Union of low-rank subspaces detector

Mohsen Joneidi[1] ✉, Parvin Ahmadi[2], Mostafa Sadeghi[2], Nazanin Rahnavard[1]

[1]*Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando 32826, USA*
[2]*Department of Electrical Engineering, Sharif University of Technology, Tehran 145888-9694, Islamic Republic of Iran*
✉ *E-mail: joneidi@knights.ucf.edu*

Abstract: The problem of signal detection using a flexible and general model is considered. Owing to applicability and flexibility of sparse signal representation and approximation, it has attracted a lot of attention in many signal processing areas. In this study, the authors propose a new detection method based on sparse decomposition in a union of subspaces model. Their proposed detector uses a dictionary that can be interpreted as a bank of matched subspaces. This improves the performance of signal detection, as it is a generalisation for detectors. Low-rank assumption for the desired signals implies that the representations of these signals in terms of some proper bases would be sparse. Their proposed detector exploits sparsity in its decision rule. They demonstrate the high efficiency of their method in the cases of voice activity detection in speech processing.

1 Introduction

Sparse approximation techniques have found wide use due to their benefits and high flexibility in many applications in image and signal processing [1, 2]. Sparse representation can efficiently extract most important features of a signal, so it provides very promising results in data compression [3], de-noising [4], blind source separation [5], signal classification [6] and so on. The methods based on exploiting the signal sparsity have two main steps. First, an over-complete dictionary [1] is selected/learned according to the structural characteristics of the set of signals, and then the target signal is decomposed over the dictionary to obtain a compact representation. Representation in terms of a few designed/learned bases can accurately capture the signal structure characteristics, which in turn leads to an improvement in the distinction between noise/interference and structured signals.

In some signal processing applications, the task is to detect the presence of a signal from its noisy measurements. For example, in speech processing, voice activity detection (VAD) is performed to distinguish speech segments from non-speech segments in an audio stream. VAD plays a critical role on increasing the capacity of transmission and speech storage by reducing the average bit rate [7].

Signal detection is an old problem in signal processing and there are some traditional signal detectors including energy detector, matched filter and matched subspace detector [8]. Matched signal detector is the most basic framework for signal detection which needs a bank of matched signals to design a detector system. However, in many applications it is preferred to replace rank-1 signals by a multirank matched subspace [8]. Matched subspace detector assumes the span of a subspace as the desired signals and rejects that part of signal which lies on the null-space of the assumed subspace. Generalised likelihood ratio test (GLRT) for matched subspace detector is the uniformly most powerful invariant (UMP-invariant) statistic for detection [8]. The subspace model for detection needs some bases as the span of desired signals which can be a set of fixed bases such as discrete Fourier transform or data-dependent bases such as principal component analysis (PCA). Although subspace model is more adaptive for signal analysis, it needs several parameters that must be either known or estimated. For example, the set of bases spanning the desired signals, the coefficients of the bases, noise covariance and

signal-to-noise ratio (SNR). Depending on the knowledge about different parameters, the optimum statistic is suggested in [9] for four situations. In the case of unknown coefficients, orthogonal projection of the observation is used to determine the coefficients of the contribution of each basis. The present paper assumes a more general model for signals in which considers a union of subspaces (UoSs).

Sparsity has been exploited widely in detection purposes, for example, abnormal event detection [10], VAD [11] and face detection [12]. A multi-criteria detection based on intelligent switching between traditional detection and sparse detection is proposed in [13]. In these works, sparsity has been used to extract features or define a heuristic criterion for detection. Compressive detector is another application of sparsity for signal detection. It is able to detect signals only by using some measurements from the original samples while the performance is not degraded dramatically [14, 15]. The goal of compressive detector is to preserve the performance of detector the same as the original detector. In this paper, we use sparsity from a different point of view. The traditional detectors are generalised to consider sparsity on the optimum decision rule and a new trade-off is suggested between sparsity (rank of a subspace) and error of projection (distance to a low-rank subspace).

In this paper, we propose a new signal detection method based on the union of low-rank subspaces (ULRS) model [16, 17]. This model is able to reveal intrinsic structure of a set of signals. The proposed detector is a generalised version of traditional detectors. In other words, imposing a union of rank-1 subspaces model for desired signals yields nothing other than the traditional matched filter banks. We investigate our detector from different points of views in order to show relation between our method and other classical detectors. We also derive a robust version of the proposed detector in order to provide robustness against outliers and gross errors. We provide theoretical investigations as well as experimental results on VAD.

The rest of this paper is organised as follows. Section 2 provides a brief background on sparse representation theory and basic concepts of detection theory. In Section 3 we describe our new signal detection method, study its performance and provide its robust version. Section 4 experimentally demonstrates the effectiveness of our proposed signal detection method. Finally, Section 5 concludes this paper with a summary of the proposed paper.

2 Theoretical background and review: basic theory of sparse decomposition

Sparse decomposition of signals based on some basis functions has attracted a lot of attention during the last decade [1]. In this approach, one wants to approximate a given signal as a linear combination of as few basis functions as possible. Each basis function is called an atom and their collection is called a dictionary [18]. The dictionary is usually over-complete, that is, the number of atoms is much more than the dimension of atoms. Specifically, let $y \in \mathbb{R}^N$ be the signal which is sparsely represented over the dictionary $D \in \mathbb{R}^{N \times K}$ with $K > N$. This amounts to the following problem:

$$\hat{x} = \underset{x}{\mathrm{argmin}} \, \|x\|_0 \text{ s.t. } y = Dx \qquad (1)$$

where $\|.\|_0$ stands for the so-called ℓ_0 pseudo-norm which counts the number of non-zero elements. Many algorithms have been introduced to solve the problem of finding the sparsest approximation of a signal in a given over-complete dictionary (for a good review see [19]). For a specified class of signals, for example, class of natural images, the dictionary should have the capability of sparsely representing the signals. In some applications, there is a predefined and fixed dictionary which is well-matched to the contents of the specific class of signals. Over-complete discrete cosine transform (DCT) dictionary for the class of natural images is an example. These non-adaptive dictionaries are favourable because of their simplicity. On the other hand, learning-based dictionary results in better matching contents of the signals [1]. Most dictionary learning algorithms are indeed a generalisation of the clustering algorithms. While in clustering each training signal is forced to assign only one atom (cluster centre), in the dictionary learning problem each signal is allowed to use more than one atom provided that it uses as fewest as few atoms as possible. The general dictionary learning problem can be stated as follows:

$$(\hat{D}, \hat{X}) = \underset{D, X}{\mathrm{argmin}} \, \|Y - DX\|_F^2 \text{ s.t.} \|x_i\|_0 \le T, \, \forall i. \qquad (2)$$

where the columns of Y contain the observed data and x_i, the columns of X, are sparse representations of the observed data. Most dictionary learning algorithms solve the above problem by alternatively minimising it over D and X. Dictionary learning algorithms differ mainly in performing the minimisation over the dictionary. Dictionary learning has an important role in the sparse decomposition-based methods. A section in the proposed method section is allocated for discussion on dictionary learning.

3 Basic theory of detection

In this section, we review signal detection theory and study some related detectors to our proposed one. First, consider the following model for detection:

$$\begin{array}{ll} \mathfrak{H}_0 : y = n, & \text{signal absence} \\ \mathfrak{H}_1 : y = s + n, & \text{signal presence} \end{array} \qquad (3)$$

where $y \in \mathbb{R}^N$ is the observation vector, $s \in \mathbb{R}^N$ is the signal of interest and $n \in \mathbb{R}^N$ is the observation noise of the model. First, we assume that the probability density functions of $p(y|H_0)$ and $p(y|\mathfrak{H}_1)$ are known. In this case, the likelihood ratio test (LRT) gives

$$\frac{p(y|\mathfrak{H}_1)}{p(y|\mathfrak{H}_0)} \underset{\mathfrak{H}_0}{\overset{\mathfrak{H}_1}{\gtrless}} \gamma \qquad (4)$$

where γ is a threshold that satisfies the desired amount of probability of false alarm. By Gaussian assumption on the noise with covariance matrix R, LRT simplifies to

$$s^H R^{-1} y \lessgtr \gamma \qquad (5)$$

where R is the covariance matrix. If it is not known in advance, it must be determined by obtaining the sample covariance matrix in the above test. Probability of detection is then equal to [20]

$$P_D(\alpha) = Q(Q^{-1}(\alpha) - \sqrt{\mathrm{SNR}}) \qquad (6)$$

in which α is the probability of false alarm and $Q(\alpha) = \int_\alpha^\infty \exp(-t^2) \mathrm{d}t$. Another well-known detector is GLRT [21] which is derived by maximising conditional densities constituting the LRT with respect to the unknown parameters. The following detection criterion is obtained by assuming the covariance matrix to be unknown:

$$\frac{|s^H \hat{R}^{-1} y|}{s^H \hat{R}^{-1} s \left(1 + \frac{1}{K} y^H \hat{R}^{-1} y\right)} \lessgtr \gamma \qquad (7)$$

where L is the number of snapshots available for $\hat{R}$ estimation. In [21], no optimality has been claimed for GLRT. However, Scharf and Friedlander [8] have shown that GLRT is UMP invariant. This is the strongest statement of optimality derived for a detector. GLRT detector may be interpreted as a projection on the null-space of the interference followed by a matched subspace detector [8]. Consider the following model for hypothesis test:

$$\begin{array}{ll} \mathfrak{H}_0 : y = C\theta + n, & \text{signal absence} \\ \mathfrak{H}_1 : y = Dx + C\theta + n, & \text{signal presence} \end{array} \qquad (8)$$

where $C \in \mathbb{R}^{N \times p}$ spans the background or interference subspace, $p < N$ and θ determine contribution of each column of C. $D \in \mathbb{R}^{N \times m}$ spans signal subspace which is to be detected, $m < N$ and x determine the contribution of each column of D. It is obvious that if $p \ge N$ or $m \ge N$, then C or D spans all the space of the signals. In other words, in this case C or D may be over-fitted for background detection and signal detection, respectively. On the other hand, restrictions of p and m may result in unreliable subspaces which are unable to fit suitable matched subspace. The role of matched subspaces detector is as follows:

$$y^H P_C^\perp P_{DC^\perp} P_C^\perp y \lessgtr \gamma \qquad (9)$$

where $P_C^\perp$ is the orthogonal projection matrix on the null-space of C and $P_{DC^\perp}$ is the part of orthogonal projection of P_D which does not account for subspace spanned by C. Fig. 1 shows the block diagram of this detector.

At the conclusion of paper, Scharf and Friedlander [8] mentioned that basis can be extracted from discrete cosine transform, wavelet transform or learned by data-dependent analysis such as PCA. Using such basis provides a matched subspace for the whole desired signals which are going to be detected. For more illustration see Fig. 2. This figure shows composites of some three-dimensional (3D) data by signal and non-signal (interference and noise) parts. Two low-rank subspaces are shown corresponding to rank-1 and rank-2 subspaces (the low-rank matched subspace) which are obtained by PCA.

The main contribution of [8] may be answering 'no' to the question, 'Can the GLRT be improved upon?' while they did not assume any prior information on the structure of the low-rank

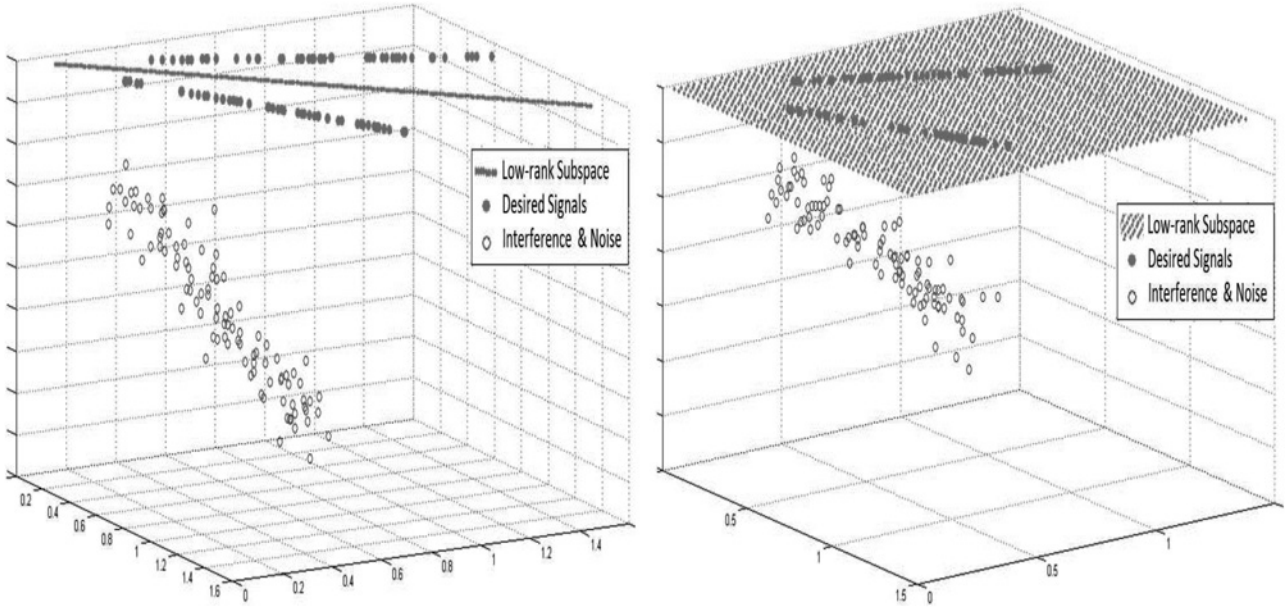**Fig. 1** *Block diagram of the matched subspace detector*

Fig. 2 *Two low-rank subspaces learned by PCA for some 3D signals*

matched filter. The structural assumption can be applied by assuming a sparse prior on the coefficients of $\boldsymbol{\theta}$ and $\boldsymbol{x}$. The proposed method of this paper suggests using the model of ULRS for signals due to its suitable fitness which has been proven in many signal processing applications. Instead of traditional analysis such as PCA, modern analysis such as the methods proposed in [22–24] can be exploited in order to recover suitable bases spanning these low-rank subspaces. Fig. 3 shows a union of matched low-rank subspaces corresponding to the data of Fig. 2.

Compressive detection is another application of the sparse theory exploited in signal detection and studied in [25, 15]. Instead of dealing with all the samples of the signal, the compressed detector works with few measurements. This detector distinguishes between two hypotheses

$$
\begin{aligned}
\mathfrak{H}_0 &: z = \Phi n, & \text{signal absence} \\
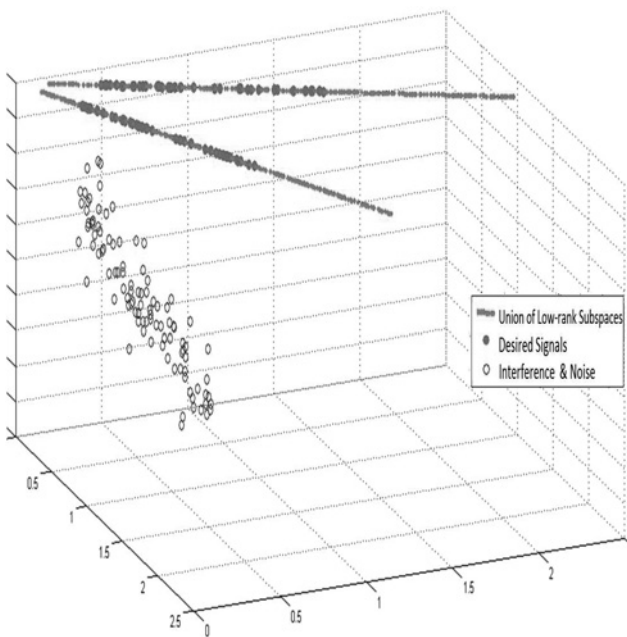\mathfrak{H}_1 &: z = \Phi(s + n), & \text{signal presence}
\end{aligned}
\tag{10}
$$

Fig. 3 *Union of rank-1 subspaces provides suitable matched subspaces*

where $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ is the measurement matrix and $\boldsymbol{z}$ is the measurement. If no further prior is known about $\boldsymbol{s}$, no optimal $\boldsymbol{\Phi}$ can be designed and random measurements yield a detector with the following performance [25]:

$$
P_D(\alpha) \cong Q\left(Q^{-1}(\alpha) - \sqrt{\frac{M}{N} \text{SNR}} \right)
\tag{11}
$$

in which the performance of the detector is degraded by factor $\sqrt{M/N}$ compared with the traditional matched filter. Having knowledge of $\boldsymbol{s} = \boldsymbol{D\theta}$ results in a compressed detector as shown in [25]

$$
P_D(\alpha) \cong Q\left(Q^{-1}(\alpha) - \sqrt{\frac{M}{K} \text{SNR}} \right)
\tag{12}
$$

in which the performance of the detector is improved by a factor of $\sqrt{M/N}$ compared with the random measurement detector. Reference [15] studied two cases about the knowledge of $\boldsymbol{D}$. The first case assumes that $\boldsymbol{D}$ is known and the second case assumes that $\boldsymbol{D}$ consists of a set of parametric basis, where the active basis of $\boldsymbol{D}$ can be recovered by a sparse coding algorithm. Recently, Eldar and Mishali [26] investigated the problem of detection of an ULRSs via compressed measurements. The compressed detector still performs worst than the matched filter by factor $\sqrt{M/N}$.

In this paper, we are going to exploit the low-rank structure characteristic of the signals to design a new detector. Our detector is not compressed and the goal is to design a generalised detector using sparsity (that is, assuming a structure) which implicitly exists in the signals. In Section 3, the proposed detector will be presented. Our detector first assumes a model according to sparse signals and then derives an optimum rule of detection.

4 Proposed approach

In this section, we introduce our model for signal detection. We want to distinguish between two hypotheses $\mathfrak{H}_0$ and $\mathfrak{H}_1$

$$
\begin{aligned}
\mathfrak{H}_0 &: y = n, & \text{signal absence} \\
\mathfrak{H}_1 &: y = Dx + e + n, & \text{signal presence}
\end{aligned}
\tag{13}
$$

where $\boldsymbol{D}$ is the dictionary which can be interpreted as a bank of matched filters, $\boldsymbol{e}$ is the error vector of the model which denotes the mismatch between the exact matched filter and the UoSs spanned by the columns of $\boldsymbol{D}$. Assume that $\boldsymbol{e}$ is a zero-mean white Gaussian noise with variance σ_e^2, that is, $\boldsymbol{e} \sim N(0, \sigma_e^2 \boldsymbol{I})$. In our method, the signal ($\boldsymbol{Dx}$) matched to the observed signal ($\boldsymbol{y}$) is unknown; so it must be determined. This section is divided into four sections. In the first section, we analyse the role of the coefficients of the linear combination ($\boldsymbol{x}$) and then describe our approach for coefficients estimation. In the second section, the performance of our proposed detection method will be analysed. Since dictionary learning is a critical issue in the model, third section is allocated for discussing on the dictionary learning. In the last section, we will explain how our method may become robust to detect signals that are contaminated by gross errors.

5 Discussion on the coefficients (*x*)

Linear combination of the dictionary atoms generates the matched signal for detection. Three cases are considered for $\boldsymbol{x}$ estimation. First, no constraint solution, second matched filter bank and third applying Gaussian distribution. First, assume that there is no constraint on $\boldsymbol{x}$, that is, orthogonal projection of the signal onto the span of the desired subspace. This method is used in matched subspace method to identify the part of signal that amounts for the desired signals [8]. The solution for it will be

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{Dx}\|_2^2 = (\boldsymbol{D}^\mathrm{T}\boldsymbol{D})^{-1}\boldsymbol{D}^\mathrm{T}\boldsymbol{y} \qquad (14)$$

This answer suffers from over-fitting as some signals that do not contain the target signal may be decomposed in terms of the atoms. More restricted constraints may alleviate this problem. Now let us assume that just one element of $\boldsymbol{x}$ is allowed to be non-zero. This constraint helps reducing over-fitting. By this assumption the problem becomes

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{Dx}\|_2^2 \quad \text{s.t.} \ \|\boldsymbol{x}\|_0 = 1 \qquad (15)$$

The solution will be zero except in the position corresponding to the atom with maximum correlation. This solution is nothing but the traditional matched filter bank. Each matched filter which has more correlation is considered as the matched signal. All correlations are sufficient statistics for the decision. If all the correlations are less than a threshold, no detection is performed.

The third scenario we study is assuming Gaussian prior on $\boldsymbol{x}$. The motivation of considering this assumption for $\boldsymbol{x}$ is to avoid over-learning and moreover having less sensitive coefficients. Estimation of $\boldsymbol{x}$ by the assumption of Gaussian distribution on $\boldsymbol{n}$ and $\boldsymbol{x}$ can be obtained as follows:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{Dx}\|_2^2 + \lambda \|\boldsymbol{x}\|_2^2 = (\boldsymbol{D}^\mathrm{T}\boldsymbol{D} + \lambda \boldsymbol{I})^{-1}\boldsymbol{D}^\mathrm{T}\boldsymbol{y} \qquad (16)$$

This solution for the coefficients of linear combination is the Ridge regression [27]. Solution (15) is the least over-learned and solution (14) is the most over-learned one. It is interesting to see how each of the solutions covers the signal space for learning. Solution (15) provides high learning for few 1D subspaces corresponding to each atom, whereas solutions (14) and (16) provide high learning for many subspaces corresponding to arbitrary selections of the atoms. Involvement of all the atoms to form the matched signal results in detection of undesired signals as the target signal due to the expansion of the matched subspaces. To keep the number of involved atoms limited, we suggest modifying problem (16) as follows:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{Dx}\|_2^2 + \lambda \|\boldsymbol{x}\|_0 \qquad (17)$$

There is a large enough value for λ such that the solution of the above problem is the same as (15). Now we show that this
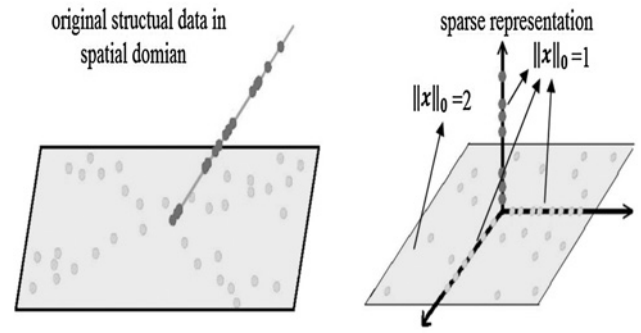

Fig. 4 *Sparse representation of some structural data whose distribution is in agreement with the one defined by (21)*

problem is the maximum a posteriori (MAP) estimation of $\boldsymbol{x}$ under multivariate independent Gaussian prior

$$p(\boldsymbol{x}, \boldsymbol{W}) \propto \sqrt{|\boldsymbol{W}|} \exp(-\boldsymbol{x}^\mathrm{T}\boldsymbol{Wx}) \qquad (18)$$

where $\boldsymbol{W}$ is a diagonal matrix. By this assumption, two unknowns must be estimated. First, we obtain the maximum likelihood (ML) estimation of $\boldsymbol{W}$

$$\boldsymbol{W}_\mathrm{ML} = \underset{\boldsymbol{W}}{\operatorname{argmax}} \ \log p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{W}) = \underset{\boldsymbol{W}}{\operatorname{argmax}} \ \boldsymbol{x}^\mathrm{T}\boldsymbol{Wx} - \log(|\boldsymbol{W}|) \qquad (19)$$

By setting the derivative with respect to $\boldsymbol{W}$ equal to zero, the solution of (19) is $\boldsymbol{W}^{-1} = \boldsymbol{xx}^\mathrm{T}$ which has no solution; however, we need only the diagonal elements of $\boldsymbol{W}$ due to the independent assumption on the entries of $\boldsymbol{x}$. Therefore, calculating the derivative with respect to only diagonal elements of $\boldsymbol{W}$ (w_{ii}) results in

$$w_{ii} = \frac{1}{x_i^2} = \lim_{\delta \to 0} \frac{1}{\delta + x_i^2} \qquad (20)$$

where δ is a small positive for avoiding division by zero. Then we insert the obtained $\boldsymbol{W}$ in (18)

$$p(\boldsymbol{x}) \propto \exp\left(-\sum_i \lim_{\delta \to 0} \frac{x_i^2}{\delta + x_i^2}\right) = \exp(-\|\boldsymbol{x}\|_0) \qquad (21)$$

Actually, $\boldsymbol{W}$ is an auxiliary parameter which is used just for more adaptation of the coefficients distribution. The obtained $\boldsymbol{W}$ results in a distribution with more probability of having orthogonal low-rank subspaces (in the space to which $\boldsymbol{x}$ belongs; for more illustration see Fig. 4). Corresponding to these orthogonal low-rank subspaces, there are non-orthogonal low-rank subspaces in the observation domain which $\boldsymbol{y}$ or $\boldsymbol{Dx}$ belongs to this space. The MAP estimation of $\boldsymbol{x}$ by prior of (21) results in the suggested problem (17) which is a generalised version of (15) from the aspect of sparsity level of the coefficients, and a generalised version of (16) from the aspect of prior distribution on the coefficients for estimation. In [4], it is proved that in a certain condition, problem (17) leads to the same solution with the following ℓ_1 regularised problem:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{Dx}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 \qquad (22)$$

6 Performance analysis

First, we define the false alarm rate and the detection alarm rate

$$\begin{aligned}
P_F &= \Pr(\mathfrak{H}_1 \text{ is chosen while } \mathfrak{H}_0 \text{ is true}) \\
P_D &= \Pr(\mathfrak{H}_1 \text{ is chosen while } \mathfrak{H}_1 \text{ is true}) \\
P_F &= \int_{p(\boldsymbol{y}|\mathfrak{H}_1) > \gamma p(\boldsymbol{y}|\mathfrak{H}_0)} p(\boldsymbol{y}|\mathfrak{H}_0)\mathrm{d}\boldsymbol{y} = \alpha.
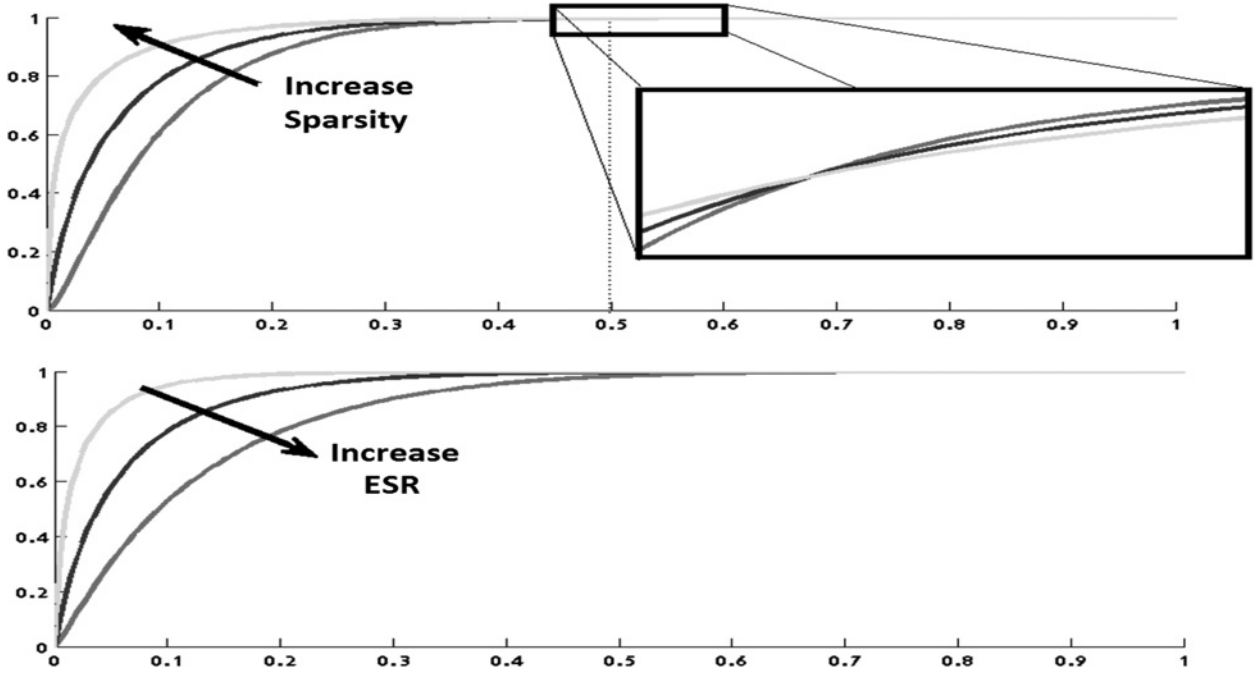\end{aligned} \qquad (23)$$

Fig. 5 *Trade-off between sparsity and error to signal ration (ESR) and its effect on the detector performance*

Parameter γ satisfies the desired amount of false alarm probability, α

$$p(\boldsymbol{y}|\mathfrak{H}_0) = \frac{1}{(2\pi\sigma_n^2)^{N/2}} \exp\left(-\frac{\|\boldsymbol{y}\|_2^2}{2\sigma_n^2}\right)$$

$$p(\boldsymbol{y}|\mathfrak{H}_1) = \frac{1}{(2\pi(\sigma_n^2 + \sigma_e^2))^{N/2}} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{Dx}\|_2^2}{2(\sigma_n^2 + \sigma_e^2)}\right) \quad (24)$$

By solving $p(\boldsymbol{y}|\mathfrak{H}_1) = \gamma p(\boldsymbol{y}|\mathfrak{H}_0)$, the threshold for decision rule can be achieved

$$t = <\boldsymbol{y},\ \boldsymbol{Dx}> \underset{\mathfrak{H}_0}{\overset{\mathfrak{H}_1}{\gtrless}} C + \|\boldsymbol{Dx}\|_2^2 - \|\boldsymbol{y}\|_2^2\left(\frac{\sigma_e^2}{\sigma_n^2}\right), \quad (25)$$

where C is a constant value depending on σ_n^2 and σ_e^2 and the desired α. The sufficient statistic for decision making is $t = <\boldsymbol{y},\ \boldsymbol{Dx}> = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{Dx}$. It is easy to show that

$$P_D(\alpha) = Q\left(Q^{-1}(\alpha) - \sqrt{\frac{\mathrm{SNR}}{1 + \mathrm{ESR}}}\right) \quad (26)$$

where $\mathrm{SNR} = \mathbb{E}\|\boldsymbol{Dx}\|_2^2/\sigma_n^2$ and $\mathrm{ESR} = \sigma_e^2/\mathbb{E}\|\boldsymbol{Dx}\|_2^2$. As can be seen, the performance of the detector is degraded by a factor of $1 + \mathrm{ESR}$. However, our detector has learned a suitable space for signals to be detected. In other words, we accept a small deterioration of the performance due to the generalisation of the detector. Flexibility of the sparse representation-based detector is the most distinguished advantage. Dictionary learning [23] is the most important issue for the methods based on sparse representation. In the sparse detector, the dictionary should be learned such that $\mathrm{ESR} \ll 1$ to avoid performance deterioration and at the same time $\mathrm{ESR} \gg \epsilon$ to avoid over-learning. In the next section, we will explain how to learn an appropriate dictionary. In (25), sparsity has no effect on the performance. Now we introduce a decision rule for detection that exploits the sparsity of the coefficients. To this end, we solve equation $p(\boldsymbol{y}|\boldsymbol{x}, \mathfrak{H}_1)p(\boldsymbol{x}) = \gamma p(\boldsymbol{y}|\mathfrak{H}_0)$ by the obtained $p(\boldsymbol{x})$ in (21). The new decision rule can be achieved as follows:

$$t = <\boldsymbol{y},\ \boldsymbol{Dx}> \underset{\mathfrak{H}_0}{\overset{\mathfrak{H}_1}{\gtrless}} C + \|\boldsymbol{Dx}\|_2^2 - \|\boldsymbol{y}\|_2^2\left(\frac{\sigma_e^2}{\sigma_n^2}\right) + \gamma\|\boldsymbol{x}\|_0 \quad (27)$$

where c is a positive constant value. As $\|\boldsymbol{x}\|_0$ increases, $\mathfrak{H}_0$ may be more probable, because the signals representation in terms of the dictionary would be sparse only for the learned signals. Similar to (26), it is easy to show that

$$P_D(\alpha) = Q\left(f(\|\boldsymbol{x}\|_0)Q^{-1}(\alpha) - \sqrt{\frac{\mathrm{SNR}}{1 + \mathrm{ESR}}}\right) \quad (28)$$

where f is an increasing homogenous function. As can be seen, the probability of detection increases (decreases) when sparsity increases (decreases) for false alarm rates smaller than 0.5 [because when $\alpha < 0.5$ then $Q^{-1}(\alpha) > 0$]. As the desired false alarm rates are often small, the probability of detection would increase in this region (it is favourable for a detector that the top-left region of its receiver operating characteristic (ROC) be close to the ideal ROC). If the representation of a signal is sparse, this signal lies in the desired low-rank subspace (that is, it meets our assumed model for the target signals). Thus the probability of detection would increase for these signals that have sparse representation in terms of the dictionary atoms, which is actually what we expect from sparsity. Fig. 5 shows the ROC of (22) with $\mathrm{SNR} = +20$ dB for different sparsity levels.

Traditional matched filter banks have the most sparsity level, but it is not practical. For instance, in VAD, it is not feasible to collect all possible voices in a bank. A small number of filters result in high ESR and low performance. Our proposed detector makes a trade-off between ESR and sparsity in order to have a good detector performance. Dictionary learning has a critical role in the trade-off which is studied as follows.

7 Learning the dictionary

In this section, we explain the role of dictionary learning in the proposed detection method. In many detection problems, the number of training signals may not be as large as the number of

possible matched filters that cover all the target signals space. By the proposed approach, we search for a dictionary learned by a set of finite number of signals that efficiently represents those signals. The dictionary should be general to be able to deal with a signal that has not been seen before. Assume that we have a set of signals ($Y \in \mathbb{R}^{N \times L}$). Dictionary learning is a function that maps Y to D where $D \in \mathbb{R}^{N \times K}$. An appropriate dictionary should have ESR $\ll 1$ to be a suitable representation for the training data and also ESR should not be too small to have a general dictionary that is not over-learned for only the training data. Two algorithms for dictionary learning are presented.

7.1 K-means algorithm

K-means method uses K centroids of clusters to characterise the training data [28]. They are determined by minimising the sum of squared errors

$$D = \underset{D}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{i \in C_k} \| y_i - d_k \|_2^2 \qquad (29)$$

where the columns of D are d_k, $k = 1, \ldots, K$. The provided dictionary assigns to each training data a centroid. K should be large enough to satisfy the desired amount of ESR. Problem (14) has to be solved to determine the coefficients so that only one of them is non-zero. This dictionary learns some points in the signal space. As the distance from these points increases, the level of learning would decrease. In other words, this dictionary is obtained by the union of spheres model. This model may not be a suitable choice for ordinary signals. The next algorithm agrees with a more appropriate model for the data. The K-singular value decomposition (SVD) learns the signal space with an ULRSs.

7.2 K-SVD algorithm

By extending the union of spheres to a union of low-dimensional subspaces, K-means algorithm is generalised to K-SVD algorithm [23]. This flexible model agrees with many signals such as images and audio signals. For example, natural images have sparse representation in terms of DCT dictionary. In other words, by combination of only a few DCT bases, it is possible to approximate the blocks of an image. The following problem provides the dictionary learned by K-SVD:

$$D = \underset{D, x_i}{\operatorname{argmin}} \sum_{i} \| y_i - D x_i \|_2^2 \text{ s.t. } \| x_i \|_0 \leq T. \qquad (30)$$
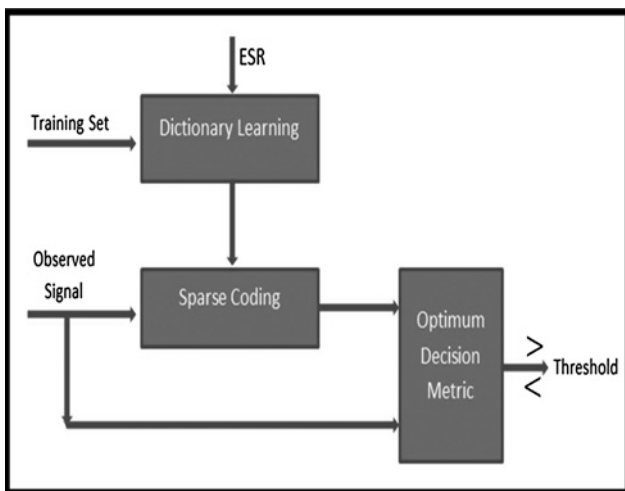

Fig. 6 *Block diagram of our proposed detector based on dictionary learning*

This algorithm is based on atom-by-atom updating over the columns of D. Recently, more efficient algorithms for atom-by-atom updating are suggested in [29]. Each arbitrary selection of few D columns characterises a cluster corresponding to a subspace. The dictionary learned by K-SVD is in agreement with the proposed problem (22). After learning, test signals that lie on the learned low-dimensional subspaces can be reconstructed and detected. In addition to dictionary learning using training signals, it is possible to design a dictionary using parametric functions [30]. Kernels of fast fourier transform (FFT) and DCT are two examples from this class of dictionaries where bases sweep the parameter of frequency. Fig. 6 shows the block diagram of the proposed detection method.

8 Robustness

Assume that a dictionary has learnt to detect face images without sun glasses. If a face image with sun glasses is given to it for detection, gross error in the region of eyes may result in a wrong detection. To solve this problem, a distribution has to be supposed that has longer tail than Gaussian. Laplace distribution is our suggestion for the error distribution. Thus $\mathfrak{H}_1$ implies that the observed signal is the combination of few atoms of D, a Laplace distributed error and a Gaussian distributed noise

$$\mathfrak{H}_1 : y = Dx + e + n \qquad (31)$$

The problem of coefficients estimation for (22) by new prior assumption has been already presented in robust statistics [31]

$$\hat{x} = \underset{x}{\operatorname{argmin}} \| y - Dx \|_H + \lambda \| x \|_1 \qquad (32)$$

where

$$\| s \|_H = \sum h(s_i)$$
$$h(s) = \begin{cases} s^2, & \text{if } |s| \leq \rho \\ \rho s, & \text{if } |s| > \rho \end{cases} \qquad (33)$$

In other words, small errors and large errors are penalised by ℓ_2 norm and ℓ_1 norm, respectively. λ is the parameter of the mixture distribution of Gaussian and Laplace. Let re-write (32) as follows:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \| y - Dx - e \|_2^2 + \lambda \| x \|_1 + \rho \| e \|_1 \qquad (34)$$

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\| y - [D \ I] \begin{bmatrix} x \\ e \end{bmatrix} \right\|_2^2 + \lambda \left(\| x \|_1 + \| \frac{\rho}{\lambda} e \|_1 \right)$$

Let us define $\tilde{e}$ as $\rho / \lambda e$

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\| y - \left[D \ \frac{\lambda}{\rho} I \right] \begin{bmatrix} x \\ \tilde{e} \end{bmatrix} \right\|_2^2 + \lambda (\| x \|_1 + \| \tilde{e} \|_1) \qquad (35)$$

By substitution of $B = [D \lambda / \rho I]$ and $z = [x^T \tilde{e}^T]^T$, we have

$$\hat{z} = \underset{z}{\operatorname{argmin}} \| y - Bz \|_2^2 + \lambda \| z \|_1 \qquad (36)$$

This problem is similar to (22) except that its dictionary is extended by scaled identity matrix. Identity matrix projects inappropriate parts of the signals onto corresponding coefficients. Inappropriate parts of the signals may be large errors or out of the desired subspace interferences or outlier data. Wright *et al.* [32] also intuitively have used the same dictionary to obtain a robust framework for face recognition. A same procedure can be pursued to learn robust dictionary by a set of unreliable data [33].
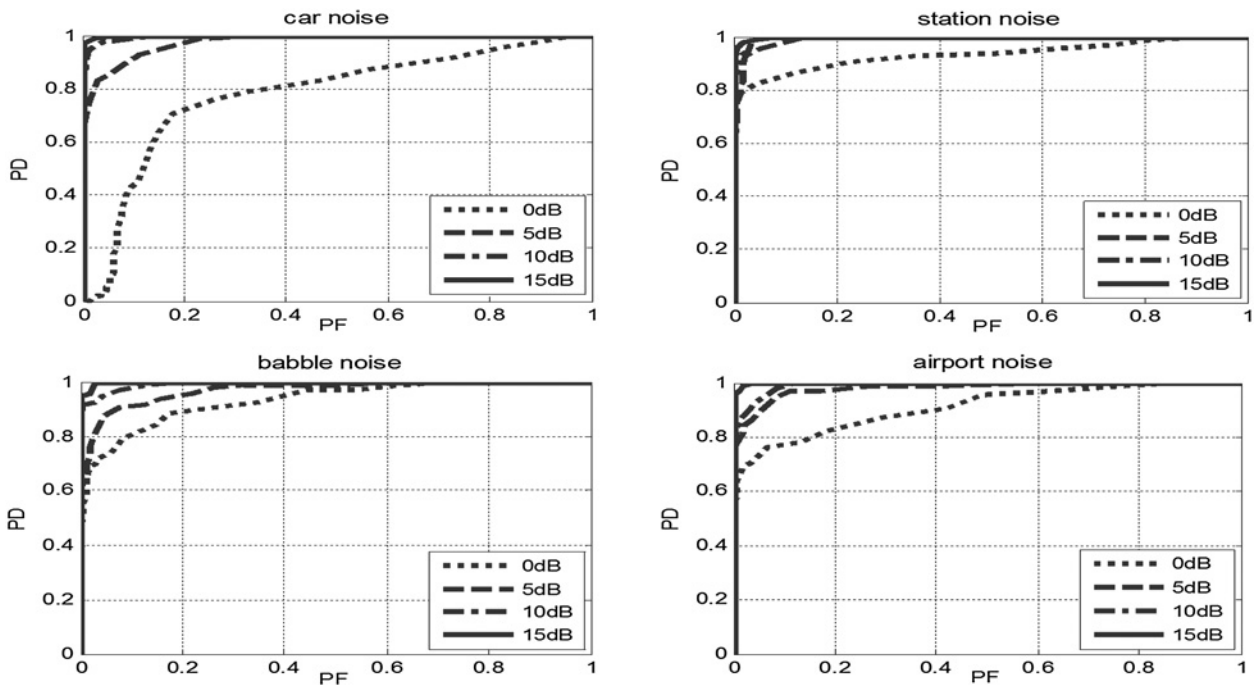
Fig. 7 *Block diagram of our proposed detector based on dictionary learning*

9 Experimental results

We evaluated the performance of our proposed method in the case study of VAD. To construct the learned dictionary, clean speech signals of noisy speech corpus (NOIZEUS) database were used [34]. In NOIZEUS database, 30 sentences were selected which include all phonemes in the American English language. The sentences were produced by three male and three female speakers and originally sampled at 25 kHz and down-sampled to 8 kHz. We divided the clean speech signals into 25 ms frames with 10 ms frame shift. After removing the silent frames, we extracted standard Mel-frequency cepstral coefficients (MFCCs) using ten Mel triangular filters, energy values computed at each of the ten Mel triangular filters, total energy (the first cepstral coefficient) and entropy from each speech frame. MFCC features capture the most relevant information of speech signal, and they are widely used in speech and speaker recognition making the VAD method easy to integrate with existing applications. Therefore, our features vector was 24D, and the total number of vectors was about 6300. By using the K-SVD algorithm, we obtained a learned dictionary with 100 atoms, which was used in the following experiments for obtaining the sparse representation based on orthogonal matching pursuit (OMP) method.
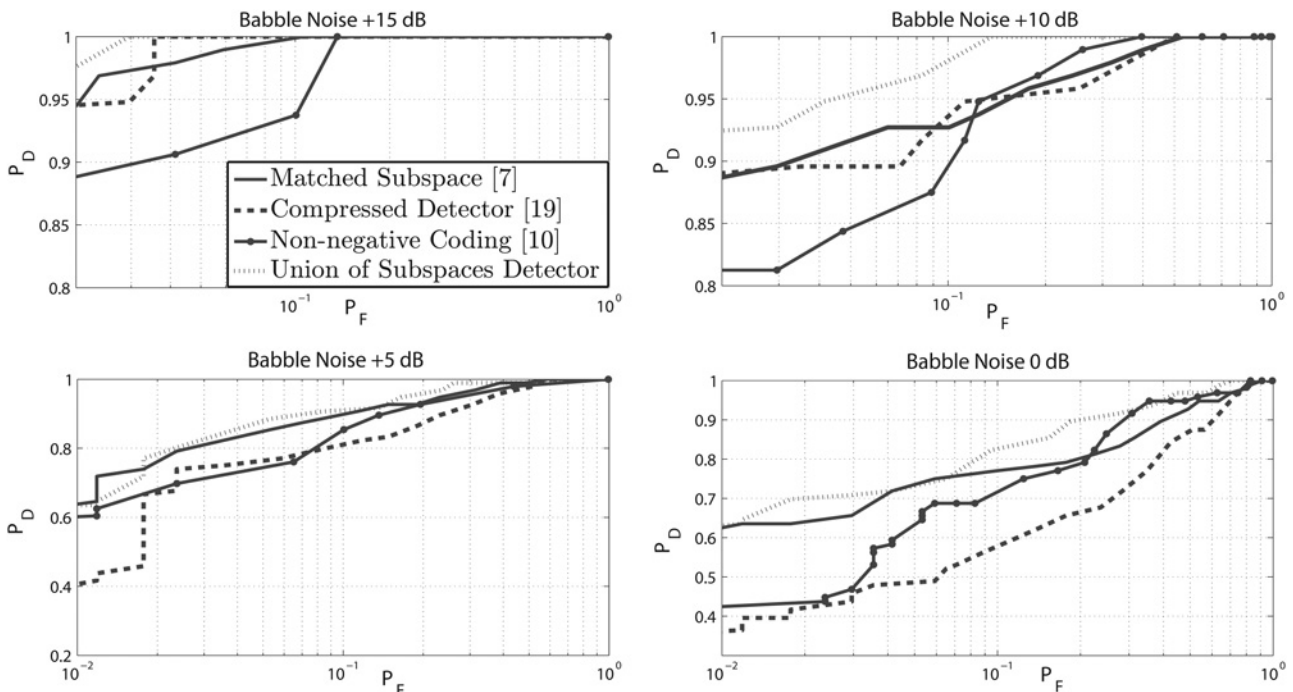

Fig. 8 *Comparing the performance of our proposed method with sparse non-negative coding-based VAD [11], matched subspace detector [8] where the bases are found by PCA and detection using the compressed measurements [20]*

To evaluate the performance of the proposed method, the speech detection probability (PD) and false alarm probability (PF) were investigated based on a reference decision. A clean test speech (sp10.wav), taken from the NOIZEUS database, was down-sampled at 8000 Hz and was used for the reference decisions. To simulate noisy environments, several noise signals as the subset of the NOIZEUS database were used. Noise signals included recordings from different places (Babble (crowd of people), Car, …) at SNRs of 0, 5, 10 and 15 dB. The ROC curves for VAD using our proposed method are illustrated in Fig. 7 which shows PD versus PF.

Sparsity in VAD has been exploited already. For example, a feature extraction is performed to suggest a decision rule for detection in [11]. We compared the result of our method with the sparsity-based VAD method proposed in [11]. As can be seen in Fig. 8, our method shows better performance in low SNR conditions.

10 Conclusion

This paper presented a new sparsity-based detector. The performance of the method was evaluated in a realistic application: VAD in speech signal processing. Our detector proposed a new trade-off for designing detectors by assuming the ULRSs model. The trade-off is between the sparsity and the error of ULRSs model denoted by ESR. In our detector the number of filter banks is proportional to the size of the dictionary. Appropriate dictionary is able to regularise the sparsity and the introduced parameter ESR. Simulation results showed that the proposed method is effective and has a high anti-noise ability due to optimum projection of signals to reliable learned low-rank subspaces.

11 References

1 Elad, M.: 'Sparse and redundant representations: from theory to applications in signal and image processing' (Springer Publishing Company, Incorporated, 2010, 1st edn.)
2 Bruckstein, A.M., Donoho, D.L., Elad, M.: 'From sparse solutions of systems of equations to sparse modeling of signals and images', *SIAM Rev.*, 2009, **51**, pp. 34–81
3 Rahmoune, A., Vandergheynst, P., Frossard, P.: 'Sparse approximation using m-term pursuit and application in image and video coding', *IEEE Trans. Image Process.*, 2012, **21**, (4), pp. 1950–1962
4 Elad, M., Aharon, M.: 'Image denoising via sparse and redundant representations over learned dictionaries', *Trans. Image Process.*, 2006, **15**, pp. 3736–3745
5 Comon, P., Jutten, C.: 'Handbook of blind source separation: independent component analysis and applications' (Academic Press, 2010, 1st edn.)
6 Mairal, J., Bach, F., Ponce, J., *et al.*: 'Supervised dictionary learning', *Adv. Neural Process. Syst.*, 2008, pp. 1033–1040
7 Benyassine, A., Shlomot, E., u Su, H., *et al.*: 'Itu-t recommendation g.729 annex b: a silence compression scheme for use with g.729 optimized for v.70 digital simultaneous voice and data applications', *IEEE Commun. Mag.*, 1997, **35**, pp. 64–73
8 Scharf, L., Friedlander, B.: 'Matched subspace detectors', *IEEE Trans. Signal Process.*, 1994, **42**, pp. 2146–2157
9 Kraut, S., Scharf, L., McWhorter, L.: 'Adaptive subspace detectors', *IEEE Trans. Signal Process.*, 2001, **49**, pp. 1–16
10 Ahmadi, P., Khoram, S., Joneidi, M., *et al.*: 'Discovering motion patterns in traffic videos using improved group sparse topical coding'. 2014 Seventh Int. Symp. on Telecommunications (IST), September 2014, pp. 343–348
11 Teng, P., Jia, Y.: 'Voice activity detection via noise reducing using non-negative sparse coding', *IEEE Signal Process. Lett.*, 2013, **20**, pp. 475–478
12 Le, T., Luu, K., Savvides, M.: 'SparCLeS: Dynamic ℓ_1 sparse classifiers with level sets for robust beard/moustache detection and segmentation', *IEEE Trans. Image Process.*, 2013, **22**, pp. 3097–3107
13 Shim, B., Song, B.: 'Multiuser detection via compressive sensing', *IEEE Commun. Lett.*, 2012, **16**, pp. 972–974
14 Duarte, M., Davenport, M., Wakin, M., *et al.*: 'Sparse signal detection from incoherent projections'. 2006 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proc., May 2006, vol. 3, p. III
15 Davenport, M., Boufounos, P., Wakin, M., *et al.*: 'Signal processing with compressive measurements', *IEEE J. Sel. Top. Signal Process.*, 2010, **4**, pp. 445–460
16 Lu, Y., Do, M.: 'A theory for sampling signals from a union of subspaces', *IEEE Trans. Signal Process.*, 2008, **56**, pp. 2334–2345
17 Liu, G., Lin, Z., Yan, S., *et al.*: 'Robust recovery of subspace structures by low-rank representation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, pp. 171–184
18 Mallat, S., Zhang, Z.: 'Matching pursuits with time-frequency dictionaries', *IEEE Trans. Signal Process.*, 1993, **41**, pp. 3397–3415
19 Tropp, J., Wright, S.: 'Computational methods for sparse solution of linear inverse problems', *Proc. IEEE*, 2010, **98**, pp. 948–958
20 Davenport, M.A., Wakin, M.B., Baraniuk, R.G.: 'Detection and estimation with compressive measurements', Technical Report, 2006
21 Kelly, E.: 'An adaptive detection algorithm', *IEEE Trans. Aerosp. Electron. Syst.*, 1986, **AES-22**, pp. 115–127
22 Sadeghi, M., Joneidi, M., Babaie-Zadeh, M., *et al.*: 'Sequential subspace finding: a new algorithm for learning low-dimensional linear subspaces'. 2013 Proc. of the 21st European Signal Processing Conf. (EUSIPCO), September 2013, pp. 1–5
23 Aharon, M., Elad, M., Bruckstein, A.: 'K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation', *IEEE Trans. Signal Process.*, 2006, **54**, pp. 4311–4322
24 Rubinstein, R., Peleg, T., Elad, M.: 'Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model', *IEEE Trans. Signal Process.*, 2013, **61**, pp. 661–677
25 Wang, Z., Arce, G., Sadler, B.: 'Subspace compressive detection for sparse signals'. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008, March 2008, pp. 3873–3876
26 Eldar, Y., Mishali, M.: 'Robust recovery of signals from a structured union of subspaces', *IEEE Trans. Inf. Theory*, 2009, **55**, pp. 5302–5316
27 Hoerl, A.E., Kennard, R.W.: 'Ridge regression: biased estimation for nonorthogonal problems', *Technometrics*, 1970, **12**, pp. 55–67
28 Jain, A.K.: 'Data clustering: 50 years beyond k-means', *Pattern Recognit. Lett.*, 2010, **31**, pp. 651–666
29 Sadeghi, M., Babaie-Zadeh, M., Jutten, C.: 'Learning overcomplete dictionaries based on atom-by-atom updating', *IEEE Trans. Signal Process.*, 2014, **62**, pp. 883–891
30 Yaghoobi, M., Daudet, L., Davies, M.: 'Parametric dictionary design for sparse coding', *IEEE Trans. Signal Process.*, 2009, **57**, pp. 4800–4810
31 Huber, P., Wiley, J., InterScience, W.: 'Robust statistics' (Wiley, New York, 1981)
32 Wright, J., Yang, A., Ganesh, A., *et al.*: 'Robust face recognition via sparse representation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, pp. 210–227
33 Amini, S., Sadeghi, M., Joneidi, M., *et al.*: 'Outlier-aware dictionary learning for sparse representation'. 2014 IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP), September 2014, pp. 1–6
34 Loizou, P.C.: 'Speech enhancement: theory and practice' (CRC Press, Inc., Boca Raton, FL, USA, 2013, 2nd edn.)