CHAPTER II
COMPRESSIVE SAMPLING

1. Traditional compression

In general, the signal measurement method is designed to cope with a signal in order to not to lose information with the uniform sampling at least twice faster than its bandwidth followed the classical Shannon/Nyquist theorem so that the quality of signal highly depends on the number of samples to store them.

However, the sampling rate in practical is out of control and beyond the current state-of-the-art is very expensive such that the samples in signal which require a good quality should be filtered by high-speed analog-to-digital converter to sample and store them. Furthermore, In the process of conventional compression, there are transformation and truncation processes, shown in Figure 1. The process depends on the types of transformations (such as *Fourier transformation*, *discrete cosine transformation*, *wavelets* etc.) which are designed for specific issues and the truncation process is the algorithm to reduce the size of compressed signal by selecting the significant coefficients.

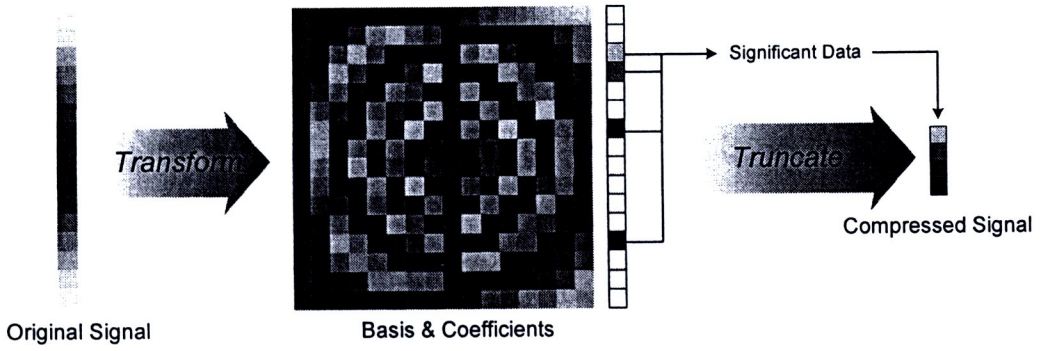

Figure 1 Traditional Compression Method

Alternatively, the new compressive sampling technique that tackles this problem is more general linear measurement scheme, acceptable to solve via the optimization tools, and significantly requisite certain signal sampling rate at below Shannon/Nyquist theorem [1].

2. Compressive sampling

This idea is proposed in 2006 by Candès, Romberg and Tao [1]. They confidently perceived the features of signal being *sparse* (see descriptions below). It is also possible to measure the signal randomly if measurement process can collect enough significant information on that non-zero information [2].

Consider a real-valued, finite-length, one-dimensional, discrete-time signal x which is viewed as $N \times 1$ an column vector in $\mathbb{R}^N$ with elements $x[n], n = 1,2,..., N$. For an image or higher-dimensional signal, it is reshaped by vectorizing into a long one-dimensional vector.

Let any signal x in $\mathbb{R}^N$ be expressed in terms of basis of $N \times 1$ vectors $\{\psi_{i=1}^N\}$ by for simplicity assuming that the basis is orthonormal and the vectors $\{\psi_i\}$ are stacked as columns of $N \times N$ basis matrix, i.e. $\Psi = [\psi_1|\psi_2|...|\psi_N]$. The considered signal x can be expressed as,

$$x = \sum_{i=1}^{N} s_i \psi_i \quad \text{or} \quad x = \Psi s, \tag{1}$$

where s is the $N \times 1$ column vector of weighting coefficients $s_i = \langle x, \psi_i \rangle = \psi_i^T x$ and the $\cdot^T$ denotes the (*Hermitian*) transpose operation. This means that the coefficient vector s is a representation of signal vector x in Ψ domain. In Figure 2, if an arbitrary signal is localized in Ψ domain by only K non-zero coefficient s_i and $(N - K)$ entries are zero, this signal is called K – sparse under *sparsity* property [2].

In actual fact, many natural and manmade are compressible in the sense, for example in many current applications that use the traditional compression method coding signal into transformed domain, JPEG and JPEG-2000 image compression are based on DCT and wavelets, audio signals based on Fourier basis, etc. However, these applications are not designed to sample and then compress all at once. Thus, the potential compressed signal should start with the possible largest number of samples N even if the number of non-zero coefficients K is small ($K << N$).

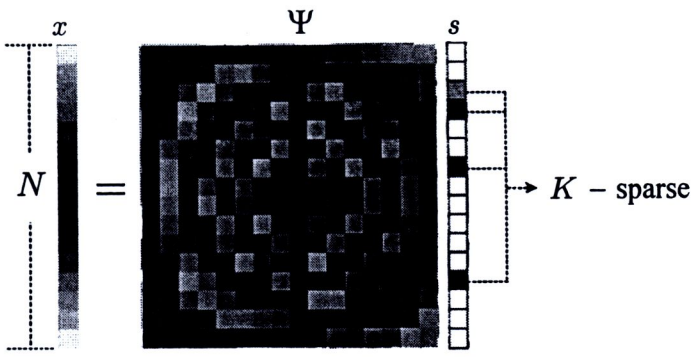

Figure 2 Represented signal in transformed domain

Otherwise, compressive sampling is more general acquisition approach that compresses the signal at one step of sample-then-compress framework directly without the other intermediate stage of taking K samples. Consider *linear measurement process* that computes the inner products between signal vector x and the collection vectors $\{\phi_j\}_{j=1}^{M}$ of measurement matrix Φ of size $M \times N$ where $M < N$ as in $y_j = \langle x, \phi_j \rangle$ stacking in to the $M \times 1$ vector y can be expressed as,

$$y = \Phi x = \Phi \Psi s = \Theta s, \tag{2}$$

where $\Theta = \Phi\Psi$ is an $M \times N$ matrix. Figure 3 is a graphic view of compressive sampling measurement process which is illustrated in the delineated scheme and Figure 4 shows the condensed scheme substituting the matrix product Θ in the equation. Note that the measurement process is such non-adaptive and thus the matrix Φ does not depend in any way on the signal vector x.

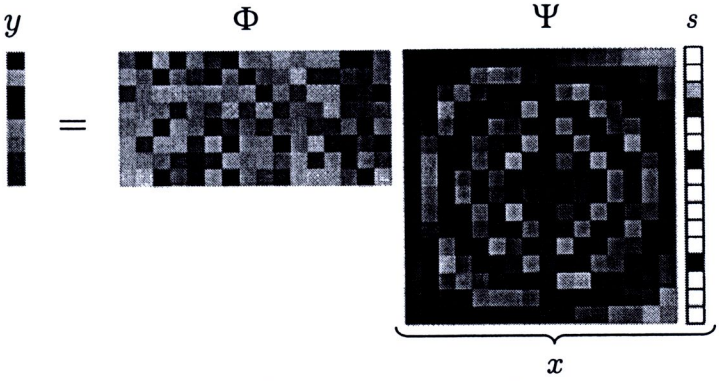

Figure 3 Delineated scheme of compressive sampling measurement process

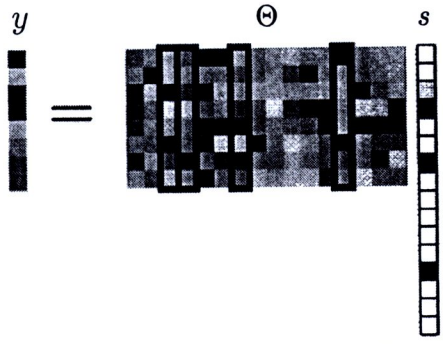

Figure 4 Condensed scheme of compressive sampling measurement process

The objective of compressive sampling measurement is to design the matrix Φ and reconstruction algorithm for $K-$sparse. In fact, there are fewer equations than unknowns, generating infinitely possible solutions. Thus, this problem cannot be easily solved unless there are other conditions.

However, the compressive sampling scheme needs the measurement matrix rows $M \approx K$ or slightly more. Because the compressed signal y is a linear combination of the non-zero coefficients s_i corresponding to the highlighted columns of matrix product Θ (Figure 4). Thus, there are only K non-zero entries and M rows seeming like K unknowns and $M \approx K$ equations which is sufficiently to find the solution.

Although the number of equations M equals or exceeds a number of unknowns, the locations of K non-zero entries are also not known. A necessary and sufficient condition which is to ensure to solve this $M \times K$ system is called the *restricted isometry property* (RIP),

$$1 - \varepsilon_{\text{RIP}} \leq \frac{\left\|\Theta \bar{s}\right\|_2}{\left\|\bar{s}\right\|_2} \leq 1 + \varepsilon_{\text{RIP}}, \tag{3}$$

where some $\varepsilon_{\text{RIP}} > 0$ and the any vector $\bar{s}$ shares the same K non-zero entries as the vector s. Note that the measurement matrix Θ must preserve the lengths of these particular $K-$sparse vectors. So far there are currently independently identically distributed (i.i.d.) random Gaussian distribution and Rademacher distribution with random ± 1 entries which has been shown to satisfy the RIP [2].

3. Reconstruction algorithms

The RIP provides the theoretical condition to guarantee that a K – sparse can be recovered by the M measurement in compressed vector y. Since $M < N$, there are infinitely many solution vectors $\bar{s}$ that satisfy $\Theta\bar{s} = y$ and thus lie on $(N-M)$-dimensional hyperplane $\mathcal{H} = \mathcal{N}(\Theta) + s$ corresponding to the null space $\mathcal{N}(\Theta)$ of the matrix Θ translated to the correct sparse solution s. Because if $\Theta s = y$ then $\Theta(s + r) = y$ for any vector r in the null space. Thus, the goal is to find the sparest signal coefficient vector s in the translated null space. A logical algorithm that is used to count the smallest number of coefficients in null space $\mathcal{H}$ defined as ℓ_0-minimization,

$$(\mathrm{P}_0) \qquad \min \|s\|_0 \text{ sub. to } \Theta s = y, \tag{4}$$

where $\|s\|_0$ is called "zero-norm" of vector s. The forms of ℓ_p-norm of the vector s as $\|s\|_p$ are expressed as,

$$\|s\|_p = \begin{cases} \left| i \,|\, s_i \neq 0, i = 1,2,\ldots,N \right| & , p = 0 \\ \sum_{i=1}^{N} \left(|s_i|^p \right)^{\frac{1}{p}} & , p > 0 \end{cases}. \tag{5}$$

The ℓ_0-minimization needs sufficiently the measurement matrix rows $M = K + 1$ with i.i.d. Gaussian measurements, this optimization can recover a K – sparse signal exactly with very high probability [3]. However, it is unfortunately hard to solve because it is an NP-complete problem that requires an exhaustive enumeration of C_K^N possible combinations for all locations of the non-zero entries in signal vector s [2].

Thus, the surprised reconstruction algorithm which is an approximate optimization from $M \geq cK \log(N/K)$ (for some constant c) i.i.d. Gaussian measurements can exactly recover the K – sparse vector and closely approximate compressible vectors stably with high probability via the ℓ_1-minimization,

$$(P_1) \qquad \min \|s\|_1 \text{ sub. to } \Theta s = y. \tag{6}$$

It is proved in [1] and rewritten in the following theorem,

Theorem 1. *Let s_0 be a discrete signal support on an unknown set T, and choose Ω of size N_ω (which is the number of samples to take in Fourier domain) uniformly at random from all set of this size, i.e. each of the $C_{N_\omega}^N$ possible subsets are equally likely. For a given accuracy parameter M, if*

$$|T| \le c_M (\log N)^{-1} |\Omega|, \tag{7}$$

for explicit value c_M and then with probability at least $1 - O(N^{-M})$, the minimizer to problem (6) is unique and equal to s_0.

Moreover, the ℓ_1-minimization is the convex optimization problem that can be conveniently solved via a linear optimization known as *simplex method* or *basis pursuit* [1, 2, 4] requiring computational complexity of about $O(N^3)$.

4. Geometrical interpretation

The view of geometry helps to visualize why the ℓ_1-minimization can recover the true K – sparse signal. From Figure 5 (a), note that K – sparse vector s in $\mathbb{R}^N$ contain many zero entries that corresponding to K-dimensional hyperplanes aligned on the coordinate axes; this means that a sparse vectors live close to the coordinate axes in $\mathbb{R}^N$. Thus, the translated null space hyperplane $\mathcal{H} = \mathcal{N}(\Theta) + s$ which is of $(N - M)$ dimensions should touch ℓ_1-ball (the capacity of ℓ_1 objective function) at the correct solution vector near the coordinate axes, which is precisely where the sparse vector s is located, shown in Figure 5(b).

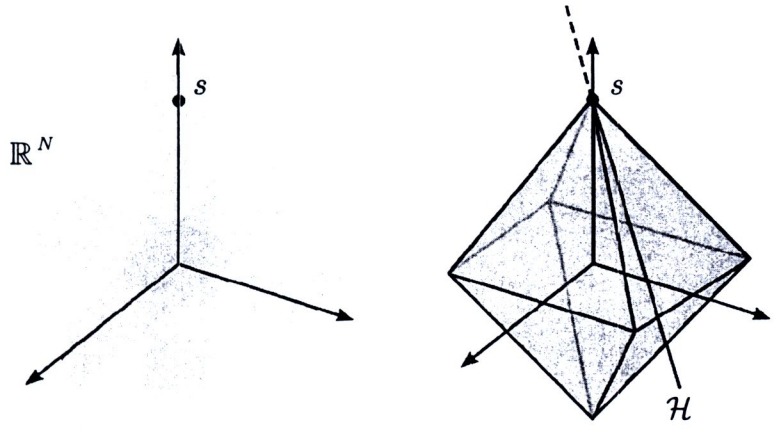

Figure 5 Sparse vector in $\mathbb{R}^N$: (a) it lies on K-dimensional hyper planes aligned with the coordinate axes. (b) ℓ_1-ball touches the translated hyperplane at K – sparse vector near the coordinate axes.

Although ℓ_1-minimization can find K – sparse signal with high probability, its performance is under the explicit probability and enough measurements. Thus, it is possible to reconstruct the fault K – sparse signal for some chances especially in fewer measurements. For example, Figure 6 shows the compressive sampling recovery via ℓ_1-minimization when varying size of measurement matrix rows M.

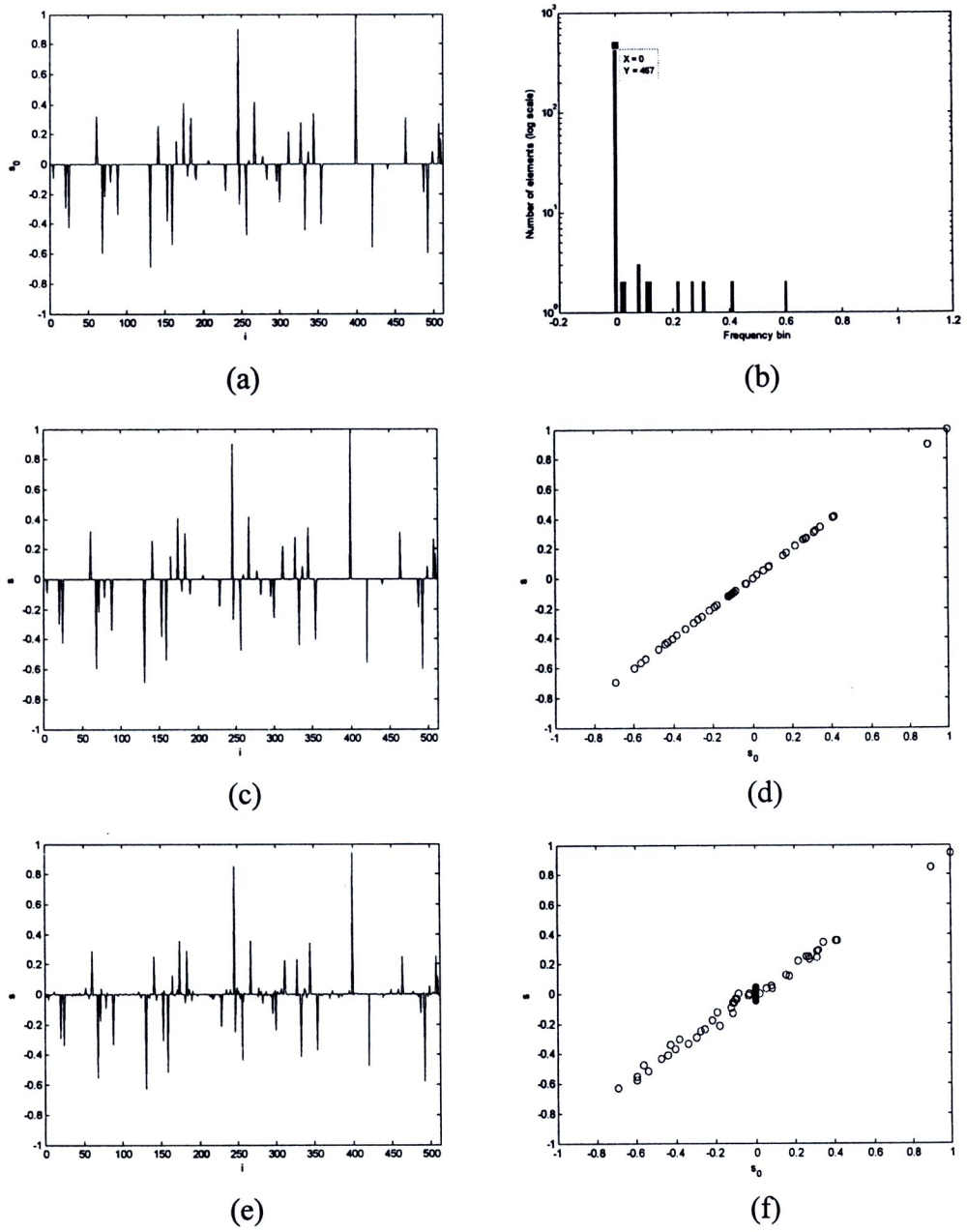

Figure 6 Spare recovery using ℓ_1-minimization: (a) original coefficient vector s_0 on the interval [-1,1], length $N = 512$, with 45 spikes, (b) its histogram of absolute coeffient vector $|s_0|$, reconstructed coefficient vector when using mesurement matrix (c) size 200×512 ((d) its scatter plot (coefficient-by-coefficient of s_0 versus its reconstruction)) and (e) size 150×512((d) its scatter plot)