

COSPARSE ANALYSIS MODELING – UNIQUENESS AND ALGORITHMS

Sangnam Nam¹ Michael E. Davies² Michael Elad³ Rémi Gribonval¹

¹ IRISA, INRIA Rennes – Bretagne Atlantique, France

² IDCOM & Joint Research Institute for Signal and Image Processing, Edinburgh University, UK

³ Department of Computer Science, The Technion – Israel Institute of Technology, Israel

ABSTRACT

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome. In this paper, we consider this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). Our work proposes a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. This paper also considers two pursuit algorithms for solving the missing data problem, an L1-based and a new greedy method. Our simulations demonstrate the appeal of the analysis model, and the success of the pursuit techniques presented.

Index Terms— Sparse Representation, Inverse Problems, Algorithms, Cosparse Analysis Model

1. INTRODUCTION

A ubiquitous problem that has found many applications, from signal processing to machine learning, is to estimate a high-dimensional vector $\mathbf{x}_0 \in \mathbf{R}^d$ from a set of possibly incomplete linear observations $\mathbf{y} \in \mathbf{R}^m$,

$$\mathbf{y} = \mathbf{M}\mathbf{x}_0. \quad (1)$$

When the number of observations m is smaller than the data dimension d , this is an ill-posed problem which admits infinitely many solutions. Solving it is hopeless unless we can use additional prior knowledge on \mathbf{x}_0 .

The assumption that \mathbf{x}_0 is *sparse* is known to be of significant help, and it is now well understood that under incoherence assumptions on the matrix \mathbf{M} , sufficiently sparse vectors \mathbf{x}_0 can be robustly estimated by the solution of:

$$\hat{\mathbf{x}} := \arg \min_{\mathbf{x}} \|\mathbf{x}\|_{\tau} \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{x}, \quad (2)$$

The authors acknowledge the support by the European Community’s FP7-FET program, SMALL project, under grant agreement no. 225913.

where $0 \leq \tau \leq 1$.

In many cases, \mathbf{x}_0 itself may not be sparse, but may still admit a sparse representation \mathbf{z}_0 in some *synthesis dictionary* \mathbf{D} , i.e., $\mathbf{x}_0 = \mathbf{D}\mathbf{z}_0$, $\mathbf{z}_0 \in \mathbf{R}^n$, in which case one will estimate it using a modified optimization problem

$$\hat{\mathbf{x}}_S := \mathbf{D}\hat{\mathbf{z}}; \quad \hat{\mathbf{z}} := \arg \min_{\mathbf{z}} \|\mathbf{z}\|_{\tau} \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{z}. \quad (3)$$

The focus of this paper is on an alternative to the above sparse synthesis data model; the *cosparse analysis data model*. In this model, there is an *analysis operator* $\Omega \in \mathbf{R}^{p \times d}$ such that the *analysis coefficients sequence* $\Omega\mathbf{x}_0$ of the signal \mathbf{x}_0 of interest is, in some sense, sparse. One could then solve, in place of (3), the *analysis-based* optimization problem,

$$\hat{\mathbf{x}}_A := \arg \min_{\mathbf{x}} \|\Omega\mathbf{x}\|_{\tau} \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{x}. \quad (4)$$

Typically the dimensions are $m \leq d \leq p, n$.

The analysis ℓ_{τ} -minimization method (4) has already been widely used in practice and studied theoretically (e.g., [1], [2], [3], [4]). However, there seems to be little theoretical discussion in the literature of the underlying cosparse analysis model. *In this paper, we highlight the different structure of this new model and its implications.* We look at the uniqueness of ‘cosparse’ recovery problem in the analysis model, and propose a new greedy algorithm which is based on the concept of the cosparse analysis data model. We present simulation results from applying the pursuit algorithms to signals conforming to the proposed model, demonstrating their appeal.

2. COSPARSE ANALYSIS DATA MODEL

The choice of the optimization principle (3) for data estimation will not provide an exact estimator for \mathbf{x}_0 , unless we assume the *k-sparse synthesis data model*:

$$\exists \mathbf{z}_0 \in \mathbf{R}^n, \quad \mathbf{x}_0 = \mathbf{D}\mathbf{z}_0, \quad \|\mathbf{z}_0\|_0 = k. \quad (5)$$

As it is now well understood, if the *k-sparse synthesis model* holds for sufficiently small k (we will also say that \mathbf{x}_0 is *synthesis-sparse* or simply *sparse*), then the optimization problem (3) with $\tau = 1$ indeed allows the recovery of \mathbf{x}_0 by

convex optimization ([5], [6]), which has bounded complexity.

Just as for the sparse synthesis data model, one can hope to estimate \mathbf{x}_0 using (4) only if \mathbf{x}_0 is ‘sufficiently cosparse’. In contrast to the synthesis model, we need to define the ℓ -*cosparse analysis data model* as follows:

$$\|\Omega\mathbf{x}_0\|_0 = p - \ell. \quad (6)$$

One can observe that the larger ℓ is, the lower dimensional \mathbf{x}_0 is. When \mathbf{x}_0 satisfies the ℓ -cosparse analysis model (6) for large ℓ , we will say informally that \mathbf{x}_0 is *analysis-cosparse*, or simply *cosparse*. The integer ℓ in (6) will be called the *cosparsity* of \mathbf{x}_0 . We highlight that the *cosparsity* of \mathbf{x}_0 can be very high while the coefficient sequence $\Omega\mathbf{x}_0$ is not sparse in the standard sparse synthesis model sense (see further discussion below).

2.1. Similarities with the Synthesis Model

Both models (synthesis and analysis) can be viewed as a more general data model called the *union of subspaces model*. In this model, it is assumed that there is a collection of subspaces $S_\gamma \subset \mathbf{R}^d$, $\gamma \in \Gamma$, such that a signal of interest \mathbf{x}_0 coincides with an element from the union $\cup_{\gamma \in \Gamma} S_\gamma$.

The cosparse analysis model is associated to a union of subspaces model as follows: given an analysis operator Ω and its rows $\omega_j \in \mathbf{R}^d$, $1 \leq j \leq p$, the vectors that satisfy the model (6) are simply in the orthogonal complement to linear combinations of ℓ rows of Ω : when the *zero* entries of the analysis coefficient vector $\Omega\mathbf{x} \in \mathbf{R}^p$ are in the index set $\Lambda \subset \llbracket 1, p \rrbracket$, we have $\Omega_\Lambda\mathbf{x} = 0$, i.e., $\mathbf{x} \in \mathcal{W}_\Lambda$, where

$$\mathcal{W}_\Lambda := \text{span}(\omega_j, j \in \Lambda)^\perp = \{\mathbf{x} : \langle \omega_j, \mathbf{x} \rangle = 0, j \in \Lambda\}. \quad (7)$$

Hence, \mathbf{x}_0 satisfies the cosparse analysis data model if it is an element of $\cup_\Lambda \mathcal{W}_\Lambda$, where the union is taken over sets of cardinality $\#\Lambda = \ell$.

Similarly, the sparse synthesis model is recognized as a union of subspaces model: leaving the details to the reader, given a dictionary $\mathbf{D} \in \mathbf{R}^{d \times n}$, we observe that \mathbf{x}_0 satisfies the sparse synthesis data model if it belongs to $\cup_T \mathcal{V}_T$ where \mathcal{V}_T 's are k -dimensional subspaces spanned by k columns from \mathbf{D} .

Being viewed as union of subspaces models, the synthesis and analysis data models may be considered to be ‘equivalent’ if the subspaces \mathcal{W}_Λ and \mathcal{V}_T are of the same dimension and the number of subspaces in the unions are the same. That is, the two data models are similar when $\ell = d - k$, $\binom{p}{\ell} \approx \binom{n}{k}$.

2.2. Differences with the Synthesis Model

Given a dictionary $\mathbf{D} \in \mathbf{R}^{d \times n}$ for the sparse synthesis model, it is easy to see that there exist signals with very sparse representations. For example, if the signal \mathbf{x}_0 is a scalar multiple of a column of \mathbf{D} , then it has a 1-sparse representation. We also

observe that for any representation $\mathbf{x}_0 = \mathbf{D}\mathbf{z}$ with $\|\mathbf{z}\|_0 > d$, there is a sparser representation $\mathbf{x}_0 = \mathbf{D}\mathbf{z}'$, $\|\mathbf{z}'\|_0 \leq d$.

In contrast, given a generic analysis operator $\Omega \in \mathbf{R}^{p \times d}$ with no non-trivial linear dependencies among its rows, the representation $\Omega\mathbf{x}_0$ of $\mathbf{x}_0 \neq 0$ cannot achieve sparsity $\|\Omega\mathbf{x}_0\|_0 \leq p - d$ when $p > d$. Indeed, *cosparsity* $\ell \geq d$ can only be achieved when $\mathbf{x}_0 = 0$ since it implies that $\mathbf{x}_0 \in \mathbf{R}^d$ is orthogonal to at least d linearly independent rows of Ω . Also, $\|\Omega\mathbf{x}_0\|_0$ can be as large as p unlike in the synthesis case. Focussing on *cosparsity*, the number ℓ of *zero* entries in $\Omega\mathbf{x}_0$ can freely range from 0 to $d - 1$. This is the reason why it is more natural to consider the notion of *cosparsity*, and this has implications in understanding the uniqueness of the ‘sparse recovery’ problem for the cosparse analysis model.

3. UNIQUENESS

We now consider the uniqueness of solutions to (1) under a cosparse analysis model constraint. Once the index set Λ such that $\Omega_\Lambda\mathbf{x}_0 = 0$ is known, we can form the linear system,

$$\begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{M} \\ \Omega_\Lambda \end{bmatrix} \mathbf{x}. \quad (8)$$

This system has a unique solution $\mathbf{x} = \mathbf{x}_0$ provided that the matrix $\begin{bmatrix} \mathbf{M} \\ \Omega_\Lambda \end{bmatrix}$ has rank d . This requires $m \geq d - \ell$.

In practice, Λ is not known, and we are interested in uniqueness guarantees of (1) under the cosparse analysis model. This is an instance of a more general uniqueness problem studied in [7, 8] for union of subspaces models. In particular [7] shows that \mathbf{M} is invertible on the union of subspaces $\cup_{\gamma \in \Gamma} S_\gamma$ if and only if \mathbf{M} is invertible on all subspaces $S_\gamma + S_\theta$ for all $\gamma, \theta \in \Gamma$. In the context of the cosparse analysis model this yields the following result:

Proposition 1. *Suppose that for some fixed $\ell > 0$, Ω and \mathbf{M} satisfy*

$$(\mathcal{W}_{\Lambda_1} + \mathcal{W}_{\Lambda_2}) \cap \text{Null}(\mathbf{M}) = \{0\} \quad (9)$$

for every pair of subsets Λ_1 and Λ_2 of $\llbracket 1, p \rrbracket$ with $\#\Lambda_i \geq \ell$, $i = 1, 2$. Then there is at most one solution \mathbf{x} that satisfies

$$\mathbf{y} = \mathbf{M}\mathbf{x} \text{ and } \|\Omega\mathbf{x}\|_0 \leq p - \ell. \quad (10)$$

To interpret the result of Proposition 1 in terms of relations between the number of measurements m and the *cosparsity* ℓ , let us now consider the case where Ω is in general position, which means that any collection of less than d rows from Ω is linearly independent. This implies that $\dim(\mathcal{W}_{\Lambda_1} + \mathcal{W}_{\Lambda_2}) \leq 2(d - \ell)$ whenever $\#\Lambda_i \geq \ell$. We further assume that (i) \mathbf{M} is of full rank, which yields $\dim(\mathcal{W}_{\Lambda_1} + \mathcal{W}_{\Lambda_2}) + \dim(\text{Null}(\mathbf{M})) \leq 2(d - \ell) + d - m$; and (ii) \mathbf{M} is ‘independent’ of Ω , implying that when $2(d - \ell) + d - m \leq d$, then (9) is satisfied. This leads to the following proposition (A general version of this for union of subspaces is given in [8, Theorem 2.3].):

Proposition 2. *Let Ω be an analysis operator in general position. Then, for almost all \mathbf{M} (with respect to the Lebesgue measure), a necessary and sufficient condition for (10) to have at most one solution is $m \geq 2(d - \ell)$.*

There are three interesting observations that we would like to draw attention to. First, while $m \geq d - \ell$ measurements are required to identify \mathbf{x}_0 when Λ is known, here we need twice as many to blindly identify Λ . Second, some readers may notice that Proposition 2 is the counter part to the following result in the sparse synthesis model: roughly speaking, in order to recover k -sparse \mathbf{x}_0 as the sparsest representation that matches the observation, one needs at least $m \geq 2k$ measurements. The proposition says that in the cosparsity analysis model, again roughly speaking, in order to recover ℓ -cosparsity \mathbf{x}_0 , at least $m \geq 2(d - \ell)$ measurements are necessary. This suggests that $d - \ell$ is playing the role of k . Third, it is interesting to note that the number of rows in Ω , p , plays no role in the uniqueness result.

4. ALGORITHMS

In this section we present two algorithms, both targeting the solution of the problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\Omega \mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{M} \mathbf{x} = \mathbf{y}. \quad (11)$$

4.1. The Analysis ℓ_1 -minimization Algorithm

The first algorithm amounts to using ℓ_1 -norm instead of ℓ_0 -norm in (11). Throughout the paper, we call this algorithm the ‘‘analysis ℓ_1 -minimization algorithm’’, or simply, ‘‘analysis- ℓ_1 ’’.

4.2. The Greedy Analysis Pursuit Algorithm

The new algorithm we present is the Greedy Analysis Pursuit (GAP) algorithm. It is an adaptation of the well-known greedy pursuit algorithm used for the synthesis model – the Orthogonal Matching Pursuit (OMP). Similar to the synthesis case, our goal is to detect the informative support of $\Omega \mathbf{x}$ – in the analysis case, this amounts to the locations of the zeros in the vector $\Omega \mathbf{x}$, so as to introduce additional constraints to the underdetermined system $\mathbf{M} \mathbf{x} = \mathbf{y}$. Note that to obtain a solution, one needs to detect at least $d - m$ of these zeros, and thus if the cosparsity $\ell > d - m$, detection of the complete set of zeros is not mandatory.

The algorithm aims to detect the elements outside the set Λ , this way carving its way towards the detection of the desired support. Therefore, the support $\hat{\Lambda}$ is initialized to be the whole set $\{1, 2, 3, \dots, p\}$, and through the iterations it is reduced towards a set of size ℓ (or less, $d - m$). A detailed description of the algorithm is available in Figure 1.

- **Task:** Approximate the solution of (11).
- **Parameters:** Given are the matrices \mathbf{M} , Ω , the vector \mathbf{y} . We also use a very small parameter $\lambda > 0$.
- **Initialization:** Set $k = 0$ and perform the following steps:
 - **Initialize Co-Support:** $\hat{\Lambda}_k = \{1, 2, 3, \dots, p\}$,
 - **Initialize Solution:** $\hat{\mathbf{x}}_k = \begin{bmatrix} \mathbf{M} \\ \sqrt{\lambda} \Omega_{\hat{\Lambda}_k} \end{bmatrix}^\dagger \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}$.
- **GAP Iterations:** $k \rightarrow k + 1$ and perform the following steps:
 - **Project:** Compute $\alpha = \Omega \hat{\mathbf{x}}_{k-1}$,
 - **Update Support:** $\hat{\Lambda}_k = \hat{\Lambda}_{k-1} \setminus \{\arg \max_{i \in \hat{\Lambda}_{k-1}} |\alpha_i|\}$,
 - **Update Solution:** $\hat{\mathbf{x}}_k = \begin{bmatrix} \mathbf{M} \\ \sqrt{\lambda} \Omega_{\hat{\Lambda}_k} \end{bmatrix}^\dagger \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}$.
 - **Stopping Criterion:** If $k \geq p - d + m$ or $\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{k-1}$ is small, stop.
- **Output:** The proposed solution is $\hat{\mathbf{x}}_k$.

Fig. 1. The Greedy Analysis Pursuit (GAP) algorithm.

5. EMPIRICAL PERFORMANCE OF THE ALGORITHMS

In this section, we show the results obtained by applying the algorithms described in Section 4 to a synthetic problem. In the experiment, $\mathbf{M} \in \mathbf{R}^{m \times d}$ was chosen to be a random Gaussian matrix, and $\Omega \in \mathbf{R}^{p \times d}$ were constructed to be the transpose of a random tight frame. Next, for a fixed cosparsity ℓ , random ℓ rows, say Λ , of Ω were selected. The target signal \mathbf{x}_0 was constructed by finding an orthonormal basis \mathbf{B} for \mathcal{W}_Λ , generating an independent random normal vector c that corresponds to the size of \mathbf{B} , and constructing \mathbf{x}_0 as $\mathbf{x}_0 = \mathbf{B}c$. The observation was obtained by $\mathbf{y} = \mathbf{M} \mathbf{x}_0$.

We have used the Matlab `cvx` package [9] with high-est precision to solve the analysis ℓ_1 optimization problem (4) and the final solution was debiased. For GAP, we have implemented the algorithm in Matlab. The parameter λ was taken to be 10^{-6} and for the stopping criterion, $\|\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{k-1}\|_2 / \|\hat{\mathbf{x}}_k\|_2 < 10^{-5}$ was used. The code is available on the web at <http://small-project.eu/software-data>.

Figure 2 shows the so-called phase transition diagrams obtained from the experiment. The gray level of each square in a diagram indicates what proportion of the corpus of 50 randomly drawn target signals \mathbf{x}_0 was correctly identified (perfect recovery), for a given set of parameters δ , ρ , and σ as described below. Black corresponds to zero recovery while white corresponds to all 50 instances of recovery. In all cases, the signal dimension d is set to 200. We varied the number m of measurements, the co-sparsity ℓ of the target signal, and

the dictionary size p according to the following formulae:

$$m = \delta d, \quad \ell = d - \rho m, \quad p = \sigma d. \quad (12)$$

δ can be interpreted as the compression ratio or the amount of measurements available in relation to the signal dimension d , and ρ as the relative dimension of the signal to m . Lastly, σ indicates the amount of redundancy of the analysis operator. A relative error of size less than 10^{-6} was counted as perfect recovery. In the case of $\sigma = 1$, the weak phase transition curve of Donoho and Tanner [10] is also plotted in red.

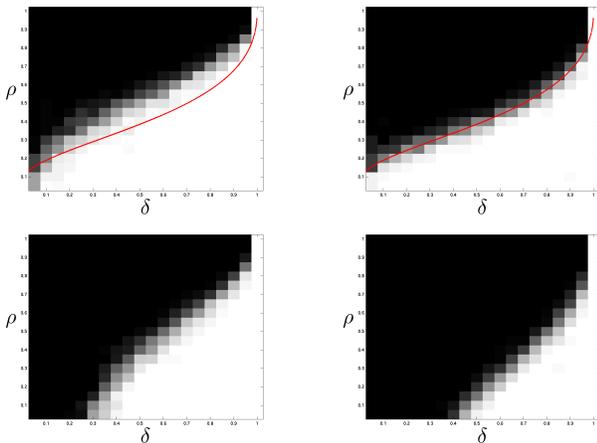


Fig. 2. Recovery Rate of GAP (left) and Analysis- ℓ_1 (right) for $d = 200$. From top to bottom, $\sigma = 1$ (non-redundant case) and 1.2 (more redundant case).

One may loosely interpret that an algorithm performs better than the other if its phase diagram has a larger bright area than that of the other. The figure shows that GAP performs better than analysis- ℓ_1 in both complete ($\sigma = 1$) and overcomplete ($\sigma = 1.2$) cases. Looking at the diagrams more closely, the recovery (white) region for GAP starts at smaller δ than that of analysis- ℓ_1 . This means that GAP can have perfect recovery with fewer measurements. We also observe that the recovery region for GAP is taller, which means that GAP can recover less cosparse vectors (larger ρ) than analysis- ℓ_1 .

We also observe that for the generic Ω used in this experiment, redundancy ($\sigma > 1$) comes with a price; below a critical under-sampling ratio δ , recovery becomes impossible. We believe that this is due to the exponentially growing number of low dimensional subspaces associated to the cosparse analysis model.

To give an idea on how the algorithms compare in terms of computational cost, for $p = d = 200$, $m = 100$, and $\ell = 190$, GAP and analysis- ℓ_1 took 77.3 and 69.7 seconds, respectively, to complete 50 runs. However, when the redundancy σ is large, GAP, as expected, took longer: for $p = 240$, $d = 200$, $m = 100$, $\ell = 190$, the run time were 322.8 and 96.8 seconds, respectively for GAP and analysis- ℓ_1 .

6. CONCLUSION

We have described a cosparse analysis data model which is an alternative to the widely used sparse synthesis data model. Although analysis representations have been studied before, this cosparse analysis model offers a new perspective on such approaches. We have also presented a uniqueness result for this model, hence provided a ground for the possibility of recovering a cosparse signal when only a partial information about the signal is available. Furthermore, we have shown by synthetic experiments that there are algorithms that achieve the perfect recovery of cosparse signals.

7. REFERENCES

- [1] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, June 2007.
- [2] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," *In Proceedings of SPIE*, vol. 7446 (Wavelets XIII), August 2009.
- [3] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, In Press, 2010.
- [4] J. Portilla, "Image restoration through 10 analysis-based sparse optimization in tight frames," *Proc. of the 16th IEEE Int. Conf. on Image Proc.*, pp. 3865–3868, 2009.
- [5] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
- [6] D.L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Aca. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
- [7] Y. M. Lu and M. N. Do, "Sampling signals from a union of subspaces," *IEEE Signal Proc. Mag.*, pp. 41–47, mar 2008.
- [8] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inform. Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.
- [9] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab Software for Disciplined Convex Programming," August 2008.
- [10] D. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *Journal of the AMS*, vol. 22, no. 1, pp. 1–53, Jan. 2009.