

SIGNAL RECOVERY FROM RANDOM MEASUREMENTS VIA ORTHOGONAL MATCHING PURSUIT

JOEL A. TROPP AND ANNA C. GILBERT

ABSTRACT. This technical report demonstrates theoretically and empirically that a greedy algorithm called Orthogonal Matching Pursuit (OMP) can reliably recover a signal with m nonzero entries in dimension d given $O(m \ln d)$ random linear measurements of that signal. This is a massive improvement over previous results for OMP, which require $O(m^2)$ measurements. The new results for OMP are comparable with recent results for another algorithm called Basis Pursuit (BP). The OMP algorithm is faster and easier to implement, which makes it an attractive alternative to BP for signal recovery problems.

1. INTRODUCTION

Let \mathbf{s} be a d -dimensional real signal with at most m nonzero components. This type of signal is called m -sparse. Let $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a sequence of measurement vectors in \mathbb{R}^d that does not depend on the signal. We use these vectors to collect N linear measurements of the signal:

$$\langle \mathbf{s}, \mathbf{x}_1 \rangle, \quad \langle \mathbf{s}, \mathbf{x}_2 \rangle, \quad \dots, \quad \langle \mathbf{s}, \mathbf{x}_N \rangle$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product. The problem of *signal recovery* asks two distinct questions:

- (1) How many measurements are necessary to reconstruct the signal?
- (2) Given these measurements, what algorithms can perform the reconstruction task?

As we will see, signal recovery is dual to sparse approximation, a problem of significant interest [MZ93, CDS01, RKD99, Mil02, Tem02].

To the first question, we can immediately respond that no fewer than m measurements will do. Even if the measurements were adapted to the signal, it would still take m pieces of information to determine all the components of an m -sparse signal. In the other direction, d nonadaptive measurements always suffice because we could simply list the d components of the signal. Although it is not obvious, sparse signals can be reconstructed with far less information.

The method for doing so has its origins during World War II. The US Army had a natural interest in screening soldiers for syphilis. But syphilis tests were expensive, and the Army realized that it was wasteful to perform individual assays to detect an occasional case. Their solution was to pool blood from groups of soldiers and test the pooled blood. If a batch checked positive, further tests could be performed. This method, called *group testing*, was subsequently studied in the computer science and statistics literatures. See [DH93] for a survey.

Very recently, a specific type of group testing has been proposed by the computational harmonic analysis community. The idea is that, by randomly combining the entries of a sparse signal, it is possible to generate a small set of summary statistics that allow us to identify the nonzero

Date: 11 April 2005. Revised September 2006.

2000 Mathematics Subject Classification. 41A46, 68Q25, 68W20, 90C27.

Key words and phrases. Algorithms, approximation, Basis Pursuit, group testing, Orthogonal Matching Pursuit, signal recovery, sparse approximation.

The authors are with the Department of Mathematics, The University of Michigan at Ann Arbor, 2074 East Hall, 530 Church St., Ann Arbor, MI 48109-1043. E-mail: {jtropp|annacg}@umich.edu. ACG has been supported by NSF DMS 0354600.

entries of the signal. The following theorem, drawn from papers of Candès–Tao [CT05] and of Rudelson–Vershynin [RV05], describes one example of this remarkable phenomenon.

Theorem 1. *Let $N \geq Km \ln(d/m)$, and choose N vectors $\mathbf{x}_1, \dots, \mathbf{x}_N$ independently from the standard Gaussian distribution on \mathbb{R}^d . The following statement is true with probability exceeding $(1 - e^{-kN})$. It is possible to reconstruct every m -sparse signal \mathbf{s} in \mathbb{R}^d from the data $\{\langle \mathbf{s}, \mathbf{x}_n \rangle\}$. The numbers K and k are universal constants.*

An important detail is that a particular choice of the Gaussian measurement vectors succeeds for every m -sparse signal with high probability. This theorem extends earlier results of Candès–Romberg–Tao [CRT06], Donoho [Don06a], and Candès–Tao [CT04].

All five of the papers [CRT06, Don06a, CT04, RV05, CT05] offer constructive demonstrations of the recovery phenomenon by proving that the original signal \mathbf{s} is the unique solution to the linear program

$$\min_{\mathbf{f}} \|\mathbf{f}\|_1 \quad \text{subject to} \quad \langle \mathbf{f}, \mathbf{x}_n \rangle = \langle \mathbf{s}, \mathbf{x}_n \rangle \quad \text{for } n = 1, 2, \dots, N. \quad (\text{BP})$$

This optimization problem provides an answer to our second question about how to reconstruct the sparse signal. Note that this formulation requires knowledge of the measurement vectors.

When we talk about (BP), we often say that the linear program can be solved in polynomial time with standard scientific software, and we cite books on convex programming such as [BV04]. This line of talk is misleading because it may take a long time to solve the linear program, even for signals of moderate length.¹ Furthermore, when off-the-shelf optimization software is not available, the implementation of optimization algorithms may demand serious effort.² Both these concerns are receiving attention from researchers. In the meantime, one might wish to consider alternate methods for reconstructing sparse signals from random measurements.

To that end, we adapted a sparse approximation algorithm called Orthogonal Matching Pursuit (OMP) [PRK93, DMA97] to handle the signal recovery problem. The major advantages of this algorithm are its ease of implementation and its speed. On the other hand, conventional wisdom on OMP has been pessimistic about its performance outside the simplest régimes. This complaint dates to a 1996 paper of DeVore and Temlyakov [DT96]. Pursuing their reasoning leads to an example of a nonrandom ensemble of measurement vectors and a sparse signal that OMP cannot identify without d measurements [CDS01, Sec. 2.3.2]. Other negative results, such as Theorem 3.10 of [Tro04] and Theorem 5 of [Don06b], echo this concern.

But these negative results about OMP are very deceptive. Indeed, the empirical evidence suggests that OMP can recover an m -sparse signal when the number of measurements N is a constant multiple of m . The goal of this work is to present a rigorous proof that OMP can perform this feat. This technical report establishes the following theorem in detail.

Theorem 2 (OMP with Gaussian Measurements). *Fix $\delta \in (0, 1)$, and choose $N \geq Km \ln(d/\delta)$. Suppose that \mathbf{s} is an arbitrary m -sparse signal in \mathbb{R}^d , and choose N measurement vectors $\mathbf{x}_1, \dots, \mathbf{x}_N$ independently from the standard Gaussian distribution on \mathbb{R}^d . Given the data $\{\langle \mathbf{s}, \mathbf{x}_n \rangle\}$ and the measurement vectors, Orthogonal Matching Pursuit can reconstruct the signal with probability exceeding $(1 - \delta)$. For this theoretical result, it suffices that $K = 16$. When m is large, it suffices to take $K \approx 4$.*

In comparison, previous positive results, such as Theorem 3.6 from [Tro04], only demonstrate that Orthogonal Matching Pursuit can recover m -sparse signals when the number of measurements N is on the order of m^2 . Theorem 2 improves massively on this earlier work.

Theorem 2 is weaker than Theorem 1 for several reasons. First, our result requires somewhat more measurements than the result for (BP). Second, the quantifiers are ordered differently. Whereas we

¹Although this claim qualifies as folklore, the literature does not currently offer a refutation that we find convincing.

²The paper [GMS05] discusses the software engineering problems that arise in optimization.

prove that OMP can recover any sparse signal given random measurements independent from the signal, the result for (BP) shows that a single set of random measurement vectors can be used to recover all sparse signals. In Section ??, we argue that these formal distinctions may be irrelevant in practice. Indeed, we believe that the large advantages of Orthogonal Matching Pursuit make Theorem 2 extremely compelling.

2. ORTHOGONAL MATCHING PURSUIT FOR SIGNAL RECOVERY

This section describes a greedy algorithm for signal recovery. This method is analogous with Orthogonal Matching Pursuit, an algorithm for sparse approximation. First, let us motivate the computational technique.

Suppose that \mathbf{s} is an arbitrary m -sparse signal in \mathbb{R}^d , and let $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a family of N measurement vectors. Form an $N \times d$ matrix Φ whose rows are the measurement vectors, and observe that the N measurements of the signal can be collected in an N -dimensional data vector $\mathbf{v} = \Phi \mathbf{s}$. We refer to Φ as the *measurement matrix* and denote its columns by $\varphi_1, \dots, \varphi_d$.

As we mentioned, it is natural to think of signal recovery as a problem dual to sparse approximation. Since \mathbf{s} has only m nonzero components, the data vector $\mathbf{v} = \Phi \mathbf{s}$ is a linear combination of m columns from Φ . In the language of sparse approximation, we say that \mathbf{v} has an m -term representation over the dictionary Φ . This perspective allows us to transport results on sparse approximation to the signal recovery problem.

In particular, sparse approximation algorithms can be used for signal recovery. To identify the ideal signal \mathbf{s} , we need to determine *which* columns of Φ participate in the measurement vector \mathbf{v} . The idea behind the algorithm is to pick columns in a greedy fashion. At each iteration, we choose the column of Φ that is most strongly correlated with the remaining part of \mathbf{v} . Then we subtract off its contribution to \mathbf{v} and iterate on the residual. One hopes that, after m iterations, the algorithm will have identified the correct set of columns.

Algorithm 3 (OMP for Signal Recovery).

INPUT:

- An $N \times d$ measurement matrix Φ
- An N -dimensional data vector \mathbf{v}
- The sparsity level m of the ideal signal

OUTPUT:

- An estimate $\hat{\mathbf{s}}$ in \mathbb{R}^d for the ideal signal
- A set Λ_m containing m elements from $\{1, \dots, d\}$
- An N -dimensional approximation \mathbf{a}_m of the data vector \mathbf{v}
- An N -dimensional residual $\mathbf{r}_m = \mathbf{v} - \mathbf{a}_m$

PROCEDURE:

- (1) Initialize the residual $\mathbf{r}_0 = \mathbf{v}$, the index set $\Lambda_0 = \emptyset$, and the iteration counter $t = 1$.
- (2) Find the index λ_t that solves the easy optimization problem

$$\lambda_t = \arg \max_{j=1, \dots, d} |\langle \mathbf{r}_{t-1}, \varphi_j \rangle|.$$

If the maximum occurs for multiple indices, break the tie deterministically.

- (3) Augment the index set $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$ and the matrix of chosen atoms $\Phi_t = [\Phi_{t-1} \quad \varphi_{\lambda_t}]$. We use the convention that Φ_0 is an empty matrix.
- (4) Solve a least-squares problem to obtain a new signal estimate:

$$\mathbf{x}_t = \arg \min_{\mathbf{x}} \|\Phi_t \mathbf{x} - \mathbf{v}\|_2.$$

(5) Calculate the new approximation of the data and the new residual:

$$\begin{aligned}\mathbf{a}_t &= \mathbf{\Phi}_t \mathbf{x}_t \\ \mathbf{r}_t &= \mathbf{v} - \mathbf{a}_t.\end{aligned}$$

(6) Increment t , and return to Step 2 if $t < m$.

(7) The estimate $\hat{\mathbf{s}}$ for the ideal signal has nonzero indices at the components listed in Λ_m . The value of the estimate $\hat{\mathbf{s}}$ in component λ_j equals the j th component of \mathbf{x}_t .

Steps 4, 5, and 7 have been written to emphasize the conceptual structure of the algorithm; they can be implemented more efficiently. It is important to recognize that the residual \mathbf{r}_t is always orthogonal to the columns of $\mathbf{\Phi}_t$. Therefore, the algorithm always selects a new atom at each step, and $\mathbf{\Phi}_t$ has full column rank.

The running time of the OMP algorithm is dominated by Step 2, whose total cost is $O(mNd)$. At iteration t , the least-squares problem can be solved with marginal cost $O(tN)$. To do so, we maintain a QR factorization of $\mathbf{\Phi}_t$. Our implementation uses the Modified Gram–Schmidt (MGS) algorithm because the measurement matrix is unstructured and dense. The book [Bjö96] provides extensive details and a survey of alternate approaches. When the measurement matrix is structured, more efficient implementations of OMP are possible; see the paper [KR06] for one example.

According to [NN94], there are algorithms that can solve (BP) with a dense, unstructured measurement matrix in time $O(N^2 d^{3/2})$. We are focused on the case where d is much larger than m or N , so there is a substantial gap between the cost of OMP and the cost of BP.

A prototype of the OMP algorithm first appeared in the statistics community at some point in the 1950s, where it was called stagewise regression. The algorithm later developed a life of its own in the signal processing [MZ93, PRK93, DMA97] and approximation theory [DeV98, Tem02] literatures. Our adaptation for the signal recovery problem seems to be new.

3. GAUSSIAN MEASUREMENT ENSEMBLES

In this report, we are concerned with Gaussian measurements only. In this section, we identify the properties of this measurement ensemble that are used to prove that the algorithm succeeds. Since Gaussian matrices are so well studied, we can make much more precise claims about them than other types of random matrices.

A Gaussian measurement ensemble for m -sparse signals in \mathbb{R}^d is a $d \times N$ matrix $\mathbf{\Phi}$, whose entries are drawn independently from the $\text{NORMAL}(0, N^{-1})$ distribution. For reference, the density function of this distribution is

$$p(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi N}} e^{-x^2 N/2}.$$

As we will see, this matrix has the following four properties:

- (G0) Independence: The columns of $\mathbf{\Phi}$ are stochastically independent.
- (G1) Normalization: $\mathbb{E} \|\boldsymbol{\varphi}_j\|_2^2 = 1$ for $j = 1, \dots, d$.
- (G2) Joint correlation: Let $\{\mathbf{u}_t\}$ be a sequence of m vectors whose ℓ_2 norms do not exceed one. Let $\boldsymbol{\varphi}$ be a column of $\mathbf{\Phi}$ that is independent from this sequence. Then

$$\mathbb{P} \{ \max_t |\langle \boldsymbol{\varphi}, \mathbf{u}_t \rangle| \leq \varepsilon \} \geq \left(1 - e^{-\varepsilon^2 N/2} \right)^m.$$

- (G3) Smallest singular value: Given an $N \times m$ submatrix \mathbf{Z} from $\mathbf{\Phi}$, the m th largest singular value $\sigma_{\min}(\mathbf{Z})$ satisfies

$$\mathbb{P} \left\{ \sigma_{\min}(\mathbf{Z}) \geq 1 - \sqrt{m/N} - \varepsilon \right\} \geq 1 - e^{-\varepsilon^2 N/2}$$

for any positive ε .

3.1. Joint Correlation. The joint correlation property (G2) is essentially a large deviation bound for sums of random variables. For the Gaussian measurement ensemble, we can leverage classical techniques to establish this property.

Proposition 4. *Let $\{\mathbf{u}_t\}$ be a sequence of m vectors whose ℓ_2 norms do not exceed one. Independently, choose \mathbf{z} to be a random vector with i.i.d. $\text{NORMAL}(0, N^{-1})$ entries. Then*

$$\mathbb{P}\{\max_t |\langle \mathbf{z}, \mathbf{u}_t \rangle| \leq \varepsilon\} \geq \left(1 - e^{-\varepsilon^2 N/2}\right)^m.$$

Proof. Let \mathbf{z} be a random vector whose entries are i.i.d. $\text{NORMAL}(0, 1)$. Define the event

$$E \stackrel{\text{def}}{=} \{\mathbf{z} : \max_t |\langle \mathbf{z}, \mathbf{u}_t \rangle| \leq \varepsilon \sqrt{N}\},$$

which is identical with the event that interests us. We will develop a lower bound on $\mathbb{P}(E)$. Observe that this probability decreases if we replace each vector \mathbf{u}_t by a unit vector pointing in the same direction. Therefore, we may assume that $\|\mathbf{u}_t\|_2 = 1$ for each t .

Geometrically, we can view $\mathbb{P}(E)$ as the Gaussian measure of m intersecting symmetric slabs. Sidak's Lemma [Bal02, Lemma 2] shows that the Gaussian measure of this intersection is no smaller than the product of the measures of the slabs. In probabilistic language,

$$\mathbb{P}(E) \geq \prod_{t=1}^m \mathbb{P}\left\{|\langle \mathbf{z}, \mathbf{u}_t \rangle| \leq \varepsilon \sqrt{N}\right\}.$$

Since each \mathbf{u}_t is a unit vector, each of the random variables $\langle \mathbf{z}, \mathbf{u}_t \rangle$ has a $\text{NORMAL}(0, 1)$ distribution on the real line. It follows that each of the m probabilities can be calculated as

$$\begin{aligned} \mathbb{P}\left\{|\langle \mathbf{z}, \mathbf{u}_t \rangle| \leq \varepsilon \sqrt{N}\right\} &= \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} e^{-x^2/2} dx \\ &\leq 1 - e^{-\varepsilon^2 N/2}. \end{aligned}$$

The final estimate is a well-known Gaussian tail bound. See [Bal02, p. 118], for example. \square

3.2. Smallest Singular Value. The singular value property (G3) follows directly from a theorem of Davidson and Szarek [DS02].

Proposition 5 (Davidson–Szarek). *Suppose that \mathbf{Z} is a tall $N \times m$ matrix whose entries are iid $\text{NORMAL}(0, N^{-1})$. Then its smallest singular value σ_{\min} satisfies*

$$\mathbb{P}\left\{\sigma_{\min}(\mathbf{Z}) \geq 1 - \sqrt{m/N} - \varepsilon\right\} \geq 1 - e^{-\varepsilon^2 N/2}.$$

It is a standard consequence of measure concentration that the minimum singular value of a Gaussian matrix clusters around its expected value (see [Led01], for example). Calculating the expectation, however, involves much more ingenuity. Davidson and Szarek produce their result with a clever application of the Slepian–Gordon lemma.

4. SIGNAL RECOVERY WITH ORTHOGONAL MATCHING PURSUIT

If we take random measurements of a sparse signal using an admissible measurement matrix, then OMP can be used to recover the original signal with high probability.

Theorem 6. *Suppose that \mathbf{s} is an arbitrary m -sparse signal in \mathbb{R}^d , and draw a random $N \times d$ Gaussian measurement matrix independent from the signal. Given the data $\mathbf{v} = \Phi \mathbf{s}$, Orthogonal Matching Pursuit can reconstruct the signal with probability exceeding*

$$\sup_{\varepsilon \in (0, \sqrt{N/m-1})} \left[1 - e^{-(\sqrt{N/m-1}-\varepsilon)^2/2}\right]^{m(d-m)} \left[1 - e^{-\varepsilon^2 m/2}\right]$$

The success probability here is best calculated numerically. Some analysis yields a slightly weaker but more serviceable corollary.

Corollary 7. *Fix $\delta \in (0, 1)$, and choose $N \geq Km \log(d/\delta)$ where K is an absolute constant. Suppose that \mathbf{s} is an arbitrary m -sparse signal in \mathbb{R}^d , and draw a random $N \times d$ Gaussian measurement matrix Φ independent from the signal. Given the data $\mathbf{v} = \Phi \mathbf{s}$, Orthogonal Matching Pursuit can reconstruct the signal with probability exceeding $(1 - \delta)$.*

The preceding theorem holds with $K = 16$ for any $m \geq 1$. As the number m of nonzero components approaches infinity, it is possible to take $K = 4 + \varepsilon$ for any positive number ε . These facts will emerge during the proof.

4.1. Proof of Theorem 6. Most of the argument follows the approach developed in [Tro04]. The main difficulty here is to deal with the nasty independence issues that arise in the stochastic setting. The primary novelty is a route to avoid these perils.

We begin with some notation and simplifying assumptions. Without loss of generality, assume that the first m entries of the original signal \mathbf{s} are nonzero, while the remaining $(d - m)$ entries equal zero. Therefore, the measurement vector \mathbf{v} is a linear combination of the first m columns from the matrix Φ . Partition the matrix as $\Phi = [\Phi_{\text{opt}} \mid \Psi]$ so that Φ_{opt} has m columns and Ψ has $(d - m)$ columns. Note that \mathbf{v} is stochastically independent from the random matrix Ψ .

For a vector \mathbf{r} in \mathbb{R}^N , define the *greedy selection ratio*

$$\rho(\mathbf{r}) \stackrel{\text{def}}{=} \frac{\|\Psi^T \mathbf{r}\|_\infty}{\|\Phi_{\text{opt}}^T \mathbf{r}\|_\infty} = \frac{\max_\psi |\langle \psi, \mathbf{r} \rangle|}{\|\Phi_{\text{opt}}^T \mathbf{r}\|_\infty}$$

where the maximization takes place over the columns of Ψ . If \mathbf{r} is the residual vector that arises in Step 2 of OMP, the algorithm picks a column from Φ_{opt} if and only if $\rho(\mathbf{r}) < 1$. In case $\rho(\mathbf{r}) = 1$, an optimal and a nonoptimal column both achieve the maximum inner product. The algorithm has no provision for choosing one instead of the other, so we assume that the algorithm always fails. The greedy selection ratio was first identified and studied in [Tro04].

Imagine that we could execute m iterations of OMP with the input signal \mathbf{s} and the restricted measurement matrix Φ_{opt} to obtain a sequence of residuals $\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_{m-1}$ and a sequence of column indices $\omega_1, \omega_2, \dots, \omega_m$. The algorithm is deterministic, so these sequences are both functions of \mathbf{s} and Φ_{opt} . In particular, the residuals are stochastically independent from Ψ . It is also evident that each residual lies in the column span of Φ_{opt} .

Execute OMP with the input signal \mathbf{s} and the full matrix Φ to obtain the actual sequence of residuals $\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_{m-1}$ and the actual sequence of column indices $\lambda_1, \lambda_2, \dots, \lambda_m$. Observe that OMP succeeds in reconstructing \mathbf{s} after m iterations if and only if the algorithm selects the first m columns of Φ in some order. We will use induction to prove that success occurs if and only if $\rho(\mathbf{q}_t) < 1$ for each $t = 0, 1, \dots, m - 1$.

In the first iteration, OMP chooses one of the optimal columns if and only if $\rho(\mathbf{r}_0) < 1$. The algorithm sets the initial residual equal to the input signal, so $\mathbf{q}_0 = \mathbf{r}_0$. Therefore, the success criterion is identical with $\rho(\mathbf{q}_0) < 1$. It remains to check that λ_1 , the actual column chosen, matches ω_1 , the column chosen in our thought experiment. Because $\rho(\mathbf{r}_0) < 1$, the algorithm selects the index λ_1 of the column from Φ_{opt} whose inner product with \mathbf{r}_0 is largest (ties being broken deterministically). Meanwhile, ω_1 is defined as the column of Φ_{opt} whose inner product with \mathbf{q}_0 is largest. This completes the base case.

Suppose that, during the first k iterations, the actual execution of OMP chooses the same columns as our imaginary invocation of OMP. That is, $\lambda_j = \omega_j$ for $j = 1, 2, \dots, k$. Since the residuals are calculated using only the original signal and the chosen columns, it follows that $\mathbf{r}_k = \mathbf{q}_k$. Repeating the argument in the last paragraph, we conclude that the algorithm identifies an optimal column if and only if $\rho(\mathbf{q}_k) < 1$. Moreover, it must select $\lambda_{k+1} = \omega_{k+1}$.

In consequence, the event on which the algorithm succeeds is

$$E_{\text{succ}} \stackrel{\text{def}}{=} \{\max_t \rho(\mathbf{q}_t) < 1\}$$

where $\{\mathbf{q}_t\}$ is a sequence of m random vectors that fall in the column span of Φ_{opt} and that are stochastically independent from Ψ . We can decrease the probability of success by placing the additional requirement that the smallest singular value of Φ_{opt} meet a lower bound:

$$\mathbb{P}(E_{\text{succ}}) \geq \mathbb{P}\{\max_t \rho(\mathbf{q}_t) < 1 \text{ and } \sigma_{\min}(\Phi_{\text{opt}}) \geq \sigma\}.$$

We will use Σ to abbreviate the event $\{\sigma_{\min}(\Phi_{\text{opt}}) \geq \sigma\}$. Applying the definition of conditional probability, we reach

$$\mathbb{P}(E_{\text{succ}}) \geq \mathbb{P}\{\max_t \rho(\mathbf{q}_t) < 1 \mid \Sigma\} \cdot \mathbb{P}(\Sigma). \quad (4.1)$$

Property (M3) controls $\mathbb{P}(\Sigma)$, so it remains to develop a lower bound on the conditional probability.

Assume that Σ occurs. For each index $t = 0, 1, \dots, m-1$, we have

$$\rho(\mathbf{q}_t) = \frac{\max_{\psi} |\langle \psi, \mathbf{q}_t \rangle|}{\|\Phi_{\text{opt}}^T \mathbf{q}_t\|_{\infty}}.$$

Since $\Phi_{\text{opt}}^T \mathbf{q}_t$ is an m -dimensional vector,

$$\rho(\mathbf{q}_t) \leq \frac{\sqrt{m} \max_{\psi} |\langle \psi, \mathbf{q}_t \rangle|}{\|\Phi_{\text{opt}}^T \mathbf{q}_t\|_2}.$$

To simplify this expression, define the vector

$$\mathbf{u}_t \stackrel{\text{def}}{=} \frac{\sigma \mathbf{q}_t}{\|\Phi_{\text{opt}}^T \mathbf{q}_t\|_2}.$$

The basic properties of singular values furnish the inequality

$$\frac{\|\Phi_{\text{opt}}^T \mathbf{q}\|_2}{\|\mathbf{q}\|_2} \geq \sigma_{\min}(\Phi_{\text{opt}}) \geq \sigma$$

for any vector \mathbf{q} in the range of Φ_{opt} . The vector \mathbf{q}_t falls in this subspace, so $\|\mathbf{u}_t\|_2 \leq 1$. In summary,

$$\rho(\mathbf{q}_t) \leq \frac{\sqrt{m}}{\sigma} \max_{\psi} |\langle \psi, \mathbf{u}_t \rangle|$$

for each index t . On account of this fact,

$$\mathbb{P}\{\max_t \rho(\mathbf{q}_t) < 1 \mid \Sigma\} \geq \mathbb{P}\left\{\max_t \max_{\psi} |\langle \psi, \mathbf{u}_t \rangle| < \frac{\sigma}{\sqrt{m}} \mid \Sigma\right\}.$$

Exchange the two maxima and use the independence of the columns of Ψ to obtain

$$\mathbb{P}\{\max_t \rho(\mathbf{q}_t) < 1 \mid \Sigma\} \geq \prod_{\psi} \mathbb{P}\left\{\max_t |\langle \psi, \mathbf{u}_t \rangle| < \frac{\sigma}{\sqrt{m}} \mid \Sigma\right\}.$$

Since every column of Ψ is independent from $\{\mathbf{u}_t\}$ and from Σ , Property (G2) of the measurement matrix yields a lower bound on each of the $(d-m)$ terms appearing in the product. It emerges that

$$\mathbb{P}\{\max_t \rho(\mathbf{q}_t) < 1 \mid \Sigma\} \geq \left(1 - e^{-\sigma^2 N/2m}\right)^{m(d-m)}.$$

We may choose the parameter

$$\sigma = 1 - \sqrt{m/N} - \varepsilon \sqrt{m/N}$$

where ε ranges between zero and $\sqrt{N/m} - 1$. This substitution delivers

$$\mathbb{P}\{\max_t \rho(\mathbf{q}_t) < 1 \mid \Sigma\} \geq \left(1 - e^{-(\sqrt{N/m} - 1 - \varepsilon)^2/2}\right)^{m(d-m)}.$$

With the foregoing choice of σ , Property (G3) furnishes

$$\mathbb{P}\{\sigma_{\min}(\Phi_{\text{opt}}) \geq \sigma\} \geq 1 - e^{-\varepsilon^2 m/2}.$$

Introduce this fact and (4.1) into the inequality (4.1). This action yields

$$\mathbb{P}(E_{\text{succ}}) \geq \left[1 - e^{-(\sqrt{N/m}-1-\varepsilon)^2/2}\right]^{m(d-m)} \left[1 - e^{-\varepsilon^2 m/2}\right] \quad (4.2)$$

for $\varepsilon \in (0, \sqrt{N/m}-1)$. The optimal value of this probability estimate is best determined numerically.

4.2. Proof of Corollary 7. We need to show that it possible to choose the number of measurements on the order of $m \ln d$ while maintaining an error as small as we like. Note that we may assume $m \geq 1$. We begin with (4.2), in which we apply the inequality $(1-x)^k \geq 1-kx$, valid for $k \geq 1$ and $x \leq 1$. Then invoke the bound $m(d-m) \leq d^2/4$ to reach

$$\mathbb{P}(E_{\text{succ}}) \geq 1 - m(d-m) e^{-(\sqrt{N/m}-1-\varepsilon)^2/2} - e^{-\varepsilon^2 m/2}. \quad (4.3)$$

where we have also discarded a positive term of higher order. We will bound the two terms on the right-hand side separately.

Fix a number $\delta \in (0, 1)$, and select $N \geq Km \ln d$, where the constant $K = K_m$ will be determined in a moment. Choose

$$\varepsilon = 1 + \left(\frac{2 \ln(d/\delta)}{m}\right)^{1/2}.$$

Clearly, ε is positive. By definition of N and ε , we have

$$\sqrt{N/m} - 1 - \varepsilon \geq \left(K^{1/2} - 2m^{-1/2}\right) \ln(d/\delta).$$

Our choice of K will ensure that the parenthesis is nonnegative, which in turn ensures that ε is not too large. The last displayed equation implies

$$d^2 e^{-(\sqrt{N/m}-1-\varepsilon)^2/2} \leq d^{2-(K^{1/2}-2m^{-1/2})} \delta^{K^{1/2}-2m^{-1/2}}.$$

To cancel the exponent two, we select $K = 4(1+m^{-1/2})^2$. With these choices, the second term on the right-hand side of (4.3) is no larger than $\delta^2/4$.

Let us move to the remaining term. We can bound ε below using the inequality $\sqrt{a}+\sqrt{b} \geq \sqrt{a+b}$.

$$\varepsilon = (\ln \varepsilon)^{1/2} + \left(\frac{2 \ln(d/\delta)}{m}\right)^{1/2} \geq \left[\ln(e(d/\delta)^{2/m})\right]^{1/2}.$$

It follows immediately that

$$e^{-\varepsilon^2 m/2} \leq \left(e(d/\delta)^{2/m}\right)^{-m/2} = e^{-m/2} d^{-1} \delta.$$

Since $m \geq 1$, we may conclude that the last term of (4.3) is no larger than $e^{-1/2} \delta$. Note that, when $\delta \geq d^{-1}$, the last term is actually smaller than $e^{-1/2} \delta^2$.

In view of these bounds,

$$\mathbb{P}(E_{\text{succ}}) \geq 1 - (0.25 + e^{-1/2})\delta > 1 - \delta.$$

Assuming that $\delta \geq d^{-1}$, the success probability actually satisfies a stronger estimate:

$$\mathbb{P}(E_{\text{succ}}) > 1 - \delta^2.$$

Moreover, the argument provides a sufficient choice of the constant:

$$K \geq 4(1+m^{-1/2})^2$$

When $m = 1$, we can take $K = 16$. It is also evident that we may select $K = 4 + \varepsilon$ for any positive ε , provided that m is large enough.

REFERENCES

- [Bal02] K. Ball. Convex geometry and functional analysis. In W. B. Johnson and J. Lindenstrauss, editors, *Handbook of Banach Space Geometry*, pages 161–193. Elsevier, 2002.
- [Bjö96] Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, 1996.
- [BV04] S. Boyd and L. Vanderberghe. *Convex Optimization*. Cambridge Univ. Press, 2004.
- [CDS01] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by Basis Pursuit. *SIAM Review*, 43(1):129–159, 2001.
- [CRT06] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information. *IEEE Trans. Inform. Theory*, 52(2):489–509, Feb. 2006.
- [CT04] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? Submitted for publication, Nov. 2004.
- [CT05] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, Dec. 2005.
- [DeV98] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, pages 51–150, 1998.
- [DH93] Ding-Zhu Du and Frank K. Hwang. *Combinatorial group testing and its applications*. World Scientific, 1993.
- [DMA97] G. Davis, S. Mallat, and M. Avellaneda. Greedy adaptive approximation. *J. Constr. Approx.*, 13:57–98, 1997.
- [Don06a] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, Apr. 2006.
- [Don06b] D. L. Donoho. For most large underdetermined systems of linear equations the minimal l_1 -norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59(6):797–829, 2006.
- [DS02] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices, and Banach spaces. In W. B. Johnson and J. Lindenstrauss, editors, *Handbook of Banach Space Geometry*, pages 317–366. Elsevier, 2002.
- [DT96] R. DeVore and V. N. Temlyakov. Some remarks on greedy algorithms. *Adv. Comput. Math.*, 5:173–187, 1996.
- [GMS05] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, 47(1):99–132, March 2005.
- [KR06] S. Kunis and H. Rauhut. Random sampling of sparse trigonometric polynomials II: Orthogonal Matching Pursuit versus Basis Pursuit. Preprint, 2006.
- [Led01] M. Ledoux. *The Concentration of Measure Phenomenon*. Number 89 in Mathematical Surveys and Monographs. American Mathematical Society, Providence, 2001.
- [Mil02] A. J. Miller. *Subset Selection in Regression*. Chapman and Hall, London, 2nd edition, 2002.
- [MZ93] S. Mallat and Z. Zhang. Matching Pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993.
- [NN94] Y. E. Nesterov and A. S. Nemirovski. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.
- [PRK93] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad. Orthogonal Matching Pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. of the 27th Annual Asilomar Conference on Signals, Systems and Computers*, Nov. 1993.
- [RKD99] B. D. Rao and K. Kreutz-Delgado. An affine scaling methodology for best basis selection. *IEEE Trans. Signal Processing*, 47(1):187–200, 1999.
- [RV05] M. Rudelson and R. Vershynin. Geometric approach to error correcting codes and reconstruction of signals. *Int. Math. Res. Not.*, 64:4019–4041, 2005.
- [Tem02] V. Temlyakov. Nonlinear methods of approximation. *Foundations of Comput. Math.*, July 2002.
- [Tro04] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, Oct. 2004.