

Compressed Sensing

Yonina C. Eldar

*Electrical Engineering Department,
Technion-Israel Institute of Technology, Haifa,
Israel, 32000*

1 Introduction

Compressed sensing (CS) is an exciting, rapidly growing field that has attracted considerable attention in electrical engineering, applied mathematics, statistics, and computer science. CS offers a framework for simultaneous sensing and compression of finite-dimensional vectors that relies on linear dimensionality reduction. Quite surprisingly, it predicts that sparse high-dimensional signals can be recovered from highly incomplete measurements by using efficient algorithms.

To be more specific, let x be an n -vector. In CS we do not measure x directly, but rather acquire $m < n$ linear measurements of the form $y = Ax$ using an $m \times n$ CS matrix A . Ideally, the matrix is designed to reduce the number of measurements as much as possible while allowing for recovery of a wide class of signals from their measurement vectors y . Thus, we would like to choose $m \ll n$.

Since A has fewer rows than columns it has a nonempty null space. This implies that for any particular signal x_0 , an infinite number of signals x yield the same measurements $y = Ax = Ax_0$. To enable recovery, we must therefore limit ourselves to a special class of input signals x .

Sparsity is the most prevalent signal structure used in CS. In its simplest form, sparsity implies that x has only a small number of nonzero values but we do not know which entries are nonzero. Mathematically, we express this condition as $\|x\|_0 \leq k$ where $\|x\|_0$ denotes the ℓ_0 -“norm” of x , which counts the number of nonzeros in x (note that $\|\cdot\|_0$ is not a true norm, since in general $\|\alpha x\|_0 \neq |\alpha| \|x\|_0$ for $\alpha \in \mathbb{R}$). More generally, CS ideas can be applied when a suitable representation of x is sparse. A signal x is k -sparse in a basis Ψ if there exists a vector $\theta \in \mathbb{R}^n$ with only $k \ll n$ nonzero entries such that $x = \Psi\theta$. As an example, the success of many compression algorithms, such as JPEG2000, is tied to the fact that natural images are often sparse in an appropriate

wavelet transform.

Finding a sparse vector x that satisfies the measurement equation $y = Ax$ can be performed by an exhaustive search over all possible sets of size k . In general, however, this is impractical; in fact, the task of finding such an x is known to be NP-hard. The surprising result at the heart of CS is that if x (or a suitable representation of x) is k -sparse, then it can be recovered from $y = Ax$ using a number of measurements m that is on the order of $k \log n$, under certain conditions on the matrix A . Furthermore, recovery is possible using polynomial-time algorithms that are robust to noise and mismodelling of x . In particular, the essential results hold when x is *compressible*, namely, when it is well approximated by its best k -term representation $\min_{\|v\|_0 \leq k} \|x - v\|$, where the norm in the objective is arbitrary.

CS has led to a fundamentally new approach to signal processing, analog-to-digital converter (ADC) design, image recovery, and compression algorithms. Consumer electronics, civilian and military surveillance, medical imaging, radar and many other applications rely on efficient sampling. Reducing the sampling rate in these applications by making efficient use of the available degrees of freedom can improve the user experience, increase data transfer, improve imaging quality, and reduce power, cost and exposure time.

2 Design of Measurement Matrices

The ability to recover x from a small number of measurements $y = Ax$ depends on the properties of the CS matrix A . In particular, A should be designed so as to enable unique identification of a k -sparse signal x . Let the support S of x be the set of indices over which x is nonzero, and denote by x_S the vector x restricted to its support. We similarly denote by A_S the columns of A corresponding to this support so that $y = Ax = A_S x_S$. When the support is known, we can recover x from y via $x_S = (A_S^T A_S)^{-1} A_S^T y$, assuming that A_S has full column rank. The difficulty in CS arises from the fact that the support of x is not known in advance. Therefore, determining conditions on A that ensure recovery is more involved.

As a first step, we would like to choose A such that every two distinct signals x, x' that

are k -sparse lead to different measurement vectors $Ax \neq Ax'$. This can be ensured if the *spark* of A satisfies $\text{spark}(A) \geq 2k + 1$ where $\text{spark}(A)$ is the smallest number of columns of A that are linearly dependent. Since $\text{spark}(A) \in [2, m + 1]$, this yields the requirement that $m \geq 2k$.

Unfortunately, computing the spark of a general matrix A has combinatorial computational complexity, since one must verify that all sets of columns of a certain size are linearly independent. Instead, one can provide (sub-optimal) recovery guarantees using the *coherence* $\mu(A)$, which is easily computable, and defined as

$$\mu(A) = \max_{1 \leq i \neq j \leq n} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2},$$

where a_i is the i th column of A . For any A ,

$$\text{spark}(A) \geq 1 + \frac{1}{\mu(A)}.$$

Therefore, if

$$k < \frac{1}{2} \left(1 + \frac{1}{\mu(A)} \right), \quad (1)$$

then for any $y \in \mathbb{R}^m$ there exists at most one k -sparse signal $x \in \mathbb{R}^n$ such that $y = Ax$.

In order to ensure stable recovery from noisy measurements $y = Ax + w$, where w represents noise, more stringent requirements on A are needed. One such condition is the *restricted isometry property* (RIP). A matrix A has the (k, δ) -RIP for $\delta \in (0, 1)$ if, for all k -sparse vectors x ,

$$(1 - \delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta)\|x\|_2^2.$$

This means that all submatrices of A of size $m \times k$ are close to an isometry, and therefore distance-preserving. Clearly if A has the $(2k, \delta)$ -RIP with $0 < \delta < 1$, then $\text{spark}(A) \geq 2k + 1$.

The RIP enables recovery guarantees that are much stronger than those based on spark and coherence. Another property used to characterize A is the null space condition. This requirement ensures that the null space of A does not contain vectors that are concentrated on a small subset of indices. If a matrix satisfies the RIP, then it also has the null space property. However, checking whether A satisfies either of these conditions has combinatorial computational complexity.

Random matrices A of size $m \times n$ with $m < n$, whose entries are independent and identically distributed (i.i.d.) with continuous distributions, have $\text{spark}(A) = m + 1$ with high probability. When the distribution has zero mean and finite variance, then in the asymptotic regime (as m and n grow) the coherence converges to $\mu(A) = 2\sqrt{\log n/m}$. Random matrices from Gaussian, Rademacher, or more generally a sub-gaussian distribution have the (k, δ) -RIP with high probability if $m = \mathcal{O}(k \log(n/k)/\delta^2)$. Similarly, it can be shown that a partial Fourier matrix with $m = \mathcal{O}(k \log^4 n/\delta^2)$ rows, namely a matrix formed from the $n \times n$ Fourier matrix by taking m of its rows uniformly at random, satisfies the RIP of order k with high probability. A similar result holds for random submatrices of orthogonal matrices.

There are also deterministic matrices that satisfy the spark and RIP conditions. For example, an $m \times n$ Vandermonde matrix constructed from n distinct scalars has spark equal to $m + 1$. Unfortunately, these matrices are poorly conditioned for large values of n , rendering the recovery problem numerically unstable. It is also possible to construct deterministic CS matrices of size $m \times n$ that have the (k, δ) -RIP for $k = \mathcal{O}(\sqrt{m} \log m / \log(n/m))$.

3 Recovery Algorithms

Many algorithms have been proposed to recover a sparse vector x from measurements $y = Ax$. When the measurements are noise free, and A satisfies the spark requirement, the unique sparse vector x can be found by solving the optimization problem

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \|x\|_0 \text{ subject to } y = Ax. \quad (2)$$

Solving (2) relies on an exhaustive search. Therefore, a variety of computationally feasible alternatives have been developed.

One popular approach to obtain a tractable problem is to replace the ℓ_0 norm by the ℓ_1 norm, which is convex. The resulting adaptation of (2), known as basis pursuit (BP), is defined by

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \|x\|_1 \text{ subject to } y = Ax.$$

This algorithm can be implemented as a linear program, making its computational complexity polynomial in n . BP is easily modified to allow for noisy measurements by changing the constraint to $\|y - Ax\|_2^2 \leq \epsilon$, where ϵ is an appropriately chosen bound on the noise magnitude. The Lagrangian relaxation of the resulting problem is given by

$$\hat{x} = \arg \min_{x \in \mathbb{R}^n} \|x\|_1 + \lambda \|y - Ax\|_2^2,$$

and is known as basis pursuit denoising (BPDN). Many fast methods have been developed in order to find BPDN solutions.

An alternative to optimization-based techniques are greedy algorithms for sparse signal recovery. These methods are iterative in nature and select columns of A according to their correlation with the measurements y . Several greedy methods can be shown to have performance guarantees that match those obtained for BPDN.

For example, the matching pursuit (MP) and orthogonal matching pursuit (OMP) algorithms proceed by finding the column a_j of A most correlated to the signal residual, where

$$j = \arg \max_i \frac{|a_i^T r|^2}{\|a_i\|_2^2}.$$

The residual r is obtained by subtracting the contribution of a partial estimate of the signal from y : $r = y - A_S x_S$ where S is the current guess of the support set. The convergence criterion used to find sparse representations consists of checking whether $y = Ax$ exactly or approximately. The difference between the two techniques is in the coefficient update stage. While in OMP in each stage all nonzero elements are chosen so as to minimize the residual error $\|y - A_S x_S\|_2^2$, in MP only the component associated with the currently selected column is updated to $a_j^T r / \|a_j\|_2^2$.

Another popular approach is known as iterative hard thresholding (IHT): starting from an initial estimate $\hat{x}_0 = 0$, the algorithm iterates a gradient descent step followed by hard thresholding, i.e.,

$$\hat{x}_i = H_k(\hat{x}_{i-1} + A^T(y - A\hat{x}_{i-1})).$$

Here $H_k(v)$ returns the k entries of v that are largest in absolute value.

Many of the CS algorithms above come with guarantees on their performance. For example,

BP and OMP recover a k -sparse vector from noiseless measurements when the matrix A satisfies (1). There also exist coherence-based guarantees designed for measurements corrupted with arbitrary noise. In general, though, results based on coherence typically suffer from the so-called square-root bottleneck: they require $m = O(k^2)$ measurements to ensure good recovery.

Stronger guarantees are available based on the RIP, which motivates the popularity of random CS matrices. In particular, OMP recovers a k -sparse vector from exact measurements if A has the $(k + 1, \delta)$ -RIP with a small enough value of δ . More generally, a sparse vector x can be recovered with small error from noisy measurements using IHT and BPDN when A has the (ck, δ) -RIP, with appropriate values of c and δ . These results also hold when x is not exactly sparse but only compressible. The recovery error in this case is proportional to that of the best k -sparse approximation of x and to the norm of the noise. Since random matrices satisfying the RIP can be constructed as long as $m = O(k \log(n/k))$ it follows that with high probability on the order of $k \log(n/k)$ measurements suffice to guarantee recovery of sparse vectors in the noise free setting and to ensure recovery with small error in the noisy case.

4 Applications

4.1 Imaging

One of the first applications of CS is the single-pixel camera. This camera uses a single photon detector (the single pixel) to measure m inner products of a desired image, represented by a vector x in \mathbb{R}^n , and a set of test vectors. Each vector represents the pattern of a digital micromirror device (DMD) which consists of n tiny mirrors that are individually oriented in a pseudo random fashion either towards the photodiode (representing a 1) or away from it (representing a 0). The incident light-field is reflected off the DMD, collected by a lens and then focused onto the photodiode which computes the inner product between x and the random DMD pattern. This process is repeated m times with different patterns. Good recovery of the underlying image has been obtained using about 60% fewer random measure-

ments than the reconstructed pixels, assuming sparsity of the image in the wavelet domain.

Another application in which the measurements are performed in the transform domain is magnetic resonance imaging (MRI). In MRI the measurements correspond to samples of the image’s 2D continuous Fourier transform. By exploiting the principles of CS, one can recover an MRI image from fewer Fourier-domain measurements assuming sparsity of the image in an appropriate transform domain. For example, MR angiography images are typically sparse in the pixel domain. The sparsity of these images can be increased by considering spatial finite differences. Brain MRI’s are known to be sparse in the wavelet domain and dynamic MRI is sparse in the temporal domain.

MRI scanning time strictly depends on the number of samples taken during acquisition. Therefore, applications of CS to MRI offer significant improvement in image acquisition speed. As current MRI scanning time lasts at least 30 minutes, rapid MRI will reduce patient discomfort and image distortion due to patient movement during acquisition. An important factor affecting the performance of CS-based MRI recovery is the sampling trajectory chosen in the frequency domain. Pure random sampling is impractical, due to hardware and physiological constraints. This directly impacts the RIP and coherence of the resulting measurement matrix. Different applications of MRI impose varying constraints on the possible trajectories which must be taken into account when designing a CS-based MRI system.

4.2 Analog-to-Digital Conversion

To date, essentially all ADCs follow the celebrated Shannon-Nyquist theorem, which states that in order to avoid information loss when converting an analog signal to digital the sampling rate must be at least twice the signal bandwidth. On going demand for data, as well as advances in radio frequency technology and the desire to improve resolution, have promoted the use of high-bandwidth signals. The resulting rates dictated by the Shannon-Nyquist theorem impose severe challenges both on the acquisition hardware and on the subsequent storage and processors.

Combining ideas of sampling theory with the

principles of CS, several new paradigms have been developed that allow sampling and processing a wide class of analog signals at sub-Nyquist rates using practical hardware architectures. One such framework is referred to as Xampling, and has led to sub-Nyquist prototypes for a variety of problems including cognitive radio, radar, ultrasound imaging, ultra wideband communication and more. Two of the hardware boards developed for cognitive radio and radar are presented in Fig. 1.

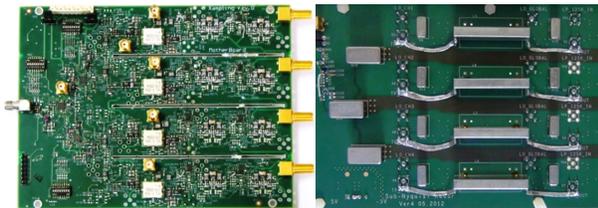


Figure 1: Sub-Nyquist hardware prototypes for cognitive radio (left) and radar (right).

In a cognitive radio setting the signal $x(t)$ is modeled as a multiband input with sparse spectra, such that its continuous-time Fourier transform is supported on N frequency intervals with individual widths not exceeding B Hz. Each interval is centered around an unknown carrier frequency f_i that is no larger than a maximum frequency f_{\max} . Using the Xampling paradigm, a sub-Nyquist prototype referred to as the modulated wideband converter (MWC) has been developed that can sample and process such signals at rates as low as $2NB$, despite the fact that the signal may be spread over a very wide frequency range. This rate is much lower than the Nyquist rate, corresponding to f_{\max} .

The MWC modulates the incoming signal with a pseudo-random periodic sequence, applies a lowpass filter to the result, and then samples the output at a low rate. The mixing operation aliases the spectrum to baseband with different weights for each frequency interval. The signal is recovered using CS techniques which account for the signal structure. The board in Fig. 1 samples signals with a Nyquist rate of 2.4GHz and spectral occupancy of 120MHz at a rate of 280MHz.

Another signal class that can be sampled at

sub-Nyquist rates are streams of pulses:

$$x(t) = \sum_{\ell=1}^L a_{\ell} h(t - t_{\ell}), \quad t \in [0, \tau],$$

where the time delays t_{ℓ} and amplitudes a_{ℓ} are unknown. Such signals arise, for example, in communication channels which introduce multipath fading, ultrasound imaging, and radar. Here again the Xampling paradigm can be used to sample and process such signals at rates as low as $2L/\tau$ irrespective of the signal's bandwidth.

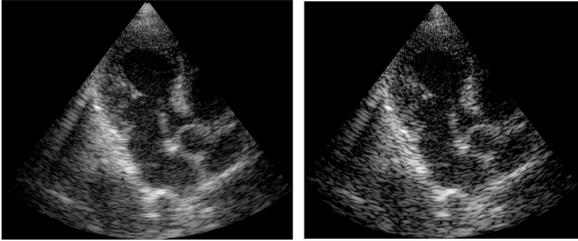


Figure 2: Sub-Nyquist ultrasound imaging. Standard image (left) and an image formed at $1/28$ of the Nyquist rate (right).

The board in Fig. 1 allows, for example, detection of radar signals at $1/30$ of the signal's Nyquist rate. Figure 2 demonstrates fast ultrasound imaging using Xampling. On the left is an ultrasound frame obtained by standard imaging techniques, while the right-hand side image is formed from samples at $1/28$ of the Nyquist rate. All the processing is performed at this low rate as well.

5 Extensions

In recent years, the area of CS has branched out to many new fronts and has worked its way into several application areas. This, in turn, necessitates a fresh look at many of the basics of CS.

A significant part of recent work on CS can be classified into three major areas. The first group consists of theory and applications related to CS matrices that are not completely random, or entirely deterministic, and that often exhibit considerable structure. This largely follows from efforts to model the way samples are acquired in

practice, which leads to sensing matrices that inherit their structure from the real world.

The second group includes signal representations that exhibit structure beyond sparsity and broader classes of signals, such as low-rank matrices and matrix completion; exploiting the distribution of the nonzero coefficients or other structured knowledge on the nonzero entries of x ; and continuous-time signals with finite or infinite-dimensional representations. In the context of analog signals, large efforts are being devoted to the development of efficient ADC prototypes that achieve sub-Nyquist sampling in practice.

Finally, a very recent trend in CS is to move away from the linear measurement model, and consider various types of nonlinear measurements. One particular example is phase retrieval problems in which the measurements have the form $y_i = |a_i^T x|^2$ for a set of vectors a_i . Note that here only the magnitude of $a_i^T x$ is measured, and not the phase. Phase retrieval problems arise in many areas of optics, where the detector can only measure the magnitude of the received optical wave. Several important applications of phase retrieval include X-ray crystallography, transmission electron microscopy and coherent diffractive imaging. Exploiting sparsity and ideas related to low-rank matrix representations results in efficient algorithms for phase retrieval with provable recovery guarantees. Another example of nonlinear measurements are quantized measurements.

Further Reading

1. A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.
2. N. Wagner, Y. C. Eldar and Z. Friedman, "Compressed Beamforming in Ultrasound Imaging," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4643–4657, Sept. 2012.
3. 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 Proc. Mag.*, vol. 25, no. 2, pp. 83–91, March 2008.
4. Y. C. Eldar and G. Kutyniok, eds, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.
5. S. Foucart and H. Rauhut, *A Mathematical In-*

Introduction to Compressive Sensing, Boston, MA: Birkhauser, 2013.

6. M. Lustig, D. Donoho and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging", *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182-1195, 2007.
7. M. Mishali and Y. C. Eldar, "Sub-Nyquist Sampling," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 98–124, 2011.
8. J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.