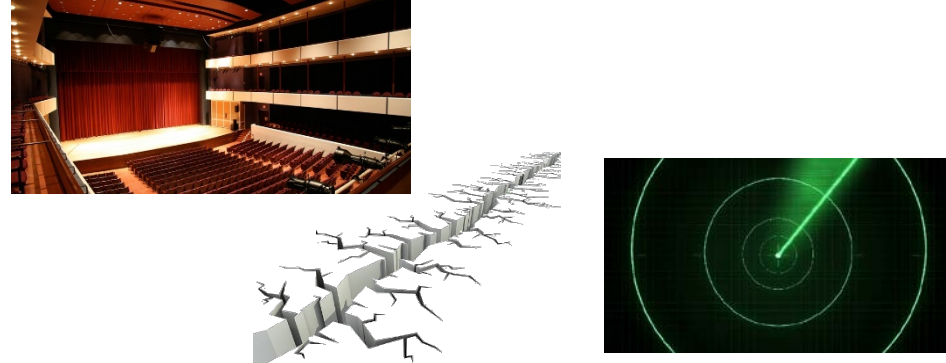


Problem definition: Blind echo retrieval

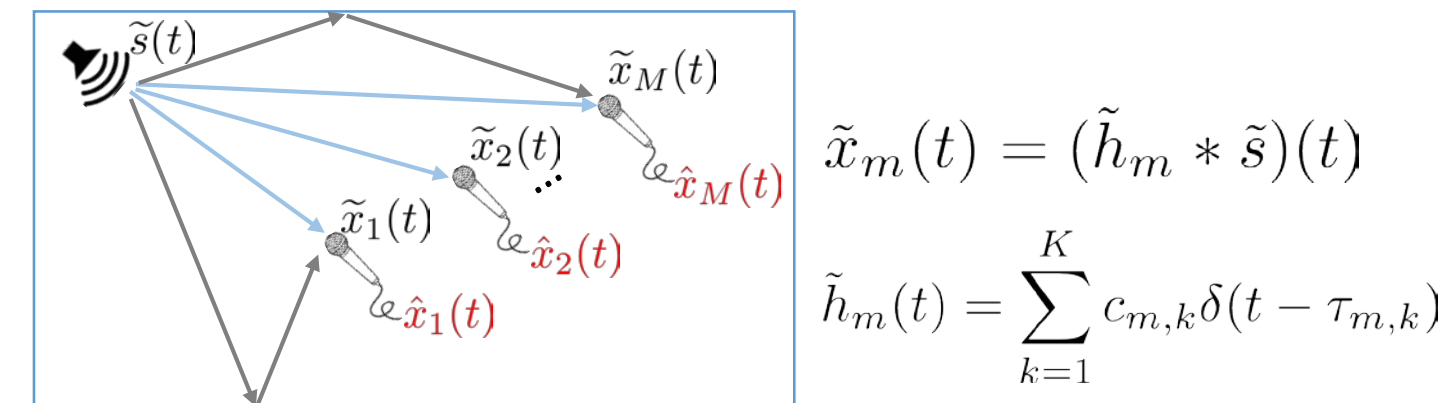
Given **M sensors** measuring delayed and attenuated copies of an ***unknown*** source signal in the **discrete-time** domain, can the echo locations and weights be recovered?

Applications:

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

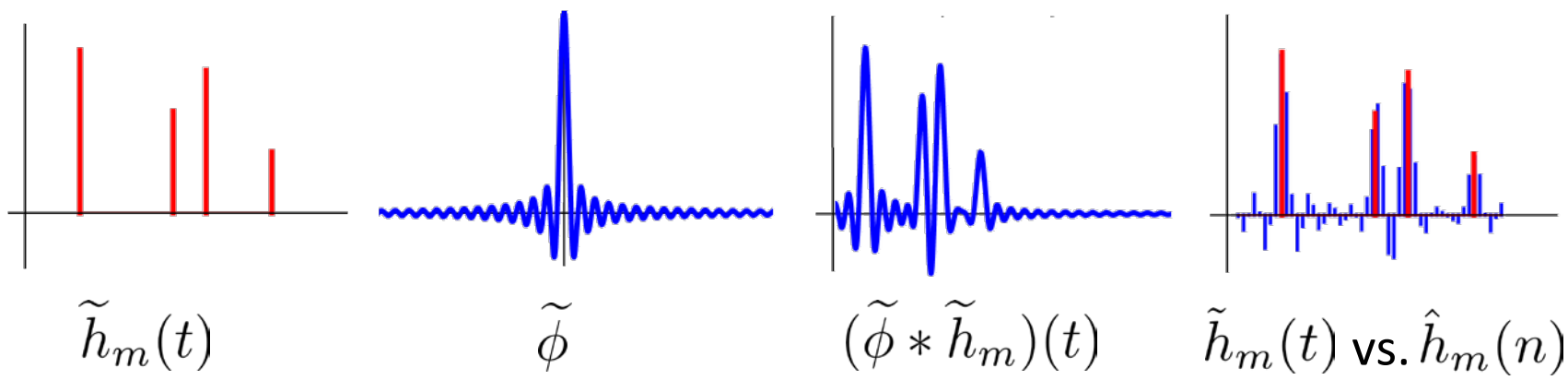


Measurements in continuous domain



Measurements in discrete domains

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



Classical methods in two steps:

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

Cross-relation (CR) [1]

$$\argmin_{\hat{h}_1(1)=1, \hat{h}_2} \left\| \hat{h}_1 \star \hat{x}_2 - \hat{h}_2 \star \hat{x}_1 \right\|_2^2 = \left\| \text{Toep}(\hat{x}_2) \hat{h}_1 - \text{Toep}(\hat{x}_1) \hat{h}_2 \right\|_2^2$$

Constrained LASSO (LASSO) [2]

$$\argmin_{\hat{h}_1(1)=1, \hat{h}_2} \left\| \text{Toep}(\hat{x}_2) \hat{h}_1 - \text{Toep}(\hat{x}_1) \hat{h}_2 \right\|_2^2 + \lambda (\|\hat{h}_1\|_1 + \|\hat{h}_2\|_1)$$

Limits:

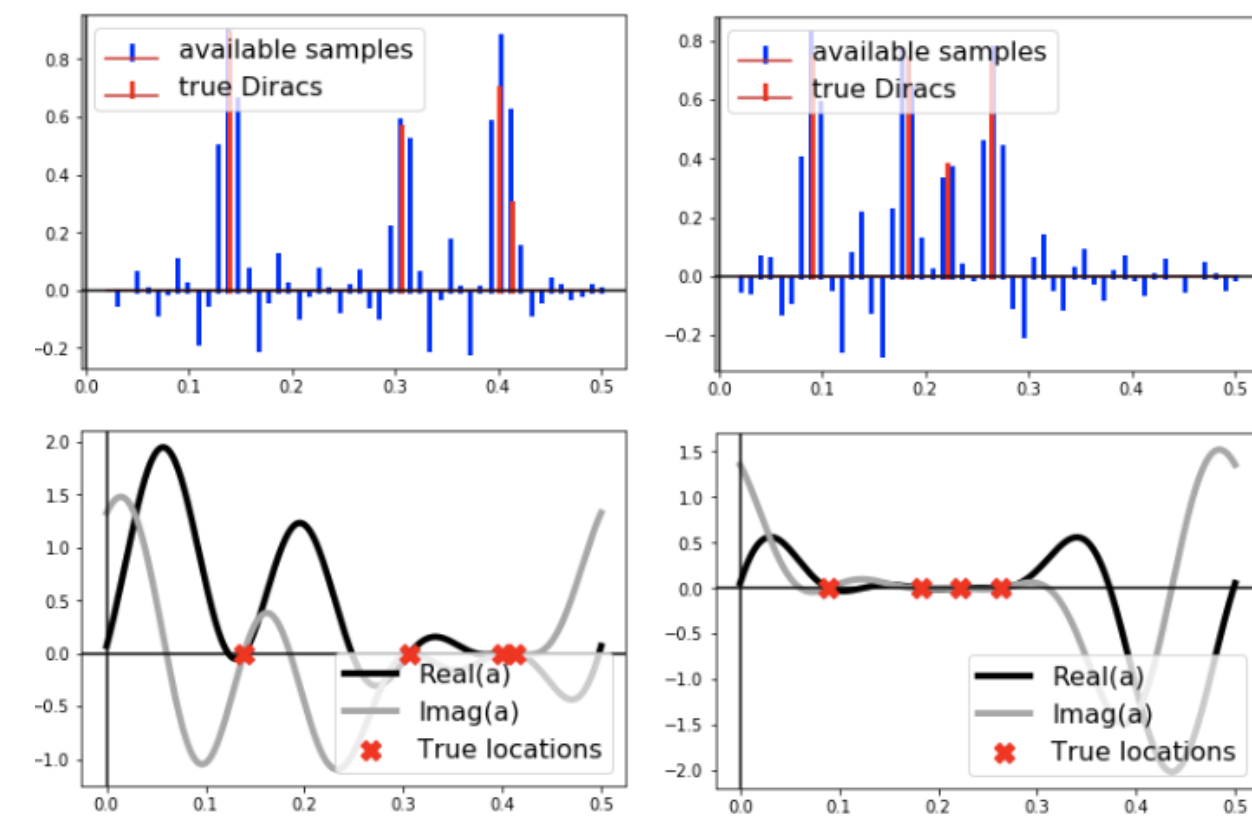
1. Require filter sizes but true filters are infinite
2. True echoes are off-grid
3. Dirac energy spillage (basis mismatch [3])

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:

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

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



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

Blind case with MULAN: MULTichannel ANnihilation

- We consider the following non-convex program:

$$\argmin_{\|\mathbf{z}\|_2^2 = \|\mathbf{a}_1\|_2^2 = \dots = \|\mathbf{a}_M\|_2^2 = 1} \sum_{m=1}^M \|\mathbf{a}_m \star (\mathbf{x}_m \odot \mathbf{z})\|_2^2 = \sum_{m=1}^M \|\text{Toep}(\mathbf{x}_m \odot \mathbf{z}) \mathbf{a}_m\|_2^2 = \sum_{m=1}^M \|\text{Toep}_0(\mathbf{a}_m) \text{Diag}(\mathbf{x}_m) \mathbf{z}\|_2^2 = C(\mathbf{a}, \mathbf{z})$$

MULAN Algorithm

Input: $\{\mathbf{x}_{1:M}(f); f \in \mathcal{F}\}$
Output: $\{\tau_{m,k}, c_{m,k}\}_{m,k=1}^{M,K}$

0. Randomly initialize \mathbf{z}
- Repeat** until convergence of $C(\mathbf{a}, \mathbf{z})$:

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

$$\mathbf{a}_m \leftarrow \argmin_{\|\mathbf{a}_m\|_2^2=1} \|\text{Toep}(\mathbf{x}_m \odot \mathbf{z}) \mathbf{a}_m\|_2^2 = \min_eig_vect(\text{Toep}(\mathbf{x}_m \odot \mathbf{z}))$$
 2. Update \mathbf{z} :

$$\mathbf{z} \leftarrow \argmin_{\|\mathbf{z}\|_2^2=1} \|\mathbf{Q} \mathbf{z}\|_2^2 = \min_eig_vect(\mathbf{Q}), \quad \mathbf{Q} = \begin{bmatrix} \text{Toep}_0(\mathbf{a}_1) \text{Diag}(\mathbf{x}_1) \\ \text{Toep}_0(\mathbf{a}_2) \text{Diag}(\mathbf{x}_2) \\ \vdots \\ \text{Toep}_0(\mathbf{a}_M) \text{Diag}(\mathbf{x}_M) \end{bmatrix}$$
3. From each \mathbf{a}_m deduce τ_m and 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 (20 in our experiments)

- Problem solved exactly if $C(\mathbf{a}, \mathbf{z}) = 0$

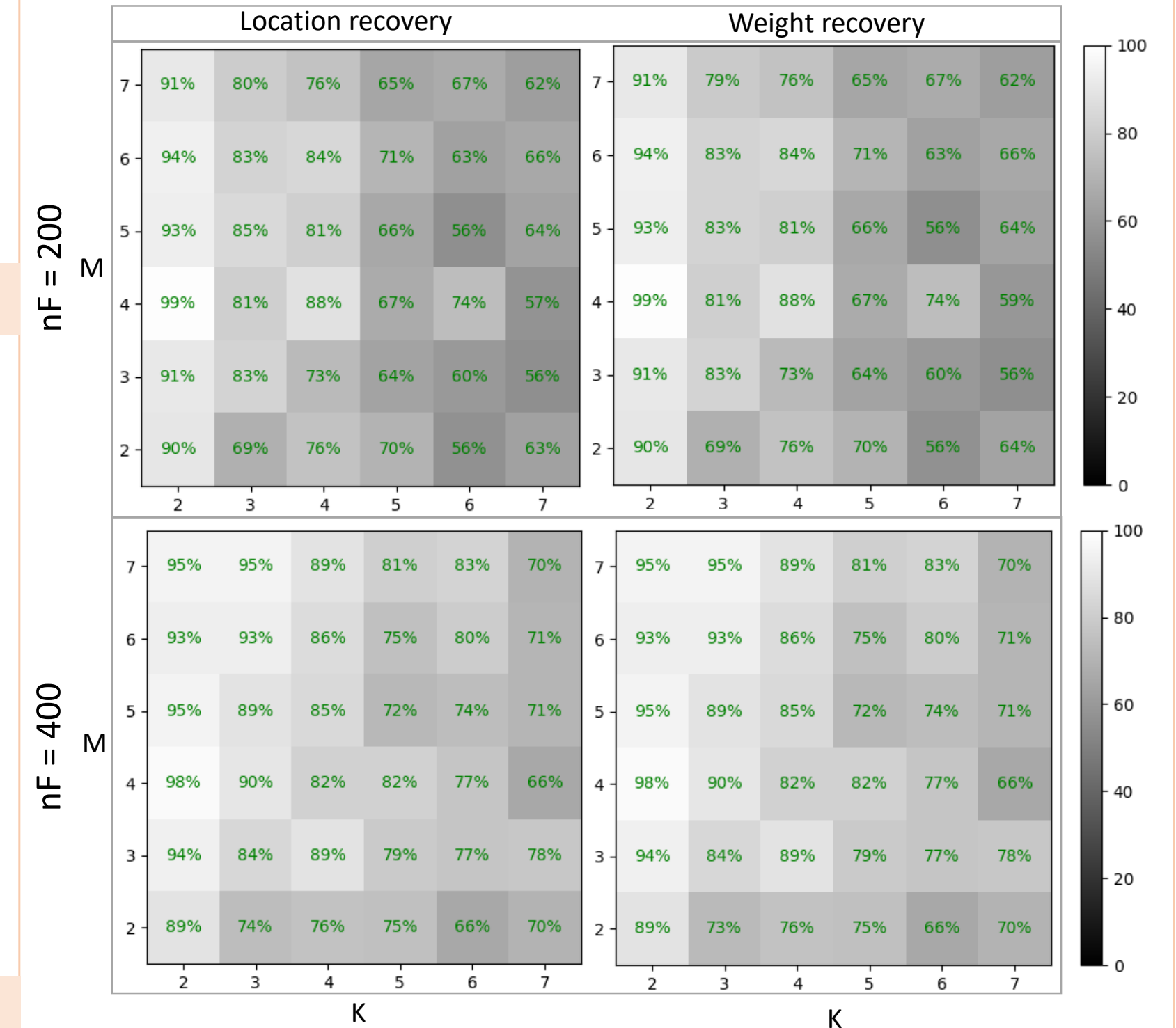
- Filters are recovered up to a **global shift** and **scaling**

- 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 rooms, microphone pair, and speech source:

case	method	full location recovery	weight RMSE
<i>on-grid</i>	CR [1]	92 %	0.0390
	LASSO [2]	13 %	0.155
	MULAN (proposed)	59 %	0.00016
<i>off-grid</i>	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.