Gridless 3D Recovery of Image Sources from Room Impulse Responses

T. Sprunck[1,2], Y. Privat[2], C. Foy[3], A. Deleforge[1]

[1]Inria, Nancy, France [2]IRMA, Strasbourg, France [3]UMRAE, Strasbourg, France

1. Motivation

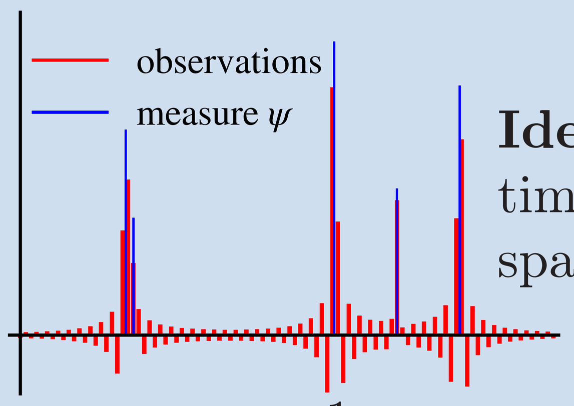
Can one hear the shape of a room ?
More precisely, given discrete filtered measurements of the propagation from an impulse sound source to a microphone antenna, can we recover the locations of the walls?

3. Hypotheses

- Rectangular cuboid rooms
- Specular reflections
- Omnidirectional sources and receivers
- Frequency-independent walls, floor, ceiling
- Fixed source and receiver responses: ideal low-pass filters
- One point source emitting a perfect impulse at $t = 0$

4. What is Super-Resolution ?

Objective: reconstruct a discrete measure $\psi = \sum_k a_k \delta_{\mathbf{r}_k}$ from linear observations $\mathbf{x} = \Gamma\psi = \int_r \gamma(r) d\psi(r) \in \mathbb{R}^D$

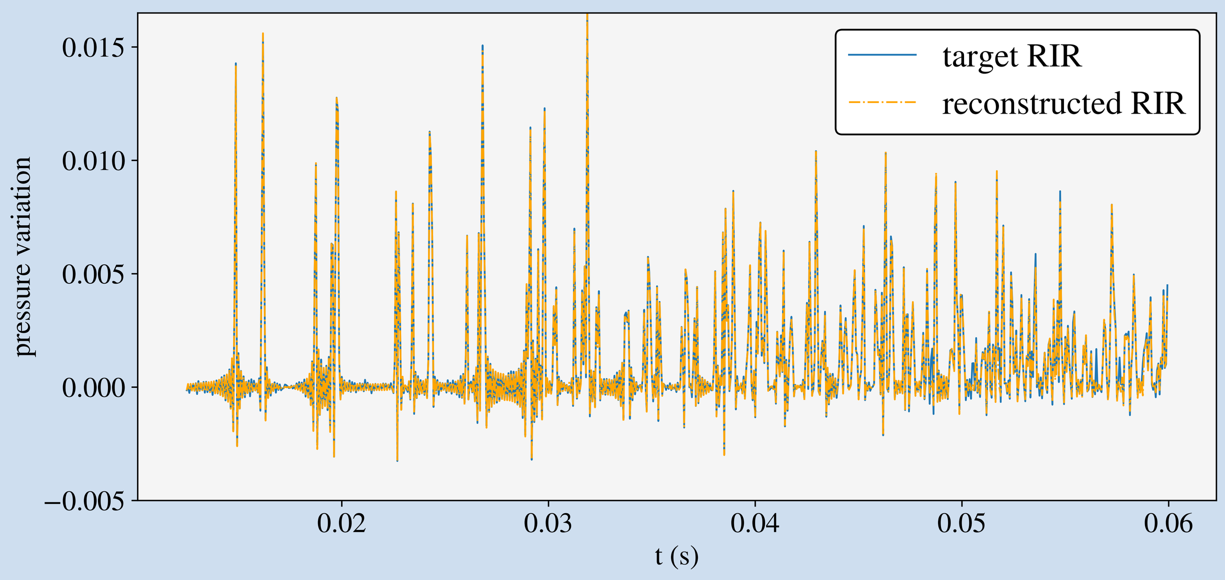


Idea: consider a relaxed convex optimization problem over the entire space of Radon measures [2] of $\mathbb{R}^3$:

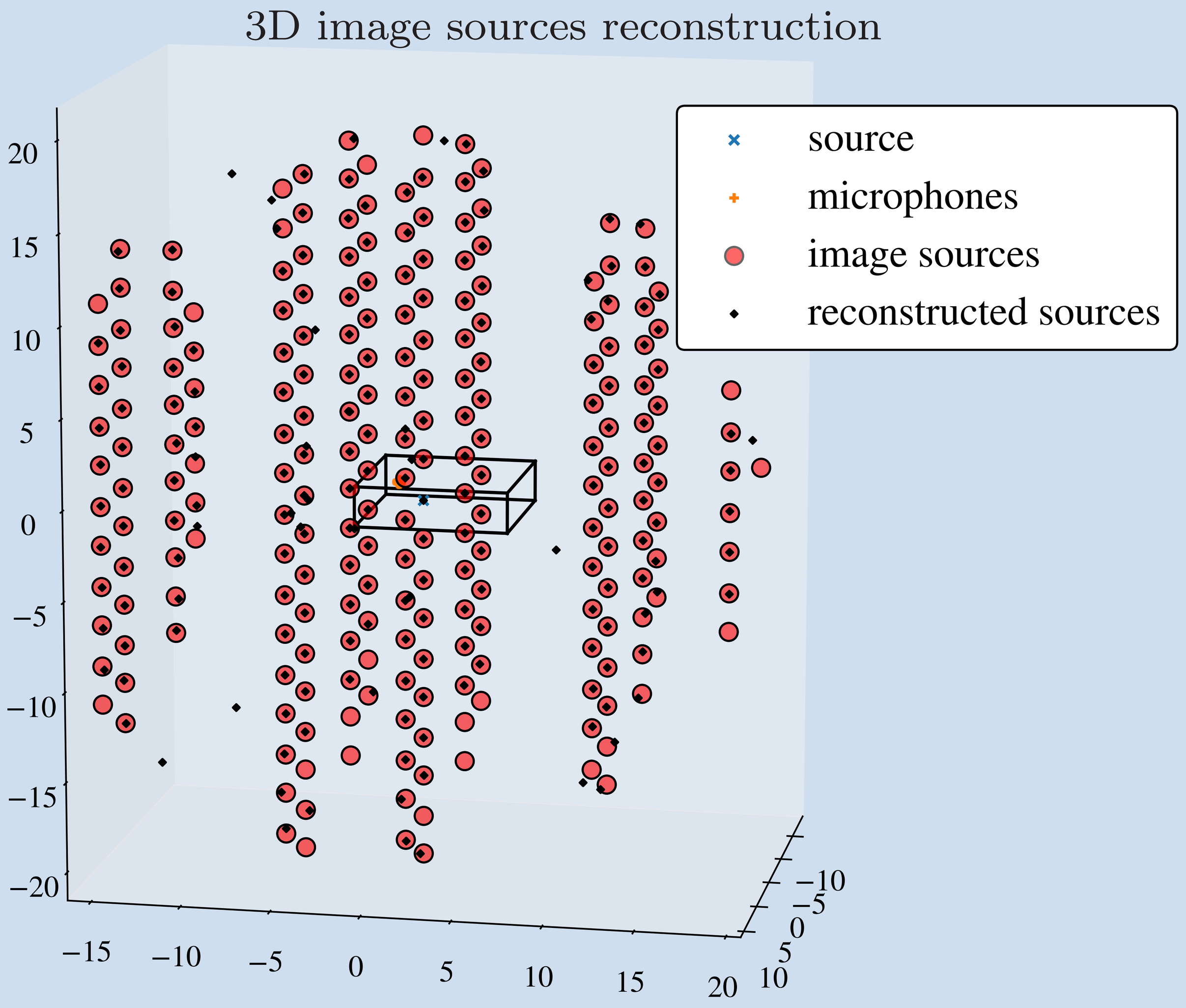
$$\min_{\psi\in\mathcal{M}(\mathbb{R}^3)} \underbrace{\frac{1}{2}\|\mathbf{x}-\Gamma\psi\|_2^2}_{\text{data compliance}} + \underbrace{\lambda\|\psi\|_{\text{TV}}}_{\text{regularization}} \qquad \text{(BLASSO)}$$

7. Example of reconstruction

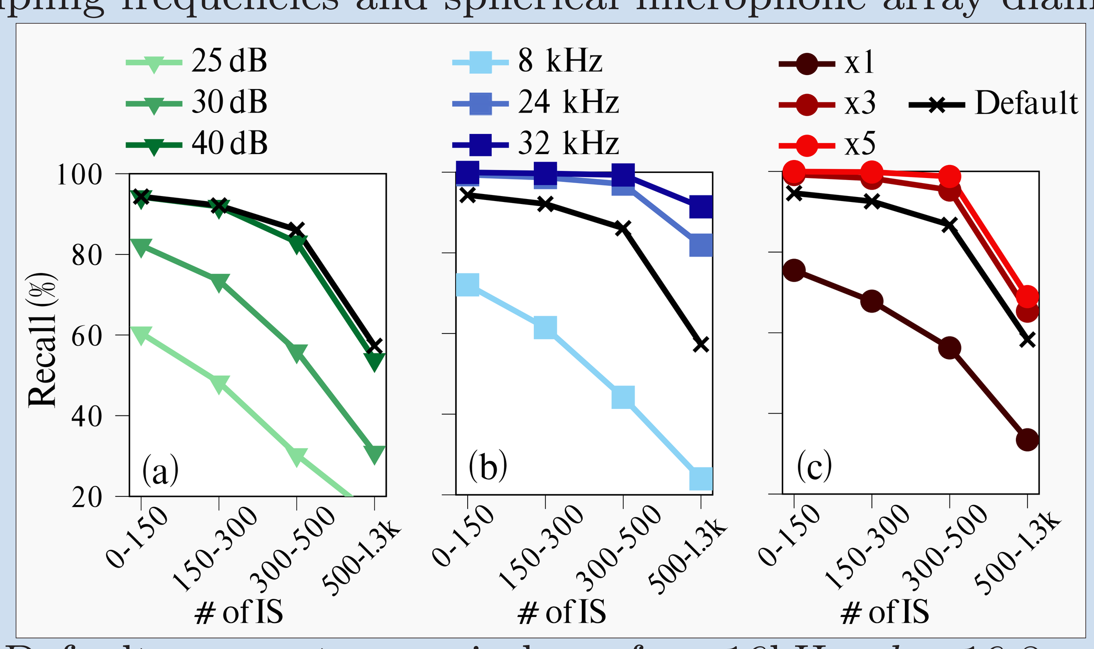
One synthetic noisy RIR and its reconstruction



3D image sources reconstruction



Recall over a RIR dataset for varying noise ratios (PSNR), sampling frequencies and spherical microphone array diameter



Default parameters: noiseless, $f_s = 16$kHz, $d = 16.8$cm
Recall thresholds : 2° angular error, 1cm radial error

2. The Image Source model

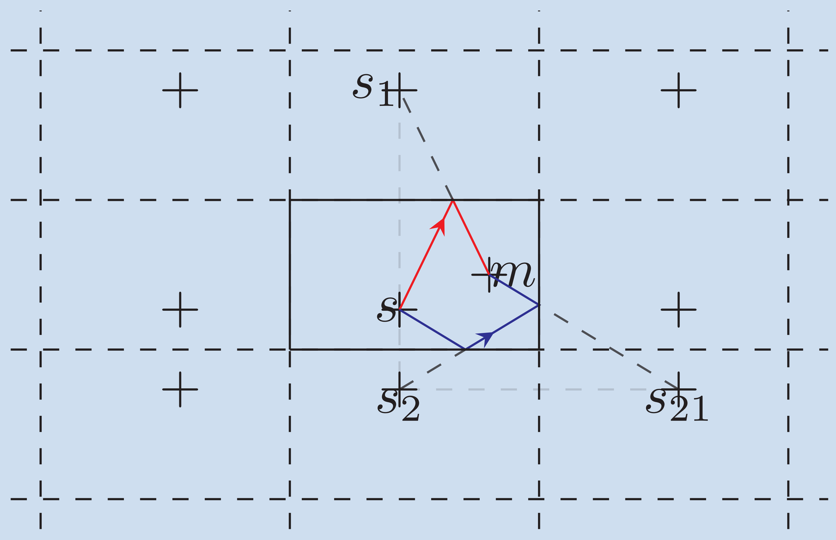
The Room Impulse Response (RIR) is the pressure field $p(\mathbf{r}, t)$ resulting from an impulse source term in the wave equation with boundary conditions :

$$\begin{cases} \frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = a_0\delta_{\mathbf{r}^{\text{src}}}(\mathbf{r})\delta_0(t) & \mathbf{r}\in\Omega \\ \mathbf{n}(\mathbf{r})\cdot\nabla p(\mathbf{r},t) + \frac{\partial}{\partial t}\beta(\mathbf{r},t) * p(\mathbf{r},t) = 0 & \mathbf{r}\in\partial\Omega \end{cases} \qquad (1)$$

In the framework of the image source model and when considering frequency independant walls, the pressure field in the room can be approximated by solving a free field, inhomogeneous equation [1]:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \sum_{k=0}^{+\infty} a_k \delta_{\mathbf{r}_k^{\text{src}}}(\mathbf{r})\,\delta_0(t). \qquad (2)$$

- Each image source corresponds to a sound reflection path
- The image sources are constructed iteratively by taking successive reflections of the original source with respect to the walls



The first order sources yield the geometry.

5. Image Source Recovery with Super-Resolution

Consider the wave equation with source term ψ:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \psi(\mathbf{r})\delta_0(t) \qquad (3)$$

The multi-channel measurements $\mathbf{x}$ of p by the microphones are given by:

$$x_{m,n} := (\kappa * p(\mathbf{r}_m^{mic}, \cdot))(n/f_s) = \int_{\mathbf{r}\in\mathbb{R}^3} \frac{\kappa\left(n/f_s - \|\mathbf{r}_m^{\text{mic}} - \mathbf{r}\|_2/c\right)}{4\pi\|\mathbf{r}_m^{\text{mic}} - \mathbf{r}\|_2}\psi(\mathbf{r})\,d\mathbf{r} \qquad (4)$$

where $\kappa : t \mapsto \text{sinc}(\pi f_s t)$ is an ideal low-pass filter.
In practice the sources remain at a minimum distance from the microphones and we define the observation operator:

$$\begin{array}{rccl} \Gamma^\varepsilon: & \mathcal{M}(\mathbb{R}^3_\varepsilon) & \longrightarrow & \mathbb{R}^{M(N+1)} \\ & \psi & \mapsto & \mathbf{x} = \left(\int_{r\in\mathbb{R}^3_\varepsilon} \frac{\kappa(n/f_s - \|r - r_m^{\text{mic}}\|_2/c)}{4\pi\|r - r_m^{\text{mic}}\|_2} d\psi(r)\right)_{\substack{1\le m\le M \\ 0\le n\le N}} \end{array} \qquad (5)$$

where M is the number of microphones, $N+1$ the number of time samples and $\mathbb{R}^3_\varepsilon = \mathbb{R}^3 \setminus \bigcup_m B(r_m^{\text{mic}}, \varepsilon)$, $\varepsilon > 0$.

6. Adapted Sliding Frank-Wolfe algorithm [3, 4]

k-th iteration: let $\psi^k = \sum_i^{N_k} a_i^k \delta_{\mathbf{r}_i^k}$ ($\mathbf{r}_i^k$ pairwise distinct):

1. Spike-finding. Find $\mathbf{r}_*^k \in \text{argmax}_{\mathbf{r}\in\mathbb{R}^3_\varepsilon} |\eta_\lambda^k(\mathbf{r})|$, $\eta_\lambda^k(\mathbf{r}) = \frac{1}{\lambda}\Gamma^*(\mathbf{x} - \Gamma\psi^k)(\mathbf{r})$. If $\|\eta_\lambda^k(\mathbf{r})\|_2 \le 1$, stop.
2. Amplitude optimization. Find $\psi^{k+1} = \sum_i^{N_k} a_i^{k+1}\delta_{\mathbf{r}_i^k} + a_{N_k+1}^{k+1}\delta_{\mathbf{r}_*^k}$ solving
$$\inf_{a_i^{k+1}\ge 0} \frac{1}{2}\|\mathbf{x} - \Gamma\psi^{k+1}\|_2^2 + \lambda\|\psi^{k+1}\|_{\text{TV}}.$$
3. Remove Dirac masses with zero amplitudes from ψ^{k+1}.

Last step (sliding): find ψ^* minimizing locally the criterion wrt. $(a, \mathbf{r})$ using as initial point $(a^{k_{\max}}, \mathbf{r}^{k_{\max}})$.

8. Conclusion

We observe a high recovery rate of low order image sources and consequently of the wall locations on synthetic data. The model and recovery method must be extended to non-rectangular, frequency dependant rooms and to unknown source and receiver responses to cover real applications.

References

[1] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65:943–950, 1976.
[2] Vincent Duval and Gabriel Peyré. Exact Support Recovery for Sparse Spikes Deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.
[3] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies. The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy. *Inverse Problems*, 2019.
[4] Pierre-Jean Bénard, Yann Traonmilin, and Jean-Francois Aujol. Fast off-the-grid sparse recovery with over-parametrized projected gradient descent. In *EUSIPCO*, pages 2206–2210, 2022.