Sparsity in Astrophysics: from Wavelets to Compressed Sensing

J.-L. Starck

Euclid
What is a good representation for data?

A signal $s$ ($n$ samples) can be represented as sum of weighted elements of a given dictionary

$$
\Phi = \{ \phi_1, \ldots, \phi_K \} \quad \text{Ex: Haar wavelet}
$$

$$
\sum_{k=1}^{K} \alpha_k \phi_k = \Phi \alpha
$$

- Fast calculation of the coefficients
- Analyze the signal through the statistical properties of the coefficients
- Approximation theory uses the sparsity of the coefficients

$$
f(x) = c_0 + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} c_{jk} \psi_{jk}(x) .
$$
The STARLET Transform
Isotropic Undecimated Wavelet Transform (a trous algorithm)

\[ \varphi = B_3 - \text{spline} \]
\[ \frac{1}{2} \psi \left( \frac{x}{2} \right) = \frac{1}{2} \varphi \left( \frac{x}{2} \right) - \varphi(x) \]
\[ h = [1, 4, 6, 4, 1]/16, \quad g = \delta - h, \quad \tilde{h} = \tilde{g} = \delta \]

\[ I(k, l) = c_{J,k,l} + \sum_{j=1}^{J} w_{j,k,l} \]
**Sparsity Model 1:** we consider a dictionary which has a fast transform/reconstruction operator:

\[ \Phi = \{\phi_1, \ldots, \phi_K\} \]

\[ s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi \alpha \]

- **Local DCT**
  - Stationary textures
  - Locally oscillatory

- **Wavelet transform**
  - Piecewise smooth
  - Isotropic structures

- **Curvelet transform**
  - Piecewise smooth, edge
How to measure sparsity?

with \( 0^0 = 0 \), \( \|
\alpha \|_0 = \sum_k \alpha_k^0 = \# \{ \alpha_k \neq 0 \} \)

Formally, the sparsest coefficients are obtained by solving the optimization problem:

(P0) Minimize \( \| \alpha \|_0 \) subject to \( s = \phi \alpha \)

It has been proposed (to relax and) to replace the \( l_0 \) norm by the \( l_1 \) norm (Chen, 1995):

(P1) Minimize \( \| \alpha \|_1 \) subject to \( s = \phi \alpha \)

It can be seen as a kind of convexification of (P0).

It has been shown (Donoho and Huo, 1999) that for certain dictionary, if there exists a highly sparse solution to (P0), then it is identical to the solution of (P1).
Morphological Diversity

\[ \phi = [\phi_1, \ldots, \phi_L], \quad \alpha = \{\alpha_1, \ldots, \alpha_L\}, \quad s = \phi \alpha = \sum_{k=1}^{L} \phi_k \alpha_k \]

**Sparsity Model 2:** we consider a signal as a sum of \( K \) components \( s_k, s = \sum_{k=1}^{K} s_k \) each of them being sparse in a given dictionary:

\[ s_k = \Phi_k \alpha_k \]
\[ s = \sum_{k=1}^{K} s_k = \sum_{k=1}^{K} \Phi_k \alpha_k = \Phi \alpha \]


**Sparsity Model 1:** we consider a dictionary which has a fast transform/reconstruction operator:

\[ \Phi = \{ \phi_1, \ldots, \phi_K \} \]

\[ s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi \alpha \]

- **Local DCT**
  - Stationary textures
  - Locally oscillatory

- **Wavelet transform**
  - Piecewise smooth
  - Isotropic structures

- **Curvelet transform**
  - Piecewise smooth, edge

**Sparsity Model 2:** Morphological Diversity:

\[ \phi = [\phi_1, \ldots, \phi_L], \quad \alpha = \{\alpha_1, \ldots, \alpha_L\}, \quad s = \phi \alpha = \sum_{k=1}^{L} \phi_k \alpha_k \]

**Sparsity Model 3:** we adapt/learn the dictionary directly from the data

Model 3 can be also combined with model 2:

Advantages of model 1 (fixed dictionary): extremely fast.

Advantages of model 2 (union of fixed dictionaries):
- more flexible to model 1.
- The coupling of local DCT+curvelet is well adapted to a relatively large class of images.

Advantages of model 3 (dictionary learning):
atoms can be obtained which are well adapted to the data, and which could never be obtained with a fixed dictionary.

Drawback of model 3 versus model 1,2:
We pay the price of dictionary learning by being less sensitive to detect very faint features.
Complexity: Computation time, parameters, etc
INVERSE PROBLEMS

\[ Y = HX + N \]
\[ X = \Phi \alpha, \text{ and } \alpha \text{ is sparse} \]

\[ \min_{\alpha} \|\alpha\|_p \quad \text{subject to} \quad \|Y - H\Phi \alpha\|^2 \leq \varepsilon \]

- Denoising
- Deconvolution
- Component Separation
- Inpainting
- Blind Source Separation
- Minimization algorithms
- Compressed Sensing
Denoising using a sparsity model

\[ Y = X + N \]

Denoising using a sparsity prior on the solution:

\[ X \text{ is sparse in } \Phi, \text{ i.e. } X = \Phi \alpha \text{ where most of } \alpha \text{ are negligible.} \]

\[ \tilde{\alpha} \in \arg \min_{\alpha} \frac{1}{2} \| Y - \Phi \alpha \|^2 + t \| \alpha \|_p^p, \quad 0 \leq p \leq 1. \]
\[ p=0 \]

\[ \tilde{\alpha} \in \arg \min_{\alpha} \frac{1}{2} \| Y - \Phi \alpha \|^2 + \frac{t^2}{2} \| \alpha \|_0 \]

\[ \Rightarrow \text{Solution via Iterative Hard Thresholding} \]

\[ \tilde{\alpha}^{(t+1)} = \text{HardThresh}_{\mu t}(\tilde{\alpha}^{(t)} + \mu \Phi^T(Y - \Phi \tilde{\alpha}^{(t)})), \quad \mu = 1/\|\Phi\|^2. \]

\[ \tilde{\alpha}_{j,k} = \text{HardThresh}_t(\alpha_{j,k}) = \begin{cases} 
\alpha_{j,k} & \text{if } |\alpha_{j,k}| \geq t, \\
0 & \text{otherwise}. 
\end{cases} \]

1st iteration solution:

\[ \tilde{X} = \Phi \text{ HardThresh}_t(\Phi^TY) = \Delta_{\Phi,t}(Y) \]

Exact for $\Phi$ orthonormal.
\[ p=1 \]

\[ \tilde{\alpha} = \arg \min_{\alpha} \frac{1}{2} \| Y - \Phi \alpha \|_2^2 + t \| \alpha \|_1 \]

\[ \Rightarrow \text{Solution via iterative Soft Thresholding} \]

\[ \tilde{\alpha}^{(t+1)} = \text{SoftThresh}_{\mu t}(\tilde{\alpha}^{(t)} + \mu \Phi^T(Y - \Phi \tilde{\alpha}^{(t)})), \mu \in (0, 2/\|\Phi\|^2). \]

\[ \tilde{\alpha}_{j,k} = \text{SoftThresh}_t(\alpha_{j,k}) = \text{sign}(\alpha_{j,k})(|\alpha_{j,k}| - t)_+ \]

1st iteration solution:

\[ \tilde{X} = \Phi \text{SoftThresh}_t(\Phi^T Y) = \Delta_{\Phi,t}(Y) \]

Exact for \( \Phi \) orthonormal.
Inverse Problems and Iterative Thresholding Minimizing Algorithm

Iterative thresholding with a varying threshold was proposed in (Starck et al, 2004; Elad et al, 2005) for sparse signal decomposition in order to accelerate the convergence. The idea consists in using a different threshold $\lambda^{(n)}$ at each iteration.

For IST:

$$\alpha^{(n+1)} = \text{HT}_{\lambda^{(n)}} \left( \alpha^{(n)} + \Phi^T A^T \left( Y - A\Phi\alpha^{(n)} \right) \right)$$

For IHT:

$$\alpha^{(n+1)} = \text{ST}_{\lambda^{(n)}} \left( \alpha^{(n)} + \Phi^T A^T \left( Y - A\Phi\alpha^{(n)} \right) \right)$$

More Refs: Vonesch et al, 2007; Elad et al 2008; Wright et al., 2008; Nesterov, 2008 and Beck-Teboulle, 2009; Blumensath, 2008; Maleki et Donoho, 2009; Dupe et al, 2010; Dupe et al 2011; Geyre et al, 2011, etc.
DECONVOLUTION

Compressed Sensing

A non linear sampling theorem

“Signals with exactly K components different from zero can be recovered perfectly from ~ K log N incoherent measurements”

Replace samples with few linear projections

\[ y = \Theta x \]

\[ M \times 1 \text{ measurements} \]

\[ \Theta \]

\[ M \times N \]

\[ N \times 1 \text{ sparse signal} \]

\[ K \text{ nonzero entries} \]

\[ K < M \ll N \]

Reconstruction via non linear processing:

\[ \min_x \|x\|_1 \quad \text{s.t.} \quad y = \Theta x \]
Soft Compressed Sensing Definition

\[ Y = \Theta X = \Theta \Phi \alpha \]

Mutual coherence:

\[ \mu_{\Theta, \Phi} = \max_{i,k} \left| \langle \Theta_i, \Phi_k \rangle \right| \]

Mutual coherence the degree of similarity between the sparsity and measurement systems.
Radio-Interferometry

\[
\Phi = \begin{cases}
\Phi = \text{Id} & \text{(Hogbom, 1974)} \\
\Phi = \text{Wavelet Transform} & \text{(Wakker, and Schwarz, 1988; Starck et al 1994)}
\end{cases}
\]

\[y = \text{FOURIER} \]

\[\Rightarrow \text{See (McEwen et al, 2011; Wenger et al, 2010; Wiaux et al, 2009; Cornwell et al, 2009; Suskimo, 2009; Feng et al, 2011).}\]
CS-Radio Astronomy

The Applications of Compressive Sensing to Radio Astronomy: I Deconvolution

Australian Square Kilometer Array Pathfinder (ASKAP) radio telescope.
CS-Radio Astronomy

Hogbom CLEAN

MEM residual
Missing Data

- Period detection in temporal series

COROT: HD170987

- Bad pixels, cosmic rays, point sources in 2D images, ...
Missing Data

Period detection in temporal series

\[ y = \Theta \times \Phi \]

Observation Mask
Measurement System

FOURIER

COROT: HD170987

Original data

PSD for Original light curve of HD170987
COROT: HD170987 with in-painting

Original data

in-painted data
Inpainting:

Transfering Spatial Data to the Earth

Transferring Data to the Earth

Observed Herschel Data During the Calibration Phase, November 2010, without any compression.

A scan (16 x 16 pixels at 40 Hz during 25 min each, we obtained 60000 images.
Map from Uncompressed Data

Official Pipeline Reconstruction: Averaging
Gamma Ray Instruments (Integral) - Acquisition with coded masks

\[ y \quad \Theta \quad \Phi \quad \alpha \]

CODED Mask
Measurement System

INTEGRAL/IBIS Coded Mask

Crab Nebula Integral Observation

Courtesy I. Caballero, J. Rodriguez (AIM/Saclay)
SVOM (future French-Chinese Gamma-Ray Burst mission)

- ECLAIRs france-chinese satellite ‘SVOM’ (launch in 2014-2015)
  Gamma-ray detection in energy range 4 - 120 keV
  Coded mask imaging (at 460 mm of the detector plane)

Physical mask pattern
(46 x 46 pixels of 11.7 mm)

ECLAIR could become the first CS-Designed Astronomical Instrument
Pseudo-3D Weak Lensing
3D Weak Lensing

The convergence $\kappa$, as seen in sources of a given redshift bin, is the linear transformation of the matter density contrast, $\delta$, along the line-of-sight (Simon et al 2009):

$$\kappa = Q \delta + N$$

with $\delta(r) \equiv \rho(r)/\bar{\rho} - 1$

$$Q_{\ell \ell} = \frac{3H_0^2\Omega_M}{2c^2} \int_{w_{\ell}}^{w_{\ell+1}} dw \frac{W^{(i)}(w) f_K(w)}{a(w)} , \quad W^{(i)}(w) = \int_0^{w^{(i)}} dw' \frac{f_K(w - w')}{f_K(w')} \left( p(z) \frac{dz}{dw} \right)_{z = z(w')}$$

where $H_0$ is the hubble parameter, $\Omega_M$ is the matter density parameter, $c$ is the speed of light, $a(w)$ is the scale parameter evaluated at comoving distance $w$, and

$$f_K(w) = \begin{cases} K^{-1/2} \sin(K^{1/2}w), & K > 0 \\ w, & K = 0 \\ (-K)^{-1/2} \sinh([-K]^{1/2}w) & K < 0 \end{cases}$$

gives the comoving angular diameter distance as a function of the comoving distance and the curvature, $K$, of the Universe.
3D Weak Lensing

\[ \kappa = Q + \delta \]

- \( M \) measurements: \( M \times N \) \((M > N)\)
- number of bins in the source plane

\( \delta \) is sparse.
- \( Q \) spreads out the information in \( \delta \) along \( \kappa \) bins.
- More unknown than measurements

\[
\min_{\delta} \| \delta \|_1 \quad s.t. \quad \frac{1}{2} \| \kappa - Q\delta \|_{\Sigma^{-1}}^2 \leq \epsilon
\]

Recent optimization method, based on proximal theory, such as Chambolle & Pock (2010) can be used to find the solution.
Reconstructions of two clusters along the line of sight, located at a redshift 0.2 and 1.0 (data binned into Nsp = 20 redshift bins, but aim to reconstruct onto Nlp = 25 redshift bins).
Sparsity/CS in Astrophysics

- Sparsity is very efficient for
  - Inverse problems (denoising, deconvolution, etc).
  - Inpainting
  - Component Separation.

- Compressed Sensing
  - New algorithms to process radio-interferommetric data
  - New approaches to analyze data.
  - New instruments design.

- Perspectives
  - PLANCK component separation.
  - Euclid