MAKING DO WITH LESS: AN INTRODUCTION TO COMPRESSED SENSING

KURT BRYAN* AND TANYA LEISE†

Abstract. This article offers an accessible but rigorous and essentially self-contained account of some of the central ideas in compressed sensing, aimed at nonspecialists and undergraduates who have had linear algebra and some probability. The basic premise is first illustrated by considering the problem of detecting a few defective items in a large set. We then build up the mathematical framework of compressed sensing, to show how combining efficient sampling methods with elementary ideas from linear algebra and a bit of approximation theory, optimization, and probability, allows the estimation of unknown quantities with far less sampling of data than traditional methods.

Key words. compressed sensing, restricted isometry property, random matrices, linear programming, sparse signal recovery

AMS subject classifications. 94A12, 68P30, 15B52, 90C05.

1. Introduction. We begin with a simple puzzle to highlight a few key ideas. Suppose we have 7 gold coins, one of which we suspect is counterfeit and so of a different mass than the other coins. Given an accurate electronic scale, can we detect the counterfeit coin by using the scale at most 3 times? If we know how much a gold coin should weigh (e.g., an American Eagle half-ounce gold coin should weigh 16.966 grams), the following strategy works. Label the coins with numbers 1 through 7. For the first weighing place coins 1, 3, 5, and 7 on the scale; for the second weighing use coins 2, 3, 6, and 7; and for the third weighing use coins 4, 5, 6, and 7. We can express these choices using a $0-1$ matrix Φ whose kth row indicates which coins to include in the kth weighing:

$$\Phi = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{1.1}$$

Observe that the kth column encodes the integer k in binary, with the digit in the kth row corresponding to 2^{k-1}. It's not hard to see that from the outcome of these three weighings we can uniquely identify any single bad coin. For example, if only the first set's mass deviates from the nominal value (here, four times the mass of a good coin), we deduce that coin #1 is the counterfeit. If both the first and second sets' mass deviate, we deduce that coin #3 is the counterfeit, and so on. In general, coding in binary which sets' mass deviates from the nominal value yields the counterfeit coin's number. In the case of N coins, a similar strategy can be used to detect a single counterfeit coin with around $n = \log_2(N)$ weighings, a considerable savings over the N measurements required by sequentially weighing each coin.

This strategy is an example of a *combinatorial group test*, an efficient means of testing a large number of items in batches in order to identify a small set of atypical items. For example, in *high throughput screening* large chemical libraries are tested

*Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN 47803; email: kurt.bryan@rose-hulman.edu; phone: (812)877-8485; fax: (812)877-8883.

†Department of Mathematics, Amherst College, Amherst, MA 01002; email: tleise@amherst.edu; phone: (413)542-5411; fax: (413)542-2550.

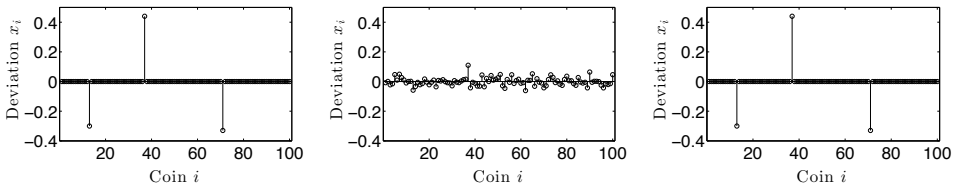

FIG. 1.1. *Left: Actual deviations x_i in mass. Center: Result of ℓ^2-regularization. Right: Result of ℓ^1-regularization.*

on a biological target with the goal of identifying a few active compounds (see [27] for more on this and other applications).

To illustrate, let us consider the counterfeit coin problem on a somewhat larger scale. Suppose we have $N = 100$ coins, numbered from 1 to 100, of which a small number may be counterfeit. Let the ith component x_i of $\mathbf{x} \in \mathbb{R}^N$ denote the *deviation* of the ith coin from the nominal mass. Specifically, let's suppose $x_{13} = -0.3, x_{37} = 0.44$, $x_{71} = -0.33$, and all other $x_i = 0$, so that coins 13, 37, and 71 are counterfeits. We seek to identify the counterfeits by weighing n subsets of the coins, with $n \ll N$. As before, we can form a $0-1$ sensing matrix Φ, now $n \times N$, whose kth row encodes which coins are included in the kth weighing. However, we are no longer assuming that at most one coin is bad, so it's not clear that the binary encoding strategy we used earlier will work. How should we design the sensing matrix?

Structured methods for choosing subsets to test in order to detect multiple deviations have been developed ([3, 20]), but we'll take a different tack: we'll choose the subsets randomly! In the present example we'll create a sensing matrix Φ by selecting $n = 20$ random subsets of coins to weigh. Specifically, let Φ be a 20×100 matrix, each entry chosen randomly and independently as 0 or 1 with equal probability (e.g., flip each coin to determine whether to include it in the current weighing). The component b_i of the vector $\mathbf{b} = \Phi\mathbf{x}$ is the deviation from nominal of the mass of the ith subset. Our goal is to recover $\mathbf{x}$ from knowledge of Φ and the measurements $\mathbf{b}$.

A matrix Φ with fewer rows than columns, like either sensing matrix above, leads to an underdetermined system $\Phi\mathbf{x} = \mathbf{b}$ which, if consistent, has infinitely many solutions. In such a situation one commonly *regularizes* the problem, by imposing additional conditions on $\mathbf{x}$ so that a unique solution $\mathbf{x} = \mathbf{x}^*$ exists. The difficulty lies in finding conditions that guarantee $\mathbf{x}^*$ is the desired solution (and not one of the infinitely many other solutions to $\Phi\mathbf{x} = \mathbf{b}$). In the present case what we have going for us is the fact that the correct solution is *sparse*, that is, most of its components are zero (since we assume most of the coins are not counterfeit or otherwise defective.)

One common choice for regularizing an underdetermined linear system is to choose the vector $\mathbf{x}^*$ that satisfies $\Phi\mathbf{x} = \mathbf{b}$ and minimizes the standard Euclidean ℓ^2 norm $\|\mathbf{x}\|_2 = (\sum_i x_i^2)^{1/2}$. This is an easy problem to solve with multivariable calculus, e.g., with Lagrange multipliers. Unfortunately, this type of regularization is quite inappropriate here: the solution vector $\mathbf{x}^*$ that minimizes $\|\mathbf{x}\|_2$ does not correspond to a few bad coins, since $x_i^* \neq 0$ for most indices i, as shown in the middle panel of Figure 1.1. Note however that $\mathbf{x}^*$ satisfies $\Phi\mathbf{x}^* = \mathbf{b}$ exactly.

More generally, the minimum ℓ^2 norm solution to a linear system $\Phi\mathbf{x} = \mathbf{b}$ is almost never sparse. This is illustrated in the left panel of Figure 1.2, in which the dashed line is a low-dimensional analogue of the hyperplane representing the set of all solutions to $\Phi\mathbf{x} = \mathbf{b}$. Imagine increasing the radius of a circle centered at the origin

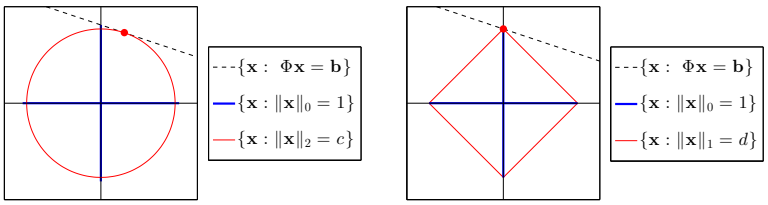

FIG. 1.2. *Left: 2-dimensional analogues of ℓ^2 and ℓ^0 balls and of $\Phi\mathbf{x} = \mathbf{b}$. Right: 2-dimensional analogues of ℓ^1 and ℓ^0 balls and of $\Phi\mathbf{x} = \mathbf{b}$. Here c and d are constants with c a bit less than d. Note that the set $\{\mathbf{x} : \|\mathbf{x}\|_0 = 1\}$ coincides with the coordinate axes.*

until it touches the line. This point of contact yields the minimum value of $\sqrt{x_1^2 + x_2^2}$ among all points on the line and is the solution to $\Phi\mathbf{x} = \mathbf{b}$ with minimum ℓ^2 norm. It's easy to see that for a typical line both components of this point are nonzero.

Since we believe the true solution is sparse, it makes sense to look for solutions to $\Phi\mathbf{x} = \mathbf{b}$ that have the fewest possible nonzero components. Let us define $\|\mathbf{x}\|_0$ as the number of nonzero components in $\mathbf{x}$, then regularize by seeking a solution to $\Phi\mathbf{x} = \mathbf{b}$ that minimizes $\|\mathbf{x}\|_0$. The quantity $\|\mathbf{x}\|_0$ is sometimes called the "ℓ^0 norm," though it is not a norm; see Exercise 3. If maximal sparsity is what we want, this is clearly the way to go. The panels in Figure 1.2 illustrate the idea in $\mathbb{R}^2$: The set of vectors that have one nonzero component, $\{\mathbf{x} : \|\mathbf{x}\|_0 = 1\}$, coincides with the coordinate axes. The solution set to $\Phi\mathbf{x} = \mathbf{b}$ thus contacts $\|\mathbf{x}\|_0 = 1$ in precisely two locations, each on a coordinate axis. These solutions are sparser than that obtained via ℓ^2 regularization, indeed they are the sparsest possible. Moreover, these ideas extend to the general problem in which $\mathbf{x}$ has multiple nonzero components. Unfortunately, finding the solution to a linear system $\Phi\mathbf{x} = \mathbf{b}$ with the fewest nonzero components is computationally intractable if the system is large; see [22].

What we need is a regularization technique that promotes sparse solutions yet remains computationally tractable. Regularizing with the ℓ^1 norm, defined as

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{N} |x_i|, \tag{1.2}$$

turns out to do the trick! For the underlying intuition see the panel on the right in Figure 1.2, which shows the graph of an ℓ^1 ball in $\mathbb{R}^2$ ($|x_1| + |x_2| = d$) that just contacts the solution set to $\Phi\mathbf{x} = \mathbf{b}$. This point is the minimum ℓ^1 norm solution and agrees with a sparse solution produced by ℓ^0 regularization.

Although minimizing $\|\mathbf{x}\|_1$ subject to linear constraints $\Phi\mathbf{x} = \mathbf{b}$ may look difficult, because $|x_i|$ is not differentiable, it turns out that this problem falls into the realm of convex optimization. Indeed, the problem can be converted to a standard problem in linear programming, as described in Section 3 of [16].

For the 100 coin problem, minimizing $\|\mathbf{x}\|_1$ subject to constraints $\Phi\mathbf{x} = \mathbf{b}$ (with a specific randomly chosen Φ) yields $x_{13} = -0.3, x_{37} = 0.44$, $x_{71} = -0.33$, and all other $x_i = 0$, *exactly*, as illustrated by the right panel in Figure 1.1! But maybe we got lucky—after all, the $n = 20$ subsets were randomly chosen. However, repeating the above experiment 1000 times (each time choosing a new random Φ) yields 980 successes, each an exact solution. If we increase the number of weighings for each experiment to $n = 25$ we obtain a perfect record, 1000 successes in 1000 trials. Decreasing the number of weighings to $n = 15$ in each experiment results in 787 perfect

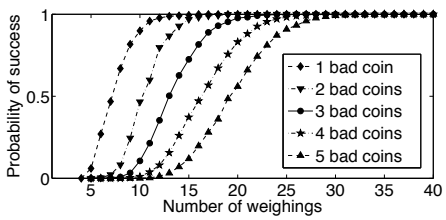

FIG. 1.3. *Success rates of ℓ^1-regularization for $N = 100$ coins, as a function of number n of weighings.*

solutions in 1000 trials, and $n = 10$ results in only 141 successes. Figure 1.3 illustrates the situation, for one to five bad coins. As the number of bad coins goes up, we need more weighings to reliably identify them.

1.1. Overview of Compressed Sensing. The counterfeit coin problem illustrates a few key features of compressed sensing (CS): We seek to recover a sparse vector $\mathbf{x} \in \mathbb{R}^N$, that is, a vector with most of its components equal to zero, from measurements made by taking $n \ll N$ inner products, $b_i = \langle \mathbf{r}_i, \mathbf{x} \rangle$, where each $\mathbf{r}_i$ is a row vector that is typically generated randomly. In the coin problem $\mathbf{r}_i$ is a random $0-1$ vector encoding which coins were used in the ith weighing. In other applications the $\mathbf{r}_i$ may have random components drawn from a normal or other distribution. To recover $\mathbf{x}$ we form an $n \times N$ matrix Φ from these row vectors $\mathbf{r}_i$ and then find the minimum ℓ^1 norm solution to the linear system $\Phi\mathbf{x} = \mathbf{b}$.

The field of CS emerged as a hot topic with the publication of seminal papers in 2006 by Emmanuel Candès, Justin Romberg, and Terence Tao [11] and by David Donoho [21]. Traditional signal processing based on Shannon's information theory focuses on uniform sampling, that is, systematically collecting data at evenly spaced points on a grid to achieve some desired resolution. Think of a digital camera taking a high-resolution photograph by recording a value at every single pixel on a finely spaced grid. This approach takes the pessimistic view that we know nothing *a priori* about the data. The leap that propels CS is the realization that most data, e.g., photos of people or landscapes, have some inherent structure that we can use to our advantage. This inherent structure can often be viewed as a type of sparsity. In contrast to the typical megapixel digital camera, a CS-designed single-pixel camera takes relatively few measurements that are equivalent to sums of randomly located pixels [24], analogous to our example of using a scale to measure the total mass of randomly selected subsets of coins. The article [27] provides additional background and references concerning problems involving sparse approximation and signal recovery, closely related to and often predating CS.

The key mathematical ideas in CS, as we will show, can be easily understood and form an elegant theory. It should be noted, however, that developing practical CS devices is extremely challenging, requiring new sensing technology able to physically mimic the action of random measurement matrices.

For overviews of CS accessible to a general audience, see [14, 28, 31, 32, 33], as well as [26] which includes a historical overview. For a wonderful audio demonstration of CS, see the website [2]. More complete mathematical developments of CS can be found in [13, 9, 4, 18], to mention a few examples from the rapidly expanding CS literature. A comprehensive listing of CS-related articles and other materials can be found at http://dsp.rice.edu/cs.

1.2. The General Setting. In the counterfeit coin problem of Section 1 we sought to recover a sparse vector $\mathbf{x} \in \mathbb{R}^N$ from an underdetermined system $\Phi\mathbf{x} = \mathbf{b}$. In other settings the vector $\mathbf{x}$ may not itself be sparse, but rather it may be the case that $\mathbf{x} = \Psi\mathbf{s}$ for some $n \times n$ orthogonal matrix Ψ and sparse vector $\mathbf{s}$; that is, $\mathbf{x}$ may have a sparse representation in an alternate basis for $\mathbb{R}^N$, spanned by the columns of Ψ. In this case the information $\mathbf{b}$ we collect leads to a system $\Phi\Psi\mathbf{s} = \mathbf{b}$, from which we wish to recover $\mathbf{s}$ (and then $\mathbf{x}$). An important topic in compressed sensing is that of *incoherence*, a lack of correlation between the sensing modality embodied by the rows of Φ and the basis formed by the columns of Ψ, that facilitates the recovery of sparse signals, a condition often met when Φ is randomly generated. In the interest of brevity we'll focus on the case in which Ψ is the identity matrix—so $\mathbf{x}$ is itself sparse—and Φ is randomly generated in some manner. For more on the notion of incoherence see [9].

We thus focus on solving a linear system of equations

$$\Phi\mathbf{x} = \mathbf{b}, \tag{1.3}$$

where the sensing matrix Φ is $n \times N$ and n may be much smaller than N. We will assume that (1.3) is consistent. In this case any solution will not be unique. However, we're going to add the additional assumption that the solution vector $\mathbf{x}$ is "k-sparse" for some value of $k \ll n$, that is, $\mathbf{x}$ has at most k nonzero components. We'll denote the set of such vectors in $\mathbb{R}^N$ by Σ_k (the dependence on N will not be explicit). A few natural questions arise:

1. Does the additional information that the solution is k-sparse nail down a unique solution? If so, for what relative values of k, n, and N? What conditions on the matrix Φ are sufficient?
2. Under what conditions will minimizing the ℓ^1 norm subject to $\Phi\mathbf{x} = \mathbf{b}$ be successful in recovering sparse solutions? Why does it work?

We'll address each of these in the following sections. Section 2 will develop conditions guaranteeing uniqueness of k-sparse solutions by examining the null space of Φ and leads to the "restricted isometry property," which is the focus of Section 3. Section 2 requires familiarity with basic linear algebra and Section 3 assumes knowledge of elementary probability. The proof that ℓ^1 minimization recovers the k-sparse solution is given in Section 4, and we make some final remarks in Section 5.

Exercise 1. Suppose that for Φ as in (1.1) we have $b_1 = 0, b_2 = 0.2$, and $b_3 = 0.2$ where $\mathbf{b} = \Phi\mathbf{x}$ and x_i denotes the deviation of the ith coin from the nominal value. If we know at most one coin is counterfeit, which one is it? Make up an example (by making appropriate choices for $\mathbf{x}$) to show that two bad coins cannot be uniquely identified.

Exercise 2. Is Σ_k a subspace of $\mathbb{R}^N$? Why or why not?

Exercise 3. Let $\|\mathbf{x}\|_0$ denote the number of nonzero components in the vector $\mathbf{x}$. Why isn't this a norm?

Exercise 4. Let $\Phi = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$ be any 1×2 ($n = 1, N = 2$) matrix in which both entries are nonzero, and consider the equation $\Phi\mathbf{x} = b$, where $\mathbf{x} \in \mathbb{R}^2$ and $b \neq 0$ is a scalar. Show that there are always two 1-sparse solutions to $\Phi\mathbf{x} = b$. Thus we can't solve $\Phi\mathbf{x} = b$ uniquely for $\mathbf{x}$ in this situation, even under the assumption that $\mathbf{x}$ is 1-sparse.

Exercise 5. Let

$$\Phi = \begin{bmatrix} 1 & 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1 & 1/\sqrt{2} \end{bmatrix}.$$

a. Suppose that for some fixed $\mathbf{b} \in \mathbb{R}^2$ the equation $\Phi\mathbf{x} = \mathbf{b}$ has a 1-sparse solution. Show this solution is unique, and so we can recover any 1-sparse solution $\mathbf{x} \in \mathbb{R}^4$.
b. Let $\mathbf{b}$ a vector in $\mathbb{R}^2$. Show that the equation $\Phi\mathbf{x} = \mathbf{b}$ can have as many as six distinct 2-sparse solutions.

Exercise 6. Consider the underdetermined linear equation $\Phi\mathbf{x} = \mathbf{b}$ where Φ is the matrix in Exercise 5, $\mathbf{x} \in \mathbb{R}^4$, and $\mathbf{b} = [0, 3]^t$ (here the superscript t denotes the transpose).

a. Verify that the vector $\mathbf{x} = [0, 0, 3, 0]^t$ is a 1-sparse solution.
b. Find the minimum norm solution to $\Phi\mathbf{x} = \mathbf{b}$ using the ℓ^2 norm. (Suggestion: Solve $\Phi\mathbf{x} = \mathbf{b}$ for x_1 and x_3 in terms of x_2 and x_4, then express $\|\mathbf{x}\|_2^2$ in terms of just x_2 and x_4 and minimize in these two variables.) What's the sparsity of this solution?
c. Find the minimum norm solution to $\Phi\mathbf{x} = \mathbf{b}$ using the ℓ^1 norm $\|\mathbf{x}\|_1$ and the same approach as part (b). Although $\|\mathbf{x}\|_1$ isn't differentiable, it's easy to find the minimum graphically after you've expressed $\|\mathbf{x}\|_1$ as a function of two variables, by plotting $\|\mathbf{x}\|_1$ as a function of x_2 and x_4.

2. Uniqueness of k-Sparse Solutions.

Our first task is to establish conditions guaranteeing that there is only one k-sparse solution to $\Phi\mathbf{x} = \mathbf{b}$. This gives us some hope of being able to distinguish it from all of the other solutions.

2.1. The Null Space of Φ.

Let $\mathbf{x}^*$ satisfy $\Phi\mathbf{x}^* = \mathbf{b}$. All other solutions to $\Phi\mathbf{x} = \mathbf{b}$ are of the form $\mathbf{x} = \mathbf{x}^* + \eta$ where $\Phi\eta = \mathbf{0}$. That is, η is in $\mathcal{N}(\Phi)$, the null space of Φ. Let us suppose that $\mathbf{x}^* \in \Sigma_k$ for some k and examine conditions under which $\mathbf{x}^*$ is certain to be the only k-sparse solution.

Suppose, to the contrary, there is another (distinct) k-sparse solution $\mathbf{x}^{**}$. We have $\Phi(\mathbf{x}^* - \mathbf{x}^{**}) = \mathbf{0}$, that is, $\mathbf{x}^* - \mathbf{x}^{**} \in \mathcal{N}(\Phi)$, but $\mathbf{x}^* - \mathbf{x}^{**}$ is not the zero vector. Observe that if $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are any vectors in Σ_k then $\mathbf{x}^* - \mathbf{x}^{**} \in \Sigma_{2k}$ (Exercise 7). We conclude that if $\Phi\mathbf{x} = \mathbf{b}$ has more than one k-sparse solution, $\mathcal{N}(\Phi)$ must contain a nonzero $2k$-sparse vector. The contrapositive of this statement yields the next lemma.

Lemma 2.1. *Suppose that $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$, that is, all nonzero elements in the null space of Φ have at least $2k+1$ non-zero components. Then any k-sparse solution to $\Phi\mathbf{x} = \mathbf{b}$ is unique.*

A simple variation on the condition $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$ is given by the following:

Lemma 2.2. *The condition $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$ holds if and only if every subset of $2k$ columns of Φ is linearly independent.*

Proof: Suppose $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$. Let ϕ_j denote the jth column of Φ and let $T \subseteq \{1, \ldots, N\}$ be any subset of indices with cardinality $|T| = 2k$. Consider the subset $\{\phi_j : j \in T\}$ of $2k$ columns of Φ. If $\sum_{j \in T} x_j \phi_j = 0$ for some scalars x_j, then $\Phi\mathbf{x} = \mathbf{0}$ for the vector $\mathbf{x}$ with components x_j for $j \in T$ and $x_j = 0$ otherwise. Thus $\mathbf{x} \in \Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$, which implies $\mathbf{x} = \mathbf{0}$. Therefore $x_j = 0$ for all j and so $\{\phi_j : j \in T\}$ is a linearly independent set.

Conversely, suppose every subset $\{\phi_j : j \in T\}$ of $2k$ columns of Φ is linearly independent. Consider any vector $\mathbf{x} \in \Sigma_{2k} \cap \mathcal{N}(\Phi)$ and let $T \subseteq \{1, \ldots, N\}$ be any set of $2k$ indices so that if $x_j \neq 0$ then $j \in T$ (and $x_j = 0$ if $j \notin T$). We use this to form a subset $\{\phi_j : j \in T\}$ of $2k$ columns of Φ that satisfies

$$\sum_{j \in T} x_j \phi_j = \Phi \mathbf{x} = \mathbf{0},$$

because $\mathbf{x} \in \mathcal{N}(\Phi)$. But $\{\phi_j : j \in T\}$ is linearly independent by assumption, so $x_j = 0$ for every $j = 1, \ldots, N$, that is, $\mathbf{x} = \mathbf{0}$. Therefore $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$.$\square$

Unfortunately, the condition $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$ could be very hard to check for any given matrix. A brute force approach based on Lemma 2.2 would require us to check the linear independence of all $\binom{N}{2k}$ subsets of columns for Φ, an essentially impossible task if N, n, and k are very large. We need to find a better strategy.

Exercise 7. Show that if $\mathbf{x}^*$ and $\mathbf{x}^{**}$ are both in Σ_k then any linear combination $c_1\mathbf{x}^* + c_2\mathbf{x}^{**}$ lies in Σ_{2k}.

Exercise 8. If $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$ holds for an $n \times N$ matrix Φ, why must we have $2k \leq n$? (Hint: use Lemma 2.2.)

Exercise 9. Let Φ be the matrix from Exercise 5. Verify that the vectors

$$\mathbf{v}_1 = [1 \quad -\sqrt{2} \quad 1 \quad 0]^t, \quad \mathbf{v}_2 = [\sqrt{2} \quad -1 \quad 0 \quad 1]^t$$

form a basis for $\mathcal{N}(\Phi)$, and use this to show that there are no nonzero 2-sparse vectors in $\mathcal{N}(\Phi)$.

Exercise 10. Suppose Φ has the property that $\mathcal{N}(\Phi) \cap \Sigma_m = \{\mathbf{0}\}$ for some m.

a. Show that if $\Phi' = c\Phi$ for some nonzero scalar c then $\mathcal{N}(\Phi') \cap \Sigma_m = \{\mathbf{0}\}$.
b. Show that if Φ' is obtained from Φ by multiplying the jth row of Φ by some nonzero scalar c_j (for one specific value of j) then $\mathcal{N}(\Phi') \cap \Sigma_m = \{\mathbf{0}\}$. Hint: Show that if $\mathbf{x}$ satisfies $\Phi'\mathbf{x} = \mathbf{0}$ then $\Phi\mathbf{x} = \mathbf{0}$ too.
c. Use part (b) to show that if Φ' is obtained from Φ by multiplying each row of Φ by some nonzero scalar (not necessarily the same for each row) then $\mathcal{N}(\Phi') \cap \Sigma_m = \{\mathbf{0}\}$.
d. Show that if Φ' is obtained by multiplying any given column of Φ by a nonzero scalar d then $\mathcal{N}(\Phi') \cap \Sigma_m = \{\mathbf{0}\}$. Hint: If Φ' is obtained by multiplying the jth column of Φ by d and $\mathbf{x}' = (x_1, x_2, \ldots, x_N)$ satisfies $\Phi'\mathbf{x}' = \mathbf{0}$ then $\mathbf{x} = (x_1, x_2, \ldots, x_{j-1}, dx_j, x_{j+1}, \ldots, x_N)$ satisfies $\Phi\mathbf{x} = \mathbf{0}$.
e. Use part (d) to argue that if Φ' is obtained from Φ by multiplying each column of Φ by some nonzero scalar (not necessarily the same for each column) then $\mathcal{N}(\Phi') \cap \Sigma_m = \{\mathbf{0}\}$.

2.2. The Restricted Isometry Property. If a brute force approach to verifying the condition $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$ isn't feasible, how do we check that it holds in any specific case, or better yet, build it into the sensing matrix Φ?

The condition $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$ requires that no nonzero vector $\mathbf{x} \in \Sigma_{2k}$ satisfies $\Phi\mathbf{x} = \mathbf{0}$. There is no loss of generality in confining our attention to unit vectors (with respect to the ℓ^2 norm), because if $\mathbf{x} \neq \mathbf{0}$ then $\Phi\mathbf{x} = \mathbf{0}$ if and only if $\Phi\mathbf{u} = \mathbf{0}$ where $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|_2$, a unit vector. We thus seek a condition to assure that no unit vector $\mathbf{u} \in \Sigma_{2k}$ satisfies $\Phi\mathbf{u} = \mathbf{0}$. One simple way to do this is to require that there exists a positive constant c_1 such that for all $2k$-sparse unit vectors $\mathbf{u}$ we have $\|\Phi\mathbf{u}\|_2^2 \geq c_1$.

This rules out $\Phi\mathbf{u} = \mathbf{0}$, because then $\|\Phi\mathbf{u}\|_2^2 = 0$. We've thus established the following lemma:

LEMMA 2.3. *If there exists a positive constant c_1 such that $c_1 \leq \|\Phi\mathbf{u}\|_2^2$ for all $2k$-sparse unit vectors then $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$.*

A variation on Lemma 2.3 turns out to be useful for analyzing the effectiveness of ℓ^1 minimization for recovering sparse solutions. First, whether or not $c_1 \leq \|\Phi\mathbf{u}\|_2^2$ holds, there will *always* exist some constant $c_2 > 0$ so that $\|\Phi\mathbf{u}\|_2^2 \leq c_2$ for all unit vectors $\mathbf{u} \in \Sigma_{2k}$. The reason is that the mapping $\mathbf{x} \to \|\Phi\mathbf{x}\|_2^2$ is continuous from $\mathbb{R}^N$ to $\mathbb{R}$ (Exercise 14) and the set of $2k$-sparse unit vectors in $\mathbb{R}^N$ is compact (Exercise 15), so $\|\Phi\mathbf{u}\|_2^2$ must attain a maximum value on this set. We can take c_2 to be this maximum value. The two conditions can be amalgamated into the statement that there exist positive constants c_1 and c_2 so that

$$c_1 \leq \|\Phi\mathbf{u}\|_2^2 \leq c_2 \tag{2.1}$$

for all unit vectors $\mathbf{u} \in \Sigma_{2k}$.

From Exercise 10 part (a), rescaling the matrix Φ by a constant doesn't change the condition $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$ (or the problem of solving $\Phi\mathbf{x} = \mathbf{b}$, if we rescale $\mathbf{b}$ too). In what follows it will be convenient to multiply (2.1) through by $2/(c_1 + c_2)$ and define $\delta = (c_2 - c_1)/(c_2 + c_1)$, where we'll always have $0 \leq \delta \leq 1$. If we redefine Φ appropriately by multiplying by $\sqrt{2/(c_1 + c_2)}$ then (2.1) yields the equivalent inequality

$$1 - \delta \leq \|\Phi\mathbf{u}\|_2^2 \leq 1 + \delta \tag{2.2}$$

for the rescaled system with Φ and $\mathbf{b}$ both rescaled by a factor $\sqrt{2/(c_1 + c_2)}$.

Note that $c_1 > 0$ excludes the possibility $\delta = 1$. This motivates the following key definition, introduced by Candès and Tao [12]:

DEFINITION 2.4. *An $n \times N$ matrix Φ satisfies the restricted isometry property (RIP) of order m if there is some constant $\delta_m \in (0, 1)$ such that*

$$1 - \delta_m \leq \|\Phi\mathbf{u}\|_2^2 \leq 1 + \delta_m \tag{2.3}$$

for all m-sparse unit vectors $\mathbf{u} \in \mathbb{R}^N$.

With Definition 2.4 can state the following variation of Lemma 2.3:

LEMMA 2.5. *If a matrix Φ satisfies the RIP of order $2k$ for some $k \geq 1$ then $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$ and any k-sparse solution to $\Phi\mathbf{x} = \mathbf{b}$ is unique.*

Again, the right hand inequality of (2.3) is not necessary for $\Sigma_{2k} \cap \mathcal{N}(\Phi) = \{\mathbf{0}\}$, but the size of the constant δ_m has implications for the effectiveness and stability of ℓ^1 minimization in recovering sparse solutions to $\Phi\mathbf{x} = \mathbf{b}$, as we'll see in Section 4.

EXERCISE 11. Let $\Phi = [1/2 \;\; 4/3]$.

a. Show the Φ satisfies the RIP (2.3) of order 1 with constant $\delta_1 = 7/9$.
b. Show that Φ does NOT satisfy the RIP of order 2. (Remember, we need $\delta_2 \in (0, 1)$.)

Exercise 12. Show that if a matrix Φ satisfies the RIP of order k, then it also satisfies the RIP of order j for any positive integer j less than k.

Exercise 13. Use the fact that any vector $\mathbf{x} \in \mathbb{R}^N$ can be written as $\mathbf{x} = \|\mathbf{x}\|_2 \mathbf{u}$, where $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|_2$ is a unit vector, to show that Definition 2.4 is equivalent to requiring that there exists some $\delta_m \in (0,1)$ such that

$$(1-\delta_m)\|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1+\delta_m)\|\mathbf{x}\|_2^2 \tag{2.4}$$

for any m-sparse vector $\mathbf{x}$ in $\mathbb{R}^N$ (not restricting $\mathbf{x}$ to be a unit vector).

2.3. Interpretation of the RIP. To interpret the RIP (2.4) from a more geometric perspective, suppose (2.4) holds for the case $m = 2k$. Let $\mathbf{x}$ and $\mathbf{x}'$ be any two vectors in Σ_k. From the left inequality in (2.4) we then have

$$\|\Phi\mathbf{x} - \Phi\mathbf{x}'\|_2 = \|\Phi(\mathbf{x}-\mathbf{x}')\|_2 \geq \sqrt{1-\delta_{2k}}\|\mathbf{x}-\mathbf{x}'\|_2 \tag{2.5}$$

since $\mathbf{x} - \mathbf{x}' \in \Sigma_{2k}$. Equation (2.5) shows that the distance between $\Phi\mathbf{x}$ and $\Phi\mathbf{x}'$ is always some fraction of the distance between the vectors $\mathbf{x}$ and $\mathbf{x}'$ themselves. The closer δ_{2k} is to zero, the more Φ behaves like an isometry (distance preserving map) on Σ_k, keeping elements of Σ_k well separated under multiplication by Φ. In other words, the images $\Phi\mathbf{x}$ for $\mathbf{x} \in \Sigma_k$ will be easy to distinguish from each other.

Unfortunately, the RIP itself isn't really any easier to verify for a given matrix Φ than the condition that subsets of columns of Φ are independent. However, the advantage of the RIP is that it can be shown to hold with high probability for large classes of matrices generated by certain random procedures, so we can be confident that these matrices will work in a compressed sensing application. We give an example of such a class of matrices in Section 3.

Exercise 14. Prove that the mapping $g : \mathbb{R}^N \to \mathbb{R}$ defined by $g(\mathbf{x}) := \|\Phi\mathbf{x}\|_2^2$ is continuous.

Exercise 15. Prove that the set K_m of m-sparse unit vectors in $\mathbb{R}^N$ is compact, i.e., closed and bounded. Suggestion: K_m is clearly bounded, so show that the complement of K_m is open (given any vector with at least $m+1$ nonzero components or that is not a unit vector, show there exists an open neighborhood of that vector that does not intersect K_m).

Exercise 16. Here's an alternative version of the RIP of Definition 2.4, based on eigenvalues. First, consider a subset $T \subseteq \{1, 2, \ldots, N\}$, say with elements T_i and cardinality $|T|$. If Φ is $n \times N$, let Φ_T be the $n \times |T|$ matrix obtained by deleting those columns of Φ whose index does not lie in T.

a. Show that the condition that there exists a constant $\delta_m \in (0,1)$ such that

$$(1-\delta_m) \leq \|\Phi_T\mathbf{u}\|_2^2 \leq (1+\delta_m) \tag{2.6}$$

for all subsets T with $|T| \leq m$ and all unit vectors $\mathbf{u} \in \mathbb{R}^{|T|}$ is equivalent to Definition 2.4.

b. It's a fact from linear algebra that if $\mathbf{M}$ is an $n \times p$ matrix and $\mathbf{u}$ is a unit vector in $\mathbb{R}^p$ then

$$\lambda' \leq \|\mathbf{M}\mathbf{u}\|_2^2 \leq \lambda'' \tag{2.7}$$

where λ' and λ'' are the smallest and largest eigenvalues of the $p \times p$ matrix $\mathbf{M}^t\mathbf{M}$ ($\mathbf{M}^t\mathbf{M}$ is symmetric positive semidefinite, so all eigenvalues are real and

nonnegative; the inequality (2.7) can be proved by writing $\mathbf{M}^t\mathbf{M} = \mathbf{QDQ}^t$ where $\mathbf{Q}$ is orthogonal). Use this fact to show that the condition of equation (2.6) in part (a) is equivalent to the statement that for any set T with $|T| \leq m$ the eigenvalues of $\Phi_T^t\Phi_T$ lie in the interval $(1-\delta_m, 1+\delta_m)$.

c. Use the condition of part (b) to show that the matrix Φ in Exercise 5 satisfies the RIP of order 2 with $\delta_2 = \sqrt{2}/2 \approx 0.707$. You'll need to form the 2×2 matrix $\Phi_T^t\Phi_T$ for all six subsets T of $\{1,2,3,4\}$ with $|T| = 2$ and compute the eigenvalues; use a computer algebra system or Matlab.

3. RIP for Normal Random Matrices. Although the RIP may be hard to verify for any specific matrix, it turns out that matrices constructed via certain random processes can be shown to possess the RIP with high probability. This is one reason why the matrices used in compressed sensing generally involve some kind of randomness. Many classes of random matrices have been shown to be suitable. The focus of this section is to examine one such class, namely, matrices with entries that are independent normally distributed random variables. The results of this section are intended to convince the reader that there are in fact matrices that possess the RIP, but the results are not necessary to understand why ℓ^1 minimization works, which is the topic of Section 4.

In the coin problem above we were able to find a small number of defective coins with high probability, by using only a fraction $q = 0.25$ of the usually necessary $N = 100$ equations. More generally, we seek conditions under which we have a good chance of identifying a k-sparse vector in $\mathbb{R}^N$ using only $n = qN$ ($q \ll 1$) linear equations. The following theorem offers an answer by telling us when the RIP will hold with high probability for random matrices of certain types.

THEOREM 3.1. *Let Φ be an $n \times N$ matrix with $n = qN$ for some $q \in (0,1)$ and entries that are independent samples of a normal random variable with mean 0 and variance $1/n$. For any fixed $\epsilon \in (0,1)$, $\delta \in (0,1)$ and $n \geq 12/\delta$ the matrix Φ will satisfy the RIP of order m with constant $\delta_m = \delta$ with probability at least $1-\epsilon$ if N is chosen large enough to satisfy the inequality*

$$-c_0 qN + (m+1/2)\ln(N) + c_1 \leq \ln(\epsilon) \tag{3.1}$$

where $c_0 = \delta^2/144 - \delta^3/1296 > 0$ and $c_1 = m\ln(\frac{36e}{m\delta}) + \frac{1}{2}\ln(q/\pi)$.

In short, for these types of random matrices the RIP of a given order can be made to hold with probability as close to one as we like, provided k, n, and N stand in a certain relation to each other.

The focus of the rest of this section is to give a simple proof of Theorem 3.1, similar to that in [4], which we break up into a few lemmas requiring no more than elementary probability and calculus. The proof hinges on the *concentration inequality* stated in Lemma 3.2.

EXERCISE 17. Show that for any fixed q, ϵ, δ in $(0,1)$, and $m \geq 1$, the inequality (3.1) will hold for all sufficiently large N. With $\delta = 0.1, \epsilon = 0.01, m = 10$, and $q = 0.25$, how large must N be for the inequality to be satisfied?

3.1. Showing that the RIP Holds with High Probability. Before stating the next lemma, we define the following function $p(n,\epsilon)$ that we will use to help bound

the probability that a matrix Φ does *not* exhibit the desired properties:

$$p(n,\epsilon) = \sqrt{\frac{n}{\pi}} e^{-n(\epsilon^2/4-\epsilon^3/6)}. \tag{3.2}$$

For any fixed $\epsilon > 0$, the function $p(n,\epsilon)$ can be made arbitrarily (and rapidly) close to zero by taking n large (see Figure 3.1). In what follows we use the notation $P(E)$ to denote the probability of an event E.

LEMMA 3.2. *Let Φ be an $n \times N$ matrix whose entries ϕ_{ij} are independent samples of a normal random variable with mean 0 and variance $1/n$. For any fixed $\epsilon \in (0,1)$ and unit vector $\mathbf{u} \in \mathbb{R}^N$ the inequality*

$$P(\big|\|\Phi\mathbf{u}\|^2 - 1\big| \geq \epsilon) \leq p(n,\epsilon)$$

holds for $n \geq 2/\epsilon$. (Note that the result doesn't actually depend on N.)

Proof: Fix $\epsilon \in (0,1)$ and let Φ be an $n \times N$ matrix that satisfies the hypotheses of the lemma. Let $\mathbf{u}$ be any fixed unit vector in $\mathbb{R}^N$ and set $\mathbf{y} = \Phi\mathbf{u}$, so $\mathbf{y} \in \mathbb{R}^n$. Recall from elementary probability that if $X_1, \ldots, X_N$ are independent normal random variables, all with mean μ and variance σ^2, then $X = \sum_j c_j X_j$ has mean $\mu \sum_j c_j$ and variance $\sigma^2 \sum_j c_j^2$. It follows that each component $y_i = \sum_{j=1}^N \phi_{ij} u_j$ of $\mathbf{y}$ is an independent normal random variable with mean 0 and variance $1/n$. Therefore $n\|\mathbf{y}\|_2^2 = \sum_{i=1}^n n y_i^2$ is a χ^2 variable with n degrees of freedom and probability density function (pdf)

$$g_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

for $x \geq 0$ (see Section 3.3 of [29]). Here Γ is the *Gamma function*, defined for any real number $\alpha > 0$ as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx. \tag{3.3}$$

For positive integers n, $\Gamma(n) = (n-1)!$, $\Gamma(1/2) = \sqrt{\pi}$, and $\Gamma(n+1/2) = \frac{(2n)!}{4^n n!}\sqrt{\pi}$.

The pdf of $\|\mathbf{y}\|_2^2$ itself is given by

$$f_n(x) = n g_n(nx) = \frac{(n/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-nx/2}. \tag{3.4}$$

The mean and variance of $\|\mathbf{y}\|_2^2$ are 1 and $2/n$ (Exercise 18), respectively, so as n increases the quantity $\|\mathbf{y}\|_2^2$ is more and more strongly "concentrated" near 1 (see Figure 3.1). However, to prove the lemma we need to quantify the last statement. Specifically, we have

$$P(\big|\|\Phi\mathbf{u}\|^2 - 1\big| \geq \epsilon) = \int_0^{1-\epsilon} f_n(x)\, dx + \int_{1+\epsilon}^\infty f_n(x)\, dx. \tag{3.5}$$

We will estimate the value of each integral on the right in (3.5). To simplify the notation, we use $\alpha = n/2$. Note that we are only interested in the case $\alpha \geq 1/\epsilon > 1$.

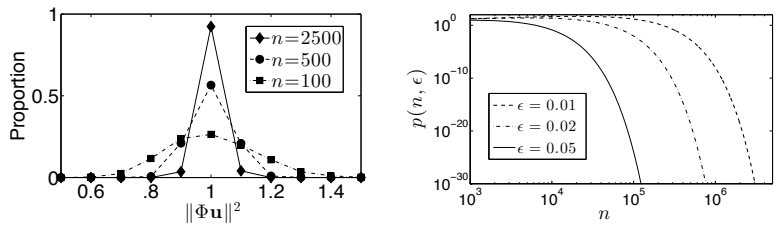

FIG. 3.1. *Illustration of Lemma 3.2. Left: A histogram with bin size 0.1 (markers indicate center of each bin) showing the proportion out of 4,000 normal random matrices* Φ *of size* $n \times N$ *for which* $\|\Phi\mathbf{u}\|_2^2$ *falls into each bin of width 0.1, where* $\mathbf{u}$ *is a fixed randomly generated unit vector and* $N = 10,000$*. The marker corresponding to interval* $[0.95, 1.05]$ *shows the proportion of matrices* Φ *that satisfied* $|\|\Phi\mathbf{u}\|_2^2 - 1| \le \epsilon = 0.05$*: 26% for* $n = 100$*, 57% for* $n = 500$*, and 92% for* $n = 2,500$*. Right: Graphs of the upper bound* $p(n,\epsilon)$ *for various values of* ϵ*.*

We begin with the first integral on the right in (3.5). For any $\alpha \ge 1$, $x^{\alpha-1}e^{-\alpha x}$ attains its maximum value on the interval $[0,\infty)$ when $x = 1 - 1/\alpha \ge 1 - \epsilon$, and is strictly increasing on the interval $(0, 1 - 1/\alpha)$. Thus $x^{\alpha-1}e^{-\alpha x}$ attains its maximum value on $[0, 1-\epsilon]$ at $x = 1-\epsilon$, and this maximum value is $M = (1-\epsilon)^{\alpha-1}e^{-\alpha(1-\epsilon)}$. It follows that for $\alpha \ge 1/\epsilon$ we have

$$\int_0^{1-\epsilon} x^{\alpha-1}e^{-\alpha x}\,dx \le (1-\epsilon)M = (1-\epsilon)^{\alpha}e^{-\alpha(1-\epsilon)}. \tag{3.6}$$

From (3.4) and by using (3.6) with $\alpha = n/2$ we then have, for $n \ge 2/\epsilon$,

$$\int_0^{1-\epsilon} f_n(x)\,dx \le \frac{(n/2)^{n/2}}{\Gamma(n/2)}e^{-n(1-\epsilon)/2}(1-\epsilon)^{n/2}. \tag{3.7}$$

Applying the bound (see [5])

$$\frac{\alpha^\alpha}{\Gamma(\alpha)} \le \frac{e^\alpha\sqrt{\alpha}}{\sqrt{2\pi}} \tag{3.8}$$

with $\alpha = n/2 \ge 1$ in (3.7) yields

$$\int_0^{1-\epsilon} f_n(x)\,dx \le \frac{\sqrt{n}}{2\sqrt{\pi}}e^{n\epsilon/2}(1-\epsilon)^{n/2}. \tag{3.9}$$

From Exercise 20(d) we have $(1-\epsilon)^{1/\epsilon} \le e^{-1-\epsilon/2}$ for any $\epsilon \in (0,1)$. As a result

$$(1-\epsilon)^\alpha = ((1-\epsilon)^{1/\epsilon})^{\alpha\epsilon} \le e^{(-1-\epsilon/2)\alpha\epsilon} = e^{-\alpha\epsilon}e^{-\alpha\epsilon^2/2}. \tag{3.10}$$

Applying the estimate (3.10) to the bound (3.9) yields

$$\int_0^{1-\epsilon} f_n(x)\,dx \le \frac{\sqrt{n}}{2\sqrt{\pi}}e^{-n\epsilon^2/4} \tag{3.11}$$

for $n \ge 2/\epsilon$. This gives us the bound we need on the first integral in (3.5).

Now we bound the second integral on the right in (3.5). From Exercise 21,

$$x^{n/2-1}e^{-nx/2} \le \left(\frac{1+\epsilon}{e}\right)^{n/2-1} e^{-\frac{1+n\epsilon/2}{1+\epsilon}x} \tag{3.12}$$

for $n \geq 2$ and $\epsilon > -1$. Integrate each side of (3.12) from $x = 1+\epsilon$ to $x = \infty$ to obtain

$$\int_{1+\epsilon}^{\infty} x^{n/2-1} e^{-nx/2}\, dx \leq (1+\epsilon)^{n/2} e^{-\frac{n}{2}(1+\epsilon)}/(1+n\epsilon/2). \tag{3.13}$$

From Exercise 20(c) we have $(1+\epsilon)^{1/\epsilon} \leq e^{1-\epsilon/2+\epsilon^2/3}$ for $\epsilon \in (0,1)$, so that from (3.13) and imitating (3.10) we obtain

$$\int_{1+\epsilon}^{\infty} x^{n/2-1} e^{-nx/2}\, dx \leq \frac{e^{-n/2-n\epsilon^2/4+n\epsilon^3/6}}{1+n\epsilon/2}. \tag{3.14}$$

Use $\alpha = n/2$ in the bound (3.8) with (3.14) to find

$$\int_{1+\epsilon}^{\infty} f_n(x)\, dx \leq \frac{\sqrt{n} e^{-n\epsilon^2/4+n\epsilon^3/6}}{(2+n\epsilon)\sqrt{\pi}}. \tag{3.15}$$

Combining equations (3.11) and (3.15) yields the estimate in the lemma (noting that $e^{n\epsilon^3/6} > 1$ and $2 + n\epsilon > 2$).□

Exercise 18. Let $\mathbf{u}$ be a unit vector in $\mathbb{R}^N$ and Φ an $n \times N$ matrix whose entries are sampled from independent normal random variables with mean 0 and variance $1/n$. Prove that the expected value of $\|\Phi\mathbf{u}\|_2^2$ equals 1 and the variance equals $2/n$.

Exercise 19. Let $\epsilon = 0.1$ in the bound on the right in (3.5). Compute the value of the function $p(n,\epsilon)$ in Lemma 3.2 (p as defined in (3.2)) for $n = 10, 10^2, 10^3, 10^4$. How large must n be before the bound on $P(\big|\|\Phi\mathbf{u}\|^2 - 1\big| \geq \epsilon)$ is below 10^{-12}? Repeat for $\epsilon = 0.01$.

Exercise 20. Prove that $(1+\epsilon)^{1/\epsilon} \leq e^{1-\epsilon/2+\epsilon^2/3}$ for all $\epsilon \in (0,1)$, as follows:

a. Show that

$$1 - \epsilon/2 + \epsilon^2/3 - \epsilon^3/4 + \epsilon^4/5 - \cdots \leq 1 - \epsilon/2 + \epsilon^2/3.$$

for $\epsilon \in (0,1)$. (Look at pairs of terms in the alternating series.)

b. Find the Taylor series for $f(x) = \ln(1+x)$ about $x = 0$ and use this to show that the left side of the displayed inequality in part (a) is the Taylor series for the real-analytic function g defined by

$$g(\epsilon) = \begin{cases} \ln(1+\epsilon)/\epsilon, & \epsilon \neq 0 \\ 1, & \epsilon = 0 \end{cases}$$

c. Exponentiate $g(\epsilon) \leq 1 - \epsilon/2 + \epsilon^2/3$ to obtain the desired inequality.

d. Imitate (a)-(c) to show that $(1-\epsilon)^{1/\epsilon} \leq e^{-1-\epsilon/2}$ for all $\epsilon \in (0,1)$ by using the Taylor series for

$$\tilde{g}(\epsilon) = \begin{cases} \ln(1-\epsilon)/\epsilon, & \epsilon \neq 0 \\ 1, & \epsilon = 0 \end{cases}$$

Exercise 21. Prove that if $\alpha \geq 1$ and $\epsilon > -1$ then

$$x^{\alpha-1} e^{-\alpha x} \leq \left(\frac{1+\epsilon}{e}\right)^{\alpha-1} e^{-\frac{1+\alpha\epsilon}{1+\epsilon}x} \tag{3.16}$$

for all $x \geq 0$. Hint: define the function

$$g(x) = \left(\frac{1+\epsilon}{e}\right)^{\alpha-1} x^{\alpha-1} e^{((1+\alpha\epsilon)/(1+\epsilon)-\alpha)x}$$

($g(x)$ is just the left side of (3.16) divided by the right side). Then show that $g(x) > 0$ for $x > 0$, that

$$\lim_{x\to 0^+} g(x) = \lim_{x\to\infty} g(x) = 0,$$

and that g has a unique critical point (a maximum) at $x = 1+\epsilon$, with value $g(1+\epsilon) = 1$.

3.2. Extending Lemma 3.2. Lemma 3.2 holds only for a fixed unit vector $\mathbf{u}$, but we need a version of that lemma's inequality that holds simultaneously for all m-sparse unit vectors, with high probability. The argument that follows—in particular, Lemmas 3.3 and 3.5—are adaptations of similar arguments in [4].

Consider a subset $T \subset \{1, 2, \ldots, N\}$ with $|T| = m$. For any such subset $T = \{T_1, \ldots, T_m\}$ let X_T denote the set of all vectors $\mathbf{x} \in \mathbb{R}^N$ that are m-sparse, with $x_i = 0$ if $i \notin T$. Let U_T denote the set of all *unit* vectors in X_T. Our goal in what follows is to extend the central inequality of Lemma 3.2 so it holds for all m-sparse unit vectors simultaneously, with high probability. We'll do this by first showing that for any fixed T the inequality holds with high probability on a finite subset $A_Q \subset U_T$ (Lemma 3.3) and then extend the result to hold for all vectors $\mathbf{u} \in U_T$ (Lemma 3.5). We then prove Theorem 3.1 by extending the result to hold for all such sets T (simultaneously), and thereby all m-sparse vectors.

In Lemmas 3.3 and 3.5 we assume that Φ is an $n \times N$ matrix whose entries are independent samples of a normal random variable with mean 0 and variance $1/n$.

LEMMA 3.3. *For any fixed subset $T \subset \{1, 2, \ldots, N\}$ with $|T| = m$, $\delta \in (0, 1)$, any subset $A_Q \subset U_T$ with cardinality $Q = |A_Q| < \infty$, and $n \geq 2/\delta$, the inequality*

$$1 - \delta \leq \|\Phi\mathbf{u}\|_2^2 \leq 1 + \delta$$

holds simultaneously for all $\mathbf{u} \in A_Q$ with probability greater than $1 - Qp(n, \delta)$.

Proof: In Lemma 3.2 let's choose $\epsilon = \delta$, so that for any $\delta \in (0, 1)$ and fixed vector $\mathbf{p}_k \in A_Q$ we have that $\left|\|\Phi\mathbf{p}_k\|_2^2 - 1\right| \geq \delta$ holds with probability less than $p(n, \delta)$, defined in (3.2), when $n \geq 2/\delta$. Equivalently, the inequality

$$1 - \delta \leq \|\Phi\mathbf{p}_k\|_2^2 \leq 1 + \delta \tag{3.17}$$

holds with probability at least $1 - p(n, \delta)$, or fails to hold with probability at most $p(n, \delta)$.

Recall Boole's inequality, also called the "union bound" [17]:

$$P(E_1 \cup E_2 \cup \cdots \cup E_Q) \leq P(E_1) + \cdots + P(E_Q), \tag{3.18}$$

where the E_k are any events, which need not be independent. Let E_k denote the event that the inequality (3.17) fails to hold for the point $\mathbf{p}_k$, where $1 \leq k \leq Q$. We conclude from (3.18) that the probability of (3.17) failing to hold for at least one of the $\mathbf{p}_k$ is less than $Qp(n, \delta)$. This means that the probability that (3.17) does in fact hold *simultaneously* for all $\mathbf{p}_k \in A_Q$ is greater than or equal to $1 - Qp(n, \delta)$, which is exactly the probability specified in the lemma.□

The following lemma tells us how to obtain a specific suitable subset $A_Q \subset U_T$ for our purposes, through a bound for the covering number of the unit sphere [34]:

LEMMA 3.4. *For each $\epsilon \in (0,1)$ and positive integer $m \geq 2$, there exists a subset $A_Q(\epsilon)$ of the unit sphere $S^{m-1} = \{\mathbf{u} \in \mathbb{R}^m : \|\mathbf{u}\|_2 = 1\}$ with at most $Q(\epsilon) := (3/\epsilon)^m$ points that satisfies the following property: for every $\mathbf{u} \in S^{m-1}$ there exists $\mathbf{p} \in A_Q(\epsilon)$ such that $\|\mathbf{u} - \mathbf{p}\|_2 < \epsilon$.*

Proof: Fix $\epsilon \in (0,1)$ and an integer $m \geq 2$. Inductively construct the set A_Q as follows: Begin by choosing any $\mathbf{p}_1 \in S^{m-1}$. After choosing $\mathbf{p}_k$, choose $\mathbf{p}_{k+1} \in S^{m-1}$ such that $\|\mathbf{p}_{k+1} - \mathbf{p}_j\|_2 \geq \epsilon$ for $j = 1, \ldots, k$. Continue selecting points until no further such points can be found; the process will stop at some finite number Q of points because S^{m-1} is compact (every open cover has a finite subcover). Define $A_Q = \{\mathbf{p}_1, \ldots, \mathbf{p}_Q\}$. We claim that for all $\mathbf{u} \in S^{m-1}$, there exists $\mathbf{p}_k \in A_Q$ such that $\|\mathbf{u} - \mathbf{p}_k\| < \epsilon$; otherwise, there would exist $\mathbf{p}_{Q+1} \in S^{m-1}$ such that $\|\mathbf{p}_{Q+1} - \mathbf{p}_k\| \geq \epsilon$ for $k = 1, \ldots, Q$, contradicting the stopping condition. Balls $B_{\epsilon/2}(\mathbf{p}_k)$ of radius $\epsilon/2$ centered at the points in A_Q are disjoint, while the union $\cup_{k=1}^{Q} B_{\epsilon/2}(\mathbf{p}_k)$ lies in the ball $B_{1+\epsilon/2}(\mathbf{0})$. The total volume of the union must be less than the volume of the containing ball, so $Q \cdot \mathrm{Vol}(B_{\epsilon/2}) \leq \mathrm{Vol}(B_{1+\epsilon/2})$. Since $\mathrm{Vol}(B_r) = r^m \mathrm{Vol}(B_1)$ in $\mathbb{R}^m$, $Q \leq (1 + 2/\epsilon)^m = ((\epsilon + 2)/\epsilon)^m \leq (3/\epsilon)^m$. $\square$

Lemma 3.3 asserts that $1 - \delta \leq \|\Phi\mathbf{u}\|_2^2 \leq 1 + \delta$ holds with high probability for all $\mathbf{u} \in A_Q$. The next lemma extends this to all $\mathbf{u} \in U_T$ by using the set $A_Q = A_Q(\epsilon)$ from Lemma 3.4 with $\epsilon = \delta/4$ (but note the norm in (3.19) below is not squared).

LEMMA 3.5. *For any fixed subset $T \subset \{1, 2, \ldots, N\}$ with $|T| = m$, $\delta \in (0,1)$, and $n \geq 4/\delta$, the inequality*

$$1 - \delta \leq \|\Phi\mathbf{u}\|_2 \leq 1 + \delta \tag{3.19}$$

holds simultaneously for all $\mathbf{u} \in U_T$ with probability greater than $\tilde{p} = 1 - Q(\delta/4)p(n, \delta/2)$.
Proof: For a fixed Φ define a constant B (which will be a random variable that depends on Φ) as

$$B = -1 + \sup_{\mathbf{u} \in U_T} \|\Phi\mathbf{u}\|_2. \tag{3.20}$$

Actually, since U_T is compact in $\mathbb{R}^N$ and the mapping $\mathbf{x} \to \|\Phi\mathbf{x}\|_2$ is continuous (see Exercises 14 and 15), we can replace "sup" with "max." This also makes it clear that $B < \infty$. From (3.20) we obviously have $\|\Phi\mathbf{u}\|_2 \leq 1 + B$ for any $\mathbf{u} \in U_T$.

If $\mathbf{x} \in X_T$ with $\mathbf{x} \neq \mathbf{0}$, then $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|_2 \in U_T$. The inequality $\|\Phi\mathbf{u}\|_2 \leq 1 + B$ is equivalent to the inequality

$$\|\Phi\mathbf{x}\|_2 \leq (1 + B)\|\mathbf{x}\|_2. \tag{3.21}$$

We will show that $B \leq \delta$ with probability at least $\tilde{p}$, yielding the right-side inequality in (3.19).

To do this, let $\mathbf{u} \in U_T$. Fix a set $A_Q(\delta/4)$ as in Lemma 3.4 and choose a point $\mathbf{p}_k \in A_Q(\delta/4)$ so that $\|\mathbf{u} - \mathbf{p}_k\|_2 \leq \delta/4$. The vector $\mathbf{u} - \mathbf{p}_k \in X_T$, so the bound (3.21) applies to show $\|\Phi(\mathbf{u} - \mathbf{p}_k)\|_2 \leq (1 + B)\delta/4$. Applying Lemma 3.3 (using $\delta/2$ in place of δ), we have $\|\Phi\mathbf{p}_k\|_2 \leq \sqrt{1 + \delta/2} \leq (1 + \delta/2)$ with probability at least $\tilde{p}$ for $n \geq 4/\delta$.

As a result, with probability at least $\tilde{p}$ we have

$$\begin{aligned}\|\Phi\mathbf{u}\|_2 &= \|\Phi\mathbf{p}_k + \Phi(\mathbf{u}-\mathbf{p}_k)\|_2 \\ &\le \|\Phi\mathbf{p}_k\|_2 + \|\Phi(\mathbf{u}-\mathbf{p}_k)\|_2 \text{ (by the triangle inequality)} \\ &\le (1+\delta/2)\|\mathbf{p}_k\|_2 + (1+B)\delta/4 \text{ (using (3.21))} \\ &\le 1+\delta/2+(1+B)\delta/4 \text{ (because } \|\mathbf{p}_k\|_2 = 1) \\ &= 1+(3/4+B/4)\delta. \end{aligned} \tag{3.22}$$

Equivalently, we have $\|\Phi\mathbf{u}\|_2 > 1+(3/4+B/4)\delta$ with probability less than $1-\tilde{p}$. Since $\|\Phi\mathbf{u}\|_2 \le 1+B$ we conclude that $1+(3/4+B/4)\delta < 1+B$ with probability less than $1-\tilde{p}$, or equivalently, $1+B \le 1+(3/4+B/4)\delta$ with probability at least $\tilde{p}$. A little rearrangement yields $B \le \frac{3\delta}{4-\delta}$ for $\delta \in (0,1)$ with probability at least $\tilde{p}$; since $\frac{3\delta}{4-\delta} \le \delta$ for $\delta \in (0,1)$, we obtain $B \le \delta$ with probability at least $\tilde{p}$, which verifies the right side inequality in (3.19).

To demonstrate the left side inequality in (3.19) start with

$$\|\Phi\mathbf{p}_k\|_2 \le \|\Phi(\mathbf{p}_k-\mathbf{u})\|_2 + \|\Phi\mathbf{u}\|_2$$

from the triangle inequality. Rearrange to obtain $\|\Phi\mathbf{u}\|_2 \ge \|\Phi\mathbf{p}_k\|_2 - \|\Phi(\mathbf{p}_k-\mathbf{u})\|_2$. Then again from Lemma 3.3 (using $\delta/2$ in place of δ), with probability at least $\tilde{p}$ we have

$$\begin{aligned}\|\Phi\mathbf{u}\|_2 &\ge \|\Phi\mathbf{p}_k\|_2 - \|\Phi(\mathbf{p}_k-\mathbf{u})\|_2 \\ &\ge (1-\delta/2) - (1+B)\|\mathbf{p}_k-\mathbf{u}\|_2 \\ &\ge (1-\delta/2) - (1+B)\frac{\delta}{4} \\ &\ge (1-\delta/2) - (1+\delta)\frac{\delta}{4} \\ &= 1-\frac{3}{4}\delta - \frac{1}{4}\delta^2, \end{aligned} \tag{3.23}$$

where we've applied the bound (3.21) to $\mathbf{p}_k - \mathbf{u}$ and used $\|\mathbf{p}_k\|_2 = 1$ and $\|\mathbf{p}_k-\mathbf{u}\|_2 \le \delta/4$. It's easy to see that $3\delta/4+\delta^2/4 < \delta$ for $\delta \in (0,1)$, so that $1-\frac{3}{4}\delta-\frac{1}{4}\delta^2 > 1-\delta$. From this and (3.23) we obtain $\|\Phi\mathbf{u}\|_2 \ge 1-\delta$ with probability at least $\tilde{p}$ for $n \ge 4/\delta$, which is the left side inequality in (3.19). □

3.3. Finishing the proof of Theorem 3.1. We can now show that if $n = qN$ (that is, we have N unknowns but want to use a fraction q of the usual requirement of N equations), then we can recover an m-sparse solution with high probability if N is large enough.

Proof of Theorem 3.1: Fix a subset $T \subset \{1,2,\ldots,N\}$ with $|T| = m$ and let $\mathbf{x} \in X_T$, $\mathbf{x} \ne \mathbf{0}$. Applying Lemma 3.5 to the vector $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|_2$ with δ replaced by $\delta/3$ shows that

$$(1-\delta/3)\|\mathbf{x}\|_2 \le \|\Phi\mathbf{x}\|_2 \le (1+\delta/3)\|\mathbf{x}\|_2 \tag{3.24}$$

holds with probability greater than $1-Q(\delta/12)p(n,\delta/6)$, that is, fails with probability less than $Q(\delta/12)p(n,\delta/6)$, assuming now that $n \ge 12/\delta$. If we square each entry in (3.25) and note that $1-\delta \le (1-\delta/3)^2$ and $(1+\delta/3)^2 \le 1+\delta$ for $\delta \in (0,1)$ we obtain

$$(1-\delta)\|\mathbf{x}\|_2^2 \le \|\Phi\mathbf{x}\|_2^2 \le (1+\delta)\|\mathbf{x}\|_2^2, \tag{3.25}$$

again, with probability greater than $1 - Q(\delta/12)p(n, \delta/6)$.

Now consider the set of all subsets $T \in \{1, 2, \ldots, N\}$ with $|T| = m$; there are $\binom{N}{m}$ such subsets. If we let E_T denote the event that equation (3.25) *fails* to hold on X_T then $P(E_T) \leq Q(\delta/12)p(n, \delta/6)$ as remarked above. We again employ Boole's inequality (3.18) to conclude that the probability P_0 of equation (3.25) failing on at least one of the sets T is bounded as follows:

$$P_0 \leq \binom{N}{m} Q(\delta/12)p(n, \delta/6) \leq \left(\frac{36eN}{m\delta}\right)^m \sqrt{\frac{n}{\pi}} e^{-n(\delta^2/144 - \delta^3/1296)}, \tag{3.26}$$

where we used the classic bound $\binom{N}{m} \leq \left(\frac{eN}{m}\right)^m$ (see Exercise 22). Finally, let's fix $n = qN$ for some $q \in (0, 1)$. In this case the right side of (3.26) yields

$$P_0 \leq N^{m+1/2} \left(\frac{36e}{m\delta}\right)^m \sqrt{\frac{q}{\pi}} e^{-qN(\delta^2/144 - \delta^3/1296)}. \tag{3.27}$$

For any fixed choice of $\delta \in (0, 1)$, $q \in (0, 1)$, and $m \geq 1$ the right side of (3.27) goes to 0 as $N \to \infty$, which shows that the probability of (3.25) failing can be made arbitrarily small by taking N sufficiently large. In fact, if we choose $\epsilon \in (0, 1)$ and take the log of both sides of the inequality $N^{m+1/2} \left(\frac{36e}{m\delta}\right)^m \sqrt{\frac{q}{\pi}} e^{-qN(\delta^2/144 - \delta^3/1296)} \leq \epsilon$, we obtain the inequality (3.1). □

Exercise 22. Show that $\binom{N}{m} \leq \left(\frac{eN}{m}\right)^m$. Suggestion: observe that $\binom{N}{m} \leq \frac{N^m}{m!}$ and $\log(m!) = \sum_{x=1}^m \log x \geq \int_1^m \log x dx$.

4. Finding Sparse Solutions with ℓ^1 Minimization. We now turn to a more quantitative analysis of why ℓ^1 minimization successfully recovers sparse signals. Let's start with the simple case of a nonzero 1-sparse vector $\mathbf{x}^*$ that satisfies $\Phi\mathbf{x}^* = \mathbf{b}$, where Φ is an $n \times N$ matrix that satisfies the RIP of order 3 with constant δ. If it turns out that $\mathbf{x}^*$ is the unique solution to the optimization problem

$$\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \Phi\mathbf{x} = \mathbf{b}, \tag{4.1}$$

then we have a reasonable way of recovering $\mathbf{x}^*$ from the measurement vector $\mathbf{b}$, for (4.1) can be cast as a standard linear programming problem [16].

The key again lies in considering the properties of the null space $\mathcal{N}(\Phi)$, because all solutions to $\Phi\mathbf{x} = \mathbf{b}$ have the form $\mathbf{x}^* + \eta$ for some $\eta \in \mathcal{N}(\Phi)$. If for each nonzero $\eta \in \mathcal{N}(\Phi)$ the function $q_\eta(t) = \|\mathbf{x}^* + t\eta\|_1$ has a unique global minimum at $t = 0$, then $\mathbf{x}^*$ will be the unique solution of (4.1). To see this, suppose $\Phi\mathbf{x}^{**} = \mathbf{b}$ with $\mathbf{x}^{**} \neq \mathbf{x}^*$. Let $\eta = \mathbf{x}^{**} - \mathbf{x}^* \neq \mathbf{0}$ and note $\eta \in \mathcal{N}(\Phi)$. Then $\|\mathbf{x}^{**}\|_1 = q_\eta(1) > q_\eta(0) = \|\mathbf{x}^*\|_1$.

Let us thus examine conditions under which $q_\eta(t) = \|\mathbf{x}^* + t\eta\|_1$ is guaranteed to have a unique global minimum at $t = 0$. Fix any nonzero $\eta = (\eta_1, \ldots, \eta_N) \in \mathcal{N}(\Phi)$. Without loss of generality, assume that $\mathbf{x}^* = (x_1^*, 0, \ldots, 0)$. If $\eta_1 = 0$, then $q_\eta(t) = |x_1^*| + |t| \sum_{j=2}^N |\eta_j|$ is clearly minimized globally at $t = 0$. Now suppose $\eta_1 \neq 0$, in which case we must examine the function $q_\eta(t) = |x_1^* + t\eta_1| + |t| \sum_{j=2}^N |\eta_j|$, whose graph will look like that shown in Figure 4.1.

There are 2 critical points where a global minimum can occur: $t = 0$ and $t = -x_1^*/\eta_1$. We require $q_\eta(0) < q(-x_1^*/\eta_1)$, which can be rewritten

$$|\eta_1| < \sum_{j=2}^N |\eta_j|. \tag{4.2}$$

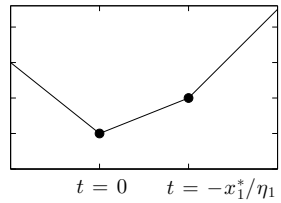

FIG. 4.1. *Graph of function* $q_\eta(t) = |x_1^* + t\eta_1| + |t| \sum_{j=2}^N |\eta_j|$ *for the case* $\eta_1 \neq 0$.

We'll show that if the RIP of order 3 holds for Φ with δ_3 sufficiently small, inequality (4.2) must hold for any $\eta \in \mathcal{N}(\Phi)$. We'll also finally make use of the upper bound part of the RIP. The argument that follows is a distillation of that in [18].

For a fixed $\eta \in \mathcal{N}(\Phi)$ define the set T_1 to consist of the indices of the two components largest in magnitude from the set $\{\eta_2, \ldots, \eta_N\}$, define T_2 to consist of the indices of the next two largest components, and so on through T_s which will contain indices of the smallest one or two elements (depending on whether N is even or odd). Let $T = \{1\} \cup T_1$ and T^c be its complement, $\cup_{j=2}^s T_j$. Let $\eta_{T_j} \in \mathbb{R}^N$ be the vector η but with all components set to zero except those corresponding to indices in T_j. Observe that $\mathbf{0} = \Phi\eta = \Phi\eta_T + \Phi\eta_{T^c}$, so $\Phi\eta_T = -\Phi\eta_{T^c}$. Also, η_T is 3-sparse. Since Φ satisfies the RIP of order 3 for some constant δ (and so also of order 2 by Exercise 12), we have

$$\begin{aligned}
|\eta_1| &\leq \|\eta_T\|_2 && (\text{since } \|\eta_T\|_2^2 = |\eta_1|^2 + \|\eta_{T_1}\|_2^2) \\
&\leq \frac{1}{\sqrt{1-\delta}} \|\Phi\eta_T\|_2 && (\text{using RIP of order 3, left side of (2.4)}) \\
&= \frac{1}{\sqrt{1-\delta}} \|\Phi\eta_{T^c}\|_2 && (\text{since } \|\Phi\eta_T\|_2 = \|\Phi\eta_{T^c}\|_2) \\
&\leq \frac{1}{\sqrt{1-\delta}} \sum_{j=2}^s \|\Phi\eta_{T_j}\|_2 && (\text{by the triangle inequality}) \\
&\leq \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \sum_{j=2}^s \|\eta_{T_j}\|_2 && (\text{using RIP of order 2, right side of (2.4)}) \qquad (4.3)
\end{aligned}$$

If $i \in T_{j+1}$ and $T_j = \{k_1, k_2\}$, then $|\eta_i| \leq |\eta_{k_1}|$ and $|\eta_i| \leq |\eta_{k_2}|$, and so $|\eta_i| \leq \frac{1}{2}(|\eta_{k_1}| + |\eta_{k_2}|) = \frac{1}{2}\|\eta_{T_j}\|_1$. We use this to bound the ℓ^2 norm of $\eta_{T_{j+1}}$ by the ℓ^1 norm of η_{T_j},

$$\|\eta_{T_{j+1}}\|_2 = \sqrt{\eta_{i_1}^2 + \eta_{i_2}^2} \leq \sqrt{\frac{1}{4}\|\eta_{T_j}\|_1^2 + \frac{1}{4}\|\eta_{T_j}\|_1^2} = \frac{1}{\sqrt{2}}\|\eta_{T_j}\|_1. \qquad (4.4)$$

Combining (4.3) and (4.4) yields

$$|\eta_1| \leq \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \sum_{j=2}^s \|\eta_{T_j}\|_2 \leq \frac{\sqrt{1+\delta}}{\sqrt{2(1-\delta)}} \sum_{j=1}^{s-1} \|\eta_{T_j}\|_1 \leq \frac{\sqrt{1+\delta}}{\sqrt{2(1-\delta)}} \sum_{j=2}^N |\eta_j|, \qquad (4.5)$$

which gives us (4.2) if we have $\frac{\sqrt{1+\delta}}{\sqrt{2(1-\delta)}} < 1$, that is, $\delta < \frac{1}{3}$. Therefore ℓ^1 minimization is guaranteed to exactly recover 1-sparse solutions if Φ satisfies the RIP of order 3 with constant $\delta_3 < 1/3$.

The following theorem generalizes this result [18]:

THEOREM 4.1. *Let Φ be an $n \times N$ matrix that satisfies the RIP of order $3k$ with constant $\delta_{3k} < \frac{1}{3}$. Then solving the ℓ^1 minimization problem (4.1) exactly recovers k-sparse signals.*

Indeed, this result can be proved with a straightforward extension of the argument above, based on considering the function $q_\eta(t) = \|\mathbf{x}^* + t\eta\|_1$ where $\mathbf{x}^*$ is a k-sparse solution to $\Phi\mathbf{x}^* = \mathbf{b}$, and defining the sets T_j to consist of $2k$-tuples of indices.

Theorems 3.1 and 4.1 are not sharp—the method works in many cases that don't satisfy the conditions in these theorems. For example, in a simulation using $k = 3$, $q = 0.25$, $\epsilon = 0.05$, and $N = 100$ we successfully recovered 3-sparse solutions in 1000 of 1000 trials with random normal matrices Φ, while Theorem 3.1 suggests we need $N \geq 8.78 \times 10^5$. A practical rule of thumb is that $n \geq 4k$, that is, $q \geq 0.25$, is often sufficient for ℓ^1 minimization to recover a k-sparse signal [12], and much effort has been directed at obtaining sharper theoretical results. For readers interested in numerically exploring some examples, [10] provides helpful Matlab code. See [25] for background on linear programming and [16] for development of basis pursuit as a means of efficient ℓ^1 minimization.

An estimate similar to that of Lemma 3.2 can be derived in the case that the entries of Φ are ± 1 signed Bernoulli variables, and so results analogous to Theorems 3.1 and 4.1 also hold; see [1]. We should point out, however, that for $0-1$ Bernoulli matrices like those used in the coin problem the RIP estimates are not quite as good, as shown in [15] and discussed in [6].

EXERCISE 23. Why didn't we assume Φ satisfies the RIP of order 2 in the simple 1-sparse solution scenario above? Redo the analysis under the assumption of the RIP of order 2, with $|T_j| = 1$, and show that it requires $\delta > 1$.

5. Beyond Sparse Signal Recovery. Actual signals are rarely sparse, as we've been assuming. As a generalization we say that $\mathbf{x}$ is "k-compressible" if $\mathbf{x}$ has k components that are "much larger" in magnitude than the remaining $N - k$ components. This might be the case in the coin problem if most of the coins are "acceptable," say lie within a small error tolerance ± 0.001 grams, but there are a few really off-mass coins that differ significantly from nominal. Many signals arising in applications are compressible in this sense, for example, photos taken by a digital camera. The CS algorithm recovers such signals almost as well as simply keeping the largest components–but without doing the extensive sensing that would be necessary to identify all of the components in order to determine which are the largest [18].

The CS algorithm can also be adapted to work for noisy signals, in which case the ℓ^1 minimization problem becomes

$$\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\Phi\mathbf{x} - \mathbf{b}\|_2 \leq \epsilon,$$

where ϵ bounds the amount of noise. This still produces good estimates in a computationally efficient manner and the problem remains stable, in that small errors in $\mathbf{b}$ have relatively little effect on the solutions produced by our algorithm [14].

The applications of CS are continually expanding, including medical imaging, communications, analog-to-information conversion, geophysical data analysis, compressive radar, and genetic screening, among others. For example, CS can greatly

reduce scan times and potentially increase resolution of magnetic resonance imaging [30]. CS microarrays combine group testing and CS principles to accurately identify genetic sequences, e.g, to detect pathogens in a water sample [19]. CS has also been used to image the rupture process of a main shock in an earthquake [35]. Many other applications are being developed (see http://dsp.rice.edu/cs). All in all, CS is proving to be a powerful and flexible paradigm and will continue to be an exciting field of mathematical research that extends into many other disciplines.

REFERENCES

[1] D. Achlioptas, *Database-friendly random projections: Johnson-Lindenstrauss with binary coins*, Journal of Computer and System Sciences 66 (2003) 671-687.

[2] L. Balzano, R. Nowak and J. Ellenberg, *Compressed sensing audio demonstration*, http://sunbeam.ece.wisc.edu/csaudio/ (accessed 5/21/11).

[3] A. Amini and F. Marvasti, *Deterministic Construction of Binary, Bipolar, and Ternary Compressed Sensing Matrices*, IEEE Trans. Inf. Theory 57 (2011), 2360-2370.

[4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, *A simple proof of the restricted isometry property for random matrices*, Constructive Approximation 28 (2008), 253-263.

[5] N. Batir, *Inequalities for the gamma function*, Arch. Math. 91 (2008), 554-563.

[6] S. R. Becker, "Practical compressed sensing : modern data acquisition and signal processing," Ph.D. dissertation, California Institute of Technology, 2011. `http://resolver.caltech.edu/CaltechTHESIS:06022011-152525054` (accessed 3/21/12).

[7] E. J. Candès. *The restricted isometry property and its implications for compressed sensing*, Compte Rendus de l'Academie des Sciences, Paris, Serie I, 346 (2008) 589-592.

[8] E. J. Candès and J. Romberg. *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. of Comput. Math. (2006) 6:227-254.

[9] E. Candès and J. Romberg, *Sparsity and incoherence in compressive sampling*, Inverse Problems 23 (2007), 969-985.

[10] E. Candès and J. Romberg, ℓ_1*-MAGIC*, http://www.acm.caltech.edu/l1magic/ (accessed 5/21/11).

[11] E. J. Candès, J. Romberg and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory 52 (2006),489-509.

[12] E. Candès and T. Tao, *Decoding by linear programing*, IEEE Trans. Inf. Theory 51 (2005), 4203-4215.

[13] E. Candès and T. Tao, *Near optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inf. Theory 52 (2006), 5406-5425.

[14] E. J. Candès and M.B. Wakin, *An introduction to compressive sampling*, IEEE Signal Processing Magazine 25 (2008), 21-30.

[15] V. Chandar, *A negative result concerning explicit matrices with the restricted isometry property*, preprint, 2008, available at `http://dsp.rice.edu/cs` (accessed 3/20/2012).

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Review 43:129-159.

[17] K. L. Chung and F. Aitsahlia, "Elementary Probability Theory," 4th ed. Springer (2003).

[18] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best k-term approximation*, Journal of the American Mathematical Society 22 (2009), 211-231.

[19] W. Dai, M. A. Sheikh, O. Milenkovic, and R. Baraniuk, *Compressive Sensing DNA Microarrays*, EURASIP Journal on Bioinformatics Systems Biology 24 (2009), 162824.

[20] R. A. DeVore, *Deterministic constructions of compressed sensing matrices*, Journal of Complexity 23 (2007) 918-925.

[21] D. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory 52 (2006), 1289-1306.

[22] D. Donoho, *For most large underdetermined systems of linear equations, the minimal ℓ_1- norm solution is also the sparsest solution*, Communications on Pure and Applied Mathematics, 59 (2006), 797-829.

[23] D. L. Donoho and P. B. Stark, *Uncertainty principles and signal recovery*, SIAM J. on Appl. Math. 49 (1989), 906-932.

[24] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, *Single-pixel imaging via compressive sampling*, IEEE Signal Processing Mag. 25 (2008), 83-91.

[25] M. C. Ferris, O. L. Mangasarian, and S. J. Wright, *Linear Programming with Matlab*, SIAM 2007.

[26] M. Fornasier and H. Rauhut, *Compressive sensing*, in Handbook of Mathematical Methods in Imaging, ed O. Scherzer, Springer (2011) pp. 187-228.

[27] A. Gilbert and P. Indyk, *Sparse recovery using sparse matrices*, Proceedings of the IEEE 98 (2010), 937-947.

[28] B. Hayes, *The best bits*, American Scientist 97 (2009), 276-280.

[29] R. V. Hogg, A. Craig, and J. W. McKean, *Introduction to Mathematical Statistics 6/E*, Pearson (2005).

[30] M. Lustig, D. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*. Magnetic Resonance in Medicine 58(2007), 1182 - 1195.

[31] D. MacKenzie, *Compressed sensing makes every pixel count*, in Whats Happening in the Mathematical Sciences Vol. 7, Amer. Math. Soc. (2009) pp. 114-127.

[32] C. Moler, *"Magic" reconstruction: Compressed sensing*, MathWorks News & Notes (2010), http://www.mathworks.com/company/newsletters/index.html (accessed 5/20/2011).

[33] J. Romberg, *Imaging via compressive sampling*, IEEE Signal Processing Mag. 25 (2008),14-20.

[34] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, in Compressed Sensing: Theory and Applications, eds Y. Eldar and G. Kutynick, Cambridge University Press (2012).

[35] H. Yao, P. Gerstoft, P. M. Shearer, and C. Mecklenbrauker,*Compressive sensing of the Tohoku-Oki Mw 9.0 earthquake: Frequency-dependent rupture modes*, Geophys. Res. Lett. 38(2011), L20310.