MULAN: A Blind and Off-Grid Method for Multichannel Echo Retrieval

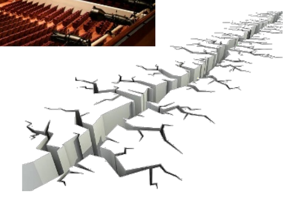
Helena Peić Tukuljac, Antoine Deleforge and Rémi Gribonval

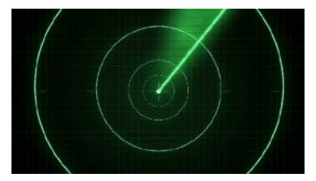
Problem definition: Blind echo retrieval

Given ***M* sensors** measuring delayed and attenuated copies of an ***unknown*** source signal in the **discrete-time** domain, can the echo locations and weights be recovered?

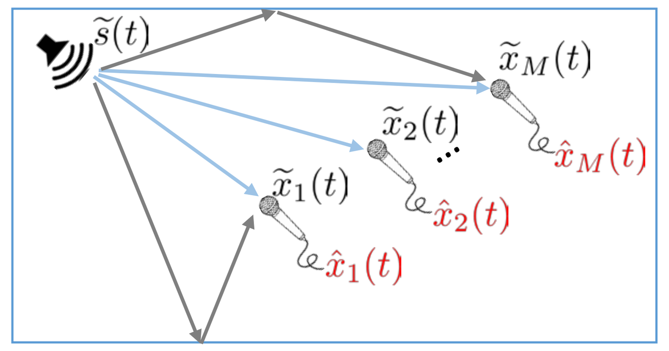
Applications:

- Room Acoustics
- Audio Signal Processing
- Seismology
- Passive sonars …





Measurements in continuous domain

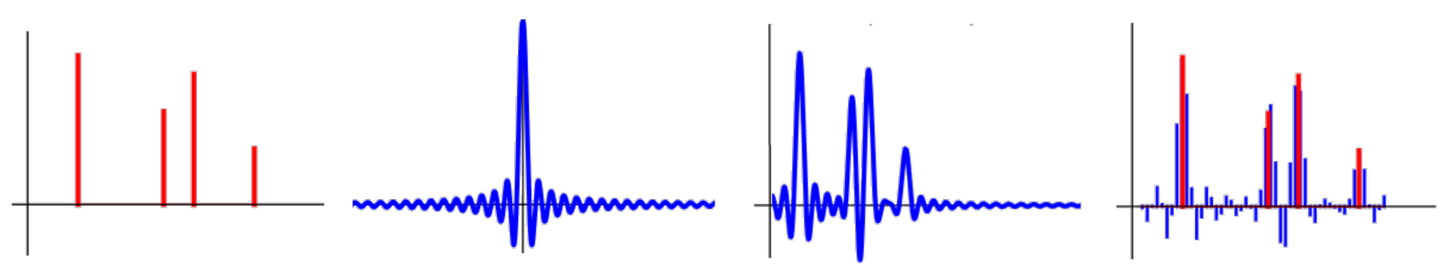


$$\tilde{x}_m(t) = (\tilde{h}_m * \tilde{s})(t)$$

$$\tilde{h}_m(t) = \sum_{k=1}^{K} c_{m,k}\delta(t - \tau_{m,k})$$

Measurements in discrete domains

Time: $\hat{\mathbf{x}}_m(n) \approx (\hat{\mathbf{h}}_m \star \hat{\mathbf{s}})(n)$ where $\begin{cases} \hat{h}_m = (\tilde{\phi} * \tilde{h}_m)(n/F_s), \\ \hat{s}(n) = \tilde{s}(n/F_s), n \in \mathbb{Z} \end{cases}$



$\tilde{h}_m(t)$ $\quad\tilde{\phi}\quad$ $(\tilde{\phi} * \tilde{h}_m)(t)$ $\quad\tilde{h}_m(t)$ vs. $\hat{h}_m(n)$

Classical methods in two steps:

1. Blind estimation of discrete-time filters
2. Echo retrieval by peak-picking on filters

Cross-relation (CR) [1]

$$\underset{\hat{h}_1(1)=1,\hat{h}_2}{\operatorname{argmin}} \left\| \hat{\boldsymbol{h}}_1 \star \hat{\boldsymbol{x}}_2 - \hat{\boldsymbol{h}}_2 \star \hat{\boldsymbol{x}}_1 \right\|_2^2 = \left\| Toep(\hat{\boldsymbol{x}}_2)\hat{\boldsymbol{h}}_1 - Toep(\hat{\boldsymbol{x}}_1)\hat{\boldsymbol{h}}_2 \right\|_2^2$$

Constrained LASSO (LASSO) [2]

$$\underset{\hat{h}_1(1)=1,\hat{h}_2}{\operatorname{argmin}} \left\| Toep(\hat{\boldsymbol{x}}_2)\hat{\boldsymbol{h}}_1 - Toep(\hat{\boldsymbol{x}}_1)\hat{\boldsymbol{h}}_2 \right\|_2^2 + \lambda(\|\hat{\boldsymbol{h}}_1\|_1 + \|\hat{\boldsymbol{h}}_2\|_1)$$

Limits:

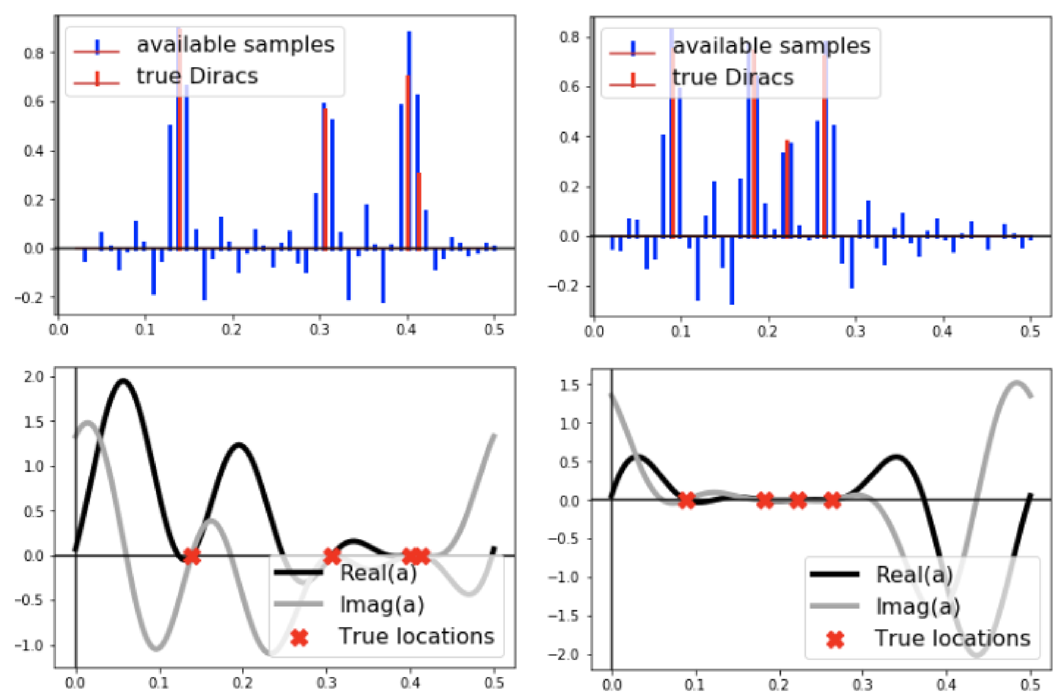
1. Requires filter sizes but true filters are infinite
2. True echoes are off-grid
3. Dirac energy spillage (basis mismatch [3])

Frequency: $x_m(f) \approx h_m(f)s(f) \approx \left(\sum_{k=1}^{K} c_{m,k}e^{-2\pi i f \tau_{m,k}}\right)s(f)$ where: $x_m(f) = \sum_{n=0}^{N-1} \hat{x}_m(n)e^{-2\pi i f n/F_s}$ (DFT)

Assuming input signal $s(f)$ is **known**, the transfer function $h_m(f)$ is given by:

$$h_m(f) = z(f)x_m(f) = \sum_{k=1}^{K} c_{m,k}e^{-2\pi f \tau_{m,k}} \quad (1) \quad \text{where: } z(f) = \frac{1}{s(f)}$$

Non-blind case with finite rate of innovation [4] and annihilating filters



- $\{h_m(f)\}_{f=f_0,f_0+\Delta_f,f_0+2\Delta_f,\ldots}$ is a sum of geometrical series with ratios $r_{m,k} = e^{-2\pi i \Delta_f \tau_{m,k}}$
- We have: $[1, -w] \star [w^0, w^1, w^2, \ldots, w^{F-1}] = \mathbf{0}_{F-1}$ for any $w \in \mathbb{C}$
- Hence: $\mathbf{a}_m \star \mathbf{h}_m = \mathbf{0}_{F-K}$ for $\boldsymbol{a}_m \equiv [1, -r_{m,1}] \star [1, -r_{m,2}] \star \ldots [1, -r_{m,K}] \in \mathbb{C}^{K+1}$
- Let $P_{\boldsymbol{a}_m}[y] = \sum_{k=0}^{K} a_{m,k}y^k \Rightarrow$ the roots of P_m are $\{r_{m,k}\}_{k=1}^{K}$!
- From $\boldsymbol{r}_m$ we deduce $\boldsymbol{\tau}_m$ and also $\boldsymbol{c}_m$ by solving (1)

Blind case with MULAN: MULtichannel ANnihilation

- We consider the following non-convex program:

$$\underset{\|\boldsymbol{z}\|_2^2=\|\boldsymbol{a}_1\|_2^2=\cdots=\|\boldsymbol{a}_M\|_2^2=1}{\operatorname{argmin}} \sum_{m=1}^{M} \|\boldsymbol{a}_m \star (\boldsymbol{x}_m \odot \boldsymbol{z})\|_2^2 = \sum_{m=1}^{M} \|Toep(\boldsymbol{x}_m \odot \boldsymbol{z})\boldsymbol{a}_m\|_2^2 = \sum_{m=1}^{M} \|Toep_0(\boldsymbol{a}_m)Diag(\boldsymbol{x}_m)\boldsymbol{z}\|_2^2 = C(\boldsymbol{a}, \boldsymbol{z})$$

MULAN Algorithm

Input: $\{\boldsymbol{x}_{1:M}(f); f \in \mathcal{F}\}$

Output: $\{\tau_{m,k}, c_{m,k}\}_{m,k=1}^{M,K}$

0. Randomly initialize $\boldsymbol{z}$

Repeat until convergence of $C(\boldsymbol{a}, \boldsymbol{z})$:

1. For each m update $\boldsymbol{a}_m$:

$$\boldsymbol{a}_m \leftarrow \underset{\|\boldsymbol{a}_m\|_2^2=1}{\operatorname{argmin}} \|Toep(\boldsymbol{x}_m \odot \boldsymbol{z})\boldsymbol{a}_m\|_2^2 = min\_eig\_vect(\,Toep(\boldsymbol{x}_m \odot \boldsymbol{z}))$$

2. Update $\boldsymbol{z}$:

$$\boldsymbol{z} \leftarrow \underset{\|\boldsymbol{z}\|_2^2=1}{\operatorname{argmin}} \|\mathbf{Q}\boldsymbol{z}\|_2^2 = min\_eig\_vect(\mathbf{Q}), \quad \mathbf{Q} = \begin{bmatrix} Toep_0(\boldsymbol{a}_1)Diag(\boldsymbol{x}_1) \\ Toep_0(\boldsymbol{a}_2)Diag(\boldsymbol{x}_2) \\ \vdots \\ Toep_0(\boldsymbol{a}_M)Diag(\boldsymbol{x}_M) \end{bmatrix}$$

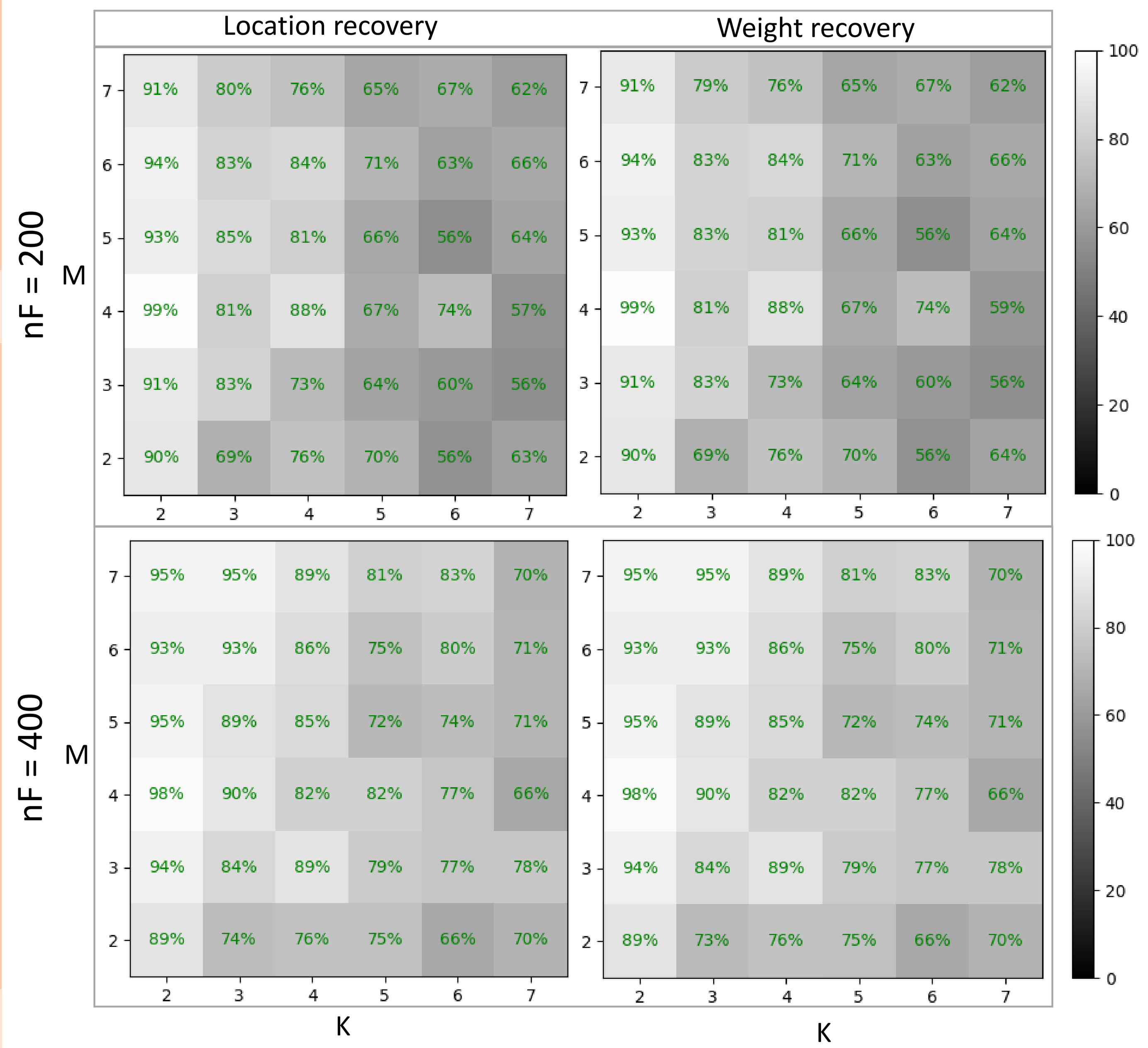
3. From each $\boldsymbol{a}_m$ deduce $\boldsymbol{\tau}_m$ and $\boldsymbol{c}_m$

4. Shift and rescale $\{\tau_{m,k}, c_{m,k}\}_{m,k=1}^{M,K}$

- **Multiple initializations** are used to fight local minima (15 in our experiments)
- Problem solved exactly if $C(\boldsymbol{a}, \boldsymbol{z}) = 0$
- Filters are recovered up to a **global shift** and **scaling**
- Non-identifiability when:
  - Filters have **common roots** (*open question*)
  - Emitted signal has too many zero frequencies
- Need at least $2K + 1$ regularly-spaced frequencies (*open question*)

Results

Percentage of full weight and location recovery vs. (M, K, F) :



Benchmark for *M = 2*, *K = 7* over 100 randomly simulated shoe-box room, microphone pair, and speech source:

case	method	full location recovery	weight RMSE
on-grid	CR [1]	**92** %	0.0390
	LASSO [2]	13 %	0.155
	MULAN (proposed)	59 %	**0.00016**
off-grid	CR [1]	1%	0.0442
	LASSO [2]	2%	0.0346
	MULAN (proposed)	**70** %	**0.00048**

References

[1] Guanghan Xu, Hui Liu, Lang Tong, and Thomas Kailath. A least-squares approach to blind channel identification. Transactions on signal processing, IEEE (1995).
[2] Yuanqing Lin, Jingdong Chen, Youngmoo Kim, and Daniel D. Lee. Blind channel identification for speech dereverberation using l1-norm sparse learning. In Advances in Neural Information Processing Systems, Curran Associates, Inc. (2008).
[3] Yuejie Chi, Louis L. Scharf, Ali Pezeshki, and A. Robert Calderbank. Sensitivity to basis mismatch in compressed sensing. Transactions on Signal Processing, IEEE (2011).
[4] Martin Vetterli, Pina Marziliano, and Thierry Blu. Sampling signals with finite rate of innovation. Transactions on Signal Processing, IEEE (2002).
[5] Yuejie Chi. Guaranteed blind sparse spikes deconvolution via lifting and convex optimization. Journal of Selected Topics in Signal Processing, IEEE (2016).

Conclusion

- Near-exact **blind** and **off-grid** echo retrieval from multichannel **discrete-time** measurements directly in the parameter space
- Future work: robustness to observation noise, filter mismatch and missing frequencies, convex relaxation of the program [5] or spectral initialization.