Recent Results on the Co-Sparse Analysis Model*

Michael Elad

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

Contributed Session: Mathematical Signal and Image Processing

The research leading to these results has received funding from:

The European Council under the European union's Seventh Framework Programme (FP/2007-2013) ERC grant Agreement ERC-SPARSE- 320649

Google Faculty Research Award

*Joint work with

Ron Rubinstein

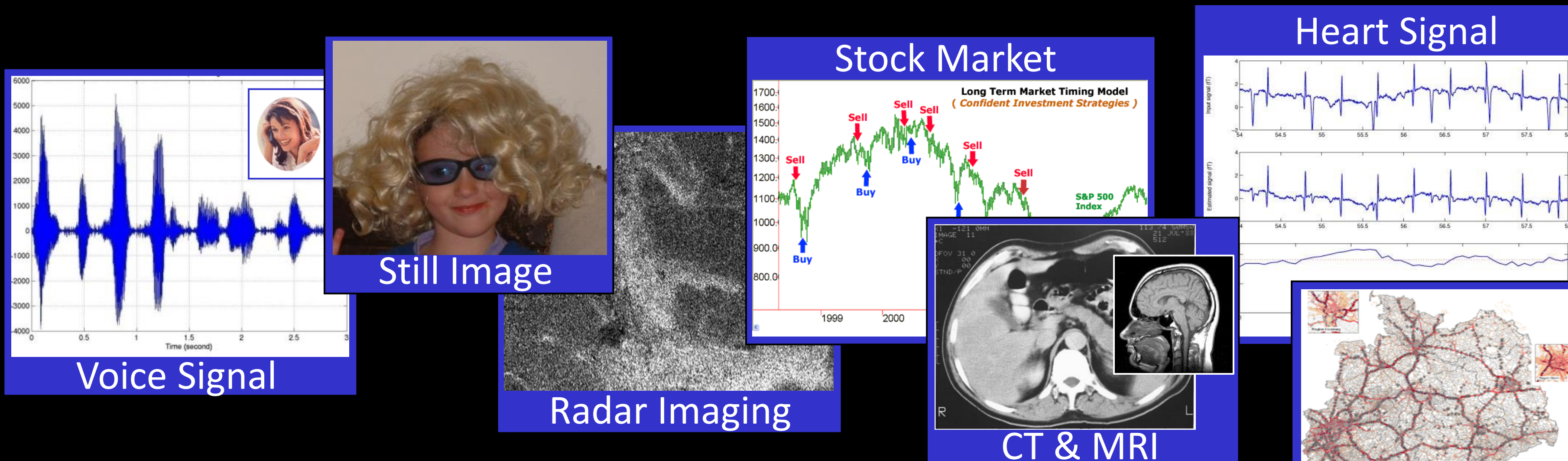
Tomer Peleg

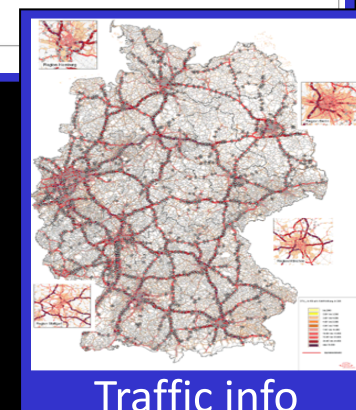
&

Remi Gribonval, 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 carries information, it must have an inner structure.
- This structure = rules the data complies with.
- Signal/image processing relies on exploiting these “rules” by adopting models.
- A model = mathematical construction describing the properties of the signal.
- In the past decade, sparsity-based models has been drawing major attention.

Sparsity-Based Models

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

Synthesis

Analysis

The attention to sp… based m…els has … to th… n, le… 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; BUT there are many things about it we do not know yet

Agenda

Part I - Background

Recalling the Synthesis Sparse Model

Part II - Analysis

Turning to the Analysis Model

Part III – THR Performance

Revealing Important Dictionary Properties

Part IV – Dictionaries

Analysis Dictionary-Learning and Some Results

Part V – We Are Done

Summary and Conclusions

Part I - Background

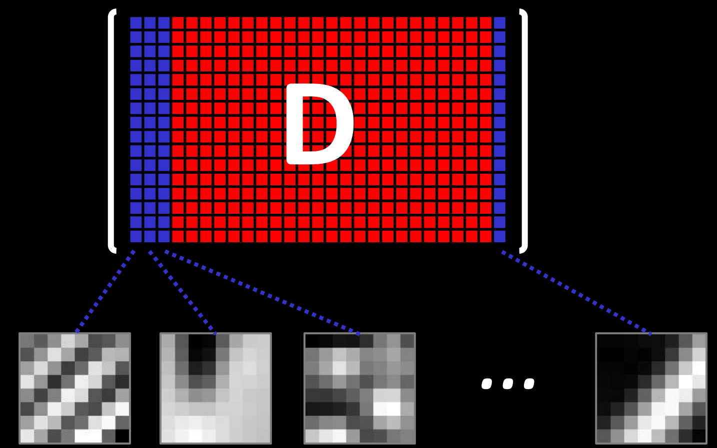
Recalling the Synthesis Sparse Model

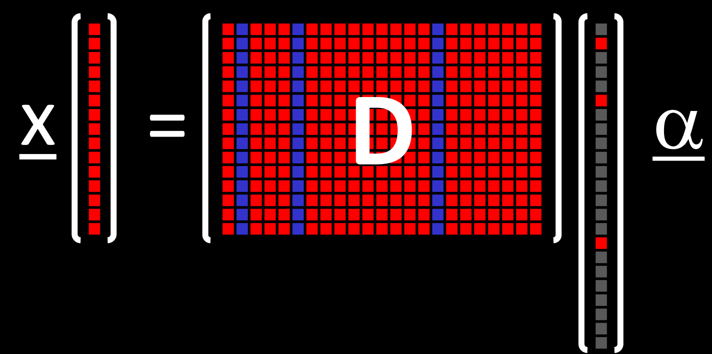
The Sparsity-Based Synthesis Model

- 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:

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

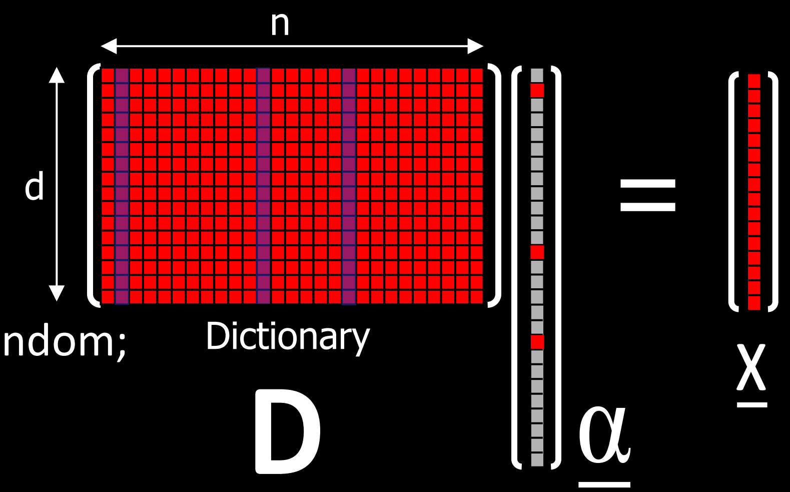
- We seek a sparsity of $\underline{\alpha}$, meaning that it is assumed to contain mostly zeros.
- We typically assume that n>d: redundancy.
- This model is typically referred to as the synthesis sparse and redundant representation model for signals.





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):

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

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

The Synthesis Model – Pursuit

- Fundamental problem: Given the noisy measurements,

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

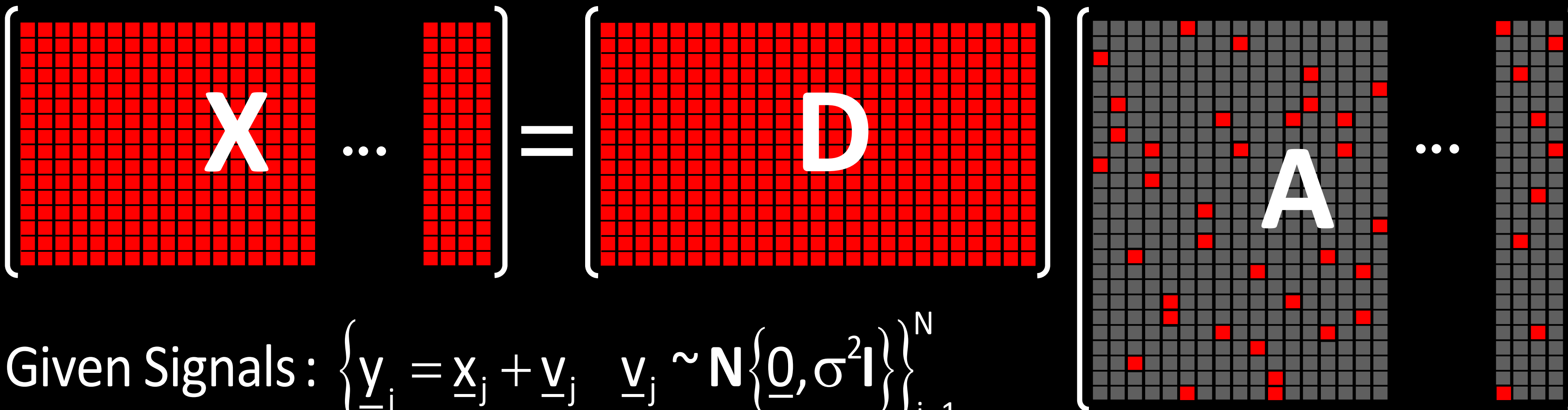
recover the clean signal $\underline{x}$ – This is a denoising task.

- This can be posed as: $\hat{\underline{\alpha}} = \underset{\underline{\alpha}}{\text{ArgMin}} \left\| \underline{y} - \mathbf{D}\underline{\alpha} \right\|_2^2 \text{ s.t. } \left\| \underline{\alpha} \right\|_0 = k \Rightarrow \hat{\underline{x}} = \mathbf{D}\hat{\underline{\alpha}}$
- 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 : $\left\{ \underline{y}_j = \underline{x}_j + \underline{v}_j \quad \underline{v}_j \sim \mathbf{N}\left\{ \underline{0}, \sigma^2 \mathbf{I} \right\} \right\}_{j=1}^{N}$

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

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)
…

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" , ACHA, Vol. 34, No. 1, Pages 30-56, January 2013.

The Analysis Model – Basics

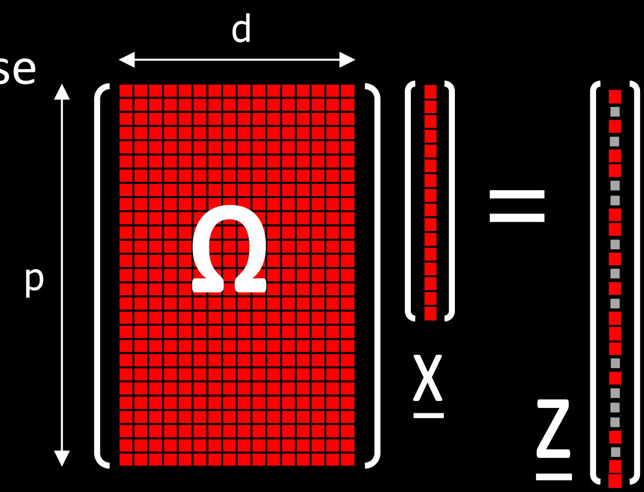
* $\text{spark}\{\mathbf{\Omega}^T\} = d+1$

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

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

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

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



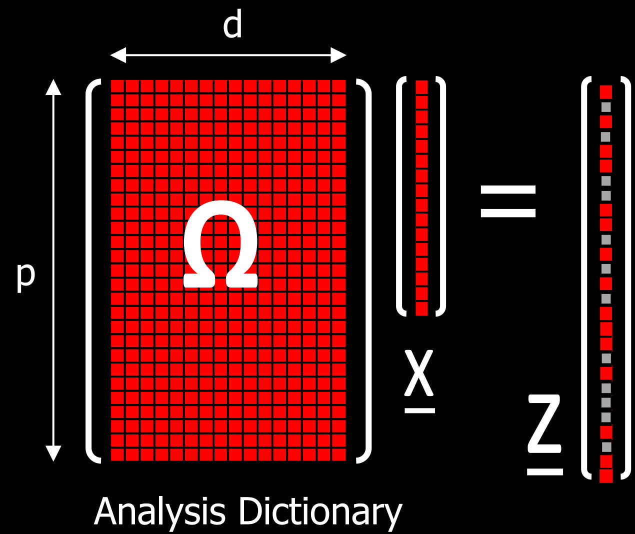
Analysis Dictionary

- This model puts an emphasis on the zeros in $\underline{z}$ for characterizing the signal, just like zero-crossings of wavelets used for defining a signal [Mallat (`91)].
- Co-Rank: Rank($\mathbf{\Omega}_\Lambda$)≤ℓ (strictly smaller if there are linear dependencies in **Ω**).
- If **Ω** is in general position*, then the co-rank and the co-sparsity are the same, and $0 \le \ell < d$, implying that we cannot expect to get a truly sparse analysis.

The Analysis Model – Bayesian View

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

	Synthesis Signals
Support:	Choose the support T (|T|=k) at random
Coef. :	Choose $\underline{\alpha}_T$ at random
Generate:	Synthesize by: $\mathbf{D}_T\underline{\alpha}_T=\underline{x}$

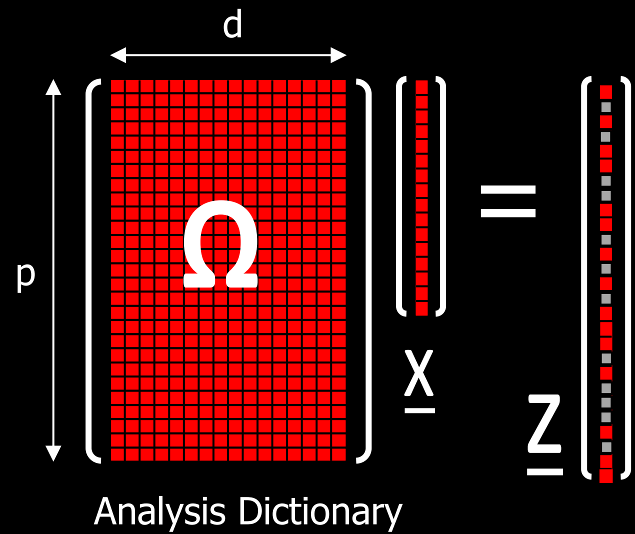


- Bottom line: an analysis signal $\underline{x}$ satisfies: $\exists\Lambda \mid |\Lambda|=\ell$ s.t. $\mathbf{\Omega}_\Lambda \underline{x}=\underline{0}$.

The Analysis Model – UoS

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

	Synthesis Signals
What is the Subspace Dimension:	k
How Many Subspaces:	$\binom{n}{k}$
Who are those Subspaces:	$\text{span}\{\mathbf{D}_T\}$



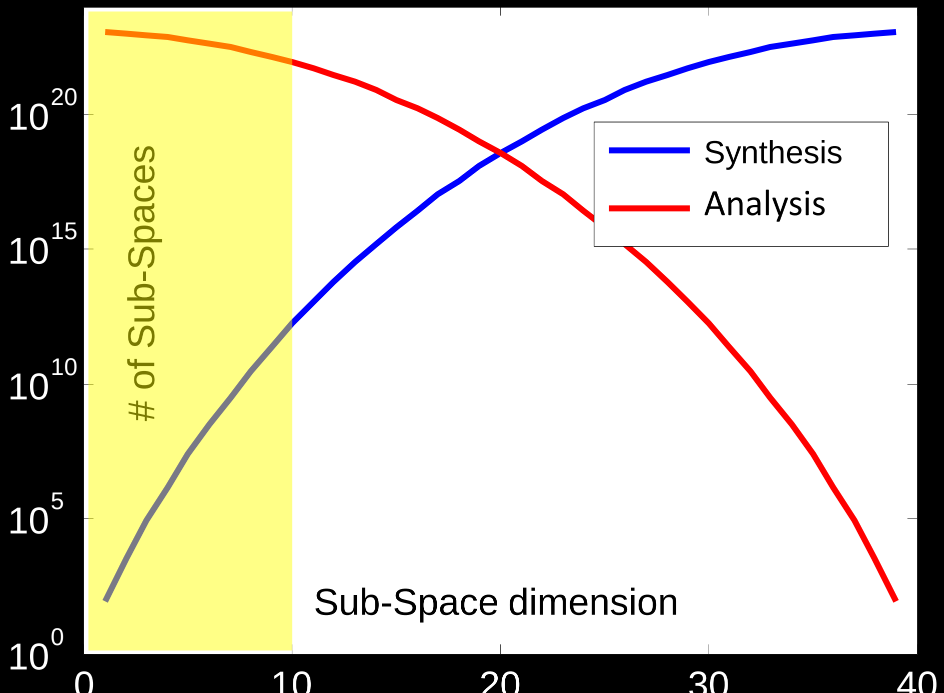
- The analysis and the synthesis models offer both a UoS construction, but these are very different from each other in general.

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.

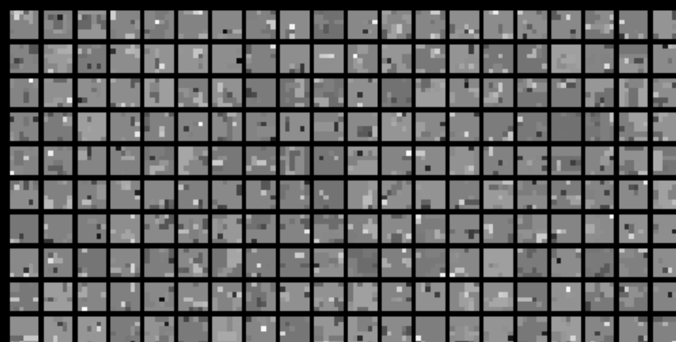
- In the general case, for d=40 and p=n=80, this graph shows the count of the number of subspaces. As can be seen, the two models are substantially different, the analysis model is much richer in low-dim., and the two complete each other.

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



The Low-Spark Ω Case

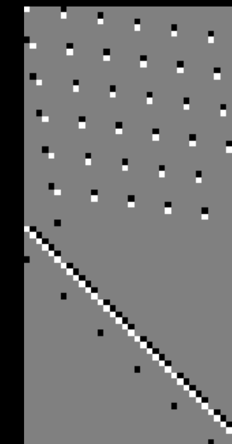
- What if spark(Ω^T)<<d ?
- For example: a TV-like operator for image-patches of size 6×6 pixels (Ω size is 72×36).
- Here are analysis-signals generated for co-sparsity (ℓ) of 32:

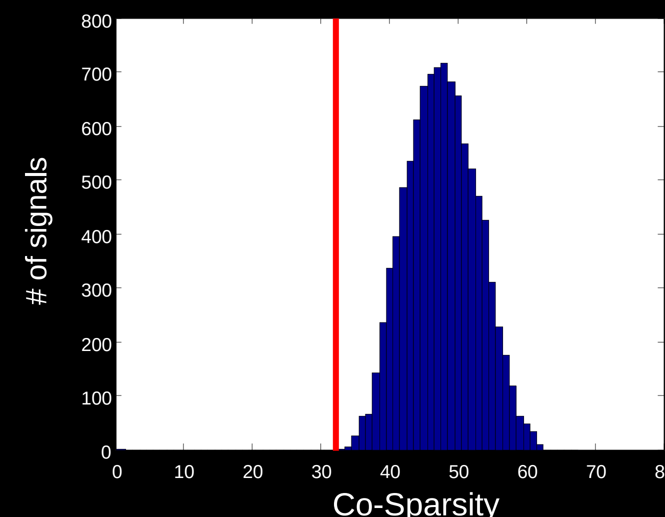


- Their true co-sparsity is higher – see graph:
- In such a case we may consider $\ell > d$, and thus

… the number of subspaces is smaller.

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





The Analysis Model – Pursuit

- Fundamental problem: Given the noisy measurements,

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

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

- This goal can be posed as:

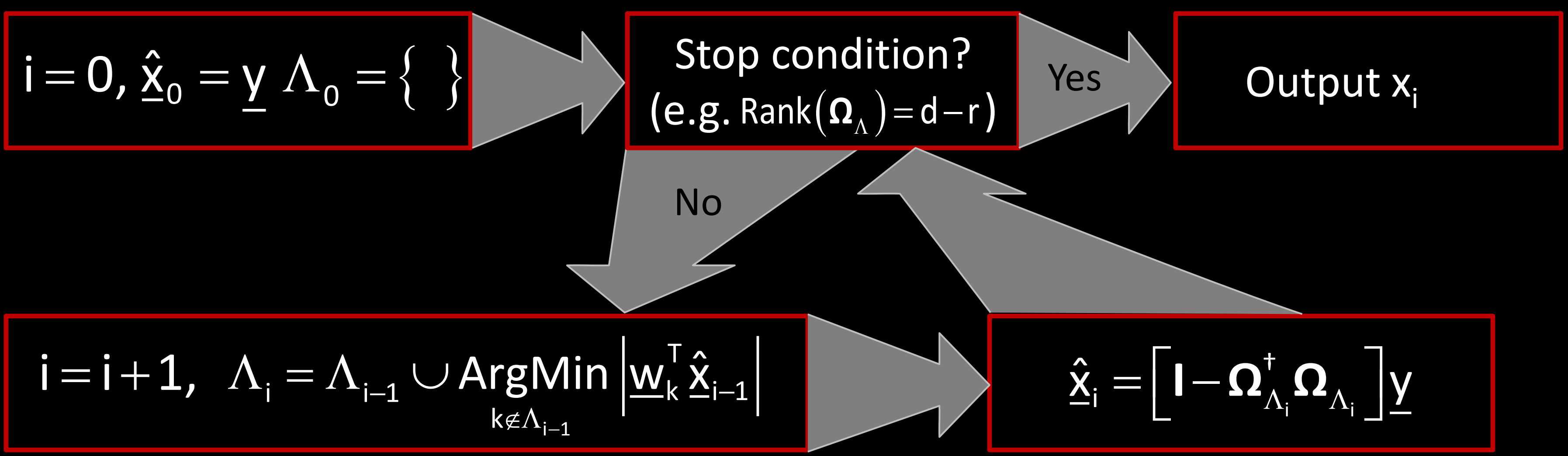
$$\hat{\underline{x}} = \underset{\underline{x},\Lambda}{\text{ArgMin}} \|\underline{y} - \underline{x}\|_2^2 \text{ s.t. } \mathbf{\Omega}_{\Lambda}\underline{x} = 0 \quad \& \quad |\Lambda| = p - \ell \text{ or } \text{rank}(\mathbf{\Omega}_{\Lambda}) = d - r$$

- 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 (BG, OBG, GAP), and Hybrid methods (AIHT, ASP, ACoSaMP, …).
- Theoretical study providing pursuit guarantees depending on the co-sparsity and properties of $\mathbf{\Omega}$. See [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

$$\underline{\hat{x}} = \underset{\underline{x},\Lambda}{\text{ArgMin}} \left\| \underline{y} - \underline{x} \right\|_2^2 \text{ s.t. } \left\{ \mathbf{\Omega}_\Lambda \underline{x} = 0 \ \ \& \ \ \text{Rank}\left(\mathbf{\Omega}_\Lambda\right) = d - r \right\}$$



The Analysis Model – Backward Greedy

Is there a similarity to a synthesis pursuit algorithm?

Synthesis OMP

$i=0$

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

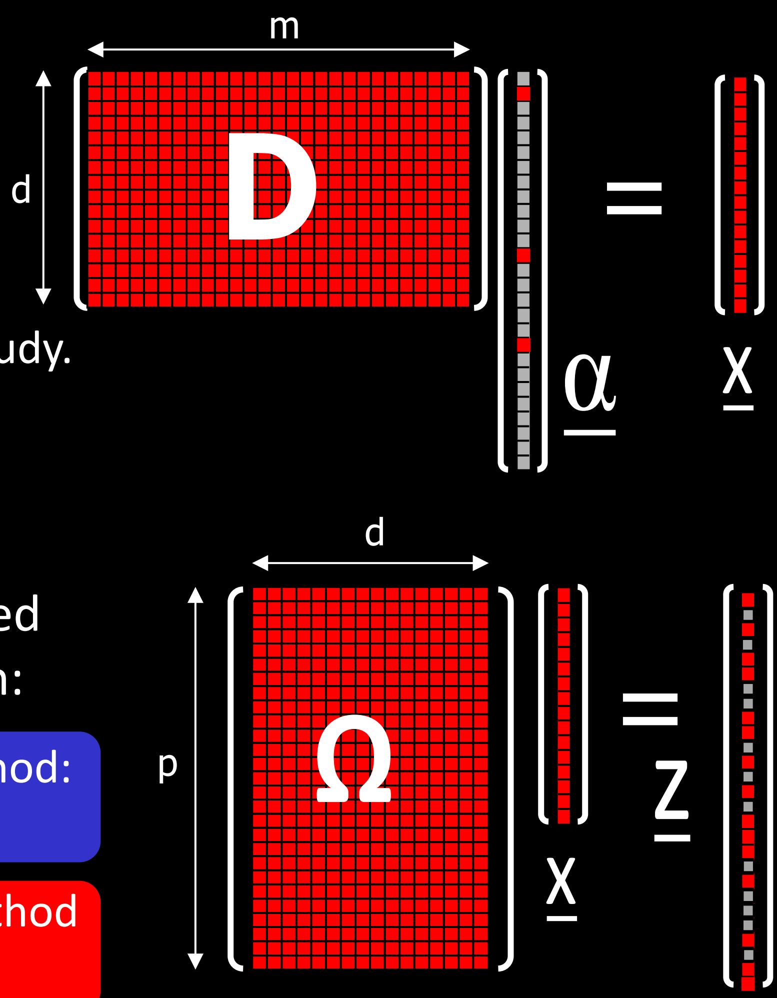
$i=i$

Other options:

- Optimized BG pursuit (OBG) [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 like algorithms [Giryes, et. al. (`12)]
- Analysis-THR [Peleg & Elad (`12)]

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 work on the analysis model so far touched on all the above. In this talk we shall focus on:
  - A theoretical study of the simplest pursuit method: Analysis-THR.
  - Developing a K-SVD like dictionary learning method for the analysis model.



Part III – THR Performance

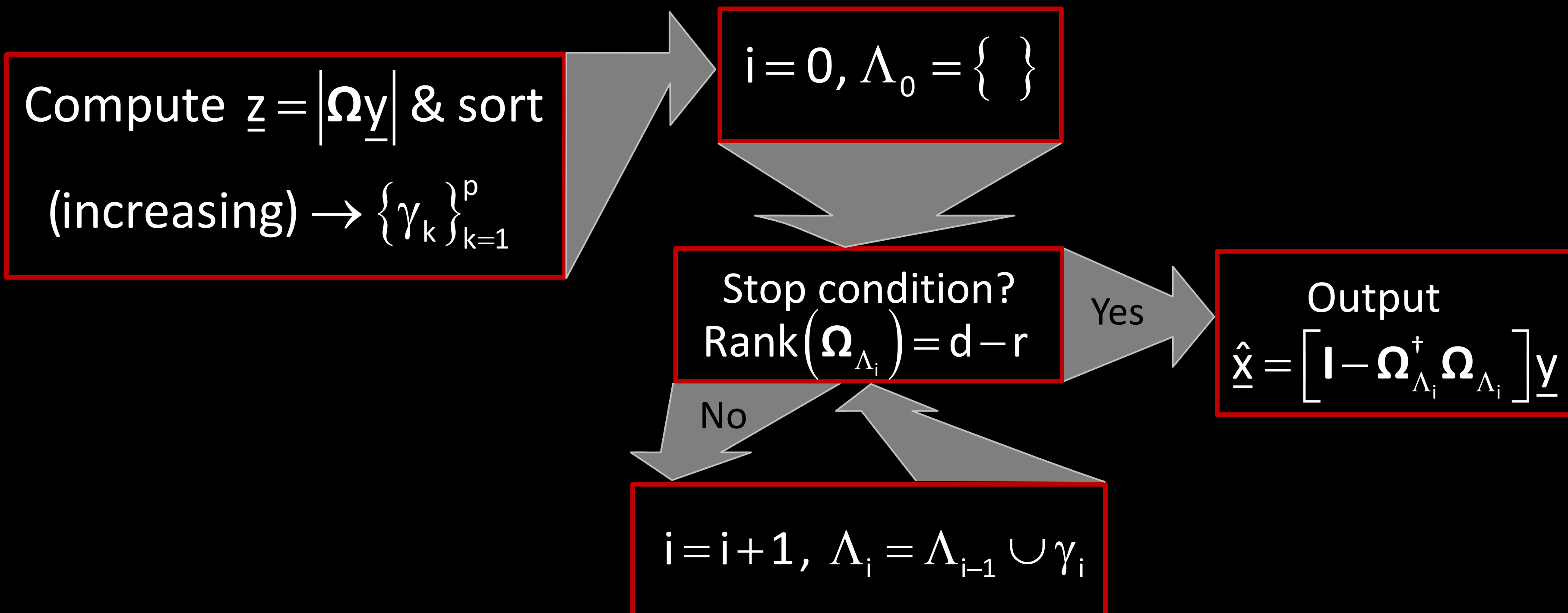
Revealing Important Dictionary Properties

1. T. Peleg and M. Elad, Performance Guarantees of the Thresholding Algorithm for the Co-Sparse Analysis Model, IEEE Transactions on Information Theory, Vol. 59, No. 3, Pages 1832-1845, March 2013.

The Analysis-THR Algorithm

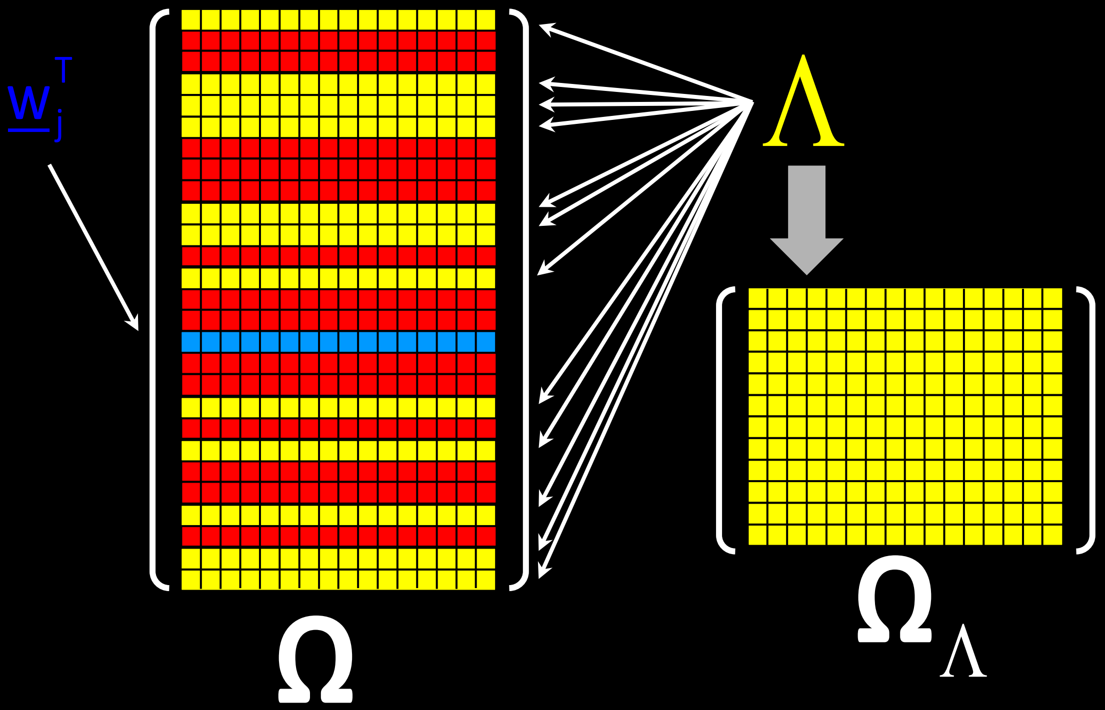
Analysis-THR aims to find an approximation for the problem

$$\hat{\underline{x}} = \underset{\underline{x},\Lambda}{\text{ArgMin}} \left\| \underline{y} - \underline{x} \right\|_2^2 \text{ s.t. } \left\{ \boldsymbol{\Omega}_\Lambda \underline{x} = 0 \;\; \& \;\; \text{Rank}\left(\boldsymbol{\Omega}_\Lambda\right) = d - r \right\}$$



The Restricted Ortho. Projection Property

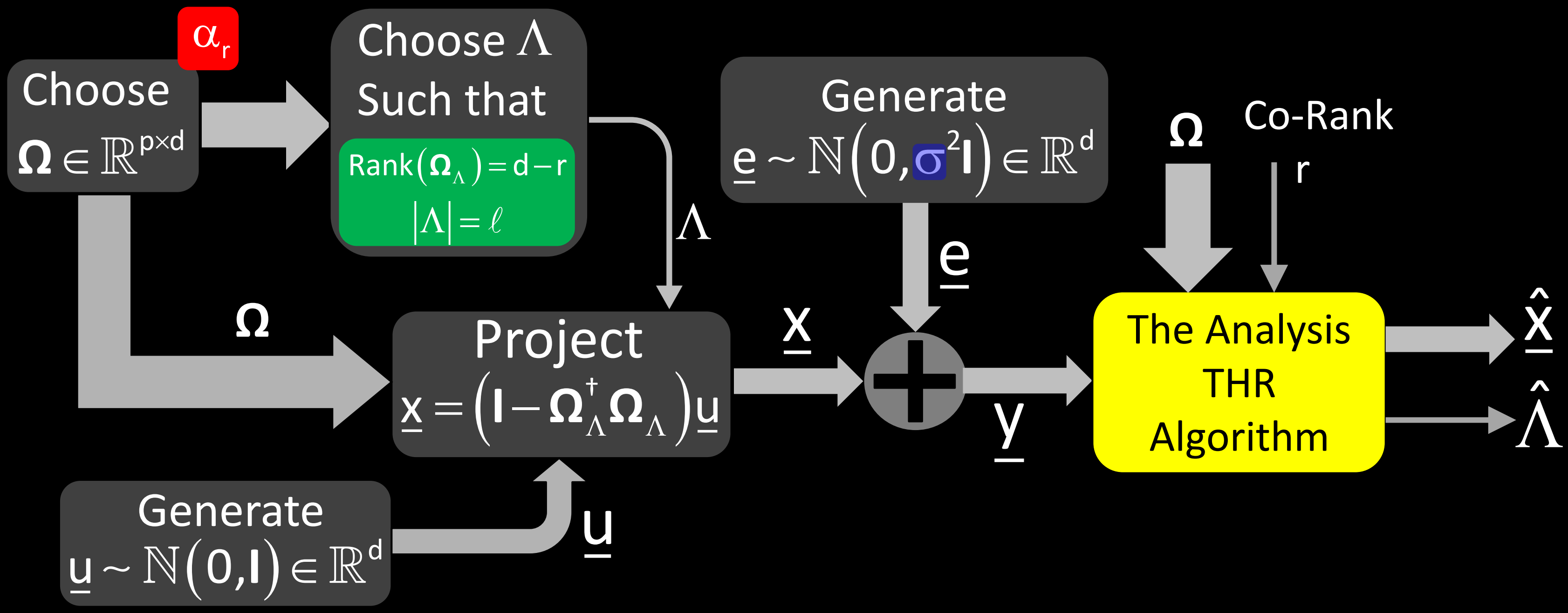
$$\alpha_r = \min_{\Lambda, j \,|\, \substack{\text{Rank}(\mathbf{\Omega}_\Lambda) = d-r \\ \text{and } j \notin \Lambda}} \left\| \left(\mathbf{I} - \mathbf{\Omega}_\Lambda^\dagger \mathbf{\Omega}_\Lambda \right) \underline{w}_j \right\|_2$$



- ROPP aims to get near orthogonality of rows outside the co-support (i.e., α_r should be as close as possible to 1).
- This should remind of the (synthesis) ERC [Tropp ('04)]:

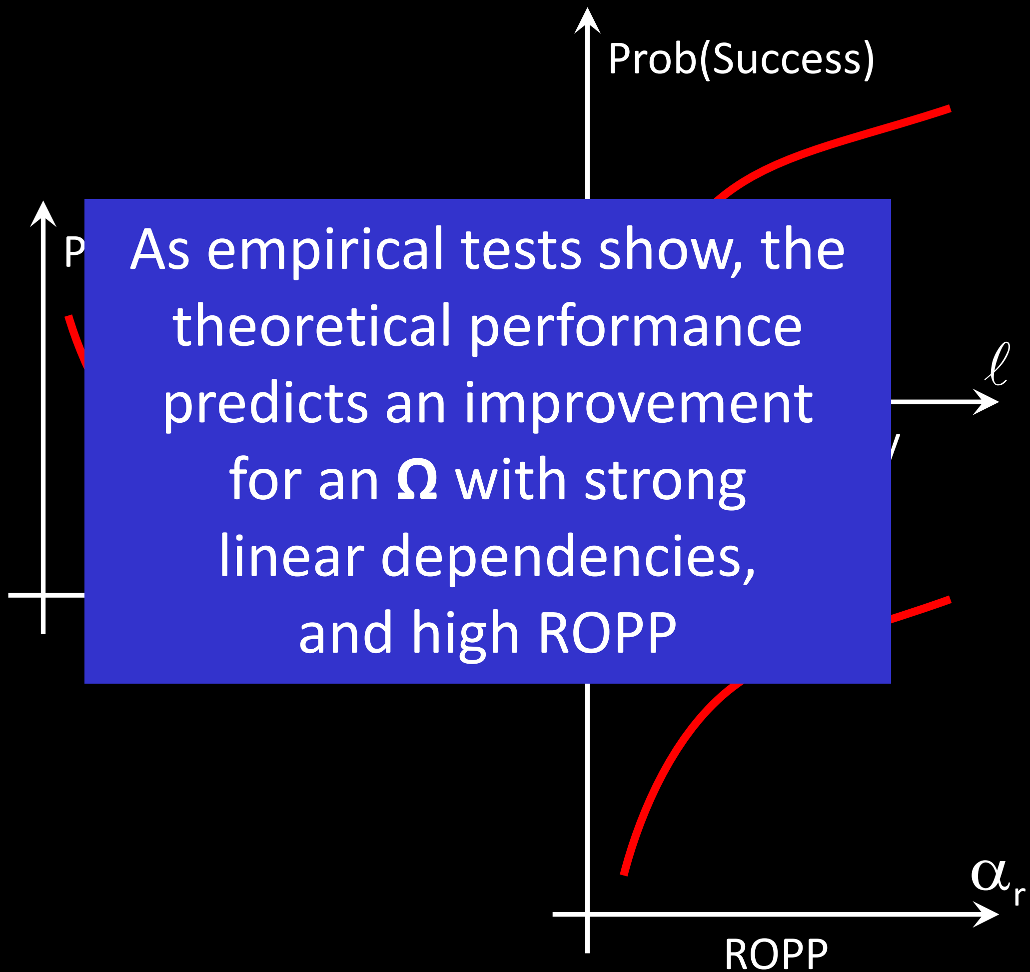
$$\max_{S, j \,|\, |S| = k \,\&\, j \notin S} \left\| \mathbf{D}_S^\dagger \underline{d}_j \right\|_1 \leq 1$$

Theoretical Study of the THR Algorithm

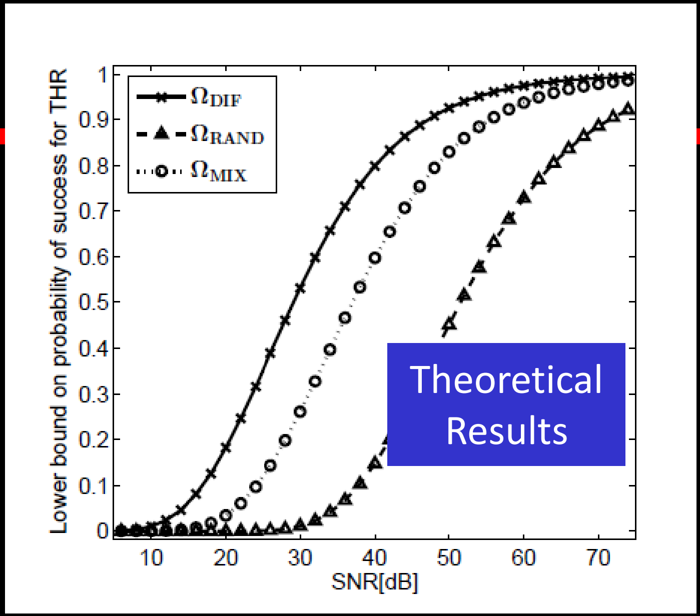


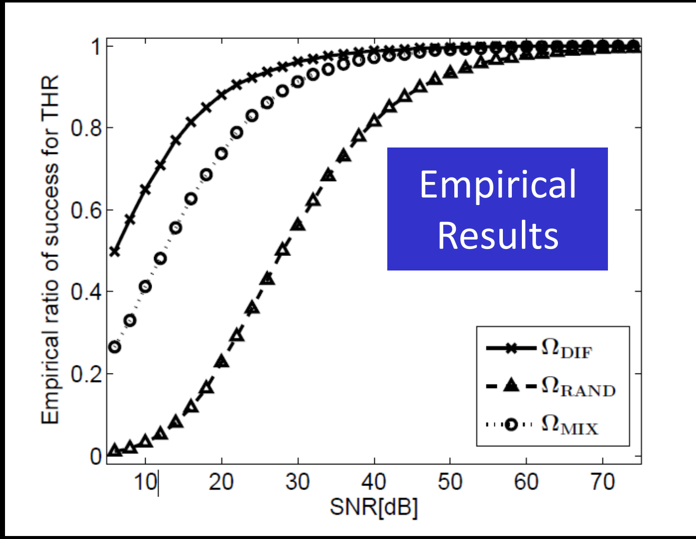
$$\Pr\left\{\text{sucess}\left(\text{i.e. } \hat{\Lambda} = \Lambda\right)\right\}$$

Implications



As empirical tests show, the theoretical performance predicts an improvement for an **Ω** with strong linear dependencies, and high ROPP

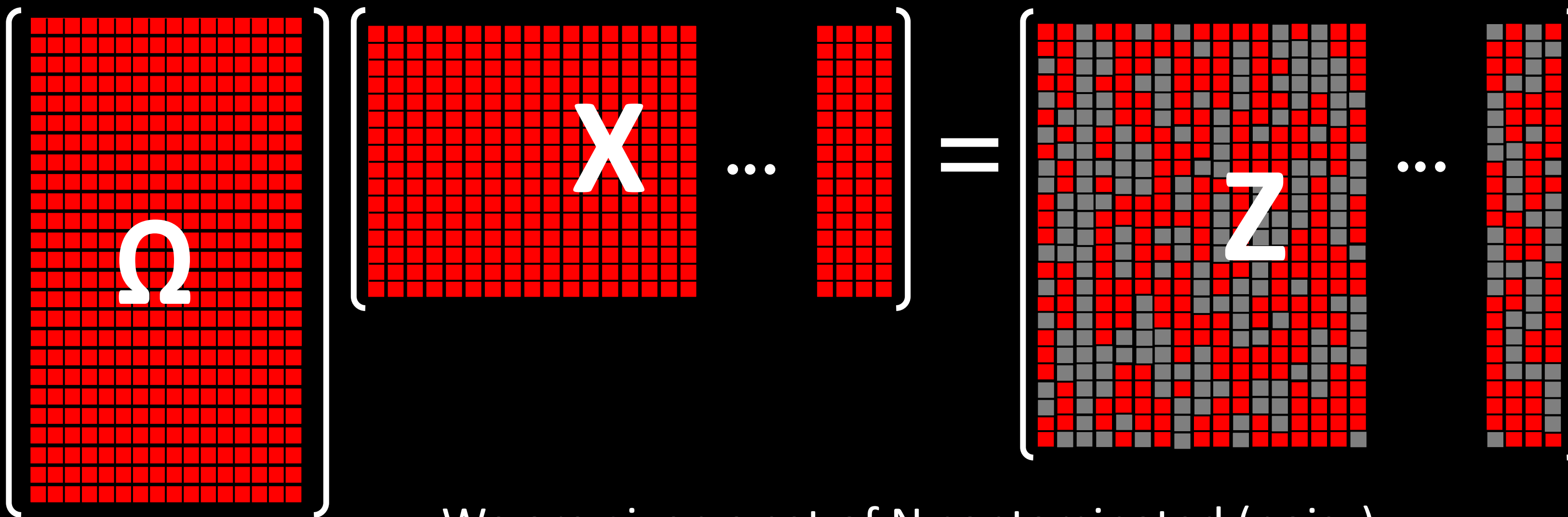




Part IV – Dictionaries
Analysis Dictionary-Learning and Some Results

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", IEEE-TSP, Vol. 61, No. 3, Pages 661-677, March 2013.

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

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

Analysis

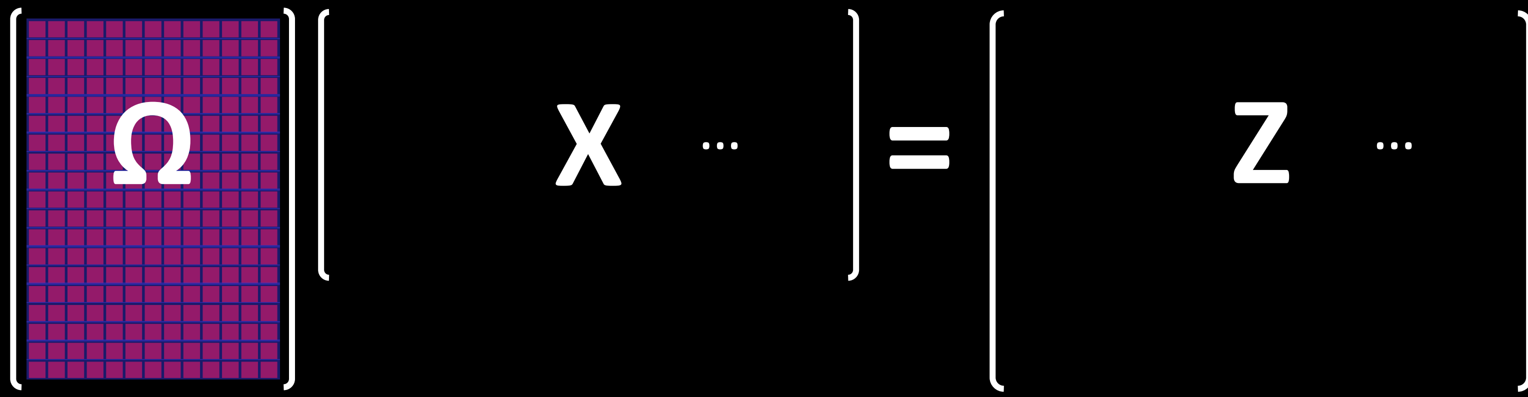
$$\underset{\mathbf{\Omega},\mathbf{X}}{\text{Min}} \left\| \mathbf{X} - \mathbf{Y} \right\|_F^2 \quad \text{s.t.} \ \forall j = 1,2,\ldots,N \quad \left\| \mathbf{\Omega} \underline{x}_j \right\|_0 \le p - \ell$$

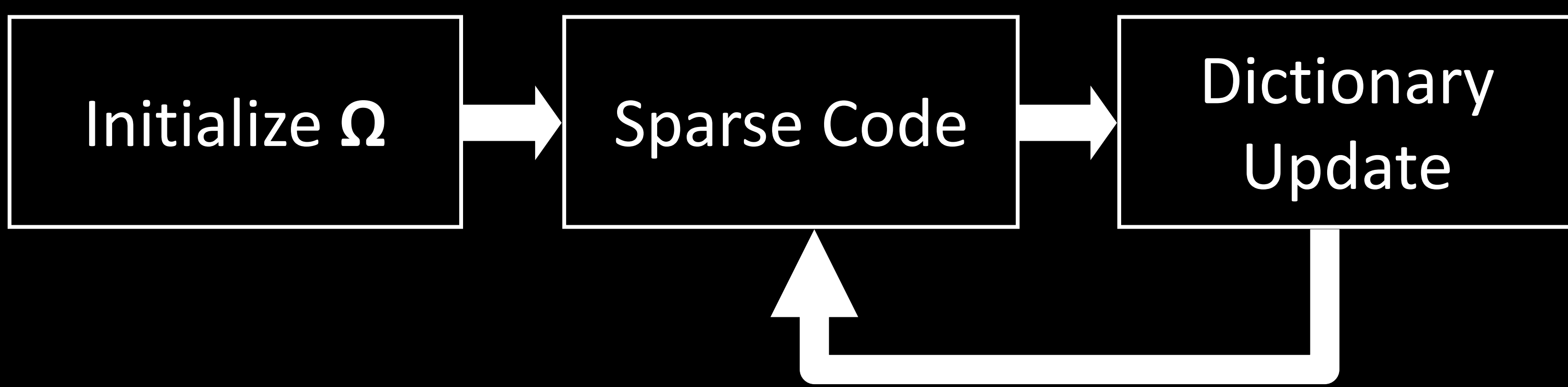
Noisy Examples

Denoised Signals are L_0 Sparse

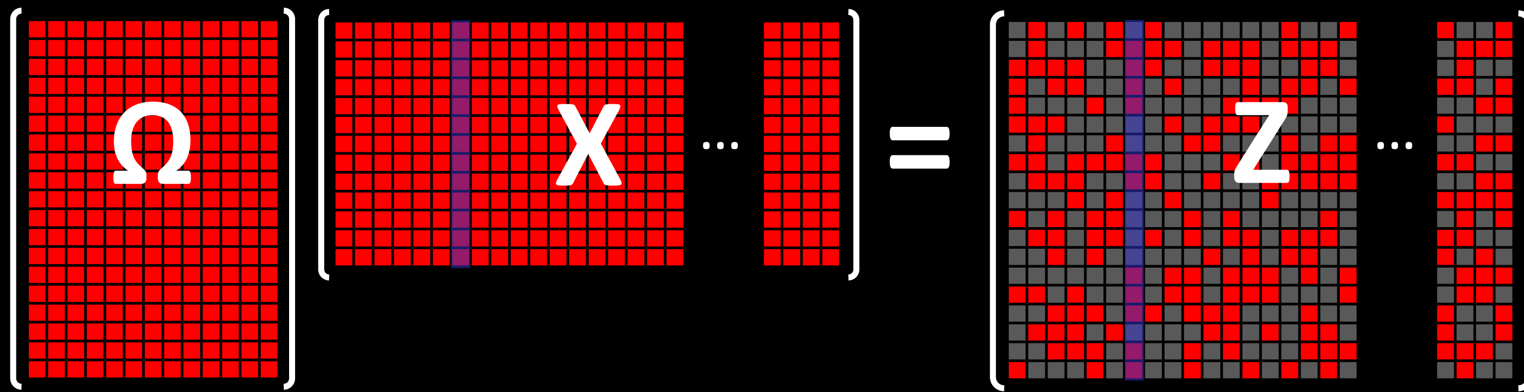
We shall adopt a similar approach to the K-SVD for approximating the minimization of the analysis goal

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





Analysis K-SVD – Sparse-Coding Stage



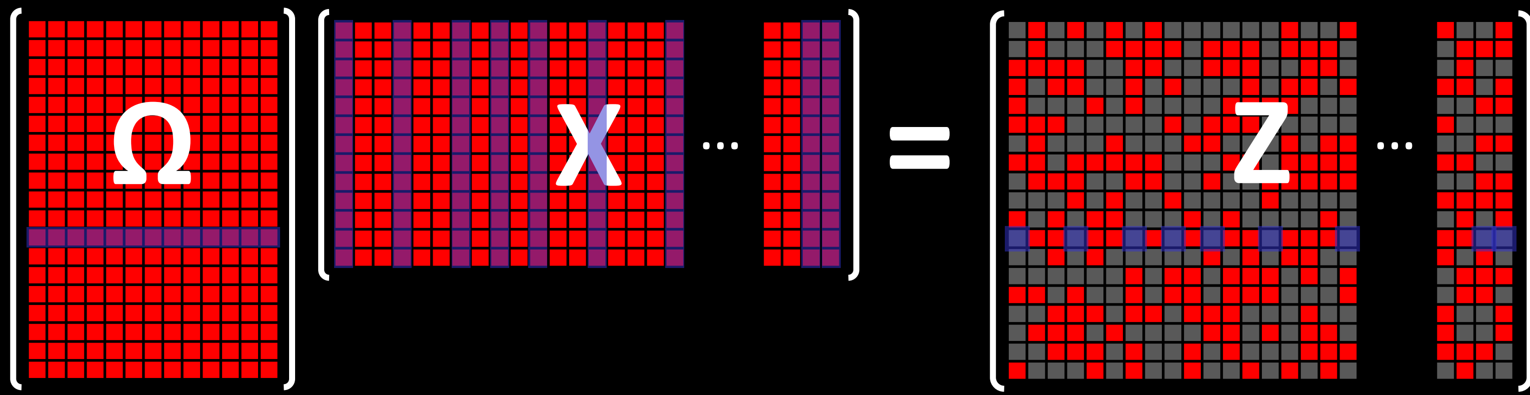
$$\underset{\Omega,X}{\text{Min}} \left\| X - Y \right\|_F^2 \quad \text{s.t.} \; \forall j = 1,2,\ldots,N \;\; \left\| \Omega \underline{x}_j \right\|_0 \le p - \ell$$

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

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

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

Analysis K-SVD – Dictionary Update Stage



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

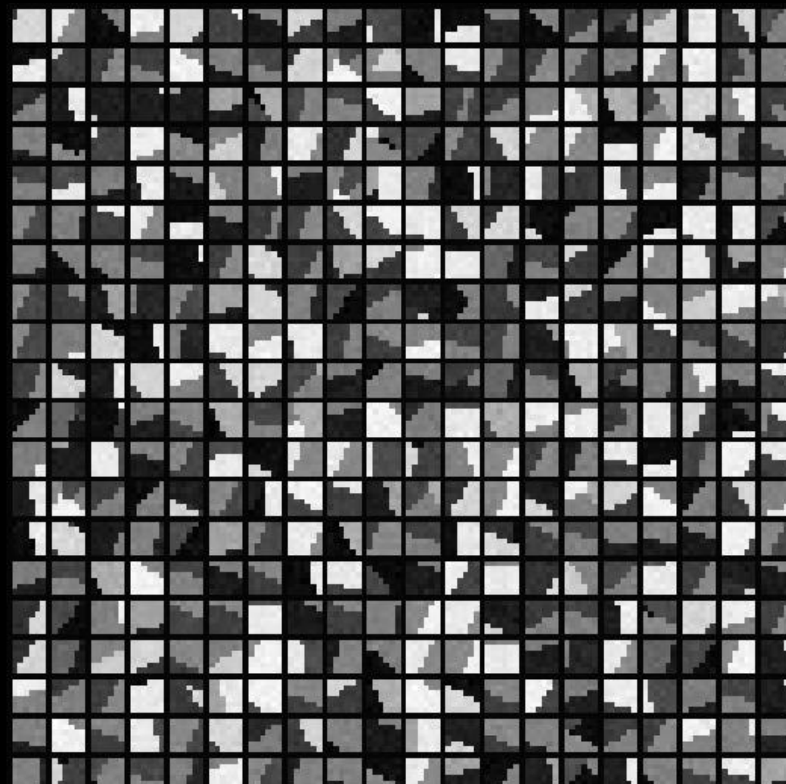
- Only signals orthogonal to the atom should get to vote for its new value.
- The known supports should be preserved.
- Improved results for applications are obtained by promoting linear dependencies within **Ω**.

Analysis Dictionary Learning – Results (1)

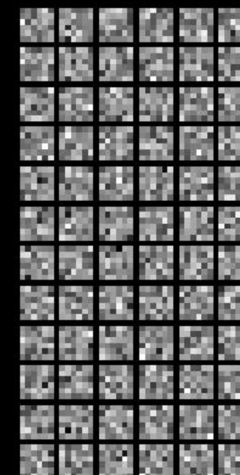
Experiment #1: Piece-Wise Constant Image

- We take 10,000 6×6 patches (+noise $\sigma=5$) to train on
- Here is what we got (we promote sparse outcome):

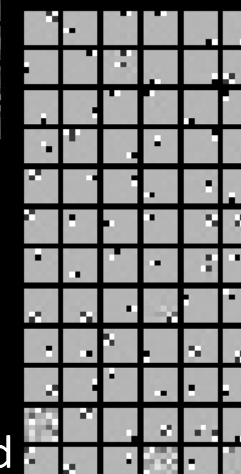
Original Image



Patches used for training



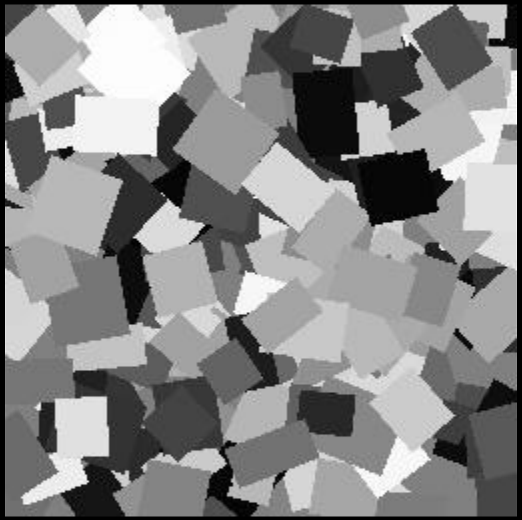
Initial Ω



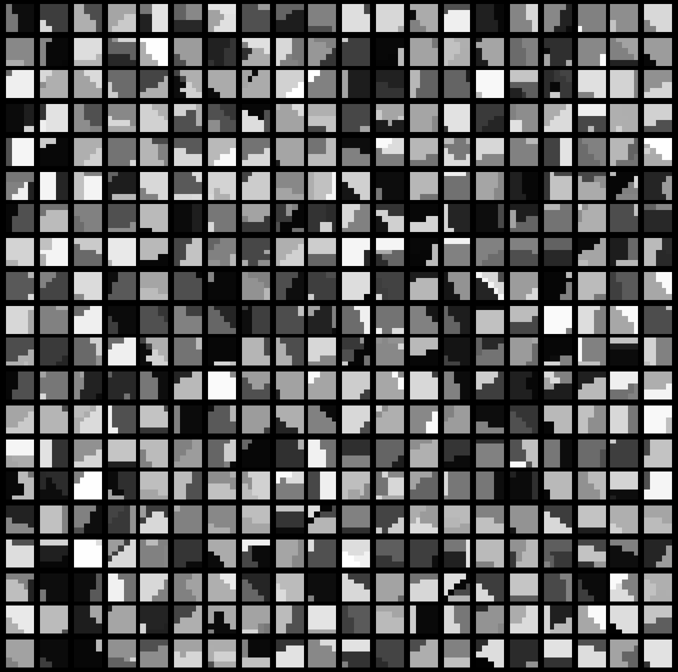
Trained (100 iterations) Ω

Analysis Dictionary Learning – Results (2)

Experiment #2: denoising of the piece-wise constant image.



256×256

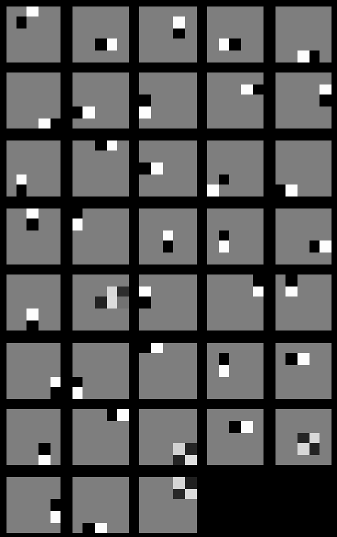


Non-flat patch examples

Analysis Dictionary Learning – Results (2)

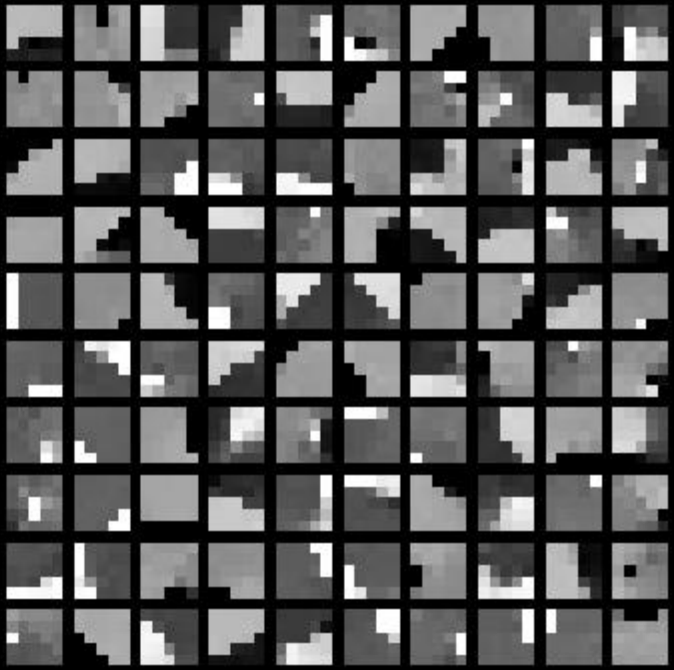
Learned dictionaries for σ=5

Analysis Dictionary



38 atoms
(again, promoting sparsity in $\mathbf{\Omega}$)

Synthesis Dictionary



100 atoms

Analysis Dictionary Learning – Results (2)

	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:

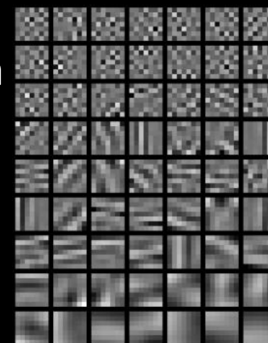
σ=5	σ=10
σ=15	σ=20

Analysis Dictionary Learning – Results (3)

Experiment #3: denoising of natural images (with σ=5)

The following results were obtained by modifying the DL algorithm to improve the ROPP

Barbara



House

Peppers

Method	Barbara	House	Peppers
Fields of Experts (2009)	37.19 dB	38.23 dB	27.63 dB
Synthesis K-SVD	**38.08 dB**	**39.37 dB**	37.78 dB
Analysis K-SVD	37.75 dB	39.15 dB	**37.89 dB**

An Open Problem: How to “Inject” linear dependencies into the learned dictionary?

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.

Today we discussed

- The differences between the two models,
- A theoretical study of the THR algorithm, &
- Dictionary learning for the analysis model.

These slides and the relevant papers can be found in
http://www.cs.technion.ac.il/~elad

Thank you for your time,

and …

Thanks to the organizers:

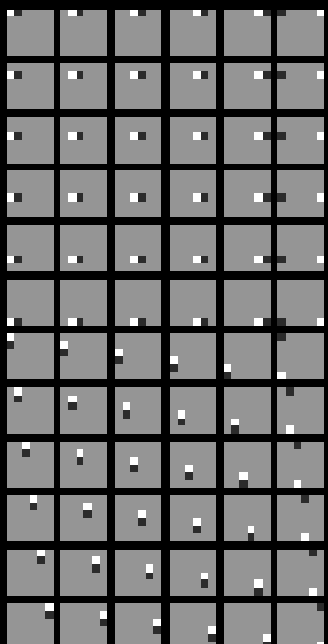
Gitta Kutyniok and Otmar Scherzer

Questions?

The Analysis Model – The Signature

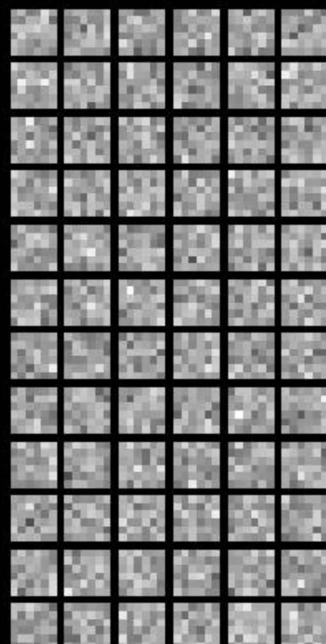
Consider two possible dictionaries:

$\boldsymbol{\Omega}_{\text{DIF}}$

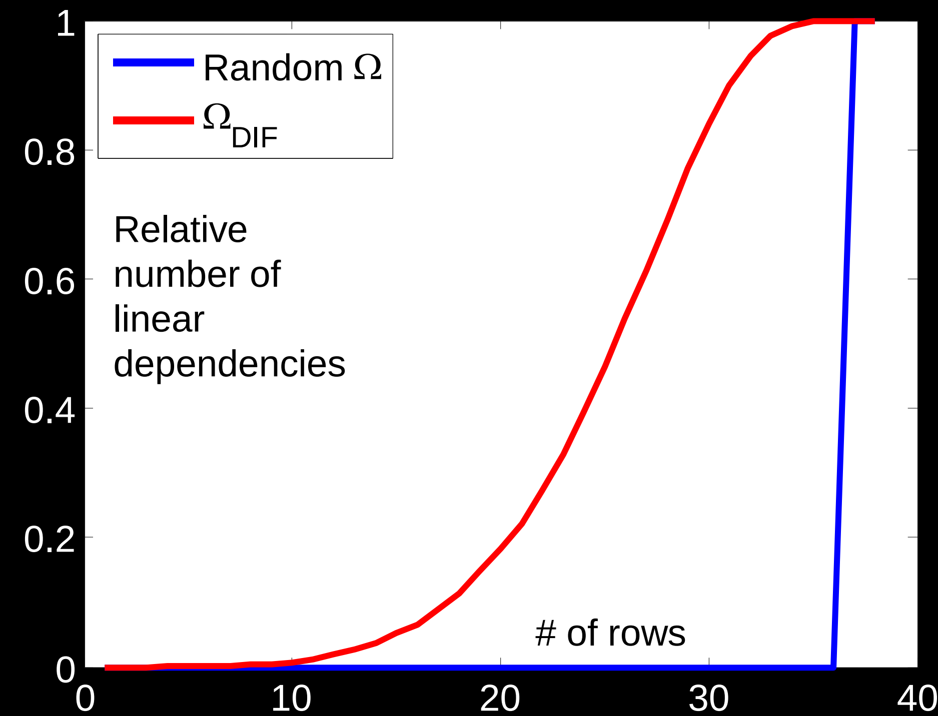


$\text{Spark}\left(\boldsymbol{\Omega}^T\right) = 4$

Random $\boldsymbol{\Omega}$



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



The Signature of a matrix is more informative than the Spark.

Is it enough?