INVITED SURVEY PAPER

A User's Guide to Compressed Sensing for Communications Systems

Kazunori HAYASHI[†a)], **Masaaki NAGAHARA**[†b)], *and* **Toshiyuki TANAKA**[†c)], *Members*

SUMMARY This survey provides a brief introduction to compressed sensing as well as several major algorithms to solve it and its various applications to communications systems. We firstly review linear simultaneous equations as ill-posed inverse problems, since the idea of compressed sensing could be best understood in the context of the linear equations. Then, we consider the problem of compressed sensing as an underdetermined linear system with a prior information that the true solution is sparse, and explain the sparse signal recovery based on ℓ_1 optimization, which plays the central role in compressed sensing, with some intuitive explanations on the optimization problem. Moreover, we introduce some important properties of the sensing matrix in order to establish the guarantee of the exact recovery of sparse signals from the underdetermined system. After summarizing several major algorithms to obtain a sparse solution focusing on the ℓ_1 optimization and the greedy approaches, we introduce applications of compressed sensing to communications systems, such as wireless channel estimation, wireless sensor network, network tomography, cognitive radio, array signal processing, multiple access scheme, and networked control.

key words: *compressed sensing, sparse signal, compressible signal, ℓ_1-norm, underdetermined system*

1. Introduction

Sparse reconstruction using the minimization of ℓ_1-norm, which we call ℓ_1 optimization, has been receiving a lot of attention triggered by studies on *compressed sensing* (also known as *compressive sensing* or *compressive sampling*) [1]–[3], where the problem is to reconstruct a finite-dimensional sparse vector based on its linear measurements of dimension smaller than the size of the unknown sparse vector. It is true that there have been several works which utilize sparsity or the ℓ_1 optimization before compressed sensing, such as Logan's phenomenon [4], the matching pursuit [5], overcomplete representations in the context of wavelet transforms [6], [7], and the least absolute shrinkage and selection operator (Lasso) [8], [9]. Moreover, many conventional signal processing techniques, such as dimensionality reduction via principal component analysis, as well as the subspace method, have utilized the fact that signals of interest can often be represented by using not all but only a few elements of a basis, which is nothing but sparsity of signals. However, the reason for the current explosion of researches related to the sparsity of signals and/or systems will be that they have successfully shown that sparse signals* can be reconstructed from a small number of non-adaptive linear measurements by using optimization algorithms with practical computational complexity. Compressed sensing is related to one of the fundamental problems in information and signal processing, and is based on the assumption of the sparsity in some transform domain, which is valid for a lot of signals around us. We are therefore sure that it is not just a passing fad and will be on every future textbook in the field of signal processing or communication theory as Shannon's sampling theory [10] is on every current textbook.

This paper provides a brief introduction to compressed sensing as well as several major algorithms to solve it and its various applications to communications systems, assuming readers to be potential users of compressed sensing and to be in the field of communications. Although there are several approaches to introduce compressed sensing, we believe that it could be best understood in the context of linear simultaneous equations, especially those in the field of communications. Thus, we firstly review the linear equations as ill-posed (i.e., over- or under-determined) inverse problems both with and without measurement noise. Then, we consider the problem of compressed sensing, namely, the inverse problem of the underdetermined linear equations with a prior knowledge that the true solution is sparse, and explain the sparse signal recovery based on ℓ_1 optimization, which plays the central role in compressed sensing, with intuitive explanations on the optimization problem. Moreover, we introduce some important properties of the sensing matrix (i.e., the matrix of coefficients of the linear equations), such as the restricted isometry property (RIP), the spark, the null space property (NSP), the neighborliness, and the mutual coherence, as well as the treatment via random matrix ensembles, which are used to provide necessary and/or sufficient conditions for the recovery of the sparse signals from the underdetermined linear system.

After the introduction of basic ideas of compressed sensing above, we explain major concrete algorithms to obtain a sparse solution, which are roughly classified into two approaches: a convex relaxation based ℓ_1 optimization approach and a greedy approach. Furthermore, we introduce various examples of applications of compressed sensing to communications systems. Since it is almost impossible to

Manuscript received August 17, 2012.
Manuscript revised November 14, 2012.

[†]The authors are with the Graduate School of Informatics, Kyoto University, Kyoto-shi, 606-8501 Japan.
a) E-mail: kazunori@i.kyoto-u.ac.jp
b) E-mail: nagahara@ieee.org
c) E-mail: tt@i.kyoto-u.ac.jp

DOI: 10.1587/transcom.E96.B.685

*In the basic framework of compressed sensing, the reconstruction of a finite dimensional vector is commonly considered. Thus, in this paper, "signals" denotes discrete signals or vectors, unless otherwise stated.

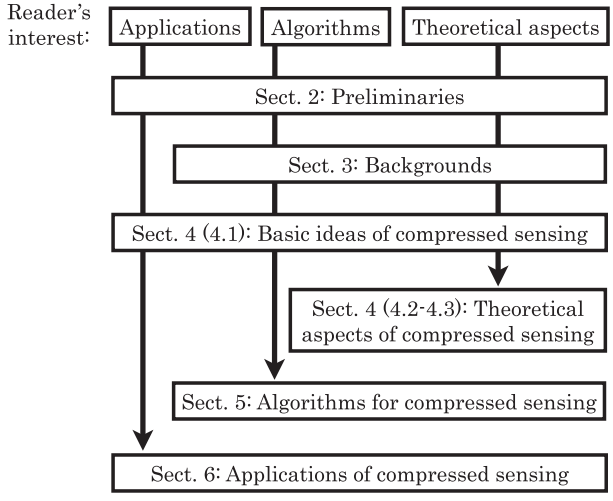

Fig. 1 Quick courses for readers, who are interested only in applications, algorithms, or some basic theoretical aspects.

cover all examples of the applications proposed so far, we focus on some topics, in which we believe successful applications can be found, such as sparse wireless channel estimation, data gathering problem in wireless sensor network, delay and loss tomographies in network, spectrum sensing in cognitive radio systems, direction-of-arrival (DOA) estimation in array signal processing, multiple access schemes, and data compression for networked control.

The rest of the paper is organized as follows. In Sect. 2, as preliminaries, we define some norms and terms to be used in the paper. In Sect. 3, we review conventional approaches to cope with ill-posed linear simultaneous equations without assuming the sparsity of the solution. In Sect. 4, a brief introduction to compressed sensing is provided more emphasizing methodologies but with some analytical aspects, which users of compressed sensing should know. Sections 5 and 6 respectively introduce several algorithms to solve the problem of compressed sensing and various applications to communications systems. After providing some surveys, tutorials and books for further studies in Sect. 7, we conclude this paper in Sect. 8.

In addition, quick courses for readers, who are interested only in applications, algorithms, or some basic theoretical aspects, are summarized in Fig. 1. Some sections of this survey can be omitted depending on reader's interest.

2. Preliminaries

2.1 ℓ_p-norm

In the analysis or algorithms of compressed sensing, we encounter various norms, while we usually use the Euclidean (ℓ_2-) norm in the conventional problems of communications. Thus, let us firstly define some norms.

The ℓ_p-norm of a vector $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}} \in \mathbb{R}^n$ is defined for $p \geq 1$ as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \tag{1}$$

where $[\cdot]^{\mathrm{T}}$ denotes the transpose.

One can use the same formula (1) to define a norm-like quantity for $0 < p < 1$. The resulting quantity $\|\cdot\|_p$ for $0 < p < 1$ is no longer a norm in the mathematical sense since it does not satisfy the triangle inequality[†] (so it is sometimes called a quasi-norm). Nevertheless, it is frequently used in the context of compressed sensing.

We can furthermore consider $\|\mathbf{x}\|_0$, which is not even a quasi-norm, defined as

$$\|\mathbf{x}\|_0 = |\operatorname{supp}(\mathbf{x})|, \tag{2}$$

where $\operatorname{supp}(\mathbf{x}) = \{i : x_i \neq 0\}$ and $|\operatorname{supp}(\mathbf{x})|$ is the cardinality of $\operatorname{supp}(\mathbf{x})$. By convention, we call it the ℓ_0-norm.

2.2 Sparse and Compressible Signals

Sparsity of signal is the central theme of this paper. A signal $\mathbf{x} \in \mathbb{R}^n$ is said to be *sparse* (or *exactly sparse*) if most of the elements are exactly equal to zero, i.e., $\|\mathbf{x}\|_0 \ll n$. A signal $\mathbf{x}$ is said to be *k-sparse* when it has at most k nonzeros, i.e., $\|\mathbf{x}\|_0 \leq k$. Also, we define the set of all k-sparse signals as

$$\Sigma_k = \{\mathbf{x} : \|\mathbf{x}\|_0 \leq k\}. \tag{3}$$

The set Σ_k is not convex, because, for some $\mathbf{x}, \mathbf{z} \in \Sigma_k$, we have $\mathbf{x} + \mathbf{z} \notin \Sigma_k$ (note that we have $\mathbf{x} + \mathbf{z} \in \Sigma_{2k}$). Also, it should be noted that a signal $\mathbf{x}$ might be k-sparse in a certain representation using some pre-determined basis $\mathbf{\Phi}$, i.e., $\mathbf{x} = \mathbf{\Phi c}$ where $\mathbf{c} \in \Sigma_k$, instead of the signal $\mathbf{x}$ itself. In some cases, such $\mathbf{x}$ is still said to be k-sparse.

In practical situations, it will be quite rare to meet exactly sparse signals. Instead, we will encounter signals most entries of which are approximately zero. The signals, which can be well-approximated by sparse signals, are called *compressible* (also called *approximately sparse* or *relatively sparse*). The compressibility of a signal $\mathbf{x}$ can be evaluated by the error induced by the best approximation with $\hat{\mathbf{x}} \in \Sigma_k$ as

$$\sigma_k(\mathbf{x})_p = \min_{\hat{\mathbf{x}} \in \Sigma_k} \|\mathbf{x} - \hat{\mathbf{x}}\|_p. \tag{4}$$

3. Linear Equations Review

We begin with the review of a linear system having m equations and n unknowns as

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{5}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, assuming that both $\mathbf{A}$ and $\mathbf{y}$ are exactly known. Hereafter, we call $\mathbf{y}$ a *measurement vector*, and $\mathbf{A}$ a *sensing matrix*. The term of linear measurement means that each element of $\mathbf{y}$ is obtained as the inner

[†]A norm should satisfy: (i) $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$, (ii) $\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|$, $\forall \alpha \in \mathbb{R}$, (iii) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

product between the unknown vector $\mathbf{x}$ and each row vector of $\mathbf{A}$. Note that we do not assume the sparsity of $\mathbf{x}$ in this section.

We firstly consider the case with $n = m$ and $\mathbf{A}$ is non-singular (if $\mathbf{A}$ is singular then it corresponds to the case with $n > m$ to be discussed later). In this case, since $\mathbf{A}^{-1}$ exists, we can obtain a unique solution $\mathbf{x}$ satisfying (5), by multiplying $\mathbf{A}^{-1}$ to both sides of (5) from the left, as

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}. \tag{6}$$

The case with $n < m$, where the number of the equations (measurements) is greater than the number of unknowns, is basically the same as the previous situation, if the measurements are not contaminated with noise. Assuming $\mathbf{A}$ to be of full column rank, the true solution is obtained by

$$\mathbf{x} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y}. \tag{7}$$

This can be easily verified by substituting into (7) the singular value decomposition (SVD) of $\mathbf{A}$ given by

$$\mathbf{A} = \mathbf{U}\begin{bmatrix} \boldsymbol{\Xi} \\ \mathbf{0}_{(m-n)\times n} \end{bmatrix}\mathbf{V}^{\mathrm{T}}, \tag{8}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices of size $m \times m$ and $n \times n$, respectively, where $\boldsymbol{\Xi}$ is a diagonal matrix of size $n \times n$ with its diagonal elements equal to the singular values of $\mathbf{A}$, and where $\mathbf{0}_{(m-n)\times n}$ is a zero matrix of size $(m-n) \times n$.

On the other hand, in the case of $n > m$, where the number of measurements is less than that of unknowns, the system (5) of linear equations is underdetermined, with its solution no longer unique. This is because there exist infinitely many vectors $\mathbf{z}$ in the null space of $\mathbf{A}$, which is defined as

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{z} : \ \mathbf{A}\mathbf{z} = \mathbf{0}\}, \tag{9}$$

such that $\mathbf{x}+\mathbf{z}$ with an arbitrary $\mathbf{z} \in \mathcal{N}(\mathbf{A})$ satisfies (5), where $\mathbf{x}$ denotes the "true" solution. Thus, we have infinitely many candidates of the solution in this case. Such a problem is called an ill-posed inverse problem. A common approach taken to choose one solution from the candidates is regularization, where the solution is chosen by penalizing (or minimizing) the norm of the candidates. A typical choice of the norm is the squared ℓ_2-norm $\|\mathbf{x}\|_2^2$, and the regularization problem is formulated as

$$\hat{\mathbf{x}}_{\mathrm{MN}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_2^2 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}, \tag{10}$$

where MN stands for "minimum norm." The optimization problem (10) can be analytically solved by the method of Lagrange multipliers. Define the Lagrange function $\mathcal{L}(\mathbf{x})$ for (10) by

$$\begin{aligned}\mathcal{L}(\mathbf{x}) &= \|\mathbf{x}\|_2^2 + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{y}) \\ &= \mathbf{x}^{\mathrm{T}}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} - \mathbf{y}^{\mathrm{T}}\boldsymbol{\lambda}.\end{aligned} \tag{11}$$

Since the Lagrange function $\mathcal{L}(\mathbf{x})$ is a quadratic function of $\mathbf{x}$, one can minimize it by considering the condition that its derivative with respect to $\mathbf{x}$ vanishes

$$\frac{\partial \mathcal{L}(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{x} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{0}, \tag{12}$$

to obtain

$$\hat{\mathbf{x}}_{\mathrm{MN}} = -\frac{1}{2}\mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda}. \tag{13}$$

By substituting $\hat{\mathbf{x}}_{\mathrm{MN}}$ into the constraint of $\mathbf{A}\mathbf{x} = \mathbf{y}$, one can specify the values of the Lagrange multipliers as

$$\boldsymbol{\lambda} = -2(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{y}. \tag{14}$$

Finally, we obtain

$$\hat{\mathbf{x}}_{\mathrm{MN}} = \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{y}, \tag{15}$$

which is called the minimum-norm solution. It should be noted here that $\hat{\mathbf{x}}_{\mathrm{MN}}$ will be different from the true solution in general (that is why we use the symbol $\hat{\mathbf{x}}_{\mathrm{MN}}$ to distinguish it with $\mathbf{x}$). Also, note that ℓ_2-norm is not the unique choice for the cost function to be minimized in the regularization formalism, even though the minimum-norm solution $\hat{\mathbf{x}}_{\mathrm{MN}}$ could be satisfactory for several practical purposes. Actually, this can be considered as a motivation to use ℓ_0- or ℓ_1-norm, which leads to compressed sensing.

So far, we have assumed noise-free cases. Measurements available in practical systems are, however, often contaminated with some form of noise and hence the equation (5) does not hold exactly. In particular, if $n < m$, the measurement vector $\mathbf{y}$ might not be included in the image (or equivalently, the column space) of $\mathbf{A}$, that is, (5) has no solution in general. A familiar way to cope with this problem is to rely on the method of least squares (LS), where the optimization problem is given by

$$\hat{\mathbf{x}}_{\mathrm{LS}} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2. \tag{16}$$

Since the cost function can be expanded as

$$\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} - 2\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{y}, \tag{17}$$

and is quadratic, one can, once again, minimize it by considering the condition that its derivative with respect to $\mathbf{x}$ vanishes

$$\frac{\partial}{\partial \mathbf{x}}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 = 2\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} - 2\mathbf{A}^{\mathrm{T}}\mathbf{y} = \mathbf{0}, \tag{18}$$

yielding the LS solution

$$\hat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y}, \tag{19}$$

where we have assumed $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ to be non-singular, and hence $n \le m$. Note that, if $\mathbf{y}$ is noise-free, $\hat{\mathbf{x}}_{\mathrm{LS}}$ coincides with the true solution in (6) or (7) for $n = m$ and $n < m$, respectively, as they have exactly the same form.

For the case with $n > m$, (19) is not available for uniquely determining the solution, because we still have infinitely many candidates of the solution, which makes the

cost function of (16) to be zero, regardless of the existence of the noise, while $\hat{\mathbf{x}}_{\mathrm{MN}}$ is applicable for this case as well. Another approach might be the utilization of the regularized LS method, which considers the optimization problem of the form

$$\hat{\mathbf{x}}_{\mathrm{rLS}} = \arg\min_{\mathbf{x}} \left(\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 \right), \tag{20}$$

where the regularization parameter λ is to control the balance between the squared error and ℓ_2-norm of the solution. Since we have

$$\frac{\partial}{\partial \mathbf{x}} \left(\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 \right) = 2\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} - 2\mathbf{A}^{\mathrm{T}}\mathbf{y} + 2\mathbf{x}, \tag{21}$$

$\frac{\partial \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_2^2}{\partial \mathbf{x}} = \mathbf{0}$ yields the regularized LS solution

$$\hat{\mathbf{x}}_{\mathrm{rLS}} = (\lambda \mathbf{I} + \mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y}. \tag{22}$$

Note that $\hat{\mathbf{x}}_{\mathrm{rLS}}$ is applicable also to the case with $n \leq m$.

4. Compressed Sensing

Compressed sensing is a method to obtain a unique solution from an underdetermined linear system taking advantage of the prior knowledge that the true solution is sparse. In this section, we consider the linear simultaneous equations

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{23}$$

assuming $n < m$ and $\|\mathbf{x}\|_0 \ll n$. Here, we call $\mathbf{A}$ a sensing matrix again, while it is also referred to as a *(overcomplete) dictionary*, with its columns being called *atoms*. Note that if $\mathbf{x}$ is sparse in some pre-determined basis as $\mathbf{x} = \mathbf{\Phi}\mathbf{c}$ where $\|\mathbf{c}\|_0 \ll n$, then we have

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{\Phi}\mathbf{c}, \tag{24}$$

and $\mathbf{A}\mathbf{\Phi}$ is regarded as the sensing matrix in what follows.

4.1 Signal Recovery via ℓ_1 Optimization

A natural and straightforward approach to obtain a sparse solution from the underdetermined system (23) will be formulated as the optimization problem of

$$\hat{\mathbf{x}}_{\ell_0} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}, \tag{25}$$

which is called ℓ_0 optimization problem. Although the problem is similar to that in (10) superficially, the problem in (25) is far more difficult to solve because of the discrete and the non-convex natures of ℓ_0-norm.

The convention used here is to replace $\|\mathbf{x}\|_0$ with $\|\mathbf{x}\|_1$, which is a convex function, as

$$\hat{\mathbf{x}}_{\ell_1} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}. \tag{26}$$

It should be noted that the problem in (26) can be obtained by replacing the squared ℓ_2-norm in the conventional approach in (10) with ℓ_1-norm. The problem in (26) is tractable, since it can be posed as a linear programming (LP) problem. Actually, if we define nonnegative vectors $\mathbf{u} = [u_1, \ldots, u_n]^{\mathrm{T}}$ and $\mathbf{v} = [v_1, \ldots, v_n]^{\mathrm{T}}$ as

$$u_i = \begin{cases} x_i & \text{if } x_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

$$v_i = \begin{cases} -x_i & \text{if } x_i < 0 \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

then we have $\mathbf{x} = \mathbf{u} - \mathbf{v}$. The ℓ_1-norm of $\mathbf{x}$ can then be represented as a linear function of $\mathbf{z} = [\mathbf{u}^{\mathrm{T}}, \mathbf{v}^{\mathrm{T}}]^{\mathrm{T}}$, as

$$\|\mathbf{x}\|_1 = \mathbf{1}_n^{\mathrm{T}}(\mathbf{u} + \mathbf{v}) = \mathbf{1}_{2n}^{\mathrm{T}}\mathbf{z}, \tag{29}$$

where $\mathbf{1}_n$ is the all-one vector of size $n \times 1$. The constraint $\mathbf{A}\mathbf{x} = \mathbf{y}$ can also be represented as a form that is linear in terms of $\mathbf{z}$ by noting

$$\mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{u} - \mathbf{v}) = [\mathbf{A}, -\mathbf{A}]\mathbf{z}. \tag{30}$$

Therefore, the problem in (26) can be rewritten as

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \mathbf{1}_{2n}^{\mathrm{T}}\mathbf{z} \ \text{ subject to } [\mathbf{A}, -\mathbf{A}]\mathbf{z} = \mathbf{y} \text{ and } \mathbf{z} \succeq \mathbf{0}, \tag{31}$$

where $\succeq$ stands for the element-wise inequality. The optimization problem in (31) has the standard structure of LP.

An alternative LP formulation is obtained by rewriting the problem (26) as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x},\mathbf{t}} \mathbf{1}_n^{\mathrm{T}}\mathbf{t} \quad \text{subject to} \quad -\mathbf{t} \preceq \mathbf{x} \preceq \mathbf{t} \text{ and } \mathbf{A}\mathbf{x} = \mathbf{y}. \tag{32}$$

It is straightforward to observe that (32) reduces to (31) by letting $\mathbf{x} = \mathbf{u} - \mathbf{v}$ and $\mathbf{t} = \mathbf{u} + \mathbf{v}$.

One might ask the validity of the relaxation of the ℓ_0-norm with the ℓ_1-norm. Although analytical justification is given in Sect. 4.2, an intuitive understanding that the ℓ_1-norm can promote a sparse solution can be obtained by the shape of the ℓ_p-ball shown in Fig. 2. In the figure, assuming $\mathbf{x} \in \mathbb{R}^2$, signal recovery methods using ℓ_1-norm and ℓ_2-norm are illustrated, where the solid lines stand for the linear constraint and the dotted lines are the ℓ_p-ball with $p = 1$ and 2 when they touch the lines corresponding to the constraint. Thus, the points of $\hat{\mathbf{x}}$ in the figure show the solutions that minimize the cost functions for the two cases ($p = 1$ and $p = 2$). From the observations, it can be understood that if we employ the ℓ_1-norm, the solution tends to be on one of the two axes depending on the linear constraint, whereas for the case with ℓ_2-norm, the solution will have nonzero values for both elements in $\hat{\mathbf{x}}$ in general.

When the measurements include noise, we can also consider other optimization problems with different constraints. If the noise is bounded, a natural choice will be

$$\hat{\mathbf{x}}_{\ell_1} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \tag{33}$$

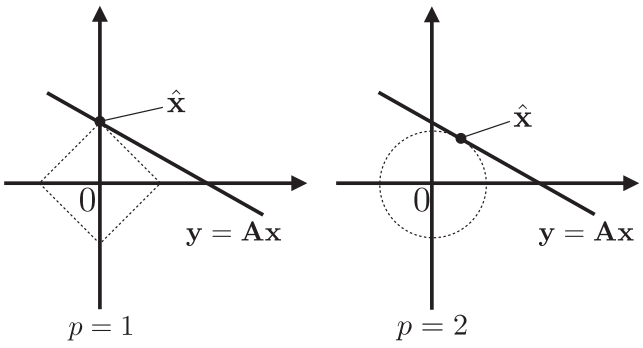

Fig. 2 Intuitive illustrations of signal recovery using ℓ_p-norm ($p = 1, 2$). Dotted line stands for ℓ_p-ball (i.e., a contour of constant value of $\|\mathbf{x}\|_p$) when it touches the linear constraint of $\mathbf{y} = \mathbf{A}\mathbf{x}$ and $\hat{\mathbf{x}}$ is the corresponding solution.

where $\epsilon > 0$ is a given constant. Note that this recovery problem can provide a certain guarantee not only for the case with the bounded noise but also for the case with Gaussian noise under some conditions on the sensing matrix [11], [12].

The constrained optimization problem in (33) can be recast into the unconstrained optimization problem

$$\hat{\mathbf{x}}_{\ell_1\text{-}\ell_2} = \arg\min_{\mathbf{x}} \left(\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1 \right), \tag{34}$$

which is sometimes called an ℓ_1-ℓ_2 optimization problem. One can also regard the problem in (34) as the modified version of the conventional approach in (20), where the squared ℓ_2-norm in the regularization term is replaced with ℓ_1-norm. With an appropriate choice of the parameter λ, the problem in (34) will yield the same solution as that in (33). However, the value of λ in the problem (34) that corresponds to the value of ϵ in the problem (33) in the sense that these two problems share the same solution is not known a priori in general. It should be noted that the Lasso estimator [8], [9] often appears in the form of the ℓ_1-ℓ_2 optimization (34) in the literature. The original definition of the Lasso in [8] is, however, a constrained optimization problem given by

$$\hat{\mathbf{x}}_{\text{Lasso}} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \ \text{ subject to } \ \|\mathbf{x}\|_1 \le t, \tag{35}$$

where $t > 0$ is a given constant. Once again, it is true that (34) and (35) are equivalent with an appropriate choice of t given λ and vice versa. While the selection of λ (or ϵ, t) has a large impact on the reconstruction performance, surprisingly enough, solutions for all λ can be obtained with almost the same computational complexity as solving for only one value of λ by a homotopy-type path-following algorithm called the least angle regression (LARS) algorithm [13], [14]. In addition, from a view point of Bayesian approach, (34) can be regarded as the maximum a posteriori (MAP) estimator when the additive noise is white Gaussian and the Laplacian prior $\propto \exp(-\lambda\|\mathbf{x}\|_1)$ is imposed on the unknown vector $\mathbf{x}$.

4.2 Properties of Sensing Matrix

4.2.1 Overview

Whether or not one can recover a sparse vector $\mathbf{x}$ from a measurement vector $\mathbf{y}$ depends on several factors, including the sensing matrix $\mathbf{A}$, the sparse vector $\mathbf{x}$ itself, and the algorithm used for recovering $\mathbf{x}$. Although these factors are not independent of each other, it has been recognized that arguing properties of the sensing matrix $\mathbf{A}$ is useful in several respects. One reason is that in many applications we have control only on the measurement processes. Another reason is that one would like to know whether the recovery with an algorithm will be successful for all (or almost all) sparse vectors $\mathbf{x}$. In this subsection, we address some important properties of the sensing matrix used in the analysis of compressed sensing.

4.2.2 Restricted Isometry Property (RIP)

One of the most well-known properties of the sensing matrix is the restricted isometry property (RIP) introduced by Candes and Tao [2]:

Definition 4.1 (Restricted Isometry Property (RIP)): A matrix $\mathbf{A}$ satisfies the RIP of order k if there exists a constant $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2 \le (1 + \delta_k)\|\mathbf{x}\|_2^2 \tag{36}$$

for any $\mathbf{x} \in \Sigma_k$.

As discussed in Sect. 3, the solution of a linear equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ is non-unique if and only if the null space $\mathcal{N}(\mathbf{A})$ of $\mathbf{A}$ is non-trivial (i.e., not equal to $\{\mathbf{0}\}$). We typically have $n > m$ in the context of compressed sensing, in which case $\mathcal{N}(\mathbf{A})$ is always non-trivial. But since we can assume that the true vector is sparse, then, for the purpose of guaranteeing uniqueness of the *sparse* solution of the linear equation $\mathbf{A}\mathbf{x} = \mathbf{y}$, it might be reasonable to think about something like

$$\mathcal{N}_k(\mathbf{A}) = \{\mathbf{z} : \mathbf{A}\mathbf{z} = \mathbf{0},\ \mathbf{z} \in \Sigma_k\} \tag{37}$$

in place of the full null space $\mathcal{N}(\mathbf{A})$. The RIP given in Definition 4.1 goes a bit further and quantifies the degree of isometry of the operation of $\mathbf{A}$ on any k-sparse vectors. Here, $\delta_k < 1$ implies $\mathcal{N}_k(\mathbf{A}) = \{\mathbf{0}\}$, and smaller values of δ_k suggest how far we are from the non-uniqueness of the solution of the linear equation $\mathbf{A}\mathbf{x} = \mathbf{y}$.

Investigating properties of the sensing matrix $\mathbf{A}$ on the set Σ_k is equivalent to studying properties of submatrices of $\mathbf{A}$ in the following sense. If we knew the support $\Lambda = \text{supp}(\mathbf{x})$ of $\mathbf{x} \in \Sigma_k$, then we could apply the method of LS in (19) with the submatrix $\mathbf{A}_\Lambda$ of $\mathbf{A}$ formed by collecting the columns of $\mathbf{A}$ whose indices are in Λ, provided that the linear system with the submatrix $\mathbf{A}_\Lambda$ is well-posed (such an estimator is called an *oracle estimator*). The RIP can then be considered as the condition that imposes well-posedness

to any submatrix $\mathbf{A}_\Lambda$ with $|\Lambda| = k$. Indeed, the RIP constant δ_k is given in terms of the maximum and minimum singular values of all the submatrices $\{\mathbf{A}_\Lambda : |\Lambda| \le k\}$.

The RIP has been used to establish sufficient conditions of the recovery for a lot of scenarios including the case with noisy measurements. As an example, we have a following theorem for noiseless recovery.

Theorem 4.1: [11] Let $\mathbf{A}$ be a sensing matrix that satisfies the RIP of order $2k$ with the constant $\delta_{2k} < \sqrt{2} - 1$. Then, the solution $\hat{\mathbf{x}}_{\ell_1}$ to (26) obeys

$$\|\hat{\mathbf{x}}_{\ell_1} - \mathbf{x}\|_2 \le C_0 \frac{\sigma_k(\mathbf{x})_1}{\sqrt{k}}, \tag{38}$$

where

$$C_0 = 2\frac{1 - (1 - \sqrt{2})\delta_{2k}}{1 - (1 + \sqrt{2})\delta_{2k}}. \tag{39}$$

Thus, for any exactly sparse signal $\mathbf{x} \in \Sigma_k$, the exact recovery is possible by the ℓ_1 optimization if the sensing matrix satisfies the RIP of order $2k$ with the constant $\delta_{2k} < \sqrt{2} - 1$.

The sufficient condition $\delta_{2k} < \sqrt{2} - 1 = 0.414$ for the exact recovery of arbitrary k-sparse signals via the ℓ_1 optimization has further been refined in subsequent studies: For example, Foucart and Lai have improved it to $\delta_{2k} < 0.4531$ [15], and Cai et al. have obtained $\delta_{2k} < 0.472$ [16] and $\delta_k < 0.307$ [17].

Although it is unquestionable that the RIP has been one of key issues in the theoretical development of compressed sensing, there are two major drawbacks with the RIP; 1) the RIP only gives a sufficient condition for the recovery, which may not be tight, and 2) it is typically very difficult to verify whether a given matrix satisfies the RIP or not, or to calculate the corresponding constant. This motivates exploration of other properties of the sensing matrix.

4.2.3 Null Space Property (NSP)

In order to guarantee the exact recovery of exactly sparse signals $\mathbf{x} \in \Sigma_k$ from $\mathbf{y} = \mathbf{A}\mathbf{x}$, it is necessary that, for any pairs of distinct vectors $\mathbf{x}, \mathbf{x}' \in \Sigma_k$, we have $\mathbf{A}\mathbf{x} \neq \mathbf{A}\mathbf{x}'$. Since $\mathbf{A}\mathbf{x} \neq \mathbf{A}\mathbf{x}'$ implies $\mathbf{A}(\mathbf{x} - \mathbf{x}') \neq \mathbf{0}$ and we have $\mathbf{x} - \mathbf{x}' \in \Sigma_{2k}$, it is proved that there exists at most one $\mathbf{x} \in \Sigma_k$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$ if and only if $\mathcal{N}(\mathbf{A}) \cap \Sigma_{2k} = \{\mathbf{0}\}$. Moreover, one of the equivalent conditions to this is that the smallest number of columns from $\mathbf{A}$ that are linearly-dependent, which is called the *spark* of $\mathbf{A}$ and is denoted by spark($\mathbf{A}$), is greater than $2k$ [18].

In order to consider the recovery of compressible signals, we let Λ a subset of $\{1, \ldots, n\}$ and Λ^c its complement set. Also, $\mathbf{h}_\Lambda$ is the same vector as $\mathbf{h}$ except for the entries with the indexes in Λ^c, which are set equal to zero. Then, the following property is useful to discuss the guarantee of the recovery of compressible signals:

Definition 4.2 (Null Space Property (NSP)): A matrix $\mathbf{A}$ satisfies the NSP of order k if there exists a constant $C > 0$ such that

$$\|\mathbf{h}_\Lambda\|_2 \le C\frac{\|\mathbf{h}_{\Lambda^c}\|_1}{\sqrt{k}} \tag{40}$$

holds for all $\mathbf{h} \in \mathcal{N}(\mathbf{A})$ and for all Λ such that $|\Lambda| \le k$.

Note that if we assume that $\mathbf{A}$ satisfies the NSP of order $2k$ and that $\mathbf{h} \in \mathcal{N}(\mathbf{A})$ is $2k$-sparse, then there exists an index set Λ such that $\|\mathbf{h}_{\Lambda^c}\|_1 = 0$, which together with the NSP implies $\mathbf{h}_\Lambda = \mathbf{0}$ to hold. Therefore, if $\mathbf{A}$ satisfies the NSP of order $2k$, $\mathbf{0}$ is the only $2k$-sparse vector in $\mathcal{N}(\mathbf{A})$, which coincides with the condition mentioned above.

One might think the definition of the NSP is somewhat strange because the ℓ_2-norm on the left-hand side is bounded not by ℓ_2-norm but by ℓ_1-norm. The definition is related to the inequality for the accuracy of the recovery defined as

$$\|\Delta(\mathbf{A}\mathbf{x}) - \mathbf{x}\|_2 \le C\frac{\sigma_k(\mathbf{x})_1}{\sqrt{k}}, \tag{41}$$

for all $\mathbf{x}$, where $\Delta(\cdot)$ denotes an arbitrary recovery algorithm and $\sigma_k(\mathbf{x})_1$ is defined in (4). Using the NSP, a necessary condition of the guarantee in the form of (41) for any recovery algorithm can be obtained as follows:

Theorem 4.2: [19] Let $\mathbf{A}$ be a sensing matrix and Δ a recovery algorithm. If the pair $(\mathbf{A}, \Delta)$ satisfies (41) then $\mathbf{A}$ satisfies the NSP of order $2k$.

Although the norm in (41) is arbitrary, it is known that if we replace the minimum ℓ_1-norm on the right-hand side of (41) with the minimum ℓ_2-norm, then the number of the measurements required to satisfy the guarantee can be close to n, which is of course not acceptable. Thus, combined with the fact that we have $\|\mathbf{x}\|_1/\sqrt{k} \le \|\mathbf{x}\|_2$ for $\mathbf{x} \in \Sigma_k$, it is reasonable to employ the form of the NSP defined above.

It should be noted that if a sensing matrix satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2} - 1$, then it also satisfies the NSP of order $2k$ [12]. Therefore, we can modify Theorem 4.1 to replace the RIP with the NSP. Thus, the NSP of order $2k$ is also sufficient to achieve the guarantee of (41) with the ℓ_1 optimization approach (26).

4.2.4 Neighborliness

As depicted in Fig. 2, success of ℓ_1 optimization problem in the case where $n = 2$, $m = 1$, and $k = 1$ is geometrically characterized as follows: Each vertex of the ℓ_1-ball corresponds to each sparsity pattern of two-dimensional 1-sparse vectors, taking into account the sign of the non-zero elements, such as $[+, 0]^{\mathrm{T}}$, $[0, -]^{\mathrm{T}}$, etc. Let us consider projection from $\mathbb{R}^2$ to a complementary subspace of $\mathcal{N}(\mathbf{A})$. Then, ℓ_1 optimization for 1-sparse vectors with a certain sparsity pattern succeeds if and only if the image of the vertex corresponding to the sparsity pattern by the projection is still a vertex of the image of the ℓ_1-ball (which in this case is actually a line segment).

This geometric interpretation can naturally be extended

to the general high-dimensional setting. Namely, ℓ_1 optimization for k-sparse vectors with a certain sparsity pattern gives the correct answers if and only if the image, via a projection onto a complementary subspace of $\mathcal{N}(\mathbf{A})$, of the $(k-1)$-dimensional face of the n-dimensional ℓ_1-ball, that is corresponding to the sparsity pattern considered, is still a face of the image of the ℓ_1-ball by the projection. In this interpretation, the image of the ℓ_1-ball by the projection is the polytope defined as the convex hull of the $2n$ points $\{\pm\mathbf{A}\mathbf{e}_1, \pm\mathbf{A}\mathbf{e}_2, \ldots, \pm\mathbf{A}\mathbf{e}_n\}$, where $\mathbf{e}_i$, $i = 1, \ldots, n$, denote the unit vectors forming the standard basis of $\mathbb{R}^n$. Then the success of ℓ_1 optimization can be related with the geometric property called the *neighborliness*, which in this case asks whether the k points chosen from the $2n$ points according to the sparsity pattern span a face of the polytope. The neighborliness property gives a necessary and sufficient condition for k-sparse vectors with a certain sparsity pattern to be exactly reconstructed via ℓ_1 optimization [20]–[23], as opposed to RIP, spark, and NSP, which only provide sufficient conditions.

4.2.5 Mutual Coherence

While the RIP, the spark, the NSP, and the neighborliness can provide guarantees of the recovery of the sparse signals, they suffer from the problem that it is very hard to verify whether a given matrix satisfies the properties or not. A simple way to provide a computable guarantee is to exploit the mutual coherence of the sensing matrix defined as follows [18], [24], [25],

Definition 4.3 (Mutual Coherence): The mutual coherence $\mu(\mathbf{A})$ of a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ is the largest absolute normalized inner product between different columns of $\mathbf{A}$

$$\mu(\mathbf{A}) = \max_{1\le i,j\le m,\ i\neq j} \frac{|\mathbf{a}_i^{\mathrm{T}}\mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2}, \tag{42}$$

where $\mathbf{a}_i$ is the i-th column of $\mathbf{A}$.

By applying the Geršgorin disk theorem to the Gram matrix $\mathbf{G} = \mathbf{A}_\Lambda^{\mathrm{T}}\mathbf{A}_\Lambda$, it is straightforward to obtain

$$\mathrm{spark}(\mathbf{A}) \ge 1 + \frac{1}{\mu(\mathbf{A})}. \tag{43}$$

Therefore, we have the following condition on $\mathbf{A}$ for the unique recovery.

Theorem 4.3: [18], [24] If

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{A})}\right), \tag{44}$$

then there exists at most one signal $\mathbf{x} \in \Sigma_k$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Some guarantees of recovery in terms of mutual coherence are also available for the case with the ℓ_1 optimization or the ℓ_1-ℓ_2 optimization [26], [27].

4.3 Random Matrix Ensembles

As we mentioned, most properties on sensing matrices are hard to evaluate on an instance of sensing matrices. It is also widely believed that it would be difficult to find a good sensing matrix in terms of those properties when n, m, k are arbitrarily given. In order to circumvent this difficulty from the theoretical side, one commonly-taken approach is to consider an ensemble of random matrices and to show that a property satisfies a condition for successful reconstruction with high probability, where the probability is based on the randomness of sensing matrices in the ensemble considered. Several insightful results have so far been obtained on the basis of considering random matrix ensembles. In many studies, ensembles of random matrices with independent and identically-distributed (i.i.d.) Gaussian-distributed elements are considered, but there are several other instances. Here, we show an example of the result on the RIP constant of random matrix ensembles:

Theorem 4.4: [28], [29] Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a Gaussian random matrix having i.i.d. elements of mean 0 and variance $1/m$ or a Bernoulli random matrix having i.i.d. elements equal to $\pm 1/\sqrt{m}$ with probability $1/2$. Let $\epsilon, \delta \in (0, 1)$ and if

$$m \ge C\delta^{-2}\left(k \ln\left(\frac{n}{k}\right) + \ln \epsilon^{-1}\right), \tag{45}$$

for a constant $C > 0$, then the constant of the RIP of $\mathbf{A}$ satisfies $\delta_k \le \delta$ with probability at least $1 - \epsilon$.

The significance of Theorem 4.4 is that the number of linear measurements to estimate any n-dimensional k-sparse vector via ℓ_1 optimization can be much smaller than n. Indeed, Theorem 4.4 states that it is sufficient to have linear measurements whose number is approximately proportional to k, with the extra penalty factor $\ln n$, with probability close to 1.

The notion of the neighborliness combined with random matrix ensembles yields a series of highly non-trivial results [20]–[23]. Among others, it has been found that ℓ_1 optimization for the exactly sparse and noiseless case exhibits phase transition in the infinite-dimensional limit, where the parameters n, m, k are sent to infinity while their ratios are kept finite. More concretely, given $\rho = k/m$ there is a critical threshold $\delta_c(\rho)$ of $\delta = m/n$ such that, with probability approaching 1 in the infinite-dimensional limit, ℓ_1 optimization gives the true answer if $\delta > \delta_c(\rho)$ and fails if $\delta < \delta_c(\rho)$. The threshold $\delta = \delta_c(\rho)$ is called the Donoho-Tanner threshold. It should be noted that the same result can be derived in several different ways, including the approach on the basis of statistical mechanics of disordered systems [30], [31] and the study on the basis of a modified version of the null space characterization [32]–[34]. For further extensions, see e.g., [35]–[42].

5. Algorithms

The original optimization problem (25) for compressed sensing is known to be NP-hard in general [43], [44]. In this section, we survey major practical algorithms to solve the problems of compressed sensing presented in Sect. 4. We mainly focus on two approaches; ℓ_1 optimization and greedy pursuits, since the approaches are well studied and widely used in practical applications of compressed sensing, and also almost all existing algorithms fall into one of the two approaches. For demonstration purpose, we have implemented all the algorithms (Algorithms 1–6) in Scilab [45] codes, which are available at [46]. Some other approaches are briefly introduced at the end of this section.

5.1 ℓ_1 Optimization

5.1.1 General-Purpose Tools or Tailored Ones?

We here consider ℓ_1 optimization problems which often arise in compressed sensing. The problems are summarized in Table 1.

An easy way to solve these problems is to use a general-purpose convex programming toolbox such as MATLAB Optimization Toolbox [56] or `cvx` [57], [58]. For example, if you want to solve the second problem in Table 1 by `cvx`, you can simply type the following commands on `cvx`:

```
cvx_begin
  variable x(n);
  minimize(norm(x,1));
  subject to
    norm(A*x-y,2)<=eps
cvx_end
```

However, when the problem is very large (i.e., n is a large number), such a general-purpose toolbox often requires a considerable length of computational time. To avoid this, one should investigate the structure of each problem and choose a suitable algorithm to exploit it. In what follows, we see algorithms that effectively work for each optimization problem listed in Table 1.

Table 1 ℓ_1 optimization problems and solvers.

Problem	Solver	Subsec.
$\min_{\mathbf{x}} \|\mathbf{x}\|_1$ subject to $\mathbf{Ax} = \mathbf{y}$	LP [47]	5.1.2
$\min_{\mathbf{x}} \|\mathbf{x}\|_1$ subject to $\|\mathbf{Ax} - \mathbf{y}\|_2 \le \epsilon$	NESTA [48]	5.1.3
	SPGL1 [49]	5.1.5
$\min_{\mathbf{x}} \left(\frac{1}{2}\|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1\right)$	FISTA [50]	5.1.4
	FPC [51]	5.1.5
	Bregman [52]	5.1.5
	GPSR [53]	5.1.5
	TwIST [54]	5.1.5
	SpaRSA [55]	5.1.5
$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{y}\|_2^2$ subject to $\|\mathbf{x}\|_1 \le t$	SPGL1 [49]	5.1.5

5.1.2 Linear Programming (LP)

We first consider the equality-constrained ℓ_1 optimization (26). As we have seen in the previous section, the optimization (26) can equivalently be reformulated as the LP problem

$$\min_{\mathbf{z}} \mathbf{1}_{2n}^{\mathrm{T}}\mathbf{z} \ \text{ subject to } \mathbf{A}_0\mathbf{z} = \mathbf{y} \text{ and } \mathbf{z} \succeq \mathbf{0}, \tag{46}$$

where $\mathbf{A}_0 = [\mathbf{A}, -\mathbf{A}]$. Several efficient schemes for solving LP have been available. Among them, one can, for example, adopt the *primal-dual interior-point* algorithm [47], [59]. We here briefly review it.

First of all, by the Karush-Kuhn-Tucker (KKT) conditions [59] (also known as the Kuhn-Tucker conditions [60]) at the optimal point $\mathbf{z}^*$ of the optimization (46), there exist two vectors $\boldsymbol{\nu}^* \in \mathbb{R}^m$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^{2n}$ such that

$$\mathbf{1}_{2n} + \mathbf{A}_0^{\mathrm{T}}\boldsymbol{\nu}^* - \boldsymbol{\lambda}^* = \mathbf{0}, \tag{47}$$

$$\lambda_i^* z_i^* = 0, \quad i = 1, 2, \ldots, 2n, \tag{48}$$

$$\mathbf{A}_0\mathbf{z}^* = \mathbf{y}, \tag{49}$$

$$\mathbf{z}^* \succeq \mathbf{0}, \quad \boldsymbol{\lambda}^* \succeq \mathbf{0}. \tag{50}$$

The primal-dual interior-point algorithm solves the nonlinear equations (47)–(49) by the Newton iteration method keeping the approximated vectors, say $\mathbf{z}[k]$, $\boldsymbol{\nu}[k]$, $\boldsymbol{\lambda}[k]$, $k = 0, 1, 2, \ldots$, at an *interior point* of the region defined by the inequalities (50), that is, $\mathbf{z}[k] \succ \mathbf{0}$ and $\boldsymbol{\nu}[k] \succ \mathbf{0}$. Because of this property, the method is also called an *interior-point method*. We may also relax the equation (48), which is called the complementary slackness condition, as

$$\lambda_i[k] z_i[k] = \frac{1}{\tau[k]}, \quad k = 0, 1, 2, \ldots, \tag{51}$$

where $\{\tau[k] : k = 0, 1, 2, \ldots\}$ is an increasing positive sequence, that is, $0 < \tau[0] < \tau[1] < \tau[2] < \cdots$.

To formulate Newton iteration, we define the primal, dual, and central residuals for the KKT conditions (47)–(49) as

$$\begin{aligned} \mathbf{r}_{\text{primal}} &:= \mathbf{A}_0\mathbf{z} - \mathbf{y}, \\ \mathbf{r}_{\text{dual}} &:= \mathbf{1}_{2n} + \mathbf{A}_0^{\mathrm{T}}\boldsymbol{\nu} - \boldsymbol{\lambda}, \\ \mathbf{r}_{\text{central}} &:= \boldsymbol{\Lambda}\mathbf{z} - \tau^{-1}\mathbf{1}_{2n}, \end{aligned} \tag{52}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_{2n}\}$ is a diagonal matrix with the diagonal elements $\{\lambda_1, \lambda_2, \ldots, \lambda_{2n}\}$. Suppose that we are given an interior point $\mathbf{p} = (\mathbf{z}, \boldsymbol{\nu}, \boldsymbol{\lambda})$, on the basis of which the residual vector $\mathbf{r}(\mathbf{p}) := (\mathbf{r}_{\text{primal}}, \mathbf{r}_{\text{dual}}, \mathbf{r}_{\text{central}})$ is determined. If the residual vector $\mathbf{r}(\mathbf{p})$ is sufficiently small, then the interior point $\mathbf{p}$ will be close to the optimal point $\mathbf{p}^* = (\mathbf{z}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*)$ which satisfies the KKT conditions (47)–(49) with (48) in place of the slackness condition (51).

Otherwise, we seek the next point $\mathbf{p} + \Delta\mathbf{p}$ such that

$$\mathbf{r}(\mathbf{p} + \Delta\mathbf{p}) \approx \mathbf{0}. \tag{53}$$

If one applies the Newton method to solve it, one first considers the Taylor expansion of the left-hand side of (53) around the interior point $\mathbf{p}$, yielding

$$\mathbf{r}(\mathbf{p}+\Delta\mathbf{p}) = \mathbf{r}(\mathbf{p}) + J_{\mathbf{r}}(\mathbf{p})\Delta\mathbf{p} + O\left(\|\Delta\mathbf{p}\|^2\right), \tag{54}$$

where $J_{\mathbf{r}}(\mathbf{p})$ is the Jacobian of $\mathbf{r}$ at $\mathbf{p}$. Linearizing this by neglecting the second- and higher-order terms, we have the linear system

$$J_{\mathbf{r}}(\mathbf{p})\Delta\mathbf{p} = -\mathbf{r}(\mathbf{p}), \tag{55}$$

or

$$\begin{bmatrix} \mathbf{A}_0 & 0 & 0 \\ 0 & \mathbf{A}_0^{\mathrm{T}} & -\mathbf{I} \\ \boldsymbol{\Lambda} & 0 & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{z} \\ \Delta\boldsymbol{\nu} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \mathbf{A}_0\mathbf{z} - \mathbf{y} \\ \mathbf{1}_{2n} + \mathbf{A}_0^{\mathrm{T}}\boldsymbol{\nu} - \boldsymbol{\lambda} \\ \boldsymbol{\Lambda}\mathbf{z} - \tau^{-1}\mathbf{1}_{2n} \end{bmatrix}, \tag{56}$$

where $\mathbf{Z} = \mathrm{diag}\{z_1, z_2, \ldots, z_{2n}\}$. By using the solution of the linear equations (55) or (56), we have the Newton iteration

$$\begin{aligned} \mathbf{p}[k+1] &= \mathbf{p}[k] + \Delta\mathbf{p}[k] \\ &= \mathbf{p}[k] - J_{\mathbf{r}}(\mathbf{p}[k])^{-1}\mathbf{r}(\mathbf{p}[k]), \quad k = 0, 1, 2, \ldots, \end{aligned} \tag{57}$$

with an initial interior point $\mathbf{p}[0]$ (i.e., $\mathbf{z}[0] \succ \mathbf{0}$ and $\boldsymbol{\lambda}[0] \succ \mathbf{0}$). At each step, we have to solve the linear equations (55) or (56) to obtain $\Delta\mathbf{p}[k]$, and this takes most of the computational time, especially when the size of the matrix $\mathbf{A}_0$ is large. In such cases, we may adopt the conjugate gradient method [61] for solving the linear equations.

In the interior-point method, we do not use (57) as it is since the next point $\mathbf{p}[k+1]$ might not be an interior point. To guarantee $\mathbf{p}[k+1]$ being an interior point, we modify the Newton iteration (57). That is, we use the step size parameter $s \in (0, 1]$ as

$$\mathbf{p}[k+1] = \mathbf{p}[k] + s\Delta\mathbf{p}[k]. \tag{58}$$

The step size s is chosen such that

1. $\mathbf{p} + s\Delta\mathbf{p}$ is an interior point, that is, $\mathbf{z} + s\Delta\mathbf{z} \succ \mathbf{0}$ and $\boldsymbol{\lambda} + s\Delta\boldsymbol{\lambda} \succ \mathbf{0}$.
2. The residual $\mathbf{r}(\mathbf{p} + s\Delta\mathbf{p})$ is sufficiently small, that is,

$$\|\mathbf{r}(\mathbf{p} + s\Delta\mathbf{p})\|_2 \le (1 - \alpha s)\|\mathbf{r}(\mathbf{p})\|_2, \tag{59}$$

where α is a sufficiently small number (e.g., $\alpha = 0.01$).

To achieve the first requirement, we choose s as follows: define the index sets

$$\mathcal{I}_{\mathbf{z}} := \{i : [\Delta\mathbf{z}]_i < 0\}, \quad \mathcal{I}_{\boldsymbol{\lambda}} := \{i : [\Delta\boldsymbol{\lambda}]_i < 0\}, \tag{60}$$

where $[\cdot]_i$ is the i-th element of a vector. Then, we can choose s as

$$s = 0.99 \min\left\{1, \{-z_i/\Delta z_i : i \in \mathcal{I}_{\mathbf{z}}\}, \{-\lambda_i/\Delta\lambda_i : i \in \mathcal{I}_{\boldsymbol{\lambda}}\}\right\}. \tag{61}$$

This s clearly satisfies the first requirement. Then, check if the second requirement is satisfied with this s. If the inequality (59) is not satisfied, change s as $s := s/a$ ($a > 1$) and check (59) again.

Finally, let us consider the stopping criterion. Assume that the primal and dual residuals, $\mathbf{r}_{\mathrm{primal}}$ and $\mathbf{r}_{\mathrm{dual}}$, are sufficiently small. Then the following value may be a measure for the precision of approximation:

$$\eta = \mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}. \tag{62}$$

This is called the surrogate duality gap. In summary, the primal-dual interior-point algorithm repeats the Newton iterations described above until the surrogate duality gap η gets smaller than a given tolerance. See Algorithm 1 for the details.

The MATLAB toolbox "ℓ_1-MAGIC" [47], [62] includes MATLAB routines for solving LP (26) with the primal-dual interior-point method. One can also use MATLAB Optimization Toolbox [56] or `cvx` [57], [58] mentioned above.

Algorithm 1 Primal-dual method for LP

Require: $\mathbf{y} \in \mathbb{R}^m$ (observed vector)
Ensure: $\mathbf{x} \in \mathbb{R}^n$ (estimated sparse vector)
Give $\mathbf{p}[0] := (\mathbf{z}[0], \boldsymbol{\nu}[0], \boldsymbol{\lambda}[0])$ and $\tau[0]$ such that

$$\mathbf{z}[0] > 0, \ \boldsymbol{\nu}[0] > 0, \ \tau[0] > 0.$$

$k := 0$.
repeat
  Solve (56) to obtain the step direction $\Delta\mathbf{p}[k]$.
  Determine the step length $s > 0$ as in Sect. 5.1.2.
  $\mathbf{p}[k+1] := \mathbf{p}[k] + s\Delta\mathbf{p}[k]$.
  Give $\tau[k+1]$ such that $\tau[k+1] > \tau[k]$.
  $\eta[k+1] := \mathbf{z}[k+1]^{\mathrm{T}}\boldsymbol{\lambda}[k+1]$.
  $k := k + 1$.
until $\max\{\|\mathbf{r}_{\mathrm{primal}}[k]\|_2, \|\mathbf{r}_{\mathrm{dual}}[k]\|_2\} \le$ EPS_FEAS **and** $\eta[k] \le$ EPS.
return $\mathbf{x} := [z_1[k], \ldots, z_n[k]]^{\mathrm{T}} - [z_{n+1}[k], \ldots, z_{2n}[k]]^{\mathrm{T}}$.

5.1.3 Nesterov's Algorithm (NESTA)

We next consider the quadratically constrained ℓ_1 optimization

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \le \epsilon, \tag{63}$$

where $\epsilon > 0$ is given (see also (33)). This problem also belongs to the class of the *second-order cone programming* (SOCP), and can be numerically solved by using a primal-dual method [63], or a log-barrier method [59]. However, these methods rely mainly on Newton steps as the method for LP mentioned above, which are problematic when we tackle a large-scale problem. In this paper we review a much faster alternative algorithm for the optimization (63) based on Nesterov's method, called Nesterov's algorithm or NESTA for short [48]. Throughout this subsection, we assume for simplicity that the row vectors of $\mathbf{A}$ are of unit length and mutually orthogonal, that is, we assume

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I}. \tag{64}$$

Before discussing NESTA for (63), let us consider a general convex optimization problem of the form

$$\min_{\mathbf{x}\in C} f(\mathbf{x}), \tag{65}$$

where f is a convex function on a closed convex subset C of the n-dimensional Euclidean space $\mathbb{R}^n$. Suppose that f is differentiable and that its gradient ∇f is Lipschitz, that is, there exists $L > 0$ such that for any $\mathbf{x}, \mathbf{v} \in C$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{v})\|_2 \le L\|\mathbf{x} - \mathbf{v}\|_2. \tag{66}$$

To solve (65) numerically, we first let a quadratic model $q_L : C \times C \to \mathbb{R}$ be defined as

$$q_L(\mathbf{x}, \mathbf{v}) := f(\mathbf{v}) + \nabla f(\mathbf{v})^{\mathrm{T}}(\mathbf{x} - \mathbf{v}) + \frac{L}{2}\|\mathbf{x} - \mathbf{v}\|_2^2. \tag{67}$$

Under the Lipshitz condition, $f(\mathbf{x}) \le q_L(\mathbf{x}, \mathbf{v})$ holds for all $\mathbf{x} \in C$, with equality if $\mathbf{x} = \mathbf{v}$, which implies that the function $q_L(\mathbf{x}, \mathbf{v})$ gives a quadratic approximation to $f(\mathbf{x})$ around the point $\mathbf{v} \in C$. By using this approximation, we can consider the following iteration for solving (65):

$$\mathbf{x}[k+1] = \arg\min_{\mathbf{x}\in C} q_L(\mathbf{x}, \mathbf{x}[k]), \quad k = 0, 1, 2, \ldots. \tag{68}$$

With the property of $q_L(\mathbf{x}, \mathbf{v})$ mentioned above, the sequence $\{\mathbf{x}[k] : k = 0, 1, \ldots\}$ via (68) satisfies

$$\begin{aligned} f(\mathbf{x}[k+1]) &\le q_L(\mathbf{x}[k+1], \mathbf{x}[k]) \\ &\le q_L(\mathbf{x}[k], \mathbf{x}[k]) \\ &\le f(\mathbf{x}[k]), \end{aligned} \tag{69}$$

for $k = 0, 1, 2, \ldots$. It follows that the iteration (68) is a descent scheme for (65). This scheme is also known as the majorization-minimization (MM) approach and the function $q_L(\mathbf{x}, \mathbf{v})$ is called a *majorizer* of $f(\mathbf{x})$ [64].

Since $q_L(\mathbf{x}, \mathbf{v})$ can be rewritten as

$$q_L(\mathbf{x}, \mathbf{v}) = \frac{L}{2}\left\|\mathbf{x} - (\mathbf{v} - L^{-1}\nabla f(\mathbf{v}))\right\|_2^2 - \frac{L}{2}\|\nabla f(\mathbf{v})\|_2^2 + f(\mathbf{v}), \tag{70}$$

the iteration (68) can alternatively be represented as

$$\begin{aligned} \mathbf{x}[k+1] &= \arg\min_{\mathbf{x}\in C} \left\|\mathbf{x} - (\mathbf{x}[k] - L^{-1}\nabla f(\mathbf{x}[k]))\right\|_2^2 \\ &= \mathcal{P}_C\left(\mathbf{x}[k] - L^{-1}\nabla f(\mathbf{x}[k])\right), \end{aligned} \tag{71}$$

where $\mathcal{P}_C : \mathbb{R}^n \to C$ is the orthogonal projection defined by

$$\mathcal{P}_C(\mathbf{v}) := \arg\min_{\mathbf{x}\in C} \|\mathbf{x} - \mathbf{v}\|_2^2, \quad \mathbf{v} \in \mathbb{R}^n. \tag{72}$$

Because of this property, the approximated optimization is called the *gradient projection method* [65], or *gradient method* [66], [67]. It is known that the sequence $\{\mathbf{x}[k] : k = 0, 1, \ldots\}$ generated by the gradient method with any initial value $\mathbf{x}[0]$ converges to the solution $\mathbf{x}^*$ of the original problem (65) with convergence rate $f(\mathbf{x}[k]) - f(\mathbf{x}^*) = O(1/k)$ [65]–[67].

In [68], Nesterov proposed an accelerated version of the gradient method, which is given in the following form:

$$\begin{aligned} \mathbf{w}[k] &= \arg\min_{\mathbf{x}\in C} q_L(\mathbf{x}, \mathbf{x}[k]), \\ \mathbf{z}[k] &= \arg\min_{\mathbf{x}\in C} \frac{L}{\sigma}p(\mathbf{x}) + \sum_{i=0}^{k} \alpha[i]\nabla f(\mathbf{x}[i])^{\mathrm{T}}(\mathbf{x} - \mathbf{x}[i]), \\ \mathbf{x}[k+1] &= \tau[k]\mathbf{z}[k] + (1 - \tau[k])\mathbf{w}[k], \end{aligned} \tag{73}$$

where $\sigma > 0$ is a convexity parameter, and where $p(\mathbf{x})$ is a *prox-function* for the feasible set C, that is, $p(\mathbf{x})$ vanishes at the prox-center

$$\mathbf{x}_{\mathrm{c}} = \arg\min_{\mathbf{x}\in C} p(\mathbf{x}), \tag{74}$$

and satisfies

$$p(\mathbf{x}) \ge \frac{\sigma}{2}\|\mathbf{x} - \mathbf{x}_{\mathrm{c}}\|_2, \tag{75}$$

for all $\mathbf{x} \in C$. In the iteration, the vector $\mathbf{z}[k]$ *keeps in mind* the previous iterations, which leads to acceleration of the gradient method. In fact, it is proved that if $\alpha[k]$ and $\tau[k]$ are chosen appropriately, for example letting

$$\alpha[k] = \frac{k+1}{2}, \quad \tau[k] = \frac{2}{k+3}, \quad k = 0, 1, 2, \ldots, \tag{76}$$

then the sequence $\{\mathbf{x}[k] : k = 0, 1, \ldots\}$ generated by the iteration (73) converges to $\mathbf{x}^*$ with quadratic convergence rate $O(1/k^2)$ [66].

In [66], Nesterov also extended the above method to a non-smooth convex function

$$f(\mathbf{x}) = \max_{\mathbf{u}\in C'} \mathbf{u}^{\mathrm{T}}\mathbf{W}\mathbf{x}, \tag{77}$$

where C' is a closed convex set in $\mathbb{R}^l$ (l is a positive integer) and where $\mathbf{W} \in \mathbb{R}^{l\times n}$. Note that in our ℓ_1 optimization (63),

$$f(\mathbf{x}) = \|\mathbf{x}\|_1 = \max_{\mathbf{u}\in C'} \mathbf{u}^{\mathrm{T}}\mathbf{x}, \tag{78}$$

with $C' = \{\mathbf{u} \in \mathbb{R}^n : \max_i |u_i| \le 1\}$. Then the optimization problem (65) becomes the following saddle point problem:

$$\min_{\mathbf{x}\in C} \max_{\mathbf{u}\in C'} \mathbf{u}^{\mathrm{T}}\mathbf{W}\mathbf{x}. \tag{79}$$

To approximate this non-smooth convex function by a smooth one, Nesterov proposed to minimize

$$f_\mu(\mathbf{x}) = \max_{\mathbf{u}\in C'} \left\{\mathbf{u}^{\mathrm{T}}\mathbf{W}\mathbf{x} - \mu p'(\mathbf{u})\right\}, \tag{80}$$

where $\mu > 0$ and where $p'(\mathbf{u})$ is a prox-function for C', that is, $p'(\mathbf{u})$ vanishes at the prox-center $\mathbf{u}'_{\mathrm{c}} \in C'$, and satisfies

$$p'(\mathbf{u}) \ge \frac{\sigma'}{2}\|\mathbf{u} - \mathbf{u}'_{\mathrm{c}}\|_2^2 \tag{81}$$

for all $\mathbf{u} \in C'$, with a positive constant σ'. The function f_μ is convex and gives a smooth approximation of the non-smooth convex function f. The parameter μ controls the degree of smoothing, and $\lim_{\mu\to 0} f_\mu = f$ holds. Indeed, Nesterov proved that f_μ is continuously differentiable, and that the gradient satisfies

$$\nabla f_\mu(\mathbf{x}) = W^{\mathrm{T}}\mathbf{u}_\mu(\mathbf{x}), \tag{82}$$

where

$$\mathbf{u}_\mu(\mathbf{x}) = \arg\max_{\mathbf{u}\in C'} \left\{\mathbf{u}^{\mathrm{T}}\mathbf{W}\mathbf{x} - \mu p'(\mathbf{u})\right\}. \tag{83}$$

It is also shown that ∇f_μ is Lipschitz with constant $L_\mu = (\mu\sigma')^{-1}\|\mathbf{W}\|^2$. Nesterov's algorithm for the non-smooth $f(\mathbf{x})$ in (77) employs the iteration (73) with the smooth approximation f_μ. For a fixed $\mu > 0$, this algorithm converges at a rate of $O(1/k^2)$ [66].

Now, let us apply Nesterov's algorithm to our ℓ_1 optimization (63). For the ℓ_1-norm objective function (78), the smooth approximation f_μ can be chosen as

$$f_\mu(\mathbf{x}) = \max_{\mathbf{u}\in C'} \left\{\mathbf{u}^{\mathrm{T}}\mathbf{x} - \frac{\mu}{2}\mathbf{u}^{\mathrm{T}}\mathbf{u}\right\} = \sum_{i=1}^{n} |x_i|_\mu, \tag{84}$$

where $|\cdot|_\mu : \mathbb{R} \to [0,\infty)$ is a function defined by

$$|x|_\mu = \begin{cases} \frac{1}{2\mu}x^2, & \text{if } |x| < \mu, \\ |x| - \frac{\mu}{2}, & \text{otherwise.} \end{cases} \tag{85}$$

This function also appears in the context of robust statistics [69], [70], and is thus known as Huber's loss function [71], [72]. Note that the prox-function $p'(\mathbf{u})$ in (80) is now chosen as $p'(\mathbf{u}) = \frac{1}{2}\|\mathbf{u}\|_2^2$. For this approximating function, the gradient ∇f_μ is given by a saturation function

$$\begin{aligned}[\nabla f_\mu(\mathbf{x})]_i &= [\mathrm{sat}_\mu(\mathbf{x})]_i \\ &= \begin{cases} \frac{1}{\mu}x_i, & \text{if } |x_i| < \mu \\ \mathrm{sgn}(x_i), & \text{otherwise,} \end{cases}\end{aligned} \tag{86}$$

where

$$\mathrm{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ -1, & \text{if } x < 0. \end{cases} \tag{87}$$

The Lipschitz constant is $L_\mu = \mu^{-1}$ in this case.

Then the problem is approximated as the following smooth constrained optimization problem:

$$\min_{\mathbf{x}\in C} f_\mu(\mathbf{x}), \quad C = \{x \in \mathbb{R} : \|\mathbf{y} - A\mathbf{x}\|_2 \leq \epsilon\}. \tag{88}$$

This is solved by Nesterov's iteration (73), which includes the minimization steps for $\mathbf{w}[k]$ and $\mathbf{z}[k]$. The KKT conditions for the minimization in $\mathbf{w}[k]$ give the following linear system:

$$\left(\mathbf{I} + \mu\lambda_1\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\mathbf{w}[k] = \mu\lambda_1\mathbf{A}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\phi}(\mathbf{x}[k]), \tag{89}$$

where

$$\begin{aligned}&\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x} - \mu\nabla f_\mu(\mathbf{x}) = \mathbf{x} - \mu\cdot\mathrm{sat}_\mu(\mathbf{x}), \\ &\lambda_1 = \max\{0, (\epsilon\mu)^{-1}\|\mathbf{y} - \mathbf{A}\boldsymbol{\phi}(\mathbf{x}[k])\|_2 - \mu^{-1}\}.\end{aligned} \tag{90}$$

By the orthonormality assumption (64), the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is a projection matrix and is thus idempotent, so that we have

$$\left(\mathbf{I} + \mu\lambda_1\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} = I - \frac{\mu\lambda_1}{\mu\lambda_1 + 1}\mathbf{A}^{\mathrm{T}}\mathbf{A}, \tag{91}$$

and hence $\mathbf{w}[k]$ is directly obtained by

$$\mathbf{w}[k] = \left(I - \frac{\mu\lambda_1}{\mu\lambda_1 + 1}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left(\mu\lambda_1\mathbf{A}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\phi}(\mathbf{x}[k])\right). \tag{92}$$

For $\mathbf{z}[k]$ in Nesterov's iteration (73), set $\sigma = 1$ and take the prox-function

$$p(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{x}[0]\|_2^2, \tag{93}$$

where $\mathbf{x}[0]$ is an initial guess for $\mathbf{x}^*$ (e.g., $\mathbf{x}[0] = \mathbf{A}^{\mathrm{T}}\mathbf{y}$). Then the KKT conditions for the minimization in $\mathbf{z}[k]$ lead to the linear system

$$\left(\mathbf{I} + \mu\lambda_2\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\mathbf{z}[k] = \mu\lambda_2\mathbf{A}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\psi}[k], \tag{94}$$

where

$$\begin{aligned}&\boldsymbol{\psi}[k] = \mathbf{x}[0] - \mu\sum_{i=0}^{k}\alpha[i]\mathrm{sat}_\mu(\mathbf{x}[i]), \\ &\lambda_2 = \max\{0, (\epsilon\mu)^{-1}\|\mathbf{y} - \mathbf{A}\boldsymbol{\psi}[k]\|_2 - \mu^{-1}\}.\end{aligned} \tag{95}$$

The inverse property in (91) gives

$$\mathbf{z}[k] = \left(I - \frac{\mu\lambda_2}{\mu\lambda_2 + 1}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left(\mu\lambda_2\mathbf{A}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\psi}[k]\right). \tag{96}$$

In NESTA, parameters $\alpha[k]$ and $\tau[k]$ are chosen as in (76), by which the sequence $\{\mathbf{x}[k] : k = 0, 1, \ldots\}$ of NESTA converges to the optimal solution of the approximated problem (88) with convergence rate $O(1/k^2)$. We describe NESTA in Algorithm 2. It is clear that under the orthonormality assumption on $\mathbf{A}$ the most expensive computation in the iteration is matrix-vector multiplication, and hence the computational time at each step has significantly been reduced compared with that of the primal-dual interior-point algorithm described in Sect. 5.1.2, which involves inversion of a matrix (or solving a linear system) in (55) or (56). We would like to mention that if one does not have

Algorithm 2 NESTA

Require: $\mathbf{y} \in \mathbb{R}^m$ {observed vector}
Ensure: $\mathbf{x} \in \mathbb{R}^n$ {estimated sparse vector}
  $\mathbf{x}[0] := \mathbf{A}^{\mathrm{T}}\mathbf{y}$. {initial guess}
  $\boldsymbol{\psi}[-1] := \mathbf{x}[0]$.
  $k := 1$.
  repeat
    $\lambda_1 := \max\left\{0, (\epsilon\mu)^{-1}\|\mathbf{y} - \mathbf{A}\boldsymbol{\phi}(\mathbf{x}[k])\|_2 - \mu^{-1}\right\}$.
    $\mathbf{w}[k] := \left(I - \frac{\mu\lambda_1}{\mu\lambda_1+1}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left(\mu\lambda_1\mathbf{A}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\phi}(\mathbf{x}[k])\right)$.
    $\boldsymbol{\psi}[k] := \boldsymbol{\psi}[k-1] - \mu\cdot\frac{k+1}{2}\cdot\mathrm{sat}(\mathbf{x}[k])$.
    $\lambda_2 := \max\left\{0, (\epsilon\mu)^{-1}\|\mathbf{y} - \mathbf{A}\boldsymbol{\psi}[k]\|_2 - \mu^{-1}\right\}$.
    $\mathbf{z}[k] := \left(I - \frac{\mu\lambda_2}{\mu\lambda_2+1}\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\left(\mu\lambda_2\mathbf{A}^{\mathrm{T}}\mathbf{y} + \boldsymbol{\psi}[k]\right)$.
    $\mathbf{x}[k+1] := \frac{2}{k+3}\mathbf{z}[k] + \frac{k+1}{k+3}\mathbf{w}[k]$.
    $k := k + 1$.
  until $|f_\mu(\mathbf{x}[k]) - f_\mu(\mathbf{x}[k-1])| \leq \mathrm{EPS}$.
  return $\mathbf{x} := \mathbf{x}[k]$.

the orthonormality assumption on $\mathbf{A}$, one has to solve the linear equations (89) and (94), which, however, can still be feasible for moderate-sized problems.

The parameter $\mu > 0$ should be chosen taking account of a trade-off between the accuracy of the approximation f_μ (note that $\lim_{\mu\to 0} f_\mu(\mathbf{x}) = \|\mathbf{x}\|_1$) and the rate of convergence (the rate is proportional to μ). MATLAB codes for NESTA are available at [73].

5.1.4 Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)

Unlike the constrained ℓ_1 optimization discussed in Sect. 5.1.2 or 5.1.3, there have been considerable researches on efficient algorithms (e.g., [51]–[55]) for the ℓ_1-ℓ_2 optimization

$$\min_{\mathbf{x}} \left(\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right), \tag{97}$$

where $\lambda > 0$ is a fixed parameter. In this subsection, we review an efficient algorithm for (97), called *fast iterative shrinkage-thresholding algorithm* (FISTA) [50].

First of all, let us consider the simplest case of (97); namely, $m = n$ and $\mathbf{A}$ is orthogonal. In this case, we have $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I}$, and

$$\begin{aligned} &\min_{\mathbf{x}} \left(\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right) \\ &\quad = \min_{\mathbf{x}} \left(\frac{1}{2}\left\|\mathbf{A}(\mathbf{x} - \mathbf{A}^{\mathrm{T}}\mathbf{y})\right\|_2^2 + \lambda \|\mathbf{x}\|_1 \right) \\ &\quad = \min_{\mathbf{x}} \left(\frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\|_2^2 + \lambda \|\mathbf{x}\|_1 \right), \end{aligned} \tag{98}$$

where $\mathbf{x}_0 = \mathbf{A}^{\mathrm{T}}\mathbf{y}$. Using the fact that the minimizer of the scalar function $g(t) = (1/2)(t - t_0)^2 + \lambda|t|$ is obtained as $t^* = \mathrm{sgn}(t_0)\max\{0, |t_0| - \lambda\}$, where sgn is defined in (87), the optimal solution of (97) in this case is

$$\mathbf{x}^* = \mathcal{S}_\lambda(\mathbf{x}_0) = \mathcal{S}_\lambda(\mathbf{A}^{\mathrm{T}}\mathbf{y}), \tag{99}$$

where $\mathcal{S}_\lambda$ is a shrinkage-thresholding function defined by

$$[\mathcal{S}_\lambda(\mathbf{x})]_i = \mathrm{sgn}(x_i)\max\{0, |x_i| - \lambda\}, \quad i = 1, 2, \ldots, n. \tag{100}$$

For a general rectangular matrix $\mathbf{A}$, we begin with an approximation model as in NESTA. For the objective function in (97), we define the following approximation model (cf. (67)):

$$\begin{aligned} Q_L(\mathbf{x}, \mathbf{v}) &= \frac{1}{2}\|\mathbf{A}\mathbf{v} - \mathbf{y}\|_2^2 + (\mathbf{x} - \mathbf{v})^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{y}) \\ &\quad + \frac{L}{2}\|\mathbf{x} - \mathbf{v}\|_2^2 + \lambda\|\mathbf{x}\|_1 \\ &= \frac{L}{2}\left\|\mathbf{x} - (\mathbf{v} - L^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{v} - \mathbf{y}))\right\|_2^2 + \lambda\|\mathbf{x}\|_1 \\ &\quad - \frac{L}{2}\left\|\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{v} - \mathbf{y})\right\|_2^2 + \frac{1}{2}\|\mathbf{A}\mathbf{v} - \mathbf{y}\|_2^2, \end{aligned} \tag{101}$$

Algorithm 3 FISTA

Require: $\mathbf{y} \in \mathbb{R}^m$ {observed vector}
Ensure: $\mathbf{x} \in \mathbb{R}^n$ {estimated sparse vector}
  $\mathbf{x}[0] := \mathbf{A}^{\mathrm{T}}\mathbf{y}$. {initial guess}
  $\mathbf{w}[1] := \mathbf{x}[0]$.
  $\beta[1] := 1$.
  $k := 1$.
  repeat
    $\mathbf{x}[k] := \mathcal{S}_{\lambda/L}\left(\mathbf{w}[k] + L^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{y} - \mathbf{A}\mathbf{w}[k])\right)$.
    $\beta[k+1] := \frac{1}{2} + \sqrt{\frac{1}{4} + \beta[k]^2}$.
    $\mathbf{w}[k+1] := \mathbf{x}[k] + \frac{\beta[k]-1}{\beta[k+1]}(\mathbf{x}[k] - \mathbf{x}[k-1])$.
    $k := k + 1$.
  until $|F(\mathbf{x}[k-1]) - F(\mathbf{x}[k-2])| \le \mathrm{EPS}$.
  return $\mathbf{x} := \mathbf{x}[k-1]$.

where $L > 0$ is a parameter chosen such that

$$L \ge \|\mathbf{A}^{\mathrm{T}}\mathbf{A}\|. \tag{102}$$

Note that $\|\mathbf{A}^{\mathrm{T}}\mathbf{A}\|$ is the Lipschitz constant of the gradient of $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$. It follows that if (102) holds, then $Q_L(\mathbf{x}, \mathbf{v})$ is a majorizer of the objective function $F(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1$, that is, $F(\mathbf{x}) \le Q_L(\mathbf{x}, \mathbf{v})$ holds for all $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$ with equality if $\mathbf{x} = \mathbf{v}$. Then we form the following iteration:

$$\begin{aligned} \mathbf{x}[k+1] &= \arg\min_{\mathbf{x}} Q_L(\mathbf{x}, \mathbf{x}[k]) \\ &= \arg\min_{\mathbf{x}} \left(\frac{L}{2}\|\mathbf{x} - \boldsymbol{\phi}(\mathbf{x}[k])\|_2^2 + \lambda\|\mathbf{x}\|_1 \right), \\ &\quad k = 0, 1, 2, \ldots, \end{aligned} \tag{103}$$

where $\boldsymbol{\phi}(\mathbf{x}) = \mathbf{x} - L^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{y})$. Since Q_L is a majorizer, we have $F(\mathbf{x}[k+1]) \le F(\mathbf{x}[k])$ for $k = 0, 1, 2, \ldots$ (see (69)).

The closed-form representation (99) of the solution of the simple optimization (98) gives the minimizer of (103) as $\mathcal{S}_{\lambda/L}(\boldsymbol{\phi}(\mathbf{x}[k]))$. As a result, the iteration can be rewritten in a compact form as

$$\mathbf{x}[k+1] = \mathcal{S}_{\lambda/L}(\mathbf{x}[k] + L^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{y} - \mathbf{A}\mathbf{x}[k])). \tag{104}$$

This iteration is also known as the iterative shrinkage-thresholding algorithm (ISTA) [74]–[77]. It is proved that if $L > 0$ is chosen to satisfy (102), then the sequence $\{\mathbf{x}[k] : k = 0, 1, \ldots\}$ by ISTA (104) converges to the optimal solution of (97) with convergence rate $O(1/k)$ [50].

FISTA (Fast ISTA) is an accelerated version of ISTA. As in NESTA (see Sect. 5.1.3), FISTA updates the solution based not only on quantities evaluated in the previous iteration, but also on two or more previously computed ones. In fact, FISTA is an extension of Nesterov's work [68] to achieve the convergence rate $O(1/k^2)$ [50]. Algorithm 3 describes FISTA, where $F(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1$. MATLAB codes related to FISTA may be found at [78].

5.1.5 Other Algorithms for ℓ_1 Optimization

A considerable number of studies have been made on fast and accurate algorithms for ℓ_1 optimization besides LP,

FISTA, and NESTA mentioned above. We here overview recent algorithms for ℓ_1 optimization.

For the ℓ_2-constrained ℓ_1 optimization (63), an efficient algorithm based on the spectral projection gradient (SPG) method was proposed in [49]. The algorithm solves iteratively the following optimization:

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 \ \text{ subject to } \ \|\mathbf{x}\|_1 \le t. \tag{105}$$

with updating the parameter t. The algorithm is called SPGL1 (L1 stands for ℓ_1).

On the other hand, Hale, Yin and Zhang proposed an iteration algorithm for the ℓ_1-ℓ_2 optimization problem (97) based on the fixed-point continuation (FPC) method. They introduced a fixed point equation $\mathbf{x} = F(\mathbf{x})$ that holds at the optimal solution $\mathbf{x} = \mathbf{x}^*$, and form $\mathbf{x}[k+1] = F(\mathbf{x}[k])$. Yin et al. proposed an algorithm for (97), called Bregman iteration, which solves at each step a subproblem, the Bregman-distance regularization, by the FPC algorithm mentioned above. Figueiredo, Nowak, and Wright also proposed in [53] an algorithm for (97), which reformulates the problem as a bound-constrained quadratic programming, which is solved by the gradient projection (GP) method with Barzilai-Borwein steps. This algorithm is called GPSR, where SR stands for sparse reconstruction. TwIST (two-step iterative shrinkage-thresholding) algorithm [54] is also a fast algorithm for (97), in which each iterate depends on the two previous iterates in IST like FISTA. Yet another algorithm for (97) is SpaRSA (sparse reconstruction by separable approximation) by Wright, Nowak, and Figueiredo, which is much like FISTA and FPC mentioned above.

5.2 Greedy Algorithms

5.2.1 Preliminaries

Let us return to the original compressed sensing problem (see (25)):

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \ \text{ subject to } \ \mathbf{A}\mathbf{x}=\mathbf{y}. \tag{106}$$

Suppose temporarily that the mutual coherence (see Definition 4.3) of $\mathbf{A}$ satisfies $\mu(\mathbf{A}) < 1$ and that we know that the optimal vector $\mathbf{x}^*$ of (106) is 1-sparse, that is, $\|\mathbf{x}^*\|_0 = 1$. Then, the solution is unique by Theorem 4.3, and the vector $\mathbf{y}$ is a scalar multiple of a column vector in $\mathbf{A}$. That is, there exists $J \in \{1,2,\dots,n\}$ and $z^* \in \mathbb{R}$ such that $\mathbf{y} = z^*\mathbf{a}_J$, where $\mathbf{a}_J$ is the J-th column vector of $\mathbf{A}$. To find the optimal index J, we define the error function $e(j)$ as

$$\begin{aligned} e(j) &= \min_z \|z\mathbf{a}_j - \mathbf{y}\|_2^2 \\ &= \min_z \left[(\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j)z^2 - 2(\mathbf{a}^{\mathrm{T}}\mathbf{y})z + \mathbf{y}^{\mathrm{T}}\mathbf{y}\right] \\ &= \min_z (\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j)\left(z - \frac{\mathbf{a}_j^{\mathrm{T}}\mathbf{y}}{\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j}\right)^2 + \mathbf{y}^{\mathrm{T}}\mathbf{y} - \frac{(\mathbf{a}_j^{\mathrm{T}}\mathbf{y})^2}{\mathbf{a}_j^{\mathrm{T}}\mathbf{a}_j} \\ &= \|\mathbf{y}\|_2^2 - \frac{(\mathbf{a}_j^{\mathrm{T}}\mathbf{y})^2}{\|\mathbf{a}_j\|_2^2}, \quad j = 1,2,\dots,n. \end{aligned} \tag{107}$$

By the assumptions, there exists $J \in \{1,2,\dots,n\}$ such that $e(J) = 0$, and the optimal solution $\mathbf{x}^* = [x_1^*,\dots,x_n^*]^{\mathrm{T}}$ is given by

$$x_j^* = \begin{cases} \frac{\mathbf{a}_J^{\mathrm{T}}\mathbf{y}}{\|\mathbf{a}_J\|_2^2}, & \text{if } \ j = J, \\ 0, & \text{otherwise.} \end{cases} \tag{108}$$

Note that finding J requires just n steps.

Then suppose that $\mu(\mathbf{A}) < \frac{1}{2M-1}$, $M > 1$, and $\|\mathbf{x}^*\|_0 = M$. Again, by Theorem 4.3, the solution is unique, and the vector $\mathbf{y}$ is a linear combination of at most M columns of $\mathbf{A}$. Since there exist $\binom{n}{M} \sim O(n^M)$ patterns of the linear combination, testing all of the candidates as above may be prohibitive in view of computational time if M is relatively large. To avoid such exhaustive search, we can employ *greedy algorithms* [6, Sect. 12.3] [18, Sect. 3.1] [79]. A greedy algorithm iteratively builds up the approximate solution of (106) by updating the support set one by one. Although greedy algorithms do not lead to the optimal solution but a local minimum in general, it may outperform the ℓ_1 optimization in some cases [25], [80]. In this subsection, we introduce major greedy algorithms to obtain the solution of the compressed sensing problem (106).

5.2.2 Matching Pursuit (MP)

One of the simplest greedy algorithms is the *matching pursuit* (MP) [5], [81], also known as the pure greedy algorithm in approximation theory [82].

MP optimizes the approximation by selecting a column vector at each step. At the first step ($k = 1$), we search for the best 1-sparse approximation $\mathbf{x}[1]$ of $\mathbf{x}$ in the sense of minimizing the residual $\mathbf{y} - A\mathbf{x}[1]$. As in the argument in Sect. 5.2.1, it can be found via

$$\begin{aligned} J[1] &= \arg\min_j \ e(j) \\ &= \arg\min_j \ \left\{\|\mathbf{y}\|_2^2 - \frac{(\mathbf{a}_j^{\mathrm{T}}\mathbf{y})^2}{\|\mathbf{a}_j\|_2^2}\right\} \\ &= \arg\max_j \ \frac{|\mathbf{a}_j^{\mathrm{T}}\mathbf{y}|}{\|\mathbf{a}_j\|_2}, \\ z[1] &= \frac{\mathbf{a}_{J[1]}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{a}_{J[1]}\|_2^2}. \end{aligned} \tag{109}$$

The first approximation $\mathbf{x}[1]$ is then given by setting the $J[1]$-th element of $\mathbf{x}[0] = \mathbf{0}$ by $z[1]$, that is, $x_{J[1]}[1] = z[1]$. This gives the residual $\mathbf{r}[1] = \mathbf{y} - \mathbf{A}\mathbf{x}[1] = \mathbf{y} - z[1]\mathbf{a}_{J[1]}$, which results in the best 1-sparse approximation of the vector $\mathbf{y}$:

$$\mathbf{y} = z[1]\mathbf{a}_{J[1]} + \mathbf{r}[1]. \tag{110}$$

At the next step ($k = 2$), MP further approximates the *residual* $\mathbf{r}[1]$ by a 1-sparse vector $z[2]\mathbf{a}_{J[2]}$ just as in the first step described above, that is,

$$J[2] = \arg\max_j \frac{|\mathbf{a}_j^{\mathrm{T}}\mathbf{r}[1]|}{\|\mathbf{a}_j\|_2}. \tag{111}$$

Algorithm 4 Matching Pursuit (MP)

Require: $\mathbf{y} \in \mathbb{R}^m$ {observed vector}
Ensure: $\mathbf{x} \in \mathbb{R}^n$ {estimated sparse vector}
  $\mathbf{x}[0] := \mathbf{0}$.
  $\mathbf{r}[0] := \mathbf{y} - \mathbf{A}\mathbf{x}[0] = \mathbf{y}$.
  $k := 0$.
  repeat
    $J := \arg\max_j \frac{|\mathbf{a}_j^{\mathrm{T}}\mathbf{r}[k]|}{\|\mathbf{a}_j\|_2}, \quad z^* := \frac{\mathbf{a}_J^{\mathrm{T}}\mathbf{r}[k]}{\|\mathbf{a}_J\|_2^2}$.
    $x_J[k+1] := x_J[k] + z^*$.
    $\mathbf{r}[k+1] := \mathbf{y} - \mathbf{A}\mathbf{x}[k+1] = \mathbf{r}[k] - z^*\mathbf{a}_J$.
    $k := k + 1$.
  until $\|\mathbf{r}[k]\|_2 \le \mathtt{EPS}$.
  return $\mathbf{x} := \mathbf{x}[k]$.

Then we obtain a 2-sparse approximation of $\mathbf{y}$:

$$\mathbf{y} = z[1]\mathbf{a}_{J[1]} + z[2]\mathbf{a}_{J[2]} + \mathbf{r}[2]. \tag{112}$$

In the same manner, we will obtain an M-sparse approximation of the vector $\mathbf{y}$ after M steps:

$$\mathbf{y} = \sum_{k=1}^{M} z[k]\mathbf{a}_{J[k]} + \mathbf{r}[M]. \tag{113}$$

It is proved in [5] that the residual sequence $\{\mathbf{r}[k] : k = 0, 1, \dots\}$ converges linearly to zero if

$$\mathrm{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\} = \mathbb{R}^m. \tag{114}$$

The number of iterations is now remarkably reduced compared with the exhaustive search mentioned above, which requires solving minimization problems $O(n^M)$ times. Algorithm 4 describes the algorithm of MP.

5.2.3 Orthogonal Matching Pursuit (OMP)

Orthogonal Matching Pursuit (OMP) is an improved version of MP [81], [83]. OMP is also known as orthogonal greedy algorithm in approximation theory [82].

At step $k \ge 1$, OMP first selects the optimal index $J[k]$ as in (109) or (111) and updates the support set as $\Lambda[k] = \Lambda[k-1] \cup \{J[k]\}$, with the initialization $\Lambda[0] = \emptyset$. OMP then updates the vector $\mathbf{x}[k]$ by orthogonally projecting the measurement vector $\mathbf{y}$ onto the subspace spanned by the column vectors $\{\mathbf{a}_j : j \in \Lambda[k]\}$. In other words, $\mathbf{x}[k]$ minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$ subject to $\mathrm{supp}(\mathbf{x}) = \Lambda[k]$. The full algorithm of OMP is described in Algorithm 5. In this algorithm, $\mathbf{x}_\Lambda$ is the vector of length $|\Lambda|$ obtained by collecting the entries of $\mathbf{x}$ corresponding to the support set Λ, and $\mathbf{A}_\Lambda$ is the submatrix of $\mathbf{A}$ of size $m \times |\Lambda|$ composed of the column vectors of $\mathbf{A}$ corresponding to Λ.

A major difference of OMP compared with MP is that OMP will never select the same index twice since the residual $\mathbf{r}[k]$ is orthogonal to the already chosen column vectors $\{\mathbf{a}_j : j \in \Lambda[k]\}$. As a result, if (114) holds, OMP will always produce an estimate $\mathbf{x}$ which satisfies $\mathbf{A}\mathbf{x} = \mathbf{y}$ after m iterations. Moreover, if at least one of the following statements is true:

Algorithm 5 Orthogonal Matching Pursuit (OMP)

Require: $\mathbf{y} \in \mathbb{R}^m$ {observed vector}
Ensure: $\mathbf{x} \in \mathbb{R}^n$ {estimated sparse vector}
  $\mathbf{x}[0] := \mathbf{0}$.
  $\mathbf{r}[0] := \mathbf{y} - \mathbf{A}\mathbf{x}[0] = \mathbf{y}$.
  $\Lambda := \emptyset$.
  $k := 0$.
  repeat
    $J := \arg\max_j \frac{|\mathbf{a}_j^{\mathrm{T}}\mathbf{r}[k]|}{\|\mathbf{a}_j\|_2}$.
    $\Lambda := \Lambda \cup \{J\}$.
    $\mathbf{x}_\Lambda[k+1] := \arg\min_{\mathbf{v}} \|\mathbf{A}_\Lambda\mathbf{v} - \mathbf{y}\|_2^2$.
    $\mathbf{r}[k+1] := \mathbf{y} - \mathbf{A}\mathbf{x}[k+1] = \mathbf{y} - \mathbf{A}_\Lambda\mathbf{x}_\Lambda[k+1]$.
    $k := k + 1$.
  until $\|\mathbf{r}[k]\|_2 \le \mathtt{EPS}$.
  return $\mathbf{x} := \mathbf{x}[k]$.

1. the mutual coherence of $\mathbf{A}$ satisfies $\mu(\mathbf{A}) < \frac{1}{2M-1}$ [84],
2. $\mathbf{A}$ satisfies the RIP (see Definition 4.1) of order $M+1$ with constant $\delta_{M+1} < \frac{1}{3\sqrt{M}}$ [85],

then OMP will recover any M-sparse vector $\mathbf{x}$ from the measurement $\mathbf{y} = \mathbf{A}\mathbf{x}$ in M iterations. A probabilistic guarantee with a random matrix $\mathbf{A}$ is also obtained in [25]; if the row vectors of $\mathbf{A}$ are drawn independently from the n-dimensional standard Gaussian distribution, then OMP will recover M-sparse vectors with high probability using only $m \sim O(M \log n)$ measurements.

5.2.4 Other Greedy Algorithms

Based on OMP, a couple of variants have been proposed. Here we briefly present an overview of the current state of greedy algorithms.

Gradient Pursuit [86]**:** The computational cost of OMP is dominated by the orthogonal projection step (see Algorithm 5):

$$\mathbf{x}_\Lambda[k+1] := \arg\min_{\mathbf{v}} \|\mathbf{A}_\Lambda\mathbf{v} - \mathbf{y}\|_2^2. \tag{115}$$

To avoid this, the step (115) is replaced by

$$\mathbf{x}_\Lambda[k+1] := \mathbf{x}_\Lambda[k] - s[k]\mathbf{A}_\Lambda^{\mathrm{T}}(\mathbf{A}_\Lambda\mathbf{x}[k] - \mathbf{y}), \tag{116}$$

where $s[k]$ is a step size and the update direction is the negative gradient of the cost function in (115). This is called the gradient pursuit. The conjugate gradient can also be adopted.

StOMP [87]**:** MP or OMP selects *just one* column from $\mathbf{A}$ at each iteration. This means that for an M-sparse vector, we need at least M iterations. To speed up the algorithm, we can select *multiple* columns at each step. More precisely, the support set update step in OMP (see Algorithm 5) can be replaced by

$$\Lambda := \Lambda \cup \left\{ j : \frac{|\mathbf{a}_j^{\mathrm{T}}\mathbf{r}[k]|}{\|\mathbf{a}_j\|_2} \ge T[k] \right\}, \tag{117}$$

where $T[k] > 0$ is a threshold parameter determining which columns are to be selected for addition to

the support set. This idea was first proposed in [87] and called the stagewise orthogonal matching pursuit (StOMP).

ROMP [88], [89]**:** Regularized orthogonal matching pursuit (ROMP) is another alternative multi-column selection method proposed in [88], [89]. In this algorithm, the index set $\{1, 2, \ldots, n\}$ is divided into different groups $\mathcal{J}_1, \mathcal{J}_2, \ldots$, such that in each group, the vectors $\{\mathbf{a}_j^{\mathrm{T}}/\|\mathbf{a}_j\|_2 : j \in \mathcal{J}_l\}$ have similar magnitudes, that is, for any $i, j \in \mathcal{J}_l$,

$$\frac{|\mathbf{a}_i^{\mathrm{T}}\mathbf{r}[k]|}{\|\mathbf{a}_i\|_2} \leq 2\frac{|\mathbf{a}_j^{\mathrm{T}}\mathbf{r}[k]|}{\|\mathbf{a}_j\|_2}. \tag{118}$$

Then the support set Λ is updated as

$$\begin{aligned} L &:= \arg\max_l \sum_{j\in\mathcal{J}_l} \frac{(\mathbf{a}_i^{\mathrm{T}}\mathbf{r}[k])^2}{\|\mathbf{a}_i\|_2^2}, \\ \Lambda &:= \Lambda \cup \mathcal{J}_L. \end{aligned} \tag{119}$$

CoSaMP [90]**:** Needell and Tropp developed a greedy algorithm for compressed sensing and proposed CoSaMP (Compressive Sampling Matching Pursuit) in [90]. As in StOMP and ROMP, CoSaMP also selects multiple indices at each step by picking up $2M$ largest components of $\mathbf{A}^{\mathrm{T}}\mathbf{r}[k]$ (cf. (117)), and merges the selected indices with the current support set. Then, CoSaMP solves the LS on the current support set as in OMP to obtain an approximation. A new step in CoSaMP is *pruning*; CoSaMP keeps the M largest entries of the LS approximation and prunes the others by setting them zero. As a result, the size of the support set is reduced to M. Algorithm 6 describes the procedure of CoSaMP. In this algorithm, $\mathcal{H}_M$ is the thresholding operator that sets all but the largest M elements of its argument to zero. Note that CoSaMP requires knowledge of the sparsity M of the solution. A similar algorithm called subspace pursuit has been proposed in [91].

IHT [92], [93]**:** Let us consider the following problem related to compressed sensing:

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \ \text{ subject to } \ \|\mathbf{x}\|_0 \leq M. \tag{120}$$

Iterative hard thresholding (IHT) [92], [93] is a greedy algorithm for this problem. IHT is very simple; the iteration rule is given as

$$\mathbf{x}[k+1] = \mathcal{H}_M\left(\mathbf{x}[k] + L^{-1}\mathbf{A}^{\mathrm{T}}(\mathbf{y} - \mathbf{A}\mathbf{x}[k])\right), \tag{121}$$

where $\mathcal{H}_M$ is the same thresholding operator as that used in the algorithm of CoSaMP. The operator $\mathcal{H}_M$ is also a nonlinear projection onto the M-sparse subset $\Sigma_M \subseteq \mathbb{R}^n$. It is worth noting that IHT looks similar to the gradient projection method (71) and ISTA (104). In fact, the iteration (121) can be rewritten using a quadratic model $\bar{Q}_L$, called a surrogate function [94], as

$$\begin{aligned} &\mathbf{x}[k+1] = \min_{\mathbf{x}} \bar{Q}_L(\mathbf{x}, \mathbf{x}[k]) \ \text{ subject to } \ \|\mathbf{x}\|_0 \leq M, \\ &\bar{Q}_L(\mathbf{x}, \mathbf{v}) = L^{-1}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 - L^{-1}\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{v}\|_2^2 + \|\mathbf{x} - \mathbf{v}\|_2^2. \end{aligned} \tag{122}$$

It is proved in [92] that if the parameter L satisfies (102), then IHT produces a sequence that converges to a local minimum of the optimization problem (120).

Algorithm 6 CoSaMP

Require: $\mathbf{y} \in \mathbb{R}^m$ and $M \in \mathbb{N}$ {observed vector and sparsity}
Ensure: $\mathbf{x} \in \mathbb{R}^n$ {estimated sparse vector}
  $\mathbf{x}[0] := \mathbf{0}$.
  $\mathbf{r}[0] := \mathbf{y} - \mathbf{A}\mathbf{x}[0] = \mathbf{y}$.
  $\Lambda := \emptyset$.
  $k := 0$.
  repeat
    $\Lambda := \Lambda \cup \mathrm{supp}\left(\mathcal{H}_{2M}(\mathbf{A}^{\mathrm{T}}\mathbf{r}[k])\right)$.
    $\mathbf{z} := \arg\min_{\mathbf{v}} \|\mathbf{A}_\Lambda \mathbf{v} - \mathbf{y}\|_2^2, \quad \mathbf{z}_{\Lambda^c} := \mathbf{0}$.
    $\mathbf{x}[k+1] := \mathcal{H}_M(\mathbf{z})$.
    $\Lambda := \mathrm{supp}(\mathbf{x}[k+1])$.
    $\mathbf{r}[k+1] := \mathbf{y} - \mathbf{A}\mathbf{x}[k+1] = \mathbf{y} - \mathbf{A}_\Lambda \mathbf{x}_\Lambda[k+1]$.
    $k := k + 1$.
  until $\|\mathbf{r}[k]\|_2 \leq \mathrm{EPS}$.
  return $\mathbf{x} := \mathbf{x}[k]$.

5.3 Alternative Ways towards Sparse Solution

Along with ℓ_1 optimization and greedy algorithms discussed above, there are a number of computational methods for solving the original compressed sensing problem (25). We here introduce some other approaches for compressed sensing.

One approach is to relax the ℓ_0-norm in (25) to the ℓ_p-norm with $p \in (0, 1)$:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_p^p \ \text{ subject to } \ \mathbf{A}\mathbf{x} = \mathbf{y}. \tag{123}$$

Clearly, this optimization is *nonconvex*. Although such a nonconvex optimization problem is much more difficult to solve than ℓ_1 optimization, the nonconvex optimization will recover sparse vectors with fewer measurements [95], [96]. Also, this approach may increase robustness against noise and lead to stability (i.e., small perturbation in the original vector $\mathbf{x}$ implies small estimation error) [97]. For numerical optimization of (123), one can adopt an iteratively-reweighted least squares [98] or an operator-splitting algorithm [99].

Another approach is to adopt the Bayesian framework; Bayesian techniques in machine learning have been adapted to compressed sensing problems. In the Bayesian framework, one assumes a prior distribution for the unknown vector that promotes sparsity [100], [101]. More recently, application of belief propagation has been proposed by assuming sparse sensing matrices [102]. The approximate message passing (AMP) [103] has also been inspired by belief propagation, but is applicable to dense sensing matrices,

provided that they are typical as realizations of some random matrix ensemble. AMP is also proven to achieve the Donoho-Tanner threshold for ℓ_1 optimization.

6. Applications

Recently, compressed sensing has been applied to various problems of communications systems. For the effective application of it, two key issues will be 1) the sparsity of signals of interest (Is it natural to assume the sparsity of them? In which domain?) and 2) the linear measurements of the signals (How can we obtain linear equations? Is it happy to reduce the number of them? How is the sensing matrix?). In this section, we introduce several examples of the problems of communications systems, where compressed sensing has been applied in a natural and beautiful manner, such as wireless channel estimation, network tomography, wireless sensor network, and cognitive radio. Applications to some other topics are also mentioned at the end of this section.

6.1 Wireless Channel Estimation

Channel estimation is one of the most important techniques in wireless communications systems, because a lot of modern communications technologies assume availability of channel state information. In current wireless communications systems, channel estimation is usually performed by sending some known signals, called pilot signals or training signals, before and/or during communications. Since these known signals do not convey user data but consume power and bandwidth, reduction of the required amount of training signals while keeping a sufficient estimation accuracy has been one of the main scopes of the study on channel estimation. Although it is true that several non-training assisted channel estimation (i.e., blind channel estimation) schemes have been proposed [104], the blind approach is not commonly used in practical communications systems so far because it usually requires high computational complexity.

In the context of training-based channel estimation, various properties of wireless channels in time, frequency, and space domains have been utilized to reduce the amount of training signals. The sparsity of channel impulse response is one of them. It is known that the impulse response of wireless channel tends to be sparse for larger bandwidth [105]–[110], although it also depends on whether the considered environment is scattering rich or not. Intuitively, this can be understood that the wideband signal reveals the actual response of the wireless channel, which has discrete nature consisting of multipath components, while it is smoothed out for narrowband communications systems. It is also known that underwater acoustic channel exhibits sparsity in both temporal (delay spread) and frequency (Doppler spread) domains [111], [112]. Therefore, several works have tried to utilize the sparsity for the channel estimation even before the birth of compressed sensing, such as [113]–[118].

Despite the fact that there are numerous preceding studies trying to exploit sparsity, it is also true that studies on sparse channel estimation have been much more accelerated recently. To cite some examples, compressed sensing approaches using OMP algorithm and the ℓ_1-ℓ_2 optimization are applied to channel estimation of multicarrier underwater acoustic communications systems and higher robustness against the Doppler effects is numerically demonstrated over the conventional schemes, such as the method of LS or subspace methods [119], [120]. Ultra-wideband channel estimation based on compressed sensing is proposed in [121] and it is shown that the proposed detector can outperform conventional correlator-based detector with only 1/3 of the sampling rate. Compressed sensing is applied for the estimation of doubly selective channels with the block transmission using cyclic prefix [122]–[124], and some sparsity-enhancing basis are proposed, while basis expansion models [125], [126] are commonly used for the estimation of time-varying channels. An optimization method of the pilot placement for sparse channel estimation in OFDM systems is proposed in [127] by using a modified version of a discrete stochastic approximation algorithm [128]. A sparse channel estimation technique using CoSaMP algorithm is applied for a two-way relay network in [129] and the improvement in mean square error (MSE) performance has been demonstrated at the cost of increased computational complexity. Excellent surveys on sparse channel estimation methods can be found in [130] and [131].

Here, we briefly review a simple approach to sparse channel estimation with a naive assumption that the channel impulse response itself is sparse in time domain (i.e., almost all taps have zero or close to zero values). However, it should be noted that, depending on wireless environments, this sparse tap model might be inappropriate and utilization of an appropriate overcomplete dictionary could significantly improve the estimation performance [121], [131]. Let $\mathbf{a} = [a_1, \ldots, a_P]^{\mathrm{T}} \in \mathbb{R}^P$ denote a vector of training signals for the channel estimation, which is inserted between data signals, and $\mathbf{x} = [x_1, \ldots, x_L]^{\mathrm{T}} \in \mathbb{R}^L$ be a vector of finite channel impulse response with $\|\mathbf{x}\|_0 \ll L$ as depicted in Fig. 3. Then, assuming $P > L$, the corresponding received signal vector $\mathbf{y} = [y_1, \ldots, y_{P-L+1}]^{\mathrm{T}}$, which is not contaminated by data signals, is written as

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \mathbf{v}, \tag{124}$$

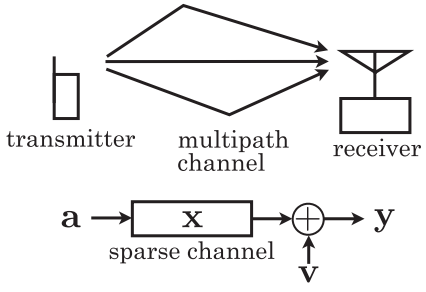

Fig. 3 Training-based wireless sparse channel estimation. The training signal vector **a** is sent through the sparse multipath channel having the impulse response of **x** and the additive white noise **v**.

where $\mathbf{X}$ is the Toeplitz channel matrix of size $(P-L+1)\times P$ defined as

$$\mathbf{X} = \begin{bmatrix} x_L & \dots & x_1 & 0 & \dots & 0 \\ 0 & x_L & & x_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \dots & 0 & x_L & \dots & x_1 \end{bmatrix}, \tag{125}$$

and where $\mathbf{v} = [v_1, \dots, v_{P-L+1}]^{\mathrm{T}}$ is an additive white noise vector. By using the channel impulse response vector $\mathbf{x}$, (124) can alternatively be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \tag{126}$$

where $\mathbf{A}$ is a Toeplitz matrix of size $(P - L + 1) \times L$ defined as

$$\mathbf{A} = \begin{bmatrix} a_L & a_{L-1} & \dots & a_1 \\ a_{L+1} & a_L & \dots & a_2 \\ \vdots & \vdots & & \vdots \\ a_P & a_{P-1} & \dots & a_{P-L+1} \end{bmatrix}. \tag{127}$$

Thus, by regarding $\mathbf{A}$ to be a sensing matrix and assuming $P - L + 1 < L$ in order to achieve higher spectral efficiency, the problem to obtain $\mathbf{x}$ from $\mathbf{y}$ in (126) can be considered as a problem of compressed sensing in (23) but with observation noise. As for the recoverability of the sparse vector for the case of random Toeplitz sensing matrix, it is shown in [132] that random Toeplitz matrices satisfy RIP under certain conditions. For example, if the entries a_i, $i = 1, \dots, P$, in $\mathbf{A}$ are i.i.d. and are equal to $\pm 1/\sqrt{P - L + 1}$ with probability $1/2$, then $\mathbf{A}$ satisfies RIP of order k with the RIP constant δ_k with probability at least $1 - \exp(-\frac{c_1(P-L+1)}{k^2})$, provided that $P - L + 1 \geq c_2 k^2 \log L$, where $c_1,\ c_2 > 0$ are functions of δ_k.

We have so far discussed application of compressed sensing to training-based channel estimation schemes for the purpose of reducing the required training signals, but blind channel estimation schemes can also benefit from compressed sensing in a different manner. Subspace method [133] is one of the most popular blind channel identification schemes, since it enables us to estimate channel impulse response up to a complex multiplicative constant, which is inherent to the problem, from the second-order statistics of the received signal. However, it also has a drawback that the length of the unknown channel impulse response has to be exactly known a priori: If otherwise, even an overestimate of the length results in breakdown of the estimation, because of the following reason. The subspace method utilizes the linear equations derived from the orthogonality between the signal subspace and the noise subspace of the sample correlation matrix of the received signal. Thus, to obtain an estimate of channel, and, in order to make the linear equations well-posed, the exact information of the length of the unknown channel response is required. If the received signal-to-noise ratio (SNR) is high, then we can easily estimate the length from the number of large eigenvalues of the correlation matrix. However, if the received SNR is not high enough, which is often the case in common wireless environments, it is usually difficult to tell the threshold of the signal and noise subspaces. In order to cope with the problem, compressed sensing approach has been introduced to the subspace method assuming that the channel impulse response is sparse, and has numerically been shown that the blind channel estimation is possible without exact information of the length (i.e., overestimate works), if around half of the channel taps are zero [134].

Note that the impulse response of wireless channel is usually modeled by a complex vector, whereas we have discussed the estimation of the real sparse vector in this paper. As for the estimation algorithm of the complex sparse vector, readers are referred to, say, [74].

6.2 Wireless Sensor Network

A typical problem setting of wireless sensor networks is as follows: Some physical phenomenon in the area of interest is measured by a lot of sensor nodes with communication capability. The data sensed by the nodes are then either sent to a central node, which performs signal processing to extract the desired information in the area and is referred to as a fusion center or a sink node, or shared by the sensor nodes, which perform distributed processing in the network. In such a problem, we can oftentimes reasonably assume some spatial and/or temporal correlation of the sensed data, which leads to the motivation to apply compressed sensing to the sensor network. Bajwa et al. [135] have been the first to introduce compressed sensing to the wireless sensor network, and since then, there have been several works on the topic. In this section, we introduce basic ideas of compressed sensing based wireless sensor networks for some network scenarios. The interested readers are referred to a good tutorial on the decentralized compression of *networked data* [136], which is written by the same authors as [135].

The first scenario considered in this section is depicted in Fig. 4, where a number of sensor nodes periodically measure the physical environment and directly send the obtained

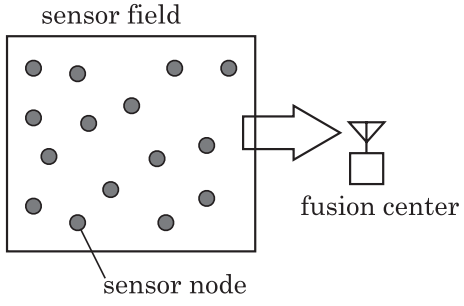

Fig. 4 Wireless sensor network with a fusion center [135]. The fusion center is assumed to be within the communication range from all sensor nodes, and sensed data at each sensor node are directly sent to the fusion center. Because of the linearity of the wireless medium, the received signal at the fusion center is a superposition of the transmitted signals from all the sensor nodes, which leads to a linear measurement in the framework of compressed sensing.

data to the fusion center. Let n and x_j, $j = 1, \ldots, n$, denote the number of sensor nodes and the measured data at the j-th sensor node, respectively. In the basic approach shown in [136], each sensor node sends its own sensed data x_j to the fusion center using m time slots ($m < n$) with m random coefficients $\{A_{i,j}\}_{i=1}^{m}$, which are locally generated by using its own node identity (ID) as the seed of a pseudo-random number generator. Since n sensor nodes are assumed to send $A_{i,j}x_j$, $j = 1, \ldots, n$, in the i-th time slot simultaneously, the received signal at the fusion center in the i-th time slot is given by

$$y_i = \sum_{j=1}^{n} A_{i,j}x_j + v_i. \tag{128}$$

Here, ignoring the impact of channel gain, we have assumed for simplicity the additive white Gaussian noise (AWGN) channel for the link between each sensor node and the fusion center, and v_i is the AWGN in the i-th time slot. It should be noted that the fusion center obtains a linear measurement of $\mathbf{s}$ thanks to the additive nature of the radio waves. At the end of the m-th time slot, the fusion center obtains a received signal vector $\mathbf{y} = [y_1, \ldots, y_m]^{\mathrm{T}}$ as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \tag{129}$$

where $\mathbf{A}$ is the $m \times n$ matrix whose (i, j)-element is $A_{i,j}$, where $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$ is the vector of sensed data, and where $\mathbf{v} = [v_1, \ldots, v_m]^{\mathrm{T}}$.

In most applications of wireless sensor networks, the sensed data vector $\mathbf{x}$ itself will not be sparse. However, because of the spatial correlation of the physical phenomenon, like temperature, pressure, or radio activity, we can usually assume that $\mathbf{x}$ is sparse in some transform domain. We specifically assume that $\mathbf{x}$ is represented by an appropriate $n \times n$ invertible transformation matrix $\mathbf{\Phi}$, as

$$\mathbf{x} = \mathbf{\Phi}\mathbf{c}, \tag{130}$$

where $\mathbf{c}$ is a sparse vector. Note that the typical choice of the transformation could be discrete Fourier transform (DFT), discrete Cosine transform (DCT), or wavelet transform, and the fusion center has to know $\mathbf{\Phi}$.

Under the above sparsity assumption, the received signal vector $\mathbf{y}$ is represented as

$$\mathbf{y} = \mathbf{A}\mathbf{\Phi}\mathbf{c} + \mathbf{v}. \tag{131}$$

Therefore, the task to estimate a sparse vector $\mathbf{c}$ from $\mathbf{y}$ will be a standard compressed sensing problem as in (24), where $\mathbf{A}\mathbf{\Phi}$ is regarded as a sensing matrix ($\mathbf{A}$ is available at the fusion center if it knows all IDs of the sensor nodes). Since the reduction of the number m of transmissions is of crucial importance to many wireless sensor networks in order to decrease the power consumption at sensor nodes, the approach of compressed sensing will be very suited for the sensor network problems. Although we have assumed the AWGN channel for the link between each sensor node and the fusion center and utilized artificially generated random coefficients, we can also utilize fading phenomena as the random coefficient generator for the sensing matrix by assuming multiple fusion centers. A similar approach can be found in [137].

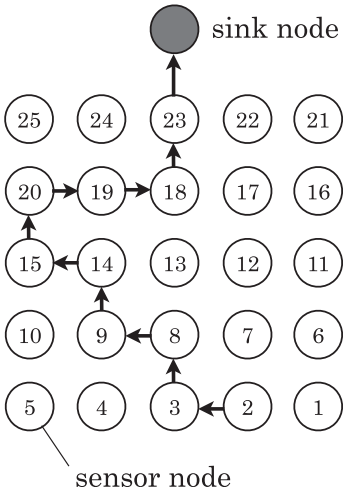

Fig. 5 Multi-hop wireless sensor network with a sink node [140]. Communication range of each sensor node is assumed to be limited, thus multi-hop routing is required to reach the sink node. Each sensor node on the route to the sink node relays the received data after adding its own measurement data multiplied by a locally generated random number.

On the other hand, if the sensing area is very large or the transmit power of the sensor node is strictly limited, we have to resort to multi-hop communication to send the sensed data from the sensor node to the fusion center (sink node) as depicted in Fig. 5. Here, we explain a typical approach to obtain linear measurements at the sink node [138]–[141]. The data gathering is performed by temporally separated m multi-hop communications with randomly determined initial (starting) nodes. For each multi-hop communication, a randomly selected starting node (in Fig. 5, the node "2" is the starting node) computes the product of its own measured data and a locally generated random coefficient, determines one of the neighboring nodes using a routing algorithm employed in the sensor network, and then sends the computed product to the determined node. At each relaying sensor node involved in the multi-hop communication, it adds the product of its own measured data and a locally generated random coefficient to the received signal, and forwards the result to the next sensor node. Thus, the received signal at the sink node for the i-th multi-hop communication is given by

$$y_i = \sum_{j \in \mathcal{P}_i} A_{i,j}x_j, \tag{132}$$

where $\mathcal{P}_i$ and $A_{i,j}$ are the set of indexes of the nodes in the i-th multi-hop communication and the random coefficient generated at the j-th node, respectively, and where x_j is the measured data at the j-th node. Note that $A_{i,j}$ takes a different value for different i or j in general. Since we also assume the spatial correlation model in (130), the received signal vector $\mathbf{y} = [y_1, \ldots, y_m]^{\mathrm{T}}$ obtained at the sink node is

given by

$$\mathbf{y} = \mathbf{A}\mathbf{\Phi}\mathbf{c}, \tag{133}$$

where $\mathbf{A}$ is an $m \times n$ matrix with the (i, j)-element equal to $A_{i,j}$ if $j \in \mathcal{P}_i$, and zero otherwise. Although this looks a standard compressed sensing problem as in the previous case, it should be noted that $\mathbf{A}$ is determined by the routing algorithm used in the sensor network and that some common routing algorithms result in high coherence of the sensing matrix. Actually, it has been reported that the existing algorithms such as the shortest path routing algorithm or the greedy routing algorithm cannot achieve so much performance gain compared with the conventional approach in terms of the reconstruction error versus the total number of transmissions in the network [138], [139]. Therefore, some random routing protocols have been proposed to achieve the requirement of incoherence [140], [141]. Also, the combined use of the linear random network coding with compressed sensing based sensor network has been investigated in [142] and [143].

In the methods described above in this section, an ideal media access control (MAC) protocol, in terms of synchronization, channel state information, or collisions, is commonly assumed. However, especially in typical sensor network scenarios, such an ideal MAC might be difficult to achieve because of the requirements of low-cost, limited bandwidth and limited battery power on sensor nodes. Thus, there are some works on the joint design of data gathering and MAC protocol for compressed sensing based sensor network such as [144] and [145].

6.3 Network Tomography

Network tomography is a term coined by Vardi [146] upon the similarity between the network inference and medical tomography. Two major forms of network tomography are link-level parameter estimation from end-to-end measurements and end-to-end traffic intensity estimation based on link-level measurements [147]. Since the original paper of compressed sensing [3] was largely motivated by the reconstruction problem of magnetic resonance imaging (MRI), which is one of the medical tomographies [148], it is quite natural to consider application of compressed sensing to network tomography [149]–[152].

Network tomography of the link-level parameter estimation is commonly performed to estimate link delay or link loss rate in order to detect failures of links or nodes inside the network. Since only a limited number of links have large delays or high loss rates in typical networks, if we consider a vector composed by delays or loss rates of all links, we can reasonably assume that the vector will be sparse or compressible. In network tomography, as shown in Fig. 6, end-to-end measurements are performed by sending probe packets from source nodes to receiver nodes through the network[†]. An end-to-end measurement from a source node to a receiver node acquires information about a route from the source node to the receiver node. In graph theory

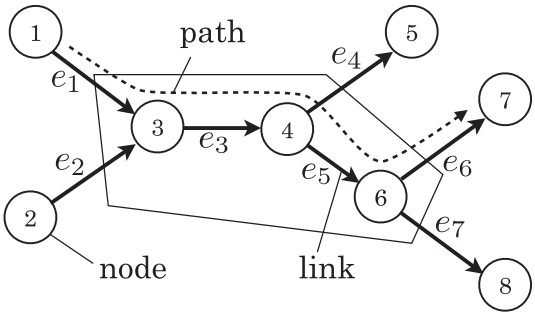

Fig. 6 Network tomography with 8 nodes and 7 links [152]. Each node stands for a terminal or a router, and each edge between two nodes is called a link. A path is a connection between two nodes composed by multiple links in general. Nodes 1 and 2 are source nodes, 5, 7, and 8 are receiver nodes, and 3, 4, and 6 are internal nodes.

terminology, such a route is called a path, which consists of multiple links connected in sequence in general. The number of measurements in network tomography is thus equal to the number of paths considered. Although the selection of paths has a large impact on the inference performance as in the case of multi-hop sensor network, we assume that the paths are given in advance for the moment.

Network tomography to estimate link delays is sometimes called *delay tomography*. A formulation on the basis of linear measurements can be established straightforwardly for delay tomography, since the overall delay of a path is a sum of the delays of all the links belonging to the path. Consider a network with n links, and assume that the number of measurements is m. Let x_j denote the delay of link e_j, $j = 1, \ldots, n$. Then the overall delay of the i-th path, $i = 1, \ldots, m$, is given by

$$y_i = \sum_{j \in \mathcal{P}_i} A_{i,j} x_j, \tag{134}$$

where $\mathcal{P}_i$ is the set of indexes of links in the i-th path, and where $A_{i,j} = 1$ if $j \in \mathcal{P}_i$ and $A_{i,j} = 0$ otherwise. Thus, by defining $\mathbf{y} = [y_1, \ldots, y_m]^{\mathrm{T}}$, $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$, and $\{\mathbf{A}\}_{i,j} = A_{i,j}$, we have

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{135}$$

which is the problem of compressed sensing with binary sensing matrix, if we assume that $\mathbf{x}$ is sparse or compressible, and that link delays in the network are stationary.

Network tomography in which link loss rates are to be estimated is sometimes called *loss tomography*. One can establish a formulation for loss tomography just as in a similar manner to delay tomography described above, if we assume that packets are lost independently on each link. Let p_j denote the loss rate at link e_j, $j = 1, \ldots, n$. Although loss rates themselves are not additive, under the independence assumption, the link success rates $(1 - p_j)$ are multiplicative, i.e., the overall packet success rate of the i-th path,

[†]This type of network tomography is called active measurement, while the inference performed using existing packets in the network is called passive measurement.

$i = 1, \ldots, m$, is given by

$$1 - q_i = \prod_{j \in \mathcal{P}_i} (1 - p_j), \tag{136}$$

where q_i is the overall packet loss rate of the i-th path. By defining $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$ and $\mathbf{y} = [y_1, \ldots, y_m]^{\mathrm{T}}$ with

$$x_j = -\log(1 - p_j), \tag{137}$$

$$y_i = -\log(1 - q_i) = -\sum_{j \in \mathcal{P}_i} x_j, \tag{138}$$

we have an expression of linear measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{139}$$

where $\mathbf{A}$ is the same matrix (routing matrix) as in (135). Note that if $\mathbf{p} = [p_1, \ldots, p_n]^{\mathrm{T}}$ is sparse or compressible, then $\mathbf{x}$ is sparse or compressible as well, because $p_j = 0$ or close to 0 means $x_j = 0$ or close to 0. Also note that it will not be possible to obtain the value of q_i (and hence y_i) from the transmission of a single probe packet, unlike the case of delay estimation. Thus, multiple probe packets are transmitted for each path, and the ratio of the number of received packets to the total number of transmitted packets is used as an estimate of $(1 - q_i)$.

In both cases, the problems come down to compressed sensing with the binary sensing matrix, which is determined by paths (routes) from source nodes to receiver nodes. Since the increase in the number of paths, which corresponds to the number of measurements, results in the increase in the number of probe packets injected into the network, it is desirable to minimize the number of paths in order not to give unnecessary load to the network. As for the number of required measurements for the reconstruction of k-sparse vectors with random binary measurements matrices, it has been known that $O(k \log \frac{n}{k})$ measurements are required [153], [154], while it has been shown that $O(k \log n)$ measurements are needed if the binary matrix is accompanied by a graph constraint [151], [155]. Moreover, a deterministic guarantee and a designed method of the routing matrix for the reconstruction of any 1-sparse signal are provided in [149], [150], taking advantage of the knowledge on compressed sensing using expander graphs [154], [156].

6.4 Cognitive Radio

Cognitive radio has been one of the typical applications of compressed sensing from its early stage. In cognitive radio networks, unlicensed cognitive radio users (secondary users) are supposed to utilize licensed frequency bands without causing harmful interference to the users in the licensed systems (primary users) taking advantage of the temporal vacancy of the primary users as depicted in Fig. 7.

As such, it is necessary for the secondary users (or cognitive radio networks) to sense the radio-spectrum environment in order to find vacant frequency bands prior to communications.

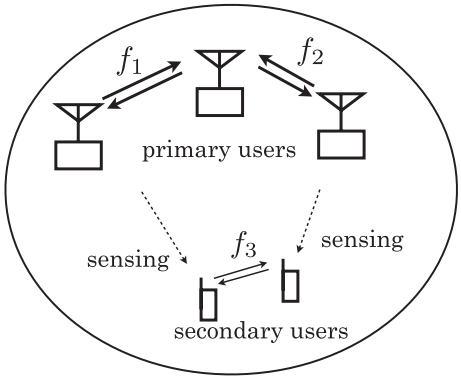

Fig. 7 A simple example of cognitive radio system, where the licensed frequency bands for the primary system are assumed to be f_1, f_2 and f_3. Since f_3 is not occupied by the primary users, the secondary users can utilize f_3 for their communications without causing any interference to the primary users, although they are not licensed users of the band.

The spectrum sensing can be a challenging task if it has to be performed over a wide frequency band, because the sampling rate to meet the Nyquist-Shannon sampling theorem can be prohibitively high. A key observation to alleviate this problem is that the occupancy of licensed bands by primary users is typically rather low [157]. While taking advantages of the sparsity of the spectrum in use at any instance of time has been the original motivation of cognitive radio, it also allows us to apply compressed sensing approach to the wide-band spectrum sensing problem, which is called *compressed spectrum sensing*, to reduce the required sampling rate [158]–[160]. Moreover, the compressed spectrum sensing has been extended to the cooperative spectrum sensing, where observations or inferences of spectrum sensing at multiple secondary users are combined in order to achieve robustness against observation noise and/or fading effects [161]–[166]. Furthermore, in order to cope with the time variant nature of the spectrum environment, compressed spectrum sensing schemes with dynamic sampling rate adjustment have been proposed [167]–[171], because the required sampling rate depends on the sparsity order of the spectrum.

In this subsection, we introduce a basic approach of the compressive spectrum sensing at a single cognitive terminal [158], [159]. A particular feature with the spectrum sensing is that, unlike other applications explained so far, analog signals have to be explicitly considered. Let $x(t)$, $t \in [0, nT_{\mathrm{s}}]$, denote a received analog wide-band signal at the terminal considered, where T_{s} is the sampling period equal to the inverse of the Nyquist rate. Thus, n samples of $x(t)$ obtained by uniform sampling with the period T_{s} are required to recover $x(t)$ without aliasing. By letting $\mathbf{x} = [x(T_{\mathrm{s}}), \ldots, x(nT_{\mathrm{s}})]^{\mathrm{T}}$ denote a vector composed of the full sample set of $x(t)$ with the Nyquist rate, the linear measurement (sampling) process in the discrete-time domain can be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{140}$$

where $\mathbf{A}$ denotes an $m \times n$ measurement matrix. Note that $m < n$ corresponds to sub-Nyquist rate sampling, and a simple choice of $\mathbf{A}$ is a selection matrix obtained by randomly retaining m rows of an $n \times n$ identity matrix. The discrete spectrum of $x(t)$ can be calculated as

$$\mathbf{c} = \mathbf{D}\mathbf{x}, \tag{141}$$

where $\mathbf{D}$ denotes the unitary DFT matrix whose (i, j)-element is given by $\frac{1}{\sqrt{n}} e^{-\sqrt{-1}\frac{2\pi}{n}ij}$, $i, j = 0, \ldots, n-1$. Thus, if the primary signal occupies the frequency band sparsely enough, we have a formulation of standard compressed sensing as

$$\mathbf{y} = \mathbf{A}\mathbf{D}^{\mathrm{H}}\mathbf{c}, \tag{142}$$

where $\mathbf{A}\mathbf{D}^{\mathrm{H}}$ is regarded as the sensing matrix and $\mathbf{c}$ is an unknown sparse vector.

In [158], not only the sparsity of the primary signal itself in the frequency domain, but also the sparsity of subband edges is utilized, where it is supposed that $x(t)$ is in the frequency range $[f_0, f_k]$ and k consecutive subbands are in the range with the boundaries of $f_0 < f_1 < \cdots < f_k$ as depicted in Fig. 8. The edge spectrum $\mathbf{z}_s$, which is defined as the derivative wavelet of $\mathbf{c}$ at scale s, is calculated by using the discrete-time signal vector $\mathbf{x}$ as

$$\mathbf{z}_s = \mathbf{\Gamma}\mathbf{D}\mathbf{\Phi}_s\mathbf{x}, \tag{143}$$

where $\mathbf{\Gamma}$ is the differentiation matrix given by

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & 0 & \ldots & \ldots & 0 \\ -1 & 1 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & -1 & 1 \end{bmatrix}, \tag{144}$$

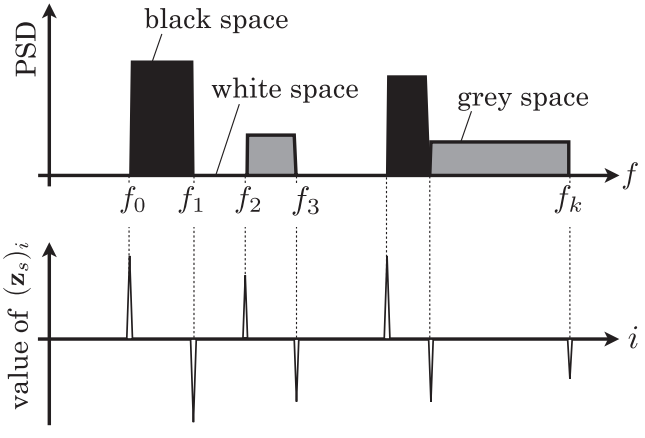

Fig. 8 Upper figure: k frequency subbands with piecewise smooth power spectrum density (PSD) [158]. Black, grey, and white spectrum spaces [175] correspond to the PSD levels of high, medium, and low, respectively. White space, as well as grey space in some cases, can be used for communications of secondary users. Lower figure: The value of each element $(\mathbf{z}_s)_i$ corresponding to the PSD in the upper figure is depicted as an intuitive illustration.

and where $\mathbf{\Phi}_s$ is a matrix representation of a wavelet smoothing function with a scale factor s. For an intuitive understanding, an example of $\mathbf{z}_s$ is depicted in Fig. 8. By assuming the piecewise smooth power spectrum density (PSD) as shown in Fig. 8, $\mathbf{z}_s$ has sparsity, and thus the linear equations

$$\mathbf{y} = \mathbf{A}(\mathbf{\Gamma}\mathbf{D}\mathbf{\Phi}_s)^{-1}\mathbf{z}_s \tag{145}$$

give another formulation of compressed sensing.

It should be noted that, while we have treated $x(t)$ as a deterministic signal in the discussion above, the received signal should commonly be treated as a random process. Therefore, the spectrum should be evaluated by using the autocorrelation function of $x(t)$, which requires calculations directly using $\mathbf{x}$ obtained by Nyquist rate sampling. This is of course not acceptable, since it spoils the advantages of the compressed spectrum sensing with sub-Nyquist rate. The problem could be settled, for example, with the approach proposed in [165], where one utilizes *analog-to-information converter (AIC)* [172], [173], which acquires samples with the rate corresponding to the information rate instead of the Nyquist rate by using a random demodulator. The framework of *Xampling* [174] assumes a signal space described not as a linear space but as a union of subspaces, and proposes a generic construction of signal-processing systems for such signals using a conventional analog-to-digital (AD) converter, by putting a carefully-designed analog preprocessor in front of the AD converter in order to compress the signal bandwidth, as well as a compressed-sensing-based nonlinear subspace detector after the AD converter. AIC can be regarded as one of the realization methods of the signal acquisition part of Xampling with hardware devices and a low-rate AD converter.

6.5 Some Other Topics

In this subsection, we briefly introduce applications of compressed sensing to some other topics in communications systems.

(1) Array Signal Processing

Direction-of-arrival (DOA) estimation of incoming waves is one of the major problems in array signal processing. Compressed sensing has been applied to the problem taking advantage of the sparsity of incoming signals in the angular domain. In [176], reduction of sampling rate at each antenna element except one reference element is achieved by the introduction of compressed sensing. Moreover, in [177], inspired by the work on the relation between compressed sensing and array signal processing in [178], [179], the problem of DOA estimation is formulated as a multiple measurement vector (MMV) problem [180]–[183], which considers the recovery of a set of sparse vectors sharing a common nonzero support on the basis of multiple measurement vectors (i.e., snapshots, in array signal processing terminology), and an algorithm named *compressive MUSIC*, which is regarded as an extension of the conventional multiple signal

classification (MUSIC) algorithm [184], has been proposed to reduce the number of required snapshots.

As for other applications to array signal processing, compressed sensing has been applied to the problem of synthesizing a desired far-field beam-pattern by controlling sensors' positions and array weights with the minimum number of sensors, which is called the maximally sparse array [185]–[188]. Also, it has been applied to the diagnosis of antenna arrays [189] and some radar applications [190]–[192], assuming sparsities of failure antenna elements and reflectivity functions of targets, respectively.

(2) Multiple Access Scheme

Compressed sensing has also been applied to problems of medium access control (MAC) layer, such as multiple access schemes [193]–[195], downlink scheduling [196], and resource allocation for feedback channels [197]. A typical approach to the application to multiple access schemes is similar to the conventional code-division multiple-access (CDMA) scheme, but with the signature codes shorter than the number of users, which can be regarded as an overloaded CDMA. A key strategy here is that, although the number n of potential users might be large, the number k of active users at one time can be typically far smaller than n and even smaller than the length m of the signature codes. Thus, taking advantage of the sparsity of active users, one can extract signals of active users by using algorithms of compressed sensing, where each column vector of the sensing matrix corresponds to the signature code of each user.

(3) Networked Control

A relatively novel application of compressed sensing is *networked control* [198]–[200]. Networked control has recently attracted a lot of attention in both communications and control systems communities. In networked control, a controller is placed away from a controlled plant and the controller should communicate with the plant over rate-limited networks, such as wireless networks. In this situation, the data should be compressed to satisfy the rate-limiting constraint. Also, we cannot use an intelligent coder/encoder such as a vector quantizer since computational delay may degrade the control stability and performance. To tackle with these problems, sparsification of control vectors to be transmitted has been proposed for predictive control systems [201], [202] and for remote control systems [203], [204] based on the notion of compressed sensing. In these studies, FISTA (see Sect. 5.1.4) or OMP (see Sect. 5.2.3) are used and proved to be effective in feedback control since the algorithms are extremely fast. Stability issues of the feedback control system are partly solved in [201], [202], but a general theorem is still open.

7. Further Studies

Since we have introduced compressed sensing from a view point of users, we have provided some theorems without proofs. Readers who are interested in more theoretical aspects are referred to the excellent books [6], [7], [9], [12], [18], as well as the original papers [1]–[3]. For applications of compressed sensing to other than communications systems, we would like to mention a special issue on applications of sparse representation and compressive sensing in the Proceedings of the IEEE [205]. Also, excellent surveys and tutorials are available in the special issue on compressive sampling in the IEEE Signal Processing Magazine [206]. Up-to-date information can be obtained through the Internet [207]–[209].

As for a recent topic on compressed sensing, an approach called *blind compressed sensing* has been recently proposed [210]. In this survey, we have assumed that the sensing matrix $\mathbf{A}$ (more precisely, the product of the sensing matrix $\mathbf{A}$ and the sparsity basis $\mathbf{\Phi}$) is known a priori by the reconstruction algorithm, but this requirement might not be met in some situations. Thus, in blind compressed sensing, the sparse signal recovery is performed without the prior knowledge of the sparsity basis but with some additional constraint, where the elements of both standard compressed sensing and dictionary learning [211]–[214], whose purpose is to find a sparsity basis for a given set of data, are combined. Such an approach might be appreciated in some applications of communications systems.

8. Conclusion

We have explained basic ideas of compressed sensing in this survey, thinking much of methodological aspects rather than theoretical ones, and assuming readers to be potential users of compressed sensing in the field of communications. We have started our discussions from the review of ill-posed linear simultaneous equations, and then, the problem of compressed sensing is described as the underdetermined linear system with a prior knowledge of the unknown vector being sparse. The ℓ_1 optimization approach and its variants are introduced as the convex relaxation of the direct ℓ_0 optimization, and some important properties and known results on the sensing matrix regarding the guarantee of the sparse signal recovery are also briefly explained. Moreover, as a guide for users of compressed sensing, several existing algorithms to solve the problem of compressed sensing are explained in detail focusing on two major approaches, namely, the ℓ_1 optimization approach and the greedy approach. Furthermore, we have introduced various examples of applications of compressed sensing to various problems of communications systems in physical (PHY), MAC and network layers.

Compressed sensing is a new paradigm to extract information taking advantage of the sparsity nature, which is possessed by various signals and systems. Since the band-limited nature can be regarded as a special case of the sparsity in the frequency domain, the framework of compressed sensing is the generalization of the conventional sampling theorem. In this regard, compressed sensing will be one of the common tools in this field in the near future.

While we have introduced several examples of applications, we believe that compressed sensing is still in its early

stage as to the application to communications systems and a lot of potential applications are yet to come. For example, not only the application of compressed sensing to classical problems but also to the design of whole communications systems from PHY to even application layers might open up new vistas in the field. It is our great pleasure if this survey could motivate readers to apply compressed sensing to their own research topics.

Acknowledgments

This work was supported in part by the Grant-in-Aids for Scientific Research No. 24560457 and No. 24560543 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan. The authors would like to thank Dr. Takahiro Matsuda, Osaka University, Japan, and Dr. Megumi Kaneko, Kyoto University, Japan, for their valuable information on applications of compressed sensing.

References

[1] D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol.52, no.4, pp.1289–1306, April 2006.
[2] E.J. Candes and T. Tao, "Decoding by linear programming," IEEE Trans. Inf. Theory, vol.51, no.12, pp.4203–4215, Dec. 2005.
[3] E.J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory, vol.52, no.2, pp.489–509, Feb. 2006.
[4] D.L. Donoho and B. Logan, "Signal recovery and the large sieve," SIAM J. Appl. Math., vol.52, no.2, pp.577–591, April 1992.
[5] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Trans. Signal Process., vol.41, no.12, pp.3397–3415, Dec. 1993.
[6] S. Mallat, A Wavelet Tour of Signal Processing, Third ed. The Sparse Way, Academic Press, 2008.
[7] J.L. Starck, F. Murtagh, and J.M. Fadili, Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity, Cambridge University Press, 2010.
[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," J.R. Statist. Soc. B, vol.58, no.1, pp.267–288, 1996.
[9] P. Buhlmann and S. van de Geer, Statistics for High-Dimensional Data: Methods, Theory and Applications, Springer, 2011.
[10] C.E. Shannon, "Communication in the presence of noise," Proc. IRE, vol.37, no.1, pp.10–21, Jan. 1949.
[11] E.J. Candès, "The restricted isometry property and its implications for compressed sensing," C.R. Math. Acad. Sci. Paris, vol.346, no.9, pp.589–592, May 2008.
[12] Y.C. Eldar and G. Kutyniok, Compressed Sensing, Cambridge University Press, 2012.
[13] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," The Annals of Statistics, vol.32, no.2, pp.407–499, April 2004.
[14] D.L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," IEEE Trans. Inf. Theory, vol.54, no.11, pp.4789–4812, Nov. 2008.
[15] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," Appl. Comput. Harmon. Anal., vol.26, no.3, pp.395–407, 2009.
[16] T.T. Cai, L. Wang, and G. Xu, "Shifting inequality and recovery of sparse signals," IEEE Trans. Signal Process., vol.58, no.3, pp.1300–1308, 2009.
[17] T.T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," IEEE Trans. Inf. Theory, vol.56, no.9, pp.4388–4394, 2010.
[18] M. Elad, Sparse and Redundant Representations, Springer, 2010.
[19] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," J. Amer. Math. Soc., vol.22, no.1, pp.211–231, Jan. 2009.
[20] D.L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," Proc. Natl. Acad. Sci. USA, vol.102, no.27, pp.9446–9451, July 2005.
[21] D.L. Donoho and J. Tanner, "Neighborliness of randomly projected simplices in high dimensions," Proc. Natl. Acad. Sci. USA, vol.102, no.27, pp.9452–9457, July 2005.
[22] R. Schneider and W. Weil, Stochastic and Integral Geometry, Springer-Verlag, Berlin, Heidelberg, 2008.
[23] D.L. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," J. Amer. Math. Soc., vol.22, no.1, pp.1–53, Jan. 2009.
[24] D.L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization," Proc. Natl. Acad. Sci. USA, vol.100, no.5, pp.2197–2202, March 2003.
[25] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Trans. Inf. Theory, vol.53, no.12, pp.4655–4666, Dec. 2007.
[26] D.L. Donoho, M. Elad, and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. Inf. Theory, vol.52, no.1, pp.6–18, Jan. 2006.
[27] Z. Ben-Haim, Y.C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," IEEE Trans. Signal Process., vol.58, no.10, pp.5030–5043, Oct. 2010.
[28] H. Rauhut, Compressive Sensing and Structured Random Matrices: Theoretical Foundations and Numerical Methods for Sparse Recovery Volume 9 of Radon Series Comp. Appl. Math., de Gruyter, 2010.
[29] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," Constr. Approx., vol.28, no.3, pp.253–263, Dec. 2008.
[30] Y. Kabashima, T. Wadayama, and T. Tanaka, "A typical reconstruction limit for compressed sensing based on l_p-norm minimization," J. Stat. Mech.: Theory and Experiment, vol.2009, pp.L09003–1–12, Sept. 2009. Erratum: *ibid.* vol.2012, E07001, 2012.
[31] Y. Kabashima, T. Wadayama, and T. Tanaka, "Statistical mechanical analysis of a typical reconstruction limit of compressed sensing," Proc. 2010 IEEE Int. Symp. Inf. Theory, pp.1533–1537, 2010.
[32] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing — Probabilistic analysis of a null-space characterization," Proc. IEEE ICASSP 2008, pp.3377–3380, 2008.
[33] M. Stojnic, "A simple performance analysis of ℓ_1 optimization in compresed sensing," Proc. IEEE ICASSP2009, pp.3021–3024, 2009.
[34] M. Stojnic, "ℓ_1 optimization and its various thresholds in compressed sensing," Proc. IEEE ICASSP2010, pp.3910–3913, 2010.
[35] M. Amin Khajehnejad, W. Xu, A. Salman Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," Proc. 2009 IEEE Int. Symp. Info. Theory, pp.483–487, 2009.
[36] D.L. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," Phil. Trans. R. Soc. A, vol.367, no.1906, pp.4273–4293, Nov. 2009.
[37] D.L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," Discrete Comput. Geom., vol.43, no.3, pp.522–541, 2010.
[38] T. Tanaka and J. Raymond, "Optimal incorporation of sparsity information by weighted ℓ_1 optimization," Proc. 2010 IEEE Int.

Symp. Info. Theory, pp.1598–1602, 2010.
[39] K. Takeda and Y. Kabashima, "Statistical mechanical assessment of a reconstruction limit of compressed sensing: Toward theoretical analysis of correlated signals," Europhysics Letters, vol.95, no.1, pp.18006–1–6, July 2011.
[40] M. Vehkapera, S. Chatterjee, and M. Skoglund, "Analysis of MMSE estimation for compressive sensing of block sparse signals," Proc. 2011 IEEE Info. Theory Workshop, pp.553–557, 2011.
[41] R. Matsushita and T. Tanaka, "Critical compression ratio of iterative reweighted ℓ_1 minimization for compressed sensing," Proc. 2011 IEEE Info. Theory Workshop, pp.568–572, 2011.
[42] S. Rangan, A.K. Fletcher, and V.K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," IEEE Trans. Inf. Theory, vol.58, no.3, pp.1902–1923, March 2012.
[43] B. Natarajan, "Sparse approximate solutions to linear systems," SIAM J. Comput., vol.24, pp.227–234, April 1995.
[44] J.A. Tropp and S.J. Wright, "Computational methods for sparse solution of linear inverse problems," Proc. IEEE, vol.98, no.6, pp.948–958, June 2010.
[45] http://www.scilab.org/
[46] http://www.msys.sys.i.kyoto-u.ac.jp/csdemo.html
[47] E. Candes and J. Romberg, "ℓ_1-MAGIC: Recovery of sparse signals via convex programming," Techical Report, Caltech, Oct. 2005.
[48] S. Becker, J. Bobin, and E. Candes, "NESTA: A fast and accurate first-order method for sparse recovery," SIAM J. Imaging Sci., vol.4, pp.1–39, 2011.
[49] E.V.D. Berg and M.P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," SIAM J. Scientific Computing, vol.31, no.2, pp.890–912, 2008.
[50] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Science, vol.2, no.1, pp.183–202, 2009.
[51] T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1-minimization: Methodology and convergence," SIAM J. Optimization, vol.19, pp.1052–6234, 2008.
[52] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1 minimization with applications to compressed sensing," SIAM J. Imaging Science, vol.1, no.1, pp.143–168, March 2008.
[53] M.A.T. Figueiredo, R.D. Nowak, and S.J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," IEEE J. Sel. Top. Signal Process., vol.1, no.4, pp.586–597, Dec. 2007.
[54] J.M. Bioucas-Dias and M.A.T. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process., vol.16, no.12, pp.2992–3004, Dec. 2007.
[55] S.J. Wright, R.D. Nowak, and M.A.T. Figueiredo, "Sparse reconstruction by separable approximation," IEEE Trans. Signal Process., vol.57, no.7, pp.2479–2493, July 2009.
[56] T. Coleman, M.A. Branch, and A. Grace, Matlab Optimization Toolbox User's Guide, Math Works, 1999.
[57] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx/, April 2011.
[58] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in Recent Advances in Learning and Control, ed. V. Blondel, S. Boyd, and H. Kimura, Lecture Notes in Control and Information Sciences, pp.95–110, Springer-Verlag Limited, 2008.
[59] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[60] R.T. Rockafellar, Convex Analysis, Princeton University Press, 1970.
[61] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed., SIAM, 2003.
[62] http://users.ece.gatech.edu/˜justin/l1magic/
[63] F. Alizadeh and D. Goldfarb, "Second-order cone programming," Mathematical Programming, vol.95, pp.3–51, 2003.
[64] M.A.T. Figueiredo, J.M. Bioucas-Dias, and R.D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," IEEE Trans. Image Process., vol.16, no.12, pp.2980–2991, Dec. 2007.
[65] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal-recovery problems," in Convex Optimization in Signal Processing and Communications, ed. D.P. Palomar and Y.C. Eldar, pp.42–88, Cambridge University Press, 2010.
[66] Y. Nesterov, "Smooth minimization of non-smooth functions," Mathematical Programming, vol.103, pp.127–152, 2005.
[67] Y. Nesterov, "Gradient methods for minimizing composite objective function," Technical Report, CORE, Universite Catholique de Louvain, 2007.
[68] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$," Doklady AN USSR, vol.269, pp.543–547, 1983. translated as Soviet Math. Docl.
[69] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, Robust Statistics: The Approach Based on Influence Functions, Wiley Series in Probability and Statistics, John Wiley & Sons, 1986.
[70] P.J. Huber and E.M. Ronchetti, Robust Statistics, 2nd ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Hokoben, NJ, USA, 2009.
[71] V.N. Vapnik, Statistical Learning Theory, John Wiley & Sons, 1998.
[72] B. Schölkopf and A.J. Smola, Learning with Kernels, MIT Press, 2002.
[73] http://www-stat.stanford.edu/˜candes/nesta/
[74] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Commun. Pure Appl. Math., vol.57, no.11, pp.1413–1457, Nov. 2004.
[75] P.L. Combettes and V.R. Wajs, "Signal recovery by proximal forward-backward splitting," Multiscale Model. Simul., vol.4, pp.1168–1200, 2005.
[76] M. Elad, "Why simple shrinkage is still relevant for redundant representations?," IEEE Trans. Inf. Theory, vol.52, no.12, pp.5559–5569, Dec. 2006.
[77] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," IEEE Signal Process. Mag., vol.27, no.3, pp.76–88, May 2010.
[78] http://sparselab.stanford.edu/
[79] T. Blumensath, M.E. Davies, and G. Rilling, "Greedy algorithms for compressed sensing," in Compressed Sensing, ed. Y.C. Eldar and G. Kutyniok, pp.348–393, Cambridge University Press, 2012.
[80] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials, II. orthogonal matching pursuit versus basis pursuit," Foundations of Computational Mathematics, vol.8, pp.737–763, 2008.
[81] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," Rec. 27th Asilomar Conf. Sign. Syst. Comput., 1993.
[82] R.A. DeVore and V.N. Temlyakov, "Some remarks on greedy algorithms," Advances in Computational Mathematics, vol.5, pp.173–187, 1996.
[83] S. Mallat, G. Davis, and Z. Zhang, "Adaptive time-frequency decompositions," Optical Engineering, vol.33, no.7, pp.2183–2191, July 1994.
[84] J.A. Tropp, "Greed is good: algorithmic results for sparse approximation," IEEE Trans. Inf. Theory, vol.50, no.10, pp.2231–2242, Oct. 2004.
[85] M.A. Davenport and M.B. Wakin, "Analysis of orthogonal match-

ing pursuit using the restricted isometry property," IEEE Trans. Inf. Theory, vol.56, no.9, pp.4395–4401, Sept. 2010.
[86] T. Blumensath and M.E. Davies, "Gradient pursuits," IEEE Trans. Signal Process., vol.56, no.6, pp.2370–2382, June 2008.
[87] D.L. Donoho, Y. Tsaig, I. Drori, and J.L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," 2006. Preprint.
[88] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," Foundations of Computational Mathematics, vol.9, pp.317–334, 2009.
[89] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," IEEE J. Sel. Top. Signal Process., vol.4, no.2, pp.310–316, April 2010.
[90] D. Needell and J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," Appl. Comput. Harmonic Anal., vol.26, no.3, pp.301–321, 2008.
[91] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," IEEE Trans. Inf. Theory, vol.55, no.5, pp.2230–2249, May 2009.
[92] T. Blumensath and M. Davies, "Iterative thresholding for sparse approximations," Journal of Fourier Analysis and Applications, vol.14, pp.629–654, 2008.
[93] T. Blumensath and M.E. Davies, "Iterative hard thresholding for compressed sensing," Applied and Computational Harmonic Analysis, vol.27, no.3, pp.265–274, 2009.
[94] K. Lange, D.R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," Journal of Computational and Graphical Statistics, vol.9, no.1, pp.1–20, March 2000.
[95] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," IEEE Signal Process. Lett., vol.14, no.10, pp.707–710, Oct. 2007.
[96] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," Inverse Problems, vol.24, pp.1–14, 2008.
[97] R. Saab and Ö. Yılmaz, "Sparse recovery by non-convex optimization — Instance optimality," Applied and Computational Harmonic Analysis, vol.29, no.1, pp.30–48, 2010.
[98] B.D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," IEEE Trans. Signal Process., vol.47, no.1, pp.187–200, Jan. 1999.
[99] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on, pp.262–265, July 2009.
[100] D.P. Wipf and B.D. Rao, "Sparse Bayesian learning for basis selection," IEEE Trans. Signal Process., vol.52, no.8, pp.2153–2164, Aug. 2004.
[101] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," IEEE Trans. Signal Process., vol.56, no.6, pp.2346–2356, June 2008.
[102] D. Baron, S. Sarvotham, and R.G. Baraniuk, "Bayesian compressive sensing via belief propagation," IEEE Trans. Signal Process., vol.58, no.1, pp.269–280, Jan. 2010.
[103] D.L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," Proc. National Academy of Sciences, vol.106, no.45, pp.18914–18919, 2009.
[104] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," Proc. IEEE, vol.86, no.10, pp.1951–1968, Oct. 1998.
[105] R.J.M. Cramer, R.A. Scholtz, and M.Z. Win, "Evaluation of an ultra-wide-band propagation channel," IEEE Trans. Antennas Propag., vol.50, no.5, pp.561–570, May 2002.
[106] A.F. Molisch, "Ultrawideband propagation channels — Theory, measurement, and modeling," IEEE Trans. Veh. Technol., vol.54, no.5, pp.1528–1545, Sept. 2005.
[107] A.F. Molisch, "Ultra-wide-band propagation channels," Proc. IEEE, vol.97, no.2, pp.353–371, Feb. 2009.
[108] N. Czink, X. Yin, H. Ozcelik, M. Herdin, E. Bonek, and B. Fleury, "Cluster characteristics in a MIMO indoor propagation environment," IEEE Trans. Wireless Commun., vol.6, no.4, pp.1465–1475, April 2007.
[109] L. Vuokko, V.M. Kolmonen, J. Salo, and P. Vainikainen, "Measurement of large-scale cluster power characteristics for geometric channel models," IEEE Trans. Antennas Propag., vol.55, no.11, pp.3361–3365, Nov. 2007.
[110] V. Raghavan, G. Hariharan, and A.M. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," IEEE J. Sel. Areas Commun., vol.1, no.3, pp.357–371, Oct. 2007.
[111] W. Li and J.C. Preisig, "Estimation of rapidly time-varying sparse channels," IEEE J. Ocean. Eng., vol.32, no.4, pp.927–939, Oct. 2007.
[112] M. Stojanovic, "OFDM for underwater acoustic communications: Adaptive synchronization and sparse channel estimation," Proc. IEEE ICASSP 2008, pp.5288–5291, March 2008.
[113] M. Kocic, D. Brady, and S. Merriam, "Reduced-complexity RLS estimation for shallow-water channels," Proc. AUV'94, pp.165–170, July 1994.
[114] S.F. Cotter and B.D. Rao, "Matching pursuit based decision-feedback equalizers," Proc. IEEE ICASSP'00, vol.5, pp.2713–2716, June 2000.
[115] S.F. Cotter and B.D. Rao, "The adaptive matching pursuit algorithm for estimation and equalization of sparse time-varying channels," Proc. Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, 2000, vol.2, pp.1772–1776, Oct. 2000.
[116] S.F. Cotter and B.D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," IEEE Trans. Commun., vol.50, no.3, pp.374–377, March 2002.
[117] W. Dongming, H. Bing, Z. Junhui, G. Xiqi, and Y. Xiaohu, "Channel estimation algorithms for broadband MIMO-OFDM sparse channel," Proc. 14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003 (PIMRC 2003), vol.2, pp.1929–1933, Sept. 2003.
[118] C.J. Wu and D.W. Lin, "Sparse channel estimation for OFDM transmission based on representative subspace fitting," Proc. IEEE VTC 2005-Spring, pp.495–499, 2005.
[119] C.R. Berger, S. Zhou, J.C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," Proc. IEEE OCEANS 2009, May 2009.
[120] C.R. Berger, S. Zhou, J.C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," IEEE Trans. Signal Process., vol.58, no.3, pp.1708–1721, March 2010.
[121] J.L. Paredes, G.R. Arce, and Z. Wang, "Ultra-wideband compressed sensing: Channel estimation," IEEE J. Sel. Top. Signal Process., vol.1, no.3, pp.383–395, Oct. 2007.
[122] G. Taubock and Hlawatsch, "Compressed sensing based estimation of doubly selective channels using a sparsity-optimized basis expansion," Proc. EUSIPCO 2008, Aug. 2008.
[123] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," IEEE J. Sel. Top. Signal Process., vol.4, no.2, pp.255–271, April 2010.
[124] M. Sharp and A. Scaglione, "A useful performance metric for compressed channel sensing," IEEE Trans. Signal Process., vol.59, no.6, pp.2982–2988, June 2011.
[125] D. Borah and B. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," IEEE Trans. Commun., vol.47, no.6, pp.862–873, June 1999.
[126] G.B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," Proc. IEEE, vol.86, no.10, pp.1969–1986, Oct. 1998.

[127] C. Qi and L. Wu, "Optimized pilot placement for sparse channel estimation in OFDM systems," IEEE Signal Process. Lett., vol.18, no.12, pp.749–752, Dec. 2011.
[128] S. Andradottir, "A global search method for discrete stochastic optimization," SIAM J. Optim., vol.6, no.2, pp.513–530, May 1996.
[129] G. Gui, Z. Chen, Q. Meng, Q. Wan, and F. Adachi, "Compressed channel estimation for sparse multipath two-way relay networks," Int. J. Phys. Sci., vol.6, no.12, pp.2782–2788, June 2011.
[130] W.U. Bajwa, J. Haupt, A.M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," Proc. IEEE, vol.98, no.6, pp.1058–1076, June 2010.
[131] C.R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," IEEE Commun. Mag., vol.48, no.11, pp.164–174, Nov. 2010.
[132] J. Haupt, W.U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," IEEE Trans. Inf. Theory, vol.56, no.11, pp.5862–5875, Nov. 2010.
[133] E. Moulines, P. Duhamel, J.F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," IEEE Trans. Signal Process., vol.43, no.2, pp.516–525, Feb. 1995.
[134] K. Hayashi, H. Matsushima, H. Sakai, E. de Carvalho, and P. Popovski, "Subspace based blind sparse channel estimation," Proc. APSIPA ASC 2012, Dec. 2012.
[135] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," Proc. IPSN'06, pp.134–1423, April 2006.
[136] J. Haupt, W.U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," IEEE Signal Process. Mag., vol.25, no.2, pp.92–101, March 2008.
[137] J. Meng, H. Li, and Z. Han, "Sparse event detection in wireless sensor networks using compressive sensing," Proc. CISS 2009, pp.181–185, March 2009.
[138] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "On the interplay between routing and signal representation for compressive sensing in wireless sensor networks," Proc. 2009 Information Theory and Applications Workshop, pp.206–215, Feb. 2009.
[139] S. Lee, S. Pattem, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega, "Compressed sensing and routing in multi-hop networks," University of Southern California CENG Technical Report, April 2009.
[140] X. Wang, Z. Zhao, Y. Xia, and H. Zhang, "Compressed sensing based random routing for multi-hop wireless sensor networks," Proc. ISCIT 2010, pp.220–225, Oct. 2010.
[141] X. Wang, Z. Zhao, Y. Xia, and H. Zhang, "Compressed sensing for efficient random routing in multi-hop wireless sensor networks," Proc. IEEE GLOBECOM Workshops 2010, pp.266–271, Dec. 2010.
[142] S. Katti, S. Shintre, S. Jaggi, D. Katabi, and M. Medard, "Real network codes," Proc. Forty-Fifth Annual Allerton Conference, pp.389–395, Sept. 2007.
[143] N. Nguyen, D.L. Jones, and S. Krishnamurthy, "Netcompress: Coupling network coding and compressed sensing for efficient data communication in wireless sensor networks," Proc. 2010 IEEE SIPS, pp.356–361, Oct. 2010.
[144] F. Fazel, M. Fazel, and M. Stojanovic, "Random access compressed sensing for energy-efficient underwater sensor networks," IEEE J. Sel. Areas Commun., vol.29, no.8, pp.1660–1670, Sept. 2011.
[145] M. Kaneko and K. Al Agha, "Compressed sensing based protocol for efficient reconstruction of sparse superimposed data in a multi-hop wireless sensor network," arXiv:1208.1410v1 [cs.OH], http://arxiv.org/abs/1208.1410, Aug. 2012.
[146] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," J. Amer. Stat. Assoc., vol.91, no.433, pp.365–377, March 1996.
[147] A. Coates, A.O. Hero, III, R. Nowak, and B. Yu, "Internet tomography," IEEE Signal Process. Mag., vol.19, no.3, pp.47–65, May 2002.
[148] M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly, "Compressed sensing MRI," IEEE Signal Process. Mag., vol.25, no.2, pp.72–82, March 2008.
[149] M.H. Firooz and S. Roy, "Network tomography via compressed sensing," Proc. IEEE GLOBECOM 2010, pp.1–5, Dec. 2010.
[150] M.H. Firooz and S. Roy, "Link delay estimation via expander graphs," arXiv:1106.0941v1 [cs.NI], http://arxiv.org/abs/1106.0941, June 2011.
[151] W. Xu, E. Mallada, and A. Tang, "Compressive sensing over graphs," Proc. IEEE INFOCOM 2011, pp.2087–2095, April 2011.
[152] T. Matsuda, M. Nagahara, and K. Hayashi, "Link quality classifier with compressed sensing based on ℓ_1-ℓ_2 optimization," IEEE Commun. Lett., vol.15, no.10, pp.1117–1119, Oct. 2011.
[153] E.J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," IEEE Trans. Inf. Theory, vol.52, no.12, pp.5406–5425, Dec. 2006.
[154] W. Xu and B. Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs," Proc. IEEE ITW'07, pp.414–419, Sept. 2007.
[155] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," IEEE Trans. Inf. Theory, vol.55, no.9, pp.4299–4308, Sept. 2009.
[156] P. Indyk and M. Ruzic, "Near-optimal sparse recovery in the l1 norm," Proc. IEEE FOCS'08, pp.199–207, Oct. 2008.
[157] D.A. Roberson, C.S. Hood, J.L. LoCicero, and J.T. MacDonald, "Spectral occupancy and interference studies in support of cognitive radio technology deployment," Proc. 2006 1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks, pp.26–35, Sept. 2006.
[158] Z. Tian and G.B. Giannakis, "Compressed sensing for wideband cognitive radios," Proc. IEEE ICASSP 2007, vol.4., pp.1357–1360, April 2007.
[159] Y.L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," Proc. IEEE ICASSP 2009, pp.2337–2340, April 2009.
[160] V. Havary-Nassab, S. Hassan, and S. Valaee, "Compressive detection for wide-band spectrum sensing," Proc. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010), pp.3094–3097, March 2010.
[161] Z. Tian, "Compressed wideband sensing in cooperative cognitive radio networks," Proc. IEEE GLOBECOM 2008, pp.1–5, Nov. 2008.
[162] Y. Wang, A. Pandharipande, Y.L. Polo, and G. Leus, "Distributed compressive wide-band spectrum sensing," Proc. 2009 IEEE Information Theory and Applications Workshop, pp.178–183, Feb. 2009.
[163] F. Zeng, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," IEEE J. Sel. Top. Signal Process., vol.5, no.1, pp.37–48, Feb. 2011.
[164] J.A. Bazerque and G.B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," IEEE Trans. Signal Process., vol.58, no.3, pp.1847–1862, March 2010.
[165] S. Cheng, V. Stankovic, and L. Stankovic, "An efficient spectrum sensing scheme for cognitive radio," IEEE Signal Process. Lett., vol.16, no.6, pp.501–504, June 2009.
[166] Z. Zhang, Z. Han, H. Li, D. Yang, and C. Pei, "Belief propagation based cooperative compressed spectrum sensing in wideband cognitive radio networks," IEEE Trans. Wireless Commun., vol.10, no.9, pp.3020–3031, Sept. 2011.
[167] C.C. Huang and L.C. Wang, "Dynamic sampling rate adjustment for compressive spectrum sensing over cognitive radio network," IEEE Wireless Commun. Lett., vol.1, no.2, pp.57–60, April 2012.
[168] Z. Zhang, H. Li, D. Yang, and C. Pei, "Space-time Bayesian compressed spectrum sensing for wideband cognitive radio networks,"

Proc. 2010 IEEE Symposium on New Frontiers in Dynamic Spectrum, pp.1–11, April 2010.
[169] W. Yin, Z. Wen, S. Li, J. Meng, and Z. Han, "Dynamic compressive spectrum sensing for cognitive radio networks," Proc. CISS 2011, pp.1–6, March 2011.
[170] X. Wang, W. Guo, Y. Lu, and W. Wang, "Adaptive compressive sampling for wideband signals," Proc. IEEE VTC Spring 2011, pp.1–5, May 2011.
[171] Z. Tian, G. Leus, and V. Lottici, "Joint dynamic resource allocation and waveform adaptation for cognitive networks," IEEE J. Sel. Areas Commun., vol.29, no.2, pp.443–454, Feb. 2011.
[172] S. Kirolos, T. Ragheb, J. Laska, M.E. Duarte, Y. Massoud, and R.G. Baraniuk, "Practical issues in implementing analog-to-information converters," Proc. 6th International Workshop on System-on-Chip for Real-Time Applications, pp.141–146, Dec. 2006.
[173] E.J. Candès and M.B. Wakin, "An introduction to compressive sampling," IEEE Signal Process. Mag., vol.25, no.2, pp.21–30, March 2008.
[174] M. Mishali, Y. Eldar, and A. Elron, "Xampling: Signal acquisition and processing in union of subspaces," IEEE Trans. Signal Process., vol.59, no.10, pp.4719–4734, Oct. 2011.
[175] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," IEEE J. Sel. Areas Commun., vol.23, no.2, pp.201–220, Feb. 2005.
[176] A.C. Gurbuz, J.H. McClellan, and V. Cevher, "A compressive beamforming method," Proc. IEEE ICASSP 2008, pp.2617–2620, March 2008.
[177] J.M. Kim, O.K. Lee, and J.C. Ye, "Compressive MUSIC: Revisiting the link between compressive sensing and array signal processing," IEEE Trans. Inf. Theory, vol.58, no.1, pp.278–301, Jan. 2012.
[178] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," Proc. IEEE ICASSP'96, vol.3, pp.1688–1691, May 1996.
[179] Y. Bresler, "Spectrum-blind sampling and compressive sensing for continuous-index signals," Proc. 2008 Information Theory and Applications Workshop, pp.547–554, Jan. 2008.
[180] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," IEEE Trans. Signal Process., vol.54, no.12, pp.4634–4643, Dec. 2006.
[181] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," IEEE Trans. Signal Process., vol.53, no.7, pp.2477–2488, July 2005.
[182] M. Mishali and Y.C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," IEEE Trans. Signal Process., vol.56, no.10, pp.4692–4702, Oct. 2008.
[183] E. van den Berg and M.P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," IEEE Trans. Inf. Theory, vol.56, no.5, pp.2516–2527, May 2010.
[184] R. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag., vol.34, no.3, pp.276–280, March 1986.
[185] G. Prisco and M. D'Urso, "Maximally sparse arrays via sequential convex optimizations," IEEE Antennas Wireless Propag. Lett., vol.11, pp.192–195, 2012.
[186] M. Carlin, G. Oliveri, and A. Massa, "A cs-based strategy for the design of shaped-beam sparse arrays," Proc. IEEE APSURSI 2011, pp.1996–1999, July 2011.
[187] G. Oliveri, F. Robol, M. Carlin, and A. Massa, "Synthesis of large sparse linear arrays by Bayesian compressive sensing," Proc. EUCAP 2011, pp.2078–2081, April 2011.
[188] G. Oliveri and A. Massa, "Bayesian compressive sampling for pattern synthesis with maximally sparse non-uniform linear arrays," IEEE Trans. Antennas Propag., vol.59, no.2, pp.467–481, Feb. 2011.
[189] M.D. Migliore, "A compressed sensing approach for array diagnosis from a small set of near-field measurements," IEEE Trans. Antennas Propag., vol.59, no.6, pp.2127–2133, June 2011.
[190] R. Baraniuk and P. Steeghs, "Compressive radar imaging," Proc. 2007 IEEE Radar Conference, pp.128–133, April 2007.
[191] C. Berger, B. Demissie, J. Heckenbach, P. Willett, and S. Zhou, "Signal processing for passive radar using OFDM waveforms," IEEE J. Sel. Top. Signal Process., vol.4, no.1, pp.226–238, Feb. 2010.
[192] Y.S. Yoon and M.G. Amin, "Imaging of behind the wall targets using wideband beamforming with compressive sensing," Proc. 2009 IEEE/SP 15th Workshop on Statistical Signal Processing, pp.93–96, Aug. 2009.
[193] R. Mao and H. Li, "A novel multiple access scheme via compressed sensing with random data traffic," Proc. IEEE WCNC 2010, pp.1–6, April 2010.
[194] T. Xue, X. Dong, and Y. Shi, "A multiple access scheme based on multi-dimensional compressed sensing," Proc. IEEE ICC 2012, pp.1–5, June 2012.
[195] S.T. Qaseem, T.Y. Al-Naffouri, and T.M. Al-Murad, "Compressive sensing based opportunistic protocol for exploiting multiuser diversity in wireless networks," Proc. IEEE PIMRC 2009, pp.1447–1451, Sept. 2009.
[196] S.R. Bhaskaran, L. Davis, A. Grant, S. Hanly, and P. Tune, "Downlink scheduling using compressed sensing," Proc. 2009 IEEE Information Theory Workshop on Networking and Information Theory, pp.201–205, June 2009.
[197] S.T. Qaseem and T.Y. Al-Naffouri, "Compressive sensing for reducing feedback in MIMO broadcast channels," Proc. IEEE ICC 2010, pp.1–5, May 2010.
[198] W. Zhang, M.S. Branicky, and S.M. Phillips, "Stability of networked control systems," IEEE Control Syst. Mag., vol.21, no.1, pp.84–99, Feb. 2001.
[199] J.P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," Proc. IEEE, vol.95, no.1, pp.138–162, Jan. 2007.
[200] A. Bemporad, M. Heemels, and M. Johansson, Networked Control Systems, Springer, Oct. 2010.
[201] M. Nagahara and D.E. Quevedo, "Sparse representations for packetized predictive networked control," IFAC 18th World Congress, pp.84–89, Aug. 2011.
[202] M. Nagahara, D.E. Quevedo, and J. Ostergaard, "Packetized predictive control for rate-limited networks via sparse representation," 51st IEEE Conference on Decision and Control (CDC), pp.1362–1367, Dec. 2012.
[203] M. Nagahara, T. Matsuda, and K. Hayashi, "Compressive sampling for remote control systems," IEICE Trans. Fundamentals, vol.E95-A, no.4, pp.713–722, April 2012.
[204] M. Nagahara, D.E. Quevedo, T. Matsuda, and K. Hayashi, "Compressive sampling for networked feedback control," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp.2733–2736, March 2012.
[205] "Special issue: Applications of sparse representation and compressive sensing," Proc. IEEE, vol.98, no.6, pp.906–1101, June 2010.
[206] R.G. Baraniuk, E. Candès, R. Nowak, and M. Vetterli, "Compressive sampling," IEEE Signal Process. Mag., vol.25, no.2, pp.12–13, March 2008.
[207] Compressive Sensing Resources, http://dsp.rice.edu/cs
[208] The Big Picture, https://sites.google.com/site/igorcarron2/cs/
[209] Nuit Blanche, http://nuit-blanche.blogspot.jp/
[210] S. Gleichman and Y.C. Eldar, "Blind compressed sensing," IEEE Trans. Inf. Theory, vol.57, no.10, pp.6958–6975, Oct. 2011.
[211] K. Engan, S.O. Aase, and J.H. Husoy, "Frame based signal compression using method of optimal directions (mod)," Proc. IEEE ISCAS'99, vol.4, pp.1–4, July 1999.
[212] M.S. Lewicki and T.J. Sejnowski, "Learning overcomplete representations," Neural Comput., vol.12, no.2, pp.337–365, Feb. 2000.

[213] K. Kreutz-Delgado, J.F. Murray, B.D. Rao, K. Engan, T.W. Lee, and T.J. Sejnowski, "Dictionary learning algorithms for sparse representation," Neural Comput., vol.15, no.2, pp.349–396, Feb. 2003.
[214] M. Aharon, M. Elad, and M. Bruckstein, "On the uniqueness of overcomplete dictionaries, and practical way to retrieve them," Linear Alg. and Its Applic., vol.416, no.1, pp.48–67, July 2006.

Kazunori Hayashi received the B.E., M.E. and Ph.D. degrees in communication engineering from Osaka University, Osaka, Japan, in 1997, 1999 and 2002, respectively. Since 2002, he has been with the Department of Systems Science, Graduate School of Informatics, Kyoto University. He is currently an Associate Professor there. His research interests include digital signal processing for communication systems. He received the ICF Research Award from the KDDI Foundation in 2008, the IEEE Globecom 2009 Best Paper Award, the IEICE Communications Society Excellent Paper Award in 2011, the WPMC'11 Best Paper Award, and the Telecommunications Advancement Foundation Award in 2012. He is a member of IEEE and ISCIE.

Masaaki Nagahara received the Bachelor's degree in engineering from Kobe University in 1998, the Master's degree and the Doctoral degree in informatics from Kyoto University in 2000 and 2003. He is currently a Lecturer at the Graduate School of Informatics, Kyoto University. His research interests include digital signal processing and digital control systems. He received Young Authors Award in 1999 and Best Paper Award in 2012 from the Society of Instrument and Control Engineers (SICE), and Transition to Practice Award from IEEE Control Systems Society in 2012. He is a member of IEEE, ISCIE, and SICE.

Toshiyuki Tanaka received the B. Eng., M. Eng., and Dr. Eng. degrees in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1988, 1990, and 1993, respectively. From 1993 to 2005, he was with the Department of Electronics and Information Engineering at Tokyo Metropolitan University, Tokyo, Japan. He is currently a professor at the Graduate School of Informatics, Kyoto University, Kyoto, Japan. He received DoCoMo Mobile Science Prize in 2003, and Young Scientist Award from the Minister of Education, Culture, Sports, Science and Technology, Japan, in 2005. His research interests are in the areas of neural networks, machine learning, and information and communication theory. He is a member of the IEEE, the Japanese Neural Network Society, the Acoustical Society of Japan, the Physical Society of Japan, and the Architectural Institute of Japan.