ISSN 0005-1179, Automation and Remote Control, 2010, Vol. 71, No. 11, pp. 2259–2282. © Pleiades Publishing, Ltd., 2010.
Original Russian Text © O.N. Granichin, D.V. Pavlenko, 2010, published in Avtomatika i Telemekhanika, 2010, No. 11, pp. 3–28.

REVIEWS

Randomization of Data Acquisition and ℓ_1-optimization (Recognition with Compression)

O. N. Granichin and D. V. Pavlenko

St. Petersburg State University, St. Petersburg, Russia

Received February 16, 2010

Abstract—A new paradigm for processing signals with sparse representation in some basis is actively developed for some time past. It relies largely on the ideas of measurement randomization and ℓ_1-optimization. The recent methods of acquisition and representation of the compressed data were christened *compressive sensing*.

DOI: 10.1134/S0005117910110019

1. INTRODUCTION

In the last century, the term "information" became popular in many domains of science and got special definitions and treatments in each of them [1]. One definition given in Wikipedia states: "*Information is something potentially perceived as representation, though not created or presented for that purpose. Information has been perceived by a conscious mind and also interpreted by it, the specific context associated with this interpretation may cause the transformation of the information into knowledge.*" The term "information" is usually confused or identified with the allied notions of "data" (received message) and "knowledge." In the present paper, understanding diversity of possible treatments and not going into details, we assume that *knowledge* enables one to restore the required *information* x from the *data* y acquired in the course of experiments or computations. One can often assume for simplicity that the essential information about the phenomenon $x \in \mathbb{X}$ under consideration is related with the available data $y \in \mathbb{Y}$ through understanding of the regularities of the phenomenon—knowledge (operator) $\Phi : \mathbb{X} \to \mathbb{Y}$:

$$y = \Phi x \ \ (= \Phi(x)).$$

If the operator Φ is invertible, then it provides an exhaustive knowledge for full restoration of x from y. It is known from the matrix algebra that $x = \Phi^{-1}y$ for the linear dependence $y, x \in \mathbb{R}^N$ and nondegenerate $N \times N$ matrix Φ.

The case where the data are subject to the action of uncontrollable disturbances

$$y = \Phi x + \xi$$

is characteristic of the open systems. From the philosophical point of view, discrimination of the term ξ emphasizes the impact on the data y of other phenomena except for those defined by x. The last relation is more natural from the practical standpoint even in the absence of a direct external action. Usually, data acquisition is a process of interaction of the examined phenomenon (reflected by the information x) with some measurement system with its characteristics integrated in ξ.

For an insignificant level of external disturbances ξ (or at their damping), the problem of restoring x from y comes to that of inverting the operator Φ, which usually is attained by increasing the number of observations: for $x \in \mathbb{R}^N$, chosen are $m > N$ and $y \in \mathbb{R}^m$.

For significant external disturbances ξ, the statistical formulation of the problem is usually used. The possibilities of its solution for $m \gg N$ were considered in detail within the framework of the traditional mathematical theory of experiment design [2]. The external disturbances are regarded as a realization of some sequence of independent random variables with the zero mean. In applications, however, this assumption is frequently violated, which may affect the traditional estimation procedures.

This seems strange at first glance, but the problem of restoring x may be solved efficiently even in the case of noncentered correlated and even nonrandom noise [3, 4] by random selection of the matrix Φ. The idea of using random regressors to suppress the effect of shifting was already suggested by R. Fisher [5] in the form of the randomized principle of experiment design. Besides the problem of experiment design where the regressors may be randomized by the experimenter, the random inputs occur in many problems of control, identification, filtration, recognition, and so on (see, for example, [6–10]).

The recurrent algorithms for estimation of x under random inputs were also considered in [6–11]. The rate of convergence of the recurrent algorithms to estimate the regression parameters under random inputs were studied in [12,13] where optimal algorithms with the best possible convergence rate were proposed.

It was concluded in [3] that **randomization** of the process of measurements enables one to

- eliminate the effect of bias,
- reduce the number of iterations and, consequently, observations.

The volumes of information processed have increased dramatically in the XXI century. To a large measure, this is due to the mass transition to processing the flows of two-dimensional (2-D) and three-dimensional (3-D) data. Complexity of the traditional methods of signal quantization grows exponentially with dimensionality. Quantization of the 1-D signals for $N = 10^3$ readings corresponds to 10^6 for the 2-D case and 10^9 for 3-D, which is extremely high. In the modern applications for the digital photo and video cameras, the traditional requirement on the desired measurement frequency (speed, Nyquist rate) is so high that too much data must be substantially compressed before being stored or transmitted. In other applications including the display systems (medical scanners and radars) and high-speed analog-digit converters, the increase in the measurement frequency proves to be too costly.

From the practical point of view, it is extremely interesting to investigate the possibilities of restoration of $x \in \mathbb{R}^N$ from $y \in \mathbb{R}^m$ for $m \ll N$, which is of course unrealizable in the general case. However, a new paradigm of **compressive sensing** came recently to the aid of the traditional signal processing theory. It allows one to restore with a sufficient accuracy the sparse information x [14,15]. During the last five years the term *compressive sensing* was brought into common use.

The main purpose of the present paper lies in explaining this new paradigm which relies on the nonadaptive linear projections retaining the signal structure and enabling one to restore the information using, for example, the methods of ℓ_1-optimization: x is determined as the solution of a problem like

$$||x||_1 = \sum_j |x[j]| \to \min \ : \quad y = \Phi x.$$

The new methodology is based on a certain—usually randomized—selection of the matrix Φ and on the fact that the vector $\hat{x}$ resulting from ℓ_1-optimization has at most m nonzero components, that is, is strongly sparse. This remarkable fact was established and used in seismology [17, 18] and, independently of these papers, in the yearly essay of one of the present authors who solved the problem of constructing the ℓ_1-optimal stabilizing controller of the nonminimum-phase plant that was initially presented in [19]. An extended explanation of the fact why only a small number—equal to the codimension of the subspace $\Phi x = 0$—of the components of the vector of solution of the problem of ℓ_1-optimization is not equal to zero was given with a detailed geometrical interpretation in "Automation and Remote Control" in 1984 [20]. Three years later, a similar result was published in [21].

A formalized statement of the problem with explanations and examples is given in Section 2. Various methods of solution are described and the arising problems analyzed in Section 3. Next, practical examples are considered and the outlooks are discussed in the conclusion. The paper is structured similar to [22].

2. FORMULATION OF THE PROBLEM

Information, data acquisition, control. The famous book of N. Wiener *Cybernetics or Control and Communication in the Animals and the Machines* [23] declares the making of a new science of *cybernetics* where the information–control relation in the phenomena of the material world plays the role of its fundamental property.

Information processing and control decision making may be schematized as follows (see Fig. 1). We assume that *the information x about an important for the researcher variable phenomenon (or control plant, or some element of the environment) mediated by some forms of communication Ψ manifests itself via the signal f.*

The signal f may be observed using some recording instrumentation or special sensors. The process of observation (data acquisition) usually is sufficiently complicated and may be treated as *application of some observation operator A to the signal f. The data y are obtained as the result of interaction with the recording instrumentation and used by the researcher in an attempt to restore (estimate) the information x by generating the estimate $\hat{x}$. The results of estimation characterize to a certain extent the changes occurring in the phenomenon of interest to the researcher and are used then to make one or another control decision.*

The stage of data acquisition plays a key part. All processes and phenomena go on in time. In reality, all signals are of analog nature. Let us denote by t the time instant and by $f(t)$ the "instantaneous" value of the signal f at the time instant t. The simplest example of the operator A is represented by acquiring a set of "instantaneous" values of the signal f at m points $t_1, t_2, \ldots, t_m$. In this case, the operator A is formally representable as an assignment to the signal f of correspondence of the vector of its convolutions with argument-shifted delta-functions

$$A(f) = \begin{pmatrix} \delta_{t_1} \times f \\ \cdots \\ \delta_{t_m} \times f \end{pmatrix},$$

where $\delta_{t_n}(t) = \delta(t - t_n)$, $n = 1, 2, \ldots, m$. The modern information theory originated from the famous *Kotel'nikov theorem* [24] (known internationally as the *Nyquist–Shannon theorem* [25–27]) stating that if an analog signal $f : \mathbb{R} \to \mathbb{R}$ from $L_2(\mathbb{R})$ has a limited spectrum, then it can be restored uniquely without losses from its discrete readings taken at a frequency greater than the doubled maximal frequency of the spectrum. This theorem in many cases enables one to identify the analog signal $f(t)$ with the corresponding set of discrete values $f[n]$, $n = 1, 2, \ldots, N$.

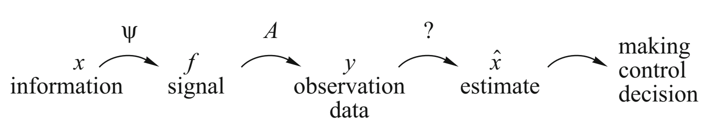

Fig. 1. Diagram of data acquisition and processing.

The instantaneous data are an idealization of a kind. The real registering instruments interact for some time with the arriving input and only then output the result which in fact is a certain integral characteristic of the signal over the registration interval. In the general case, the results of measuring the signal f from the space of functions $L_2(\mathbb{R})$ (data acquisition) may be treated as a scalar product in $L_2(\mathbb{R})$ with some observation function a which is a convolution of f with a.

In many practical applications, the original notion of information x may be described mush simpler than the actual signals f observed by the researcher. For example, to make a decision in some control system, one needs to know that a signal in the form of an acoustic or electromagnetic wave appeared in the registering channel. Of interest is a one-bit answer to the simple *yes/no* question, whereas the arriving and registered signal may have complex form and be distributed in time and space (multidimensional vector). The rapidly progressing new paradigm of information processing relies namely on such specificity.

Compressed signals. Let us consider a real one-dimensional finite-length discrete signal f.[1] Its values make up an $N\times 1$ column vector in $\mathbb{R}^N$ with the elements $f[n],\ n = 1, 2, \dots, N$. (We consider the 2-D images or high-dimensionality signals as vectorized in a long one-dimensional vector.) Any signal in $\mathbb{R}^N$ may be expanded in some basis of the $N \times 1$-dimensional vectors $\{\psi_j\}_{j=1}^N$. We assume for simplicity that this basis is orthonormalized. With the use of an $N \times N$ matrix of the basis $\Psi = (\psi_1, \psi_2, \dots, \psi_N)$ with the columns of vectors $\{\psi_j\}$, the signal f is representable as

$$f = \Psi x = \sum_{j=1}^{N} x[j]\psi_j, \tag{1}$$

where x is an $N \times 1$ column vector of the weight coefficients $x[j] = \langle x, \psi_j\rangle = \psi_j^{\mathrm{T}} x$ and T stands for transposition. Obviously, f and x are equivalent representations of the signal. Usually, f is called the representation in the time (or space) domain, and x, in the spectral or transformed Ψ-domain.

The signal f is referred to as *s-sparse* if it is a linear combination only of s basic vectors, that is, only s components $x[j]$ in (1) are other than zero and the rest $(N - s)$ of them are zero.

Of interest is the case where $s \ll N$. For such s-sparse signals, one may assume that it is namely their representation in the corresponding Ψ-domain that is the essential information that they carry. This information is defined uniquely by two sets of s natural (indices) and real (values) numbers.

Along with the definition of the s-sparse signals, we use a more general notion of the compressed signal. The signal f is called *compressed* if it has a representation like (1) where only several components $x[j]$ are large enough, the rest of them being small.

Remarks. It deserves noting that the requirement of the sparse signal representation in some basis is frequent but not obligatory. It was omitted in [28] and replaced by the requirement for existence of a manifold on which f lies. In other works ([29], for example), the requirement on sparseness in some basis is relaxed to a sparse representation in some redundant dictionary. Apart from sparse representation, additional constraints such as generation of this representation by a random Markov field [30] are sometimes applied to the signal. A case where vectors y and x are parameterized by an infinite countable or even noncountable set was discussed in [31].

Transforming coding and its inefficiency. The fact that the compressed signals are well approximated by the s-sparse representations underlies the transforming coding with which many encounter even in the everyday life using the JPEG, MP3, and MPEG formats of images, audio, and video. In the data acquisition systems (for example, digital photo and video cameras), the transforming coding is pivotal: the assembly of the transformation coefficients $\{x[j]\}$ is calculated from the ob-

[1] Use of the telecommunication term *signal* is a tribute to tradition. In real fact, the matter may concern data of any nature.

tained full N-sample of the signal f via $x = \Psi^{\mathrm{T}} f$, then s large components are localized in x, the rest $N - s$ of them being rejected (see, for example, [32] for detail). As the result, coded are s values themselves and the numbers of their positions in the vector x.

One of the following orthogonal transforms is usually employed in image processing: discrete Fourier transform (DFT), discrete cosine transform (DCT) or discrete wavelet-transform (DWT).

We recall that DFT of the N-dimensional signal f set down as a vector is defined as a vector $\hat{f}$ with the components $\hat{f}[k] = \sum\limits_{j=0}^{N-1} f[j] e^{-2\pi \mathrm{i} jk/N}$, where i is the imaginary unit. It is an invertible linear transformation with matrix:

$$\Upsilon = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-\frac{2\pi \mathrm{i}}{N}} & e^{-\frac{4\pi \mathrm{i}}{N}} & e^{-\frac{6\pi \mathrm{i}}{N}} & \dots & e^{-\frac{2\pi \mathrm{i}}{N}(N-1)} \\ 1 & e^{-\frac{4\pi \mathrm{i}}{N}} & e^{-\frac{8\pi \mathrm{i}}{N}} & e^{-\frac{12\pi \mathrm{i}}{N}} & \dots & e^{-\frac{2\pi \mathrm{i}}{N}2(N-1)} \\ 1 & e^{-\frac{6\pi \mathrm{i}}{N}} & e^{-\frac{12\pi \mathrm{i}}{N}} & e^{-\frac{18\pi \mathrm{i}}{N}} & \dots & e^{-\frac{2\pi \mathrm{i}}{N}3(N-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-\frac{2\pi \mathrm{i}}{N}(N-1)} & e^{-\frac{2\pi \mathrm{i}}{N}2(N-1)} & e^{-\frac{2\pi \mathrm{i}}{N}3(N-1)} & \dots & e^{-\frac{2\pi \mathrm{i}}{N}(N-1)^2} \end{pmatrix}.$$

DCT is an orthogonal transform closely related with the Fourier transform but operating with real numbers. It is defined by a similar matrix with the elements $\cos\, k\left(j + \dfrac{1}{2}\right)\dfrac{\pi}{N}$.

DWT is based on the convolution of the signal with filters of high and low frequencies with successive double sparsing of the number of readings. At that, the sparsed result of convolution with the low-frequency filter may be regarded as a reduced copy of the original signal which was purged of noise and over which the given operation is repeated. DWT is also a linear operator.

The corresponding two-dimensional invertible linear transforms are defined in a similar manner. For example, the two-dimensional Fourier transform of the vector $N_1 \times N_2$ is defined as follows:

$$\hat{f}[k_1, k_2] = \sum_{j_1=0}^{N_1-1} \sum_{j_2=0}^{N_2-1} f[j_1, j_2] e^{-2\pi \mathrm{i} j_1 k_1/N_1} e^{-2\pi \mathrm{i} j_2 k_2/N_2}.$$

It is true for the majority of the real-world images that application of these transforms provides transformed signals with a substantially smaller number of components.

Unfortunately, the transforming coding where a full sample is first obtained and then compression is performed has three inherent disadvantages:

- First, the initial size of the sample N may be extremely large even if the size s of the resulting assembly of components is small.
- Second, all N transformation coefficients $\{x[j]\}$ must be calculated even though their majority, except for s coefficients, will be rejected.
- Third, the positions of s large coefficients must be additionally coded.

Problem of compressive sensing. The aforementioned disadvantages of the transforming coding are eliminated by using the compressive sensing (CS) owing to the direct determination of the compressed representation of the signal without the intermediate determination of the N-sample [14, 15].

Let us consider the general linear measurement process calculating $m < N$ internal scalar products of f and a collection of vectors $\{a_i\}_{i=1}^m$:

$$y[i] = \langle a_i, f \rangle.$$

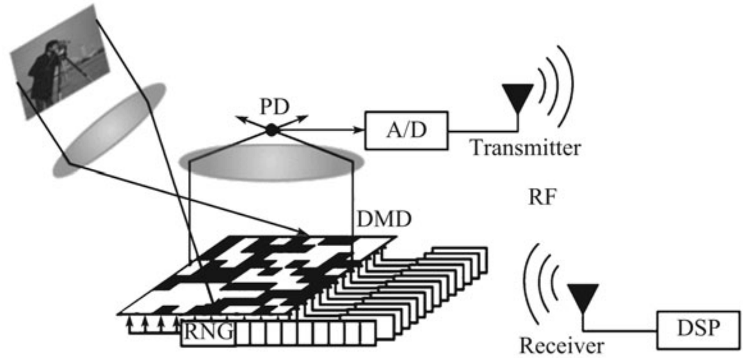

Fig. 2. Scheme of the single-pixel camera [34].

We collect the results $y[i]$ in an $m \times 1$ column vector y and generate the rows of the $m \times N$ matrix A from the transposed measurement vectors a_i^{T}. By substituting Ψ from (1), the expression of y may be rearranged in

$$y = Af = A\Psi x = \Phi x, \tag{2}$$

where $\Phi = A\Psi$ is the $m \times N$ matrix.

Another possible approach to measurement is described in [33] where each sensor measures only one signal bit, the sign of projection of f on a_i:

$$y[i] = \mathrm{sgn}\langle a_i, f\rangle.$$

Instead of condition (2), in this case consideration is given to the condition $y^{\mathrm{T}}\Phi x \geq 0$.

We notice that the considered measurement processes are nonadaptive in the sense that the matrix A is fixed and independent of the signal f.

The problem of compressive sensing lies in designing

- a universal measurement matrix A such that the essential information about any s-sparse (or compressed) signal is not damaged at reducing the dimensionality from $f \in \mathbb{R}^N$ to $y \in \mathbb{R}^m$, and
- a reconstruction algorithm restoring x—and, consequently, f—only from $m \sim s$ measurements (or approximately the same number of measurements as the transmitted number of coefficients at the traditional transforming coding).

One needs not to regard (2) as a process of "multiplying" the signal f by the matrix A. It deserves noting that in the CS paradigm this measurement process is not usually carried out separately, but is part of the physical process of data acquisition, which allows one to do without excessive computer power and save power in the sensing devices. Particular form of the matrix A is important in the calculations for signal restoration from the acquired compressed data. For example, if in a problem there is a technical opportunity for direct determination of the coefficients of the Fourier transform of f, then the corresponding vector y is obtained directly from a selected assembly of m indices, and the rows a_i^{T} of the matrix A from (2) are the corresponding rows of the matrix of the Fourier transform. Another illustrative example may be found at the WEB site `http://dsp.rice.edu/cscamera` where consideration is given to a single-pixel compressing digital camera [34, 35] obtaining directly (in analog manner by using only lenses and mirrors) m random projections of the desired image without preliminary collection of the values (data acquisition) of all its N pixels. As can be seen in Fig. 2, the light waves emerging from the picture f are reflected by a special device of digital micromirrors (DMD) consisting of an array of N tiny mirrors. Such DMD devices can be found in many computer projectors and projection TV sets. The reflected light is collected by the second lens and focused on a single-pixel photodiode (PD). Each micromirror can be oriented independently either in the direction of the photodiode (corresponds to one) or aside (corresponds to zero). To collect the measurements, pseudorandom orientations of the micromirrors creating the measurement vector a_i of ones and zeros are established using a random number generator (RNG). The photodiode voltage is then equal to $y[i]$ which is the internal product of a_i and the image f, that is, the reflected rays are "summed optically." To obtain all components of y, the process is repeated m times. In this way, we obtain the measurement vector y of (2), multiplication by matrix being a part and parcel of a special physical process of measurement.

3. SOLUTION

Design of a universal measurement matrix. The measurement matrix A should enable reconstruction of the signal f of length N from a smaller number of measurements $m < N$ (from the vector y). Since $m < N$, this problem is ill-conditioned. However, if f is s-sparse and one is aware of the locations of the s other-than-zero components x, then the problem may be solvable for $m \geq s$. The necessary and sufficient condition for solvability of this simplified problem lies in that the inequalities

$$\lambda^{-1}||z||_2 \leq ||\Phi z||_2 \leq \lambda||z||_2 \tag{3}$$

must be satisfied for some $0 < \lambda < \infty$ and any nonzero vector z where the same s nonzero components as in x are specified, that is, the matrix $\Phi = A\Psi$ must retain the lengths of such specific in a sense s-sparse vectors. (Here and below, $\|z\|_\rho = \left(\sum_{j=1}^{N} |z[j]|^\rho\right)^{\frac{1}{\rho}}$, $\rho = 1, 2$.) In the general case, of course, the positions of the s other-than-zero components of x are unknown. However, satisfaction of condition (3) for arbitrary $2s$-sparse vectors z suffices for stable solution of the problem for any s-sparse vector x. One can make sure of the validity of this fact using the proof by contradiction. The following unique decoding rule may be used: among all vectors x such that $y = \Phi x$, we select that one with the least number of nonzero coefficients. Let a problem have two different solutions x' and x''. Obviously, they have at most s nonzero components each. Since x', x'' are s-sparse vectors, $\bar{x} = x' - x''$ is a $2s$-sparse vector. In virtue of linearity, $\Phi\bar{x} = \Phi x' - \Phi x'' = 0$ and, therefore, we have $\bar{x} = 0$ in virtue of (3), that is, $x' = x''$, which is a contradiction.

The literature on CS most frequently uses another condition allied to (3) which is called the *restricted isometry property* (RIP) [36]: the $m \times N$ matrix Φ has RIP(δ, m) with the parameters $\delta \in (0, 1)$ and $m \in \mathbb{N}$ if

$$\sqrt{1-\delta} \leq \frac{||\Phi z||_2}{||z||_2} \leq \sqrt{1+\delta} \tag{4}$$

is satisfied for any nonzero m-sparse vector z. Condition (3) is called in the literature the *modified RIP* (MRIP).

Along with RIP-like properties, CS makes use of the condition

$$\mu(A, \Psi) = \sqrt{N} \max_{i,j} \frac{|\langle a_i, \psi_j \rangle|}{||a_i||_2}$$

for smallness of mutual dependence $\mu(A, \Psi)$ of the rows $\{a_i\}$ of the matrix A and the columns $\{\psi_j\}$ of the matrix Ψ which is called the *noncoherence* of A and Ψ. To satisfy it, the rows of the matrix A should represent the columns of Ψ (and vice versa) as a linear combination with the majority of the coefficients being zeros. CS uses widely the fact that the random matrices A with great probability are strongly noncoherent with any fixed basis Ψ. For example, if the rows of the matrix A represent a random sample of an orthonormalized basis obtained by orthogonalization of N random vectors selected uniformly and independently from the unit sphere, then $\mu(A, \Psi) \approx \sqrt{2 \log N}$. A sufficient condition for the possibility of highly probable precise restoration from m observations with the matrix A was obtained in [37] for the s-sparse vectors:

$$m \geq c\,\mu(A, \Psi)^2\, s \log N, \tag{5}$$

where c is a positive constant. For the above random matrix A, we have $m \geq 2cs(\log N)^2$.

As was indicated in [38], systems with $m \approx 4s$ prove to work well in practice.

The conditions $\mathrm{RIP}(\delta, 2s)$ and $\mathrm{MRIP}(\lambda, 2s)$ are nonrobust in the sense that they are not sufficient for restoration of

- an arbitrary s-sparse signal under noisy observations y

or

- a compressed signal with small nonzero $(N - s)$ components.

In these cases, the conditions $\mathrm{RIP}(\delta, 3s)$ or $\mathrm{MRIP}(\lambda, 3s)$ are sufficient in the sense (see [38]) that

$$||x - \hat{x}||_1 \leq \text{const}\; ||x - x^\star||_1,$$

where $\hat{x}$ is the result of restoration of x and $x^\star$ is the vector obtained from x by zeroing all components, with the exception of s greatest-in-magnitude components. In this case, the condition for great degree of noncoherence is overconservative because the permissible value of $\mu(A, \Psi)$ is related with $\mathrm{RIP}(\delta, 3s)$ as $\delta = (m - 1)\mu(A, \Psi)$.

Direct construction of the measurement matrix A such that $\Phi = A\Psi$ features RIP requires verification of condition (4) for each of $C_N^s = \dfrac{N!}{s!(N-s)!}$ possible combinations of the positions of s other-than-zero components in the vector z of length N. However, it turned out that RIP may be satisfied with high probability just by taking a random matrix as A (randomization of observations). At that, the vector of measurement results y is an assembly of m different linear combinations of the components of f with randomly selected weights.

The random $m \times N$ measurement matrix A with independent identically distributed (i.i.d.) elements $a[i, j]$ and normal distribution density with zero means and variance $1/m$

$$a[i, j] \sim \mathcal{N}\left(0, \frac{1}{m}\right)$$

has two interesting and useful properties [14]:

- if $0 < \delta < 1$ and

$$m \geq c_1\, s \log(N/s), \tag{6}$$

  then, the matrix A satisfies $\mathrm{RIP}(\delta, m)$ with the probability $\geq 1 - 2e^{-c_2 m}$, where $c_1, c_2 > 0$ are small constants depending only on δ (consequently, s-sparse and compressible signals of length N can be restored with high probability only from $m \ll N$ random measurements);
- the matrix A is **universal** in the sense that at reduction of dimensionality from $f \in \mathbb{R}^N$ to $y \in \mathbb{R}^m$ not only essential information about any s-sparse (or compressed) signal is not damaged, but also the matrix $\Phi = A\Psi$ is random with normally distributed i.i.d. elements; therefore, Φ will feature $\mathrm{RIP}(\delta, m)$ with the same high probability independently of the method of selecting the orthonormalized basis Ψ.

The result about satisfying condition $\mathrm{RIP}(\delta, m)$ with high probability from property (6) of the random matrix A relies largely on the earlier publications [39, 40].

Other random measurement matrices such as random sample of the i.i.d. elements $a[i,j]$ from the symmetrical Bernoulli distribution

$$\mathrm{P}(a[i,j] = \pm 1/\sqrt{m}) = \frac{1}{2}$$

or a similar distribution

$$a[i,j] = \begin{cases} +\sqrt{3/m} & \text{with the probability } \dfrac{1}{6} \\ 0 & \text{with the probability } \dfrac{2}{3} \\ -\sqrt{3/m} & \text{with the probability } \dfrac{1}{6} \end{cases}$$

may be used for CS.

It was shown that property (6) can be easily derived from the Johnson–Lindenstrauss lemma [42] according to which any set of n points in the d-dimensional Euclidean space may be embedded in the m-dimensional Euclidean space so that $m \sim \log n$ and is independent of d and all mutual pairwise distances are approximately retained [41]. For the aforementioned three last methods of generation of the random matrix A, its universality was substantiated and particular conditions for selection of the constants c_1 and c_2 were given in [41]:

$$c_2 \leq c_0(\delta/2) - c_1 \left(1 + \frac{1 + \log(12/\delta)}{\log(N/s)}\right), \quad c_0(\delta/2) = \frac{\delta^2}{16} - \frac{\delta^3}{48}.$$

The proof of the property $\mathrm{RIP}(\delta, m)$ for the random matrix Φ in [41] follows the lines of [43] in using the possibility of covering each of the C_N^m possible m-dimensional subsets of the set $\{z:\ \|z\|_2 = 1\}$ by at most $(12/\delta)^m$ spheres of the radius $\delta/4$ and relies on the proof that the expectation

$$E\|\Phi z\|_2^2 = \|z\|_2^2 \ \ \forall z \in \mathbb{R}^N$$

observation and the random variables $\|\Phi z\|_2^2$ are strictly concentrated about $\|z\|_2^2$:

$$\mathrm{Prob}\left\{ |\|\Phi z\|_2^2 - \|z\|_2^2| \geq \frac{\delta}{2}\|z\|_2^2 \right\} \leq 2e^{-mc_0(\delta/2)}.$$

Other randomized methods of generation of the measurement matrix are also proposed in the literature (see, for example, [15, 38]): uniform random sample of N columns on the unit sphere in $\mathbb{R}^m$; random sample of the projector P and its normalization $A = \sqrt{\dfrac{N}{m}}P$ or choice by some other sub-Gaussian distribution. Satisfaction of $\mathrm{RIP}(\delta, m)$ is supported in each of these cases by condition (6) with an appropriate constant c_1 depending on the selected method of generating the random matrix A.

Storage of $m \times N$ elements of the matrix A requiring $\mathcal{O}(mN)$ memory units becomes a challenge with increase in N. For simple rules of generation of A, its impact may be reduced by using, for example, a "repeated" generator of pseudorandom numbers and again calculating each time A. The random "cuts" of m rows from the $N \times N$ matrices of the discrete Fourier transform (DFT) or the Vandermonde matrices corresponding to interpolation at different N points offer other examples of the matrices satisfying the RIP and MRIP conditions. The size of memory for such random matrices may be reduced to $\mathcal{O}(m \log N)$. If the rows of matrix A are a random sample of transposed columns of an orthonormalized basis obtained by orthogonalization of N random vectors selected uniformly and independently from the unit sphere $\mathbb{R}^N$, then to satisfy the property RIP(δ, m) with high precision, it suffices to take m:

$$m \geq Cs\,(\log N)^4.$$

The difficulty of verifying RIP for larger m is its most substantial disadvantage. Active attempts being made to replace it by more constructive properties. An interesting example of such possibility is described in [44].

Design of a signal reconstruction algorithm. The algorithm of signal reconstruction must reconstruct the signal f of length N or, which is the same, its corresponding sparse vector of coefficients x from m measurements (vector y), random measurement matrix A (or random law of its generation) and basis Ψ. Since $m < N$ in (2), for the s-sparse signals there are infinitely many x' satisfying $\Phi x' = y$, which is due to the fact that if $\Phi x = y$, then $\Phi(x + r) = y$ for any vector r from the zero subspace $\mathcal{N}(\Phi)$ of the matrix Φ. Therefore, the aim of the signal reconstruction algorithm is to determine a vector of the coefficients of a sparse representation of the signal in $(N-m)$-dimensional shifted zero subspace $\mathcal{H} = \mathcal{N}(\Phi) + x$.

Reconstruction by minimization of the ℓ_2-norm. A classical approach to the inverse problems of the type at hand lies in determining in the shifted zero subspace $\mathcal{H}$ a vector with the least norm (power) ℓ_2

$$\hat{x} = \arg\min ||x'||_2 \quad : \qquad \Phi x' = y.$$

Solution of this optimization problem may be conveniently represented in the form of the least-squares method (LSM) as follows:

$$\hat{x} = \Phi^{\mathrm{T}}(\Phi\Phi^{\mathrm{T}})^{-1}y.$$

Unfortunately, the result of ℓ_2-minimization almost never is an s-sparse vector and has many other-than-zero elements.

Reconstruction by minimization of the ℓ_0-norm. Since the ℓ_2-norm measures the signal power and not its sparseness, it is possible to consider the problem of minimization of the ℓ_0-norm which calculates the number of nonzero components

$$||x||_0 = |\{i | x[i] \neq 0\}|.$$

(Strictly speaking, ℓ_0 is not a norm because it does not satisfy the property of uniformity, but the term "ℓ_0-norm" is widely used.)

By using only $m = s + 1$ random i.i.d. measurements [45], one can restore an s-sparse signal with high probability from the solution of the modified optimization problem

$$\hat{x} = \arg\min ||x'||_0 \quad : \quad \Phi x' = y.$$

Unfortunately, the problem of optimization of the ℓ_0-norm is nonconvex and of the combinatorial type, the computational procedures of its solution are numerically unstable and NP-hard and require an enormous enumeration of all C_N^s possible variants of allocation of the nonzero elements in x.

Reconstruction by minimization of the ℓ_1-norm. It may seem strange, but optimization based on the ℓ_1-norm

$$\hat{x} = \arg\min ||x'||_1 \quad : \quad \Phi x' = y,$$

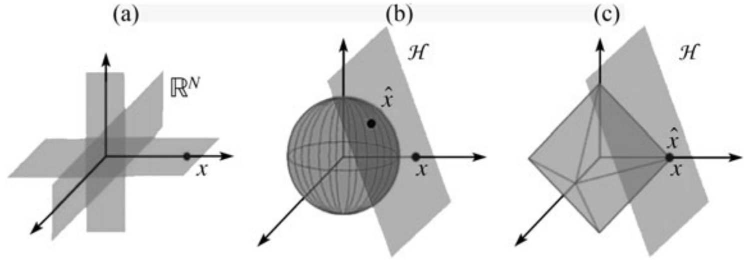

Fig. 3. (a) The space of $2s$-sparse vectors in $\mathbb{R}^3$ consists of three planes having two coordinate axes each; (b) ℓ_2-minimization determines the nonsparse vector $\hat{x}$ distant from x; (c) ℓ_1-minimization determines the sparse point $\hat{x}$ of contact of the ℓ_1-sphere with the hyperplane $\mathcal{H}$ coinciding with x with high probability.

enables one to reconstruct precisely and with high probability the s-sparse signals and approximate well the compressed signals using only $m \geq c_1\, s \log(N/s)$ i.i.d. random measurements [14, 15]. It is a problem of convex optimization reducible to that of linear programming known as the basis pursuit [46]:

$$\langle c, \tilde{x}\rangle \to \min \quad : \quad \tilde{\Phi}\tilde{x} = y, \quad \tilde{x} \geq 0,$$

where

$$c = (1, 1, \ldots, 1)^{\mathrm{T}}, \quad \tilde{x} = \begin{pmatrix} x_+ \\ x_- \end{pmatrix}, \quad \tilde{\Phi} = (\Phi, -\Phi).$$

Solution may be based on the method of internal point (computational complexity of the order of $\mathcal{O}(N^3)$) or the simplex method (theoretical exponential complexity) which in practice proved to be relatively fast.

Geometrical interpretation. Geometrical representation of the problem of compressive sensing in $\mathbb{R}^N$ illustrates why solution of the problem of ℓ_2-optimization does not give aid in restoration of the original sparse signal, but the signal can be reconstructed with the use of ℓ_1-optimization.

The set of all s-sparse vectors x in $\mathbb{R}^N$ is, as shown in Fig. 3a for $N = 3$, a strongly nonlinear space consisting of all s-dimensional hyperplanes stretching along the coordinate axes. The shifted zero subspace $\mathcal{H} = \mathcal{N}(\Phi) + x$ is oriented at a random angle defined by randomization in the matrix Φ (see Fig. 3b). The result of minimization of the ℓ_2-norm is a point on $\mathcal{H}$ which is nearest to the origin and can be determined by extending the hypersphere (ℓ_2-sphere) until touching $\mathcal{H}$. Owing to the random orientation of $\mathcal{H}$, this point $\hat{x}$ which is nearest to the origin will be with high probability far from the coordinate axes and, consequently, neither s-sparse nor close to the correct answer x. On the contrary, the ℓ_1-sphere in Fig. 3c is a convex combination of points lying on the coordinate axes. Therefore, at expansion of the ℓ_1-sphere in the three-dimensional space, the point lying on the coordinate axes ($s = 1$) and coinciding precisely with the point of location of the desired sparse vector x contacts first with the shifted two-dimensional zero subspace $\mathcal{H}$. (In practice, $N, m, s \gg 3$ and the intuitive three-dimensional reasoning should not lead astray.)

Other reconstruction algorithms. In the practical problems, reduction of the problem of ℓ_1-optimization to that of linear programming is not satisfactory because of great computational complexity. Other methods using the ℓ_1 norm such as the ℓ_1 regularized LSM or the so-called LASSO approach proposed in [47] and, independently, in [46] as the basis pursuit de-noising (BPDN)

$$\hat{x}_\lambda = \arg\min ||\Phi x' - y||_2^2 + \lambda ||x'||_1$$

are considered in the literature along with this problem. Another variant based on passing to the dual problem was proposed in [48].

Smoothed ℓ_0-norm (approximation of ℓ_0-norm). A method based on approximating the delta-function $\delta(t)$ by a smooth function $e^{-\frac{t^2}{2\sigma^2}}$ for $\sigma \to 0$:

$$||x||_0 = N - \sum_i \delta(x[i]) \approx N - \sum_i e^{-\frac{(x[i])^2}{2\sigma^2}} = N - F(x, \sigma)$$

was proposed in [49]. The nonconvex optimization problem

$$\hat{x}_\sigma = \arg\max F(x', \sigma) \quad : \quad \Phi x' = y$$

is solved for each σ. Methods of local optimization can determine solution only in the case of a good initial approximation. It was proposed in [49] instead of solving this problem to take as $\hat{x}_\sigma$ for each value of σ the result of several iterations of the algorithm of projection of the gradient [50]. Solution of the problem of restoration is obtained as the limit of $\hat{x}_\sigma$ for $\sigma \to 0$.

Minimization-based reconstruction of the nonconvex regularizers. Other ℓ_ρ-norms that are used for $0 < \rho < 1$ in addition to the ℓ_1-norm [51–53] are distinguished for better approximation of the ℓ_0-norm. In particular, an example where $\ell_{1/2}$ is substantially superior to ℓ_1 in terms of restoration performance may be found in [51]. Their weak point lies in that they are not convex functions, which may result in hitting undesirable local minima. This disadvantage is suppressed by using methods like graduated non-convexity [52].

Iterative algorithms. Some publications suggest to use the iterative weighing algorithms for restoration of the original signal. The weighted ℓ_ρ-norms for approximation of the ℓ_0-norm $(\sum_i w_i(x[i])^\rho)^{\frac{1}{\rho}}$ with ρ equal to one or two were considered in [54–56]. For example, for $\rho = 2$ the problem of minimization of the weighted norm

$$\hat{x}_k = \arg\min (x')^{\mathrm{T}} W_k x' \quad : \quad \Phi x' = y,$$

where $W_k = \mathrm{diag}(w_k[1], \ldots, w_k[N])$, is solved in [54] analytically as follows:

$$\hat{x}_k = {W_k}^{-1} \Phi^{\mathrm{T}} (\Phi {W_k}^{-1} \Phi^{\mathrm{T}})^{-1} y.$$

Then,

$$w_{k+1}[i] = (\hat{x}_k[i]^2 + \epsilon)^{p/2-1},$$

where $\epsilon > 0$ is a small value, are selected iteratively as the following set of the weight coefficients. The solution $\hat{x}$ is established as the limit $\hat{x}_k$.

To solve the problem of restoration of the original signal, [57–61] proposed a number of iterative algorithms like iterative shrinkage thresholding (IST):

$$\hat{x}_{k+1} = (1 - \beta)\hat{x}_k + \beta H \left(\hat{x}_k + \Phi^{\mathrm{T}}(y - \Phi \hat{x}_k) \right),$$

where H is the so-called denoising function, Moreau proxi-map [61–63] as a rule having the form

$$H(u) = \arg\min_t (\mathrm{dist}(t, u) + \lambda L(t)).$$

Usually, selected is the distance $\mathrm{dist}(t, u) = ||t - u||_2^2$, and the second addend is directly proportional to the regularizer $L(t)$, the penalty function for nonsatisfaction of some property—$||t||_1$ (Huber function), for example.

More efficient modifications of IST proposed in [60,61] are represented by the fast IST algorithm and the two steps IST (TwIST) algorithm based on the method of accelerating the optimization algorithm proposed in [50] under the name of "heavy ball" in the paper it is mentioned under the name of two-step iterative method (TwSIM).

A similar algorithm was considered in [64] for image restoration from a set of the coefficients of Fourier transform. Selected as H was $H(u) = \Upsilon(F(\Upsilon^{-1}(u)))$, where Υ is the discrete Fourier transform and F is the original image denoising algorithm. In distinction to IST, it was proposed in [64] to add to $\hat{x}_k$ at each iteration an additional random disturbance whose irregular part is filtered by the F algorithm with parameters corresponding to the added disturbance.

Robust algorithms are among the fastest tools for restoration of x from the measurements $y = \Phi x$. The simplest of them—matching pursuit (MP) [46, 65]—at each step specifies in y the most correlated row Φ. This simple idea underlies more complicated robust algorithms having as a rule the following structure:

(1) Initialize $x_0 = 0$, $r_0 = y$, $k = 1$, $\Lambda_0 = \emptyset$.
(2) Estimate the residue $r_k = y - \Phi x_{k-1}$.
(3) Establish from r_k, Φ, and y a new estimate x_k. Each algorithm has its own such step, but usually is as follows:
   (a) calculate by any method one or more current indices $\{\lambda_{k1}, \ldots, \lambda_{kn_k}\}$ of the nonzero components of the vector x that were not yet included in Λ_k, $\Lambda_k = \Lambda_{k-1} \cup \{\lambda_{k1}, \ldots, \lambda_{kn_k}\}$;
   (b) compile the matrix Φ_k of the columns of Φ with the corresponding indices from Λ_k;
   (c) estimate the nonzero values of the vector x_k (we denote $\hat{x}_k = x_k[\Lambda_k]$): $\hat{x}_k = \arg\min ||\Phi_k \hat{x}_k - y||_2$. The components with numbers not belonging to Λ_k are assumed to be zero.
(4) $k := k + 1$, go to step (2) if the stop condition is not met: $||r_k|| <$ *threshold*.

In one of the most popular robust algorithms of the MP family, the orthogonal MP (OMP) [66, 67], a set consisting of the single index

$$\arg\max_k |\langle r_k, \phi_i \rangle|,$$

where ϕ_i is the ith column of the matrix Φ, is taken as Λ_k at step (3a).

The regularized OMP (ROMP) [68] selects the set of indices at step (3a) using two additional steps:

- calculate the set J of s indices with the greatest scalar value $\langle r_k, \phi_i \rangle$;
- of all subsets I of the set J such that

$$|\langle r_k, \phi_i \rangle| \le 2|\langle r_k, \phi_j \rangle| \ \forall i, j \in I,$$

  take that for which $(\Phi^{\mathrm{T}} r)[I]$ has the greatest ℓ_2-norm.

The compressive sampling matching pursuit (CoSaMP) [69] was inspired by the ROMP algorithm and repeats it almost literally except that it takes not all indices i for which $|\langle r_k, \phi_i \rangle|$ exceeds the threshold but exactly $2s$ greatest indices. The vector resulting from the solution of

$$\arg\min ||\Phi_k \hat{x}_k - y||_2$$

by zeroing all components except for s components having the greatest magnitude is taken as the estimate x_k.

The subspace pursuit method [70] is CoSaMP where only s and not $2s$ indices are added at step (3a). In the stagewise OMP (StOMP) algorithm [71], at step (3a) selected is the set of indices i such that

$$|\langle r_k, \phi_i \rangle| > threshold_k,$$

where $threshold_k$ is the threshold defined at the kth step. In the DThresh method [72], the condition for selection of indices is as follows:

$$|\langle r_k, \phi_i \rangle| \geq \frac{\delta||r^i||}{\sqrt{s}}.$$

The tree-based OMP (TOMP) was proposed in [73] to seek for a sparse representation of image by the wavelets.

The lattice matching pursuit (LaMP) algorithm was proposed in [30] for the case where the signal is generated by a random Markov field. Its scheme differs from the general scheme in that, instead of accumulating the indices of the nonzero components in Λ_k, the set of indices is completely re-evaluated according to the most probable configuration of the given Markov field. The algorithms of gradient pursuit and conjugate gradient pursuit were described in [74], and of that of sequentially sparse matching pursuit (SSAMP), in [75].

A method of restoration based on the belief propagation (BP) algorithm used in the graphic models for preset statistical dependencies between the components $x[i]$ was proposed in [76]. In the presence of cycles in such models, the BP algorithm (called in this case the cyclic BP) may converge to a local minimum or not converge at all. However, in the practical problems it usually gives a very good approximation. The combinatorial restoration algorithms are described in [77].

Along with the iterative combinatorial CS algorithm based on using a connectivity matrix of a certain regular graph as the special-form measurement matrix and having the computational complexity of the order of $\mathcal{O}\left(N \log \frac{N}{s}\right)$, [78] also compares in detail different approaches. Another table for comparison of the efficiency of the main signal restoration algorithms is given in [79].

4. PRACTICAL EXAMPLES

(1) Let consideration be given to a spectrally sparse wide-spectrum signal

$$f(t) = \sum_{j=0}^{N} x[j] \exp^{i2\pi jt/N}, \quad t = 0, \ldots, N-1,$$

where N is a very large number, but the number of nonzero components $x[j]$ is less than or equal to s which is assumed to be relatively small, $s \ll N$. Let the signal $f(t)$ be a sum of five sinusoidal functions and noise:

$$\begin{aligned} f(t) = 10.7\sin(2\pi \times 50t) + 20\sin(2\pi \times 120t) + 31.5\sin(2\pi \times 200t) \\ + 23\sin(2\pi \times 300t) + 25\sin(2\pi \times 450t) + error \end{aligned}$$

for $t \in [0, 1)$.

Let us consider a problem where five ($s = 5$) unknown active frequencies from the interval $[0, 500]$ and the corresponding amplitudes. We notice that a set of unknown active frequencies is not necessarily a subset of the sequence of integers. According to the Nyquist–Shannon theory, having defined the possible bandwidth $[0, 500]$, one must take readings with the frequency $2 \times 500 = 1000$, that is, at least $N = 1000$ readings must be taken for the interval $[0, 1)$.

The new CS paradigm with a high probability ensures determination of the information about frequencies and amplitudes from a relatively small $m \sim s\log(N/s)$ sample of values of $f(t)$. For the example under consideration, this value is bounded by $4s\log_2(N/s)$ and can be taken approximately equal to

$$m \approx 4 \times 5 \times \log_2(1000/5)(= 153).$$

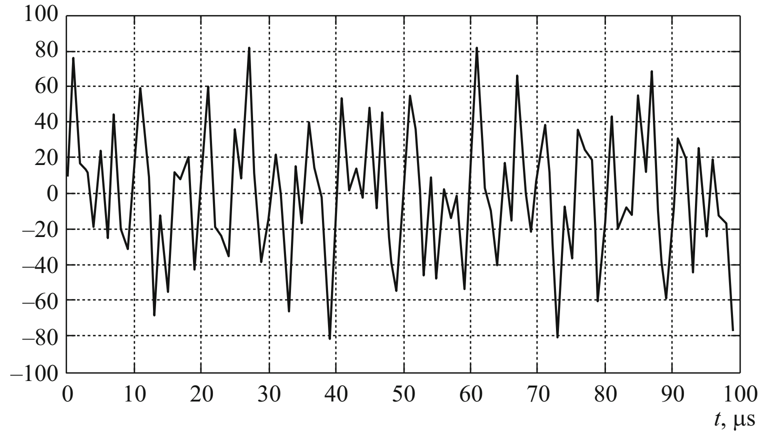

Fig. 4. 5-sparse noisy signal.

Additionally, these values need not be selected in a special manner, the restoration algorithm will work effectively almost with any their set of an appropriate size. Let us assume by way of illustration that the measurements follow the law

$$y = \Phi x,$$

where a random matrix from ± 1 is selected as Φ.

Figure 4 depicts an example of the signal $f(t)$ for $s = 5$ with additional random normally distributed noise having variance 0.1. Figure 5a shown the signal $f(t)$ in the spectral domain. The ℓ_1-regularized least squares method

$$\hat{x} = \min_x ||y - \Phi x||_2^2 + \lambda ||x||_1$$

was used for approximate restoration for $\lambda = 0.01$. The result of such restoration in the spectral domain is shown in Fig. 5b. Despite the substantial reduction in the number of readings (data) and adverse impact of noise together with the approximate nature of the algorithm, the highest-value frequencies correspond to their original counterparts.

(2) We follow [34, 35] and as the second example consider the aforementioned single-pixel compressing digital camera receiving directly m random linear measurements without preliminary acquisition of the values of all N pixels of the original image (see Fig. 2). The light waves f emerging from the photograph of a person with camera pass through the first lens and get into the array of two-position micromirrors reflecting light either in the direction of the second lens or aside. The light flow arriving to the second lens is focused at one point of the PD (optically summed), and then this "sum" goes to the analog-digit converter (A/D) and coded by a set of bits as one pixel which is transmitted finally through the data channel. The process is repeated m times to obtain all components of y. Therefore, we obtain the measurement vector

$$y = Af,$$

multiplication by matrix being done owing to a special physical measurement process, rather than in a digital computing device. In the receiver (DSP), the measurement vector is decoded.

Figures 6b and 6c shows the restored images obtained from the original picture (Fig. 6a) with the use of single-pixel camera. This restoration used random measurements in amounts of 20% and 40% as compared with the amount of pixels in the original picture. As was supposed, the images

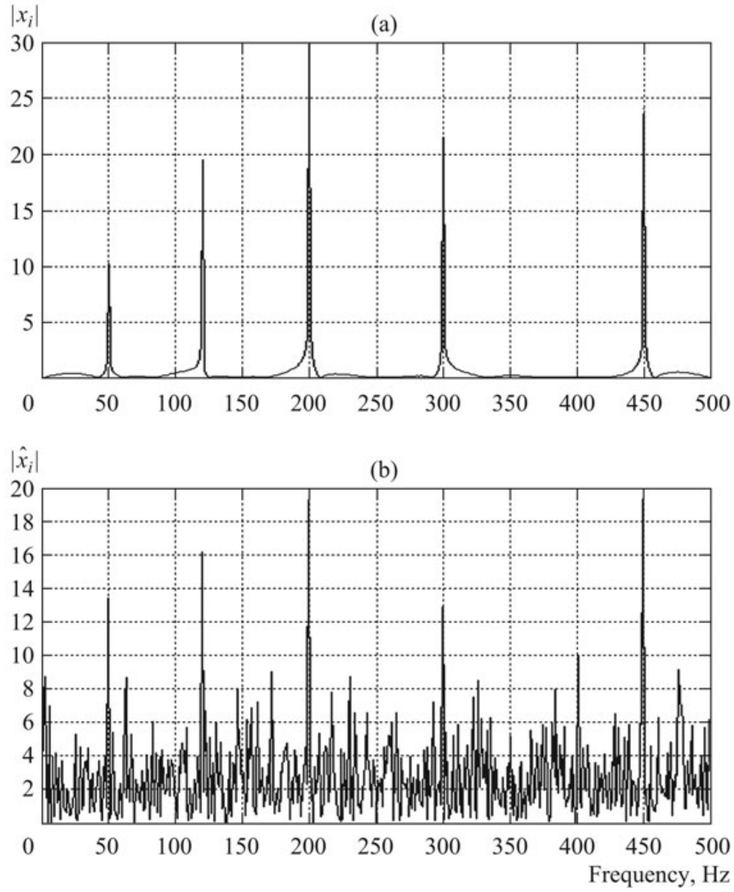

Fig. 5. Determination of frequencies and amplitudes of the 5-sparse signal. (a) The original and (b) restored signals in the spectral domain.

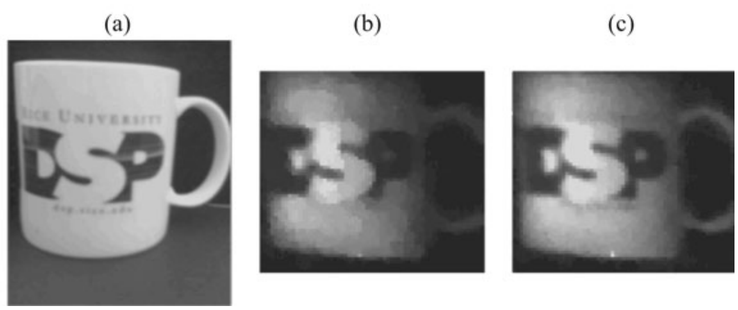

Fig. 6. (a) The ordinary black-and-white digital image ($N = 64 \times 64 = 4096$ pixels); (b) image reconstructed from $m = 800$ random measurements by the single-pixel camera (20 % of measurements); (c) $m = 1600$ (40 % of measurements).

(Figs. 6a–6c) are not identical, quality of (Fig. 6b) is lower, but the relevant details are rendered correctly.

In addition to the requirement on lower amount of measurements, the single-pixel camera may reflect also the waves of such length for which it would be difficult or expensive to construct a great array of sensors. This device may also acquire data through time intervals enabling reconstruction of video [34].

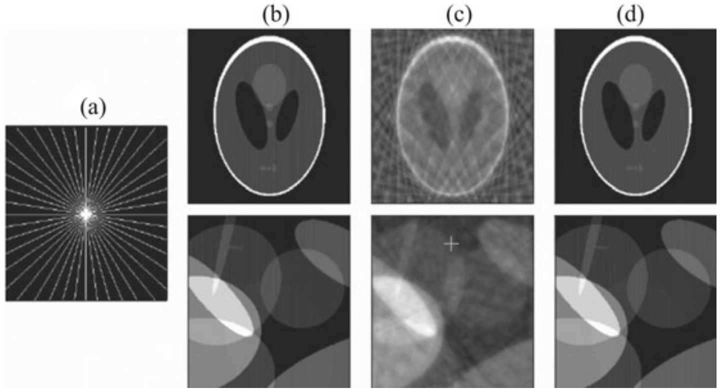

Fig. 7. Precise restoration of the test "Shepp–Logan phantom" images and the set of ellipses. (a) Measurable coefficients in the spectral domain make up radial rays; (b) original test image; (c) restoration with the use of ℓ_2-minimization; (d) restoration with the use of ℓ_1-minimization.

(3) One of the first popular problems where the paradigm of compressive sensing was used to advantage was the medical magnetic resonance tomography (MRT). Models where only some of the coefficients of the two-dimensional Fourier transform that are on the rays emerging from the origin of the spectral domain (Fig. 7a) are measurable were described in [80, 81]. Obviously, such situation obeys the equation

$$y = \Phi x + \xi,$$

where x are all Fourier coefficients represented as vector and the matrix $\Phi = (\phi_0^{\mathrm{T}}, \ldots, \phi_m^{\mathrm{T}})^{\mathrm{T}}$ is constructed from the corresponding unit vectors ϕ_i. The problem lies in restoring the original image of the examined human organ. Examples of restoration of the test images by means of ℓ_2-optimization and ℓ_1-optimization are shown in Figs. 7c and 7d.

The denoising filter F was used in [64]. Therefore, the condition for no noise on the image I is as follows: $F(I) \approx I$, which is rather probable for the desired image. The paper uses a method allied to IST using the denoising function

$$H(x) = \Upsilon(F(\Upsilon^{-1}(x))),$$

where Υ is the discrete two-dimensional Fourier transform represented in the matrix form and x are the Fourier coefficients represented by an appropriate vector.

(4) The fact that the compression algorithms are much less resource-intensive as compared with those of reconstruction proves to be useful at using compressive sensing for compression and restoration. This situation is exactly opposite to that existing no, for example, in the compression of video. The operation of multiplication by a fixed matrix A in (2) is relatively simple (as compared, for example, with the traditional methods of video coding) and may be used for fast coding before data transmission from a low-power device such as the cellular telephone to a high-performance device (server). In such cases, it is advisable to take the matrix A either random or problem-oriented.

The existing standards of video compression define only the decoding algorithm disregarding the method of coding. The single constraint imposed on the compressing algorithm lies in that the result of compression must be decoded by a decoding algorithm. High degree of compression with retention of good quality may require about ten runs of the coding algorithm, high computing power, and time.

In the case of transmitting through the communication channels the video data in the conference mode, the computational complexity is not an overburden for the coder, but in the traditional transmission compression algorithms even in the online mode the coder's computational requirements may exceed those of the decoder by the factor of five to ten.

The situation in compressive sensing is quite the contrary, which allows one to transmit information between mobile devices of limited computing power by compressing at the mobile device, transmitting to the server, and using the server powerful resources to restore information, compress it anew by the classical algorithms, and then transmit to another mobile device.

For example, [82] describes application of the concept of compressive sensing to video compression through simple multiplication by a new matrix A_k of each frame represented by a vector. If all frames are put one to one in a pile, then we get a three-dimensional array f of numbers corresponding to brightness of each pixel or of three-dimensional vectors denoting color. According to [82], this representation is well coded by the three-dimensional wavelets because the difference between the neighbor frames usually is small. The measurement matrix in this case will be block-diagonal:

$$A = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_l \end{pmatrix},$$

and we obtain

$$y = Af.$$

As was already noted, in the case of video f admits sparse representation by the three-dimensional wavelets $f = \Psi x$. By denoting $\Phi = A\Psi$ we obtain the classical problem of restoration of the sparse x: $y = \Phi x$. A pile f of the same frames is obtained by multiplying the determined vector x by A.

(5) The problem of deblurring arises, for example, in astronomy [83]. The blurred image may be regarded as the result of some blurring operator A (convolution of the original image with some window). It was assumed in [83] that this operator is known but, possibly, is not reversible. It is known, however, that the image admits a sparse representation in some basis Ψ.

The case of an unknown blurring convolution operator and noisy measurements

$$y = g \times x + \xi$$

was discussed in [84]. For a number of orthogonal transforms (DFT, DCT, wavelet-transform), the majority of real-world images has sparse representation in this basis. It is one of the key ideas of using the property of sparseness in the problems of high definition.

Another idea is used in [84]. If some blocks are similar, it is advisable to process them jointly. If the corresponding coefficients of the orthogonal transformation of such blocks are piled together, then a 3D array results to which the one-dimensional orthogonal transformation of the third dimension is applied. Solution is sought by minimizing the weighted sum of the error and ℓ_0-norm of the resulting vector using an algorithm of the IST type with $L(t) = \|t\|_1$.

(6) The potentialities of the CS paradigm in learning are described in [85].

The problem of classification lies in classifying some input vector f with one of a few classes d that are known in advance. Learning lies in forming a classification rule on the basis of the learning sample of the pairs (f_k, d_k). At that, for an input close to f_k the classifier should provide an output vector close to d_k.

The input data may be multidimensional vectors of appreciable dimensions, which may hinder construction of a good classifier because of great computational burden, on the one hand, and, on the other hand, with a greater probability lead to defects such as overfitting when f classifies well the examples, but poorly all other, even allied, vectors. Different methods are used to reduce dimensionality. A method based on compressive sensing was proposed in [85]. It turns out that at deciding to what class to assign the input signal, in the case of a priori information about signal sparseness, it is possible to do without signal restoration and confine oneself to solving a similar problem in a space of substantially smaller dimensionality.

Let us consider the problem of classification of high-dimensionality vectors f having a sparse representation in some basis. Let us assume that they may be reasonably divided into two classes denoted by ± 1. The idea relies on the fact that the projections using the random matrix A retain in a sense the distances with a higher probability. Linear separability is an important property of classification. It is retained if the distances are retained. Therefore, if there exists a hyperplane separating the classes, then with a greater probability there will be also a hyperplane separating their projections $y = Af$. As the result, it is suggested to train a classifier separating the vectors $y_k = Af_k$ where the pairs (f_k, d_k), $d_k \in \{-1, +1\}$ are the learning sample. Use of the classifier implies multiplication by the matrix A and classification of the resulting vector.

It deserves noting an important distinction of this variant of using compressive sensing: the compressed vector **needs no restoration**, its processing takes place immediately in the compressed form.

The problem of classification where the object represented by the vector u may immediately belong to more than one of many classes is considered in [86]. The situation of membership to more than one class is described by the vector x of zeros and ones where $x[j]$ is equal to one if u belongs to the class j or zero, otherwise. It is also assumed that the number of classes to which each object belongs is small as compared with the total number of classes. The vector x may be, for example, as follows:

$$x = (0, 0, \ldots, 0, 1, 0, \ldots, 0, 1, 1, 0, \ldots, 0).$$

The classical approach to the construction of such classifier lies in adjusting the parameters θ of the map $f(u, \theta)$—the classifier—with the aim of minimizing the mean risk functional

$$F(\theta) = \sum_k (f(u_k, \theta) - x_k)^2,$$

where the pairs (u_k, x_k) are some learning sample. In the case of a great number of classes, the vectors x_k have a corresponding great dimension, which makes the problem a computational burden.

The new approach described in [86] suggests to adjust the parameters θ by means of a test base of the pairs (u_k, y_k), where $y_k = \Phi x_k$ and Φ is a random matrix, and not of the pairs (u_k, x_k). As soon as such classifier is constructed and some object u is fed to the input, y is calculated as $f(u, \theta)$, then the problem of restoration of the sparse x is solved under the condition that $y = \Phi x$ using robust algorithms because of interest are the classes themselves and not the precise values of x. Thus, complexity of computations is transferred from the phase of learning to that of use.

(7) Other applications are exemplified by detection of metal artefacts in stomatological applications [87], analog-information transformation [88], or reception of the radar images [89].

5. CONCLUSIONS

Acquisition of information on the basis of compressive sensing may be more efficient than the traditional sampling of rare or compressed signals. The popular estimate by the least square method is inadequate in CS for good signal reconstruction. Therefore, other type of convex optimization are used. The domain of application of compressive sensing have recently come far beyond the limits of the coding/decoding theory and embraces now the problems of image classification and processing. In some applications, the actual arriving signals require no restoration at all, and the data are processed only in the compressed form.

The main part of the present paper focused on the discrete signals f, but the paradigm of compressive sensing is applied also to the s-sparse (or compressed) analog signals $f(t)$ that can be represented or approximated using only s of N possible elements from some continuous basis or dictionary $\{\psi_j(t)\}_{j=1}^N$. Whereas each basic element $\psi_j(t)$ may have a great scatter of frequencies (and, consequently, high Nyquist rate), the analog signal $f(t)$ has only s degrees of freedom and, therefore, may be measured at a substantially smaller number of points [88, 90].

REFERENCES

1. Granichin, O.N. and Kiyaev, V.I., *Informatsionnye tekhnologii v upravlenii* (Information Technologies in Control), Moscow: Binom, 2008.
2. Ermakov, S.M. and Zhiglyavskii, A.A., *Matematicheskaya teoriya optimal'nogo eksperimenta* (Mathematical Theory of Optimal Experiments), Moscow: Nauka, 1987.
3. Granichin, O.N. and Polyak, B.T., *Randomizirovannye algoritmy otsenivaniya i optimizatsii pri pochti proizvol'nykh pomekhakh* (Randomized Algorithms of Estimation and Optimization under Almost Arbitrary Noise), Moscow: Nauka, 2003.
4. Granichin, O.N., Linear Regression and Filtering under Nonstandard Assumptions (Arbitrary Noise), *IEEE Trans. Automat. Control*, 2004, vol. 49, no. 10, pp. 1830–1835.
5. Fisher, R.A., *The Design of Experiments*, Edinburgh: Oliver and Boyd, 1935.
6. Ljung, L., *System Identification: Theory for the User*, Upper Saddle River: PTR Prentice Hall, 2nd ed., 1999.
7. Tsypkin, Ya.Z., *Informatsionnaya teoriya identifikatsii* (Informational Identification Theory), Moscow: Nauka, 1995.
8. Györfi, L., Stochastic Approximation from Ergodic Sample for Linear Regression, *Z. Wahrscti. Verw. Geb.*, 1980, vol. 54, pp. 47–55.
9. Krieger, A. and Masry, E., Convergence Analysis of Adaptive Linear Estimation for Dependent Stationary Processes, *IEEE Trans. Inform. Theory*, 1988, vol. 34, pp. 177–182.
10. Young, R.S., *Recursive Estimation and Time-Series Analysis. An Introduction*, Berlin: Springer, 1984.
11. Albert, A., *Regression and the Moore–Penrose Pseudoinverse*, New York: Academic, 1972.
12. Polyak, B.T. and Tsypkin, Ya.Z., Adaptive Estimation Algorithms (Converhence, Optimality, Stability), *Autom. Remote Control*, 1979, no. 3, pp. 378–389.
13. Polyak, B.T., New Method of Stochastic Approximation Type, *Autom. Remote Control*, 1990, no. 7, pp. 937–946.
14. Candes, E., Romberg, J., and Tao, T., Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information, *IEEE Trans. Inform. Theory*, 2006, vol. 52, no. 2, pp. 489–509.
15. Donoho, D., Compressed Sensing, *IEEE Trans. Inform. Theory*, 2006, vol. 52, no. 4, pp. 1289–1306.
16. Kashin, B.S. and Temlyakov, V.N., Remark on the Problem of Compressed Measurement, *Mat. Zametki*, 2007, vol. 82, no. 6, pp. 829–837.
17. Claerbout, J.F. and Muir, F., Robust Modeling with Erratic Data, *Geophysics*, 1973, vol. 38, no. 5, pp. 826–844.

18. Santosa, F. and Symes, W.W., Linear Inversion of Band-limited Reflection Seismograms, *SIAM J. Sci. Statist. Comput.*, 1986, vol. 7, no. 4, pp. 1307–1330.

19. Granichin, O.N., Optimal Control of Linear Plant with Nonregular Bounded Noise, in *Abstracts of All-Union Conf. "Theory of Adaptive Systems and Its Applications,"* Moscow: VINITI, 1983, p. 26.

20. Barabanov, A.E. and Granichin, O.N., Optimal Controller for a Linear Plant with Bounded Noise, *Autom. Remote Control*, 1984, no. 5, pp. 578–584.

21. Dahlem, M. and Pearson, J.B., ℓ-1 optimal Feedback Controllers for MIMO Discrete Systems, *IEEE Trans. Automat. Control*, 1987, vol. 32, no. 4, pp. 314–322.

22. Baraniuk, R.G., Compressive Sensing, *IEEE Signal Process. Magaz.*, 2007, vol. 52, no. 2, pp. 118–120,124.

23. Wiener, N., *Cybernetics or Control and Communication in the Animals and the Machines*, New York: MIT Press, 1961. Translated under the title *Kibernetika ili upravlenie i svyaz' v zhibotnom i mashine*, Moscow: Nauka, 1983.

24. Kotel'nikov, V.A., On the Throughput of Ether and Wire in Electrical Communication, in *All-Union Energy Committee. Materials for the First All-Union Conference on Technical Reconstruction of Communication and Development of Low-current Industry*, 1933. Paper reprinted in *Usp. Phys. Nauk*, vol. 176, no. 7, pp. 762–770.

25. Kupfmuller, K., Uber die Dynamik der Selbsttatigen Verstarkungsregler, *Elektr. Nachrichtentechnik*, 1928, vol. 5, no. 11, pp. 459–467.

26. Nyquist, H., Certain Topics in Telegraph Transmission Theory, *Trans. AIEE*, 1928, vol. 47, pp. 617–644.

27. Shannon, C.E., Communication in the Presence of Noise, *Proc. IRE*, 1949, vol. 37, no. 1, pp. 10–21.

28. Wakin, M., Manifold-based Signal Recovery and Parameter Estimation from Compressive Measurements, *Preprint of Colorado School of Mines*, 2008.

29. Rauhut, H., Schnass, K., and Vandergheynst, P., Compressed Sensing and Redundant Dictionaries, *IEEE Trans. Inf. Theory*, 2008, vol. 54, no. 5, pp. 2210–2219.

30. Cevher, V., Duarte, M., Hegde, C., and Baraniuk, R., Sparse Signal Recovery Using Markov Random Fields, in *Proc. Workshop Neural Information Process. Systems (NIPS)*, 2008.

31. Mishali, M. and Eldar, Y.C., Reduce and Boost: Recovering Arbitrary Sets of Jointly Sparse Vectors, *IEEE Trans. Signal Process.*, 2008, vol. 56, no. 10, pp. 4692–4702.

32. Taubman, D.S. and Marcellin, M.V., *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, Norwell: Kluwer, 2001.

33. Boufounos, P.T. and Baraniuk, R.G., 1-Bit Compressive Sensing, in *42nd Conf. Inf. Sc. Systems (CISS)*, Princeton, 2008, pp. 19–21.

34. Takhar, D., Bansal, V., Wakin, M., et al., A Compressed Sensing Camera: New Theory and an Implementation Using Digital Micromirrors, in *Proc. Comput. Imaging IV SPIE Electronic Imaging*, San Jose, 2006.

35. Duarte, M.F., Davenport, M.A., Takhar, D., et al., Single-Pixel Imaging via Compressive Sampling, *IEEE Signal Process. Magaz.*, 2008, vol. 25, no. 2, pp. 83–91.

36. Candes, E.J., The Restricted Isometry Property and Its Implications for Compressed Sensing, *Comptes Rendus de l'Acad. des Sci., Serie I*, 2008, vol. 346, pp. 589–592.

37. Candes, E. and Romberg, J., Sparsity and Incoherence in Compressive Sampling, *Inverse Problems*, 2007, vol. 23, no. 3, pp. 969–985.

38. Candes, E. and Wakin, M., People Hearing without Listening: An Introduction to Compressive Sampling, *IEEE Signal Process. Magaz.*, 2008, vol. 25, no. 2, pp. 21–30.

39. Kashin, B.S., Diameters of Some Finite-dimensional Sets and Classes of Smooth Functions, *Izv. Akad. Nauk SSSR, Mat.*, 1977, vol. 42, no. 2, pp. 334–351.

40. Garnaev, A.Yu. and Gluskin, E.D., On Diameters of the Euclidean Sphere, *Dokl. Akad. Nauk SSSR*, 1984, vol. 277, no. 5, pp. 200–204.

41. Baraniuk, R., Davenport, M., DeVore, R., and Wakin, M., A Simple Proof of the Restricted Isometry Property for Random Matrices (aka "The Johnson–Lindenstrauss Lemma Meets Compressed Sensing"), *Constr. Approx.*, 2007, vol. 28, no. 3, pp. 253–263.

42. Johnson, W. and Lindenstrauss, J., Extensions of Lipschitz Mapping into a Hilbert Space, in *Conf. Modern Anal. Probabil.*, 1984, pp. 189–206.

43. Achlioptas, D., Database-friendly Random Projections, in *Proc. ACM SIGACT-SIGMOD-SIGART Symp. Principles Database Systems*, 2001, pp. 274–281.

44. Juditsky, A. and Nemirovski, A., On Verifiable Sufficient Conditions for Sparse Signal Recovery via ℓ_1 Minimization, *Rreprint*, 2010, Available at `http://arxiv.org/abs/0809.2650`.

45. Baron, D., Wakin, M.B., Duarte, M., Sarvotham, S., and Baraniuk, R.G., Distributed Compressed Sensing, 2005, Available at: `http://dsp.rice.edu/cs/DCS112005.pdf`.

46. Chen, S.S., Donoho, D.L., and Saunders, M.A., Atomic Decomposition by Basis Pursuit, *SIAM J. Sci. Comput.*, 2001, vol. 43, no. 1, pp. 129–159.

47. Tibshirani, R., Regression Shrinkage and Selection via the Lasso, *J. Royal. Statist. Soc. B*, 1996, vol. 58, no. 1, pp. 267–288.

48. Chretien, S., An Alternating l_1 Approach to the Compressed Sensing Problem, *IEEE Signal Process. Lett.*, 2010, vol. 17, no. 2, pp. 181–184.

49. Mohimani, H., Babaie-Zadeh, M., and Jutten, C., A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed ℓ_0 Norm, *IEEE Trans. Signal Process.*, 2009, vol. 57, no. 1, pp. 289–301.

50. Polyak, B.T., *Vvedenie v optimizatsiyu*, Moscow: Nauka, 1983. Translated into English under the title *Introduction to Optimization*, New York: Optimization Software, 1987.

51. Chartrand, R., Exact Reconstructions from Surprisingly Little Data, *IEEE Signal Process. Lett.*, 2007, vol. 14, pp. 707–710.

52. Chartrand, R., Fast Algorithms for Nonconvex Compressive Sensing: MRI Reconstruction from Very Few Data, *IEEE Int. Sympos. Biomedical Imaging (ISBI)*, 2009, pp. 1349–1360.

53. Saab, R., Chartrand, R., and Yilmaz, O., Stable Sparse Approximations via Nonconvex Optimization, in *Proc. 33rd Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, 2008, pp. 3885–3888.

54. Chartrand, R. and Yin, W., Iteratively Reweighted Algorithms for Compressive Sensing, in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Las Vegas, 2008.

55. Daubechies, I., DeVore, R., Fornasier, M., and Gunturk, S., Iteratively Re-weighted Least Squares Minimization: Proof of Faster than Linear Rate for Sparse Recovery, in *42nd Ann. Conf. Inf. Sci. Systems*, 19–21 March 2008, pp. 26–29.

56. Wipf, D. and Nagarajan, S., *Iterative Reweighted 1 and 2 Methods for Finding Sparse Solutions*, UC San Francisco, Technical Report, 2008.

57. Bredies, K. and Lorenz, D.A., Iterated Hard Shrinkage for Minimization Problems with Sparsity Constraints, *SIAM J. Sci. Comput.*, 2008, vol. 30, no. 2, pp. 657–683.

58. Herrity, K.K., Gilbert, A.C., and Tropp, J.A., Sparse Approximation via Iterative Thresholding, in *ICASSP 2006 Proc.*, 14–16 May 2006, vol. 3, pp. 624–627.

59. Ma, J., Improved Iterative Curvelet Thresholding for Compressed Sensing, *Preprint*, 2009, Available at: `http://dsp.rice.edu/sites/dsp.rice.edu/files/cs/ISTcs2.pdf`.

60. Beck, A. and Teboulle, M., A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, *SIAM J. Imaging Sci.*, 2009, vol. 2, no. 1, pp. 183–202.

61. Bioucas-Dias, J. and Figueiredo, M., A New TwIST: Two-step Iterative Shrinkage/Thresholding Algorithms for Image Restoration, *IEEE Trans. Image Process.*, 2007, vol. 16, no. 12, pp. 2992–3004.

62. Combettes, P.L. and Wajs, V.R., Signal Recovery by Proximal Forward-Backward Splitting, *SIAM J. Multiscale Modeling Simulation*, 2005, vol. 4, no. 4, pp. 1168–1200.

63. Combettes, P.L. and Pesquet, J.-Ch., Proximal Splitting Methods in Signal Processing, *Preprint*, 2010, Available at: `http://arxiv4.library.cornell.edu/abs/0912.3522`.

64. Egiazarian, K., Foi, A., and Katkovnik, V., Compressed Sensing Image Reconstruction via Recursive Spatially Adaptive Filtering, in *Proc. IEEE Int. Conf. Image Process., ICIP 2007*, San Antonio, 16–19 September 2007, pp. 549–552.

65. Mallat, S. and Zhang, Z., Matching Pursuit with Time-Frequency Dictionaries, *IEEE Trans. Signal Process.*, 1993, vol. 41, no. 12, pp. 3397–3415.

66. Tropp, J. and Gilbert, A.C., Signal Recovery from Partial Information via Orthogonal Matching Pursuit, 2005, Available at: `www-personal.umich.edu/~jtropp/papers/TG06-Signal- Recovery.pdf`.

67. Davenport, M.A. and Wakin, M.B., Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property, *Preprint*, 2010, Available at: `http://arxiv.org/abs/0909.0083`.

68. Needell, D. and Vershynin, R., Uniform Uncertainty Principle and Signal Recovery via Regularized Orthogonal Matching Pursuit, *Foundat. Comput. Math.*, 2009, vol. 9, pp. 317–334.

69. Needell, D. and Tropp, J.A., CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples, *Appl. Comp. Harmonic Anal.*, 2008, vol. 26, pp. 301–321.

70. Dai, W. and Milenkovic, O., Subspace Pursuit for Compressive Sensing Signal Reconstruction, *IEEE Trans. Inf. Theory*, 2009, vol. 55, no. 5, pp. 2230–2249.

71. Donoho, D.L., Tsaig, Y., Drori, I., and Starck, J.-L., Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit, *Stanford Univ. Tech. Rep.*, 2006, Available at: `http://stat.stanford.edu/ idrori/StOMP.pdf`.

72. Cohen, A., Dahmen, W., and DeVore, R., Instance Optimal Decoding by Thresholding in Compressed Sensing, *Preprint*, 2008, Available: `http://www.igpm.rwth-aachen.de/Download/reports/pdf/IGPM289.pdf`.

73. La, C. and Minh, N.D., Tree-based Orthogonal Matching Pursuit Algorithm for Signal Reconstruction, in *IEEE Int. Conf. Image Process.*, 8–11 Oct., 2006, pp. 1277–1280.

74. Blumensath, T. and Davies, M.E., Gradient Pursuits, *IEEE Trans. Signal Process.*, 2008, vol. 56, no. 6, pp. 2370–2382.

75. Berinde, R. and Indyk, P., *Sequential Sparse Matching Pursuit*, Allerton: MIT, 2009, Available at: `http://people.csail.mit.edu/indyk/ssmp.pdf`.

76. Baron, D., Sarvotham, S., and Baraniuk, R.G., Bayesian Compressive Sensing via Belief Propagation, 2008, Available at: `http://arxiv.org/abs/0812.4627.`

77. Cormode, G. and Muthukrishnan, S., *Combinatorial Algorithms for Compressed Sensing*, Berlin: Springer, 2006.

78. Jafarpour, S., Xu, W., Hassibi, B., and Calderbank, R., Efficient and Robust Compressed Sensing Using Optimized Expander Graphs, *IEEE Trans. Info. Theory*, 2009, vol. 55, no. 9, pp. 4299–4308.

79. Berinde, R., Indyk, P., and Ruzic, M., Practical Near-optimal Sparse Recovery in the L1 Norm, in *46th Annual Allerton Conf. Communication, Control, Computing*, Allerton, 2008, pp. 198–205.

80. Lustig, M., Donoho, D.L., Santos, J.M., and Pauly, J.M., Compressed Sensing MRI, *IEEE Signal Process. Magaz.*, 2008, vol. 25, no. 2, pp. 72–82.

81. Candes, E.J. and Romberg, J., Practical Signal Recovery from Random Projections, in *Proc. SPIE Computational Imaging*, 2005, vol. 5674, pp. 76–86.

82. Wakin, M., Laska, J., Duarte, M., Baron, D., Sarvotham, S., Takhar, D., Kelly, K., and Baraniuk, R., Compressive Imaging for Video Representation and Coding, in *Proc. Picture Coding Sympos.—PCS 2006*, Beijing, 2006, Available at: `http://www.mines.edu/mwakin/papers/ pcs-camera.pdf`.

83. Ma, J. and Le Dimet, F.X., Deblurring from Highly Incomplete Measurements for Remote Sensing, *IEEE Trans. Geoscience Remote Sensing*, 2009, vol. 47, no. 3, pp. 792–802.

84. Katkovnik, V. and Egiazarian, K., Nonlocal Image Deblurring: Variational Formulation with Nonlocal Collaborative L0 Norm Prior, in *Int. Workshop Local Non-Local Approximation in Image Process.*, 19–21 Aug., 2009, pp. 46–53.

85. Calderbank, R., Jafarpour, S., and Robert Schapire, R., Compressed Learning: Universal Sparse Dimensionality Reduction and Learning in the Measurement Domain, *Preprint*, 2009, Available at: `http://dsp.rice.edu/files/cs/cl.pdf`.

86. Hsu, D., Langford, J., Kakade, S., and Zhang, T., Multi-label Prediction via Compressed Sensing, *Preprint*, 2009, Available at: `http://arxiv.org/abs/0902.1284`.

87. Jiyoung, C., Minwoo, K., Won, S., and Jong, C.Y., Compressed Sensing Metal Artifact Removal in Dental CT, in *IEEE Int. Sympos. Biological Imaging*, 2009, pp. 334–337.

88. Kirolos, S., Laska, J., Wakin, M., et al., Analog-to-Information Conversion via Random Demodulation, in *Proc. IEEE Dallas Circuits Systems Workshop*, 2006, pp. 71–74.

89. Baraniuk, R. and Steeghs, P., Compressive Radar Imaging, in *IEEE Radar Conf.*, 2007, pp. 128–133.

90. Vetterli, M., Marziliano, P., and Blu, T., Sampling Signals with Finite Rate of Innovation, *IEEE Trans. Signal Process.*, 2002, vol. 50, no. 6, pp. 1417–1428.

This paper was recommended for publication by B.T. Polyak, a member of the Editorial Board