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

MS67: Sparse and Redundant Representations for Image Reconstruction and Geometry Extraction
Sunday May 20 4:30PM – 6:30PM

*Joint work with


Ron Rubinstein


Tomer Peleg

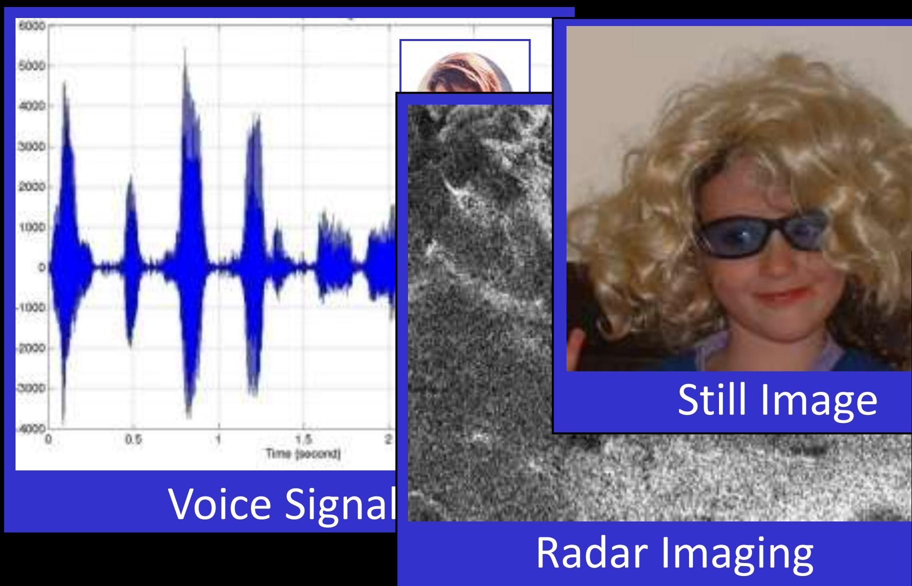

Remi Gribonval

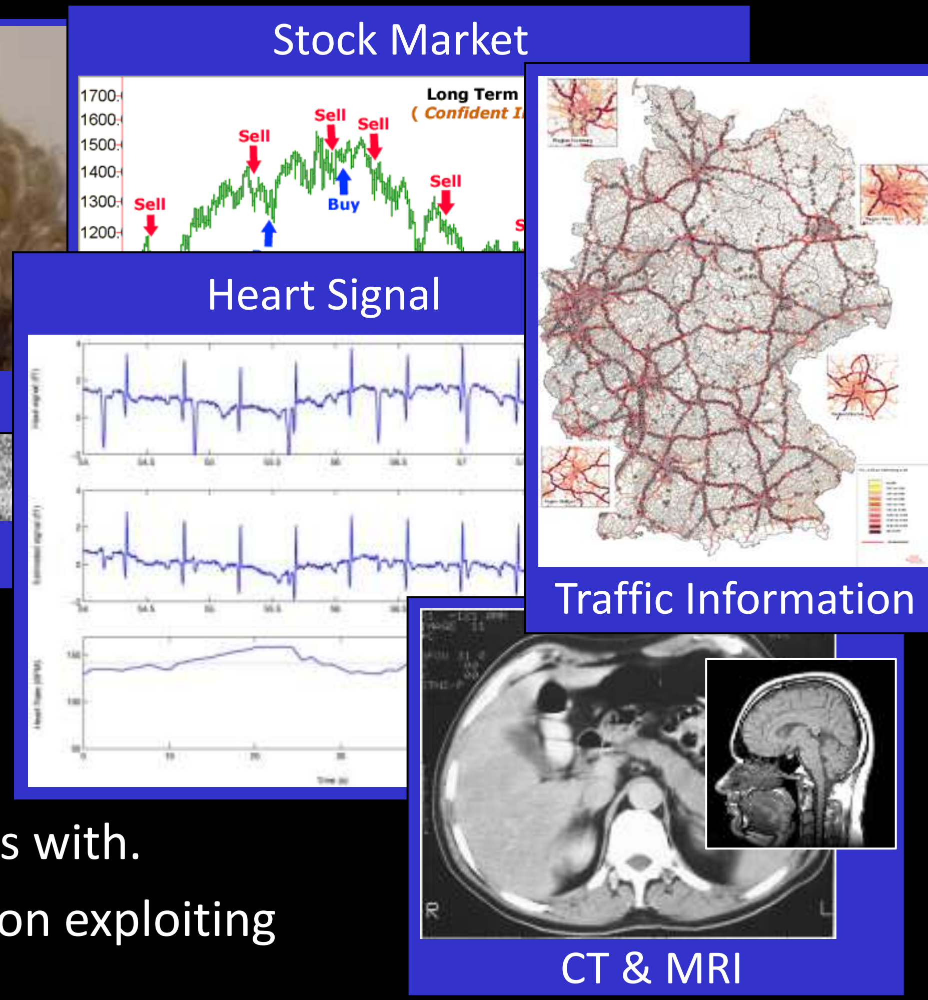
and

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

Technion
Israel Institute of Technology

Informative Data → Inner Structure

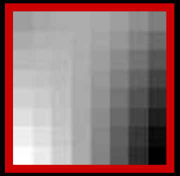




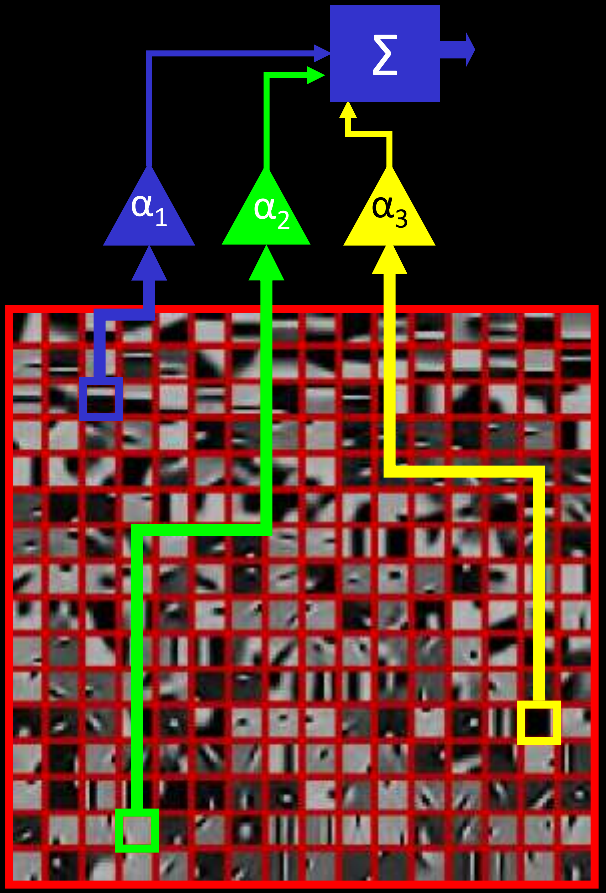
- It does not matter what is the data you are working on – if it is carrying information, it has an inner structure.
- This structure = rules the data complies with.
- Signal/image processing heavily relies on exploiting these "rules" by adopting **models**.

Sparse & redundant Repres. Modeling

- 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 (at least) 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 know better !! The two are VERY DIFFERENT

However …

Sparsity and Redundancy can be Practiced in (at least) two different ways

Synthesis
as presented above

Analysis

The attention to s… …ity-based …dels ha… …ly to… …ion le… …the an… …sis almost untouched.

This Talk's Message:

The co-sparse analysis model is a very appealing alternative to the synthesis model, it has a great potential for signal modeling.

Part I - Background

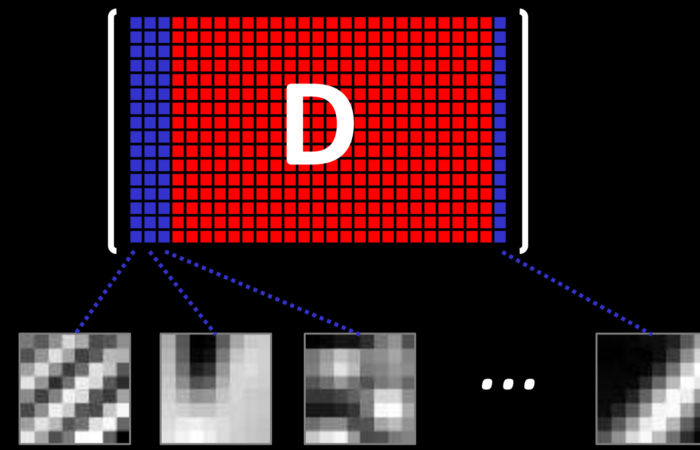
Recalling the Synthesis Sparse Model

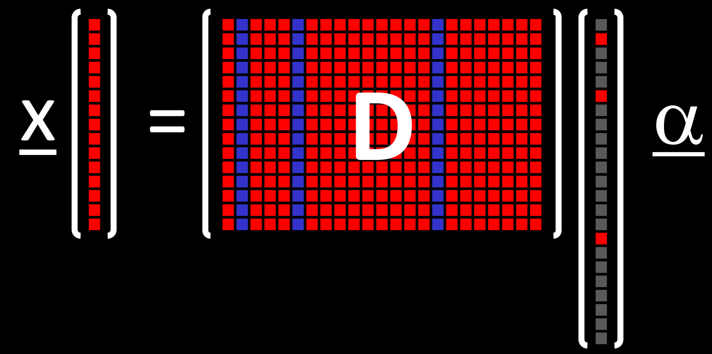
The Sparsity-Based Synthesis Model

- 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:

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

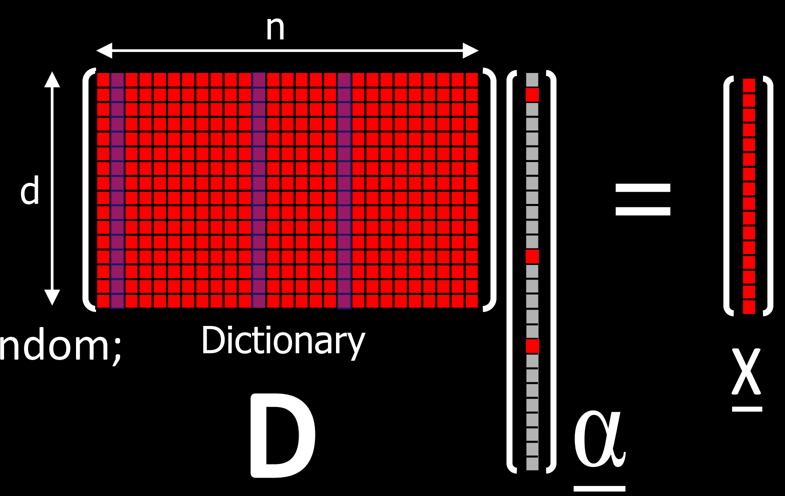
- We seek a sparsity of $\underline{\alpha}$, meaning that it is assumed to contain mostly zeros.
- This model is typically referred to as the synthesis sparse and redundant representation model for signals.
- This model became very popular and very successful in the past decade.





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):

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

- 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.

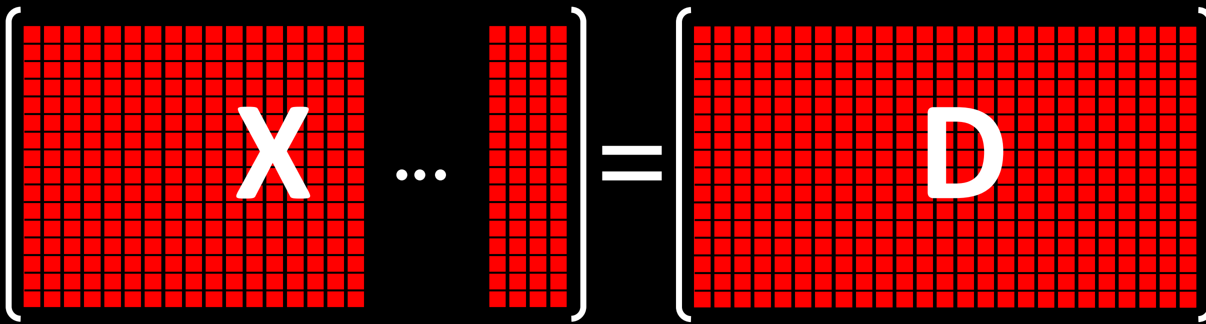
- This can be posed as: $\hat{\underline{\alpha}} = \underset{\underline{\alpha}}{\mathrm{ArgMin}} \left\| \underline{y} - \mathbf{D}\underline{\alpha} \right\|_2^2 \text{ s.t. } \left\| \underline{\alpha} \right\|_0 = k \;\Rightarrow\; \hat{\underline{x}} = \mathbf{D}\hat{\underline{\alpha}}$

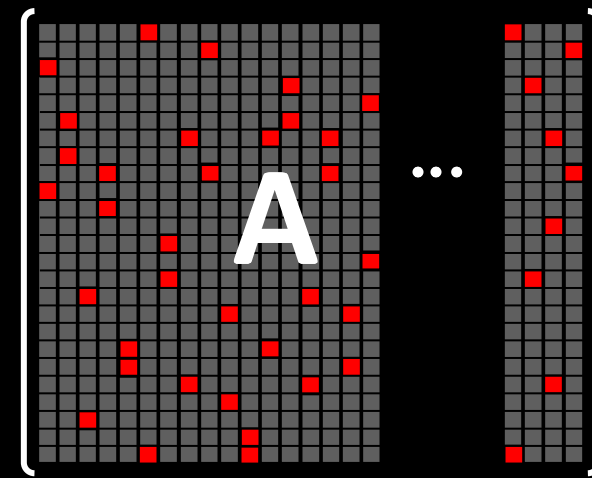
- While this is a (NP-) hard problem, its approximated solution can be obtained by
  - Use L_1 instead of L_0 (Basis-Pursuit)
  - Greedy methods (MP, OMP, LS-OMP)
  - Hybrid methods (IHT, SP, CoSaMP).

  } 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





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

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

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)
…

Part II - Analysis
Turning to 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" , Submitted to ACHA, June 2011.

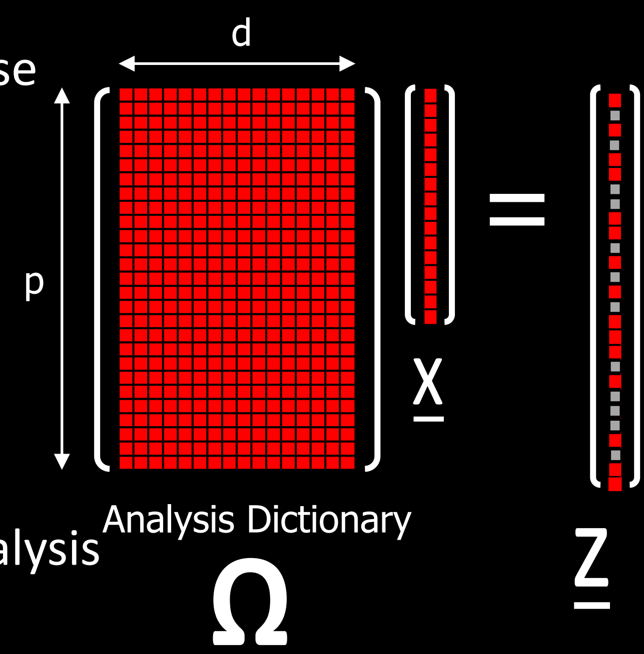
The Analysis Model – Basics

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

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

- Co-sparsity: ℓ - the number of zeros in $\underline{z}$.
- Co-Support: Λ - the rows that are orthogonal to $\underline{x}$

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



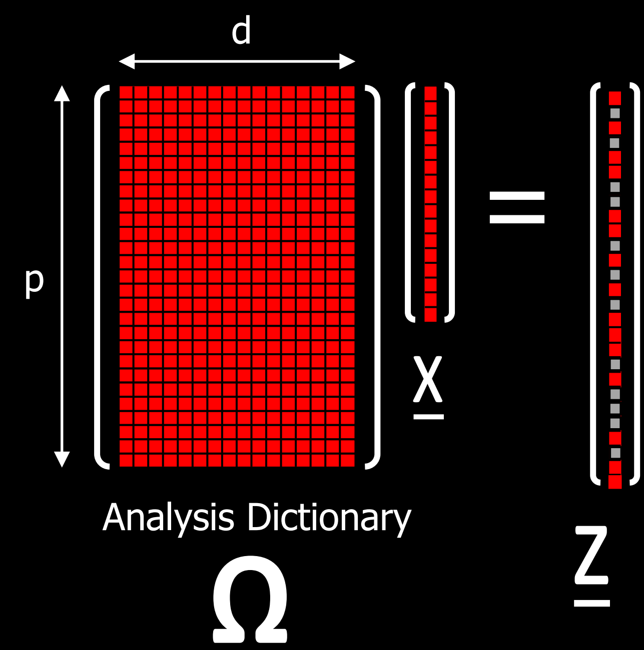
- This model puts an emphasis on the zeros in the analysis representation, $\underline{z}$, rather then the non-zeros, in characterizing the signal. This is much like the way zero-crossings of wavelets are used to define a signal [Mallat (`91)].
- If Ω is in general position*, then $0 \le \ell < d$ and thus we cannot expect to get a truly sparse analysis representation – Is this a problem? Not necessarily!

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

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. $\mathbf{\Omega}_\Lambda$: $\underline{x}=\left[\mathbf{I}-\mathbf{\Omega}_\Lambda^\dagger\mathbf{\Omega}_\Lambda\right]\underline{v}$

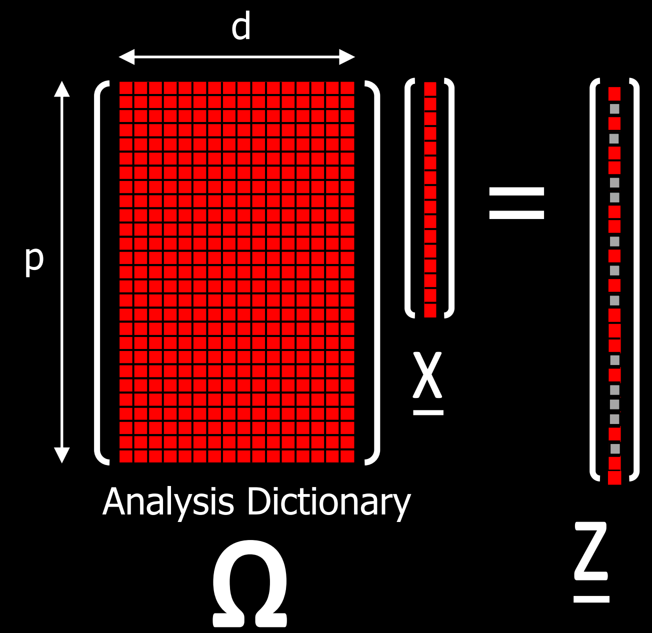


- Bottom line: an analysis signal $\underline{x}$ satisfies: $\exists\Lambda \,\Big|\, |\Lambda|=\ell$ s.t. $\mathbf{\Omega}_\Lambda\underline{x}=\underline{0}$

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:	$\text{span}\{\mathbf{D}_T\}$	$\text{span}^{\perp}\{\mathbf{\Omega}_{\Lambda}\}$

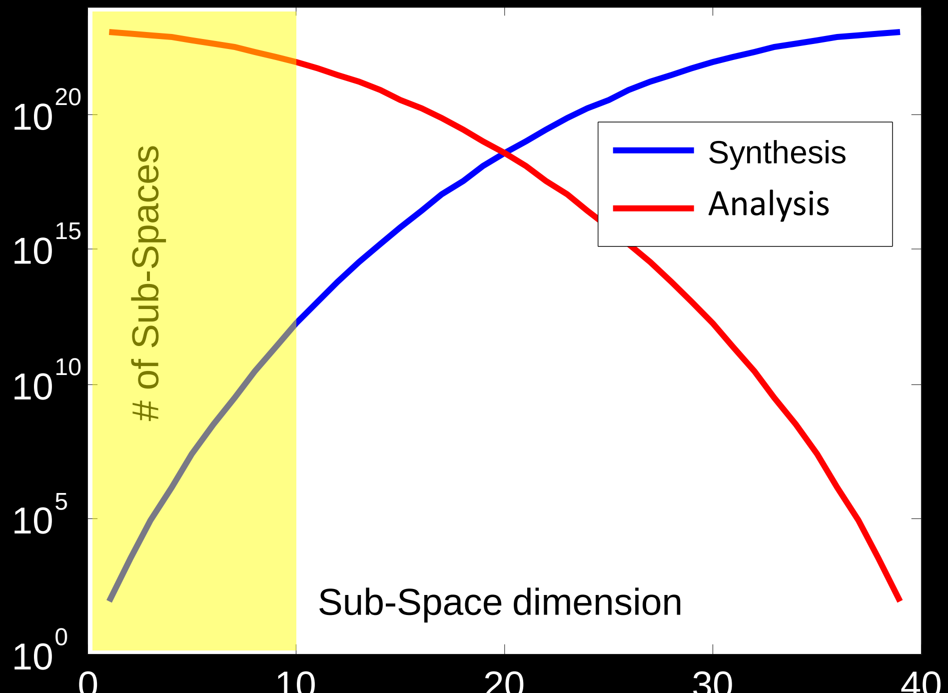


- The analysis and the synthesis models offer both a UoS construction, but these are very different from each other in general.

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:

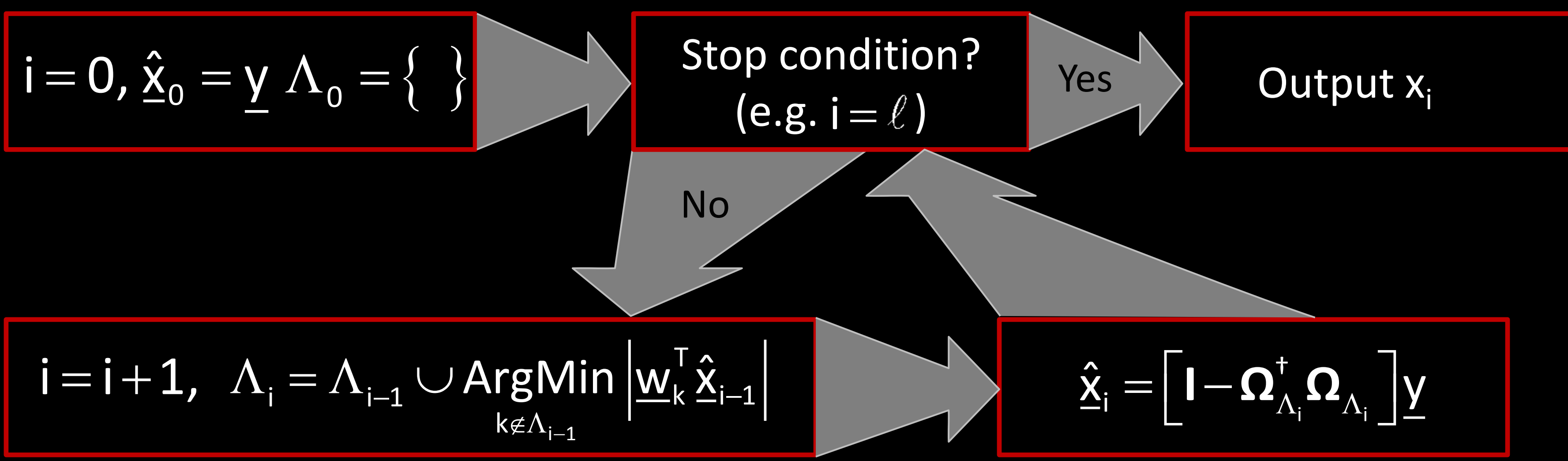
$$\hat{\underline{x}} = \underset{\underline{\alpha}}{\text{ArgMin}} \left\| \underline{y} - \underline{x} \right\|_2^2 \text{ s.t. } \left\| \mathbf{\Omega} \underline{x} \right\|_0 = p - \ell$$

- 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)], [Peleg & Elad ('12)].

The Analysis Model – Backward Greedy

BG finds one row at a time from Λ for approximating the solution of

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



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

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

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

The Analysis Model – Backward Greedy

Is there a similarity to a synthesis pursuit algorithm?

Synthesis OMP

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)]

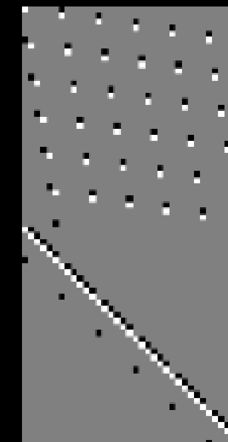
The Low-Spark Ω Case

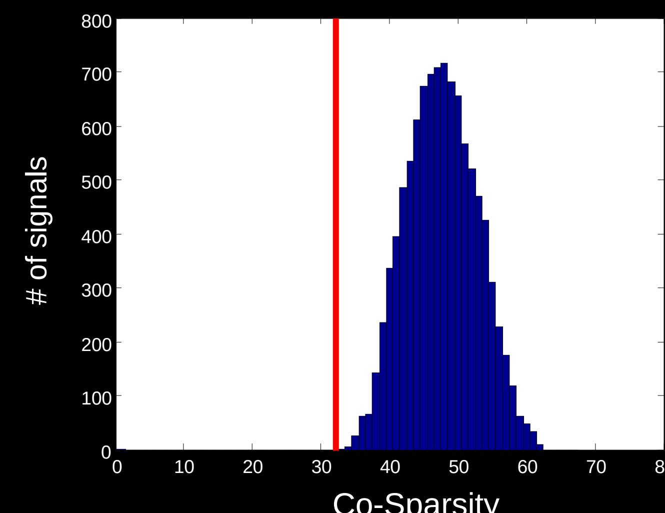
- What if spark(Ω^T)<<d ?
- For example: a TV-like operator for image-patches of size 6×6 pixels (Ω size is 72×36).
- Here are analysis-signals generated for co-sparsity (ℓ) of 32:

- 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.

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



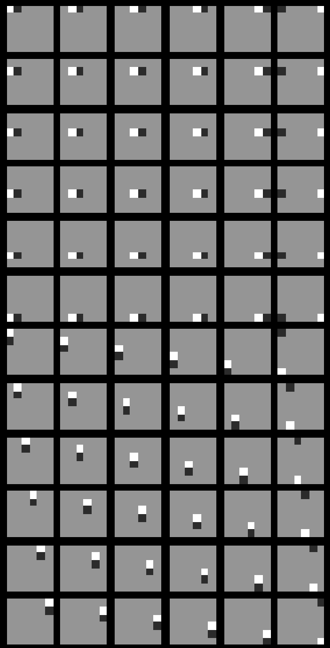


The Analysis Model – The Signature

SKIP?

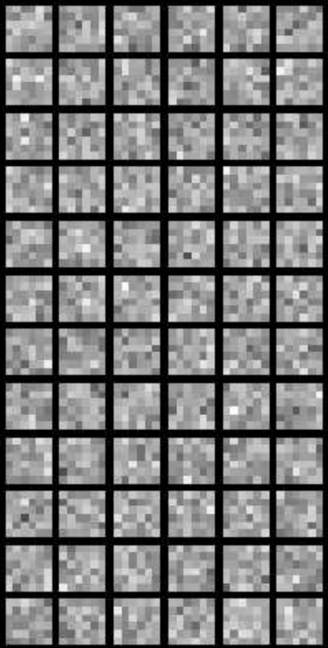
Consider two possible dictionaries:

Ω_{DIF}

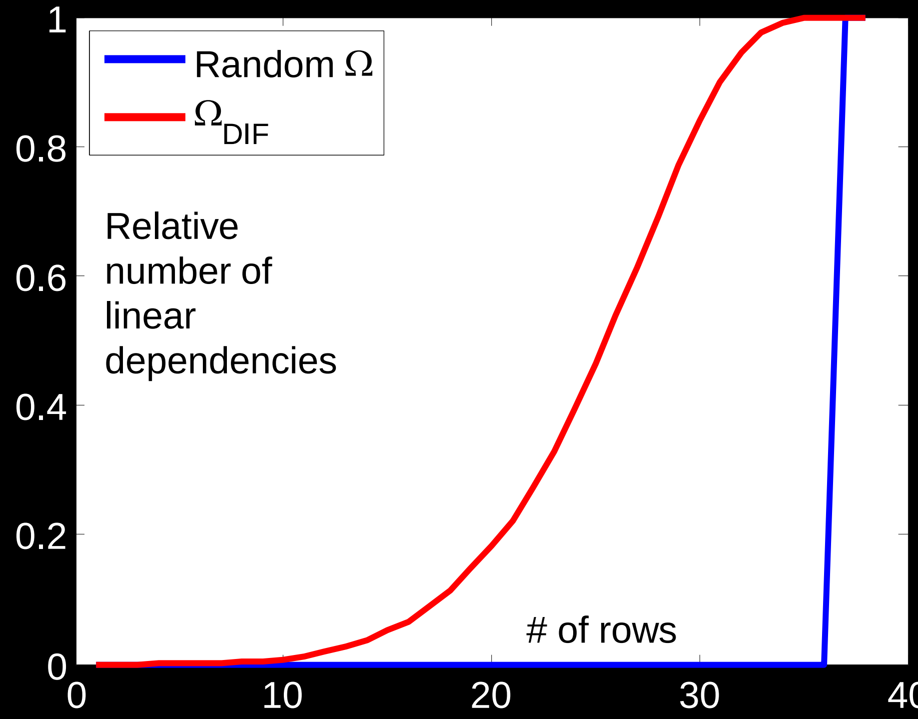


$Spark(\Omega^T) = 4$

Random Ω



$Spark(\Omega^T) = 37$

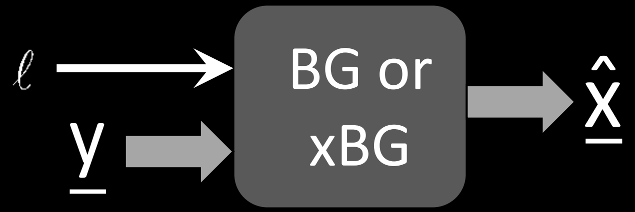


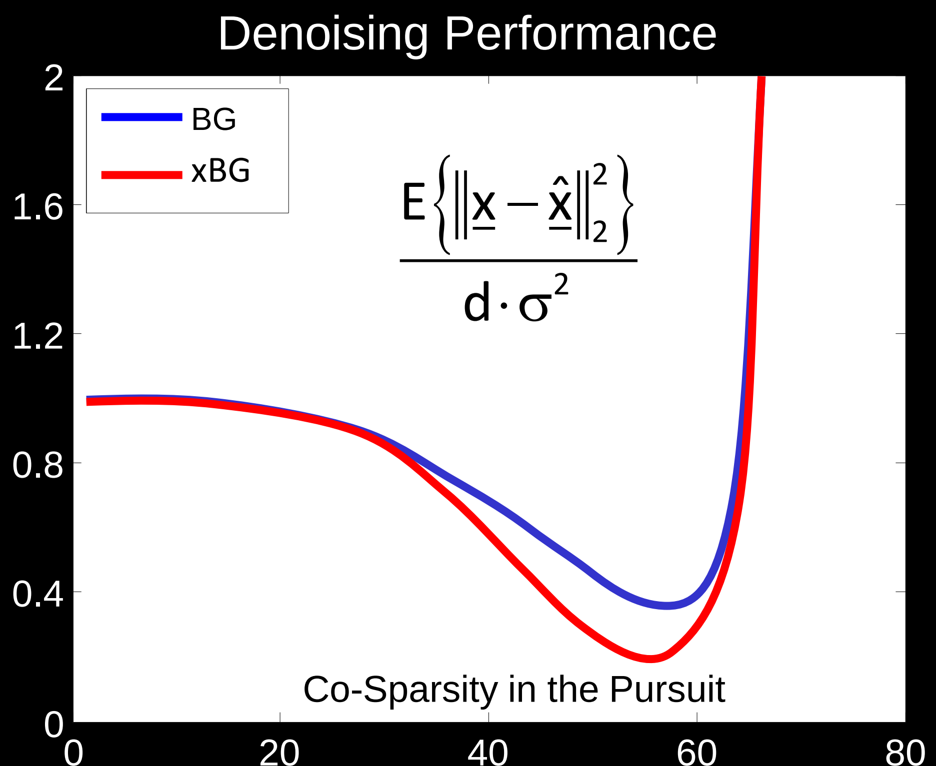
The Signature of a matrix is more informative than the Spark

The Analysis Model – Pursuit Results

SKIP?

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

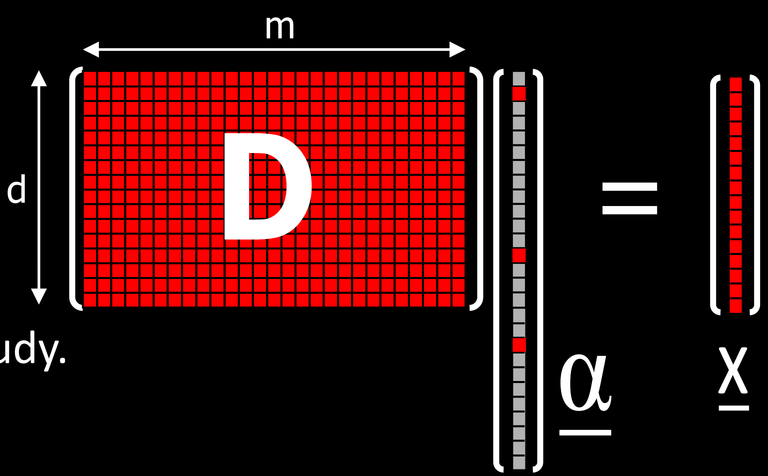


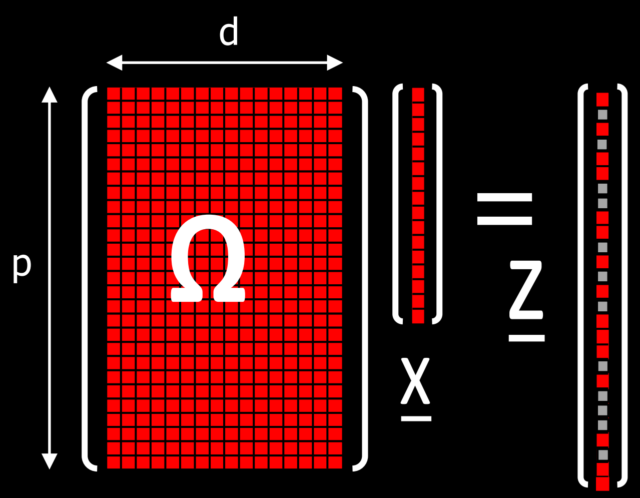


- We see an effective denoising, attenuating the noise by a factor ~0.3. This is achieved for an effective co-sparsity of ~55.

Synthesis vs. Analysis – Summary

- The two align for p=n=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.



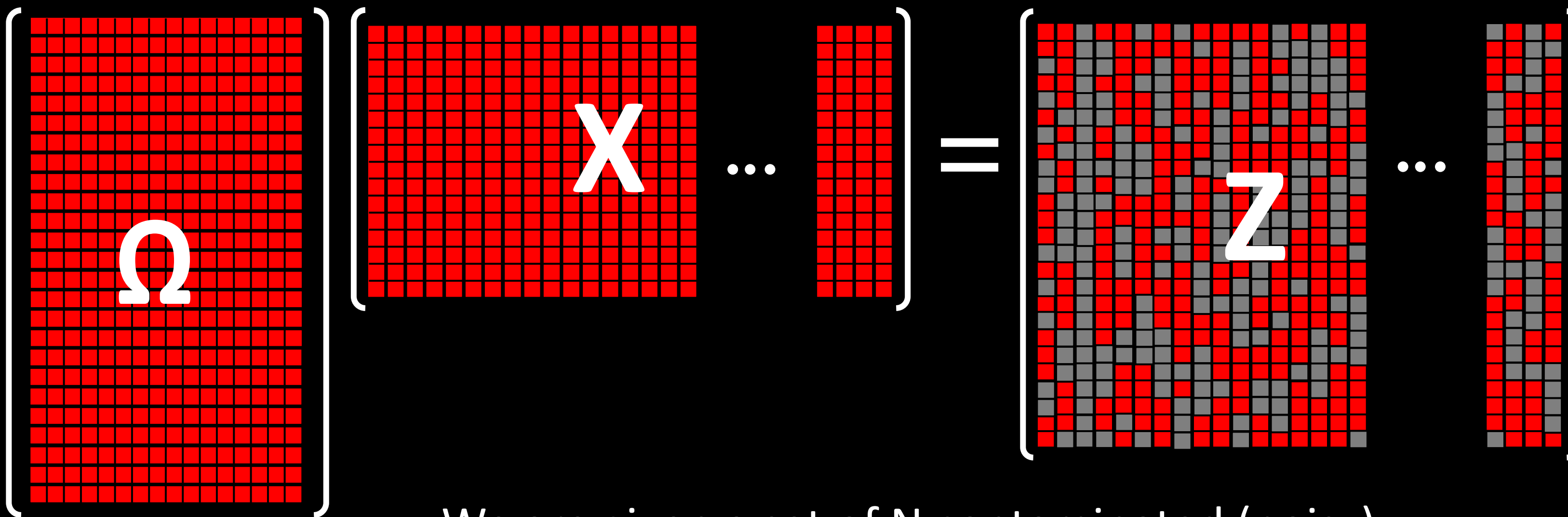


Part III – Dictionaries

Analysis Dictionary-Learning and Some Results

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 IEEE-TSP.

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, **Ω**

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

Analysis Dictionary Learning – Goal

Synthesis

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

Analysis

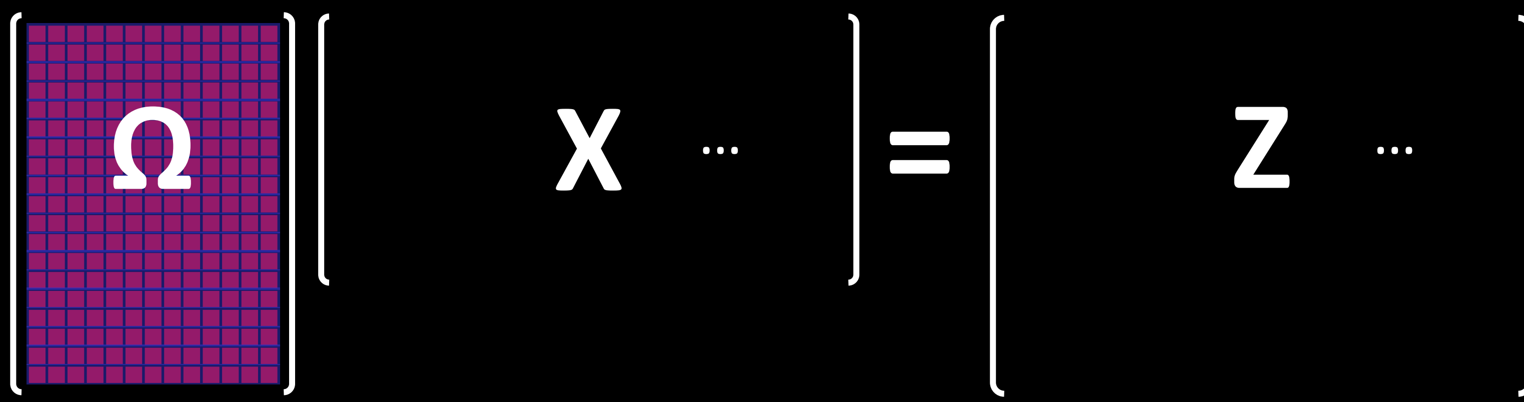
$$\underset{\mathbf{\Omega},\mathbf{X}}{\text{Min}} \left\|\mathbf{X}-\mathbf{Y}\right\|_F^2 \quad \text{s.t.} \quad \forall j=1,2,\ldots,N \quad \left\|\mathbf{\Omega}\underline{x}_j\right\|_0 \le p-\ell$$

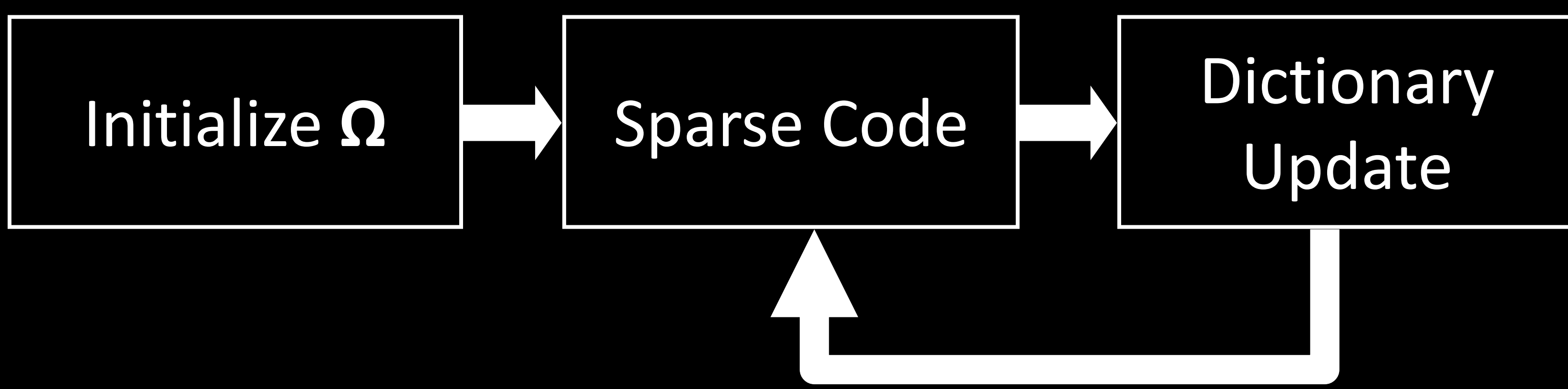
Noisy Examples

Denoised Signals are L_0 Sparse

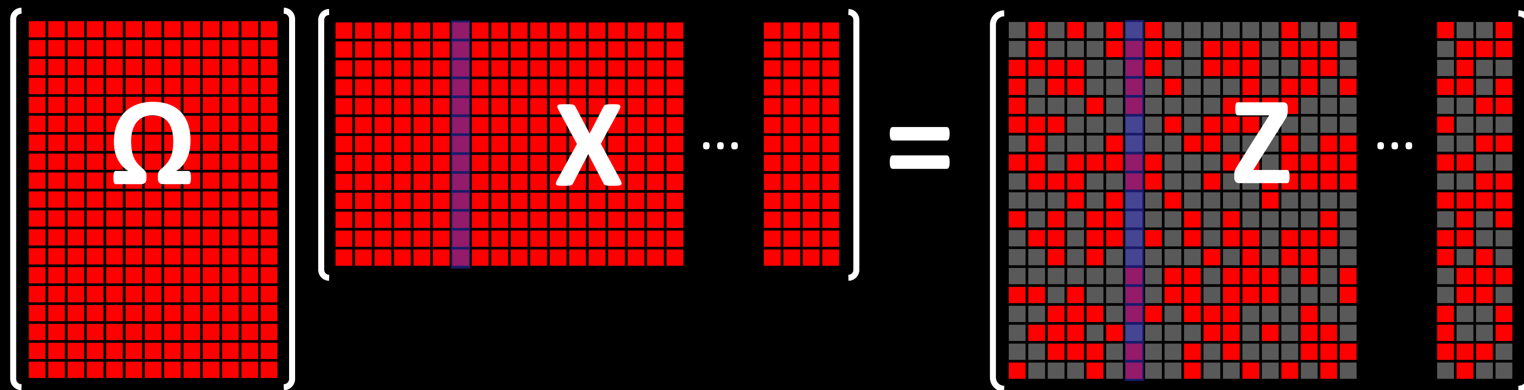
We shall adopt a similar approach to the K-SVD for approximating the minimization of the analysis goal

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





Analysis K-SVD – Sparse-Coding Stage



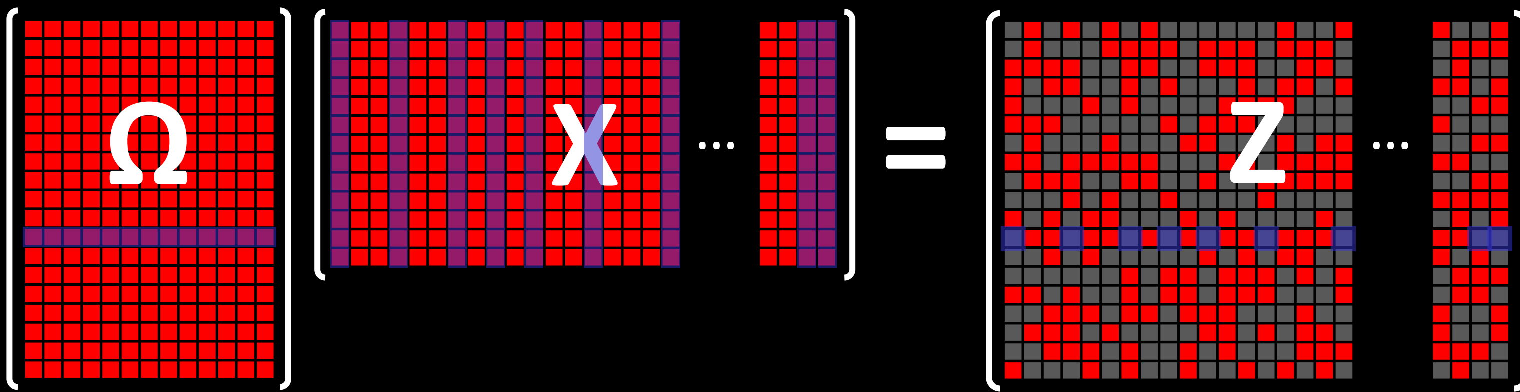
$$\underset{\Omega, X}{\text{Min}} \ \|X - Y\|_F^2 \ \ \text{s.t.}\ \forall j = 1,2,\ldots,N \ \ \|\Omega \underline{x}_j\|_0 \le p - \ell$$

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

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

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

Analysis K-SVD – Dictionary Update Stage



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

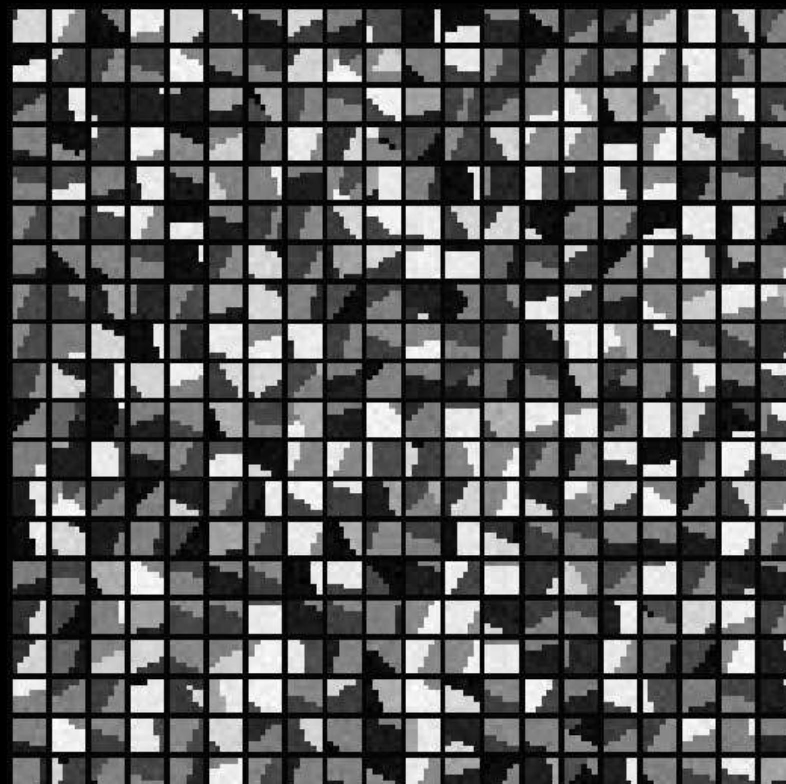
- Only signals orthogonal to the atom should get to vote for its new value.
- The known supports should be preserved.

Analysis Dictionary Learning – Results (1)

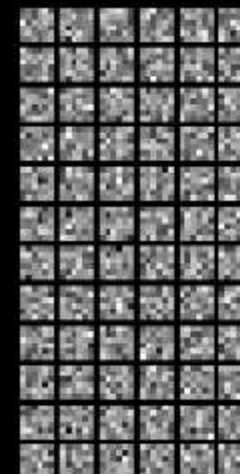
Experiment #1: Piece-Wise Constant Image

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

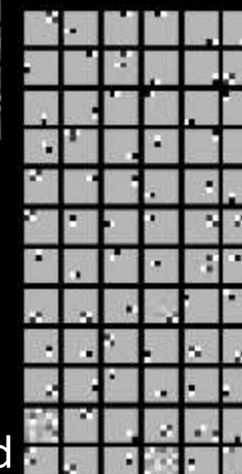
Original Image



Patches used for training



Initial Ω



Trained (100 iterations) Ω

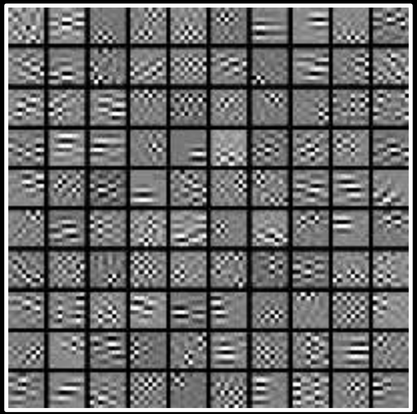
Analysis Dictionary Learning – Results (2)

Experiment #2: 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

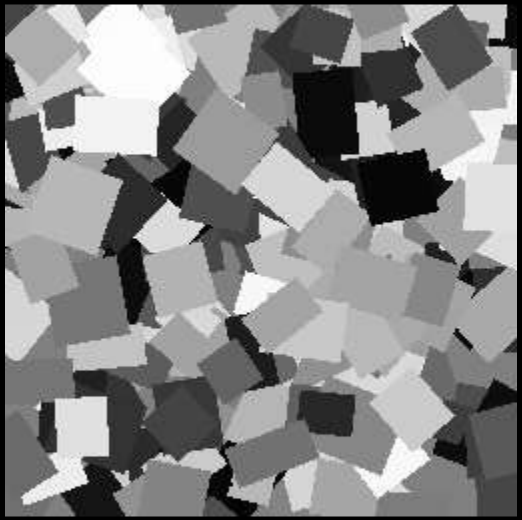
Original Images



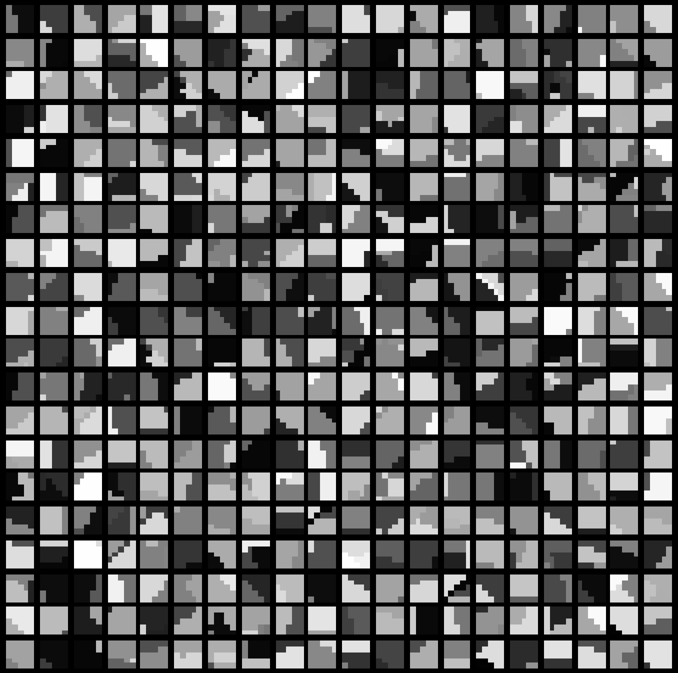
Trained Ω
(100 iterations)

Analysis Dictionary Learning – Results (3)

Experiment #3: denoising of piece-wise constant images



256×256



Non-flat patch examples

Analysis Dictionary Learning – Results (3)

	BM3D		Synthesis K-SVD		Sparse Analysis K-SVD	
Average subspace dimension	n/a		2.42	2.03	**1.75**	**1.74**
			1.79	1.69	**1.51**	**1.43**
Patch denoising: error per element	n/a		2.91	5.37	**1.97**	**4.38**
			7.57	10.29	**6.81**	**9.62**
Image PSNR [dB]	40.66	35.44	43.68	38.13	**46.02**	**39.13**
	32.23	30.32	34.83	32.02	**35.03**	**31.97**

Cell Legend:

$\sigma=5$	$\sigma=10$
$\sigma=15$	$\sigma=20$

Part V – 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