HAL
archives-ouvertes.fr

Joint estimation of sound source location and boundary impedance with physics-driven cosparse regularization

Nancy Bertin, Srđan Kitić, Rémi Gribonval

▶ To cite this version:

HAL Id: hal-01247227
https://hal.archives-ouvertes.fr/hal-01247227v3

JOINT ESTIMATION OF SOUND SOURCE LOCATION AND BOUNDARY IMPEDANCE WITH PHYSICS-DRIVEN COSPARSE REGULARIZATION

N. Bertin⋆   *S. Kitić*†   *R. Gribonval*†

† Inria, Centre Inria Rennes - Bretagne Atlantique, France
⋆IRISA - CNRS UMR 6074 - Inria Rennes, France

ABSTRACT

Indoor acoustic source localization can be efficiently performed by modeling the sound propagation in the room, and by solving the arising inverse problem by means of cosparse regularization and convex optimization techniques. However, previous methods relying on this approach used to assume the knowledge of a number of room characteristics: its geometry, the walls' absorption or reflexion properties, as well as the speed of sound. In this paper, we show that this model, and the corresponding algorithms, can be extended to the case where the specific acoustic impedance of the boundary is unknown. The proposed method allows to jointly estimate the boundary impedances and the sound pressure in the room, without any preliminary calibration phase, from the only knowledge of the room geometry. Validated on simulation, this new algorithm constitutes a important step towards practical applicability of sound field cosparse modeling.

Index Terms— source localization, reverberation, cosparsity, acoustic impedance, wave equation

1. INTRODUCTION

Acoustic source localization is the problem of determining the position of one or more sources of sound based solely on microphone recordings. The valuable piece information can be exploited in multiple contexts and applications: tracking of sound sources [1, 2, 3], robotics [4, 5, 6], seismic imaging [7, 8] and medical imaging [9, 10]. In speech and sound enhancement, a source position estimate can be used to perform noise reduction, by means of beamforming techniques [11, 12, 13, 14]. Noise, in this case, can include the reverberated part of the sound emitted by the source of interest. Reverberation is considered a nuisance, against which one has to be "robust", if not able to suppress it.

Recent work [15], including our own [16, 17], has explored the idea that reverberation actually contains some useful information about the sources or the environment. Provided an appropriate model (a physics-driven cosparse model described in section 2), we showed that accurate source localization can be performed by exploiting the acoustic multipath, by means of scalable algorithms with affordable computational cost and memory requirements. However, these algorithms required good knowledge of the room characteristics: its geometry, the nature of the propagation medium and the boundary conditions. Such information is rarely available in practice, which may limit the usefulness of this approach in real applications.

Fortunately, while estimating the sound speed in air is important, in many cases an initial guess (for instance, from a temperature

measurement in a steady-state room) can be accurate enough, and it has be showed that it can be learnt in the cosparse framework in some cases [18]. On the other hand, guessing the specific acoustic impedance of the walls, floor and ceiling of the room is much more difficult. In practice, one may consult standardized tables of absorption properties of different structures from construction engineering literature (with the need of knowing exactly which materials were used), or to physically measure the acoustic impedance (*e.g.* [19, 20]) of the concerned room, which requires specific hardware setup and calibration. Those two approaches are inflexible when considering different acoustic environments, and possibly inaccurate. A less demanding approach is to use only a microphone array and a known sound source, as presented in [21]. Such measurement would allow to tune the model for later use with unknown sources in the same room, but still requires a preliminary calibration step.

Our goal is to go a step beyond aforementioned approaches: we want a method that can simultaneously estimate the specific impedance *and* the acoustic pressure/sound source component, thus also performing source localization, while not being given any other information but the geometry of the room. In this paper, we show that the previously developed cosparse model and optimization algorithms can be adapted to cope with such situation, namely with unknown boundary conditions. We present an algorithm able to jointly estimate the source location and the impedance at the boundaries (section 3), and we validate it experimentally (section 4).

2. PHYSICS-DRIVEN COSPARSE MODELING OF THE ACOUSTIC SCENE

We formulate the source localization problem as follows: given an array of m omnidirectional microphones with a known geometry, we want to determine the locations of k point monopole sound sources, within an enclosed spatial environment. To this end, we first introduce the physics knowledge we rely on, and how a simple discretization leads to recasting source localization as an ill-posed inverse problem that can be regularized from a sparsity assumption, and practically solved with convex optimization approaches.

2.1. From the wave equation to inverse problems

A microphone array captures samples of the sound pressure field p, which obeys the well-known *wave equation*:

$$\Box_\omega p = \triangle_\mathbf{r} p(\mathbf{r}, \mathrm{t}) - \frac{1}{c(\mathbf{r}, \mathrm{t})^2} \frac{\partial^2 p(\mathbf{r}, \mathrm{t})}{\partial \mathrm{t}^2} = z(\mathbf{r}, \mathrm{t}), \qquad (1)$$

were p is defined on the *spacetime* product space $\omega := (\mathbf{r}, \mathrm{t}) \in \Omega = \Gamma \times \mathcal{T}$, $\triangle$ is the *Laplace operator*, $\Box$ the *D'Alembert operator*, z is the *source term*, and $c(\mathbf{r}, \mathrm{t})$ denotes the *speed of propagation*. The wave equation defines a unique pressure field as soon as it

is accompanied by appropriate initial and boundary conditions, *i.e.*, some conditions on the pressure at the limits $\partial\Omega$ of the domain Ω, that can be represented under the form of an operator B such that $B_{\partial\Omega}p = z_{\partial\Omega}$. Then, the full physics-driven model can be encompassed in the concised form of an operator A:

$$\square_{\omega}p = z, \ B_{\partial\Omega}p = z_{\partial\Omega} \ \Leftrightarrow \ Ap = z. \tag{2}$$

We then recast the localization problem as an inverse problem. The continuous domain Ω is replaced by a set of discrete coordinates of dimension $\mathsf{n} := \mathsf{st}$, where s and t represent the numbers indicating finesse of the spatial and temporal grids, respectively. The discrete observation data $y \rightarrow \mathbf{y} \in \mathbb{R}^{\mathsf{mt}}$ is obtained by downsampling a discretized signal $p \rightarrow \mathbf{p} \in \mathbb{R}^{\mathsf{n}}$ by means of a row-reduced identity matrix $\mathbf{M} \in \mathbb{R}^{\mathsf{mt}\times\mathsf{n}}$. Once discretized, the differential operator A is represented in matrix form as $\mathbf{A} \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$. In our case, this discretization is performed by means of a Finite-Domaine-Time-Difference (FDTD) Standard Leap Frog (SLF) method, which corresponds to second-order centered finite differences in space and time [22]. Likewise, the right hand side z of (2) is discretized into $\mathbf{z} \in \mathbb{R}^{\mathsf{n}}$, and $\mathbf{Ap} = \mathbf{z}$ holds.

Localization as an inverse problem boils down to recovering $\mathbf{z}$ from measurement $\mathbf{y} = \mathbf{Mp}$, and knowing that $\mathbf{Ap} = \mathbf{z}$ (with a fully known operator $\mathbf{A}$ at this stage). This problem is obviously ill-posed, but can be regularized and solved as soon as we assume the sparsity of the source term z, naturally emerging from the reasonable assumption that there are a small number of sources, compared to the number of voxels in the discretized space Γ. The *sparse analysis (aka cosparse)* [23] optimization problem to solve is:

$$\underset{\mathbf{p}}{\text{minimize}} \ f_r(\mathbf{Ap}) + f_d(\mathbf{Mp} - \mathbf{y}) + f_c(\mathbf{Ap}). \tag{3}$$

Here, $f_r(\cdot)$ denotes a measure of sparsity, $f_d(\cdot)$ denotes a measure of data-fidelity in the discrete context, and $f_c(\cdot)$ denotes a penalty that enforces consistency with boundary and/or initial conditions. The matrix $\mathbf{A}$ is called *the analysis operator*.

2.2. Solving the inverse problem through convex optimization

With appropriate choices of the cost functions, the optimization problem (3) is a convex problem that can be practically addressed by first-order optimization algorithms such as the Simultaneous Direction Method of Multipliers (SDMM) scheme [24], or variants thereof. It is the case, for instance, by choosing $f_r(\cdot)$ as the structured sparsity-inducing *joint $\ell_{2,1}$-norm* [25]:

$$\|\mathbf{z}\|_{2,1} = \sum_{\mathsf{i}} \sqrt{\sum_{\mathsf{j}} |z_{\mathsf{i},\mathsf{j}}|^2}, \tag{4}$$

where $z_{\mathsf{i},\mathsf{j}}$ denotes the $(\mathsf{i},\mathsf{j})^{\text{th}}$ element obtained by transforming the vector $\mathbf{z}$ into a matrix $\mathbf{Z}$ whose columns are jointly sparse subvectors; and $f_d(\cdot)$ and $f_c(\cdot)$ as the characteristic functions of the set of vectors whose norm is bounded by a tolerance parameter (*e.g.* ε):

$$\chi_{\ell_2 \leq \varepsilon}(\mathbf{v}) := \begin{cases} 0 & \|\mathbf{v}\|_2 \leq \varepsilon, \\ +\infty & \text{otherwise.} \end{cases} \tag{5}$$

We showed that the problem can be accurately solved by means of a warm-started, weighted SDMM strategy fully described in [17], with known speed of sound, known geometry, and uniform Neumann boundary condition. It has also the advantage of showing remarkable computational properties, making it tractable even in 3D, and able to computationally benefit from more observations (unlike its counterpart "sparse synthesis" approach). The interested reader may refer to the abovementioned article [17] for further details.

3. SOURCE LOCALIZATION WITH UNKNOWN BOUNDARY CONDITIONS

After reminding how we can model and solve the source localization problem in a perfectly known environment, we now relax the assumption of knowing the (spatial) boundary conditions, and turn it into variables to be jointly estimated with $\mathbf{z}$.

3.1. Boundary modeling

Here, we will consider the adaptation of *Mur's absorbing boundary condition* [26], which is stated as follows:

$$\frac{\partial p}{\partial \mathsf{t}} + c\xi \nabla p \cdot \mathbf{n} = 0, \tag{6}$$

where the parameter ξ is termed *specific acoustic impedance* and controls the reflection coefficient through the relation:

$$R(\theta) = \frac{\xi \cos\theta - 1}{\xi \cos\theta + 1}, \tag{7}$$

where θ is the incidence angle. This implies that for large values of ξ, the reflected wave is almost identical to the incident wave - equivalent to the well-known Neumann (*hard wall*) condition, while for the values of ξ close to 1 this condition models an absorbing boundary. For $\xi = 0$, adapted Mur's boundary condition is equivalent to constant Dirichlet's (*soft wall*) boundary condition. As such, this model encompasses classical, idealistic boundary conditions as well as it offers the possibility to represent more complex and realistic situations. In general, coefficient ξ is frequency-dependent, but, for the sake of simplicity, we take it to be constant along frequencies.

3.2. Discretization of the boundary

Now, in order to make the new unknowns appear in the formulation of the optimization problem to solve, we need to explicitly model the discretization of the boundary terms. We assume here a 2D spatial domain Γ (extension to 3D is straightforward). In the SLF scheme, pressure inside the *discrete* spatial domain $\Gamma \setminus \partial\Gamma$ (excluding the boundary), and corresponding to time samples $\mathsf{t} > 2$, is discretized using the 7-point stencil as follows:

$$\frac{\partial^2 p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}}}{\partial x^2} + \frac{\partial^2 p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}}}{\partial y^2} - \frac{1}{c^2}\frac{\partial^2 p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}}}{\partial t^2} =$$
$$\frac{p_{\mathsf{i}-1,\mathsf{j}}^{\mathsf{t}} - 2p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}} + p_{\mathsf{i}+1,\mathsf{j}}^{\mathsf{t}}}{\mathsf{d}_{\mathsf{x}}^2} + \frac{p_{\mathsf{i},\mathsf{j}-1}^{\mathsf{t}} - 2p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}} + p_{\mathsf{i},\mathsf{j}+1}^{\mathsf{t}}}{\mathsf{d}_{\mathsf{y}}^2}$$
$$- \frac{1}{c^2}\frac{p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}+1} - 2p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}} + p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}-1}}{\mathsf{d}_{\mathsf{t}}^2} + \mathrm{O}(\max(\mathsf{d}_{\mathsf{x}},\mathsf{d}_{\mathsf{y}},\mathsf{d}_{\mathsf{t}})^2) \tag{8}$$

where d_{x}, d_{y} and d_{t} denote the discretized spatial and temporal step sizes, respectively. Neglecting the $\mathrm{O}(\cdot)$ term yields a convenient *explicit* scheme [27] to compute $p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}+1}$ using pressure values at the previous two discrete time instances ($p_{(\cdot,\cdot)}^{\mathsf{t}}$ and $p_{(\cdot,\cdot)}^{\mathsf{t}-1}$). Formulas for boundary nodes are obtained by substituting a non-existent spatial point in the previous scheme by the expressions obtained from discretized boundary conditions. For the frequency-independent acoustic absorbing boundary condition, *e.g.* the missing point behind the right "wall" is evaluated as:

$$p_{\mathsf{i}+1,\mathsf{j}}^{\mathsf{t}} = p_{\mathsf{i}-1,\mathsf{j}}^{\mathsf{t}} + \frac{\mathsf{d}_{\mathsf{p}}}{\mathsf{cd}_{\mathsf{t}}\xi_{\mathsf{i},\mathsf{j}}}\left(p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}-1} - p_{\mathsf{i},\mathsf{j}}^{\mathsf{t}+1}\right), \tag{9}$$

which leads to the explicit expression:

$$p_{i,j}^{t+1} = \left[2(1 - 2\lambda^2)p_{i,j}^t + \lambda^2(p_{i,j+1}^t + p_{i,j-1}^t) + 2\lambda^2 p_{i-1,j}^t \right.$$
$$\left. - \left(1 - \frac{\lambda}{\xi_{i,j}}\right)p_{i,j}^{t-1}\right] \bigg/ \left(1 + \frac{\lambda}{\xi_{i,j}}\right), \quad (10)$$

where $\lambda = cd_t/d_x = cd_t/d_y = cd_t/d_z$, assuming uniform spatial discretization, for simplicity. When corners (and edges in 3D) are considered, the condition is applied to all directions where the stencil points are missing.

3.3. Optimization problem reformulation

By grouping coefficients involving $\xi_{\cdot,\cdot}^{-1}$ in these expressions, we can represent the boundary part $\mathbf{A}_{\partial\Gamma}$ of the analysis operator as the sum of the following two component matrices:

$$\mathbf{A}_{\partial\Gamma} = \underbrace{\begin{bmatrix} \underline{\mathbf{A}}_{\partial\Gamma_1} & & & \\ & \underline{\mathbf{A}}_{\partial\Gamma_1} & & \\ & & \ddots & \\ & & & \underline{\mathbf{A}}_{\partial\Gamma_1} \end{bmatrix}}_{\mathbf{A}_{\partial\Gamma_1} \in \mathbb{R}^{bt \times n}}$$

$$+ \ \mathtt{diag}\,(\mathbf{Sb}) \underbrace{\begin{bmatrix} \underline{\mathbf{A}}_{\partial\Gamma_2} & & & \\ & \underline{\mathbf{A}}_{\partial\Gamma_2} & & \\ & & \ddots & \\ & & & \underline{\mathbf{A}}_{\partial\Gamma_2} \end{bmatrix}}_{\mathbf{A}_{\partial\Gamma_2} \in \mathbb{R}^{bt \times n}}. \quad (11)$$

The vector $\mathbf{b} = [\xi_{1,1} \ \xi_{2,1} \ \ldots \ \xi_{i,j} \ \ldots]^{-\mathsf{T}} \in \mathbb{R}^{\tilde{b}}$ contains inverse acoustic impedances, *i.e. specific acoustic admittances*. The number of elements in $\mathbf{b}$ is equal to $\tilde{b} = b + (d-2)e + (d-1)c$, where b is the number of spatial elements corresponding to discretized boundary, e is the number of edges, c the number of corner nodes and d is the number of spatial dimensions (*i.e.* $d = 2$ for 2D). The blocks $\underline{\mathbf{A}}_{\partial\Gamma_1}$ and $\underline{\mathbf{A}}_{\partial\Gamma_2}$ contain fixed coefficients of $\mathbf{A}_{\partial\Gamma}$, while the matrix $\mathbf{S} \in \mathbb{R}^{bt \times \tilde{b}}$ is a row-wise concatenation of blocks $\tilde{\mathbf{S}} \in \mathbb{R}^{b \times \tilde{b}}$. Each individual block $\tilde{\mathbf{S}}$ is almost an identity matrix, except for the edge and corner nodes. We also define the matrix $\mathbf{A}_{\Gamma \setminus \partial\Gamma}$, which is the analysis operator $\mathbf{A}$ *without* the boundary rows, and the matrix $\mathbf{A}_0$ which is the part of $\mathbf{A}$ defining initial conditions.

The admittance vector $\mathbf{b}$ is parametrizing enclosure boundaries, which, for rooms, may be composed of walls, floor, ceiling, windows etc. At least on macroscopic scale, these structures are approximately homogeneous. Hence, we will suppose that $\mathbf{b}$ admits a *piecewise constant* model (provided we take care of the ordering of elements within $\mathbf{b}$). This weak assumption usually holds in practice, unless the discretization is very crude.

Now, having introduced the involved matrices and adopted assumptions, we proceed to formulating the optimization problem:

$$\underset{\mathbf{p},\mathbf{b}}{\text{minimize}} \ f_r(\mathbf{A}_{\Gamma \setminus \partial\Gamma}\mathbf{p}) + \chi_{\ell_2 \leq \sigma}(\mathbf{A}_0\mathbf{p}) + \chi_{\ell_2 \leq \varepsilon}(\mathbf{Mp} - \mathbf{y})$$

$$+ \|\mathbf{b}\|_{\mathrm{TV}} + \lambda\|\mathbf{A}_{\partial\Gamma}\mathbf{p}\|_2^2,$$

subject to $\mathbf{A}_{\partial\Gamma} = \mathbf{A}_{\partial\Gamma_1} + \mathtt{diag}\,(\mathbf{Sb})\,\mathbf{A}_{\partial\Gamma_2}$ and $\mathbf{b} \succ \mathbf{0}$, (12)

where $\|\mathbf{b}\|_{\mathrm{TV}} = \|\nabla\mathbf{b}\|_1$ denotes the total variation norm, which is known to promote piecewise constant solutions.

3.4. The CALAIS algorithm

A bilinear form is embedded into the product $\mathbf{A}_{\partial\Gamma}\mathbf{p}$, and thus, (12) is a biconvex problem. Although we could apply alternating minimization, we chose here to use a biconvex version of ADMM, which is known to exhibit better empirical performance [28]. The conceptual pseudocode (in Algorithm 1), is termed *Cosparse Acoustic Localization, Acoustic Impedance estimation and Signal recovery (CALAIS)*.

Algorithm 1 *CALAIS*

Require: $\mathbf{y}$, $\mathbf{M}$, $\mathbf{A}_{\partial\Gamma_1}$, $\mathbf{A}_{\partial\Gamma_2}$, $\mathbf{A}_0$, $\mathbf{A}_{\Gamma \setminus \partial\Gamma}$, $\mathbf{b}^{(0)}$, $\mathbf{S}$, ε, λ

1: $\mathbf{u}^{(0)} = \mathbf{0}$, $\mathbf{A}_{\partial\Gamma}^{(0)} = \mathbf{A}_{\partial\Gamma_1} + \mathtt{diag}\left(\mathbf{Sb}^{(0)}\right)\mathbf{A}_{\partial\Gamma_2}$

2: $\mathbf{p}^{(i+1)} = \operatorname{argmin}_\mathbf{p} f_r(\mathbf{A}_{\Gamma \setminus \partial\Gamma}\mathbf{p}) + \lambda\|\mathbf{A}_{\partial\Gamma}^{(i)}\mathbf{p} + \mathbf{u}^{(i)}\|_2^2$,
   s.t. $\|\mathbf{A}_0\mathbf{p}\|_2 \leq \sigma$, $\|\mathbf{Mp} - \mathbf{y}\|_2 \leq \varepsilon$

3: $\mathbf{b}^{(i+1)} = \operatorname{argmin}_\mathbf{b} \|\mathbf{b}\|_{\mathrm{TV}} + \lambda\|\mathtt{diag}\left(\mathbf{A}_{\partial\Gamma_2}\mathbf{p}^{(i+1)}\right)\mathbf{Sb} + \mathbf{A}_{\partial\Gamma_1}\mathbf{p}^{(i+1)} + \mathbf{u}^{(i)}\|_2^2$ s.t. $\mathbf{b} \succeq \mathbf{0}$

4: $\mathbf{A}_{\partial\Gamma}^{(i+1)} = \mathbf{A}_{\partial\Gamma_1} + \mathtt{diag}\left(\mathbf{Sb}^{(i+1)}\right)\mathbf{A}_{\partial\Gamma_2}$

5: $\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + \mathtt{diag}\left(\mathbf{A}_{\partial\Gamma_2}\mathbf{p}^{(i+1)}\right)\mathbf{Sb}^{(i+1)} + \mathbf{A}_{\partial\Gamma_1}\mathbf{p}^{(i+1)}$

6: **if** convergence **then**

7:    terminate

8: **else**

9:    $i \leftarrow i + 1$

10:    go to 2

11: **end if**

12: **return** $\hat{\mathbf{p}} = \mathbf{p}^{(i+1)}$, $\hat{\mathbf{b}} = \mathbf{b}^{(i+1)}$, $\hat{\mathbf{A}}_{\partial\Gamma} = \mathbf{A}_{\partial\Gamma}^{(i+1)}$

Note that evaluating the minimizers of the intermediate convex problems (steps 2 and 3) is here left to "black-box" approaches, *e.g.* we used the aforementioned Weighted SDMM algorithm (see [17]). To accelerate computation, Weighted SDMM is always warm-started by the estimate obtained in the preceding CALAIS iteration.

4. SIMULATIONS

4.1. Experimental setup

All experiments are conducted in a simulated two dimensional (rectangular) room of size 15×15m, with acquisition time set to $t = 1$s and speed of sound $c = 343$m/s. With spatial step size of $1m^2$, this corresponds to ~ 110000 discrete points in space and time. For the experiments where we avoid the so-called "inverse crime", a different grid is used for generating data: in this case, the spatial step size is $0.25m^2$, which yields a discrete model of size ~ 450000, and the sources may be "off-the-grid" when performing reconstruction. White noise sound sources emit during the whole acquisition period $(0, \tau]$, except for $t = 0$, due to the homogeneous initial conditions. The specific acoustic impedance parameters are set to ξ_1 and ξ_2, for each pair of opposite "walls". Two series of experiments are performed, with the varying number of microphones m ($3 \leq m \leq 90$) and sources k ($1 \leq k \leq 18$), each repeated 20 times.

The first series: We assume idealized (inverse crime) conditions, *i.e.* a perfectly accurate geometric model. In this setting we conducted both the experiments with noiseless, and with noisy measurement data. The impedances are set per pair of opposite "2D walls", as follows: one pair is always modeled as *hard wall* ($\xi_1 = 100$), whereas the other is modeled either as *soft wall* ($\xi_2 = 0.3$) or as a highly-absorbing surface ($\xi_2 = 0.9$).

The second series: This is the non-inverse crime setup: we keep the previous settings, but use the fine grid to generate the data
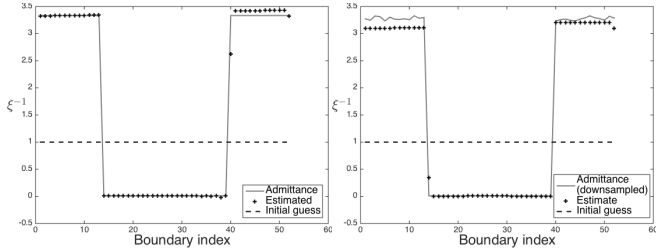
Fig. 1: Original and estimated acoustic admittances in the inverse crime (left) and non-inverse crime (right) setting.

and the crude grid to solve the inverse problem (observations $\mathbf{y}$ are temporally low-pass filtered to reduce aliasing). Moreover, we add white Gaussian measurement noise (AWGN) such that *per-sample* SNR is 20dB. Finally, we weaken the piecewise constant model by corrupting the vector $\boldsymbol{\xi}$ with AWGN, distributed as $\mathcal{N}(0, 0.01)$.

For simplicity, we assumed that the number of sources is known - when this is not the case, one could detect source locations from $\hat{\mathbf{z}}_{\Gamma \backslash \partial \Gamma} = \mathbf{A}_{\Gamma \backslash \partial \Gamma} \hat{\mathbf{p}}$ using a magnitude threshold. Performance is evaluated in terms of minimal Root Mean Square Error (RMSE) per sound source. In the inverse crime setting, localization is considered successful when this error is zero. In the non-inverse crime setting, an error on the order of crude grid's stepsize is tolerated. In all experiments, we used $\lambda = 0.1$ and the initial estimate $\mathbf{b}^{(0)} = \mathbf{1}$ (the vector of all ones). The stopping criterion is based on the relative distance between the objective function (12) values at two successive iterations. The data fidelity parameter ε is set to 0, as well as the initial condition constraint parameter σ, in all experiments. Thus, the noise variance and model error are assumed unknown beforehand.

4.2. Results

In the first series of experiments, the CALAIS algorithm achieves very high overall localization accuracy. In fact, it performs almost identically as our standard physics-driven cosparse localization (comparable to results in [17]), for which the impedance parameters are known. The results are unaffected by the change of impedance value ξ_2, suggesting that, as long as there is *any* multipath diversity in the system, it is able to exploit it. For all these reasons, we do not present localization results for the inverse crime case (an example of the admittance estimation is given in Figure 1 - left).

In the second (more realistic) case, where the models are inaccurate and the measurements are noisy, localization accuracy deteriorates. Fortunately, "perfect" localization accuracy is preserved when the number of sources is low ($k = 1$ and $k = 2$). In these cases, median results in Figure 2 indicate that localization is possible provided that the number of microphones is above 12. Moreover, even though median RMSE is usually above the tolerance for $k = 3$, it is still lower than two crude grid's spatial stepsizes, suggesting that the sources are localized in their true, immediate neighborhood. Finally, if localization is successful (*w.r.t.* the RMSE tolerance), the admittances are also accurately estimated (Figure 1, right - the ground truth admittance has been downsampled for better visualization).

5. CONCLUSION AND PERSPECTIVES

Physics-driven cosparse approach to sound localization offers attractive features, but its applicability was previously limited by the
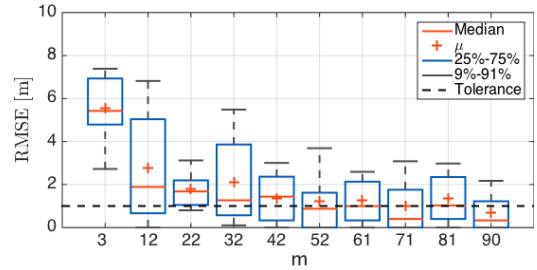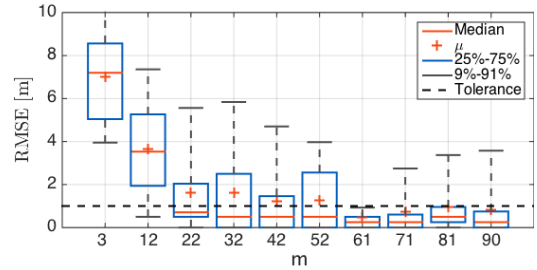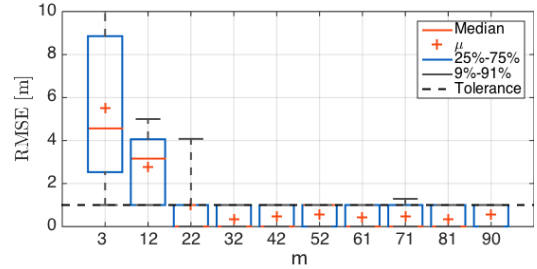

Fig. 2: RMSE *vs* number of microphones m in the non-inverse crime setting. From top to bottom: $k = 1$, $k = 2$ and $k = 3$ sources.

need for knowledge of some room characteristics. Here, we argued that, as long as a piecewise constant model of the specific acoustic impedance approximately holds, this parameter can be estimated in parallel to localization and recovery. To our best knowledge, there have been no attempts so far to perform blind estimation of these physical parameters, by exploiting spatial sparsity of sound sources. The results could be potentially improved if, in addition, a source model was assumed (*e.g.* bandlimited signals). We have chosen to apply a biconvex ADMM heuristics, but we envision the possibility of developing a *lifting scheme*, successfully exploited in phase retrieval problems (albeit, with an increase in problem dimension) [29, 30]. Had we considered an unknown speed of sound, we feel that a multiconvex reformulation of the optimization problem would have allowed us to jointly estimate the source locations, speed of sound and specific acoustic impedance.

However, even with the now open possibility to estimate the speed of sound and wall impedance, the approach remains *semi-blind* in the sense that the geometry of the room still needs to be known beforehand. Recent approaches have demonstrated that such a "room shape estimation" is reachable with known impedance [31], but the *fully blind* localization with completely unknown environment remains an open problem. An equally important issue is computational cost of 3D wavefield estimation in temporal domain, which governs the computational complexity of our algorithm.

6. REFERENCES

[1] S. Affes, S. Gazor, and Y. Grenier, "Robust adaptive beamforming via lms-like target tracking," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 1994, vol. 4, pp. IV–269.

[2] D.B. Ward, E. Lehmann, R.C. Williamson, et al., "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 826–836, 2003.

[3] P.-J. Chung, J.F. Böhme, and A.O. Hero, "Tracking of multiple moving sources using recursive em algorithm," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 50–60, 2005.

[4] J.-M. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2003, vol. 2, pp. 1228–1233.

[5] K. Nakadai, H. G Okuno, H. Kitano, et al., "Real-time sound source localization and separation for robot audition.," in *INTERSPEECH*, 2002.

[6] A. Deleforge and R. Horaud, "The cocktail party robot: Sound source separation and localisation with an active binaural head," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 2012, pp. 431–438.

[7] S. Miron, N. Le Bihan, and J.I. Mars, "Vector-sensor music for polarized seismic sources localization," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 74–84, 2005.

[8] J.C. Chen, K. Yao, and R.E. Hudson, "Source localization and beamforming," *Signal Processing Magazine, IEEE*, vol. 19, no. 2, pp. 30–39, 2002.

[9] J.-F. Synnevåg, A. Austeng, and S. Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 54, no. 8, pp. 1606–1613, 2007.

[10] F.H. Rose, C.D. McGregor, and P. Mathur, "Method and apparatus for stone localization using ultrasound imaging," Jan. 30 1990, US Patent 4,896,673.

[11] X. Zhang and J.H.L. Hansen, "CSA-BF: A constrained switched adaptive beamformer for speech enhancement and recognition in real car environments," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 733–745, 2003.

[12] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.

[13] I. Chiba, T. Takahashi, and Y. Karasawa, "Transmitting null beam forming with beam space adaptive array antennas," in *Vehicular Technology Conference, 1994 IEEE 44th*. IEEE, 1994, pp. 1498–1502.

[14] J. Benesty, S. Makino, and J. Chen, *Speech enhancement*, Springer Science & Business Media, 2005.

[15] I. Dokmanić and M. Vetterli, "Room helps: Acoustic localization with finite elements," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Ieee, 2012, pp. 2617–2620.

[16] S. Kitić, N. Bertin, and R. Gribonval, "Hearing behind walls: localizing sources in the room next door with cosparsity," in *IEEE International Conference o Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 3087–3091.

[17] S. Kitić, L. Albera, N. Bertin, and R. Gribonval, "Physics-driven inverse problems made tractable with cosparse regularization," *IEEE Transactions on Signal Processing*, vol. 64, pp. 335–348, 2016.

[18] C. Bilen, S. Kitić, N. Bertin, and R. Gribonval, "Sparse acoustic source localization with blind calibration for unknown medium characteristics," in *iTwist-2nd international-Traveling Workshop on Interactions between Sparse models and Technology*, 2014.

[19] G.P. Nava, Y. Yasuda, Y. Sato, and S. Sakamoto, "On the in situ estimation of surface acoustic impedance in interiors of arbitrary shape by acoustical inverse methods," *Acoustical science and technology*, vol. 30, no. 2, pp. 100–109, 2009.

[20] J.-P. Dalmont, "Acoustic impedance measurement, part i: A review," *Journal of Sound and Vibration*, vol. 243, no. 3, pp. 427–439, 2001.

[21] N. Antonello, T. van Waterschoot, M. Moonen, and P.A. Naylor, "Evaluation of a numerical method for identifying surface acoustic impedances in a reverberant room," in *Proc. of the European Congress and Exposition on Noise Control Engineering*, 2015, pp. 1–6.

[22] R.J. LeVeque, *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*, vol. 98, Siam, 2007.

[23] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[24] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[25] R. Jenatton, J.-Y. Audibert, and F. Bach, "Structured variable selection with sparsity-inducing norms," *The Journal of Machine Learning Research*, vol. 12, pp. 2777–2824, 2011.

[26] K. Kowalczyk and M. van Walstijn, "Modeling frequency-dependent boundaries as digital impedance filters in fdtd and k-dwm room acoustics simulations," *Journal of the Audio Engineering Society*, vol. 56, no. 7/8, pp. 569–583, 2008.

[27] J. Strikwerda, *Finite difference schemes and partial differential equations*, SIAM, 2004.

[28] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[29] E.J. Candes, Y.C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Review*, vol. 57, no. 2, pp. 225–251, 2015.

[30] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *arXiv preprint arXiv:1501.06864*, 2015.

[31] T. Nowakowski, N. Bertin, R. Gribonval, J. De Rosny, and L. Daudet, "Membrane Shape And Boundary Conditions Estimation Using Eigenmode Decomposition," in *ICASSP - IEEE International Conference on Acoustics, Speech and Signal Processing*, Shanghai, China, 2016.