A Compressed Introduction to Compressed Sensing:
Combining Sparsity and Sampling

Laurent Jacques
TELE / EPL / UCL

November 27, 2009
SEQUEL - Lille Nord Europe
Outline

- Introduction
- Short introduction to “Sparsity”
- Classical Compression Scheme
- Sensing and Compressing
- Non-linear Reconstruction Methods
- Sensing Matrix Market
- Other Reconstruction Methods
- Some Applications
- Conclusion
Introduction

* Torrent of information to record, to store and to compress: from Science, Technology, private life, ...

* More and more constraints on sensors: miniaturization, power consumption, low communication, robustness, networking, ...

* Current codecs sometimes too heavy to be embedded in light sensors
Torrent of information to record, to store and to compress: from Science, Technology, private life, ...

More and more constraints on sensors: miniaturization, power consumption, low communication, robustness, networking, ...

Current codecs sometimes too heavy to be embedded in light sensors

**Question:**

Can we *sample* and *compress* in the same time? (even for smaller compression ratio) (and reconstruct the signal?)
Torrent of information to record, to store and to compress: from Science, Technology, private life, ...

More and more constraints on sensors: miniaturization, power consumption, low communication, robustness, networking, ...

Current codecs sometimes too heavy to be embedded in light sensors

**Question:**

Can we *sample* and *compress* in the same time? (even for smaller compression ratio) (and reconstruct the signal?)

**YES!** Since 2006, [Candès, Tao, Romberg, Donoho, DeVore, Cohen, Gribonval, ...]
Signal Sparsity (1/2)

* Signals have structures, features, edges, ...

3-D data

Speech signal

biology

Video

astronomy
Signal Sparsity (2/2)

* Given $x \in \mathbb{R}^N$ (e.g. $N = \text{number of pixels, samples, voxels} \ldots$)
* There is a sparsity basis (e.g. Wavelets, DCT, Fourier)

$$\Psi = (\Psi_1, \ldots, \Psi_D) \in \mathbb{R}^{N \times D}$$

where $x$ has the representation

$$x = \sum_{j=1}^{D} \alpha_j \Psi_j = \Psi \alpha$$

$\alpha \in \mathbb{R}^D$ is the coefficient vector
Signal Sparsity (2/2)

* Given \( x \in \mathbb{R}^N \) (e.g. \( N = \) number of pixels, samples, voxels ...)

* There is a sparsity basis (e.g. Wavelets, DCT, Fourier)

\[
\Psi = (\Psi_1, \cdots, \Psi_D) \in \mathbb{R}^{N \times D}
\]

where \( x \) has the \textit{representation}

\[
x = \sum_{j=1}^{D} \alpha_j \Psi_j = \Psi \alpha
\]

\( \alpha \in \mathbb{R}^D \) is the coefficient vector

* True sparsity: \( \alpha \) has \( K \) non-zero elements

* Compressibility: \( |\alpha_{(k)}| \leq C k^{-p} \) (sparsely approximable)
Classical Compression Transform
Classical Compression Transform

Wavelet basis $\Psi$
Classical Compression Transform

$\mathbf{x}$

Thresholding:
Keep $K$ first

Wavelet basis $\Psi$

$\alpha$
Classical Compression Transform

Wavelet basis $\Psi$

Thresholding:
Keep $K$ first

$x_K = \Psi \alpha_K$

$K = N/10$
Classical Compression Transform

* But: wasteful process!
Classical Compression Transform

* But: wasteful process!

* Recording of $N$ samples for $x$
  + Computing $\alpha$ $\Rightarrow O(N \log N)$ at best!
  + Ordering $\alpha$ $\Rightarrow O(N \log N)$
  + Thresholding by keeping $K \ll N$ values
But: wasteful process!

- Recording of $N$ samples for $x$
  - Computing $\alpha \Rightarrow O(N \log N)$ at best!
  - Ordering $\alpha \Rightarrow O(N \log N)$
  - Thresholding by keeping $K \ll N$ values

Can we avoid that?

Can we just sample $M$ values with $K < M \ll N$?

YES!
Sensing and Compressing

* Use a general Sensing model:

\[ M \text{ linear questions (projections) about } x: \]
Sensing and Compressing

* Use a general Sensing model:

\[ M \text{ linear questions (projections) about } x: \]

\[ y_i = \Phi_i^T x = \langle \Phi_i, x \rangle, \ 1 \leq i \leq M, \quad \text{or } y = \Phi x \in \mathbb{R}^M \]

with a sensing matrix: \( \Phi = (\Phi_1, \cdots, \Phi_M)^T \in \mathbb{R}^{M \times N} \)
Sensing and Compressing

- Use a general Sensing model:

  $M$ linear questions (projections) about $x$:

  \[ y_i = \Phi_i^T x = \langle \Phi_i, x \rangle, \ 1 \leq i \leq M, \quad \text{or} \quad y = \Phi x \in \mathbb{R}^M \]

  with a sensing matrix: \( \Phi = (\Phi_1, \cdots, \Phi_M)^T \in \mathbb{R}^{M \times N} \)

- Use another prior information:

  $x$ is sparse or compressible (i.e. well represented by sparsity)

  \[ y = \Phi x = (\Phi \Psi)\alpha = \Theta \alpha \]
Sensing and Compressing

* Use a general Sensing model:

\[ M \text{ linear questions (projections) about } x: \]

\[ y_i = \Phi_i^T x = \langle \Phi_i, x \rangle, \quad 1 \leq i \leq M, \quad \text{or } y = \Phi x \in \mathbb{R}^M \]

with a sensing matrix: \( \Phi = (\Phi_1, \cdots, \Phi_M)^T \in \mathbb{R}^{M \times N} \)

* Use another prior information:

\[ x \text{ is } \text{sparse or compressible} \quad (\text{i.e. well represented by sparsity}) \]

\[ y = \Phi x = (\Phi \Psi)\alpha = \Theta\alpha \]

* \textit{non-linear} reconstructions: convex optimization, greedy methods, ...
More Realistic Sensing

- Realistic Compressed Sensing:
  \[ y = \Theta \alpha + n \]
  where \( n = \) additive white noise (gaussian, uniform, GGD, ...)
  e.g. of bounded power \( \|n\|_2 \leq \epsilon \)

- \( y = \) sensor output! (no access to \( x \), or \( \alpha \))

- Important: we know only: \((y, \Phi) + \) a noise bound
Non-Linear Reconstructions (1/5)

- $y$ has $M < N$ components, but $x$ is sparse!
Non-Linear Reconstructions (1/5)

\* \( y \) has \( M < N \) components, but \( x \) is sparse!
\* If the support \( S \) of \( x \) was known (without noise, \( \Psi = \text{Id} \)):

\[
y = \Phi_S x_S \quad \Rightarrow \quad x_S = (\Phi_S^T \Phi_S)^{-1} \Phi_S^T y
\]

if \( \Phi_S^T \Phi_S \) not singular. Condition: \( M > K \)
Non-Linear Reconstructions (1/5)

- $y$ has $M < N$ components, but $x$ is sparse!
- If the support $S$ of $x$ was known (without noise, $\Psi = \text{Id}$):

  $$y = \Phi_S x_S \Rightarrow x_S = (\Phi_S^T \Phi_S)^{-1} \Phi_S^T y$$

  if $\Phi_S^T \Phi_S$ not singular. Condition: $M > K$
- What if $S$ is unknown? Impose the RIP! (Restricted Isometry Property)
Non-Linear Reconstructions (1/5)

* $y$ has $M < N$ components, but $x$ is sparse!
* If the support $S$ of $x$ was known (without noise, $\Psi = \text{Id}$):
  \[ y = \Phi_S x_S \Rightarrow x_S = (\Phi_S^T \Phi_S)^{-1} \Phi_S^T y \]
  if $\Phi_S^T \Phi_S$ not singular.  Condition: $M > K$

* What if $S$ is unknown? Impose the RIP! (Restricted Isometry Property)
  For all $K$-sparse $u$, there exists $0 \leq \delta_K < 1$ such
  \[ (1 - \delta_K) \| u \|_2^2 \leq \| \Phi u \|_2^2 \leq (1 + \delta_K) \| u \|_2^2 \]
  i.e., we foresee all possible subsets $S$ of $K$ elements.
Non-Linear Reconstructions (2/5)

* **Theorem** [Candes, Tao, Romberg, Donoho]:

  If $\Phi$ is RIP of order $2K$, i.e. $\delta_{2K} < 1$, then, for any $K$-sparse signal $x$ sensed by $y = \Phi x$, $x$ is **perfectly** recovered by the ideal program:

  $$x^* = \arg\min_u \|u\|_0 \text{ s.t. } y = \Phi u$$
**Non-Linear Reconstructions** (2/5)

* **Theorem** [Candes, Tao, Romberg, Donoho]:

*If* \( \Phi \) *is RIP of order* \( 2K \), *i.e.* \( \delta_{2K} < 1 \), *then, for any* \( K \)-sparse signal \( x \) *sensed by* \( y = \Phi x \), *x is perfectly recovered by the ideal program:*  

\[
x^* = \arg \min_u \|u\|_0 \text{ s.t. } y = \Phi u
\]

*Proof:* \( (1 - \delta_{2K})\|x - x^*\|_2^2 \leq \|\Phi x - \Phi x^*\|_2^2 = \|y - y\|_2^2 = 0! \)
Non-Linear Reconstructions (2/5)

* **Theorem** [Candes, Tao, Romberg, Donoho]:

If $\Phi$ is RIP of order $2K$, i.e. $\delta_{2K} < 1$, then, for any $K$-sparse signal $x$ sensed by $y = \Phi x$, $x$ is **perfectly** recovered by the ideal program:

$$x^* = \text{arg min}_u \|u\|_0 \text{ s.t. } y = \Phi u$$

**Proof:** $(1 - \delta_{2K})\|x - x^*\|_2^2 \leq \|\Phi x - \Phi x^*\|_2^2 = \|y - y\|_2^2 = 0!$

* **Problem:** NP-complete program (Combinatorial complexity) [Natarajan 95]
Non-Linear Reconstructions (2/5)

* **Theorem** [Candes, Tao, Romberg, Donoho]:

If $\Phi$ is RIP of order $2K$, i.e. $\delta_{2K} < 1$, then, for any $K$-sparse signal $x$ sensed by $y = \Phi x$, $x$ is perfectly recovered by the ideal program:

$$x^* = \arg \min_u \|u\|_0 \text{ s.t. } y = \Phi u$$

**Proof**: $(1 - \delta_{2K})\|x - x^*\|_2^2 \leq \|\Phi x - \Phi x^*\|_2^2 = \|y - y\|_2^2 = 0!$

* **Problem**: NP-complete program (Combinatorial complexity) [Natarajan 95]

* But **Relax!**, for $\delta_{2K} < \sqrt{2} - 1$, this works (Basis Pursuit):

$$x^* = \arg \min_u \|u\|_1 \text{ s.t. } y = \Phi u$$

[Chen, Donoho, Tropp, Candes, ...]

and this is Convex Programming! (LP, int. point., Proximal, ...)

[Chen, Donoho, Tropp, Candes, ...]
Non-Linear Reconstructions (3/5)

\[ x = x_{\text{BP}} \]

\[ \Phi u = y \]

\[ \ell_1 \text{-ball} \]

\[ \ell_2 \text{-ball} \]
With noise: (Basis Pursuit DeNoise - BPDN)

\[ x^* = \arg \min_u \|u\|_1 \quad \text{s.t.} \quad \|y - \Phi u\|_2 \leq \epsilon \]

Stability:

\[ \|x - x^*\|_2 \leq C\epsilon + D\|x - x_K\|_1/\sqrt{K} \]

Measurement error

Compressibility error, i.e. 0 if \( x \) is \( K \) sparse
Non-Linear Reconstructions (5/5)

\[ x^* \]

\[ \| \Phi u - y \|_2 \leq \epsilon \]

\[ \ell_1 \text{-ball} \]
RIP matrices? (link with J.-L. Lemma)

* When do we have \((1 - \delta_K)\|u\|_2^2 \leq \|\Phi u\|_2^2 \leq (1 + \delta_K)\|u\|_2^2\) ?

(for \(u\) \(K\)-sparse)
RIP matrices? (link with J.-L. Lemma)

✶ When do we have $(1 - \delta_K)\|u\|_2^2 \leq \|\Phi u\|_2^2 \leq (1 + \delta_K)\|u\|_2^2$ ?
(for $u$ $K$-sparse)

✶ Use Measure Concentration theory [R. Baraniuk et al, 07]:

✶ For any $x$, and $\Phi$ random such that (for some $c > 0$):

$$P_{\Phi} \left[ \left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| > \epsilon \|x\|_2^2 \right] \leq 2 e^{-cm \epsilon^2}$$

✶ The set of $K$-sparse signals of unit norm can be $\epsilon$ - covered by a finite set of no more than $(\frac{N}{K}) (\frac{3}{\epsilon})^K \leq (\frac{3eN}{K\epsilon})^K$ elements

✶ Union bounds on the probability of failure + covering

✶ Result: $M \geq O(\epsilon^{-2}K \log N/K)$
RIP matrices? (link with J.-L. Lemma)

- When do we have \((1 - \delta_K)\|u\|_2^2 \leq \|\Phi u\|_2^2 \leq (1 + \delta_K)\|u\|_2^2\) \(\text{? (for } u \text{ } K\text{-sparse})\)

- Use Measure Concentration theory [R. Baraniuk et al, 07]:
  - For any \(x\), and \(\Phi\) random such that (for some \(c > 0\)):
    \[
P_{\Phi} \left[\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon \|x\|_2^2\right] \leq 2e^{-cm\epsilon^2}
    \]
  - The set of \(K\)-sparse signals of unit norm can be \(\epsilon\)-covered by a finite set of no more than \(\binom{N}{K}(\frac{3}{\epsilon})^K \leq (\frac{3eN}{K\epsilon})^K\) elements
  - Union bounds on the probability of failure + covering
  - \textbf{Result: } \(M \geq O(\epsilon^{-2}K \log N/K)\)

- Similar developments for Johnson-Lindenstrauss Lemma
- General result for any subset by Mendelson et al [07, 08]
Sensing Matrix Market  (1/3)

- Random (sub) Gaussian Ensemble: $\Phi \in \mathbb{R}^{M \times N}$

  $\Phi_{ij} \sim_{\text{iid}} N(0, 1/M)$

  RIP-$K$ w.v.h.p. if $M \geq O(K \log N/K)$

  and optimal: if $\Psi$ ONB, $\Theta = \Phi \Psi$ is RIP-$K$
Sensing Matrix Market (1/3)

* Random (sub) Gaussian Ensemble: $\Phi \in \mathbb{R}^{M \times N}$

$$\Phi_{ij} \sim_{\text{iid}} \mathcal{N}(0, 1/M)$$

RIP-$K$ w.v.h.p. if $M \geq O(K \log N/K)$

and optimal: if $\Psi$ ONB, $\Theta = \Phi \Psi$ is RIP-$K$

Problem:
- big storage, slow coding, ...

Partial solutions:
- Bernoulli: $\Phi_{ij} \sim_{\text{iid}} \pm 1/\sqrt{M}$
- Pseudo-randomness: $\Phi$ is stored in few seeds (LFSR)
Sensing Matrix Market (2/3)

Random Fourier Ensemble: \( J \subset \{1, \cdots, N\}, \#J = M \)

\[ \Phi = \mathcal{J}\mathcal{F}, \quad \mathcal{F} \in \mathbb{R}^{N \times N}, \quad \mathcal{J}v = v_J \]

\[ M \geq O(K \log^4 N) \]
Sensing Matrix Market (2/3)

- Random Fourier Ensemble: \( J \subset \{1, \cdots, N\}, \#J = M \)

\[
\Phi = JF, \quad F \in \mathbb{R}^{N \times N}, \quad Jv = v_J
\]

\[
M \geq O(K \log^4 N)
\]

pros:
- fast, since Fourier
- analog counterpart in optics (Fourier lens)
- low storage (\(M\) positions)

cons:
- optimal only in \(\Psi = \text{Id}\)
Random Basis Ensemble: $\Phi = \mathcal{J}\mathcal{U}$

$M \geq O(\mu(\mathcal{U}, \Psi)^2 K \log^4 N)$

with $\mu(\mathcal{U}, \Psi)$ the coherence between $\mathcal{U}$ and $\Psi$.

e.g. $\mathcal{U} = \text{Noiselets}$, optimal with wavelets (Haar, Daubechies, ...)

Random Convolutions [J. Romberg]

$\Phi = \mathcal{J} \mathcal{F}^* D \mathcal{F}$

$M \geq O(K \log^5 N)$

and many others (spr. spectrum, Hadamard, Toeplitz, ...
Other reconstruction methods (1/3)

Other convex programs:

\[
x^* = \arg \min_u \|u\|_1 + \frac{\lambda}{2} \|y - \Phi u\|_2^2
\]

[M. J. Wainwright 06]

\[
x^* = \arg \min_u \|y - \Phi u\|_2 \text{ s.t. } \|u\|_1 \leq \tau
\]

(Lasso)

\[
x^* = \arg \min_u \|u\|_1 \text{ s.t. } \|\Phi^*(y - \Phi u)\|_\infty \leq \epsilon
\]

(Dantzig Sel.)

[Candes, Tao, 07]

GGD noise and quantization distortion [L. Jacques, D. Hammond, M.J. Fadili, 09]

\[
x^* = \arg \min_u \|u\|_1 \text{ s.t. } \|y - \Phi u\|_p \leq \epsilon \quad (\text{BPDQ}_p)
\]

(oversampling allows high \(p\))

\[2 \leq p \leq \infty\]

Many variants exist (e.g. mixed norms, Poisson noise, sparse noise, speckle, ...)

\[
(\overset{\text{oversampling allows high } p}{\text{oversampling allows high } p})
\]

Many variants exist (e.g. mixed norms, Poisson noise, sparse noise, speckle, ...)

\[2 \leq p \leq \infty\]
Other reconstruction methods (2/3)

Outside convexity: (with $0 < p < 1$)

\[
x^* = \arg \min_u \|u\|_p^p \\
\text{s.t. } \|y - \Phi u\|_2 \leq \epsilon
\]
Other reconstruction methods \((2/3)\)

Outside convexity: (with \(0 < p < 1\))

\[
x^* = \arg \min_u \|u\|^p_p \\
\text{s.t. } \|y - \Phi u\|_2 \leq \epsilon
\]

\[
x = x_{\ell_1} = x_{\ell_q}
\]

\(\mathbb{R}^2\)

\(\Phi u = y\)

\(\ell_0\)-ball

\(\ell_1\)-ball

\(\ell_q\)-ball
Other reconstruction methods \((2/3)\)

Outside convexity: \((\text{with } 0 < p < 1)\)

\[
x^* = \arg\min_u \|u\|^p_p \\
\text{s.t. } \|y - \Phi u\|_2 \leq \epsilon
\]

Practically: reweighting - \(\ell_1\)

\[
u^{(n+1)} = \arg\min_u \|W^{(n)} u\|_1 \\
\text{s.t. } \|y - \Phi u\|_2 \leq \epsilon
\]

with \(W_{ij}^{(n)} = \delta_{ij} (|u_i^{(n)}|^{1-p} + \eta)^{-1}\),

\(W^{(0)} = \text{Id, } \eta > 0\).

for fixed point: \(\|W^{(n)} u\|_1 \simeq \|u\|^p_p\)

[Chartrand, Candes, Wakin, Boyd]
Other reconstruction methods (3/3)

Greedy Algorithm:

- Matching Pursuit, Orthogonal MP, Homotopy, ...
- CoSaMP, Iterative Hard Thresholding, subspace Pursuit, ...
Other reconstruction methods (3/3)

Greedy Algorithm:

- Matching Pursuit, Orthogonal MP, Homotopy,
- CoSaMP, Iterative Hard Thresholding, subspace Pursuit, ...

Other priors:

- Minimizing sparsity in gradient (TV-norm)

\[
x^* = \arg \min_{u} \|u\|_{TV} \text{ s.t. } \|y - \Phi u\|_2 \leq \epsilon, \quad \|u\|_{TV} = \sum_i \|\nabla u_i\|.
\]

- Logan-Shepp phantom
- Fourier Sampling: 22 lines
- IFFT + 0s
- BPDN - TV

- adding positivity constraint, ...
Some CS Applications ...
Rice Single-pixel Camera

[Duarte, Davenport, Takbar, Laska, Sun, Kelly, Baraniuk, 2008]
EPFL CS CMOS Camera

- Analog Random Convolution
- Initial Pseudo-Random Sequence
- LFSR
- Shift Register
- Current Summation
- Combined LFSRs
- Op-Amps
- Multiplexer
- ADC $Q_\Delta$

\[
Q_\Delta[y_i = \sum_j a_{r(i)-j} x_j]
\]

$M=N/3$, 11 bit + noise: 27dB

[Jacques, Vanderheyndt, Bibet, Majidzadeh, Schmid, Leblebici, ICASSP 2009]
CS for UWB Pulse Trains

\( \Phi = \text{R. Basis (Hadamard)} + \text{Spr. Spect.} \)

\( \Psi = \text{Redundant basis of pulses} \)

[Naini, Gribonval, Jacques, Vanderghynst, ICASSP, 2009]

Previous and Related works: Eldar, Laska, Baraniuk, ....
“Analog-to-Information Converter”

Avg. oversampling M/K for 25 dB vs K/N

Input SNR 30 dB
Radio-Interferometry

As for MRI, sampling in the Fourier domain:

\[ \Phi = \mathcal{J} \mathcal{F} \]
Conclusion: Take Away Messages

- Sampling and Compressing at the same stage is possible!
- Asymmetry: light linear coding / heavy non-linear decoding
- CS is not there to beat existing codecs, e.g. JPEG2000
  \[ \text{Best Codec} \ < \ CS \ <\ < \ Full \ Sampling \]
- CS offers also sampling democracy (scalability)
Conclusion: Take Away Messages

✴ Sampling and Compressing at the same stage is possible!
✴ Asymmetry: light linear coding / heavy non-linear decoding
✴ CS is not there to beat existing codecs, e.g. JPEG2000

\[
\text{Best Codec} < \text{CS} << \text{Full Sampling}
\]
✴ CS offers also sampling democracy (scalability)

✴ **Future:**
✴ avoid the RIP (e.g. null space properties, phase transition, ...)
✴ continue to improve decoding programs
✴ behavior under additional signal priors (TV, positivity, ...)
✴ or under different noises (e.g. Poisson, impulsive, quantization, ...)
✴ more and more CS hardware to come ....
**Links (Science 2.0.)**

* Rice CS Resources page: [http://www-dsp.rice.edu/cs](http://www-dsp.rice.edu/cs)

* Igor Carron’s “Nuit Blanche” blog: [http://nuit-blanche.blogspot.com](http://nuit-blanche.blogspot.com)

  1 CS post/day!
Thank you!
References:


