Probabilistic Engineering Mechanics 61 (2020) 103082

Contents lists available at ScienceDirect

Probabilistic Engineering Mechanics

journal homepage: www.elsevier.com/locate/probengmech

Sparse representations and compressive sampling approaches in engineering mechanics: A review of theoretical concepts and diverse applications

Ioannis A. Kougioumtzoglou *, Ioannis Petromichelakis, Apostolos F. Psaros

Department of Civil Engineering and Engineering Mechanics, Columbia University, 500 W 120th St, New York, NY 10027, United States

ARTICLE INFO

Keywords:
Sparse representations
Compressive sampling
Engineering mechanics
Uncertainty quantification
Incomplete data

ABSTRACT

A review of theoretical concepts and diverse applications of sparse representations and compressive sampling (CS) approaches in engineering mechanics problems is provided from a broad perspective. First, following a presentation of well-established CS concepts and optimization algorithms, attention is directed to currently emerging tools and techniques for enhancing solution sparsity and for exploiting additional information in the data. These include alternative to ℓ_1-norm minimization formulations and iterative re-weighting solution schemes, Bayesian approaches, as well as structured sparsity and dictionary learning strategies. Next, CS-based research work of relevance to engineering mechanics problems is categorized and discussed under three distinct application areas: a) *inverse problems in structural health monitoring*, b) *uncertainty modeling and simulation*, and c) *computationally efficient uncertainty propagation*. Notably, the vast majority of problems in all three areas share the challenge of "incomplete data", addressed by the versatile CS framework. In this regard, incomplete data may manifest themselves in various different forms and can correspond to missing or compressed data, or even refer generally to insufficiently few function evaluations. The primary objective of this review paper relates to identifying and presenting significant contributions in each of the above three application areas in engineering mechanics, with the goal of expediting additional research and development efforts. To this aim, an extensive list of 248 references is provided, composed almost exclusively of books and archival papers, which can be readily available to a potential reader.

1. Introduction

The problem of determining the current and predicting the future states of a system based on knowledge of a limited number of data points has been a persistent challenge in a wide range of scientific fields. Advancements in this direction have led to various significant theoretical results, which have unequivocally revolutionized modern science. One of the most characteristic examples relates to the development of representations based on Fourier series [1]. This trigonometric series expansion of periodic functions has served as the starting point for various efficient expansion and representation schemes (e.g., [2]). During the past fifteen years, research efforts have focused on identifying and exploiting low-dimensional representations of high-dimensional data, as well as on establishing conditions guaranteeing unique representation in the low-dimensional space. This has triggered the birth of the currently expanding field of compressive sampling (CS) (e.g., [3,4]), as well as the rejuvenation of the more general field of sparse representations (e.g., [5,6]).

From a historical perspective, there have been several examples and early observations suggesting that signal reconstruction is possible by utilizing a smaller number of samples than the minimum dictated by the Shannon–Nyquist (SN) theorem (e.g., [2,7]). Indicatively, Carathéodory showed in [8,9] that a signal expressed as a sum of any k sinusoids can be recovered based on knowledge of its values at zero time and at any other $2k$ time points. Further, Beurling [10] discussed the possibility of extrapolating in a nonlinear manner and determining the complete Fourier transform of a signal assuming that only part of the Fourier transform is known. Dorfman [11] studied the combinatorial group testing problem and provided one of the first sparse signal recovery problem formulations. Also, Logan [12] showed that it is possible to reconstruct a band-limited corrupted signal by an ℓ_1-norm minimization approach. These early, seemingly paradoxical, results were further supported by relevant studies in the field of geophysics [13–15] (see also [16]) pertaining to the analysis of seismic signals of spike train form due to the layered structure of geological formations. It was shown that these sparse spike trains can be recovered accurately based on incomplete and noisy measurements.

Nevertheless, it can hardly be disputed that sparse representations theory and tools have been revitalized in recent years due to the pioneering work in [17–19], which provided bounds on the number

* Corresponding author.
E-mail address: ikougioum@columbia.edu (I.A. Kougioumtzoglou).

https://doi.org/10.1016/j.probengmech.2020.103082
Received 11 February 2020; Received in revised form 27 May 2020; Accepted 22 June 2020
Available online 25 June 2020
0266-8920/© 2020 Elsevier Ltd. All rights reserved.

of measurements required for the recovery of high-dimensional data under the condition that the latter possess a low-dimensional representation in a transformed domain. The aforementioned theoretical results, coupled with potent numerical algorithms from the well-established field of convex optimization, have led to numerous impactful contributions in a wide range of application areas. In this regard, CS-related theoretical advancements and diverse applications associated with the fields of signal and image processing, biomedicine, communication systems and sensor networks, information security and pattern recognition have been well-documented in dedicated books (e.g., [3,4,20–26]), special issues (e.g., [27]) and review papers (e.g., [28–43]).

More recently, the field of engineering mechanics has also benefited from the advent of sparse representations and CS approaches in conjunction with uncertainty quantification and health monitoring of diverse systems and structures. However, to the best of the authors' knowledge, there are currently no review papers providing a comprehensive discussion and a broad perspective on the aforementioned developments in engineering mechanics. In fact, although there have been a couple of relevant efforts reported previously, these focus either on specific and relatively narrow application domains, or are cross-disciplinary in nature and lack any focus on a specific research field. Indicatively, the authors in [44] focus exclusively on reviewing polynomial chaos expansions coupled with CS approaches as applied in stochastic mechanics problems, whereas reference [45] discusses the problem of CS-based governing dynamics modeling of complex systems with applications in interdisciplinary science and engineering.

In this regard, in an effort to address this gap in the literature and to complement some of the previous works by incorporating more recent developments, this review paper focuses on sparse representations and CS approaches in the field of engineering mechanics. Specifically, in Section 2, following a presentation of well-established CS concepts and optimization algorithms, attention is directed to currently emerging tools and techniques for enhancing solution sparsity and for exploiting additional information in the data. These include alternative to ℓ_1-norm minimization formulations and iterative re-weighting solution schemes, Bayesian approaches, as well as structured sparsity and dictionary learning strategies. Next, in Section 3, a rather broad perspective is provided on CS-related contributions to engineering mechanics, and relevant research work is categorized under three distinct application areas: (a) inverse problems in structural health monitoring, (b) uncertainty modeling and simulation, and (c) computationally efficient uncertainty propagation. Notably, the vast majority of problems in all three areas share the challenge of "incomplete data", addressed by the versatile CS framework. In this regard, incomplete data may manifest themselves in various different forms and can correspond to missing or compressed data, or even refer generally to insufficiently few function evaluations. Further, concluding remarks are presented in Section 4. It is noted that the primary objective of this review paper relates to identifying and discussing significant contributions in each of the above three application areas in engineering mechanics, with the goal of expediting additional research and development efforts. To this aim, an extensive list of 248 references is provided, composed almost exclusively of books and archival papers, which can be readily available to a potential reader.

2. Theoretical concepts and algorithmic aspects

In this section, the basic theoretical concepts and algorithmic aspects related to sparse representations and CS tools are reviewed. To enhance the pedagogical merit of the paper and motivate the reader, a simple example is provided first where the necessity for CS tools arises naturally. Next, the problem of approximating a sparse signal is formulated as an optimization problem and solved via a brute-force approach. The need for more computationally efficient tools is discussed and relevant methodologies are presented. Further, the critical question regarding performance guarantees and the number of required measurements is addressed. Lastly, methodologies for exploiting additional information present in the data and for enhancing solution sparsity are also discussed. The interested reader is also directed to the books in Refs. [3–6] for a more detailed presentation from a signal processing and a mathematics perspectives.

2.1. Motivation

The SN theorem states that a bandlimited continuous-time signal can be exactly reconstructed by a set of uniformly spaced measurements sampled at the Nyquist rate; i.e., at a frequency double the maximum frequency present in the signal (e.g., [2,7]). Although the SN theorem has impacted significantly the signal processing field and related applications, in many cases the dictated minimum number of measurements can be prohibitive from a computational cost perspective [3]. In this regard, the acquired signals are often compressed by utilizing an appropriate change of basis. In this new basis, the expansion coefficient vector has only few nonzero elements; thus, yielding significant savings in terms of required storage capacity (see, for instance, related "lossy" compression techniques [46]).

In this context, a vector $\boldsymbol{x}$ of length n is referred to as k-sparse if at most k out of its n components are nonzero; that is,

$$\|\boldsymbol{x}\|_0 \leq k \tag{1}$$

where $\|\cdot\|_p$ denotes the ℓ_p-norm defined as [6]

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \tag{2}$$

for $0 < p < \infty$. Note that, according to the definition of Eq. (2), $\|\cdot\|_0$ (also called the cardinality of $\boldsymbol{x}$) is not a proper norm. However, it can be defined as the limit of $\|\cdot\|_p$ for $p \to 0$; that is,

$$\|\boldsymbol{x}\|_0 = \sum_{i=1}^{n} \lim_{p\to 0} |x_i|^p = \sum_{i=1}^{n} I(x_i) \tag{3}$$

In Eq. (3), $I(x_i)$ is the indicator function, which takes the value 0 if $x_i = 0$ and 1 otherwise [6]. Moreover, if the coefficient vector is k-sparse, then the original signal is characterized as sparse, or, in other words, it exhibits a sparse representation in this particular expansion basis. Accordingly, a vector is referred to as compressible (or approximately sparse), if it can be approximated satisfactorily by a sparse vector. Typical examples of sparse, compressible, and dense (i.e., not sparse) vectors are shown in Fig. 1 to illustrate the differences between them. In passing, it is worth noting that there also exist generalizations of the concept of a vector norm to that of a matrix norm. Indicatively, the Frobenius norm (given as the sum of squares of the elements of a matrix) and the nuclear norm (given as the sum of singular values of a matrix) can be construed as generalizations of the ℓ_2- and the ℓ_1-norms, respectively, to account for matrices (e.g., [47,48]).

Next, motivated by the aforementioned sparsity, inherent in a wide range of signals in various applications, it is natural to pose the question whether it is possible to bypass the potentially cumbersome task of recording the signal at a rate dictated by the SN theorem. In this manner, acquisition of the signal directly in compressed form by employing a sub-Nyquist rate would circumvent the computationally costly two-step process of capturing and storing the complete signal first, and then compressing it by discarding the redundant information. For tutorial effectiveness, consider the continuous-time signal

$$y(t) = 10\cos(30t + \pi) + 12\cos\left(20t + \frac{\pi}{8}\right) + 4\cos\left(6t + \frac{2\pi}{3}\right) \tag{4}$$

where $0 \leq t < 2\pi$. According to the SN theorem [2,7], exact reconstruction requires the signal to be sampled at minimum 60 points in the time domain. However, the signal of Eq. (4) can be represented exactly in the frequency domain by 6 coefficients only, i.e., 2 for each dimension of the real and imaginary components of the harmonics at 6, 20 and 30 rad/s. Clearly, in this case, the frequency domain representation of the signal can be construed as significantly more "compact" (or,

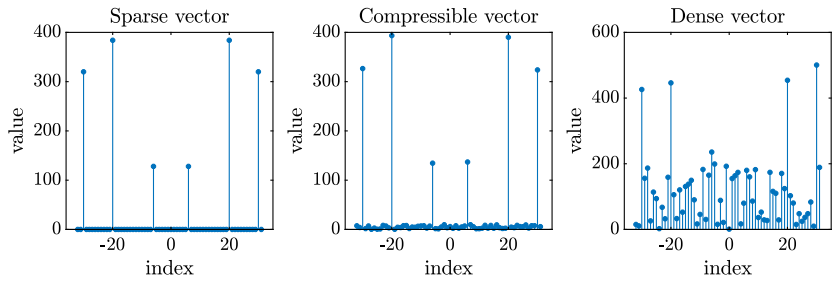


Fig. 1. Typical examples of sparse, compressible and dense (i.e., not sparse) vectors.

in other words, sparser) than the corresponding representation in the time domain. Considering this seemingly inefficient signal representation in one domain (where typically measurements are acquired) and the considerably more compact representation in a different one, it would be advantageous to develop a methodology for reconstructing (exactly or approximately) signals that are known to have a sparse representation in some basis by collecting the smallest possible number of measurements.

This problem takes the form of an underdetermined linear system of equations, i.e.,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 \tag{5}$$

where $\boldsymbol{y}$ is a vector containing $m < n$ measurements of the original signal, $\boldsymbol{x}_0$ is the coefficient vector, and $\boldsymbol{A}$ can be either fixed, or written as $\boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{D}$, with $\boldsymbol{D}$ being the basis matrix and $\boldsymbol{\Phi}$ an $m \times n$ matrix (also known as CS matrix [49] as it randomly deletes rows of $\boldsymbol{D}$). In passing, the original complete signal is denoted by $\boldsymbol{y}_0$ and is given as $\boldsymbol{y}_0 = \boldsymbol{D}\boldsymbol{x}_0$. For the specific aforementioned example of Eq. (4), the $\boldsymbol{y}$ vector represents the measurements in the time domain, while $\boldsymbol{D}$ represents the Fourier basis matrix. Obviously, the underdetermined system of Eq. (5) has either no solution, or an infinite number of solutions. Further, since most real-life signals are corrupted with noise and are characterized by measurement errors, the reconstruction tools should exhibit robustness in their implementation. In fact, the reconstruction tools should also exhibit stability in their performance and be capable of addressing even cases of almost-sparse signals.

2.2. Regularization of the underdetermined system of equations

It has been shown that although the system of Eq. (5) is amenable in general to an infinite number of solutions, it can be regularized (i.e., constrained) so that only one solution is relevant (e.g., [4]). Specifically, it can take the form of a constrained optimization problem for which the objective function relates to the sparsity of the signal (expressed via the ℓ_0-norm); and feasible solutions are only the ones satisfying Eq. (5), which acts as the constraint. This can be written as

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \quad \text{subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{6}$$

Further, it has been proved [6] that utilizing $m \geq 2k$ measurements Eq. (6) yields a unique solution $\boldsymbol{x}$, equal to the k-sparse coefficient vector $\boldsymbol{x}_0$. This result defines a measurement bound (see also Sections 2.4–2.7 for more general results and discussion) dictating the required number of measurements for a certain sparsity degree of $\boldsymbol{x}_0$, and is based on the necessary and sufficient condition (e.g., [6])

$$2k \leq \text{krank}(\boldsymbol{A}) \tag{7}$$

In Eq. (7), krank($\boldsymbol{A}$) is the largest number s for which every subset of s column vectors of $\boldsymbol{A}$ is linearly independent. The bound $m \geq 2k$ can be easily obtained by utilizing the condition of Eq. (7) in conjunction with the inequality

$$\text{krank}(\boldsymbol{A}) \leq \text{rank}(\boldsymbol{A}) \leq m \tag{8}$$

that holds for any arbitrary matrix [6]. Note that krank($\boldsymbol{A}$) is NP-hard (NP standing for nondeterministic polynomial time) to compute, i.e., even for relatively small matrices $\boldsymbol{A}$ there is no available algorithm for computing it efficiently [6]. Therefore, it is often impossible to verify numerically the condition of Eq. (7). Alternatively, the mutual coherence of $\boldsymbol{A}$ is easier to evaluate from an algorithmic point of view. This is defined as

$$\mu(\boldsymbol{A}) = \max_{i \neq j} \left| \left\langle \frac{\boldsymbol{a}_i}{\|\boldsymbol{a}_i\|}, \frac{\boldsymbol{a}_j}{\|\boldsymbol{a}_j\|} \right\rangle \right| \tag{9}$$

and can be used to determine whether the ℓ_0-norm minimization problem of Eq. (6) has a unique solution coinciding with $\boldsymbol{x}_0$. In Eq. (9), $\boldsymbol{a}_i$, $\forall i \in \{1, \ldots, n\}$, represents the ith column vector of $\boldsymbol{A}$, and $\langle \cdot, \cdot \rangle$ denotes an inner product. Thus, $\mu(\boldsymbol{A})$ can be construed as a metric of the independence degree of the basis vectors of $\boldsymbol{A}$. In this regard, considering also that krank($\boldsymbol{A}$) $\geq 1/\mu(\boldsymbol{A})$ [6], satisfying the condition

$$2k \leq \frac{1}{\mu(\boldsymbol{A})} \tag{10}$$

is sufficient for obtaining $\boldsymbol{x}_0$ as the unique solution of the ℓ_0-norm minimization problem of Eq. (6) with $m \geq 2k$ [50]. It becomes clear that matrices $\boldsymbol{A}$ with higher mutual coherence values $\mu(\boldsymbol{A})$ are better choices within the context of CS and sparse approximations. More importantly, it has been shown that matrices $\boldsymbol{A}$, satisfying conditions such as those of Eqs. (7) and (10), can be constructed by random submatrices of bounded orthonormal systems (e.g., Fourier, wavelet and Legendre bases with randomly deleted rows [4]). Indicatively, regarding the ℓ_0-norm minimization problem of Eq. (6), any signal with a k-sparse Fourier coefficient vector can be exactly reconstructed by $2k$ randomly selected measurements in the time domain. Thus, the signal of Eq. (4) can be reconstructed by utilizing 12 measurements only as compared to the 60 measurements required by the SN theorem [4].

Further, as mentioned in Section 2.1, measurement errors and noise are often unavoidable in practice, while the exact sparsity degree of the coefficient vector may not be known a priori. Clearly, this motivates the reformulation of the ℓ_0-norm minimization problem of Eq. (6) to account for measurement errors and for approximately sparse coefficient vectors. In this regard, Eq. (6) is cast in the form

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \quad \text{subject to } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \epsilon \tag{11}$$

where ϵ is a pre-specified discrepancy from the measurement vector $\boldsymbol{y}$. It can be readily seen that the equality constraint of Eq. (6) is replaced in Eq. (11) by an inequality constraint; thus, enlarging the set of feasible solutions and enhancing the flexibility and robustness of the technique from a practical implementation perspective. Notably, theoretical results referring to necessary and sufficient conditions for sparse signal reconstruction, similar to the ones described by Eqs. (7) and (10), are also available for the formulation of Eq. (11) [5]. In Sections 2.4–2.5, approximate solution techniques are presented for the ℓ_0-norm minimization problem of Eq. (11), which is also referred to in the literature as "noisy" ℓ_0-norm minimization problem.

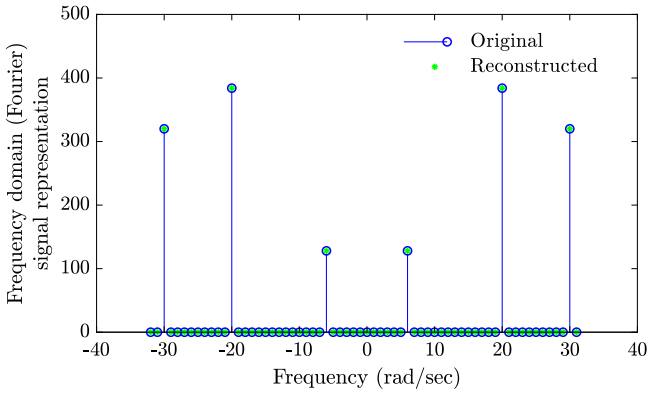


Fig. 2. Frequency domain representation of the original (blue circles) and the reconstructed (green dots) signals of Eq. (4).

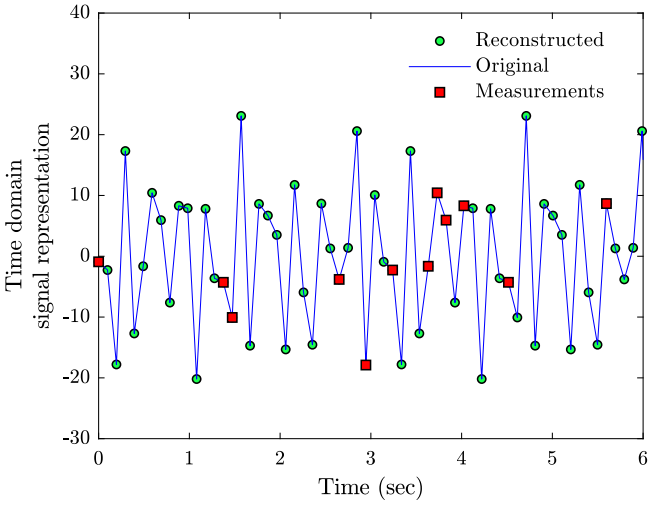


Fig. 3. Time domain representation of the original (blue line) and the reconstructed (green circles) signals of Eq. (4); the set of randomly sampled measurements (red squares) is also included.

2.3. A brute-force solution approach for the ℓ_0-norm minimization problem

A rather brute-force approach for solving the ℓ_0-norm minimization problem of Eq. (6) relates to employing an exhaustive search by considering all possible supports (i.e., all possible combinations of the nonzero components locations) for the estimated vector $\boldsymbol{x}$ and by checking if a solution to $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ exists for each and every support. In this regard, Fig. 2 shows the reconstructed Fourier coefficient vector for the signal of Eq. (4) by solving the ℓ_0-norm minimization problem of Eq. (6) via exhaustive search. Further, in Fig. 3 the original and the reconstructed signals in the time domain are plotted. It is seen that the reconstructed signal by utilizing only 12 measurements (i.e., 20% of the number required by the SN theorem) matches perfectly that of Eq. (4).

Although it becomes clear, based on the simple example of Eq. (4), that a CS-based signal reconstruction is advantageous due to the considerably smaller number of required measurements, a brute-force solution approach by exhaustive search becomes computationally prohibitive for an increasing size of $\boldsymbol{A}$. This is readily understood by considering the number of supports to be examined in an exhaustive search solution framework, which is equal to $\binom{n}{k}$. In fact, it has been shown that solving the ℓ_0-norm minimization problem of Eq. (6) (or alternatively, Eq. (11)) is NP-hard [6,51].

To address this challenge, two main categories of solution approaches are presented and discussed in the following sections. Specifically, in Section 2.4 methods for solving approximately the ℓ_0-norm minimization problems of Eqs. (6) and (11) are described, and in Section 2.5 methods for solving a relaxed form of the ℓ_0-norm minimization problem are presented. In passing, it is worth noting that, as mentioned in Section 1, a wide range of application areas have benefited from the advent of CS-based theoretical concepts and related numerical solution schemes. This is anticipated considering the fact that various seemingly unrelated problems originating from different disciplines can be cast as underdetermined systems of equations of the form of Eq. (5). In this regard, the solution approaches discussed in the ensuing sections are quite general and have been adopted in a rather straightforward manner by diverse research fields, including the field of engineering mechanics, which is the focus of this paper.

2.4. Approximate solutions to the ℓ_0-norm minimization problem: Greedy methods

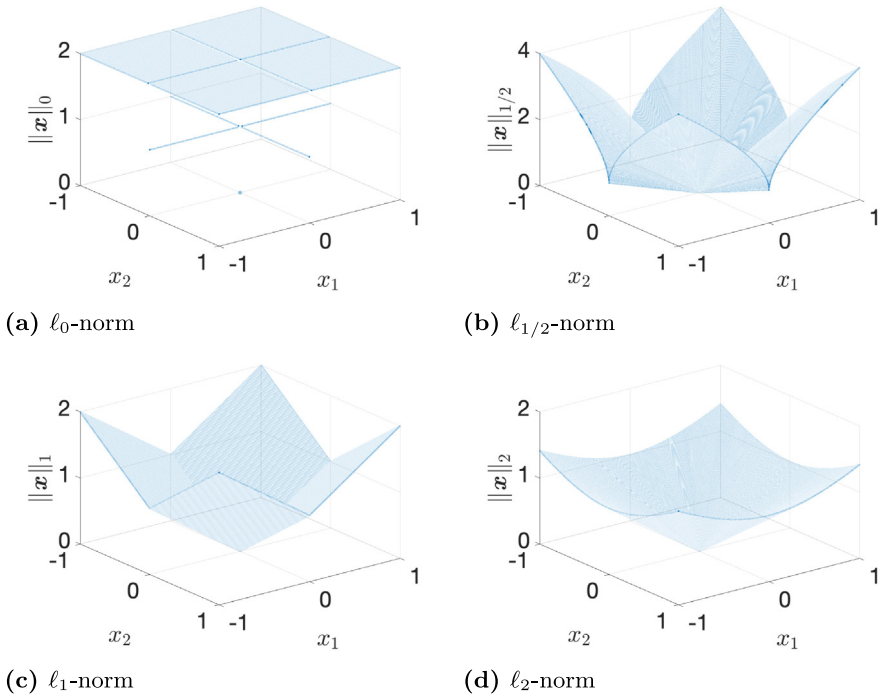
An approximate approach for solving the ℓ_0-norm problem of Eq. (6) (or Eq. (11)) relates to, first, determining the support of the coefficient vector (i.e., the optimal locations of the nonzero coefficients) and, second, estimating the values of the nonzero coefficients. In this regard, aiming to obtain a globally optimal solution, greedy methods make a sequence of decisions based on certain local optimality conditions. More specifically, greedy methods can be broadly divided into two categories [3]. The first relates to greedy pursuits, which construct the support of an initially empty coefficient vector $\boldsymbol{x}$ in an iterative manner. This is done by adding in each iteration a column vector of matrix $\boldsymbol{A}$ that best fits the measurements $\boldsymbol{y}$ according to a pre-specified criterion, and subsequently, by estimating the coefficients of the selected support. The second relates to thresholding-type methods, which iteratively obtain an estimate $\boldsymbol{x}$ and employ "hard" thresholding to retain only the largest k coefficients. Concisely, the main difference between these two categories is that greedy pursuits construct the solution $\boldsymbol{x}$ gradually by adding nonzero coefficients to it, whereas thresholding methods provide an estimate for the solution in each iteration cycle, and subsequently, remove the relatively small coefficients as dictated by the prescribed threshold.

Greedy methods have been developed in conjunction with diverse applications in various fields such as signal processing and statistics. In this regard, a vast amount of algorithms with similar characteristics is available in the literature, while multiple extensions and modifications have been proposed over the years [3,6]. Typical examples of greedy pursuits include the matching pursuit (MP) [52] and the orthogonal matching pursuit (OMP) [53] algorithms, whereas indicative popular thresholding methods include the iterative hard thresholding [54] and the compressive sampling matching pursuit [55] algorithms. In general, it can be argued that greedy methods are attractive due to the fact that their numerical implementation is rather straightforward, and that the associated computational cost is kept at a reasonable level under the condition that the sparsity degree k is relatively low. Further, in many cases theoretical performance guarantees are available (e.g., [3]), while empirical studies can be performed to analyze and compare the performance of different algorithms (see also Section 2.7). The interested reader is also directed to [3,5] and references therein for more details.

In the following, the OMP algorithm, which is one of the most widely utilized greedy methods, is presented in more detail [4]. The input to OMP is the m-length measurement vector $\boldsymbol{y}$, the $m \times n$ matrix $\boldsymbol{A}$ and the sparsity degree k that the coefficient vector $\boldsymbol{x}_0$ is anticipated to exhibit. At the $j = 0$ iteration, the algorithm starts with an empty support $\mathcal{S}^{(0)}$, a zero vector $\boldsymbol{x}^{(0)}$, and thus, with a residual vector equal to

$$\boldsymbol{r}^{(0)} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(0)} = \boldsymbol{y} \tag{12}$$

Next, the column vector of matrix $\boldsymbol{A}$ with the maximum correlation with the residual $\boldsymbol{r}^{(0)}$ is selected. In this manner, the corresponding location in the support of $\boldsymbol{x}^{(1)}$ becomes active. Subsequently, for the



(a) ℓ_0-norm (b) $\ell_{1/2}$-norm

(c) ℓ_1-norm (d) ℓ_2-norm

Fig. 4. Two-dimensional representation of various norms restricted in the domain $[-1,1]\times[-1,1]$.

selected support $\mathcal{S}^{(1)}$, the value of the nonzero component is obtained by employing ordinary least squares and minimizing $\left\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}^{(1)}\right\|_2$. The above scheme is repeated in an iterative manner until convergence. The convergence criterion may relate, for instance, to the quantity $\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2$ becoming smaller than a prescribed threshold, or to the number of nonzero elements in $\boldsymbol{x}$ exceeding a given value. It is noted that, at every iteration of the algorithm, the vector $\boldsymbol{x}$ is updated as an orthogonal projection of $\boldsymbol{y}$ onto the subspace of the selected vectors from $\boldsymbol{A}$. The fact that all the components of $\boldsymbol{x}$ may change at a given iteration constitutes the most pronounced difference with the related MP algorithm. Specifically, at the jth iteration, the MP algorithm adds a single component in $\boldsymbol{x}$ according to the criterion of maximum reduction of the residual $\boldsymbol{r}^{(j+1)}$, without affecting the rest of the vector elements.

Further, indicative OMP enhancements include efficient implementation schemes for performing the matrix–vector multiplications $\boldsymbol{A}^T\boldsymbol{r}^{(j)}$ to obtain correlation degree estimates at each iteration j (e.g., based on fast Fourier transform [5]), reduction of the least squares related computational cost (e.g., based on QR factorization [56]), approaches for adding not only one, but multiple elements to the support at each iteration (e.g., Stagewise OMP [57]), and various alternative strategies for selecting the next location to be added to the support $\mathcal{S}$ (e.g., least squares OMP [5]).

Also, it is worth noting that the aforementioned algorithms (and OMP in particular), which aim at solving the ℓ_0-norm minimization problem directly, perform satisfactorily when the sparsity k of the coefficient vector is small compared to its size n. However, they can become computationally prohibitive in cases, for example, of high-dimensional problems exhibiting low sparsity degrees.

2.5. Relaxation of the ℓ_0-norm minimization problem: ℓ_1-norm minimization

As highlighted in Section 2.3, the ℓ_0-norm minimization problem in Eq. (6) is NP-hard to solve. To address this challenge, a popular solution approach relates to utilizing a surrogate for approximating the non-convex objective function. To this aim, convex surrogates appear particularly attractive, primarily due to the plethora of well-established theoretical tools and efficient numerical algorithms for analyzing and solving convex optimization problems.

In this regard, there are several properties rendering the ℓ_1-norm an efficacious convex surrogate for the non-convex ℓ_0-norm. Specifically, focusing on the hypercube B_∞ defined as $B_\infty = \{\boldsymbol{x}|\,\|\boldsymbol{x}\|_\infty \leq 1\}$ (i.e., each dimension of $\boldsymbol{x}$ is restricted within the interval $[-1,1]$), the ℓ_1-norm is the largest convex function not exceeding $\|.\|_0$; that is, $\|\boldsymbol{x}\|_1 = \sup\{h(\boldsymbol{x}) \mid h \text{ is convex and } h(\boldsymbol{x}) \leq \|\boldsymbol{x}\|_0\ \forall \boldsymbol{x} \in B_\infty\}$. This property can be proved formally (e.g., [4]) and such a function is referred to typically as convex envelope. In Fig. 4, the two-dimensional case is depicted for tutorial effectiveness, where it can be observed that the ℓ_1-norm (Fig. 4c) is indeed the convex envelope of the ℓ_0-norm (Fig. 4a). Further, it is seen that the $\ell_{1/2}$-norm (Fig. 4b) is non-convex (this holds for all ℓ_p-norms with $0 < p < 1$), whereas the ℓ_2-norm (Fig. 4d) is not the largest convex underestimator of the ℓ_0-norm; the latter holds true for all ℓ_p-norms with $p > 1$ (e.g., [5]). Next, utilizing the ℓ_1-norm surrogate, the optimization problem in Eq. (6) is replaced by

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_1 \quad \text{subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{13}$$

which is a convex optimization problem that can be solved much more efficiently than the one in Eq. (6) [3,6]. Eq. (13) is also referred to in the literature as convex relaxation of the ℓ_0-norm minimization problem.

A graphical demonstration of the property of the ℓ_1-norm minimization to recover sparse vectors is provided in Fig. 5. Specifically, the set of feasible vectors $\boldsymbol{x}$ related to the optimization problem in Eq. (13) form the affine subspace

$$S = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\} = \{\boldsymbol{x}_0\} + \text{null}(\boldsymbol{A}) \tag{14}$$

where null($\boldsymbol{A}$) denotes the nullspace of matrix $\boldsymbol{A}$. The ℓ_1-norm minimization determines the point in S with the smallest ℓ_1-norm. This procedure can be illustrated by considering the ℓ_1-ball

$$B_1 = \{\boldsymbol{x} \mid \|\boldsymbol{x}\|_1 \leq 1\} \tag{15}$$

of radius one in $\mathbb{R}^n$, which contains all vectors $\boldsymbol{x}$ with objective function in Eq. (13) at most equal to one. Scaling B_1 by s yields the set of vectors $\boldsymbol{x}$ with objective function at most equal to s. In this regard, initializing the process by setting the value of s equal to zero, B_1 is expanded gradually by increasing s. The ℓ_1-norm minimizer is obtained when sB_1 first reaches the affine subspace S and this intersection point $\boldsymbol{x}_0$ is the solution to the optimization problem in Eq. (13). Considering the

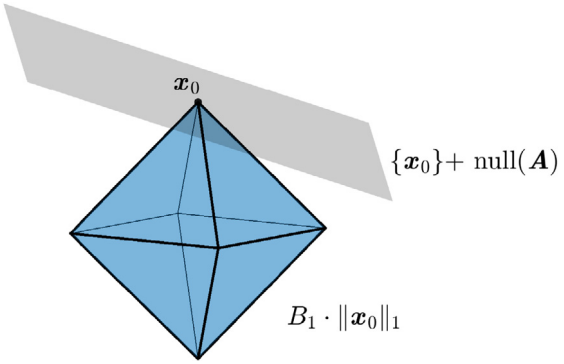


Fig. 5. Graphical demonstration of the ℓ_1-norm property of recovering sparse vectors: sparse vector x_0, feasible set $\{x_0\} + \text{null}(A)$ and ℓ_1-ball of radius $\|x_0\|_1$.

geometry of the ball B_1, it is readily seen that the possible solution points belong either to the vertices or to the edges of B_1, which correspond, indeed, to sparse vectors.

Motivated by early observations in the 1960 s [12] related to its efficiency in recovering sparse vectors, ℓ_1-norm minimization has been employed over the past few decades in various applications based on rather heuristic arguments. In fact, it was not until the first decade of the twenty-first century that conditions guaranteeing sparse recovery were formally stated and proved. Specifically, it was shown in [50,58] that the condition of Eq. (10) guarantees that the ℓ_1-norm minimization problem in Eq. (13) has a unique solution provided that the columns of A have unit ℓ_2-norms. Moreover, this solution coincides with the unique solution of the ℓ_0-norm minimization problem in Eq. (6). Notably, it was shown in [59] that the mutual coherence of any matrix $A \in \mathbb{R}^{m\times n}$ with unit ℓ_2-norm columns is bounded by

$$\mu(A) \geq \sqrt{\frac{n-m}{m(n-1)}} \tag{16}$$

Eq. (16) is referred to in the literature as Welch's bound (e.g., [5]). Taking into account Eqs. (10) and (16) yields a bound estimate of the form $m \geq Ck^2$ on the number m of measurements required for the recovery of a k-sparse vector via ℓ_1-norm minimization; C denotes a constant factor. Note, however, that this bound on m appears quite conservative from a practical perspective, and that a significantly smaller number of measurements (than of the order of k^2) can be adequate for successful sparse recovery based on ℓ_1-norm minimization.

In this regard, a condition that leads to considerably "tighter" bounds on the number m of measurements guaranteeing exact ℓ_1-norm recovery of any k-sparse vector, relates to the restricted isometry property (RIP). According to [60], a matrix A satisfies the RIP of order k, with constant δ, if

$$\forall\ x\ k\text{-sparse}, \quad (1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \tag{17}$$

Also, the order k RIP constant $\delta_k(A)$ is the smallest number δ satisfying the inequality in Eq. (17). Further, it was shown in [17] that if $y = Ax_0$ with $\|x_0\|_0 = k$ and $\delta_{2k}(A) < \sqrt{2}-1$, then x_0 is the unique optimal solution of the optimization problem in Eq. (13). Next, focusing on the RIP of Gaussian random matrices, i.e., matrices A with independent $\mathcal{N}(0, 1/m)$ random variables as entries, it was shown (e.g., [17,61]) that a k-sparse vector can be reconstructed via ℓ_1-norm minimization by employing only $m \geq C'k\log(n/k)$ random measurements, where C' is a constant factor. This is a substantially tighter bound on m as compared to $m \geq Ck^2$ derived by employing Eqs. (10) and (16). Note also that the bound $m \geq C'k\log(n/k)$ allows for (k, m, n) to scale proportionally [18], a result that contributed to the rapid development of the CS field. One of the tightest known bounds of this kind related to Gaussian matrices was derived in [19]. This requires only $m \geq 8k\log(n/k) + 12k$ measurements and does not involve any unknown constants. However, notwithstanding their considerable theoretical value, in many cases the practical merit of such measurement bounds is limited by the fact that the sparsity degree k of the target vector is unknown, in general, a priori. Thus, alternative approaches are required for assessing the performance of sparse recovery algorithms (see Section 2.7). In passing, it is mentioned for completeness that two other relevant properties, which are actually utilized in the proof of exact ℓ_1-norm recovery under the RIP [17], are the nullspace and the restricted strong convexity properties. Unfortunately, the problem of verifying any of these conditions for a given matrix A is NP-hard [62].

2.6. Approximate solutions to the ℓ_1-norm minimization problem: Convex optimization methods

The ℓ_1-norm minimization problem of Eq. (13), also known as basis pursuit, can be cast in the form

$$\min_x \|x\|_1 \quad \text{subject to } \|y - Ax\|_2 \leq \epsilon \tag{18}$$

to account for measurement error as also explained in Section 2.1. To facilitate its numerical solution, Eq. (18) can be equivalently written as an unconstrained problem in the form

$$\min_x \|y - Ax\|_2^2 + \lambda\|x\|_1 \tag{19}$$

The problem of Eq. (19) is also referred to in the literature as basis pursuit denoising (BPDN) [63], or least absolute shrinkage and selection operator (LASSO) [64].

Two main challenges to be addressed for solving, in a computationally efficient manner, the convex optimization problem of Eq. (19) relate to scalability and non-differentiability. Second-order convex optimization algorithms, such as interior-point methods or quasi-Newton schemes [65], are advantageous in the sense that they require typically relatively few iterations to converge. Note, however, that for an n-variate problem each iteration involves the solution of an $n \times n$ linear system, incurring a computational cost of $O(n^3)$ per iteration. In various signal processing and statistical learning applications (as well as in many engineering mechanics applications discussed in Section 3), the number n of variables can reach the order of millions; thus, rendering the computational cost of even a single iteration prohibitively large. Therefore, attention has been directed to alternative algorithms, which utilize only first-order information about the objective function.

A standard first-order method in convex optimization is the gradient descent method, which in its basic form and under certain smoothness conditions exhibits a convergence rate of $O(1/j)$, where j is the iteration number. Nevertheless, the objective function in Eq. (19) involves the non-differentiable ℓ_1-norm. To address this challenge related to the evaluation of the gradient, subgradient methodologies can be employed; however, these are typically characterized by a relatively poor convergence rate of the order $O(1/\sqrt{j})$ [6].

Further, the proximal gradient (PG) method (e.g., [66]) exhibits considerable efficiency in solving optimization problems, where the objective function consists of the sum of a smooth convex function $f(x)$ (with ∇f being Lipschitz continuous; see [67] for a definition of Lipschitz continuity) and a non-differentiable convex function $g(x)$, such as in Eq. (19). In this regard, the proximal operator is defined as

$$\text{prox}_g(z) = \arg\min_x \left\{ g(x) + \frac{1}{2}\|x - z\|_2^2 \right\} \tag{20}$$

and the update formulas at iteration j take the form

$$\begin{aligned} z^{(j)} &= x^{(j)} - \frac{1}{L}\nabla f\left(x^{(j)}\right) \\ x^{(j+1)} &= \text{prox}_{g/L}\left(z^{(j)}\right) \end{aligned} \tag{21}$$

where L is typically equal to the Lipschitz constant of ∇f. The jth PG iteration in Eq. (21) can be construed as a two-step update formula, where, first, an ordinary gradient descent step $z^{(j)}$ decreasing the smooth function f is determined and, second, the step $x^{(j+1)}$

is chosen in a manner that it both reduces the value of the non-differentiable function g and remains close to $\boldsymbol{z}^{(j)}$ via the introduction of the term $\left\|\boldsymbol{x}-\boldsymbol{z}^{(j)}\right\|_2^2$ in the proximal operator of Eq. (20). The strong convexity of $\|.\|_2^2$ guarantees that the PG step has a unique solution [68]. Also, compared to the subgradient method, the PG algorithm yields a convergence rate of $O(1/j)$, which corresponds to the standard case of no non-differentiable terms. Clearly, although the PG method exhibits a relatively high convergence rate, the minimization of a non-differentiable function at each iteration can be computationally demanding. Nevertheless, the non-differentiable ℓ_1-norm function in Eq. (19) yields a proximal operator (also known as soft thresholding operator) in closed-form, i.e.,

$$[\text{prox}_{\lambda\|.\|_1}(\boldsymbol{z})]_i = \text{soft}(z_i,\lambda) = \text{sign}(z_i)\max(|z_i|-\lambda,0) \tag{22}$$

This leads to the iterative soft-thresholding algorithm (ISTA) [69], which is widely utilized in ℓ_1-norm minimization approaches. Moreover, it has been shown that the theoretically optimal convergence rate for first-order optimization methods is $O(1/j^2)$ and this can be achieved by Nesterov's accelerated gradient method (AGM) [70]. Note that the fundamental concept in AGM, which relates to introducing a momentum term in the gradient descent update formula, can be readily used in conjunction with the ISTA. This has led to the computationally efficient fast iterative shrinkage-thresholding algorithm (FISTA) [71] with a convergence rate of $O(1/j^2)$.

A potential shortcoming of ISTA and FISTA relates to parameter λ, which needs to be tuned for solving the problem in Eq. (19). The least angle regression (LARS) algorithm [21,72] constitutes an alternative approach with the advantageous feature of computing the entire solution path directly; that is, the solution of Eq. (19) corresponding to a given range of λ values is provided as the output of a single run of the algorithm. LARS follows a procedure somewhat similar to the OMP, and thus, it can be argued that its performance deteriorates for problems with increasing dimensionality. It is worth mentioning that there exist various other algorithms for solving the BPDN problem in Eq. (19). Indicatively, these include coordinate descent algorithms, which update a single coordinate at each iteration, and primal–dual algorithms (e.g., [73]).

Alternative solution schemes, which deviate considerably from the standard ℓ_1-norm formulation of Eq. (19) and aim at further enhancing the sparsity of the solution, are presented and discussed in Section 2.8.

2.7. Performance analysis

Successful CS-based reconstruction of a sparse coefficient vector $\boldsymbol{x}_0$ relies on a priori knowledge of the minimum possible number m, where m is the size of the measurement vector $\boldsymbol{y}$. Of course, the selected solution algorithm, the basis matrix $\boldsymbol{D}$ and the CS matrix $\boldsymbol{\Phi}$ affect the reconstruction accuracy as well. In this regard, this section focuses on presenting both theoretical results and practical approaches for addressing the above points related to CS performance assessment.

The problem is typically posed in the literature as determining the minimum number m for exact reconstruction of an arbitrary coefficient vector $\boldsymbol{x}_0$ with sparsity degree k, given a fixed matrix $\boldsymbol{A}$ (strong case). Alternatively, the weak case relates to determining the minimum number m for exact reconstruction of a specific coefficient vector $\boldsymbol{x}_0$ with sparsity degree k by appropriately constructing a matrix $\boldsymbol{A}$; see also [74].

In this regard, the potential of combinatorial geometry has been explored for providing precise measurement bounds (e.g., [74]; see also [75] for alternative approaches). More specifically, consider the ℓ_1-ball of Fig. 5, which represents a convex polytope $\mathcal{C}$ in $\mathbb{R}^n$. A polytope generalizes the definition of a three-dimensional polyhedron to n dimensions and is characterized by a number $f_0(\mathcal{C})$ of 0-dimensional faces (i.e., vertices), a number $f_1(\mathcal{C})$ of 1-dimensional faces (i.e., edges), and, in general, a number $f_k(\mathcal{C})$ of k-dimensional faces (e.g., [74]). Obviously, multiplying a matrix $\boldsymbol{A}$ of size $m\times n$ by a vector of size n projects the vector onto a lower-dimensional space, and thus, the number of k-dimensional faces of $\boldsymbol{A}\mathcal{C}$ can only be less than or equal to $f_k(\mathcal{C})$, i.e.,

$$f_k(\boldsymbol{A}\mathcal{C}) \le f_k(\mathcal{C}) \quad \text{for} \quad k\ge 0 \tag{23}$$

Notably, it has been proved (e.g., [76]) that the ratio of face counts of the projected polytope $\boldsymbol{A}\mathcal{C}$ over the original polytope $\mathcal{C}$ is equal to the probability of exact reconstruction of $\boldsymbol{x}_0$ by solving the ℓ_1-norm minimization problem of Eq. (13). Also, it has been shown that, for the weak case [76,77] and for matrices $\boldsymbol{A}$ with independent identically distributed $\mathcal{N}(0,1)$ Gaussian random entries and provided a sufficiently large number m, the fraction $f_k(\boldsymbol{A}\mathcal{C})/f_k(\mathcal{C})$ approaches 1 as the problem dimension n approaches infinity. Various other similar results have been obtained referring to, indicatively, the reconstruction of a non-negative coefficient vector by solving a special form of Eq. (13) (e.g., [74]), cases of employing other than ℓ_1-norm regularization methods (e.g., [78]), cases of finite n values (e.g., [79]), as well as the strong reconstruction case for which the condition $f_k(\boldsymbol{A}\mathcal{C}) = f_k(\mathcal{C})$ must hold.

However, notwithstanding empirical evidence indicating that exact reconstruction with a similarly small number m of measurements is possible (e.g., [77]), theoretical results on precise measurement bounds for cases of non-Gaussian matrices $\boldsymbol{A}$ have been scarce; see also [75]. In this regard, an alternative approach relates to associating the reconstruction performance with certain properties of matrix $\boldsymbol{A}$, such as krank($\boldsymbol{A}$), $\mu(\boldsymbol{A})$, nullspace property, and RIP; see also Sections 2.2–2.5. Concisely, exact reconstruction of the coefficient vector $\boldsymbol{x}_0$ is guaranteed (at least with high probability; see, for instance, [60]) if a condition, such as Eq. (17), pertaining to the sparsity degree k and to a property of $\boldsymbol{A}$ is satisfied. In this context, it is noted that there has been extensive research during the past decade (e.g., [4]) on identifying properties of $\boldsymbol{A}$ with direct relation to the performance of the reconstruction problem. Of course, verifying such conditions for a given matrix $\boldsymbol{A}$, or constructing a matrix $\boldsymbol{A}$ adhering to prescribed properties, are nontrivial challenges. For example, Eq. (7) represents a necessary and sufficient condition guaranteeing exact reconstruction of $\boldsymbol{x}_0$ by solving either the ℓ_0-norm or the ℓ_1-norm minimization problems of Eqs. (6) and (13), respectively; however, it is NP-hard to verify [62]. On the other hand, it is rather straightforward to construct matrices $\boldsymbol{A}$ with low mutual coherence $\mu(\boldsymbol{A})$ dictated by the sufficient condition of Eq. (10) (e.g., [4,58]); however, the measurement bound obtained is rather conservative (referred to as "pessimistic" in the CS literature [74]). Further, the nullspace property guaranteeing exact reconstruction of $\boldsymbol{x}_0$ by solving the ℓ_1-norm minimization problem of Eq. (13) is also NP-hard to verify [4]. Regarding matrices $\boldsymbol{A}$ satisfying the RIP of Eq. (17), these can be constructed by employing, for instance, random submatrices of bounded orthonormal systems [4], and thus, RIP has been used in various practical problems. Nevertheless, the RIP-based measurement vector size m is also a pessimistic bound; see also [60,74] for a discussion and [80] for related improvements.

Although the aforementioned results and conditions are characterized by theoretical rigor and have been catalytic for the advancement of CS, alternative rather empirical approaches are necessary for addressing more general cases. Indicatively, these include the tasks of tuning a certain algorithm (i.e., selecting an optimal set of parameters; e.g., [81]) and comparing performances of different reconstruction algorithms [82], as well as cases of coefficient vectors exhibiting structured sparsity (see Section 2.8.3); and thus, alternative algorithms are required for exploiting this additional information [83]. In this regard, empirical measurement bounds are often constructed in practice in the form of a phase diagram, i.e., a diagram depicting the transition from accurate recovery to recovery with significant error. In fact, such a diagram not only provides a required number of measurements as a function of problem size n and sparsity degree k, but also illustrates the behavior of the reconstruction error with varying values of m and k. Of particular importance to applications is the width of the transition

zone from accurate to inaccurate reconstruction, which has been shown to be sharper for increasing values of n (e.g., [78]). It is worth noting that phase diagrams, as a tool for assessing the performance of CS methodologies, appear versatile in addressing a wide range of diverse problems in a rather straightforward manner.

Phase diagrams are typically constructed with the aid of synthetic data. Specifically, for a fixed coefficient vector length n, synthetic vectors $\boldsymbol{x}_0$ are constructed randomly with varying sparsity degree values k, and reconstruction is attempted based on measurement vectors $\boldsymbol{y}$ of varying sizes m (see [74] for more details). The above procedure is applied for every possible combination of (m, k) with successful reconstruction indicated when the error associated with the estimate $\boldsymbol{x}$, i.e.,

$$err = \frac{\|\boldsymbol{x} - \boldsymbol{x}_0\|_2}{\|\boldsymbol{x}_0\|_2} \tag{24}$$

is below a certain threshold (e.g., $< 10^{-5}$). Finally, the mean reconstruction success rate is plotted for each and every pair of m/n (undersampling ratio or degree of underdeterminacy) and k/m (sparsity ratio). In Fig. 6, indicative results are plotted for various values of m/n and k/m, with $n = 100$ and $\boldsymbol{A}$ being the Fourier basis matrix with randomly deleted rows. The mean success rate has been evaluated based on 200 reconstruction runs by employing the standard basis pursuit SPGL1 algorithm [84]. The region corresponding to reconstruction error less than 10^{-5} with high probability is shown with blue color, whereas yellow indicates the region corresponding to inaccurate reconstruction with high probability. The transition zone lies in between. To provide an illustrative example, it is seen that for a k-sparse coefficient vector $\boldsymbol{x}_0$ with $n = 100$ and $k = 20$, $m = 30$ measurements are adequate to yield successful reconstruction with high probability.

2.8. Enhancing sparsity and exploiting additional information in the data

In this section, attention is directed to currently emerging tools and techniques for enhancing solution sparsity and for exploiting additional information in the data. These include alternative to ℓ_1-norm minimization formulations and iterative re-weighting solution schemes, Bayesian approaches, as well as structured sparsity and dictionary learning strategies.

2.8.1. Alternative to ℓ_1-norm minimization formulations and iterative re-weighting solution schemes

As discussed in Sections 2.2–2.3, although the ℓ_0-norm formulation of Eq. (6) leads to sparse coefficient vectors based on minimal number of measurements, there is no known algorithm for solving it efficiently. In this regard, although convex ℓ_1-norm relaxations of the ℓ_0-norm problem have been proposed to address this challenge (see Sections 2.5–2.6), it has been shown that alternative, mostly non-convex, proxies of ℓ_0-norm exhibit enhanced sparsity-promoting behavior in comparison to ℓ_1-norm. Indicatively, the difference of the convex ℓ_1- and ℓ_2-norms has been considered in [85,86], leading to an overall non-convex Lipschitz-continuous metric denoted as ℓ_{1-2}-norm. The related minimization problem can be solved, for instance, by the difference of convex functions algorithm [87].

An alternative, more general, formulation of the sparse vector recovery problem relates to replacing the ℓ_0-norm minimization criterion by an ℓ_p-norm criterion as

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_p^p \quad \text{subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{25}$$

where $0 < p \leq 1$, and to employing an iterative re-weighting solution scheme. In fact, although the formulation in Eq. (25) is non-convex (for $p \neq 1$), it can be solved efficiently by iteratively minimizing a convex function, such as the re-weighted ℓ_2-norm. In this context, the focal underdetermined system solver (FOCUSS) [88] has been one of the first such research efforts followed by a number of relevant contributions [89–91] addressing the problem of Eq. (25) in conjunction with iterative re-weighting solution schemes. Further, important theoretical results, similar to the RIP (see Section 2.5), have been established [92–95] providing conditions guaranteeing equivalence between Eqs. (25) and (6).

It is worth mentioning that the iteratively-reweighted-least-squares (IRLS) method, initially introduced for robust statistical estimation applications [96,97], has also received significant attention with the advent of CS [4,98]. The IRLS solves a least squares problem iteratively considering $\|\boldsymbol{x}\|_1 = \boldsymbol{x}^T\boldsymbol{X}^{-1}\boldsymbol{x}$ and $\boldsymbol{X} = \text{diag}(|\boldsymbol{x}|)$. In other words, IRLS re-weights the ℓ_2-norm iteratively to approximate an ℓ_1-norm minimization function. In a similar manner, FOCUSS re-weights the ℓ_2-norm iteratively to approximate the ℓ_p-norm minimization function of Eq. (25) with $0 < p < 1$. Note that FOCUSS has been widely utilized in early studies of the dictionary learning problem as well (see Section 2.8.4).

In the following, attention is directed to a solution scheme initially proposed in [99], which re-weights the ℓ_1-norm iteratively to approximate an ℓ_0-norm minimization function. This solution scheme, referred to as IRℓ_1 in the ensuing analysis, aims at enhancing the sparsity exhibited by the ℓ_1-norm formulation while preserving convexity. The rationale relates to minimizing the influence of the nonzero coefficients magnitude, similarly to the ℓ_0-norm. In this regard, a number of positive weights $w_1, \ldots, w_n$ are introduced, and the "weighted" ℓ_1-norm minimization problem is formulated as

$$\min_{\boldsymbol{x}} \|\boldsymbol{W}\boldsymbol{x}\|_1 \quad \text{subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{26}$$

where $\boldsymbol{W} = \text{diag}(w_1, \ldots, w_n)$. It is noted that the solution of the problem in Eq. (26) does not coincide, in general, with the solution of the original problem in Eq. (13). In fact, the weights w_i can be construed as parameters to be appropriately selected for improving the reconstruction performance. Since the weights w_i are introduced to counteract the influence of the coefficients magnitude, it is evident that their optimal values should be inversely proportional to the magnitudes, i.e.,

$$w_i = \begin{cases} \frac{1}{|x_{0,i}|} & \text{if } x_{0,i} \neq 0 \\ \infty & \text{if } x_{0,i} = 0 \end{cases} \tag{27}$$

Taking into account Eq. (27), the problem in Eq. (26) is guaranteed to yield the correct solution $\boldsymbol{x}$ under the assumption that $m \geq k$ [99].

Clearly, since $\boldsymbol{x}_0$ is the unknown vector to be determined, it is not possible to select the weights according to Eq. (27). Nevertheless, large weights are used in practice to discourage nonzero coefficients, whereas small weights are used to encourage nonzero coefficients. To provide an illustrative example, consider the 3-dimensional problem shown in Fig. 7, where the target vector is $\boldsymbol{x}_0 = [0, 0, 1]^T$, $\boldsymbol{A} = [6, -2, 3]$, and the plane in gray color represents the affine subspace $\{\boldsymbol{x}_0\}+\text{null}(\boldsymbol{A})$, i.e., the set of points $\boldsymbol{x} \in \mathbb{R}^3$ satisfying $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{y}$. It is observed in Fig. 7a that the plane $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ intersects with the interior of the ℓ_1-ball of radius 1 centered at the origin. Hence, the ℓ_1-norm minimization approach discussed in Section 2.5 (see Fig. 5 and the subsequent discussion) recovers the incorrect vector $\bar{\boldsymbol{x}} = [0.5, 0, 0]^T$ shown in Fig. 7b. Next, considering a weighting matrix $\boldsymbol{W} = \text{diag}(5, 5, 1)$, the weighted ℓ_1-norm minimization of Eq. (26) correctly recovers $\boldsymbol{x}_0$ as shown in Fig. 7c, which depicts the weighted ℓ_1-ball B_1^W of radius 1, defined as $B_1^W = \{\boldsymbol{x} \mid \|\boldsymbol{W}\boldsymbol{x}\|_1 \leq 1\}$. It is worth noting that any choice of the weights with $w_1 > 2w_3$ and $3w_2 > 2w_3$ in the above example yields a sufficiently modified (weighted) ℓ_1-ball for recovering $\boldsymbol{x}_0$. The fact that there may exist a wide range of possible candidate values for the weights $w_1, \ldots, w_n$ has motivated the development of an iterative algorithm in [99]. Initially, the weights are set equal to 1 and the formulation degenerates to Eq. (13), which yields a first approximation $\boldsymbol{x}^{(0)}$ of the target vector. Next, the weights are updated according to

$$w_i^{(j+1)} = \frac{1}{|x_i^{(j)}| + \epsilon} \quad \forall \; i = 1, \ldots, n \tag{28}$$

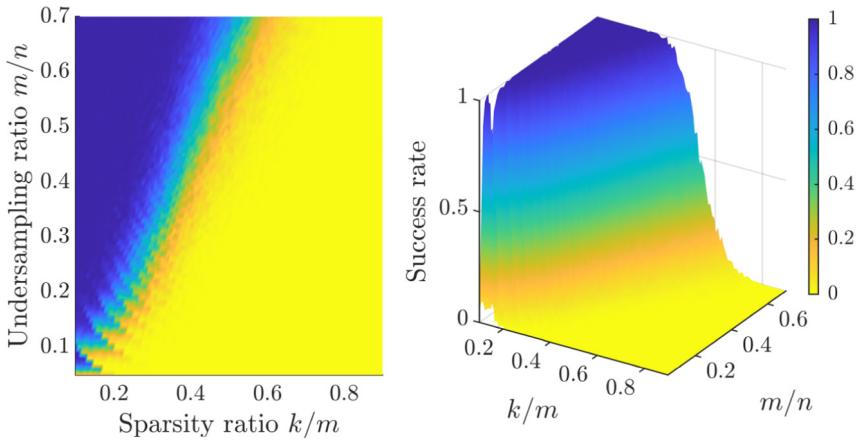


Fig. 6. Phase diagram corresponding to random sub-matrices of a Fourier basis matrix and to reconstruction using the SPGL1 algorithm. The brightness of each point represents the observed success rate, ranging from certain failure (yellow) to certain success (blue). The z-axis corresponds to the average success rate over 200 runs; the y-axis corresponds to the ratio showing the degree of the problem underdeterminacy, whereas the y-axis corresponds to the ratio showing the sparsity degree of the coefficient vector. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

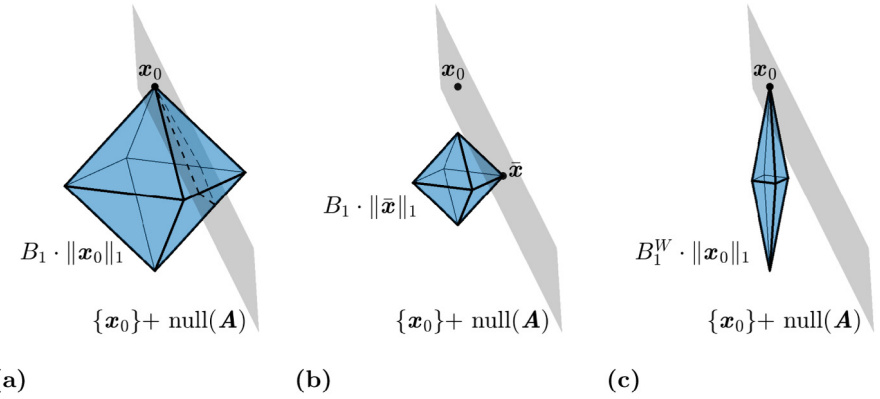


Fig. 7. Weighted ℓ_1-norm minimization for improved sparse signal recovery. (a) Sparse vector $\boldsymbol{x}_0$, feasible set $\{\boldsymbol{x}_0\} + \text{null}(\boldsymbol{A})$ and ℓ_1-ball of radius $\|\boldsymbol{x}_0\|_1$. (b) There exists vector $\bar{\boldsymbol{x}} \neq \boldsymbol{x}_0$ with $\|\bar{\boldsymbol{x}}\|_1 < \|\boldsymbol{x}_0\|_1$. (c) Weighted ℓ_1-ball; there is no $\bar{\boldsymbol{x}} \neq \boldsymbol{x}_0$ with $\|\boldsymbol{W}\bar{\boldsymbol{x}}\|_1 \leq \|\boldsymbol{W}\boldsymbol{x}_0\|_1$.

where a small value ϵ is introduced to avoid division by zero. This IRℓ_1 scheme has been shown to correctly converge to the sparse solution vector in a relatively small number of iterations. Further, in comparison to the original ℓ_1-norm minimization, a smaller number of measurements m is required in general [99]. Moreover, any ℓ_1-norm minimization method (see Section 2.6) can be employed for the solution of Eq. (26).

In passing, note that the weights used in FOCUSS take the form $w_i^{(j+1)} = 1/x_i^{(j)}$, whereas the coefficients converging to zero are removed and constrained to be identically zero. Moreover, an IRLS scheme was proposed in [94], where the value of ϵ in the update formula of the weights (see Eq. (28)) is gradually reduced with increasing iteration number. This scheme, referred to as ϵ-regularized IRLS, exhibits similar performance to the IRℓ_1 in sparse vector recovery, and appears to outperform the standard IRLS in the sense that it requires significantly fewer measurements, especially as the value of p decreases and approaches 0 [94].

2.8.2. Bayesian CS approaches

This section presents the fundamental concepts of an alternative class of methodologies addressing the CS problem from a Bayesian perspective [100]. In this regard, the prior belief that $\boldsymbol{x}_0$ is sparse is expressed via an appropriately chosen probability density function (PDF), whereas the objective is to provide a posterior PDF for the values of the estimate $\boldsymbol{x}$ by utilizing a small number m of measurements $\boldsymbol{y}$, where $m < n$. The Bayesian CS approach exhibits two significant advantages over the standard CS techniques. First, in contrast to the deterministic estimates obtained for the sparse vector $\boldsymbol{x}$ in the traditional CS framework, Bayesian CS yields a posterior PDF. Clearly, this provides a tool for uncertainty quantification associated with the reconstructed vector $\boldsymbol{x}$. Second, instead of a priori selecting a fixed random matrix $\boldsymbol{\Phi}$ following standard CS (see Section 2.1), the posterior PDF can be employed for determining the CS matrix $\boldsymbol{\Phi}$ adaptively. This is achieved in an iterative manner by selecting at each cycle the next row of $\boldsymbol{\Phi}$ that minimizes the reconstruction uncertainty (see [100]).

In Bayesian CS, the residual vector $\boldsymbol{r}$ (see Eq. (12)) is modeled as a zero-mean Gaussian vector with covariance matrix $\sigma^2\boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix of size m. This choice yields a Gaussian likelihood function of the form

$$p(\boldsymbol{y}|\boldsymbol{x},\sigma^2) = (2\pi\sigma^2)^{-\frac{m}{2}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2\right) \tag{29}$$

Comparing with the standard CS, Eq. (29) corresponds to the first term of Eq. (19) and can be construed as a measure of the reconstruction accuracy for given $\boldsymbol{x}$ and σ^2. Next, a sparsity-promoting prior PDF is required for $\boldsymbol{x}$. A popular choice is the Laplace PDF [101] of the form

$$p(\boldsymbol{x}|\lambda) \propto \prod_{i=1}^{n} \frac{\lambda}{2}\exp\left(-\lambda|x_i|\right) = \left(\frac{\lambda}{2}\right)^n \exp\left(-\lambda\|\boldsymbol{x}\|_1\right) \tag{30}$$

where λ is the coefficient of the penalty factor in Eq. (19). It is noted that the Bayesian formulation corresponding to the standard CS problem of Eq. (19) aims at determining the maximum a posteriori (MAP) value of $\boldsymbol{x}$ by using the likelihood function of Eq. (29) in conjunction with the Laplace prior of Eq. (30). This naturally raises the question

of whether the Bayesian approach can be adapted for determining the complete posterior PDF $p(\boldsymbol{x}|\boldsymbol{y})$. Unfortunately, the Laplace prior in Eq. (30) is not conjugate to the Gaussian likelihood in Eq. (29), and thus, the Bayesian inference problem cannot be solved to yield the posterior PDF in closed-form. The interested reader is directed to [102] for more details about conjugacy in Bayesian inference.

In this regard, there have been efforts for addressing this issue within the context of sparse Bayesian learning [103] by introducing a technique typically referred to as relevance vector machine (RVM). Specifically, two distinct PDFs are utilized, i.e., a zero-mean Gaussian prior on each element of $\boldsymbol{x}$ of the form

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) \propto \prod_{i=1}^{n} \mathcal{N}(x_i|0,\alpha_i^{-1}) \tag{31}$$

and a Gamma prior for each element α_i of $\boldsymbol{\alpha}$ given by

$$p(\boldsymbol{\alpha}|\beta,\gamma) \propto \prod_{i=1}^{n} \Gamma(\alpha_i|\beta,\gamma) \tag{32}$$

In Eqs. (31) and (32), $\boldsymbol{\alpha}$ represents hyperparameters, whereas β and γ are parameters that need to be tuned. Hence, a marginalization over the hyperparameters $\boldsymbol{\alpha}$, yields the overall prior on $\boldsymbol{x}$ as

$$p(\boldsymbol{x}|\beta,\gamma) \propto \prod_{i=1}^{n} \int_0^{\infty} \mathcal{N}(x_i|0,\alpha_i^{-1})\Gamma(\alpha_i|\beta,\gamma)\mathrm{d}\alpha_i \tag{33}$$

Since the Gamma PDF $\Gamma(\alpha_i|\beta,\gamma)$ is the conjugate prior of the Gaussian PDF $\mathcal{N}(x_i|0,\alpha_i^{-1})$ with respect to α_i, the integrals appearing in the product of Eq. (33) can be evaluated in closed-form yielding the Student-t distribution [103]. The PDF of Eq. (33) is plotted in Fig. 8a, where it is seen that nonzero probability values are concentrated primarily around the origin and along the axes; thus, indicating that sparse vectors are more probable than dense vectors. The Laplace prior of Eq. (30), plotted in Fig. 8b, exhibits similar features. In contrast, a product of independent Gaussian random variables, plotted in Fig. 8c, does not exhibit probability concentration along the axes.

Several other sparsity inducing priors have been proposed in the sparse Bayesian learning literature. Indicatively, the spike-and-slab approach introduced in [104], where spike refers to the probability of a particular coefficient being zero and slab relates to the prior distribution of the coefficients, has been utilized for Bayesian variable selection [105] and for penalized likelihood estimation [106]. More recently, the horseshoe distribution was proposed in [107] (see also [108] for a review survey), which exhibits advanced performance in terms of robustness and adaptivity to different sparsity patterns, and is amenable to analytical mathematical treatment.

The hierarchical structure discussed so far leads eventually to a convenient representation of the complete posterior PDF $p(\boldsymbol{x}|\boldsymbol{y})$ as multivariate Gaussian with mean vector and covariance matrix given by

$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{A}^T\boldsymbol{y} \tag{34}$$

and

$$\boldsymbol{\Sigma} = \left(\sigma^{-2}\boldsymbol{A}^T\boldsymbol{A} + \mathrm{diag}(\alpha_1,\ldots,\alpha_n)\right)^{-1} \tag{35}$$

respectively. Therefore, the Bayesian CS formulation leads to the problem of estimating the hyperparameters σ and $\boldsymbol{\alpha} = [\alpha_1,\ldots,\alpha_n]^T$. This can be achieved with the aid of standard Bayesian tools such as Markov chain Monte Carlo [109] and variational inference [110]. Nevertheless, it can be argued that the standard solution approach in Bayesian CS is the RVM [103], which is a type II maximum-likelihood approach exhibiting both satisfactory accuracy and computational efficiency [100]. Specifically, the objective relates to estimating the values of $\boldsymbol{\alpha}$ and σ that maximize the logarithm of the marginal likelihood, where marginalization is performed over $\boldsymbol{x}$. This can be accomplished by implementing an expectation maximization (EM) algorithm, which leads to closed-form recursive formulas for the iterative solution of the unknown hyperparameters $\boldsymbol{\alpha}$ and σ (see [103] and [100] for more details). From a computational cost perspective, the evaluation of Eq. (35) involves the inversion of an $n \times n$ matrix, an operation of complexity $O(n^3)$. This limitation has been addressed in [111,112] by developing a fast RVM algorithm with complexity $O(nk^2)$.

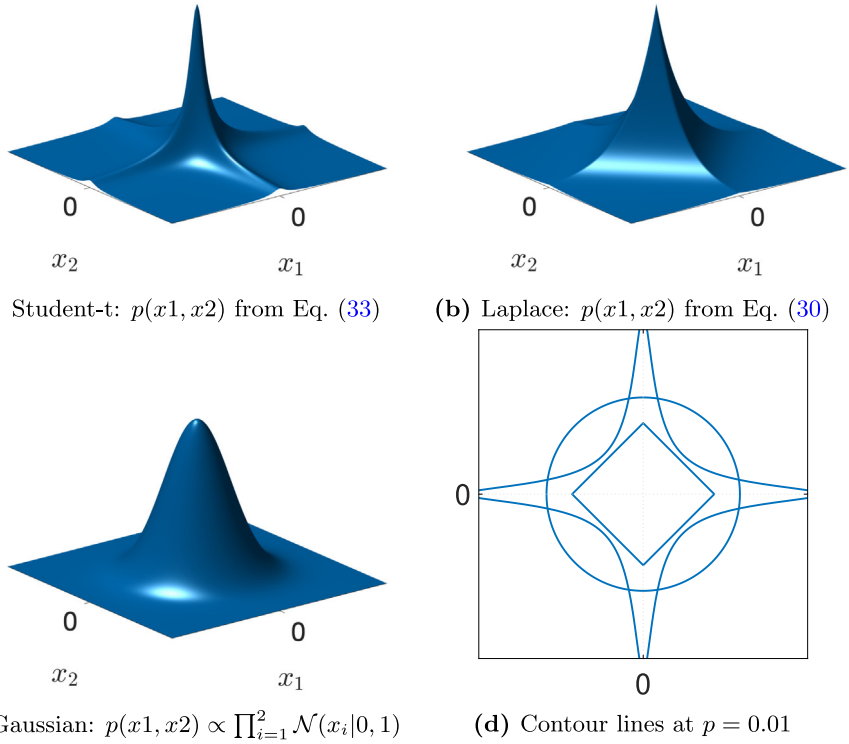
2.8.3. Structured sparsity

Solving the optimization problems of Eqs. (6) and (13) by employing the convex and non-convex approaches described in Sections 2.4–2.5, respectively, the position of each coefficient in the coefficient vector $\boldsymbol{x}_0$ is not taken into account. However, due to the physics of the specific problem, $\boldsymbol{x}_0$ may exhibit not only sparsity, but also additional patterns. This situation is referred to in the literature as structured sparsity [113]. In this regard, typical examples of structured sparsity include group sparsity (e.g., [114]), according to which the coefficients of $\boldsymbol{x}_0$ are clustered in disjoint or overlapping groups; hierarchical sparsity, according to which the coefficients are divided into parents and children that are jointly zero or nonzero (see, for instance, the wavelet tree sparsity in [115]); and, more generally, graph sparsity, according to which underlying relationships between coefficients are described by a graph structure with the aid of nodes, representing the coefficients, and edges, representing the relationships between them (e.g., [116]).

Notably, within the context of sparse reconstruction, structured sparsity serves as additional information to be exploited for further reducing the required number of measurements. In fact, it has been shown that modifying the regularization method leads to improvement in reconstruction accuracy for a given number of measurements (e.g., [113]). Approaches for exploiting structured sparsity include both convex (e.g., [114]) and non-convex (e.g., [116,117]) formulations with a varying degree of success depending on the type of information available.

An indicative example of a greedy, non-convex, approach proposed in [116] is StructOMP, which can be construed as a generalization of the OMP algorithm described in Section 2.4. In StructOMP, the input consists not only of the m-length measurement vector $\boldsymbol{y}$ and the $m \times n$ matrix $\boldsymbol{A}$, but also of the group structure that the coefficient vector is anticipated to exhibit. Specifically, a block set is defined that contains possible disjoint or overlapping groups, while each block is assigned a value that describes its complexity; i.e., generalizing the notion of sparsity to account for group structures (see [116] for more details). For example, in standard sparse vectors, each component of the coefficient vector is considered to have complexity 1, and thus, if a given coefficient is active, the coefficient vector will be less sparse by 1. In other words, every coefficient is equally penalized and encouraged to be zero. In group sparse vectors treated within the StructOMP framework, if a given block is active, the complexity of the overall coefficient vector increases by the complexity of that group, and thus, blocks are not evenly penalized. Next, similarly to the standard OMP, the algorithm, first, selects which block reduces $\|\boldsymbol{r}\|_2$ (see Eq. (12)) per unit increase of complexity the most (this block is considered to provide the maximum progress to the algorithm), and, second, assigns values to the coefficients of the selected block via least squares regression. Subsequently, the algorithm locates the next block corresponding to the maximum progress and terminates either when $\|\boldsymbol{r}\|_2$ becomes smaller than a prescribed threshold, or when the complexity of $\boldsymbol{x}$ becomes larger than a prescribed value. In general, StructOMP is straightforward to implement and can address cases of overlapping groups as well. Theoretical results regarding its performance (e.g., Structured-RIP; see also Section 2.7) can be found in [116].

Further, a notable convex approach is the elastic net, proposed in [118] in the context of grouped variable selection in regression analysis. Specifically, as argued in [118], the two most widely used regularization techniques, namely, ℓ_1-norm penalization (referred to as LASSO [64] in the statistical and as BPDN [63] in the signal processing communities) and ℓ_2-norm penalization (used for coefficient vector shrinkage and also known as ridge regression or Tikhonov regularization) are unable to perform variable selection, shrinkage



(a) Student-t: $p(x1, x2)$ from Eq. (33) **(b)** Laplace: $p(x1, x2)$ from Eq. (30)

(c) Gaussian: $p(x1, x2) \propto \prod_{i=1}^{2} \mathcal{N}(x_i|0, 1)$ **(d)** Contour lines at $p = 0.01$

Fig. 8. Various probability density functions that may be employed as priors. (a), (b) Nonzero probability values are primarily concentrated around the origin and along the axes encouraging sparse solutions. (c) There is no probability density concentration along the axes. (d) Contour lines of the three density functions at PDF value $p = 0.01$.

and variable grouping, simultaneously. Thus, a rather straightforward approach relates to combining ℓ_1-norm (that promotes sparsity and leads to shrinkage) with ℓ_2-norm (that promotes grouping and leads to shrinkage) penalizations, yielding the minimization problem

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda_1 \|\boldsymbol{x}\|_1 + \lambda_2 \|\boldsymbol{x}\|_2^2 \tag{36}$$

The formulation of Eq. (36), referred to as naive elastic net in the original paper [118], can be recast as a LASSO problem (see Eq. (19)); thus, a solution estimate $\hat{\boldsymbol{x}}$ can be obtained by using the LARS-EN algorithm, which is a modified version of LARS (see also Section 2.6). However, Eq. (36) leads to an undesirably high degree of shrinkage due to the combined shrinkage effects of the ℓ_1- and ℓ_2-norms. Therefore, scaling of the solution estimate is typically applied in the form

$$\boldsymbol{x} = (1 + \lambda_2)\hat{\boldsymbol{x}} \tag{37}$$

where $\boldsymbol{x}$ is known as the elastic net solution estimate. Overall, elastic net exhibits the significant advantage of successfully promoting group sparsity, even in cases where no information about group structures in $\boldsymbol{x}_0$ is available; see also Fig. 9 for a visual representation of the elastic net unit-norm ball in $\mathbb{R}^3$.

Next, another widely used regularization technique that promotes group sparsity is the ℓ_1/ℓ_p penalty, which encourages sparse solutions at the group level, but not within the groups (e.g., [114]). In this regard, the minimization problem becomes

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \sum_{g \in \mathcal{G}} d_g \left\|\boldsymbol{x}_g\right\|_p^p \tag{38}$$

where $\boldsymbol{x}_g$ represents the coefficients of $\boldsymbol{x}$ that belong to group $g \in \mathcal{G}$, with $\mathcal{G}$ being the set of all groups, and d_g represent positive scalar weights. The approach described by Eq. (38) is referred to in the literature as group-LASSO, with typical p values being 2 and ∞ [113]. Indicatively, in Fig. 9 the three-dimensional ℓ_1/ℓ_2-norm ball is shown and compared with the elastic net unit-norm ball. Group-LASSO has been shown to improve reconstruction performance as compared to standard LASSO [119]. This is under the condition that there is a priori knowledge of coefficients $\boldsymbol{x}_0$ forming disjoint groups, with either simultaneously active or simultaneously inactive coefficients. Further, to account for the impact of uneven group sizes, rather sophisticated approaches exist for appropriately selecting the weights d_g (see, for instance, [120]). Clearly, if groups in $\mathcal{G}$ are allowed to overlap, more complex coefficient structures can be formed, such as hierarchical and graph structures [113]. In fact, a direct extension of group-LASSO relates to considering groups in $\mathcal{G}$ defined as intersections of complements of overlapping groups [121]. An alternative approach, commonly referred to as latent group-LASSO [122], considers groups in $\mathcal{G}$ defined as unions of overlapping groups. Interestingly, the latter approach can be construed as a convex relaxation of StructOMP; see [123] for more details and comparisons.

2.8.4. Dictionary learning strategies

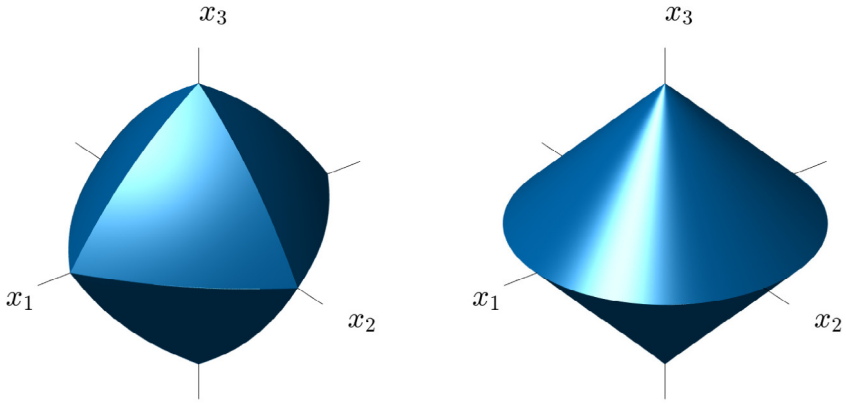
This section focuses on approaches addressing the problem of determining an optimal matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ in Eq. (5) based on a training set of available signals $\{\boldsymbol{y}_i\}_{i=1}^N$. These approaches are collectively referred to in the literature as dictionary learning and seek for a proper basis $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ (also termed overcomplete dictionary), promoting the sparse representation of signals with similar characteristics to the training set.

In the following, $\boldsymbol{Y} \in \mathbb{R}^{m \times N}$ denotes the matrix with the training vectors $\{\boldsymbol{y}_i\}_{i=1}^N$ as its columns, and $\boldsymbol{X} \in \mathbb{R}^{n \times N}$ represents the matrix with the corresponding representation vectors $\{\boldsymbol{x}_i\}_{i=1}^N$ as its vectors, where $\boldsymbol{y}_i = \boldsymbol{A}\boldsymbol{x}_i$ for all $i = 1, \ldots, N$. Also, the columns of matrix $\boldsymbol{A}$ are referred to herein as dictionary atoms, following the established terminology in the literature. In this regard, the associated optimization problem can be formulated quite generally in the form

$$\min_{\boldsymbol{A},\boldsymbol{X}} \ \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_F^2 \quad \text{subject to} \quad \|\boldsymbol{x}_i\|_0 \le k_0 \ \ \forall\, i = 1, \ldots, N \tag{39}$$

where $\|.\|_F$ denotes the Frobenius norm, and k_0 is an integer denoting a prespecified target sparsity degree. Eq. (39) can be equivalently cast in the form

$$\min_{\boldsymbol{A},\boldsymbol{X}} \sum_{i=1}^{N} \|\boldsymbol{x}_i\|_0 \quad \text{subject to} \ \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_F^2 \le \epsilon \tag{40}$$



(a) Elastic net unit-norm ball **(b)** ℓ_1/ℓ_2-norm ball

Fig. 9. Comparison between the elastic net unit-norm ball (a) and the ℓ_1/ℓ_2-norm ball (b) in $\mathbb{R}^3$. In the elastic net ball, curved contours encourage grouping of coefficients, whereas sharp edges and vertices promote sparsity. In the ℓ_1/ℓ_2-norm ball, sparsity is promoted between the groups $g_1 = \{x_1, x_2\}$ and $g_2 = \{x_3\}$, however, no particular direction is encouraged within the groups.

for a fixed value ϵ. Next, to address the non-convexity of the ℓ_0-norm, an ℓ_1-norm relaxation can be introduced yielding an objective function of the form

$$\min_{\boldsymbol{A},\boldsymbol{X}} \sum_{i=1}^{N} \frac{1}{2} \left\| \boldsymbol{y}_i - \boldsymbol{A}\boldsymbol{x}_i \right\|_2^2 + \lambda \left\| \boldsymbol{x}_i \right\|_1 \tag{41}$$

Although the objective function in Eq. (41) is not jointly convex with respect to variables $\boldsymbol{A}$ and $\boldsymbol{X}$, it becomes convex with respect to one variable when the other is kept fixed [124]. This motivates a two-step iterative solution approach, adopted by the vast majority of researchers in dictionary learning. The approach entails a sparse coding step, i.e., the determination of the representation vectors in $\boldsymbol{X}$, followed by a dictionary update step for $\boldsymbol{A}$. A closer examination of Eq. (41) shows that the ℓ_1-norm constraint on $\boldsymbol{x}_i$ tends to reduce the values of the nonzero elements of $\boldsymbol{x}_i$, which forces the elements of $\boldsymbol{A}$ to increase arbitrarily in the dictionary update step. This undesirable effect can be meliorated by restraining the columns of $\boldsymbol{A}$ to have ℓ_2-norms less or equal to one.

Dictionary learning approaches have also been developed within a Bayesian framework (see Section 2.8.2). Indicatively, according to [125] the dictionary $\boldsymbol{A}$ is determined based on maximization of the likelihood PDF $p(\boldsymbol{Y}|\boldsymbol{A})$. Two fundamental assumptions are introduced in [125]. First, independence is assumed between the training samples, and second, the prior PDF $p(\boldsymbol{X})$ is chosen in a manner that the elements of each representation vector $\boldsymbol{x}_i$ are zero-mean independent identically distributed random variables following the Laplace distribution. These assumptions lead to a similar to Eq. (41) formulation, which can be solved efficiently by employing a steepest descent approach for the sparse coding step and a closed-form formula for the dictionary update step (see also [126–128]). Furthermore, the method of optimal directions [129–131] can be construed as a modification of the aforementioned approach, which provides a simpler dictionary update formula and allows for the adoption of more sophisticated techniques (e.g., OMP or FOCUSS) for the sparse coding step; see also [132–134] for an alternative related approach, which relies on a MAP setting. In general, the dictionary learning techniques available in the literature utilize various different solution approaches for the dictionary update and/or the sparse coding steps. For instance, the OMP [135] (see Section 2.4) can be utilized for the problems in Eq. (39) or Eq. (40), or the ISTA [136,137] and the LARS [124] (see Section 2.6) can be employed for the problem in Eq. (41).

Early work in dictionary learning has also been inspired by vector quantization (VQ) clustering [135,138]. In VQ clustering, a set of descriptive vectors $\{\boldsymbol{d}_j\}_{j=1}^{K}$ is learned and each training sample is represented by one of those vectors, typically the closest one in an ℓ_2-norm sense. This approach can be construed as an extreme case of sparse representation, where only one atom of dictionary $\boldsymbol{A}$ is selected for representing $\boldsymbol{y}$; thus, yielding a 1−sparse representation vector $\boldsymbol{x}$. Note that in the general sparse representation framework discussed so far, each signal is represented as a linear combination of more than one atoms of $\boldsymbol{A}$. Further, the K-means algorithm, also referred to as the generalized Lloyd algorithm [139], is routinely utilized in the VQ training procedure. This motivates its use for addressing the dictionary learning problem. Interestingly, the K-means algorithm is a two step procedure, which, first, determines the 1−sparse representation vectors, and second, updates the dictionary (or codebook in VQ terminology), in a similar manner as the two step dictionary update framework discussed previously. In this regard, a dictionary learning technique referred to as K-SVD was presented in [135], where the sparse coding step is performed by employing OMP (this choice is not restrictive), and the dictionary update is performed by sequentially updating each column of $\boldsymbol{A}$ based on singular value decomposition (SVD) to minimize the approximation error. Although it is not guaranteed to converge, and its convergence performance depends on the robustness of the adopted sparse coding algorithm [135], K-SVD has exhibited highly satisfactory accuracy in various applications such as image denoising [138].

The dictionary learning approaches discussed so far can be categorized as "batch" algorithms in the sense that the complete training set is provided as an input and the ensemble of the training samples is processed at each iteration. Clearly, this affects the computational efficiency of these algorithms, especially in applications involving large training sets. To address this challenge, an online dictionary learning algorithm was presented in [124,140], which processes the training signals one at a time, or in mini-batches. This algorithm utilizes a LARS solution approach for the problem in Eq. (41) regarding the sparse coding step, and a block-coordinate descent method with warm restarts for the dictionary update step; see also [141]. Compared with the standard batch algorithms, this online dictionary learning technique has been shown to exhibit enhanced performance for both large and small training sets [124]. Moreover, under certain rather strong conditions, convergence to a stationary point is guaranteed [124]. In passing, it is noted that all dictionary learning algorithms address a non-convex problem, and thus, are susceptible to being trapped in local minima [124,135,138].

3. Diverse applications in engineering mechanics

Sparse representations and CS approaches have impacted significantly the field of engineering mechanics over the past few years. In this

section, relevant research work is categorized under three distinct application areas, whereas a concerted effort is made to highlight the links and interconnections between the theoretical concepts presented in Section 2 and the specific engineering mechanics applications discussed below.

The first application area relates to inverse problems in the field of structural health monitoring, and specifically to the development of techniques for structural system identification and damage detection subject to incomplete data. In fact, applications related to efficient data compression and storage at the sensors level, and to fast data transmission, have proved quite advantageous for real-time structural health monitoring. Also, exploitation of the inherently sparse data structure of vibration response measurements has benefited the development of efficacious system identification and damage detection schemes.

The second application area relates to uncertainty modeling and simulation under incomplete data. In particular, CS-based techniques have been developed within the context of stochastic processes to address problems in engineering mechanics related to spectral analysis, statistics estimation and Monte Carlo simulation under sparse measurements.

The third application area relates to developing computationally efficient uncertainty propagation techniques for determining the response statistics of diverse systems in engineering mechanics. The rationale relates to employing CS tools for evaluating the system response, which is represented by appropriately chosen sparse expansions. In this manner, the associated computational cost is reduced, and thus, the solution technique can be applied to higher-dimensional problems.

It is interesting to note that although the aforementioned application areas appear relatively unrelated to each other, the theoretical concepts and mathematical tools utilized (and described in Section 2) are surprisingly similar in their implementation. This is due to the fact that problems in all three areas share the challenge of incomplete data. Of course, incomplete data may manifest themselves in various different forms and can correspond to missing or compressed data, or even refer generally to insufficiently few function evaluations. Ultimately, however, in all herein considered applications, the mathematical formulation yields an undetermined linear system of the form of Eq. (5), which can be addressed by the versatile CS machinery discussed in Section 2.

3.1. Inverse problems in structural health monitoring: Structural system identification and damage detection under incomplete data

One of the first applications of sparse representations and CS theory in the field of structural health monitoring has been the analysis of sparse and/or incomplete data acquired by diverse sensor technologies. In this regard, many applications have focused on developing efficient data compression schemes for real-time structural health monitoring. The rationale relates to acquiring the signal directly in a compressed format. Clearly, this circumvents the computational burden of compressing it locally at the sensor and bypasses the need for sensors with high storage capacity. This entails the utilization of CS in conjunction with an appropriate compression basis (in which the signal has a sparse representation) for reconstructing data series with far higher resolution than those originally captured. Notably, in the problem of data compression the CS efficiency can be optimized by appropriately designing the sampling matrix $\boldsymbol{\Phi}$, or, equivalently, matrix $\boldsymbol{A}$ in Eq. (5). However, this is not the case when the problem of limited and/or missing data is considered. Indicatively, practical reasons for the occurrence of limited data include data loss due to both equipment failure (e.g., damaged sensors) and sensor thresholding limitations. Numerous other issues including sensor maintenance, bandwidth limitations, usage and data acquisition restrictions, as well as data corruption may also lead to missing data. It becomes clear that applying CS theory to the problem of missing data for signal reconstruction differs primarily in one respect as compared to data compression; that is, missing data are not necessarily intentional. Obviously, this removes control over one important step of compressive sampling, i.e., the arrangement of the sampling matrix $\boldsymbol{\Phi}$. Indeed, as mentioned in Section 2.2, a number of bases with randomly deleted rows, such as Fourier, satisfy the requirements of Eqs. (7) and (10) for sparse reconstruction with high probability. Unfortunately, the missing data may not be uniformly distributed over the record; thus, regular or large gaps of missing data can lead to matrices $\boldsymbol{A}$ with less incoherent basis vectors. Clearly, this additional challenge highlights the need for assessing the performance of the various CS tools in conjunction with matrices $\boldsymbol{A}$ that do not (strictly) conform to theoretical conditions such as the RIP (see also Section 2.7).

One of the first CS applications for data compression can be found in [142], where the authors utilized bridge vibration data and employed orthogonal expansion bases (e.g., Fourier and wavelets) in conjunction with an ℓ_1-norm minimization formulation. The approach was subsequently applied for signal reconstruction related to the problem of data loss in a wireless sensor network during transmission of data between the wireless sensor nodes and the base station [143]. In a similar context, CS was employed in [144] for data loss recovery associated with a fast-moving wireless sensing technique for structural health monitoring of bridges without interrupting traffic.

Further, following pioneering contributions in signal processing (see Section 2.8.2), a Bayesian CS methodology was proposed in [145], which, in contrast to the standard approaches delineated in Sections 2.3–2.6, provides also with an estimate of the signal reconstruction uncertainty. Specifically, this Bayesian treatment yields posterior distributions $p(\boldsymbol{x}|\boldsymbol{y})$ for the basis coefficients of Eq. (5), which can be used eventually for suppressing the basis terms whose contribution to the reconstructed signal is minimal. The methodology was further enhanced and its reconstruction robustness was improved in [146], where its performance was assessed with regard to recovery of lost data during wireless transmission.

Next, by proposing a matrix reshape scheme, a low-rank representation of large-scale structural seismic and typhoon responses was identified in [147], which proved to be beneficial for efficient data compression. The scheme was coupled in [148] with a nuclear norm minimization algorithm for recovering of multi-channel structural response time-histories with randomly missing data. The same authors exploited CS tools in [149] for efficient transmission and recovery of large-scale image data related to structural system and civil infrastructure health diagnosis. Furthermore, the relatively recently proposed concept of group sparsity (see also Section 2.8.3) was employed in [150] for reconstructing incomplete vibration data measured by sensors placed at various different locations of the structure, while in [151] a dictionary learning strategy (see also Section 2.8.4) was proposed for under-sampled acoustic emission signal reconstruction.

Finally, it is worth mentioning that Refs. [152,153] focus on practical implementation of CS algorithms in wireless sensor networks, and provide relevant discussions about optimal configurations and energy efficiency aspects. Moreover, a hybrid sensor network configuration was proposed in [154] (see also [155]), based on fusion of a minimal number of tethered sensors with wireless nodes, for improving the information content of the transmitted data.

In the remainder of the section, attention is directed to system identification and damage detection methodologies, which exploit the capabilities of the CS machinery. In this regard, the work in [156] constitutes one of the first research efforts to employ CS-based data analysis for estimating the damage condition of a structure. In [157], a CS-based scheme was proposed and applied for determining the degradation of a pipe-soil interaction model, where the damage identification task was treated as a pattern classification problem. In a relatively different context, a standard ℓ_1-norm optimization approach was proposed in [158] for identifying the distribution of moving vehicle loads on cable-stayed bridges. Further, a scheme was devised in [159] based on a combination of blind feature extraction and sparse representation classification, in conjunction with a modal analysis treatment, for locating the structural damage and assessing its severity. The same authors

proposed an output-only identification approach in [160] by coupling CS with blind source separation schemes for determining the mode shape matrix of the structural model. In [161] the approach was modified to account for video camera based vibration measurements. Along similar lines, in [162] the mode shapes of a multi-degree-of-freedom (MDOF) structural system were identified based on under-sampled vibration data collected by wireless sensors; see also [163] for a formulation of the mode shape identification problem based on atomic norm minimization.

In [164,165] a sensitivity-based model updating scheme in conjunction with ℓ_1-norm minimization was proposed for identifying localized damage in structures based on incomplete modal information; see also [166,167] for some related work. In this context, several authors highlighted the limitations of employing a Tikhonov regularization strategy (see also Section 2.8.3), typically used in sensitivity-based model updating, for addressing the resulting underdetermined problem. In particular, to address issues related to over-smoothing resulting from Tikhonov regularization and to promote the sparseness of the damage identification problem, various ℓ_1-norm regularization schemes were proposed in [168–173]. In [174] the authors combined CS for signal reconstruction with auto-regressive and Wiener filter based methods for structural damage detection and localization, while in [175] the ill-posedness of the inverse damage identification problem was addressed by adding an ℓ_1-norm regularization term in the objective function.

Furthermore, in [176,177] a spectral identification technique was developed for determining the parameters of nonlinear and time-variant structural systems based on available input–output (excitation-response) realizations. A significant advantage of the technique relates to the fact that it can readily account for the presence of fractional derivative terms in the system governing equations, as well as for the cases of non-stationary, incomplete and/or noise-corrupted data. Specifically, the technique relies on recasting the governing equations as a set of multiple-input-multiple-output systems in the wavelet domain. Next, an ℓ_1-norm minimization procedure based on CS theory is employed for determining the wavelet coefficients of the available incomplete non-stationary input–output data. Finally, these wavelet coefficients are utilized to reconstruct the non-stationary incomplete signals, and consequently, to determine system related time- and frequency-dependent wavelet-based frequency response functions and associated parameters. The technique can be construed as a generalization of the multiple-input-single-output methodology pioneered by Bendat and co-workers (e.g., [178]) to account for non-stationary and incomplete data, as well as for fractional derivative modeling.

Moreover, in [179,180] a power spectrum blind multi-coset sampling approach was proposed for operational modal analysis applications involving wireless sensor networks. In comparison with a CS treatment, the performance of the multi-coset sampling approach appeared rather insensitive to the signal sparsity degree. In [181] a dictionary learning approach (see also Section 2.8.4) was applied for nonlinear structural system identification and for determining the underlying governing equations based on available input–output data; see also [182,183] for indicative applications of sparsity-based algorithms utilizing dictionaries in damage detection problems. It is worth mentioning that CS concepts have also been used for structural system impact force identification. Indicatively, in [184] a hybrid ℓ_1/ℓ_2-norm minimization approach for promoting group sparsity was proposed (see Section 2.8.3); see also [185,186] for some relevant references.

3.2. Uncertainty modeling and simulation under incomplete data

CS-based techniques have also been developed within the context of stochastic processes to address problems in stochastic engineering mechanics related to spectral analysis, statistics estimation and Monte Carlo simulation under incomplete available data. Specifically, Kougioumtzoglou and co-workers relied on CS theory for stationary and non-stationary stochastic process power spectrum estimation subject to missing data [187]. This was done in conjunction with an ℓ_1-norm optimization algorithm for obtaining a sparse representation of the signal in the selected basis (i.e., Fourier or wavelets). Notably, the underlying stochastic process power spectrum can be estimated in a direct manner by utilizing the determined expansion coefficients; thus, circumventing the computational cost related to reconstructing the signal in the time domain.

The technique was enhanced in [188] by utilizing an adaptive basis re-weighting scheme for increasing further the sparsity of the solution (see also Section 2.8.1), and was applied in [189] for structural system response and reliability analysis under missing data. The rationale relates to applying CS to multiple process records iteratively, and to utilizing the cumulative information from all records for the purpose of seeking a sparse representation in an average sense over an ensemble. By introducing this iterative process to alter basis coefficients, a significant gain in spectral estimation accuracy was observed as compared to standard CS. In a similar context, an ℓ_p-norm $(0 < p < 1)$ optimization algorithm was proposed in [190] for promoting solution sparsity (see also Section 2.8.1). Regarding the effect of the chosen norm on the power spectrum estimation error, it was shown that the $\ell_{1/2}$-norm provides almost always a sparser solution than the ℓ_1-norm. This was corroborated by various examples considering stationary, non-stationary and two-dimensional processes related to sea wave, wind, and material properties spectra, respectively. It was also observed that the reconstruction accuracy of the technique is further enhanced when coupled with the aforementioned adaptive basis re-weighting scheme.

The above developments have found recently diverse applications in marine engineering. Indicatively, a methodology based on $\ell_{1/2}$-norm minimization (see Section 2.8.1) was proposed for efficient processing and joint time–frequency analysis of relatively long water wave records by enabling reconstruction of data recorded at a very low (sub-Nyquist) sampling rate [191]. Further, a CS technique relying on adaptive basis re-weighting (see Section 2.8.1 and [188]) was developed in [192] for extrapolating in the spatial domain and estimating the space–time characteristics of a sea state based on data collected at very few spatially sparse points (e.g., wave buoys). This is of considerable importance to a number of marine engineering applications involving three-dimensional waves interacting with marine structures, such as optimizing arrays of wave energy converters. Furthermore, a novel approach for measuring the sea surface elevation on vertical breakwaters was developed in [193]. Note that this is not a trivial problem since alternative typically used approaches, such as ultrasonic probes and image processing, exhibit limitations related to signal distortion and high computational cost, respectively. In this regard, the authors in [193] relied on pressure measurements and on a CS-based reconstruction algorithm in conjunction with a generalized harmonic wavelet basis. Specifically, the proposed approach leads to an ℓ_1-norm based constrained optimization scheme, which utilizes the known values of the free surface data to reconstruct all other missing data while adhering at the same time to prescribed upper and lower bounds at all time instants. The approach was also used in [194] as a validation tool for supporting the veracity of the analytically derived probability distribution of the nonlinear wave crest height on a vertical breakwater.

In [195] a Bayesian CS approach (see also Section 2.8.2) was proposed for estimating profiles of soil properties based on sparse measurement data. The approach is capable of quantifying the uncertainty of the statistical estimates as well, while its performance was assessed in [196] against alternative widely used techniques for interpolation of spatially varying and sparsely measured geo-data. From a random field simulation perspective, the approach was coupled in [197] with a Karhunen–Loève expansion for generating random field samples within a Monte Carlo simulation context. It was further generalized in [198] for simulation of cross-correlated random fields in the spatial domain, and in [199] to account for non-stationary and non-Gaussian random fields. Also, it was shown in [200] that the approach can be employed for random field simulation without the need for "detrending" first, while a bootstrap approach for statistical inference of random field auto-correlation structure was proposed in [201] based on a combination of Bayesian CS and Karhunen–Loève expansion.

3.3. Computationally efficient uncertainty propagation in engineering mechanics

Addressing the challenge of uncertainty propagation in engineering mechanics relates to the development of analytical and numerical methodologies for stochastic response analysis of engineering systems. Specifically, ever-increasing computational capabilities, novel signal processing techniques, and advanced experimental setups have contributed to a highly sophisticated mathematical modeling of the system governing equations. In general, these take the form of high-dimensional stochastic (partial/fractional) differential equations to be solved for evaluating the system response statistics; see also [202,203] for a broad perspective. In this regard, a wide range of solution techniques rely on appropriate (stochastic) representations and expansions of the system response quantities of interest (e.g., displacements, stresses, etc.), where the objective is to determine the expansion coefficients accurately and in a computationally efficient manner. Recently, the potential sparsity of such expansions has been exploited and CS-based strategies have been proposed for reducing the associated computational cost and for extending the range of applicability of the techniques to problems of higher dimensions.

In this context, a rather popular solution technique in stochastic mechanics relates to the use of polynomial chaos expansions (e.g., [204–206]). This entails the expansion of the system response quantity on a basis of (multivariate) polynomials that are orthogonal with respect to the joint PDF of the input. Recently, polynomial chaos expansions have been coupled with CS concepts and tools for efficient representation and determination of the system response (e.g., [44]). This has been motivated not only by theoretical results showing that multivariate functions possess sparse expansions in orthogonal polynomial bases (e.g., [207]), but also by the typically observed structured sparsity (see also Section 2.8.3) in the polynomial chaos expansions of various problems; that is, coefficients corresponding to low polynomial orders tend to be larger than coefficients corresponding to higher orders.

One of the first research efforts to explore the sparsity-promoting properties of the ℓ_1-norm, in conjunction with a LARS algorithm (see Section 2.6) for automatically detecting the significant coefficients of the polynomial chaos expansion, can be found in [208]. Further, in [209] the polynomial chaos expansion was combined with standard CS for efficiently constructing a solution representation of elliptic stochastic partial differential equations. Applications of the technique to address diverse problems in the fields of molecular biology, astrodynamics, and computational fluid dynamics can be found in [210,211], and [212], respectively.

Following the aforementioned relatively standard implementation of the CS approach, a weighting scheme was proposed in [213] for further promoting sparsity in the recovery of the expansion coefficients (see also Section 2.8.1), while an adaptive re-weighting ℓ_1-norm minimization scheme was applied in [214] for the solution of stochastic partial differential equations. Note that in several cases, such as in [215], the construction of the weighting matrix $\boldsymbol{W}$ in Eq. (26) can be based on a priori information and on theoretical results about the decay of the polynomial chaos coefficients. Moreover, additional information in the form of response derivative estimates may be available. In this context, gradient-enhanced ℓ_1-norm minimization schemes were proposed in [216,217] for accelerating the determination of the polynomial coefficients. Further, it is worth mentioning that alternative optimization algorithms based on ℓ_p-norm, $p < 1$, [218,219] and on ℓ_{1-2}-norm [220] (see Section 2.8.1) were also employed for increasing the sparsity of the obtained polynomial chaos expansion coefficient vector (see also Section 2.8.1).

More recently, dictionary learning approaches (see Section 2.8.4) and iterative basis updating schemes were proposed for increasing the approximation accuracy and for decreasing the required number of expansion coefficients. In this regard, anisotropic basis sets with more terms in important dimensions were constructed in [221] in an adaptive manner, while an incremental algorithm was employed in [222] for promoting sparsity by exploring sub-dimensional expansions. Also, by resorting to CS concepts, a basis adaptation technique was developed in [223] yielding a sparse polynomial chaos expansion. Specifically, a two-step optimization algorithm was devised, which calculates the coefficients and the input projection matrix of a low dimensional polynomial chaos expansion with respect to an optimally rotated basis. Further, to reduce the number of samples necessary for recovering the expansion coefficients, importance sampling and coherence-optimal sampling strategies were developed in [224], and applied in [225] in conjunction with adaptive global bases; see also [226,227] for relevant work. It is worth mentioning that Bayesian CS (see also Section 2.8.2) has also been used in conjunction with polynomial chaos expansions for basis selection and uncertainty quantification regarding the basis significance (e.g., [228,229]).

Of course, polynomial chaos expansions are not the only response representations that have been employed in conjunction with CS strategies for efficient uncertainty propagation. Indicatively, sparse wavelet-based expansions were employed in [230], and were coupled with importance sampling schemes for determining the expansion coefficients in an efficient manner. In [231], a problem-dependent basis in conjunction with a Karhunen–Loève representation was proposed for enhancing the sparsity of the coefficient vector. In a relatively different context, CS was applied in [232] for the computationally efficient calculation of high-dimensional integrals arising in quantum mechanics. In particular, by interpreting the integrand as a tensor in a suitable tensor product space, its entries were determined by utilizing an ℓ_1-norm minimization in conjunction with few only function evaluations. Next, by employing a rank reduction strategy, the high-dimensional integrand was cast in the form of a sum of low dimensional functions to be integrated by a standard Gauss–Hermite quadrature rule.

Further, Kougioumtzoglou and coworkers have recently adapted, extended, and applied the Wiener path integral methodology, which originates from theoretical physics (e.g., [233–235]), for the stochastic response analysis and optimization of diverse engineering dynamical systems (e.g., [236–242]). Specifically, it has been shown that the joint response transition PDF of stochastically excited dynamical systems can be expressed exactly as a functional integral over all possible paths that the response process may follow [236,237]. Notably, a diverse class of problems, such as systems endowed with fractional derivative terms [238] or characterized by singular diffusion matrices [243], structures exhibiting various nonlinear behaviors [240], as well as systems subject to non-white, non-Gaussian and non-stationary excitation processes [244], can be readily addressed by the versatile Wiener path integral formalism.

Nevertheless, the analytical evaluation of the path integral is, in general, a highly challenging task, and thus, approximate solution techniques are typically required. In this regard, the standard approach, which is referred to in the theoretical physics literature as the semi-classical approximation, relates to accounting in the path integral only for the path associated with the maximum probability of occurrence (also known as the most probable path). Therefore, evaluating the path integral degenerates to obtaining the most probable path and to determining its probability. However, obtaining analytically in explicit form the (dependent on boundary conditions) most probable path is generally impossible. Thus, a variational problem is solved numerically for determining a specific point of the joint response PDF. Accordingly, for an M-DOF system corresponding to $2M$ stochastic dimensions (M displacements and M velocities), and discretizing the effective PDF domain using N points in each dimension, the number of required "measurements" (i.e., number of boundary value problems to be solved numerically) becomes N^{2M}. Clearly, this demonstrates the high computational cost related to a brute-force implementation. However, it has been shown recently that this "data acquisition" process can be coupled with versatile expansion schemes, compressive sampling techniques and group sparsity concepts. Specifically, by utilizing (time-variant) sparse representations of the response PDF (e.g., monomial or

wavelet bases) and by exploiting the group structure of the expansion coefficients (see Section 2.8.3), the response PDF of relatively high-dimensional nonlinear systems can be determined in a computationally efficient manner [239,245,246].

4. Concluding remarks

A review of CS theoretical concepts and numerical tools in conjunction with diverse applications in engineering mechanics has been attempted from a broad perspective. In this regard, a concerted effort has been made to highlight the links and interconnections between the CS theory and algorithms presented in Section 2 and the plethora of applications in engineering mechanics discussed in Section 3. Hopefully, the extensive list of readily available references can serve as a compass for navigating the interested researcher though the multitude of CS concepts and applications, even beyond the scope of this paper.

It is anticipated that the currently rapid progress in data science and machine learning will facilitate further the development, enhancement and application of CS-based techniques in engineering mechanics. Indicatively, the work in [247] constitutes an interesting effort towards this direction, where deep learning is employed in conjunction with sparse regression and ℓ_1-norm minimization for simultaneous reduced-order modeling and data-driven identification of the governing equations of the dynamical system. Such approaches may prove in the near future indispensable for the materialization and practical implementation of emerging concepts in data-driven uncertainty quantification and health monitoring of diverse engineering systems and structures; see, for instance, the concept of a digital twin mirroring the physical system and tracking its temporal evolution (e.g., [248]).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

I. A. Kougioumtzoglou gratefully acknowledges the support by the CMMI Division of the National Science Foundation, USA (Award number: 1724930).

References

[1] J. Fourier, Theorie Analytique De La Chaleur, Par M. Fourier, Chez Firmin Didot, père et fils, 1822.
[2] J.G. Proakis, D.G. Manolakis, Introduction to Digital Signal Processing, Prentice Hall Professional Technical Reference, 1988.
[3] Y.C. Eldar, G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.
[4] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Birkhäuser Basel, 2013.
[5] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer Science & Business Media, 2010.
[6] I. Rish, G. Grabarnik, Sparse Modeling: Theory, Algorithms, and Applications, CRC Press, 2014.
[7] C.E. Shannon, Communication in the presence of noise, Proc. IRE 37 (1949) 10–21.
[8] C. Carathéodory, Über den variabilitätsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehmen, Math. Ann. 64 (1907) 95–115.
[9] C. Carathéodory, Über den variabilitätsbereich Der Fourier'schen konstanten von positiven harmonischen funktionen, Rend. Circ. Mat. Palermo (1884-1940) 32 (1911) 193–217.
[10] A. Beurling, Sur les Intégrales de Fourier absolument convergentes et leur application à une transformation fonctionelle, in: Ninth Scandinavian Mathematical Congress, 1938, pp. 345–366.
[11] R. Dorfman, The detection of defective members of large populations, Ann. Math. Stat. 14 (1943) 436–440.
[12] B. Logan, Properties of High-Pass Functions (Doctoral Thesis), Electrical Engineering Department, Columbia University, New York, 1965.
[13] H.L. Taylor, S.C. Banks, J.F. McCoy, Deconvolution with the $\ell 1$ norm, Geophysics 44 (1979) 39–52.
[14] S. Levy, P.K. Fullagar, Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution, Geophysics 46 (1981) 1235–1243.
[15] C. Walker, T.J. Ulrych, Autoregressive recovery of the acoustic impedance, Geophysics 48 (1983) 1338–1350.
[16] F.J. Herrmann, M.P. Friedlander, O. Yilmaz, Fighting the curse of dimensionality: compressive sensing in exploration seismology, IEEE Signal Process. Mag. 29 (2012) 88–100.
[17] E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59 (2006) 1207–1223.
[18] D.L. Donoho, et al., Compressed sensing, IEEE Trans. Inform. Theory 52 (2006) 1289–1306.
[19] M. Rudelson, R. Vershynin, On sparse reconstruction from fourier and Gaussian measurements, Comm. Pure Appl. Math. 61 (2008) 1025–1045.
[20] P. Bühlmann, S. Van De Geer, Statistics for High-Dimensional Data: Methods, Theory and Applications, Springer Science & Business Media, 2011.
[21] R. Tibshirani, M. Wainwright, T. Hastie, Statistical Learning with Sparsity: The Lasso and Generalizations, Chapman and Hall/CRC, 2015.
[22] J.-L. Starck, F. Murtagh, J. Fadili, Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis, Cambridge university press, 2015.
[23] Q. Zhang, B. Li, Dictionary learning in visual computing, Synth. Lect. Image Video Multimedia Process. 8 (2015) 1–151.
[24] H. Boche, Compressed Sensing and its Applications : MATHEON Workshop 2013, Birkhäuser, 2015.
[25] H. Boche, Compressed Sensing and its Applications : Second International MATHEON Conference 2015, Birkhäuser, 2017.
[26] H. Boche, Compressed Sensing and its Applications : Third International MATHEON Conference 2017, Birkhäuser, 2019.
[27] R.G. Baraniuk, E. Candes, M. Elad, Y. Ma, Applications of sparse representation and compressive sensing, Proc. IEEE 98 (2010) 906–909.
[28] M.F. Duarte, Y.C. Eldar, Structured compressed sensing: From theory to applications, IEEE Trans. Signal Process. 59 (2011) 4053–4085.
[29] D. Craven, B. McGinley, L. Kilmartin, M. Glavin, E. Jones, Compressed sensing for bioelectric signals: A review, IEEE J. Biomed. Health Inf. 19 (2014) 529–540.
[30] L.P. Yaroslavsky, Compression, restoration, Resampling,'Compressive sensing': Fast transforms in digital imaging, J. Opt. 17 (2015) 073001.
[31] D. Thapa, K. Raahemifar, V. Lakshminarayanan, Less is more: Compressive sensing in optics and image science, J. Modern Opt. 62 (2015) 415–429.
[32] Z. Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, A survey of sparse representation: Algorithms and applications, IEEE Access 3 (2015) 490–530.
[33] Y. Zhang, L.Y. Zhang, J. Zhou, L. Liu, F. Chen, X. He, A review of compressive sensing in information security field, IEEE Access 4 (2016) 2507–2519.
[34] N. Vaswani, J. Zhan, Recursive recovery of sparse signal sequences from compressive measurements: a review, IEEE Trans. Signal Process. 64 (2016) 3523–3549.
[35] G. Kumar, K. Baskaran, R. Blessing, M. Lydia, A comprehensive review on the impact of compressed sensing in wireless sensor networks, Int. J. Smart Sens. Intell. Syst. 9 (2016).
[36] R.E. Carrillo, A.B. Ramirez, G.R. Arce, K.E. Barner, B.M. Sadler, Robust compressive sensing of sparse signals: A review, EURASIP J. Adv. Signal Process. 2016 (2016) 108.
[37] Y.V. Parkale, S.L. Nalbalwar, Application of compressed sensing (CS) for ECG signal compression: a review, in: Proceedings of the International Conference on Data Engineering and Communication Technology, Springer, 2017, pp. 53–65.
[38] M. Sandilya, S. Nirmala, Compressed sensing trends in magnetic resonance imaging, Eng. Sci. Technol. Int. J. 20 (2017) 1342–1352.
[39] S. Cheng, Z. Cai, J. Li, Approximate sensory data collection: A survey, Sensors 17 (2017) 564.
[40] M. Rani, S. Dhok, R. Deshmukh, A systematic review of compressive sensing: concepts, implementations and applications, IEEE Access 6 (2018) 4875–4894.
[41] H. Djelouat, A. Amira, F. Bensaali, Compressive sensing-based iot applications: a review, J. Sens. Actuator Netw. 7 (2018) 45.
[42] Y. Wang, D. Meng, M. Yuan, Sparse recovery: From vectors to tensors, Natl. Sci. Rev. 5 (2017) 756–767.
[43] E. Sejdić, I. Orović, S. Stanković, Compressive sensing meets time–frequency: An overview of recent advances in time–frequency processing of sparse signals, Digit. Signal Process. 77 (2018) 22–35.
[44] J. Hampton, A. Doostan, Compressive sampling methods for sparse polynomial chaos expansions, in: Handbook of Uncertainty Quantification, Springer International Publishing, 2017, pp. 1–29.
[45] W.-X. Wang, Y.-C. Lai, C. Grebogi, Data based identification and prediction of nonlinear and complex dynamical systems, Phys. Rep. 644 (2016) 1–76.
[46] K. Sayood, Introduction to Data Compression, Newnes, 2012.
[47] C.D. Meyer, Matrix Analysis and Applied Linear Algebra, Vol. 71, Siam, 2000.
[48] S. Friedland, L.-H. Lim, Nuclear norm of higher-order tensors, Math. Comp. 87 (2018) 1255–1281.
[49] E.J. Candès, Compressive sampling, in: Proceedings of the International Congress of Mathematicians, Vol. 3, Madrid, Spain, 2006, pp. 1433–1452.

[50] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization, Proc. Natl. Acad. Sci. 100 (2003) 2197–2202.
[51] B.K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 24 (1995) 227–234.
[52] S.G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (1993) 3397–3415.
[53] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, IEEE, 1993, pp. 40–44.
[54] T. Blumensath, M.E. Davies, Iterative thresholding for sparse approximations, J. Fourier Anal. Appl. 14 (2008) 629–654.
[55] D. Needell, J.A. Tropp, Cosamp: iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (2009) 301–321.
[56] G.M. Davis, S.G. Mallat, Z. Zhang, Adaptive time-frequency decompositions, Opt. Eng. 33 (1994) 2183–2192.
[57] D.L. Donoho, I. Drori, Y. Tsaig, J.-L. Starck, Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit, Department of Statistics, Stanford University, 2006.
[58] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inform. Theory 49 (2003) 3320–3325.
[59] L. Welch, Lower bounds on the maximum cross correlation of signals (corresp.), IEEE Trans. Inform. Theory 20 (1974) 397–399.
[60] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory (2005).
[61] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, Constr. Approx. 28 (2008) 253–263.
[62] A.M. Tillmann, M.E. Pfetsch, The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing, IEEE Trans. Inform. Theory 60 (2014) 1248–1259.
[63] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (2001) 129–159.
[64] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1996) 267–288.
[65] J. Nocedal, S. Wright, Numerical Optimization, Springer Science & Business Media, 2006.
[66] P.L. Combettes, J.-C. Pesquet, Proximal splitting methods in signal processing, in: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, 2011, pp. 185–212.
[67] H.H. Sohrab, Basic Real Analysis, Springer, 2003.
[68] N. Parikh, S. Boyd, et al., Proximal algorithms, Found. Trends Optim. 1 (2014) 127–239.
[69] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure Appl. Math. 57 (2004) 1413–1457.
[70] Y.E. Nesterov, A method for solving the convex programming problem with convergence rate o (1/K$\hat{2}$), in: Soviet Mathematics Doklady, Vol. 269, 1983, pp. 543–547.
[71] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (2009) 183–202.
[72] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al., Least angle regression, Ann. Statist. 32 (2004) 407–499.
[73] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vision 40 (2011) 120–145.
[74] D.L. Donoho, J. Tanner, Precise undersampling theorems, Proc. IEEE 98 (2010) 913–924.
[75] D. Amelunxen, M. Lotz, M.B. McCoy, J.A. Tropp, Living on the edge: phase transitions in convex programs with random data, Inf. Inference: J. IMA 3 (2014) 224–294.
[76] D.L. Donoho, High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension, Discrete Comput. Geom. 35 (2006) 617–652.
[77] D. Donoho, 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 367 (2009) 4273–4293.
[78] D.L. Donoho, J. Tanner, Counting the faces of randomly-projected hypercubes and orthants, with applications, Discrete Comput. Geom. 43 (2010) 522–541.
[79] D.L. Donoho, J. Tanner, Exponential bounds implying construction of compressed sensing matrices, error-correcting codes, and neighborly polytopes by random sampling, IEEE Trans. Inform. Theory 56 (2010) 2002–2016.
[80] T.T. Cai, A. Zhang, Sharp RIP bound for sparse signal and low-rank matrix recovery, Appl. Comput. Harmon. Anal. 35 (2013) 74–93.
[81] A. Maleki, D.L. Donoho, Optimally tuned iterative reconstruction algorithms for compressed sensing, IEEE J. Sel. Top. Sign. Proces. 4 (2010) 330–341.
[82] J.D. Blanchard, J. Tanner, Performance comparisons of greedy algorithms in compressed sensing, Numer. Linear Algebra Appl. 22 (2015) 254–282.
[83] C. Hegde, P. Indyk, L. Schmidt, A fast approximation algorithm for tree-sparse recovery, in: 2014 IEEE International Symposium on Information Theory, IEEE, 2014, pp. 1842–1846.
[84] E. Van Den Berg, M.P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2008) 890–912.
[85] E. Esser, Y. Lou, J. Xin, A method for finding structured sparse solutions to nonnegative least squares problems with applications, SIAM J. Imaging Sci. 6 (2013) 2010–2046.
[86] P. Yin, Y. Lou, Q. He, J. Xin, Minimization of 1-2 for compressed sensing, SIAM J. Sci. Comput. 37 (2015) A536–A563.
[87] P.D. Tao, The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems, Ann. Oper. Res. 133 (2005) 23–46.
[88] I.F. Gorodnitsky, B.D. Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, IEEE Trans. Signal Process. 45 (1997) 600–616.
[89] M.A. Figueiredo, R.D. Nowak, A bound optimization approach to wavelet-based image deconvolution, in: IEEE International Conference on Image Processing 2005, Vol. 2, IEEE, 2005, pp. II–782.
[90] M.A. Figueiredo, J.M. Bioucas-Dias, R.D. Nowak, Majorization–minimization algorithms for wavelet-based image restoration, IEEE Trans. Image Process. 16 (2007) 2980–2991.
[91] Z. Xu, H. Zhang, Y. Wang, X. Chang, Y. Liang, L 1/2 regularization, Sci. China Inf. Sci. 53 (2010) 1159–1169.
[92] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett. 14 (2007) 707–710.
[93] R. Chartrand, V. Staneva, Restricted isometry properties and nonconvex compressive sensing, Inverse Problems 24 (2008) 035020.
[94] R. Chartrand, W. Yin, Iteratively reweighted algorithms for compressive sensing, in: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2008, pp. 3869–3872.
[95] R. Saab, R. Chartrand, O. Yilmaz, Stable sparse approximations via nonconvex optimization, in: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2008, pp. 3885–3888.
[96] E. Schlossmacher, An iterative technique for absolute deviations curve fitting, J. Amer. Statist. Assoc. 68 (1973) 857–859.
[97] P.W. Holland, R.E. Welsch, Robust regression using iteratively reweighted least-squares, Commun. Statist.-Theory Methods 6 (1977) 813–827.
[98] I. Daubechies, R. DeVore, M. Fornasier, C.S. Güntürk, Iteratively reweighted least squares minimization for sparse recovery, Comm. Pure Appl. Math. 63 (2010) 1–38.
[99] E.J. Candes, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted ℓ 1 minimization, J. Fourier Anal. Appl. 14 (2008) 877–905.
[100] S. Ji, Y. Xue, L. Carin, et al., Bayesian compressive sensing, IEEE Trans. Signal Process. 56 (2008) 2346.
[101] S. Babacan, R. Molina, A.K. Katsaggelos, Bayesian compressive sensing using Laplace priors, IEEE Trans. Image Process. 19 (2009) 53–63.
[102] W.M. Bolstad, J.M. Curran, Introduction to Bayesian Statistics, John Wiley & Sons, 2016.
[103] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.
[104] T.J. Mitchell, J.J. Beauchamp, Bayesian variable selection in linear regression, J. Amer. Statist. Assoc. 83 (404) (1988) 1023–1032.
[105] H. Ishwaran, J. Rao, et al., Spike and slab variable selection: frequentist and Bayesian strategies, Ann. Statist. 33 (2) (2005) 730–773.
[106] V. Ročková, E.I. George, The spike-and-slab lasso, J. Amer. Statist. Assoc. 113 (521) (2018) 431–444.
[107] C.M. Carvalho, N.G. Polson, J.G. Scott, The horseshoe estimator for sparse signals, Biometrika 97 (2) (2010) 465–480.
[108] A. Bhadra, J. Datta, N.G. Polson, B. Willard, et al., Lasso meets horseshoe: A survey, Statist. Sci. 34 (3) (2019) 405–427.
[109] W.R. Gilks, S. Richardson, D. Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman and Hall/CRC, 1995.
[110] D.M. Blei, A. Kucukelbir, J.D. McAuliffe, Variational inference: A review for statisticians, J. Amer. Statist. Assoc. 112 (2017) 859–877.
[111] A.C. Faul, M.E. Tipping, Analysis of sparse Bayesian learning, in: Advances in Neural Information Processing Systems, 2002, pp. 383–389.
[112] M.E. Tipping, A.C. Faul, Fast marginal likelihood maximisation for sparse Bayesian models., in: AISTATS, 2003.
[113] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, Structured sparsity through convex optimization, Statist. Sci. 27 (2012) 450–468.
[114] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (2006) 49–67.
[115] C. Chen, J. Huang, Compressive sensing MRI with wavelet tree sparsity, in: Advances in Neural Information Processing Systems, 2012, pp. 1115–1123.
[116] J. Huang, T. Zhang, D. Metaxas, Learning with structured sparsity, J. Mach. Learn. Res. 12 (2011) 3371–3412.
[117] R.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hegde, Model-based compressive sensing, IEEE Trans. Inf. Theory 56 (2010) 1982–2001.
[118] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2005) 301–320.
[119] J. Huang, T. Zhang, The benefit of group sparsity, Ann. Statist. 38 (2010) 1978–2004.

[120] F.R. Bach, Consistency of the group lasso and multiple kernel learning, J. Mach. Learn. Res. 9 (2008) 1179–1225.
[121] R. Jenatton, J.-Y. Audibert, F. Bach, Structured variable selection with sparsity-inducing norms, J. Mach. Learn. Res. 12 (2011) 2777–2824.
[122] G. Obozinski, L. Jacob, J.-P. Vert, Group lasso with overlaps: The latent group lasso approach, 2011, arXiv preprint arXiv:1110.0413.
[123] J. Mairal, R. Jenatton, G. Obozinski, F. Bach, Convex and network flow optimization for structured sparsity, J. Mach. Learn. Res. 12 (2011) 2681–2720.
[124] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, J. Mach. Learn. Res. 11 (2010) 19–60.
[125] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by v1?, Vis. Res. 37 (1997) 3311–3325.
[126] B.A. Olshausen, D.J. Field, Natural image statistics and efficient coding, Network: Comput. Neural Syst. 7 (1996) 333–339.
[127] M.S. Lewicki, B.A. Olshausen, Probabilistic framework for the adaptation and comparison of image codes, J. Opt. Soc. Amer. A 16 (1999) 1587–1601.
[128] M.S. Lewicki, T.J. Sejnowski, Learning overcomplete representations, Neural Comput. 12 (2000) 337–365.
[129] K. Engan, S.O. Aase, J. Husoy, Method of optimal directions for frame design, in: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99, Vol. 5, IEEE, 1999, pp. 2443–2446.
[130] K. Engan, B.D. Rao, K. Kreutz-Delgado, Frame design using FOCUSS with method of optimal directions (MOD), in: Proceedings of the Norwegian Signal Processing Symposium, Vol. 99, 1999, pp. 65–69.
[131] K. Engan, S.O. Aase, J.H. Husøy, Multi-frame compression: Theory and design, Signal Process. 80 (2000) 2121–2140.
[132] K. Kreutz-Delgado, J.F. Murray, B.D. Rao, K. Engan, T.-W. Lee, T.J. Sejnowski, Dictionary learning algorithms for sparse representation, Neural Comput. 15 (2003) 349–396.
[133] J.F. Murray, K. Kreutz-Delgado, An improved FOCUSS-based learning algorithm for solving sparse linear inverse problems, in: Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers, Vol. 1, IEEE, 2001, pp. 347–351.
[134] K. Kreutz-Delgado, B.D. Rao, FOCUSS-based dictionary learning algorithms, in: Wavelet Applications in Signal and Image Processing VIII, Vol. 4119, International Society for Optics and Photonics, 2000, pp. 459–474.
[135] M. Aharon, M. Elad, A. Bruckstein, et al., K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (2006) 4311.
[136] K. Gregor, Y. LeCun, Learning fast approximations of sparse coding, in: Proceedings of the 27th International Conference on International Conference on Machine Learning, Omnipress, 2010, pp. 399–406.
[137] A. Fawzi, M. Davies, P. Frossard, Dictionary learning for fast classification based on soft-thresholding, Int. J. Comput. Vis. 114 (2015) 306–321.
[138] I. Tosic, P. Frossard, Dictionary learning: what is the right representation for my signal?, IEEE Signal Process. Mag. 28 (2011) 27–38.
[139] A. Gersho, R.M. Gray, Vector Quantization and Signal Compression, Vol. 159, Springer Science & Business Media, 2012.
[140] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 689–696.
[141] D. Bertsekas, Nonlinear Programming, Athena Scientific, 1999.
[142] Y. Bao, J.L. Beck, H. Li, Compressive sampling for accelerometer signals in structural health monitoring, Struct. Health Monit. 10 (2011) 235–246.
[143] Y. Bao, H. Li, X. Sun, Y. Yu, J. Ou, Compressive sampling-based data loss recovery for wireless sensor networks used in civil structural health monitoring, Struct. Health Monit. 12 (2013) 78–95.
[144] Y. Bao, Y. Yu, H. Li, X. Mao, W. Jiao, Z. Zou, J. Ou, Compressive sensing-based lost data recovery of fast-moving wireless sensing for structural health monitoring, Struct. Control Health Monit. 22 (2015) 433–448.
[145] Y. Huang, J.L. Beck, S. Wu, H. Li, Robust Bayesian compressive sensing for signals in structural health monitoring, Comput.-Aided Civ. Infrastruct. Eng. 29 (2014) 160–179.
[146] Y. Huang, J.L. Beck, S. Wu, H. Li, Bayesian compressive sensing for approximately sparse signals and application to structural health monitoring signals for data loss recovery, Probab. Eng. Mech. 46 (2016) 62–79.
[147] Y. Yang, S. Nagarajaiah, Y.-Q. Ni, Data compression of very large-scale structural seismic and typhoon responses by low-rank representation with matrix reshape, Struct. Control Health Monit. 22 (2015) 1119–1131.
[148] Y. Yang, S. Nagarajaiah, Harnessing data structure for recovery of randomly missing structural vibration responses time history: sparse representation versus low-rank structure, Mech. Syst. Signal Process. 74 (2016) 165–182.
[149] Y. Yang, S. Nagarajaiah, Robust data transmission and recovery of images by compressed sensing for structural health diagnosis, Struct. Control Health Monit. 24 (2017) e1856.
[150] Y. Bao, Z. Shi, X. Wang, H. Li, Compressive sensing of wireless sensors based on group sparse optimization for structural health monitoring, Struct. Health Monit. 17 (2018) 823–836.
[151] Z. Pang, M. Yuan, M.B. Wakin, A random demodulation architecture for sub-sampling acoustic emission signals in structural health monitoring, J. Sound Vib. 431 (2018) 390–404.
[152] S. O'Connor, J. Lynch, A. Gilbert, Compressed sensing embedded in an operational wireless sensor network to achieve energy efficiency in long-term monitoring applications, Smart Mater. Struct. 23 (2014) 085014.
[153] Z. Zou, Y. Bao, H. Li, B.F. Spencer, J. Ou, Embedding compressive sensing-based data loss recovery algorithm into wireless smart sensors for structural health monitoring, IEEE Sens. J. 15 (2014) 797–808.
[154] R. Klis, E.N. Chatzi, Data recovery via hybrid sensor networks for vibration monitoring of civil structures, Int. J. Sustain. Mater. Struct. Syst. 2 (2015) 161–184.
[155] R. Klis, E.N. Chatzi, Vibration monitoring via spectro-temporal compressive sensing for wireless sensor networks, Struct. Infrastruct. Eng. 13 (2017) 195–209.
[156] D. Mascareñas, A. Cattaneo, J. Theiler, C. Farrar, Compressed sensing techniques for detecting damage in structures, Struct. Health Monit. 12 (2013) 325–338.
[157] Y. Wang, H. Hao, Damage identification scheme based on compressive sensing, J. Comput. Civ. Eng. 29 (2013) 04014037.
[158] Y. Bao, H. Li, Z. Chen, F. Zhang, A. Guo, Sparse $\ell 1$ optimization-based identification approach for the distribution of moving heavy vehicle loads on Cable-stayed bridges, Struct. Control Health Monit. 23 (2016) 144–155.
[159] Y. Yang, S. Nagarajaiah, Structural damage identification via a combination of blind feature extraction and sparse representation classification, Mech. Syst. Signal Process. 45 (2014) 1–23.
[160] Y. Yang, S. Nagarajaiah, Output-only modal identification by compressed sensing: non-uniform low-rate random sampling, Mech. Syst. Signal Process. 56 (2015) 15–34.
[161] Y. Yang, C. Dorn, T. Mancini, Z. Talken, S. Nagarajaiah, G. Kenyon, C. Farrar, D. Mascareñas, Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-nyquist), video measurements, J. Sound Vib. 390 (2017) 232–256.
[162] J.Y. Park, M.B. Wakin, A.C. Gilbert, Modal analysis with compressive measurements, IEEE Trans. Signal Process. 62 (2014) 1655–1670.
[163] S. Li, D. Yang, G. Tang, M.B. Wakin, Atomic norm minimization for modal analysis from random and compressed samples, IEEE Trans. Signal Process. 66 (2018) 1817–1831.
[164] E.M. Hernandez, Identification of isolated structural damage from incomplete spectrum information using l1-norm minimization, Mech. Syst. Signal Process. 46 (2014) 59–69.
[165] E.M. Hernandez, Identification of localized structural damage from highly incomplete modal information: Theory and experiments, J. Eng. Mech. 142 (2015) 04015075.
[166] X.-Q. Zhou, Y. Xia, S. Weng, L1 regularization approach to structural damage detection using frequency data, Struct. Health Monit. 14 (2015) 571–582.
[167] C.B. Smith, E.M. Hernandez, Detection of spatially sparse damage using impulse response sensitivity and LASSO regularization, Inverse Probl. Sci. Eng. 27 (2019) 1–16.
[168] C. Zhang, Y. Xu, Comparative studies on damage identification with tikhonov regularization and sparse regularization, Struct. Control Health Monit. 23 (2016) 560–579.
[169] C. Zhang, J.-Z. Huang, G.-Q. Song, L. Chen, Structural damage identification by extended kalman filter with $\ell 1$-norm regularization scheme, Struct. Control Health Monit. 24 (2017) e1999.
[170] R. Hou, Y. Xia, X. Zhou, Structural damage detection based on l1 regularization using natural frequencies and mode shapes, Struct. Control Health Monit. 25 (2018) e2107.
[171] R. Hou, Y. Xia, Y. Bao, X. Zhou, Selection of regularization parameter for l1-regularized damage detection, J. Sound Vib. 423 (2018) 141–160.
[172] X. Zhou, R. Hou, Y. Wu, Structural damage detection based on iteratively reweighted $\ell 1$ regularization algorithm, Adv. Struct. Eng. 22 (2019) 1479–1487.
[173] L. Wang, Z.-R. Lu, Sensitivity-free damage identification based on incomplete modal data, sparse regularization and alternating minimization approach, Mech. Syst. Signal Process. 120 (2019) 43–68.
[174] M. Jayawardhana, X. Zhu, R. Liyanapathirana, U. Gunawardana, Compressive sensing for efficient health monitoring and effective damage detection of structures, Mech. Syst. Signal Process. 84 (2017) 414–430.
[175] J. Guo, L. Wang, I. Takewaki, Modal-based structural damage identification by minimum constitutive relation error and sparse regularization, Struct. Control Health Monit. 25 (2018) e2255.
[176] I.A. Kougioumtzoglou, K.R.M. dos Santos, L. Comerford, Incomplete data based parameter identification of nonlinear and time-variant oscillators with fractional derivative elements, Mech. Syst. Signal Process. 94 (2017) 279–296.
[177] K.R.M. dos Santos, O. Brudastova, I.A. Kougioumtzoglou, Spectral identification of nonlinear multi-degree-of-freedom structural systems with fractional derivative terms based on incomplete non-stationary data, Struct. Saf. 86 (2020) 101975.
[178] J.S. Bendat, Nonlinear Systems Techniques and Applications, John Wiley & Sons, 1998.
[179] K. Gkoktsi, A. Giaralis, Assessment of sub-nyquist deterministic and random data sampling techniques for operational modal analysis, Struct. Health Monit. 16 (2017) 630–646.

[180] K. Gkoktsi, A. Giaralis, A multi-sensor sub-nyquist power spectrum blind sampling approach for low-power wireless sensors in operational modal analysis applications, Mech. Syst. Signal Process. 116 (2019) 879–899.
[181] Z. Lai, S. Nagarajaiah, Sparse structural system identification method for nonlinear dynamic systems with hysteresis/inelastic behavior, Mech. Syst. Signal Process. 117 (2019) 813–842.
[182] Z. Lai, S. Nagarajaiah, Semi-supervised structural linear/nonlinear damage detection and characterization using sparse identification, Struct. Control Health Monit. 26 (2019) e2306.
[183] D. Sen, A. Aghazadeh, A. Mousavi, S. Nagarajaiah, R. Baraniuk, Sparsity-based approaches for damage detection in plates, Mech. Syst. Signal Process. 117 (2019) 333–346.
[184] A. Rezayat, V. Nassiri, B. De Pauw, J. Ertveldt, S. Vanlanduit, P. Guillaume, Identification of dynamic forces using group-sparsity in frequency domain, Mech. Syst. Signal Process. 70 (2016) 756–768.
[185] Q. Li, Q. Lu, A hierarchical Bayesian method for vibration-based time domain force reconstruction problems, J. Sound Vib. 421 (2018) 190–204.
[186] B. Qiao, Z. Mao, J. Liu, Z. Zhao, X. Chen, Group sparse regularization for impact force identification in time domain, J. Sound Vib. 445 (2019) 44–63.
[187] L. Comerford, I.A. Kougioumtzoglou, M. Beer, Compressive sensing based stochastic process power spectrum estimation subject to missing data, Probab. Eng. Mech. 44 (2016) 66–76.
[188] L.A. Comerford, M. Beer, I.A. Kougioumtzoglou, Compressive sensing based power spectrum estimation from incomplete records by utilizing an adaptive basis, in: 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES), IEEE, 2014, pp. 117–124.
[189] L. Comerford, H. Jensen, F. Mayorga, M. Beer, I. Kougioumtzoglou, Compressive sensing with an adaptive wavelet basis for structural system response and reliability analysis under missing data, Comput. Struct. 182 (2017) 26–40.
[190] Y. Zhang, L. Comerford, I.A. Kougioumtzoglou, M. Beer, Lp-norm minimization for stochastic process power spectrum estimation subject to incomplete data, Mech. Syst. Signal Process. 101 (2018) 361–376.
[191] V. Laface, I.A. Kougioumtzoglou, G. Malara, F. Arena, Efficient processing of water wave records via compressive sensing and joint time-frequency analysis via harmonic wavelets, Appl. Ocean Res. 69 (2017) 1–9.
[192] G. Malara, I.A. Kougioumtzoglou, F. Arena, Extrapolation of random wave field data via compressive sampling, Ocean Eng. 157 (2018) 87–95.
[193] V. Laface, G. Malara, A. Romolo, F. Arena, I.A. Kougioumtzoglou, Compressive sensing-based reconstruction of sea free-surface elevation on a vertical wall, J. Waterw. Port Coast. Ocean Eng. 144 (2018) 04018010.
[194] V. Laface, G. Malara, I.A. Kougioumtzoglou, A. Romolo, F. Arena, Nonlinear wave crest distribution on a vertical breakwater, Coast. Eng. 138 (2018) 227–234.
[195] Y. Wang, T. Zhao, Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling, Géotechnique 67 (2016) 523–536.
[196] Y. Wang, O.V. Akeju, T. Zhao, Interpolation of spatially varying but sparsely measured geo-data: A comparative study, Eng. Geol. 231 (2017) 200–217.
[197] Y. Wang, T. Zhao, K.-K. Phoon, Direct simulation of random field samples from sparsely measured geotechnical data with consideration of uncertainty in interpretation, Can. Geotech. J. 55 (2017) 862–880.
[198] T. Zhao, Y. Wang, Simulation of cross-correlated random field samples from sparse measurements using Bayesian compressive sensing, Mech. Syst. Signal Process. 112 (2018) 384–400.
[199] S. Montoya-Noguera, T. Zhao, Y. Hu, Y. Wang, K.-K. Phoon, Simulation of non-stationary non-Gaussian random fields from sparse measurements using Bayesian compressive sampling and Karhunen-Loève expansion, Struct. Saf. 79 (2019) 66–79.
[200] Y. Wang, T. Zhao, Y. Hu, K.-K. Phoon, Simulation of random fields with trend from sparse measurements without detrending, J. Eng. Mech. 145 (2018) 04018130.
[201] Y. Wang, T. Zhao, K.-K. Phoon, Statistical inference of random field auto-correlation structure from multiple sets of incomplete and sparse measurements using Bayesian compressive sampling-based bootstrapping, Mech. Syst. Signal Process. 124 (2019) 217–236.
[202] J. Li, J. Chen, Stochastic Dynamics of Structures, John Wiley & Sons, 2009.
[203] M. Grigoriu, Stochastic Systems: Uncertainty Quantification and Propagation, in: Springer Series in Reliability Engineering, 2012.
[204] R.G. Ghanem, P.D. Spanos, Stochastic Finite Elements: A Spectral Approach, Courier Dover Publications, 2003.
[205] D. Xiu, G.E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2002) 619–644.
[206] G. Stefanou, The stochastic finite element method: Past, present and future, Comput. Methods Appl. Mech. Engrg. 198 (2009) 1031–1051.
[207] B. Adcock, S. Brugiapaglia, C.G. Webster, Compressed sensing approaches for polynomial approximation of high-dimensional functions, in: Compressed Sensing and Its Applications, Birkhäuser-Springer, 2017, pp. 93–124.
[208] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, J. Comput. Phys. 230 (2011) 2345–2367.
[209] A. Doostan, H. Owhadi, A non-adapted sparse approximation of PDEs with stochastic inputs, J. Comput. Phys. 230 (2011) 3015–3034.
[210] H. Lei, X. Yang, B. Zheng, G. Lin, N.A. Baker, Constructing surrogate models of complex systems with enhanced sparsity: quantifying the influence of conformational uncertainty in biomolecular solvation, Multiscale Model. Simul. 13 (2015) 1327–1353.
[211] B.A. Jones, N. Parrish, A. Doostan, Postmaneuver collision probability estimation using sparse polynomial chaos expansions, J. Guid. Control Dyn. 38 (2015) 1425–1437.
[212] S. Salehi, M. Raisee, M.J. Cervantes, A. Nourbakhsh, Efficient uncertainty quantification of stochastic CFD problems using sparse polynomial chaos and compressed sensing, Comput. & Fluids 154 (2017) 296–321.
[213] L. Fagiano, M. Khammash, Simulation of stochastic systems via polynomial chaos expansions and convex optimization, Phys. Rev. E 86 (2012) 036702.
[214] X. Yang, G.E. Karniadakis, Reweighted $\ell 1$ minimization method for stochastic elliptic differential equations, J. Comput. Phys. 248 (2013) 87–108.
[215] J. Peng, J. Hampton, A. Doostan, A weighted $\ell 1$-minimization approach for sparse polynomial chaos expansions, J. Comput. Phys. 267 (2014) 92–111.
[216] J. Peng, J. Hampton, A. Doostan, On polynomial chaos expansion via gradient-enhanced $\ell 1$-minimization, J. Comput. Phys. 310 (2016) 440–458.
[217] L. Guo, A. Narayan, T. Zhou, A gradient enhanced $\ell 1$-minimization for sparse approximation of polynomial chaos expansions, J. Comput. Phys. 367 (2018) 49–64.
[218] L. Guo, Y. Liu, L. Yan, Sparse recovery via ℓq-minimization for polynomial chaos expansions, Numer. Math.: Theory Methods Appl. 10 (2017) 775–797.
[219] H. Rauhut, C. Schwab, Compressive sensing Petrov-Galerkin approximation of high-dimensional parametric operator equations, Math. Comp. 86 (2017) 661–700.
[220] L. Yan, Y. Shin, D. Xiu, Sparse approximation using $\ell 1$-$\ell 2$ minimization and its application to stochastic collocation, SIAM J. Sci. Comput. 39 (2017) A229–A254.
[221] J.D. Jakeman, M.S. Eldred, K. Sargsyan, Enhancing $\ell 1$-minimization estimates of polynomial chaos expansions using basis selection, J. Comput. Phys. 289 (2015) 18–34.
[222] N. Alemazkoor, H. Meidani, Divide and conquer: an incremental sparsity promoting compressive sampling approach for polynomial chaos expansions, Comput. Methods Appl. Mech. Engrg. 318 (2017) 937–956.
[223] P. Tsilifis, X. Huan, C. Safta, K. Sargsyan, G. Lacaze, J.C. Oefelein, H.N. Najm, R.G. Ghanem, Compressive sensing adaptation for polynomial chaos expansions, J. Comput. Phys. 380 (2019) 29–47.
[224] J. Hampton, A. Doostan, Compressive sampling of polynomial chaos expansions: convergence analysis and sampling strategies, J. Comput. Phys. 280 (2015) 363–386.
[225] J. Hampton, A. Doostan, Basis adaptive sample efficient polynomial chaos (BASE-PC), J. Comput. Phys. 371 (2018) 20–49.
[226] J.D. Jakeman, A. Narayan, T. Zhou, A generalized sampling and preconditioning scheme for sparse approximation of polynomial chaos expansions, SIAM J. Sci. Comput. 39 (2017) A1114–A1144.
[227] P. Diaz, A. Doostan, J. Hampton, Sparse polynomial chaos expansions via compressed sensing and D-optimal design, Comput. Methods Appl. Mech. Engrg. 336 (2018) 640–666.
[228] K. Sargsyan, C. Safta, H.N. Najm, B.J. Debusschere, D. Ricciuto, P. Thornton, Dimensionality reduction for complex models via Bayesian compressive sensing, Int. J. Uncertain. Quantif. 4 (2014).
[229] G. Karagiannis, B.A. Konomi, G. Lin, A Bayesian mixed shrinkage prior procedure for spatial-stochastic basis selection and evaluation of gpc expansions: applications to elliptic SPDEs, J. Comput. Phys. 284 (2015) 528–546.
[230] D. Schiavazzi, A. Doostan, G. Iaccarino, Sparse multiresolution regression for uncertainty propagation, Int. J. Uncertain. Quantif. 4 (2014).
[231] H. Liang, Q. Sun, Q. Du, Data-driven compressive sensing and applications in uncertainty quantification, J. Comput. Phys. 374 (2018) 787–802.
[232] P. Rai, K. Sargsyan, H. Najm, Compressed sparse tensor based quadrature for vibrational quantum mechanics integrals, Comput. Methods Appl. Mech. Engrg. 336 (2018) 471–484.
[233] R. Feynman, A. Hibbs, Quantum Mechanics and Path Integrals, McGraw-Hill, 1965.
[234] M. Chaichian, A. Demichev, Path Integrals in Physics: Volume I Stochastic Processes and Quantum Mechanics, Institute of Physics Publishing, 2018.
[235] H.S. Wio, Path Integrals for Stochastic Processes: An Introduction, World Scientific, 2013.
[236] I.A. Kougioumtzoglou, P.D. Spanos, An analytical wiener path integral technique for non-stationary response determination of nonlinear oscillators, Probab. Eng. Mech. 28 (2012) 125–131.
[237] I.A. Kougioumtzoglou, P.D. Spanos, Nonstationary stochastic response determination of nonlinear systems: a wiener path integral formalism, J. Eng. Mech. 140 (2014) 04014064.
[238] A. Di Matteo, I.A. Kougioumtzoglou, A. Pirrotta, P.D. Spanos, M. Di Paola, Stochastic response determination of nonlinear oscillators with fractional derivatives elements via the wiener path integral, Probab. Eng. Mech. 38 (2014) 127–135.
[239] A.F. Psaros, I.A. Kougioumtzoglou, I. Petromichelakis, Sparse representations and compressive sampling for enhancing the computational efficiency of the wiener path integral technique, Mech. Syst. Signal Process. 111 (2018) 87–101.

[240] I. Petromichelakis, A.F. Psaros, I.A. Kougioumtzoglou, Stochastic response determination and optimization of a class of nonlinear electromechanical energy harvesters: a wiener path integral approach, Probab. Eng. Mech. 53 (2018) 116–125.

[241] A.F. Psaros, Y. Zhao, I.A. Kougioumtzoglou, An exact closed-form solution for linear multi-degree-of-freedom systems under Gaussian white noise via the wiener path integral technique, Probab. Eng. Mech. (2020) 103040.

[242] A.F. Psaros, I.A. Kougioumtzoglou, Functional series expansions and quadratic approximations for enhancing the accuracy of the wiener path integral technique, J. Eng. Mech. 146 (2020) 04020065.

[243] I. Petromichelakis, A.F. Psaros, I.A. Kougioumtzoglou, Stochastic response determination of nonlinear structural systems with singular diffusion matrices: A wiener path integral variational formulation with constraints, Probab. Eng. Mech. 60 (2020) 103044.

[244] A.F. Psaros, O. Brudastova, G. Malara, I.A. Kougioumtzoglou, Wiener path integral based response determination of nonlinear systems subject to non-white, non-Gaussian, and non-stationary stochastic excitation, J. Sound Vib. 433 (2018) 314–333.

[245] I.A. Kougioumtzoglou, A. Di Matteo, P.D. Spanos, A. Pirrotta, M. Di Paola, An efficient wiener path integral technique formulation for stochastic response determination of nonlinear MDOF systems, J. Appl. Mech. 82 (2015) 101005.

[246] A.F. Psaros, I. Petromichelakis, I.A. Kougioumtzoglou, Wiener path integrals and multi-dimensional global bases for non-stationary stochastic response determination of structural systems, Mech. Syst. Signal Process. 128 (2019) 551–571.

[247] K. Champion, B. Lusch, J.N. Kutz, S.L. Brunton, Data-driven discovery of coordinates and governing equations, 2019, arXiv preprint arXiv:1904.02107.

[248] R. Ganguli, S. Adhikari, The digital twin of discrete dynamic systems: initial approaches and future challenges, Appl. Math. Model. 77 (2020) 1110–1128.