Blind Speech Deconvolution via Pretrained Polynomial Dictionary and Sparse Representation

Jian Guan[1], Xuan Wang[1⋆], Shuhan Qi[1], Jing Dong[2], and Wenwu Wang[3]

[1]Shenzhen Graduate School
Harbin Institute of Technology, Shenzhen, 518055, China
{j.guan,wangxuan,shuhanqi}@cs.hitsz.edu.cn
[2] College of Electrical Engineering and Control Science
Nanjing Tech University, Nanjing, 211800, China
jingdong@njtech.edu.cn
[3] Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, GU2 7XH, United Kingdom
w.wang@surrey.ac.uk

Abstract. Blind speech deconvolution aims to estimate both the source speech and acoustic channel from the convolutive reverberant speech. The problem is ill-posed and underdetermined, which often requires prior knowledge for the estimation of the source and channel. In this paper, we propose a blind speech deconvolution method via a pretrained polynomial dictionary and sparse representation. A polynomial dictionary learning technique is employed to learn the dictionary from room impulse responses, which is then used as prior information to estimate the source and the acoustic impulse responses via an alternating optimization strategy. Simulations are provided to demonstrate the performance of the proposed method.

Keywords: blind deconvolution, speech dereverberation, polynomial dictionary learning, acoustic channel estimation

1 Introduction

Blind speech deconvolution aims to esimate the source speech and acoustic room impulse responses (RIRs) from the observed reverberant speech. This is a common problem which can be benificial for several applications, such as automatic speech recognition (ASR) [17], sound reproduction [3], and hearing aids [10]. In addition, it also has been used for multimedia affective computing [18, 14].

The reverberant speech $\mathbf{y} \in \mathbb{R}^L$ of a single-input and single-output (SISO) acoustic system is generated by the linear convolution of the source speech $\mathbf{s} \in \mathbb{R}^M$ and the acoustic RIR $\mathbf{h} \in \mathbb{R}^N$, which can be expressed as follows

$$\mathbf{y} = \mathbf{s} * \mathbf{h} + \mathbf{w}, \tag{1}$$

⋆ Corresponding author

where $*$ denotes the convolution operation, $\mathbf{w}$ is the system noise, and $L = M + N - 1$. In matrix form, (1) can be written as

$$\mathbf{y} = \mathcal{S}\mathbf{h} = \mathcal{H}\mathbf{s} + \mathbf{w}, \tag{2}$$

where $\mathcal{S} \in \mathbb{R}^{L\times N}$ and $\mathcal{H} \in \mathbb{R}^{L\times M}$ are the linear convolution matrices constructed from $\mathbf{s}$ and $\mathbf{h}$, respectively. Assuming both $\mathbf{s}$ and $\mathbf{h}$ are unknown, only given the observation $\mathbf{y}$, the estimation of $\mathbf{s}$ and $\mathbf{h}$ becomes blind, which is an ill-posed and undetermined inverse problem [2]. There are unlimited possible combinations of $\mathbf{s}$ and $\mathbf{h}$ which satisfy (1). To address this problem, prior information (such as sparsity of signals [13] [15] [16] or acoustic impulse responses [7]) is usually exploited to reduce the solution space for the estimation of $\mathbf{s}$ and $\mathbf{h}$.

In our previous work [7], we take into account the sparsity of a sparse SISO acoustic system as the prior knowledge to solve the blind speech deconvolution problem. The RIR of an acoustic system usually consists of three parts: direct path, early reflections and late reverberation. In [7], we assume the acoustic system has a very low reverberation level, so that the late reflection is negligible, and the RIR $\mathbf{h}$ of such an acoustic system can be seen as sparse. However, the proposed method in [7] cannot be used to fully recover the RIR of an acoustic system with a higher level of reverberation, as the sparse prior knowledge is not applicable for such an acoustic system. In some works, the pretrained dictionaries are used to exploit the prior knowledge of signals to be estimated in blind deconvolution. In [9], a learned dictionary is employed in blind deconvolution for infrared spectrum restoration, where each spectrum can be sparsely represented by the overcomplete dictionary. In [8], adaptive dictionary learning is applied for single image deblurring, where each image patch is encoded by the sparse representation of an overcomplete dictionary.

In this paper, we focus on blind speech deconvolution for an SISO acoustic system with a high level of reverbertion, and propose a blind speech deconvolution method by utilizing a dictionary learned from the acoustic RIRs. However, it is challenging to learn the dictionary from the signals with time delays (e.g. acoustic RIRs), as conventional dictionary learning methods cannot be applied directly to process the signals with time delays. In order to address this problem, we use a polynomial dictionary learning method [6] to learn the dictionary for the blind speech deconvolution, where the polynomial dictionary method is first proposed in our previous work [6]. In [6], we developed a polynomial dictionary learning technique to deal with the signals with time lags, and the proposed method employed polynomial matrices for acoustic RIRs modeling, where each RIR is split into several segments, and each segment can be seen as a FIR and modeled by a polynomial. Therefore, a polynomial dictionary can be learned, and the acoustic RIRs can then be sparsely represented by the learned polynomial dictionary. Here, such a polynomial dictionary is trained to provide the prior information by sparsely representing the acoustic RIR needed to be estimated for blind speech deconvolution. By using the polynomial dictionary, we introduce a blind speech deconvolution model with polynomial sparse representation, and propose an alternating optimization method to estimate the source speech and

acoustic RIR in two steps, by fixing one and updating the other iteratively. To the best of our knowledge, we are the first to introduce and apply the polynomial dictionary technique for blind speech deconvolution.

The rest of the paper is organized as follows: Section 2 briefly reviews the previous work about blind speech deconvolution and introduces the polynomial sparse representation. Section 3 presents the proposed model and method in details. Section 4 shows the simulations result; Section 5 concludes the paper and discusses the potential future work.

2 Preliminaries

2.1 Sparse Blind Speech Deconvolution

In [7], we assume an acoustic system is sparse, by taking into account the sparsity of RIR $\mathbf{h}$, a sparse blind speech deconvolution model is proposed as follows

$$F(\mathbf{s},\mathbf{h}) = \|\mathbf{s} * \mathbf{h} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{h}\|_1 + r(\mathbf{s}), \tag{3}$$

where the L2-norm is the data fidelity term, the regularization $r(\mathbf{s})$ is an indicator function [11] which accounts for the dynamic range of $\mathbf{s}$, and the L1-regularization takes into account the sparsity of $\mathbf{h}$, where λ is the penalty parameter. (3) can be reformulated as

$$\begin{aligned} F(\mathbf{s},\mathbf{h}) &= \|\mathcal{S}\mathbf{h} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{h}\|_1 + r(\mathbf{s}) \\ &= \|\mathcal{H}\mathbf{s} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{h}\|_1 + r(\mathbf{s}). \end{aligned} \tag{4}$$

An alternating optimization method is used in [7] to address the sparse blind speech deconvolution problem (3).

2.2 Polynomial Sparse Representation

In [6], we have developed a polynomial dictionary learning technique to process the acoustic signals with time delays, where the signals can be sparsely represented by a pretrained polynomial dictionary. First, the training acoustic signals are modeled by a polynomial matrix, and the polynomial dictionary is learned by using the method in [6]. Then, the learned polynomial dictionary is used to sparsely approximate the test acoustic signals.

Assuming $\mathbf{D}(z)$ is the learned polynomial dictionary, and $\mathbf{h}$ is the acoustic signal needed to be sparsely represented, then $\mathbf{h}$ can be modeled by a polynomial matrix $\mathbf{H}(z)$. Here, $\mathbf{H}(z)$ can be sparsely approximated as follows

$$\mathbf{H}(z) = \mathbf{D}(z)\mathbf{X}, \tag{5}$$

where $\mathbf{X}$ is the sparse representation matrix. (3) can be reformulated to the following equation

$$\underline{\mathbf{H}} = \underline{\mathbf{D}}\mathbf{X}, \tag{6}$$

where $\underline{\mathbf{H}}$ and $\underline{\mathbf{D}}$ denote the concatenated coefficient matrices of $\mathbf{H}(z)$ and $\mathbf{D}(z)$, respectively, defined as

$$\underline{\mathbf{H}} = [\mathbf{H}(0); \ldots; \mathbf{H}(\ell); \ldots; \mathbf{H}(\mathcal{L}-1)], \tag{7}$$

$$\underline{\mathbf{D}} = [\mathbf{D}(0); \ldots; \mathbf{D}(\ell); \ldots; \mathbf{D}(\mathcal{L}-1)], \tag{8}$$

where $\mathbf{H}(\ell)$ and $\mathbf{D}(\ell)$ are the coefficient matrices of $\mathbf{H}(z)$ and $\mathbf{D}(z)$ at lag ℓ, respectively. $\mathcal{L}$ denotes the maximum time lags of each polynomial element of the polynomial matrix.

In this paper, the polynomial dictionary $\mathbf{D}(z)$ is learned from the training acoustic RIRs, and the desired RIR $\mathbf{h}$ can be sparsely approximated by the pretrained dictionary $\mathbf{D}(z)$.

First, $\mathbf{h}$ is modeled by $\mathbf{H}(z)$, and reformulated as $\underline{\mathbf{H}}$. Then, $\underline{\mathbf{H}}$ can be sparsely represented by the linear combination of the new dictionary $\underline{\mathbf{D}}$ with $\mathbf{X}$, where $\mathbf{X}$ can be calculated by optimizing the following loss

$$\begin{aligned} &\min_{\mathbf{X}} \|\underline{\mathbf{H}} - \underline{\mathbf{D}}\mathbf{X}\|_F^2 \\ &\text{subject to } \forall i,\ \|\mathbf{x}_i\|_0 \leq \kappa, \end{aligned} \tag{9}$$

where $\mathbf{x}_i$ is the ith column from $\mathbf{X}$, and κ is sparsity which is the number of nonzero entries. Here, the orthogonal matching pursuit (OMP) algorithm [12] can be used to solve (9).

Once $\mathbf{X}$ is obtained, $\underline{\mathbf{H}}$ can be sparsely represented by using (6), so that the desired RIR $\mathbf{h}$ can be resembled by a "vectorization" operation, which aims to recover $\mathbf{h}$ from $\mathbf{H}(z)$ or $\underline{\mathbf{H}}$. The vectorization operation is defined as

$$\mathbf{h} = vec(\mathbf{H}(z)) = vec(\underline{\mathbf{H}}) = vec(\underline{\mathbf{D}}\mathbf{X}), \tag{10}$$

where vec is the vectorization operator.

3 Blind Speech Deconvolution with Polynomial Sparse Representation

3.1 Proposed Model

By employing the polynomial dictionary to sparsely approximate the RIR $\mathbf{h}$, we propose the blind speech deconvolution model as follows

$$F(\mathbf{s}, \mathbf{h}) = \|\mathbf{s} * \mathbf{h} - \mathbf{y}\|_2^2 + \gamma \|\mathbf{H}(z) - \mathbf{D}(z)\mathbf{X}\|_2^2 + \beta \sum \|\mathbf{x}_i\|_0 + r(\mathbf{s}), \tag{11}$$

where $\|\mathbf{s} * \mathbf{h} - \mathbf{y}\|_2^2$ represents the data fidelity term, and $*$ is the convolution operator. $\mathbf{H}(z)$ is the polynomial matrix used to model the RIR $\mathbf{h}$, and $\mathbf{D}(z)$ is the pretrained polynomial dictionary. Here, $\|\mathbf{H}(z) - \mathbf{D}(z)\mathbf{X}\|_2^2$ denotes the sparse representation of RIR $\mathbf{h}$ with $\mathbf{D}(z)$, and $\mathbf{X}$ is the representation coefficient matrix which is also unknown. γ and β are the penalty parameters. $r(\mathbf{s})$ is the regularization on $\mathbf{s}$, which is an indicator function accounts for the dynamic

range of source speech. It is used to reduce the solution space for $\mathbf{s}$ estimation. Note that, $r(\mathbf{s})$ is defined in our previous work [7].

According to (5) and (6), the proposed model (11) can be recast as

$$F(\mathbf{s},\mathbf{h}) = \|\mathbf{s} * \mathbf{h} - \mathbf{y}\|_2^2 + \gamma\|\underline{\mathbf{H}} - \underline{\mathbf{D}}\mathbf{X}\|_2^2 + \beta\sum\|\mathbf{x}_i\|_0 + r(\mathbf{s}). \tag{12}$$

In addition, (12) can be further reformulated as

$$\begin{aligned} F(\mathbf{s},\mathbf{h}) &= \|\mathcal{S}\mathbf{h} - \mathbf{y}\|_2^2 + \gamma\|\underline{\mathbf{H}} - \underline{\mathbf{D}}\mathbf{X}\|_2^2 + \beta\sum\|\mathbf{x}_i\|_0 + r(\mathbf{s}) \\ &= \|\mathcal{H}\mathbf{s} - \mathbf{y}\|_2^2 + \gamma\|\underline{\mathbf{H}} - \underline{\mathbf{D}}\mathbf{X}\|_2^2 + \beta\sum\|\mathbf{x}_i\|_0 + r(\mathbf{s}). \end{aligned} \tag{13}$$

An alternating optimization strategy is employed to solve (13), by fixing one and updating the other iteratively, and the proposed method includes two steps: **RIR h Estimation** and **Signal s Estimation**, as described next.

3.2 Proposed Alternating Optimization Method

RIR h Estimation In this step, by fixing signal $\mathbf{s}$, the optimization of cost function (13) is reduced to

$$\mathbf{h}^{(k+1)} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathcal{S}^{(k)}\mathbf{h} - \mathbf{y}\|_2^2 + \gamma\|\underline{\mathbf{H}} - \underline{\mathbf{D}}\mathbf{X}\|_2^2 + \beta\sum\|\mathbf{x}_i\|_0, \tag{14}$$

where $\mathcal{S}^{(k)}$ is the linear convolution matrix constructed from $\mathbf{s}^{(k)}$ at the kth iteration. Note that, $\mathbf{s}^{(0)}$ is initialized by using the observation $\mathbf{y}$. In order to optimize (14), we split (14) into two sub-optimization problems.

First, we start by initializing $\mathbf{h}$ with $\mathbf{h}_0$, as $\mathbf{h}^{(0)} = \mathbf{h}_0$. Here, $\mathbf{h}_0$ is a randomly generated Gaussian vector with zero mean and unit variance. Therefore, (14) can be reduced to the following sub-optimization problem

$$\mathbf{X}^{(k+1)} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\underline{\mathbf{H}}^{(k)} - \underline{\mathbf{D}}\mathbf{X}\|_2^2 + \beta\sum\|\mathbf{x}_i\|_0, \tag{15}$$

where $\underline{\mathbf{H}}^{(k)}$ is the concatenated coefficient matrix of $\mathbf{H}^{(k)}(z)$ used to model $\mathbf{h}^{(k)}$ at the kth iteration. Here, the sparse representation $\mathbf{X}$ can be calculated by optimizing (15), using the OMP algorithm.

Then, once $\mathbf{X}^{(k+1)}$ is obtained, (14) can be reduced to another sub-optimization problem, which is

$$\mathbf{h}^{(k+1)} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathcal{S}^{(k)}\mathbf{h} - \mathbf{y}\|_2^2 + \gamma\|\underline{\mathbf{H}} - \underline{\mathbf{D}}\mathbf{X}^{(k+1)}\|_2^2. \tag{16}$$

With a vectorization operation, (16) can be further rewritten as

$$\mathbf{h}^{(k+1)} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathcal{S}^{(k)}\mathbf{h} - \mathbf{y}\|_2^2 + \gamma\|\mathbf{h} - vec(\underline{\mathbf{D}}\mathbf{X}^{(k+1)})\|_2^2. \tag{17}$$

Here, the CVX toolbox [5] is used to optimize (17) for $\mathbf{h}$ estimation. After obtaining $\mathbf{h}^{(k+1)}$, we can update $\mathcal{H}^{(k+1)}$ for $\mathbf{s}$ estimation at the current iteration.

Signal s Estimation In this step, by fixing RIR $\mathbf{h}$, the optimization of (13) is reduced to the following form

$$\mathbf{s}^{(k+1)} = \underset{\mathbf{s}}{\operatorname{argmin}} \|\mathcal{H}^{(k+1)}\mathbf{s} - \mathbf{y}\|_2^2 + r(\mathbf{s}). \tag{18}$$

Here, we use a variable metric forward-backward method (VMFB) [4] to minimize (18) for $\mathbf{s}$ estimation.

Once $\mathbf{s}$ is updated, then $\mathcal{S}^{(k+1)}$ can be constructed by using $\mathbf{s}^{(k+1)}$ for updating $\mathbf{h}$ at the next iteration until satisfies the stopping criterion. Here, the stopping criterion is defined as $\|\mathbf{s}^{(k+1)} - \mathbf{s}^{(k)}\|_2^2 \leq \epsilon$, where ϵ is set as 10^{-6}. The proposed method is given in Algorithm 1.

Algorithm 1

Input: observation $\mathbf{y}$, polynomial dictionary $\mathbf{D}(z)$, penalty parameter γ, number of iterations I_k, stopping threshold ϵ

Output: $\mathbf{s}_{opt}$, $\mathbf{h}_{opt}$

Initialization: $\mathbf{s}^{(0)} = \mathbf{y}$, construct $\mathbf{S}^{(0)}$ from $\mathbf{s}^{(0)}$, $\mathbf{h}^{(0)} = \mathbf{h}_0$, model $\mathbf{h}^{(0)}$ by $\mathbf{H}^0(z)$ and reformulate as $\underline{\mathbf{H}}^0$, reformulate $\mathbf{D}(z)$ as $\underline{\mathbf{D}}$, $\gamma = 0.001$, $I_k = 1200$, $\epsilon = 10^{-6}$.

Iterations:

for $k = 1, \ldots, I_k$ **do**

RIR h Estimation:

Update $\mathbf{X}^{(k+1)}$ by solving (15), using OMP algorithm.

Sparsely approximate $\underline{\mathbf{H}}$ as $\underline{\mathbf{D}}\mathbf{X}^{(k+1)}$, and conduct vectorization operation as $vec(\underline{\mathbf{D}}\mathbf{X}^{(k+1)})$.

Update $\mathbf{h}^{(k+1)}$ by optimizing (17), using CVX toolbox.

Construct $\mathcal{H}^{(k+1)}$ from $\mathbf{h}^{(k+1)}$.

Signal s Estimation:

Update $\mathbf{s}^{(k+1)}$ by minimizing (18), using VMFB as in [7].

Construct $\mathcal{S}^{(k+1)}$ from $\mathbf{s}^{(k+1)}$.

Stopping criterion: If $\|\mathbf{s}^{(k+1)} - \mathbf{s}^{(k)}\|_2^2 <= \epsilon$, then $\mathbf{s}_{opt} = \mathbf{s}^{(k+1)}$, $\mathbf{h}_{opt} = \mathbf{h}^{(k+1)}$, and break, else continue.

end for

4 Simulations and Results

In this section, we carry out several experiments by using the proposed method for blind speech deconvolution. The polynomial dictionary used is pretrained from the simulated acoustic RIRs, and the proposed method is applied to deconvolve the reverberant speech with different levels of noise. Here, reconstruction error is used to evaluate the performance for the signal and RIR estimation, defined as

$$R_{err} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{a}_i - \hat{\mathbf{a}}_i)^2, \tag{19}$$

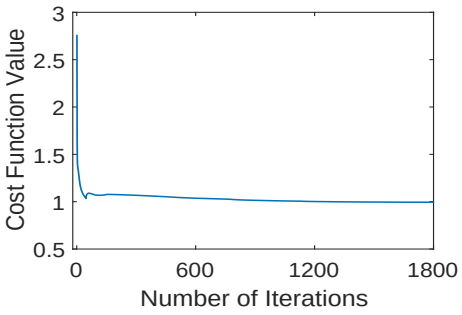

Fig. 1: Convergence of the proposed method.

where n is the length of $\mathbf{a}$, $\mathbf{a}$ and $\hat{\mathbf{a}}$ represent the original signal and estimated signal respectively, both normalized as $\mathbf{a} = \frac{\mathbf{a}}{max(|\mathbf{a}|)}$. Note that, $\mathbf{a}$ can denote either signal $\mathbf{s}$ or RIR $\mathbf{h}$.

4.1 Experimental Setup

The polynomial dictionary is learned from 1200 acoustic RIRs, which are simulated in a small room size of $8 \times 10 \times 3$ by room image model [1], where the test RIR $\mathbf{h}$ is generated in the same way. The length of the simulated RIR is 1200. 5 speech signals are selected from TIMIT database as source signals, where the length of each signal is 500. The first 500 samples from the observation $\mathbf{y}$ (discounting the silence part) are used to initialize the signal $\mathbf{s}$.

4.2 Results and Analysis

First, we conduct experiments to illustrate the performance of the proposed method for blind speech deconvolution, and compared it with the method in [7]. Here, no noise is added to the observation $\mathbf{y}$.

Figure 1 shows the change of cost function value with iterations. We can see that the proposed method can converge within 1800 iterations. Note that, the value becomes stable after 1200 iterations, so that the number of iterations is set to be 1200 in the following experiments. Figure 2 illustrates the deconvolution performance of the proposed method, as compared with the method in [7]. As can be seen from Figure 2, both methods can resemble the source signal very well, as shown in the subplots (c) and (e). However, the RIR estimated by the method in [7] is more sparse, where the late part is almost smoothed, as shown in the subplot (f). This is reasonable, as the method in [7] only considers the sparsity of an acoustic system, and imposes an L1-norm constraint on RIR for $\mathbf{h}$ estimation. In contrast, the proposed method can achieve a better performance for RIR estimation, which can fully recover the RIR including the late reflections, as shown in the subplot (d). This is because that the polynomial dictionary used to sparsely approximate the RIR can provide useful prior information for the RIR estimation.

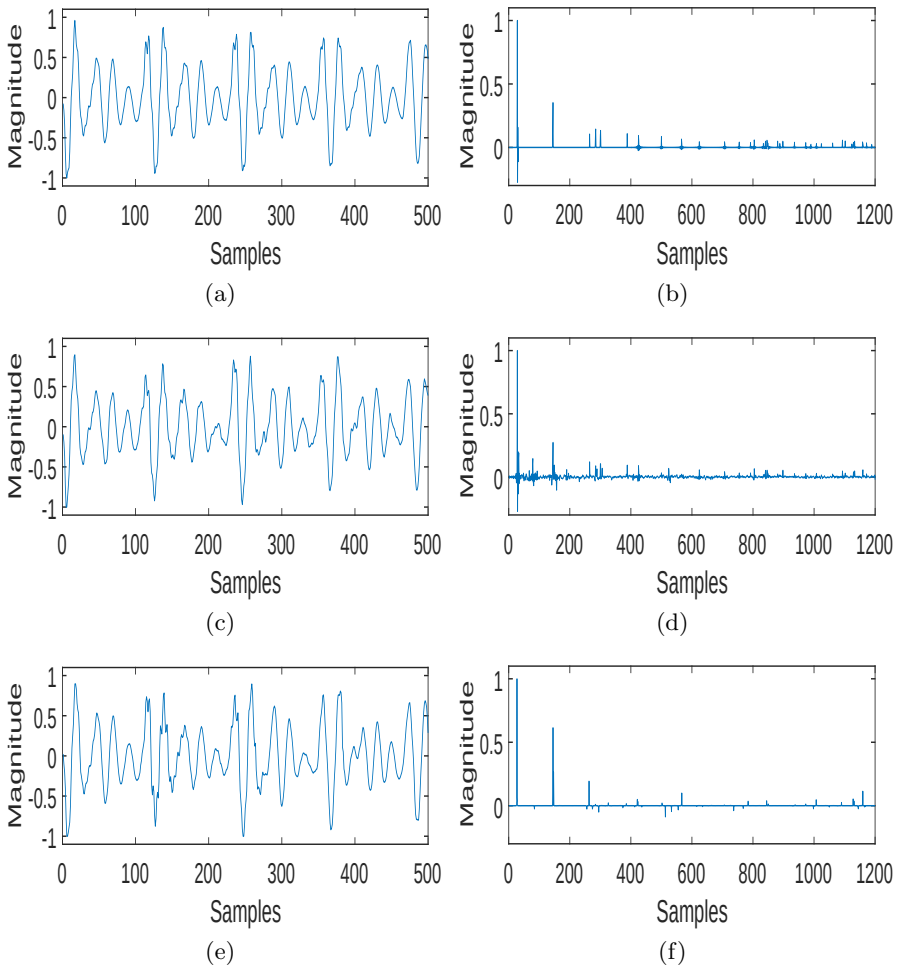

Fig. 2: Illustration of blind speech deconvolution by using the proposed method, and the method in [7]: (a) and (b) are the original source signal and RIR respectively; (c) and (d) are the estimated source and RIR respectively, by the proposed method; (e) and (f) are the estimated source and RIR respectively, by the method in [7].

Then, another experiment is carried out to further evaluate the performance of the proposed method, where white Gaussian noise with zero mean and variance chosen to achieve different signal-to-noise ratios (SNRs) is added to the observation $\mathbf{y}$, i.e., for SNR at 10 dB, 20 dB and 30 dB. In addition, we also test the observation $\mathbf{y}$ without noise interference. 5 speech signals and one RIR are tested, and 20 realizations are conducted for each noise level. Table 1 shows the average reconstruction errors.

Table 1: Performance comparison in terms of reconstruction error for signal and RIR estimation at different noise levels (SNRs). Note that, N denotes no noise is added to the observation $\mathbf{y}$.

Noise levels (dB)		10	20	30	N
Proposed method	Signal error	0.2840	**0.1628**	**0.1456**	**0.1302**
	RIR error	0.0588	**0.0306**	**0.0256**	**0.0139**
Method in [7]	Signal error	0.1838	0.1685	0.1662	0.1621
	RIR error	0.0346	0.0351	0.0353	0.0365

From Table 1, we can see that the proposed method can provide better performance for signal and RIR estimation with low levels of noise (e.g., for SNR at 20 dB and 30 dB), as well as without noise interference. However, the method in [7] obtains a better estimation result for a high level of noise (e.g., for SNR at 10 dB).

5 Conclusions and Future Work

In this paper, we introduced a blind speech deconvolution model with polynomial sparse representation term, where a polynomial dictionary was used to provide the prior information for blind speech deconvolution. A polynomial dictionary learning technique was employed to train the polynomial dictionary and sparsely approximate the acoustic RIR for blind speech deconvolution. In order to estimate both the signal and RIR, an alternating optimization method was then proposed to address the blind speech deconvolution problem. The proposed method was applicable for the acoustic system with different levels of reverberation and can be used to fully recover the RIR with late reflections, as the pretrained polynomial dictionary can better fit the sparse representation of the RIRs from different kind of acoustic systems. Simulations demonstrated the validity of the proposed method, and the results shown that the proposed method can provide better performance for blind speech deconvolution, especially, for the RIR estimation. In our future work, we will learn polynomial dictionary from real acoustic RIRs, and apply the proposed method for real acoustic systems.

6 Acknowledgements

The work was initially conducted when J. Guan was visiting the Centre for Vision, Speech and Signal Processing (CVSSP), the University of Surrey. This work was supported in part by International Exchange and Cooperation Foundation of Shenzhen City, China (No. GJHZ20150312114149569).

References

1. Allen, J.B., Berkley, D.A.: Image method for efficiently simulating small-room acoustics. The Journal of the Acoustical Society of America 65(4), 943–950 (1979)

2. Benichoux, A., Vincent, E., Gribonval, R.: A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors. In: Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP). pp. 6108–6112. IEEE (2013)
3. Betlehem, T., Abhayapala, T.D.: A modal approach to soundfield reproduction in reverberant rooms. In: Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP). vol. 3, pp. 289 –292. IEEE (2005)
4. Chouzenoux, E., Pesquet, J.C., Repetti, A.: A block coordinate variable metric forward–backward algorithm. Journal of Global Optimization pp. 1–29 (2016)
5. Grant, M., Boyd, S., Grant, M., Boyd, S., Blondel, V., Boyd, S., Kimura, H.: Cvx: Matlab software for disciplined convex programming, version 2.1. Recent Advances in Learning and Control pp. 95–110 (2014)
6. Guan, J., Dong, J., Wang, X., Wang, W.: A polynomial dictionary learning method for acoustic impulse response modeling. In: Latent Variable Analysis and Signal Separation, Lecture Notes in Computer Science, vol. 9237, pp. 211–218. Springer (2015)
7. Guan, J., Wang, X., Wang, W., Huang, L.: Sparse blind speech deconvolution with dynamic range regularization and indicator function. Circuits, Systems, and Signal Processing pp. 1–16 (2017)
8. Hu, Z., Huang, J.B., Yang, M.H.: Single image deblurring with adaptive dictionary learning. In: Proc. of International Conference on Image Processing (ICIP). pp. 1169–1172. IEEE (2010)
9. Liu, H., Liu, S., Huang, T., Zhang, Z., Hu, Y., Zhang, T.: Infrared spectrum blind deconvolution algorithm via learned dictionaries and sparse representation. Applied Optics 55(10), 2813–2818 (2016)
10. Löllmann, H.W., Vary, P.: Low delay noise reduction and dereverberation for hearing aids. EURASIP Journal on advances in signal processing 2009(1), 437–807 (2009)
11. Parikh, N., Boyd, S.: Proximal algorithms. Foundations and Trends in Optimization 1(3), 127–239 (2014)
12. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.: Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In: Proc. of the 27th Asilomar Conference on Signals, Systems and Computers. pp. 40–44. IEEE (1993)
13. Repetti, A., Pham, M.Q., Duval, L., Chouzenoux, E., Pesquet, J.C.: Euclid in a taxicab: Sparse blind deconvolution with smoothed regularization. IEEE Signal Processing Letters 22(5), 539–543 (2015)
14. Schuller, B.: Affective speaker state analysis in the presence of reverberation. International Journal of Speech Technology 14(2), 77–87 (2011)
15. Selesnick, I.: Sparse deconvolution (an MM algorithm). [Online]. Available: `http://cnx.org/content/m44991/` (2012)
16. Wang, L., Chi, Y.: Blind deconvolution from multiple sparse inputs. IEEE Signal Processing Letters 23(10), 1384–1388 (2016)
17. Wang, L., Nakagawa, S., Kitaoka, N.: Blind dereverberation based on cmn and spectral subtraction by multi-channel LMS algorithm. In: Proc. of INTERSPEECH 2008. pp. 1032–1035 (2008)
18. Zhao, S., Yao, H., Gao, Y., Ji, R., Ding, G.: Continuous probability distribution prediction of image emotions via multitask shared sparse regression. IEEE Transactions on Multimedia 19(3), 632–645 (2017)