Fusion of Orthogonal Matching Pursuit and Least Squares Pursuit for Robust Sparse Recovery

Nicolae Cleju, Iulian Ciocoiu
Faculty of Electronics, Telecommunications and Information Technology
"Gheorghe Asachi" Technical University of Iasi
700506, Iasi, Romania
Email: {nikcleju, iciocoiu}@etti.tuiasi.ro

***Abstract*—We propose two approaches for obtaining a more robust algorithm for sparse signal recovery based on combining two existing algorithms, Orthogonal Matching Pursuit and the related Least Squares Pursuit. The first approach relies on averaging the gradient vectors used in the two algorithms for atom selection at every iteration. The second approach is to use the atom selected by the algorithm which has the highest confidence in its choice. Simulation results show more robust and well-balanced recovery performance, even in testing scenarios which are adverse to one or the other of the underlying algorithms.**

I. Introduction

Sparse signal recovery from compressed measurements [1], [2] is an optimization problem that has been a prominent research topic in signal processing throughout the last two decades. Consider an unknown signal $\mathbf{x} \in R^n$ that has a sparse representation γ in some known basis or overcomplete dictionary $D \in \mathbb{R}^{n \times N}$, $\mathbf{x} = D\gamma, \|\gamma\|_0 = k$. The signal x is known only through a smaller set of m linear measurements, comprising the measurement vector $\mathbf{y}$, with $m < n$. Arranging the linear measurements as rows in an $m \times n$ acquisition matrix A, this can be described as

$$\mathbf{y} = A\mathbf{x} = AD\gamma, \text{ with } \|\gamma\|_0 = k \tag{1}$$

To recover the original signal x from the knowledge of y, one proceeds by finding the most sparse solution γ to the undetermined system (1) via the optimization problem

$$\mathbf{y} = D \cdot \arg\min \|\gamma\|_0 \text{ s.t. } y = AD\gamma. \tag{2}$$

Multiple variations of this formulation exist, that take into account possible additive noise in the measurements, or use a different norm for quantifying the sparsity level (e.g. the ℓ_1 norm [3]). For most algorithms used in practice, under certain incoherence assumptions on the effective dictionary AD [4], [5], if the decomposition vector γ is sufficiently sparse it can be guaranteed that it will be recovered as the solution of the minimization problem, and thus the true signal $\mathbf{x}$ can be reconstructed.

Of particular interest for this paper are two algorithms for solving (2) directly, the well known Orthogonal Matching Pursuit algorithm (OMP) [6], [7] and the related Least Squares Pursuit [8]. The two algorithms share a common iterative framework, with differences only in how the constraints are handled within an iteration. In this paper we investigate two heuristic methods for fusing them into a single, more robust approach, that combines the advantages of both and is capable of achieving more balanced recovery performance overall. We start by presenting the two algorithms below.

Algorithm 1 Orthogonal Matching Pursuit (OMP)

Iteration $k \leftarrow 0$
$T^0 \leftarrow \varnothing$
repeat
  Update solution via projection on the support

$$\gamma^k \leftarrow \arg\min_{\gamma} \|y - A\gamma\|_2^2 \text{ subject to } \gamma_{T_c^{(k)}} = 0$$

  Compute the gradient of the data-fidelity error:

$$\mathbf{g} = A^T(y - A\gamma^k)$$

  Add to the support the atom corresponding to largest absolute value of the gradient

$$T^{k+1} \leftarrow T^k \cup \arg\max_i |\mathbf{g}_i|$$

until stop criterion
Output γ^k

A. Orthogonal Matching Pursuit

Orthogonal Matching Pursuit (OMP) is a well known sparse recovery algorithm which operates in a greedy fashion, selecting one new atom at every iteration and adding it to the estimated support. The estimated support T^k increases by 1 atom at every iteration, until the support set is large enough final solution is found. Every new atom is selected as follows:

1. Project orthogonally the measurement vector $\mathbf{y}$ on the current support T^k. This results in a candidate vector γ^k which is k-sparse, but has a non-zero data-fidelity error, $\|\mathbf{y} - A\gamma^k\|_2^2 > 0$ (if the data-fidelity error is 0, then γ^k is the optimal solution of (2) and OMP terminates).
2. Compute the gradient of the data-fidelity error, $\mathbf{g} = A^T(\mathbf{y} - A\gamma^k)$, and select the atom corresponding to the largest absolute element in $\mathbf{g}$, which brings the largest expected decrease in the error term. This step can also be understood as computing the correlation of the representation residual $(\mathbf{y} - A\gamma^k)$ with all the atoms in A and choosing the most similar atom.

Algorithm 2 Least Squares Pursuit

Iteration $k \leftarrow 0$
$T^0 \leftarrow \varnothing$
repeat
  Update solution via projection on the affine set
$$\gamma^k \leftarrow \arg\min_{\gamma} \|\gamma_{T_c^{(k)}}\|_2^2 \text{ subject to } y = A\gamma$$
  Compute the gradient of the sparsity error
$$\mathbf{g} \leftarrow \gamma_{T_c^{(k)}}$$
  Add to the support the atom corresponding to largest absolute value of the gradient
$$T^{k+1} \leftarrow T^k \cup \arg\max_i |\mathbf{g}_i|$$
until stop criterion
Output γ^k

The algorithm is summarized in Algorithm 1. Note that the algorithm is described slightly different than usually in the literature, in that an iteration starts from a support set T^k and ends with finding the updated support T^{k+1}. This better points out the similarity with the Least Squares Pursuit algorithm, which follows a similar procedure.

B. Least Squares Pursuit

Least Squares Pursuit (LSP) [8] is a recent alternative to OMP, with a different atom selection rule. Given a support set T^k at the beginning of iteration k, the solution γ^k at the current step does not necessarily have to be computed via orthogonal projection on the set of atoms from T^k, as OMP does. The solution γ^k seeks to minimize two separate error terms: the sparsity error $\|\gamma^k_{T_c^{(k)}}\|_2^2$, and the data-fidelity error $\|y - A\gamma\|_2^2$. Here, $\gamma^k_{T_c^k}$ denotes the restriction of γ^k to the elements outside the support T^k. Thus, the sparsity error is measured as the energy of γ^k that falls outside the current support T^k. The data-fidelity error is measured as the energy of the reconstruction residual.

The orthogonal projection stage of OMP is equivalent to minimizing the data-fidelity error, while keeping the sparsity error strictly equal to 0 (via projecting strictly on the atoms from the support). LSP proposes to exchange the two terms: minimize the sparsity error, while keeping the data-fidelity error strictly equal to 0, via projection on the affine solution space of the equation $\mathbf{y} = A\gamma$. The atom selection rule follows the same exchange: while OMP chooses the largest absolute value from the gradient of the data-fidelity error, LSP chooses the largest absolute value from the gradient of the sparsity error. The full algorithm description thus mirrors OMP, except for the exchanged error terms. Its description is presented in Algorithm 2.

II. Heuristic fusion strategies

Given the two algorithms OMP and LSP, we investigates some fusion strategies aiming to merge both approaches into a single, more-robust algorithm. We are motivated by the following two observations:

- OMP and LSP follow the same iterative procedure, as can be observed from the descriptions in Algorithm 1 and Algorithm 2. The difference lies in the operations *within* an iteration, where the sparsity and data-fidelity error terms are exchanged. Thus, it makes sense to attempt to combine the atom selection rule of both algorithms into a single more robust version, while keeping the common iterative procedure unchanged, i.e. every iteration k starts from a support set T^k and ends with the final support set T^{k+1} with one additional atom.
- The theoretical conditions that guarantee successful recovery for OMP and LSP are different in nature, and thus the two algorithms may behave differently in practice. In particular, one well-known sufficient condition for the success of exact sparse signal recovery with OMP is [5]:
$$\|D^*_{T^c}(D^*_T)^\dagger\|_{\infty,\infty} < 1. \tag{3}$$
whereas a similar condition for LSP is [8]:
$$\|{N_{T_c}}^\dagger N_T\|_{\infty,\infty} < 1, \tag{4}$$
where N is a (transposed) tight frame, the rows of N spanning the null space of D.
A particular difference is that the LSP condition depends only on the null space of the dictionary, and thus should be robust to different scaling of the dictionary rows, for example. On the contrary, the sufficient condition for OMP depends on the actual values of the dictionary atoms and their scalings. Thus, the two algorithms may have different robustness in practice for certain types of signals, and it is worth merging their approaches into a more global robust solution.

The two fusion strategies are described below.

A. Using the average gradient

A first fusion strategy is based on selecting the atom which performs best in both OMP and LSP algorithm, jointly. Let $\mathbf{g}_{OMP}$ designate the gradient vector in OMP and $\mathbf{g}_{LSP}$ the corresponding gradient in LSP (see Algorithm 1 and 2).

Define an average gradient as
$$\mathbf{g}^{avg} = \lambda \cdot \mathbf{g}_{OMP} + (1 - \lambda) \cdot \mathbf{g}_{LSP}, \tag{5}$$
for some weighting parameter λ. The new atom selected for the support is chosen as the largest absolute value of the average gradient:
$$T^{k+1} \leftarrow T^k \cup \arg\max |\mathbf{g}^{avg}_i|$$

This is the same rule as in OMP and LSP, but operating on their averaged gradient. The selected atom is not necessarily one that would be chosen by either OMP or LSP if they would run separately. If the OMP best choice is a poor candidate for LSP and vice-versa, then the selected atom may well be one of the second-best choices, but which performs reasonably well in both algorithms simultaneously. This increases the selection robustness.

B. Using the most confident candidate

A second fusion strategy aims at picking the candidate atom selected by either OMP or LSP, depending on which algorithm is more "confident" in its choice. We define the confidence of OMP or LSP at iteration k as the ratio between the largest and the second largest absolute values of the gradient:

$$C = \frac{\arg\max |\mathbf{g}_i|}{\arg\max_{i \neq \arg\max |\mathbf{g}_i|} |\mathbf{g}_i|}$$

The selected atom is then either the OMP selection or the LSP selection, depending which one has a higher confidenre

$$T^{k+1} \leftarrow T^k \cup \begin{cases} \arg\max |\mathbf{g}_i^{OMP}|, & \text{if } C^{OMP} > C^{LSP} \\ \arg\max |\mathbf{g}_i^{LSP}|, & \text{else} \end{cases}$$

The idea is that if the largest absolute value of the gradient is close to the second largest value, then the selection rule is less robust in case of random errors of the measurement vector, of the dictionary, or if the signal is only approximate sparse. On the contrary, if the largest value stands out with a large margin over the next best one, than the choice will likely be less influenced by errors. This is captured by the confidence value defined before, which ranges from a value of 1 (least confident, the two largest values are identical) to ∞ (most confident, the second-best value is close to 0). In the end, we use the atom which has the largest confidence.

III. Simulation results

In order to test the sparse recovery performance of the fusion approaches proposed, we run a set of tests with various conditions. The first test uses random Gaussian dictionaries of size 200×200, which are more favorable to the LSP algorithm. The second test uses random dictionaries with repeating identical atoms, which is a scenario more favorable for OMP. We want to observe if the fusion approaches are able to mitigate the downsides, and provide consistent good results in both cases.

The first test uses random Gaussian dictionaries of size 200×200. Exact-sparse data are created as random linear combinations of atoms from the dictionary. The data are the acquired with a random Gaussian projection matrix, and then reconstructed with the algorithms under test. The results are displayed in Fig.1 as a function of the relative compression factor $\delta = \frac{m}{n}$ and relative sparsity $\rho = \frac{k}{m}$, where k is the sparsity of the generated signals, m is the number of measurements (rows of the acquisition matrix), and $n = 200$ is the fixed signal dimension. For each value pair (δ, ρ) a number of 100 sparse signals are generated, acquired and reconstructed. A signal is considered perfectly recovered if its relative error is less than $1e^{-6}$. The percentage of perfectly recovered signals is represented by the gray level: white indicates 100% recovered signals, black indicates 0%, and gray levels indicate intermediate percentages.

The results are displayed as a grid approximation of a phase-transition image [9], using a discrete set of values δ, ρ covering the range [0,1] in steps of 0.1. The level of gray indicates the percentage of successfully recovered signals.

As can be seen in Fig. 1, the proposed approaches achieve better performance then OMP, with the first one (averaging the gradients) being essentially identical to LSP.

For the second test, we create exact-sparse signals from dictionaries in which some of the atoms are repeated multiple times, thus illustrating the case of severely coherent dictionaries. These dictionaries negatively impact the LSP algorithm in the following way: while neither of the identical atoms is selected in the support, the minimization of the sparsity error distributes small coefficient values for all of them, since ℓ_2 minimization promotes many small values rather than a single large-valued coefficient. This makes the identical atoms much less likely to be selected in the support, compared to the other atoms which are not repeated, thus introducing bias and negatively impacting the correct atom choice of the algorithm. On the other hand, OMP is not affected by the presence of multiple identical atoms, since all of them have the same correlation with the residual, and the signal is reconstructed identicaly no matter which one of the identical atoms is chosen. For both algorithms, when one of the atoms is selected in the support, all its remaining clones do not impact the reconstruction any further, since the orthogonal projection in both algorithms ensures that identical atoms cannot be chosen again.

The results of this test are presented in Fig. 2, for a dictionary of size 60×1060, but with some 30 atoms repeated around 33 times. As expected, in this case LSP is more affected than OMP and achieves worse behavior. Similar to the first test, we note that the fusion algorithm based on averaging the gradients is able to mitigate the downsides, obtaining recovery performance similar to OMP. The second fusion approach, based on the highest confidence, does not perform very well in this test.

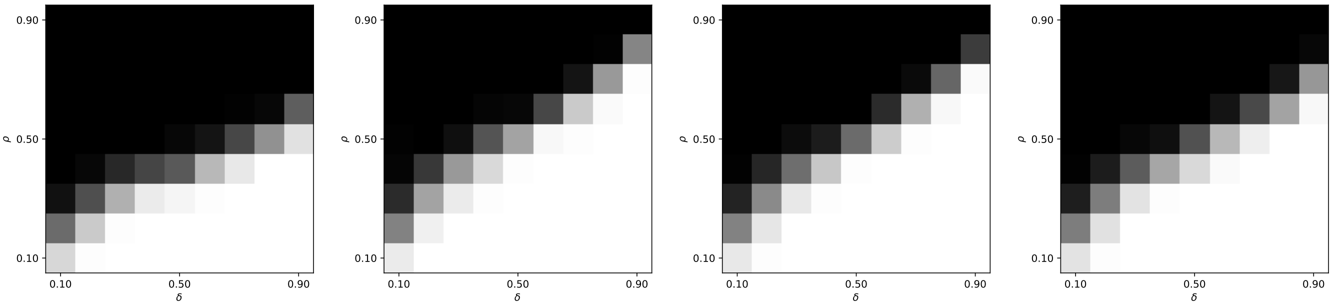
The weighting factor used for the average gradient fusion approach is $\lambda = 0.5$, but similar results were obtained within a larger range around this value.

IV. Conclusions

This paper presents two fusion approaches for combining two similar sparse recovery algorithms, Orthogonal Matching Pursuit and Least Squares Pursuit, into a more robust recovery solution. The first approach proposed is based on selecting the atom which provides maximum benefit for both algorithms simultaneously, while the second approach is to use the atom selected by the algorithm which has larger confidence in its choice.

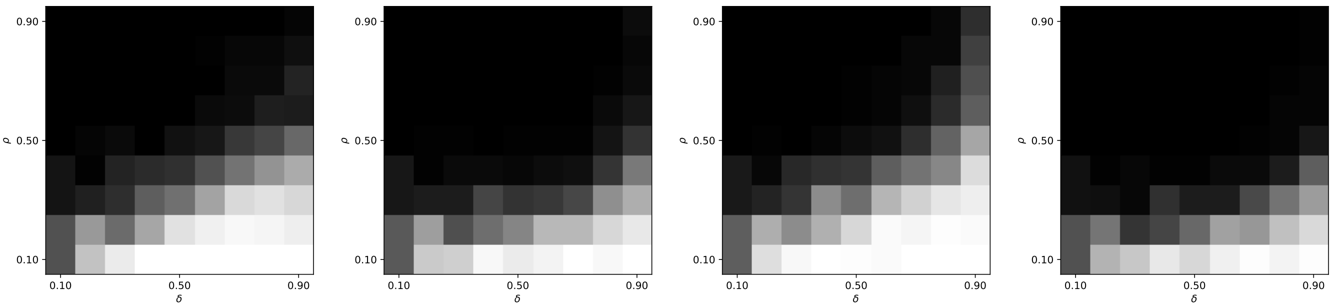
The proposed approaches are tested in two scenarios, with random dictionaries which are favorable for either LSP and OMP, respectively. The results indicate that the fusion algorithms achieve a more balanced performance, being more robust against the adversities in each individual scenario. This suggests that such a combination of different algorithms may provide a better overall recovery process in the general case.

Further work should include more thorough testing of different scenarios, and different types of dictionaries, with



(a) OMP (b) LSP (c) Fusion: average gradient (d) Fusion: largest confidence

Fig. 1: **Test 1.** Percentage of exact-sparse signals reconstructed perfectly with different recovery algorithms, using random Gaussian dictionaries: (a) Orthogonal Matching Pursuit, (b) Least Squares Pursuit, (c) Fusion approach: average gradient, (d) Fusion approach: largest confidence. White indicates 100% recovered signals and black 0%.



(a) OMP (b) LSP (c) Fusion: average gradient (d) Fusion: largest confidence

Fig. 2: **Test 2.** Percentage of exact-sparse signals reconstructed perfectly with different recovery algorithms, using dictionaries with repeated atoms: (a) Orthogonal Matching Pursuit, (b) Least Squares Pursuit, (c) Fusion approach: average gradient, (d) Fusion approach: largest confidence. White indicates 100% recovered signals and black 0%.

special care to avoid introducing dictionary bias in favor of one algorithm or the other, which can alter the interpretation of the results.

ACKNOWLEDGMENT

This work was supported by a grant of Ministery of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P1-1.1-PD-2016-0127, within PNCDI III.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

[3] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, 2005.

[4] E. J. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007.

[5] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, 2004.

[6] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annual Asilomar Conf. on Signals, Systems, and Computers*, 1993, pp. 40–44.

[7] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, pp. 57–98, 1997.

[8] N. Cleju, "Least squares pursuit for sparse signal recovery," in *2017 International Symposium on Signals, Circuits and Systems (ISSCS)*, 2017, pp. 1–4.

[9] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1906, pp. 4273–4293, 2009.