NOISE AWARE ANALYSIS OPERATOR LEARNING FOR APPROXIMATELY COSPARSE SIGNALS

Mehrdad Yaghoobi[†], Sangnam Nam[‡], Rémi Gribonval[‡], and Mike E. Davies[†]

[†] Institute for Digital Communications (IDCom), the University of Edinburgh, EH9 3JL, UK

[‡] INRIA, Centre Inria Rennes - Bretagne Atlantique, 35042 Rennes Cedex, France

yaghoobi@ieee.org, sangnam.nam@inria.fr, remi.gribonval@inria.fr, mike.davies@ed.ac.uk

ABSTRACT

This paper investigates analysis operator learning for the recently introduced cosparse signal model that is a natural analysis complement to the more traditional sparse signal model. Previous work on such analysis operator learning has relied on access to a set of clean training samples. Here we introduce a new learning framework which can use training data which is *corrupted by noise* and/or is only *approximately cosparse*. The new model assumes that a p-cosparse signal exists in an epsilon neighborhood of each data point. The operator is assumed to be uniformly normalized tight frame (UNTF) to exclude some trivial operators. In this setting, a bi-level optimization algorithm is introduced to learn a suitable analysis operator.

Index Terms— Sparse Approximation, Analysis Framework, Cosparse Signal Model, Analysis Operator Learning, Douglas-Rachford Splitting

1. INTRODUCTION

The linear sparse model is a suitable low-dimensional model for many natural signals. A discrete signal $\mathbf{y} \in \mathbb{R}^m$ is called sparse when we can represent $\mathbf{y}$ using a vector $\mathbf{x} \in \mathbb{R}^n$, with few non-zero elements, and a linear overcomplete transform $\mathbf{\Phi} \in \mathbb{R}^{m \times q}$ $m < q$, called the *dictionary*, as follows

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}. \qquad (1)$$

This has also been called the *synthesis* parsimony model [1]. As $\mathbf{\Phi}$ is overcomplete, the model is underdetermined and for many reasons a convex sparsity penalty, *i.e.* $\ell_1(\mathbf{x}) := \|\mathbf{x}\|_1 = \sum |x_i|$, has been used to recover the sparse coefficients $\mathbf{x}$.

The formulation (1) is an ideal model which often does not precisely fit to the real world signals either because the signal is noisy or is only *approximately* sparse. A standard approach to compensate the effect of noise and/or model mismatch, is to assume that the noise is Gaussian and the signal is within an ϵ neighborhood of the observation. The linear constraint can then be relaxed to give the following convex program, called Basis Pursuit Denoising (BPDN) [2],

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s. t. } \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|^2 \le \epsilon. \qquad (2)$$

A better approximation is sometimes possible by subsequently projecting $\mathbf{y}$ onto the span of the columns of $\mathbf{\Phi}$ in the support of $\mathbf{x}$: a technique known as debiasing.

1.1. Dictionary Learning

In many instances the appropriate dictionary $\mathbf{\Phi}$ is not known and must be learnt from training data $\mathbf{Y} = [\mathbf{y}_i]_{i \in \mathcal{I}} \in \mathbb{R}^{m \times L}$ [3]. One approach to synthesis dictionary learning is to solve (2) in a matrix form and apply a constraint on the dictionary to exclude trivial solutions, as follows:

$$\min_{\mathbf{X}, \mathbf{\Phi} \in \mathcal{D}} \|\mathbf{X}\|_1 \text{ s. t. } \|\mathbf{Y} - \mathbf{\Phi}\mathbf{X}\|^2 \le \epsilon. \qquad (3)$$

Such an optimization has been dealt with in [4]. As $\mathbf{\Phi}$ does not appear in the objective, many bi-level gradient based methods have difficulties with this formulation and they use a Lagrangian multiplier version of (3), see for example [5–8] and references therein.

1.2. Cosparse and Approximately Cosparse Models

An alternative low-dimensional signal model is the *analysis* parsimony model [1]. In this setting, there exists an overcomplete analysis linear operator $\mathbf{\Omega} \in \mathbb{R}^{N \times m}$, which maps the signals $\mathbf{y} \in \mathbb{R}^m$ to low-dimensional subspaces of a higher-dimensional *analysis space* $\mathbb{R}^N$ [9]. In other words, $\mathbf{z} = \mathbf{\Omega}\mathbf{y}$, where $\mathbf{z}$ has many zero elements, *i.e.* $\|\mathbf{z}\|_0 = N - p$, where p is the number of zero components. The integer p, which is called the cosparsity of $\mathbf{y}$ [9], has an important role in the analysis of signal recovery in linear inverse problems.

Similar to the synthesis framework, most of the real world signals are not *exactly* cosparse: some noise corrupts the observations and there are also some model mismatches. It suggests an approximate cosparse model like $\mathbf{z} \approx \mathbf{\Omega}\mathbf{y}$ where $\mathbf{z}$ is cosparse and the accuracy of the approximation depends on the noise/model mismatch level. In other words, the model assumes that some cosparse signal $\widetilde{\mathbf{y}}$, such that $\mathbf{z} := \mathbf{\Omega}\widetilde{\mathbf{y}}$ has many zero elements, is within an ϵ neighborhood of $\mathbf{y}$, *i.e.* $\|\widetilde{\mathbf{y}} - \mathbf{y}\|^2 \le \epsilon$ [10].

1.3. Analysis Operator Learning (AOL)

Unlike the synthesis framework, there is very little work on adapting the analysis model to a set of sample data $\mathbf{Y}$. Recent works that have begun to investigate this problem are [11–13]. The main difficulty in the analysis operator learning is avoiding trivial solutions emerging from the learning process. Two ad hoc techniques [11] and [13] have been presented that do not directly exclude trivial solutions but appear to converge to some acceptable fixed points when the algorithm is suitably initialized. Alternatively a constrained optimization approach has been presented in [12] that avoids most trivial solutions. One restriction of the AOL method of [12] is that it requires noiseless co-sparse data that exactly follow the model $\mathbf{z} = \mathbf{\Omega}\mathbf{y}$. Removing this restriction is the goal of the present paper.

2. CONSTRAINED ANALYSIS OPERATOR LEARNING

Cosparse signals, by definition, have a large number of zero elements in the analysis space. One proposal for adapting an operator to a set of signals $\mathbf{Y}$ is to maximize the number of zeros of $\mathbf{\Omega Y}$ or equivalently minimized the number of non-zero elements. The algorithm presented in [12] uses this approach along with the ℓ_1 convex surrogate function for measuring sparsity. If we do not apply a constraint on the operator then trivial solutions such as $\mathbf{\Omega} = \mathbf{0}$ will invariably exist. It is explained in [12] why some simple constraints like row norm, row norm plus full-rank and the tight frame constraints do not resolve the issue. Instead, a constraint $\mathcal{C}$ called Uniform Normalized Tight Frame (UNTF) was proposed to simply exclude such trivial solutions. The constrained AOL can then be reformulated as,

$$\min_{\mathbf{\Omega}} \|\mathbf{\Omega Y}\|_1 \ \text{s.t.} \ \mathbf{\Omega} \in \mathcal{C} \tag{4}$$

where,

$$\mathcal{C} = \{\mathbf{\Omega} \in \mathbb{R}^{N \times m} : \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{I}, \ \forall i \ \|\omega_i\|_2 = c\}, \tag{5}$$

and ω_i is the ith row of $\mathbf{\Omega}$. As $\|\cdot\|_1$ is continuous and $\mathcal{C}$ is not convex, (4) may have many separate *local optima*. A projected sub-gradient iterative method is presented in [12] to find a local minimum. For convenience the procedure is summarized in Algorithm 1, where $\overline{\text{sgn}}$ is an entrywise version (*i.e.* $\{\overline{\text{sgn}}(\mathbf{A})\}_{ij} = \overline{\text{sgn}}(\mathbf{A}_{ij})$) of the sign function which is defined by $\overline{\text{sgn}}(a) = \text{sgn}(a)$ when $a \neq 0$ and $\overline{\text{sgn}}(a) = [-1, 1]$ otherwise. The notations $\mathcal{P}_{UN}$ (resp. $\mathcal{P}_{TF}$) refer to projections onto the set of uniformly normalized (resp. tight frame) matrices, cf [12].

The simulations in [12] show that when the signals are cosparse enough, *i.e.* many zero elements in each $\mathbf{\Omega y}_i$, generic analysis operators can be recovered using this framework.

2.1. Noise Aware AOL (NAAOL) Formulation

When the signal $\mathbf{y}$ is noise corrupted or only approximately cosparse an approximate signal $\widehat{\mathbf{y}}$, also called the denoised signal, can be found using an analysis denoising formulation as follows,

$$\min_{\widehat{\mathbf{y}}} \|\mathbf{\Omega}\widehat{\mathbf{y}}\|_1 + \frac{\lambda}{2}\|\widehat{\mathbf{y}} - \mathbf{y}\|_2^2, \tag{6}$$

where parameter λ regulates the level of noise/model missmatch. Such a noisy formulation motivates us to further extend the AOL framework to consider noise and approximate cosparsity. This leads to an optimization problem with two parameters $\widehat{\mathbf{Y}}$ and $\mathbf{\Omega}$ as follows,

$$\min_{\mathbf{\Omega}, \widehat{\mathbf{Y}}} \|\mathbf{\Omega}\widehat{\mathbf{Y}}\|_1 + \frac{\lambda}{2}\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F^2 \quad \text{s.t.} \quad \mathbf{\Omega} \in \mathcal{C}. \tag{7}$$

We use the constraint $\mathcal{C}$ defined in (5), since the new formulation still has the same set of trivial solutions. Problem (7) is a non-convex optimization problem on the intersection of two manifolds, which is generally difficult to solve. A practical algorithm to find a *local minimum* of (7) is presented next.

2.2. Noise Aware AOL Algorithm

It is clear that the NAAOL problem can be simplified to the AOL problem, if $\widehat{\mathbf{Y}}$ is known. Alternatively, if we know the optimal operator $\mathbf{\Omega}$, $\widehat{\mathbf{Y}}$ can easily be found using a convex program. However, the size of the problem makes such an optimization challenging. The main difficulties are that ℓ_1 is not differentiable and the existence

---

Algorithm 1 Projected Subgradient for AOL (Operator Update)

Input: $\mathbf{Y}$, $K_{\max}$, Ω_{in}, stepsize η, threshold $\epsilon \ll 1$
1: **initialization:** $k = 1$, $\Omega^{[0]} = \mathbf{0}$, $\Omega^{[1]} = \Omega_{in}$
2: **while** $\epsilon \leq \|\Omega^{[k]} - \Omega^{[k-1]}\|_F$ and $k \leq K_{\max}$ **do**
3:    $\Omega_G = \partial f(\Omega^{[k]}) = \overline{\text{sgn}}(\Omega^{[k]}\mathbf{Y})\mathbf{Y}^T$
4:    $\Omega^{[k+1]} = \mathcal{P}_{UN}\left\{\mathcal{P}_{TF}\left\{\Omega^{[k]} - \eta\,\Omega_G\right\}\right\}$
5:    $k = k + 1$
6: **end while**
7: **output:** $\Omega_{out} = \Omega^{[k-1]}$

---

of $\mathbf{\Omega}$ inside the ℓ_1 penalty does not allow us to use conventional methods for solving problems like (2). We here use the Douglas-Rachford Splitting (DRS) technique to efficiently solve (7) when $\mathbf{\Omega}$ is fixed: see [14] for a brief overview. This technique has been used for approximation with the Total Variation (TV) penalty and analysis sparse approximation in [15][1]. We here only present a simple version of the DRS technique, tailored for this problem. A new parameter $\widehat{\mathbf{Z}} := \mathbf{\Omega}\widehat{\mathbf{Y}}$ is defined, such that the optimization with respect to $\widehat{\mathbf{Z}}$ is easier. Now the problem is a constrained (but still convex) program with two parameters $\widehat{\mathbf{Z}}$ and $\widehat{\mathbf{Y}}$ as follows,

$$\min_{\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}} \|\widehat{\mathbf{Z}}\|_1 + \frac{\lambda}{2}\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F^2 \ \text{s.t.} \ \widehat{\mathbf{Z}} = \mathbf{\Omega}\widehat{\mathbf{Y}}. \tag{8}$$

The Augmented Lagrangian (AL) [17] method is applied to solve (8). In the Lagrangian multiplier method we use the dual parameter $\mathbf{B} \in \mathbb{R}^{N \times L}$ and add a penalty term, $\left\langle \mathbf{B}, \mathbf{\Omega}\widehat{\mathbf{Y}} - \widehat{\mathbf{Z}} \right\rangle$. In the AL, we also add an extra quadratic penalty related to the constraint and derive the new objective $\psi(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \mathbf{B})$ as follows,

$\psi(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \mathbf{B})$
$$= \|\widehat{\mathbf{Z}}\|_1 + \frac{\lambda}{2}\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F^2 + \gamma\left\langle \mathbf{B}, \mathbf{\Omega}\widehat{\mathbf{Y}} - \widehat{\mathbf{Z}} \right\rangle + \frac{\gamma}{2}\|\mathbf{\Omega}\widehat{\mathbf{Y}} - \widehat{\mathbf{Z}}\|_F^2$$
$$= \|\widehat{\mathbf{Z}}\|_1 + \frac{\lambda}{2}\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F^2 + \frac{\gamma}{2}\|\mathbf{B} + \mathbf{\Omega}\widehat{\mathbf{Y}} - \widehat{\mathbf{Z}}\|_F^2 - \frac{\gamma}{2}\|\mathbf{B}\|_F^2$$

where $0 < \gamma \in \mathbb{R}^+$ is a constant parameter. According to the duality property, the solution of $\max_{\mathbf{B}} \min_{\widehat{\mathbf{Z}}, \widehat{\mathbf{Y}}} \psi(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \mathbf{B})$ coincides with the solution of (8). Using the DRS method, we iteratively optimize a convex/concave surrogate objective $\psi_s(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \mathbf{B}, \mathbf{B}^{[n]}) = \psi(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \mathbf{B}) - \|\mathbf{B} - \mathbf{B}^{[n]}\|_F^2$, where $\mathbf{B}^{[n]}$ is the current estimation of $\mathbf{B}$. The fixed points of the iterative updates of $\psi_s(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \mathbf{B}, \mathbf{B}^{[n]})$ are the same as $\psi(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \mathbf{B})$, as the extra term $\|\mathbf{B} - \mathbf{B}^{[n]}\|_F^2$ vanishes in any fixed points. $\psi_s(\widehat{\mathbf{Y}}, \widehat{\mathbf{Z}}, \mathbf{B}, \mathbf{B}^{[n]})$ is convex with respect to $\widehat{\mathbf{Z}}$ and $\widehat{\mathbf{Y}}$ and concave with respect to $\mathbf{B}$. We can thus iteratively update each of the parameters, while keeping the rest fixed, as follows,

$$\widehat{\mathbf{Y}}^{[n+1]} = (\lambda\mathbf{I} + \gamma\,\mathbf{\Omega}^T\mathbf{\Omega})^{-1}(\lambda\mathbf{Y} + \gamma\,\mathbf{\Omega}^T(\widehat{\mathbf{Z}}^{[n]} - \mathbf{B}^{[n]}))$$
$$\widehat{\mathbf{Z}}^{[n+1]} = \mathcal{S}_{\frac{1}{\gamma}}\left\{\mathbf{\Omega}\widehat{\mathbf{Y}}^{[n+1]} + \mathbf{B}^{[n]}\right\} \tag{9}$$
$$\mathbf{B}^{[n+1]} = \mathbf{B}^{[n]} + (\mathbf{\Omega}\widehat{\mathbf{Y}}^{[n+1]} - \widehat{\mathbf{Z}}^{[n+1]}),$$

where $\mathcal{S}_\alpha$, with an $\alpha > 0$, is the entrywise soft-threshold operator defined by $\mathcal{S}_\alpha(\beta) = \beta - \alpha\,\text{sgn}(\beta)$ if $|\beta| \geq \alpha$ and 0 otherwise.

---

[1] [15] derives the formulation by incorporating the Bregman distance. However, it has been shown that the new method, called Alternating Bregman Splitting method, is the same as DRS applied to the dual problems [16].
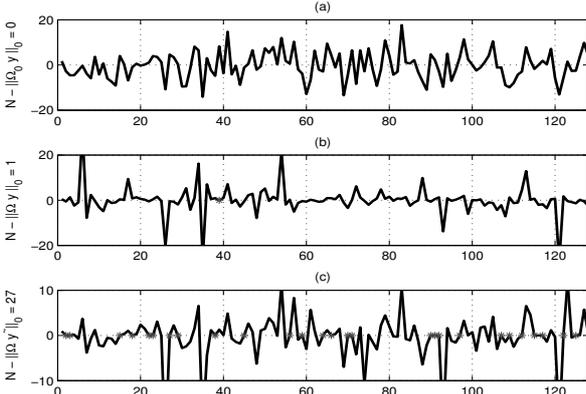
Algorithm 2 Noise Aware Analysis Operator Learning (NAAOL)

Input: $\mathbf{Y}$, $K_{\max}$, Ω_{in}, η, $\epsilon \ll 1$, λ, γ

1: **initialization:** $k = 1$, $\Omega^{[0]} = \mathbf{0}$, $\Omega^{[1]} = \Omega_{in}$
2: **while** $\epsilon \leq \|\Omega^{[k]} - \Omega^{[k-1]}\|_F$ and $k \leq K_{\max}$ **do**
3:     Operator update from Algorithm 1.
4:     **while** not converged **do**
5:        Parameters update with (9).
6:     **end while**
7:     $k = k + 1$
8: **end while**
9: **output:** $\Omega_{out} = \Omega^{[k-1]}$

Fig. 2. The cosparsities of $\mathbf{y}$ and $\widehat{\mathbf{y}}$ respectively with the operators learned with AOL and NAAOL.

Fig. 1. Signals in the analysis space $\mathbb{R}^N$, $N = 128$. The coefficients with almost zero magnitude, *i.e.* less than 0.01, are indicated with stars. The cosparsity in each case is: (a) $p = 0$, (b) $p = 1$, and (c) $p = 27$.
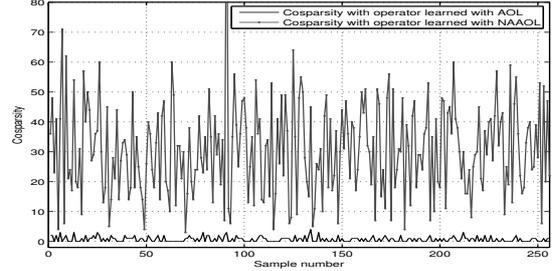
Note that the update formula for $\widehat{\mathbf{Y}}$ involves a matrix inversion, which is computationally expensive. Here $\mathbf{\Omega}$ is a tight frame and the matrix inversion is thus significantly simplified using the fact that $\mathbf{\Omega}^T\mathbf{\Omega}$ is identity. In this case, the operator $(\lambda\mathbf{I} + \gamma\,\mathbf{\Omega}^T\mathbf{\Omega})^{-1}$ is simply a scaling with $\frac{1}{\lambda + \gamma}$.

We iterate (9) for a number of iterations or until the parameters cease to change significantly. Although the convergence of the iterative updates (9) can individually be investigated, it can also be deduced using the fact that it is a special case of DRS, which converges under mild conditions. Such an iterative algorithm only finds the solution of (8) and needs to be combined with the Operator Update (OU) of Algorithm 1. A pseudocode for the full NAAOL is presented in Algorithm 2.

3. SIMULATION RESULTS

For our experiments we used a set of face images, which are centred and cropped [18]. Such images can be modeled as approximately piecewise smooth signals. A pseudo-random admissible $\mathbf{\Omega}^{[0]} \in \mathbb{R}^{128 \times 64}$ has been used as an initial analysis operator and a training set of size $L = 131072$ of 8×8 image patches ($m = 64$) from 13 different faces, have been used to learn the operators ($N = 128$).

We applied both of the AOL and the NAAOL algorithms to demonstrate how much the cosparsities of the training samples increase using the noise aware formulation. AOL was iterated $K_{\max} = 4000$ times and NAAOL iterated for $K_{\max} = 10$ iterations, while the inner-loop, *i.e.* Algorithm 1, was iterated 100 times.

A plot of the analysis coefficients for an example $\mathbf{y}$ along with its corresponding cosparsities, with three different $\mathbf{\Omega}$, are presented in Figure 1. The initial operator $\Omega_0 = \Omega^{[0]}$ has been used with $\mathbf{y}$ in (a). Not surprisingly, the signal is not cosparse with this arbitrary operator ($p = 0$). In (b), the same plot is drawn using the learned operator with AOL. Although some coefficients are small, most are not zero, and $p = 1$. In the last plot, we have shown the analysis coefficients for $\widehat{y}$ using the learned operator with NAAOL. It is clear that the cosparsity has been increased significantly (from $p = 1$ to $p = 27$). We have further plotted the cosparsities of the first 256 training samples $\mathbf{y}$'s using the learned operator found by AOL and corresponding approximations $\widehat{\mathbf{y}}$'s, which are found by NAAOL, in Figure 2. This figure also shows, the operator learning (7) using the noise aware formulation (6) results in much greater cosparsity.

The aim of second experiment is to denoise the image using a learned operator and a TV-type operator. We keep using the settings of the first experiment. The learned operator and the finite difference operator can now be used to denoise a corrupted version of another face from the database, using (6). The original face is shown in Figure 3 (a) and the noisy version with additive *i.i.d* Gaussian noise, is shown in (b). Denoising was performed using two different regularization settings: $\lambda = 0.3$, 0.1. The bottom two rows show the denoised images using the TV-type operator and the learned dictionary. We can visually conclude that the two operators successfully denoise the corrupted images with some slight differences. The results with the learned operators are smoother (this is mostly visible on a screen rather than a printed copy of the paper).

As the initial goal was to increase the cosparsities of the signals, we have also shown the cosparsities of different patches of the selected face image. The horizontal axis presents the index number of the patches. To compare the cosparsity using these operators, we have plotted their differences and its average in the bottom plot. Negative values here demonstrates the cases when the finite difference operator is a better operator than the learned operator. The average, which is indeed negative, is plotted as a horizontal line. As a result, although the learned operator performs reasonably well, the finite difference still provides cosparser signals.

4. CONCLUSION

This paper presents a novel framework to learn an analysis operator which can be used for low-dimensional modeling of sample data. The new framework has the feature that it can account for the effects of noise and model mismatch within the training data. This fact improves the robustness of the previous AOL [12]. The proposed framework is based on alternatively updating the operator and the approximations of signals. The operator can be updated using the
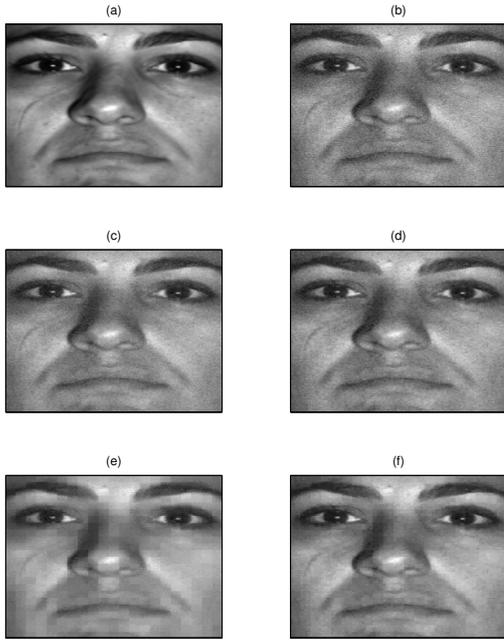
Fig. 3. Face image denoising. Top row: original face (left), noisy face (right). Denoising results using (6). Middle row: $\lambda = 0.3$ using the learned analysis operator (left) and the finite difference operator (right). Bottom row: same as middle row with $\lambda = 0.1$.

previously proposed algorithm for AOL while a new fast algorithm for solving the signal approximation has been proposed.

Two preliminary experiments are presented. The result of the first experiment demonstrates the ability of the new algorithm to compensate for model mismatch in the proposed framework. In the second experiment, we demonstrated the operator learning in a denoising application. Although the finite difference operator, which is known to be very successful for denoising of piecewise constant images, still performs better than the learned operator, the proposed operator learning scheme shows a promising result. For the signals where we do not know of any suitable operator, the proposed technique may provide some insight into this difficult low-dimensional modeling problem.
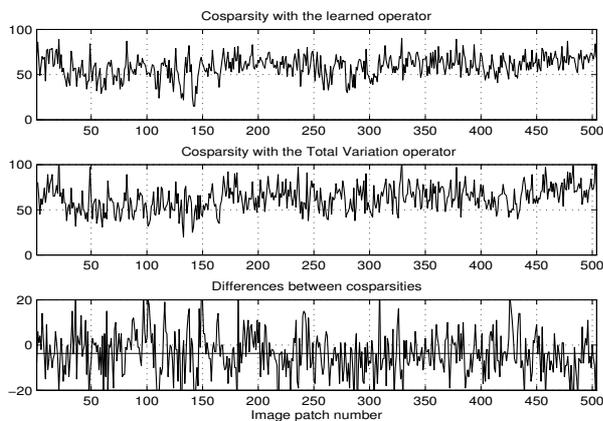
Fig. 4. Cosparsities of image paches (a)-(b) and a comparison (c).

5. REFERENCES

[1] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.

[2] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[3] "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, March 2011.

[4] M. Yaghoobi and M. Davies, "Dictionary learning for sparse representations: A Pareto curve root finding approach," in *Lecture notes in computer science, LVA/ICA*. 2010, pp. 410–417, Springer-Verlag.

[5] M. Yaghoobi, T. Blumensath, and M. Davies, "Dictionary learning for sparse approximations with the majorization method," *IEEE Trans. on Signal Processing*, vol. 57, no. 6, pp. 2178–2191, 2009.

[6] Rémi Gribonval and Karin Schnass, "Dictionary identifiability - sparse matrix-factorisation via ℓ_1 minimisation," *Information Theory, IEEE Transactions on*, vol. 56, no. 7, pp. 3523–3539, July 2010.

[7] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[8] Q. Geng, H. Wang, and J. Wright, "On the local correctness of l1 minimization for dictionary learning," Tech. Rep. abs/1101.5672, CoRR, 2011.

[9] S. Nam, M.E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," Submitted, 2011.

[10] E.J. Candès, Y.C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59 – 73, 2011.

[11] G. Peyré and J. Fadili, "Learning analysis sparsity priors," in *Proc. of Sampta'11*, 2011.

[12] M. Yaghoobi, S. Nam, R. Gribonval, and M. Davies, "Analysis operator learning for overcomplete cosparse representations," in *the European Signal Processing Conference (EUSIPCO)*, August 2011.

[13] B. Ophir, M. Elad, N. Bertin, and M.D. Plumbley, ," in *In Proceedings of EUSIPCO*, September 2011.

[14] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293 –318, 1992.

[15] T. Goldstein, , and S. Osher, "The split Bregman method for l1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323 – 343, 2009.

[16] S. Setzer, *Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage*, Scale Space and Variational Methods in Computer Vision. Springer Berlin / Heidelberg, 2009.

[17] R.T. Rockafellar, "Augmented lagrange multiplier functions and duality in nonconvex programming," *SIAM J. Control*, vol. 12, pp. 268 – 286, 1974.

[18] K.C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *PAMI*, vol. 27, no. 5, pp. 684–698, May 2005.