Sparsity & Co.:

An Overview of *Analysis* vs *Synthesis* in Low-Dimensional Signal Models

Rémi Gribonval

INRIA, France
Outline

• Sparsity: via Analysis or Synthesis ?
• An Lp perspective $0 < p \leq 1$ with localized frames
• An L0 perspective: introducing co-sparsity
• Cosparse recovery ?
• What’s next ?
Sparse Signal / Image Processing

denoising

inpainting

+ Compression,
Source Localization, Separation,
Compressed Sensing ...
Sparse Atomic Decompositions

\[ x \approx Dz \]

Signal Image (Overcomplete) dictionary of atoms Representation Coefficients
Supporting Evidence: Sparsifying \textit{Transforms}

- **natural image**
- **audio signal**
- **sparse coefficients**

\textbf{Sparsity & Co. - Rémi GRIBONVAL}

[SPARS11 - Edinburgh]

[June 28th 2011]
Transforms = Atomic Decompositions?

Instead of showing this

\[ x = Dz \]

Signal of Interest

We have shown this

\[ \Omega x = z \]
Transforms = Atomic Decompositions?

Dictionary $\mathbf{D}$

$\Omega$ is a tight frame

$x = \mathbf{D}\Omega x.$

$x$: sparse in $\mathbf{D}$

$\Omega x$: sparse

$\mathbf{D}^T$
Transforms = Atomic Decompositions?

\[ x = D\Omega x. \]

- Yes ... but some troubling facts:
  - infinitely many synthesis representations
  - only one analysis representation
Transforms = Atomic Decompositions?

Yes ... but some troubling facts:

✓ infinitely many synthesis representations
✓ only one analysis representation

By the way, what do we mean by «sparse»?
Analysis vs Synthesis: Lp sparsity in frames

with M. Nielsen
Frames in Hilbert spaces

• Frame = energy preserving analysis transform

\[ A \cdot \|x\|^2 \leq \|D^T x\|^2 \leq B \cdot \|x\|^2, \quad \forall x \in \mathcal{H}. \]

• Canonical dual frame \( \Omega = D^+ := D^T (DD^T)^{-1} \)
  ✓ perfect reconstruction property

\[ x = D\Omega x, \quad \forall x \in \mathcal{H} \]

✓ minimum energy coefficients

\[ \|\Omega x\|_2 = \min_{z:Dz=x} \|z\|_2 \quad \text{minimum Lp norms?} \]
Measures of sparsity

- Lp (quasi)-norms
  - p=0: $\|z\|_0 := \#\{j: z_j \neq 0\}$
  - p>0: $\|z\|_p := \sum_j |z_j|^p$

- Lp = sparsity-inducing for $0 \leq p \leq 1$
Analysis vs Synthesis sparsity

• For a frame $D$ and its canonical dual

$$\|\Omega x\|_2 = \min_{z: Dz = x} \|z\|_2$$

• Norm associated to sparsest synthesis coefficients

$$|x|_p := \inf_{z: Dz = x} \|z\|_p \leq \|\Omega x\|_p$$

• Converse? $\|\Omega x\|_p \leq C|x|_p$?
Analysis vs Synthesis equivalence: localized frames

• Notations

- Atoms = columns of dictionary $D = [d_j]$, canonical dual $\Omega$

• Theorem

- If: $C_q := \sup_j \| \Omega d_j \|_q < \infty$

- Then $\forall p, q \leq p \leq 2$

$$|x|_p \leq \| \Omega x \|_p \leq C_q |x|_p, \quad \forall x$$

✓ Minimum L2 norm coefficients = near Lp sparsest!

Geometry of Lp balls

• **Synthesis viewpoint**

\[ \{ x : \| x \|_p \leq 1 \} \]

• **Analysis viewpoint**

\[ \{ x : \| \Omega x \|_p \leq 1 \} \]

• **D = 5 random unit atoms**
Geometry of $L_p$ balls

- **Synthesis viewpoint**
  \[ \{ x : |x|_p \leq 1 \} \]

- **Analysis viewpoint**
  \[ \{ x : \| \Omega x \|_p \leq 1 \} \]

- $D = 5$ random unit atoms

- $C_p \approx 50$
Geometry of Lp balls

- **Synthesis viewpoint**
  \[ \{ x : |x|_p \leq 1 \} \]

- **Analysis viewpoint**
  \[ \{ x : \|\Omega x\|_p \leq 1 \} \]

- **Different sizes**: analysis ball smaller than synthesis one

- **Different shapes**: analysis ball has more peaks than synthesis one
  
  - For p=1, see Elad & al 2007

- **D = Dirac \cup\ DCT**

- \[ C_p \approx 35 \]
Transforms = Atomic Decompositions ?

\[ x = D\Omega x. \]

\[ \Omega x : \text{sparse} \]

\[ D \text{ : Tight frame} \]

\[ \Omega = D^T \]
Transforms = Atomic Decompositions ?

\[ \text{Dictionary } D \quad \text{Operator } \Omega = D^T \]

\[ x = D\Omega x. \]

\[ x: \text{ sparse in } D \quad \Omega x: \text{ sparse} \]

- Yes for \textit{localized} frames with Lp norm \(0 < p < 2\)
Transforms = Atomic Decompositions ?

Yes for localized frames with $L^p$ norm $0 < p < 2$

But ...

$x = D \Omega x.$

$x$: sparse in $D$  \quad  $\Omega x$: sparse

Dictionary $D$  \quad  Operator $\Omega$  \quad  $= D^T$
Geometry of sparse coefficients?

Coefficient Domain

Sparse coefficient

Signal Domain

$x = Dz$

Synthesis
Dictionary
Geometry of sparse coefficients?
Transforms = Atomic Decompositions ?

Generic Analysis Operators

• For \textbf{generic} tight frame with n=2d we have

\[ ||\Omega x||_0 \leq d \quad \Rightarrow \quad x = 0 \]

✓ No signal has truly sparse analysis coefficients

• But:

✓ The fact that \( ||\Omega x||_0 < n \) is a \textbf{model} on signal \( x \)

✓ Many analysis operators of interest are \textbf{not generic}
  
  • ex: Casazza, Heinecke, Krahmer, and Kutyniok. Optimally Sparse Frames. 2010. \textit{(Session #1)}
Analysis vs Synthesis: Cosparsity

with S. Nam, M. Davies, M. Elad
Introducing the cosparse model

- **Cosparse analysis model**
  - Analysis operator $\Omega$
  - Representation $\Omega x$
  - Zeroes of the representation

- **Sparse synthesis model**
  - Synthesis dictionary $D$
  - Representation $z$ s.t. $x = Dz$
  - Nonzeroes of the representation
Introducing the cosparse model

- **Cosparse analysis model**
  - Analysis operator $\Omega$
  - Representation $\Omega x$
  - **Zeroes** of the representation

- **Sparse synthesis model**
  - Synthesis dictionary $D$
  - Representation $z$ s.t. $x = Dz$
  - **Nonzeroes** of the representation

\[ \Omega \in \mathbb{R}^{l \times d} \quad \Omega x \quad \xrightarrow{\text{zeros}} \quad x \quad \text{cosparsity} = l \]

\[ \mathbb{R}^{k \times d} \quad z \quad \xrightarrow{\text{nonzeroes}} \quad x \quad \text{sparsity} = k \]

**Footnotes:**
- $\Omega$: Analysis operator
- $x$: Representation
- $\xrightarrow{\text{zeros}}$: Zeroes of the representation
- $\xrightarrow{\text{nonzeroes}}$: Nonzeroes of the representation
- $\text{cosparsity} = l$: Codimension of subspace
- $\text{sparsity} = k$: Dimension of subspace
Introducing the cosparse model

- **Cosparse analysis model**
  - Analysis operator $\Omega$
  - Representation $\Omega x$
  - **Zeroes** of the representation

- **Sparse synthesis model**
  - Synthesis dictionary $D$
  - Representation $z$ s.t. $x = Dz$
  - **Nonzeroes** of the representation

\[ x \in \Omega \quad \text{and} \quad \text{cosparsity} = \ell = \text{codimension of subspace} \]

\[ x = Dz \quad \text{and} \quad \text{sparsity} = k = \text{dimension of subspace} \]
Co-sparsity vs Sparsity

- **Cosparseity**
  - operator \( \Omega : n \times d \)
  - number of zeroes = co-dimension
    \[ \ell := n - \| \Omega x \|_0 \]
  - dimension of subspace
    \[ d - \ell \]
  - number of subspaces
    \( \binom{n}{\ell} \)

- **Sparsity**
  - dictionary \( D : d \times n \)
  - number of nonzeros = dimension
    \[ k := \| z \|_0, \ x = Dz \]
  - dimension of subspace \( k \)
  - number of subspaces
    \( \binom{n}{k} \)
Example 1: Undecimated wavelets

- **Sparse model: wavelet expansions**
  - *support* = location of significant wavelet coefficients
  - a single singularity = a large *footprint*
    - Dragotti & Vetterli 2003

- **Cosparse model ?**
  - *cosupport* = zero-crossings
    - Logan 1977, Mallat 1991

- **Two-scale relations, etc.**
  - linear dependencies
  - allows larger cosparsivity
    - Selesnick & Figueiredo 2009

\[ x = \sum_{j,k} z_{j,k} \psi_{j,k} \]
Example 1: Undecimated wavelets

- **Sparse model: wavelet expansions**
  - \textit{support} = location of significant wavelet coefficients
  - a single singularity = a large \textit{footprint}
  - Dragotti & Vetterli 2003

- **Cosparse model ?**
  - \textit{cosupport} = zero-crossings
  - Logan 1977, Mallat 1991

- **Two-scale relations, etc.**
  - linear dependencies
  - allows larger cosparsity \( \ell > d \)
  - Selesnick & Figueiredo 2009

\[
x = \sum_{j,k} z_{j,k} \psi_{j,k}
\]
Example 2: finite difference operator

• Finite-difference operator = cousin of TV norm
  - Rudin, Osher, Fatemi 1992

\[ x = (x_{i,j}) \]

\[ \Omega_{DIF} x \]

\[ i, j \quad i + 1, j \]

✓ cosupport = edges with equal pixel values

✓ not a frame!

• Loops
  ✦ linear dependencies between rows
  ✦ allows larger cosparsity
Example 2: finite difference operator

• Finite-difference operator = cousin of TV norm
  • Rudin, Osher, Fatemi 1992

\[ x = (x_{i,j}) \]

\[ \Omega_{\text{DIF}} x \]

✓ cosupport = edges with equal pixel values
✓ not a frame!

• Loops
  ✦ linear dependencies between rows
  ✦ allows larger cosparsity \( \ell > d \)
Cosparse recovery ?
Cosparse models and inverse problems
Cosparse models and inverse problems

Coefficient Domain

Signal Domain

Compressed Sensing Domain

Sparse coefficient

Synthesis Dictionary

$x = Dz$

Measurement System

$y = Mx$

$y$
Cosparse models and inverse problems

Coefficient Domain

<table>
<thead>
<tr>
<th>Range ( \Omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>z=( \Omega x )</td>
</tr>
<tr>
<td>Analysis Operator</td>
</tr>
</tbody>
</table>

Signal Domain

<table>
<thead>
<tr>
<th>VS.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthesis Dictionary ( x=Dz )</td>
</tr>
</tbody>
</table>

Compressed Sensing Domain

<table>
<thead>
<tr>
<th>Measurement System</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y=Mx )</td>
</tr>
</tbody>
</table>

y
Cosparse models and inverse problems

Analysis sparsity = form of «structured» sparsity
less sparse leads to lower dimension
Optimization principles / algorithms

- **Idealized problem**
  - Starck & al 2003
  - Portilla 2009
  - Selesnick & Figueiredo 2009
  - Afonso, Bioucas-Dias & Figueiredo 2010 *(Session #9)*

- **Convex relaxation**
  - Elad & al 2007
  - Candès & al 2010

\[
\hat{x}_{A-L0} := \arg \min_{x: y = Mx} \|\Omega x\|_0
\]

\[
\hat{x}_{A-L1} := \arg \min_{x: y = Mx} \|\Omega x\|_1
\]

- **Greedy analysis pursuit (GAP) ~ analysis-OMP**
  - Nam & al 2011 *(Session #9)*

- **Iterative cosparse projections ~ analysis-IHT**
  - Gyries & al 2011
Results: Generic Analysis Operator

\[
M = m \times d \\
\delta = \frac{m}{d} \\
\rho = \frac{d - \ell}{m} \\
\Omega
\]
Results: **Generic Analysis Operator**

\[ M = \begin{bmatrix} m \times d \end{bmatrix} \quad \delta = \frac{m}{d} \quad \rho = \frac{d - \ell}{m} \]

For GENERIC operators, there is a (high) lower bound on achievable undersampling in a Compressed Sensing Scenario ...
Co-sparsity vs Sparsity

- **Co-sparsity**
  - operator
    \[ \Omega : n \times d \]
  - number of zeroes = co-dimension
    \[ \ell := n - \| \Omega x \|_0 \]
  - dimension of subspace
    \[ d - \ell \]
  - number of subspaces
    \[ \binom{n}{\ell} \]

- **Sparsity**
  - dictionary
    \[ D : d \times n \]
  - number of nonzeros = dimension
    \[ k := \| z \|_0, \quad x = Dz \]
  - dimension of subspace
    \[ k \]
  - number of subspaces
    \[ \binom{n}{k} \]
Counting subspaces: **Generic Operators**

**Remark:** sparse / cosparse models describe combinatorially many subspaces with only $n \times d$ parameters.
Counting subspaces: Generic Operators

Remark:
sparse / cosparse models describe combinatorially many subspaces with only $n \times d$ parameters

Few small-dimensional subspaces

(Too) many small-dimensional subspaces

Number of subspaces of dimension $k$ in $\mathbb{R}^d$

$$\log(\#\text{subspaces})/d$$

- **Synthesis model, $n/d=2$**
- **Analysis model, $p/d = 2, l=d-k$**
Results: finite difference operator

Sampling locations of Fourier transform of 256x256 image (4.63% of total)
Results with finite difference operator
Results with \textit{finite difference operator}

Linear dependencies = fewer small-dimensional subspaces
Conclusions
Summary

• **Traditional Sparse Model**
  - **Synthesis dictionary of atoms**
    \[ x = Dz = \sum_i z_i d_i \quad \|z\|_0 \ll \text{dimension} \]
  - «Lego» model: building blocks
  - Low-dimension = few atoms

• **Cosparse Analysis Model**
  - **Analysis operator**
    \[ \langle \omega_i, x \rangle = 0 \quad \text{for many indices} \]
    \[ \|\Omega x\|_0 \ll \text{dimension} \]
  - «Carving out» model: constraints
  - Low-dimension = many constraints
    * Ex: coupling with laws of physics
    \[ (\Delta x)|_{\hat{\Omega}} = 0 \]
Take-home message

• Revisited viewpoint on «transforms vs dictionaries»
  ✓ Concept of cosparsity, contrasted with sparsity
    ✦ Union of subspace models
    ✦ Different relations between number of subspaces and dimension
    ✦ Different role of linear dependencies: seem desirable for inverse problems
  ✓ Co-sparse recovery guarantees with inverse problems
    ✦ Cosparse model more naturally fits «sparse analysis» algorithms
    ✦ New algorithms, recovery guarantees, empirically outperform analysis-L1

• References
What’s next?

- Recovery guarantees for GAP & Analysis-L1
  - (Nam & al, Session #9)

- Learning/designing analysis operators
  - (Rubinstein & Elad, Yaghoobi & al, Session #19)
  - Fadili & Peyré 2011, Ophir & al 2011

- Hybrid sparse/cosparse models
  - (Afonso & al, Session #9)
What’s next?

@SPARS

?
What’s next?
What’s next ?
• Joint work with
  - Morten Nielsen (Aalborg University)
  - Sangnam Nam (INRIA, France)
  - Mike Davies (University of Edinburgh, UK)
  - Miki Elad (The Technion, Israel)

• Design:
  - Jules Espiau (INRIA, France)

• Funding:
  - EU FET-Open
  - small-project.eu

• Join the team for a postdoc! remi.gribonval@inria.fr
  (ERC StG 2011 «PLEASE»)
Bibliography (1)

Bibliography (2)