STRUCTURED SPARSE CODING FOR IMAGE DENOISING OR PATTERN DETECTION

Sofia Karygianni and Pascal Frossard

Signal Processing Laboratory (LTS4)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
Email:{sofia.karygianni, pascal.frossard}@epfl.ch

ABSTRACT

Sparsity has been one of the major drives in signal processing in the last decade. Structured sparsity has also lately emerged as a way to enrich signal priors towards more meaningful and accurate representations. In this paper we propose a new structured sparsity signal model that allows for the decomposition of signals into structured molecules. We define the molecules to be linear combinations of atoms in a dictionary and we create a decomposition scheme that allows for their identification in noisy signals while being robust to small errors in the internal molecule structure. We show the effectiveness of our scheme for recovering and identifying corrupted or occluded signals on both synthetic and real data.

Index Terms— structure, molecules, coefficients, sparsity

1. INTRODUCTION

Sparsity has been introduced in signal processing as a way to obtain compact representations of signals with few significant components chosen in (overcomplete) dictionaries [1]. The pioneer work of Olshausen and Field [2] also suggested that sparsity could be the mechanism employed by V1 for achieving efficient representations of natural images. Vast research efforts have been deployed to design algorithms that compute sparse decompositions of signals, such as the l_0-regularization problem [1, 3] and its convex l_1-relaxation, the Lasso [4, 5].

Sparsity is however not always a sufficient prior; the dependencies among the atoms in the signal representation should be also taken into account in order to define effective signal models. Group sparsity was introduced as a way to enforce a pre-specified structure in the decomposition: the atoms of the dictionary are partitioned into groups and the elements of each group are encouraged to appear simultaneously in the signal decomposition by an l_1 - l_2 regularization term [6]. Moving a step further, the atoms can also obey a predefined hierarchical structure, imposed by suitably chosen

---

sets of groups [7]. Several approaches have also considered general overlapping patterns by either carefully designing the groups [8, 9] or by suitably manipulating the overlapping parts among them [10]. Finally, there are also works that describe the statistical dependencies between the atoms in a dictionary with graphical models. For example, Markov Random Fiels (MRFs) are employed for modeling the underlying dependencies in [11, 12, 13] . The resulting structure model is essentially a probability distribution function that compares the different possible supports of atoms in the signal representation.

The above methods identify structures in terms of their support and, as a result, they cannot distinguish among patterns with the same support but different energy distributions among components. This can however be important for signal recognition for example. We propose here a new signal model that exactly targets a better modeling of higher level structures. Our model assumes that signals are made of molecules, which are combinations of atoms with a specific energy distribution. Molecules are able to model more complex signal patterns than common group sparsity priors while keeping a simple form with linear combinations of atoms. We also introduce a new decomposition scheme that allows for the identification of molecule structures in signals. Its novelty resides mainly in the explicit inclusion of an error matrix that allows for small errors in the structure of molecules. Our scheme manages to perform high level decomposition of signals while staying robust to small deviations from the predefined structures. We provide experimental results on both synthetic and real signals that show the effectiveness of our scheme for recovering corrupted or occluded structures in signals.

2. PROBLEM FORMULATION

We consider a new structured signal model based on molecules, which are sets of atoms with a given energy distribution, i.e. the relative energy of each atom in each molecule instance is the same. This can be realized by defining molecule prototypes as linear combinations of atoms, while each molecule instance is any multiple of one prototype. Given a possibly

overcomplete dictionary $D \in \mathbb{R}^{N \times K}$, the molecule set is thus expressed as a matrix $C \in \mathbb{R}^{K \times M}$, where each column specifies the energy distribution among atoms in a given molecule, i.e., the molecule matrix is defined as $C = [\bar{c}_1 \, \bar{c}_2 \, \cdots \, \bar{c}_M]$ where $\bar{c}_i \in \mathbb{R}^{K \times 1}$. The support of the molecule defined by $\bar{c}_i$ is $s_i = \{j \in [1, N], c_{ij} > 0\}$ and the corresponding molecule is $m_i = D\bar{c}_i$. We assume that each m_i is normalized. A signal x that is a combination of molecules can then be written as $x = DCa$ where $D = [\bar{d}_1 \, \bar{d}_2 \, \cdots \, \bar{d}_K], \bar{d}_i \in \mathbb{R}^{N \times 1}$, is the dictionary of elementary atoms and a is a positive coefficient vector.

The matrix C serves as a rich prior about the signal. However, this comes at the expense of reduced robustness to noise since the traditional signal decomposition techniques using the dictionary of molecules may lead to inaccurate representations in case of small deviations from the predefined energy distribution in the noisy signal. Hence we propose a new decomposition technique that incorporates an error matrix to capture errors inside the structure of molecules. We denote this error matrix as $E \in \mathbb{R}^{K \times M}$.

Finally, given a signal $x \in \Re^N$, its sparse decomposition into molecules, namely $a \in \mathbb{R}^M$, as well as the related error matrix E can be found by solving the following optimization scheme:

$$\{\hat{E}, \hat{a}\} = \arg\min_{a, E} \{||x - D * (C + E) * a||_2 + \lambda_1 * ||a||_1$$
$$+ \lambda_2 * ||W. * E||_1\} \quad (1)$$

subject to $a_i \geq 0, \forall i \in [1, M]$. The optimization in Eq. (1) contains the unstructured error term expressed by $||x - D * (C + E) * a||_2$, that measures which part of the signal cannot fit into the structure model. The remaining two terms in Eq. (1) express sparsity and energy constraints on a and E respectively. The first one enforces sparse decompositions of signals in molecules. The latter is a constraint on the error E expressed by the term $||W. * E||_1$. The weight matrix W is a mask that ensures that the errors appear mostly on atoms belonging to the support of molecules. This is achieved by setting to one the entries of W in the support of each molecule and to arbitrary high values the other entries. This constraint enforces the decomposition to stay close to the signal model by preventing random atoms to appear in the signal. Finally the regularization parameters λ_1, λ_2 control the relative weight of each constraint and depend on the target application.

3. ALGORITHM

The cost function in the optimization problem given in Eq. (1) is not convex on both (a, E). However, considering that a is fixed, the optimization over E is convex since the involved l_2 and l_1 norms are convex functions of affine maps on E. Similarly, when E is fixed, the optimization over a is convex too. Therefore, we can use an alternate minimization procedure to solve for both (a, E). Although alternate optimization cannot guarantee the convergence to a global minimum, it usually converges to a good solution in practice.

To be more specific, we firstly set E to zero and solve the optimization in Eq. (1) only for a. This first step is actually equivalent to finding a sparse decomposition of the signal in the molecule dictionary directly. The solution serves as an initialization for a and is denoted as a_0. Then, by fixing a to a_0, we solve Eq. (1) for E. Its solution E_0 is used as the value of E for the next step. After completing this initialization procedure, we alternate over the optimization for a and E until no significant changes occur on any of the minimization problems. In this alternating phase, for the optimization over a, we employ the method of reweighted l_1-minimization described in [14] as it leads to a more effective approximation of the desired l_0 norm. All the convex minimization subproblems involved in the above procedure are solved with the SeDuMi interior point solver in CVX [15]. Their computational complexity depends on the size of the used dictionary as well as the number of non-zero atoms in the molecules. The pseudocode of the proposed scheme is shown in Figure 3.

Algorithm 1 Structured sparse decomposition with molecules (A0)

Input: $x, D, C, W, \lambda_1, \lambda_2, \epsilon$
    Initialization
1: $a_0 = \arg\min_a \{||x - D * C * a||_2 + \lambda_1 * ||a||_1\}, a \geq 0$
2: $E_0 = \arg\min_E \{||x - D*(C+E)*a||_2 + \lambda_2 * ||W.*E||_1\}$
3: $a_p = a_0$
4: $E_p = E_0$
    Alternate optimization
5: **while** true **do**
6:     $WA = 1./a_p$
7:     $\hat{a} = \arg\min_a \{||x - D*(C+E_p)*a||_2 + \lambda_1 * ||WA.*a||_1\}, a \geq 0$
8:     $\hat{E} = \arg\min_E \{||x - D*(C+E)*\hat{a}||_2 + \lambda_2 * ||W.*E||_1\}$
9:     **if** $(||\hat{a} - a_p|| + ||\hat{E} - E_p|| < \epsilon)$ **then**
10:         return
11:     **else**
12:         $a_p = \hat{a}$
13:         $E_p = \hat{E}$
14:     **end if**
15: **end while**
Output: $\hat{a}, \hat{E}$

---

4. RESULTS

The proposed scheme (A0) is a method to decompose signals into higher level patterns, the molecules. Its goal is to achieve a simple decomposition algorithm that imposes a rich but flexible prior on the signals. To prove its effectiveness we
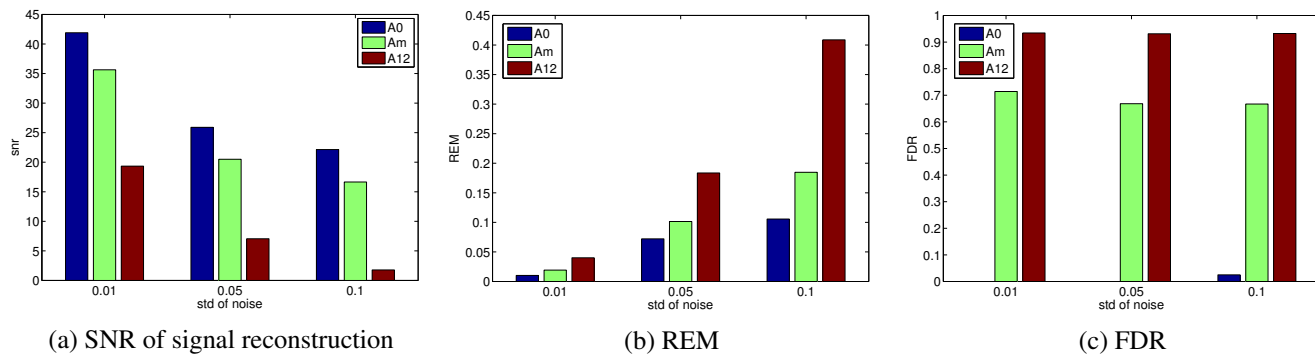
(a) SNR of signal reconstruction   (b) REM   (c) FDR

Fig. 1. The denoising result for synthetic data and different noise levels. In (a) we can see the SNR of the different signal reconstructions while in (b) we have the error in the recovery of molecule coefficients. In (c) we plot the FDR for all the schemes.

compare it against the l_1-l_2 group norm (the algorithm is denoted as A12 in this paper) [10], a well known approach for incorporating prior information about groups in sparse coding. Each group $g_i \in \mathcal{G}$ is defined according to the support of the corresponding molecule m_i and the equivalent optimization problem is:

$$\hat{b} = \arg\min_b \{||x - D * b||_2 + \lambda \sum_{g_i \in \mathcal{G}} ||b_{g_i}||\} \qquad (2)$$

The vector of coefficients b is the signal decomposition in the atom level and b_{g_i} is its restriction on g_i. The decomposition $\hat{a}$ in groups (or equivalently molecules in our case) is found as the l_2 norm of the coefficients in each group i.e., $\hat{a}_i = ||\hat{b}_{g_i}||$.

Since molecules are linear combinations of atoms, an alternative option is to introduce them directly into the dictionary and perform sparse coding with the dictionary of molecules. This approach is similar to the sparse coding step in [16]. In the following, we also compare our scheme with sparse coding with l_1 regularization on the molecule dictionary (we denote this algorithm as Am), i.e., that is the outcome of:

$$\hat{a} = \arg\min_a \{||x - D_m * a||_2 + \lambda||a||_1\} \qquad (3)$$

where $D_m = DC$ is the molecule dictionary.

The performance of the algorithms is compared using various measures. To quantify the performance in terms of the signal recovery we compute both the reconstruction error of the signal approximation, i.e., $||x - \hat{x}||$ where $\hat{x}$ is the signal reconstruction based on $\hat{a}$ or $\hat{b}$, as well as the error in the recovery of molecule coefficients (REM) defined as $||a - \hat{a}||$.

Moreover, we are also interested in the accuracy of the molecule detection. Therefore, we also provide results for the false discovery rate (FDR), which is the percentage of the detected molecules that is incorrect. We finally note that the true positive rate is almost always equal to one.

4.1. Denoising on Synthetic Data

Firstly, we have investigated the performance of our scheme for denoising synthetic image data. We have created a dictionary of 30 molecules by randomly combining 5 atoms of a DCT dictionary. Each testing signal is then composed of one molecule and some Gaussian noise. The average results for 30 testing signals across different levels of noise are shown in Figure 1.

From the bar plots, we can verify that the inclusion of the molecule structure in the decomposition certainly improves the denoising performance. The SNR of the recovered signals as well as the recovered energy in the molecule level are always better with our scheme A0. The detection of the molecules is also more accurate as it is shown from the negligible FDR. This is happening because the explicit inclusion of the error matrix in the decomposition allows for flexibility that compensates the effect of noise. The high values for the REM in A12 are due to the lack of internal structure in the groups. As a result, A12 is not able to distinguish between the two different levels of energy: the one of the atoms inside a molecule and the one at the molecule level; the molecule coefficients cannot be recovered accurately.

4.2. Occlusion on Synthetic data

We have further investigated the performance of our scheme in case of occlusion, i.e., in case existing structures in signals are incomplete. We have created a molecule dictionary with highly overlapping molecules. To be more specific, we have used a dictionary of 30 molecules divided into 10 groups. In each group, molecules have 4 out of 11 atoms in common with coefficients that result in a correlation equal to 60% for the molecules in the group.

Each testing signal is then composed of one molecule missing 50% of its uncommon part plus Gaussian noise. The results for different levels of noise are shown in Figure

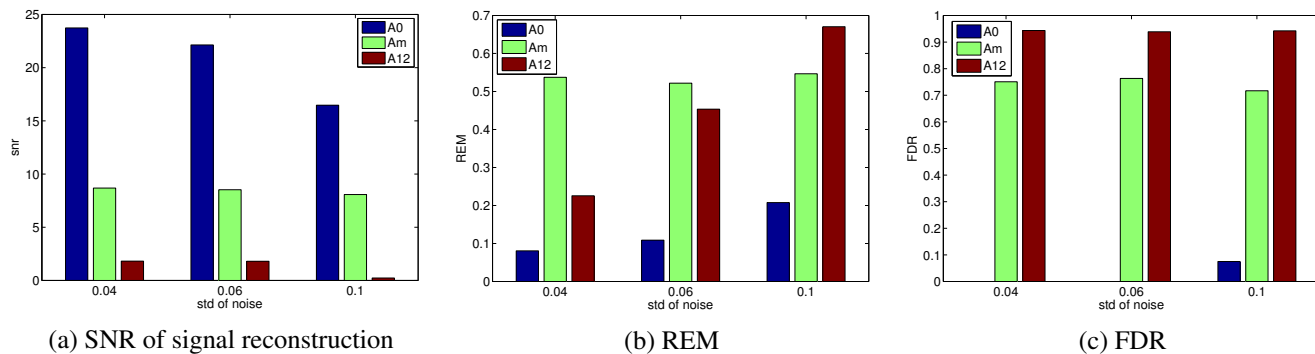(a) SNR of signal reconstruction   (b) REM   (c) FDR

Fig. 2. The results for the case of occluded synthetic data across different noise levels. The performance of A0 is significantly better than the other schemes as it detects more accurately the correct molecules.

2. Again, our scheme achieves better reconstruction performance and better structure detection. It is important to notice the poor performance of A12 in terms of reconstruction due to the inability of the l_1-l_2 group norm to complete missing parts of structures; as a result the occluded atoms cannot be recovered. At the same time, the molecule dictionary gets easily confused by the common support of the molecules resulting in high FDR rates and high REM due to the distribution of the signal energy over the wrongly detected molecules.

4.3. Occlusion on MNIST

We have also tried our scheme with data from the MNIST dataset [17], downsampled to 14×14 images. For the atom dictionary we have used an overcomplete dictionary of 323 Gaussian, anisotropic atoms with a mother function $\phi(x,y) = A \exp(-(x/4)^2 - y^2)$. In order to create our molecule dictionary and the structure matrix C, we have randomly chosen one instance of each digit and approximated it using l_1 sparse coding.

For testing the performance under occlusion, we have created a testing set with 30 signals, each consisting of one molecule missing a specific percentage of its structure. The results for different missing percentages are shown in Figure 3. In this case, the difference between A0 and Am is not very significant in terms of the reconstruction of signals. This is due to the low coherence of the molecule dictionary in combination with the absence of any additional noise. However, we have benefits in the molecule detection performance as it is obvious from the zero FDR for A0 in plot (c) of Figure 3.

5. CONCLUSIONS

In this paper we have presented a new structured signal model based on the notion of molecules, which are defined as linear combinations of atoms in a dictionary. We have also introduced a sparse decomposition scheme with structure flexibility provided by the explicit inclusion of constrained errors in
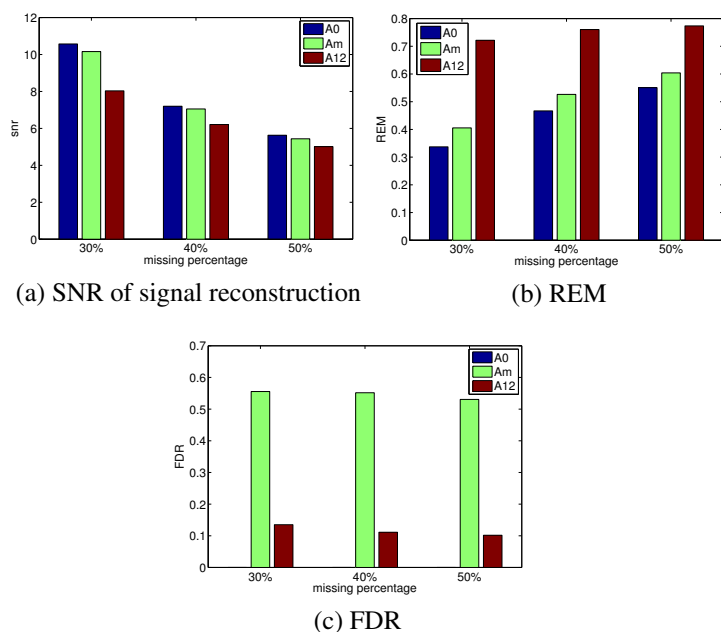

(a) SNR of signal reconstruction   (b) REM

(c) FDR

Fig. 3. The SNR, REM and FDR for occlusion on MNIST digits.

molecules. We have tested our scheme on both synthetic and real signals and proved that it offers significant benefits to the reconstruction of noisy or occluded signals especially in the case of highly correlated molecules. The extension to more severe deformations of structures as well as the learning of structures from data remain as future work.

6. REFERENCES

[1] Stephane G. Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, 1993.

[2] Bruno A. Olshausen and David J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by v1," *Vision Research*, 1997.

[3] Joel A. Tropp and Anna C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, 2007.

[4] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, 1994.

[5] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, 1998.

[6] Ming Yuan, Ming Yuan, Yi Lin, and Yi Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Series B*, 2006.

[7] Peng Zhao, Guilherme Rocha, and Bin Yu, "The composite absolute penalties family for grouped and hierarchical variable selection," *Annals of Statistics*, 2009.

[8] Rodolphe Jenatton, Jean-Yves Audibert, and Francis Bach, "Structured variable selection with sparsity-inducing norms," *Journal of Machine Learning Reasearch*, 2011.

[9] Junzhou Huang, Tong Zhang, and Dimitris Metaxas, "Learning with structured sparsity," in *International Conference on Machine Learning*, 2009.

[10] Laurent Jacob, Guillaume Obozinski, and Jean-Philippe Vert, "Group lasso with overlap and graph lasso," in *International Conference on Machine Learning*, 2009.

[11] Tomer Peleg, Yonina C. Eldar, and Michael Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *IEEE Transactions on Signal Processing*, 2012.

[12] Pierre J. Garrigues and Bruno A. Olshausen, "Learning horizontal connections in a sparse coding model of natural images," in *NIPS*, 2008.

[13] Volkan Cevher, Marco F. Duarte, Chinmay Hegde, and Richard G. Baraniuk, "Sparse signal recovery using markov random fields.," in *NIPS*, 2008.

[14] Emmanuel J. Candes, Michael B. Wakin, and Stephen P. Boyd, "Enhacing sparsity by reweighted $l1$ minimization," *Journal of Fourier Analysis and Applications*, 2008.

[15] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sept. 2013.

[16] Ron Rubinstein, Michael Zibulevsky, and Michael Elad, "Double sparsity: learning sparse dictionaries for sparse signal approximation.," *IEEE Transactions on Signal Processing*, 2010.

[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998.