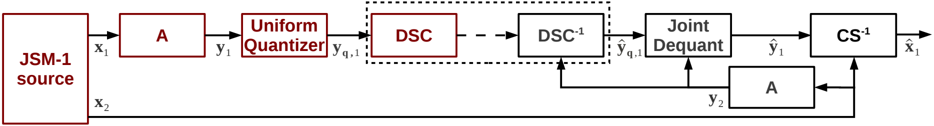
G. Coluccia *et al.*, "Lossy compression of distributed sparse sources: a practical scheme", Proc. Eusipco 2011.

Distributed Source Coding



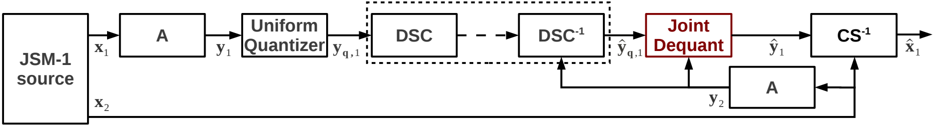
- DSC refers to the problem of compressing correlated i.i.d. sources X and Y without cooperation at their encoders
- Asymmetric case:
  - Y is available at the decoder
  - X is compressed at rate $H(X|Y)$ and recovered using Y as side information
- Lossless setup: uses channel codes (syndrome coding) at rate depending on the correlation between X and Y
- Lossy setup: adds a quantization stage

Proposed Algorithm – Quantization, CS, coding



- Linear measurement with Gaussian sensing matrix A
- Uniform scalar quantizer with step size Δ
- Slepian-Wolf coding stage
  - syndrome-base turbo codes

Proposed Algorithm – Joint Dequantization



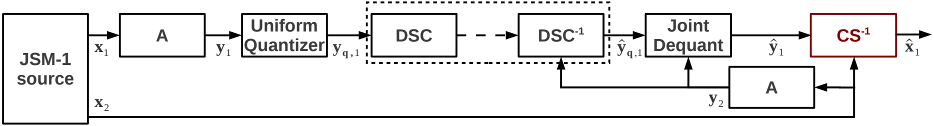
- Side information x_2 can be exploited for joint dequantization
  - Reconstruction: linear combination of midpoint reconstruction and x_2:

$$\widehat{y}_1 = \frac{\sigma_q^2}{\sigma_q^2 + \sigma^2} y_2 + \frac{\sigma^2}{\sigma_q^2 + \sigma^2} \widehat{y}_{q,1}$$

with clipping to quantization interval

$$\left[\widehat{y}_{q,1} - \frac{\Delta}{2} < \widehat{y}_1 < \widehat{y}_{q,1} + \frac{\Delta}{2} \right]$$

Proposed Algorithm – Joint Reconstruction

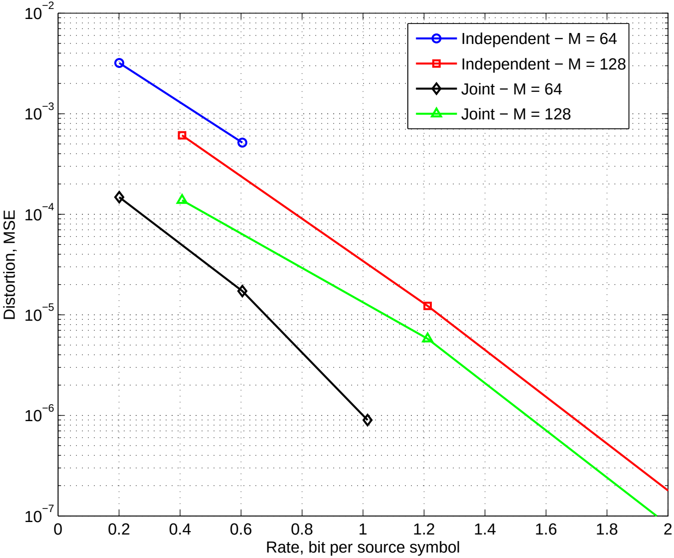


- Side information x_2 can be exploited for joint reconstruction

  1. Generate initial separate reconstruction $\widehat{x}_1$
  2. Estimate $\widehat{x}_C$ comparing x_2 and $\widehat{x}_1$
  3. Subtract $\widehat{y}_C = A\widehat{x}_C$ from the measurement vector $\widehat{y}_1$
  4. Reconstruct $\alpha x_{I,1}$: sparser than original x_1

Joint vs. Independent Reconstruction

$N = 512$, $K = 16$, $\Phi = \text{DCT}$, $\alpha = 10^{-2}$



Theoretical results: rate gain

- Result:
  - For $K, M, N \to \infty$ the elements of y_1 weakly converge to a Gaussian distribution
  - The entropy-constrained *operational* RD curve is

$$\lim_{R \to +\infty} \lim_{K,M,N \to +\infty} \frac{1}{\sigma_{y_1}^2} 2^{2R} D_{y_1}^{ec}(R) = \frac{\pi e}{6}$$

with $\sigma_{y_1}^2 = \sigma_A^2 \left[K_C \sigma_{\theta_C}^2 + \alpha^2 K_j \sigma_{\theta_j}^2 \right]$

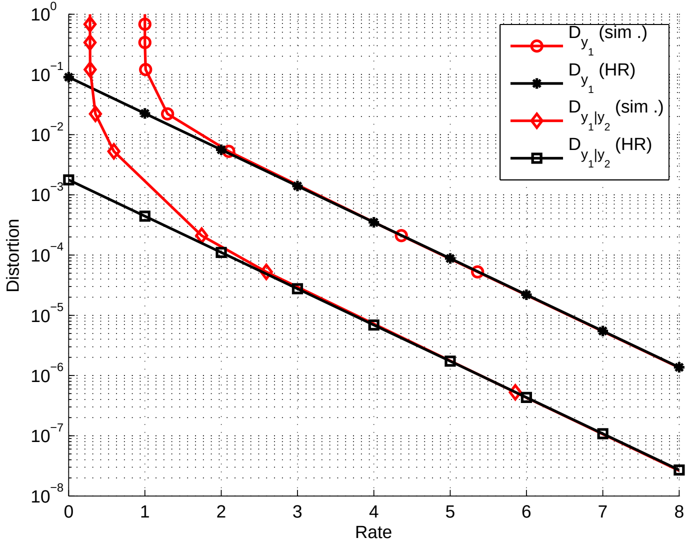
  - Using y_1 as side information, we have a rate gain R^* such that

$$\lim_{R \to +\infty} \lim_{K,M,N \to +\infty} \frac{1}{\sigma_{y_1}^2} 2^{2(R+R^*)} D_{y_1|y_2}^{ec}(R) = \frac{\pi e}{6}$$

with $R^* = -\frac{1}{2} \log_2 \left\{ 1 - \left[\left(1 + \alpha^2 \frac{K_1}{K_C} \frac{\sigma_{\theta_1}^2}{\sigma_{\theta_C}^2} \right) \left(1 + \alpha^2 \frac{K_2}{K_C} \frac{\sigma_{\theta_2}^2}{\sigma_{\theta_C}^2} \right) \right]^{-1} \right\}$

Validation: measurements domain

$N = 512$, $K = 16$, $\Phi = \mathrm{DCT}$, $\alpha = 10^{-1}$, $M = 128$



Theoretical results: reconstruction distortion

- Reconstruction performance depend on quantization noise added to measurements
- For CS reconstruction, we use the ideal oracle estimator

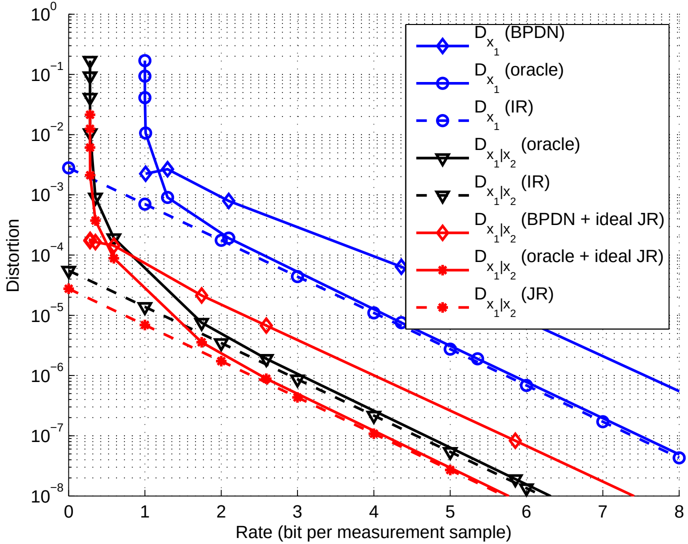
$$D_{\hat{x}_1}(R) = \frac{K_C + K_1}{N} D_{y_1}(R) \qquad D_{\hat{x}_1|x_2}(R) = \frac{K_C + K_1}{N} D_{y_1|y_2}(R)$$

- If we use the side information also for joint reconstruction, we can achieve an additional rate gain given by

$$R^{\mathrm{JR}} = \frac{1}{2} \log_2 \left(1 + \frac{K_C}{K_1}\right)$$

Validation: reconstruction

$N = 512$, $K = 16$, $\Phi = \mathrm{DCT}$, $\alpha = 10^{-1}$, $M = 128$



Distributed sensing: Distributed reconstruction

Distributed reconstruction algorithms

Classical distributed compressed sensing paradigm

- A sensor network acquires the compressed data;
- A fusion center collects the data and performs joint reconstruction.

Distributed compressed sensing with no fusion center

- Is it possible to perform reconstruction with no fusion center?
- Goal: distribute the reconstruction task over the network, taking into account the sensors' limited computational power.

Problem Statement

Model:

- a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - nodes in $\mathcal{V}$ are sensors
  - edges in $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ are available communication links
- set of observations

$$y_v = A_v x_0 + \xi_v \qquad v \in \mathcal{V}$$

  - $x_0 \in \Sigma_k \subseteq \mathbb{R}^n$ unknown signal
  - $A_v \in \mathbb{R}^{m \times n}$ with $m \ll n$
  - ξ_v i.i.d. Gaussian noise

LASSO regression: all data are collected in a fusion center

$$\widehat{x} = \operatorname{argmin} \sum_{v \in \mathcal{V}} \|y_v - A_v x\|_2^2 + 2\alpha \|x\|_1$$

Novel Distributed Approach

$$\min \sum_{v \in \mathcal{V}} \underbrace{\|y_v - A_v x\|_2^2 + 2\alpha \|x\|_1}_{f_v(x)}$$

Idea: Develop an iterative procedure that envisages cooperation among sensors.

At each step, the sensors

- move towards the minimum of their own LASSO functionals $f_v(x)$;
- improve their current estimate by sharing information with their neighbors.

Numerical results

Experiments

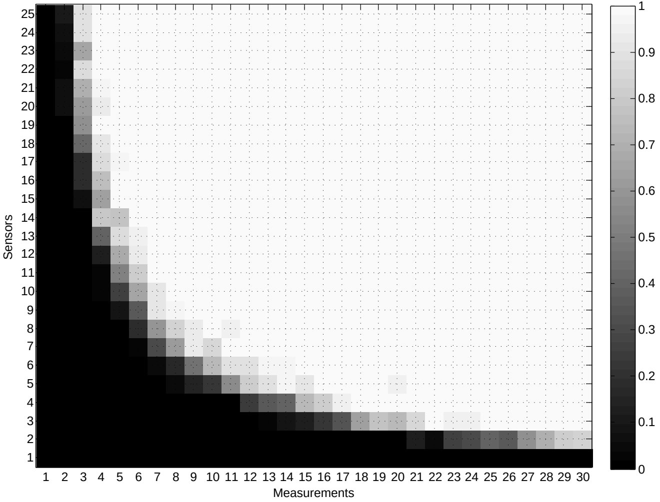
signal length n, number of measurements per node m, number of nodes $|\mathcal{V}|$

- signal: x is generated by choosing k components uniformly among the n elements
- sensing matrix: $A_v(i,j) \sim \mathrm{N}(0, 1/\sqrt{m})$
- consensus matrix: $P_{v,w} = 1/N$, $\forall (v,w) \in \mathcal{V} \times \mathcal{V}$

Declare **success** if

$$\sum_{v \in \mathcal{V}} \|x_0 - x_v^\star\|_2^2 / (n|\mathcal{V}|) < 10^{-4}$$

Reconstruction probability



$n = 150$, $k = 15$.

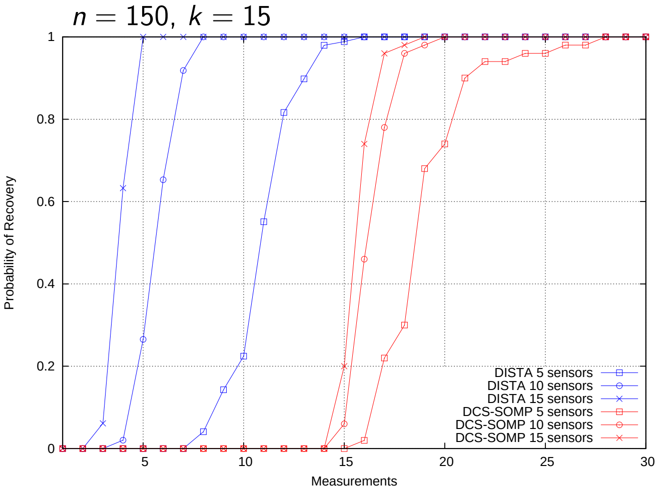
DCS-SOMP

DCS - Simultaneous Orthogonal Matching Pursuit

- Derived from OMP [Tropp & Gilbert '05]
- Joint support recontruction:
  - average the magnitudes of the projection of the residuals;
  - identify the maximal value and selects a new element of the support.
- Given the support, each sensor independently reconstructs the signal

DCS-SOMP requires at least k measurements per sensor

DISTA vs DCS-SOMP

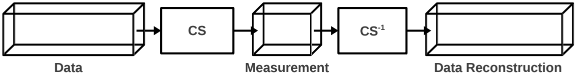


- DCS-SOMP requires at least k measurements per sensor

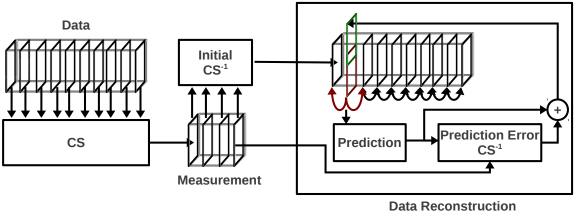
Reconstruction of “big” images

Progressive sensing and reconstruction of multidimensional signals

- Trivial approach, complexity of BP is infeasible



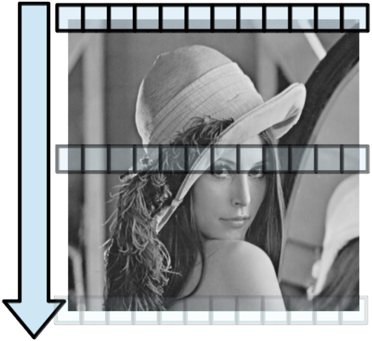
- Proposed iterative architecture



  - CS exploits correlation on a "subset" of dimensions
  - Iterative joint reconstruction based on linear predictors exploits correlation on "orthogonal" dimensions

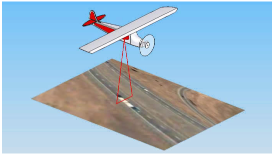
2D progressive imaging

- Devices equipped with 1D array of detectors
- Array moves in orthogonal direction

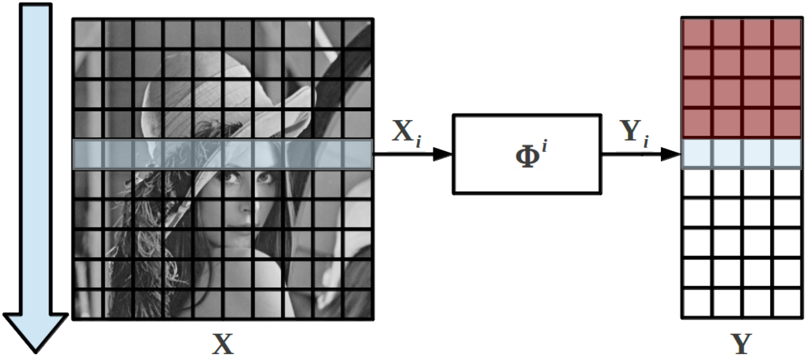


Flatbed Scanners

Airborne Imagers

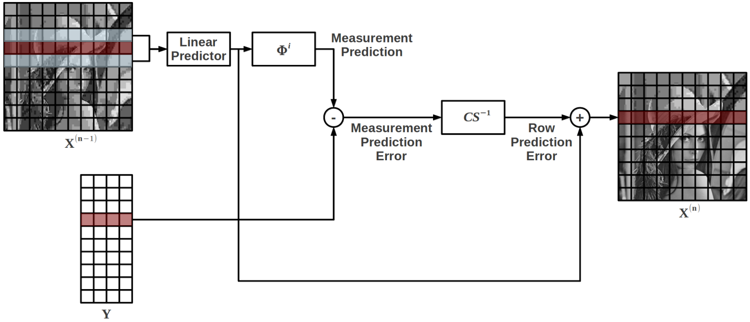


Acquisition: separate row CS measurement



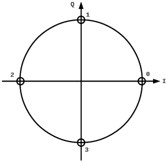
- M/N_{COL} dimensionality reduction
- No transform coding and sorting of coefficients

Reconstruction: refinement of row i at iteration (n)

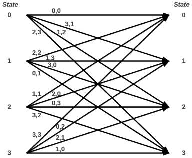


- Each row is predicted from the adjacent ones
- Measurement prediction error is computed and reconstructed
- Row estimate is obtained as predictor + prediction error
- Complexity reduction $\sim O(N_{\text{ROW}}^2)$

Flatbed scanner (bilevel images)



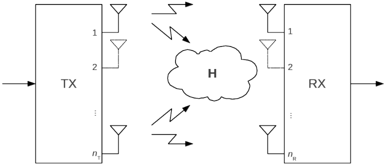
(a) Constellation



(b) Trellis

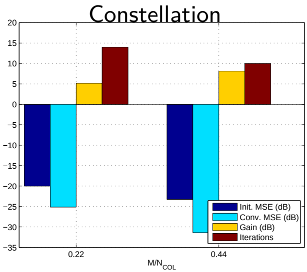
Lorem ipsum dolor sit amet, consectetur
adipiscing elit. Pellentesque ac feugiat arcu.
Nam bibendum nisl sit amet neque
pellentesque luctus in vel massa. Sed congue
tellus arcu, vel eleifend nisl. Etiam quis magna
orci, ac dictum est. Nunc cursus suscipit dui,
id venenatis dolor consequat eu. Duis
bibendum felis vitae lacus euismod luctus.
Mauris sem purus, scelerisque vitae imperdiet
a, iaculis sed nibh. Nam commodo tellus eget

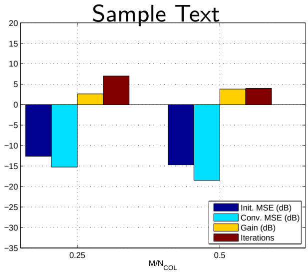
(c) Sample Text

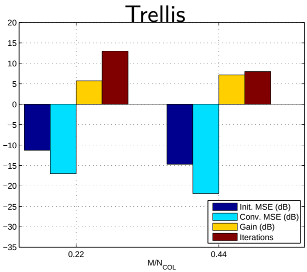


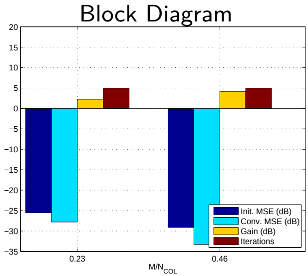
(d) Block Diagram

Flatbed scanner (bilevel images)



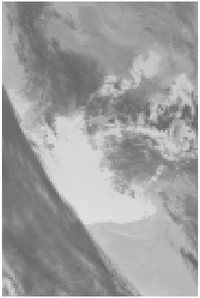




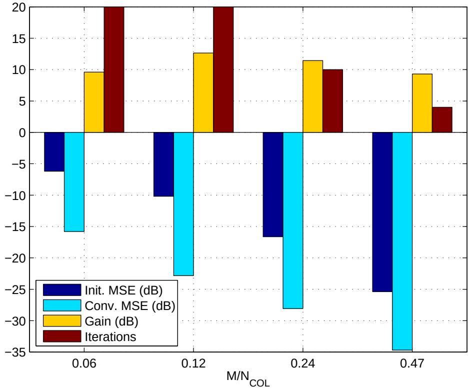


- Larger M $\Rightarrow$ faster convergence, lower MSE
- Sparse image, $5 \div 10$ dB gain, ~ 10 iterations
- Dense image, 3 dB gain, ~ 5 iterations

AIRS sensor (1501 spectral channels)



- $\sim$ 10 dB MSE gain
- 5 $\div$ 10 iterations



Conclusions

A number of open problems:

- Optimal joint reconstruction for multiple sources
- Distributed reconstruction
  - Performance bounds
  - Extensions (noise, innovation, faulty nodes, ...)
  - New formulations
- Error resilient representations of sparse signals
  - scalable, multiple descriptions, ...
- Improved reconstruction of big signals
  - Better signal models than sparsity