

A Performance Guarantee for Orthogonal Matching Pursuit Using Mutual Coherence

Mohammad Emadi, Ehsan Miandji and Jonas Unger

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-147092>

N.B.: When citing this work, cite the original publication.

The original publication is available at www.springerlink.com:

Emadi, M., Miandji, E., Unger, J., (2018), A Performance Guarantee for Orthogonal Matching Pursuit Using Mutual Coherence, *Circuits, systems, and signal processing*, 37(4), 1562-1574. <https://doi.org/10.1007/s00034-017-0602-x>

Original publication available at:

<https://doi.org/10.1007/s00034-017-0602-x>

Copyright: Springer Verlag (Germany)

<http://www.springerlink.com/?MUD=MP>



A Performance Guarantee for Orthogonal Matching Pursuit Using Mutual Coherence

Mohammad Emadi[†] · Ehsan Miandji[†] ·
Jonas Unger

the date of receipt and acceptance should be inserted later

Abstract In this paper we present a new performance guarantee for the Orthogonal Matching Pursuit (OMP) algorithm. We use mutual coherence as a metric for determining the suitability of an arbitrary overcomplete dictionary for exact recovery. Specifically, a lower bound for the probability of correctly identifying the support of a sparse signal with additive white Gaussian noise and an upper bound for the mean square error is derived. Compared to previous work, the new bound takes into account the signal parameters such as dynamic range, noise variance, and sparsity. Numerical simulations show significant improvements over previous work and a much closer correlation to empirical results of OMP.

Keywords Compressed Sensing · Sparse Representation · Orthogonal Matching Pursuit · Sparse Recovery

1 Introduction

Estimating a sparse signal from noisy, and possibly random, measurements is now a well-studied field in signal processing Elad (2010). There has been a large body of research dedicated to proposing efficient algorithms to tackle this problem, which we briefly overview in this section. A few applications where sparse signal recovery can be used include: estimating the direction of arrival in radar arrays Emadi and

M. Emadi
Department of Electrical and Computer Engineering,
Cornell University, Ithaca, NY, USA
Present affiliation: Qualcomm Technologies Inc., San Jose, CA, USA
E-mail: memadi@qti.qualcomm.com

E. Miandji and J. Unger
Department of Science and Technology, Linköping University, Sweden
E-mail: ehsan.miandji@liu.se

J. Unger
Department of Science and Technology, Linköping University, Sweden
E-mail: jonas.unger@liu.se

[†] Equal contributor

Sadeghi (2013); Liu et al (2011); Malioutov et al (2005), imaging Duarte et al (2008), source separation Bofill and Zibulevsky (2001); Castrodad et al (2011), and inverse problems in image processing Elad and Aharon (2006); Miandji et al (2015); Yang et al (2010). Indeed the conditions under which a sparse recovery algorithm achieves exact reconstruction is of critical importance. Having such knowledge allows for designing efficient systems that take into account the effect of different signal parameters in the recovery process. In this paper we will propose a new performance guarantee for a popular sparse recovery method, namely Orthogonal Matching Pursuit (OMP). We start by formulating the problem.

Let $\mathbf{s} \in \mathbb{R}^N$ be an unknown variable that we would like to estimate from the measurements

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{w}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is a deterministic matrix and $\mathbf{w} \in \mathbb{R}^M$ is a noise vector, often assumed to be white Gaussian noise with mean zero and covariance $\sigma^2 \mathbf{I}$, where \mathbf{I} is the identity matrix. The matrix \mathbf{A} is called a *dictionary*. We consider the case when \mathbf{A} is *overcomplete*, i.e. $N > M$; hence uniqueness of the solution of (1) cannot be guaranteed. However, if most elements of \mathbf{s} are zero, we can limit the space of possible solutions, or even obtain a unique one, by solving

$$\hat{\mathbf{s}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon, \quad (2)$$

where ϵ is a constant related to \mathbf{w} . The location of nonzero entries in \mathbf{s} is known as the *support* set, which we denote by A . In some applications, e.g. estimating the direction of arrival in antenna arrays Malioutov et al (2005), correctly identifying the support is more important than the accuracy of values in $\hat{\mathbf{s}}$. When the correct support is known, the solution of the least squares problem $\|\mathbf{y} - \mathbf{A}_A \mathbf{x}_A\|_2^2$ gives $\hat{\mathbf{s}}$, where \mathbf{A}_A is formed using the columns of \mathbf{A} indexed by A .

Solving (2) is an NP-hard problem and several greedy algorithms have been proposed to compute an approximate solution of (2), see e.g. Mallat and Zhang (1993); Needell and Tropp (2009); Needell and Vershynin (2010); Pati et al (1993). In contrast to greedy methods, convex relaxation algorithms replace the ℓ_0 pseudo-norm in (2) with an ℓ_1 norm, leading to a convex optimization problem known as the Basis Pursuit (BP) Chen et al (1998):

$$\hat{\mathbf{s}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon, \quad (3)$$

While convex relaxation methods require weaker conditions for exact recovery Donoho et al (2006); Tropp (2006), they are computationally more expensive than greedy methods, specially when $N \gg M$ Hsieh et al (2012); Marvasti et al (2012). Many algorithms have been proposed to solve (3), see e.g. van den Berg and Friedlander (2009); Efron et al (2004); Figueiredo et al (2007); Wright et al (2009). Thanks to convexity of (3), a unique solution can be guaranteed; however, it is not guaranteed that the solution of (3) is equivalent to the true solution of (2). Such equivalence conditions have been established in e.g. Donoho (2006); Donoho and Huo (2001); Donoho et al (2006); Tropp (2006).

A common variant of (3) is known as the Lasso problem Tibshirani (1994):

$$\hat{\mathbf{s}} = \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \lambda, \quad (4)$$

where λ is a parameter related to SNR. Under certain conditions imposed on the dictionary, solutions of (3) and (4) are unique and the sparsest Tropp (2006). However, as it will be described in the next section, algorithms for solving (3) and (4) suffer from high computational burden.

The most important aspect of a sparse recovery algorithm is the uniqueness of the obtained solution. Many metrics have been proposed to evaluate the suitability of a dictionary for exact recovery. A few examples include Mutual Coherence (MC) Donoho and Huo (2001), cumulative coherence Tropp (2006), the spark Donoho and Elad (2003), exact recovery coefficient Tropp (2004), and restricted isometry constant Candès et al (2006). Among these metrics, MC is the most efficient to compute and has shown to provide acceptable performance guarantees Ben-Haim et al (2010). The mutual coherence of a dictionary \mathbf{A} , denoted $\mu_{\max}(\mathbf{A})$, is the maximum absolute cross correlation of its columns Donoho and Huo (2001):

$$\mu_{i,j}(\mathbf{A}) = \langle \mathbf{A}_i, \mathbf{A}_j \rangle, \quad (5)$$

$$\mu_{\max}(\mathbf{A}) = \max_{1 \leq i \neq j \leq N} |\mu_{i,j}(\mathbf{A})|, \quad (6)$$

where we have assumed, as with the rest of the paper, that $\|\mathbf{A}_i\|_2 = 1$, $i \in \{1, \dots, N\}$.

As mentioned earlier, greedy algorithm are significantly faster than convex relaxation methods. Among greedy methods for solving (2), OMP provides a better trade-off between the computation complexity and the accuracy of the solution Elad (2010); Eldar and Kutyniok (2012); Marvasti et al (2012); Pope (2009). This method computes a matrix multiplication with the complexity of $\mathcal{O}(NL)$ in each iteration, while the computational complexity of ℓ_1 algorithms are in the order of $\mathcal{O}(N^2L)$ (using linear programming) or $\mathcal{O}(N^3)$ (using an interior-point method). On the other hand, OMP is a heuristic algorithm with theoretical guarantees that are not as accurate as those for ℓ_1 methods. The most recent coherence-based results regarding the convergence of OMP is reported in Ben-Haim et al (2010), where the authors also compare to commonly used ℓ_1 algorithms.

In this paper, we will improve the results of Ben-Haim et al (2010) and derive a new performance guarantee for the OMP algorithm based on MC. Specifically, a lower bound for the probability of correctly identifying the support of a sparse signal and an upper bound for the resulting Mean Square Error (MSE) is derived. The new probability bound, unlike previous work, takes into account signal parameters such as dynamic range, sparsity, and the noise characteristics. To achieve this, we treat elements of the sparse signal as centered independent and identically distributed random variables with an arbitrary distribution. Our main contribution, namely Theorem 1, is presented in section 2. We will analytically and numerically compare our results with Ben-Haim et al (2010). Our numerical results show significant improvements with respect to the probability of successful support recovery of a sparse signal. Most importantly, our numerical results match the empirically obtained results of OMP more closely. Section 3 will present the numerical results in more detail, followed by our conclusion in section 4.

2 OMP CONVERGENCE ANALYSIS

In this section we present and prove the main result of the paper, namely Theorem 1. Numerical results will be presented in section 3.

Theorem 1 Let $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{w}$, where $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\tau = \|\mathbf{s}\|_0$ and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Then OMP identifies the true support with lower bound probability of

$$\left(1 - 2N \exp\left(\frac{-s_{\min}^2}{\frac{8\tau}{N}\left(\tau\gamma^2 + \frac{\sigma^2}{N}\right) + \frac{4s_{\min}}{3}\left(\gamma + \frac{\beta}{N}\right)}\right)\right) \left(1 - N\sqrt{\frac{2}{\pi}}\frac{\sigma}{\beta}e^{-\beta^2/2\sigma^2}\right), \quad (7)$$

where $s_{\min} = \min(|\mathbf{s}_i|)$, $\gamma = \mu_{\max}s_{\max}$, and β is a positive constant satisfying $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$. Moreover, if the support is correctly identified, then the MSE of the estimated coefficients, $\hat{\mathbf{s}}$, is bounded from above by

$$\frac{\tau\beta^2}{(1 - \mu_{\max}(\tau - 1))^2}. \quad (8)$$

Let us compare Theorem 1 with the theoretical results reported in Ben-Haim et al (2010). We observe that, if we set $\beta \triangleq \sigma\sqrt{2(1 + \alpha)}\log N$, the second term in (7) becomes equivalent to the probability of success reported for OMP in Ben-Haim et al (2010). However, the analysis of Ben-Haim et al. Ben-Haim et al (2010) imposes a condition that is dependent on signal parameters such as s_{\min} and τ . Therefore, the probability of success is dependent on the satisfaction of the aforementioned condition. In contrast, Theorem 1 does not impose any conditions on signal parameters. In fact, the effect of various signal parameters is modeled probabilistically by the first term of (7). Moreover, in our analysis, the first term of (7) is a function of s_{\max} , among other parameters. This parameter, together with s_{\min} , define the dynamic range of the signal. In other words, our analysis takes the dynamic range of the signal into account. This is in contrast with Ben-Haim et al (2010), where only s_{\min} is taken into account by the condition imposed on parameters.

The following lemmas will provide us with the necessary tools for the proof of Theorem 1. The proof of the lemmas are postponed to the Appendix.

Lemma 1 Define $\Gamma_j = |\langle \mathbf{A}_j, \mathbf{A}\mathbf{s} + \mathbf{w} \rangle|$, where $j \in \{1, \dots, N\}$, then for some constant $\xi \geq 0$ we have

$$\Pr\{\Gamma_j \geq \xi\} \leq 2 \exp\left(\frac{-\xi^2}{2(N\nu + c\xi/3)}\right), \quad (10)$$

where

$$\mathbf{x}_n = \mu_{j,n}\mathbf{s}_n + \frac{1}{N}\langle \mathbf{A}_j, \mathbf{w} \rangle \quad (11)$$

$$|\mathbf{x}_n| \leq c, \quad n = 1, \dots, N \quad (12)$$

$$\mathbb{E}\{\mathbf{x}_n^2\} \leq \nu, \quad n = 1, \dots, N \quad (13)$$

The following lemma explicitly formulates the upper bounds c and ν introduced in (12) and (13).

Lemma 2 Let $\mathbf{x}_n = \mu_{j,n}\mathbf{s}_n + N^{-1}\langle \mathbf{A}_j, \mathbf{w} \rangle$, for any $n \in \{1, \dots, N\}$ and a fixed index $j \in \{1, \dots, N\}$. Assume that $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, and $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$, $\forall j \in \{1, \dots, N\}$. Then,

$$|\mathbf{x}_n| \leq \mu_{\max}s_{\max} + \frac{\beta}{N}, \quad (14)$$

$$\mathbb{E}\{\mathbf{x}_n^2\} \leq \frac{1}{N}\tau s_{\max}^2 \mu_{\max}^2 + \frac{\sigma^2}{N^2}. \quad (15)$$

We can now state the proof of Theorem 1.

Proof (Proof of Theorem 1) It is shown in Ben-Haim et al (2010) that when $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$, $\forall j \in \{1, \dots, N\}$, then OMP identifies Λ if

$$\min_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_\Lambda \mathbf{s}_\Lambda + \mathbf{w} \rangle| \geq \max_{k \notin \Lambda} |\langle \mathbf{A}_k, \mathbf{A}_\Lambda \mathbf{s}_\Lambda + \mathbf{w} \rangle|. \quad (16)$$

Using the triangle inequality on the left-hand side of (16) we have

$$\begin{aligned} & \min_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_\Lambda \mathbf{s}_\Lambda + \mathbf{w} \rangle| \\ &= \min_{j \in \Lambda} |\mathbf{s}_j + \langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| \\ &\geq \min_{j \in \Lambda} |\mathbf{s}_j| - \max_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle|. \end{aligned} \quad (17)$$

From (16) and (17), we can see that the OMP algorithm identifies the true support if

$$\begin{cases} \max_{k \notin \Lambda} \{\Gamma_k\} < \min_{j \in \Lambda} \frac{|\mathbf{s}_j|}{2}, \\ \max_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| < \min_{j \in \Lambda} \frac{|\mathbf{s}_j|}{2}. \end{cases} \quad (18)$$

Using (18) we can define the probability of error for OMP as

$$\begin{aligned} \Pr\{\text{error}\} &= \\ & \Pr \left\{ \max_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| \geq \frac{s_{\min}}{2} \right\} \\ &+ \Pr \left\{ \max_{k \notin \Lambda} \{\Gamma_k\} \geq \frac{s_{\min}}{2} \right\}, \end{aligned} \quad (19)$$

with the upper bound

$$\begin{aligned} \Pr\{\text{error}\} &\leq \\ & \sum_{j \in \Lambda} \Pr \left\{ |\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| \geq \frac{s_{\min}}{2} \right\} \\ &+ \sum_{k \notin \Lambda} \Pr \left\{ \Gamma_k \geq \frac{s_{\min}}{2} \right\}. \end{aligned} \quad (20)$$

For the first term on the right-hand side of (20), excluding the summation over indices in Λ , from Lemma 1 we have

$$\begin{aligned} & \Pr_{j \in \Lambda} \left\{ |\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| \geq \frac{s_{\min}}{2} \right\} \\ &\leq 2 \exp \left(\frac{-s_{\min}^2}{8((\tau - 1)\nu + s_{\min}c/6)} \right). \end{aligned} \quad (21)$$

Note that unlike Lemma 1, the dictionary \mathbf{A} in (21) is supported on $\Lambda \setminus \{j\}$, i.e. all the indices in the true support excluding j . Therefore the term $(\tau - 1)$, instead of

N , appears in the denominator of (21). From Lemma 2, the upper bounds c and ν , defined in (12) and (13) respectively, are

$$|\mathbf{x}_n| \leq \mu_{\max} s_{\max} + \frac{\beta}{N}, \quad (22)$$

$$E \left\{ \mathbf{x}_n^2 \right\} \leq \frac{\tau-1}{N} \mu_{\max}^2 s_{\max}^2 + \frac{\sigma^2}{N^2}. \quad (23)$$

Combining (22) and (23) with (21) yields

$$\begin{aligned} & \Pr_{j \in \Lambda} \left\{ \left| \langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle \right| \geq \frac{s_{\min}}{2} \right\} \\ & \leq 2 \exp \left(\underbrace{\frac{-s_{\min}^2}{\frac{8(\tau-1)}{N} \left((\tau-1)\gamma^2 + \frac{\sigma^2}{N} \right) + \frac{4s_{\min}}{3} \left(\gamma + \frac{\beta}{N} \right)}}_{P_1} \right), \end{aligned} \quad (24)$$

where we have defined $\gamma = \mu_{\max} s_{\max}$ for notational brevity.

Applying the same procedure on the second term on the right-hand side of (20), excluding the summation, yields

$$\begin{aligned} & \Pr_{k \notin \Lambda} \left\{ \Gamma_k \geq \frac{s_{\min}}{2} \right\} \\ & \leq 2 \exp \left(\underbrace{\frac{-s_{\min}^2}{\frac{8\tau}{N} \left(\tau\gamma^2 + \frac{\sigma^2}{N} \right) + \frac{4s_{\min}}{3} \left(\gamma + \frac{\beta}{N} \right)}}_{P_2} \right), \end{aligned} \quad (25)$$

Substituting (24) and (25) into (20) we obtain

$$\Pr\{\text{error}\} \leq \tau P_1 + (N - \tau) P_2 \leq N P_2, \quad (26)$$

where the last inequality follows since $P_2 > P_1$. So far we have assumed that $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta, \forall j$. The probability of success is the joint probability of $\Pr \{ |\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta \}$ and the inverse of (26). The former can be bounded as follows

$$\Pr \{ |\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta \} = 1 - 2Q \left(\frac{\beta}{\sigma} \right) \geq 1 - \underbrace{\sqrt{\frac{2}{\pi}} \frac{\sigma}{\beta} e^{-\beta^2/2\sigma^2}}_{P_3}, \quad (27)$$

where $Q(x)$ is the Gaussian tail probability. Since $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$ should hold for all $j \in \{1, \dots, N\}$ we have

$$\Pr_{j=1, \dots, N} \{ |\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta \} \geq (1 - P_3)^N \geq 1 - N P_3, \quad (28)$$

Inverting the probability event in (26) and combining with (28) yields (7).

To prove (8), we proceed similar to Ben-Haim et al (2010). Using the triangle inequality we have:

$$\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2 \leq \|\hat{\mathbf{s}}_{orc} - \mathbf{s}\|_2^2 + \|\hat{\mathbf{s}}_{orc} - \hat{\mathbf{s}}\|_2^2, \quad (29)$$

where $\hat{\mathbf{s}}_{orc}$ is the oracle estimator, i.e. an estimator that knows the true support of \mathbf{s} , *a priori*. If the OMP algorithm identifies the true support, then the second term on right-hand side of (29) will be zero. For the first term we have

$$\begin{aligned} \|\hat{\mathbf{s}}_{orc} - \mathbf{s}\|_2^2 &= \left\| (\mathbf{A}_{\Lambda_0}^T \mathbf{A}_{\Lambda_0})^{-1} \mathbf{A}_{\Lambda_0}^T \mathbf{w} \right\|_2^2 \\ &\leq \left\| (\mathbf{A}_{\Lambda_0}^T \mathbf{A}_{\Lambda_0})^{-1} \right\|_2^2 \sum_{j \in \Lambda_0} |\langle \mathbf{A}_j, \mathbf{w} \rangle|^2 \\ &\leq \left\| (\mathbf{A}_{\Lambda_0}^T \mathbf{A}_{\Lambda_0})^{-1} \right\|_2^2 \tau \beta^2, \end{aligned} \quad (30)$$

The term $\left\| (\mathbf{A}_{\Lambda_0}^T \mathbf{A}_{\Lambda_0})^{-1} \right\|_2$ will be smaller than the maximum eigenvalue of $(\mathbf{A}_{\Lambda_0}^T \mathbf{A}_{\Lambda_0})^{-1}$, which is equal to inverse of the minimum eigenvalue of $\mathbf{A}_{\Lambda_0}^T \mathbf{A}_{\Lambda_0}$. According to the Gershgorin circle theorem Golub and Van Loan (1996), this number will be larger than $(1 - \mu_{\max}(\tau - 1))$. Substituting this into (30) completes the proof.

3 Numerical Results

In this section we compare numerical results of Theorem 1 and Ben-Haim et al. Ben-Haim et al (2010) with the empirically obtained results of the OMP algorithm. Throughout this section we will refer to the analysis of Ben-Haim et al Ben-Haim et al (2010) as CBPG. Moreover, we will only compare the probability of success, i.e. equation (7), since extensive results on MSE have been reported in Ben-Haim et al (2010).

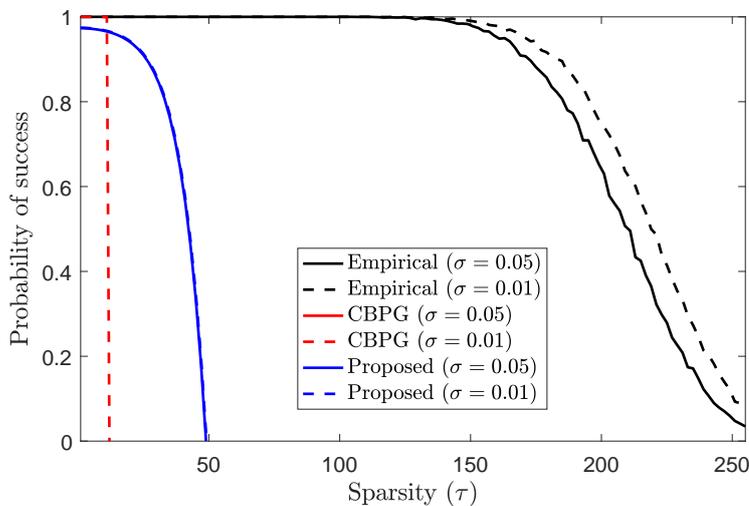


Fig. 1 Probability of successful recovery of the support compared to the sparsity of the signal. Parameters used are $N = 2048$, $M = 1024$, $s_{\min} = 0.5$, $s_{\max} = 1$, and $\mu_{\max} = 0.0313$. The CBPG method refers to Ben-Haim et al. Ben-Haim et al (2010).

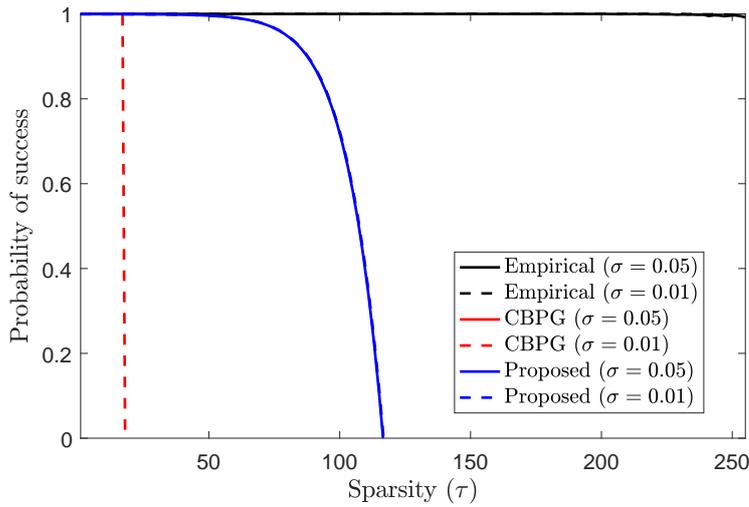


Fig. 2 Same as Fig. 1 but with $N = 4096$ and $M = 2048$. Here we have $\mu_{\max} = 0.0221$.

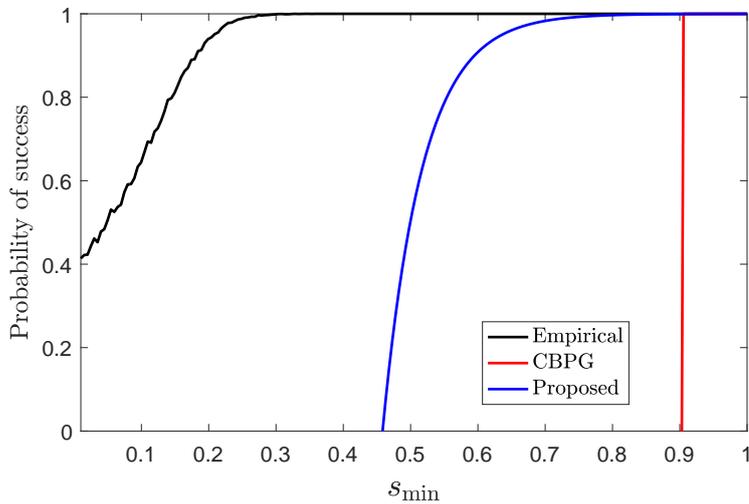


Fig. 3 Parameters are $N = 1024$, $M = 512$, $\tau = 5$, $\sigma = 0.05$, $s_{\max} = 1$.

All the empirical results are obtained by performing the OMP algorithm 5000 times using a random signal with additive white Gaussian noise in each trial. The probability of success is the ratio of successful trials to the number of trials. We use the same dictionary as CBPG, defined $\mathbf{A} = [\mathbf{I}, \mathbf{H}]$, where \mathbf{H} is the Hadamard matrix. To construct the sparse noisy signal in each trials we proceed as follows: The support of the signal is chosen uniformly at random from the set $\{1, 2, \dots, N\}$, i.e. $A \in \{1, 2, \dots, N\}$. The nonzero elements located at A are drawn randomly from a uniform distribution on the interval $[s_{\min}, s_{\max}]$, multiplied randomly by $+1$ or -1 . Once the sparse signal is constructed, the input to the OMP algorithm is obtained by evaluating (1).

In order to facilitate the comparison of our results with CBPG, we need to fix the value of β . To do this, we empirically calculate β as $\max_{\mathbf{w}} \max_j |\langle \mathbf{A}_j, \mathbf{w} \rangle|$, where the maximum over \mathbf{w} is computed using 10^4 samples from $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Note that CBPG uses another constant, termed α , which is related to β by the definition $\beta \triangleq \sigma \sqrt{2(1 + \alpha) \log N}$, see Ben-Haim et al (2010) for more details. Hence we can calculate α from the empirically obtained value of β .

In Figures 1 and 2 we consider the effect of sparsity on the probability of success, where each figure considers a different signal dimensionality. The parameters used are similar to CBPG; specifically, we set $s_{\min} = 0.5$, $s_{\max} = 1$, and consider two values for noise variance: $\sigma^2 = 0.0025$ and $\sigma^2 = 0.0001$. While CBPG fails for larger noise variance, Theorem 1 produces valid results for both noise variances, with probability curves that are very close to each other; this shows that Theorem 1 has less sensitivity to noise, a result that is closer to empirical evaluation of OMP. Moreover, Theorem 1 reports high probabilities for a much larger range of values for τ . Most importantly, we see that the shape of the probability curves are similar to that of empirical results, while CBPG behaves like a step function. This is due to the fact that the condition imposed by CBPG is not satisfied for a large range of τ values.

In Fig. 2, we double the size of the signal. In this case the value of mutual coherence and β decreases (we use the same noise variances). Hence we expect that the probability of success increases in all cases. While the probability of success improves a lot for the empirical curve, the results of CBPG does not change considerably. This is due to fact that the condition imposed by CBPG is very sensitive to signal parameters; specifically, it is not satisfied for large values of τ . On the other hand, our analysis shows less sensitivity to τ when the signal dimensionality is increased. This behavior is expected since the empirical results also demonstrate such patterns in the probability of success.

Finally, Fig. 3 presents the effect of signal dynamic range on the probability of support recovery. For this test we set $N = 1024$, $M = 512$, $\tau = 5$, $\sigma^2 = 0.0025$, and $s_{\max} = 1$, while varying $s_{\min} \in [0.01, 1]$. Here we also see that Theorem 1 achieves results that match empirical results more closely compared to what is obtained using CBPG. This implies that Theorem 1 leads to valid results for signals with much higher dynamic range than CBPG. The MATLAB source code is provided for further analysis of the results, see section ‘‘Additional Files’’.

4 Conclusion

We presented a new bound for the probability of correctly identifying the support of a noisy sparse signal using the OMP algorithm. This result was accompanied by an upper bound for MSE. Compared to previous work, specifically Ben-Haim et al. Ben-Haim et al (2010), our analysis replaces a sharp condition with a probabilistic bound. Comparisons to empirical results obtained by OMP show a much improved correlation than previous work. Indeed the probability bound can be improved as the distance to empirical results is still significant.

Appendix

Proof (proof of Lemma 1) Expanding Γ_j , we can show that

$$\begin{aligned}\Gamma_j &= |\langle \mathbf{A}_j, \mathbf{A}\mathbf{s} + \mathbf{w} \rangle| \\ &= \left| \sum_{m=1}^M \mathbf{A}_{m,j} \left(\sum_{n=1}^N \mathbf{A}_{m,n} \mathbf{s}_n + \mathbf{w}_m \right) \right| \\ &= \left| \sum_{n=1}^N \left\{ \sum_{m=1}^M \mathbf{A}_{m,j} \mathbf{A}_{m,n} \mathbf{s}_n + \frac{1}{N} \sum_{m=1}^M \mathbf{A}_{m,j} \mathbf{w}_m \right\} \right|.\end{aligned}\quad (31)$$

Using (5) we have that

$$\Gamma_j = \left| \sum_{n=1}^N \left\{ \mu_{j,n} \mathbf{s}_n + \frac{1}{N} \langle \mathbf{A}_j, \mathbf{w} \rangle \right\} \right| = \left| \sum_{n=1}^N \mathbf{x}_n \right|. \quad (32)$$

As mentioned in section 1, we assume that the elements of the sparse vector \mathbf{s} are centered random variables. Hence, the elements of \mathbf{s} are either zero or zero-mean random variables, implying that $\mathbb{E}\{\mathbf{s}_n\} = 0$ for all $n = 1, \dots, N$. Together with the fact that $\mathbb{E}\{\mathbf{w}\} = 0$, we have

$$\mathbb{E}\{\mathbf{x}_n\} = \mu_{j,n} \mathbb{E}\{\mathbf{s}_n\} + N^{-1} \mathbb{E}\{\langle \mathbf{A}_j, \mathbf{w} \rangle\} = 0, \quad (33)$$

for all $n = 1, \dots, N$. According to Bernstein's inequality Bennett (1962), if $\mathbf{x}_1, \dots, \mathbf{x}_N$ are independent real random variables with mean zero, where $\mathbb{E}\{\mathbf{x}_n^2\} \leq \nu$, and $\Pr\{|\mathbf{x}_n| < c\} = 1$, then

$$\begin{aligned}\Pr \left\{ \left| \sum_{n=1}^N \mathbf{x}_n \right| \geq \xi \right\} &\leq 2 \exp \left(\frac{-\xi^2}{2 \left(\sum_{n=1}^N \mathbb{E}\{\mathbf{x}_n^2\} + c\xi/3 \right)} \right) \\ &\leq 2 \exp \left(\frac{-\xi^2}{2(N\nu + c\xi/3)} \right),\end{aligned}\quad (34)$$

where (34) follows using (13). This completes the proof.

Proof (proof of Lemma 2) Equation (14) follows trivially from the triangle inequality. For (15) we have

$$\mathbb{E}\{\mathbf{x}_n^2\} = \frac{1}{N} \sum_{n=1}^N \mathbb{E}\{\mathbf{x}_n^2\} \quad (35)$$

$$= \frac{1}{N} \sum_{n=1}^N \mathbb{E} \left\{ \mu_{j,n}^2 \mathbf{s}_n^2 + \frac{1}{N^2} \langle \mathbf{A}_j, \mathbf{w} \rangle^2 + \frac{2}{N} \mu_{j,n} \mathbf{s}_n \langle \mathbf{A}_j, \mathbf{w} \rangle \right\} \quad (36)$$

$$= \frac{1}{N} \sum_{n=1}^N \mu_{j,n}^2 \mathbb{E}\{\mathbf{s}_n^2\} + \frac{1}{N^2} \mathbb{E}\{\langle \mathbf{A}_j, \mathbf{w} \rangle^2\} + \underbrace{\frac{2}{N} \mathbb{E}\{\mu_{j,n} \mathbf{s}_n \langle \mathbf{A}_j, \mathbf{w} \rangle\}}_0 \quad (37)$$

$$\leq \frac{\tau}{N} \mu_{\max}^2 \mathbf{s}_{\max}^2 + \frac{1}{N^2} \mathbb{E}\{\langle \mathbf{A}_j, \mathbf{w} \rangle^2\}, \quad (38)$$

where the last term in (37) is zero since $E\{\mathbf{w}\} = 0$, which implies $E\{\langle \mathbf{A}_j, \mathbf{w} \rangle\} = 0$. Moreover, we have

$$\begin{aligned} E\left\{\langle \mathbf{A}_j, \mathbf{w} \rangle^2\right\} &= E\left\{\left(\sum_{m=1}^M \mathbf{A}_{m,j} \mathbf{w}_m\right)\left(\sum_{m=1}^M \mathbf{A}_{m,j} \mathbf{w}_m\right)\right\} \\ &= \sum_{m=1}^M \mathbf{A}_{m,j} \mathbf{A}_{m,j} E\{\mathbf{w}_m \mathbf{w}_m\} = \sigma^2. \end{aligned} \quad (39)$$

combining (38) and (39) completes the proof.

References

- Ben-Haim Z, Eldar Y, Elad M (2010) Coherence-based performance guarantees for estimating a sparse vector under random noise. *Signal Processing, IEEE Transactions on* 58(10):5030–5043, DOI 10.1109/TSP.2010.2052460
- Bennett G (1962) Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association* 57(297):33–45
- van den Berg E, Friedlander MP (2009) Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing* 31(2):890–912, DOI 10.1137/080714488
- Bofill P, Zibulevsky M (2001) Underdetermined blind source separation using sparse representations. *Signal Processing* 81(11):2353–2362, DOI [http://doi.org/10.1016/S0165-1684\(01\)00120-7](http://doi.org/10.1016/S0165-1684(01)00120-7)
- Candès EJ, Romberg JK, Tao T (2006) Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics* 59(8):1207–1223, DOI 10.1002/cpa.20124
- Castrodad A, Xing Z, Greer JB, Bosch E, Carin L, Sapiro G (2011) Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 49(11):4263–4281, DOI 10.1109/TGRS.2011.2163822
- Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* 20:33–61
- Donoho D (2006) Compressed sensing. *Information Theory, IEEE Transactions on* 52(4):1289–1306, DOI 10.1109/TIT.2006.871582
- Donoho D, Huo X (2001) Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on* 47(7):2845–2862, DOI 10.1109/18.959265
- Donoho D, Elad M, Temlyakov V (2006) Stable recovery of sparse overcomplete representations in the presence of noise. *Information Theory, IEEE Transactions on* 52(1):6–18, DOI 10.1109/TIT.2005.860430
- Donoho DL, Elad M (2003) Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. *Proceedings of the National Academy of Sciences* 100(5):2197–2202, DOI 10.1073/pnas.0437847100, <http://www.pnas.org/content/100/5/2197.full.pdf>
- Duarte MF, Davenport MA, Takbar D, Laska JN, Sun T, Kelly KF, Baraniuk RG (2008) Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine* 25(2):83–91, DOI 10.1109/MSP.2007.914730

- Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least angle regression. *Annals of Statistics* 32(2):407–499
- Elad M (2010) *Sparse and Redundant Representations*. Springer New York, DOI 10.1007/978-1-4419-7011-4
- Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing* 15(12):3736–3745, DOI 10.1109/TIP.2006.881969
- Eldar YC, Kutyniok G (eds) (2012) *Compressed Sensing Theory and Applications*. Cambridge University Press, Cambridge Books Online
- Emadi M, Sadeghi K (2013) DOA estimation of multi-reflected known signals in compact arrays. *Aerospace and Electronic Systems, IEEE Transactions on* 49(3):1920–1931, DOI 10.1109/TAES.2013.6558028
- Figueiredo MAT, Nowak RD, Wright SJ (2007) Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing* 1(4):586–597, DOI 10.1109/JSTSP.2007.910281
- Golub G, Van Loan C (1996) *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press
- Hsieh SH, Lu CS, Pei SC (2012) Fast omp: Reformulating omp via iteratively refining l2-norm solutions. In: *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pp 189–192, DOI 10.1109/SSP.2012.6319656
- Liu ZM, Huang ZT, Zhou YY (2011) Direction-of-arrival estimation of wideband signals via covariance matrix sparse representation. *Signal Processing, IEEE Transactions on* 59(9):4256–4270, DOI 10.1109/TSP.2011.2159214
- Malioutov D, Cetin M, Willsky A (2005) A sparse signal reconstruction perspective for source localization with sensor arrays. *Signal Processing, IEEE Transactions on* 53(8):3010–3022, DOI 10.1109/TSP.2005.850882
- Mallat S, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on* 41(12):3397–3415, DOI 10.1109/78.258082
- Marvasti F, Amini A, Haddadi F, Soltanolkotabi M, Khalaj BH, Aldroubi A, Sanei S, Chambers J (2012) A unified approach to sparse signal processing. *EURASIP Journal on Advances in Signal Processing* 2012:44
- Miandji E, Kronander J, Unger J (2015) Compressive image reconstruction in reduced union of subspaces. *Computer Graphics Forum* 34(2):33–44, DOI 10.1111/cgf.12539
- Needell D, Tropp J (2009) CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis* 26(3):301–321, DOI <http://dx.doi.org/10.1016/j.acha.2008.07.002>
- Needell D, Vershynin R (2010) Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *Selected Topics in Signal Processing, IEEE Journal of* 4(2):310–316, DOI 10.1109/JSTSP.2010.2042412
- Pati Y, Rezaifar R, Krishnaprasad PS (1993) Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, pp 40–44 vol.1, DOI 10.1109/ACSSC.1993.342465
- Pope G (2009) *Compressive sensing: A summary of reconstruction algorithms*. Master’s thesis, ETH Zürich

- Tibshirani R (1994) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58:267–288
- Tropp J (2004) Greed is good: algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on* 50(10):2231–2242, DOI 10.1109/TIT.2004.834793
- Tropp J (2006) Just relax: convex programming methods for identifying sparse signals in noise. *Information Theory, IEEE Transactions on* 52(3):1030–1051, DOI 10.1109/TIT.2005.864420
- Wright S, Nowak R, Figueiredo M (2009) Sparse reconstruction by separable approximation. *Signal Processing, IEEE Transactions on* 57(7):2479–2493, DOI 10.1109/TSP.2009.2016892
- Yang J, Wright J, Huang TS, Ma Y (2010) Image super-resolution via sparse representation. *IEEE Transactions on Image Processing* 19(11):2861–2873, DOI 10.1109/TIP.2010.2050625

Additional Files

The source code for generating Figures 1, 2, and 3 can be downloaded from <https://liu.box.com/s/fgq6calflqys487kxgyifbpb5x36vz0>.