A Performance Guarantee for Orthogonal Matching Pursuit Using Mutual Coherence

Mohammad Emadi · Ehsan Miandji · Jonas Unger

Received: 29 December 2016 / Revised: 25 June 2017 / Accepted: 27 June 2017 / Published online: 27 July 2017
© Springer Science+Business Media, LLC 2017

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 the 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

Mohammad Emadi and Ehsan Miandji have contributed equally to this work.

Mohammad Emadi
memadi@qti.qualcomm.com

Ehsan Miandji
ehsan.miandji@liu.se
Jonas Unger
jonas.unger@liu.se

1 Qualcomm Technologies Inc., San Jose, CA, USA
2 Department of Science and Technology, Linköping University, Linköping, Sweden
1 Introduction

Estimating a sparse signal from noisy, and possibly random, measurements is now a well-studied field in signal processing [13]. 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 [16,20,21], imaging [11], source separation [3,5], and inverse problems in image processing [14,24,34]. 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 \( s \in \mathbb{R}^N \) be an unknown variable that we would like to estimate from the measurements

\[
y = As + w,
\]

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

\[
\hat{s} = \min_x \|x\|_0 \quad \text{s.t.} \quad \|y - Ax\|_2^2 \leq \epsilon,
\]

where \( \epsilon \) is a constant related to \( w \). The location of nonzero entries in \( s \) is known as the support set, which we denote by \( \Lambda \). In some applications, e.g. estimating the direction of arrival in antenna arrays [21], correctly identifying the support is more important than the accuracy of values in \( \hat{s} \). When the correct support is known, the solution of the least squares problem \( \|y - A_{\Lambda}x_{\Lambda}\|_2^2 \) gives \( \hat{s} \), where \( A_{\Lambda} \) is formed using the columns of \( A \) indexed by \( \Lambda \).

Solving (2) is an NP-hard problem, and several greedy algorithms have been proposed to compute an approximate solution of (2), see e.g. [22,25–27]. In contrast to greedy methods, convex relaxation algorithms replace the \( \ell_0 \) pseudo-norm in (2) with an \( \ell_1 \) norm, leading to a convex optimization problem known as the basis pursuit (BP) [6]:

\[
\hat{s} = \min_x \|x\|_1 \quad \text{s.t.} \quad \|y - Ax\|_2^2 \leq \epsilon,
\]

While convex relaxation methods require weaker conditions for exact recovery [9, 31], they are computationally more expensive than greedy methods, specially when \( N \gg M \) [19,23]. Many algorithms have been proposed to solve (3), see e.g. [12,17, ...
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. [7–9,31].

A common variant of (3) is known as the Lasso problem [29]:

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

where \(\lambda\) is a parameter related to SNR. Under certain conditions imposed on the dictionary, solutions of (3) and (4) are unique and the sparsest [31]. 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) [8], cumulative coherence [31], the spark [10], exact recovery coefficient [30], and restricted isometry constant [4]. Among these metrics, MC is the most efficient to compute and has shown to provide acceptable performance guarantees [1]. The mutual coherence of a dictionary \(A\), denoted \(\mu_{\text{max}}(A)\), is the maximum absolute cross-correlation of its columns [8]:

\[
\mu_{i,j}(A) = \langle A_i, A_j \rangle, \quad (5)
\]

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

where we have assumed, as with the rest of the paper, that \(\|A_i\|_2 = 1, i \in \{1, \ldots, N\}\).

As mentioned earlier, greedy algorithms 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 [13,15,23,28]. This method computes a matrix multiplication with the complexity of \(O(NL)\) in each iteration, while the computational complexity of \(\ell_1\) algorithms is in the order of \(O(N^2L)\) (using linear programming) or \(O(N^3)\) (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 \(\ell_1\) methods. The most recent coherence-based results regarding the convergence of OMP is reported in [1], where the authors also compare to commonly used \(\ell_1\) algorithms.

In this paper, we will improve the results of [1] 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 Sect. 2. We will analytically and numerically compare our results with [1]. 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 Sect. 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 Sect. 3.

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

\[
\left( 1 - 2N \exp \left( \frac{-s_{\min}^2}{8\tau N (\tau \gamma^2 + \frac{\sigma^2}{N})} \right) \right) \left( 1 - N \sqrt{\frac{2}{\pi}} \frac{\sigma}{\beta} e^{-\beta^2/2\sigma^2} \right),
\]

(7)

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

\[
\frac{\tau \beta^2}{(1 - \mu_{\max}(\tau - 1))^2}.
\]

(8)

Let us compare Theorem 1 with the theoretical results reported in [1]. 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 [1]. However, the analysis of Ben-Haim et al. [1] imposes a condition that is dependent on signal parameters such as \( s_{\min} \) and \( \tau \). 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 \( s_{\min} \), defines 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 [1], 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 is postponed to Appendix.

**Lemma 1** Define \( \Gamma_j = |\langle A_j, As + w \rangle| \), where \( j \in \{1, \ldots, N\} \), then for some constant \( \xi \geq 0 \) we have

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

(9)
where
\[ x_n = \mu_{j,n}s_n + \frac{1}{N} \langle A_j, w \rangle \]  
\[ |x_n| \leq c, \quad n = 1, \ldots, N \]  
\[ E \{ x_n^2 \} \leq \nu, \quad n = 1, \ldots, N \]  
\[ E \{ x_n^2 \} \leq \frac{1}{N} \tau s_{\max}^2 + \sigma^2 \]  
\[ \mu_{\max} \]

The following lemma explicitly formulates the upper bounds \( c \) and \( \nu \) introduced in (11) and (12).

**Lemma 2** Let \( x_n = \mu_{j,n}s_n + N^{-1} \langle A_j, w \rangle \), for any \( n \in \{1, \ldots, N\} \) and a fixed index \( j \in \{1, \ldots, N\} \). Assume that \( w \sim N(0, \sigma^2 I) \), and \( |\langle A_j, w \rangle| \leq \beta, \forall j \in \{1, \ldots, N\} \). Then,

\[ |x_n| \leq \mu_{\max} s_{\max} + \frac{\beta}{N}, \]  
\[ E \{ x_n^2 \} \leq \frac{1}{N} \tau s_{\max}^2 + \sigma^2 \]  
\[ \nu \]

We can now state the proof of Theorem 1.

**Proof** (Proof of Theorem 1) It is shown in [1] that when \( |\langle A_j, w \rangle| \leq \beta, \forall j \in \{1, \ldots, N\} \), then OMP identifies \( \Lambda \) if

\[ \min_{j \in \Lambda} |\langle A_j, A_{\Lambda}s_{\Lambda} + w \rangle| \geq \max_{k \notin \Lambda} |\langle A_k, A_{\Lambda}s_{\Lambda} + w \rangle|. \]  
\[ \min \]

Using the triangle inequality on the left-hand side of (15), we have

\[ \min_{j \in \Lambda} |\langle A_j, A_{\Lambda}s_{\Lambda} + w \rangle| \]
\[ = \min_{j \in \Lambda} |s_j + \langle A_j, A_{\Lambda \setminus \{j\}}s_{\Lambda \setminus \{j\}} + w \rangle| \]
\[ \geq \min_{j \in \Lambda} |s_j| - \max_{j \in \Lambda} |\langle A_j, A_{\Lambda \setminus \{j\}}s_{\Lambda \setminus \{j\}} + w \rangle|. \]
\[ \min \]

From (15) and (16), we can see that the OMP algorithm identifies the true support if

\[ \left\{ \begin{array}{l}
\max_{k \notin \Lambda} \{ \Gamma_k \} < \min_{j \in \Lambda} \frac{|s_j|}{2}, \\
\max_{j \in \Lambda} |\langle A_j, A_{\Lambda \setminus \{j\}}s_{\Lambda \setminus \{j\}} + w \rangle| < \min_{j \in \Lambda} \frac{|s_j|}{2}.
\end{array} \right. \]  
\[ \min \]

Using (17), we can define the probability of error for OMP as

\[ \Pr\{\text{error}\} = \Pr \left\{ \max_{j \in \Lambda} |\langle A_j, A_{\Lambda \setminus \{j\}}s_{\Lambda \setminus \{j\}} + w \rangle| \geq \frac{s_{\min}}{2} \right\} + \Pr \left\{ \max_{k \notin \Lambda} \{ \Gamma_k \} \geq \frac{s_{\min}}{2} \right\}, \]  
\[ \min \]
with the upper bound

\[
\Pr\{\text{error}\} \leq \sum_{j \in A} \Pr\left\{ \|A_j, A_{\Lambda \setminus \{j\}}s_{\Lambda \setminus \{j\}} + w\| \geq \frac{s_{\min}}{2} \right\} \\
+ \sum_{k \notin A} \Pr\left\{ \Gamma_k \geq \frac{s_{\min}}{2} \right\}.
\]  \hspace{1cm} (19)

For the first term on the right-hand side of (19), excluding the summation over indices in \(A\), from Lemma 1 we have

\[
\Pr_{j \in A} \left\{ \|A_j, A_{\Lambda \setminus \{j\}}s_{\Lambda \setminus \{j\}} + w\| \geq \frac{s_{\min}}{2} \right\} \\
\leq 2 \exp \left( \frac{-s_{\min}^2}{8((\tau - 1)\nu + s_{\min}c/6)} \right).
\]  \hspace{1cm} (20)

Note that unlike Lemma 1, the dictionary \(A\) in (20) 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 (20). From Lemma 2, the upper bounds \(c\) and \(\nu\), defined in (11) and (12), respectively, are

\[
|x_n| \leq \mu_{\max}s_{\max} + \frac{\beta}{N},
\]  \hspace{1cm} (21)

\[
E\{x_n^2\} \leq \frac{\tau - 1}{N}\mu_{\max}^2s_{\max}^2 + \frac{\sigma^2}{N^2}.
\]  \hspace{1cm} (22)

Combining (21) and (22) with (20) yields

\[
\Pr_{j \in A} \left\{ \|A_j, A_{\Lambda \setminus \{j\}}s_{\Lambda \setminus \{j\}} + w\| \geq \frac{s_{\min}}{2} \right\} \\
\leq 2 \exp \left( \frac{-s_{\min}^2}{8((\tau - 1)\nu + s_{\min}c/6)} \right),
\]  \hspace{1cm} (23)

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 (19), excluding the summation, yields

\[
\Pr_{k \notin A} \left\{ \Gamma_k \geq \frac{s_{\min}}{2} \right\} \\
\leq 2 \exp \left( \frac{-s_{\min}^2}{8(\tau\nu + \frac{\sigma^2}{N}) + 4s_{\min}c(\gamma + \frac{\beta}{N})} \right),
\]  \hspace{1cm} (24)
Substituting (23) and (24) into (19), we obtain

$$\Pr\{\text{error}\} \leq \tau P_1 + (N - \tau) P_2 \leq NP_2,$$

(25)

where the last inequality follows since $P_2 > P_1$. So far we have assumed that $|\langle A_j, w \rangle| \leq \beta$, $\forall j$. The probability of success is the joint probability of $\Pr\{|\langle A_j, w \rangle| \leq \beta\}$ and the inverse of (25). The former can be bounded as follows

$$\Pr\{|\langle A_j, w \rangle| \leq \beta\} = 1 - 2 Q\left(\frac{\beta}{\sigma}\right) \geq 1 - \sqrt{\frac{2}{\pi}} \frac{\beta e^{-\beta^2/2\sigma^2}}{P_3},$$

(26)

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

$$\Pr_{j=1, \ldots, N}\{|\langle A_j, w \rangle| \leq \beta\} \geq (1 - P_3)^N \geq 1 - NP_3,$$

(27)

Inverting the probability event in (25) and combining with (27) yields (7).

To prove (8), we proceed similar to [1]. Using the triangle inequality, we have:

$$\|s - \hat{s}\|_2^2 \leq \|\hat{s}_{orc} - s\|_2^2 + \|\hat{s}_{orc} - \hat{s}\|_2^2,$$

(28)

where $\hat{s}_{orc}$ is the oracle estimator, i.e. an estimator that knows the true support of $s$, a priori. If the OMP algorithm identifies the true support, then the second term on right-hand side of (28) will be zero. For the first term, we have

$$\|\hat{s}_{orc} - s\|_2^2 = \left\| (A_{A_0}^T A_{A_0})^{-1} A_{A_0}^T w \right\|_2^2 \leq \left\| (A_{A_0}^T A_{A_0})^{-1} \right\|_2^2 \sum_{j \in A_0} |\langle A_j, w \rangle|^2$$

$$\leq \left\| (A_{A_0}^T A_{A_0})^{-1} \right\|_2^2 \tau \beta^2,$$

(29)

The term $\left\| (A_{A_0}^T A_{A_0})^{-1} \right\|$ will be smaller than the maximum eigenvalue of $(A_{A_0}^T A_{A_0})^{-1}$, which is equal to inverse of the minimum eigenvalue of $A_{A_0}^T A_{A_0}$. According to the Gershgorin circle theorem [18], this number will be larger than $(1 - \mu_{\text{max}}(\tau - 1))$. Substituting this into (29) completes the proof.

3 Numerical Results

In this section, we compare numerical results of Theorem 1 and Ben-Haim et al. [1] with the empirically obtained results of the OMP algorithm. Throughout this section, we will refer to the analysis of Ben-Haim et al. [1] as CBPG. Moreover, we will only
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. [1].

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

compare the probability of success, i.e. Eq. (7), since extensive results on MSE have been reported in [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 $A = [I, H]$, where $H$ is the Hadamard
Parameters are $N = 1024$, $M = 512$, $\tau = 5$, $\sigma = 0.05$, $s_{\text{max}} = 1$

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 $\{1, 2, \ldots, N\}$, i.e. $\Lambda \in \{1, 2, \ldots, N\}$. The nonzero elements located at $\Lambda$ are drawn randomly from a uniform distribution on the interval $[s_{\text{min}}, s_{\text{max}}]$, 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 $\beta$. To do this, we empirically calculate $\beta$ as $\max \max_{w} |\langle A_j, w \rangle|$, where the maximum over $w$ is computed using $10^4$ samples from $w \sim \mathcal{N}(0, \sigma^2 I)$. Note that CBPG uses another constant, termed $\alpha$, which is related to $\beta$ by the definition $\beta \triangleq \sigma \sqrt{2(1 + \alpha)} \log N$, see [1] for more details. Hence, we can calculate $\alpha$ from the empirically obtained value of $\beta$.

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 $s_{\text{min}} = 0.5$, $s_{\text{max}} = 1$ 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 $\tau$. Most importantly, we see that the shape of the probability curves is 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 $\tau$ values.

In Fig. 2, we double the size of the signal. In this case, the value of mutual coherence and $\beta$ 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 do not change considerably. This is due to the fact that the condition imposed by CBPG is very sensitive to signal parameters; specifically, it is not satisfied for large values of \( \tau \). On the other hand, our analysis shows less sensitivity to \( \tau \) 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 \( s_{\text{max}} = 1 \), while varying \( s_{\text{min}} \in [0.01, 1] \). 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 Sect. 5.

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. [1], our analysis replaces a sharp condition with a probabilistic bound. Comparisons to empirical results obtained by OMP show a much improved correlation than in the 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 \( \Gamma_j \), we can show that

\[
\Gamma_j = \left| \langle A_j, As + w \rangle \right|
\]

\[
= \left| \sum_{m=1}^{M} A_{m,j} \left( \sum_{n=1}^{N} A_{m,n}s_n + w_m \right) \right|
\]

\[
= \left| \sum_{n=1}^{N} \left( \sum_{m=1}^{M} A_{m,j} A_{m,n}s_n + \frac{1}{N} \sum_{m=1}^{M} A_{m,j} w_m \right) \right|. \tag{30}
\]

Using (5), we have that

\[
\Gamma_j = \left| \sum_{n=1}^{N} \left( \mu_{j,n}s_n + \frac{1}{N} \langle A_j, w \rangle \right) \right| = \left| \sum_{n=1}^{N} x_n \right|. \tag{31}
\]

As mentioned in Sect. 1, we assume that the elements of the sparse vector \( s \) are centered random variables. Hence, the elements of \( s \) are either zero or zero-mean random variables, implying that \( \text{E}\{s_n\} = 0 \) for all \( n = 1, \ldots, N \). Together with the
fact that $E\{w\} = 0$, we have

$$E\{x_n\} = \mu_{j,n}E\{s_n\} + N^{-1}E\{\langle A_j, w \rangle\} = 0,$$

(32)

for all $n = 1, \ldots, N$. According to Bernstein’s inequality [2], if $x_1, \ldots, x_N$ are independent real random variables with mean zero, where $E\{x_n^2\} \leq \nu$, and $Pr\{|x_n| < c\} = 1$, then

$$Pr\left\{ \left| \sum_{n=1}^{N} x_n \right| \geq \xi \right\} \leq 2 \exp \left( -\frac{\xi^2}{2 \left( \sum_{n=1}^{N} E\{x_n^2\} + c\xi/3 \right)} \right) \leq 2 \exp \left( -\frac{\xi^2}{2(N\nu + c\xi/3)} \right),$$

(33)

where (33) follows using (12). This completes the proof.

Proof (proof of Lemma 2) Equation (13) follows trivially from the triangle inequality. For (14), we have

$$E\{x_n^2\} = \frac{1}{N} \sum_{n=1}^{N} E\{x_n^2\}$$

(34)

$$= \frac{1}{N} \sum_{n=1}^{N} \left[ \mu_{j,n}^2 s_n^2 + \frac{1}{N^2} \langle A_j, w \rangle^2 + \frac{2}{N} \mu_{j,n} s_n \langle A_j, w \rangle \right]$$

(35)

$$= \frac{1}{N} \sum_{n=1}^{N} \mu_{j,n}^2 E\{s_n^2\} + \frac{1}{N^2} E\{\langle A_j, w \rangle^2\} + \frac{2}{N} E\{\mu_{j,n} s_n \langle A_j, w \rangle\}$$

(36)

$$\leq \frac{\tau}{N} \mu_{\text{max}}^2 s_{\text{max}}^2 + \frac{1}{N^2} E\{\langle A_j, w \rangle^2\},$$

(37)

where the last term in (36) is zero since $E\{w\} = 0$, which implies $E\{\langle A_j, w \rangle\} = 0$. Moreover, we have

$$E\{\langle A_j, w \rangle^2\} = E\left\{ \left( \sum_{m=1}^{M} A_{m,j} w_m \right) \left( \sum_{m=1}^{M} A_{m,j} w_m \right) \right\}$$

$$= \sum_{m=1}^{M} A_{m,j} A_{m,j} E\{w_m w_m\} = \sigma^2.$$

(38)

Combining (37) and (38) completes the proof.
References

1. Z. Ben-Haim, Y. Eldar, M. Elad, Coherence-based performance guarantees for estimating a sparse vector under random noise. IEEE Trans. Signal Process. 58(10), 5030–5043 (2010). doi: 10.1109/TSP.2010.2052460

2. G. Bennett, Probability inequalities for the sum of independent random variables. J. Am. Stat. Assoc. 57(297), 33–45 (1962)

3. P. Bofill, M. Zibulevsky, Underdetermined blind source separation using sparse representations. Signal Process. 81(11), 2353–2362 (2001). doi: 10.1016/S0165-1684(01)00120-7

4. E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. 59(8), 1207–1223 (2006). doi: 10.1002/cpa.20124

5. A. Castrodad, Z. Xing, J.B. Greer, E. Bosch, L. Carin, G. Sapiro, Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. 49(11), 4263–4281 (2011). doi: 10.1109/TGRS.2011.2163822

6. S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. 20, 33–61 (1998)

7. D. Donoho, Compressed sensing. IEEE Trans. Inf. Theory 52(4), 1289–1306 (2006). doi: 10.1109/TIT.2006.871582

8. D. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inf. Theory 47(7), 2845–2862 (2001). doi: 10.1109/18.959265

9. D. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inf. Theory 52(1), 6–18 (2006). doi: 10.1109/TIT.2005.860430

10. D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. Proc. Natl. Acad. Sci. 100(5), 2197–2202 (2003). doi: 10.1073/pnas.0437847100. http://www.pnas.org/content/100/5/2197.full.pdf

11. M.F. Duarte, M.A. Davenport, D. Takbar, J.N. Laska, T. Sun, K.F. Kelly, R.G. Baraniuk, Single-pixel imaging via compressive sampling. IEEE Signal Process. Mag. 25(2), 83–91 (2008). doi: 10.1109/MSP.2007.914730

12. B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression. Ann. Stat. 32(2), 407–499 (2004)

13. M. Elad, Sparse and Redundant Representations (Springer, New York, 2010). doi: 10.1007/978-1-4419-7011-4

14. M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. 15(12), 3736–3745 (2006). doi: 10.1109/TIP.2006.881969

15. Y.C. Eldar, G. Kutyniok (eds.), Compressed Sensing Theory and Applications (Cambridge University Press, Cambridge, 2012)

16. M. Emadi, K. Sadeghi, DOA estimation of multi-reflected known signals in compact arrays. EEE Trans. Aerosp. Electron. Syst. 49(3), 1920–1931 (2013). doi: 10.1109/TAES.2013.6558028

17. M.A.T. Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. IEEE J. Sel. Top. Signal Process. 1(4), 586–597 (2007). doi: 10.1109/JSTSP.2007.910281

18. G. Golub, C. Van Loan, Matrix Computations. Johns Hopkins Studies in the Mathematical Sciences (Johns Hopkins University Press, Baltimore, Ann Arbor, MI, 1996)

19. S.H. Hsieh, C.S. Lu, S.C. Pei, Fast omp: reformulating omp via iteratively refining l2-norm solutions, in 2012 IEEE Statistical Signal Processing Workshop (SSP), 2012, pp. 189–192. doi: 10.11109/SSP.2012.6319656

20. Z.M. Liu, Z.T. Huang, Y.Y. Zhou, Direction-of-arrival estimation of wideband signals via covariance matrix sparse representation. IEEE Trans. Signal Process. 59(9), 4256–4270 (2011). doi: 10.1109/TSP.2011.2159214

21. D. Malioutov, M. Cetin, A. Wilsky, A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans. Signal Process. 53(8), 3010–3022 (2005). doi: 10.1109/TSP.2005.850882

22. S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process. 41(12), 3397–3415 (1993). doi: 10.1109/78.258082

23. F. Marvasti, A. Amini, F. Haddadi, M. Soltanolkotabi, B.H. Khalaj, A. Aldroubi, S. Sanei, J. Chambers, A unified approach to sparse signal processing. EURASIP J. Adv. Signal Process. 2012, 44 (2012)
24. E. Miandji, J. Kronander, J. Unger, Compressive image reconstruction in reduced union of subspaces. Comput. Graph. Forum 34(2), 33–44 (2015). doi: 10.1111/cgf.12539
25. D. Needell, J. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. 26(3), 301–321 (2009). doi: 10.1016/j.acha.2008.07.002
26. D. Needell, R. Vershynin, Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. IEEE J. Sel. Top. Signal Process. 4(2), 310–316 (2010). doi: 10.1109/JSTSP.2010.2042412
27. Y. Pati, R. Rezaifar, P.S. Krishnaprasad, 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, 1993, pp. 40–44. doi: 10.1109/ACSSC.1993.342465
28. G. Pope, Compressive sensing: a summary of reconstruction algorithms, Master’s thesis, ETH Zürich, 2009
29. R. Tibshirani, Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B 58, 267–288 (1994)
30. J. Tropp, Greed is good: algorithmic results for sparse approximation. IEEE Trans. Inf. Theory 50(10), 2231–2242 (2004). doi: 10.1109/TIT.2004.834793
31. J. Tropp, Just relax: convex programming methods for identifying sparse signals in noise. IEEE Trans. Inf. Theory 52(3), 1030–1051 (2006). doi: 10.1109/TIT.2005.864420
32. E. van den Berg, M.P. Friedlander, Probing the pareto frontier for basis pursuit solutions. SIAM J. Sci. Comput. 31(2), 890–912 (2009). doi: 10.1137/080714488
33. S. Wright, R. Nowak, M. Figueiredo, Sparse reconstruction by separable approximation. IEEE Trans. Signal Process. 57(7), 2479–2493 (2009). doi: 10.1109/TSP.2009.2016892
34. J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation. IEEE Trans. Image Process. 19(11), 2861–2873 (2010). doi: 10.1109/TIP.2010.2050625