The CS theory tells us that when certain conditions hold, namely that the functions $\{\phi_m\}$ cannot sparsely represent the elements of the basis $\{\psi_n\}$ (a condition known as *incoherence* of the two dictionaries [20–22, 91]) and the number of measurements M is large enough, then it is indeed possible to recover the set of large $\{\alpha(n)\}$ (and thus the signal x) from a similarly sized set of measurements y. This incoherence property holds for many pairs of bases, including for example, delta spikes and the sine waves of a Fourier basis, or the Fourier basis and wavelets. Significantly, this incoherence also holds with high probability between an arbitrary fixed basis and a randomly generated one.

2.8.3 Methods for signal recovery

Although the problem of recovering x from y is ill-posed in general (because $x \in \mathbb{R}^N$, $y \in \mathbb{R}^M$, and $M < N$), it is indeed possible to recover *sparse* signals from CS measurements. Given the measurements $y = \Phi x$, there exist an infinite number of candidate signals in the shifted nullspace $\mathcal{N}(\Phi) + x$ that could generate the same measurements y (see Section 2.4.1). Recovery of the correct signal x can be accomplished by seeking a *sparse* solution among these candidates.

Recovery via ℓ_0 optimization

Supposing that x is exactly K-sparse in the dictionary Ψ, then recovery of x from y can be formulated as the ℓ_0 minimization

$$\widehat{\alpha} = \arg\min \|\alpha\|_0 \quad \text{s.t. } y = \Phi\Psi\alpha. \tag{2.9}$$

Given some technical conditions on Φ and Ψ (see Theorem 2.1 below), then with high probability this optimization problem returns the proper K-sparse solution α, from which the true x may be constructed. (Thanks to the incoherence between the two bases, if the original signal is sparse in the α coefficients, then no other set of sparse signal coefficients α' can yield the same projections y.) We note that the recovery program (2.9) can be interpreted as finding a K-term approximation to y from the columns of the dictionary $\Phi\Psi$, which we call the *holographic basis* because of the complex pattern in which it encodes the sparse signal coefficients [21].

In principle, remarkably few incoherent measurements are required to recover a K-sparse signal via ℓ_0 minimization. Clearly, more than K measurements must be taken to avoid ambiguity; the following theorem establishes that $K + 1$ random measurements will suffice. (Similar results were established by Venkataramani and Bresler [92].)

Theorem 2.1 *Let Ψ be an orthonormal basis for $\mathbb{R}^N$, and let $1 \leq K < N$. Then the following statements hold:*

1. *Let Φ be an $M \times N$ measurement matrix with i.i.d. Gaussian entries with $M \geq 2K$. Then with probability one the following statement holds: all signals $x = \Psi\alpha$ having expansion coefficients $\alpha \in \mathbb{R}^N$ that satisfy $\|\alpha\|_0 = K$ can be recovered uniquely from the M-dimensional measurement vector $y = \Phi x$ via the ℓ_0 optimization (2.9).*

2. *Let $x = \Psi\alpha$ such that $\|\alpha\|_0 = K$. Let Φ be an $M \times N$ measurement matrix with i.i.d. Gaussian entries (notably, independent of x) with $M \geq K+1$. Then with probability one the following statement holds: x can be recovered uniquely from the M-dimensional measurement vector $y = \Phi x$ via the ℓ_0 optimization (2.9).*

3. *Let Φ be an $M \times N$ measurement matrix, where $M \leq K$. Then, aside from pathological cases (specified in the proof), no signal $x = \Psi\alpha$ with $\|\alpha\|_0 = K$ can be uniquely recovered from the M-dimensional measurement vector $y = \Phi x$.*

Proof: See Appendix A.

The second statement of the theorem differs from the first in the following respect: when $K < M < 2K$, there will necessarily exist K-sparse signals x that cannot be uniquely recovered from the M-dimensional measurement vector $y = \Phi x$. However, these signals form a set of measure zero within the set of *all* K-sparse signals and can safely be avoided if Φ is randomly generated independently of x.

Unfortunately, as discussed in Section 2.5.2, solving this ℓ_0 optimization problem is prohibitively complex. Yet another challenge is robustness; in the setting of Theorem 2.1, the recovery may be very poorly conditioned. In fact, *both* of these considerations (computational complexity and robustness) can be addressed, but at the expense of slightly more measurements.

Recovery via ℓ_1 optimization

The practical revelation that supports the new CS theory is that it is not necessary to solve the ℓ_0-minimization problem to recover α. In fact, a much easier problem yields an equivalent solution (thanks again to the incoherency of the bases); we need only solve for the ℓ_1-sparsest coefficients α that agree with the measurements y [20–22, 24–27, 29]

$$\widehat{\alpha} = \arg\min \|\alpha\|_1 \quad \text{s.t. } y = \Phi\Psi\alpha. \tag{2.10}$$

As discussed in Section 2.5.2, this optimization problem, also known as *Basis Pursuit* [80], is significantly more approachable and can be solved with traditional linear programming techniques whose computational complexities are polynomial in N.

There is no free lunch, however; according to the theory, more than $K+1$ measurements are required in order to recover sparse signals via Basis Pursuit. Instead, one typically requires $M \geq cK$ measurements, where $c > 1$ is an *oversampling factor*. As an example, we quote a result asymptotic in N. For simplicity, we assume that

Appendix A
Proof of Theorem 2.1

We first prove Statement 2, followed by Statements 1 and 3.

Statement 2 (Achievable, $M \geq K+1$): Since Ψ is an orthonormal basis, it follows that entries of the $M \times N$ matrix $\Phi\Psi$ will be i.i.d. Gaussian. Thus without loss of generality, we assume Ψ to be the identity, $\Psi = I_N$, and so $y = \Phi\alpha$. We concentrate on the "most difficult" case where $M = K+1$; other cases follow similarly.

Let Ω be the index set corresponding to the nonzero entries of α; we have $|\Omega| = K$. Also let Φ_Ω be the $M \times K$ mutilated matrix obtained by selecting the columns of Φ corresponding to the indices Ω. The measurement y is then a linear combination of the K columns of Φ_Ω. With probability one, the columns of Φ_Ω are linearly independent. Thus, Φ_Ω will have rank K and can be used to recover the K nonzero entries of α.

The coefficient vector α can be uniquely determined if no other index set $\widehat{\Omega}$ can be used to explain the measurements y. Let $\widehat{\Omega} \neq \Omega$ be a different set of K indices (possibly with up to $K-1$ indices in common with Ω). We will show that (with probability one) y is not in the column span of $\Phi_{\widehat{\Omega}}$, where the column span of the matrix A is defined as the vector space spanned by the columns of A and denoted by colspan(A).

First, we note that with probability one, the columns of $\Phi_{\widehat{\Omega}}$ are linearly independent and so $\Phi_{\widehat{\Omega}}$ will have rank K. Now we examine the concatenation of these matrices $\left[\Phi_\Omega \;\; \Phi_{\widehat{\Omega}}\right]$. The matrix $\left[\Phi_\Omega \;\; \Phi_{\widehat{\Omega}}\right]$ cannot have rank K unless $\text{colspan}(\Phi_\Omega) = \text{colspan}(\Phi_{\widehat{\Omega}})$, a situation that occurs with probability zero. Since these matrices have $M = K+1$ rows, it follows that $\left[\Phi_\Omega \;\; \Phi_{\widehat{\Omega}}\right]$ will have rank $K+1$; hence the column span is $\mathbb{R}^{K+1}$.

Since the combined column span of Φ_Ω and $\Phi_{\widehat{\Omega}}$ is $\mathbb{R}^{K+1}$ and since each matrix has rank K, it follows that $\text{colspan}(\Phi_\Omega) \cap \text{colspan}(\Phi_{\widehat{\Omega}})$ is a $(K-1)$-dimensional linear subspace of $\mathbb{R}^{K+1}$. (Each matrix contributes one additional dimension to the column span.) This intersection is the set of measurements in the column span of Φ_Ω that could be confused with signals generated from the vectors $\widehat{\Omega}$. Based on its dimensionality, this set has measure zero in the column span of Φ_Ω; hence the probability that α can be recovered using $\widehat{\Omega}$ is zero. Since the number of sets of K indices is finite, the probability that there exists $\widehat{\Omega} \neq \Omega$ that enables recovery of α is zero.

Statement 1 (Achievable, $M \geq 2K$): We first note that, if $K \geq N/2$, then with probability one, the matrix Φ has rank N, and there is a unique (correct) reconstruction. Thus we assume that $K < N/2$. The proof of Statement 1 follows similarly to the proof of Statement 2. The key fact is that with probability one, all subsets of up to $2K$ columns drawn from Φ are linearly independent. Assuming this holds, then for two index sets $\Omega \neq \widehat{\Omega}$ such that $|\Omega| = |\widehat{\Omega}| = K$, colspan($\Phi_\Omega$) $\cap$ colspan($\Phi_{\widehat{\Omega}}$) has dimension equal to the number of indices common to both Ω and $\widehat{\Omega}$. A signal projects to this common space only if its coefficients are nonzero on exactly these (fewer than K) common indices; since $\|\alpha\|_0 = K$, this does not occur. Thus every K-sparse signal projects to a unique point in $\mathbb{R}^M$.

Statement 3 (Converse, $M \leq K$): If $M < K$, then there is insufficient information in the vector y to recover the K nonzero coefficients of α; thus we assume $M = K$. In this case, there is a single explanation for the measurements only if there is a single set Ω of K linearly independent columns *and* the nonzero indices of α are the elements of Ω. Aside from this pathological case, the rank of subsets $\Phi_{\widehat{\Omega}}$ will generally be less than K (which would prevent robust recovery of signals supported on $\widehat{\Omega}$) or will be equal to K (which would give ambiguous solutions among all such sets $\widehat{\Omega}$). □