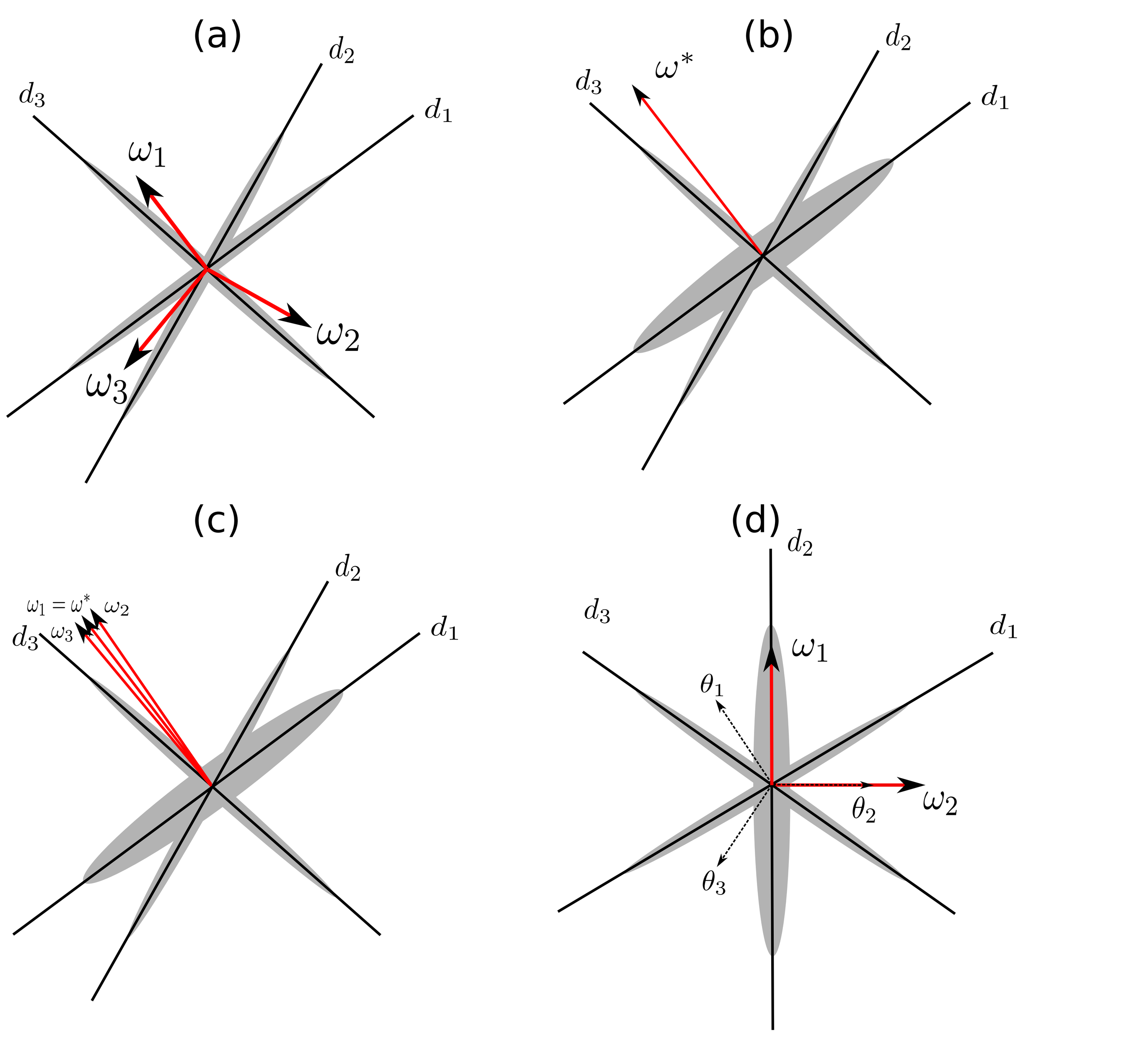
Relaxed Analysis Operator Learning

Mehrdad Yaghoobi and Mike Davies

Institute for Digital Communications (IDCom), the University of Edinburgh, EH9 3JL, UK
yaghoobi@ieee.org, mike.davies@ed.ac.uk

Abstract — *The problem of analysis operator learning can be formulated as a constrained optimisation problem. This problem has been approximately solved using projected gradient or geometric gradient descent methods. We will propose a relaxation for the constrained analysis operator learning in this poster. The relaxation has been suggested here to, a) reduce the computational complexity of the optimisation and b) include a larger set of admissible operators. We will show here that an appropriate relaxation can be useful in presenting a projection-free optimisation algorithm, while preventing the problem to become ill-posed. The relaxed optimisation objective is not convex and it is thus not always possible to find the global optimum. However, when a rich set of training samples are given, we empirically show that the desired synthetic analysis operator is recoverable, using the introduced gradient descent or conjugate gradient algorithms.*



Constrained Analysis Operator Learning Formulation

The aim of analysis operator learning is to find an operator $\mathbf{\Omega} \in \mathbb{R}^{a\times n}$, adapted to a set of observations of the signals $\mathbf{Y} = [\mathbf{y}_i] \in \mathbb{R}^{n\times l}, \mathbf{y}_i = x_i + n_i$, where $\mathbf{\Omega Y}$ is sparse [1], see figure (a). In this setting, $\mathbf{Y}$ is called (*approximately*) cosparse. With ℓ_1 sparsity measure, a formulation for finding $\mathbf{\Omega}$ is as follows,

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

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

- Row norm (UN) constraints: the optimum solution is obtained by repeating the best row ω^*, i.e., $\mathbf{\Omega}_1^* := [\omega_i = \omega^*]_{i\in[1,a]}^T$, see figure (b).
- Row norm + full rank constraints: the optimum solutions have very small condition numbers, *e.g.* $\mathcal{P}_{\mathcal{C}_F}\{\epsilon \mathbf{A} + \mathbf{\Omega}_1^*\}$, where $\mathcal{P}_{\mathcal{C}_F}$, $\mathbf{A}$ and ϵ respectively are row normalisation, a random Gaussian matrix and a very small constant, see figure (c).
- Tight frame (TF) constraint: the optimum solutions are the zero-padded bases, see figure (d).
- Proposed constraint: Uniform Normalised Tight Frame (UNTF):

$$\mathcal{C} = \{\mathbf{\Omega} \in \mathbb{R}^{a\times n} \ : \ \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{I}, \ \forall i \ \|\omega_i\|_2 = \sqrt{\frac{a}{n}}\}.$$

- Suggested solver: Alternating Minimisation with a Projected Subgradient type algorithm for updating $\mathbf{\Omega}$.
- **Issues:**
  1. No analytical way to project onto UNTF → **no convergence** proof.
  2. Projection onto TF needs a full SVD calculation → **expensive implementation** and non-scalable algorithm.
  3. ℓ_1 term is not differentiable → **slow convergence** of the projected subgradient algorithm.

Relaxed Analysis Operator Learning

Suggested Relaxation and a New Formulation

- **Relaxing the objective:** using a convex, but continuously differentiable, sparsity constraint $g(\mathbf{\Omega Y}) = \sum_{i,j} \zeta(\{\mathbf{\Omega Y}\}_{i,j})$, for ζ defined as $\zeta(x) = |x| - s\ln(1+|x|/s), \ s \in \mathbb{R}^+, s \ll 1$ [2].
- **Relaxing the constraint:** using quartic constraints $\|\mathbf{\Omega}^T\mathbf{\Omega} - \mathbf{I}\|_F^2 \le \epsilon_{TF}$ and $\left(\omega_i^T\omega_i - \frac{m}{n}\right)^2 \le \epsilon_{UN}, \quad \forall i \in [1,n]$.
- **Relaxed AOL Formulation:** An unconstrained objective is generate by using two Lagrange multipliers γ and θ [3]:

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

Relaxed Analysis Operator Learning with Alternating Minimisation

initialisation: $\mathbf{\Omega}^{[0]}$, $\mathbf{X}^{[0]} = \mathbf{Y}$, $k = 0$,
while not converged **do**

$$\mathbf{\Omega}^{[k+1]} = \operatorname{argmin}_{\mathbf{\Omega}\in\mathcal{C}} g(\mathbf{\Omega X}^{[k]}) + \tfrac{\gamma}{4}\|\mathbf{\Omega}^T\mathbf{\Omega} - \mathbf{I}\|_F^2 + \tfrac{\theta}{4}\textstyle\sum_i \left(\omega_i^T\omega_i - \frac{m}{n}\right)^2,$$

(using a Gradient Descent or a Conjugate Gradient method)

$$\mathbf{X}^{[k+1]} = \operatorname{argmin}_{\mathbf{X}} g(\mathbf{\Omega}^{[k+1]}\mathbf{X}) + \tfrac{\lambda}{2}\|\mathbf{Y}-\mathbf{X}\|_F^2$$

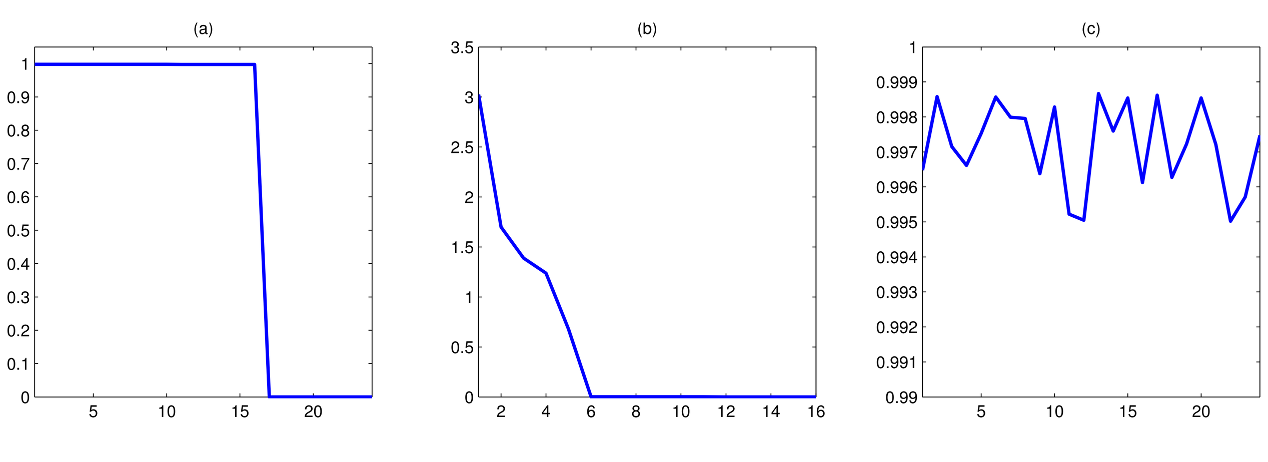
(Convex program! Using a parameter splitting and Augmented Lagrangian MM)

$k = k+1$
end while.

Simulations and Summary

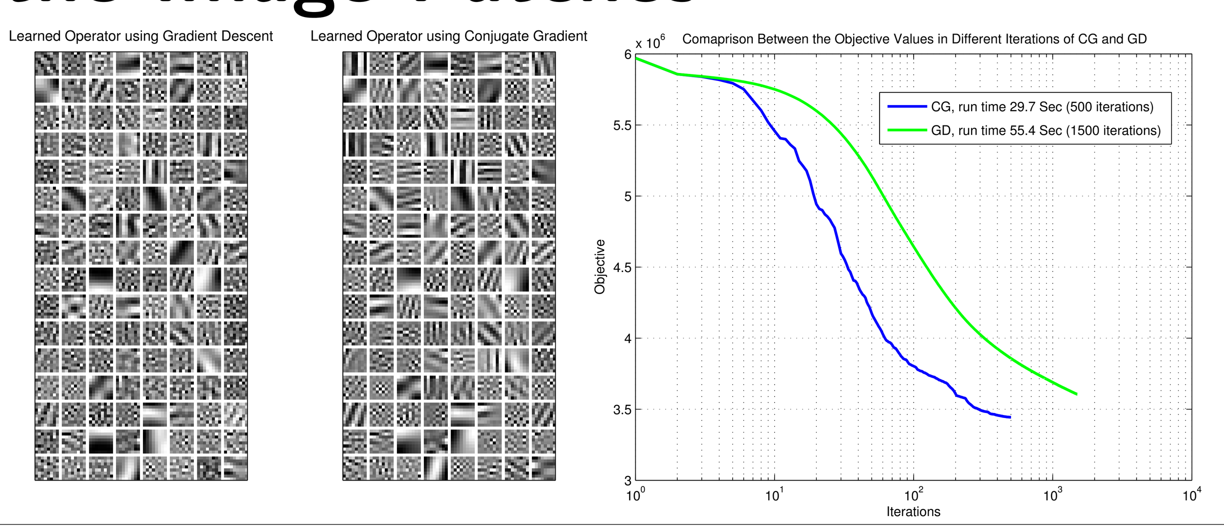
Relaxed Synthetic Operator Recovery

- Learning $\Omega \in \mathbb{R}^{24\times 16}$ from $\mathbf{Y} \in \mathbb{R}^{16\times 576}$, $q = 10$ cosparse exemplars with respect to a reference UNTF operator $\mathbf{\Omega}_0$.
- Sorted ℓ_2 norms of the rows of learned operator using a **TF** penalty $(\theta = 0)$ (left).
  ⇒ The learned operator is a zero-padded orthogonal basis.
- Singular values of the learned operator using a **UN** penalty $(\gamma = 0)$ (middle).
  ⇒ The learned operator is a rank deficient operator.
- Normalised inner-products between the rows of the synthetic reference operator $\mathbf{\Omega}_0$ and the corresponding rows in the learned operator using a **UNTF** penalty (right).
  ⇒ The reference operator is approximately recovered.



Relaxed AOL for the Image Patches

- Learning an operator for the 8×8 image patches, using a set of standard images ('Barbara', 'boat', 'Lena', 'fingerprint', 'flinstones', 'house' and 'peppers') as the image exemplars. $\mathbf{\Omega}$ was two times overcomplete and we used 4192 training samples.
- A noise-less setting has been used for learning, *i.e.* $\mathbf{X} = \mathbf{Y}$.
- Gradient Descent and (Nonlinear) Conjugate Gradient methods, with a line-search, have been used for the optimisation based upon $\mathbf{\Omega}$.
- Although each iteration of CG is computationally more costly, it converges faster in practice. The total simulation time for CG was less than GD.



Summary

- The proposed relaxed analysis operator learning helps us to have faster learning algorithm, while allowing a larger set of admissible operators.
- UN and TF penalties are necessary for a reliable operator learning. Over-relaxation of these constraints may lead us to some trivial solutions.
- The relaxed formulation allows us to use CG or GD methods for the operator update stage of the AOL.
- The proposed relaxed AOL algorithm converges significantly faster, as we can now learn an operator for the image patches in less than a minute, using a standard workstation!

Acknowledgement

This work was supported by EU FP7, FET-Open grant number 225913 and EPSRC grant EP/J015180/1. MED acknowledges support of his position from the Scottish Funding Council and their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership in Engineering and Mathematics.

[1] M. Yaghoobi, S. Nam, R. Gribonval and M. E. Davies, *"Constrained Overcomplete Analysis Operator Learning for Cosparse Signal Modelling"*, submitted, `http://arxiv.org/abs/1205.4133`.
[2] M. Elad, B. Matalon and M. Zibulevsky *"Coordinate and Subspace Optimization Methods for Linear Least Squares with Non-quadratic Regularization"*, ACHA, vol.23, num. 3, pp 346-367, 2007.
[3] M. Yaghoobi and M. E. Davies, *"Relax and Accelerate: A Generalised Analysis Operator Learning Algorithm"*, in preparation.