The Analysis (Co-)Sparse Model
Origin, Definition, Pursuit, Dictionary-Learning and Beyond

Michael Elad
The Computer Science Department
The Technion – Israel Institute of technology
Haifa 32000, Israel

*Joint work with
Ron Rubinstein
Tomer Peleg
Remi Gribonval
Sangnam Nam, Mark Plumbley, Mike Davies, Raja Giryes, Boaz Ophir, Nancy Bertin

Workshop 1224: Applied Harmonic Analysis and Sparse Approximation
June 10-16, 2012
Introduction

Why Models for Signals?
What is the Analysis Model?
- It does not matter what is the data you are working on – if it is carrying information, it has an inner structure.
- This structure = rules the data complies with.
- Signal/image processing heavily relies on these rules.
Who Needs Models?

- Models are central in signal and image processing.
- They are used for various tasks – sampling, IP, separation, compression, detection, ...
- A model is a set of mathematical relations that the data is believed to satisfy.

Effective removal of noise relies on a proper modeling of the signal.
There are many different ways to mathematically model signals and images with varying degrees of success.

The following is a partial list of such models (for images):

Good models should be simple while matching the signals:

**Which Model to Use?**

- Principal-Component-Analysis
- Anisotropic diffusion
- Markov Random Field
- Wiener Filtering
- DCT and JPEG
- Huber-Markov
- Wavelet & JPEG-2000
- Piece-Wise-Smooth
- C2-smoothness
- Besov-Spaces
- UoS
- Total-Variation
- Local-PCA
- Mixture of Gaus.
The fields of signal & image processing are essentially built of an evolution of models and ways to use them for various tasks.
A Model Based on Sparsity & Redundancy

A Sparsity-Based Model

Signal Processing
- Wavelet & Frames
- Multi-Scale Analysis
- Signal Transforms

Machine Learning

Mathematics
- Approximation Theory
- Linear Algebra
- Harmonic Analysis
- Optimization

Signal Processing Applications
- Blind Source Separation
- Compression
- Anomaly Detection
- Denoising
- Inpainting
- Texture-Synthesis
- Classification
- CT-Reconstruction
- Super-Resolution
- Image Fusion
- Demosaicing
- Deblurring

Mathematics Applications
- Approximation Theory
- Linear Algebra
- Harmonic Analysis
- Optimization

Applications
- Denoising
- Compression
- Image Fusion
- Deblurring

The Analysis (Co-)Sparse Model: Definition, Pursuit, Dictionary-Learning and Beyond
By: Michael Elad
What is This Model?

- Task: model image patches of size 10×10 pixels.
- We assume that a dictionary of such image patches is given, containing 256 atom images.
- The sparsity-based model assumption: every image patch can be described as a linear combination of few atoms.
The attention to sparsity-based models has been given mostly to the synthesis option, leaving the analysis almost untouched.

For a long-while these two options were confused, even considered to be (near)-equivalent.

Well ... now we (think that we) know better !! The two are VERY DIFFERENT.
The co-sparse analysis model is a very appealing alternative to the synthesis model, with a great potential for leading us to a new era in signal modeling.
Part I - Background

Recalling the Synthesis Sparse Model, the K-SVD, and Denoising
The Sparsity-Based Synthesis Model

- We assume the existence of a synthesis dictionary $D \in \mathbb{R}^{d \times n}$ whose columns are the atom signals.
- Signals are modeled as sparse linear combinations of the dictionary atoms:
  \[ x = D\alpha \]
- We seek a sparsity of $\alpha$, meaning that it is assumed to contain mostly zeros.
- This model is typically referred to as the synthesis sparse and redundant representation model for signals.
- This model became very popular and very successful in the past decade.
The Synthesis Model – Basics

- The synthesis representation is expected to be sparse: \( \| \alpha \|_0 = k \ll d \)

- Adopting a Bayesian point of view:
  - Draw the support \( T \) (with \( k \) non-zeroes) at random;
  - Choose the non-zero coefficients randomly (e.g. iid Gaussians); and
  - Multiply by \( D \) to get the synthesis signal.

- Such synthesis signals belong to a Union-of-Subspaces (UoS):
  \[ x \in \bigcup_{|T|=k} \text{span}\{D_T\} \quad \text{where} \quad D_T \alpha_T = x \]

- This union contains \( \binom{n}{k} \) subspaces, each of dimension \( k \).
The Synthesis Model – Pursuit

- Fundamental problem: Given the noisy measurements,

\[ y = x + v = D\alpha + v, \quad v \sim N\{0, \sigma^2 I\} \]

recover the clean signal \( x \) – This is a denoising task.

- This can be posed as: \( \hat{\alpha} = \text{ArgMin}_{\alpha} \| y - D\alpha \|^2 \text{ s.t. } \|\alpha\|_0 = k \Rightarrow \hat{x} = D\hat{\alpha} \)

- While this is a (NP-) hard problem, its approximated solution can be obtained by
  - Use \( L_1 \) instead of \( L_0 \) (Basis-Pursuit)
  - Greedy methods (MP, OMP, LS-OMP)
  - Hybrid methods (IHT, SP, CoSaMP).

- Theoretical studies provide various guarantees for the success of these techniques, typically depending on \( k \) and properties of \( D \).
The Synthesis Model – Dictionary Learning

Given Signals: \( \{ y_j = x_j + v_j \mid v_j \sim N(0, \sigma^2 I) \}_{j=1}^{N} \)

\[
\min_{D, A} \left\| DA - Y \right\|_F^2 \\
\text{s.t. } \forall j = 1, 2, \ldots, N \left\| \alpha_j \right\|_0 \leq k
\]

Example are linear combinations of atoms from D

Each example has a sparse representation with no more than k atoms

Field & Olshausen (‘96)
Engan et. al. (‘99)
... 
Gribonval et. al. (‘04)
Aharon et. al. (‘04)
...
The Synthesis Model – K-SVD  
Aharon, Elad & Bruckstein (’04)

Initialize $D$
- e.g. choose a subset of the examples

Sparse Coding
- Use OMP or BP

Dictionary Update
- Column-by-Column by SVD computation

Recall: the dictionary update stage in the K-SVD is done one atom at a time, updating it using ONLY those examples who use it, while fixing the non-zero supports.
This method (and variants of it) leads to state-of-the-art results.
Part II – Analysis?
Source of Confusion

Synthesis and Analysis Denoising

$$\min_{\alpha} \|\alpha\|_p \quad \text{s.t.} \quad \|D\alpha - y\|_2 \leq \varepsilon$$

Synthesis denoising

$$\min_{x} \|\Omega x\|_p \quad \text{s.t.} \quad \|x - y\|_2 \leq \varepsilon$$

Analysis Alternative

These two formulations serve the signal denoising problem, and both are used frequently and interchangeably with $D = \Omega^\dagger$.
Case 1: \( D \) is square and invertible

**Synthesis**

\[
\begin{align*}
\min_{\alpha} \|\alpha\|_p & \quad \text{s.t.} \quad \|D\alpha - y\|_2 \leq \varepsilon \\
\text{Define} \quad x &= D\alpha \\
\text{and thus} \quad D^{-1}x &= \alpha
\end{align*}
\]

**Analysis**

\[
\begin{align*}
\min_{x} \|\Omega x\|_p & \quad \text{s.t.} \quad \|x - y\|_2 \leq \varepsilon \\
\text{Define} \quad D^{-1} &= \Omega \\
\min_{x} \|D^{-1}x\|_p & \quad \text{s.t.} \quad \|x - y\|_2 \leq \varepsilon
\end{align*}
\]
Case 1: \( D \) is square and invertible

**Synthesis**

\[
\min_{\alpha} \| \alpha \|_p \quad \text{s.t.} \quad \| D \alpha - y \|_2 \leq \varepsilon
\]

**Analysis**

\[
\min_{x} \| \Omega x \|_p \quad \text{s.t.} \quad \| x - y \|_2 \leq \varepsilon
\]

Define \( x = D \alpha \)

and thus \( \Omega = D \Sigma^{-1} \)

The Two are Exactly Equivalent

\[
\min_{x} \| D^{-1} x \|_p \quad \text{s.t.} \quad \| x - y \|_2 \leq \varepsilon
\]
Case 2: Redundant $\mathbf{D}$ and $\mathbf{\Omega}$

\[ \alpha = \mathbf{\Omega} x \]
\[ \Rightarrow \mathbf{\Omega}^T \alpha = \mathbf{\Omega}^T \mathbf{\Omega} x \]
\[ \Rightarrow \left( \mathbf{\Omega}^T \mathbf{\Omega} \right)^{-1} \mathbf{\Omega}^T \alpha = \mathbf{\Omega}^+ \alpha = x \]

Define $\alpha = \mathbf{\Omega} x$

and thus $\mathbf{\Omega}^+ \alpha = x$

Analysis

\[
\begin{align*}
\min_{\mathbf{x}} & \quad \|\mathbf{\Omega} \mathbf{x}\|_p^p \\
\text{s.t.} & \quad \|\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon
\end{align*}
\]
Case 2: Redundant $D$ and $\Omega$

**Synthesis**

$$\min_{\alpha} \|\alpha\|_p \text{ s.t. } \|D\alpha - y\|_2 \leq \varepsilon$$

Define $D = \Omega^+$

$$\min_{\alpha} \|\alpha\|_p \text{ s.t. } \|\Omega^+\alpha - y\|_2 \leq \varepsilon$$

**Analysis**

$$\min_{x} \|\Omega x\|_p \text{ s.t. } \|x - y\|_2 \leq \varepsilon$$

Define $\alpha = \Omega x$

and thus $\Omega^+\alpha = x$
Case 2: Redundant $D$ and $\Omega$

**Synthesis**

\[
\min_{\alpha} \|\alpha\|_p \quad \text{s.t.} \quad \|D\alpha - y\|_2 \leq \varepsilon
\]

**Analysis**

\[
\min_{x} \|\Omega x\|_p \quad \text{s.t.} \quad \|x - y\|_2 \leq \varepsilon
\]

Define $x$ and thus $x = \Omega x$

Thus $\Omega^\dagger\alpha = x$

Exact Equivalence again?
Not Really !

\[ \alpha = \Omega x \]
\[ \Rightarrow \Omega^T \alpha = \Omega^T \Omega x \]
\[ \Rightarrow \left( \Omega^T \Omega \right)^{-1} \Omega^T \alpha = \Omega^+ \alpha = x \]

We should require
\[ \Omega x = \alpha = \Omega \Omega^+ \alpha \]

The vector \( \alpha \) defined by \( \alpha = \Omega x \) must be spanned by the columns of \( \Omega \). Thus, what we actually got is the following analysis-equivalent formulation

\[
\begin{align*}
\min_{\alpha} \| \alpha \|_p^p \quad \text{s.t.} \quad \| \mathbf{D} \alpha - y \|_2 \leq \varepsilon \quad \& \quad \alpha = \Omega \Omega^+ \alpha
\end{align*}
\]

which means that \textit{analysis} \( \neq \) \textit{synthesis} in general.
So, Which is Better? Which to Use?

- Our paper [Elad, Milanfar, & Rubinstein (‘07)] was the first to draw attention to this dichotomy between analysis and synthesis, and the fact that the two may be substantially different.

- We concentrated on $p=1$, showing that
  - The two formulations refer to very different models,
  - The analysis is much richer, and
  - The analysis model may lead to better performance.

- In the past several years there is a growing interest in the analysis formulation (see recent work by Portilla et. al., Figueiredo et. al., Candes et. al., Shen et. al., Nam et. al., Fadiliy & Peyré, Kutyniok et. al., Ward and Needel, ...).

- Our goal: better understanding of the analysis model, its relation to the synthesis, and how to make the best of it in applications.
Part III - Analysis

A Different Point of View Towards the Analysis Model

The Analysis Model – Basics

- The analysis representation $z$ is expected to be sparse
  \[ \| \Omega x \|_0 = \| z \|_0 = p - \ell \]

- **Co-sparsity**: $\ell$ - the number of zeros in $z$.

- **Co-Support**: $\Lambda$ - the rows that are orthogonal to $x$
  \[ \Omega_{\Lambda} x = 0 \]

- If $\Omega$ is in **general position***, then $0 \leq \ell < d$ and thus we cannot expect to get a truly sparse analysis representation – Is this a problem? Not necessarily!

- This model puts an emphasis on the zeros in the analysis representation, $z$, rather then the non-zeros, in characterizing the signal. This is much like the way zero-crossings of wavelets are used to define a signal [Mallat (‘91)].

* $\text{spark}\{\Omega^T\} = d + 1$
The Analysis Model – Bayesian View

- Analysis signals, just like synthesis ones, can be generated in a systematic way:

<table>
<thead>
<tr>
<th></th>
<th>Synthesis Signals</th>
<th>Analysis Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Support:</strong></td>
<td>Choose the support ( T ) ((</td>
<td>T</td>
</tr>
<tr>
<td><strong>Coeff.:</strong></td>
<td>Choose ( \alpha_T ) at random</td>
<td>Choose a random vector ( \nu )</td>
</tr>
<tr>
<td><strong>Generate:</strong></td>
<td>Synthesize by: ( D_T \alpha_T = x )</td>
<td>Ortho ( \nu ) w.r.t. ( \Omega_\Lambda ): ( x = \left[ I - \Omega_\Lambda^T \Omega_\Lambda \right] \nu )</td>
</tr>
</tbody>
</table>

- Bottom line: an analysis signal \( x \) satisfies: \( \exists \Lambda \ | |\Lambda| = \ell \) s.t. \( \Omega_\Lambda x = 0 \)
The Analysis Model – UoS

- Analysis signals, just like synthesis ones, leads to a union of subspaces:

<table>
<thead>
<tr>
<th></th>
<th>Synthesis Signals</th>
<th>Analysis Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is the Subspace Dimension:</strong></td>
<td>k</td>
<td>d - \ell</td>
</tr>
<tr>
<td><strong>How Many Subspaces:</strong></td>
<td>\binom{n}{k}</td>
<td>\binom{p}{\ell}</td>
</tr>
<tr>
<td><strong>Who are those Subspaces:</strong></td>
<td>\text{span}{D_T}</td>
<td>\text{span}^\perp{\Omega_A}</td>
</tr>
</tbody>
</table>

- The analysis and the synthesis models offer both a UoS construction, but these are very different from each other in general.
The Analysis Model – Count of Subspaces

- Example: p=n=2d:
  - Synthesis: k=1 (one atom) – there are $2d$ subspaces of dimensionality 1.
  - Analysis: $\ell = d-1$ leads to \( \binom{2d}{d-1} \gg O(2^d) \) subspaces of dimensionality 1.

- In the general case, for d=40 and p=n=80, this graph shows the count of the number of subspaces. As can be seen, the two models are substantially different, the analysis model is much richer in low-dim., and the two complete each other.

- The analysis model tends to lead to a richer UoS. Are these good news?
The Analysis Model – Pursuit

- Fundamental problem: Given the noisy measurements,
  \[ y = x + v, \quad \exists |\Lambda| = \ell \quad \text{s.t.} \quad \Omega_{\Lambda} x = 0, \quad v \sim \mathcal{N}\{0, \sigma^2 I\} \]
  recover the clean signal \( x \) – This is a denoising task.

- This goal can be posed as:
  \[
  \hat{x} = \text{ArgMin} \| y - x \|_2^2 \quad \text{s.t.} \quad \| \Omega x \|_0 = p - \ell
  \]

- This is a (NP-) hard problem, just as in the synthesis case.

- We can approximate its solution by \( L_1 \) replacing \( L_0 \) (BP-analysis), Greedy methods (OMP, ...), and Hybrid methods (IHT, SP, CoSaMP, ...).

- Theoretical studies should provide guarantees for the success of these techniques, typically depending on the co-sparsity and properties of \( \Omega \). This work has already started [Candès, Eldar, Needell, & Randall (’10)], [Nam, Davies, Elad, & Gribonval, (’11)], [Vaiter, Peyré, Dossal, & Fadili, (’11)], [Giryes et. al. (’12)].
The Analysis Model – Backward Greedy

BG finds one row at a time from \( \Lambda \) for approximating the solution of

\[
\hat{x} = \underset{\alpha}{\text{ArgMin}} \| y - x \|_2^2 \quad \text{s.t.} \quad \| \Omega x \|_0 = p - \ell
\]

\[
i = 0, \quad \hat{x}_0 = y, \quad \Lambda_0 = \{ \}
\]

Stop condition? 
(e.g. \( i = \ell \))

Yes

Output \( x_i \)

No

\[
i = i + 1, \quad \Lambda_i = \Lambda_{i-1} \cup \text{ArgMin} \left| w_k^T \hat{x}_{i-1} \right|
\]

\[
\hat{x}_i = \left[ I - \Omega_{\Lambda_i}^\dagger \Omega_{\Lambda_i} \right] y
\]
The Analysis Model – Backward Greedy

Is there a similarity to a synthesis pursuit algorithm?

\[
i = 0, \quad r_0 = y, \quad \Lambda_0 = \{ \} \quad \text{Stop condition? (e.g. } i = \ell \text{)}
\]

\[
i = i + 1, \quad \Lambda_i = \Lambda_{i-1} \cup \text{ArgMax} |d_k^T r_{i-1}| \quad \text{No}
\]

\[
\begin{align*}
r_i & = [I - D_{\Lambda} D_{\Lambda}^+] y \\
n & = \text{Output } = y - r_i
\end{align*}
\]
The Analysis Model – Backward Greedy

Is there a similarity to a synthesis pursuit algorithm?

Other options:
- A Gram-Schmidt acceleration of this algorithm.
- Optimized BG pursuit (xBG) [Rubinstein, Peleg & Elad (‘12)]
- Greedy Analysis Pursuit (GAP) [Nam, Davies, Elad & Gribonval (‘11)]
- Iterative Cosparse Projection [Giryes, Nam, Gribonval & Davies (‘11)]
- $L_p$ relaxation using IRLS [Rubinstein (‘12)]
- CoSaMP, SP, IHT and IHP analysis algorithms [Giryes et. al. (‘12)]

Stop condition? 

No

Yes

Output $x$ = $y - r_i$

$i = 0$

$i = i + 1$
The Analysis Model – Low-Spark $\Omega$

- What if $\text{spark}(\Omega^T) << d$?
- For example: a TV-like operator for image-patches of size 6×6 pixels ($\Omega$ size is 72×36).
- Here are analysis-signals generated for co-sparsity ($\ell$) of 32:

- Their true co-sparsity is higher – see graph:
- In such a case we may consider $\ell > d$, and thus
- ... the number of subspaces is smaller.

\[ \Omega = \begin{bmatrix} \text{Horizontal Derivative} \\ \text{Vertical Derivative} \end{bmatrix} = \begin{bmatrix} \text{Horizontal} \\ \text{Vertical} \end{bmatrix} \]

Horizontal Derivative
Vertical Derivative

Co-Sparsity

# of signals
The Analysis Model – The Signature

Consider two possible dictionaries:

Ω_{DIF}

Spark(Ω^T) = 4

Random Ω

Spark(Ω^T) = 37

The Signature of a matrix is more informative than the Spark
The Analysis Model – Low-Spark $\Omega$ – Pursuit

- An example – performance of BG (and xBG) for these TV-like signals:
- 1000 signal examples, SNR=25.

We see an effective denoising, attenuating the noise by a factor $\sim 0.2$. This is achieved for an effective co-sparsity of $\sim 55$. 

![Denoising Performance Graph]

$E\left\{\left\| x - \hat{x} \right\|_2^2\right\}$

$\frac{d \cdot \sigma^2}{2}$

Co-Sparsity in the Pursuit
Synthesis vs. Analysis – Summary

- The two align for $p=m=d$: non-redundant.

- Just as the synthesis, we should work on:
  - Pursuit algorithms (of all kinds) – Design.
  - Pursuit algorithms (of all kinds) – Theoretical study.
  - Dictionary learning from example-signals.
  - Applications ...

- Our experience on the analysis model:
  - Theoretical study is harder.
  - The role of inner-dependencies in $\Omega$?
  - Great potential for modeling signals.
Part IV – Dictionaries

Analysis

Dictionary-Learning by K-SVD-Like Algorithm


Goal: given a set of signals, find the analysis dictionary $\Omega$ that best fit them.
Analysis Dictionary Learning – The Signals

We are given a set of $N$ contaminated (noisy) analysis signals, and our goal is to recover their analysis dictionary, $\Omega$

$$\begin{align*}
\{y_j &= x_j + v_j, \quad \exists |\Lambda_j| = \ell \quad \text{s.t.} \quad \Omega_{\Lambda_j} x_j = 0, \quad v \sim N\{0, \sigma^2 I\}\}_j^{N} \end{align*}$$
The Analysis (Co-)Sparse Model: Definition, Pursuit, Dictionary-Learning and Beyond

By: Michael Elad

Analysis Dictionary Learning – Goal

Min\(_{D,A}\) \(\|DA - Y\|_F^2\) s.t. \(\forall j = 1,2,\ldots,N\) \(\|\alpha_j\|_0 \leq k\)

Min\(_{\Omega,X}\) \(\|X - Y\|_F^2\) s.t. \(\forall j = 1,2,\ldots,N\) \(\|\Omega x_j\|_0 \leq p - \ell\)

Noisy Examples

Denoised Signals are \(L_0\) Sparse

We shall adopt a similar approach to the K-SVD for approximating the minimization of the analysis goal.
The Analysis (Co-)Sparse Model: Definition, Pursuit, Dictionary-Learning and Beyond

By: Michael Elad

Analysis K-SVD – Outline [Rubinstein, Peleg & Elad (‘12)]

\[
\begin{bmatrix}
\Omega \\
\end{bmatrix}
\begin{bmatrix}
X & \ldots \\
\end{bmatrix}
= 
\begin{bmatrix}
Z & \ldots \\
\end{bmatrix}
\]

Initialize \( \Omega \) \rightarrow Sparse Code \rightarrow Dictionary Update
Analysis K-SVD – Sparse-Coding Stage

\[
\min_{\Omega, x} \|X - Y\|_F^2 \quad \text{s.t.} \quad \forall j = 1, 2, \ldots, N \quad \|\Omega x_j\|_0 \leq p - \ell
\]

Assuming that \(\Omega\) is fixed, we aim at updating \(X\)

\[
\hat{x}_j = \arg\min_x \|x - y_j\|_2^2 \quad \text{s.t.} \quad \|\Omega x\|_0 \leq p - \ell
\]

These are \(N\) separate analysis-pursuit problems. We suggest to use the BG or the xBG algorithms.
Analysis K-SVD – Dictionary Update Stage

\[
\begin{pmatrix}
\Omega \\
\vdots
\end{pmatrix}
\begin{pmatrix}
X \\
\vdots
\end{pmatrix} =
\begin{pmatrix}
Z \\
\vdots
\end{pmatrix}
\]

\[
\begin{align*}
\min_{\Omega,x} & \quad \|X - \Omega x\|^2_F \\
\text{s.t.} & \quad \forall j = 1, 2, \ldots, N \quad \|\Omega x_j\|_0 \leq p - \ell
\end{align*}
\]

- Only signals orthogonal to the atom should get to vote for its new value.
- The known supports should be preserved.
After the sparse-coding, $\Lambda_j$ are known. We now aim at updating a row (e.g. $w_k^T$) from $\Omega$.

We use only the signals $S_k$ that are found orthogonal to $w_k$.

Each example should keep its co-support $\Lambda_j \setminus k$.

Avoid trivial solution.

Minimize $\|X_k - Y_k\|_2^2$ subject to:

- $\forall j \in S_k, \Omega_j x_j = 0$
- $w_k^T X_k = 0$
- $\|w_k\|_2 = 1$

Each of the chosen examples should be orthogonal to the new row $w_k$. 
Analysis Dictionary – Dic. Update (3)

This problem we have defined is too hard to handle.

Intuitively, and in the spirit of the K-SVD, we could suggest the following alternative:

\[
\min_{w_k, x_k} \left\| X_k - Y_k \right\|_2^2 \quad \text{s.t.} \quad \begin{cases}
\forall j \in S_k : \Omega_j x_j = 0 \\
w_k^T X_k = 0 \\
\left\| w_k \right\|_2 = 1
\end{cases}
\]

**WRONG!**
A better approximation for our original problem is

\[
\min_{w_k, x_k} \left\| x_k - y_k \right\|_2^2 \quad \text{s.t.} \quad \forall j \in S_k, \Omega_{j-1} y_j = 0, \quad \left\| w_k \right\|_2 = 1
\]

which is equivalent to

\[
\min_{w_k} \left\| w_k^T y_k \right\|_2^2 \quad \text{s.t.} \quad \left\| w_k \right\|_2 = 1
\]

The obtained problem is a simple eigenvalue approximation problem, easily given by SVD.
Part V – Results

For Dictionary-Learning and Image Denoising
Synthetic experiment #1: TV-Like $\Omega$

- We generate 30,000 TV-like signals of the same kind described before ($\Omega$: 72×36, $\ell$=32)
- We apply 300 iterations of the Analysis K-SVD with BG (fixed $\ell$), and then 5 more using the xBG
- Initialization by orthogonal vectors to randomly chosen sets of 35 examples
- Additive noise: SNR=25. Atom detected if: $1 - \left| w^T \hat{w} \right| < 0.01$

Even though we have not identified $\Omega$ completely (~92% this time), we got an alternative feasible analysis dictionary with the same number of zeros per example, and a residual error within the noise level.
Synthetic experiment #1: TV-Like $\Omega$

Original Analysis Dictionary

Learned Analysis Dictionary
Synthetic experiment #2: Random $\Omega$

- Very similar to the above, but with a random (full-spark) analysis dictionary $\Omega$: $72 \times 36$
- Experiment setup and parameters: the very same as above
- In both algorithms: replacing BG by $x$BG (in both experiments) leads to a consistent descent in the relative error, and better recovery results.

As in the previous example, even though we have not identified $\Omega$ completely (~80% this time), we got an alternative feasible analysis dictionary with the same number of zeros per example, and a residual error within the noise level.
Experiment #3: Piece-Wise Constant Image

- We take 10,000 patches (+noise $\sigma=5$) to train on

- Here is what we got:

Original Image

Patches used for training

Initial $\Omega$

Trained (100 iterations) $\Omega$
Analysis Dictionary Learning – Results (4)

Experiment #4: The Image “House”

- We take 10,000 patches (+noise $\sigma=10$) to train on
- Here is what we got:

- Initial $\Omega$
- Trained $\Omega$ (100 iterations)
- Patches used for training
- Original Image

We take 10,000 patches (+noise $\sigma=10$) to train on.

Here is what we got:

- Initial $\Omega$
- Trained $\Omega$ (100 iterations)
- Patches used for training
- Original Image
Experiment #5: A set of Images

- We take 5,000 patches from each image to train on.
- Block-size 8×8, dictionary size 100×64. Co-sparsity set to 36.
- Here is what we got:

Original Images

Localized and oriented atoms

Trained $\Omega$
(100 iterations)
Back to Image Denoising – (1)

Non-flat patch examples

256×256
Synthesis K-SVD Dictionary Learning:

- Training set – 10,000 noisy non-flat 5x5 patches.
- Initial dictionary – 100 atoms generated at random from the data.
- 10 iterations – sparsity-based OMP with k=3 for each patch example.
  (dimension 4, 3 atoms + DC) + K-SVD atom update.

Patch Denoising – error-based OMP with $\varepsilon^2=1.3\sigma^2$.

Image Reconstruction – Average overlapping patch recoveries.
Back to Image Denoising – (3)

- **Analysis K-SVD Dictionary Learning**
  - Training set – 10,000 noisy non-flat 5x5 patches.
  - Initial dictionary – 50 rows generated at random from the data.
  - 10 iterations – rank-based OBG with r=4 for each patch example + constrained atom update (sparse zero-mean atoms).
  - Final dictionary – keep only 5-sparse atoms.

- **Patch Denoising** – error-based OBG with $\varepsilon^2=1.3d\sigma^2$.

- **Image Reconstruction** – Average overlapping patch recoveries.
Learned dictionaries for $\sigma=5$

Analysis Dictionary

Synthesis Dictionary

- 38 atoms
- 100 atoms
Back to Image Denoising – (5)

<table>
<thead>
<tr>
<th></th>
<th>BM3D</th>
<th>Synthesis K-SVD</th>
<th>Analysis K-SVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average subspace dimension</td>
<td>n/a</td>
<td>2.42</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.03</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.79</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.69</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Cell Legend:

<table>
<thead>
<tr>
<th>σ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>σ=5</td>
<td>σ=10</td>
</tr>
<tr>
<td>σ=15</td>
<td>σ=20</td>
</tr>
</tbody>
</table>
### Back to Image Denoising – (6)

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>BM3D</th>
<th>Synthesis K-SVD</th>
<th>Analysis K-SVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma=5$</td>
<td><img src="image1.png" alt="BM3D_5" /></td>
<td><img src="image2.png" alt="Synthesis_5" /></td>
<td><img src="image3.png" alt="Analysis_5" /></td>
</tr>
<tr>
<td>Scaled to [0,20]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma=10$</td>
<td><img src="image4.png" alt="BM3D_10" /></td>
<td><img src="image5.png" alt="Synthesis_10" /></td>
<td><img src="image6.png" alt="Analysis_10" /></td>
</tr>
<tr>
<td>Scaled to [0,40]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Part VI – We Are Done

Summary and Conclusions
Today ...

Sparsity and Redundancy are practiced mostly in the context of the synthesis model.

Yes, the analysis model is a very appealing (and different) alternative, worth looking at.

Is there any other way?

In the past few years there is a growing interest in this model, deriving pursuit methods, analyzing them, designing dictionary-learning, etc.

What next?

So, what to do?

• Deepening our understanding
• Applications?
• Combination of signal models ...  

More on these (including the slides and the relevant papers) can be found in http://www.cs.technion.ac.il/~elad
The Analysis Model is Exciting Because

- It poses mirror questions to practically every problem that has been treated with the synthesis model.
- It leads to unexpected avenues of research and new insights – E.g. the role of the coherence in the dictionary.
- It poses an appealing alternative model to the synthesis one, with interesting features and a possibility to lead to better results.
- Merged with the synthesis model, such constructions could lead to new and far more effective models.
Thank You all!

And thanks are also in order to the organizers,
Ingrid Daubechies,
Gitta Kutyniok,
Holger Rauhut, and
Thomas Strohmer

More on these (including the slides and the relevant papers) can be found in http://www.cs.technion.ac.il/~elad