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

Mathematisches Forschungsinstitut Oberwolfach

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

*Joint work with


Ron Rubinstein


Tomer Peleg


Remi Gribonval

and

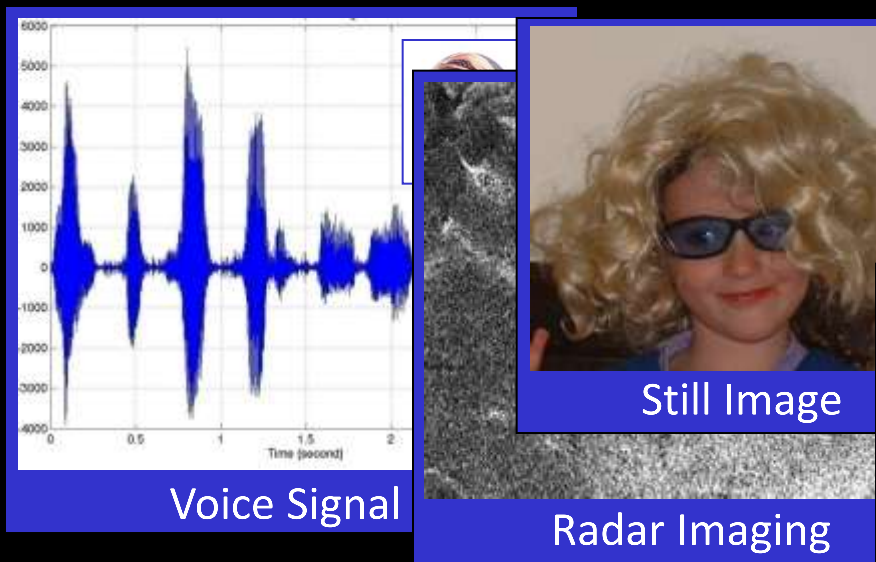
Sangnam Nam, Mark Plumbley, Mike Davies, Raja Giryes, Boaz Ophir, Nancy Bertin

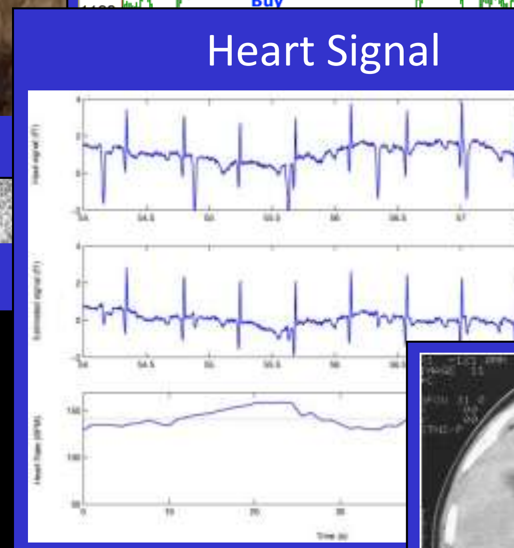
Technion
Israel Institute of Technology

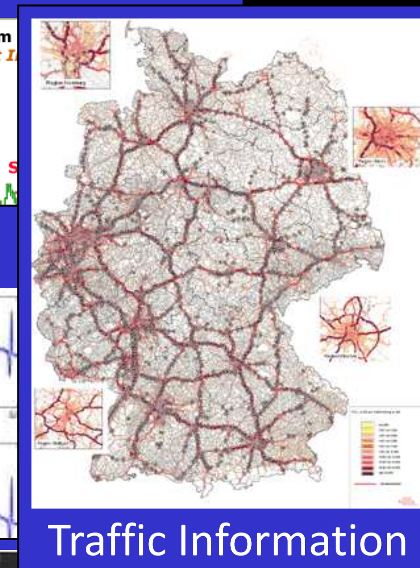
Introduction

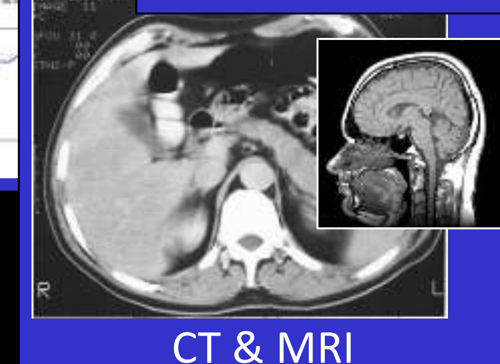
Why Models for Signals?
What is the Analysis Model?

Informative Data → Inner Structure



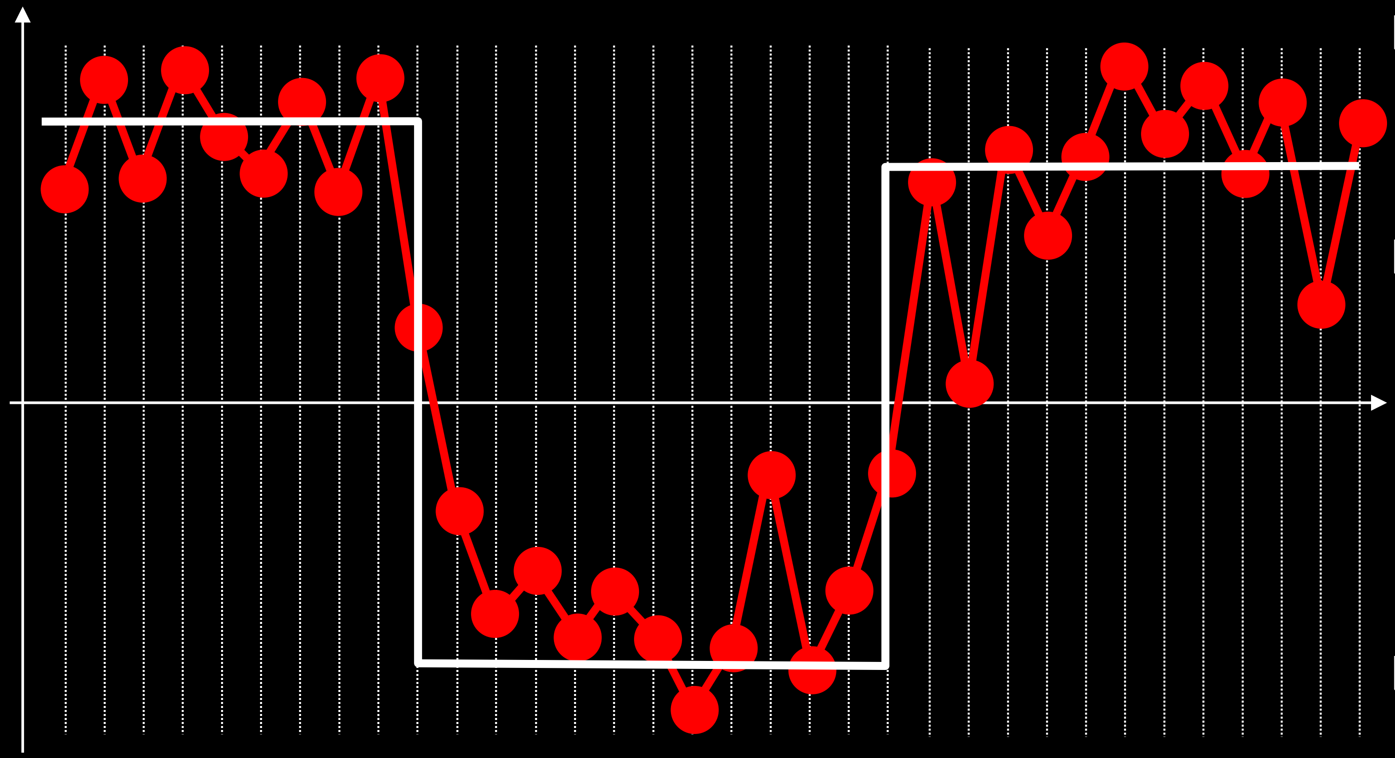






- ❑ It does not matter what is the data you are working on – if it is carrying information, it has an inner structure.
- ❑ This structure = rules the data complies with.
- ❑ Signal/image processing heavily relies on these rules.

Who Needs Models?



Effective removal of noise relies on a proper modeling of the signal

- 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.

Which Model to Use?

- 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:

Simplicity ⟷ Reliability

Principal-Component-Analysis

Anisotropic diffusion

Markov Random Field

Wienner Filtering

DCT and JPEG

Huber-Markov

Wavelet & JPEG-2000

Piece-Wise-Smooth

C2-smoothness

Besov-Spaces

UoS

Total-Variation

Local-PCA

Mixture of Gaus.

Research in Signal/Image Processing

Model

Problem (Application)

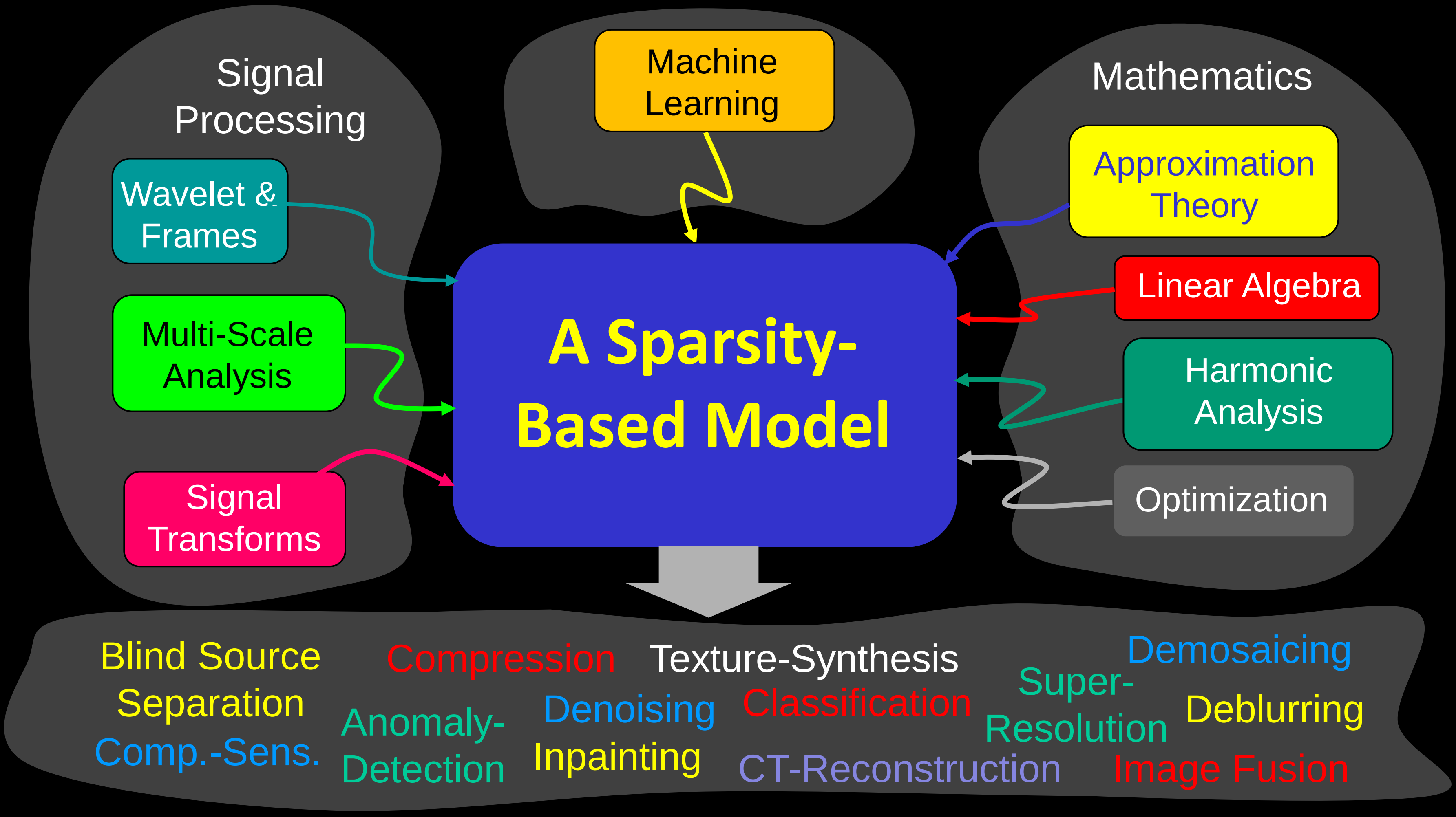
Signal

Numerical Scheme & Math. Statements

A New Research Paper is Born

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



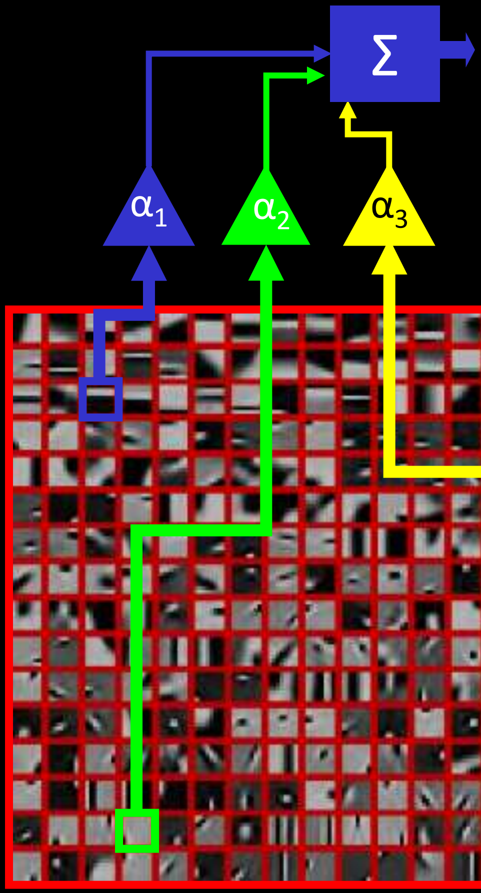
What is This Model?

- 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.

→ Chemistry of Data



However …

Sparsity and Redundancy can be Practiced in two different ways

Synthesis
as presented above

Analysis

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

This Talk is About the Analysis Model

Part I – Recalling the Sparsity-Based Synthesis Model

Part II – Analysis Model – Source of Confusion

Part III – Analysis Model – a New Point of View

Part IV – Analysis K-SVD Dictionary Learning

Part V – Some Preliminary Results and Prospects

The message:

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

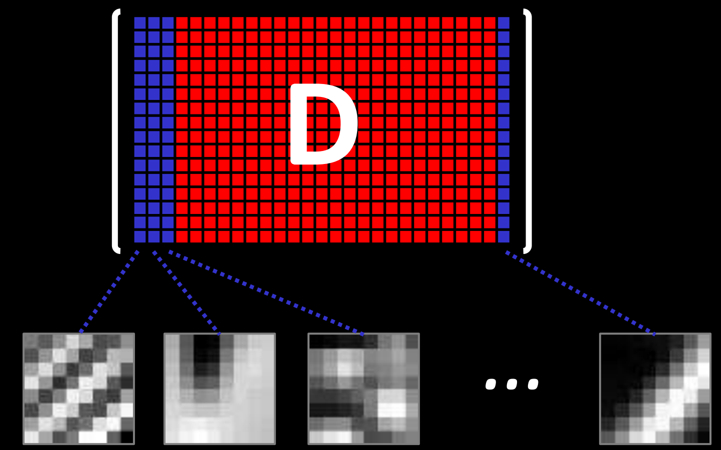
Recalling the Synthesis Sparse Model, the K-SVD, and Denoising

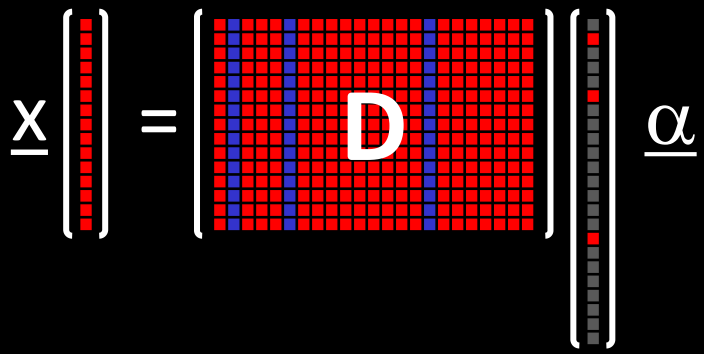
The Sparsity-Based Synthesis Model

- We assume the existence of a synthesis dictionary $\mathbf{D} \in \mathbb{R}^{d \times n}$ whose columns are the atom signals.
- Signals are modeled as sparse linear combinations of the dictionary atoms:

$$\boxed{\underline{x} = \mathbf{D}\underline{\alpha}}$$

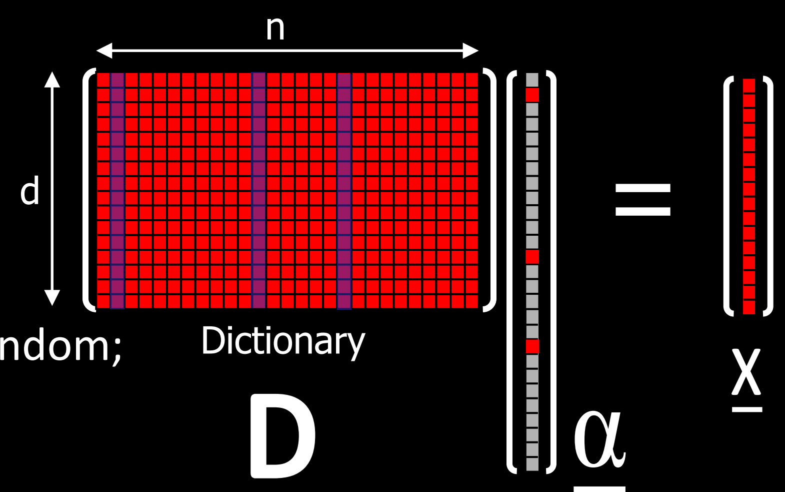
- We seek a sparsity of $\underline{\alpha}$, meaning that it is assumed to contain mostly zeros.
- This model is typically referred to as the synthesis sparse and redundant representation model for signals.
- This model became very popular and very successful in the past decade.





The Synthesis Model – Basics

- The synthesis representation is expected to be sparse: $\|\underline{\alpha}\|_0 = k << d$
- 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):

$$\underline{x} \in \bigcup_{|T|=k} \operatorname{span}\{\mathbf{D}_T\} \quad \text{where} \quad \mathbf{D}_T \underline{\alpha}_T = \underline{x}$$

- This union contains $\binom{n}{k}$ subspaces, each of dimension k.

The Synthesis Model – Pursuit

- Fundamental problem: Given the noisy measurements,

$$\underline{y} = \underline{x} + \underline{v} = \mathbf{D}\underline{\alpha} + \underline{v}, \qquad \underline{v} \sim \mathbf{N}\left\{\underline{0}, \sigma^2 \mathbf{I}\right\}$$

  recover the clean signal $\underline{x}$ – This is a denoising task.

- This can be posed as: $\hat{\underline{\alpha}} = \underset{\underline{\alpha}}{\mathrm{ArgMin}} \left\| \underline{y} - \mathbf{D}\underline{\alpha} \right\|_2^2 \text{ s.t. } \left\| \underline{\alpha} \right\|_0 = k \Rightarrow \hat{\underline{x}} = \mathbf{D}\hat{\underline{\alpha}}$

- 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).

  Pursuit Algorithms

- Theoretical studies provide various guarantees for the success of these techniques, typically depending on k and properties of **D**.

The Synthesis Model – Dictionary Learning

$$\left[\ \mathbf{X}\ \cdots\ \right] = \left[\ \mathbf{D}\ \right]\left[\ \mathbf{A}\ \cdots\ \right]$$

Given Signals : $\left\{\underline{y}_j = \underline{x}_j + \underline{v}_j \quad \underline{v}_j \sim \mathbf{N}\left\{\underline{0}, \sigma^2 \mathbf{I}\right\}\right\}_{j=1}^{N}$

$$\min_{\mathbf{D},\mathbf{A}} \left\|\mathbf{DA}-\mathbf{Y}\right\|_F^2 \ \text{s.t.}\ \forall j=1,2,\ldots,N \quad \left\|\underline{\alpha}_j\right\|_0 \le k$$

Example are linear combinations of atoms from **D**

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

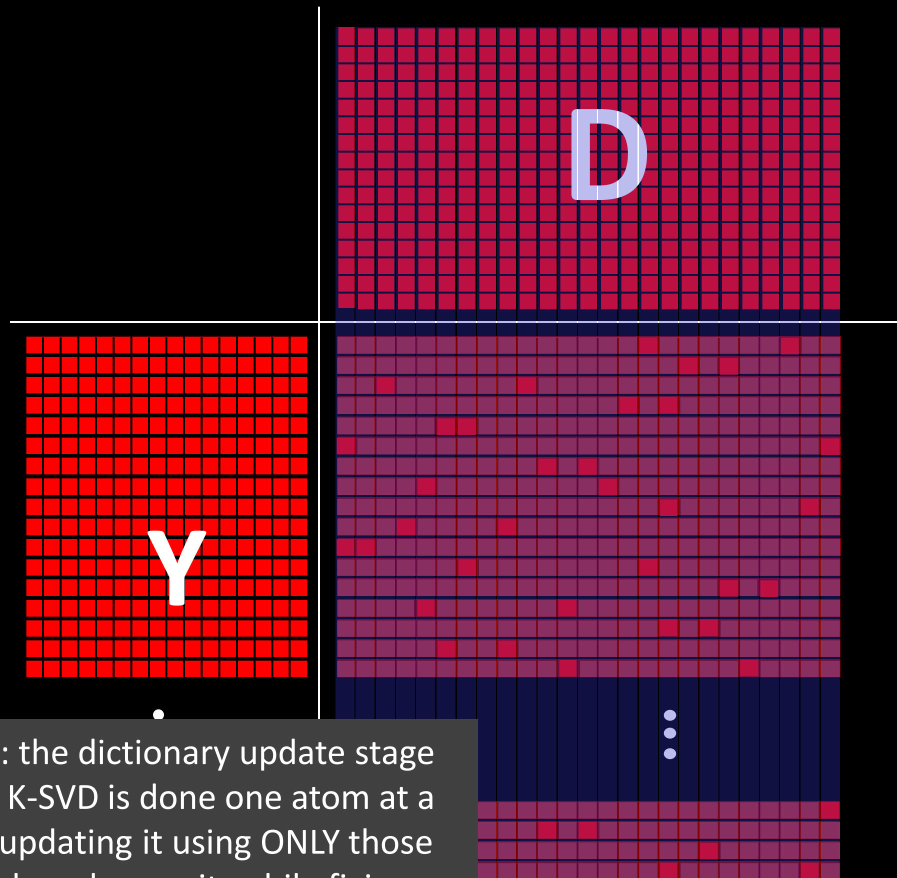
Field & Olshausen (\`96)
Engan et. al. (\`99)
…
Gribonval et. al. (\`04)
Aharon et. al. (\`04)
…

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

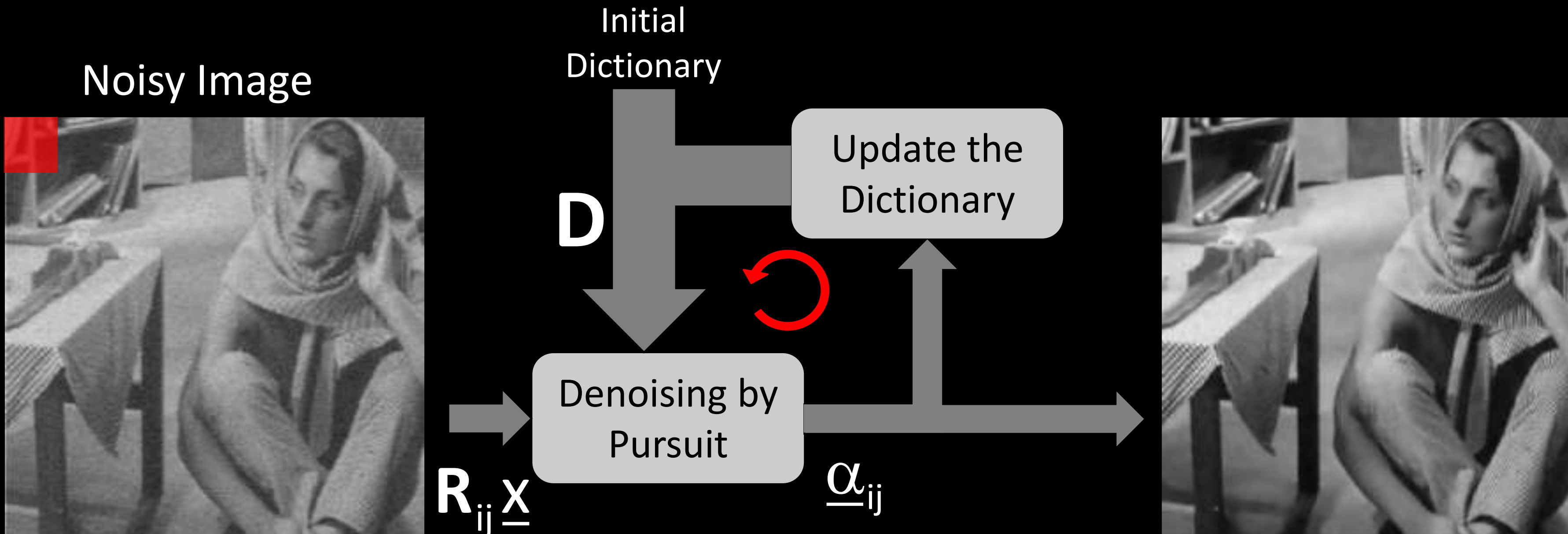


Recall: the dictionary update stage in the K-SVD is done one atom at a time, updating it using ONLY those examples who use it, while fixing the non-zero supports.

Synthesis Model – Image Denoising Elad & Aharon (`06)

$$\hat{\underline{x}} = \underset{\underline{x},\{\underline{\alpha}_{ij}\}_{ij},\mathbf{D}}{\text{ArgMin}} \ \frac{1}{2}\left\|\underline{x}-\underline{y}\right\|_2^2 + \mu\sum_{ij}\left\|\mathbf{R}_{ij}\underline{x}-\mathbf{D}\underline{\alpha}_{ij}\right\|_2^2 \ \text{ s.t. } \ \left\|\underline{\alpha}_{ij}\right\|_0^0 \le L$$

Noisy Image



Reconstructed Image

This method (and variants of it) leads to state-of-the-art results.

Part II – Analysis?
Source of Confusion

M. Elad, P. Milanfar, and R. Rubinstein, "Analysis Versus Synthesis in Signal Priors", *Inverse Problems*. Vol. 23, no. 3, pages 947-968, June 2007.

Synthesis and Analysis Denoising

$$\min_{\underline{\alpha}} \|\underline{\alpha}\|_p^p \;\; \text{s.t.} \;\; \|\mathbf{D}\underline{\alpha} - \underline{y}\|_2 \le \varepsilon$$

Synthesis denoising

$$\min_{\underline{x}} \|\mathbf{\Omega}\underline{x}\|_p^p \;\; \text{s.t.} \;\; \|\underline{x} - \underline{y}\|_2 \le \varepsilon$$

Analysis Alternative

These two formulations serve the signal denoising problem, and both are used frequently and interchangeably with $\mathbf{D}=\mathbf{\Omega}^{\dagger}$

Case 1: **D** is square and invertible

Synthesis

$$\underset{\underline{\alpha}}{\text{Min}} \left\| \underline{\alpha} \right\|_p^p \text{ s.t. } \left\| \mathbf{D}\underline{\alpha} - \underline{y} \right\|_2 \le \varepsilon$$

Define $\underline{x} = \mathbf{D}\underline{\alpha}$

and thus $\mathbf{D}^{-1}\underline{x} = \underline{\alpha}$

Analysis

$$\underset{\underline{x}}{\text{Min}} \left\| \mathbf{\Omega}\underline{x} \right\|_p^p \text{ s.t. } \left\| \underline{x} - \underline{y} \right\|_2 \le \varepsilon$$

Define $\mathbf{D}^{-1} = \mathbf{\Omega}$

$$\underset{\underline{x}}{\text{Min}} \left\| \mathbf{D}^{-1}\underline{x} \right\|_p^p \text{ s.t. } \left\| \underline{x} - \underline{y} \right\|_2 \le \varepsilon$$

Case 1: **D** is square and invertible

Synthesis

$$\underset{\underline{\alpha}}{\text{Min}} \left\|\underline{\alpha}\right\|_p^p \text{ s.t. } \left\|\mathbf{D}\underline{\alpha}-\underline{y}\right\|_2 \le \varepsilon$$

Analysis

$$\underset{\underline{x}}{\text{Min}} \left\|\mathbf{\Omega}\underline{x}\right\|_p^p \text{ s.t. } \left\|\underline{x}-\underline{y}\right\|_2 \le \varepsilon$$

Define $\underline{x}$ … $\mathbf{\Omega}$

and thus **D** …

The Two are Exactly Equivalent

$$\underset{\underline{x}}{\text{Min}} \left\|\mathbf{D}^{-1}\underline{x}\right\|_p^p \text{ s.t. } \left\|\underline{x}-\underline{y}\right\|_2 \le \varepsilon$$

Case 2: Redundant **D and Ω**

$$\underline{\alpha} = \mathbf{\Omega}\underline{x}$$
$$\Rightarrow \mathbf{\Omega}^T\underline{\alpha} = \mathbf{\Omega}^T\mathbf{\Omega}\underline{x}$$
$$\Rightarrow \left(\mathbf{\Omega}^T\mathbf{\Omega}\right)^{-1}\mathbf{\Omega}^T\underline{\alpha} = \mathbf{\Omega}^\dagger\underline{\alpha} = \underline{x}$$

Analysis

$$\underset{\underline{x}}{\text{Min}} \left\|\mathbf{\Omega}\underline{x}\right\|_p^p \text{ s.t. } \left\|\underline{x} - \underline{y}\right\|_2 \le \varepsilon$$

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

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

Case 2: Redundant **D and Ω**

Synthesis

$$\underset{\underline{\alpha}}{\text{Min}} \left\|\underline{\alpha}\right\|_p^p \text{ s.t. } \left\|\mathbf{D}\underline{\alpha}-\underline{y}\right\|_2 \le \varepsilon$$

Define $\mathbf{D}=\mathbf{\Omega}^{\dagger}$

$$\underset{\underline{\alpha}}{\text{Min}} \left\|\underline{\alpha}\right\|_p^p \text{ s.t. } \left\|\mathbf{\Omega}^{\dagger}\underline{\alpha}-\underline{y}\right\|_2 \le \varepsilon$$

Analysis

$$\underset{\underline{x}}{\text{Min}} \left\|\mathbf{\Omega}\underline{x}\right\|_p^p \text{ s.t. } \left\|\underline{x}-\underline{y}\right\|_2 \le \varepsilon$$

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

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

Case 2: Redundant **D and Ω**

Synthesis

$$\underset{\underline{\alpha}}{\text{Min}} \left\|\underline{\alpha}\right\|_p^p \text{ s.t. } \left\|\mathbf{D}\underline{\alpha}-\underline{y}\right\|_2 \le \varepsilon$$

Analysis

$$\underset{\underline{x}}{\text{Min}} \left\|\mathbf{\Omega}\underline{x}\right\|_p^p \text{ s.t. } \left\|\underline{x}-\underline{y}\right\|_2 \le \varepsilon$$

Defin… ne $\underline{\alpha} = \mathbf{\Omega}\underline{x}$

…us $\mathbf{\Omega}^{\dagger}\underline{\alpha} = \underline{x}$

Exact Equivalence again ?

$$\underset{\underline{\alpha}}{\text{Min}} \left\|\underline{\alpha}\right\|_p^p \text{ s.t. } \left\|\mathbf{\Omega}^{\dagger}\underline{\alpha}-\underline{y}\right\|_2 \le \varepsilon$$

Not Really !

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

$$\Rightarrow \mathbf{\Omega}^T\underline{\alpha} = \mathbf{\Omega}^T\mathbf{\Omega}\underline{x}$$

$$\Rightarrow \left(\mathbf{\Omega}^T\mathbf{\Omega}\right)^{-1}\mathbf{\Omega}^T\underline{\alpha} = \mathbf{\Omega}^{\dagger}\underline{\alpha} = \underline{x}$$

We should require

$$\mathbf{\Omega}\underline{x} = \underline{\alpha} = \mathbf{\Omega}\mathbf{\Omega}^{\dagger}\underline{\alpha}$$

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

$$\underset{\underline{\alpha}}{\text{Min}} \left\|\underline{\alpha}\right\|_p^p \text{ s.t. } \left\|\mathbf{D}\underline{\alpha} - \underline{y}\right\|_2 \le \varepsilon \quad \& \quad \underline{\alpha} = \mathbf{\Omega}\mathbf{\Omega}^{\dagger}\underline{\alpha}$$

which means that analysis $\neq$ 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

1. S. Nam, M.E. Davies, M. Elad, and R. Gribonval, "Co-sparse Analysis Modeling - Uniqueness and Algorithms" , ICASSP, May, 2011.
2. S. Nam, M.E. Davies, M. Elad, and R. Gribonval, "The Co-sparse Analysis Model and Algorithms" , to appear in ACHA.

The Analysis Model – Basics

- The analysis representation $\underline{z}$ is expected to be sparse

$$\left\|\boldsymbol{\Omega}\underline{x}\right\|_0 = \left\|\underline{z}\right\|_0 = p - \ell$$

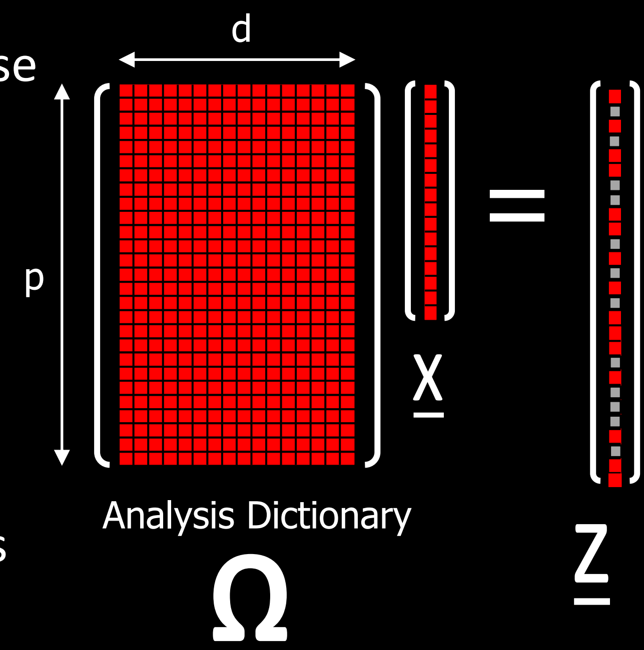
- Co-sparsity: ℓ - the number of zeros in $\underline{z}$.

- Co-Support: Λ - the rows that are orthogonal to $\underline{x}$

$$\boldsymbol{\Omega}_{\Lambda}\underline{x} = \underline{0}$$

- If $\boldsymbol{\Omega}$ is in general position*, then $0 \le \ell < d$ 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, $\underline{z}$, 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)].

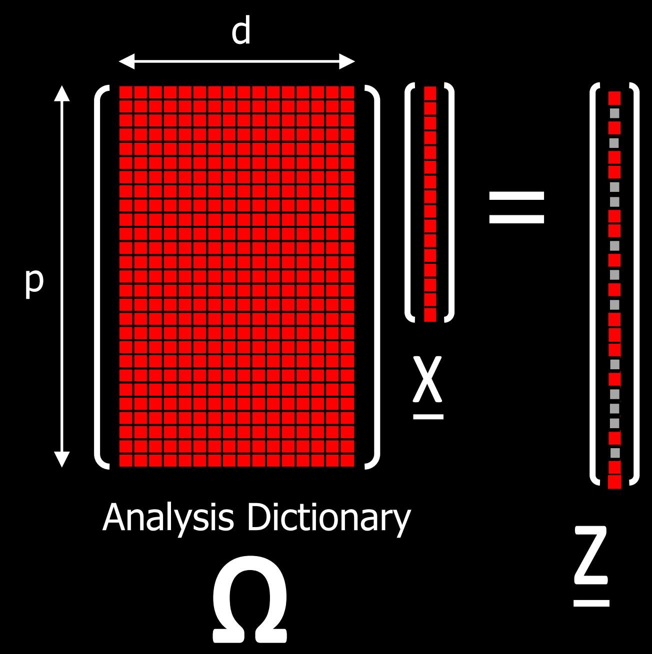


* $\text{spark}\left\{\boldsymbol{\Omega}^T\right\} = d+1$

The Analysis Model – Bayesian View

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

	Synthesis Signals	Analysis Signals
Support:	Choose the support T ($\|T\|$=k) at random	Choose the co-support Λ ($\|\Lambda\|=\ell$) at random
Coef. :	Choose $\underline{\alpha}_T$ at random	Choose a random vector $\underline{v}$
Generate:	Synthesize by: $\mathbf{D}_T\underline{\alpha}_T=\underline{x}$	Orhto $\underline{v}$ w.r.t. $\mathbf{\Omega}_\Lambda$: $\underline{x}=\left[\mathbf{I}-\mathbf{\Omega}_\Lambda^{\dagger}\mathbf{\Omega}_\Lambda\right]\underline{v}$

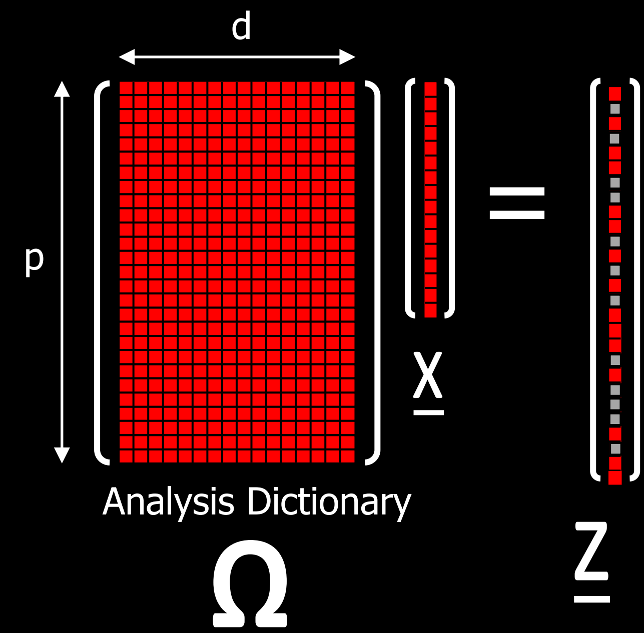


- Bottom line: an analysis signal $\underline{x}$ satisfies: $\exists\Lambda \;\Big|\; |\Lambda|=\ell$ s.t. $\mathbf{\Omega}_\Lambda\underline{x}=\underline{0}$

The Analysis Model – UoS

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

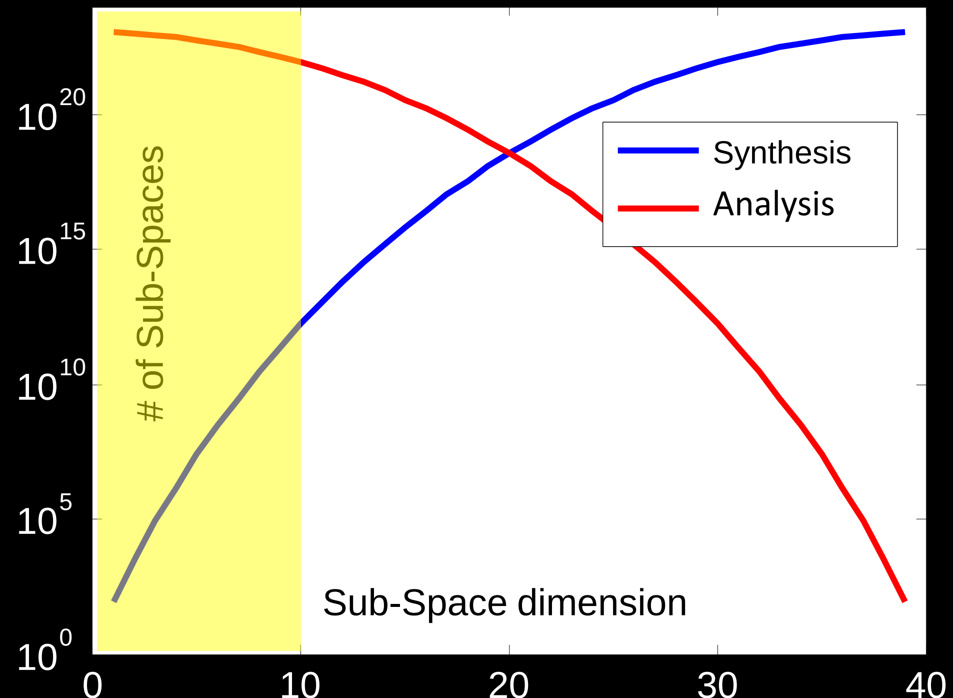
	Synthesis Signals	Analysis Signals
What is the Subspace Dimension:	k	d-ℓ
How Many Subspaces:	$\binom{n}{k}$	$\binom{p}{\ell}$
Who are those Subspaces:	$\text{span}\{\mathbf{D}_T\}$	$\text{span}^{\perp}\{\mathbf{\Omega}_{\Lambda}\}$



- 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: ℓ=d-1 leads to $\binom{2d}{d-1}$>>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,

$$\underline{y} = \underline{x} + \underline{v}, \quad \exists|\Lambda| = \ell \text{ s.t. } \mathbf{\Omega}_\Lambda \underline{x} = \underline{0}, \quad \underline{v} \sim \mathbf{N}\left\{\underline{0}, \sigma^2 \mathbf{I}\right\}$$

  recover the clean signal $\underline{x}$ – This is a denoising task.

- This goal can be posed as: $$\hat{\underline{x}} = \underset{\underline{\alpha}}{\text{ArgMin}} \left\|\underline{y} - \underline{x}\right\|_2^2 \text{ s.t. } \left\|\mathbf{\Omega}\underline{x}\right\|_0 = p - \ell$$

- 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 $\mathbf{\Omega}$. 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 Λ for approximating the solution of $\hat{\underline{x}} = \underset{\underline{\alpha}}{\text{ArgMin}} \left\| \underline{y} - \underline{x} \right\|_2^2 \ \text{s.t.} \ \left\| \mathbf{\Omega} \underline{x} \right\|_0 = p - \ell$

$i = 0, \ \hat{\underline{x}}_0 = \underline{y} \ \ \Lambda_0 = \{ \ \}$

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

Yes

Output x_i

No

$i = i + 1, \ \ \Lambda_i = \Lambda_{i-1} \cup \underset{k}{\text{ArgMin}} \left| \underline{w}_k^T \hat{\underline{x}}_{i-1} \right|$

$\hat{\underline{x}}_i = \left[\mathbf{I} - \mathbf{\Omega}_{\Lambda_i}^{\dagger} \mathbf{\Omega}_{\Lambda_i} \right] \underline{y}$

The Analysis Model – Backward Greedy

Is there a similarity to a synthesis pursuit algorithm?

Synthesis OMP

$i=0,\ \underline{r}_0=\underline{y}\ \Lambda_0=\{\ \}$

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

Yes

Output $=\underline{y}-\underline{r}_i$

No

$i=i+1,\ \ \Lambda_i=\Lambda_{i-1}\cup \mathrm{Arg\,Max}\left|\underline{d}_k^T\underline{r}_{i-1}\right|$

$\underline{r}_i=\left[\mathbf{I}-\mathbf{D}_\Lambda\mathbf{D}_\Lambda^+\right]\underline{y}$

The Analysis Model – Backward Greedy

Is there a similarity to a synthesis pursuit algorithm?

Synthesis OMP

$i = 0$ … $= \underline{y} - \underline{r}_i$

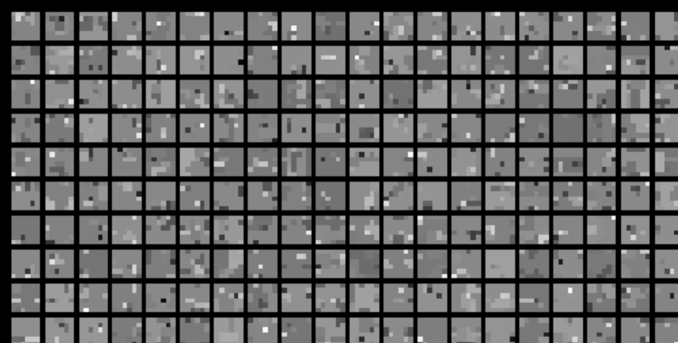
$i =$ …

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)]

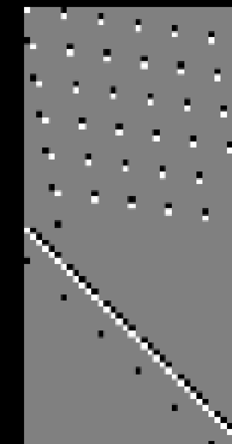
The Analysis Model – Low-Spark Ω

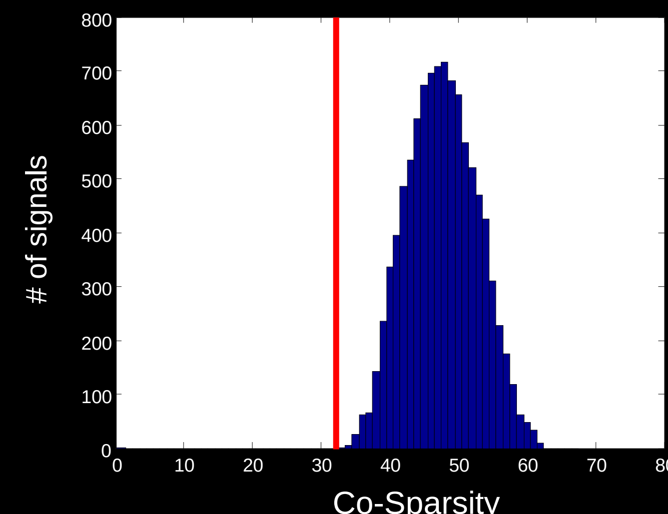
- What if spark(Ω^T)<<d ?
- For example: a TV-like operator for image-patches of size 6×6 pixels (Ω size is 72×36).
- Here are analysis-signals generated for co-sparsity (ℓ) of 32:



- 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.

$$\Omega = \begin{bmatrix} \text{Horizontal} \\ \text{Derivative} \\ \hdashline \text{Vertical} \\ \text{Derivative} \end{bmatrix} =$$

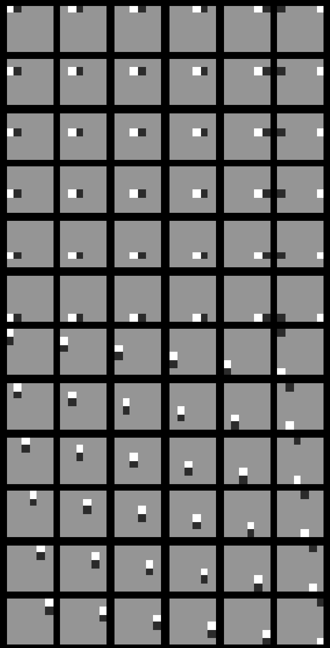




The Analysis Model – The Signature

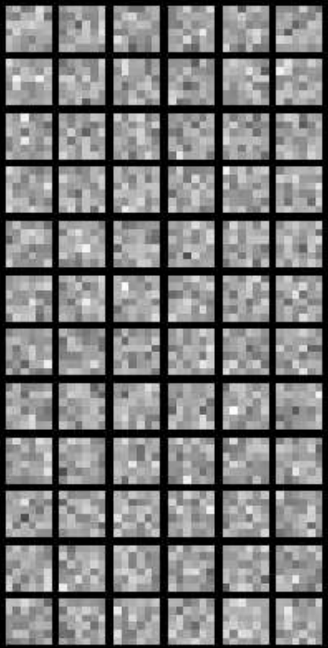
Consider two possible dictionaries:

Ω_{DIF}

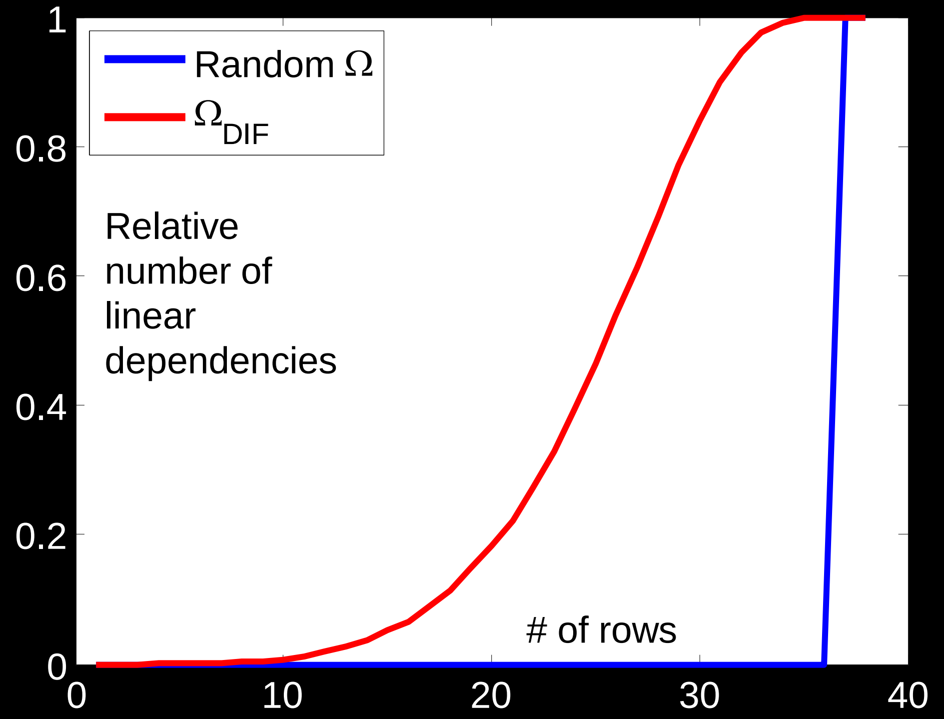


$$\mathrm{Spark}\left(\Omega^T\right)=4$$

Random Ω



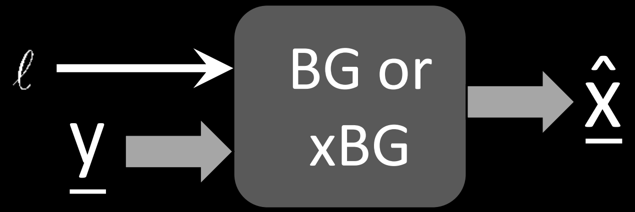
$$\mathrm{Spark}\left(\Omega^T\right)=37$$

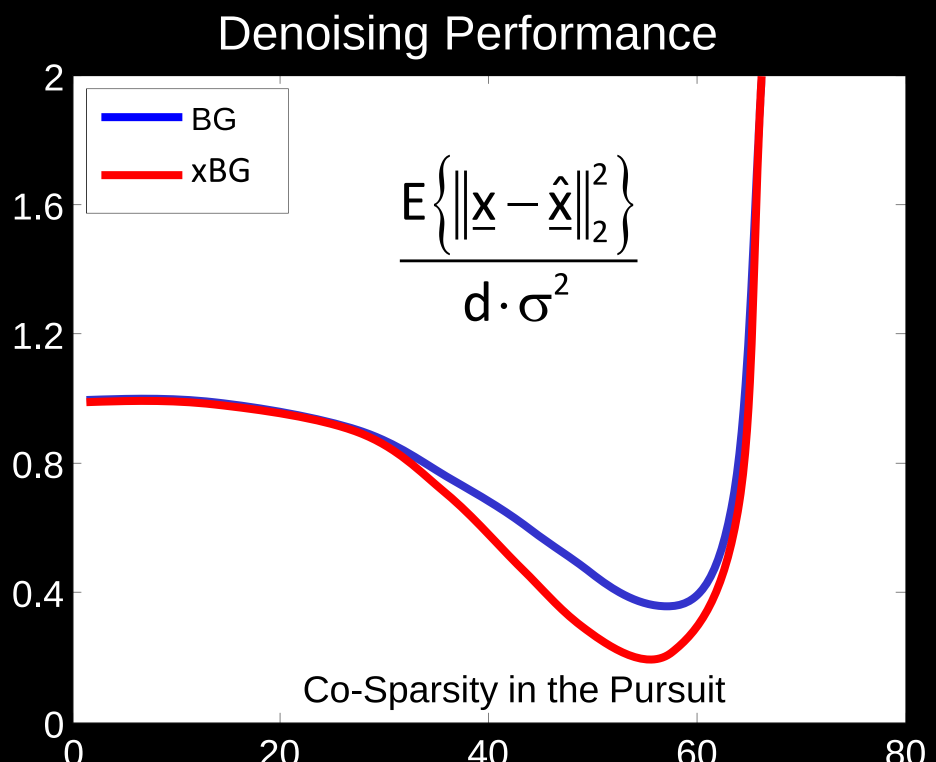


The Signature of a matrix is more informative than the Spark

The Analysis Model – Low-Spark Ω – Pursuit

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

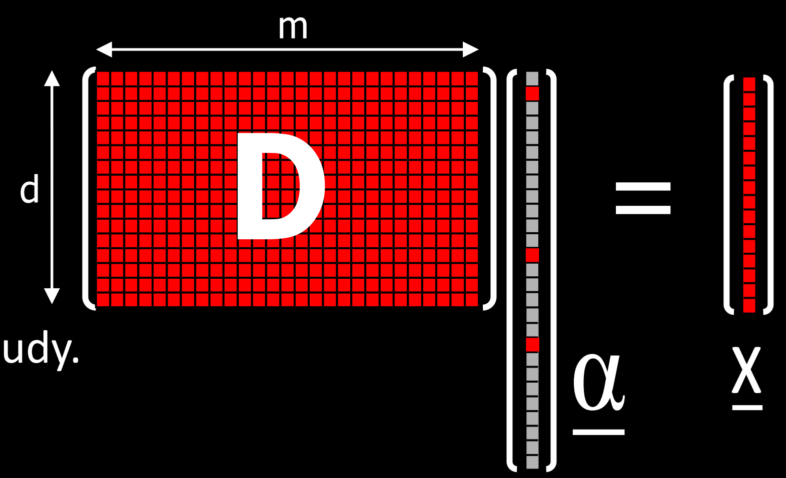


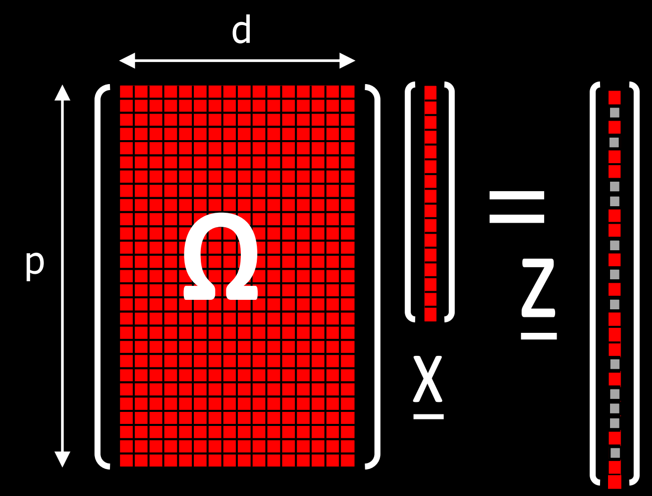


- We see an effective denoising, attenuating the noise by a factor ~0.2. This is achieved for an effective co-sparsity of ~55.

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 Ω ?
  - Great potential for modeling signals.





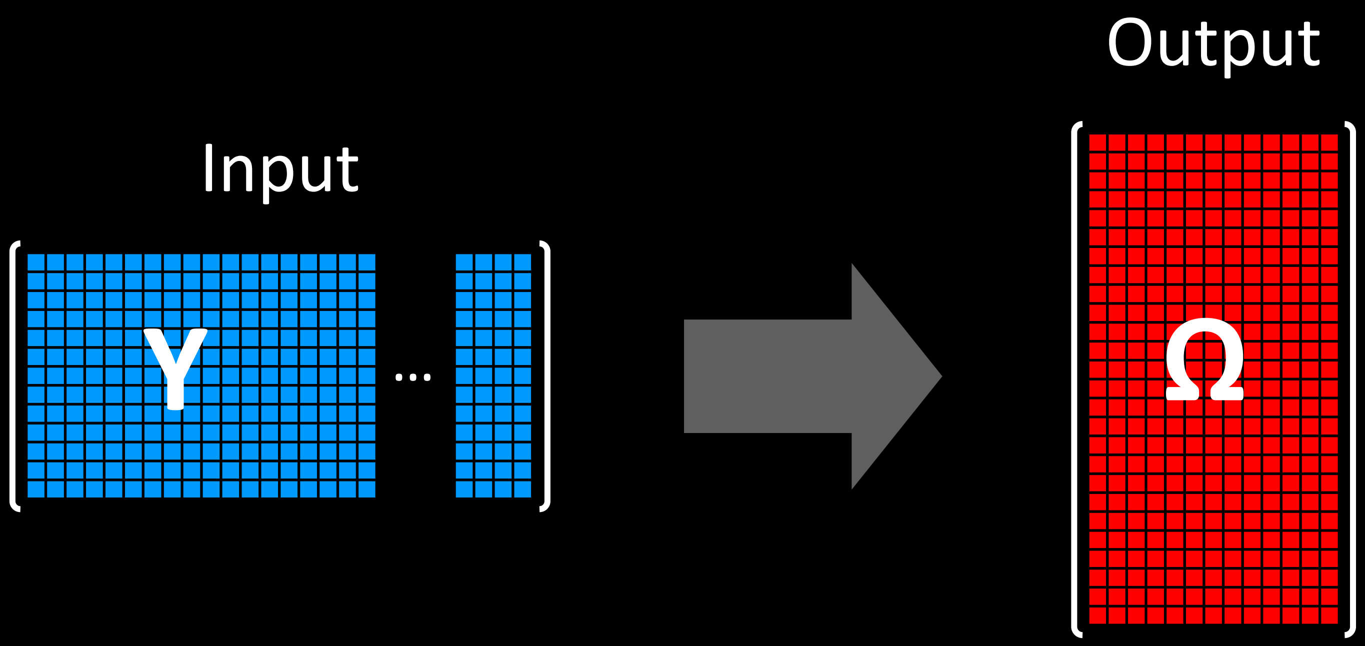
Part IV – Dictionaries

Analysis Dictionary-Learning by K-SVD-Like Algorithm

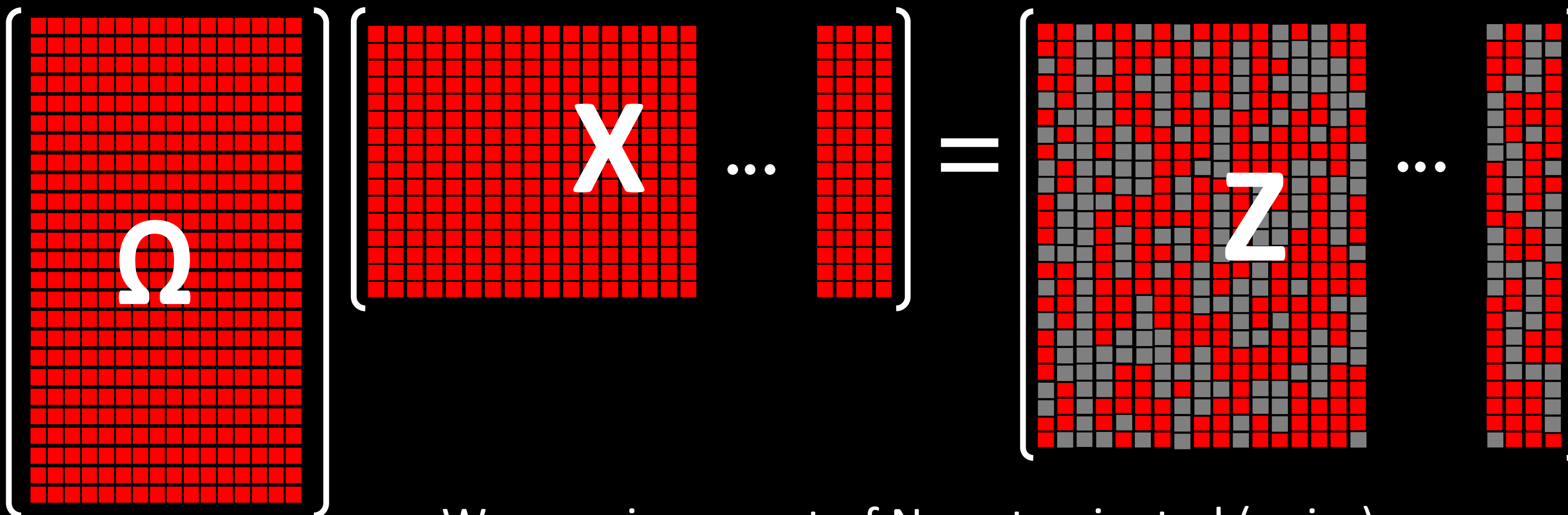
1. B. Ophir, M. Elad, N. Bertin and M.D. Plumbley, "Sequential Minimal Eigenvalues - An Approach to Analysis Dictionary Learning", EUSIPCO, August 2011.
2. R. Rubinstein T. Peleg, and M. Elad, "Analysis K-SVD: A Dictionary-Learning Algorithm for the Analysis Sparse Model", submitted to IEEE-TSP.

Analysis Dictionary Learning – The Goal

Goal: given a set of signals, find the analysis dictionary $\mathbf{\Omega}$ 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, $\mathbf{\Omega}$

$$\left\{\underline{y}_j = \underline{x}_j + \underline{v}_j,\quad \exists\left|\Lambda_j\right| = \ell \ \text{ s.t. } \ \mathbf{\Omega}_{\Lambda_j}\underline{x}_j = \underline{0},\quad \underline{v} \sim \mathbf{N}\left\{\underline{0}, \sigma^2\mathbf{I}\right\}\right\}_{j=1}^{N}$$

Analysis Dictionary Learning – Goal

Synthesis

$$\underset{\mathbf{D},\mathbf{A}}{\text{Min}} \left\|\mathbf{DA}-\mathbf{Y}\right\|_F^2 \quad \text{s.t. } \forall j=1,2,\ldots,N \quad \left\|\underline{\alpha}_j\right\|_0 \leq k$$

Analysis

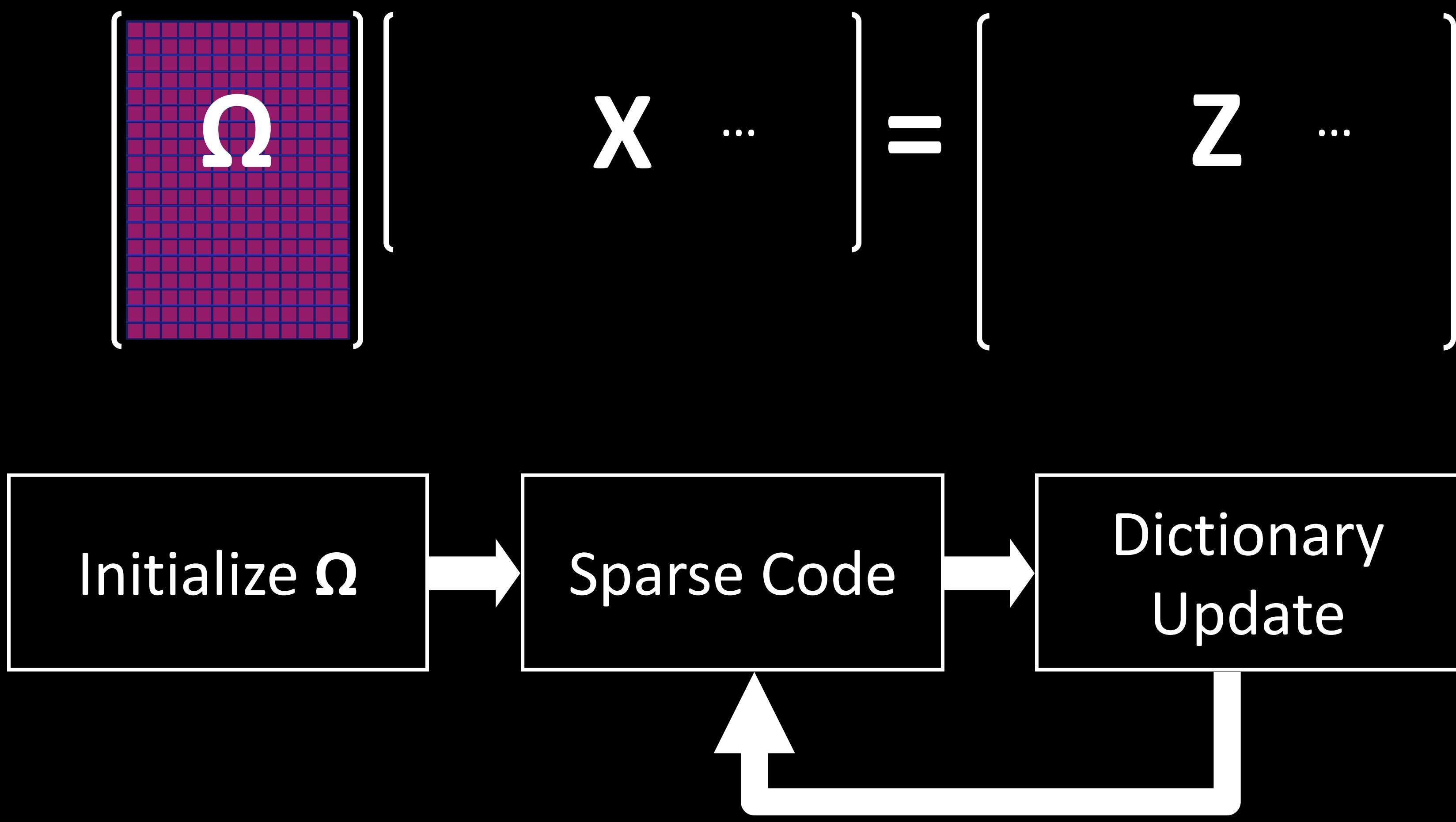
$$\underset{\boldsymbol{\Omega},\mathbf{X}}{\text{Min}} \left\|\mathbf{X}-\mathbf{Y}\right\|_F^2 \quad \text{s.t. } \forall j=1,2,\ldots,N \quad \left\|\boldsymbol{\Omega}\underline{x}_j\right\|_0 \leq p-\ell$$

Noisy Examples

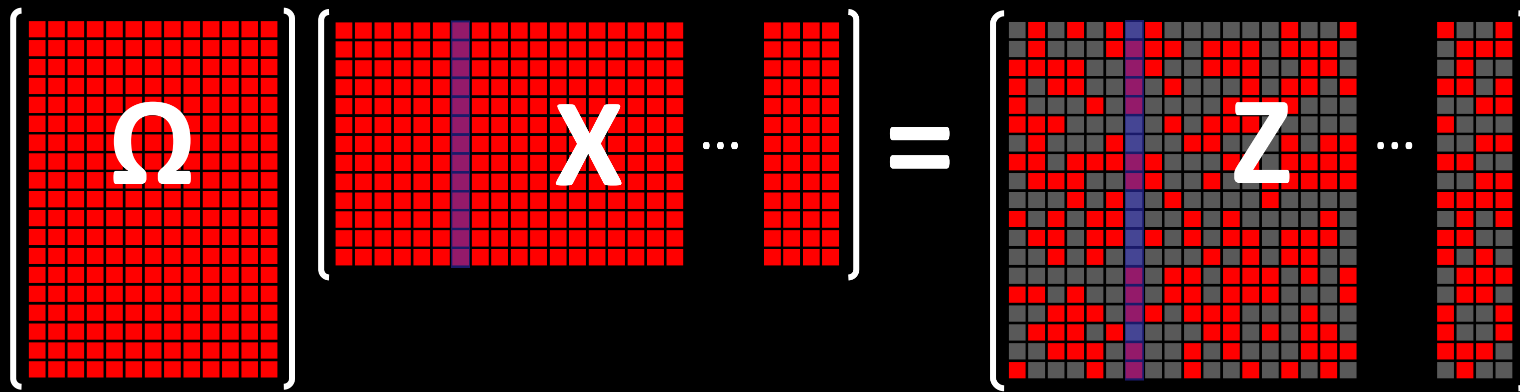
Denoised Signals are L_0 Sparse

We shall adopt a similar approach to the K-SVD for approximating the minimization of the analysis goal

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



Analysis K-SVD – Sparse-Coding Stage



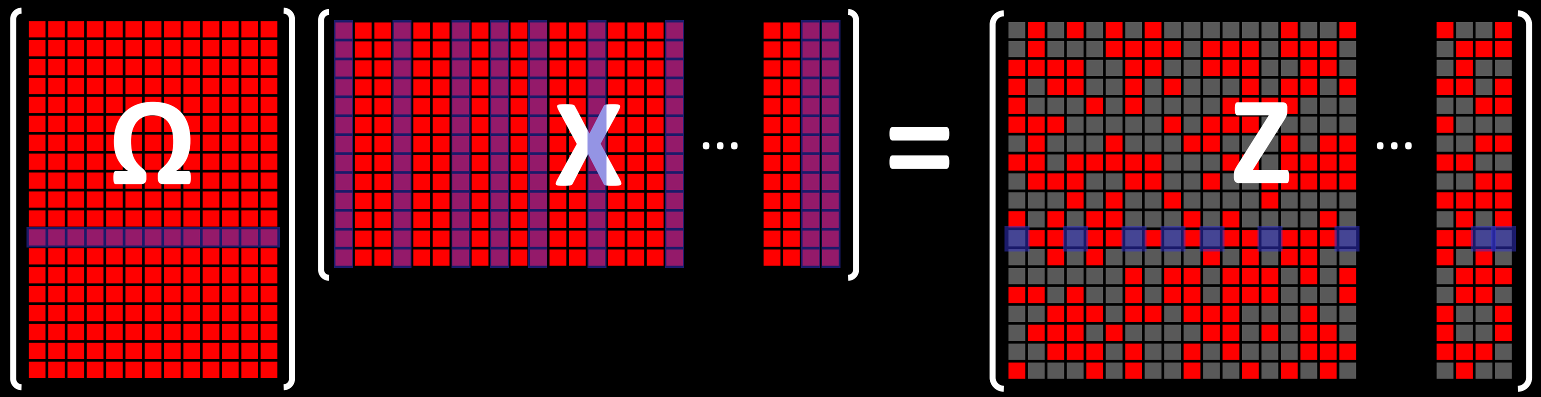
$$\underset{\mathbf{\Omega},\mathbf{X}}{\text{Min}} \left\|\mathbf{X}-\mathbf{Y}\right\|_F^2 \ \text{ s.t. } \forall j=1,2,\ldots,N \ \left\|\mathbf{\Omega}\underline{x}_j\right\|_0 \le p-\ell$$

Assuming that $\mathbf{\Omega}$ is fixed, we aim at updating $\underline{X}$

$$\left\{ \hat{\underline{x}}_j = \underset{\underline{x}}{\text{ArgMin}} \left\|\underline{x}-\underline{y}_j\right\|_2^2 \ \text{ s.t. } \left\|\mathbf{\Omega}\underline{x}\right\|_0 \le p-\ell \right\}_{j=1}^{N}$$

These are N separate analysis-pursuit problems. We suggest to use the BG or the xBG algorithms.

Analysis K-SVD – Dictionary Update Stage



$$\underset{\mathbf{\Omega},\mathbf{X}}{\text{Min}} \left\|\mathbf{X}-\mathbf{Y}\right\|_F^2 \ \text{ s.t. } \forall j=1,2,\ldots,N \ \left\|\mathbf{\Omega}\underline{x}_j\right\|_0 \le p-\ell$$

- Only signals orthogonal to the atom should get to vote for its new value.
- The known supports should be preserved.

Analysis Dictionary – Dic. Update (2)

After the sparse-coding, Λ_j are known. We now aim at updating a row (e.g. $\underline{w}_k^T$) from $\mathbf{\Omega}$

We use only the signals S_k that are found orthogonal to $\underline{w}_k$

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

$$\min_{\underline{w}_k, \mathbf{X}_k} \left\| \mathbf{X}_k - \mathbf{Y}_k \right\|_2^2 \quad \text{s.t.} \quad \begin{cases} \forall j \in S_k \quad \mathbf{\Omega}_j \underline{x}_j = \underline{0} \\ \underline{w}_k^T \mathbf{X}_k = \underline{0} \\ \left\| \underline{w}_k \right\|_2 = 1 \end{cases}$$

Avoid trivial solution

Each of the chosen examples should be orthogonal to the new row $\underline{w}_k$

Analysis Dictionary – Dic. Update (3)

$$\underset{\underline{w}_k,\mathbf{X}_k}{\text{Min}} \ \left\|\mathbf{X}_k - \mathbf{Y}_k\right\|_2^2 \ \text{ s.t. } \begin{Bmatrix} \forall j \in S_k \quad \mathbf{\Omega}_j \underline{x}_j = \underline{0} \\ \underline{w}_k^T \mathbf{X}_k = \underline{0} \\ \left\|\underline{w}_k\right\|_2 = 1 \end{Bmatrix}$$

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

$$\underset{\underline{w}_k,\mathbf{X}_k}{\text{Min}} \ \left\|\mathbf{X}_k - \left[\cdots\mathbf{\Omega}_j\right]\mathbf{Y}_k\right\|_2^2 \ \text{ s.t. } \begin{Bmatrix} \underline{w}_k^T \mathbf{X}_k = \underline{0} \\ \left\|\underline{w}_k\right\|_2 = 1 \end{Bmatrix}$$

WRONG!

Analysis Dictionary – Dic. Update (4)

A better approximation for our original problem is

$$\underset{\underline{w}_k,\mathbf{X}_k}{\text{Min}} \left\|\mathbf{X}_k - \mathbf{Y}_k\right\|_2^2 \quad \text{s.t.} \quad \left\{ \begin{array}{l} \forall j \in S_k \quad \cancel{\boldsymbol{\Omega}_j \underline{x}_j = \underline{0}} \\ \underline{w}_k^T \mathbf{X}_k = \underline{0} \\ \left\|\underline{w}_k\right\|_2 = 1 \end{array} \right\}$$

which is equivalent to

$$\underset{\underline{w}_k,}{\text{Min}} \left\|\underline{w}_k^T \mathbf{Y}_k\right\|_2^2 \quad \text{s.t.} \quad \left\|\underline{w}_k\right\|_2 = 1$$

The obtained problem is a simple eigenvaue approximation problem, easily given by SVD

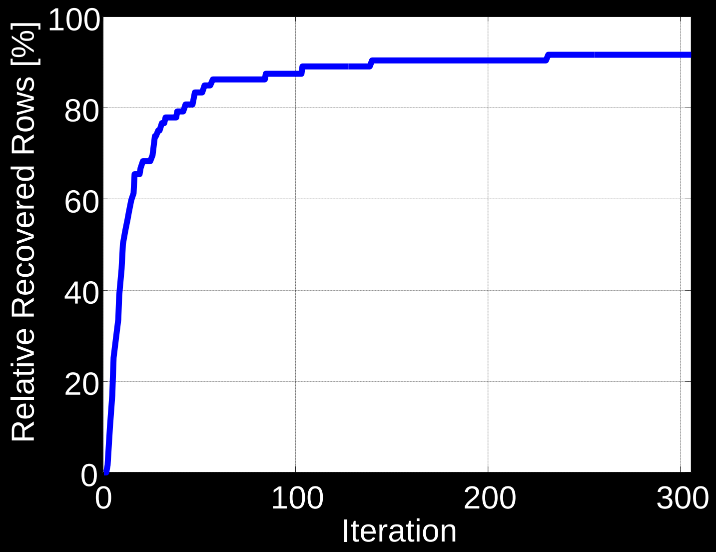
Part V – Results

For Dictionary-Learning and Image Denoising

Analysis Dictionary Learning – Results (1)

Synthetic experiment #1: TV-Like Ω

- We generate 30,000 TV-like signals of the same kind described before (Ω: 72×36, ℓ=32)
- We apply 300 iterations of the Analysis K-SVD with BG (fixed ℓ), 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: $1-\left|\underline{w}^T\underline{\hat{w}}\right|<0.01$

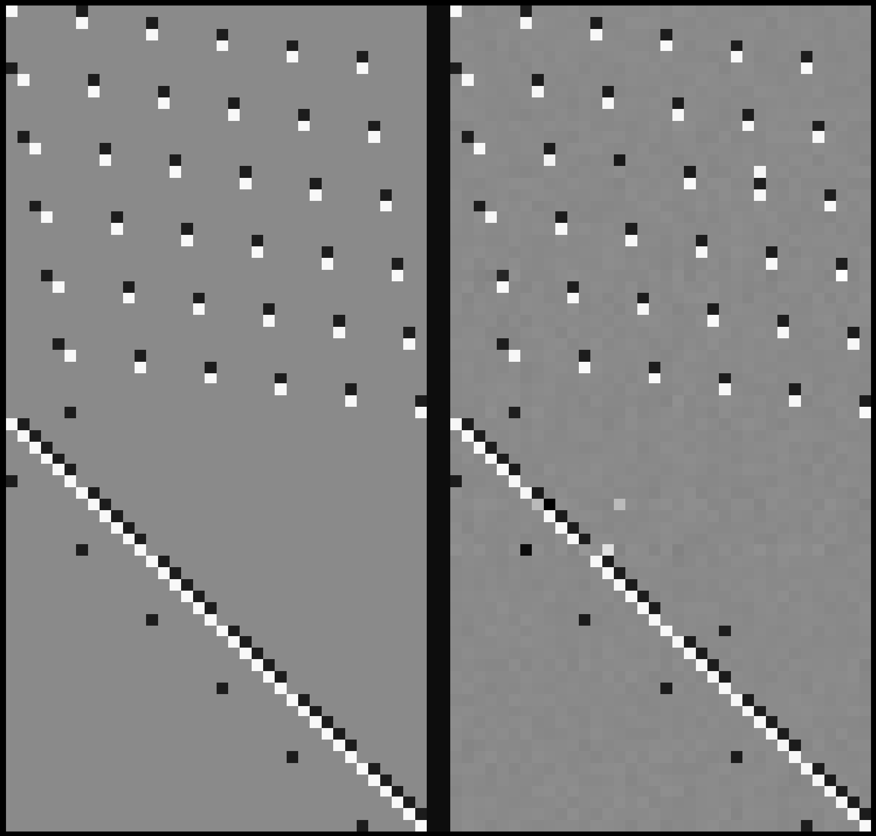


Even though we have not identified Ω completely (~92% this time), we got an alternative feasible analysis dictionary with the same number of zeros per example, and a residual error within the noise level.

Analysis Dictionary Learning – Results (1)

Synthetic experiment #1: TV-Like Ω

Original Analysis Dictionary

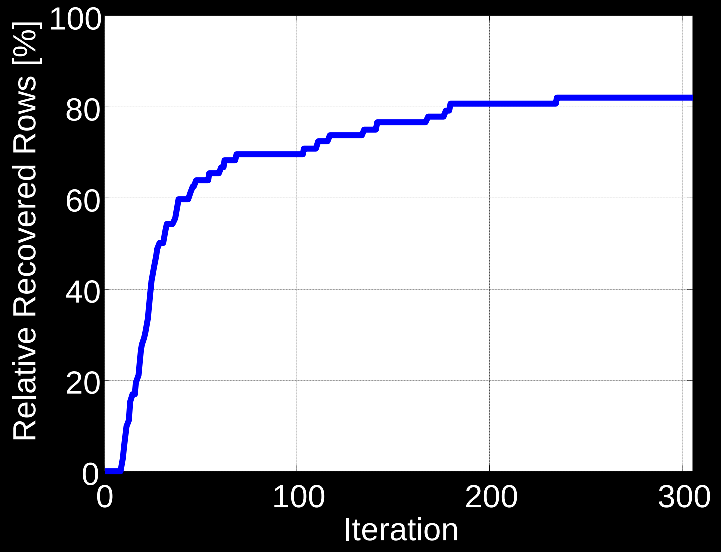


Learned Analysis Dictionary

Analysis Dictionary Learning – Results (2)

Synthetic experiment #2: Random $\boldsymbol{\Omega}$

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



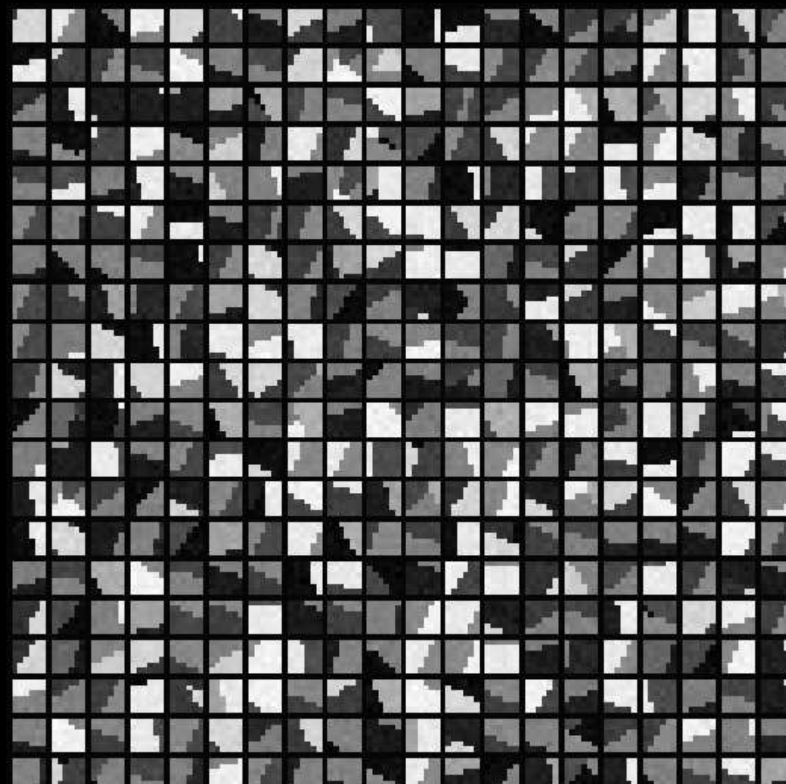
As in the previous example, even though we have not identified $\boldsymbol{\Omega}$ completely (~80% this time), we got an alternative feasible analysis dictionary with the same number of zeros per example, and a residual error within the noise level.

Analysis Dictionary Learning – Results (3)

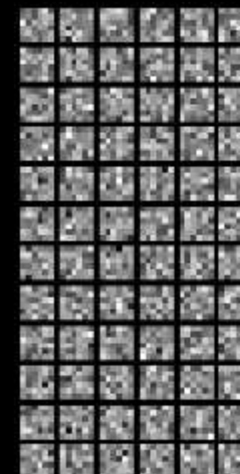
Experiment #3: Piece-Wise Constant Image

- We take 10,000 patches (+noise $\sigma=5$) to train on
- Here is what we got:

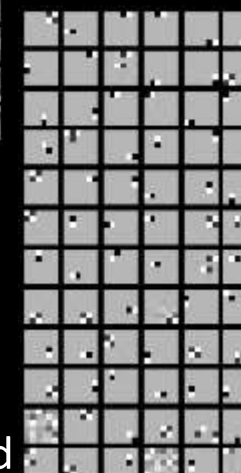
Original Image



Patches used for training



Initial Ω



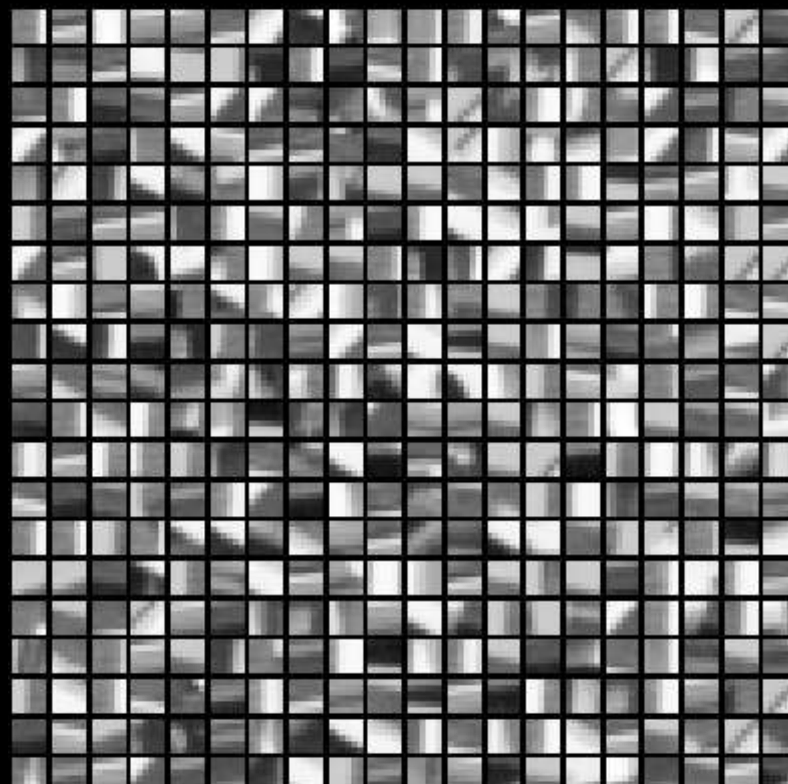
Trained (100 iterations) Ω

Analysis Dictionary Learning – Results (4)

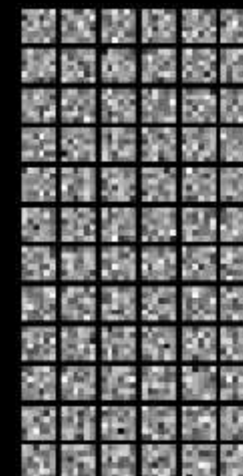
Experiment #4: The Image "House"

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

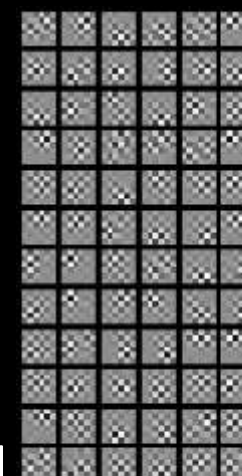
Original Image



Patches used for training



Initial Ω



Trained (100 iterations) Ω

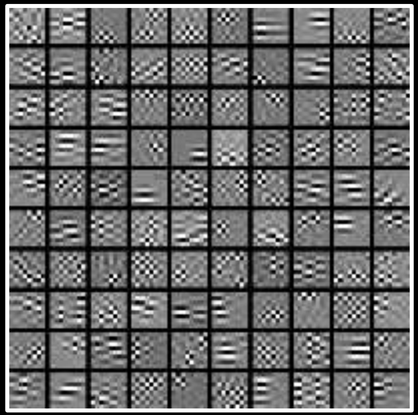
Analysis Dictionary Learning – Results (5)

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:

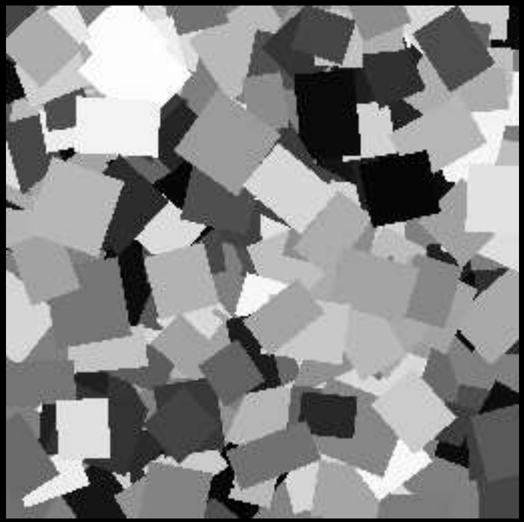
Localized and oriented atoms

Original Images

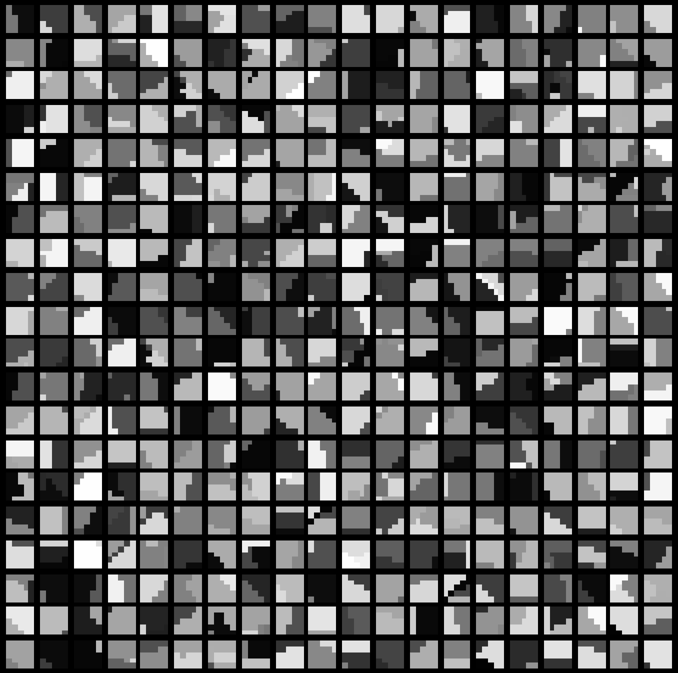


Trained Ω
(100 iterations)

Back to Image Denoising – (1)



256×256



Non-flat patch examples

Back to Image Denoising – (2)

- **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 $\varepsilon^2=1.3d\sigma^2$.

- **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.

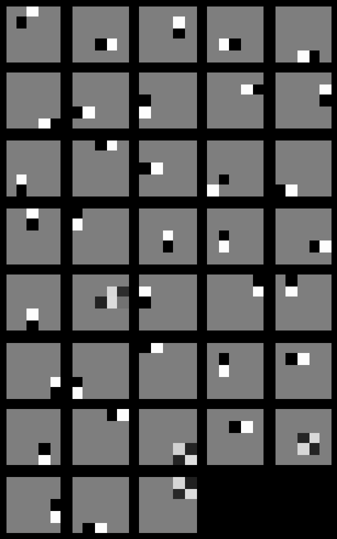
❑ **Patch Denoising** – error-based OBG with $\varepsilon^2=1.3d\sigma^2$.

❑ **Image Reconstruction** – Average overlapping patch recoveries.

Back to Image Denoising – (4)

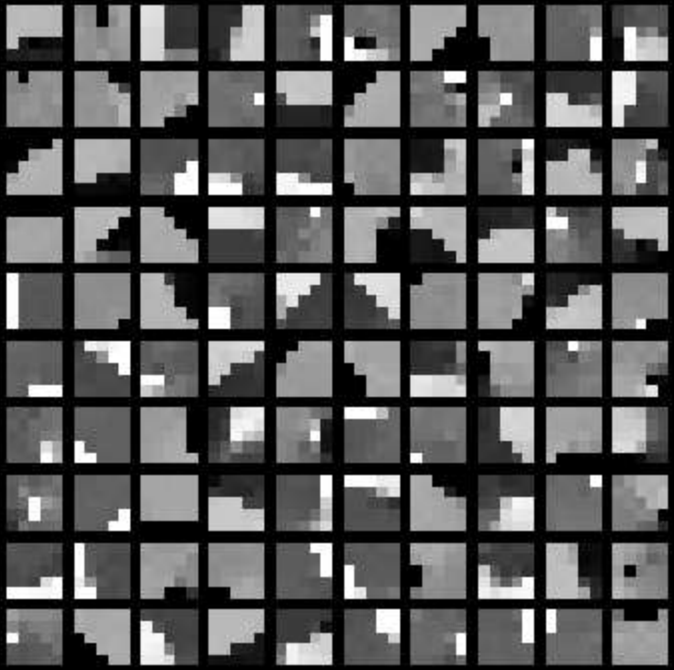
Learned dictionaries for σ=5

Analysis Dictionary



38 atoms

Synthesis Dictionary



100 atoms

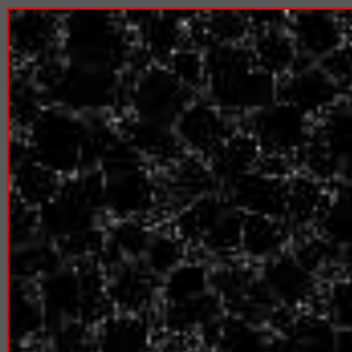
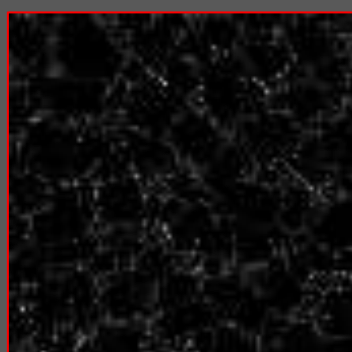
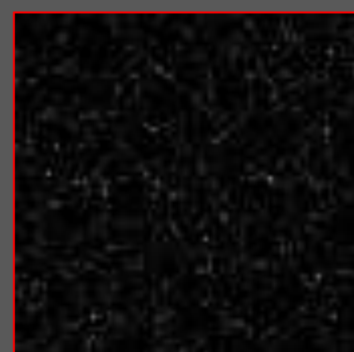
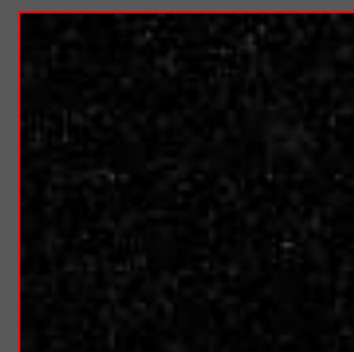
Back to Image Denoising – (5)

	BM3D	Synthesis K-SVD		Analysis K-SVD	
Average subspace dimension	n/a	2.42	2.03	**1.75**	**1.74**
		1.79	1.69	**1.51**	**1.43**

Cell Legend:

σ=5	σ=10
σ=15	σ=20

Back to Image Denoising – (6)

	BM3D	Synthesis K-SVD	Analysis K-SVD
σ=5 Scaled to [0,20]			
σ=10 Scaled to [0,40]			

Part VI – We Are Done

Summary and Conclusions

Today …

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

Is there any other way?

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

So, what to do?

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

What next?

- 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

Questions?

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