Signal Space CoSaMP for Sparse Recovery with Redundant Dictionaries

Mark A. Davenport
Georgia Institute of Technology
mdav@gatech.edu

Deanna Needell
Claremont McKenna College
dneedell@cmc.edu

Michael B. Wakin
Colorado School of Mines
mwakin@mines.edu

Compressive sensing (CS) offers the promise that we can acquire a vector \( \mathbf{x} \in \mathbb{C}^n \) via only \( m \ll n \) linear measurements provided that \( \mathbf{x} \) is sparse or compressible. Specifically, CS considers the problem where we obtain measurements of the form \( \mathbf{y} = \mathbf{A\alpha} + \mathbf{e} \), where \( \mathbf{A} \) is an \( m \times n \) sensing matrix and \( \mathbf{e} \) is a noise vector. If \( \mathbf{x} \) is sparse or compressible and \( \mathbf{A} \) satisfies certain conditions, then CS provides a mechanism to recover the signal \( \mathbf{x} \) from the measurement vector \( \mathbf{y} \) efficiently and robustly.

Typically, however, signals of practical interest are not themselves sparse, but rather have a sparse expansion in some dictionary \( \mathbf{D} \). By this we mean that there exists a sparse coefficient vector \( \mathbf{\alpha} \) such that the signal \( \mathbf{x} \) can be expressed as \( \mathbf{x} = \mathbf{D}\mathbf{\alpha} \). One could then ask the simple question: How can we account for this signal model in CS? In some cases, there is a natural way to extend the standard CS formulation—since we can write the measurements as \( \mathbf{y} = \mathbf{A}\mathbf{D}\mathbf{\alpha} + \mathbf{e} \) we can use standard CS techniques to first obtain an estimate \( \hat{\mathbf{\alpha}} \) of the sparse coefficient vector. We can then synthesize an estimate \( \hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{\alpha}} \) of the original signal. Unfortunately, this is a rather restrictive way to proceed for two main reasons: (i) the application of standard CS results to this problem will require that the matrix given by the product \( \mathbf{AD} \) satisfy certain properties that will not be satisfied for many interesting choices of \( \mathbf{D} \), as discussed further below, and (ii) we are not really interested in recovering \( \mathbf{\alpha} \) per se, but rather in obtaining an accurate estimate of \( \mathbf{x} \). It may be possible to recover \( \mathbf{x} \) in situations where recovering \( \mathbf{\alpha} \) is impossible, and even if we could apply standard CS results to ensure that our estimate of \( \mathbf{\alpha} \) is accurate, for some poorly conditioned dictionaries \( \mathbf{D} \) this would not necessarily translate into a recovery guarantee for \( \mathbf{x} \).

We consider an alternative approach to this problem and develop an algorithm for which we can provide guarantees on the recovery of \( \mathbf{x} \) while making no direct assumptions concerning our choice of \( \mathbf{D} \). Before we describe our approach, however, it will be illuminating to see precisely what goes wrong in an attempt to extend the standard CS formulation. Towards this end, let us return to the case where \( \mathbf{x} \) is itself sparse (when \( \mathbf{D} = \mathbf{I} \)). In this setting, there are many possible algorithms that have been proposed for recovering an estimate of \( \mathbf{x} \) from measurements of the form \( \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \), including \( \ell_1 \)-minimization approaches and greedy/iterative methods such as iterative hard thresholding (IHT), orthogonal matching pursuit (OMP), and compressive sampling matching pursuit (CoSaMP). For any of these algorithms, it can be shown that \( \mathbf{x} \) can be accurately recovered from the measurements \( \mathbf{y} \) if the matrix \( \mathbf{A} \) satisfies the RIP with a sufficiently small constant \( \delta_2 \). Importantly, if \( \mathbf{A} \) is generated randomly with a number of rows roughly proportional to the sparsity level \( k \) then with high probability \( \mathbf{AD} \) will satisfy the RIP of order \( k \).

We now return to the case where \( \mathbf{x} \) is sparse with respect to a dictionary \( \mathbf{D} \). If \( \mathbf{D} \) is unitary (i.e., if the dictionary is an orthonormal basis), then it is straightforward to show that for a fixed \( \mathbf{D} \), if we choose a random \( \mathbf{A} \), then with high probability \( \mathbf{AD} \) will satisfy the RIP. Thus, standard CS algorithms can be used to accurately recover \( \mathbf{\alpha} \), and because \( \mathbf{D} \) is unitary, the signal space recovery error \( \|\mathbf{x} - \hat{\mathbf{x}}\|_2 \) will exactly equal the coefficient space recovery error \( \|\mathbf{\alpha} - \hat{\mathbf{\alpha}}\|_2 \).

Unfortunately, this approach won’t do in cases where \( \mathbf{D} \) is not unitary and especially in cases where \( \mathbf{D} \) is highly redundant/overcomplete. For example, \( \mathbf{D} \) might represent the overcomplete DFT, the Undecimated Discrete Wavelet Transform, a redundant Gabor dictionary, or a union of orthonormal bases. The challenges that we must confront when dealing with \( \mathbf{D} \) of this form include:

- Redundancy in \( \mathbf{D} \) will mean that in general, the representation of a vector \( \mathbf{x} \) in the dictionary is not unique—there may exist many possible coefficient vectors \( \mathbf{\alpha} \) that can be used to synthesize \( \mathbf{x} \).
- Coherence (correlations) between the columns of \( \mathbf{D} \) can make it difficult for \( \mathbf{AD} \) to satisfy the RIP.
- As noted above, if the dictionary \( \mathbf{D} \) is poorly conditioned, the signal space recovery error \( \|\mathbf{x} - \hat{\mathbf{x}}\|_2 \) could differ substantially from the coefficient space recovery error \( \|\mathbf{\alpha} - \hat{\mathbf{\alpha}}\|_2 \), further complicating any attempt to understand how well we can recover \( \mathbf{x} \) by appealing to results concerning the recovery of \( \mathbf{\alpha} \).

All of these problems essentially stem from the fact that extending standard CS algorithms in an attempt to recover \( \mathbf{\alpha} \) is a coefficient-focused recovery strategy. By trying to go from the measurements \( \mathbf{y} \) all the way back to the coefficient vector \( \mathbf{\alpha} \), one encounters all the problems above due to the lack of orthogonality of the dictionary. In contrast, we propose a signal-focused recovery strategy for CS. Our algorithm employs the model of sparsity in an arbitrary dictionary \( \mathbf{D} \) but directly obtains an estimate of the signal \( \mathbf{x} \), and we provide guarantees on the quality of this estimate in signal space. Our algorithm is a modification of CoSaMP, and in cases where \( \mathbf{D} \) is unitary, our “Signal-Space CoSaMP” algorithm reduces to standard CoSaMP. However, our analysis requires comparatively weaker assumptions. Our bounds require only that \( \mathbf{A} \) satisfy the \( \mathbf{D} \)-RIP—this is a different and less-restrictive condition to satisfy than requiring \( \mathbf{AD} \) to satisfy the RIP. The algorithm does, however, require the existence of a near-optimal scheme for projecting a vector \( \mathbf{x} \) onto the set of signals admitting a sparse representation in \( \mathbf{D} \). While the fact that we require only an approximate projection is a significant relaxation of the requirements of traditional algorithms like CoSaMP (which require exact projections), showing that a practical algorithm can provably compute the required near-optimal projection remains a significant open problem. Nevertheless, as we will demonstrate, various practical algorithms do lead to empirically superior performance, suggesting that this challenge might not be insurmountable. We will discuss these issues and compare our treatment with related work such as Blumensath’s Projected Landweber Algorithm (based on IHT) and works that employ an assumption of “analysis sparsity.”