Recent Results on the Co-Sparse Analysis Model*

Michael Elad

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

The research leading to these results has received funding from:



The European Council under the erc 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

NIPS Workshop – December 7th, 2012



Workshop on Analysis **Operator Learning vs. Dictionary Learning:** Fraternal Twins in Sparse Modeling



Informative Data \rightarrow 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.



Traffic info

Sparsity-Based Models





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



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 D∈ R^{d×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 <u>α</u>, 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: $\|\underline{\alpha}\|_0 = \mathbf{k} \ll \mathbf{d}$
- □ 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):

 $\underline{\mathbf{x}} \in \bigcup_{|\mathsf{T}|=k} \operatorname{span} \{ \mathbf{D}_{\mathsf{T}} \} \text{ where } \mathbf{D}_{\mathsf{T}} \underline{\alpha}_{\mathsf{T}} = \underline{\mathbf{x}}$ $\square \text{ This union contains } \binom{\mathsf{n}}{\mathsf{k}} \text{ subspaces, each of dimension k.}$



7

n

Dictionary

d

The Synthesis Model – Pursuit

□ Fundamental problem: Given the noisy measurements,

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

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

 \Box This can be posed as: $\underline{\hat{\alpha}} = \operatorname{ArgMin} \left\| \underline{y} - \mathbf{D} \underline{\alpha} \right\|_{2}^{2}$ s.t. $\left\| \underline{\alpha} \right\|_{0} = k \implies \underline{\hat{x}} = \mathbf{D} \underline{\hat{\alpha}}$

While this is a (NP-) hard problem, its approximated solution can be obtained by

- Use L₁ instead of L₀ (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



 $\min_{\mathbf{D},\mathbf{A}} \|\mathbf{D}\mathbf{A} - \mathbf{Y}\|$

s.t.
$$\forall j = 1, 2, \dots, N \|\underline{\alpha}_j\|_0 \leq k$$

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", to appear in ACHA, June 2011.



The Analysis Model – Basics

□ The analysis representation <u>z</u> is expected to be sparse

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

Co-sparsity: ℓ - the number of zeros in <u>z</u>.

Co-Support: Λ - the rows that are orthogonal to <u>x</u>

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

- □ This model puts an emphasis on the zeros in <u>z</u> for characterizing the signal, just like zero-crossings of wavelets used for defining a signal [Mallat (`91)].
- **Co-Rank:** Rank(Ω_{Λ}) $\leq \ell$ (strictly smaller if there are linear dependencies in Ω).



The Co-Sparse Analysis Model:

Recent Results By: Michael Elad



d

* spark $\{ \mathbf{\Omega}^{\mathsf{T}} \}$

Analysis Dictionary

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: $\mathbf{D}_{T} \underline{\alpha}_{T} = \underline{\mathbf{x}}$



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



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:	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: $\ell = d-1$ 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 $oldsymbol{\Omega}$ Case

What if spark(Ω^T) << d ?



... the number of subspaces is smaller.





The Analysis Model – Pursuit

□ Fundamental problem: Given the noisy measurements,

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

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

□ This goal can be posed as:

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

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

- □ We can approximate its solution by L₁ replacing L₀ (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 Ω. 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 $\hat{\mathbf{x}} = \operatorname{ArgMin}_{\underline{x},\Lambda} \left\| \underline{y} - \underline{x} \right\|_{2}^{2}$ s.t. $\left\{ \mathbf{\Omega}_{\Lambda} \underline{x} = 0 \quad \& \operatorname{Rank}(\mathbf{\Omega}_{\Lambda}) = d - r \right\}$





The Analysis Model – Backward Greedy



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

OMP

Synthesis vs. Analysis – Summary

C

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 II – THR Performance Revealing Important Dictionary Properties

1. T. Peleg and M. Elad, Performance Guarantees of the Thresholding Algorithm for the Co-Sparse Analysis Model, to appear in IEEE Transactions on Information Theory.



The Analysis-THR Algorithm

Analysis-THR aims to find an approximation for the problem $\underline{\hat{x}} = \operatorname{ArgMin}_{\underline{x},\Lambda} \|\underline{y} - \underline{x}\|_{2}^{2}$ s.t. $\{\mathbf{\Omega}_{\Lambda} \underline{x} = 0 \& \operatorname{Rank}(\mathbf{\Omega}_{\Lambda}) = d - r\}$

Compute
$$\underline{z} = |\Omega \underline{y}| \& \text{sort}$$

(increasing) $\rightarrow \{\gamma_k\}_{k=1}^p$

Stop condition?

Rank $(\Omega_{\Lambda_i}) = d - r$

No

 $i = i+1, \Lambda_i = \Lambda_{i-1} \cup \gamma_i$

 $i = i+1, \Lambda_i = \Lambda_{i-1} \cup \gamma_i$



The Restricted Ortho. Projection Property

$$\alpha_{\mathsf{r}} = \min_{\substack{\Lambda, j \mid \text{Rank}(\boldsymbol{\Omega}_{\Lambda}) = \mathsf{d} - \mathsf{r} \\ \text{and } j \notin \Lambda}} \left\| \left(\mathbf{I} - \boldsymbol{\Omega}_{\Lambda}^{\dagger} \boldsymbol{\Omega}_{\Lambda} \right) \underline{\mathbf{w}}_{j} \right\|_{2}$$





Theoretical Study of the THR Algorithm



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



Implications



ROPP



Ø

10

20

30

40

SNR[dB]

50

Ø 0.2



Recent Results

 $-\Omega_{\rm DIF}$

 $- \blacktriangle - \Omega_{RAND}$ ••**•**•• Ω_{MIX}

70

60

Part IV – Dictionaries Analysis Dictionary-Learning and Some Results

- 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", to appear in 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

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



Analysis Dictionary Learning – Goal

SynthesisMin
$$\mathbf{D}, \mathbf{A}$$
 $\|\mathbf{D}\mathbf{A} - \mathbf{Y}\|_{F}^{2}$ s.t. $\forall j = 1, 2, \dots, N$ $\|\underline{\alpha}_{j}\|_{0} \leq k$ AnalysisMin
 $\mathbf{\Omega}, \mathbf{X}$ $\|\mathbf{X} - \mathbf{Y}\|_{F}^{2}$ s.t. $\forall j = 1, 2, \dots, N$ $\|\mathbf{\Omega}\underline{\mathbf{X}}_{j}\|_{0} \leq \mathbf{p} - \ell$ Noisy ExamplesDenoised Signals are L₀ 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





$$\underset{\boldsymbol{\Omega},\mathbf{X}}{\text{Min}} \left\| \mathbf{X} - \mathbf{Y} \right\|_{F}^{2} \text{ s.t. } \forall \mathbf{j} = 1, 2, \dots, N \left\| \mathbf{\Omega} \underline{\mathbf{X}}_{\mathbf{j}} \right\|_{0} \leq \mathbf{p} - \ell$$

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

 $\left\| \underline{\hat{\mathbf{x}}}_{j} = \operatorname{ArgMin}_{\mathbf{x}} \left\| \underline{\mathbf{x}} - \underline{\mathbf{y}}_{j} \right\|_{2}^{2} \quad \text{s.t.} \left\| \mathbf{\Omega} \underline{\mathbf{x}} \right\|_{0} \le \mathbf{p} - \ell \right\}$

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

$$\mathbf{\tilde{V}}$$

Analysis K-SVD – Dictionary Update Stage



$$\underset{\boldsymbol{\Omega},\boldsymbol{\mathsf{x}}}{\text{Min}} \left\| \boldsymbol{\mathsf{X}} - \boldsymbol{\mathsf{Y}} \right\|_{F}^{2} \text{ s.t. } \forall j = 1, 2, \dots, N \left\| \boldsymbol{\Omega} \underline{\mathsf{x}}_{j} \right\|_{0} \leq p - \ell$$

- 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

 \Box We take 10,000 6×6 patches (+noise σ =5) to train on

Here is what we got (we promote sparse outcome):



Original Image



Patches used for training





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 $\sigma \text{=} \text{5}$







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: $\sigma=$



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

-	-			_		-
			23			
3	14					33
**					h	11
88	8		然		98	
		1	11	1	1	
	n		X			-
$\langle ' \rangle$	M	11	11		2	з,
2	1	ł		1	-	1
3	5		1	iii		





Peppers

Method	Barbara	House	Peppers
Fields of Experts	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

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

Today we discussed

These slides and the relevant papers can be found in

http://www.cs.technion.ac.il/~elad

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

So, what to do?



Thank you for your time, and ... Thanks to the organizers: Martin Kleinsteuber **Francis Bach Remi Gribonval** John Wright **Simon Hawe**

Questions?



The Analysis Model – The Signature

Consider two possible dictionaries:







The Signature of a matrix is more informative than the Spark. Is it enough?

