Blind System Identification Using Sparse Learning for TDOA Estimation of Room Reflections

Konrad Kowalczyk, Emanuël A. P. Habets, *Senior Member, IEEE*, Walter Kellermann, *Fellow, IEEE*, and Patrick A. Naylor, *Senior Member, IEEE*

***Abstract*—Localization of early room reflections can be achieved by estimating the time-differences-of-arrival (TDOAs) of reflected waves between elements of a microphone array. For an unknown source, we propose to apply sparse blind system identification (BSI) methods to identify the acoustic impulse responses, from which the TDOAs of temporally sparse reflections are estimated. The proposed time- and frequency-domain adaptive algorithms based on crossrelation formulation are regularized by incorporating an l_1-norm sparseness constraint, which is realized using a split Bregman method. These algorithms are shown to outperform standard crossrelation-based BSI techniques when estimating TDOAs of reflections in the presence of background noise.**

***Index Terms*—Blind system identification, Bregman method, crossrelation error, sparse learning, time delay estimation.**

I. Introduction

INFORMATION about the directions of arrival of the source signal and room reflections can be applied to improve the performance of several audio applications, such as signal enhancement [1] and room-compensated audio reproduction [2]. Typically the time-difference-of-arrival (TDOA) of the direct path is estimated with the aim to localize a speaker or an interfering sound source, see [3] for an overview of available techniques. Recently reflection localization has gained more interest in the context of inferring room geometry [4] and room-aware sound reproduction [2].

Localization of reflections can be performed using the TDOAs of the reflected waves between different microphone signals, which can be conveniently computed from (blindly) estimated acoustic impulse responses (AIRs). In particular, the crossrelation-based single-input multiple-output (SIMO) blind system identification (BSI), which exploits spatial differences between multiple microphone signals, has received significant attention as it does not require any assumption about the source signal and can be formulated as an adaptive algorithm [5], [6]. Furthermore, adaptive BSI algorithms have been shown in [3] to outperform crosscorrelation-based techniques, such as generalized crosscorrelation (GCC), for source localization in noisy and reverberant environments. As the power of the direct-path signal is typically higher than the power of the early reflections, the estimation of the latter becomes more challenging, especially in the presence of additional noise.

In this letter, we aim to increase the robustness to noise of BSI methods for TDOA estimation of room reflections by incorporating a sparse AIR model. The l_1-regularized crossrelation formulation is obtained by means of a split Bregman iteration method [7], [8], which is known to be robust to background noise. The presented sparse BSI problem formulation is advantageous since the early part of acoustic impulse responses has a sparse structure [9] so that for TDOA estimation of most significant reflections only the knowledge of sparse AIR components is needed. Sparse BSI has also been applied to speech extraction [10] and dereverberation [11]. This letter presents a method to formulate adaptive BSI algorithms with sparse learning that operate in time and frequency domains. In particular, two algorithms are derived by extending the multichannel least mean squares (MCLMS) [5] and the normalized multichannel frequency-domain LMS (NMCFLMS) [6] algorithms to sparse BSI. Other algorithms can be derived analogously.

II. Problem Formulation

In the assumed reverberation model, the pth microphone signal $x_p(n)$ of a P-element microphone array is given by

$$x_p(n) = \sum_{l=0}^{L-1} h_{p,l}\, s(n-l) + \nu_p(n), \tag{1}$$

where $s(n)$ is the original unknown source signal, $\nu_p(n)$ denotes the spatially white background noise, and $\mathbf{h}_p = [h_{p,0}, h_{p,1}, \ldots, h_{p,L-1}]^T$ denotes the pth impulse response of length L, which can be linearly decomposed into

$$\mathbf{h}_p = \mathbf{h}_p^{\mathrm{dp}} + \sum_{r=1}^{Q} \mathbf{h}_p^{r} + \mathbf{h}_p^{\mathrm{rev}}, \tag{2}$$

where $\mathbf{h}_p^{\mathrm{dp}}$ and $\mathbf{h}_p^{r}$ model the fractional propagation time delays $\tau_{r,p}$ and attenuation factors of the direct-path signal $r=0$ and early room reflections $r = 1, \ldots, Q$, respectively, and $\mathbf{h}_p^{\mathrm{rev}}$ denotes the filter that models late reverberation. In this work, we aim to estimate the TDOA of the rth room reflection between the pth and qth microphone. Such a TDOA is defined as

$$\hat{\psi}_{r,pq} = |\hat{\tau}_{r,p} - \hat{\tau}_{r,q}|, \tag{3}$$

where $\hat{\tau}_{r,p}$ and $\hat{\tau}_{r,q}$ denote the estimated time delays of reflection r. Note that the task of matching the peaks between different AIRs corresponding to the same reflection [4] is beyond the scope of this letter.

Manuscript received March 25, 2013; revised April 29, 2013; accepted April 29, 2013. Date of publication May 01, 2013; date of current version May 17, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Augusto Sarti.

K. Kowalczyk and E. A. P. Habets are with the International Audio Laboratories Erlangen (a joint institution between the University of Erlangen-Nuremberg and Audio Department, Fraunhofer IIS), 91058 Erlangen, Germany (e-mail: konrad.kowalczyk@iis.fraunhofer.de; e.habets@ieee.org)

W. Kellermann is with the Chair of Multimedia Communication and Signal Processing (LMS), University of Erlangen-Nuremberg, 91058 Erlangen, Germany (e-mail: wk@lnt.de).

P. A. Naylor is with the Department of Electrical and Electronic Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, U.K. (e-mail: p.naylor@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2013.2261059

III. Blind Sparse System Identification

To find sparse solutions of an unknown system, l_1-norm regularization can be incorporated into the generic BSI cost function $J(\mathbf{h})$, which can be formulated as the following constrained problem

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \{J(\mathbf{h}) + G(\mathbf{h})\} \qquad \text{s.t.} \qquad \|\mathbf{h}\|_2^2 = 1, \quad (4)$$

where $\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T, \ldots, \hat{\mathbf{h}}_P^T]^T$, $\hat{\mathbf{h}}_p = [\hat{h}_{p,0}, \hat{h}_{p,1}, \ldots, \hat{h}_{p,L-1}]^T$ are the estimates of AIRs, and $\|\mathbf{h}\|_2^2 = 1$ ensures that trivial zero solutions, i.e., $\mathbf{h} = \mathbf{0}_{PL\times 1}$, are avoided. Such a generic BSI cost function is given by [5]

$$J(\mathbf{h}) = \frac{E(\mathbf{h})}{\|\mathbf{h}\|_2^2}, \quad (5)$$

with the *a priori* error $E(\mathbf{h})$ defined as a sum of crossrelation errors for each microphone pair pq ($p, q = 1, \ldots, P, p \neq q$)

$$E(\mathbf{h}) = \sum_{p=1}^{P-1} \sum_{q=p+1}^{P} \|\mathbf{x}_p^T \mathbf{h}_q - \mathbf{x}_q^T \mathbf{h}_p\|_2^2, \quad (6)$$

whereas the l_1-norm sparseness cost function is given by

$$G(\mathbf{h}) = \rho|\mathbf{h}|_1 = \rho \sum_{p=1}^{P} |\mathbf{h}_p|_1, \quad (7)$$

where the uniform regularization parameter ρ defines the relation between the sum of pairwise squares of the crossrelation errors and the sparseness of blindly identified AIRs.

Solving (4) using an adaptive algorithm requires calculating higher-order moments (at least the first-order gradient) of the cost function $J(\mathbf{h}) + G(\mathbf{h})$. Both $J(\mathbf{h})$ and $G(\mathbf{h})$ are convex functions but $G(\mathbf{h})$ is non-differentiable and therefore (4) cannot be solved directly. In order to formulate an adaptive algorithm, the l_1- and l_2-norm components need to be decoupled. For convenience of derivations, let us first omit the unit-norm constraint and reformulate (4) as

$$\min_{\mathbf{h},\mathbf{d}} \{J(\mathbf{h}) + G(\mathbf{d})\} \quad \text{s.t.} \quad \|\mathbf{d} - \mathbf{\Phi}\mathbf{h}\|_2^2 = 0, \quad (8)$$

where $\mathbf{\Phi}$ denotes any linear operator, and $\mathbf{d} = [\mathbf{d}_1^T, \mathbf{d}_2^T, \ldots, \mathbf{d}_P^T]^T$ is an auxiliary variable vector with elements $\mathbf{d}_p = [d_{p,0}, d_{p,1}, \ldots, d_{p,L-1}]^T$. Equation (8) can be further reformulated into an unconstrained optimization problem using a quadratic penalty function as

$$\min_{\mathbf{h},\mathbf{d}} \{J(\mathbf{h}) + G(\mathbf{d}) + \lambda\|\mathbf{d} - \mathbf{\Phi}\mathbf{h}\|_2^2\}, \quad (9)$$

where λ is a Lagrange multiplier. The split Bregman iteration method [7] can then be applied such that a set of unconstrained problems and Bregman updates is obtained, which is given by

$$(\hat{\mathbf{h}}^{k+1}, \mathbf{d}^{k+1}) = \arg\min_{\mathbf{h},\mathbf{d}} \{J(\mathbf{h}) + G(\mathbf{d}) + \lambda\|\mathbf{d} - \mathbf{\Phi}\mathbf{h} - \mathbf{b}^k\|_2^2\} \quad (10a)$$

$$\mathbf{b}^{k+1} = \mathbf{b}^k + \mathbf{\Phi}\hat{\mathbf{h}}^{k+1} - \mathbf{d}^{k+1}, \quad (10b)$$

where $\mathbf{b} = [\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_P^T]^T$, $\mathbf{b}_p = [b_{p,0}, b_{p,1}, \ldots, b_{p,L-1}]^T$ denotes the so-called Bregman variable vector, and k is an iteration index. Note that this way the error introduced by the constraint is added back in the Bregman update (10b), which makes this technique particularly robust against noise. Since $G(\mathbf{d}) = \rho|\mathbf{d}|_1$ is convex but non-differentiable, (10a) can finally be split into the l_1- and l_2-norm components which can be solved using iterative minimization techniques with respect to $\mathbf{h}$ and $\mathbf{d}$, respectively, i.e.,

$$\hat{\mathbf{h}}^{k+1} = \arg\min_{\mathbf{h}} \{J(\mathbf{h}) + \lambda\|\mathbf{d}^k - \mathbf{\Phi}\mathbf{h} - \mathbf{b}^k\|_2^2\} \text{ s.t. } \|\mathbf{h}\|_2^2 = 1, \quad (11a)$$

$$\mathbf{d}^{k+1} = \arg\min_{\mathbf{d}} \{\rho|\mathbf{d}|_1 + \lambda\|\mathbf{d} - \mathbf{\Phi}\hat{\mathbf{h}}^{k+1} - \mathbf{b}^k\|_2^2\}, \quad (11b)$$

with the unit-norm constraint explicitly written in (11a).

IV. Adaptive BSI With Sparse Learning

An adaptive BSI algorithm using sparse learning can be realized by iteratively solving (11a), (11b), and (10b), respectively. The first optimization problem (11a) is differentiable and therefore can be solved using an adaptive gradient-based algorithm. In this letter, we extend two adaptive BSI algorithms, namely, the multichannel least mean squares (MCLMS) [5] and the normalized multichannel frequency-domain LMS (NMCFLMS) [6] algorithms by incorporating an l_1-norm regularization.

A. Sparse Multichannel Least Mean Squares Algorithm

Let us replace the iteration index k with the time index n and denote the pth microphone signal vector at n as

$$\mathbf{x}_p^n = [x_p(n), x_p(n-1), \ldots, x_p(n-L+1)]^T. \quad (12)$$

1) Adaptive Sparse Crossrelation-Based Update: The sparse MCLMS (SMCLMS) update is obtained by taking the gradient of the cost function given by (11a), i.e.,

$$\begin{aligned}\hat{\mathbf{h}}^{n+1} &= \hat{\mathbf{h}}^n - \mu \frac{\partial\left(J^n(\mathbf{h}) + \lambda\|\mathbf{d}^n - \mathbf{\Phi}\mathbf{h} - \mathbf{b}^n\|_2^2\right)}{\partial \mathbf{h}} \\ &= \hat{\mathbf{h}}^n - \mu \frac{\partial J^n(\mathbf{h})}{\partial \mathbf{h}} - \mu\lambda \frac{\partial\left(\|\mathbf{d}^n - \mathbf{\Phi}\mathbf{h} - \mathbf{b}^n\|_2^2\right)}{\partial \mathbf{h}}, \end{aligned} \quad (13)$$

which after setting $\mathbf{\Phi} = \mathbf{I}$ and taking the respective derivatives can be expressed as

$$\hat{\mathbf{h}}^{n+1} = \hat{\mathbf{h}}^n - \frac{\mu\left[\frac{\partial E^n(\mathbf{h})}{\partial \mathbf{h}} - 2J^n \cdot \hat{\mathbf{h}}^n + 2\lambda(\mathbf{b}^n - \mathbf{d}^n + \hat{\mathbf{h}}^n)\right]}{\|\hat{\mathbf{h}}^n\|_2^2}. \quad (14)$$

Enforcing the unit constraint $\|\mathbf{h}\|_2^2 = 1$ by normalizing the AIR estimate at each iteration n, we obtain

$$\hat{\mathbf{h}}^{n+1} = \frac{\hat{\mathbf{h}}^n - 2\mu\left[\tilde{\mathbf{R}}^n \cdot \hat{\mathbf{h}}^n - E^n \cdot \hat{\mathbf{h}}^n - \lambda(\mathbf{b}^n - \mathbf{d}^n + \hat{\mathbf{h}}^n)\right]}{\left\|\hat{\mathbf{h}}^n - 2\mu\left[\tilde{\mathbf{R}}^n \cdot \hat{\mathbf{h}}^n - E^n \cdot \hat{\mathbf{h}}^n - \lambda(\mathbf{b}^n - \mathbf{d}^n + \hat{\mathbf{h}}^n)\right]\right\|_2}, \quad (15)$$

where μ is the step size, the sum of squares of crossrelation errors E^n is given by (6), and $\tilde{\mathbf{R}}^n$ is a $PL \times PL$ matrix with crossrelation matrices as elements (e.g., $\tilde{\mathbf{R}}_{x_p x_q}^n = \mathbf{x}_p^n (\mathbf{x}_q^n)^T$) :

$$\tilde{\mathbf{R}}^n = \begin{bmatrix} \sum_{p\neq 1} \tilde{\mathbf{R}}_{x_p x_p}^n & -\tilde{\mathbf{R}}_{x_2 x_1}^n & \cdots & -\tilde{\mathbf{R}}_{x_P x_1}^n \\ -\tilde{\mathbf{R}}_{x_1 x_2}^n & \sum_{p\neq 2} \tilde{\mathbf{R}}_{x_p x_p}^n & \cdots & -\tilde{\mathbf{R}}_{x_P x_2}^n \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{\mathbf{R}}_{x_1 x_P}^n & -\tilde{\mathbf{R}}_{x_2 x_P}^n & \cdots & \sum_{p\neq P} \tilde{\mathbf{R}}_{x_p x_p}^n \end{bmatrix}. \quad (16)$$

2) Updating Auxiliary and Bregman Variables: Having solved the minimization problem in (11a) using (15) at each n, we can reduce the computational complexity of the algorithm by updating (11b) and (10b) only every M samples, i.e., only at those time instances when $(n+1) \bmod M = 0$. First note that the cost function in (11b) exhibits no coupling between the elements of $\mathbf{d}$, thus the optimum value of $\mathbf{d}$ for (11b) can be conveniently calculated using the so-called shrinkage operator [7]. For $\mathbf{\Phi} = \mathbf{I}$, this optimum solution is given by

$$\mathbf{d}^{n+1} = \text{shrink}\left(\hat{\mathbf{h}}^{M\lfloor\frac{(n+1)}{M}\rfloor} + \mathbf{b}^{M(\lfloor\frac{(n+1)}{M}\rfloor-1)},\ \frac{\rho}{2\lambda}\right), \quad (17)$$

where $\lfloor . \rfloor$ denotes the floor operator, and the shrinkage operator is applied to compute each element of $\mathbf{d}$ separately:

$$d_{p,l}^{n+1} = \text{sign}\left(\hat{h}_{p,l}^{M\lfloor\frac{(n+1)}{M}\rfloor} + b_{p,l}^{M(\lfloor\frac{(n+1)}{M}\rfloor-1)}\right) \times \max\left(\left|\hat{h}_{p,l}^{M\lfloor\frac{(n+1)}{M}\rfloor} + b_{p,l}^{M(\lfloor\frac{(n+1)}{M}\rfloor-1)}\right| - \frac{\rho}{2\lambda}, 0\right) \quad (18)$$

for $p = 1, \ldots, P$ and $l = 0, 1, \ldots, L-1$. In the final step, the Bregman update needs to be performed every M samples:

$$\mathbf{b}^{n+1} = \mathbf{b}^{M(\lfloor\frac{(n+1)}{M}\rfloor-1)} + \hat{\mathbf{h}}^{M\lfloor\frac{(n+1)}{M}\rfloor} - \mathbf{d}^{M\lfloor\frac{(n+1)}{M}\rfloor}. \quad (19)$$

For clarity of presentation, the consecutive processing steps of the SMCLMS algorithm are summarized as Algorithm 1.

Algorithm 1 SMCLMS

initialize: $n = 0$, $\hat{\mathbf{h}}^0 = [1/\sqrt{P}, \mathbf{0}_{L-1}^T, \ldots, 1/\sqrt{P}, \mathbf{0}_{L-1}^T]^T$, $\mathbf{k}^0 = \mathbf{b}^0 = \mathbf{0}_{LP}^T$
 do
  Calculate **(16)** and **(6)** with **(12)** for $p = 1, \ldots, P$
  Calculate **(15)**
  if $(n+1)\,\mathbf{mod}\,M = 0$ **then**
   Calculate **(17)**
   Calculate **(19)**
  end if
  $n = n + 1$
 while nth input sample exists

B. Sparse Normalized Multichannel Frequency-Domain LMS Algorithm

In the following, we present a frequency-domain l_1-norm regularized BSI algorithm. Note that the l_1-norm regularization is a time-domain phenomenon, since we search for solutions which are sparse in time. Therefore, to develop the frequency-domain formulation of (11a), we can first calculate a nonsparse solution in the frequency domain, i.e,

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} J(\mathbf{h}) \quad \text{s.t.} \quad \|\mathbf{h}\|_2^2 = 1, \quad (20)$$

and the l_1-regularization is accounted for after transforming the solution of (20) back to the time domain. Thus the standard NMCFLMS computations are performed first according to [6]

$$(\hat{\underline{\mathbf{h}}}_p^{10})^{m+1} = (\hat{\underline{\mathbf{h}}}_p^{10})^m - \mu[(\boldsymbol{\mathcal{P}}_p)^m + \delta\mathbf{I}]^{-1}\sum_{q=1}^{P}(\boldsymbol{\mathcal{D}}_q^*)^m(\underline{\boldsymbol{e}}_{pq}^{01})^m, \quad (21)$$

$$(\underline{\boldsymbol{\mathcal{P}}}_p)^m = \theta(\underline{\boldsymbol{\mathcal{P}}}_p)^{m-1} + (1-\theta)\sum_{q=1,q\neq p}^{P}(\underline{\boldsymbol{\mathcal{D}}}_q^*)^m(\underline{\boldsymbol{\mathcal{D}}}_q)^m, \quad (22)$$

where m is the block index, $\theta = [1 - 1/(3L)]^L$ denotes the forgetting factor, δ is the regularization parameter, $(\hat{\mathbf{h}}_p^{10})^m = [(\hat{\mathbf{h}}_p^T)^m, \mathbf{0}_{L\times 1}^T]^T$ is the pth AIR estimate with zero padding, and all frequency-domain values are generally denoted with an underscore, e.g., $(\hat{\underline{\mathbf{h}}}_p)^m = \mathbf{F}_L(\hat{\mathbf{h}}_p)^m$, where $\mathbf{F}_L$ is an $L \times L$ DFT matrix. Note that the length of the block is selected as $2L$ with an overlap of 50%, and the mth block of the pth microphone signal is given by

$$(\mathbf{x}_p)^m = [x_p(mL - L), \ldots, x_p(mL + L - 1)]^T. \quad (23)$$

A diagonal matrix with diagonal elements given by the DFT of $(\mathbf{x}_q)^m$ is denoted with $(\underline{\boldsymbol{\mathcal{D}}}_q)^m$, and the frequency-domain crossrelation error is given by [6]

$$(\underline{\boldsymbol{e}}_{pq}^{01})^m = \boldsymbol{\mathcal{W}}_{2L\times L}^{01}\boldsymbol{\mathcal{W}}_{L\times 2L}^{01}\Big[(\underline{\boldsymbol{\mathcal{D}}}_p)^m\boldsymbol{\mathcal{W}}_{2L\times L}^{10}(\hat{\underline{\mathbf{h}}}_q)^m - (\underline{\boldsymbol{\mathcal{D}}}_q)^m\boldsymbol{\mathcal{W}}_{2L\times L}^{10}(\hat{\underline{\mathbf{h}}}_p)^m\Big] \quad (24)$$

with the window matrices $\boldsymbol{\mathcal{W}}_{2L\times L}^{01} = \mathbf{F}_{2L}\mathbf{W}_{2L\times L}^{01}\mathbf{F}_L^{-1} = 2\left(\boldsymbol{\mathcal{W}}_{L\times 2L}^{01}\right)^H$, $\boldsymbol{\mathcal{W}}_{2L\times L}^{10} = \mathbf{F}_{2L}\mathbf{W}_{2L\times L}^{10}\mathbf{F}_L^{-1}$, and $\mathbf{W}_{L\times 2L}^{01} = [\mathbf{0}_{L\times L}, \mathbf{I}_{L\times L}]$ and $\mathbf{W}_{2L\times L}^{10} = [\mathbf{I}_{L\times L}, \mathbf{0}_{L\times L}]^T$. In the second step, the solution (21) is transformed back to the time domain, and l_1-regularization is introduced as

$$\hat{\mathbf{h}}_p^{m+1} = \mathbf{W}_{L\times 2L}^{10}\mathbf{F}_{2L}^{-1}(\hat{\underline{\mathbf{h}}}_p^{10})^{m+1} + \lambda\frac{\partial\left(\|\mathbf{d}_p^m - \mathbf{h}_p - \mathbf{b}_p^m\|_2^2\right)}{\partial\mathbf{h}_p}, \quad (25)$$

which after normalizing the AIR estimates yields the following final update rule for the sparse NMCFLMS algorithm:

$$\hat{\mathbf{h}}_p^{m+1} = \frac{(1-2\mu\lambda)\mathbf{W}_{L\times 2L}^{10}\mathbf{F}_{2L}^{-1}(\hat{\underline{\mathbf{h}}}_p^{10})^{m+1} - 2\mu\lambda(\mathbf{b}_p^m - \mathbf{d}_p^m)}{\left\|(1-2\mu\lambda)\mathbf{W}_{L\times 2L}^{10}\mathbf{F}_{2L}^{-1}(\hat{\underline{\mathbf{h}}}_p^{10})^{m+1} - 2\mu\lambda(\mathbf{b}_p^m - \mathbf{d}_p^m)\right\|_2}, \quad (26)$$

where $\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T, \hat{\mathbf{h}}_2^T, \ldots, \hat{\mathbf{h}}_P^T]^T$ for all channels and the 50% overlap-save is realized using a window matrix $\mathbf{W}_{L\times 2L}^{10} = [\mathbf{I}_{L\times L}, \mathbf{0}_{L\times L}]$. Further two steps include computing the auxiliary and Bregman variables using the following relations

$$\mathbf{d}^{m+1} = \text{shrink}\left(\hat{\mathbf{h}}^{m+1} + \mathbf{b}^m,\ \frac{\rho}{2\lambda}\right), \quad (27)$$

$$\mathbf{b}^{m+1} = \mathbf{b}^m + \hat{\mathbf{h}}^{m+1} - \mathbf{d}^{m+1}. \quad (28)$$

For clarity, the processing steps of the full SNMCFLMS algorithm are summarized as Algorithm 2.

Algorithm 2 SNMCFLMS

initialize: $m = 0$, $(\hat{\underline{\mathbf{h}}})^0 = 1/\sqrt{P}\cdot\mathbf{1}_{LP}^T$, $\mathbf{d}^0 = \mathbf{b}^0 = \mathbf{0}_{LP}^T$
 do
  Calculate **(24)**, **(22),** and **(21)** with **(23)** for $p = 1, \ldots, P$
  Calculate **(26)** for $p = 1, \ldots, P$
  Calculate **(27)**
  Calculate **(28)**
  $m = m + 1$
 while mth block of input samples exists

V. Performance Evaluation

The TDOA estimation performance of the proposed sparse algorithms was compared to standard (non-sparse) counterparts through a set of 50 Monte Carlo simulations. In the experiments, a 15 s duration white Gaussian noise was used as a source signal, which was captured using 2- and 4-element uniform linear arrays (with microphone spacing $g = 0.2$ m) positioned near the center of a $5 \times 6.3 \times 3$ m room simulated using the image-source method. The source was located at

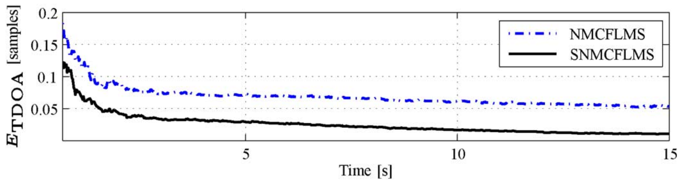

Fig. 1. TDOA error as a function of time (after initial convergence of 0.6 s) for SNMCFLMS and NMCFLMS; $P = 4$ and $\mathrm{SNR} = 40$ dB.

TABLE I
RMS TDOA ERROR E_{TDOA} IN SAMPLES (TOP PANEL) AND THE CORRESPONDING DOA ERROR E_ϕ IN DEGREES (BOTTOM PANEL)

	MCLMS		SMCLMS		NMCFLMS		SNMCFLMS	
SNR	P=2	P=4	P=2	P=4	P=2	P=4	P=2	P=4
0	-	-	-	-	-	-	0.231	0.184
5	-	-	-	-	-	-	0.115	0.078
10	-	-	-	0.151	0.325	0.175	0.039	0.028
20	0.185	0.142	0.105	0.095	0.078	0.038	0.033	0.017
40	0.072	0.02	0.04	0.023	0.074	0.053	0.026	0.009
0	-	-	-	-	-	-	0.47	0.37
5	-	-	-	-	-	-	0.23	0.16
10	-	-	-	1.84	0.66	0.35	0.08	0.06
20	2.25	1.73	1.28	1.16	0.16	0.08	0.07	0.03
40	0.88	0.24	0.49	0.28	0.15	0.11	0.05	0.02

(1.75,0.05,2) m and the reflection coefficient was set to 0.82 for all walls, resulting in a reverberation time of 360 ms. For the SMCLMS algorithm, the following parameters were set: $f_s = 8$ kHz, AIR filter length $L = 128$, and $\mu = 0.8$, whereas for the SNMCFLMS algorithm, experiments were performed for $f_s = 48$ kHz, $L = 1024$, $\mu = 0.8$, $\theta = 0.98$, and $\delta = ((P-1)/(2LP)) \sum_{q=1}^{P} \mathrm{tr}\{(\underline{\mathcal{D}}_q^*)^0 (\underline{\mathcal{D}}_q)^0\}$. The sparseness parameters of SMCLMS were determined empirically: $\lambda = 15 \cdot 10^{-4}$ and $\rho = 8/P \cdot 10^{-7}$ for $\mathrm{SNR} \geq 20$ dB (and ρ was increased by a factor $1/\lambda$ for lower SNRs). The parameters of SNMCFLMS were set to $\lambda = 5 \cdot 10^{-5}$ and $\rho = 8/P \cdot 10^{-5}$ for $\mathrm{SNR} \geq 20$ dB (and we set $\lambda = 0.4$ for lower SNR to emphasize the sparseness constraint). The root mean square TDOA error E_{TDOA} (in samples) between the peaks of simulated and estimated AIRs for the direct-path and 6 first-order reflections for all microphone pairs, was used as a performance measure for different SNR values. The corresponding direction of arrival (DOA) error, calculated as $E_\phi = \arcsin(cE_{\mathrm{TDOA}}/(f_s g))$, where c is the wave speed, is also provided. Note that the peak matching between AIRs was resolved by the AIR simulator, and the DOA errors increase with decreasing microphone spacing and sampling frequency. The results for both experiments are presented in Table I; the TDOA error as a function of time for one scenario is depicted in Fig. 1.

As can be observed, sparse algorithms in general lead to a smaller TDOA estimation error. In particular, they are significantly better at low SNR, where the AIRs estimated with non-sparse algorithms become very noisy, and eventually the reflection peaks cannot be found in such estimates (see Table I and Fig. 2).

VI. CONCLUSIONS

This letter presents a method for crossrelation-based blind identification of sparse systems for accurate TDOA estimation of early room reflections. Two adaptive time- and frequency-domain algorithms are proposed, where l_1-norm regularization is incorporated into the BSI formulation using a split Bregman method; in principle, other sparse BSI algorithms can be formulated in a similar fashion. The proposed methods are shown to outperform their non-sparse counterparts, particularly so in noisy acoustic conditions, and enable the robust performance even at very low SNR, which cannot be otherwise achieved.

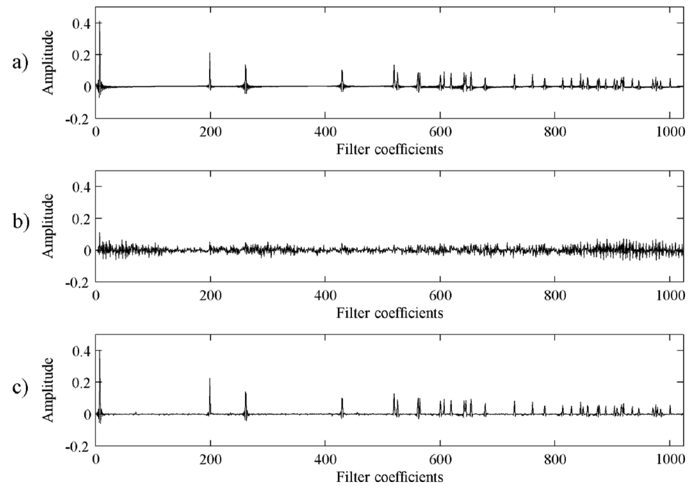

Fig. 2. Acoustic impulse responses for $P = 2$ and $\mathrm{SNR} = 10$ dB: a) normalized simulated AIR for $p = 2$; b) AIR estimated using NMCFLMS; c) AIR estimated using SNMCFLMS.

REFERENCES

[1] Y. Peled and B. Rafaely, "Method for dereverberation and noise reduction using spherical microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2010, pp. 113–116.
[2] T. Betlehem and T. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Amer.*, vol. 117, pp. 2100–2111, 2005.
[3] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Appl. Signal Process.*, vol. Special issue on advances in multimicrophone speech processing, pp. 1–19, 2006.
[4] F. Antonacci, J. Filos, M. Thomas, E. Habets, A. Sarti, P. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 10, pp. 2683–2695, Dec. 2012.
[5] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Process.*, vol. 82, pp. 1127–1138, Aug. 2002.
[6] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.
[7] T. Goldstein and S. Osher, "The split Bregman algorithm for L1 regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.
[8] Y. Wang, W. Yin, and Y. Zhang, "A Fast Algorithm for Image Deblurring with Total Variation Regularization," Tech. Rep., CAAM, 2007.
[9] K. Wagner and M. Doroslovacki, "Towards analytical convergence analysis of proportionate-type NLMS algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Apr. 2008, pp. 3825–3828.
[10] M. Yu, W. Ma, J. Xin, and S. Osher, "Multi-channel l1 regularized convex speech enhancement model and fast computation by the split Bregman method," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 2, pp. 661–675, Feb. 2012.
[11] Y. Lin, J. Chen, Y. Kim, and D. Lee, "Blind channel identification for speech dereverberation using l1-norm sparse learning," in *Neural Inform. Process. Syst. (NIPS)*, 2007.