A Union of Low-Rank Subspaces Detector

Mohsen Joneidi, Parvin Ahmadi, Mostafa Sadeghi
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
joneidi@sipincom.com, parvinahmadi@ee.sharif.ir, m.saadeghi@gmail.com

Abstract

Sparse signal representation and approximation has received a lot of attention during the last few years. This is due to its applicability and high performance in many signal processing areas. In this paper, we propose a new detection method based on sparse decomposition in a union of subspaces (UoS) model. Our 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 generalization for detectors. Low-rank assumption for the desired signals implies that the representations of these signals in terms of proper basis vectors would be sparse. Our proposed detector also exploits sparsity in its decision rule. We demonstrate the high efficiency of our method in the cases of voice activity detection in speech processing.

Keywords: signal detection, sparse representation (SR), dictionary learning, union of subspaces model
1. Introduction

In digital signal processing applications, sparse representation has attracted many attentions due to its benefits and high flexibility [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. In methods based on exploiting the signal sparsity, 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, i.e., the one that contains as few as possible non-zero coefficients. 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.

In this paper, we propose a new signal detection method based on the union of low-rank subspaces (ULRS) model [7, 8]. This model is able to reveal intrinsic structure of a set of signals. The proposed detector is a generalized 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 view in order to show relation between of 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 the paper is organized 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 the paper with a summary of the proposed work.

2. Theoretical background and review

In this section we review theoretical background for theory of sparse decomposition and theory of detection.

2.1 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 as possible basis functions. Each basis function is called an atom and their collection is called a dictionary [9]. This dictionary is usually over-completed, i.e., the number of atoms is (much) more than the dimension of atoms.
Specifically, let $y \in \mathbb{R}^n$ be the signal which is going to be sparsely represented over the dictionary $D \in \mathbb{R}^{n \times K}$ with $K > n$. This amounts to solving the following problem:

$$P_0: \min_{x} ||x||_0 \quad \text{subject to} \quad y = Dx$$

(1)

where $||.||_0$ stands for the so-called $\ell_0$ pseudo-norm which counts the number of nonzero 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 [10]).

For a specified class of signals, e.g. 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 specified class of signals, over-complete DCT dictionary for the class of natural images is an example. These non-adaptive dictionaries are favorable because of their simplicity. In other hand, learning based dictionary results in better matching the contents of the signals [1]. Most dictionary learning algorithms are indeed a generalization of the clustering algorithms. While in clustering each training signal is forced to assign only one atom (cluster center), in the dictionary learning problem each signal is allowed to use more than one atom provided that it uses as fewest as possible atoms. The general dictionary learning problem can be stated as the following:

$$\min_{D, X} ||Y - DX||_F^2 \quad \text{st: } \forall i: ||x_i||_0 \leq T$$

(3)

The various dictionary learning algorithms solve the above problem by alternatively minimizing it over $D$ and $X$. The dictionary learning algorithms differ mainly in performing
the minimization over the dictionary. Dictionary learning has an important role in the sparse decomposition based method. A subsection in the proposed method section is allocated for discussion on dictionary learning.

2.2 Basic theory of Detection

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

\[ \mathcal{H}_0 : \ y = n \quad \text{Signal absence} \]
\[ \mathcal{H}_1 : \ y = s + n \quad \text{Signal presence} \]  \hspace{2cm} (4)

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 function of \( p(y|\mathcal{H}_0) \) and \( p(y|\mathcal{H}_1) \) are known. In this case the likelihood ratio test (LRT) gives

\[ \frac{p(y|\mathcal{H}_1)}{p(y|\mathcal{H}_0)} \leq \frac{\gamma}{\mathcal{H}_1} \quad (5) \]

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

\[ s^H R^{-1} y \leq \gamma_1 \quad (6) \]

where \( R \) is the covariance matrix. If it's not known in advance, it must be determined by obtaining sample covariance matrix in the above test. Probability of detection is then equal to:
\[ P_D(\alpha) = Q(Q^{-1}(\alpha) - \sqrt{\text{SNR}}) \]  
(7)

in which \( \alpha \) is the probability of false alarm and \( Q(\alpha) = \int_\alpha^{+\infty} \frac{1}{\sqrt{\pi}} \exp(-t^2) \, dt \).

Another well-known detector is Generalized LRT [11] (GLRT) which is derived by maximizing conditional densities constituting the likelihood ratio test with respect to unknown parameters. The following detection criterion is obtained by assuming covariance matrix to be unknown

\[ \frac{|s^H \hat{R}^{-1} y|}{s^H \hat{R}^{-1}s(1 + \frac{1}{K^2} \hat{R}^{-1} s s^H \hat{R}^{-1} y)} \leq \gamma \]  
(8)

where \( K \) is the number of snapshots available for \( \hat{R} \) estimation. In [11] no optimality has been claimed for GLRT. However, Scharf and Friedlander have shown that GLRT is uniformly most powerful (UMP) invariant [12]. This is the strongest statement of optimality derived for a detector. GLRT detector may be interpreted as a projection on null space of interference followed by a matched subspace detector [12]. Consider the following model for hypothesis test

\[ \mathcal{H}_0 : y = C\theta + n \]  
(9)
\[ \mathcal{H}_1 : y = Dx + C\theta + n \]

where \( C \in \mathbb{R}^{N \times p} \) pans the background or interference subspace, \( p < N \) and \( \theta \) determines contribution of each columns of \( C \). \( D \in \mathbb{R}^{N \times m} \) spans signal subspace which is to be detected, \( m < N \) and \( x \) determines contribution of each column of \( D \). It is obvious that if \( p \geq N \) or \( m \geq N \) \(^1\) 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.

\(^1\)Assume \( C \) and \( D \) are full rank in this case.
On the other hand, restriction 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

$$\mathbf{y}^H \mathbf{P}^\perp \mathbf{P}_{DC}^\perp \mathbf{P}^\perp \mathbf{y} \leq \gamma$$ (10)

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

![Figure 1: Block diagram of matched subspace detector](image)

At the conclusion of paper [12] authors mentioned that basis can be extracted from Discrete Cosine Transform, Wavelet Transform or learned by data dependent analysis like Principal Component Analysis (PCA). Using such basis provides a matched subspace for whole desired signals which are going to be detected. For more illustration refer to figure 2. This figure shows composites of some 3D data by signal and non-signal (interference and noise) parts. Two low-rank subspaces are shown corresponding to rank-1 and rank-2 subspace (the low-rank matched subspace) which are obtained by PCA.
Figure 2. Two low-rank subspaces learned by PCA for some 3D signals

The main contribution of [12] may be answering 'no' to the question, 'Can the GLRT be improved upon?' while they did not assume any prior on structure of the low-rank matched filter. The structural assumption can be applied by assuming a sparse prior on the coefficients of $\theta$ and $x$. The recent new model of sparse representation has attracted much attention due to its flexibility and conformity for many signals. The proposed method of this paper suggests using the model of union of low-rank subspaces for signals due to its suitable fitness which has been proven in many signal processing applications. Instead of
traditional analysis like PCA, modern analysis like the methods proposed in [13] and [14] can be exploited in order to recover suitable bases spanning the union of low-rank subspaces. Figure 3 shows a union of matched low-rank subspaces corresponding to the data of figure 2.

![Figure 3. A union of rank-1 subspaces provides suitable matched subspaces](image)

Compressive detection is another application of sparse theory exploited in signal detection and studied in [15] and [16]. Instead of dealing with the whole samples of the signal, compressed detector works with few measurements. This detector distinguish between two hypothesis,

\[ H_0 : z = \Phi n \]
\[ H_1 : z = \Phi (s + n) \]

where \( \Phi \in \mathbb{R}^{M \times N} \) is the measurement matrix and \( z \) is the measurement. If no further prior is known about \( s \), no optimal \( \Phi \) can be designed, and a random measurement yield a detector with the following performance
in which, the performance of the detector is degraded by factor of $\sqrt{M/N}$ compared to traditional matched filter. Having knowledge of $s = D\theta$ results in a compressed detector that as shown in [15]

$$P_D(\alpha) \approx Q(Q^{-1}(\alpha) - \frac{M}{\sqrt{N\text{SNR}}})$$ (7)

in which, the performance of the detector is improved by a factor of $\sqrt{N/K}$ compared to random measurement detector. Reference [16] studied two cases about knowledge of D. The first case assumes that D is known. The second case assumes that D consists of a set of parametric basis, where the active basis of D can be recovered by a sparse coding algorithm. Recently, [17] investigated the problem of detection of a union of low-rank subspaces via compressed measurements. Compressed detector is still performs worse than matched filter by a factor of $\sqrt{M/K}$ (we know K>M).

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

3. The Proposed Approach
In this section, we introduce our model for signal detection. We want to distinguish between two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$: 

\[
\mathcal{H}_0 : \ y = n \\
\mathcal{H}_1 : \ y = Dx + e + n
\]  

(11)

where, $D$ is the dictionary which can be interpreted as a bank of matched filters. $e$ is the error vector of the model which denotes the mismatch between the exact matched filter and the union of subspaces spanned by the columns of $D$. Assume that $e$ is a zero-mean white Gaussian noise with variance $\sigma_e^2$, i.e. $e \sim N(0, \sigma_e^2)$. In our method, the signal $(Dx)$ matched to the observed signal $(y)$ is unknown; so it must be determined.

This section is divided into four subsections. In the first subsection, we analyze the role of the coefficients of the linear combination $(x)$ and then describe our approach for coefficients estimation. The main contribution of our paper is in this part. In the second subsection, the performance of our proposed detection method will be analyzed. Since dictionary learning is a critical issue in the model, third subsection is allocated for discussing on dictionary learning. In the last subsection we will explain how our method may become robust to detect signals that are contaminated by gross errors.

3.1 A discussion on the coefficients $(x)$

Linear combination of the dictionary atoms generates the matched signal for detection. But a question must be answered before. How we are allowed to combine the atoms for making the matched signal. To answer this question, first, we consider three scenarios in which in all of them the noise distribution is assumed to be Gaussian. First assume that
there is no constraint on \( x \) (orthogonal projection of the signal onto the span of the atoms).

The answer for it will be,

\[
x = \arg\min_x ||y - Dx||_2 = (D^TD)^{-1}D^Ty
\]  

(8)

This answer suffers from over-learning as some signals that do not contain the target signal may be decomposed in terms of the atoms. More restrict constraints may alleviate this problem. Now let assume that just one element of \( x \) is allowed to be none zero. This constraint helps reducing over-learning. By this assumption the problem becomes,

\[
x = \arg\min_x ||y - Dx||_2 \text{ st: } ||x||_0 = 1
\]  

(9)

The solution for the coefficients will be zero except one of them that have more 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 the sufficient statistic for the decision. If all correlations are less than a threshold, no detection is done.

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

\[
x = \min_x ||y - Dx||_2^2 + \lambda ||x||_2^2 = (D^TD + \lambda I)^{-1}D^Ty
\]  

(10)
This solution for the coefficients of linear combination is the Ridge regression \[18\]. Solution (9) is the least over-learned and solution (8) is the most over-learned one. It is interesting to see how each of the solutions covers the signal space for learning. Solution (9) provides high learning for few one dimensional subspaces corresponding to each atom, while solutions (8) and (10) provide high learning for many subspaces corresponding to arbitrary selection of atoms. Involvement of all atoms to form the matched signal result in detection of undesired signals as the target signal due to the expansion of learned region in the signals space. To keep the number of involved atoms limited, we suggest modifying problem (10) as follows:

\[
x = \min_x \| y - Dx \|^2 + \lambda \| x \|_0
\]  

\[(11)\]

There is a value for \( \lambda \) such that the solution of the above problem is the same as (9).

Now we show that this problem is the MAP estimation of \( x \) under multivariate independent Gaussian prior,

\[
p(x, W) \propto \sqrt{|W|} \exp(-x^TWx)
\]  

\[(12)\]

where \( W \) is a diagonal matrix. By this assumption, two unknowns must be estimated. First we obtain the ML estimation of \( W \):

\[
W_{ML} = \max_W (y|x, W) = \min_W x^T W x - \ln(|W|) \rightarrow W^{-1} = xx^T
\]  

\[(13)\]

This equation has no solution, but as we want only the diagonal elements of \( W \) (because we have assumed multivariate independent Gaussian prior), we calculate derivative respect to only diagonal elements of \( W \) \( (w_{ii}) \):
$w_{il} = \frac{1}{x_i^2 + \delta}$

(14)

$\delta$ is a small positive for avoiding division by zero. Then we put the obtained $W$ in (12):

$p(x) \propto exp(-\|x\|_0)$

(15)

Figure 4. Sparse representation of some structural data whose distribution is in agreement with the one defined by Eq. (15)

Actually $W$ is a hidden parameter which is used just for more adaptation of the coefficients distribution. The obtained $W$ results in a distribution with more probability of having orthogonal low rank subspaces (in the space to which $x$ 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 (e.g., $Dx$ belongs to this space).

The MAP estimation of $x$ by the prior of (15) results in the suggested problem (11). Problem (11) is a generalized version of (9) from aspect of sparsity level of the coefficients,
and a generalized version of (10) from aspect of prior distribution on the coefficients for estimation. In [1], it is proved that in a certain condition, problem (11) leads to the same solution with the following regularized norm 1 problem:

\[ \mathbf{x} = \min_{\mathbf{x}} \| \mathbf{y} - D\mathbf{x} \|_2^2 + \lambda \| \mathbf{x} \|_1 \]  \hspace{1cm} (16)

### 3.2 Performance analysis

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

\[ P_F = Pr(H1 \text{ chosen when } H0 \text{ true}) \]  \hspace{1cm} (17)

\[ P_D = Pr(H1 \text{ chosen when } H1 \text{ true}) \]

\[ P_F = \int_{p(\mathbf{y}|\mathcal{H}_1) > P(\mathbf{y}|\mathcal{H}_0)} p(\mathbf{y}|\mathcal{H}_0) d\mathbf{y} = \alpha \]

Parameter \( \gamma \) satisfies the desired amount of \( P_F \) (\( \alpha \)).

\[ p(\mathbf{y}|\mathcal{H}_0) = \frac{1}{(2\pi \sigma_n^2)^{N/2}} \exp\left(-\frac{\|\mathbf{y}\|_2^2}{2\sigma_n^2}\right) \]  \hspace{1cm} (18)

\[ p(\mathbf{y}|\mathcal{H}_1) = \frac{1}{(2\pi (\sigma_n^2 + \sigma_e^2))^{N/2}} \exp\left(-\frac{\|\mathbf{y} - D\mathbf{x}\|_2^2}{2(\sigma_n^2 + \sigma_e^2)}\right) \]

As can be seen in equation (19) \( t = (\mathbf{y}, D\mathbf{x}) \) is the sufficient statistic. By solving \( p(\mathbf{y}|\mathcal{H}_1) = \gamma \cdot p(\mathbf{y}|\mathcal{H}_0) \), the threshold for decision rule can be achieved.
\[
\mathcal{H}_0 \\
\mathcal{H}_1
\]
\[
t = \langle y, Dx \rangle \leq C + \|Dx\|_2^2 - \|y\|_2^2 \left( \frac{\sigma_e^2}{\sigma_n^2} \right)
\]

where C is a constant value depending on \( \sigma_n^2 \) and \( \sigma_e^2 \) and the desired \( \gamma \). It is easy to show that:

\[
P_D(\alpha) = Q(Q^{-1}(\alpha) - \frac{\text{SNR}}{\sqrt{1 + \text{ESR}}})
\]

where \( \text{SNR} = \frac{E(\|Dx\|_2^2)}{\sigma_n^2} \) and \( \text{ESR} = \frac{\sigma_e^2}{E(\|Dx\|_2^2)} \). As can be seen, the performance of the detector is degraded by a factor of \( 1 + \text{ESR} \). But 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 generalization of the detector. Flexibility of the sparse representation based detector is the most distinguished advantage. Dictionary learning [13] is the most important issue for the methods based on sparse representation. In the sparse detector, dictionary should be learned such that \( \text{ESR} \ll 1 \) to avoid performance deterioration and at the same time \( \text{ESR} \gg \epsilon \) to avoid over-learning. In the next section we will explain how to learn an appropriate dictionary. In (19), sparsity of the coefficients does not have any 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 \( p(y|x, \mathcal{H}_1)p(x) = \eta \cdot p(y|\mathcal{H}_0) \) by the obtained \( p(x) \)in the equation (15). New decision rule can be achieved as follows:

\[
\mathcal{H}_0 \\
\mathcal{H}_1
\]
\[
t = \langle y, Dx \rangle \leq C - \|y\|_2^2 \left( \frac{\sigma_e^2}{\sigma_n^2} \right) + \|Dx\|_2^2 + \gamma \|x\|_0
\]
where $\gamma$ is a positive value which depends on parts of the distribution of $\mathbf{x}$ not mentioned in (15). As $\|\mathbf{x}\|_0$ increases, $\mathcal{H}_0$ may be more probable, because the signals representation in terms of dictionary would be sparse only for the learned signals. Similar to (33), it is easy to show that:

$$P_d(\alpha) = Q \left( f(\|\mathbf{x}\|_0) Q^{-1}(\alpha) - \frac{\sqrt{\text{SNR}}}{\sqrt{1 + ESR}} \right)$$

(22)

where $f$ is an increasing homogenous function. As can be seen, the probability of detection increases (decreases) when sparsity increases\(^2\) (decreases) for false alarm rates smaller than $\frac{1}{2}$ (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 favorable for a detector that the top-left region of its 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. Figure 5 shows the ROC of (22) with SNR=+20dB for different sparsity levels.

\(^2\)That is, the $\|\mathbf{x}\|_0$ decreases.
Traditional matched filter banks have the most sparsity level, but it is not applicable for many signals detection. For instance, in voice activity detection, it is not possible to collect all possible voices in a bank. Small number of filters results 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. Learning dictionary has a critical role in the trade-off which is studied in the following.

4.3 Learning the dictionary
In this section we explain the role of dictionary learning in the proposed detection method. In many detection problems, 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 speech signals that efficiently represents the speech signals. The dictionary should be general to be able to deal with a speech signal that is not ever seen before. Assume that we have a set of signals \(X \in R^{n \times L}\) that the dictionary must be adapted for them. Dictionary learning is a function that map \(X \rightarrow D\) where \(D \in R^{n \times K}, n < K \ll L\). 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-leaned for only the training data. Two algorithms for dictionary learning are presented.

- K-means algorithm

  K-means method uses K centroids of clusters, to characterize the training data [19]. They are determined by minimizing the sum of squared errors.

  \[
  D = min_D \sum_{k=1}^{K} \sum_{i \in C_k} \|x_t - d_{k}\|^2_2
  \]  
  \[
  (23)
  \]

  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 desired amount of ESR. Problem (14) has to be solved to determine coefficients that only one of them is none 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 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 KSVD learns the signal space with a union of low-rank subspaces.

- **K-SVD algorithm**

By extending the union of spheres to a union of low-dimensional subspaces, K-means algorithm is generalized to K-SVD algorithm [13]. 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:

$$D = \min_{D} \sum_{i} \|x_i - D\alpha_i\|_2^2 \text{ st: } \|\alpha_i\|_0 \leq T \quad (24)$$

Each arbitrary selection of few D columns characterizes a cluster corresponding to a subspace. The dictionary learned by K-SVD is in agreement with the proposed problem (19). After learning, test signals that lie on the learned low dimensional subspaces can be reconstructed and detected. Figure 6 show the block diagram of the proposed detection method.

- **Designing parametric dictionary**

In addition to dictionary learning using training signals, it is possible to design a dictionary using parametric functions as the bases [20]. Kernels of FFT and DCT are two examples from this class of dictionaries where bases sweep the parameter of frequency. The major limitation for the number of atoms comes from keeping atoms uncorrelated. Maximum correlation of the atoms (probably comes from neighbor atoms) or coherency of
the dictionary ($\mu$) is a criterion to determine the number of atoms. FFT and DCT are orthonormal dictionaries thus $\mu = 0$.

![Diagram](image)

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

### 4.4 Robustness

Gaussian assumption to model the error may not be an appropriate choice, because Gaussian distribution does not agree with gross errors. Thus, these errors have high effect on the cost function for the estimation of the coefficients. For example assume that a dictionary has been 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 to be supposed for the error
distribution. Thus $\mathcal{H}_1$ implies that the observed signal is the combination of few atoms of D, a Laplace distributed error and a Gaussian distributed noise.

$$\mathcal{H}_1 : y = Dx + e + n$$ (25)

The problem of coefficients estimation for (41) by new prior assumption has been already presented in robust statistics[21]:

$$x = \min_x \|y - Dx\|_H + \rho \|x\|_1$$ (26)

where,

$$\|s\|_H = \sum h(s_i), \quad \rho(s) = \begin{cases} s^2 & |s| < \lambda \\ \lambda |s| & |s| > \lambda \end{cases}$$ (27)

In other words, small errors and large errors are penalized by norm 2 and norm 1, respectively. This is not a surprising result because combination of a Laplace and a Gaussian source is distributed like Gaussian near zero and like Laplace away from zero. $\lambda$ is the parameter of the mixture distribution of Gaussian and Laplace. Let re-write (42) as follows:

$$x = \min_x \|y - Dx - e\|_2 + \rho \|x\|_1 + \lambda \|e\|_1$$ (28)

$$x = \min_x \|y - \begin{bmatrix} D \frac{\rho}{\lambda} \\ 1 \end{bmatrix} \begin{bmatrix} x \\ e \end{bmatrix}\|_2 + \rho(\|x\|_1 + \|e\|_1)$$ (29)

By Substitution of $B = \begin{bmatrix} D \frac{\rho}{\lambda} \\ 1 \end{bmatrix}$ and $z = \begin{bmatrix} x \\ e \end{bmatrix}$, we have:

$$z = \min_z \|y - Bz\|_2 + \rho \|z\|_1$$ (30)
This problem is similar to (26) except that its dictionary is extended by scaled identity matrix. This problem provides coefficients to estimate the matched signal for robust sparse detector. 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. Authors of [22] also intuitively have used the same dictionary to obtain a robust framework for face recognition.

In the next section, experimental results on voice activity detection and linear frequency modulated signal detection will be presented.

4. 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 NOIZEUS database were used. In NOIZEUS database, thirty 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 (MFCC) using 10 Mel triangular filters, energy values computed at each of the 10 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. So our features vector was 24-dimensional, 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 OMP method.

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 0dB, 5dB, 10dB, and 15dB. The ROC Curves for VAD using our proposed method are illustrated in Fig. 7 which shows PD versus PF.
Figure 7. ROC Curves for VAD using our proposed method in different noises
Figure 8 shows the trade-off between ESR and sparsity made by our proposed method in order to have good detector performance. As it can be seen, by increasing the amount of $l^0$ norm (decreasing the sparsity), ESR decreases. According to theoretical trade-off explained in Fig. 8, a proper level of sparsity can be designed.

![Figure 8. ESR versus sparsity for data in VAD](image)

There are also some other methods recently proposed for voice activity detection based on sparse representation [23-26]. In [23], the Bregman iteration is used to exploit the sparse structure of speech signals, and a simple threshold is selected for detecting the speech from noise. In [24], more complex statistical models are assumed for the sparse representation. The basic idea of [25] is to use features extracted from a noise-reduced representation of original audio signals based on non-negative sparse coding and then employing a conditional random field (CRF). Authors of [26] showed that the signal detection can be performed by evaluating the SR of a signal based on the Bayesian framework.
In the all mentioned methods, obtaining the SR of a signal can be viewed as an intermediate level for extracting features for a detection or classification task. In these methods, first some suitable features are extracted for detection based on sparse representation, and then a method (in [23] a simple thresholding method, in [24] a more complex statistical models, in [25] CRF, and in [26] modeling none-zero elements of sparse coefficients) is used for detecting the speech from noise based on those features. On the other hand, our proposed detector assumes that the model agrees with sparsity based signal processing, which first implies the assumption of a suitable hypothesis test and then a general detector is introduced.

The sparsity-based proposed method in [25] works based on the assumption that the features of an audio signal is approximately the sum of speech features and noise features. This limits the speech features that can be used in their VAD method to only those features satisfying this assumption like magnitude spectrum. So some speech features, for example MFCC features which capture the most relevant information of speech signal, cannot be used. Also their proposed method needs to train the noise signals in addition to the clean speech signals. We compared the result of our method with the sparsity-based VAD method proposed in [25]. As can be seen in Fig. 9, our method shows better performance in low SNR conditions.
5. Conclusion

This paper presented a new sparsity-based detector. The performance of the method was evaluated in a realistic application: voice activity detection in speech signal processing. Our detector proposed a new trade-off for designing detectors by assuming union of low-rank subspaces model. The trade-off is between the sparsity and error of union of low-rank
subspaces model denoted by ESR. In our detector the number of filter banks is proportional to the size of dictionary. Proper dictionary is able to regularize 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.

References

[1] Elad M., "Sparse and Redundant Representations", Springer, 2010.
[2] Bruckstein A.M., DonohoD.L., Elad M., "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images", SIAM Review, vol. 51, no. 1, pp. 34-81, 2009.
[3] Rahmoune, A., Vanderghynst, P., Frossard, P., "Sparse Approximation Using M-Term Pursuit and Application in Image and Video Coding", Image Processing, IEEE Transactions on , vol.21, no.4, pp.1950,1962, April 2012.
[4] Elad M., Aharon M., "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries", Image Processing, IEEE Transactions on, vol.15, no.12, pp.3736,3745, Dec. 2006.
[5] Hand book of blind source separation
[6] -Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Supervised dictionary learning. Advancesin Neural Information Processing Systems (NIPS), 21, 2009c.
[7] Lu Y.M. and DoM.N., “A theory for sampling signals from a union of subspaces”, IEEE Trans. on Signal Processing, vol. 56, no. 6, pp. 2334-2345, June 2008.
[8] Liu G., Lin Zh., Yan Sh., Sun J., Yu Y.,Ma Y., "Robust Recovery of Subspace Structures by Low-Rank Representation", Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.35, no.1, pp.171,184, Jan. 2013.
[9] Mallat S. and Zhang Z., “Matching pursuits with time-frequency dictionaries”, IEEE Trans. On Signal Proc., vol. 41, no. 12, pp. 3397-3415, 1993.

[10] Tropp J. A. and Wright S. J., “Computational methods for sparse solution of linear inverse problems”, Proceedings of the IEEE, vol. 98, no. 6, pp. 948-958, 2010.

[11] Kelly E. J., "An adaptive detection algorithm", IEEE Trans. Aerosp. Electron. Systems, vol. 22, no. 1, pp. 115-127, 1986.

[12] Scharf L. L., and Friedlander B., "Matched subspace detectors", IEEE Transactions on Signal Processing, vol. 42, pp. 2146-2157, Aug. 1994.

[13] Aharon M., Elad M., and Bruckstein A., “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation”, IEEE Trans. Signal. Proc, vol. 54, pp. 4311–4322, November 2006.

[14] Rubinstein R., Peleg T. and Elad M., "Analysis K-SVD: A Dictionary-Learning Algorithm for the Analysis Sparse Model", IEEE Trans. on Signal Processing, vol. 61, no. 3, pp. 661-677, March 2013.

[15] Wang, Z., Arce, G., Sadler, B.: Subspace compressive detection for sparse signals. In: IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 3873–3876 (2008).

[16] Mark A. Davenport, Petros T. Boufounos, Michael B. Wakin, and Richard G. Baraniuk. "Signal processing with compressive measurements." to appear in Journal of Selected Topics in Signal Processing, 2010.

[17] Y. C. Eldar and M. Mishali. Robust recovery of signals from a union of subspaces. IEEE Transaction on Information Theory, Volume:55 Issue:11, 2009.

[18] Hoerl A. E. and Kennard R. W., "Ridge Regression: Biased Estimation for Nonorthogonal Problems", American Statistical Association and the American Society for Quality, 1970.

[19] Jain A. K., "Data Clustering: 50 Years Beyond K-Means", Pattern Recognition Letters, 2009.

[20] Yaghoobi M., Daudet L., and Davies M., “Parametric dictionary design for sparse coding”, in Workshop of Signal Processing with Adaptive Sparse Structured Representations (SPARS’09), 2009.

[21] Huber P., "Robust Statistics", Wiley 1981.
[22] Wright J., Yang A.Y., Ganesh A., Sastry S.S., Yi Ma, "Robust Face Recognition via Sparse Representation", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, issue 2, 2009.

[23] You D., Han J., GuibinZheng, TieranZheng, “Sparse Power Spectrum Based Robust Voice Activity Detector”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.

[24] Deng Sh., Han J., “Voice activity detection based on conjugate subspace matching pursuit and likelihood ratio test”, EURASIP Journal on Audio, Speech, and Music Processing, 2011.

[25] Teng P., Jia Y., “Voice Activity Detection Via Noise Reducing Using Non-Negative Sparse Coding”, IEEE Signal Processing Letters, vol. 20, no. 5, May 2013.

[26] Deng Sh.W., Han J.Q., “Statistical voice activity detection based on sparse representation over learned dictionary”, Digital Signal Processing, vol. 23, issue 4, pp. 1228-1232, July 2013.