Compressive Sensing Based Robust Signal Sampling

Lianlin Li (Corresponding author)

Department of Petroleum Engineering, Texas A&M University

College Station, TX, USA, 77843

&

Department of Electromagnetic Theory and Application

Institute of Electronics, Chinese Academy of Sciences, Beijing, China, 100080

Tel: 1-484-802-9596   E-mail: lianlinli1980@gmail.com

Fang Li

Institute of Electronics, Chinese Academy of Sciences, Beijing, China, 100080

Tel: 86-010-588-874-81   E-mail: fli@ie.ac.cn

Abstract

Signal processing methods have been changed substantially over the last several decades. Traditional sampling theorem of Shannon-Nyquist states that the sampling rate must be at least twice the maximum frequency presented in the signal; however, sampling at the Nyquist rate is inefficient because the signals of interest contain only a small number of significant frequencies relative to the band limit, although the locations of the frequencies may not be known a priori. Recently, compressive sensing (CS) has made a paradigmatic step in the way information is presented, stored, transmitted and recovered, by which we can acquire and reconstruct sparse signals from sub-Nyquist incoherent measurements. Three key ingredients of successfully implementing compressed sampling technique are sparsible/compressible probed signal, reliable hardware design, and low-cost computational algorithm. In this paper, we focused on two aspects about the robust sampling of sparsible/compressible signal, in particular, the design of compressed sampling hardware and the robust reconstruction via sparse Bayesian analysis. Primary results showed the high performance of proposed strategies.

Keywords: Compressive sensing, Signal, Sampling

1. Introduction

Signal processing methods have been changed substantially over the last several decades. Traditional Shannon-Nyquist based sampling strategy states that the sampling rate must be at least twice the maximum frequency presented in the signal. However, sampling at the Nyquist rate is inefficient in many cases, for example, the signals of interest contain only a small number of significant frequencies relative to the band limit. Recently, researchers in the growing vigorously field of compressive sensing (CS) have made a step forward by relating a signal's sparsity structure with its acquisition and have showed that one can recover a sparse signal from a few linear non-adaptive measurements (Candes et al., 2006; Candes and Romberg, 2007; Candes and Tao, 2006; Donoho, 2006; Romberg, 2008; Li et al., 2009). Accordingly, CS has made a paradigmatic step in the way information is presented, stored, transmitted and recovered, by which we can acquire and reconstruct sparse signals from their sub-Nyquist incoherent measurements. Roughly speaking, CS can be an effective approach in situations such as: (a) The conventional Nyquist rate of sampling is too high to implement; for example, the current analogy-to-digital conversion (ADC) technology based on uniform sampling in the time or spatial domain is limited to within the order of 1GHz. (b) The number of sensors is limited due to implementation constraints or cost. (c) The sensing process may be slow so that only a small number of measurements can be

collected.

Besides sparsity, other two key ingredients of implementation CS technique for practical applications are reliable hardware realization and robust signal reconstruction. It has been shown that the random matrix with entries independently drawn at random from a Gaussian distribution of zero mean and unit variance (or other distributions such as Bernoulli, and so on) can ensure the exact recovery of the signal which is sparse in arbitrary orthobase with overwhelming probability (Candes et al., 2006; Candes and Romberg, 2007; Candes and Tao, 2006; Donoho, 2006). Following this theory, the well-known single-pixel camera was constructed by Baraniuk et al. Later on, many efforts to design the universal CS measurement instruments have been done by many researchers, for example, the chip-level Analogy-to-Information converter (Laska et al., 2007), the single-shot compressive spectral imager (Gehm et al., 2007; Marcia et al., 2009), the random lens (Fergus et al., 2006), and so on. Unfortunately, these CS measurements cannot be usually used in practice (at least cannot used for the real-time purpose) because of its time-consuming data collection and the difficulty of physical realization. To address this drawback, the random convolution based sampling strategy has been investigated from the aspects of theoretical analysis (Romberg, 2008) and empirical attempt (Tropp et al., 2006). Application of a random filter for compressive sensing was first mentioned by Tropp et al. who proposed two equivalent realization structures of a random filter: 1) convolution with a random waveform in the time domain, and 2) multiplication with random weights in the frequency domain, both followed by equal interval down-sampling. Following the Romberg's theory, Jacques et al constructed the CMOS compressed imaging by using a lot of shift registers in a pseudo-random configuration.

In this paper we will firstly focus on the random-convolution based sampling strategy and its hardware realization. Theoretical and experimental results from the time-reversal in random media have shown that exploiting the multi-scattering of acoustic and electromagnetic wave in random media can dramatically enhance the imaging resolution. Motivated by this observation, we will mainly study the application of wave scattering/propagation in random medium in low-cost hardware design for signal sampling and diffraction tomography. The proposed strategy has the following two advantages: 1) the generation of the measurement ensemble is computationally efficient and requires less memory; and 2) the proposed sampling hardware is universal from the viewpoint of compressive sensing, in particular, it can be applied to several transform domains and leads to simple implementations. It is noted that there are several similar works by exploiting the multi-path effect for complex signal environment, for example, the super-resolution migration imaging has been numerically investigated when the point-like clutters with known characteristics appears (Cheney, 2008), the random lens has been constructed to obtain super-resolution optical imaging from sparse measurements (Marcia et al., 2009), and so on. Secondly, we will discuss the robust reconstruction of sparse signals used, in particular, to address sparse signal reconstruction under some unreliable measurements. Recovering information from the maximum number of dependable observations with some unreliable ones is critical for robust signal sampling. We will investigate this problem from the viewpoint of sparse Bayesian estimation.

2. Random Convolution Sampling Theory

In a nutshell, compressed sampling (CS) acquires a signal of interest indirectly by correcting a small number of its projections rather than evenly sampling it at the Nyquist rate which can be prohibitively high for broadband signals encountered in many applications. This new signal acquisition paradigm has revolutionized that way digital data are traditionally acquired. The sparsity is one of key issues involved in current CS theory, which express the idea that the "information rate" of a continuous-time signal may be much smaller than that suggested by its bandwidth. In other words, a discrete-time signal depends on a number of degree of freedom which is relatively much smaller than its length; actually, many natural signals are sparse or compressible in the sense that they have sparse or approximately sparse representations when expressed in an appropriate basis or dictionary. Empirically, a vast amount of facts showed that many sub-Nyquist sampled signal sparse or compressible in some transformed domain can be exactly reconstructed through the lp-norm ($0 \le p \le 1$) based recovery techniques, which is traced back to the work of Logan in 1965, Santosa and Symes in 1986, and Donoho and Stark in 1989. However, the sparse related concepts were generally agreed since the foundation of so-called compressed sampling (CS) developed by Candes, Tao, Romberg and Donoho in 2006 and 2007. Since then, a burst of intensive research activities in the new field of CS have been inspired in the past several years.

Consider a discrete signal $\mathbf{x} \in \Box^n$ which itself may or may not be sparse in the canonical basis but is sparse or approximately sparse in an appropriate basis Ψ, i.e., $\mathbf{x} = \Psi\theta$, where θ is sparse or approximately sparse. These measurements are realized by projecting the signal onto a series of test waveforms. For instance, if a discrete signal $\mathbf{x}$ is a n-dimensional vector and $\{\varphi_i, i = 1, 2, ..., m\}$ is m test vectors with the same length as x, then the

measured data are given by

$$y_i = \langle \mathbf{x}, \boldsymbol{\varphi}_i \rangle, i = 1, \ldots, m \tag{1}$$

where $\langle \ \rangle$ stands for the inner product of two vectors. Using matrix notation the sensing process is described by $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{n}$, where $\mathbf{n}$ means the "small perturbation in the data" referring either to signals that are not exactly sparse but nearly sparse, or to presence of noise in the sampling process.

In the literatures the sensing matrix $\boldsymbol{\Phi}$ required by CS theory is often chosen as a random matrix for it keeps small coherence with $\boldsymbol{\Psi}$. However, it is also known that the aforementioned random structures are inefficient in terms of realization and computation, especially for large-scale problems. So far, many sensing matrices with certain structures have been investigated by many authors and it is noted that the random convolution based approach is the most popular due to its outstanding advantages in the aspect of hardware design and algorithm consideration (Bajwa et al., 2007; Romberg, 2008; Rivenson and Stern, 2009; Tropp et al.,2006). Mathematically, the random convolution can be formulated by $y(t) = x(t) \otimes h(t)$, where the input signal $x(t)$ and output signal $y(t)$ is related by the random filter response $h(t)$. Apparently, one key issue is that what kind of $h(t)$ can be used to sample the sparse signal using as less number of measurement as possible. In terms of classical linear algebra, the received signal $y(t)$ can be expressed as $\mathbf{y} = n^{-\frac{1}{2}}\mathbf{S}_\Omega \cdot \mathbf{F}^* \cdot \boldsymbol{\Sigma} \cdot \mathbf{F} \cdot \mathbf{x} \equiv \boldsymbol{\Phi} \cdot \mathbf{x}$, where $\mathbf{F}$ is a discrete Fourier matrix with size n by n, $\mathbf{x} = \left[x(t_1), \cdots, x(t_n) \right]^T$. The entries of diagonal matrix $\boldsymbol{\Sigma} = diag\left(H(f_1), \ldots, H(f_n) \right)$ are unit magnitude complex numbers with random phase. One choice of the sampling matrix $\mathbf{S}_\Omega$ from Romberg's theorem is: to generate an i.i.d. sequence of Bernoulli random variables $\{ t_k : k = 1, 2, \cdots, n \}$, each of which takes a value of 1 with probability m/n, and sample the locations selected from $t_k \in \Omega = \{ k : t_k = 1 \}$. The Romberg's theorem universally works because the generated CS measurement matrix $\boldsymbol{\Phi}$ will be incoherent with any fixed orthonormal matrix $\boldsymbol{\Psi}$ with 'overwhelming' probability.

Although the important success of Romberg's theorem in design of digital circuit and numerical simulation, it is very hard to construct the analogy hardware that satisfies the unite amplitude-response within the working bandwidth. Here, we will investigate the CS theorem for convolution with white noise waveforms followed by deterministic subsampling. Performance of this framework is quite simple, straightforward and not costly. Only small adjustments are required to transform a traditional sensing system to a compressive sensing one. Practical examples include the famous coded aperture, radar system, CMOS compressed imaging, and so on.

The basic idea behind Romberg's work is that the randomness is designed in the frequency domain, where the spectrum of the random waveform has unit amplitude and independent random phase such that the random waveform is orthogonal with its shift which makes the convolution matrix orthogonal. In contrast, in our model the convolution matrix is not orthogonal and some frequency components of the interested signal will be filtered out. Accordingly, our model is not suitable for sensing signals which are sparse in the frequency domain; however, this sacrifice leads to an advantage of system realization. We showed that the suitability of the white convolution system for sensing a sparse signal depends on the coherence between the signal representation and the Fourier basis. Another important contribution is that the randomly selected sampling strategy is not necessary. We proved that subsampling at fixed locations also works well for the random convolution framework (Xiang et al., 2009).

For practical consideration, we assume to sense a continuous signal with limited control parameters. The signal is band-limited or approximated by a limited-resolution observation. For simplicity, we suppose the bandwidth of the signal to be 1. Since $y(t) = h(t) \otimes x(t)$, the full sampled data can be given by $y(k) = \int h(k - \tau) x(\tau) d\tau$. For $x(t) = \sum_{k=1} x(k) \operatorname{sinc}(\pi(t - k))$, we are interested in recovery of the Nyquist samples $\{ x(k) \}$. Denote $\hat{h}(k) = h(t) * \operatorname{sinc}(\pi t)|_{t=k}$, then $y(k) = \sum_l x(l) \hat{h}(k - l)$. It comes back to the discrete framework of convolution. However, the entries of the random sequence $\hat{h}(k), k = 1, \ldots, n$ are not independent of one another in general. Since $E\left(\hat{h}(k), \hat{h}(l) \right) = \delta_{k,l}$, where $\delta_{k,l}$ is the Dirac function, we can still demonstrate the continuous problem by our discrete model, $\hat{h}(k), k = 1, \ldots, n$, which is independent when generated from a Gaussian distribution. Accordingly, we can extend our convolution model into the continuous case.

In summary, we analyzed the random convolution framework, i.e., convolving the tested signal with a white random waveform, followed by subsampling at fixed locations in the measurement domain, i.e., equal interval sampling. As an effect of the reduction in the freedom of randomness, the linear CS convolution system needs more measurements than the circular one with the same size. It also becomes inefficient when the waveform length is too short relative to the signal length. Indeed, in some applications the bandwidth of the random waveform is shorter than the tested signal. However, even in this case one may have super-resolution results when the original signal is sparse enough or when more prior information is used.

3. Application of Random Convolution Sampling Theory

The basic requirement of random convolution based sampling theory is to construct a system whose system response likes random number. In many applied fields such as communication, radar, navigation, etc., the point-spread response of system under complex wave environment usually is treated as random function in view of complex behaviour of propagation path and energy lossy. Especially, the amplitude-frequency response of it will be constant for ideal lossless system, which is required by Romberg's theory. Motivated by this observation, we will investigate wave phenomena in the random medium environment, and will explore its possibility of decreasing the number of signal sampled.

Before proceeding to details, let us qualitatively consider the simplified situation as shown in Figure 1. A source transmitting pulse of $s(t)$ is located at origin of coordinate and surrounded by rectangle-ring random media while a receiver is a detector is located at $\boldsymbol{\rho}_r$ outside of this ring. It should be pointed out that in this work the structure is constructed by the dielectric (or metal) square cylinders with known electrical/geometrical parameters $\varepsilon_r = 8$ randomly located around probed obstacles. Assuming that the signal from transmitter arrives at the receiver through N independent propagation paths {d_i, i=1,2,…,N}, then the echoes recorded by the receiver can be approximately represented as

$$y(t) \approx \sum_{i=1}^{N} a_i s\left(t - \frac{d_i}{c}\right)$$

where c is the light velocity, a_i is decay coefficient for the ith propagation path. Readily we get the system response of random media as

$$H(\omega) = \sum_{i=1}^{N} a_i \exp\left(j\frac{d_i}{c}\omega\right) \qquad (2)$$

Theoretically the expected response required by random convolution theorem can be achieved by adjusting the value of $\{a_i, d_i, i = 1, 2, ...\}$ (equivalently, controlling random media); however, this is a challenge optimization problem due to nonlinear interaction between wave (whether electromagnetic or acoustic) and media, and should be deeply investigated in the near future.

From the viewpoint of electromagnetic/acoustic inverse problems, the above problem can be looked as the so-called inverse source problem; correspondingly the inverse scattering problem will be considered below, in particular, the diffraction tomography. Referring to Figure 1(b), three point targets are surrounded by the random media, while the transmitter and receiver are set to be outside the random medium. The location and reflectivity of ith target are $\boldsymbol{\rho}_i$ and σ_i (i=1, 2,3), respectively, furthermore, we will assume there is no interaction between these three targets. Now our basic goal is to determine the distribution of reflectivity (i.e., find σ_i at $\boldsymbol{\rho}_i$ (i=1, 2,3)) from the time-domain data collected by the receiver located at $\boldsymbol{\rho}_r$.

The signal arriving at the ith objects is

$$s_i(t) \approx \sum_{j=1}^{N_{1t}} a_{it,j} s\left(t - \frac{d_{it,j}}{c}\right) \approx h_{it}(t) \otimes s(t),$$

where $h_{it}(t) = \square^{-1}\left[\sum_{j=1}^{N_{1t}} a_{it,j} \exp\left(j\frac{d_{it,j}}{c}\omega\right)\right]$, N_{it} is the number of wave paths connecting transmitter and ist object, $a_{it,j}$ and $d_{it,j}$ are the amplitude decay and electrical path of jth path, respectively. Now, the recorded wave scattered from 1st object can be expressed as

$$s_i'(t) = \sigma_i \sum_{j=1}^{N_{1r}} a_{ir,j} s_i \left(t - \frac{d_{ir,j}}{c} \right) = \sigma_i h_{ir} \otimes s_i(t) = \sigma_i h_{ir} \otimes h_{it} \otimes s(t)$$

where $h_{ir}(t) = \square^{-1} \left[\sum_{j=1}^{N_{1r}} a_{ir,j} \exp \left(j \frac{d_{ir,j}}{c} \omega \right) \right]$, among N_{ir} wave paths connecting the ist object and receiver,

$a_{ir,j}$ and $d_{ir,j}$ are the amplitude decay and electrical path of jth path, respectively. Finally, the echo collected by receiver becomes

$$y(t) = \sum_i \sigma_i h_{ir} \otimes h_{it} \otimes s(t)$$

which can be rewritten as in the frequency domain,

$$y(\omega) = \sum_i \sigma_i F_i(\omega) \tag{3}$$

where $F_i(\omega) = H_{ir}(\omega) H_{it}(\omega) \tilde{s}(\omega)$. Ideally $\{F_i(\omega), i = 1, 2, ...\}$ should be the orthogonal basis in the sense of probability due to random sequences of $\{a_{ir,j}\}, \{d_{ir,j}\}, \{a_{it,j}\}$ and $\{d_{it,j}\}$, which is exactly what the random convolution theory needs. We can empirically show $\int F_i(\omega) F_j(\omega) d\omega \approx C\delta_{i,j}$ through numerical simulation, where C is a constant, if the distance between any two pixels is larger than $0.25\lambda_0$, which means the use of random media can realize the super-resolution tomography $(0.25\lambda_0)$ by the measurement configuration of one pair of transmitter/receiver, for convenience, call it compressive diffraction tomography.

3.1 Compressed sampling filter

We call a compressed-sampling filter (CSF) one for which the function relating the input signal to the output signal is pseudo-random (Li et al., 2009). In spirit of wave propagation in random media, one microwave circuit based CSFs working within the 2.0GHz to 4GHz to enable signal acquisition with sub-Nyquist sampling has been constructed, analyzed and tested. The results have empirically shown that by using the proposed architecture the S-sparse n-dimensional signal can be exactly reconstructed with $O(S \log n)$ real-valued measurements or $O(S \log(n/S))$ complex-valued measurements with overwhelming probability. The proposed CSF (with size of 96mm by 60mm) is shown in Fig. 2 (a), and the measured amplitude-frequency and phase-frequency responses of this CSF are provided in Fig. 2(b) and Fig.2 (c), respectively. So far there is no rigorous theoretical formulation for designing such microwave structure; however, one can design it under the guide of enriching the signal multi-scattering or multi-path effect to the maximum extent.

To investigate the performance of the proposed CSF, in particular, how many measurements *M* are required to reconstruct exactly the *K*-sparse *N*-dim signal by the proposed structure, the methodology commonly used in the field of compressive sampling is exploited to check empirically the relation between *K*, *N* and the number of measurements *M*. To end this, assuming the length of unknown sparse signal x is 400, and the sampling ratio is 1/3-Nyquist ratio. For each of 200 trials we randomly generate such sufficiently sparse vectors x (choosing the nonzero locations uniformly over the support in random and their values from N(0,1/400)). The graph presented in Fig. 3 shows that the success ratio for complex-valued data and real-valued data in recovering the true sparse signal, which shows that by the 1/3-Nyquist sampling, one realize the exact reconstruction of 120-sparse 400-dim signal for complex-valued measurements and 30-sparse 400-dim signal for real-valued measurements. Fig.3 show that the for complex-valued measurements, the measurements with the order of O(Klog(N/K)) is enough to exactly reconstruct K-sparse N-dim signal; for real-valued measurements, the required measurements is order of O(Klog(N)). Moreover, this conclusion for signal which is sparse in the DCT and Harr-wavelet domain still exists. It should be pointed out that though the constructed CSF not satisfied the requirements from sampling theory studied previously, in particular, nonuniform amplitude-frequency response due to energy lossy dependent on frequency, and many other reasons, non-conjugate symmetry phase-frequency response, and so on, from the presented empirical results one can realize the exact reconstruction from the sub-Nyquist measurement by the proposed structure.

Reference to Figure 4, a simple example is demonstrated to show the application of proposed CSF in signal

reconstruction, where the original signal which is the combination of two differential Gaussian pulses $x_1(t) = t \exp\left(-3\pi \dfrac{t^2}{\tau^2}\right)$ and a sine-modulated differential Gaussian pulse $x_2(t) = \sin(\omega_0 t) \exp\left(-4\pi \dfrac{t^2}{\tau^2}\right)$, where $\tau = 0.6 ns$ and $\omega_0 = 2*10^9$ rad are the input signal of the CSF, and the measured output signal is shown in Figure 4(b). The 1/3-Nyquist complex-valued measurements are used to reconstruct the original time-domain sparse signal and the reconstructed result Figure 4(a).

3.2 Compressive diffraction tomography

Another important application of random convolution theory is the well-known diffraction tomography (DT), where the electrical and geometrical properties of an object can be reconstructed from the diffracted field (Devaney, 1984; Pan and Kak, 1983). Under the assumption of weak scattering, one can use Born or Rytov approximations to derive the so-called Fourier diffraction projection theorem (FDPT), which relates the 2D Fourier transform of the object function to the measured data (Devaney, 1984; Pan and Kak, 1983). The input data of the FDPT algorithm and its variants should satisfy the so-called Nyquist-Shannon theorem to obtain the successful reconstruction; traditionally, the full-view and full-angle single-frequency measurements tightly packed over the sphere with radius k (k denotes the working wavenumber of homogeneous background media) are imperative. On the other hand, due to the use of Fourier transform the theoretical resolution limit of FDPT algorithm or its invariants are 0.5λ (λ is the working wavelength) (Devaney, 1984; Pan and Kak, 1983). Two appealing problems for the DT are how to design work system to relieve time-consuming data collection and how to improve the imaging resolution. In (Cui et al., 2004) by taking the rich information of decaying wave in the region of near field into account, an super-resolution imaging from the near-field measurement has been proposed within the framework of Born-based inversion. The further insight into the mechanism which leads to sub-wavelength resolution imaging from far-field measurement by means of multiple scattering has been carried out experimentally and theoretically in (Chen and Chew, 1998) and (Cui et al., 2004). On the other hand, the *a prior* information is that the gradient-magnitude image is highly sparse has been exploited to obtain perfect 0.5λ resolution reconstruction from few-view and limited-angle Fourier data (LaRoque et al., 2008). In this paper we will focus on the use of random media to realize the "compressive" measurement in order to obtain super-resolution reconstruction from highly sparse data.

Here we restrict ourselves into two-dimensional scalar scattering problem as shown in Figure 1(b). The complex wave environment where the probed obstacles are located has been constructed; in particular, the random background medium to enforce the significant multi-path effect has been added. Under the assumption of Born approximation for weak scattering, the measurement signal $E_s(k_j)$ (j=1,2,…,M), representing the scattered fields whose working wavenumber is k_j and is observed at $\mathbf{r}_R$ due to incident field from the line source located at $\mathbf{r}_S$, can be expressed as

$$E_s(k_j) = k_j^2 \int\limits_{Dinv} G(\mathbf{r}_R, \mathbf{r}'; k_j) E_{in}(\mathbf{r}_S, \mathbf{r}'; k_j) \chi(\mathbf{r}') d\mathbf{r}' \tag{4}$$

where $\chi(\mathbf{r})$ is assumed to be frequency-independent contrast function of probed objects $E_{in}(\mathbf{r}_S, \mathbf{r}'; k_j)$ is the incident wave due to the line source at $\mathbf{r}_S$, i.e.. $G(\mathbf{r}_R, \mathbf{r}'; k_j)$ is the Green's function for the random background media, which can be obtained and stored via in-situ measurement or numerical computation. Now (4) can be reformulated in the form of the matrix-vector, namely,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \qquad \mathbf{n}, \mathbf{y} \in \square^{M \times 1}, \ \mathbf{A} \in \square^{M \times N}, \ \mathbf{x} \in \square^{N \times 1} \tag{5}$$

where N is the total number for unknowns to be reconstructed, the entries of $\mathbf{y}$ come from $E_s(\mathbf{r}_R, \mathbf{r}_S; k_j)$, the entries of $\mathbf{x}$ are $\chi(\mathbf{r}_j')$, $\mathbf{n}$ denotes the error due to measurement noise and discretization transformation from (4) to (5), the entries of $\mathbf{A}$ are $\Delta k_j G(\mathbf{r}_R, \mathbf{r}_i; k_j) E_{in}(\mathbf{r}_S, \mathbf{r}_i; k_j)$ ($i = 1, 2, \cdots, N$). Further, if $\mathbf{x}$ can be sparsely represented in some orthogonal basis $\mathbf{\Psi}$ such as the wavelet transform, the discrete cosine transform (DCT), and so on, i.e., $\mathbf{x} = \mathbf{\Psi}\mathbf{\theta}$, the equation (2.2) can be rewritten as

$$\mathbf{y} = \mathbf{A}\mathbf{\Psi}\mathbf{\theta} + \mathbf{n}, \qquad \mathbf{n}, \mathbf{y} \in \square^{M \times 1}, \ \mathbf{A} \in \square^{M \times N}, \ \mathbf{\theta} \in \square^{N \times 1} \tag{6}$$

For the purpose of the super-resolution reconstruction from the measurements of single receiver/transmitter pair, the formed complex-valued matrix $\mathbf{A}$ is highly underdetermined, i.e., $M \square N$. Obviously, the highly underdetermined system of linear equations (6) will generate infinite solutions. If one desires to narrow the choice to one well-defined solution, additional constraint should be imposed to guarantee a unique solution. To attack this, define the general optimization problem as

$$\hat{\mathbf{x}} = \min_{\bar{x}} J(\mathbf{x}), \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{Ax}\|_2 \leq \|\mathbf{n}\|_2 < \sigma \tag{7}$$

where $J(\mathbf{x})$ is introduced to evaluate the desirability of a would-be solution $\mathbf{x}$. Usually, the squared l2-norm $J(\mathbf{x}) = \|\mathbf{x}\|_2$, a measure of energy, is employed to obtain a unique solution. It is well known that the measure of energy is not optimal, even be a disaster in some cases. Fortunately, all most of obstacles to be probed for many applications such as medical imaging, geophysics, optical imaging, and so on, are sparse, or compressible in some framework such as DCT, wavelet, and so on. Consequently, the alternative to l2-norm regularization is so-called sparsity-promoted one, i.e., $J(\mathbf{x}) = \|\mathbf{\Psi x}\|_1$.

In the following a set of numerical simulations are provided to demonstrate above approach, where the investigation domain with size 2.4m by 2.4m is divided into the subgrids with size $0.1\lambda_0$ by $0.1\lambda_0$, where $\lambda_0 = 1.2m$ is the central working wavelength, and the used working wavelength ranging from $0.8\lambda_0$ to $1.2\lambda_0$ with space of $0.004\lambda_0$. The probed objects (see Fig. 5(a)) are assumed to be sparse in Harr-wavelet domain and their reconstructions are shown in Figs. 5(b) and 5(c), where Fig. 5(b) is the result for diffraction tomography by l2-norm constraint while 5(c) is for l1-norm regularization.

From Figs. 5, one can find that (a) the use of random media can realize the diffraction tomography even if just one pair of transmitter and receiver are employed which is impossible task if the objects are located at free space; (b) Although the reconstruction for the l2-constraint and random media background media can also yield visually reasonable reconstructions, its results are much poorer than the reconstruction with l1-norm constraints optimization, which can provide us super-resolution reconstruction. In summary, by compressive diffraction tomography (i.e., the random-media background plus sparse-enhance optimization), one can obtain exactly super-resolution reconstruction from the highly sparse data.

4. Robust Sparse Signal Reconstruction

In this section we will briefly discuss the algorithm of reconstructing sparse signal involved in previous sections. Generally, the considered problem formally can be stated via linear inverse problem with l1-norm based sparse constraint, in particular,

$$\min \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{Ax} = \mathbf{b} + \mathbf{n} \tag{8}$$

where the linear measurement operator $\mathbf{A}$ mapping $\mathbf{x} \in X \subseteq \square^n$ to $\mathbf{b} \in B \subseteq \square^m$, specifically, $\mathbf{A} : X \rightarrow B$. The noise $\mathbf{n}$ follows the Gaussian distribution with zero mean and covariance $\mathbf{R}$, i.e., $\mathbf{n} \square \mathcal{N}(\mathbf{0}, \mathbf{R})$. In practice, due to non-idealities $\mathbf{A}$ in the analogy and digital implementation of CS, measurement mismatch or/and random malfunctioning of the instrument for measuring $\mathbf{b}$, and other possible reasons, we have to explore the so-called robust optimization problem, in particular,

$$\min \|\mathbf{x}\|_1 \quad \text{s.t.,} \quad (\mathbf{A} + \delta\mathbf{A})\mathbf{x} = (\mathbf{b} + \delta\mathbf{b}) + \mathbf{n}, \quad (\delta\mathbf{A}, \delta\mathbf{b}) \in U \tag{9}$$

where U we introduce uncertainty set, or the set of admissible disturbance of $(\mathbf{A}, \mathbf{b})$ for constraint on perturbation $(\delta\mathbf{A}, \delta\mathbf{b})$.

In this paper we assume that the perturbation matrix $\delta\mathbf{A}$ and observations $\delta\mathbf{b}$ are sparse, which means just a limited number of entries of $\delta\mathbf{A}$ or/and $\delta\mathbf{b}$ are non-zero and non-ignorable. Correspondingly, the resulting problem of (9) can be reformulated as

$$\left(\delta\hat{\mathbf{A}}, \delta\hat{\mathbf{b}}, \hat{\mathbf{x}}\right) =$$
$$\arg\min_{\delta\mathbf{A}, \delta\mathbf{b}, \mathbf{x}} \left[\left\|\mathbf{b} + \delta\mathbf{b} - (\mathbf{A} + \delta\mathbf{A})\mathbf{x}\right\|_{\mathbf{R}^{-1}}^2 + \gamma_1 \|\mathbf{x}\|_1 + \gamma_2 \|\delta\mathbf{A}\|_{1,1} + \gamma_3 \|\delta\mathbf{b}\|_1 \right] \tag{10}$$

where $\|\delta\mathbf{A}\|_{1,1} = \sum_{i,j} |\delta a_{i,j}|$, γ_1, γ_2 and γ_3 are three regularization factors to be determined carefully.

We will provide new insight into problem of (9) within the framework of Bayesian analysis, whose distinguished advantages include 1) to do uncertainty analysis and 2) to estimate all of parameters involved adaptively. For the sake of simplicity, we assume $\delta\mathbf{A} = \mathbf{0}$. The first task of Bayesian analysis is to assign suitable probability distribution for uncertainty $\delta\mathbf{b}$ and unknown signal $\mathbf{x}$. Traditionally, the Gaussian distribution is the favourite assumption for modelling uncertainty in many research areas such as pattern recognition, computer vision, data clustering, and interpretation problems. However, even when such justification is absent, the Gaussian remains a popular choice due to its attractive analytical properties. For example, log likelihoods are often quadratic in parameters of interest, which facilitates optimization, and a Gaussian prior can be combined with a Gaussian likelihood analytically in Bayesian inference. However, as mentioned previously, both of $\delta\mathbf{b}$ and $\mathbf{x}$ are not controlled by a Gaussian, then the model based on the Gaussian assumption favours a Gaussian estimate that may interpret the data set in a misleading way. The results in an incorrect model for the data set, especially when possible outliers exist in the data set. To model this case, the Laplacian distribution about the noise is commonly explored; however, due to computational intractability, the so-called hierarchical Gaussian model is used instead (Tipping and Faul, 2003; Ji et al., 2008), in particular,

$$\delta\mathbf{b} \sim N\left(\mathbf{0}, \mathbf{B}\right), \text{ and } \mathbf{x} \sim N\left(\mathbf{0}, \mathbf{\Lambda}\right)$$

where $\mathbf{B} = diag\left(\beta_i, i=1,2,...,m\right)$, $\mathbf{\Lambda} = diag\left(\gamma_i, i=1,2,...,n\right)$, and the Gamma distribution is used to model the induced hyperparameters of $\mathbf{B}$ and $\mathbf{\Lambda}$, i.e.,

$$\beta_i \sim Gamma\left(1, b_i\right) = d_i \exp\left(-b_i\beta_i\right),$$

and

$$\gamma_i \sim Gamma\left(\gamma_i \mid 1, d_i\right) = d_i \exp\left(-d_i\gamma_i\right).$$

Now, the estimation of unknown $\mathbf{x}$ and induced hyperparameters of $\mathbf{B}$ and $\mathbf{\Lambda}$ can be readily made through the so-called type-I likelihood estimation, namely (Tipping and Faul, 2003; Li and Li., 2011; Ji et al., 2008),

$$\hat{\mathbf{x}} = \max_{\mathbf{x}} \Pr\left(\mathbf{x} \mid \mathbf{y}, \mathbf{\Lambda}, \mathbf{B}\right) \tag{11}$$

where the the posterior probability of $\mathbf{x}$ follows

$$\Pr\left(\mathbf{x} \mid \mathbf{y}, \mathbf{\Lambda}, \mathbf{B}\right) \propto \Pr\left(\mathbf{y} \mid \mathbf{x}, \mathbf{\Lambda}, \mathbf{B}\right) \Pr\left(\mathbf{x} \mid \mathbf{\Lambda}\right) = N\left(\mathbf{\mu_x}, \mathbf{\Sigma_x}\right) \tag{12}$$

Where $\mathbf{\mu_x} = \mathbf{\Sigma_x}\mathbf{A}^T\mathbf{B}^{-1}\mathbf{y}$ and $\mathbf{\Sigma_x} = \left(\mathbf{\Lambda}^{-1} + \mathbf{A}^T\mathbf{B}^{-1}\mathbf{A}\right)^{-1}$.

Furthermore, we can estimate hyper-parameters of $\mathbf{B}$ and $\mathbf{\Lambda}$ under the type-II likelihood estimation, namely,

$$\left(\mathbf{\Lambda}, \mathbf{B}\right) = \max_{\mathbf{\Lambda},\mathbf{B}} L \tag{13}$$

where $L = \Pr\left(\mathbf{\Lambda}, \mathbf{B} \mid \mathbf{y}\right) \propto \Pr\left(\mathbf{y} \mid \mathbf{\Lambda}, \mathbf{B}\right) \Pr\left(\mathbf{\Lambda}\right) \Pr\left(\mathbf{B}\right)$,

$$\Pr\left(\mathbf{y} \mid \mathbf{\Lambda}, \mathbf{B}\right) = N\left(\mathbf{0}, \mathbf{\Sigma_b}\right),$$

and

$$\mathbf{\Sigma_b} = \mathbf{B} + \mathbf{A}\mathbf{\Lambda}\mathbf{A}^T = \left[\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{A}\mathbf{\Sigma_x}\mathbf{A}^T\mathbf{B}^{-1}\right]^{-1}.$$

After some basic manipulations, we can get

$$L \propto -\frac{1}{2}\log\left|\mathbf{\Sigma_y}\right| - \frac{1}{2}\|\mathbf{y}\|_{\mathbf{\Sigma_y^{-1}}}^2 + \sum_{i=1}^{N}\log\left[b_i\exp\left(-b_i\gamma_i\right)\right] + \sum_{i=1}^{N}\log\left[d_i\exp\left(-d_i\beta_i\right)\right] \tag{14}$$

Taking the gradient of L with respect to $\mathbf{\Lambda}, \mathbf{B}$, and setting them to be zeros, one can arrive at the closed-form

solutions to $\mathbf{\Lambda}$ and $\mathbf{B}$ as (Li and Li., 2011)

$$\gamma_i = \mu_{x,i}^2 + \Sigma_{x,ii} \tag{15}$$

and

$$\beta_i = \left(\mathbf{y} - \mathbf{A}\mathbf{\mu_x}\right)_i^2 + \left(\mathbf{A}\mathbf{\Sigma_x}\mathbf{A}^T\right)_{i,i} \tag{16}$$

As discussed above, we can see that the proposed approach within framework of Bayesian analysis belongs to iterative one, in particularly, repeatedly computing (12), (15) and (16), which is computation demanding.

5. Conclusion

Compressive sensing (CS) has made a paradigmatic step in the way information is presented, stored, transmitted and recovered, by which we can acquire and reconstruct sparse signals from sub-Nyquist incoherent measurements. Three key ingredients of successfully implementing compressed sampling technique are sparsible/compressible probed signal, reliable hardware design, and low-cost computational algorithm. In this paper, we have investigated two aspects about the robust sampling of sparse signal, in particular, the design of compressed sampling hardware and the robust reconstruction via sparse Bayesian analysis.

A key to the success of CS is the design of the measurement ensemble satisfying the so-called restricted isometric property (RIP) or other rules. Theoretical and experimental results have shown that very limited random measurements of probed signals can suffice to provide valuable information of the signal subspace. On the other hand, the well-known time reversal theory shows that exploiting the multi-scattering of acoustic wave in random media can dramatically enhance the imaging resolution. In this paper, we first discussed the relationship between the wave scattering/propagation in complex medium and the classical compressive sensing theory; afterwards, study the design of circuit-level CS hardware prototype. The proposed variable sampling approach has the following advantages: 1) the generation of the measurement ensemble is computationally efficient and requires less memory; and 2) the proposed sampling hardware is universal from the viewpoint of compressive sensing, in particular, it can be applied to several transform domains and leads to simple implementations. In addition, another contribution of this paper is the development of robust sparse Bayesian analysis to address the robust optimization problem involved in sparse signal processing.

References

Bajwa, W. U., J. D. Haupt, G. M. Raz, S. J. Wright, & R. D. Nowak. (2007). Toeplitz-structured compressed sensing matrices, IEEE/SP 14th Workshop on Statistical Signal Processing, 294-298, Madison, WI, Aug

Candes, E., J. Romberg, & T. Tao. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory*, *Vol. 52*, No. 2, 489-509. http://dx.doi.org/10.1109/TIT.2005.862083

Candes, E, & J. Romberg. (2007). Sparsity and incoherence in compressive sampling, *Inverse Problems*, *Vol. 23*, 969-986. http://dx.doi.org/10.1088/0266-5611/23/3/008

Candes, E, & T. Tao. (2006). Near-optimal signal recovery from random projections and universal encoding strategies, *IEEE Trans. Inform. Theory*, *Vol. 52,* 5406-5425. http://dx.doi.org/10.1109/TIT.2006.885507

Carin L., D. Liu, & B. Guo. (2008). In Situ compressive sensing for multi-static scattering: imaging and the restricted isometry property, Preprint.

Cui T. J., W. C. Chew, X. X. Yin, & W. Hong. (2004). Study of resolution and super resolution in electromagnetic imaging for half-space problem. *IEEE Trans. Antenna and Propagation*, *Vol. 52,* 1398-1411. http://dx.doi.org/10.1109/TAP.2004.829847

Chen F. C., & W. C. Chew. (1998). Experimental verification of super resolution in nonlinear inverse scattering, *Appl. Phys. Lett.*, *Vol. 72*, 3080-3082. http://dx.doi.org/10.1063/1.121547

Cheney M., & R. Bonneau. (2004). Imaging that exploits multipath scattering from point scatterers. *Inverse Problem*, *20*, 1691-1711. http://dx.doi.org/10.1088/0266-5611/20/5/023

Devaney A. (1984). Geophysical diffraction tomography, *IEEE Trans. Geosci. Remote Sensing*, *Vol. 22*, 3-13. http://dx.doi.org/10.1109/TGRS.1984.350573

Donoho, D. (2006). Compressed sensing, *IEEE Trans. Inform. Theory*, *Vol. 52*, No. 4, 1289-1306. http://dx.doi.org/10.1109/TIT.2006.871582

Fergus, R., A. Torralba, & W. T. Freeman. (2006). Random lens imaging, *MIT-CSAIL-TR-2006-058*.

Gehm, M. E., R. John, D. J. Brady, R. M. Willett, & T. J. Schulz. (2007). Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express*, *Vol. 15*, No. 21, 14013-14027. http://dx.doi.org/10.1364/OE.15.014013

Jacques, L., P. Vandergheynst, A. Bibet, V. Majidzadeh, A. Schmid, & Y. Leblebici. (2009). CMOS compressed imaging by random convolution. [Online] Available*: http://www.dsp.ece.rice.edu/cs*

Ji S., Y. Xue, & L. Carin. (2008). Bayesian compressive sensing. *IEEE Trans. Signal Process, Vol.56*, No.6, 2346-2356. http://dx.doi.org/10.1109/TSP.2007.914345

Laska, J. N., S. Kirolos, M. F. Duarte, T. Ragheb, R. G. Baraniuk, & Y. Massoud. (2007). Theory and implementation of an analogy-to-information conversion using random demodulation. *Proc. IEEE Int. Symposium on Circuits and Systems*, New Orleans, LA, May

LaRoque S. J., E. Y. Sidky, & X. Pan. (2008). Accurate image reconstruction from few-view and limited-angle data in diffraction tomography. *J. Opt. Soc. Am. A*, *Vol. 25*, No.7, 1772-1782. http://dx.doi.org/10.1364/JOSAA.25.001772

Li L., W. Zhang, Y. Xiang, & Fang Li. (2009). The compressed-sampling filter (CSF). *Progress in Electromagnetic Research B, Vol. 17,* 255-273. http://dx.doi.org/10.2528/PIERB09081006

Marcia, R. F., T. H. Zachary, & R. M. Willett. (2009). Compressive coded aperture imaging. *SPIE-IS&T Electronic Imaging*, *Vol. 7246,* 72460G-2

Pan S. X., & A. C. Kak. (1983). A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered back propagation. *IEEE Trans. Acoust., Speech, Signal Process*, *Vol. 31*, 1262–1275. http://dx.doi.org/10.1109/TASSP.1983.1164196

Romberg, J. (2008). Compressive sensing by random convolution. *SIAM J. Imaging Sciences*, *Vol.4*, No.4, 1098-1128

Rivenson, Y., & A. Stern. (2009). Compressed imaging with separable sensing operator. *IEEE Signal Processing Letters*, *Vol. 16*, No. 6, 449-452. http://dx.doi.org/10.1109/LSP.2009.2017817

Tropp, J. A., M. B. Wakin, M. F. Duarte, D. Baron, & R. G. Baraniuk. (2006). Random filters for compressive sampling and reconstruction. *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, Toulouse, France, May

Tipping M., & A. Faul. (2003). Fast marginal likelihood maximization for sparse Bayesian models. *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, C. M. Bishop and B. J. Frey, Eds.

Xiang, Yin, L. Li, & Fang Li. (2010). Compressive sensing by white random convolution. *Available on eprint arXiv:0909.2737*.

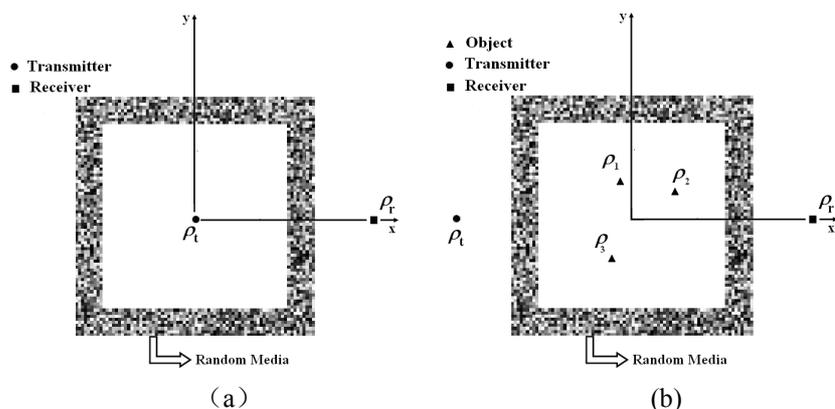（a）   (b)

Figure 1. (a) The sketch map of compressive sampling based inverse source, used for equation (2); (b) The sketch map of compressive sampling based inverse scattering, used for equation (3)

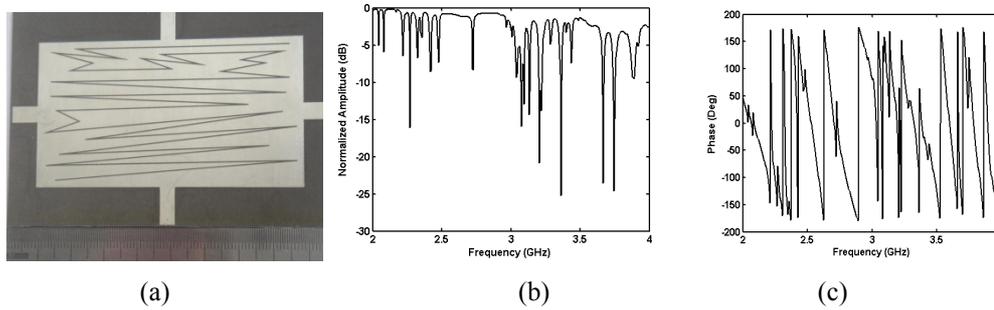(a)   (b)   (c)

Figure 2. (a)The photo of the proposed CSF based on the random microwave structure, and its (b) the amplitude-frequency response and (c) the phase-frequency response
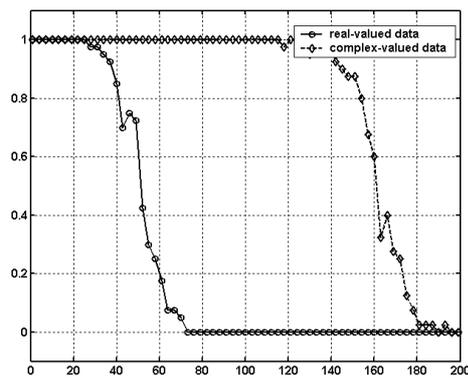

Figure 3. Probability of success of developed CSF shown in Fig. 2 in the recovery of the sparest signal when the length of unknown signal is 400 and 1/3-Nyquist measurements, where x-axis denotes the cardinality of the solution, y-axis denotes the probability of success of the developed CSF shown in Fig.2
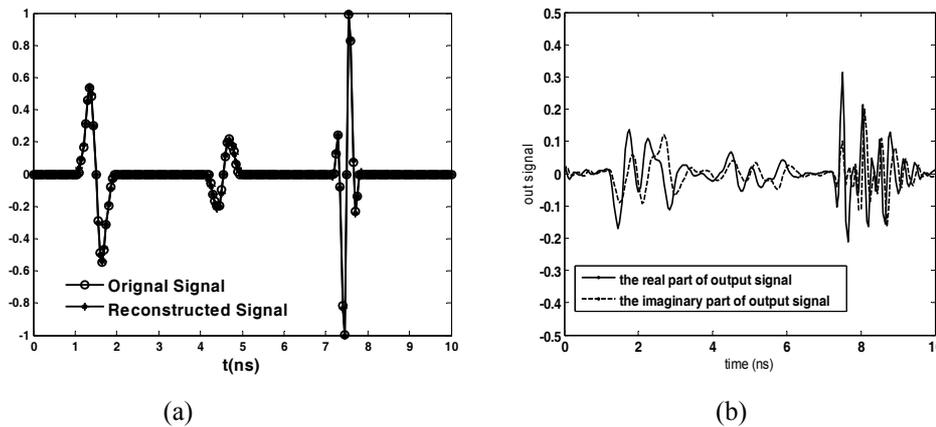

(a)   (b)

Figure 4. (a) the original and reconstructed signal, (b) the real and imaginary part of output signal of the developed CSF shown in Fig.2
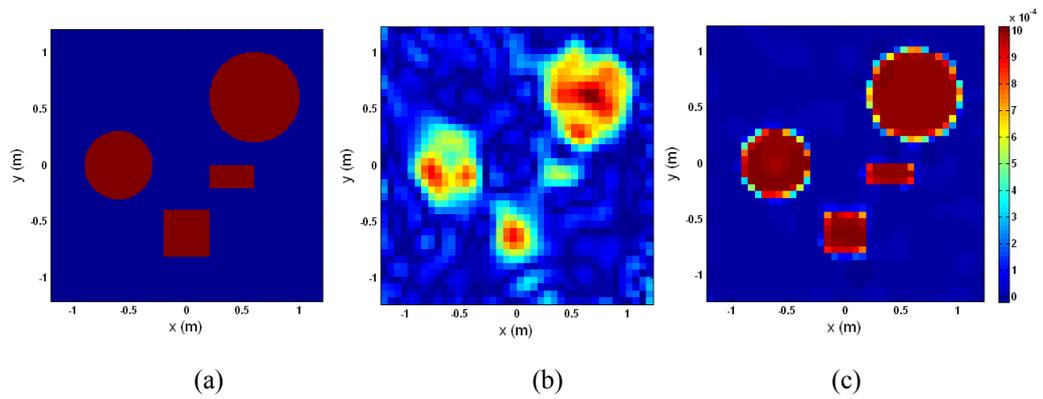
Figure 5. Diffraction tomography reconstructions with Haar-wavelet transformation from sparse data (a) true objects, (b) the reconstruction via l2-norm constraint optimization and (c) reconstruction via l1-norm constraint optimization