www.czasopisma.pan.pl www.journals.pan.pl

ARCHIVES OF ACOUSTICS
Vol. **41**, No. 2, pp. 245–254 (2016)

DOI: 10.1515/aoa-2016-0024

A Signal Subspace Speech Enhancement Approach Based on Joint Low-Rank and Sparse Matrix Decomposition

Chengli SUN$^{(1),(2)}$, Jianxiao XIE$^{(1)}$, Yan LENG$^{(3)}$

$^{(1)}$ *School of information, Nanchang Hangkong University*
Nanchang, 330063, China; e-mail: xiejianxiao@126.com

$^{(2)}$ *Science and Technology on Avionics Integration Laboratory*
Shanghai, China

$^{(3)}$ *College of Physics and Electronics, Shandong Normal University*
East Wenhua Road 88, 250014, Ji'nan, China

(*received April 13, 2015; accepted January 18, 2016*)

Subspace-based methods have been effectively used to estimate enhanced speech from noisy speech samples. In the traditional subspace approaches, a critical step is splitting of two invariant subspaces associated with signal and noise via subspace decomposition, which is often performed by singular-value decomposition or eigenvalue decomposition. However, these decomposition algorithms are highly sensitive to the presence of large corruptions, resulting in a large amount of residual noise within enhanced speech in low signal-to-noise ratio (SNR) situations. In this paper, a joint low-rank and sparse matrix decomposition (JLSMD) based subspace method is proposed for speech enhancement. In the proposed method, we firstly structure the corrupted data as a Toeplitz matrix and estimate its effective rank value for the underlying clean speech matrix. Then the subspace decomposition is performed by means of JLSMD, where the decomposed low-rank part corresponds to enhanced speech and the sparse part corresponds to noise signal, respectively. An extensive set of experiments have been carried out for both of white Gaussian noise and real-world noise. Experimental results show that the proposed method performs better than conventional methods in many types of strong noise conditions, in terms of yielding less residual noise and lower speech distortion.

Keywords: subspace speech enhancement; singular value decomposition; joint low-rank and sparse matrix decomposition.

1. Introduction

Speech is a significant form of human communication, but it is always seriously impacted by the inevitable background noise in real world conditions, which lets the noisy speech have low auditory quality and also makes performance of speech processing system degradate or even fail. Therefore, designing a suitable speech enhancement algorithm to reduce the negative effect of noise is of great importance (Vaseghi, 2006).

In the past 50 years, speech enhancement has attracted plenty of attention. According to the processing domain, traditional speech enhancement methods can roughly fall into three categories including time, frequency, and time-frequency domain methods. Time domain methods include the parametric model based method (Gannot *et al.*, 1998; Vaseghi, 2006; Kim *et al.*, 2000) and subspace method (Dendrinos *et al.*, 1991; Hu, Loizou, 2003). Frequency domain methods include Wiener filtering method (Ephraim, Malah, 1984), Markov model (Jax, Vary, 2003), and spectral subtraction (Boll, 1979). Time-frequency domain methods include the wavelet transform method (Mallat, 1999), auditory masking method (Virag, 1999), and constrained low-rank and sparse matrix decomposition (CLSMD) method (Sun *et al.*, 2014).

Due to its capability of preserving speech intelligibility, subspace based speech enhancement algorithms have recently drawn a great attention. The principle of the subspace methods is to separate the noisy speech observation space into a signal subspace and a noise subspace, and the enhanced speech was constructed using only the components of the signal within the signal subspace. In the subspace-based algorithms, subspace decomposition is a critical step for subspace separation, which is often performed via Karhunen-Loeve transform (KLT) (Ephraim, Van Trees, 1995) or singular value decomposition (SVD) (Moor, 1993). Thus, a key issue in developing a subspace-based model is the way of splitting and refining the signal and noise subspace in an optimal way. Abolhassani *et al.* (2007) introduced the variance of the reconstruction error (VRE) criterion to optimise the subspace selection for speech enhancement. Saadoune *et al.* (2014) seeked to optimise the subspace decomposition model by incorporating the psychoacoustic properties of the human auditory system into the subspace filter to reconstruct the enhanced signal. Jin (Jin, Scordilis, 2006) proposed a modified subspace method by using a linear prediction (LP) residual technique to construct the LP coefficient matrix for speech enhancement.

At the same time, many efforts were made to improve the performance of subspace-based algorithms, and the excellent noise reduction capabilities of subspace filtering techniques are confirmed by several studies. The existing subspace-based speech enhancement methods still suffer from the problem of low decomposition accuracy in the presence of large noise, resulting in a high remainder noise within enhanced speech in strong noise cases. In this paper, we propose a new subspace-based method for speech enhancement based on the principle of joint low-rank and sparse matrix decomposition (JLSMD). The proposed method differs from the previous subspace based methods in its decomposition pattern. The main idea behind our method is motivated by the recent development of low-rank and sparse models (LSMs) theory. According to the LSMs theory, if a given corrupted data matrix X has an underlying low-rank structure, yet corrupted by sparse additive noises, we denote these two ingredients as L and S. The underlying low-rank component L can be effectively recovered by solving a convex optimisation problem, even if the noise is arbitrary in magnitude. In the time domain, owing to the short-time stability of voice, speech signals within different time-frames can be assumed to have a low-rank structure. On the other hand, due to the randomness of noise, background noise is more variable than speech and thus can be regarded as being high-rank and sparse. Thus JLSMD theory can be exploited to recover the underlying enhanced speech from noisy signal.

LSMs technique has been initially used in computer vision applications, such as moving object detection (Xu *et al.*, 2012), traffic anomalies detection (Mardani, Mateos, 2013), image restoration, and alignment (Peng *et al.*, 2012), etc. It can be also applied in music information retrieval system for separation of singing voice from the musical accompaniment. More recently, LSMs theory was introduced into the speech enhancement task in our previous work (Sun *et al.*, 2014), where a constrained low-rank and sparse matrix decomposition (CLSMD) algorithm is designed for noise reduction. In the time-frequency domain, since (white) noise amplitude spectra within different time frames are usually highly correlated with each other, noise signal can be assumed as a low-rank component. On the other hand, speech signal can be regarded as relatively sparse in audio recordings. Therefore, by means of CLSMD, the noisy speech spectrogram can be decomposed into a low-rank part corresponding to noise and a sparse part corresponding to speech.

It should be pointed out that the JLSMD based subspace method is different from the CLSMD based approach (Sun *et al.*, 2014). Both of them are closely related in the sense of performing noise reduction using LSMs technique, but they vary in the assumptions and working pattern. The former assumes that speech is low-rank and noise is sparse, while the latter is just the opposite, i.e., the noise is low-rank and speech is sparse. Besides, CLSMD is a batch learning algorithm working in the time-frequency domain, while JLSMD is performed in the time-domain and is able to work frame-by-frame. Hence, the JLSMD is more suitable for real-time speech denoising task.

This paper is organised as follows. Section 2 introduces related work for the proposed method. Section 3 introduces JLSMD based subspace decomposition algorithm. In the Sec. 4, we describe the JLSMD based signal subspace speech enhancement system. Some implementation details and experimental results are described in the Sec. 5. Finally, the conclusions are given in the Sec. 6.

2. Related work

2.1. SVD-based subspace speech enhancement method

Let us consider the problem of enhancement of a speech signal contaminated by an independent additive noise. Suppose that a noisy signal vector $y(t) \in R^N$ is the sum of a clean signal vector $x(t) \in R^N$ and a noise signal vector $d(t) \in R^N$,

$$y(t) = x(t) + d(t), \tag{1}$$

where N represents the frame length. Arranging the N-dimensional vectors into a $(N - l + 1) \times l$ matrix with Toeplitz structure, we can get

$$Y = X + D, \tag{2}$$

where l represents a positive integer and $l = N/3$. Specifically, the form of observation matrix Y is written as

$$Y = \begin{bmatrix} y(l-1) & y(l-2) & \cdots & y(0) \\ y(l) & y(l-1) & \cdots & y(1) \\ \vdots & \vdots & \vdots & \vdots \\ y(N-1) & y(N-2) & \cdots & y(N-l) \end{bmatrix}. \tag{3}$$

Assuming that the rank of matrix Y is p, the optimal enhanced speech matrix $\widehat{X}$ can be estimated according to the following least-square criterion.

$$\widehat{X} = \min_{\widehat{X}} \left\| \widehat{X} - X \right\|_F^2 \qquad \text{rank}(\widehat{X}) \le p, \tag{4}$$

where $\|\bullet\|_F$ is the Frobenius norm.

If $d(t)$ is a white Gaussian noise, it satisfies the conditions $D^T D = \sigma_d^2 I$ and $X^H D = 0$, where σ_d^2 represents the variance of noise. The optimal solution of (4) can be obtained by SVD of Y

$$Y = U \sum V^H, \tag{5}$$

$$\widehat{X}_p = \sum_{k=1}^{p} \sigma_k u_k v_k^T, \tag{6}$$

where U and V are the left and right singular vector matrix of Y; $\sum$ is the diagonal eigenvalue matrix composed of the singular values σ_k; u_k, and v_k are the column vector of U and V, respectively.

The above low-rank matrix $\widehat{X}_p$ represents the original speech matrix X in the sense of least-square minimisation. This may get the optimal estimate when the noise is small, independent, and identically distributed Gaussian.

If $d(t)$ is a colored noise signal, the conditions $D^T D = \sigma_d^2 I$ and $X^H D = 0$ are no longer valid (Golub, Van Loan, 1989). In this case, we can seek a prewhitening matrix W to make sure the equation $E\left[(DW)^T DW\right] = I$ is fulfilled. Indeed, the noise signal can be prewhitened by a multiplication by

$$YW = XW + DW. \tag{7}$$

If noise matrix D is available, we can factor $D = QR$ via QR decomposition, where Q is a standard orthogonal matrix ($Q^T Q = I$) and R is the Cholesky factor of $D^T D$. From $Q = DR^{-1}$, we can obtain $W = R^{-1}$ satisfying the condition $(DR^{-1})^T(DR^{-1}) = Q^T Q = I$.

To sum up, if the additive noise is a colored noise, the transformed matrix $\overline{Y} = YR^{-1}$ should be used instead of Y. Then the traditional subspace decomposition is used to decompose the transformed matrix $\overline{Y}$. After the SVD modification by (6), a corresponding dewhitening operation (a postmultiplication by the matrix R) of $\widehat{X}_p$ should be included by

$$Z_p = \widehat{X}_p R. \tag{8}$$

Note that Z_p obtained by the above step is not a Toeplitz matrix. To let Z_p be Toeplitz-structured, we should reformat it by arithmetic averaging along the diagonals of $\widehat{X}_p$ (Tufts, Kumaresan, 1982; Tufts *et al.*, 1982).

2.2. Low-rank and sparse models

Principal component analysis (PCA) (Jolliffe, 2002) has been attracting much attention due to its wide applications to pattern recognition and computer vision. It seeks to accurately estimate the low-dimensional subspace with the given high-dimensional data via SVD. As it is well known, however, the PCA method is sensitive to non-Gaussian noises and outliers, which is often the case in real problems due to the mechanism of data acquisition.

To address this robustness issue, Wright *et al.* proposed the robust PCA technique (Wright *et al.*, 2009). The goal of robust PCA is to recover a low rank matrix $L \in R^{m\times n}$ from the corrupted observed data matrix $Y \in R^{m\times n}$

$$Y = L + S, \tag{9}$$

where the matrix $S \in R^{m\times n}$ is assumed to be sparsely supported and random in amplitude. This can be achieved by solving the following optimisation problem:

$$\min\nolimits_{L,S} \text{rank}(L) + \gamma \left\| S \right\|_0 \quad \text{subject to} \quad Y = L + S, \tag{10}$$

where l_0-norm $\|S\|_0$ counts the number of nonzero elements in the matrix S and γ is a balance parameter. Formula (10) is a highly nonconvex optimisation problem, and we can not solve it directly. Fortunately, this problem can be converted into the following convex optimisation:

$$\min\nolimits_{L,S} \|L\|_* + \gamma \left\| S \right\|_1 \quad \text{subject to} \quad Y = L + S, \tag{11}$$

where $\|L\|_*$ represents nuclear norm (Candes, Plan, 2010), which is the sum of all singular values $\|L\|_* = \sum\limits_i \sigma_i(L)$, and $\|S\|_1$ is l_1-norm which is defined as the sum of absolute values of the matrix entries. This problem is known to have a stable solution if L and S are sufficiently incoherent (Candes *et al.*, 2011), i.e., L is exactly low-rank and S is exactly sparse.

If the decomposition with predefined constrains $\text{rank}(L) \le p$ and $\|S\|_0 \le r$ is allowed, the low-rank matrix recovery problem can be solved by minimising the following decomposition error (Toh, Yun, 2010):

$$\begin{gathered} \min\nolimits_{L,S} \left\| Y - L - S \right\|_F^2, \\ \text{subject to} \quad \text{rank}(L) \le p \\ \text{and} \quad \|S\|_0 \le r. \end{gathered} \tag{12}$$

In practice, p and r are preferred to be restricted in order to control the model complexity of L and S.

3. Joint low-rank and sparse matrix decomposition algorithm

Splitting of signal and noise subspaces via decomposition algorithms, which are often done by SVD or eigenvalue decomposition, is a critical step in subspace approaches. As mentioned above, these decomposition algorithms are brittle to the presence of large corruptions and suffer from a low decomposition accuracy. In this work, we propose a new subspace decomposition algorithm based on the JLSMD, which is robust in strong noise conditions and less sensitive to the large interferences.

Problem 1 (P1). Suppose we are given a noisy data matrix Y, and know that it may be decomposed as $Y = L + S + G$, where L is a low-rank matrix corresponding to clean speech, S is a sparse matrix corresponding to the noise signal, and G is a decomposition error matrix. Let p and r be the rank constraint of L and sparse constraint of S, respectively. The P1 is to recover the underlying speech matrix L from corrupted audio data.

Through minimising the decomposition error G, the P1 can be solved by the following optimisation:

$$\begin{aligned} &\min\nolimits_{L,S} \|Y - L - S\|_F^2 , \\ &\text{subject to} \ \ \operatorname{rank}(L) \le p \\ &\text{and} \ \ \|S\|_0 \le r. \end{aligned} \tag{13}$$

The above formula can be solved by alternatively solving the following two formulas until convergence:

$$\begin{aligned} L_t &= \arg\min\nolimits_{\operatorname{rank}(L)\le p} \|Y - L - S_{t-1}\|_F^2 , \\ S_t &= \arg\min\nolimits_{\|S\|_0 \le r} \|Y - L_t - S\|_F^2 . \end{aligned} \tag{14}$$

Although the above formula is a nonconvex optimisation, its global solutions exist (Wright *et al.*, 2009).

In (Wright *et al.*, 2009) they use a singular value hard thresholding of $Y - S_{t-1}$ to update L_t and entry-wise hard thresholding of $Y - L_t$ to update S_t, respectively.

$$\begin{aligned} L_t &= \sum_{i=1}^{p} \lambda_i U_i V_i^T , \\ SVD(Y - S_{t-1}) &= U \Lambda V^T , \\ S_t &= P_C (Y - L_t), \end{aligned} \tag{15}$$

where C is a nonzero subset of the first k largest entries of $|Y - L_t|$, and $P_C(x)$ is equivalent to projecting x onto the entry set C (Candes, Terence, 2010), which is defined as ($P_C(X) = X$, $X \in C$). Due to the sparse matrix S being from partial observation, we get more a accurate S by the thresholding functions.

Since in the proposed method the distribution of the outlier should be sparse and random enough (Zhou *et al.*, 2013), we use the entry-wise hard thresholding function to estimate S (Chang *et al.*, 2000), which is $\varphi_u(x) = x \cdot 1(|x| > u)$. This gets the input, if it is larger than the threshold; if not, it is set to zero. The formula is as

$$S_t = (X - L_t) \odot [(X - L_t) > u], \quad u > 0, \tag{16}$$

where the operation $\odot$ is an element-wise multiplication.

There is a problem about the computation used with the SVD for updating the matrix L_t, because the SVD requires much computation time. Therefore, in the paper a method of fast low-rank approximation (Zhou, Tao, 2011) is proposed, which is bilateral random projections (BRP). Given a matrix $X \in R^{m \times n}$ the formula

$$L = Y_1 \left(A_2^T Y_1\right)^{-1} Y_2^T \tag{17}$$

is rank-p approximation of X, where $Y_1 = X A_1$, $Y_2 = X^T A_2$, $A_1 \in R^{n \times p}$, and $A_2 \in R^{m \times p}$ are Gaussian random matrices. In (Fazel *et al.*, 2008) matrix X can be recovered from L, and it can reduce the time cost.

Thus the speech and noise components can be decomposed into the low-rank and sparse subspace respectively by the JLSMD based subspace decomposition.

We have the following optimisation algorithm for JLSMD.

Algorithm 1. JLSMD based subspace decomposition algorithm.

```
Input: $p, \varepsilon, max, u$;
Output: $L = L_t, S = S_t$;
Initialise: $L_0 = X, S_0 = 0, t = 0$;
While not converged do
    $t = t + 1$;
    $A_1 = randn(n, p)$;
    $A_2 = randn(m, p)$;
    $Y_1 = L_{t-1} A_1$;
    $Y_2 = L_{t-1}^T A_2$;
    $L_t = Y_1 (A_2^T Y_1)^{-1} Y_2^T$;
    %Update the low rank matrix
    $R_t = L_{t-1} - L_t + S_{t-1}$;
    $S_t = R_t \odot (R_t > u)$;
    %Update the sparse matrix
    If $\frac{\|L_0 - L_t - S_t\|_F^2}{\|X\|_F^2} \le \varepsilon$ or $t == max$
      converged=1;
      break;
    end
    $L_t = L_t + R_t - S_t$;
end while
```

3.1. Convergence of JLSMD

In this section we will show the convergence properties of JLSMD, of which the objective value $G = \|X - L - S\|_F^2$ (error G) converges to a local minimum. From (Zhou, Tao, 2011) we can get that the value $G_t = \|X - L_t - S_t\|_F^2$ keeps decreasing and converges to a local minimum. Meanwhile we can get the conclusion that the solutions L and S respectively converge to local optimums with the linear rate less than 1, and the converge speeds will be slowed by augmenting

$$\begin{aligned} &\text{for } L: \frac{\|\Delta_L\|_F}{\|L+\Delta_L\|_F}, \quad \Delta_L = (S+G) - P_C(S+G), \\ &\text{for } S: \frac{\|\Delta_S\|_F}{\|L+\Delta_S\|_F}, \quad \Delta_S = (L+G) - P_M(L+G), \end{aligned} \tag{18}$$

$$M = \{H \in R^{m\times n} : rank(H) = p\} \tag{19}$$

but it will ruin the convergence unless $\|G\|_F \gg \|S\|_F$ or $\|G\|_F \gg \|L\|_F$. The derivations of the formulas can be found in (Zhou, Tao, 2011).

Therefore, JLSMD is able to get the approximated sparse matrix and low-rank matrix when G is not overwhelming.

3.2. The main parameters in JLSMD

There are two important parameters in JLSMD method, one is the affecting rank parameter e and the other one is the sparse constraint parameter u. When we use the analysis-by-synthesis approach to determine the effective rank p, the value e could affect the veracity of p, while the parameter u controls the sparsity of S. The bigger value u is, the less noise the noisy speech signal has. Then we should choose the more precise parameter u to estimate the sparse matrix S. For a better speech enhacement performance, the parameters (e, u) should be selected.

4. Speech enhancement based on JLSMD

Regarding the basic theories of subspace signal enhancement, when a speech signal is infected with an additive noise, its singular values are changed (Zehtabian *et al.*, 2010). Thus we derive an approach of JLSMD to remove the noise (sparse component) from the singular values, and more precisely recover the clean speech. In this section, we give the implementation details of the proposed method for white noise and colored noise, respectively.

4.1. In the white noise case

Figure 1 shows the block diagram of the enhanced speech based on JLSMD for white noise. At first, the noisy speech signal is divided into frames in the time domain. Then we transform each frame of the noisy speech into Toeplitz matrix. Next we determine the effective rank p with the analysis-by-synthesis approach (Bakamides *et al.*, 1991). The noisy speech matrix Y is decomposed into the low-rank matrix L with the rank p and the sparse matrix S by the JLSMD optimisation algorithm, where the matrix L is the enhanced speech matrix. Thus we get L and remove the sparse matrix S which is noise signal component. But L is not the Toeplitz matrix, we average all the diagonal elements of L to let it became a Toeplitz matrix form. Finally, the enhanced speech is constructed by taking the inverse transform of constructing Toeplitz matrix and least-squares overlap-add synthesis (Quatieri, 2002).

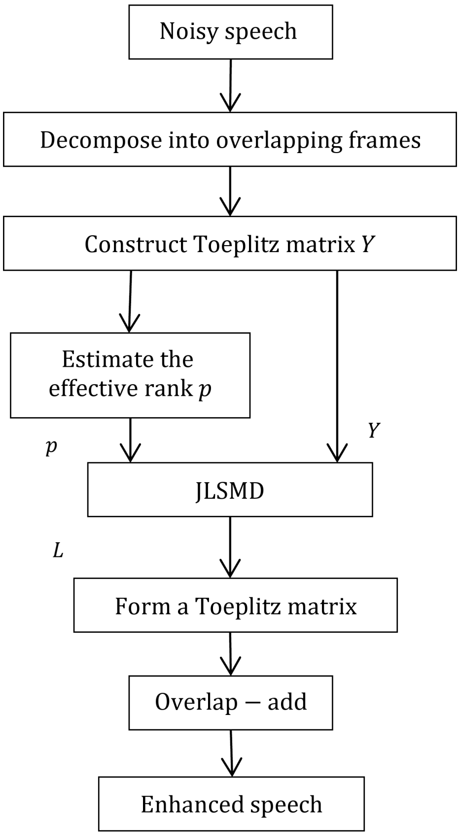
Noisy speech
Decompose into overlapping frames
Construct Toeplitz matrix Y
Estimate the effective rank p
p
Y
JLSMD
L
Form a Toeplitz matrix
Overlap – add
Enhanced speech

Fig. 1. Block diagram of the enhanced speech based on JLSMD for white noise.

4.2. In the colored noise case

Figure 2 shows the block diagram of the enhanced speech based on JLSMD for the colored noise. As compared with the white noise case, before the step of estimating the effective rank, we add the step of performing a prewhitening of noisy speech signal Y. After the procedure of JLSMD, we will perform a dewhitening of L. The next procedure is the same as in the white noise case.

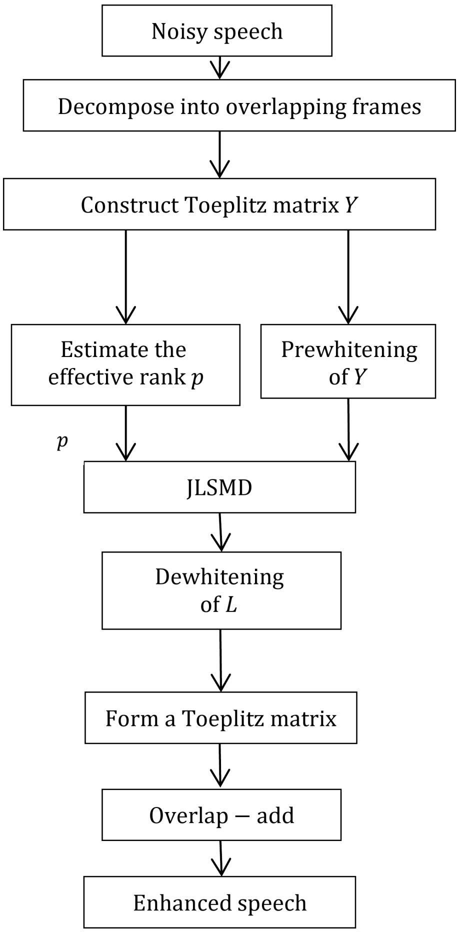

Fig. 2. Block diagram of the enhanced speech based on JLSMD for the colored noise.

4.3. Estimate the effective rank value

The input of JLSMD needs to know the rank value p for the underlying speech matrix X (Chambers, 1977). In this paper, we estimate this parameter using a synthesis approach (Bakamides *et al.*, 1991) in which consecutive reconstructions are performed and the resulting error power is compared to the noise variance in order to get the best approximation of the original clean speech signal. The number of the singular values corresponding to the reconstruction error power as close as possible to the noise variance gives the best estimate of p.

Assuming that noise signal $d(n)$ and speech signal $x(n)$ are uncorrelated and taking the ensemble averages of the squares of (1), we get

$$E\left[y^2(n)\right] = E\left[x^2(n)\right] + E\left[d^2(n)\right]. \qquad (20)$$

If $d(n)$ and $x(n)$ are zero-mean signals, the above formula can be simplified as

$$\sigma_y^2 = \sigma_x^2 + \sigma_d^2, \qquad (21)$$

where $d(n)$ represents the noise signal, σ_d^2 is the noise variance. The signal variance σ_x^2 is the power of the clean signal. Assuming that $x_k(n)$ represents the synthesised signal using the first k singular values, the following quantity is defined as

$$E_k = E\left[y^2(n)\right] - E\left[x_k^2(n)\right], \qquad (22)$$

where E_k represents the power difference between the noisy signal $y(n)$ and the synthesised signal $x_k(n)$. If the synthesised signal $x_k(n)$ is close to the clean signal $x(n)$, that is $x_k(n) \approx x(n)$, $E_k \approx E\left[d^2(n)\right]$, which is the noise variance, $x_k(n)$ is closer to the signal $x(n)$ and formula $\left(E_k - E\left[d^2(n)\right]\right)$ is closer to zero. In this case the formula $\left(E_k - E\left[d^2(n)\right]\right)$ can be used to determine the effective rank of matrix X. Thus $\left(E_k - E\left[d^2(n)\right]\right)$ can be written as

$$E_k - \sigma_d^2 = E\left[x^2(n)\right] - E\left[x_k^2(n)\right]. \qquad (23)$$

When k is less than the actual rank $p(k < p)$, the energy of the synthesised signal is less than that of the clean signal $\left(E\left[x_k^2(n)\right] < E\left[x^2(n)\right]\right)$, because $x_k(n)$ does not contain all the clean signal. When k is more than the rank $p(k > p)$, the energy of the synthesised signal is more than the clean signal $\left(E\left[x_k^2(n)\right] > E\left[x^2(n)\right]\right)$. Because of the addition of noise, the value $\left(E^k - \sigma_d^2\right)$ is negative. This implies that when $k = p$, the value of $\left(E^k - \sigma_d^2\right)$ must be zero. From what has been discussed above, the effective rank of matrix X can be determined through observing whether the value of $\left(E^k - \sigma_d^2\right)$ is zero. And it is defined as:

$$e(k) = \left|E_k - \sigma_d^2\right|, \qquad k = 1, 2, \ldots, l, \qquad (24)$$

where $e(k)$ represents the noise error estimation. At this time the minimum of $e(k)$ is used for detection of the effective rank of X. Therefore the precise value of the rank $p = k$ can be obtained by determining the fixed parameter e.

5. Experiments

For the performance evaluation of the proposed JLSMD method, we choose a total of 30 sentences (sp01∼sp30) taken from NOIZEUS database (Hu, Loizou, 2008) whose sampling frequency is 8 kHz. All measurement scores are averaged over the 30 test sentences. The noisy speeches used for evaluating are taken from five types of noise sources, including "white", "pink", "F16", "Hfchannel", and "Exhibition" noises at 0 dB, 5 dB, 10 dB, and 15 dB SNR, respectively. The frame length is 264 sample points (33 ms) with 50% frame overlap. The column and row sizes of the Toeplitz sample matrix X adopted in JLSMD are 88 and 176, respectively. We repeat the experiments ten times, then calculate PESQ-MOS and segmental-SNR (segSNR) scores on the average.

5.1. Influence of the error level e on the performance of speech enhancement

In this section, we will examine the influence of e on the performance of speech enhancement. The speech signal corrupted by Gaussian white noise and pink noise are used for the experiments at different SNR levels. In the proposed method we fix the threshold $u = 0.09$, increase e from -3 to 3 at the interval of 1, and set $e = -\sigma_d^2/12$.

Table 1 shows PESQ-MOS scores for the white noise (0 dB to 15 dB) in the case of e from -3 to 3 and $e = -\sigma_d^2/12$, where the PESQ-MOS and segSNR scores "0" represent that the parameter e causes the method to produce an error "index exceeds matrix dimensions", from which we infer that parameter e is highly related to the SNR (or noise energy). From the Table 1 we can know that the parameter $e = -3$ can obtain the best performance at 0 dB, the parameter $e = -2$ can obtain the best performance at 5 dB, and the parameter $e = 0$ gets the best performance at 10 dB and 15 dB, respectively. Meanwhile, we can see that the best e value obtained will change as the noise energy varies. Thus we set $e = -\sigma_d^2/12$ for all the cases where σ_d^2 is the noise energy.

Table 1. PESQ-MOS scores for the white noise (0 dB to 15 dB) in the case of e from -3 to 3 and $e = -\sigma_d^2/12$.

e	0 dB	5 dB	10 dB	15 dB
-3	2.0600	2.4185	0	0
-2	2.0313	2.4577	0	0
-1	2.0122	2.4077	0	0
0	1.9698	2.3369	2.8051	3.1644
1	1.9011	2.2497	2.7265	2.8899
2	1.8277	2.1281	2.4766	2.6008
3	1.7807	1.9945	2.2047	2.4506
$-\sigma_d^2/12$	2.0521	2.4657	2.8561	3.1827

Table 2 shows segSNR scores for the white noise (0 dB to 15 dB) in the case of e from 3 to -3 and $e = -\sigma_d^2/12$, from which we also find that when we set $e = -\sigma_d^2/12$, the highest segSNR is obtained. Therefore, we set the parameter e as $-\sigma_d^2/12$ in JLSMD method in the white noise case.

Table 2. segSNR scores for the white noise (0 dB to 15 dB) in the case of e from -3 to 3 and $e = -\sigma_d^2/12$.

e	0 dB	5 dB	10 dB	15 dB
-3	1.4430	2.3160	0	0
-2	1.3931	2.7264	0	0
-1	1.3694	3.9254	0	0
0	1.2424	3.9048	6.5603	8.9028
1	1.0427	3.6216	5.7790	7.0296
2	0.8548	3.2723	4.8747	5.8892
3	0.8457	2.9027	4.4607	5.1568
$-\sigma_d^2/12$	1.4692	3.9634	6.5863	9.0824

Tables 3 and 4 show the PESQ-MOS and segSNR scores for the colored noise in the case of e from -3 to 2 and $e = -\sigma_d^2/12$, respectively. Since the segSNR scores have less variation as the parameter e increases, we can deem that $e = -\sigma_d^2/12$ is appropriate in the JLSMD method in the colored noise.

Table 3. PESQ-MOS scores for the colored noise (0 dB to 15 dB) in the case of e from -3 to 2 and $e = -\sigma_d^2/12$.

e	0 dB	5 dB	10 dB	15 dB
-3	2.0695	2.4115	2.7461	3.0883
-2	2.0412	2.4122	2.7568	3.0910
-1	2.0210	2.4090	2.7627	3.1130
0	2.0060	2.3833	2.7551	3.0810
1	1.9827	2.3658	2.7313	3.0720
2	1.9590	2.3301	2.7136	3.0540
$-\sigma_d^2/12$	2.0682	2.4196	2.7685	3.1357

Table 4. segSNR scores for the colored noise (0 dB to 15 dB) in the case of e from -3 to 2 and $e = -\sigma_d^2/12$.

e	0 dB	5 dB	10 dB	15 dB
-3	-1.2221	1.2772	4.0219	6.5562
-2	-1.0776	1.4007	4.0473	6.6662
-1	-1.0598	1.5341	4,1078	6.7460
0	-0.9762	1.6540	4.2134	6.8132
1	-0.8656	1.7341	4.3215	6.9326
2	-0.8843	1.7556	4.3623	6.9753
$-\sigma_d^2/12$	-1.0663	1.3942	4.0621	6.7524

5.2. Influence of the sparse constraint u on performance

In this section, we will examine the influence from the sparse constraint u on the performance of speech enhancement in the case of Gaussian white noise and colored noise (pink noise). We experimentally set that the parameter $u = 0.09$ in the white noise case and $u = 0.4$ in the colored noise case.

5.3. Performance in the Gaussian white noise

In this section we will evaluate the performance of JLSMD method in the Gaussian white noise situation. Five well-known enhancement approaches are compared with the proposed method, including KLT – a generalised subspace approach of KLT (Hu, Loizou, 2003), SSboll – spectral subtraction (Boll, 1979), SSVD – original subspace approach of SVD (Dendrinos *et al.*, 1991), MMSE – minimum mean-square error short-time spectral amplitude estimator (Ephraim, Malah, 1984), Wiener-Wiener filter based on tracking *a priori* SNR using Decision-Directed method (Plapous *et al.*, 2006), and CLSMD – a constrained low-rank and sparse matrix decomposition algorithm (Sun *et al.*, 2014). For the JLSMD method we set $= -\sigma_d^2/12$, $u = 0.09$ and max $= 50$.

Tables 5 and 6 show the comparison of performance in terms of PESQ-MOS and segSNR scores. The larger the PESQ-MOS and segSNR scores are, the better the performance is. From Tables 5 and 6 we can see that the proposed method JLSMD has got the highest PESQ-MOS and segSNR scores among all the contrastive methods, except at 0 dB where CLSMD has the highest segSNR scores. This manifests that the JLSMD method has good speech enhancement performance in the white noise condition.

Table 5. segSNR scores for comparison of different methods in the case of the white noise.

Method	0 dB	5 dB	10 dB	15 dB
KLT	1.1005	3.5290	5.9374	8.0588
MMSS	−0.4641	0.7874	2.1194	3.3749
SSboll	−3.5192	−2.2135	−1.0612	0.0514
Wiener	−2.1775	−1.3365	−0.1136	1.0583
SSVD	0.4106	2.9652	5.4880	7.9579
CLSMD	**2.1466**	3.6858	4.8563	5.6401
JLSMD	1.4688	**3.9700**	**6.5965**	**9.0741**

Table 6. PESQ-MOS scores for comparison of different methods in the case of the white noise.

Method	0 dB	5 dB	10 dB	15 dB
KLT	1.9945	2.3975	2.7441	3.0788
MMSS	1.4551	1.7107	2.1402	2.4981
SSboll	1.6510	1.9436	2.1790	2.4626
Wiener	1.6048	1.9707	2.3049	2.5118
SSVD	1.7132	2.1944	2.5968	2.9725
CLSMD	2.0095	2.3845	2.5877	2.7207
JLSMD	**2.0556**	**2.4785**	**2.8446**	**3.1980**

As compared with SSVD, the proposed method has higher PESQ-MOS and segSNR scores, which shows that the proposed method has markedly improved in the low and high SNR situations. The results imply that the proposed method has significant advantages over this traditional method.

5.4. Performance in the colored noise and some real-world noises

In this section we will evaluate the performance of the aforementioned speech enhancement methods in the colored noise, including pink noise and some real-world noises. For the JLSMD method we set $= -\sigma_d^2/12$, $u = 0.4$ and max $= 50$.

Tables 7 and 8 summarise the performance comparison of PESQ-MOS and segSNR scores. In terms of segSNR, it is observed that JLSMD achieved the highest segSNR scores at 10 dB and 15 dB in the F16 noise condition. Thus we can deem that the proposed method is slightly better in noise reduction case, which needs to be improved in the near future.

Table 7. segSNR scores for comparison of different methods in different noise types.

noise	method	0 dB	5 dB	10 dB	15 dB
Pink	KLT	0.7753	**3.0972**	**5.3781**	**7.4626**
	MMSS	−0.6958	0.6341	1.9652	3.1325
	SSboll	−1.3319	−0.2226	1.1602	2.2906
	Wiener	**0.9521**	1.8636	2.3352	2.9784
	SSVD	−3.2675	−0.6419	2.1777	5.0420
	CLSMD	0.0866	1.7255	3.0208	3.7591
	JLSMD	−1.0672	1.4034	4.0446	6.7695
F16	KLT	0.6788	**2.9767**	5.2685	7.4231
	MMSS	−0.7410	0.6536	1.9044	3.0734
	SSboll	−1.1499	−0.0636	1.1943	2.3623
	Wiener	**1.1499**	2.0009	2.6108	3.0081
	SSVD	−0.2814	2.1453	4.9328	7.7252
	CLSMD	−0.3976	1.5103	3.0328	3.6692
	JLSMD	0.1948	2.6306	**5.2773**	**7.8629**
Hfchannel	KLT	0.7043	**3.2218**	**5.6871**	**7.8663**
	MMSS	−2.1817	−0.7791	0.7254	1.9933
	SSboll	−3.0755	−1.4375	0.1127	1.5833
	Wiener	**1.0176**	1.9490	2.6083	3.0045
	SSVD	−1.5944	1.0135	3.6927	6.3480
	CLSMD	0.7115	2.6017	4.2193	5.2412
	JLSMD	−0.4876	2.0446	4.7303	7.3973
Exhibition	KLT	−0.7101	**1.6229**	**4.2136**	**6.7344**
	MMSS	−2.3439	−0.8301	0.8341	2.2832
	SSboll	−2.2941	−0.7509	0.6080	1.8858
	Wiener	**−0.4744**	0.4790	1.5603	2.3289
	SSVD	−1.8785	0.9330	3.4139	6.1403
	CLSMD	−1.7303	0.3608	2.2839	3.7397
	JLSMD	−1.6766	1.0386	3.6229	6.4200

In terms of PESQ-MOS it is observed that JLSMD method obtains the highest PESQ-MOS scores in most of noise cases except at 0 dB in the pink noise condition, where CLSMD is better than JLSMD, and at 15 dB in the pink noise conditions, where KLT is slightly better than JLSMD. It is observed that the proposed method is adept in improving the overall quality of the enhanced speech, which can make the enhanced speech sound more confortable.

Table 8. PSEQ-MOS scores for comparision of different methods in different noise types.

noise	method	0 dB	5 dB	10 dB	15 dB
Pink	KLT	2.0687	2.4121	2.7548	**3.1258**
	MMSS	1.9211	1.3030	2.5694	2.7894
	SSboll	1.8367	2.1344	2.5027	2.7715
	Wiener	1.8312	2.2279	2.5724	2.8487
	SSVD	1.6869	2.0543	2.5733	2.6990
	CLSMD	**2.0916**	2.4143	2.6350	2.7636
	JLSMD	2.0712	**2.4283**	**2.7703**	3.1212
F16	KLT	2.0528	2.4223	2.7550	3.1269
	MMSS	2.0332	2.3485	2.6231	2.8269
	SSboll	1.8933	2.2068	2.5602	2.8275
	Wiener	1.8210	2.2625	2.6199	2.8881
	SSVD	1.6870	2.1395	2.5574	2.9407
	CLSMD	1.9856	2.3135	2.5978	2.7225
	JLSMD	**2.1690**	**2.5255**	**2.8521**	**3.1969**
Hfchannel	KLT	1.9298	2.2846	2.6363	2.9316
	MMSS	1.7677	2.1285	2.4918	2.7633
	SSboll	1.6881	2.0043	2.3303	2.6764
	Wiener	1.6944	2.1144	2.4890	2.7785
	SSVD	1.6643	2.0635	2.4255	2.7764
	CLSMD	1.9331	2.2958	2.5628	2.7044
	JLSMD	**1.9580**	**2.3436**	**2.7071**	**3.0563**
Exhibition	KLT	1.5453	2.0555	2.3917	2.7689
	MMSS	1.5597	2.0096	2.3538	2.6546
	SSboll	1.5400	1.9701	2.3470	2.6390
	Wiener	1.1555	1.6836	2.1119	2.4862
	SSVD	1.1597	1.7662	2.1561	2.5669
	CLSMD	1.4421	1.9325	2.2944	2.5610
	JLSMD	**1.5654**	**2.0721**	**2.4107**	**2.7753**

As compared with the CLSMD method, JLSMD obtains the higher PESQ-MOS scores, except at 0 dB in the pink noise conditions, where CLSMD is better than JLSMD. But CLSMD gets higher segSNR scores in 0 dB and 5 dB conditions. These results indicate that CLSMD has more noise reduction capability in strong noise environments, while JLSMD is better in improving the overall quality of the enhanced speech.

As compared with the SSVD method, the proposed method also gets higher scores in terms of segSNR and PESQ-MOS. The result shows that the proposed method has improved the performance of the original subspace method SSVD in the coloured noise.

6. Conclusions

In this paper, we presented a signal subspace speech enhancement based on JLSMD. The new subspace decomposition algorithm based on JLSMD is less sensitive to the large interferences as compared with traditional algorithms, and can significantly reduce noise. Experiments demonstrate that the proposed method is good at improving the overall enhanced speech quality, especially in low SNRs and white noise.

Simultaneously it should be pointed out that JLSMD method has improved the original subspace method based on SVD and can wipe out more residual noise. In the future research work we will devote more efforts to improving the noise reduction performance in the colored noise.

Acknowledgments

This article was jointly supported by the National Natural Science Foundation of China (No. 61362031, 61472138, 61462064, 61401259, 61472138), China Postdoctoral Science Foundation (No. 2015M582128), Jiangxi Province Natural Science Foundation (No. 20142BAB207002), China's Aviation Science (No. 20145556011), the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information Ministry of Education (No. 0920140122006), and the Innovation Fund Designated for Graduate Students of Jiangxi Province (No. YC2014-S392).

References

1. Abolhassani A.H., Selouani S.-A., O'Shaughnessy D. (2007), *Speech enhance-ment using PCA and variance of the reconstruction error model identification*, Automatic Speech Recognition & Understanding.
2. Bakamides S., Dendrinos M., Carayannis G. (1991), *SVD analysis by synthesis of harmonic signals*, IEEE Trans. Signal Processing, **39**, 472–477.
3. Boll S.F. (1979), *Suppression of acoustic noise in speech using spectral subtraction*, IEEE Trans. Acoust. Speech Signal Process, **27**, 113–120.
4. Candes E.J., Plan Y. (2010), *Matrix Completion With Noise.*
5. Candes E.J., Terence T. (2010), *The power of convex relaxation: near-optimal matrix completion*, IEEE Transactions on Information Theory, **56**, 2053–2080.
6. Candes E.J., Li X., Ma Y., Wright J. (2011), *Robust Principal Component Analysis?*, Journal of the ACM, **58**, 1–37.
7. Chang S.G., Yu B., Vetterli M. (2000), *Adaptive Wavelet Thresholding for Image Denoising and Compression*, IEEE Transactions on Information Theory, **9**, 1532–1547.
8. Chambers J. (1977), *Computational method for data analysis*, New York, Wiley.
9. Dendrinos M., Bakamides S., Carayannis G. (1991), *Speech enhancement from noise: A regenerative approach*, Speech Communication, **10**, 45–57.

10. Ephraim Y., Malah D. (1984), *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*, IEEE Trans. Acoust. Speech Signal Process, **ASSP-32**, 109–1121.

11. Ephraim Y., Van Trees H. (1995), *A signal subspace approach for speech enhancement*, IEEE Trans. Speech Audio Process., **3**, 251–266.

12. Fazel M., Candes E., Recht B., Parrilo P. (2008), *Compressed sensing and robust recovery of low rank matrices*, [in:] Asilomar Conf. Signals, Systems, and Computers, Pacific Grove, CA.

13. Gannot S., Burshtein D., Weinstein E. (1998), *Iterative and Sequential Kalman filter based speech enhancement algorithms*, IEEE Trans. Acoust. Speech Signal Process, **6**, 373–385.

14. Golub G., Van Loan C. (1989), *Matrix computations*, 2nd ed, Baltimore, MD: The Johns Hopkins University Press.

15. Hu Y., Loizou P.C. (2003), *A Generalized Subspace Approach for Enhancing Speech Corrupted by Colored Noise*, IEEE Trans. on Speech and Audio Processing, **11**, 334–341.

16. Hu Y., Loizou P. (2008), *Evaluation of objective quality measures for speech enhancement*, IIEEE Trans. Speech Audio Process., **16**, 229–238.

17. Jax P., Vary P. (2003), *Artificial bandwidth extension of speech signals using MMSE estimation based on a hidden Markov medol*, TEEE International Conference on Acoudtics, Speech, and Signal Processing, **8**, 680–683.

18. Jin W., Scordilis M.S. (2006), *Speech enhancement by residual domain constrained optimization*, Speech Communication, **148**.

19. Jolliffe I.T. (2002), *Principal Component Analysis*, Springer, New York.

20. Kim J.B., Lee K.Y., Lee C.W. (2000), *On the applications of the interacting multiple model algorthm for enhancing noisy speech*, IEEE Trans. Acoust. Speech Signal Process, **8**, 349–352.

21. Mallat S. (1999), *A Wavelet Tour of Signal Processing*, California: Academic press 2nd Edition.

22. Mardani M., Mateos G. (2013), *Recovery of low-rank plus compressed sparse matrices with application to unveiling traffic anomalies*, IEEE Trans. Inf. Theory, **59**.

23. Moor B. (1993), *The singular value decomposition and long and short spaces of noisy matrix*, IEEE Transactions on Signal Processing, **41**, 9, 2826–2838.

24. Peng Y., Ganesh A., Wright J., Xu W., Ma Y. (2012), *RASL: Robust Alignment by Sparse and Low-rank Decomposition for Linearly Correlated Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence.

25. Plapous C., Marro C., Scalart P. (2006), *Improved Signal-to-Noise Ratio Estimation for Speech Enhancement*, IEEE Transactions on Acoustics, Speech, and Signal Processing, **14**, 2098–2108.

26. Quatieri T. (2002), *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice Hall, Upper Saddle River, NJ.

27. Saadoune A., Selouani A., Selouani S.A. (2014), *Perceptual subspace speech enhancement using variance of the reconstruction error*, Digital Signal Processing, **24**.

28. Sun C., Zhang Q., Wang M. (2014), *A novel speech enhancement method based on constrained low-rank and sparse matrix decomposition*, Speech Communication, pp. 44–55.

29. Toh K., Yun S. (2010), *An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems*, Pacific J. Optim., pp. 615–640.

30. Tufts D., Kumaresan R. (1982), *Esimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood*, Proc. IEEE, **70**.

31. Tufts D., Kumaresan R., Kirsteins I. (1982), *Data adaptive signal estimation by singular value decomposition of a data matrix*, Proc. IEEE, **70**, 684–685.

32. Vaseghi S.V. (2006), *Advanced Digital Signal Processing and Noise Reduction*, Third Edition, John Wiley & Sons Ltd.

33. Virag N. (1999), *Single channel speech enhancement based on masking properties of the human auditory system[J]*, IEEE Trans. Acoust. Speech Signal Process, **7**, 126–323.

34. Wright J., Peng Y., Ma Y. (2009), *Robust Principal Component Analysis: Exact Recovery of Corrupted Low-rank Matrices by Convex Optimization*, [in:] NIPS.

35. Xu H., Caramanis C., Sanghavi S. (2012), *Robust PCA via outlier pursuit*, IEEE Transactions on Information Theory, **58**, 3047–3064.

36. Zehtabian A., Hassanpour H., Zehtabian S. (2010), *A novel speech enhancement approach based on singular value decomposition and genetic algorithm*, International Conference of Soft Computing and Pattern Recognition, pp. 430–435.

37. Zhou X., Yang C., Yu W. (2013), *Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation*, IEEE Trans. on Pattern Analysis and Machine Intelligence, **35**, 597–610.

38. Zhou T., Tao D. (2011), *GoDec: Randomized Low-rank & Sparse Matrix Decomposition in Noisy Case*, [in:] Proceedings of the 28 th International Conference on Machine Learning, Bellevue, WA, USA.