Orthogonal Matching Pursuit: A Brownian Motion Analysis

Alyson K. Fletcher, *Member, IEEE*, and Sundeep Rangan, *Member, IEEE*

***Abstract*—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 $n \to \infty$. 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 $k_{\min} \leq k \leq k_{\max}$ but is unknown, $m = 2k_{\max}\log(n-k_{\min})$ 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 with a similar SNR scaling.**

***Index Terms*—Compressed sensing, detection, lasso, orthogonal matching pursuit (OMP), random matrices, sparse approximation, sparsity, subset selection.**

I. Introduction

Suppose $\mathbf{x} \in \mathbb{R}^n$ is a sparse vector, meaning its number of nonzero entries k is smaller than n. The *support* of $\mathbf{x}$ 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 $\mathbf{x}$ from linear measurements of the form

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

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a known measurement matrix, $\mathbf{y} \in \mathbb{R}^m$ represents a vector of measurements and $\mathbf{w} \in \mathbb{R}^m$ is a vector of measurement 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 $\mathbf{A}$ [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 $\binom{n}{k}$ possible support sets of $\mathbf{x}$. Thus, most attention has focused on approximate methods. One simple and popular approximate algorithm is orthogonal matching pursuit (OMP) [7]–[9]. OMP is a greedy method that identifies the location of one nonzero entry of $\mathbf{x}$ at a time. A version of the algorithm will be described in detail in Section II. The best known analysis of the detection performance of OMP for large random matrices is due to Tropp and Gilbert [10], [11]. Among other results, Tropp and Gilbert show that when $\mathbf{A}$ has i.i.d. Gaussian entries, the measurements are noise-free ($\mathbf{w} = 0$), and the number of measurements scales as

$$m \geq (1+\delta)4k\log(n) \tag{2}$$

for some $\delta > 0$, the OMP method will recover the correct sparse pattern of $\mathbf{x}$ with a probability that approaches one as n and $k \to \infty$. The analysis uses a deterministic sufficient condition for success on the matrix $\mathbf{A}$ based on a greedy selection ratio introduced in [12]. A similar deterministic condition on $\mathbf{A}$ was presented in [13], and a condition using the restricted isometry property was given in [14].

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, does a bit better than proving this conjecture. We show that the scaling in measurements

$$m \geq (1+\delta)2k\log(n-k) \tag{3}$$

is sufficient for asymptotic reliable recovery with OMP provided both $n-k$ and $k \to \infty$. Theorem 1 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 $\mathbf{w}$. 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 $\mathbf{w}$, provided the signal-to-noise ratio (SNR) goes to infinity.

On the other hand, it should be pointed out that our analysis is slightly less complete than the results in [10] since we only provide asymptotic guarantees without convergence rates on the probability of error. In addition, unlike deterministic analyses based on conditions such as in [12] or [14], our results hold only pointwise over the unknown vector as opposed to providing a

Manuscript received May 30, 2011; revised October 21, 2011; accepted October 24, 2011. Date of publication November 22, 2011; date of current version February 10, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bogdan Dumitrescu. This work was supported in part by a University of California President's Postdoctoral Fellowship. The material in this paper was presented in part at the Conference on Neural Information Processing Systems, Vancouver, BC, Canada, December 2009.

A. K. Fletcher is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: alyson@eecs.berkeley.edu).

S. Rangan is with the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn, NY 11201 USA (e-mail: srangan@poly.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2176936

uniform guarantee. See the remarks after Theorem 1 for more details.

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 [15]. The lasso method, which will be described in detail in Section IV, is based on a convex relaxation of the optimal detection problem. The best analysis of sparsity pattern recovery with lasso is due to Wainwright [16], [17]. He showed in [16] that under a similar high SNR assumption, the scaling (3) in number of measurements is both necessary and sufficient for asymptotic reliable sparsity pattern detection.[1] The lasso method is often more complex than OMP, but it is widely believed to offset this disadvantage with superior performance [10]. Our results show that, at least for sparsity pattern recovery under our asymptotic assumptions, OMP performs at least as well as lasso.[2] Hence, the additional complexity of lasso for these problems may not be warranted.

Neither lasso nor OMP is the best known approximate algorithm for sparsity pattern recovery. For example, where there is no noise in the measurements, the lasso minimization (16) can be replaced by

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{v}\in\mathbb{R}^n} \|\mathbf{v}\|_1, \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{v}.$$

A well-known analysis due to Donoho and Tanner [18] shows that, for i.i.d. Gaussian measurement matrices, this minimization will recover the correct vector with

$$m \asymp 2k\log(n/m) \tag{4}$$

when $k \ll n$. 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 [19] and subspace pursuit algorithm of Dai and Milenkovic [20] achieve a scaling similar to (4). Other variants of OMP include the stagewise OMP [21] and regularized OMP [22], [23]. 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 address the "nasty independence issues" noted by Tropp and Gilbert [10] by providing a tighter bound on the maximum correlation of the incorrect vectors. Specifically, in each iteration of the OMP algorithm, there are $n-k$ possible incorrect vectors that the algorithm can choose. Since the algorithm runs for k iterations, there is a 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 of 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 II describes the OMP algorithm. Our main result, Theorem 1, is stated in Section III. A comparison to lasso is provided in Section IV, and we suggest some future problems in Section VII. The proof of the main result is somewhat long and given in Section VIII. The main result was first reported in the conference version [24]. This paper provides full proofs and more discussion on threshold selection and issues of dynamic range.

II. Orthogonal Matching Pursuit With Threshold Termination

Let $\mathbf{x} \in \mathbb{R}^n$ be a sparse vector and let

$$I_{\text{true}} = \{j : x_j \neq 0\} \tag{5}$$

be its support. The set I_{true} will also be called the *sparsity pattern*, and we let $k = |I_{\text{true}}|$, which is the number of nonzero entries of $\mathbf{x}$. The problem is to detect the sparsity pattern of $\mathbf{x}$ from a measurement vector $\mathbf{y} \in \mathbb{R}^m$ of the form (1). To this end, we analyze the following algorithm, which is a variant of the classical OMP method [7]–[9].

Algorithm 1 (OMP): Given a vector $\mathbf{y} \in \mathbb{R}^m$, a measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, and threshold level $\mu > 0$, compute an estimate $\hat{I}_{\text{OMP}}$ of the sparsity pattern of $\mathbf{x}$ as follows.

1) Initialize $t = 0$ and $\hat{I}(t) = \emptyset$.
2) Compute $\mathbf{P}(t)$, the projection operator onto the orthogonal complement of the span of $\{\mathbf{a}_i, i \in \hat{I}(t)\}$.
3) For each j, compute

$$\rho(t,j) = \frac{\left|\mathbf{a}_j'\mathbf{P}(t)\mathbf{y}\right|^2}{\|\mathbf{P}(t)\mathbf{y}\|^2}$$

and let

$$[\rho^*(t), i^*(t)] = \max_{j=1,\ldots,n} \rho(t,j) \tag{6}$$

where $\rho^*(t)$ is the value of the maximum and $i^*(t)$ is an index that achieves the maximum. Here, $\mathbf{a}_j$ is the jth column of $\mathbf{A}$.
4) If $\rho^*(t) > \mu$, set $\hat{I}(t+1) = \hat{I}(t) \cup \{i^*(t)\}$. Also, increment $t = t+1$ and return to step 2.
5) Otherwise stop. The final estimate of the sparsity pattern is $\hat{I}_{\text{OMP}} = \hat{I}(t)$.

The algorithm produces a sequence of estimates $\hat{I}(t)$, $t = 0, 1, 2, \ldots$, of the sparsity pattern I_{true}, adding one index in each iteration. As in standard OMP, in each iteration t, the measured vector $\mathbf{y}$ is projected onto the orthogonal complement of the span of the vectors in the current support set $\hat{I}(t)$. The projection, denoted $\mathbf{P}(t)\mathbf{y}$ represents the residual of $\mathbf{y}$ not in the subspace spanned by the current set of detected vectors. The algorithm then identifies an index $i^*(t)$ corresponding to the column

[1]Sufficient conditions under weaker conditions on the SNR are more subtle [17]: 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]Recall that our result is a sufficient condition for success whereas the matching condition for lasso is both necessary and sufficient.

$\mathbf{a}_j$ of $\mathbf{A}$ with the maximum correlation with the residual. If the number k of nonzero components were known *a priori* the algorithm would be run for exactly k iterations. However, we consider a simple variant of OMP where the algorithm is continued until the maximum correlation falls below a threshold level μ. In this way, the algorithm provides an estimate for both the number k of nonzero components and their indices.

Note that since $\mathbf{P}(t)$ is the projection onto the orthogonal complement of the span of $\{\mathbf{a}_j, j \in \hat{I}(t)\}$, for all $j \in \hat{I}(t)$ we have $\mathbf{P}(t)\mathbf{a}_j = 0$. Hence, $\rho(t,j) = 0$ for all $j \in \hat{I}(t)$, and therefore the algorithm will not select the same vector twice.

It should be noted that the algorithm above only provides an estimate, $\hat{I}_{\rm OMP}$, of the sparsity pattern of $I_{\rm true}$. Using $\hat{I}_{\rm OMP}$, one can estimate the vector $\mathbf{x}$ in a number of ways. For example, one can take the least-squares estimate

$$\widehat{\mathbf{x}} = \arg\min \|\mathbf{y} - \mathbf{A}\mathbf{v}\|^2 \tag{7}$$

where the minimization is over all vectors $\mathbf{v}$ such $v_j = 0$ for all $j \notin \hat{I}_{\rm OMP}$. The estimate $\widehat{\mathbf{x}}$ is the projection of the noisy vector $\mathbf{y}$ onto the space spanned by the vectors $\mathbf{a}_i$ with i in the sparsity pattern estimate $\hat{I}_{\rm OMP}$. This paper only analyzes the sparsity pattern estimate $\hat{I}_{\rm OMP}$ itself, and not the vector estimate $\widehat{\mathbf{x}}$.

III. Asymptotic Analysis

We analyze the OMP algorithm with threshold-based termination from 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 $\mathbf{x} \in \mathbb{R}^n$ be a deterministic vector. Also assume:

(a) The sparsity level $k = k(n)$ (i.e., number of nonzero entries in $\mathbf{x}$) satisfies

$$k(n) \in [k_{\min}(n), k_{\max}(n)] \tag{8}$$

for some deterministic sequences $k_{\min}(n)$ and $k_{\max}(n)$ with $k_{\min}(n) \to \infty$ as $n \to \infty$ and $k_{\max}(n) < (n/2)$ for all n.

(b) The number of measurements $m = m(n)$ is a deterministic sequence satisfying

$$m \geq (1+\delta) 2k_{\max} \log(n - k_{\min}) \tag{9}$$

for some $\delta > 0$.

(c) The minimum component power $x_{\min}^2$ satisfies

$$\lim_{n\to\infty} k x_{\min}^2 = \infty \tag{10}$$

where

$$x_{\min} = \min_{j \in I_{\rm true}} |x_j| \tag{11}$$

is the magnitude of the smallest nonzero entry of $\mathbf{x}$.

(d) The powers of the vectors $\|\mathbf{x}\|^2$ satisfy

$$\lim_{n\to\infty} \frac{1}{(n-k)^\epsilon} \log\left(1 + \|\mathbf{x}\|^2\right) = 0 \tag{12}$$

for all $\epsilon > 0$.

(e) The vector $\mathbf{y}$ is a random vector generated by (1) where $\mathbf{A}$ and $\mathbf{w}$ have i.i.d. Gaussian entries with zero mean and variance $1/m$.

Assumption 1(a) provides a range on the sparsity level k. As we will see below in Section V, bounds on this range are necessary for proper selection of the threshold level $\mu > 0$.

Assumption 1(b) is the 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 $k_{\max} = k_{\min} = k$, we obtain the simpler scaling law

$$m \geq (1+\delta) 2k \log(n-k). \tag{13}$$

We have contrasted this scaling law with the Tropp-Gilbert scaling law (2) in Section I. We will also compare it to the scaling law for lasso in Section IV.

Observe that for some values of k and n, the number of measurements m in (13) may exceed the signal dimension n. In such cases, no "compression" in the sense of compressed sensing [4] would be observed. This is not necessarily a shortcoming of the analysis, but rather reflects that OMP may not be optimal and may require a tall sensing matrix $\mathbf{A}$.

Assumption 1(c) is critical and places constraints on the smallest component magnitude. If there were no noise, this condition could be ignored since one can generally "re-scale" the problem by multiplying $\mathbf{x}$ and $\mathbf{y}$ with a sufficiently large number so that the condition is always satisfied. However, in the presence of noise, the condition is required to insure that the components are sufficiently above the noise level.

In fact, as discussed in [25], Assumptions 1(c) and (e) together imply that the signal-to-noise ratio (SNR) goes to infinity. Specifically, if we define the SNR as

$$\mathsf{SNR} = \frac{\mathbf{E}\|\mathbf{A}\mathbf{x}\|^2}{\mathbf{E}\|\mathbf{w}\|^2}$$

then under Assumption 1(e) it can be easily checked that

$$\mathsf{SNR} = \|\mathbf{x}\|^2. \tag{14}$$

Since $\mathbf{x}$ has k nonzero entries, $\|\mathbf{x}\|^2 \geq k x_{\min}^2$, and therefore condition (10) requires that $\mathrm{SNR} \to \infty$. For this reason, we will call our analysis of OMP a high-SNR analysis. The analysis of OMP with SNR that remains bounded from above is an interesting open problem.

The reason we consider this infinite SNR limit, instead of a noise-free analysis, is to properly compare against the results for lasso in [16], [17], [26] with similar SNR scalings.

Assumption 1(d) is technical and simply requires that the SNR does not grow too quickly with n. Note that even if $\mathrm{SNR} = O(k^\alpha)$ for any $\alpha > 0$, Assumption 1(d) will be satisfied.

Assumption 1(e) states that our analysis concerns large Gaussian measurement matrices $\mathbf{A}$ and Gaussian noise $\mathbf{w}$.

Our main result is as follows.

Theorem 1: Under Assumption 1, there exists a sequence of threshold levels $\mu = \mu(n)$ such that the OMP method in Algorithm 1 will asymptotically detect the correct sparsity pattern in that

$$\lim_{n\to\infty} \Pr(\hat{I}_{\rm OMP} \neq I_{\rm true}) = 0. \tag{15}$$

Moreover, the threshold levels μ can be selected simply as a function of $k_{\min}$, $k_{\max}$, n, m, and δ.

Theorem 1 provides our main scaling law for OMP. The proof is given in Section VIII.

Before discussing the result, it is important to point out two slight technical limitations in the result: First, unlike the Tropp and Gilbert result [10], our result is only asymptotic. In contrast, [10] provides an explicit bound on the rate of convergence of the limit (15). Deriving a similar bound for this result would require a more refined analysis and is a possible avenue for future work.

Secondly, the probability of success in (15) is for a *given* sequence of vectors $\mathbf{x}$. The result is therefore somewhat weaker than compressed sensing results based on deterministic conditions such as in [12] or [14]. Those results guarantee OMP's success if the sensing matrix Φ satisfies certain conditions on its mutual incoherence or restricted isometry constant. As a result, a probability bound that a randomly-generated sensing matrix would satisfy such conditions would provide a probability that OMP will succeed for *all* vectors $\mathbf{x}$ in a class—a stronger result than Theorem 1.

IV. Comparison to Lasso Performance

It is useful to compare the scaling law (13) to the number of measurements required by the widely used lasso method described for example in [15]. The lasso method finds an estimate for the vector $\mathbf{x}$ in (1) by solving the quadratic program

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{v}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{v}\|^2 + \mu\|\mathbf{v}\|_1 \tag{16}$$

where $\mu > 0$ 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 [27]. While the optimization (16) is convex, the running time of lasso is significantly longer than OMP unless $\mathbf{A}$ 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 [16], [17]. 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.

V. 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 the estimate 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, $\rho^*(t) = 0$ 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 whether $\rho^*(t) > \mu$ for some threshold μ.

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 $k \in [k_{\min}, k_{\max}]$ under which there exists a threshold such that the OMP algorithm will terminate in the correct number of iterations. The larger the number of measurements m, the wider one can make the range $[k_{\min}, k_{\max}]$.

The detailed formula for the threshold level selection is given in the proof of Theorem 1, specifically Section VIII-B. Ignoring the infinitesimal constant δ, the proof shows that any threshold level μ in the interval

$$\mu \in \left(\frac{2}{m}\log(n - k_{\min}), \frac{1}{k_{\max}}\right) \tag{17}$$

will guarantee asymptotic reliable recovery. The assumption (9) on the number of measurements m in the hypothesis of the theorem guarantees that the interval in (17) will be nonempty.

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 may miss some small entries, but it may result in an overall better estimate by not introducing too many erroneous entries or entries 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.

VI. Numerical Simulations

To verify the above analysis, we simulated the OMP algorithm with fixed signal dimension $n = 100$ and different sparsity levels k, numbers of measurements m, and randomly generated vectors $\mathbf{x}$.

In the first experiment, $\mathbf{x} \in \mathbb{R}^n$ was generated with k randomly placed nonzero values, with all the nonzero entries having the same magnitude $|x_j| = C$ for some $C > 0$. Following Assumption 1(e), the measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ and noise vector $\mathbf{w} \in \mathbb{R}^m$ were generated with i.i.d. $\mathcal{N}(0, 1/m)$ entries. Using (14) and the fact that $\mathbf{x}$ has k nonzero entries with power C^2, the SNR is given by

$$\mathsf{SNR} = \|\mathbf{x}\|^2 = kC^2$$

so the SNR can be controlled by varying C.

Fig. 1 plots the probability that the OMP algorithm incorrectly detected the sparsity pattern for different values of k and m. The probability is estimated with 1000 Monte Carlo simulations per (k, m) pair. The threshold level μ was selected as the midpoint of the interval (17) with $k = k_{\min} = k_{\max}$. Recall that the interval is nonempty when (9) is satisfied for some $\delta > 0$. If the interval is empty (i.e., the hypotheses of the theorem are not satisfied), we set μ to the lower point of the interval $\mu = (2/m)\log(n - k)$.

The solid curve in Fig. 1 is the theoretical number of measurements in (13) from Theorem 1 that guarantees exact sparsity recovery. The formula is theoretically valid as $n \to \infty$ and $\mathrm{SNR} \to \infty$. At bounded problem sizes with finite SNR, the probability of error for m satisfying (13) will, in general, be nonzero. However, Fig. 1 shows that, at least with no noise (i.e., $\mathrm{SNR} = \infty$) and a moderate problem size of $n = 100$, the probability of error for OMP is indeed low for values of m greater than the theoretical level. Specifically, when $\mathrm{SNR} = \infty$, the probability of error is between 3 and 7% for most values of k.

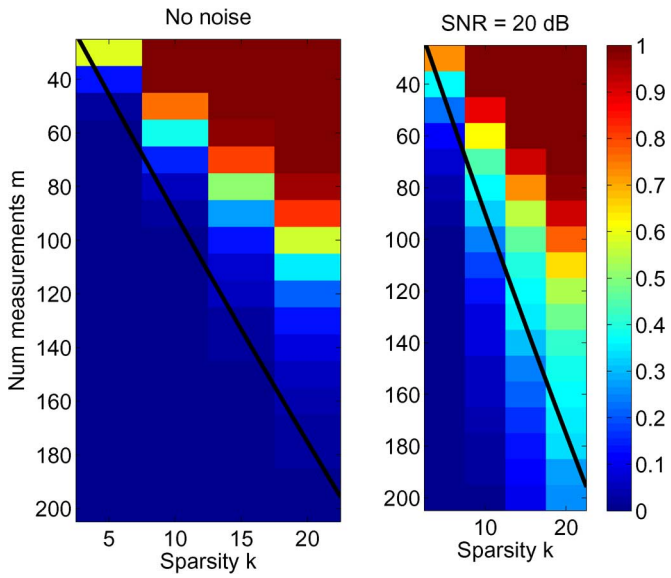

Fig. 1. OMP performance prediction. The colored bars show the probability of sparsity pattern misdetection based on 1000 Monte Carlo simulations of the OMP algorithm. The signal dimension is fixed to $n = 100$ and the error probability is plotted against the number of measurements m and sparsity level k. The solid black curve shows the theoretical number of measurements $m = 2k\log(n-k)$ sufficient for asymptotic reliable detection.

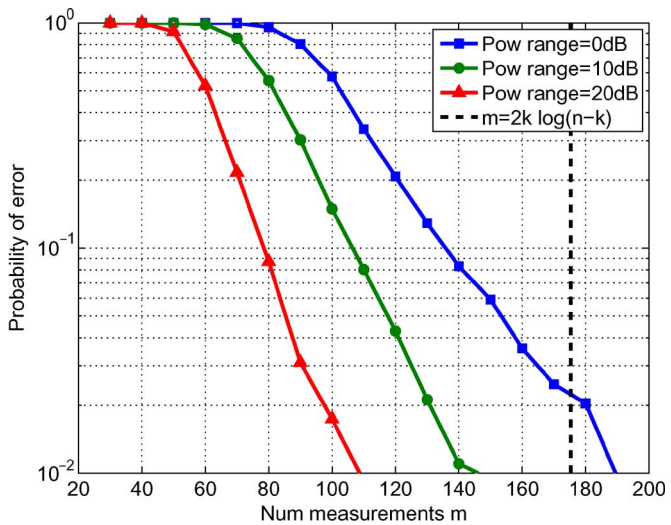

Fig. 2. OMP performance and dynamic range. Plotted is the probability of sparsity pattern detection as a function of the number of measurements for random vectors $\mathbf{x}$ with various dynamic ranges. In all cases, $n = 100$, $k = 20$ and $\mathrm{SNR} = \infty$.

When the SNR is 20 dB, the probability of error is higher—between 28 and 35%. Thus, even at finite problem sizes and high SNRs, the formula may provide a reasonable sufficient condition on the number of measurements for OMP to succeed.

Theorem 1 is only a *sufficient condition*. It is possible that for some $\mathbf{x}$, OMP could require a number of measurements less than predicted by (13). That is, the number of measurements (13) may not be *necessary*.

To illustrate such a case, we consider vectors with a nonzero dynamic range of component magnitudes. Fig. 2 shows the probability of sparsity pattern detection as a function of m for vectors $\mathbf{x}$ with different dynamic ranges. Specifically, the k nonzero entries of $\mathbf{x}$ were chosen to have powers uniformly distributed in a range of 0, 10, and 20 dB. In this simulation, we used $k = 20$ and $n = 100$, so the sufficient condition predicted by (13) is $m \approx 175$. When the dynamic range is 0 dB, all the nonzero entries have equal magnitude, and the probability of error at the value $m = 175$ is approximately 2%. However, with a dynamic range of 10 dB, the same probability of error can be achieved with $m \approx 135$ measurements, a value significantly below the sufficient condition in (13). With a dynamic range of 20 dB, the number of measurements decreases further to $m \approx 95$.

This possible benefit of dynamic range in OMP-like algorithms has been observed in [28]–[30] and in sparse Bayesian learning [31], [32]. A valuable line of future research would be to see if this benefit can be quantified. That is, it would be useful to develop a sufficient condition tighter than (13) that accounts for the dynamic range of the signals.

VII. Conclusions and Future Work

We have provided an improved scaling law on the number of measurements for asymptotic reliable sparsity pattern detection with OMP. Most importantly, the 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 element in the sparsity pattern. It would be useful if partial support recovery results such as those in [33]–[35] can be obtained for OMP.

VIII. Proof of Theorem 1

A. Proof Outline

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 alternate "genie" algorithm. A similar alternate algorithm is analyzed in [28] as well.

1) Initialize $t = 0$ and $I_{\mathrm{true}}(t) = \emptyset$.
2) Compute $\mathbf{P}_{\mathrm{true}}(t)$, the projection operator onto the orthogonal complement of the span of $\{\mathbf{a}_i, i \in I_{\mathrm{true}}(t)\}$.
3) For all $j = 1, \ldots, n$, compute

$$\rho_{\mathrm{true}}(t,j) = \frac{\left|\mathbf{a}_j' \mathbf{P}_{\mathrm{true}}(t)\mathbf{y}\right|^2}{\left\|\mathbf{P}_{\mathrm{true}}(t)\mathbf{y}\right\|^2} \tag{18}$$

and let

$$[\rho^*_{\mathrm{true}}(t), i^*(t)] = \max_{j \in I_{\mathrm{true}}} \rho_{\mathrm{true}}(t,j). \tag{19}$$

4) If $t < k$, set $I_{\mathrm{true}}(t+1) = I_{\mathrm{true}}(t) \cup \{i^*(t)\}$. Increment $t = t+1$ and return to step 2.
5) Otherwise stop. The final estimate of the sparsity pattern is $I_{\mathrm{true}}(k)$.

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 μ for the stopping condition. Also, in the maximization in (19), the genie algorithm searches over only the correct indices $j \in I_{\rm true}$. Hence, this genie algorithm can never select an incorrect index $j \notin I_{\rm true}$. Also, as in the regular OMP algorithm, the genie algorithm will never select the same vector twice for almost all vectors $\mathbf{y}$. Therefore, after k iterations, the genie algorithm will have selected all the k indices in $I_{\rm true}$ and terminate with correct sparsity pattern estimate

$$I_{\rm true}(k) = I_{\rm true}$$

with probability one.

The reason to consider the sequences $\mathbf{P}_{\rm true}(t)$ and $I_{\rm true}(t)$ instead of $\mathbf{P}(t)$ and $\hat{I}(t)$ is that the quantities $\mathbf{P}_{\rm true}(t)$ and $I_{\rm true}(t)$ depend only on the vector $\mathbf{y}$ and the columns $\mathbf{a}_j$ for $j \in I_{\rm true}$. The vector $\mathbf{y}$ also only depends on $\mathbf{a}_j$ for $j \in I_{\rm true}$ and the noise vector $\mathbf{w}$. Hence, $\mathbf{P}_{\rm true}(t)$ and $I_{\rm true}(t)$ are statistically independent of all the columns $\mathbf{a}_j$, $j \notin I_{\rm true}$. This property will be essential in bounding the "false alarm" probability to be defined shortly.

Now suppose the following two conditions hold:

$$\min_{t=0,\ldots,k-1} \max_{j \in I_{\rm true}} \rho_{\rm true}(t,j) > \mu \tag{20a}$$

$$\max_{t=0,\ldots,k} \max_{j \notin I_{\rm true}} \rho_{\rm true}(t,j) < \mu. \tag{20b}$$

The first condition (20a) would imply that, in each of the first t iterations, $t = 0, \ldots, k-1$, there is at least *one* correct index $j \in I_{\rm true}$ where the correlation $\rho_{\rm true}(t,j)$ exceeds the threshold μ. The second condition (20b) would imply that for all iterations $t = 0, \ldots, k$, and all incorrect indices $j \notin I_{\rm true}$, the correlation $\rho_{\rm true}(t,j)$ is below the threshold μ. If both (20a) and (20b) are met, a simple induction argument shows that the regular OMP algorithm, Algorithm 1, will follow the "genie" algorithm. That is, the regular OMP algorithm will terminate in k iterations and the OMP algorithm will output $\mathbf{P}(t) = \mathbf{P}_{\rm true}(t)$, $\hat{I}(t) = I_{\rm true}(t)$, and $\rho(t,j) = \rho_{\rm true}(t,j)$ for all t and j. This will in turn result in the OMP algorithm detecting the correct sparsity pattern

$$\hat{I}_{\rm OMP} = I_{\rm true}.$$

So, it suffices to show that the two events in (20a) and (20b) occur with high probability.

To this end, define the following two probabilities:

$$p_{\rm MD} = \Pr\left(\min_{t=0,\ldots k-1} \max_{j \in I_{\rm true}} \rho_{\rm true}(t,j) \le \mu \right) \tag{21}$$

$$p_{\rm FA} = \Pr\left(\max_{t=0,\ldots k} \max_{j \notin I_{\rm true}} \rho_{\rm true}(t,j) \ge \mu \right). \tag{22}$$

Both probabilities are implicitly functions of n. The first term $p_{\rm MD}$ can be interpreted as a "missed detection" probability, since it corresponds to the event that the maximum correlation energy $\rho_{\rm true}(t,j)$ on the correct vectors $j \in I_{\rm true}$ falls below the threshold. We call the second term $p_{\rm FA}$ the "false alarm" probability since it corresponds to the maximum energy on one of the "incorrect" indices $j \notin I_{\rm true}$ exceeding the threshold.

The above arguments show that

$$\Pr(\hat{I}_{\rm OMP} \ne I_{\rm true}) \le p_{\rm MD} + p_{\rm FA}.$$

So we need to show that there exists a sequence of thresholds $\mu = \mu(n) > 0$, such that $p_{\rm MD} \to 0$ and $p_{\rm FA} \to 0$ as $n \to \infty$. We will define the threshold level in Section VIII-B. Sections VIII-C and VIII-D then prove that $p_{\rm MD} \to 0$ with this threshold. The difficult part of the proof is to show $p_{\rm FA} \to 0$. This part is proven in Section VIII-G after some preliminary results in Sections VIII-E and VIII-F.

B. Threshold Selection

Select an $\epsilon > 0$ and threshold sequence $\mu(n)$ satisfying

$$\mu(n) \in \left(\frac{2(1+\epsilon)}{m} \log(n - k_{\rm min}), \frac{1}{(1+\epsilon)k_{\rm max}} \right). \tag{23}$$

To make such a selection, we must find an $\epsilon > 0$ such that the interval on the right hand side of (23) is nonempty for all n. To this end, given $\delta > 0$ in (9), let $\epsilon > 0$ such that

$$\frac{1+\delta}{1+\epsilon} \ge 1 + \epsilon. \tag{24}$$

Then, (9) and (24) show that

$$\frac{1}{(1+\epsilon)k_{\rm min}} > \frac{2(1+\epsilon)}{m} \log(n - k_{\rm min})$$

so the interval in (23) is indeed nonempty. Also, observe that since $k \ge k_{\rm min}$, (23) implies that

$$\mu \ge \frac{2(1+\epsilon)}{m} \log(n-k). \tag{25}$$

Similarly, since $k \le k_{\rm max}$, (23) shows that

$$\mu \le \frac{1}{(1+\epsilon)k}. \tag{26}$$

C. Decomposition Representation and Related Bounds

To bound the missed detection probability, it is easiest to analyze the OMP algorithm in two separate subspaces: the span of the vectors $\{\mathbf{a}_j, j \in I_{\rm true}\}$, and its orthogonal complement. This subsection defines some notation for this orthogonal decomposition and proves some simple bounds. The actual limit of the missed detection probability will then be evaluated in the next subsection, Section VIII-D.

Assume without loss of generality $I_{\rm true} = \{1, 2, \ldots, k\}$, so that the vector $\mathbf{x}$ is supported on the first k elements. Let Φ be the $m \times k$ matrix formed by the k correct columns:

$$\Phi = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k].$$

Also, let $\mathbf{x}_{\rm true} = [x_1, x_2, \ldots, x_k]'$ be the vector of the k nonzero entries so that

$$\mathbf{A}\mathbf{x} = \Phi \mathbf{x}_{\rm true}. \tag{27}$$

Now rewrite the noise vector $\mathbf{w}$ as

$$\mathbf{w} = \Phi \mathbf{v} + \mathbf{w}^{\perp} \tag{28}$$

where

$$\mathbf{v} = (\Phi'\Phi)^{-1}\Phi'\mathbf{w}, \quad \mathbf{w}^{\perp} = \mathbf{w} - \Phi\mathbf{v}. \tag{29}$$

The vectors $\Phi\mathbf{v}$ and $\mathbf{w}^{\perp}$ are, respectively, the projections of the noise vector $\mathbf{w}$ onto the k-dimensional range space of Φ and its orthogonal complement. Combining (27) with (28), we can rewrite (1) as

$$\mathbf{y} = \Phi\mathbf{z} + \mathbf{w}^{\perp} \tag{30}$$

where

$$\mathbf{z} = \mathbf{x}_{\text{true}} + \mathbf{v}. \tag{31}$$

We begin by computing the limit of the norms of the measurement vector $\mathbf{y}$ and the projected noise vector $\mathbf{w}^{\perp}$.

Lemma 1: The limits

$$\lim_{n\to\infty} \frac{\|\mathbf{y}\|^2}{1+\|\mathbf{x}\|^2} = 1$$
$$\lim_{n\to\infty} \|\mathbf{w}^{\perp}\|^2 = 1$$

hold almost surely and in probability.

Proof: The vector $\mathbf{w}$ is Gaussian, zero mean and white with variance $1/m$ per entry. Therefore, its projection, $\mathbf{w}^{\perp}$, will also be white in the $(m-k)$-dimensional orthogonal complement of the range of Φ with variance $1/m$ per dimension. Therefore, by the strong law of large numbers

$$\lim_{n\to\infty} \|\mathbf{w}^{\perp}\|^2 = \lim_{n\to\infty} \frac{m-k}{m} = 1$$

where the last step follows from the fact that (9) implies that $(k/m) \to 0$.

Similarly, it is easily verified that since $\mathbf{A}$ and $\mathbf{w}$ have i.i.d. Gaussian entries with variance $1/m$, the vector $\mathbf{y}$ is also i.i.d. Gaussian with per-entry variance $(\|\mathbf{x}\|^2+1)/m$. Again, the strong law of large numbers shows that

$$\lim_{n\to\infty} \frac{\|\mathbf{y}\|^2}{1+\|\mathbf{x}\|^2} = 1.$$

■

We next need to compute the minimum singular value of Φ.

Lemma 2: Let $\sigma_{\min}(\Phi)$ and $\sigma_{\max}(\Phi)$ be the minimum and maximum singular values of Φ, respectively. Then

$$\lim_{n\to\infty} \sigma_{\min}(\Phi) = \lim_{n\to\infty} \sigma_{\max}(\Phi) = 1$$

where the limits are in probability.

Proof: Since the matrix Φ has $\mathcal{N}(0, 1/m)$ i.i.d. entries, the Marčenko-Pastur theorem [36] states that

$$\lim_{n\to\infty} \sigma_{\min}(\Phi) = \lim_{n\to\infty} 1 - \sqrt{k/m}$$
$$\lim_{n\to\infty} \sigma_{\max}(\Phi) = \lim_{n\to\infty} 1 + \sqrt{k/m}$$

where the limits are in probability. The result now follows from (9) which implies that $(k/m) \to 0$ as $n \to \infty$. ■

We can also bound the singular values of submatrices of Φ. Given a subset $I \subseteq \{1, 2, \ldots, k\}$, let Φ_I be the submatrix of Φ formed by the columns $\mathbf{a}_i$ for $i \in I$. Also, let $\mathbf{P}_I$ be the projection onto the orthogonal complement of the span of the set $\{\mathbf{a}_i, i \in I\}$. We have the following bound.

Lemma 3: Let I and J be any two disjoint subsets of indices such that

$$I \cup J = \{1, 2, \ldots, k\}.$$

Then

$$\sigma_{\min}\left(\Phi_J'\mathbf{P}_I\Phi_J\right) \geq \sigma_{\min}^2(\Phi).$$

Proof: The matrix $\mathbf{S} = [\Phi_I\ \Phi_J]$ is identical to Φ except that the columns may be permuted. In particular $\sigma_{\min}(\mathbf{S}) = \sigma_{\min}(\Phi)$. Therefore

$$\begin{aligned}\mathbf{S}'\mathbf{S} &= \begin{bmatrix}\Phi_I'\Phi_I & \Phi_I'\Phi_J \\ \Phi_J'\Phi_I & \Phi_J'\Phi_J\end{bmatrix} \\ &\geq \sigma_{\min}^2(\mathbf{S})I \\ &= \sigma_{\min}^2(\Phi)I \\ &\geq \begin{bmatrix}0 & 0 \\ 0 & \sigma_{\min}^2(\Phi)I\end{bmatrix}.\end{aligned}$$

The Schur complement (see, for example, [37]) now shows that

$$\Phi_J'\Phi_J - \sigma_{\min}^2(\Phi)I \geq \Phi_J'\Phi_I\left(\Phi_I'\Phi_I\right)^{-1}\Phi_I'\Phi_J$$

or equivalently

$$\Phi_J'\left(I - \Phi_I\left(\Phi_I'\Phi_I\right)^{-1}\Phi_I'\right)\Phi_J \geq \sigma_{\min}^2(\Phi)I.$$

The result now follows from the fact that

$$\mathbf{P}_I = I - \Phi_I\left(\Phi_I'\Phi_I\right)^{-1}\Phi_I'.$$ ■

We also need the following tail bound on chi-squared random variables.

Lemma 4: Suppose X_i, $i = 1, 2, \ldots$, is a sequence of real-valued, scalar Gaussian random variables with $X_i \sim \mathcal{N}(0, 1)$. The variables need not be independent. Let M_k be the maximum

$$M_k = \max_{i=1,\ldots,k} |X_i|^2.$$

Then

$$\limsup_{k\to\infty} \frac{M_k}{2\log(k)} \leq 1$$

where the limit is in probability.

Proof: See for example [28]. ■

This bound permits us to bound the minimum component of $\mathbf{z}$.

Lemma 5: Let $z_{\min}$ be the minimum component value

$$z_{\min} = \min_{j=1,\ldots,k} |z_j|. \tag{32}$$

Then

$$\liminf_{n\to\infty} \frac{z_{\min}}{x_{\min}} \geq 1$$

where the limit is in probability and $x_{\min}$ is defined in (11).

Proof: Since $\mathbf{w}$ is zero mean and Gaussian, so is $\mathbf{v}$ as defined in (29). Also, the covariance of $\mathbf{v}$ is bounded above by

$$\begin{aligned}\mathbf{E}[\mathbf{v}\mathbf{v}'] &\stackrel{(a)}{=} (\Phi'\Phi)^{-1}\Phi'\left(\mathbf{E}[\mathbf{w}\mathbf{w}']\right)\Phi'(\Phi'\Phi)^{-1} \\ &\stackrel{(b)}{=} \frac{1}{m}(\Phi'\Phi)^{-1} \\ &\stackrel{(c)}{\leq} \frac{1}{m}\sigma_{\min}^{-2}(\Phi)I_n\end{aligned}$$

where I_n is the $n \times n$ identity matrix and (a) follows from the definition of $\mathbf{v}$ in (29); (b) follows from the assumption that $\mathbf{E}[\mathbf{w}\mathbf{w}'] = (1/m)\, I_m$; and (c) is a basic property of singular values. This implies that for every $i \in \{1, 2, \ldots, k\}$

$$\mathbf{E}|v_i|^2 \leq \frac{1}{m}\sigma_{\min}^{-2}(\Phi).$$

Applying Lemma 4 shows that

$$\limsup_{k\to\infty} \frac{m v_{\max}^2 \sigma_{\min}^2(\Phi)}{2\log(k)} \leq 1 \tag{33}$$

where

$$v_{\max} = \max_{i=1,\ldots,k} |v_i|.$$

Therefore

$$\begin{aligned}&\lim_{n\to\infty} \frac{v_{\max}^2}{x_{\min}^2} \\ &= \lim_{n\to\infty} \left(\frac{m v_{\max}^2}{2\log(k)}\right)\left(\frac{2\log(k)}{m x_{\min}^2}\right) \\ &\stackrel{(a)}{\leq} \lim_{n\to\infty} \left(\frac{m v_{\max}^2 \sigma_{\min}^2(\Phi)}{2\log(k)}\right)\left(\frac{2\log(k)}{m x_{\min}^2}\right) \\ &\stackrel{(b)}{\leq} \lim_{n\to\infty} \frac{2\log(k)}{m x_{\min}^2} \\ &\stackrel{(c)}{\leq} \lim_{n\to\infty} \frac{2\log(n-k)}{m x_{\min}^2} \\ &\stackrel{(d)}{\leq} \lim_{n\to\infty} \frac{1}{(1+\delta)k x_{\min}^2} \\ &\stackrel{(e)}{=} 0\end{aligned}$$

where all the limits are in probability and (a) follows from Lemma 2; (b) follows from (33); (c) follows from the fact that $k < (n/2)$ and hence $k < n - k$; (d) follows from (9); and (e) follows from (10). Now, for $j \in \{1, 2, \ldots, k\}$

$$|z_j| = |x_j + v_j| \geq |x_j| - |v_j|$$

and therefore

$$z_{\min} \geq x_{\min} - v_{\max}.$$

Hence

$$\frac{z_{\min}}{x_{\min}} \geq 1 - \frac{v_{\max}}{x_{\min}} \to 1$$

where again the limit is in probability. ■

D. Probability of Missed Detection

With the bounds in the previous section, we can now show that the probability of missed detection goes to zero. The proof is similar to Tropp and Gilbert's proof in [10] with some modifications to account for the noise.

For any $t \in \{0, 1, \ldots, k\}$, let $J(t) = I_{\text{true}} \cap I_{\text{true}}(t)^c$, which is the set of indices $j \in I_{\text{true}}$ that are *not* yet detected in iteration t of the genie algorithm in Section VIII-A. Then

$$\Phi\mathbf{z} = \Phi_{I_{\text{true}}(t)}\mathbf{z}_{I_{\text{true}}(t)} + \Phi_{J(t)}\mathbf{z}_{J(t)} \tag{34}$$

where (using the notation of the previous subsection), Φ_I denotes the submatrix of Φ formed by the columns with indices $i \in I$, and $\mathbf{z}_I$ denotes the corresponding subvector.

Now since $\mathbf{P}_{\text{true}}(t)$ is the projection onto the orthogonal complement of the span of $\{\mathbf{a}_i, i \in I_{\text{true}}(t)\}$

$$\mathbf{P}_{\text{true}}(t)\Phi_{I_{\text{true}}(t)} = 0. \tag{35}$$

Also, since $\mathbf{w}^\perp$ is orthogonal to $\mathbf{a}_i$ for all $i \in I_{\text{true}}$ and $I_{\text{true}}(t) \subseteq I_{\text{true}}$,

$$\mathbf{P}_{\text{true}}(t)\mathbf{w}^\perp = \mathbf{w}^\perp. \tag{36}$$

Therefore,

$$\begin{aligned}\mathbf{P}_{\text{true}}(t)\mathbf{y} &\stackrel{(a)}{=} \mathbf{P}_{\text{true}}(t)(\Phi\mathbf{z} + \mathbf{w}^\perp) \\ &\stackrel{(b)}{=} \mathbf{P}_{\text{true}}(t)\left(\Phi_{J(t)}\mathbf{z}_{J(t)} + \mathbf{w}^\perp\right) \\ &\stackrel{(c)}{=} \mathbf{P}_{\text{true}}(t)\Phi_{J(t)}\mathbf{z}_{J(t)} + \mathbf{w}^\perp\end{aligned} \tag{37}$$

where (a) follows from (30); (b) follows from (34) and (35); and (c) follows from (36).

Now using (36) and the fact that $\mathbf{w}^\perp$ is orthogonal to $\mathbf{a}_i$ for all $i \in I_{\text{true}}$, we have

$$\mathbf{a}_i'\mathbf{P}_{\text{true}}(t)\mathbf{w}^\perp = \mathbf{a}_i'\mathbf{w}^\perp = 0 \tag{38}$$

for all $i \in I_{\text{true}}$. Since the columns of $\Phi_{J(t)}$ are formed by vectors $\mathbf{a}_i$ with $i \in I_{\text{true}}$

$$\Phi_{J(t)}'\mathbf{P}_{\text{true}}(t)\mathbf{w}^\perp = 0. \tag{39}$$

Combining (39) and (37)

$$\|\mathbf{P}_{\text{true}}(t)\mathbf{y}\|^2 = \left\|\mathbf{P}_{\text{true}}(t)\Phi_{J(t)}\mathbf{z}_{J(t)}\right\|^2 + \|\mathbf{w}^\perp\|^2. \tag{40}$$

Now for all t, we have that

$$\begin{aligned}&\max_{j\in I_{\text{true}}} \rho_{\text{true}}(t, j) \\ &\stackrel{(a)}{=} \frac{1}{\|\mathbf{P}_{\text{true}}(t)\mathbf{y}\|^2} \max_{j\in I_{\text{true}}} \left|\mathbf{a}_j'\mathbf{P}_{\text{true}}(j)\mathbf{y}\right|^2 \\ &\stackrel{(b)}{=} \frac{1}{\|\mathbf{P}_{\text{true}}(t)\mathbf{y}\|^2} \max_{j\in J(t)} \left|\mathbf{a}_j'\mathbf{P}_{\text{true}}(j)\mathbf{y}\right|^2 \\ &\stackrel{(c)}{=} \frac{1}{\|\mathbf{P}_{\text{true}}(t)\mathbf{y}\|^2} \left\|\Phi_{J(t)}'\mathbf{P}_{\text{true}}(j)\mathbf{y}\right\|_\infty^2 \\ &\stackrel{(d)}{\geq} \frac{1}{|J(t)|\,\|\mathbf{P}_{\text{true}}(t)\mathbf{y}\|^2} \left\|\Phi_{J(t)}'\mathbf{P}_{\text{true}}(j)\mathbf{y}\right\|_2^2\end{aligned}$$

$$\begin{aligned}
&\overset{(e)}{=} \frac{\left\|\Phi'_{J(t)}\mathbf{P}_{\text{true}}(j)\Phi_{J(t)}\mathbf{z}_{J(t)}\right\|_2^2}{|J(t)|\,\|\mathbf{P}_{\text{true}}(t)\mathbf{y}\|^2} \\
&\overset{(f)}{=} \frac{\left\|\Phi'_{J(t)}\mathbf{P}_{\text{true}}(j)\Phi_{J(t)}\mathbf{z}_{J(t)}\right\|_2^2}{|J(t)|\left(\left\|\mathbf{P}_{\text{true}}(t)\Phi_{J(t)}\mathbf{z}_{J(t)}\right\|^2+\|\mathbf{w}^{\perp}\|^2\right)} \\
&\overset{(g)}{\geq} \frac{\sigma_{\min}\left(\Phi'_{J(t)}\mathbf{P}_{\text{true}}(j)\Phi_{J(t)}\right)\left\|\mathbf{z}_{J(t)}\right\|_2^2}{|J(t)|\left(\sigma^2_{\max}(\Phi)\left\|\mathbf{z}_{J(t)}\right\|^2+\|\mathbf{w}^{\perp}\|^2\right)} \\
&\overset{(h)}{\geq} \frac{\sigma^4_{\min}(\Phi)\|\mathbf{z}_{J(t)}\|_2^2}{|J(t)|\left(\sigma^2_{\max}(\Phi)\left\|\mathbf{z}_{J(t)}\right\|^2+\|\mathbf{w}^{\perp}\|^2\right)} \\
&\overset{(i)}{\geq} \frac{\sigma^4_{\min}(\Phi)z^2_{\min}}{\sigma^2_{\max}(\Phi)kz^2_{\min}+\|\mathbf{w}^{\perp}\|^2} \qquad (41)
\end{aligned}$$

where (a) follows from the definition of $\rho_{\text{true}}(t,j)$ in (18); (b) follows from the fact that $\mathbf{P}_{\text{true}}(t)\mathbf{a}_j = 0$ for all $j \in I_{\text{true}}(t)$ and hence the maximum will occur on the set $j \in I_{\text{true}} \cap I_{\text{true}}(t)^c = J(t)$; (c) follows from the fact that $\Phi_{J(t)}$ is the matrix of the columns $\mathbf{a}_j$ with $j \in J(t)$; (d) follows the bound that $\|\mathbf{v}\|_2^2 \leq d\|\mathbf{v}\|_\infty^2$ for any $\mathbf{v} \in \mathbb{R}^d$; (e) follows (37) and (39); (f) follows from (40); (g) follows from the fact that $\mathbf{P}_{\text{true}}(t)$ is a projection operator and hence

$$\sigma_{\max}\left(\mathbf{P}_{\text{true}}(t)\Phi_{J(t)}\right) \leq \sigma_{\max}\left(\Phi_{J(t)}\right) \leq \sigma_{\max}(\Phi);$$

follows from Lemma 3; and (i) follows from the bound

$$\left\|\mathbf{z}_{J(t)}\right\|^2 \geq |J(t)|\, z^2_{\min}$$

and $|J(t)| \leq k$. Therefore

$$\begin{aligned}
&\liminf_{n\to\infty} \min_{t=0,\ldots,k-1} \max_{j\in I_{\text{true}}} \frac{1}{\mu}\rho_{\text{true}}(t,j) \\
&\overset{(a)}{\geq} \liminf_{n\to\infty} \frac{1}{\mu}\frac{\sigma^4_{\min}(\Phi)z^2_{\min}}{\sigma^2_{\max}(\Phi)kz^2_{\min}+\|\mathbf{w}^{\perp}\|^2} \\
&\overset{(b)}{\geq} \liminf_{n\to\infty} \frac{1}{\mu}\frac{z^2_{\min}}{kz^2_{\min}+1} \\
&\overset{(c)}{\geq} \liminf_{n\to\infty} \frac{1}{\mu}\frac{x^2_{\min}}{kx^2_{\min}+1} \\
&\overset{(d)}{\geq} \liminf_{n\to\infty} \frac{1}{k\mu} \\
&\overset{(e)}{\geq} 1+\epsilon \qquad (42)
\end{aligned}$$

where (a) follows from (41), (b) follows from Lemmas 1 and 2; (c) follows from Lemma 5; (d) follows from the assumption of the theorem that $kx^2_{\min} \to \infty$; and (e) follows from (26). The definition of p_{MD} in (21) now shows that

$$\lim_{n\to\infty} p_{\text{MD}} = 0.$$

E. Bounds on Normalized Brownian Motions

Let $B(t)$ be a standard Brownian motion. Define the *normalized Brownian motion* $S(t)$ as the process

$$S(t) = \frac{1}{\sqrt{t}}B(t), \qquad t > 0. \qquad (43)$$

We call the process normalized since

$$\mathbf{E}\,|S(t)|^2 = \frac{1}{t}\mathbf{E}\,|B(t)|^2 = \frac{t}{t} = 1.$$

We first characterize the autocorrelation of this process.

Lemma 6: If $t > s$, the normalized Brownian motion has autocorrelation

$$\mathbf{E}\,[S(t)S(s)] = \sqrt{s/t}.$$

Proof: Write

$$S(t) = \frac{1}{\sqrt{t}}\left(B(s) + B(t) - B(s)\right).$$

Thus

$$\begin{aligned}
\mathbf{E}\,[S(t)S(s)] &= \frac{1}{\sqrt{st}}\mathbf{E}\,[(B(s) + (B(t) - B(s))\,B(s)] \\
&\overset{(a)}{=} \frac{1}{\sqrt{st}}\mathbf{E}\left[B(s)^2\right] \\
&\overset{(b)}{=} \frac{s}{\sqrt{st}} = \sqrt{\frac{s}{t}}
\end{aligned}$$

an where (a) follows from the orthogonal increments property of Brownian motions; and (b) follows from the fact that $B(s) \sim \mathcal{N}(0,s)$. ∎

We now need the following standard Gaussian tail bound.

Lemma 7: Suppose X is a real-valued, scalar Gaussian random variable, $X \sim \mathcal{N}(0,1)$. Then

$$\Pr\left(X^2 > \mu\right) \leq \frac{1}{\sqrt{\pi}\mu}\exp(-\mu/2).$$

Proof: See for example [38]. ∎

We next provide a simple bound on the maximum of sample paths of $S(t)$.

Lemma 8: For any $0 < a < b$, let

$$S_{\max}(a,b) = \sup_{t\in[a,b]} |S(t)|.$$

Then, for any $\mu > 0$

$$\Pr\left(S^2_{\max}(a,b) > \mu\right) \leq \frac{2b}{a\mu\sqrt{\pi}}\exp\left(-\frac{a\mu}{2b}\right).$$

Proof: Since $S(t)$ and $S(-t)$ are identically distributed

$$\Pr\left(S^2_{\max}(a,b) > \mu\right) \leq 2\Pr\left(\sup_{t\in[a,b]} S(t) > \sqrt{\mu}\right). \qquad (44)$$

So, it will suffice to bound the probability of the single-sided event $\sup S(t) > \sqrt{\mu}$. For $t \geq 0$, define $B_a(t) = B(a+t) - B(a)$. Then, $B_a(t)$ is a standard Brownian motion independent of $B(a)$. Also

$$\begin{aligned}
&\sup_{t\in[a,b]} S(t) > \sqrt{\mu} \\
&\quad\Rightarrow \sup_{t\in[a,b]} \frac{1}{\sqrt{t}}B(t) > \sqrt{\mu} \\
&\quad\Rightarrow \sup_{t\in[a,b]} B(t) > \sqrt{a\mu} \\
&\quad\Rightarrow B(a) + \sup_{t\in[0,b-a]} B_a(t) > \sqrt{a\mu}.
\end{aligned}$$

Now, the reflection principle (see, for example, [39]) states that for any y

$$\Pr\left(\max_{t\in[0,b-a]} B_a(t) > y\right) = 2\Pr\left(\sqrt{b-a}Y > y\right)$$

where Y is a unit-variance, zero-mean Gaussian. Also, $B(a) \sim \mathcal{N}(0,a)$, so if we define $X = (1/\sqrt{a})\,B(a)$, then $X \sim \mathcal{N}(0,1)$. Since $B(a)$ is independent of $B_a(t)$ for all $t \geq 0$, we can write

$$\Pr\left(\sup_{t\in[a,b]} S(t) > \sqrt{\mu}\right) \leq 2\Pr(\sqrt{a}X + \sqrt{b-a}Y > \sqrt{a\mu}) \tag{45}$$

where X and Y are independent zero mean Gaussian random variables with unit variance. Now $\sqrt{a}X + \sqrt{b-a}Y$ has variance

$$\mathbf{E}\left[(\sqrt{a}X + \sqrt{b-a}Y)^2\right] = a + b - a = b.$$

Applying Lemma 7 shows that (45) can be bounded by

$$\Pr\left(\sup_{t\in[a,b]} S(t) > \sqrt{\mu}\right) \leq \frac{b}{a\mu\sqrt{\pi}} \exp\left(-\frac{a\mu}{2b}\right).$$

Substituting this bound in (44) proves the lemma. ∎

Our next lemma improves the bound for large μ.

Lemma 9: There exist constants C_1, C_2, and C_3 such that for any $0 < a < b$ and $\mu > C_3$

$$\Pr\left(S^2_{\max}(a,b) > \mu\right) \leq (C_1 + C_2\log(b/a))\, e^{-\mu/2}.$$

Proof: Fix any integer $n > 0$, and define $t_i = a\,(b/a)^{i/n}$ for $i = 0, 1, \ldots, n$. Observe that t_is partition the interval $[a,b]$ in that

$$a = t_0 < t_1 < \cdots < t_n = b.$$

Also, let $r = b/a$. Then, $t_{i+1}/t_i = (b/a)^{1/n} = r^{1/n}$. Applying Lemma 8 to each interval in the partition,

$$\begin{aligned}\Pr\left(S^2_{\max}(a,b) > \mu\right) &\leq \sum_{i=1}^{n-1} \Pr\left(S^2_{\max}(t_i,t_{i+1}) > \mu\right) \\ &\leq \frac{nr^{1/n}}{\mu\sqrt{\pi}} \exp\left(-\frac{r^{-1/n}\mu}{2}\right). \end{aligned} \tag{46}$$

Now, let $\delta > 0$, and for $\mu > \delta$, let

$$n = \left\lceil -\frac{\log(r)}{\log(1-\delta/\mu)} \right\rceil. \tag{47}$$

Then

$$r^{-1/n} \geq 1 - \delta/\mu \tag{48}$$

and hence

$$\exp\left(-\frac{r^{-1/n}\mu}{2}\right) \leq e^{\delta/2} e^{-\mu/2}. \tag{49}$$

Also, (47) implies that

$$n \leq 1 - \frac{\log(r)}{\log(1-\delta/\mu)} \leq 1 + \frac{\mu}{\delta}\log(r) \tag{50}$$

where we have used the fact that $\log(1-x) < -x$ for $x > 0$. Combining the bounds (48) and (50) yields

$$\frac{nr^{1/n}}{\mu} \leq \left(1 + \frac{\mu}{\delta}\log(r)\right)\frac{1}{\mu-\delta}. \tag{51}$$

Now, pick any $\delta > 0$ and let $C_3 = 2\delta$. Then if $\mu > C_3 = 2\delta$, (51) implies that

$$\frac{nr^{1/n}}{\mu^2} \leq \frac{1}{\delta}\left(1 + 2\log(r)\right). \tag{52}$$

Substituting (49) and (52) into (46) shows that

$$\Pr\left(S_{\max}(a,b) > \mu\right) \leq (C_1 + C_2\log(r))\, e^{-\mu/2}$$

where

$$C_1 = \frac{e^{\delta/2}}{\sqrt{\pi}\delta}, \quad C_2 = \frac{2e^{\delta/2}}{\sqrt{\pi}\delta}.$$

The result now follows from the fact that $r = b/a$. ∎

F. Bounds on Sequences of Projections

We can now apply the results in the previous subsection to bound the norms of sequences of projections. Let $\mathbf{y} \in \mathbb{R}^m$ be any deterministic vector, and let $\mathbf{P}(i)$, $i = 0, 1, \ldots, k$ be a deterministic sequence of orthogonal projection operators on $\mathbb{R}^m$. Assume that the sequence $\mathbf{P}(i)$ is *decreasing* in that $\mathbf{P}(i)\mathbf{P}(j) = \mathbf{P}(i)$ for $j > i$.

Lemma 10: Let $\mathbf{a} \in \mathbb{R}^m$ be a Gaussian random vector with unit variance, and define the random variable

$$M = \max_{i=0,\ldots,k} \frac{|\mathbf{a}'\mathbf{P}(i)\mathbf{y}|^2}{\|\mathbf{P}(i)\mathbf{y}\|^2}.$$

Then there exist constants C_1, C_2, and $C_3 > 0$ (all independent of the problem parameters) such that $\mu > C_3$ implies

$$\Pr(M > \mu) \leq (C_1 + C_2\log(r))\, e^{-\mu/2}$$

where $r = \|\mathbf{P}(1)\mathbf{y}\|^2 / \|\mathbf{P}(n)\mathbf{y}\|^2$.

Proof: Define

$$z_i = \frac{\mathbf{y}'\mathbf{P}(i)\mathbf{a}}{\|\mathbf{P}(i)\mathbf{y}\|}$$

so that

$$M = \max_{i=0,\ldots,k} |z_i|^2.$$

Since each z_i is the inner product of the Gaussian vector $\mathbf{a}$ with a fixed vector, the scalars $\{z_i, i = 0, 1, \ldots, k\}$ are jointly Gaussian. Since $\mathbf{a}$ has mean zero, so do the z_is.

To compute the cross-correlations, suppose that $j \geq i$. Then

$$\begin{aligned}\mathbf{E}\left[z_i z_j\right] &= \frac{1}{\|\mathbf{P}(i)\mathbf{y}\|\,\|\mathbf{P}(j)\mathbf{y}\|}\mathbf{E}\left[\mathbf{y}'\mathbf{P}(i)\mathbf{a}\mathbf{a}'\mathbf{P}(j)\mathbf{y}\right] \\ &\overset{(a)}{=} \frac{1}{\|\mathbf{P}(i)\mathbf{y}\|\,\|\mathbf{P}(j)\mathbf{y}\|}\mathbf{y}'\mathbf{P}(i)\mathbf{P}(j)\mathbf{y} \\ &\overset{(b)}{=} \frac{1}{\|\mathbf{P}(i)\mathbf{y}\|\,\|\mathbf{P}(j)\mathbf{y}\|}\mathbf{y}'\mathbf{P}(i)\mathbf{y} \\ &= \frac{\|\mathbf{P}(i)\mathbf{y}\|}{\|\mathbf{P}(j)\mathbf{y}\|}\end{aligned}$$

where (a) uses the fact that $\mathbf{E}[\mathbf{a}\mathbf{a}'] = I_m$; and (b) uses the descending property that $\mathbf{P}(i)\mathbf{P}(j) = \mathbf{P}(i)$. Therefore, if we let $t_i = \|\mathbf{P}(i)\mathbf{y}\|^2$, we have the cross correlations

$$\mathbf{E}[z_i z_j] = \sqrt{t_i/t_j} \tag{53}$$

for all $j \geq i$. Also observe that since the projection operators are decreasing, so are the t_js. That is, for $j \geq i$

$$t_i = \|\mathbf{P}(i)\mathbf{y}\|^2 \overset{(a)}{=} \|\mathbf{P}(i)\mathbf{P}(j)\mathbf{y}\|^2 \overset{(b)}{\leq} \|\mathbf{P}(j)\mathbf{y}\|^2 = t_j$$

where again (a) uses the decreasing property; and (b) uses the fact that $\mathbf{P}(i)$ is a projection operator and norm nonincreasing.

Now let $S(t)$ be the normalized Brownian motion in (43). Lemma 6 and (53) show that the Gaussian vector

$$\mathbf{z} = (z_0, z_1, \ldots, z_k)$$

has the same covariance as the vector of samples of $S(t)$

$$\mathbf{s} = (S(t_0), S(t_1), \ldots, S(t_k))\,.$$

Since they are also both zero-mean and Gaussian, they have the same distribution. Hence, for all μ,

$$\begin{aligned}\Pr(M > \mu) &= \Pr\left(\max_{i=0,\ldots,k} |z_i|^2 > \mu\right) \\ &= \Pr\left(\max_{i=0,\ldots,k} |S(t_i)|^2 > \mu\right) \\ &\leq \Pr\left(\sup_{t\in[t_k,t_0]} |S(t)|^2 > \mu\right)\end{aligned}$$

where the last step follows from the fact that the t_is are decreasing and hence $t_k \geq t_i \geq t_0$ for all $i \in \{0, 1, \ldots, k\}$. The result now follows from Lemma 9. ∎

G. Probability of False Alarm

Recall that all the projection operators $\mathbf{P}_{\text{true}}(t)$ and the vector $\mathbf{y}$ are statistically independent of the vectors $\mathbf{a}_j$ for $j \notin I_{\text{true}}$. Since the entries of the matrix $\mathbf{A}$ are i.i.d. Gaussian with zero mean and variance $1/m$, the vector $m\mathbf{a}_j$ is Gaussian with unit variance. Hence, Lemma 10 shows that there exist constants C_1, C_2, and C_3 such that for any $\lambda > C_3$

$$\Pr\left(\max_{t=0,\ldots,k} m\frac{|\mathbf{a}_j\mathbf{P}_{\text{true}}(t)\mathbf{y}|^2}{\|\mathbf{P}_{\text{true}}(t)\mathbf{y}\|^2} \geq \lambda\right) \leq Be^{-\lambda/2} \tag{54}$$

where $j \notin I_{\text{true}}$ and

$$B = C_1 + C_2 \log\left(\frac{\|\mathbf{P}_{\text{true}}(0)\mathbf{y}\|^2}{\|\mathbf{P}_{\text{true}}(k)\mathbf{y}\|^2}\right). \tag{55}$$

Therefore,

$$\begin{aligned}p_{\text{FA}} &\overset{(a)}{=} \Pr\left(\max_{t=1,\ldots,k}\max_{j\notin I_{\text{true}}} \rho_{\text{true}}(t,j) > \mu\right) \\ &\overset{(b)}{\leq} (n-k)\max_{j\notin I_{\text{true}}} \Pr\left(\max_{t=1,\ldots,k} \rho_{\text{true}}(t,j) > \mu\right) \\ &\overset{(c)}{=} (n-k)\max_{j\notin I_{\text{true}}} \Pr\left(\max_{t=1,\ldots,k} \frac{|\mathbf{a}_j\mathbf{P}_{\text{true}}(t)\mathbf{y}|^2}{\|\mathbf{P}_{\text{true}}(t)\mathbf{y}\|^2} > \mu\right) \\ &\overset{(d)}{\leq} (n-k)Be^{-m\mu/2} \\ &\overset{(e)}{\leq} (n-k)Be^{-(1+\epsilon)\log(n-k)} \\ &= \frac{1}{(n-k)^{\epsilon}}B\end{aligned} \tag{56}$$

where (a) follows from the definition of p_{FA} in (22); (b) uses the union bound and the fact that I^c_{true} has $n-k$ elements; (c) follows from the definition of $\rho_{\text{true}}(t,j)$ in (18); (d) follows from (54) under the condition that $\mu m > C_3$; and (e) follows from (25). By (9) and the hypothesis of the theorem that $n-k \to \infty$,

$$\mu m = (1+\delta)2\log(n-k) \to \infty \text{ as } n \to \infty.$$

Therefore, for sufficiently large n, $\mu m > C_3$ and (56) holds.

Now, since $I_{\text{true}}(0) = \emptyset$, $\mathbf{P}_{\text{true}}(0) = I$ and therefore

$$\mathbf{P}_{\text{true}}(0)\mathbf{y} = \mathbf{y}. \tag{57}$$

Also, $I_{\text{true}}(k) = I_{\text{true}}$ and so $\mathbf{P}_{\text{true}}(k)$ is the projection onto the orthogonal complement of the range of Φ. Hence $\mathbf{P}_{\text{true}}(k)\Phi = 0$. Combining this fact with (30) and (36) shows

$$\mathbf{P}_{\text{true}}(k)\mathbf{y} = \mathbf{w}^{\perp}. \tag{58}$$

Therefore

$$\begin{aligned}&\liminf_{n\to\infty} p_{\text{FA}} \\ &\quad\overset{(a)}{\leq} \liminf_{n\to\infty} \frac{1}{(n-k)^{\epsilon}}B \\ &\quad\overset{(b)}{\leq} \liminf_{n\to\infty} \frac{1}{(n-k)^{\epsilon}}\left(C_1 + C_2\log\left(\frac{\|\mathbf{P}_{\text{true}}(0)\mathbf{y}\|^2}{\|\mathbf{P}_{\text{true}}(k)\mathbf{y}\|^2}\right)\right) \\ &\quad\overset{(c)}{=} \liminf_{n\to\infty} \frac{1}{(n-k)^{\epsilon}}\left(C_1 + C_2\log\left(\frac{\|\mathbf{y}\|^2}{\|\mathbf{w}^{\perp}\|^2}\right)\right) \\ &\quad\overset{(d)}{=} \liminf_{n\to\infty} \frac{1}{(n-k)^{\epsilon}}\left(C_1 + C_2\log\left(1+\|\mathbf{x}\|^2\right)\right) \\ &\quad\overset{(e)}{=} 0\end{aligned}$$

where (a) follows from (56); (b) follows from (55); (c) follows from (57) and (58); (d) follows from Lemma 1; and (e) follows from (12). This completes the proof of the theorem.

ACKNOWLEDGMENT

The authors thank V. K. Goyal for comments on an earlier draft and M. Vetterli for his support, wisdom, and encouragement.

REFERENCES

[1] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1999.
[2] A. Miller, *Subset Selection in Regression*, ser. Monographs on Statistics and Applied Probability, 2nd ed. New York: Chapman and Hall/CRC, 2002, no. 95.
[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[5] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[6] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.
[7] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Contr.*, vol. 50, no. 5, pp. 1873–1896, Nov. 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. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.
[9] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decomposition," *Opt. Eng.*, vol. 33, no. 7, pp. 2183–2191, Jul. 1994.
[10] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[11] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit: The Gaussian case," Calif. Inst. Tech., Appl. Comput. Math. 2007-01, Aug. 2007.
[12] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[13] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
[14] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.
[15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
[16] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy recovery of sparsity," Univ. of California, Berkeley, Dept. of Stat., Berkeley, CA, Tech. Rep., May 2006 [Online]. Available: http://arxiv.org/abs/math/0605740,
[17] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.
[18] D. L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 1–53, Jan. 2009.
[19] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harm. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.
[20] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
[21] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," preprint, Mar. 2006.
[22] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, Jun. 2009.
[23] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.
[24] A. K. Fletcher and S. Rangan, "Orthogonal matching pursuit from noisy measurements: A new analysis," in *Proc. Neural Inf. Process. Syst.*, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, Eds., Vancouver, Canada, Dec. 2009.
[25] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.
[26] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.
[27] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999.
[28] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework," Mar. 2009 [Online]. Available: http://arxiv.org/abs/0903.1022,
[29] A. K. Fletcher, S. Rangan, and V. K. Goyal, "A sparsity detection framework for on-off random access channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, Jun.–Jul. 2009, pp. 169–173.
[30] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Ranked sparse signal support detection," Oct. 2011 [Online]. Available: http://arxiv.org/abs/1110.6188.
[31] D. Wipf and B. Rao, "Comparing the effects of different weight distributions on finding sparse representations," in *Proc. Neural Inf. Process. Syst.*, Vancouver, Canada, Dec. 2006.
[32] Y. Jin and B. Rao, "Performance limits of matching pursuit algorithms," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, Canada, Jun. 2008, pp. 2444–2448.
[33] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.
[34] G. Reeves, "Sparse signal sampling using noisy linear projections," Univ. Calif. Berkeley, Dept. Elect. Eng. Comp. Sci., Tech. Rep. UCB/EECS-2008-3, Jan. 2008.
[35] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5111–5130, Oct. 2010.
[36] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.
[37] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985, reprinted with corrections 1987.
[38] J. Spanier and K. B. Oldham, *An Atlas of Functions*. Washington, DC: Hemisphere, 1987.
[39] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed. New York: Springer-Verlag, 1991.

Alyson K. Fletcher (S'03–M'04) received the B.S. degree in mathematics from the University of Iowa, Iowa City. She received the M.S. degree in electrical engineering in 2002, and the M.A. and Ph.D. degrees in mathematics, and electrical engineering, respectively, both in 2006, from the University of California, Berkeley.

Her research interests include signal processing, information theory, machine learning, and neuroscience.

Dr. Fletcher is a member of SWE, SIAM, and Sigma Xi. In 2005, she received the University of California Eugene L. Lawler Award, the Henry Luce Foundation's Clare Boothe Luce Fellowship, the Soroptimist Dissertation Fellowship, and University of California President's Postdoctoral Fellowship.

Sundeep Rangan (M'02) received the B.A.Sc. degree from the University of Waterloo, Canada, and the M.S. and Ph.D. degrees from the University of California, Berkeley, all in electrical engineering.

He held postdoctoral appointments at the University of Michigan, Ann Arbor, and Bell Labs. In 2000, he co-founded (with four others) Flarion Technologies, a spin-off of Bell Labs, that developed Flash OFDM, one of the first cellular OFDM data systems. In 2006, Flarion was acquired by Qualcomm Technologies, where he was a Director of Engineering involved in OFDM infrastructure products. He joined the Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, in 2010, where he is currently an Associate Professor. His research interests are in wireless communications, signal processing, information theory, and control theory.