Constrained Overcomplete Analysis Operator Learning for Cosparse Signal Modelling

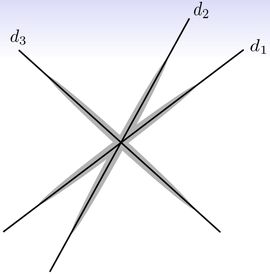
Mehrdad Yaghoobi,
Sangnam Nam, Remi Gribonval, and Mike E. Davies

IMA Conference on Numerical Linear Algebra and Optimisation,
Birmingham, UK. September 11th, 2012.

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

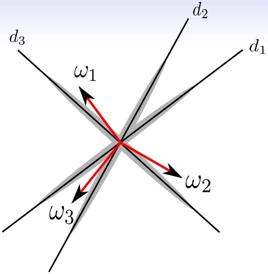
The signal $\mathbf{y}$ follows the model, if there exists a (linear) analysis operator $\Omega \in \mathbb{R}^{n \times m}$, $n \geq m$ that sparsifies $\mathbf{y}$,

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

$\|\mathbf{z}\|_0 = n - p$, 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 optimisation 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 minimising the sparsity promoting operator,

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

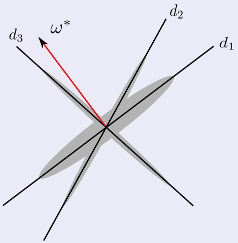
where $\mathcal{C}$ is a constraint, to exclude the trivial solutions, *e.g.* $\Omega = \mathbf{0}$.

Insufficient Constraints

Row norm constraints

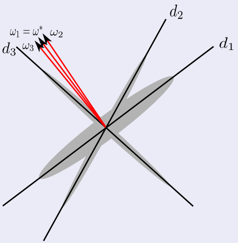
$$\forall i, \ \|\omega_i\|_2 = c$$

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



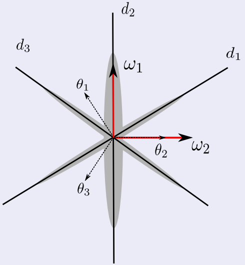
Row norm + full rank constraints

A randomly perturbed Ω from Ω_1 , *i.e.* row normalised $\Omega_1 + \mathbf{N}$, 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 Normalised 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.
- There exist some practical methods 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.

Projected Subgradient Algorithm for AOL

Motivation

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

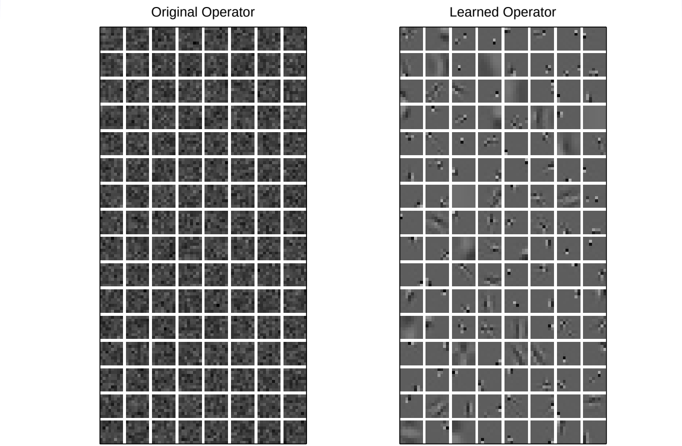
Projected Subgradient Type Algorithm for AOL

1: **initialisation:** $k = 1$, 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:** $\Omega_{out} = \Omega^{[k-1]}$.

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



Issues with the Projected Subgradinet Algorithm: Some Proposed Relaxations

- No analytical way to project onto UNTF $\rightarrow$ **no convergence** proof.
- Projection onto TF needs a full SVD calculation $\rightarrow$ **expensive implementation** and non-scalable algorithm.
- ℓ_1 term is not differentiable $\rightarrow$ **slow convergence** of the projected subgradient algorithm.

Relaxed AOL Formulation

1. **Relaxing the objective:** using a convex, but differentiable sparsity constraint $g(\Omega \mathbf{Y})$, where g is an entrywise function defined as,

$$g(x) = |x| - s \ln(1 + |x|/s), \; s \in \mathbb{R}^+$$

2. **Relaxing the constraint:** using quartic constraints $\|\Omega^T \Omega - \mathbf{I}\|_F^2 \leq \epsilon_{TF}$ and $\|\omega_i^T \omega_i - \frac{m}{n}\|_2^2 \leq \epsilon_{UN}, \quad \forall i \in [1, n]$

Relaxed Analysis Operator Learning

Relaxed Analysis Operator Learing Formulation

An unconstrained objective is generate by using two Lagrange multipliers γ and λ:

$$f(\mathbf{\Omega}) = g(\mathbf{\Omega Y}) + \frac{\gamma}{4}\|\mathbf{\Omega}^T\mathbf{\Omega} - \mathbf{I}\|_F^2 + \frac{\lambda}{4}\sum_i \left\{\|\omega_i^T\omega_i - \frac{m}{n}\|_2^2\right\}.$$

- f is differentiable and it is also convex, if we restrict its domain to $\mathcal{C}_c = \{\mathbf{\Omega} : \mathbf{\Omega}^T\mathbf{\Omega} - \mathbf{I} \succeq \mathbf{0},\ \forall i,\ (\omega_i^T\omega_i - \frac{m}{n}) \geq 0\}$.

Gradient Descent Algorithm for AOL

A variable step-size gradient descent, with line search, can be used to minimise $f(\mathbf{\Omega})$, where the gradient of f can easily be found by:

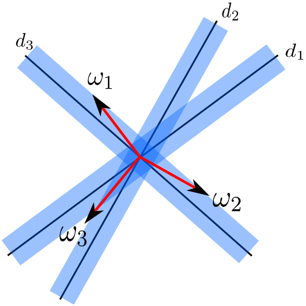
$$\nabla f = \left[\frac{\mathbf{Z}_{i,j}}{s + |\mathbf{Z}_{i,j}|}\right]_{i,j}\mathbf{Y}^T + \gamma\left(\mathbf{\Omega\Omega}^T - \mathbf{I}\right)\mathbf{\Omega} + \lambda\left[\omega_i\left(\omega_i^T\omega_i - \frac{m}{n}\right)\right]_i^T$$

$$\mathbf{Z} := \mathbf{\Omega Y}$$

Noise Aware Analysis Operator Learning

Approximately Cosparse Exemplars

- Training data $\mathbf{Y}$ is approximately cosparse, $\mathbf{Y} = \mathbf{Y}_c + \mathbf{N}$, where $\mathbf{N}$ is **noise** or **model mismatch** and $\mathbf{Y}_c$ is cosparse.
- The goal is to find an operator $\mathbf{\Omega}$, such that $\mathbf{\Omega Y}_c$ has many zeros.
- The issue is that **we do not know $\mathbf{Y}_c$ precisely!** A solution is to somehow approximate it.
- This is indeed very similar to the **dictionary learning** problem, where we do not know the sparse coefficients.



Noise Aware Analysis Operator Learning: Formulation and Algorithm

Noise Aware Analysis Operator Learning

$$\min_{\mathbf{\Omega},\widehat{\mathbf{Y}}} \|\mathbf{\Omega}\widehat{\mathbf{Y}}\|_1 + \frac{\theta}{2}\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F^2 \quad \text{s.t.} \quad \mathbf{\Omega} \in \mathcal{C}.$$

Solving by alternating minimisation technique.

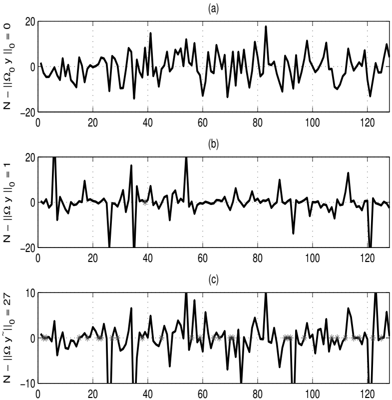
- Optimisation based on $\mathbf{\Omega}$: similar to noise-less AOL.
- Optimisation based on $\widehat{\mathbf{Y}}$: a convex program. $\rightarrow$ Douglas-Rachford Splitting (DRS) technique was used to efficiently solve the program.
- Algorithm usually converges after a few number of alternating minimisations.
- For the optimisation base on $\widehat{\mathbf{Y}}$, the ℓ_1 penalty can be relaxed, similar to the operator update step, and the new convex program can be solved using a gradient descent algorithm with a line search.

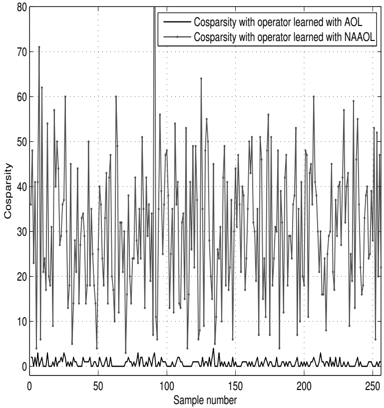
An Operator for the Face Images: Setting

- Learning an Ω for the image face patches from the Yale face database.
- $L = 16384$, 8×8 image patches were randomly selected from different faces.

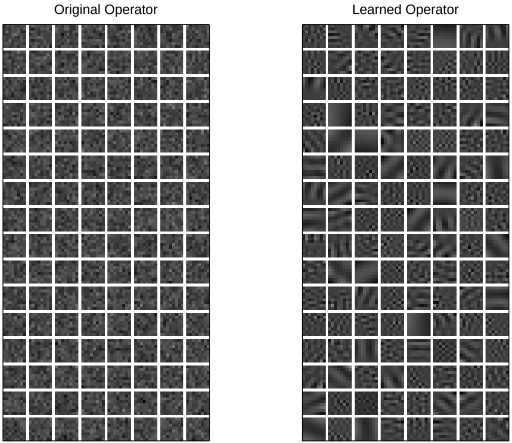
An Operator for Face Images: Cosparsity Comparison

- The analysis coefficients $\mathbf{z} = \mathbf{\Omega y}$ and cosparsities were calculated, using $\mathbf{\Omega}_0$, $\mathbf{\Omega}_{AOL}$ and $\mathbf{\Omega}_{NAAOL}$.



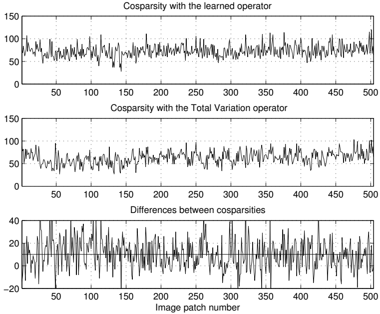


Learned Operator



Face Images Denoising: TV v.s. Learned Operator

- TV operator for comparison.
- Two different regularisation parameters, $\lambda = 0.3$ & 0.1 .



Conclusion and Future Work

Conclusion:

- The constrained analysis operator learning is a useful technique to find a suitable analysis operator.
- The proposed constraint can be relaxed to reduce the complexity of the optimisation algorithm, while including some **approximately UNTF** operators.
- The simulation results emphasis on the fact that we should use the correct analysis operator, i.e. TV or oscillatory operators.
- The convergence of the relaxed AOL is guaranteed, as its objective has a bounded curvature and its sublevel set is conpact.

Future Work:

- Investigating the **local identifiability** of operators in this framework.
- More investigations on the **structures** of the learned operators.

Thanks for your attention.