

Sparse Blind Speech Deconvolution with Dynamic Range Regularization and Indicator Function

Jian Guan · Xuan Wang ·
Wenwu Wang · Lei Huang

Received: date / Accepted: date

Abstract Blind deconvolution is an ill-posed problem. To solve such a problem, prior information, such as, the sparseness of the source (i.e. input) signal or channel impulse responses, is usually adopted. In speech deconvolution, the source signal is not naturally sparse. However, the direct impulse and early reflections of the impulse responses of an acoustic system can be considered as sparse. In this paper, we exploit the channel sparsity and present an algorithm for speech deconvolution, where the dynamic range of the convolutive speech is also used as the prior information. In this algorithm, the estimation of the impulse response and the source signal is achieved by alternating between two steps, namely, the ℓ_1 regularized least squares optimization and a proximal operation. As demonstrated in our experiments, the proposed method provides superior performance for deconvolution of a sparse acoustic system, as compared with two state-of-the-art methods.

Keywords Sparse blind deconvolution · ℓ_1 regularized least squares · Speech deconvolution · Sparse channel estimation

Jian Guan · Xuan Wang (✉)

Computer Application Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, 518055, China
E-mail: {j.guan,wangxuan}@cs.hitsz.edu.cn

Wenwu Wang

Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK
E-mail: w.wang@surrey.ac.uk

Lei Huang

College of Information Engineering, Shenzhen University, Shenzhen, 518060, China
E-mail: dr.lei.huang@ieee.org

1 Introduction

Deconvolution aims to estimate unknown source signal $\mathbf{x} \in \mathbf{R}^N$ from observation $\mathbf{y} \in \mathbf{R}^M$, assuming the system impulse response $\mathbf{h} \in \mathbf{R}^L$ is known, and

$$\mathbf{y} = \mathbf{x} * \mathbf{h} + \mathbf{w}, \quad (1)$$

where $*$ denotes the convolution operation, \mathbf{w} is the system noise, and $M = N + L - 1$.

The deconvolution problem becomes blind, when both signal \mathbf{x} and impulse response \mathbf{h} are unknown, and only the observation \mathbf{y} is given. In this case, both \mathbf{x} and \mathbf{h} need to be estimated. The blind deconvolution problem is ill-posed, which is an underdetermined problem [5], as there are many possible combinations of estimates $(\hat{\mathbf{x}}, \hat{\mathbf{h}})$ that satisfy model (1). This is similar to a related problem, i.e. convolutive blind source separation [3, 41, 42, 44], where multiple sources need to be estimated from the observed mixtures [31] [32]. The applications of blind deconvolution include communications [20], image deblurring [2, 25, 26, 45] or recovery [6], spectrum restoration [27], speech modeling [1], geophysics [22], and astronomy [8].

In order to reduce the possible solution space, additional prior knowledge (such as the sparsity of the sources) can be exploited. Recently, sparse blind deconvolution (SBD) has been widely studied, where the signal \mathbf{x} is assumed to be sparse or approximately sparse. Usually, alternating minimization algorithms are used to estimate signal \mathbf{x} and impulse response \mathbf{h} by minimizing the cost function, which often comprises a data fidelity term and a regularization term (penalty term). The regularization term is adopted to exploit the prior information, such as the sparsity of the sources [36, 37, 40], as considered in [36] for seismic signals, in [10] [21] for spike signals, and in [22] for images. In blind source separation, however, apart from the sparsity that is often assumed for the underdetermined case [28] [43], statistical independence between the sources is also widely exploited for estimating the sources and the mixing channels [9, 18, 19, 29, 30, 34, 39].

In speech processing, considering a single-input and single-output (SISO) acoustic system, the observation \mathbf{y} of model (1), which can be seen as a reverberant speech signal, is a convolution of a source signal \mathbf{x} and room impulse response (RIR) \mathbf{h} . Although speech sources tend to follow super-Gaussian (Laplacian) distributions, they are not naturally sparse, unless they are converted into a transform domain using a pre-defined or learned dictionary. The RIR usually includes three parts: direct sound, early reflections, and late reverberation. In a room environment with a relatively low level of reverberation, the late reverberation becomes negligible which renders \mathbf{h} to be relatively sparse. As a result, by considering the sparsity of the acoustic system, SBD method can be used to estimate both the speech signal and RIR.

Exploiting only the sparsity prior, however, is insufficient to address the ill-posed blind deconvolution problem, as there are no strong theoretical guarantees on the identifiability of the sources and RIR unless some additional

assumptions are enforced [11, 12, 25]. To this end, we consider the dynamic range of the source signal as additional prior information to help reduce the solution space. In practice, however, the source signal is unknown. To address this issue, we use the dynamic range of the observation as an approximation, which is then incorporated into an indicator function and used as a regularization term in our optimization framework.

In this paper, we introduce a sparse blind deconvolution model, where not only the sparsity of the acoustic system is considered, but also the dynamic range of the reverberant speech is taken into account to reduce the possible solution space for the estimation of \mathbf{h} and \mathbf{x} . As a result, the proposed model is the sum of a squared ℓ_2 norm data fidelity term and two regularization terms: ℓ_1 norm constraint term and an indicator function. The ℓ_1 norm regularization term accounts for the sparsity of RIR and the indicator function accounts for the dynamic range of the signal. An alternating minimization algorithm is then proposed to solve this SBD problem in two steps: namely, sparse RIR estimation and signal estimation. These two steps can be considered as a combination of an ℓ_1 regularized least squares problem with a proximal operation. In our proposed iterative minimization algorithm, the ℓ_1 regularized least squares (*l1_ls*) algorithm [24] and the variable metric forward-backward algorithm [13–15, 35, 36] are employed to estimate \mathbf{h} and \mathbf{x} , respectively.

The rest of the paper is organized as follows: Section II presents the model formulation and details of our proposed algorithm for estimating the RIR and the source signal. Section III shows the experiments and results. Section IV concludes the paper and discusses the potential future work.

2 Proposed Sparse Blind Deconvolution Method

2.1 Proposed Model

In matrix form, (1) can be represented as

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{w} = \mathbf{H}\mathbf{x} + \mathbf{w}, \quad (2)$$

where $\mathbf{X} \in \mathbf{R}^{M \times L}$ and $\mathbf{H} \in \mathbf{R}^{M \times N}$ are the convolution Toeplitz matrices constructed from \mathbf{x} and \mathbf{h} , respectively. For example, the first row and first column of \mathbf{X} are $[x_1, 0, \dots, 0]$ and $[x_1, \dots, x_N, 0, \dots, 0]^T$, respectively.

By exploiting the prior information of both the source signal and RIR, we propose the following blind deconvolution model for the sparse acoustic system

$$F(\mathbf{x}, \mathbf{h}) = f(\mathbf{x}, \mathbf{h}) + g(\mathbf{h}) + r(\mathbf{x}), \quad (3)$$

where $f(\mathbf{x}, \mathbf{h}) = \|\mathbf{x} * \mathbf{h} - \mathbf{y}\|_2^2$ is a squared ℓ_2 norm data fidelity term which accounts for model (1), $g(\mathbf{h}) = \lambda \|\mathbf{h}\|_1$ is an ℓ_1 regularizer which exploits the sparsity of \mathbf{h} , and λ is a regularization parameter. $r(\mathbf{x})$ is the regularization term, which considers additional prior information of the source signal \mathbf{x} . Here, $r(\mathbf{x})$ is an indicator function, where the dynamic range of the observation \mathbf{y} is used to reduce the possible solution space for the estimation of the source signal \mathbf{x} .

2.2 Proposed Optimization Methods

We use an alternating minimization strategy to optimize the cost function (3), where the sparse RIR \mathbf{h} and source signal \mathbf{x} are updated in an alternating manner, by fixing one, and updating the other.

2.2.1 Sparse RIR \mathbf{h} Estimation

In this step, \mathbf{h} is estimated by fixing the source signal \mathbf{x} . In this case, the minimization of the cost function (3) is reduced to the following form

$$\begin{aligned}\mathbf{h}^{(k+1)} &= \underset{\mathbf{h}}{\operatorname{argmin}} f(\mathbf{x}^{(k)}, \mathbf{h}) + g(\mathbf{h}) \\ &= \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{X}^{(k)}\mathbf{h} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{h}\|_1,\end{aligned}\quad (4)$$

where $\mathbf{x}^{(k)}$ is the estimated source signal \mathbf{x} at the k^{th} step, $\mathbf{X}^{(k)}$ is the linear convolution matrix constructed from $\mathbf{x}^{(k)}$, and $\mathbf{h}^{(k+1)}$ is the estimated \mathbf{h} at the $(k+1)^{\text{th}}$ step.

The optimization of model (4) is a LASSO problem [38], which can be solved by the algorithms, such as *l1Ls* [24], alternating direction method of multipliers (ADMM) [7], proximal gradient method [33], and many others from the CVX toolbox [17].

Here we adapt the interior-point method *l1Ls* as the solver to estimate \mathbf{h} , as discussed next. First, as the cost function (4) is not differentiable, it is transformed to a convex quadratic problem with a linear inequality constraint, as follows

$$\begin{aligned}\mathbf{h}^{(k+1)} &= \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{X}^{(k)}\mathbf{h} - \mathbf{y}\|_2^2 + \lambda \sum_{i=1}^L u_i, \\ \text{subject to } & -u_i \leq h_i \leq u_i, i = 1, \dots, L\end{aligned}\quad (5)$$

where $\mathbf{h} \in \mathbf{R}^L$ and $\mathbf{u} \in \mathbf{R}^L$ are the variables to be found. Then, a logarithmic barrier function $\Phi(\mathbf{h}, \mathbf{u})$ is introduced for the bound constraints $-u_i \leq h_i \leq u_i$ in the following equation

$$\Phi(\mathbf{h}, \mathbf{u}) = -\sum_{i=1}^L \log(u_i + h_i) - \sum_{i=1}^L \log(u_i - h_i), \quad (6)$$

where $\Phi(\mathbf{h}, \mathbf{u})$ is defined over domain $\Phi = \{(\mathbf{h}, \mathbf{u}) \in \mathbf{R}^L \times \mathbf{R}^L, \text{ and } |h_i| < u_i, i = 1, \dots, L\}$. By applying this logarithmic barrier, the optimization problem with the inequality constraint in (5) can be further transformed to an equivalent convex function

$$\phi_t(\mathbf{h}, \mathbf{u}) = t\|\mathbf{X}^{(k)}\mathbf{h} - \mathbf{y}\|_2^2 + t\sum_{i=1}^L \lambda u_i + \Phi(\mathbf{h}, \mathbf{u}), \quad (7)$$

where $t > 0$ is a parameter controlling the approximation accuracy.

Finally, a truncated Newton interior-point method [24] can then be used to minimize ϕ_t , which allows us to update \mathbf{h} and \mathbf{u} as follows

$$(\mathbf{h}^{(k,j+1)}, \mathbf{u}^{(k,j+1)}) = (\mathbf{h}^{(k,j)}, \mathbf{u}^{(k,j)}) + \mu(\Delta\mathbf{h}, \Delta\mathbf{u}), \quad (8)$$

where (k, j) are the indices corresponding to the j^{th} inner loop of the solver $l1\text{-}ls$ in the k^{th} iteration, μ is a step size, and $(\Delta\mathbf{h}, \Delta\mathbf{u})$ represents the Newton step, which is the solution of the following Newton system

$$\mathcal{H}[\Delta\mathbf{h}, \Delta\mathbf{u}]^T = -\mathbf{g}, \quad (9)$$

where $\mathcal{H} \in \mathbf{R}^{2L \times 2L}$ and $\mathbf{g} \in \mathbf{R}^{2L}$ are the Hessian matrix and gradient vector at the current iteration, respectively. Note that, \mathcal{H} is symmetric and positive definite, which can be expressed as

$$\begin{aligned} \mathcal{H} &= t\nabla^2 \|\mathbf{X}^{(k)}\mathbf{h} - \mathbf{y}\|_2^2 + \nabla^2 \Phi_t(\mathbf{h}, \mathbf{u}) \\ &= \begin{bmatrix} 2t\mathbf{X}^{(k)T}\mathbf{X}^{(k)} + \mathbf{D}_1 & \mathbf{D}_2 \\ \mathbf{D}_2 & \mathbf{D}_1 \end{bmatrix}, \end{aligned} \quad (10)$$

where ∇ denotes the differential operator, and

$$\mathbf{D}_1 = \text{diag}\left(\frac{2(u_1^2 + h_1^2)}{(u_1^2 - h_1^2)^2}, \dots, \frac{2(u_L^2 + h_L^2)}{(u_L^2 - h_L^2)^2}\right) \in \mathbf{R}^L \quad (11)$$

$$\mathbf{D}_2 = \text{diag}\left(\frac{-4u_1h_1}{(u_1^2 - h_1^2)^2}, \dots, \frac{-4u_Lh_L}{(u_L^2 - h_L^2)^2}\right) \in \mathbf{R}^L. \quad (12)$$

Here, $\text{diag}(a_1, \dots, a_n)$ represents a diagonal matrix whose diagonal elements are a_1, \dots, a_n . The gradient vector \mathbf{g} is formed as

$$\mathbf{g} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}, \quad (13)$$

where

$$\begin{aligned} \mathbf{g}_1 &= \nabla_{\mathbf{h}} \phi_t(\mathbf{h}, \mathbf{u}) \\ &= 2t\mathbf{X}^{(k)T}(\mathbf{X}^{(k)}\mathbf{h} - \mathbf{y}) + \begin{bmatrix} 2h_1/(u_1^2 - h_1^2) \\ \vdots \\ 2h_L/(u_L^2 - h_L^2) \end{bmatrix} \in \mathbf{R}^L \end{aligned} \quad (14)$$

$$\mathbf{g}_2 = \nabla_{\mathbf{u}} \phi_t(\mathbf{h}, \mathbf{u}) = t\lambda\mathbf{1} - \begin{bmatrix} 2u_1/(u_1^2 - h_1^2) \\ \vdots \\ 2u_L/(u_L^2 - h_L^2) \end{bmatrix} \in \mathbf{R}^L. \quad (15)$$

In each iteration, the PCG algorithm [23] is employed to compute $(\Delta\mathbf{h}, \Delta\mathbf{u})$ based on \mathcal{H} and \mathbf{g} . Then, (8) can be applied to update (\mathbf{h}, \mathbf{u}) for the next iteration, until the optimal estimate of \mathbf{h} is obtained.

2.2.2 Source Signal \mathbf{x} Estimation

In this step, by fixing the sparse RIR \mathbf{h} , the minimization of (3) is reduced to

$$\mathbf{x}^{(k+1)} = \underset{\mathbf{x}}{\operatorname{argmin}} f(\mathbf{x}, \mathbf{h}^{(k+1)}) + r(\mathbf{x}), \quad (16)$$

where $f(\mathbf{x}, \mathbf{h}^{(k+1)}) = \|\mathbf{H}^{(k+1)}\mathbf{x} - \mathbf{y}\|_2^2$. Although the sparsity of RIR has been considered for the estimation of \mathbf{h} , it does not theoretically guarantee the identifiability of the source and RIR [11] [12]. To facilitate the estimation of \mathbf{x} , we incorporate the regularization term $r(\mathbf{x})$ which enables the additional information of the source \mathbf{x} to be exploited to further reduce the solution space for \mathbf{x} , such as the energy information considered in [5] [16].

Here, we consider the dynamic range of the signal as the additional constraint $r(\mathbf{x})$. Ideally, the dynamic range of the source \mathbf{x} should be used to limit the solution space of \mathbf{x} . In practice, however, the source signal \mathbf{x} is unknown. To address this issue, we employ the dynamic range and the energy of the observed signal \mathbf{y} to approximate the prior information for \mathbf{x} . This is based on the observation that \mathbf{h} is usually bounded and decaying to zero after a certain amount of time (e.g. hundreds of milliseconds, depending on the reverberation level), as a result, for a bounded input \mathbf{x} , the output (i.e. the observed signal) \mathbf{y} will be also bounded, subject to a scaling factor.

More specifically, we assume $r(\mathbf{x})$ to be an indicator function defined as $\mathcal{C} = \{\mathbf{x} \in [\mathbf{y}_{min}, \mathbf{y}_{max}]^N \text{ and } \|\mathbf{x}\|_2 \leq \delta\}$, where δ is a threshold used to bound the energy of the source signal, set as $\|\mathbf{y}\|_2$, \mathbf{y}_{min} and \mathbf{y}_{max} are the minimum and maximum values \mathbf{y} respectively, indicating the dynamic range of the signal. In principle, the accuracy of the dynamic range and energy information will affect the estimation result. This will be illustrated by an example in Section 3.4, where we show how $r(\mathbf{x})$ influences the deconvolution results.

A variable metric forward-backward algorithm [13–15, 35, 36] is employed to optimize model (16). For each iteration k , this algorithm consists of a partial gradient step (forward step) on $f(\cdot, \mathbf{h}^{(k)})$ followed by a proximal step (backward step) on r .

First, as in [14] [35], we define the proximity operator used in the proximal step as

$$\operatorname{prox}_{\mathbf{U}, \psi}(\mathbf{z}) = \underset{\mathbf{z}' \in \mathbf{R}^S}{\operatorname{argmin}} \psi(\mathbf{z}') + 1/2 \|\mathbf{z}' - \mathbf{z}\|_{\mathbf{U}}^2, \quad (17)$$

where $\operatorname{prox}_{\mathbf{U}, \psi}(\mathbf{z})$ is assumed to be the \mathbf{U} -weighted proximity operator of ψ at \mathbf{z} , and $\mathbf{U} \in \mathbf{R}^{S \times S}$ is a symmetric positive definite (SPD) matrix, where $\psi : \mathbf{R}^S \rightarrow (-\infty, +\infty]$ is a proper, lower semicontinuous, and convex function. For $\forall \mathbf{z} \in \mathbf{R}^S$, $\|\mathbf{z}\|_{\mathbf{U}} = \langle \mathbf{z}, \mathbf{U}\mathbf{z} \rangle^{1/2}$, where $\langle \cdot, \cdot \rangle$ denotes the scalar product. Note that, if ψ is an indicator function for a convex set C , i.e.,

$$\psi(\mathbf{z}) = \begin{cases} 0, & \mathbf{z} \in C \\ +\infty, & \text{otherwise} \end{cases}, \quad (18)$$

then $prox_{\mathbf{U}, \psi}(\mathbf{z})$ is the weighted Euclidean projection onto C , which means $prox_{\mathbf{U}, \psi}(\mathbf{z})$ is the closest point in the set C to \mathbf{z} in terms of the weighted Euclidean distance.

As a result, by optimizing (16) with the variable metric forward-backward method, \mathbf{x} can be updated as

$$\mathbf{x}^{(k,i+1)} = \underbrace{prox_{(\gamma^{(k,i)})^{-1} \mathbf{A}_{k,r}}}_{\text{backward step}} \underbrace{(\mathbf{x}^{(k,i)} - \gamma^{(k,i)} \mathbf{A}_k^{-1} \nabla f)}_{\text{forward step}}, \quad (19)$$

where (k, i) are the indices corresponding to the i^{th} proximal operation in the k^{th} iteration, $\gamma^{(k,i)}$ is the step size, and \mathbf{A}_k is used to induce the proximity operator, which can be designed as in [14] [35] or [36]. Here, \mathbf{A}_k is updated as

$$\mathbf{A}_k = \|(\mathbf{H}^{(k+1)})^T (\mathbf{H}^{(k+1)})\|_F \mathbf{I}_N, \quad (20)$$

and ∇f is the partial gradient of $f(\cdot, \mathbf{h}^{(k)})$ with respect to $\mathbf{x}^{(k,i)}$, which is defined as

$$\nabla f = (\mathbf{H}^{(k+1)})^T (\mathbf{H}^{(k+1)}) \mathbf{x}^{(k,i)} - (\mathbf{H}^{(k+1)})^T \mathbf{y}. \quad (21)$$

Denoting the optimal solution to (19) as $\mathbf{x}^{(k, I_k)}$, we can update the estimated \mathbf{x} at the $(k+1)^{\text{th}}$ step as $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k, I_k)}$. The proposed algorithm is summarized in Algorithm 1.

Algorithm 1

Input: observation \mathbf{y}

Initialization: $k = 0$, $\mathbf{x}^{(0)} = \mathbf{x}^0$, construct \mathbf{X}^0 from $\mathbf{x}^{(0)}$, $\mathbf{X}^{(k)} = \mathbf{X}^0$, $\epsilon = 10^{-6}$.

Iterations:

for $k = 0, 1, \dots$

Step 1: Sparse RIR \mathbf{h} Estimation:

 Update $\mathbf{h}^{(k+1)}$ by solving (4), using (8), (9), (10), and (13),

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

 Update \mathbf{A}_k by (20).

Step 2: Signal \mathbf{x} Estimation:

for $i = 0, \dots, I_k - 1$

$\tilde{\mathbf{x}}^{(k,i)} = \mathbf{x}^{(k,i)} - \gamma^{(k,i)} \mathbf{A}_k^{-1} \nabla f$,

$\mathbf{x}^{(k,i+1)} = prox_{(\gamma^{(k,i)})^{-1} \mathbf{A}_{k,r}}(\tilde{\mathbf{x}}^{(k,i)})$,

end for

 Update $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k, I_k)}$,

 Construct $\mathbf{X}^{(k+1)}$ from $\mathbf{x}^{(k+1)}$.

Stopping criterion: If $\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2 \leq \epsilon$, then $\hat{\mathbf{x}} = \mathbf{x}^{(k+1)}$, $\hat{\mathbf{h}} = \mathbf{h}^{(k+1)}$, and break.

end for

Output: estimated $\hat{\mathbf{x}}$ and $\hat{\mathbf{h}}$

3 Experiments and Results

3.1 Experimental Setup

We apply our method for speech dereverberation, where the reverberant speech \mathbf{y} was generated by the linear convolution of a sparse RIR \mathbf{h} and a source signal \mathbf{x} . The performance indices used in our experiments are the mean absolute error (MAE) and mean squared error (MSE), defined respectively as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\bar{e}_i - \hat{e}_i|, \quad (22)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\bar{e}_i - \hat{e}_i)^2, \quad (23)$$

which measure the ℓ_1 norm and ℓ_2 norm residual errors between the original signal $\bar{\mathbf{e}}$ and the estimated signal $\hat{\mathbf{e}}$ respectively (both normalized as $\mathbf{e} = \frac{\mathbf{e}}{\max(|\mathbf{e}|)}$), and N is the length of the signal. In practice, \mathbf{e} can be either \mathbf{x} or \mathbf{h} .

The RIRs were generated in a $4 \times 4 \times 3 \text{ m}^3$ room using the room image model [4], with a sampling frequency at 16 kHz. Then, we selected the early reflection, and removed the late part of \mathbf{h} . The length of \mathbf{h} is $L = 600$. An example of such an \mathbf{h} after normalization is given in Figure 1(b). The source signals \mathbf{x} were selected from the TIMIT database, and 9 different speech signals including 5 males and 4 females were used in our experiments. The sampling frequency was 16 kHz, and the length of the sources was $N = 500$. For each k , the maximum iteration of $l1_ls$ was set as 500, and $I_k = 100$. The stopping criterion we used in all tests is $\|\mathbf{h}^{(k+1)} - \mathbf{h}^{(k)}\|_2^2 \leq \epsilon$, where ϵ is set as 10^{-6} . Several experiments have been carried out. First, we evaluate the robustness of the proposed method to the variation in initialization, and also evaluate the performance of several alternative optimization methods for estimating \mathbf{h} . Then, we compare the proposed method with two state-of-the-art SBD methods, i.e. SOOT [36] and ALPA [1]. Moreover, we illustrate how the regularization term $r(\mathbf{x})$ influences the estimation of the source signal and RIR. Finally, we evaluate the performance of the proposed method for deconvolution in a relatively high level of reverberation.

3.2 Robustness to Initialization

The initial \mathbf{x}^0 was obtained by adding white Gaussian noise of zero mean and unit variance to the source signal \mathbf{x} . Different noise levels were tested in initializations. When a high level of noise is added to the source signal, the initialization tends to be similar to the initialization with random white Gaussian noise. On the other hand, when the noise level is very low, the blind deconvolution problem tends to become a deconvolution problem, which is

Table 1: Performance comparison in terms of MAE for four different algorithms used in our proposed alternating minimization algorithm to illustrate the influence of different initializations.

		MAE (dB)			
Noise level (dB)		-5	0	5	10
Signal x	Proposed	-8.89	-9.06	-9.08	-9.10
	ADMM	-8.58	-8.67	-8.76	-8.77
	Proximal	-8.44	-8.85	-8.87	-8.88
	CVX	-8.62	-8.77	-8.77	-8.77
RIR h	Proposed	-21.75	-21.91	-21.98	-22.04
	ADMM	-20.54	-20.61	-20.63	-20.69
	Proximal	-20.72	-20.87	-20.99	-21.03
	CVX	-21.42	-21.56	-21.56	-21.56
Observation y	Proposed	-11.53	-11.69	-11.72	-11.73
	ADMM	-11.09	-11.15	-11.27	-11.28
	Proximal	-10.92	-11.34	-11.37	-11.37
	CVX	-11.09	-11.24	-11.25	-11.25

Table 2: Performance comparison in terms of MSE for four different algorithms used in our proposed alternating minimization algorithm to illustrate the influence of different initializations.

		MSE (dB)			
Noise level (dB)		-5	0	5	10
Signal x	Proposed	-15.30	-15.60	-15.65	-15.67
	ADMM	-14.76	-14.87	-15.04	-15.06
	Proximal	-14.52	-15.21	-15.25	-15.26
	CVX	-14.79	-15.06	-15.06	-15.06
RIR h	Proposed	-27.50	-27.97	-28.14	-28.34
	ADMM	-24.97	-25.03	-25.06	-25.34
	Proximal	-25.23	-25.53	-25.75	-25.94
	CVX	-26.61	-27.04	-27.05	-27.05
Observation y	Proposed	-19.09	-19.44	-19.49	-19.51
	ADMM	-18.39	-18.48	-18.67	-18.67
	Proximal	-18.07	-18.79	-18.84	-18.84
	CVX	-18.32	-18.58	-18.58	-18.58

similar to the case where the source signal \mathbf{x} is given. More specifically, \mathbf{x}^0 was obtained by adding noise to \mathbf{x} at 4 different levels from -5 dB to 10 dB in a step of 5 dB. In order to test the robustness of the proposed method, 9 speech signals and 3 room impulse responses were used in our experiments, leading to 27 realizations for each initialization. The MAE and MSE were calculated as the performance indices.

Table 3: Performance comparison among the proposed method, the smooth approximate ℓ_1/ℓ_2 based SOOT method and the alternating ℓ_p - ℓ_2 projections based ALPA method.

		MSE (dB)			MAE (dB)		
Noise level (dB)		10	20	30	10	20	30
Signal x	Proposed	-11.66	-18.12	-18.81	-6.82	-10.16	-10.60
	SOOT	-5.66	-4.23	-4.99	-3.66	-3.09	-3.33
	ALPA	-10.54	-18.49	-20.31	-6.25	-10.32	-11.37
RIR h	Proposed	-26.68	-29.23	-29.44	-20.52	-22.16	-22.34
	SOOT	-19.93	-22.16	-22.23	-16.51	-17.11	-17.02
	ALPA	-25.09	-27.25	-29.60	-20.16	-21.27	-21.97
Observation y	Proposed	-24.96	-26.22	-25.96	-13.84	-14.83	-14.75
	SOOT	-4.85	-4.85	-5.11	-3.76	-3.72	-3.59
	ALPA	-21.97	-28.91	-31.61	-12.39	-15.94	-17.62

Meanwhile, as the proposed method used an alternating minimization strategy. For comparison purpose, in the step of \mathbf{h} estimation in Algorithm 1, we also tested the robustness of other optimization methods including ADMM [7], proximal gradient method [33], and the default method from the CVX toolbox [17] to evaluate the robustness of these compared methods in this experiment.

The results are shown in Table 1 and Table 2. From these two tables, it can be observed that the performance of our proposed method is better than those of the other three methods used in the alternating minimization. In addition, our proposed method is not so sensitive to the variations in initialization. Therefore, 0 dB noise is added to the initialization of \mathbf{x} in the subsequent experiments.

3.3 Comparison with Other SBD Algorithms

Our proposed method is also compared with another two SBD methods: the SOOT algorithm [36] and ALPA algorithm [1]. In [36], a smoothed ℓ_1/ℓ_2 constraint is imposed on the seismic signal, and the dynamic range information of both \mathbf{h} and \mathbf{x} are used as the prior information. To make a fair comparison, we only used the dynamic range of \mathbf{y} as the prior information in SOOT, as in our proposed method. The ALPA algorithm is an ℓ_p norm based alternating optimization algorithm, where the sparsity constraint based on the ℓ_p norm is imposed on the excitation for speech modeling. Here in our experiment, the sparsity constraint for both SOOT and ALPA was imposed on RIR. A source signal and a RIR were selected from the previous experiment as the ground truth of this experiment. Three different noise levels (i.e. 10, 20, and 30 dB) were tested. In both algorithms, the same set of parameters was used for different noise levels. For each noise level, 200 realizations were run.

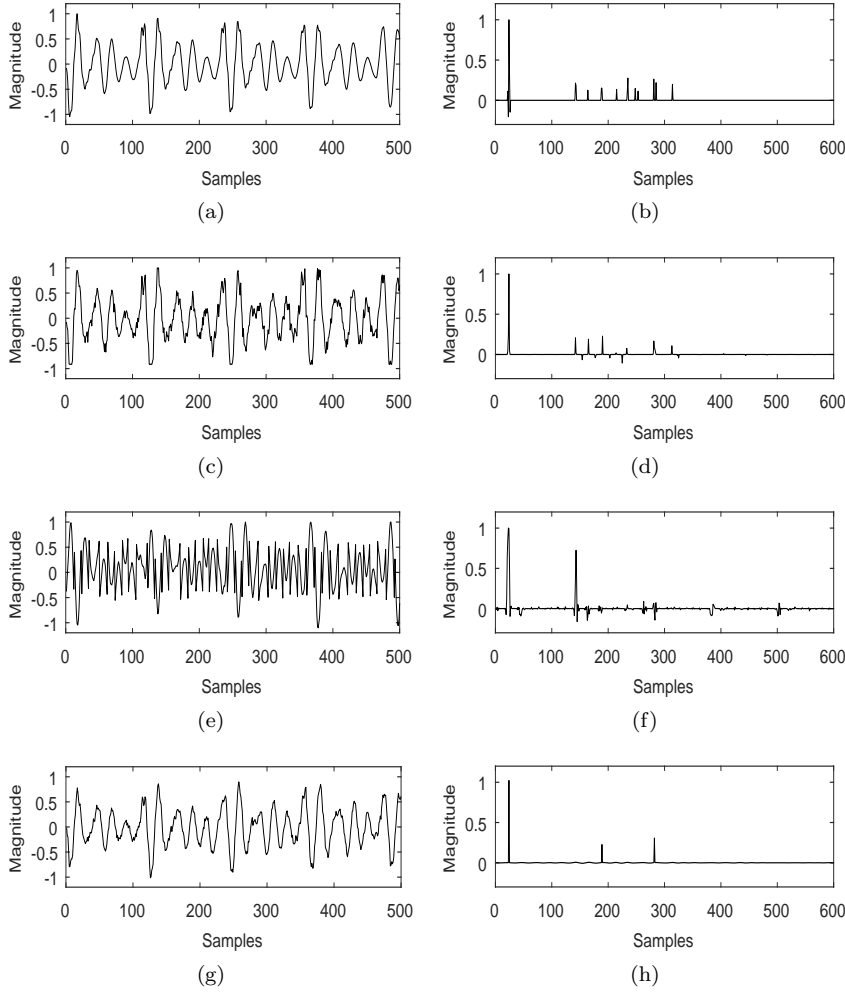


Fig. 1: An example of the signal and RIR estimated by the proposed method, SOOT, and ALPA at 20 dB noise level: the original source signal and RIR are in (a) and (b), and the source signal and RIR estimated by the proposed method in (c) and (d); by SOOT in (e) and (f); by ALPA in (g) and (h), respectively.

Table 3 shows the results in terms of MAE and MSE. It can be seen that the proposed method offers better performance as compared with the SOOT method. Note that, without the dynamic range as the prior information, the source signal and RIR estimated by SOOT would have scale ambiguity. The error of the RIR estimation is reduced due to the sparsity constraint imposed on the RIR. In addition, without the dynamic range constraint, the estimation

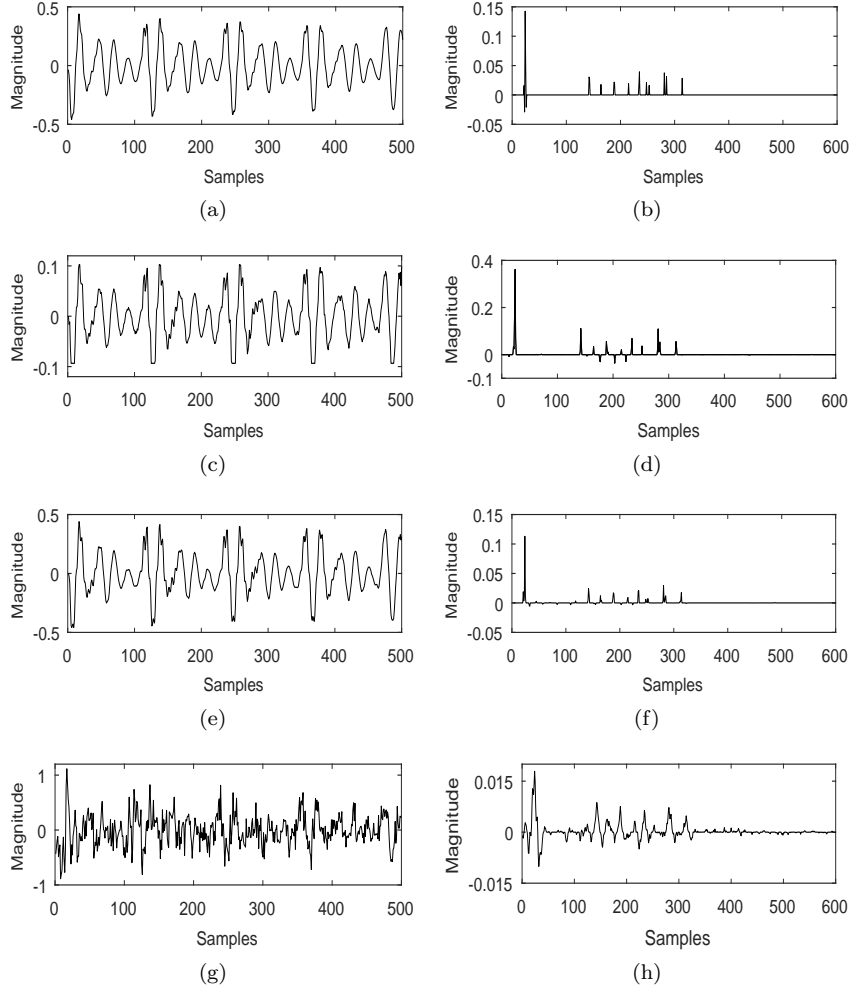


Fig. 2: Illustration of how $r(\mathbf{x})$ influences the estimation results of the source and RIR: the original source signal and RIR (unnormalized) are in (a) and (b), respectively; the estimated source and RIR in (c) and (d) where the observation \mathbf{y} is used to define $r(\mathbf{x})$; in (e) and (f) where the original source \mathbf{x} is used to define $r(\mathbf{x})$; in (g) and (h) where $r(\mathbf{x})$ is not applied.

of the source signal would be unstable and much worse as compared with that given by our proposed method. As compared with ALPA, our proposed method gives better performance for the RIR estimation at all noise levels, and a better performance in the estimation of the source signal with a relatively high level of noise. However, the ALPA algorithm tends to give slightly better results for a low level of noise (e.g. for SNR at 30 dB). An example is given in Figure 1

to show the source signal and RIR estimated by the proposed method, SOOT and ALPA, respectively. The examples were randomly selected from the 200 realizations at 20 dB noise level. We can see that the source signals estimated by both the proposed method and the ALPA algorithm are close to the original source signal. Our proposed method gives better estimate for RIR as compared with SOOT and ALPA. The results by SOOT without the dynamic range constraint are not comparable with those by the proposed method and ALPA.

3.4 The Impact of Regularization $r(\mathbf{x})$

In our proposed method, the dynamic range and energy of \mathbf{y} are used as the prior information to define the regularization term $r(\mathbf{x})$. Here, an example is given to illustrate how $r(\mathbf{x})$ influences the estimation results. Experiments were carried out under three different situations in which $r(\mathbf{x})$ was defined by the observation \mathbf{y} as in our proposed method, by the ground truth signal \mathbf{x} , or not used as a regularization. The RIR used for generating the observations is the same as the one used in the previous experiment.

From Figure 2, we can see that the source signal and RIR can be recovered very well with the use of $r(\mathbf{x})$. Especially, the scale ambiguity can be mitigated when $r(\mathbf{x})$ was defined by using the ground truth source signal, and this can be observed by comparing the subplots (e) and (f) with (a) and (b). When $r(\mathbf{x})$ was defined by using the observation \mathbf{y} , the estimated source signal and RIR resemble the ground truth source signal and RIR very well even though there are some scale ambiguities associated with the estimated source signal and RIR, which can be observed by comparing the subplots (c) and (d) with (a) and (b). When the regularization term $r(\mathbf{x})$ was not applied, the obtained estimates (as shown in the subplots (g) and (h)) appear to be much more noisy as compared with the ground truth and those obtained by incorporating $r(\mathbf{x})$. This is mainly because the optimization may be trapped in an undesired local optimum, where the source signal and RIR cannot be identified. In the ALPA algorithm [1], a normalization strategy was used to address this problem, where \mathbf{h} was normalized during the alternating optimization in each iteration. In comparison, with our proposed regularization $r(\mathbf{x})$, the prior information about the source \mathbf{x} was exploited and thus, providing more accurate estimation result. In addition, the scale ambiguity issue can be better mitigated using the function $r(\mathbf{x})$.

3.5 Deconvolution Performance in a High Level of Reverberation

We also give an example to show how the proposed method performs when the reverberation level is relatively high, e.g. the RIR \mathbf{h} is generated with a reverberation time 900 ms. Figure 3 (b) shows the ground truth of a shortened version of the RIR with 480 samples from the early part and 520 samples from the late part. The observation \mathbf{y} was added with 20 dB white Gaussian noise.

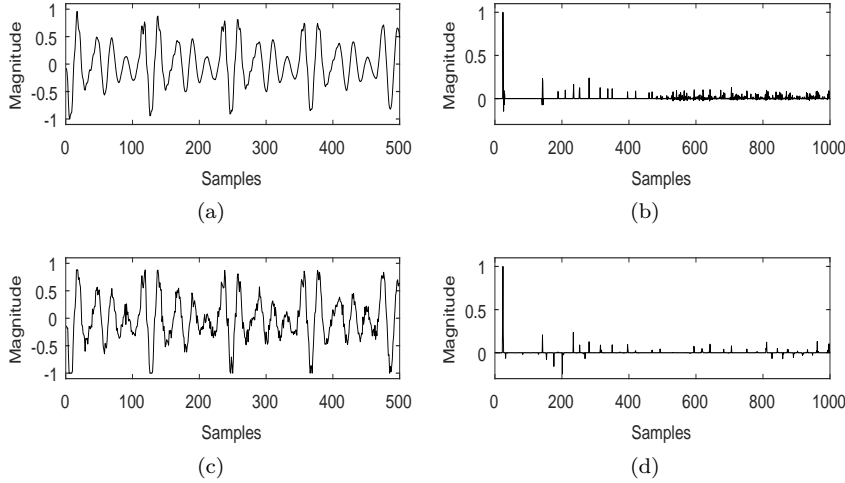


Fig. 3: Estimation results of the source and RIR for a high level of reverberation (i.e. reverberation time at 900ms) with 20 dB noise: (a) and (b) are the original source signal and RIR, respectively; (c) and (d) are the estimated source signal and RIR, respectively.

Figure 3 shows the comparison between the ground truth and the estimation result for the source signal and the RIR.

We can see from Figure 3 that the estimated source is similar to the original source, however, the RIR tends to be more sparse as compared with the original RIR. This is not surprising considering the fact that the sparsity constraint has been applied to the RIR in the proposed cost function, and as a result, a sparse RIR has been obtained from the noisy observations. To fully recover the RIR including the late part, a joint sparsity and dense model could be used where the sparsity and density can be applied for the early and late part of the RIR respectively. This is however beyond the scope of current work.

4 Conclusion

We have proposed a sparse blind deconvolution method by taking into account the sparsity of RIR and the dynamic range of the source signal in the cost function, which includes an ℓ_2 norm least squares data fidelity term, an ℓ_1 norm regularization term and an indicator function. An alternating minimization strategy is employed to optimize the cost function, where the l_1 s and proximal methods are used to solve the two sub-optimization problems. Experimental results show that our proposed method provides better estimates for both the source signal and the impulse responses of the sparse acoustic system compared with two baselines. By considering the dynamic range of the signal,

more accurate estimation of the source signal and the RIR can be obtained, where the scale ambiguity issue can also be mitigated.

The proposed method could be extended to include late reverberation by using a joint sparsity and density model, such as a mixed ℓ_1 and ℓ_2 norm to exploit the sparsity and density of RIR respectively, which is an interesting direction for future work.

Acknowledgements The work was conducted when J. Guan was visiting the University of Surrey, and supported in part by International Exchange and Cooperation Foundation of Shenzhen City, China (No. GJHZ20150312114149569). W. Wang was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) Grant Number EP/K014307 and the MOD University Defence Research Collaboration in Signal Processing. The authors thank the anonymous reviewers for their helpful suggestions, and Dr Mark Barnard for proofreading the revised manuscript.

References

1. Adiga, A., Seelamantula, C.S.: An alternating $\ell_p - \ell_2$ projections algorithm (ALPA) for speech modeling using sparsity constraints. In: Proc. of IEEE International Conference on Digital Signal Processing (DSP), pp. 291–296 (2014)
2. Ahmed, A., Recht, B., Romberg, J.: Blind deconvolution using convex programming. *IEEE Transactions on Information Theory* **60**(3), 1711–1732 (2014)
3. Alinaghi, A., Jackson, P.J., Liu, Q., Wang, W.: Joint mixing vector and binaural model based stereo source separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**(9), 1434–1448 (2014)
4. 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)
5. Benichoux, A., Vincent, E., Gribonval, R.: A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors. In: Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 26–31 (2013)
6. Bolte, J., Combettes, P.L., Pesquet, J.C.: Alternating proximal algorithm for blind image recovery. In: Proc. of IEEE International Conference on Image Processing (ICIP), pp. 1673–1676 (2010)
7. 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)
8. Campisi, P., Egiazarian, K.: *Blind Image Deconvolution: Theory and Applications*. CRC press (2007)
9. Chai, R., Naik, G., Nguyen, T.N., Ling, S., Tran, Y., Craig, A., Nguyen, H.: Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system. *Journal of Biomedical and Health Informatics* (2016). DOI 10.1109/JBHI.2016.2532354
10. Chi, Y.: Guaranteed blind sparse spikes deconvolution via lifting and convex optimization. *IEEE Journal of Selected Topics in Signal Processing* **10**(4), 782–794 (2016)
11. Choudhary, S., Mitra, U.: Fundamental limits of blind deconvolution part I: Ambiguity kernel. *arXiv preprint arXiv:1411.3810* (2014)
12. Choudhary, S., Mitra, U.: Fundamental limits of blind deconvolution part II: Sparsity-ambiguity trade-offs. *arXiv preprint arXiv:1503.03184* (2015)
13. Chouzenoux, E., Pesquet, J.C., Repetti, A.: Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. *Journal of Optimization Theory and Applications* **162**(1), 107–132 (2014)
14. Chouzenoux, E., Pesquet, J.C., Repetti, A.: A block coordinate variable metric forward-backward algorithm. *Journal of Global Optimization* **66**(3), 457–485 (2016)
15. Combettes, P.L., Pesquet, J.C.: Proximal Splitting Methods in Signal Processing pp. 1–25 (2009). URL <http://arxiv.org/abs/0912.3522>

16. Donoho, D.L.: On minimum entropy deconvolution. *Applied Time Series Analysis II* pp. 565–609 (1981)
17. 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)
18. Guo, Y., Huang, S., Li, Y., Naik, G.R.: Edge effect elimination in single-mixture blind source separation. *Circuits, Systems, and Signal Processing* **32**(5), 2317–2334 (2013)
19. Guo, Y., Naik, G.R., Nguyen, H.: Single channel blind source separation based local mean decomposition for biomedical applications. In: *Proc. of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6812–6815 (2013)
20. Haykin, S.: *Blind Deconvolution*. Prentice Hall (1994)
21. Kaarensen, K.F.: Deconvolution of sparse spike trains by iterated window maximization. *IEEE Transactions on Signal Processing* **45**(5), 1173–1183 (1997)
22. Kaarensen, K.F., Tøft, T.: Multichannel blind deconvolution of seismic signals. *Geophysics* **63**(6), 2093–2107 (1998)
23. Kelley, C.: Iterative methods for linear and nonlinear equations. *SIAM Frontiers in Applied Mathematics* **16** (1995)
24. Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: An interior-point method for large-scale ℓ_1 -regularized least squares. *IEEE Journal of Selected Topics in Signal Processing* **1**(4), 606–617 (2007)
25. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 233–240 (2011)
26. Li, X., Ling, S., Strohmer, T., Wei, K.: Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *arXiv preprint arXiv:1606.04933* (2016)
27. 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)
28. Luo, Y., Wang, W., Chambers, J.A., Lambetharan, S., Proudler, I.: Exploitation of source nonstationarity in underdetermined blind source separation with advanced clustering techniques. *IEEE Transactions on Signal Processing* **54**(6), 2198–2212 (2006)
29. Naik, G., Al-Timemy, A., Nguyen, H.: Transradial amputee gesture classification using an optimal number of sEMG sensors: an approach using ICA clustering. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **24**(8), 837–846 (2016)
30. Naik, G., Selvan, S., Nguyen, H.: Single-channel EMG classification with ensemble-Empirical-Mode-Decomposition-Based ICA for diagnosing neuromuscular disorders. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **24**(7), 734–743 (2016)
31. Naik, G.R.: Enhancement of the ill-conditioned original recordings using novel ICA technique. *International Journal of Electronics* **99**(7), 899–906 (2012)
32. Naik, G.R., Kumar, D.K.: Estimation of independent and dependent components of non-invasive EMG using fast ICA: validation in recognising complex gestures. *Computer Methods in Biomechanics and Biomedical Engineering* **14**(12), 1105–1111 (2011)
33. Parikh, N., Boyd, S.: Proximal algorithms. *Foundations and Trends in Optimization* **1**(3), 127–239 (2014)
34. Pendharkar, G., Naik, G.R., Nguyen, H.T.: Using blind source separation on accelerometry data to analyze and distinguish the toe walking gait from normal gait in ITW children. *Biomedical Signal Processing and Control* **13**, 41–49 (2014)
35. Repetti, A., Chouzenoux, E., Pesquet, J.C.: A preconditioned forward-backward approach with application to large-scale nonconvex spectral unmixing problems. In: *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1498–1502 (2014)
36. 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)
37. Selesnick, I.: Sparse deconvolution (an MM algorithm). [Online]. Available: <http://cnx.org/content/m44991/> (2012)

-
38. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288 (1996)
 39. Vincent, E., Gribonval, R., Févotte, C.: Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing* **14**(4), 1462–1469 (2006)
 40. Wang, L., Chi, Y.: Blind deconvolution from multiple sparse inputs. *IEEE Signal Processing Letters* **23**(10), 1384–1388 (2016)
 41. Wang, W., Jafari, M.G., Sanei, S., Chambers, J.A.: Blind separation of convolutive mixtures of cyclostationary signals. *International Journal of Adaptive Control and Signal Processing* **18**(3), 279–298 (2004)
 42. Wang, W., Sanei, S., Chambers, J.A.: Penalty function-based joint diagonalization approach for convolutive blind separation of nonstationary sources. *IEEE Transactions on Signal Processing* **53**(5), 1654–1669 (2005)
 43. Xu, T., Wang, W., Dai, W.: Sparse coding with adaptive dictionary learning for under-determined blind speech separation. *Speech Communication* **55**(3), 432–450 (2013)
 44. Yu, Y., Wang, W., Han, P.: Localization based stereo speech source separation using probabilistic time-frequency masking and deep neural networks. *EURASIP Journal on Audio, Speech, and Music Processing* **2016**(1), 1–18 (2016)
 45. Zhang, H., Wipf, D., Zhang, Y.: Multi-observation blind deconvolution with an adaptive sparse prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(8), 1628–1643 (2014)