

SPARSE LINEAR REGRESSION VIA GENERALIZED ORTHOGONAL LEAST-SQUARES

Abolfazl Hashemi and Haris Vikalo

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

ABSTRACT

The Orthogonal Least Squares (OLS) algorithm sequentially selects columns of the coefficient matrix to greedily find an approximate sparse solution to an underdetermined system of linear equations. In this paper, conditions under which OLS recovers sparse signals from a low number of random linear measurements with probability arbitrarily close to one are stated. Moreover, a computationally efficient generalization of Orthogonal Least-Squares which relies on a recursive relation between the components of the optimal solution to select L columns at each step and solve the resulting over-determined system of equations is proposed. This generalized OLS algorithm is empirically shown to outperform existing greedy algorithms broadly used in literature.

Index Terms— linear regression, compressed sensing, greedy algorithm, orthogonal least-squares

1. INTRODUCTION

The problem of finding sparse solution to an underdetermined system of linear equations arises in a number of practical scenarios. Examples include compressed sensing [1], sparse channel estimation in communication systems [2], compressive DNA microarrays [3] as well as a number of other applications in signal processing and machine learning [4–7]. Consider the linear measurement model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^n$ denotes the vector of observations, $\mathbf{A} \in \mathbb{R}^{n \times m}$ is the coefficient matrix (i.e., a collection of features) assumed to be full rank, $\mathbf{e} \in \mathbb{R}^n$ is the additive observation noise vector, and $\mathbf{x} \in \mathbb{R}^m$ is a vector known to have at most k non-zero components (i.e., k is the sparsity level of \mathbf{x}). We are interested in finding a sparse approximation to \mathbf{x} ; in particular, we would like to solve the so-called l_0 -constrained least-squares

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\mathbf{x}\|_0 \leq k. \quad (2)$$

The number of possible locations of non-zero entries in \mathbf{x} scales combinatorially with n which renders (2) computationally challenging; in fact, the problem is NP-hard. To enable computationally efficient search for sparse \mathbf{x} approximating

(1), the non-convex l_0 -norm-constrained optimization (2) can be replaced by a sparsity-promoting l_1 -norm optimization

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \varepsilon, \quad (3)$$

where $\varepsilon > 0$ is a predetermined measure of noise power. In the noise-free scenario where \mathbf{e} in (1) and ε in (3) are both zero and where \mathbf{A} satisfies certain properties, it is known that a sufficiently sparse \mathbf{x} can be reconstructed exactly [8]. However, while the convexity of l_1 -norm enables finding the optimal solution to the reformulated sparse vector recovery problem, the complexity of doing so (by means of, e.g., iterative shrinkage-thresholding algorithms such as [9], or alternating direction method of multipliers [10]) is often prohibitive when one deals with high dimensional data. For this reason, a number of fast greedy heuristics that attempt to solve (2) directly by successively identifying columns of \mathbf{A} which correspond to non-zero components of \mathbf{x} have been proposed [11, 12]. Among those, particular attention has been paid to the orthogonal matching pursuit (OMP) algorithm [13, 14] which has an intuitive geometric interpretation and is characterized by high speed and competitive performance; numerous modifications of OMP that explore the trade-off between accuracy and speed have been proposed in literature [15, 16]. A related Orthogonal Least-Squares (OLS) method [17], proposed as an identification algorithm for parameter estimation of generally multivariable non-linear systems which are linear in parameters, has recently been employed in compressed sensing [18]. In general, OLS outperforms OMS in settings where the columns of \mathbf{A} are non-orthogonal but it does so at a moderate increase in complexity. The existing analysis and performance guarantees for OLS are limited to the case of non-random measurements [18–20].

In this paper, we provide a result establishing that in the noiseless scenario where the coefficient matrix is drawn at random from a Gaussian or a Bernoulli distribution, with $\mathcal{O}(k \log(m))$ linear random measurements OLS guarantees recovery of \mathbf{x} with high probability. This result is comparable to those previously provided for OMP [13, 21]. Moreover, we propose a generalization of OLS, the Generalized Orthogonal Least-Squares (GOLS), an efficient algorithm which relies on a recursive relation between the components of the optimal solution to (1) to select a pre-determined number of columns and provide performance superior to existing methods.

2. PERFORMANCE GUARANTEE FOR ORTHOGONAL LEAST-SQUARES

The OLS algorithm sequentially projects columns of \mathbf{A} onto a residual vector and each time selects the column that leads to the smallest residual norm. Specifically, OLS chooses a new index j_s as

$$j_s = \operatorname{argmin}_{j \in \mathcal{I}} \left\| \mathbf{y} - \mathbf{A}_{\mathcal{S}_{i-1} \cup \{j\}} \mathbf{A}_{\mathcal{S}_{i-1} \cup \{j\}}^\dagger \mathbf{y} \right\|_2,$$

where \mathcal{I} is the set of indices that are not yet selected. This procedure is computationally more expensive than OMP since in addition to solving the least-square problem to update the residual vector, orthogonal projection of each column needs to be found at each step of OLS. Perhaps in part due to this increase in complexity, OLS has not played as prominent role in sparse signal recovery literature as OMP did.

Note that the performances of OLS and OMP are identical when the columns of \mathbf{A} are orthogonal.¹ It is beneficial to further clarify the difference between OMP and OLS. In each iteration of OMP, an element that best correlates with the current residual is chosen. OLS, on the other hand, selects a column that has the largest portion which is inexpressible by previously selected columns which, in turn, minimizes related approximation error.

The following theorem states that for Gaussian and Bernoulli matrices with normalized columns, which are often considered in compressed sensing problems, in the noiseless scenario OLS is with high probability capable of the exact recovery of sparse signals if the number of measurements grows linearly with the sparsity level and logarithmically with the dimension of the unknown signal.

Theorem 2.1. *Suppose that $\mathbf{x} \in \mathbb{R}^m$ is an arbitrary sparse vector with sparsity level k . Consider a random matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ such that its entries are drawn uniformly and independently from either $\mathcal{N}(0, 1/n)$ or $\{+1/\sqrt{n}, -1/\sqrt{n}\}$. Given the noiseless observation $\mathbf{y} = \mathbf{A}\mathbf{x}$, the OLS algorithm can recover \mathbf{x} in k iterations with probability of success exceeding $1 - \delta$ if $n = \mathcal{O}(k \log(m/\delta))$ for some $0 < \delta < \gamma$, where γ is a positive constant which is independent of n , m , and k .*

The proof, which exploits the fact that the columns of \mathbf{A} are spherically symmetric random vectors and relies on Johnson-Lindenstrauss lemma, is omitted for brevity.

3. GENERALIZED ORTHOGONAL LEAST-SQUARES

To formulate generalized OLS, we start by establishing a recursive relation between the components of the optimal solution to the l_0 -constrained least-squares problem. Let

¹In fact, orthogonality of the columns of \mathbf{A} leads to a modular objective function in (2), implying optimality of both methods.

\mathbf{B}_i denote the sub-matrix of \mathbf{A} constructed by selecting i of its columns and let $\mathbf{P}_i = \mathbf{B}_i \mathbf{B}_i^\dagger$ denote the projection matrix onto the span of the columns of \mathbf{B}_i , where $\mathbf{B}_i^\dagger = (\mathbf{B}_i^T \mathbf{B}_i)^{-1} \mathbf{B}_i^T$ is the Moore-Penrose pseudo-inverse of \mathbf{B}_i . Then, after appending \mathbf{B}_i with another column vector \mathbf{a} to form $\mathbf{B}_{i+1} = [\mathbf{B}_i \ \mathbf{a}]$, we can write

$$\begin{aligned} \mathbf{P}_{i+1} &= \mathbf{B}_{i+1} (\mathbf{B}_{i+1}^T \mathbf{B}_{i+1})^{-1} \mathbf{B}_{i+1}^T \\ &= [\mathbf{B}_i \ \mathbf{a}] \begin{bmatrix} \mathbf{B}_i^T \mathbf{B}_i & \mathbf{B}_i^T \mathbf{a} \\ \mathbf{a}^T \mathbf{B}_i & \mathbf{a}^T \mathbf{a} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B}_i^T \\ \mathbf{a}^T \end{bmatrix} \\ &\stackrel{(a)}{=} [\mathbf{B}_i \ \mathbf{P}_i^\perp \mathbf{a}] \begin{bmatrix} (\mathbf{B}_i^T \mathbf{B}_i)^{-1} & 0 \\ 0 & (\mathbf{a}^T \mathbf{P}_i^\perp \mathbf{a})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{B}_i^T \\ \mathbf{a}^T \mathbf{P}_i^\perp \end{bmatrix} \\ &\stackrel{(b)}{=} \mathbf{P}_i + \frac{\mathbf{P}_i^\perp \mathbf{a} \mathbf{a}^T \mathbf{P}_i^\perp}{\|\mathbf{P}_i^\perp \mathbf{a}\|_2^2} \end{aligned} \quad (4)$$

where (a) follows after the LDU decomposition of the intermediate matrix inverse, i.e., we use the identity [22]

$$\begin{bmatrix} \mathbf{A} & \mathbf{E} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1} \mathbf{E} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & 0 \\ 0 & \mathbf{\Delta}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{C} \mathbf{A}^{-1} & \mathbf{I} \end{bmatrix}$$

where we identify $\mathbf{A} = \mathbf{B}_i^T \mathbf{B}_i$, $\mathbf{E} = \mathbf{B}_i^T \mathbf{a}$, $\mathbf{C} = \mathbf{a}^T \mathbf{B}_i$, $\mathbf{D} = \mathbf{a}^T \mathbf{a}$, and $\mathbf{\Delta} = \mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{E}$, and introduce $\mathbf{P}_i^\perp = \mathbf{I} - \mathbf{P}_i$. The identity (b) follows from the idempotent property of the projection matrix. Alternatively, we write (4) as

$$\mathbf{P}_{i+1}^\perp = \mathbf{P}_i^\perp - \frac{\mathbf{P}_i^\perp \mathbf{a} \mathbf{a}^T \mathbf{P}_i^\perp}{\|\mathbf{P}_i^\perp \mathbf{a}\|_2^2}. \quad (5)$$

Note that, (5) is related to order-recursive least-squares [23]. However, this specific derivation makes it suitable for iterative sparse reconstruction applications.

Now, the OLS algorithm in each step selects a column with index j_s from the set \mathcal{I} of the previously non-selected columns according to

$$\begin{aligned} j_s &= \operatorname{argmin}_{j \in \mathcal{I}} \left\| \mathbf{y} - \mathbf{A}_{\mathcal{S}_{i-1} \cup \{j\}} \mathbf{A}_{\mathcal{S}_{i-1} \cup \{j\}}^\dagger \mathbf{y} \right\|_2 \\ &\stackrel{(a)}{=} \operatorname{argmin}_{j \in \mathcal{I}} \|(\mathbf{I} - \mathbf{P}_i) \mathbf{y}\|_2^2 \\ &= \operatorname{argmin}_{j \in \mathcal{I}} \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{P}_i \mathbf{y} - \mathbf{y}^T \mathbf{P}_i^T \mathbf{y} + \mathbf{y}^T \mathbf{P}_i^T \mathbf{P}_i \mathbf{y} \\ &\stackrel{(b)}{=} \operatorname{argmin}_{j \in \mathcal{I}} \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{P}_i \mathbf{y} \\ &\stackrel{(c)}{=} \operatorname{argmax}_{j \in \mathcal{I}} \mathbf{y}^T \mathbf{P}_{i-1} \mathbf{y} + \mathbf{y}^T \frac{\mathbf{P}_{i-1}^\perp \mathbf{a}_j \mathbf{a}_j^T \mathbf{P}_{i-1}^\perp}{\|\mathbf{P}_{i-1}^\perp \mathbf{a}_j\|_2^2} \mathbf{y} \\ &\stackrel{(d)}{=} \operatorname{argmax}_{j \in \mathcal{I}} \frac{\|\mathbf{y}^T \mathbf{P}_{i-1}^\perp \mathbf{a}_j\|_2^2}{\|\mathbf{P}_{i-1}^\perp \mathbf{a}_j\|_2^2} = \operatorname{argmax}_{j \in \mathcal{I}} \left| \mathbf{y}^T \frac{\mathbf{P}_{i-1}^\perp \mathbf{a}_j}{\|\mathbf{P}_{i-1}^\perp \mathbf{a}_j\|_2} \right| \end{aligned} \quad (6)$$

where (a) follows from the definition of \mathbf{P}_i , (b) is due to \mathbf{P}_i^\perp being an idempotent projection matrix, (c) follows from eq.

(4), and (d) is due to the fact that $\mathbf{y}^T \mathbf{P}_{i-1} \mathbf{y}$ is not a function of the optimization variable.

We propose a straightforward extension of OLS which selects multiple (say, L) columns of \mathbf{A} in each step rather than choosing a single column, ultimately replacing the underdetermined $n \times m$ system of equations by an overdetermined $Lk \times k$ one. This strategy is motivated by the observation that the candidate columns whose projection onto the space orthogonal to that spanned by the previously selected columns is strongly correlated with the observation vector but not chosen in the current step of OLS will likely be selected in subsequent steps of the algorithm; therefore, selecting several “good” candidates in each step accelerates the selection procedure and enables sparse reconstruction with fewer steps (and, therefore, fewer calculations of the mutual correlations needed to perform the selection). More specifically, the proposed generalized OLS algorithm performs the following: in each step, the algorithm select L columns of matrix \mathbf{A} such that their normalized projection onto the orthogonal complement of the subspace spanned by previously chosen columns have the highest correlation with the observation vector among the non-selected columns. After such columns are identified, we update the orthogonal projection matrix by repeatedly applying (5) L times. We continue until a stopping criterion is met. Generalized orthogonal least-squares algorithm is formalized as Algorithm 1.

3.1. Computational complexity

To analyze the computational complexity of GOLS, note that its first step involves a matrix-vector multiplication and a vector inner product; the computational cost of these two is dominated by the former and thus requires $\mathcal{O}(mn^2)$ operations. The other operations are those in Step 3 where a matrix-vector multiplication and a matrix addition, needing $\mathcal{O}(n^2)$ calculation, need to be repeated $\min\{k, \lfloor \frac{n}{L} \rfloor\}$ times. Therefore, the aggregate cost of operations in this step is $\mathcal{O}(mkn^2)$. Finally, finding the estimate $\hat{\mathbf{x}}$ entails solving a least-squares problem which can be implemented with a small cost $\mathcal{O}(kn)$ by relying on a QR factorization of $\mathbf{A}_{\mathcal{S}_k}$. Therefore, the total complexity of the algorithm is $\mathcal{O}(mkn^2 + kn)$.

4. SIMULATION RESULTS

To evaluate the algorithm, we compared its performance with four other sparse recovery algorithms as a function of the sparsity level k . In particular, we considered OMP, OLS, l_1 -norm minimization [8], and Least Absolute Shrinkage and Selection Operator (LASSO) [24]. As typically done in benchmarking tests [25], we used CVX [26] to implement the l_1 minimization and LASSO. The tuning parameter in LASSO is found by means of 10-fold cross validation. We draw entries of the coefficient matrix \mathbf{A} from two distributions. First, we generate entries of \mathbf{A} by drawing independently from a Gaus-

Algorithm 1 Generalized Orthogonal Least-Squares

Input: observation \mathbf{y} , coefficient matrix \mathbf{A} , sparsity k
Output: recovered support \mathcal{S}_k , estimated signal $\hat{\mathbf{x}}_k$
Initialize: $\mathcal{S}_0 = \emptyset, \mathbf{P}_0^\perp = \mathbf{I}, \mathcal{I} = \{1, 2, \dots, m\}$

for $i = 1$ to $\min\{k, \lfloor \frac{n}{L} \rfloor\}$ **do**

1. Select $\{i_{s_1}, \dots, i_{s_L}\}$ corresponding to L largest terms: $\left| \mathbf{y}^T \frac{\mathbf{P}_{i-1}^\perp \mathbf{a}_j}{\|\mathbf{P}_{i-1}^\perp \mathbf{a}_j\|_2} \right|$ for $j \in \mathcal{I}$
2. $\mathcal{S}_i = \mathcal{S}_{i-1} \cup \{i_{s_1}, \dots, i_{s_L}\}, \mathcal{I} = \mathcal{I} \setminus \mathcal{S}_i$
3. $\mathbf{D} = \mathbf{P}_{i-1}^\perp$

for $l = 1$ to L **do**

$$\mathbf{d} = \frac{\mathbf{D} \mathbf{a}_{j_{s_l}}}{\|\mathbf{D} \mathbf{a}_{j_{s_l}}\|_2}, \mathbf{D} = \mathbf{D} - \mathbf{d} \mathbf{d}^T$$

end for

$$\mathbf{P}_i^\perp = \mathbf{D}$$

end for

$$\hat{\mathbf{x}}_k = \mathbf{A}_{\mathcal{S}_k}^\dagger \mathbf{y}$$

sian distribution with zero-mean and variance n^{-1} . We then consider two different scenarios for this choice of coefficient matrix: (1) the non-zero elements of \mathbf{x} are independent and identically distributed normal random variables, and (2) the non-zero components of \mathbf{x} are drawn uniformly from alphabet $\{+1, -1\}$. Second, the entries of \mathbf{A} are drawn independently and uniformly from $\{+1/\sqrt{n}, -1/\sqrt{n}\}$. In all settings, the locations of non-zero entries of \mathbf{x} are drawn uniformly at random. The number of equations is $n = 64$, the dimension of \mathbf{x} is $m = 128$; the experiment is repeated 1000 times. Performance of each algorithm is characterized by three metrics: (i) exact recovery rate (ERR), defined as the fraction of the correctly recovered signal components, (ii) mean-square error (MSE), measuring the distance between the unknown signal and its estimate, and (iii) the running time of the algorithm. Results for the Gaussian coefficient matrices are illustrated in Fig. 1 and Fig. 2. Fig. 1 shows the performance of the algorithms for non-zero values of \mathbf{x} being normally distributed while Fig. 2 corresponds to the second scenario. Fig. 3 shows the performance of the methods for \mathbf{A} being constructed according to the second option while the non-zero values of \mathbf{x} are normally distributed. As can be seen from Fig. 1 and Fig. 3, the generalized OLS (GOLS) outperforms all the competing methods in terms of the exact recovery rate, and is better than OLS and OMP in terms of the MSE. Moreover, the runtimes of GOLS is 2nd only to OMP but the accuracy of the latter is significantly worse than that of GOLS. In the case of $\{+1, -1\}$ non-zero entries of \mathbf{x} studied in Fig. 2, l_1 -norm/LASSO methods perform the best (and are the slowest) while the GOLS offers reasonably accurate performance at relatively high speed.

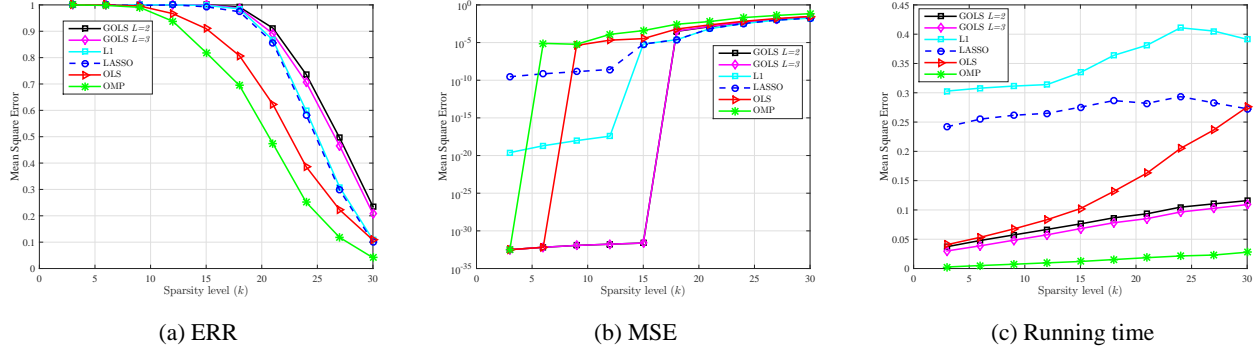


Fig. 1: Performance comparison of GOLDS, OLS, OMP, l_1 -norm minimization and LASSO for $n = 64$, $m = 128$, \mathbf{A} having Gaussian $\mathcal{N}(0, 1/n)$ entries, and the k non-zero components of \mathbf{x} drawn from $\mathcal{N}(0, 1)$ distribution.

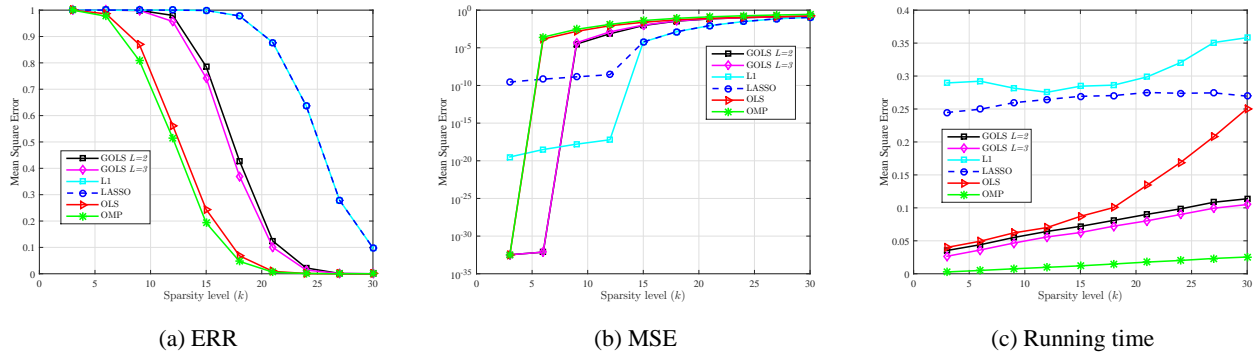


Fig. 2: Performance comparison of GOLDS, OLS, OMP, l_1 -norm minimization and LASSO for $n = 64$, $m = 128$, \mathbf{A} having Gaussian $\mathcal{N}(0, 1/n)$ entries, and the k non-zero components of \mathbf{x} randomly and equally likely set to 1 or -1 .

5. CONCLUSION

We show that for Gaussian and Bernoulli coefficient matrices, Orthogonal Least-Squares (OLS) is with high probability guaranteed to recover any sparse signal from a low number of random linear measurements. Moreover, we introduced a greedy algorithm for sparse linear regression that generalizes OLS and forms the subset of features (i.e., columns of a coefficient matrix in an underdetermined system of equations) by sequentially selecting multiple candidate columns. Since multiple indices are selected without additional cost, the running time of the algorithm is reduced compared to OLS. Thus, generalized OLS is more favorable than convex optimization based methods whose complexity grows faster with the dimension of the problem, i.e., n and m . Simulation studies demonstrate that the generalized orthogonal least-squares algorithm outperforms competing greedy methods, OLS and OMP, while being computationally more efficient than l_1 -norm minimization and LASSO.

6. REFERENCES

- [1] David L Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] Somsubhra Barik and Haris Vikalo, "Sparsity-aware sphere decoding: algorithms and complexity analysis," *Signal Processing, IEEE Transactions on*, vol. 62, no. 9, pp. 2212–2225, 2014.
- [3] Farzad Parvaresh, Haris Vikalo, Sidhant Misra, and Babak Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed dna microarrays," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 3, pp. 275–285, 2008.
- [4] Michael Lustig, David Donoho, and John M Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [5] Michael Elad, Mario AT Figueiredo, and Yi Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, 2010.
- [6] Moshe Mishali and Yonina C Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 375–391, 2010.
- [7] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, IEEE, 2009, pp. 2790–2797.

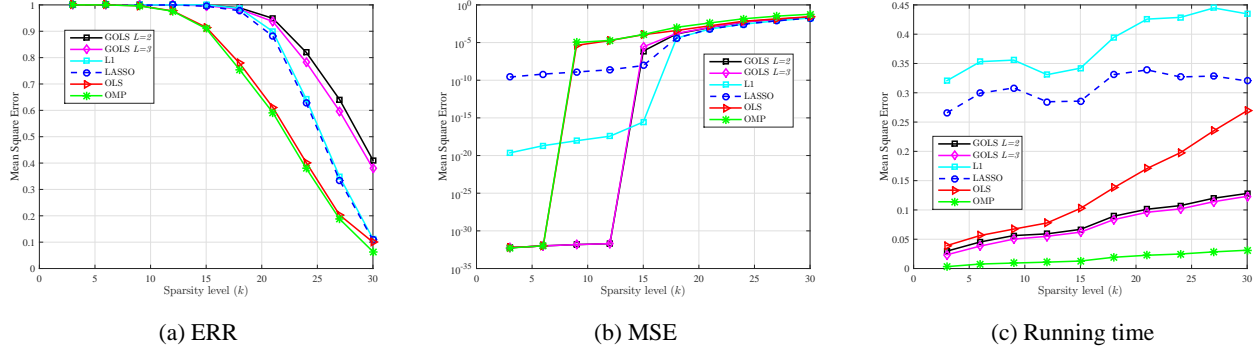


Fig. 3: Performance comparison of GOLs, OLS, OMP, l_1 -norm minimization and LASSO for $n = 64$, $m = 128$, \mathbf{A} having uniformly i.i.d entries from $\{+1/\sqrt{n}, -1/\sqrt{n}\}$, and the k non-zero components of \mathbf{x} drawn from $\mathcal{N}(0, 1)$ distribution.

- [8] Emmanuel J Candès, Justin Romberg, and Terence Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.
- [9] Amir Beck and Marc Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [10] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [11] Joel A Tropp, “Greed is good: Algorithmic results for sparse approximation,” *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [12] Deanna Needell and Roman Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 310–316, 2010.
- [13] Joel A Tropp and Anna C Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [14] Yagyensh Chandra Pati, Ramin Rezaifar, and PS Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.
- [15] David L Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck, “Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit,” *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1094–1121, 2012.
- [16] 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.
- [17] Sheng Chen, Stephen A Billings, and Wan Luo, “Orthogonal least squares methods and their application to non-linear system identification,” *International Journal of control*, vol. 50, no. 5, pp. 1873–1896, 1989.
- [18] Charles Soussen, Rémi Gribonval, Jérôme Idier, and Cédric Herzet, “Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares,” *Information Theory, IEEE Transactions on*, vol. 59, no. 5, pp. 3158–3174, 2013.
- [19] Cédric Herzet, Charles Soussen, Jérôme Idier, and Rémi Gribonval, “Exact recovery conditions for sparse representations with partial support information,” *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7509–7524, 2013.
- [20] Cédric Herzet, Angélique Drémeau, and Charles Soussen, “Relaxed recovery conditions for omp/ols by exploiting both coherence and decay,” *Information Theory, IEEE Transactions on*, vol. 62, no. 1, pp. 459–470, 2016.
- [21] Sundeep Rangan and Alyson K Fletcher, “Orthogonal matching pursuit from noisy random measurements: A new analysis,” in *Advances in Neural Information Processing Systems*, 2009, pp. 540–548.
- [22] Thomas Kailath, Ali H Sayed, and Babak Hassibi, *Linear estimation*, vol. 1, Prentice Hall Upper Saddle River, NJ, 2000.
- [23] Steven M Kay, *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*, vol. 3, Pearson Education, 2013.
- [24] Robert Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [25] Wei Dai and Olgica Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [26] Michael Grant and Stephen Boyd, “Cvx: Matlab software for disciplined convex programming,” .