A MAP FRAMEWORK FOR SUPPORT RECOVERY OF SPARSE SIGNALS USING ORTHOGONAL LEAST SQUARES

Shorya Consul, Abolfazl Hashemi, and Haris Vikalo

Department of Electrical and Computer Engineering
University of Texas at Austin, Austin, TX 78712, USA

ABSTRACT

We propose the maximum a posteriori accelerated orthogonal least-squares (MAP-AOLS) algorithm, a novel greedy scheme for accurate reconstruction of a sparse binary signal from its compressed measurements. The algorithm leverages the distributions of the sensing matrix, signal, and noise to find a support set that is optimal in the maximum a posteriori (MAP) sense. This stands in contrast to existing greedy orthogonal least squares (OLS) methods that perform reconstruction without fully exploiting all the available statistical information. In each iteration of the proposed algorithm, the distributions of the sensing matrix, noise, and signal with respect to the support set are used to identify and select the column of the sensing matrix with the largest likelihood ratio of the alternate and null hypotheses. Our extensive simulations demonstrate superiority of MAP-AOLS over existing greedy algorithms with only a minor increase in computational costs. Moreover, the proposed scheme has significantly lower computational complexity than traditional OLS.

Index Terms— Compressed sensing, maximum a posteriori estimation, orthogonal least-squares, greedy algorithm

1. INTRODUCTION

The problem of estimating sparse signals from their compressed measurements is encountered in many areas, including compressive sensing [1, 2], sparse channel estimation [3], and compressed DNA microarrays [4], to name a few. To formalize it, assume a linear measurement model of the form

$$\mathbf{y} = A\mathbf{x} + \mathbf{v}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^n$ denotes the vector of measurements, $A \in \mathbb{R}^{n \times m}$ is the sensing matrix $(n < m)$, $\mathbf{x} \in \mathbb{R}^m$ denotes the sparse signal and $\mathbf{v} \in \mathbb{R}^n$ is the measurement noise. The problem of reconstructing $\mathbf{x}$ can be formulated as an ℓ_0-minimization problem [5, 6]

$$\min_{\mathbf{x}} \quad ||\mathbf{x}||_0 \quad \text{s.t.} \quad ||\mathbf{y} - A\mathbf{x}||^2 < \epsilon,$$

known to be NP-hard; high complexity of finding the exact solution to the sparse recovery problem has spurred development of several classes of computationally efficient albeit suboptimal algorithms. These algorithms can broadly be classified in two categories:

Basis pursuit: Relaxing the ℓ_0 norm to an ℓ_1 norm enables use of convex optimization techniques including LASSO [7] and iterative shrinkage-thresholding [8]. Exact recovery guarantees for the basis pursuit algorithm have been established in both noiseless and noisy settings [9, 10]. However, implementations of these schemes via interior-point methods have polynomial complexity of $\mathcal{O}(m^3)$, limiting their feasibility in practice. While being faster, formulations based on the fast iterative shrinkage-thresholding algorithm (FISTA) and alternating direction method of multiplier (ADMM) remain computationally costly in large-scale settings.

Greedy algorithms: These schemes are significantly faster than basis pursuit algorithms, typically terminating in $O(K)$ iterations where $K < n$ denotes the sparsity level. The most widely used among greedy methods is orthogonal matching pursuit (OMP) [11], an algorithm which in each step seeks to identify and add to the support the column of A having the largest correlation with the residual. Numerous modifications to this simple scheme have been proposed, including gOMP [12] and CoSaMP [2]. An alternative to OMP is the Orthogonal Least Squares (OLS) algorithm [13]. OLS seeks to minimize the residual error with each selected column of A and has been shown to outperform OMP in settings where the columns of the sensing matrix are non-orthogonal [13], while incurring a minor increase in computational complexity.

Note that the only side information aforementioned algorithms utilize is the sparsity level of $\mathbf{x}$. However, in many practical settings other prior information exists and may potentially be exploited to improve performance of sparse reconstruction schemes. This observation motivated development of weighted ℓ_1-minimization [14] and fast Bayesian matching pursuit (FBMP) [15] that incorporate a priori information about the sparse signal $\mathbf{x}$ and the sensing matrix A. These methods were followed by the maximum a posteriori framework for the OMP algorithm in [1]. However, similar extensions do not exist for OLS, motivating the framework we outline in the current paper; given the superior performance of classical OLS over OMP, one expects that the novel Bayesian OLS would be advantageous over Bayesian OMP.

The paper considers reconstruction of sparse binary signals and presents an iterative procedure for greedy selection of the columns of the sensing matrix with the largest log-MAP ratio. To this end, we employ the accelerated OLS [6] strategy which relies on an iterative approach to significantly improve the speed of the classical OLS algorithm. Note that, as pointed out in [14], the assumption that the entries of $\mathbf{x}$ are equally likely sparse is not always true. While [14] attempts to incorporate prior sparsity information via weights assigned to the basis pursuit problem, the MAP framework proposed in the current paper seamlessly incorporates such information in the form of a prior on $\mathbf{x}$. This paves the way for a natural extension where $\mathbf{x}$ is non-binary, but with elements from a known distribution.

Notation and assumption. We assume the linear measurement model (1) where the entries of A are independent, identically distributed Gaussian $\mathcal{N}(0, \frac{1}{n})$ and $\mathbf{v} \in \mathbb{R}^n$ is independent Gaussian noise $\mathcal{N}(0, \sigma^2 I)$. It is furhter assumed that the sparsity of the signal is known to be $k \ll n$. We wish to determine the support set $\mathcal{I}$. For greedy schemes, we shall denote the support set determined in the k^{th} iteration as $\mathcal{S}_k$.

2. MAP-AOLS ALGORITHM

In this section, we describe the proposed framework. We first focus on the case of binary signals with no prior and derive a log-MAP ratio from the distribution of the sensing matrix. Then we extend our results to the case where prior support information is available.

2.1. Binary signals

We focus our attention on the specific case of binary signals, i.e., $x_i = 1$ for all $i \in \mathcal{I}$, and $x_i = 0$ otherwise. The selection criterion for OLS in the k^{th} iteration can be expressed as [6]

$$j_s = \text{argmax}_{j \in \mathcal{I} \setminus S_{k-1}} \mathbf{y}^\top \frac{P_{k-1}^\perp \mathbf{a}_j}{\|P_{k-1}^\perp \mathbf{a}_j\|_2}. \tag{2}$$

For the binary signal, this criterion can equivalently be written as

$$\begin{aligned} j_s &= \text{argmax}_{j \in \mathcal{I} \setminus S_{k-1}} (\sum_i x_i \mathbf{a}_i^\top + \mathbf{v}^\top) \frac{P_{k-1}^\perp \mathbf{a}_j}{\|P_{k-1}^\perp \mathbf{a}_j\|_2} \\ &= \text{argmax}_{j \in \mathcal{I} \setminus S_{k-1}} (\sum_{i \notin S_{k-1} \cup j} x_i \mathbf{a}_i^\top + \mathbf{v}^\top) \frac{P_{k-1}^\perp \mathbf{a}_j}{\|P_{k-1}^\perp \mathbf{a}_j\|_2} \\ &\qquad + x_j \|P_{k-1}^\perp \mathbf{a}_j\|_2, \end{aligned}$$

where the second equality stems from the idempotent property of projection matrices, i.e., ${P_{k-1}^\perp}^2 = P_{k-1}^\perp$. The modification in the summation follows from the definition of $P_{k-1}^\perp$. As the distribution of $\mathbf{a}_j$ is spherically symmetric, we can rotate $P_{k-1}^\perp \mathbf{a}_j$ to a standard unit basis vector, e.g., $\mathbf{u}_1$. This implies that j_s has the same distribution as

$$\text{argmax}_{j \in \mathcal{I} \setminus S_{k-1}} \sum_{i \notin S_{k-1} \cup j} x_i a_i(1) + v(1) + x_j \|P_{k-1}^\perp \mathbf{a}_j\|_2.$$

Let $\mathcal{H}_0$ and $\mathcal{H}_1$ denote the hypotheses that $x_j = 0$ and $x_j = 1$, respectively. Moreover, let us define $z_j^k = \sum_{i \notin S_{k-1} \cup j} x_i a_i(1) + v(1) + x_j \|P_{k-1}^\perp \mathbf{a}_j\|_2$. Therefore,

$$\mathcal{H}_0 : z_j^k = \sum_{i \in \mathcal{I} \setminus S_{k-1}} a_i(1) + v(1),$$

and

$$\mathcal{H}_1 : z_j^k = \sum_{i \in \mathcal{I} \setminus S_{k-1}} a_i(1) + v(1) + \|P_{k-1}^\perp \mathbf{a}_j\|_2.$$

Under $\mathcal{H}_0$, z_j^k is simply a sum of independent Gaussian random variables and thus has a Gaussian distribution with zero mean and variance $\sigma_0^2 = \frac{K-(k-1)}{n} + \sigma_n^2$, where K denotes the sparsity level of $\mathbf{x}$. By a similar argument, the conditional distribution of z_j^k under $\mathcal{H}_1$ is a Gaussian with variance $\frac{K-k}{n} + \sigma_n^2$. The following lemma specifies the distribution of $\|P_{k-1}^\perp \mathbf{a}_j\|_2$ needed for the MAP-based selection.

Lemma 1. The cumulative distribution function $\|P_{k-1}^\perp \mathbf{a}_j\|_2$ is given by

$$F_{\|P_{k-1}^\perp \mathbf{a}_j\|_2}(x) = \frac{\gamma(\frac{n-k+1}{2}, \frac{nx^2}{2})}{\Gamma(\frac{n-k+1}{2})}.$$

Proof. Let $\mathbf{b}_{k-1} = P_{k-1}\mathbf{a}_j \in \mathcal{L}_{k-1}$ where $\mathcal{L}_{k-1}$ is the subspace spanned by the columns of A selected and placed in $\mathcal{S}_{k-1}$. There exists an orthonormal set of basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{k-1}\}$ that spans $\mathcal{L}_{k-1}$. Let U denote a unitary matrix such that $\{U\mathbf{u}_1, U\mathbf{u}_2, ..., U\mathbf{u}_{k-1}\} = \{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{k-1}\}$ where $\mathbf{e}_i$ denote the standard basis vectors. Invoking the spherical symmetry of the Gaussian distribution once again, we note that the distribution of $\mathbf{b}_{k-1}$ is identical to that of $U\mathbf{b}_{k-1}$ which is simply the sum of the first $(k-1)$ elements of $\mathbf{a}_j$. Hence, $\|P_{k-1}\mathbf{a}_j\|_2$ has a scaled chi distribution with $k-1$ degrees of freedom. A similar argument can be applied to conclude that $\|\mathbf{a}_j\|_2$ has a chi distribution with n degrees of freedom [1]. We obtain the desired result from the identity $\|P_{k-1}^\perp \mathbf{a}_j\|_2^2 = \|\mathbf{a}_j\|_2^2 - \|P_{k-1}\mathbf{a}_j\|_2^2$. The factor of n in the numerator follows from the fact that each element of $\mathbf{a}_j$ is a Gaussian random variable with variance $\frac{1}{n}$. ∎

Given the distribution of $\|P_{k-1}^\perp \mathbf{a}_j\|_2$, we can find the distribution of z_j^k under $\mathcal{H}_1$ using the law of total expectation. However, this is analytically intractable. Instead, we observe that

$$\mathbb{E}[\|P_{k-1}^\perp \mathbf{a}_j\|_2] = \sqrt{\frac{2}{n}} \frac{\Gamma(\frac{n-k+2}{2})}{\Gamma(\frac{n-k+1}{2})}.$$

It can be shown the limit of this mean as n grows large is $\sqrt{\frac{n-k+1}{n}}$. This result is useful as in many applications the sparsity level $K \ll n$ Hence, we shall approximate the distribution of z_j^k under $\mathcal{H}_1$ as a Gaussian with mean $\mu_k = \sqrt{\frac{n-k+1}{n}}$ and variance $\sigma_1^2 = \frac{K-k}{n} + \sigma_n^2$. We now proceed to compute the log-MAP ratio as

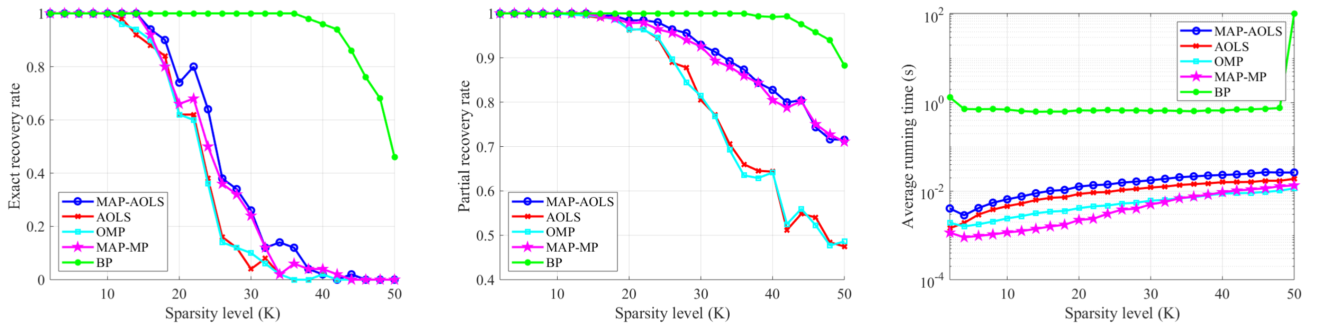
$$\begin{aligned} \Lambda_j^k &= \log\left(\frac{P(z_j^k|x_j = 1)}{P(z_j^k|x_j = 1}\frac{P(x_j = 1)}{P(x_j = 0)}\right) \\ &= \frac{(z_j^k)^2}{2\sigma_0^2} - \frac{(z_j^k - \mu_k)^2}{2\sigma_1^2} + \log\left(\frac{P(x_j = 1)}{P(x_j = 0)}\right) \\ &= \frac{(z_j^k)^2}{2(\frac{K-k+1}{n} + \sigma_n^2)} - \frac{(z_j^k - \mu_k)^2}{2(\frac{K-k}{n} + \sigma_n^2)} + \log\left(\frac{P(x_j = 1)}{P(x_j = 0)}\right). \end{aligned} \tag{3}$$

The last term can be dropped under the assumption that all the indices are equally likely to be part of the support set. Instead of optimizing (2), the proposed MAP-OLS algorithm at each iteration greedily selects the column which maximizes the log-MAP ratio (3) . The remaining steps remain unchanged from the standard OLS algorithm. However, the method implemented in this paper is based on the accelerated OLS (AOLS) algorithm (see Algorithm 1) which speeds-up the column selection procedure of standard OLS by iteratively computing and updating a set of orthogonal vectors. This reduces computational complexity from $\mathcal{O}(Kmn^2)$ to $\mathcal{O}(Kmn)$ by obviating the need to compute the projection $P_{k-1}^\perp \mathbf{a}_j$ for all the columns of A in every iteration.

The quantity of interest z_j^k is the ℓ_2-norm of the corresponding $\mathbf{q}_j$ in Algorithm 1, i.e., $\|\mathbf{q}_j^{(k)}\|_2 = z_j^k$. The threshold, ϵ, provides an alternate stopping criterion for the algorithm.

2.2. Binary signals with prior support information

In the derivation of the log-MAP ratio [cf. (3)], we made an assumption at the final step that all the support indices are equally likely. This is a common assumption in literature. However, as suggested in [12], this may not always hold. Adopting the conventions in [12], we model the signal $\mathbf{x}$ as having indices that form two sets K_1 and K_2 with the probability of each element being (independently) nonzero in the sets taken to be p_1 and p_2, respectively. The log-MAP ratio provides a natural venue for incorporating this information. However, in this model we no longer have information about the sparsity

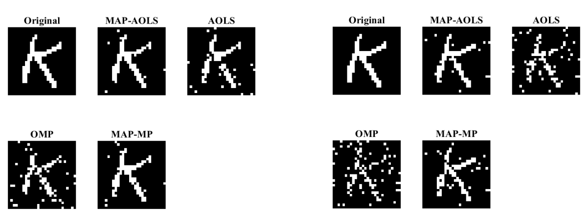


(a) Exact recovery rate (b) Partial recovery rate (c) Running time

Fig. 1: *Performance comparison of different sparse reconstruction schemes on a simulated data from noise-free linear random measurements with* $m = 256$ *and* $n = 128$.

Algorithm 1 Accelerated Orthogonal Least Squares (AOLS)

Input: $\mathbf{y}$, A, K, threshold (ϵ)
Output: Support $\mathcal{S}_K$, signal estimate $\hat{\mathbf{x}}_\mathbf{K}$
Initialize: $i = 0$, $\mathbf{r}_0 = \mathbf{y}$, $\mathcal{S}_0 = \phi$, $\mathbf{t}_j^{(i)} = \mathbf{a}_j$, $\mathbf{q}_j = \frac{\mathbf{a}_j^\top \mathbf{y}}{\mathbf{a}_j^\top \mathbf{a}_j}\mathbf{a}_j, \forall j$
while $\|\mathbf{r}_i\|_2 > \epsilon$ and $i \leq K$ **do**
  Select j_s corresponding to the largest Λ_j^k
  $i \leftarrow i + 1$
  $\mathcal{S}_i = \mathcal{S}_{i-1} \cup \{j_s\}$
  $\mathbf{u}_i = \mathbf{q}_{j_s}$, $\mathbf{r}_i = \mathbf{r}_{i-1} - \mathbf{u}_i$
  $\mathbf{t}_j^i = \mathbf{t}_j^{i-1} - \frac{\mathbf{t}_j^{i-1\top}\mathbf{u}_i}{\|\mathbf{u}_i\|_2^2}\mathbf{u}_i$
end while
$\hat{\mathbf{x}}_K = A_{\mathcal{S}_K}^\dagger \mathbf{y}$



(a) noise-free measurements (b) $\sigma_n^2 = 4.5 \times 10^{-3}$

Fig. 2: *Sparse image reconstruction on a sample image from EMNIST dataset [16].*

level K. Therefore, we propose to use $K = |K_1|p_1 + |K_2|p_2$ as a proxy for the sparsity level. The strong law of large numbers renders this a good approximation for large m, which is the regime in which the sparse recovery problem is usually considered. It is also noteworthy that a more general model with multiple sets could easily be adopted with negligible increase in computational complexity. Under this two-set model, the log-MAP ratio in the k^{th} iteration can be written as

$$\Lambda_j^k = \frac{(z_j^k)^2}{2(\frac{K-k+1}{n} + \sigma_n^2)} - \frac{(z_j^k - \mu_k)^2}{2(\frac{K-k}{n} + \sigma_n^2)} + \log(\frac{p_1}{1-p_1})1_{\{j \in K_1\}} + \log(\frac{p_2}{1-p_2})1_{\{j \in K_2\}}. \quad (4)$$

2.3. Connection to OLS

The expression for the log-MAP ratio suggests that the heuristic criterion for index selection in OLS is not optimal in the MAP sense. However, if $\sigma_1^2 = \sigma_0^2 = \sigma^2$ and the prior terms are ignored, i.e., all support indices are equally likely to be non-zero, the log-MAP ratio simplifies to

$$\Lambda_j^k = \frac{(z_j^k)^2}{2\sigma^2} - \frac{(z_j^k - \mu_k)^2}{2\sigma^2} = \frac{2z_j^k\mu_k - \mu_k^2}{2\sigma^2}.$$

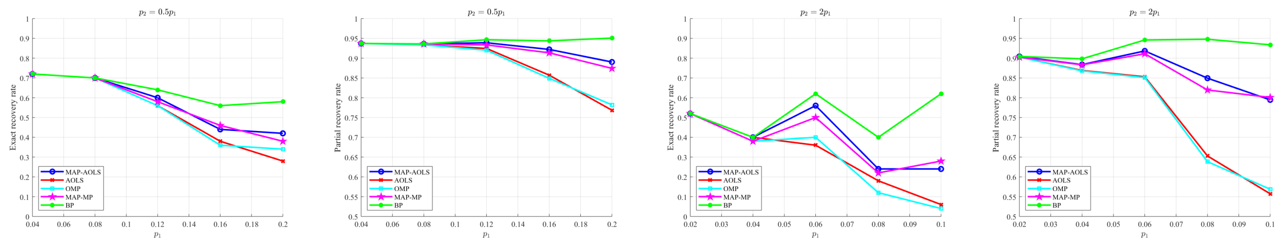
Under this assumption (i.e. $\sigma_1^2 = \sigma_0^2 = \sigma^2$), the MAP-AOLS selection criterion coincides with that of OLS. Note, however, that for the case of binary signals this assumption does not hold. Nonetheless, $\sigma_1^2 \approx \sigma_0^2$ in the low-SNR regime (i.e., when the noise power σ_n^2 is large).

3. SIMULATIONS

The performance of the proposed MAP-AOLS algorithm is compared to that of AOLS, OMP [11], MAP-MP [1], and Basis Pursuit (BP). As is typical of benchmarking tests, BP was implemented using CVX [17, 18]. The exact recovery rate (ERR) and partial recovery rate (PRR) are used as metrics to characterize the accuracy of support recovery of each algorithm. The average running time serves as a measure of computational complexity of each algorithm. These metrics were evaluated by averaging the results over 50 independent trials. In each simulation, the sensing matrix $A \in \mathbb{R}^{n \times m}$ is randomly generated with its i.i.d. elements drawn from the Gaussian distribution $\mathcal{N}(0, \frac{1}{n})$. Threshold ϵ was set to 10^{-13} for all the algorithms.

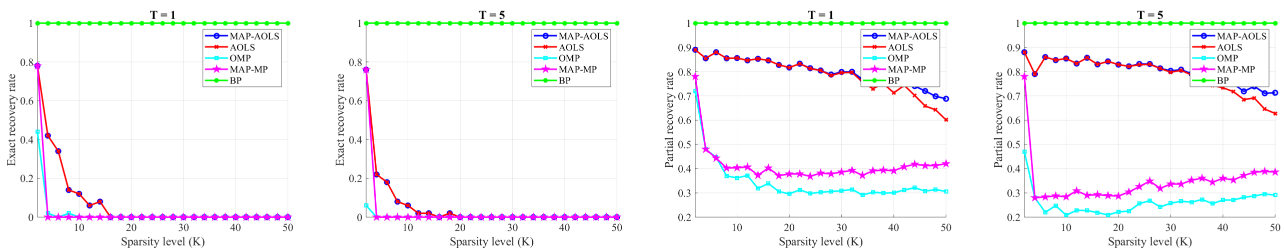
3.1. Reconstruction of binary signals in absence of noise

Fig. 1(a) and Fig. 1(b) illustrate the ERR and PRR for the reconstruction from noise-free measurements for $m = 256$ and $n = 128$. For each sparsity level, K support indices are chosen uniformly to form $\mathbf{x}$. MAP-AOLS and MAP-MP lead to the highest ERR for each value of K and an even more substantial improvement in the PRR over other greedy algorithms. The corresponding average running time comparison in Fig. 1(c) show that MAP-AOLS is only slightly slower than other greedy algorithms despite being considerably more accurate.



(a) ERR, $p_2 = 0.5p_1$ (b) PRR, $p_2 = 0.5p_1$ (c) ERR, $p_2 = 2p_1$ (d) PRR, $p_2 = 2p_1$

Fig. 3: *Accuracy of support detection with prior information. Support indices were chosen from* K_1 *with probability* p_1 *and* K_2 *with probability* $p_2 1$*, with* $|K_1| = 64$*,* $|K_2| = 192$*,* $m = 256$ *and* $n = 128$*.*



(a) ERR, $T = 1$ (b) ERR, $T = 5$ (c) PRR, $T = 1$ (d) PRR, $T = 5$

Fig. 4: *Accuracy of support detection with hybrid dictionaries. Dictionary columns were set as* $\mathbf{A}_j = \frac{b_j + t_j \mathbf{1}}{||b_j + t_j \mathbf{1}||_2}$ *where* $t_j \sim \mathcal{U}(0, T)$ *with* $m = 256$ *and* $n = 128$*.*

3.2. Reconstruction of binary sparse image

Next, we evaluate the performance of the proposed algorithm in the context of recovering a sparse binary image. A 28×28 image from the EMNIST dataset [16] is converted into a binary image by thresholding it at a pixel value of 150. Then, a 392×784 sensing matrix with i.i.d. entries drawn from a Gaussian distribution $\mathcal{N}(0, \frac{1}{392})$ is used to generate 392 linear random measurements. Finally, sparse recovery is performed separately with both noise-free and noisy measurements where the entries of the additive noise vector are assumed to be white Gaussian $\mathcal{N}(0, 0.0045)$. A single instance of this reconstruction is displayed in Fig. 2 where it is clear that MAP-AOLS delivers the best reconstruction performance; similar results are obtained in other instances of the task. Note that we exclude comparison with BP from this experiment due to its high computational cost.

3.3. Binary signal reconstruction with support prior

We consider the noise-free measurements, and set model parameters to $m = 256$ and $n = 128$. However, we now assume a two-set model for the support indices where, without a loss of generality, we select the first $|K_1|$ indices to be part of K_1 and the remaining indices to be part of K_2. For the simulations, we have chosen $|K_1| = \frac{m}{4} = 64$. To maintain the sparse structure of the signal, we would like $p_1|K_1| + p_2|K_2|$ to be small. The recovery rates for various combinations of (p_1, p_2) are shown in Fig 3.

On one hand, small values of p_1 lead to small K and the effect of each "missed" support index is more significant, leading to lower accuracy. On the other hand, recovery rates deteriorate for higher values of K. These conflicting effects explain the non-monotonic property of the observed trends. Despite this phenomenon, the high values of PRR attained by the greedy algorithms indicate that they recover most of the support indices. MAP-AOLS and MAP-MP demonstrate a significant improvement compared to AOLS and OMP in terms of both ERR and PRR (particularly in terms of the latter). They further exhibit slower drop-off at higher values of K.

3.4. Recovery with hybrid dictionaries

Finally, we analyze the performance of the algorithms when A is a hybrid dictionary [19]. Hybrid dictionaries are frequently used to examine scenarios where the amount of dependencies among columns of A is large (i.e., matrices with large mutual incoherence properties). To this end, we set $\mathbf{A}_j = \frac{b_j + t_j \mathbf{1}}{||b_j + t_j \mathbf{1}||_2}$ where $t_j \sim \mathcal{U}(0, T)$; other parameters remain unchanged. Note that T controls the extent of correlation of the columns of A; as T increases, the dictionary becomes more correlated. Fig 4 indicates that AOLS and MAP-AOLS are significantly more robust with respect to correlation in A compared to other greedy schemes.

4. CONCLUSION

We proposed a MAP framework for recovery of binary sparse signals which builds upon the OLS algorithm to enable identification of the support that is optimal in the MAP sense. The proposed algorithm was compared to AOLS, OMP, MAP-MP and BP in both noise-free and noisy settings; the results demonstrate that MAP-AOLS outperforms competing greedy schemes in both scenarios. The MAP-AOLS algorithm further exhibits superior performance when reconstructing sparse signals with non-uniform support.

The presented analysis is restricted to binary signals and could be extended to signals with active elements from known distributions. It is of interest to pursue techniques for further reduction of the computational complexity.

5. REFERENCES

[1] Namyoon Lee, "Map support detection for greedy sparse signal recovery algorithms in compressive sensing," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 4987–4999, 2016.

[2] Deanna Needell and Joel A Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[3] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 164–174, 2010.

[4] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed dna microarrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 275–285, 2008.

[5] Mudassir Masood and Tareq Y Al-Naffouri, "Sparse reconstruction using distribution agnostic bayesian matching pursuit," *IEEE Trans. Signal Process.*

[6] Abolfazl Hashemi and Haris Vikalo, "Accelerated orthogonal least-squares for large-scale sparse reconstruction," *Digital Signal Processing*, vol. 82, pp. 91–105, 2018.

[7] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[8] Ingrid Daubechies, Michel Defrise, and Christine De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[9] Emmanuel J Candes and Terence Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[10] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[11] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[12] Jian Wang, Seokbeop Kwon, and Byonghyo Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, 2012.

[13] Thomas Blumensath and Mike E Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares," 2007.

[14] M Amin Khajehnejad, Weiyu Xu, A Salman Avestimehr, and Babak Hassibi, "Weighted 1 minimization for sparse recovery with prior information," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*. IEEE, 2009, pp. 483–487.

[15] Philip Schniter, Lee C Potter, and Justin Ziniel, "Fast bayesian matching pursuit," in *Information Theory and Applications Workshop, 2008*. IEEE, 2008, pp. 326–333.

[16] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and André van Schaik, "Emnist: an extension of mnist to handwritten letters," *arXiv preprint arXiv:1702.05373*, 2017.

[17] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," `http://cvxr.com/cvx`, Mar. 2014.

[18] Michael Grant and Stephen Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited, 2008, `http://stanford.edu/~boyd/graph_dcp.html`.

[19] Charles Soussen, Rémi Gribonval, Jérôme Idier, and Cédric Herzet, "Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3158–3174, 2013.

[20] Jian Wang and Ping Li, "Recovery of sparse signals using multiple orthogonal least squares.," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2049–2062, 2017.

[21] Rachel Ward, "Compressed sensing with cross validation," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5773–5782, 2009.