Joint $L1 - L2$ Regularisation for Blind Speech Deconvolution

Jian Guan[1], Xuan Wang[1⋆], Zongxia Xie[2], Shuhan Qi[1], and Wenwu Wang[3]

[1]Shenzhen Graduate School
Harbin Institute of Technology, Shenzhen, 518055, China
{j.guan,wangxuan,shuhanqi}@cs.hitsz.edu.cn
[2] School of Computer Software
Tianjin University, 30072, China
caddiexie@hotmail.com
[3] Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, GU2 7XH, United Kingdom
w.wang@surrey.ac.uk

Abstract. The purpose of blind speech deconvolution is to recover both the original speech source and the room impulse response (RIR) from the observed reverberant speech. This can be beneficial for speech intelligibility and speech perception. However, the problem is ill-posed, which often requires additional knowledge to solve. In order to address this problem, prior informations (such as the sparseness of signal or acoustic channel) are often exploited. In this paper, we propose a joint $L1 - L2$ regularisation based blind speech deconvolution method for a single-input and single-output (SISO) acoustic system with a high level of reverberation, where both the sparsity and density of the room impulse responses (RIR) are considered, by imposing an $L1$ and $L2$ norm constraint on their early and late part respectively. By employing an alternating strategy, both the source signal and early part in the RIR can be well reconstructed while the late part of the RIR can be suppressed.

Keywords: blind deconvolution, joint regularisation, reverberation suppression, acoustic channel estimation

1 Introduction

With the rapid development of internet and widespread use of intelligent terminals, speech becomes an important approach for human-machine interaction. However, the speech intelligibility and speech quality are often degraded by room reverberation, which leads to the decrease of the performance of many applications, such as automatic speech recognition (ASR) system [13], hands-free teleconferencing [12], hearing-aids [9]. Moreover, reverberation also has detrimental effects on listening comprehension and perception of affective speech signals

⋆ Corresponding author

with emotion and interest [11, 15]. Therefore, speech dereverberation is of great importance in speech signal processing.

In our recent work, we focus on blind speech deconvolution for an SISO acoustic system, where the reverberant speech can be considered as the convolution of a source signal with the RIR. Our aim is to estimate both the source signal and the RIR from the reverberant speech, which is an ill-posed inverse problem. To solve this problem, we need to exploit prior information (e.g. sparsity) to reduce the possible solution space. However, speech signal is not naturally sparse in time domain unless converted to a transform domain by using the pre-defined dictionary (e.g. the DFT matrix) or the learned dictionary. In this paper, we instead take into account the prior information provided by the RIR of an acoustic system. The RIR usually includes three parts: direct path, early reflections and late reverberation, where the direct path and early reflections have a distinctive characteristic compared to the late reverberation, and can be considered as sparse in the time domain, whereas the late part of RIR is dense. A room that has a relatively low level of reflections can be assumed as a sparse acoustic system. In [8], we have proposed a sparse blind speech deconvolution method for such an acoustic system, where the sparsity of the RIR is exploited by imposing an $L1$ norm constraint on the RIR and a constraint on the dynamic range of the source speech in the cost function. An alternating minimisation s-trategy has been employed for the estimation of the source speech and the sparse RIR. Although the method in [8] works well for sparse RIR, it is limited in a relatively high level of reverberation, as the dense late part of the RIR has not been properly considered.

In this work, both the sparsity and density of RIR are taken into account as the prior information for blind speech deconvolution. We propose a joint $L1-L2$ regularisation based blind deconvolution model, where an $L1$ norm constraint is imposed to exploit the sparsity of the early reflections, and an $L2$ norm constraint is used to capture the density of the late reflections. Moreover, an indicator function based regularisation is applied to the source signal to account for the dynamic range of the signal. A minimisation method is proposed to optimise the cost function, using two steps: *h-step* and *x-step*, by fixing one, updating the other, in an alternating manner.

This paper is organised as follows: Section 2 briefly reviews the previous work about sparse blind speech deconvolution; Section 3 describes the proposed joint $L1-L2$ norm based method; Section 4 presents the evaluation results; and Section 5 gives the conclusion.

2 Background and previous work

2.1 Blind speech deconvolution

A reverberant speech $\boldsymbol{y} \in \mathbf{R}^{N+L-1}$ can be considered as the convolution of the source speech signal $\boldsymbol{x} \in \mathbf{R}^{N}$ and the RIR $\boldsymbol{h} \in \mathbf{R}^{L}$. Blind speech deconvolution of an SISO acoustic system can be modelled as follows

$$\boldsymbol{y} = \boldsymbol{x} * \boldsymbol{h}, \tag{1}$$

where * denotes the convolution operator. Given $\boldsymbol{y}$ as an input, the aim is to estimate the unknown source $\boldsymbol{x}$ and the RIR $\boldsymbol{h}$.

Equation (1) can be rewritten in a matrix form as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} = \boldsymbol{H}\boldsymbol{x}, \tag{2}$$

where $\boldsymbol{X} \in \mathbf{R}^{(N+L-1)\times L}$ and $\boldsymbol{H} \in \mathbf{R}^{(N+L-1)\times N}$ are the linear convolution matrices constructed from $\boldsymbol{x}$ and $\boldsymbol{h}$ respectively.

2.2 Previous work

In [8], we assume that an acoustic system has a relatively low level of reflections, so that the RIR $\boldsymbol{h}$ can be considered as sparse.

By exploiting the sparsity of RIR $\boldsymbol{h}$, we proposed a blind speech deconvolution model as follows

$$F(\boldsymbol{x}, \boldsymbol{h}) = \|\boldsymbol{x} * \boldsymbol{h} - \boldsymbol{y}\|_2^2 + \gamma\|\boldsymbol{h}\|_1 + \boldsymbol{r}(\boldsymbol{x}), \tag{3}$$

where $\|\boldsymbol{x} * \boldsymbol{h} - \boldsymbol{y}\|_2^2$ denotes the data fidelity term, $\gamma\|\boldsymbol{h}\|_1$ is the $L1$ norm constraint, which accounts for the sparsity of RIR $\boldsymbol{h}$, and γ is the regularisation parameter. Note that, only using sparsity prior is insufficient to guarantee the identifiability for the ill-posed blind deconvolution problem [3] [4]. An additional regularisation $\boldsymbol{r}(\boldsymbol{x})$ is considered to further reduce the solution space for the estimation of $\boldsymbol{x}$. Here, $\boldsymbol{r}(\boldsymbol{x})$ is defined as an indicator function [6] [14], which accounts for the dynamic range of the source signal.

An alternating optimisation strategy is adopted to address (3) in two steps: sparse RIR estimation and signal estimation, where $\boldsymbol{h}$ and $\boldsymbol{x}$ are estimated iteratively, by fixing one, and updating the other.

Sparse RIR estimation In this step, given $\boldsymbol{x}$, $\boldsymbol{h}$ can be estimated by optimising the following problem

$$\boldsymbol{h}^{(k+1)} = \underset{\boldsymbol{h}}{\operatorname{argmin}} \|\boldsymbol{X}^{(k)}\boldsymbol{h} - \boldsymbol{y}\|_2^2 + \gamma\|\boldsymbol{h}\|_1, \tag{4}$$

where $\boldsymbol{X}^{(k)}$ is the linear convolution matrix constructed from $\boldsymbol{x}^{(k)}$ in the kth iteration. Here, (4) is an $L1$ regularised least-squares problem, which can be optimised by existing alogrithms, such as *l1_ls* [10], ADMM [2], and the CVX toolbox [7]. In [8], we use *l1_ls* to update $\boldsymbol{h}$ at each iteration, once we obtain $\boldsymbol{h}^{(k+1)}$, the convolution matrix $\boldsymbol{H}^{(k+1)}$ can then be constructed, and used to update $\boldsymbol{x}$ at this iteration.

Signal estimation In this step, $\boldsymbol{x}$ is updated by minimising the following function

$$\boldsymbol{x}^{(k+1)} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{H}^{(k+1)}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \boldsymbol{r}(\boldsymbol{x}), \tag{5}$$

where a variable metric forward-backward (VMFB) method [5] is applied to optimise (5).

As shown in [8], this method works well for sparse acoustic system, but is limited when dealing with late reverberation. In this paper, we propose a joint regularisation model by using an $L1$ and $L2$ norm constraint to exploit the sparsity and density of RIR, respectively, as detailed next.

3 Joint $L1 - L2$ norm based blind speech deconvolution

We assume $\boldsymbol{h}_e$ and $\boldsymbol{h}_l$ are the early and late reflections of the RIR $\boldsymbol{h}$ of an acoustic system respectively, here the direct path of the RIR is included in $\boldsymbol{h}_e$. By taking into account the characteristics of both the early and late part of $\boldsymbol{h}$, our proposed joint $L1 - L2$ regularisation blind deconvolution model is formulated as follows

$$F(\boldsymbol{x}, \boldsymbol{h}) = \|\boldsymbol{x} * \boldsymbol{h} - \boldsymbol{y}\|_2^2 + \gamma\|\boldsymbol{h}_e\|_1 + \lambda\|\boldsymbol{h}_l\|_2^2 + \boldsymbol{r}(\boldsymbol{x}), \tag{6}$$

where $\|\boldsymbol{x} * \boldsymbol{h} - \boldsymbol{y}\|_2^2$ is the data fidelity term, the $L1$ and $L2$ norms account for the sparsity of $\boldsymbol{h}_e$ and the density of $\boldsymbol{h}_l$, and γ and λ are the regularisation parameters of these two terms respectively. $\boldsymbol{r}(\boldsymbol{x})$ is an indicator function considering the dynamic range of the source signal. As the source signal $\boldsymbol{x}$ is unknown, $\boldsymbol{r}(\boldsymbol{x})$ is approximated using the observation $\boldsymbol{y}$ as in [8].

Note that, it is reasonable to formulate the blind deconvolution problem by exploiting the early and late part of $\boldsymbol{h}$ separately as (6). The early reflections are usually caused by the speech signal arriving at the microphone within 5 ms to 100 ms after the direct sound. The length of $\boldsymbol{h}_e$ can be calculated once the sampling frequency is given, or estimated by some methods according to reverberation time RT60. In this paper, we simply assume that the early reflection is within the beginning 30 ms of $\boldsymbol{h}$. Typically, for a sampling frequency at 16 kHz, the length of $\boldsymbol{h}_e$ would be 480 samples.

The proposed joint norm model (6) can be expressed in matrix form as follows

$$F(\boldsymbol{x}, \boldsymbol{h}) = \|\boldsymbol{X}\boldsymbol{h} - \boldsymbol{y}\|_2^2 + \gamma\|\boldsymbol{h}_e\|_1 + \lambda\|\boldsymbol{h}_l\|_2^2 + \boldsymbol{r}(\boldsymbol{x}), \tag{7}$$

or, equally as

$$F(\boldsymbol{x}, \boldsymbol{h}) = \|\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \gamma\|\boldsymbol{h}_e\|_1 + \lambda\|\boldsymbol{h}_l\|_2^2 + \boldsymbol{r}(\boldsymbol{x}). \tag{8}$$

Furthermore, Equation (7) can be rewritten as

$$F(\boldsymbol{x}, \boldsymbol{h}) = \| \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{D_l} \end{bmatrix} \boldsymbol{h} - \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} \|_2^2 + \gamma\|\boldsymbol{D_e}\boldsymbol{h}\|_1 + \boldsymbol{r}(\boldsymbol{x}), \tag{9}$$

where $\boldsymbol{D_e} \in \mathbf{R}^{L\times L}$ and $\boldsymbol{D_l} \in \mathbf{R}^{L\times L}$ are the matrices defined as $\boldsymbol{D_e} = \begin{bmatrix} \boldsymbol{I_e} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$ and $\boldsymbol{D_l} = \sqrt{\lambda}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I_l} \end{bmatrix}$ respectively. Here, $\boldsymbol{I_e}$ and $\boldsymbol{I_l}$ are the identity matrices, which are designed according to the length of the early and late part of $\boldsymbol{h}$ respectively.

Let $\boldsymbol{\mathcal{X}} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{D_l} \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0} \end{bmatrix}$, where $\boldsymbol{\mathcal{X}} \in \mathbf{R}^{(N+2L-1)\times L}$ and $\mathbf{y} \in \mathbf{R}^{N+2L-1}$, then we can recast (9) as the following cost function

$$F(\boldsymbol{x}, \boldsymbol{h}) = \|\boldsymbol{\mathcal{X}}\boldsymbol{h} - \mathbf{y}\|_2^2 + \gamma\|\boldsymbol{D_e}\boldsymbol{h}\|_1 + \boldsymbol{r}(\boldsymbol{x}). \tag{10}$$

It can be seen that the reformulated cost function (10) is similar to (3), so that a similar alternating optimisation method as in [8] can be used to optimise (10).

In *h-step*, $\boldsymbol{x}$ is fixed, by using (10), $\boldsymbol{h}$ is updated by minimising the following loss

$$\boldsymbol{h}^{(k+1)} = \underset{\boldsymbol{h}}{\operatorname{argmin}} \|\boldsymbol{\mathcal{X}}^{(k)}\boldsymbol{h} - \mathbf{y}\|_2^2 + \gamma\|\boldsymbol{D_e}\boldsymbol{h}\|_1, \tag{11}$$

where $\boldsymbol{\mathcal{X}}^{(k)}$ is constructed by $\boldsymbol{X}^{(k)}$ and $\boldsymbol{D_l}$ obtained in the kth iteration. Here, (11) can be seen as a special case of $L1$ norm regularised least-squares optimisation problem, hence we can use the CVX toolbox to optimise (11). After obtaining $\boldsymbol{h}^{(k+1)}$, $\boldsymbol{H}^{(k+1)}$ can be constructed for updating $\boldsymbol{x}$ at this iteration.

In *x-step*, $\boldsymbol{h}$ is fixed, by using (8), then $\boldsymbol{x}$ can be updated by using the same cost function as (5). So that, we can also use the VMFB algorithm as in [8] to optimise (5) for the estimating of signal $\boldsymbol{x}$.

Algorithm 1

Input: $\boldsymbol{y}$, γ, λ, I_K
Initialisation: initialise $\boldsymbol{x}^{(0)}$ by using $\boldsymbol{y}$, construct $\boldsymbol{X}^{(0)}$ from $\boldsymbol{x}^{(0)}$, construct $\boldsymbol{D_e}$ and $\boldsymbol{D_l}$, initialise $\boldsymbol{\mathcal{X}}^{(0)}$ and $\mathbf{y}$ as $\boldsymbol{\mathcal{X}}^{(0)} = \begin{bmatrix} \boldsymbol{X}^{(0)} \\ \boldsymbol{D_l} \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0} \end{bmatrix}$ respectively, $I_K = 500$, $\epsilon = 10^{-6}$.
Iterations:
for $k = 1, \ldots, I_K$ **do**
$\boldsymbol{h} - \boldsymbol{step}$:
Update $\boldsymbol{h}^{(k+1)}$ by solving (11), using CVX toolbox.
Construct $\boldsymbol{H}^{(k+1)}$ from $\boldsymbol{h}^{(k+1)}$.
$\boldsymbol{x} - \boldsymbol{step}$:
Update $\boldsymbol{x}^{(k+1)}$ by solving (5), using VMFB as in [8].
Construct $\boldsymbol{X}^{(k+1)}$ from $\boldsymbol{x}^{(k+1)}$.
Update $\boldsymbol{\mathcal{X}}^{(k+1)}$, as $\boldsymbol{\mathcal{X}}^{(k+1)} = \begin{bmatrix} \boldsymbol{X}^{(k+1)} \\ \boldsymbol{D_l} \end{bmatrix}$.
Stopping criterion: If $\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|_2^2 <= \epsilon$, then $\boldsymbol{x}_{opt} = \boldsymbol{x}^{(k+1)}$, $\boldsymbol{h}_{opt} = \boldsymbol{h}^{(k+1)}$, and break, else continue.
end for
Output: $\boldsymbol{x}_{opt}$, $\boldsymbol{h}_{opt}$

Compared with the model (3), where only the $L1$ penalty parameter γ needs to be tuned, in (6), an additional parameter λ needs to be controlled. In this work, the penalty parameters are found empirically, and we will show how the penalty parameters impact on the estimation of $\boldsymbol{x}$ and $\boldsymbol{h}$ in Section 4. Here, the proposed algorithm is given in Algorithm 1.

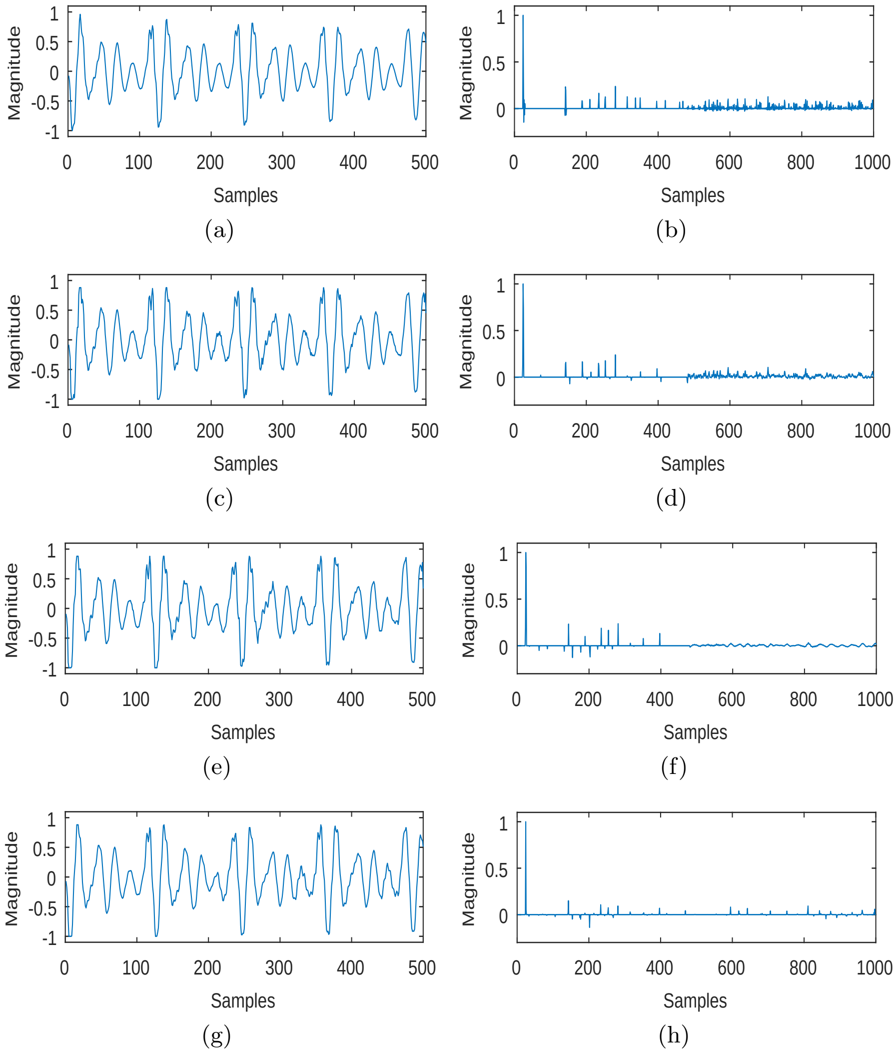

Fig. 1: Illustration of blind speech deconvolution by using the proposed method with different λ values, as compared with the method in [8]: (a) and (b) are the original signal and RIR respectively; (c) and (d) are the estimated source and RIR respectively, by the proposed method (for $\lambda = 0.005$); (e) and (f) are the estimated source and RIR respectively, by the proposed method (for $\lambda = 0.5$); (g) and (h) are the estimated source and RIR respectively, by the method in [8].

4 Simulations and results

4.1 Experimental setup

Experiments were conducted to demonstrate the validity of the proposed method for speech dereverberation. The reverberant speech $\boldsymbol{y}$ was generated by the linear convolution of source speech $\boldsymbol{x}$ and RIR $\boldsymbol{h}$, simulated in a $8 \times 10 \times 3$ room by a room image model [1]. Here, we assumed that the early reflections were generated within 30 ms. Therefore, $\boldsymbol{h}_e$ was taken as the first 480 samples of $\boldsymbol{h}$, and $\boldsymbol{h}_l$ was selected as the following 520 samples from $\boldsymbol{h}$. The overall length of $\boldsymbol{h}$ was 1000. Six speech signals (3 males and 3 females) from the TIMIT database were used as the source signals, each of length 500 samples. The observation $\boldsymbol{y}$ was used to initialise $\boldsymbol{x}$ by taking the first 500 samples from $\boldsymbol{y}$ (discounting the silence part) added with 0 dB white Gaussian noise. The maximum iteration number of the proposed method was set as 500, and the stopping criterion was $\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|_2^2 <= \epsilon$, where ϵ was set as 10^{-6}. The values of the penalty parameters γ and λ were set according to the experiments to be discussed later.

4.2 Performance indices

The reconstruction error is used as the performance index to evaluate the estimation accuracy of the source signal and RIR estimation, defined as

$$R_{err} = 20\log_{10}\left(\sqrt{\frac{1}{M}\sum_{i=1}^{M}(\hat{e}_i - \overline{e}_i)^2}\right), \tag{12}$$

where R_{err} represents the residual error (in decibels) between the original signal $\overline{\boldsymbol{e}}$ and the estimated one $\hat{\boldsymbol{e}}$ (both normalized as $\boldsymbol{e} = \frac{\boldsymbol{e}}{max(|\boldsymbol{e}|)}$), and M is the length of the signal. Here, $\boldsymbol{e}$ can be either $\boldsymbol{x}$ or $\boldsymbol{h}$.

In addition, a reverberation index is used to evaluate reverberation suppression by calculating the ratio of the energy in the early reflections to the energy in late reflections, which is

$$RR = 10log_{10}\frac{\|\boldsymbol{h}_e\|_2^2}{\|\boldsymbol{h}_l\|_2^2}. \tag{13}$$

4.3 Results and analysis

First, experiments were conducted to illustrate the performance of the proposed method as compared with the method in [8], for deconvolving the reverberant speech as generated in Section 4.1. Here, two experiments were carried out to deconvolve the same reverberant signal by using the proposed joint norm method, where the $L1$ penalty parameter was set to be the same, which was $\gamma = 0.0038$; whereas the $L2$ penalty parameters were different, which were $\lambda = 0.005$ and $\lambda = 0.5$ respectively. With different values of λ, we are able to show how the $L2$ penalty parameter impacts on the deconvolution performance, especially, for RIR estimation. The result is given in Figure 1.

From Figure 1, we can see that both of the estimated source signals resemble the original source. The early parts of the estimated RIR (as shown in the subplots (d) and (f)) are also similar to that of the ground truth RIR in (b). It is worth noting that the late reflections are reduced to a lower level as shown in the

Table 1: Comparison of the reconstruction errors (dB) for different penalty parameters.

$\boldsymbol{\lambda}$	0.0001	0.005	0.01	0.1	0.5	1	2
$\boldsymbol{\gamma = 0.0038}$							
$\boldsymbol{x}$ **error**	-19.03	-19.11	-19.11	-18.97	-18.72	-18.53	-18.44
$\boldsymbol{h}$ **error**	-31.34	-31.46	-31.38	-30.94	-30.75	-30.77	-30.80
$\boldsymbol{\gamma = 0.01\gamma_{max}}$							
$\boldsymbol{x}$ **error**	-18.94	-19.01	-18.99	-18.79	-19.76	-19.64	-19.28
$\boldsymbol{h}$ **error**	-30.22	-29.94	-29.47	-29.96	-31.12	-30.99	-30.69

subplot (f), which means the late reflections can be suppressed by increasing the $L2$ norm penalty parameter. This is beneficial for dereverberation or enhancement of an acoustic system. Also note that, the RIR can be well recovered by the proposed joint norm method with a smaller $L2$ penalty parameter as shown in the subplot (d), whereas the estimated RIR in the subplot (h) by the method in [8] is more sparse when compared with the original RIR in (b). This is because the method in [8] only used a sparsity constraint for the RIR $\boldsymbol{h}$ estimation.

Another experiment was carried out to show how the estimation results were influenced by the penalty parameters, where 6 source signals and one RIR as mentioned in Section 4.1 were used to generate 6 reverberant speech signals for the deconvolution tests. In order to see how λ influences the RIR estimation, especially for the suppression of the late part of the estimated RIR, the value of γ was fixed to $\gamma = 0.0038$ first, which was found empirically; then a warm starting strategy [14] was used for $\boldsymbol{h}$ estimation in each iteration, by setting $\gamma = 0.01\gamma_{max}$, where $\gamma_{max} = \|{\mathcal{X}^{(k)}}^T \mathbf{y}\|_\infty$. Table 1 shows the comparison of the average reconstruction errors with different parameter values.

From Table 1, we can see that the signal reconstruction error is increasing when the parameter λ is increased from 0.005 to 2 for a fixed $L1$ parameter γ, meanwhile, the RIR reconstruction error is also increasing when λ is increased from 0.005 to 0.5, which indicates λ is effective only in a certain range. When the warm starting strategy is used, the reconstruction errors of signal and RIR are fluctuating when λ is increasing, and the smallest reconstruction error can be obtained when $\lambda = 0.5$.

Note that, for a fixed γ, the variation of the RIR reconstruction error is larger than that of the signal within a certain range of λ (e.g. from 0.005 to 0.5). The increase in the reconstruction error is due to larger $L2$ penalty parameters used to suppress the late reflections.

In order to evaluate the suppression for late reflections in the estimated RIR $\boldsymbol{h}$, we carried out another experiment on the same reverberant speech signals as used in the last experiment, where different penalty parameters were tested. Figure 2 shows the variation of the reverberation indices of the estimated RIRs with the change in λ. We can see from the figure that the reverberation indices

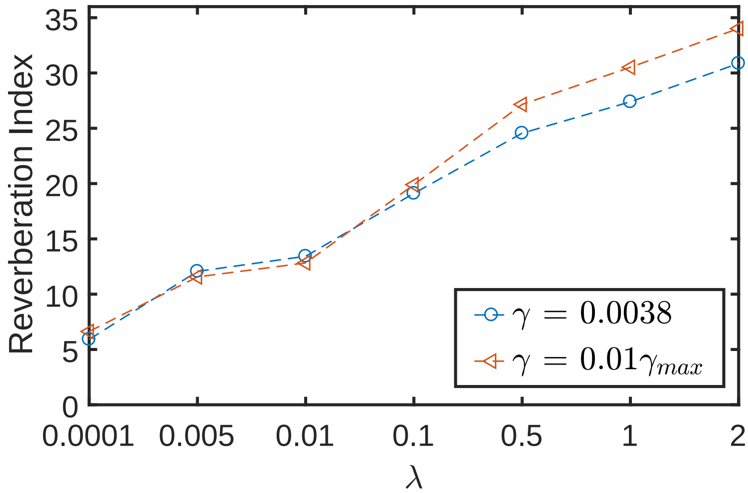

Fig. 2: The reverberation index variation for different penalty parameters, whereas the reverberation index of the original RIR is $RR_{htrue} = 6.31$.

are increasing when λ is increased, and all the reverberation indices of estimated RIRs are larger than that of the original RIR ($RR_{true} = 6.31$), which means the late part of the estimated $\boldsymbol{h}$ can be better suppressed by increasing the $L2$ norm penalty parameter. Note that, when λ is larger than a certain value, such as 0.5, the late reflections will be smoothed as shown in Figure 1 (f).

5 Conclusions and future work

We proposed a joint $L1 - L2$ norm blind deconvolution method for reverberant speech deconvolution, both the sparsity and density of RIR were taken into account. The experiments shown the proposed method not only can reconstruct the source speech and early part of RIR, but also can suppress the late part of the estimated RIR. Future work will evaluate the method with noise interference, and also automate the selection of the penalty parameters based on the reverberation level.

6 Acknowledgements

The work was 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. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning 3(1), 1–122 (2011)
3. Choudhary, S., Mitra, U.: Fundamental limits of blind deconvolution part I: Ambiguity kernel. arXiv preprint arXiv:1411.3810 (2014)
4. Choudhary, S., Mitra, U.: Fundamental limits of blind deconvolution part II: Sparsity-ambiguity trade-offs. arXiv preprint arXiv:1503.03184 (2015)
5. Chouzenoux, E., Pesquet, J.C., Repetti, A.: A block coordinate variable metric forward–backward algorithm. Journal of Global Optimization 66(3), 457–485 (2016)
6. Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Fixed-point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer (2011)
7. 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
8. 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)
9. Hu, Y., Kokkinakis, K.: Effects of early and late reflections on intelligibility of reverberated speech by cochlear implant listeners. The Journal of the Acoustical Society of America 135(1), EL22–EL28 (2014)
10. Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: An interior-point method for large-scale l1-regularized least squares. IEEE Journal of Selected Topics in Signal Processing 1(4), 606–617 (2007)
11. Klatte, M., Lachmann, T., Meis, M., et al.: Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. Noise and Health 12(49), 270 (2010)
12. Mosayyebpour, S., Esmaeili, M., Gulliver, T.A.: Single-microphone early and late reverberation suppression in noisy speech. IEEE Transactions on Audio, Speech, and Language Processing 21(2), 322–335 (2013)
13. Nakatani, T., Miyoshi, M., Kinoshita, K.: One microphone blind dereverberation based on quasi-periodicity of speech signals. In: NIPS. pp. 1417–1424 (2003)
14. Parikh, N., Boyd, S.: Proximal algorithms. Foundations and Trends in Optimization 1(3), 127–239 (2014)
15. Schuller, B.: Affective speaker state analysis in the presence of reverberation. International Journal of Speech Technology 14(2), 77–87 (2011)