Analysis Operator Learning for Overcomplete Co-sparse Representations

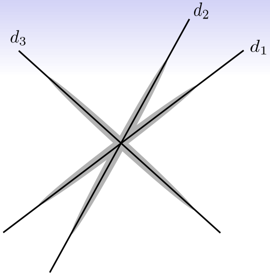
Mehrdad Yaghoobi*, Sangnam Nam, Remi Gribonval, and Mike E. Davies

European Signal Processing Conference,
EUSIPCO11 conference, Barcelona, Sep. 1st, 2011.

Outline:

Analysis Framework : An Introduction

- A low dimensional signal model.
- A special type of the union of subspaces signal model.
- Has many applications in, for example, denoising, compressed sensing and inverse problems to improve the overall performance.



Analysis Model

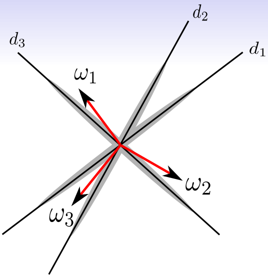
The signal $\mathbf{y}$ follows the model, if there exists a (linear) analysis operator $\Omega \in \mathbb{R}^{n \times m}$, $n \geq m$ that sparsifies $\mathbf{y}$,

$$\mathbf{z} = \Omega \mathbf{y}.$$

$\|\mathbf{z}\|_0 = n - p$, where $p > 0$ is called the co-sparsity of $\mathbf{y}$.

Analysis Operator Learning (AOL) Formulation

- A set of samples $\mathbf{Y} = [\mathbf{y}_1 \ \dots \ \mathbf{y}_i \ \dots \ \mathbf{y}_L]$ is given.
- The goal is to find a suitable analysis operator Ω such that $\|\Omega\mathbf{Y}\|_0$ is small.
- The objective is non-smooth $\Rightarrow$ not suitable for optimization with variational techniques.
- A relaxation is to select the sum of absolute values operator, *i.e.* $\|\cdot\|_1 = \sum_{ij} |\{\cdot\}_{ij}|$.



Formulation

The learned operator can be found by minimizing the sparsity promoting operator,

$$\min_{\Omega} \|\Omega\mathbf{Y}\|_1 \ \text{ s.t. } \ \Omega \in \mathcal{C}$$

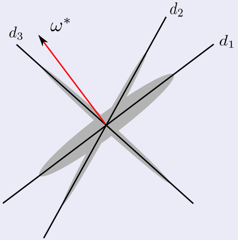
where $\mathcal{C}$ is a constraint, to exclude the trivial solutions, *e.g.* $\Omega = \mathbf{0}$.

Insufficient Constraints

Row norm constraints

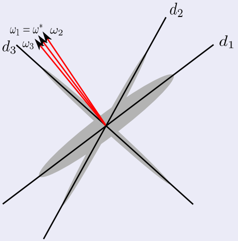
$$\forall i, \ \|\omega_i\|_2 = c$$

Rank one Ω_1 is found by repeating the best (almost) orthogonal direction ω^* to columns of $\mathbf{Y}$.



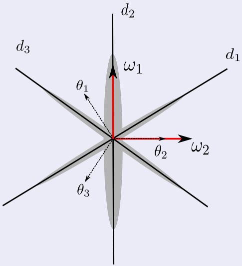
Row norm + full rank constraints

A randomly perturbed Ω from Ω_1 , *i.e.* row normalized $\Omega_1 + \mathbf{N}$, has a full rank and it is still not suitable.



Tight frame constraints

It resolves the issue in a complete setting. In the overcomplete cases, it includes zero-padded orthobases.



Proposed Constraint

Uniform Normalized Tight Frame (UNTF):

Definition: $\mathcal{C} = \{\Omega \in \mathbb{R}^{n \times m} : \Omega^T \Omega = \mathbf{I} \;\&\; \forall i \; \|\omega_i\|_2 = \sqrt{\frac{m}{n}}\}$

Pros and Cons:

- Zero-padded orthobases are not UNTF.
- Efficient methods exist to project onto the TF and the UN manifolds. However, there is no analytical way to find the projection onto the UNTF!
- There is no easy way to find the global optimum, using $\mathcal{C}$ as the constraint.

Motivation

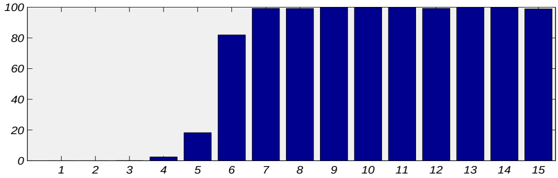
Minimization of a convex objective subject to the intersection of two manifolds $\Rightarrow$ a variant of projected subgradient algorithm is a good candidate.

Projected Subgradient Algorithm for AOL

1: **initialization:** $k = 1$, K_{max}, $\Omega^{[0]} = \mathbf{0}$, $\Omega^{[1]} = \Omega_{in}$, γ, $\epsilon \ll 1$
2: **while** $\epsilon \leq \|\Omega^{[k]} - \Omega^{[k-1]}\|_F$ and $k \leq K_{max}$ **do**
3: $\quad \Omega_G = \partial f(\Omega^{[k]})$
4: $\quad \Omega^{[k+1]} = \mathcal{P}_{UN}\left\{\mathcal{P}_{TF}\left\{\Omega^{[k]} - \gamma\Omega_G\right\}\right\}$
5: $\quad k = k + 1$
6: **end while**
7: **output:** $\Omega_{out} = \Omega^{[k-1]}$.

Exact Operator Recovery

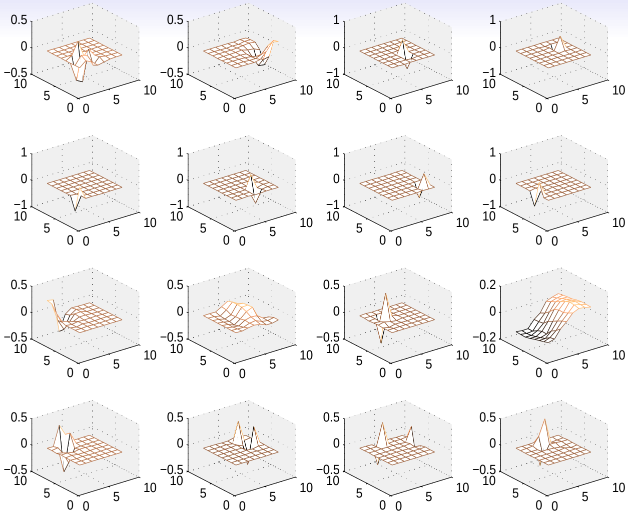
- A pseudo-random UNTF operator $\Omega_0 \in \mathbb{R}^{24\times 16}$ was used to generate $N = 768$ training samples.
- For each cosparsity p, a random normal vector was selected in the orthogonal complement space of p randomly selected rows of Ω_0.
- The simulation was started with a different pseudo-random admissible Ω_{in} and iterated 50000 times.
- The average recovery of the rows of Ω_0, for different cosparsities and 100 trials, is shown as a function of the cosparsity of the signals.



AOL for the Piecewise Constant Images

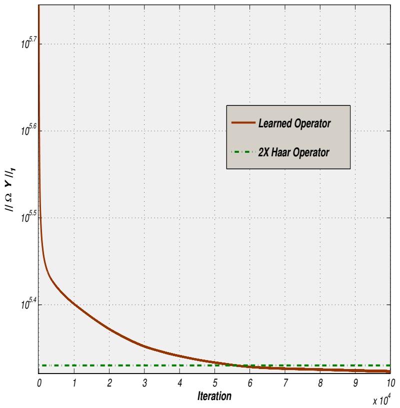
- Finding an Ω for the image patches of size 8×8.
- A 512×512 Shepp-Logan phantom image was used as the training image.
- $N = 16384$ image patches was randomly chosen from the training image.
- A pseudo-random UNTF operator $\Omega_0 \in \mathbb{R}^{128 \times 64}$ was used as the initial operator and the algorithm iterated 100,000 times!

AOL for the Piecewise Constant Images: the First 16 Learned Rows



A Comparison with Another UNTF

- Some rows have similarities with the finite difference operator rows → finite difference operator is not a UNTF.
- An alternative is to use (orthonormal) Haar wavelet as the mother basis to generate a union of orthobases.
- The union of a Haar wavelet and a circularly shifted version, was selected for comparison.



Do we get any better operator by initializing with the generated Haar based operator? NO. It is indeed a local minimum for the proposed AOL program.

Conclusion and Future Work

Conclusion:

- The proposed analysis operator learning technique showed promising results in the exact operator recovery.
- Although the proposed constraint may not be the most relevant constraint, it works well with the piecewise constant images.
- Although each iteration of the AOL algorithm is not computationally expensive, it converges very slow.

Future Work:

- Alternative constraints.
- Better optimization techniques.
- Deriving an explicit formulation for the recovery of an operator.
- Noise aware analysis operator learning.

Thanks for your attention.

Local Identifiability

Definition

Let an analysis operator Ω_0 exist that the set of given training samples $\mathbf{Y}$ are cosparse. It is called "locally identifiable", if it is a local optimum of the proposed optimization problem.

- An admissible point Ω_0 is a local minimum of $\|\Omega\mathbf{Y}\|_1$, if any perturbation of Ω_0 in the tangent space of UNTF, increases the objective.
- We can then show the local optimality of Ω_0 by showing $\Delta = 0$ is the only solution of,

$$\min_{\Delta} \|(\Omega_0 + \Delta)\mathbf{Y}\|_1 \ \text{s.t.} \quad \Delta^T\Omega_0 + \Omega_0^T\Delta = 0$$

$$\forall i \ \ \langle \omega_{0i}, \delta_i \rangle = 0.$$