Sparse Astronomical Data Analysis

Jean-Luc Starck

Collaborators: J. Bobin., F. Sureau

CEA Saclay
What is a good representation for data?

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

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

Dictionary (basis, frame)

$$\Phi = \{ \phi_1, \ldots, \phi_K \}$$

Atoms

Ex: Haar wavelet

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

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

\[ \varphi = B_3 - \text{spline}, \quad \frac{1}{2} \psi \left( \frac{x}{2} \right) = \frac{1}{2} \varphi \left( \frac{x}{2} \right) - \varphi(x) \]

\[ h = [1, 4, 6, 4, 1]/16, \quad g = \delta - h, \quad \tilde{h} = \tilde{g} = \delta \]

\[ I(k, l) = c_{j,k,l} + \sum_{j=1}^{J} w_{j,k,l} \]
Inverse Problem Regularization using Sparsity

\[ Y = AX + N \]

Between all possible solutions, we want the one which has the sparsest representation in the dictionary \( \Phi \). It leads to the following optimization problem:

\[
\min_{\alpha_1, \ldots, \alpha_T} \frac{1}{2\sigma^2} \| Y - A\Phi\alpha \|^2 + \lambda \sum_{i=1}^{T} \| \alpha_i \|_p^p, \quad 0 \leq p < 2.
\]

\[ X = \Phi\alpha \]

A sparse model can be interpreted in a Bayesian framework

Assuming the coefficients \( \alpha \) of the solution in the dictionary \( \Phi \) follow a leptokurtic PDF with heavy tails such as the generalized Gaussian distribution form:

\[
\text{pdf}_\alpha(\alpha_1, \ldots, \alpha_T) \propto \prod_{i=1}^{T} \exp \left( -\lambda \| \alpha_i \|_p^p \right), \quad 0 \leq p < 2.
\]
Denoising using a sparsity model

\[ Y = X + N \]

Denoising using a sparsity prior on the solution:

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

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

\[ \Rightarrow \text{Solution via Iterative \textbf{Hard} Thresholding} \]
\[ \tilde{\alpha}^{(t+1)} = \text{HardThresh}_{\mu t}(\tilde{\alpha}^{(t)} + \mu \Phi^{T}(Y - \Phi \tilde{\alpha}^{(t)})), \mu = 1/\|\Phi\|^{2}. \]
\[ \tilde{\alpha}_{j,k} = \text{HardThresh}_{t}(\alpha_{j,k}) = \begin{cases} \alpha_{j,k} & \text{if } |\alpha_{j,k}| \geq t, \\ 0 & \text{otherwise}. \end{cases} \]

1st iteration solution:
\[ \tilde{X} = \Phi \text{ HardThresh}_{t}(\Phi^{T}Y) = \Delta_{\Phi,t}(Y) \]

Exact for \( \Phi \) orthonormal.

\textbf{Soft prior} \hspace{1cm} \textbf{Noise Modeling}
\[ p=1 \]
\[ \tilde{\alpha} = \arg\min_{\alpha} \frac{1}{2} \| Y - \Phi \alpha \|^2 + t \| \alpha \|_1 \]

\[ \Rightarrow \text{Solution via iterative Soft Thresholding} \]
\[ \tilde{\alpha}^{(t+1)} = \text{SoftThresh}_{\mu t}(\tilde{\alpha}^{(t)} + \mu \Phi^T (Y - \Phi \tilde{\alpha}^{(t)})), \mu \in (0, 2/\|\Phi\|^2). \]
\[ \tilde{\alpha}_{j,k} = \text{SoftThresh}_t(\alpha_{j,k}) = \text{sign}(\alpha_{j,k})(|\alpha_{j,k}| - t)_+ \]

1st iteration solution:
\[ \tilde{X} = \Phi \text{ SoftThresh}_t(\Phi^T Y) = \Delta_{\Phi,t}(Y) \]

Exact for \( \Phi \) orthonormal.
Sparsity and FERMI data


very low source intensity (~$10^{-1}$ or even lower)

FERMI dedicated wavelets: 2D-1D Wavelets

$$\psi(x, y, z) = \psi^{(xy)}(x, y) \psi^{(z)}(z)$$

$$w_{j_1, j_2}[k_x, k_y, k_z] = D \ast \bar{\psi}_{j_1}^{(xy)} \ast \bar{\psi}_{j_2}^{(z)}(x, y, z)$$
Results – Multi-Spectral Image Restoration
(MS-VST + [2D+1D] Wavelet)


A 35-source grid:
5 lines x 7 columns
Thresh. Level = 4σ
Inverse Problems and Iterative Thresholding Minimizing Algorithm

\[
\min_{\alpha_1, \ldots, \alpha_T} \frac{1}{2\sigma^2} \|Y - A\Phi\alpha\|_2^2 + \lambda \sum_{i=1}^{T} \|\alpha_i\|_p^p, \ 0 \leq p < 2.
\]

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:
\[
\alpha^{(n+1)} = HT_{\lambda^{(n)}} \left( \alpha^{(n)} + \Phi^T A^T \left( Y - A\Phi\alpha^{(n)} \right) \right)
\]

For IHT:
\[
\alpha^{(n+1)} = ST_{\lambda^{(n)}} \left( \alpha^{(n)} + \Phi^T A^T \left( Y - A\Phi\alpha^{(n)} \right) \right)
\]

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; etc.
DECONVOLUTION SIMULATION
DECONVOLUTION

The Curse of Missing Data

- Period detection in temporal series
  COROT: HD170987

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


\[ \Theta_\Lambda = \text{Id}_\Lambda \]

\[ \min_{\alpha} \|\alpha\|_{\ell_0} \quad \text{s.t.} \quad y = Mx \]

Where M is the mask: 
- M(i,j) = 0 \implies \text{missing data}
- M(i,j) = 1 \implies \text{good data}

\[ x^{(n+1)} = S_{\Phi,\lambda(n)} \left\{ x^{(n)} + M \left( y - x^{(n)} \right) \right\} \]

Iterative Hard Thresholding with a decreasing threshold.

MCAlab available at: http://www.greyc.ensicaen.fr/~jfadili
Inpainting:


![Original map](image1)
![Masked map](image2)
![Inpainted map](image3)

Power spectrum: 0.3% error
Bispectrum: 1% error
Interpolation of Missing Data: Sparse Inpainting

\[ \min_{\alpha} \| \alpha \|_p \quad \text{subject to} \quad Y = M \Phi \alpha \]

Where \( M \) is the mask: 
- \( M(i,j) = 0 \) ==> missing data
- \( M(i,j) = 1 \) ==> good data

Sparse-Inpainting preserves the ISW and the Weak Lensing signal.


Theoretical justification through the sampling theory of Compressed Sensing?
PLANCK and Sparsity

• Successor of WMAP (better resolution, better sensitivity, more channels)
• Launched on May 14, 2009
• Two instruments LFI and HFI
• Nine Temperature maps at 30, 44, 70, 100, 143, 217, 353, 545, 857 GHz + Polarization

==> DATA Released in 2013
The Cosmic Microwave Background (CMB) is a relic radiation (with a temperature equals to 2.726 Kelvin) emitted 13 billion years ago when the Universe was about 370000 years old.
PLANCK SIMULATED POLARIZED DATA:

Magnitude: \[ P = \sqrt{Q^2 + V^2} \]
Orientation: \[ \alpha = \arctan \left( \frac{U}{Q} \right) \]
The importance of Source Separation
Extra foregrounds are superimposed with the CMB !!!
Point sources, galactic foregrounds, ... etc
Is Sparse models adapted to PLANCK Data?

- Components are sparse in a wavelet dictionary
**Isotropic Undecimated Wavelet on the Sphere**


\[ \hat{\psi}_{\frac{l}{2^j}} (l, m) = \hat{\phi}_{\frac{l}{2^{j-1}}} (l, m) - \hat{\phi}_{\frac{l}{2^j}} (l, m) \]

\[ \hat{H}_j(l, m) = \begin{cases} \frac{\hat{\phi}_{\frac{l}{2^{j+1}}} (l, m)}{\phi_{\frac{l}{2^j}} (l, m)} & \text{if } l < \frac{l}{2^{j+1}} \text{ and } m = 0 \\ 0 & \text{otherwise} \end{cases} \]

\[ \hat{G}_j(l, m) = \begin{cases} \frac{\hat{\psi}_{\frac{l}{2^{j+1}}} (l, m)}{\phi_{\frac{l}{2^j}} (l, m)} & \text{if } l < \frac{l}{2^{j+1}} \text{ and } m = 0 \\ \frac{1}{2^j} & \text{if } l \geq \frac{l}{2^{j+1}} \text{ and } m = 0 \\ 0 & \text{otherwise} \end{cases} \]

\[ \hat{c}_{j+1}(l, m) = \hat{H}_j(l, m) \hat{c}_j(l, m) \]

\[ \hat{w}_{j+1}(l, m) = \hat{G}_j(l, m) \hat{c}_j(l, m) \]

\[ c_0(\vartheta, \varphi) = c_J(\vartheta, \varphi) + \sum_{j=1}^{J} w_j(\vartheta, \varphi) \]
Curvelet functions on the sphere
Dictionaries

- Spherical Harmonics
- Wavelets
- Curvelets
Polarized Dictionary: E/B Polarized Wavelet

Curvelet and E/B Mode Decomposition

E-Curvelet Coefficient Backprojection
Polarized Data Denoising

\[ Q(\theta, \phi) = \sum_{l,m} c_{J,l,m}^E Z_{l,m}^+ + i c_{J,l,m}^B Z_{l,m}^- + \sum_{l,m} \sum_j \tilde{w}_{j,l,m}^E Z_{l,m}^+ + i \tilde{w}_{j,l,m}^B Z_{l,m}^- \]

\[ U(\theta, \phi) = \sum_{l,m} c_{J,l,m}^B Z_{l,m}^+ - i c_{J,l,m}^E Z_{l,m}^- + \sum_{l,m} \sum_j \tilde{w}_{j,l,m}^B Z_{l,m}^+ - i \tilde{w}_{j,l,m}^E Z_{l,m}^- \]

Where

\[ \tilde{w}_{j,k}^E = \delta(w_{j,k}^E) \]

\[ \tilde{w}_{j,k}^B = \delta(w_{j,k}^B) \]

Hard thresholding corresponds to the following non linear operation:

\[ \tilde{w}_{j,k} = \begin{cases} w_{j,k} & \text{if } |w_{j,k}| \geq T_j \\ 0 & \text{otherwise} \end{cases} \]
Polarized Data Denoising
MRS Version V2.0 available since June 2010
Wavelet, Ridgelet and Curvelet on the Sphere:
Software available at: http://jstarck.free.fr/mrs.html

1. Wavelet transforms
   • Continuous Wavelet Transform (Mexican Hat)
   • Orthogonal Wavelets
   • Undecimated isotropic wavelet transform (Spline, Meyer and Needlet filters).
   • Pyramidal wavelet transform
2. Ridgelet and Curvelet Transforms
3. Denoising using Wavelets and Curvelets
4. Gaussianity tests: Skewness, Kurtosis, Moment of order 5 and 6, Max, Higher Criticism
5. Astrophysical Component Separation (ICA on the Sphere): JADE, Fast ICA, GMCA.

Polarized Spherical Wavelets and Curvelets: SparsePol/Version 1.0
Software available at: http://jstarck.free.fr/mrspb.html


Semi-Blind Source Separation

Standard approaches amount to model each observation as a linear mixture of elementary components (i.e. CMB, SZ, Synchrotron, Free-Free, Dust ...) :

$$\forall i; \ x_i = \sum_j a_{ij} s_j + n_j$$

Which can be recast as:

$$X = A S + N$$

Blind source separation: The objective is to estimate both A and S simultaneously !!

- CMB, SZ and Free-Free emission : their electromagnetic spectrum is well known (i.e. the related columns of A are known and fixed).
- Synchrotron emission : rank-1 assumption / its electromagnetic spectrum is a power law with an unknown spectral index.

We have nine channels and we search for nine sources: 4 sources are modeled and 5 are not modeled.
Sparse Component Separation: the GMCA Method


Source: \( S = [s_1, s_n] \)
Data: \( X = [x_1, ..., x_n] = AS + N \)

We define a dictionary \( \phi \)

\[ \{S\} = \text{Argmin}_S \left\{ \sum_j \lambda_j \|s_j W\|_1 + \|X - AS\|^2_{F,\Sigma} \right\} \]

\[ \{A\} = \text{Argmin}_A \|X - AS\|^2_{F,\Sigma} \]

A and S are estimated alternately and iteratively in two steps:

1) Estimate S assuming A is fixed (iterative thresholding):

2) Estimate A assuming S is fixed (a simple least square problem):
GMCA has the lowest spectral residuals at low $l$.
Plot shows $C_l$ of (reconstructed CMB – input CMB) evaluated at high galactic latitudes.

Sam Leach (SISSA), June 19, 2008, WG2 meeting, Munich
Planck provides data that do not share the same resolution:

Planck beam

Spherical harmonic

30GHz
44GHz
70GHz
100GHz
143GHz
Component Separation: more problems

More formally:

\[ \forall i; x_i = b_i \ast \left( \sum_j a_{ij} s_j \right) + n_i \]

Globally:

\[ X = \mathcal{H}(AS) + N \]

where \( \mathcal{H} \) is the multichannel convolution operator

The mixture model no more holds! \( \mathcal{H} \) is singular!

Spectral behavior varies spatially for some components (dust, synchroton):
Component Separation

=> Use Wavelets to work at different resolutions:

Planck beam

=> Assume the mixing matrix varies smoothly

Partitionning of the Wavelet Scales
LOCAL GMCA

Undecimated Isotropic Wavelet Transform + Block Partitioning

j=1

12*32 Blocks

j=2

12*16 Blocks

j=3

12*4 Blocks

j=4

12 Blocks

Global
Wavelet-Vaguelette GMCA Decomposition

Inverse problem

\[ y = Kf + n \quad \xrightarrow{\text{WVD}} \quad f = \sum_j \sum_k \langle Kf, \Psi_{j,k} \rangle \psi_{j,k} \quad \text{with} \quad K^* \Psi_{j,k} = \psi_{j,k} \]

\[ \tilde{f} = \sum_j \sum_k \Delta(\langle y, \Psi_{j,k} \rangle) \psi_{j,k} \]

Multi-channel WVD

\[ Y_i = \sum_j K_i(A_jX)_i + N_i \quad \xrightarrow{\text{WVD}} \quad \tilde{X}_s = \left[ \sum_j \sum_k \tilde{A}_j^+ \left\langle Y_i, \Psi_{j,k}^{(i)} \right\rangle \psi_{j,k} \right]_s \]

\[ \beta_{j,k} = \left\langle Y_i, \Psi_{j,k}^{(i)} \right\rangle = A_j \quad \alpha_{j,k} = \left\langle X_s, \psi_{j,k} \right\rangle \]

GHz

The sparse GMCA solution is obtained by minimizing:

\[ \min_{\alpha_j, A_j} \sum_j \frac{1}{2\sigma^2} \|\beta_j - A_j \alpha_j\|^2 \quad \text{s.t.} \quad \alpha \text{ is sparse} \]
Input CMB map

Input CMB Map
**CMB map estimation**

L-GMCA estimate

Estimate CMB Map
CMB map estimation

Residual map for Global GMCA

Residual per latitude at 60min - Global GMCA
Residual per latitude at 60min – LGMCA

Residual map for L-GMCA
CMB map estimation

Residual per latitude

Residual per band at 60min

Angle in degrees

Standard deviation in mK

0.000  0.005  0.010  0.015

0  50  100  150  200

0.000  0.005  0.010  0.015
Sparsity in Astrophysics

Conclusions

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

- Perspectives
  - PLANCK component separation.
  - Euclid