Orthogonal Matching Pursuit From Noisy Random Measurements: A New Analysis

Authored by:

Sundeep Rangan
Alyson K. Fletcher

Abstract

Orthogonal matching pursuit (OMP) is a widely used greedy algorithm for recovering sparse vectors from linear measurements. A well-known analysis of Tropp and Gilbert shows that OMP can recover a k-sparse n-dimensional real vector from m = 4k log(n) noise-free random linear measurements with a probability that goes to one as n goes to infinity. This work shows strengthens this result by showing that a lower number of measurements, m = 2k log(n-k), is in fact sufficient for asymptotic recovery. Moreover, this number of measurements is also sufficient for detection of the sparsity pattern (support) of the vector with measurement errors provided the signal-to-noise ratio (SNR) scales to infinity. The scaling m = 2k log(n-k) exactly matches the number of measurements required by the more complex lasso for signal recovery.

1 Paper Body

A well-known analysis of Tropp and Gilbert shows that orthogonal matching pursuit (OMP) can recover a k-sparse n-dimensional real vector from m = 4k log(n) noise-free linear measurements obtained through a random Gaussian measurement matrix with a probability that approaches one as n ? ?. This work strengthens this result by showing that a lower number of measurements, m = 2k log(n ? k), is in fact sufficient for asymptotic recovery. More generally, when the sparsity level satisfies kmin ? k ? kmax but is unknown, m = 2kmax log(n ? kmin) measurements is sufficient. Furthermore, this number of measurements is also sufficient for detection of the sparsity pattern (support) of the vector with measurement errors provided the signal-to-noise ratio (SNR) scales to infinity. The scaling m = 2k log(n ? k) exactly matches the number of measurements required by the more complex lasso method for signal recovery in a similar SNR scaling.

1 Introduction Suppose x ? Rn is a sparse vector, meaning its number of nonzero components k is smaller than n. The support of x is the locations of the nonzero entries and is sometimes called its sparsity pattern. A common sparse

estimation problem is to infer the sparsity pattern of x from linear measurements of the form y = Ax + w, (1) where A ? Rm?n is a known measurement matrix, y ? Rm represents a vector of measurements and w ? Rm is a vector of measurements errors (noise). Sparsity pattern detection and related sparse estimation problems are classical problems in nonlinear signal processing and arise in a variety of applications including wavelet-based image processing [1] and statistical model selection in linear regression [2]. There has also been considerable recent interest in sparsity pattern detection in the context of compressed sensing, which focuses on large random measurement matrices A [3?5]. It is this scenario with random measurements that will be analyzed here.

Optimal subset recovery is NP-hard [6] and usually involves searches over all the nk possible support sets of x. Thus, most attention has focused on approximate methods for reconstruction. One simple and popular approximate algorithm is orthogonal matching pursuit (OMP) developed in [7?9]. OMP is a simple greedy method that identifies the location of one nonzero component of x at a time. A version of the algorithm will be described in detail below in Section 2. The best known ?

This work was supported in part by a University of California President?s Postdoctoral Fellowship and the ? Centre Bernoulli at Ecole Polytechnique F?ed?erale de Lausanne.

1

analysis of the performance of OMP for large random matrices is due to Tropp and Gilbert [10, 11]. Among other results, Tropp and Gilbert show that when the number of measurements scales as m ? (1 + ?)4k log(n)

(2)

for some ? ¿ 0, A has i.i.d. Gaussian entries, and the measurements are noise-free (w = 0), the OMP method will recover the correct sparse pattern of x with a probability that approaches one as n and k ? ?. Deterministic conditions on the matrix A that guarantee recovery of x by OMP are given in [12]. However, numerical experiments reported in [10] suggest that a smaller number of measurements than (2) may be sufficient for asymptotic recovery with OMP. Specifically, the experiments suggest that the constant 4 can be reduced to 2. Our main result, Theorem 1 below, proves this conjecture. Specifically, we show that the scaling in measurements m ? (1 + ?)2k log(n ? k) (3)

is also sufficient for asymptotic reliable recovery with OMP provided both n ? k and k ? ?. The result goes further by allowing uncertainty in the sparsity level k. We also improve upon the Tropp?Gilbert analysis by accounting for the effect of the noise w. While the Tropp?Gilbert analysis requires that the measurements are noise-free, we show that the scaling (3) is also sufficient when there is noise w, provided the signal-to-noise ratio (SNR) goes to infinity. The main significance of the new scaling (3) is that it exactly matches the conditions for sparsity pattern recovery using the well-known lasso method. The lasso method, which will be described in detail in Section 4, is based on a convex relaxation of the optimal detection problem. The best analysis of the sparsity pattern recovery with lasso is due to Wainwright [13, 14]. He showed in [13] that under a similar high SNR assumption, the scaling (3) in number of measure-

ments is both necessary and sufficient for asymptotic reliable sparsity pattern detection.1 Now, although the lasso method is often more complex than OMP, it is widely believed that lasso has superior performance [10]. Our results show that at least for sparsity pattern recovery with large Gaussian measurement matrices in high SNR, lasso and OMP have identical performance. Hence, the additional complexity of lasso for these problems is not warranted. Of course, neither lasso nor OMP is the best known approximate algorithm, and our intention is not to claim that OMP is optimal in any sense. For example, where there is no noise in the measurements, the lasso minimization (14) can be replaced by b = arg min kvk1 , s.t. y = Av. x v?Rn

A well-known analysis due to Donoho and Tanner [15] shows that, for i.i.d. Gaussian measurement matrices, this minimization will recover the correct vector with m ? 2k log(n/m)

(4)

when k ? n. This scaling is fundamentally better than the scaling (3) achieved by OMP and lasso. There are also several variants of OMP that have shown improved performance. The CoSaMP algorithm of Needell and Tropp [16] and subspace pursuit algorithm of Dai and Milenkovic [17] achieve a scaling similar to (4). Other variants of OMP include the stagewise OMP [18] and regularized OMP [19]. Indeed with the recent interest in compressed sensing, there is now a wide range of promising algorithms available. We do not claim that OMP achieves the best performance in any sense. Rather, we simply intend to show that both OMP and lasso have similar performance in certain scenarios. Our proof of (3) follows along the same lines as Tropp and Gilbert?s proof of (2), but with two key differences. First, we account for the effect of the noise by separately considering its effect in the ?true? subspace and its orthogonal complement. Second and more importantly, we provide a tighter bound on the maximum correlation of the incorrect vectors. Specifically, in each iteration of the 1 Sufficient conditions under weaker conditions on the SNR are more subtle [14]: the scaling of SNR with n determines the sequences of regularization parameters for which asymptotic almost sure success is achieved, and the regularization parameter sequence affects the sufficient number of measurements.

2

OMP algorithm, there are n ? k possible incorrect vectors that the algorithm can choose. Since the algorithm runs for k iterations, there are total of k(n ? k) possible error events. The Tropp and Gilbert proof bounds the probability of these error events with a union bound, essentially treating them as statistically independent. However, here we show that energies on any one of the incorrect vectors across the k iterations are correlated. In fact, they are precisely described by samples on a certain normalized Brownian motion. Exploiting this correlation we show that the tail bound on error probability grows as n ? k, not k(n ? k), independent events. The outline of the remainder of this paper is as follows. Section 2 describes the OMP algorithm. Our main result, Theorem 1, is stated in Section 3. A comparison to lasso is provided in Section 4, and we suggest some future problems in Section 6. The proof of the main result is sketched in Section 7.

2 Orthogonal Matching Pursuit To describe the algorithm, suppose we wish to determine the vector x from a vector y of the form (1). Let Itrue = { j : xj 6= 0 }, (5) which is the support of the vector x. The set Itrue will also be called the sparsity pattern. Let k = —Itrue —, which is the number of nonzero components of x. The OMP algorithm produces a ? sequence of estimates I(t), t = 0, 1, 2, . . ., of the sparsity pattern Itrue , adding one index at a time. In the description below, let aj denote the jth column of A. Algorithm 1 (Orthogonal Matching Pursuit) Given a vector y ? Rm , a measurement matrix A ? Rm?n and threshold level ? ¿ 0, compute an estimate I?OMP of the sparsity pattern of x as follows: ? = ?. 1. Initialize t = 0 and I(t) 2. Compute P(t), the projection operator onto the orthogonal complement of the span of ? {ai , i ? I(t)}. 3. For each j, compute ?(t, j) =

—a?j P(t)y—2 , kP(t)yk2

and let [?? (t), i? (t)] = max ?(t, j), j=1,...,n

(6)

where ?? (t) is the value of the maximum and i? (t) is an index which achieves the maximum. ? + 1) = I(t) ? ? {i? (t)}. Also, increment t = t + 1 and return to step 2. 4. If ?? (t) ¿ ?, set I(t ? 5. Otherwise stop. The final estimate of the sparsity pattern is I?OMP = I(t). ? Note that since P(t) is the projection onto the orthogonal complement of aj for all j ? I(t), ? ? P(t)aj = 0 for all j ? I(t). Hence, ?(t, j) = 0 for all j ? I(t), and therefore the algorithm will not select the same vector twice. The algorithm above only provides an estimate, I?OMP , of the sparsity pattern of Itrue . Using I?OMP , one can estimate the vector x in a number of ways. For example, one can take the least-squares estimate, b = arg min ky ? Avk2 x (7) b is where the minimization is over all vectors v such vj = 0 for all j 6? I?OMP . The estimate x the projection of the noisy vector y onto the space spanned by the vectors ai with i in the sparsity pattern estimate I?OMP . However, this paper only analyzes the sparsity pattern estimate I?OMP itself, b. and not the vector estimate x 3

3 Asymptotic Analysis We analyze the OMP algorithm in the previous section under the following assumptions. Assumption 1 Consider a sequence of sparse recovery problems, indexed by the vector dimension n. For each n, let x ? Rn be a deterministic vector and let k = k(n) be the number of nonzero components in x. Also assume: (a) The sparsity level, k = k(n) satisfies k(n) ? [kmin (n), kmax (n)],

(8)

for some deterministic sequences kmin (n) and kmax (n) with kmin (n) ? ? as n ? ? and kmax (n) ¡ n/2 for all n. (b) The number of measurements m = m(n) is a deterministic sequence satisfying m ? (1 + ?)2kmax log(n ? kmin),

(9)

for some ? ¿ 0. (c) The minimum component power x2min satisfies lim kx2min = ?,

(10)

xmin = min —xj —,

(11)

n??

where j?Itrue

is the magnitude of the smallest nonzero component of x. (d) The powers of the vectors kxk2 satisfy

1 log 1 + kxk2 = 0. ? n?? (n ? k) lim

for all ? ¿ 0.

(12)

(e) The vector y is a random vector generated by (1) where A and w have i.i.d. Gaussian components with zero mean and variance of 1/m. Assumption 1(a) provides a range on the sparsity level, k. As we will see below in Section 5, bounds on this range are necessary for proper selection of the threshold level ? ¿ 0. Assumption 1(b) is our the main scaling law on the number of measurements that we will show is sufficient for asymptotic reliable recovery. In the special case when k is known so that kmax = kmin = k, we obtain the simpler scaling law m ? (1 + ?)2k log(n ? k).

(13)

We have contrasted this scaling law with the Tropp?Gilbert scaling law (2) in Section 1. We will also compare it to the scaling law for lasso in Section 4. Assumption 1(c) is critical and places constraints on the smallest component magnitude. The importance of the smallest component magnitude in the detection of the sparsity pattern was first recognized by Wainwright [13,14,20]. Also, as discussed in [21], the condition requires that signalto-noise ratio (SNR) goes to infinity. Specifically, if we define the SNR as SNR =

EkAxk2 , kwk2

then under Assumption 1(e), it can be easily checked that SNR = kxk2 .

Since x has k nonzero components, kxk2 ? kx2min , and therefore condition (10) requires that SNR ? ?. For this reason, we will call our analysis of OMP a high-SNR analysis. The analysis of OMP with SNR that remains bounded above is an interesting open problem. 4

Assumption (d) is technical and simply requires that the SNR does not grow too quickly with n. Note that even if SNR = O(k ?) for any ? ¿ 0, Assumption 1(d) will be satisfied. Assumption 1(e) states that our analysis concerns large Gaussian measurement matrices A and Gaussian noise w. Theorem 1 Under Assumption 1, there exists a sequence of threshold levels ? = ?(n) such that the OMP method in Algorithm 1 will asymptotically detect the correct sparsity pattern in that

lim Pr I?OMP 6= Itrue = 0. n??

Moreover, the threshold levels ? can be selected simply as a function of kmin , kmax , n, m and ?. Theorem 1 provides our main result and shows that the scaling law (9) is sufficient for asymptotic recovery.

4 Comparison to Lasso Performance It is useful to compare the scaling law (13) to the number of measurements required by the widelyused lasso method described for example in [22]. The lasso method finds an estimate for the vector x in (1) by solving the quadratic program b = arg min ky ? Avk2 + ?kvk1 , x

(14)

v?Rn

where ? ¿ 0 is an algorithm parameter that trades off the prediction error with the sparsity of the solution. Lasso is sometimes referred to as basis pursuit denoising [23]. While the optimization (14) is convex, the running time of lasso is significantly longer than OMP unless A has some particular structure [10]. However, it is generally believed that lasso has superior performance. The best analysis of lasso for sparsity pattern recovery for large random matrices is due to Wainwright [13, 14]. There, it is shown that with an i.i.d. Gaussian measurement matrix and white Gaussian noise, the condition (13) is necessary for asymptotic reliable detection of the sparsity pattern. In addition, under the condition (10) on the minimum component magnitude, the scaling (13) is also sufficient. We thus conclude that OMP requires an identical scaling in the number of measurements to lasso. Therefore, at least for sparsity pattern recovery from measurements with large random Gaussian measurement matrices and high SNR, there is no additional performance improvement with the more complex lasso method over OMP.

5 Threshold Selection and Stopping Conditions In many problems, the sparsity level k is not known a priori and must be detected as part of the estimation process. In OMP, the sparsity level of estimated vector is precisely the number of iterations conducted before the algorithm terminates. Thus, reliable sparsity level estimation requires a good stopping condition. When the measurements are noise-free and one is concerned only with exact signal recovery, the optimal stopping condition is simple: the algorithm should simply stop whenever there is no more error. That is ?? (t) = 0 in (6). However, with noise, selecting the correct stopping condition requires some care. The OMP method as described in Algorithm 1 uses a stopping condition based on testing if ?? (t) ¿ ? for some threshold ?. One of the appealing features of Theorem 1 is that it provides a simple sufficient condition under which this threshold mechanism will detect the correct sparsity level. Specifically, Theorem 1 provides a range k ? [kmin , kmax] under which there exists a threshold that the OMP algorithm will terminate in the correct number of iterations. The larger the number of measurements, m, the greater one can make the range [kmin , kmax]. The formula for the threshold level is given in (20). Of course, in practice, one may deliberately want to stop the OMP algorithm with fewer iterations than the ?true? sparsity level. As the OMP method proceeds, the detection becomes less reliable and it is sometimes useful to stop the algorithm whenever there is a high chance of error. Stopping early 5

may miss some small components, but may result in an overall better estimate by not introducing too many erroneous components or components with too much noise. However, since our analysis is only concerned with exact sparsity pattern recovery, we do not consider this type of stopping condition.

6 Conclusions and Future Work We have provided an improved scaling law on the number of measurements for asymptotic reliable sparsity pattern detection with OMP. This scaling law exactly matches the scaling needed by lasso under similar conditions. However, much about the performance of OMP is still not fully understood. Most importantly, our analysis is limited to high SNR. It would be interesting to see if reasonable sufficient conditions can be derived

for finite SNR as well. Also, our analysis has been restricted to exact sparsity pattern recovery. However, in many problems, especially with noise, it is not necessary to detect every component in the sparsity pattern. It would be useful if partial support recovery results such as [24?27] can be obtained for OMP. Finally, our main scaling law (9) is only sufficient. While numerical experiments in [10, 28] suggest that this scaling is also necessary for vectors with equal magnitude, it is possible that OMP can perform better than the scaling law (9) when the component magnitudes have some variation; this is demonstrated numerically in [28]. The benefit of dynamic range in an OMP-like algorithm has also been observed in [29] and sparse Bayesian learning methods in [30, 31].

7 Proof Sketch for Theorem 1 7.1 Proof Outline Due to space considerations, we only sketch the proof; additional details are given in [28]. The main difficulty in analyzing OMP is the statistical dependencies between iterations in the OMP algorithm. Following along the lines of the Tropp?Gilbert proof in [10], we avoid these difficulties by considering the following ?genie? algorithm. A similar alternate algorithm is analyzed in [29]. 1. Initialize t = 0 and Itrue (t) = ?. 2. Compute Ptrue (t), the projection operator onto the orthogonal complement of the span of {ai , i ? Itrue (t)}. 3. For all j = 1, . . . , n, compute ?true (t, j) =

—a?j Ptrue (t)y—2 , kPtrue (t)yk2

(15)

and let [??true (t), i? (t)] = max ?true (t, j). j?Itrue

(16)

4. If t ¡ k, set Itrue (t + 1) = Itrue (t) ? {i? (t)}. Increment t = t + 1 and return to step 2. 5. Otherwise stop. The final estimate of the sparsity pattern is Itrue (k). This ?genie? algorithm is identical to the regular OMP method in Algorithm 1, except that it runs for precisely k iterations as opposed to using a threshold ? for the stopping condition. Also, in the maximization in (16), the genie algorithm searches over only the correct indices j ? Itrue . Hence, this genie algorithm can never select an incorrect index j 6? Itrue . Also, as in the regular OMP algorithm, the genie algorithm will never select the same vector twice for almost all vectors y. Therefore, after k iterations, the genie algorithm will have selected all the k indices in Itrue and terminate with correct sparsity pattern estimate Itrue (k) = Itrue with probability one. So, we need to show that true OMP algorithm behaves identically to the ?genie? algorithm with high probability. 6

To this end, define the following two probabilities:

pMD = Pr max min ?true (t, j) ? ? t=0,...k?1 j?Itrue

pFA = Pr max max ?true (t, j) ? ? t=0,...k j6?Itrue

(17) (18)

Both probabilities are implicitly functions of n. The first term, pMD , can be interpreted as a ?missed detection? probability, since it corresponds to the event that the maximum correlation energy ?true (t, j) on the correct vectors j ? Itrue falls below the threshold. We call the second term pFA the ?false alarm? probability since it corresponds to the maximum energy on one of the ?incorrect? indices j 6? Itrue exceeding the threshold. A simple induction argument shows that if there are no missed detections or false alarms, the true OMP algorithm

will select the same vectors as the ?genie? algorithm, and therefore recover the sparsity pattern. This shows that

Pr I?OMP 6= Itrue ? pMD + pFA . So we need to show that there exists a sequence of thresholds ? = ?(n) ¿ 0, such that pMD and pFA ? 0 as n ? ?. To set this threshold, we select an ? ¿ 0 such that 1+? ? 1 + ?, 1+? where ? is from (9). Then, define the threshold level ? = ?(n) =

(19)

2(1 + ?) log(n ? kmin). m

(20)

7.2 Probability of Missed Detection The proof that pMD ? 0 is similar to that of Tropp and Gilbert?s proof in [10]. The key modification is to use (10) to show that the effect of the noise is asymptotically negligible so that for large n, y ? Ax = ?xtrue .

(21)

This is done by separately considering the components of w in the span of the vectors aj for j ? Itrue and its orthogonal complement. One then follows the Tropp?Gilbert proof for the noise-free case to show that max ?true (t, j) ?

j?Itrue

1 k

for large k. Hence, using (9) and (20) one can then show lim inf max

n?? j?Itrue

1 ?true (t, j) ? 1 + ?, ?

which shows that pMD ? 0. 7.3 Probability of False Alarm This part is harder. Define

a?j Ptrue (t)y , kPtrue (t)yk so that ?true (t, j) = —z(t, j)—2 . Now, Ptrue (t) and y are functions of w and aj for j ? Itrue . Therefore, they are independent of aj for any j 6? Itrue . Also, since the vectors aj have i.i.d. Gaussian components with variance 1/m, conditional on Ptrue (t) and y, z(t, j) is normal with variance 1/m. Hence, m?true (t, j) is a chi-squared random variable with one degree of freedom. z(t, j) =

Now, there are k(n ? k) values of ?true (t, k) for t = 1, . . . , k and j 6? Itrue . The Tropp?Gilbert proof bounds the maximum of these k(n ? k) value by the standard tail bound max

max ?true (t, j) ?

j6?Itrue t=1,...,k

2 4 2 log(k(n ? k)) ? log(n2) = log(n). m m m 7

To improve the bound in this proof, we exploit the fact that for any j, the values of z(t, j) are correlated. In fact, we show that the values z(t, j), t = 1, . . . , k are distributed identically to points on a normalized Brownian motion. Specifically, let W (s) be a standard linear Brownian motion and let S(s) be the normalized Brownian motion 1 S(s) = ? B(s), s ¿ 0. s

(22)

We then show that, for every j, there exists times s1 , . . . , sk with 1 ? s1 ¡ ? ? ? ¡ sk ? 1 + kxk2 such that the vector z(j) = [z(1, j), . . . , z(k, j)] is identically distributed to [S(s1), . . . , S(sj)]. Hence,

max —z(t, j)—2 = max —S(sj)—2 ?

t=1,...,k
t=1,...,k
sup s?[1,1+kxk2]
—S(s)—2 .

The right-hand side of the sample path can then be bounded by the reflection principle [32]. This yields an improved bound, max
max ?true (t, j) ?
j6?Itrue t=1,...,k
2 log(n ? k). m
Combining this with (20) shows lim inf max
n?? j?Itrue
1 1 ?true (t, j) ? , ? 1+?
which shows that pFA ? 0.

2 References

[1] S. Mallat. A Wavelet Tour of Signal Processing. Academic Press, second edition, 1999. [2] A. Miller. Subset Selection in Regression. Number 95 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, New York, second edition, 2002. [3] E. J. Cand‘es, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory, 52(2):489? 509, February 2006. [4] D. L. Donoho. Compressed sensing. IEEE Trans. Inform. Theory, 52(4):1289?1306, April 2006. [5] E. J. Cand‘es and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. Inform. Theory, 52(12):5406?5425, December 2006. [6] B. K. Natarajan. Sparse approximate solutions to linear systems. SIAM J. Computing, 24(2):227?234, April 1995. [7] S. Chen, S. A. Billings, and W. Luo. Orthogonal least squares methods and their application to non-linear system identification. Int. J. Control, 50(5):1873?1896, November 1989. [8] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Conf. Rec. 27th Asilomar Conf. Sig., Sys., & Comput., volume 1, pages 40?44, Pacific Grove, CA, November 1993. [9] G. Davis, S. Mallat, and Z. Zhang. Adaptive time-frequency decomposition. Optical Eng., 37(7):2183?2191, July 1994. [10] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inform. Theory, 53(12):4655?4666, December 2007. [11] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit: The Gaussian case. Appl. Comput. Math. 2007-01, California Inst. of Tech., August 2007. 8

[12] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. IEEE Trans. Inform. Theory, 50(10):2231?2242, October 2004. [13] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy recovery of sparsity. Technical report, Univ. of California, Berkeley, Dept. of Statistics, May 2006. arXiv:math.ST/0605740 v1 30 May 2006. [14] M. J. Wainwright.

Sharp thresholds for high-dimensional and noisy sparsity recovery using ?1 - constrained quadratic programming (lasso). IEEE Trans. Inform. Theory, 55(5):2183?2202, May 2009. [15] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. J. Amer. Math. Soc., 22(1):1?53, January 2009. [16] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harm. Anal., 26(3):301?321, July 2008. [17] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing: Closing the gap between performance and complexity. arXiv:0803.0811v1 [cs.NA]., March 2008. [18] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. preprint, March 2006. [19] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. Found. Comput. Math., 9(3):317?334, June 2008. [20] M. J. Wainwright. Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. Technical Report 725, Univ. of California, Berkeley, Dept. of Statistics, January 2007. [21] A. K. Fletcher, S. Rangan, and V. K. Goyal. Necessary and sufficient conditions for sparsity pattern recovery. IEEE Trans. Inform. Theory, 55(12), December 2009. To appear. Original submission available online [33]. [22] R. Tibshirani. Regression shrinkage and selection via the lasso. J. Royal Stat. Soc., Ser. B, 58(1):267?288, 1996. [23] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. SIAM J. Sci. Comp., 20(1):33?61, 1999. [24] M. Akc?akaya and V. Tarokh. Noisy compressive sampling limits in linear and sublinear regimes. In Proc. Conf. on Inform. Sci. & Sys., pages 1?4, Princeton, NJ, March 2008. [25] M. Akc?akaya and V. Tarokh. Shannon theoretic limits on noisy compressive sampling. arXiv:0711.0366v1 [cs.IT]., November 2007. [26] G. Reeves. Sparse signal sampling using noisy linear projections. Technical Report UCB/EECS-2008-3, Univ. of California, Berkeley, Dept. of Elec. Eng. and Comp. Sci., January 2008. [27] S. Aeron, M. Zhao, and V. Saligrama. On sensing capacity of sensor networks for the class of linear observation, fixed SNR models. arXiv:0704.3434v3 [cs.IT]., June 2007. [28] A. K. Fletcher and S. Rangan. Sparse support recovery from random measurements with orthogonal matching pursuit. Manuscript available online at http://www.eecs.berkeley.edu/?alyson/Publications/FletcherRangan OMP.pdf, October 2009. [29] A. K. Fletcher, S. Rangan, and V. K. Goyal. On?off random access channels: A compressed sensing framework. arXiv:0903.1022v1 [cs.IT]., March 2009. [30] Y. Jin and B. Rao. Performance limits of matching pursuit algorithms. In Proc. IEEE Int. Symp. Inform. Th., pages 2444?2448, Toronto, Canada, June 2008. [31] D. Wipf and B. Rao. Comparing the effects of different weight distributions on finding sparse representations. In Proc. Neural Information Process. Syst., Vancouver, Canada, December 2006. [32] I. Karatzas and S. E. Shreve. Brownian Motion and Stochastic Calculus. Springer-Verlag, New York, NY, 2nd edition, 1991. [33] A. K. Fletcher, S. Rangan, and V. K. Goyal. Necessary and sufficient conditions on sparsity pattern recovery. arXiv:0804.1839v1 [cs.IT]., April 2008.

9