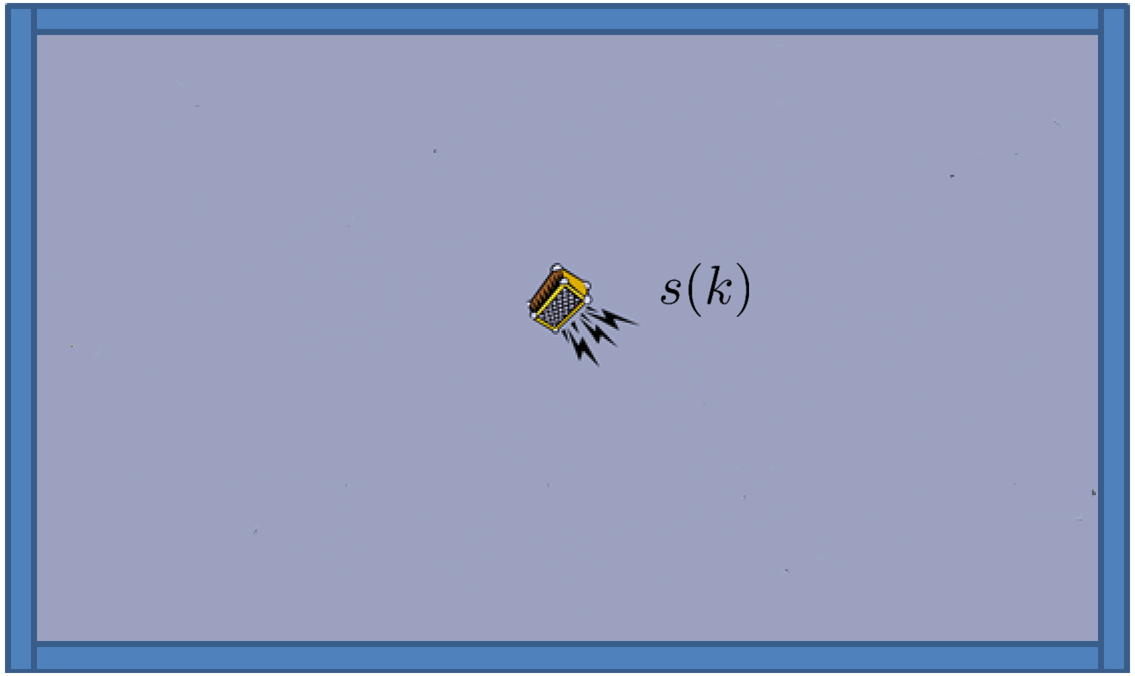
Room Impulse Response Estimation by Iterative Reweighted L_1-Norm

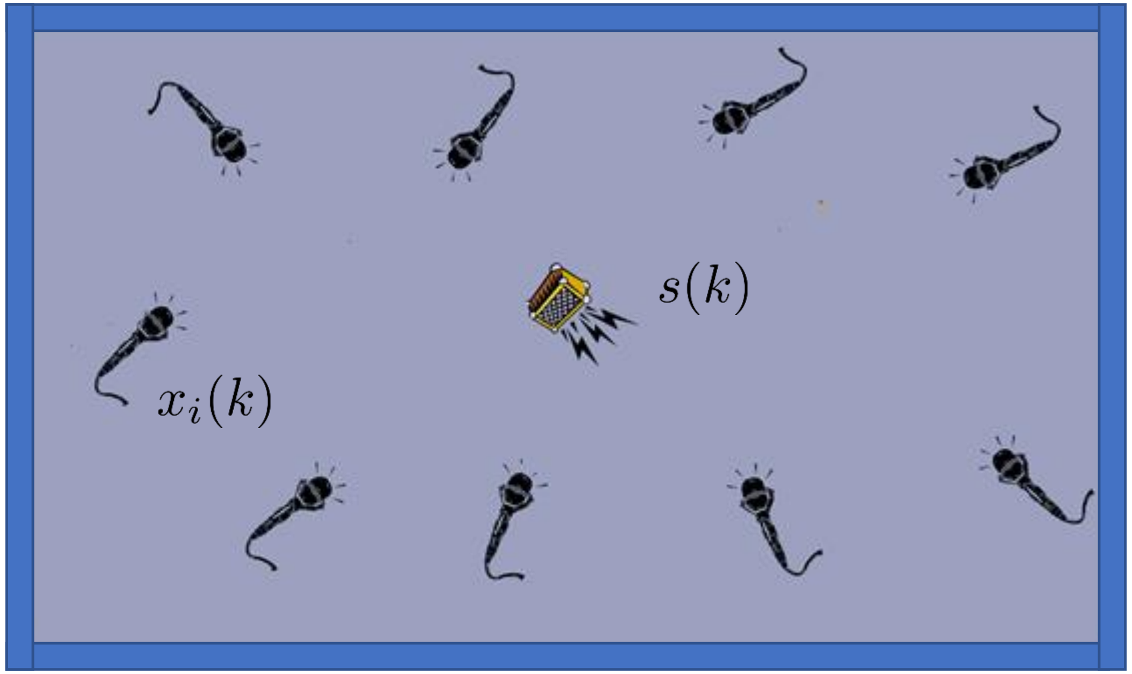
Danilo Greco

Problem statement



$s(k)$: transmitted audio signal

Problem statement

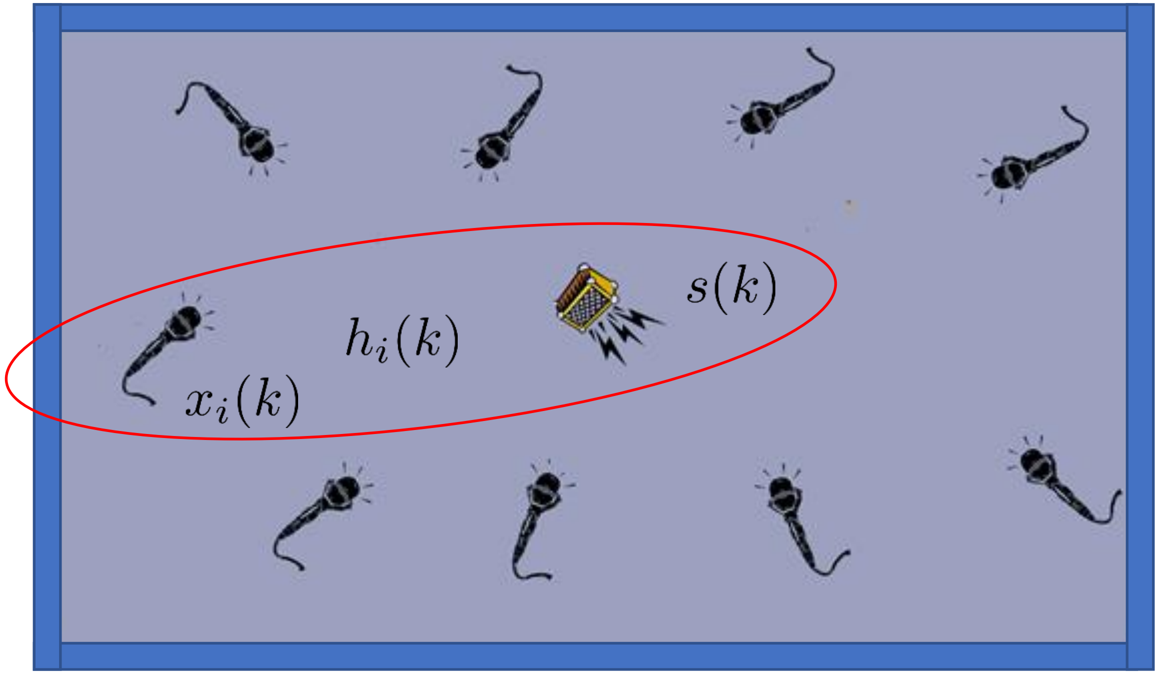


$s(k)$: transmitted audio signal

$x_i(k)$: received signal at *i*-th microphone

Problem statement

$$x_i(k) = h_i(k) * s(k), \qquad i = 1, \cdots, M$$



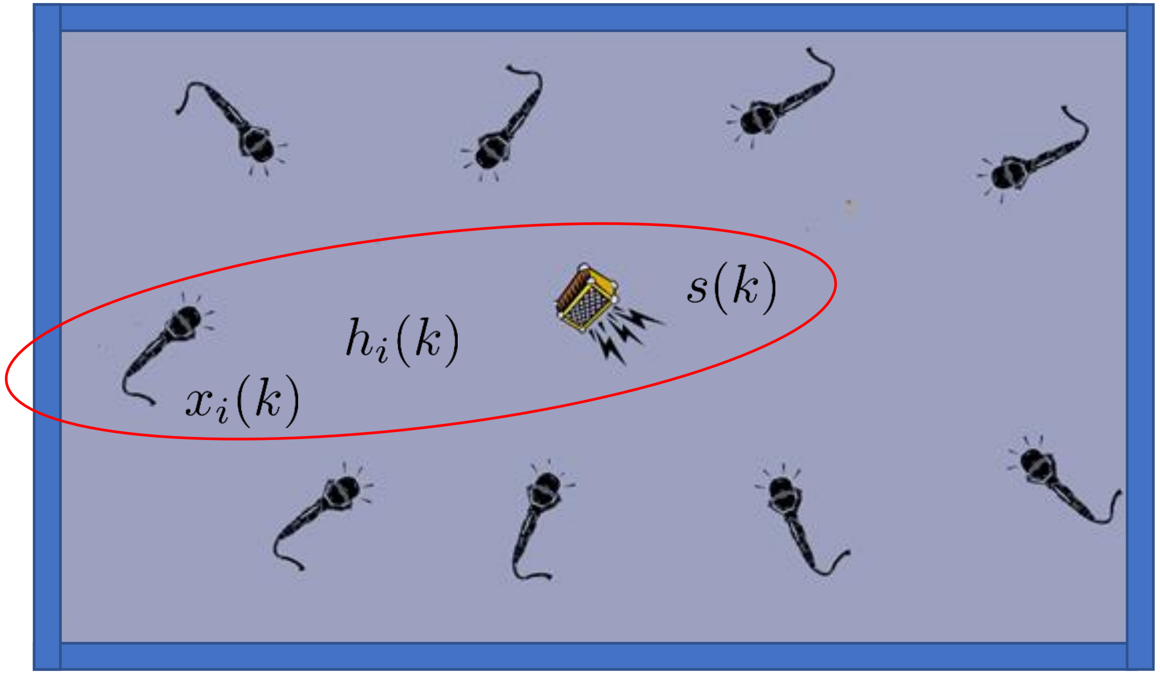
$s(k)$: transmitted audio signal

$x_i(k)$: received signal at *i*-th microphone

$h_i(k)$: room impulse response between source and *i*-th microphone

Problem statement

$$x_i(k) = h_i(k) * s(k), \qquad i = 1, \cdots, M$$



$s(k)$: transmitted audio signal

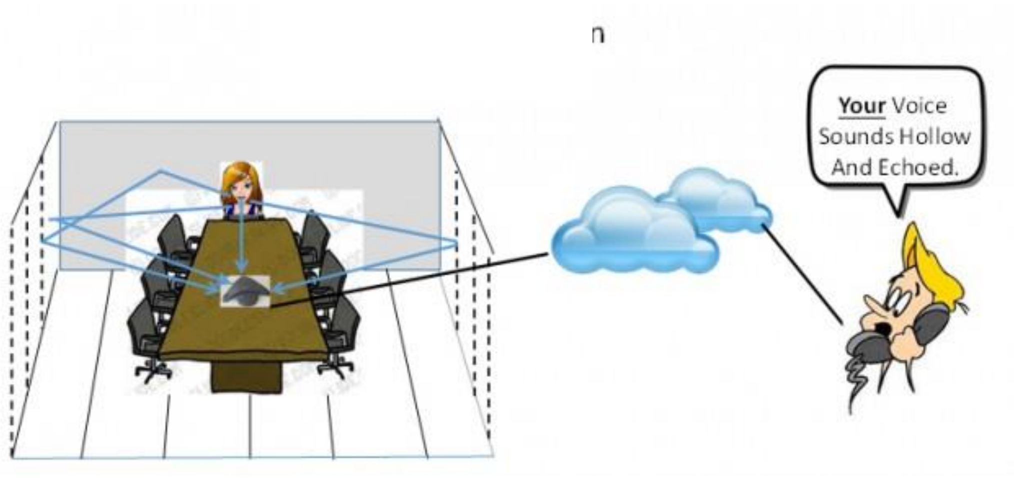
$x_i(k)$: received signal at *i*-th microphone

$h_i(k)$: room impulse response between source and *i*-th microphone

<u>Estimate</u> $h_i(k)$ <u>given</u> $x_i(k)$ for $i = 1, \cdots, M$

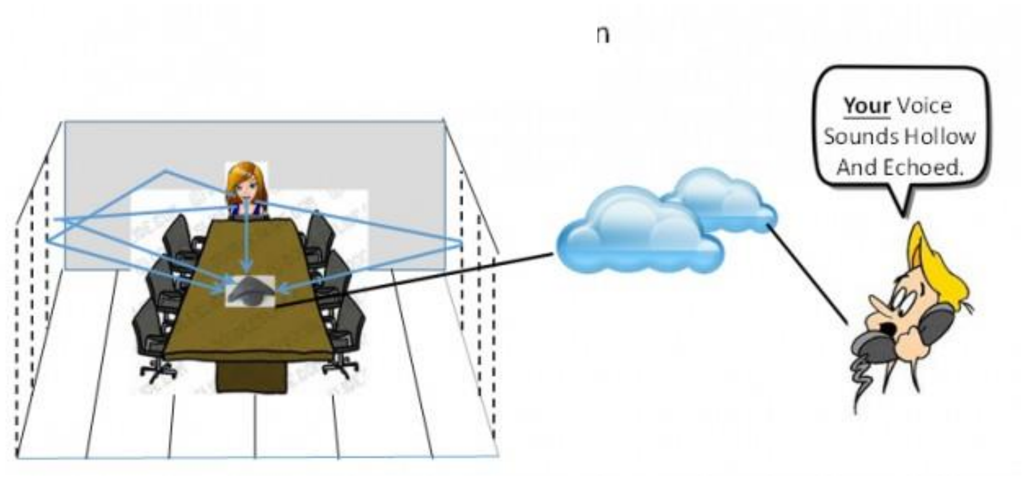
Applications

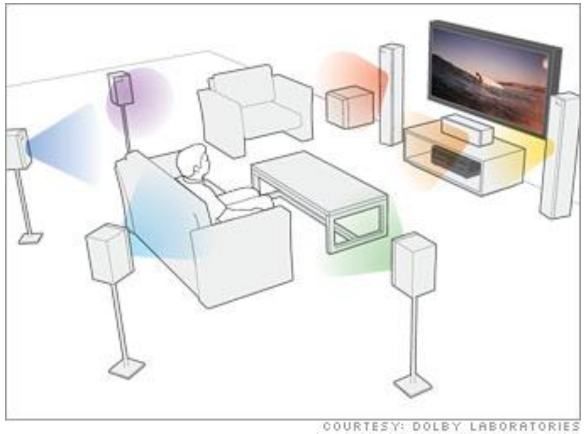
Dereverberation
Speech enhancement



Applications

Dereverberation
Speech enhancement

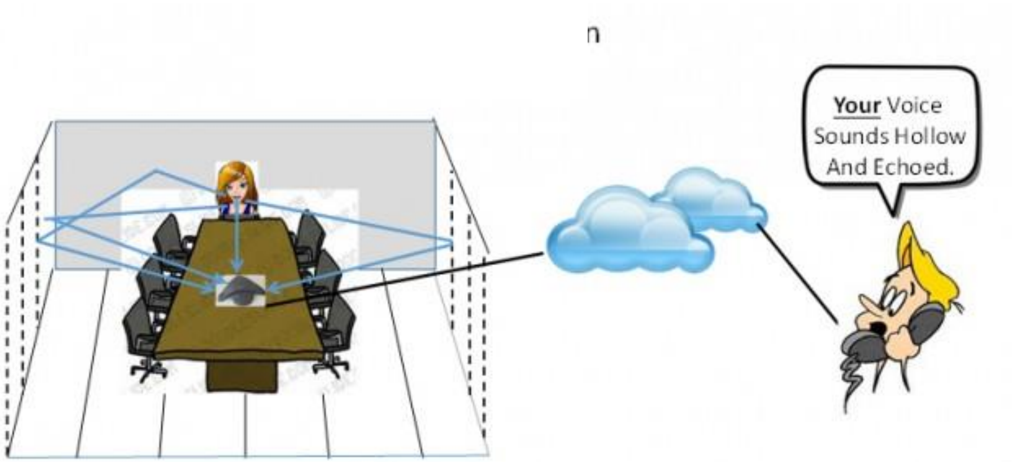


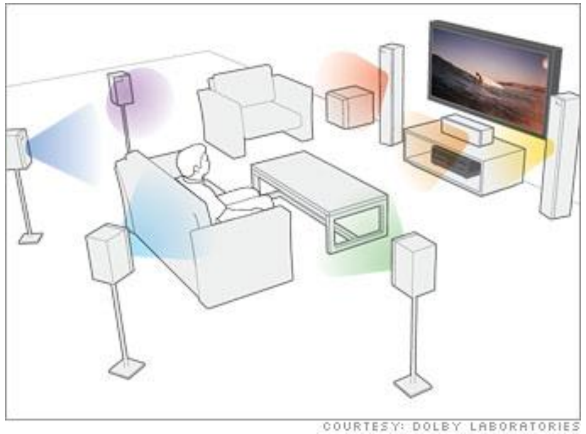


Room aware sound reproduction

Applications

Dereverberation
Speech enhancement





Room aware sound reproduction

Room Geometry estimation
[Dokmanic et al. PNAS 2013]

A short taxonomy on RIR estimation

TX signal $s(k)$

Known

Not known

$$x_i(k) = h_i(k) * s(k), \qquad i = 1, \cdots, M$$

A short taxonomy on RIR estimation

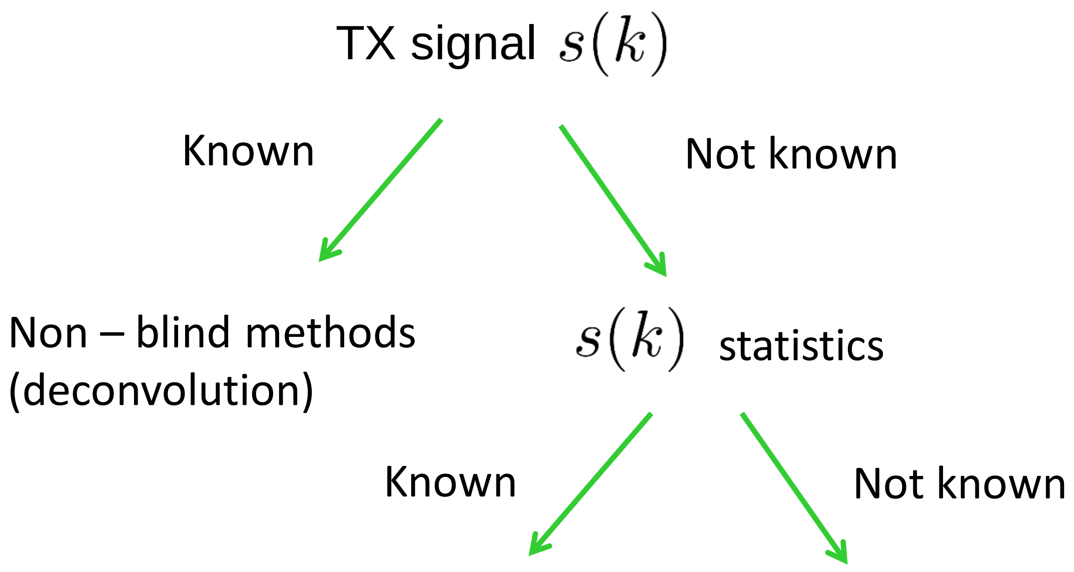
TX signal $s(k)$

Known → Non – blind methods (deconvolution)

Not known →

$$x_i(k) = h_i(k) * s(k), \qquad i = 1, \cdots, M$$

A short taxonomy on RIR estimtion



$$x_i(k) = h_i(k) * s(k), \qquad i = 1, \cdots, M$$

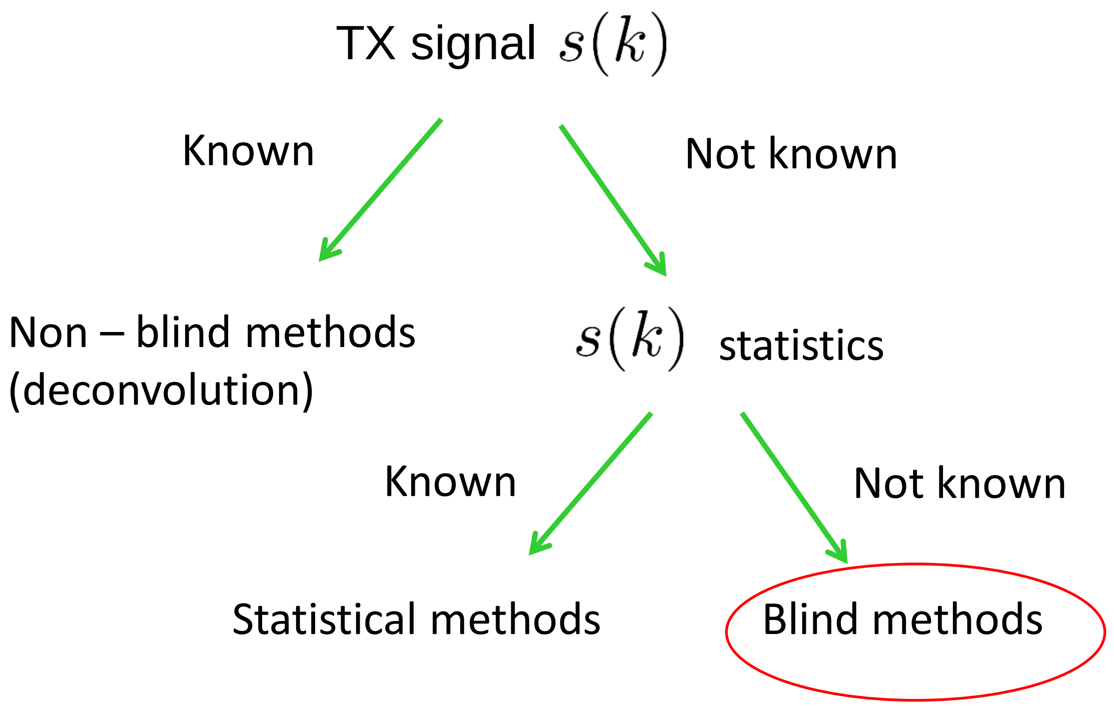
A short taxonomy on RIR estimation

TX signal $s(k)$

- Known → Non – blind methods (deconvolution)
- Not known → $s(k)$ statistics
  - Known → Statistical methods
  - Not known →

$$x_i(k) = h_i(k) * s(k), \qquad i = 1, \cdots, M$$

A short taxonomy on RIR estimation



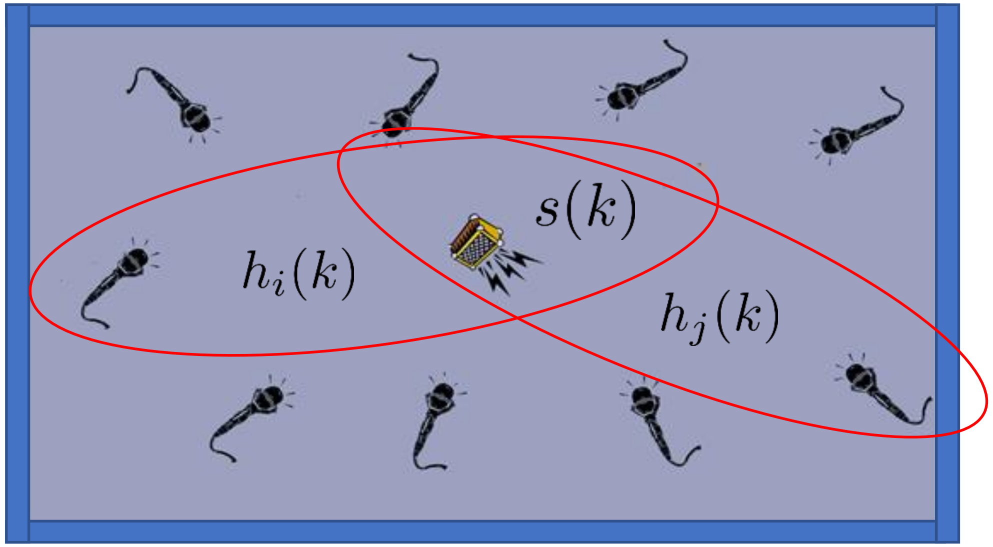
$$x_i(k) = h_i(k) * s(k), \qquad i = 1, \cdots, M$$

Single Input Multi Ouput
Blind Channel Identification

The Cross – Relation Identity (1)

For every couple of microphones:

$$h_i(k) * h_j(k) * s(k) = h_j(k) * h_i(k) * s(k), \qquad i \neq j$$

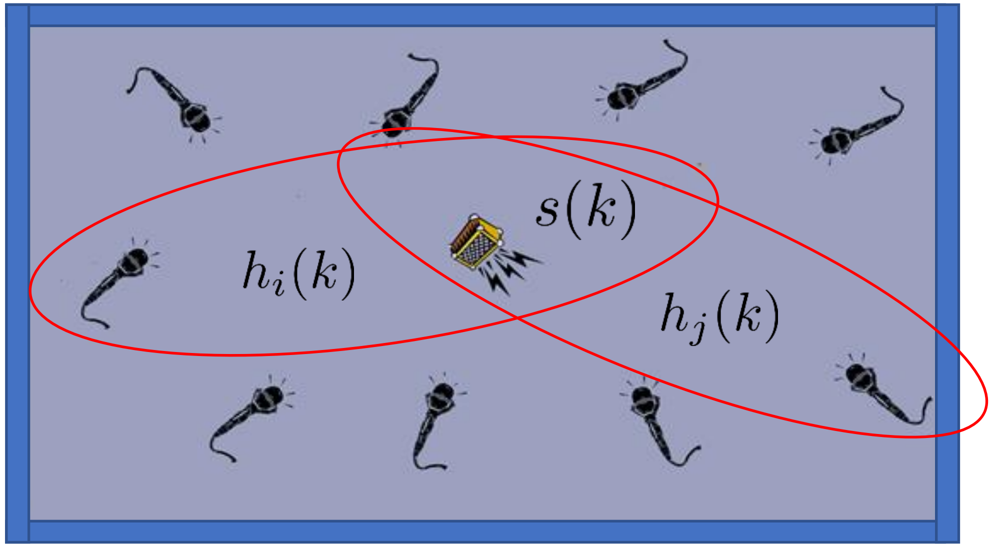


The Cross – Relation Identity (1)

<u>For every couple of microphones:</u>

$$h_i(k) * h_j(k) * s(k) = h_j(k) * h_i(k) * s(k), \qquad i \neq j$$

Swap the indices i, j



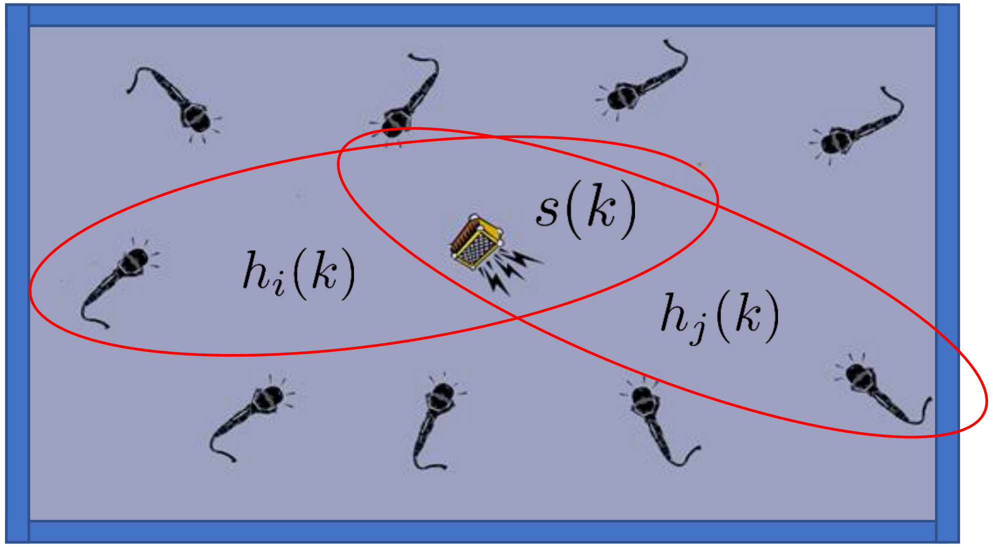
The Cross – Relation Identity (1)

For every couple of microphones:

$$h_i(k) * h_j(k) * s(k) = h_j(k) * h_i(k) * s(k), \qquad i \neq j$$

$$h_i(k) * x_j(k) = h_j(k) * x_i(k)$$

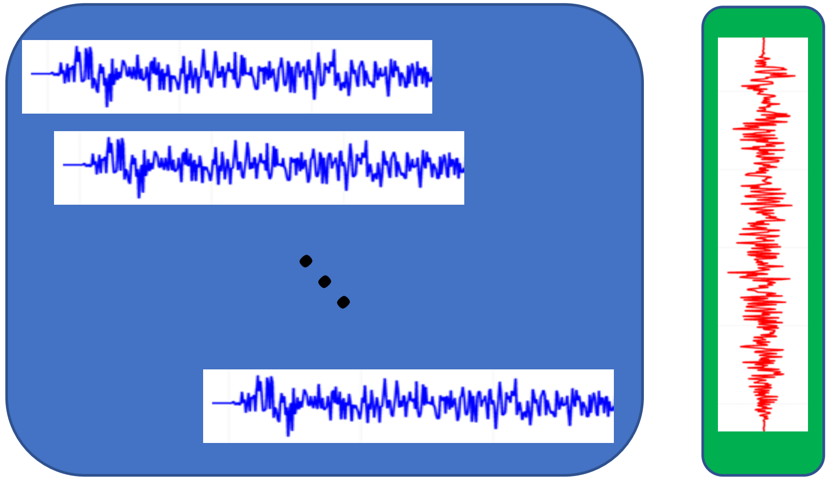
$$i \neq j$$



The Cross – Relation Identity (2)

$$x_j(k) * h_i(k) = \mathbf{X}_j \, \mathbf{h}_i$$

Shift to matrix form

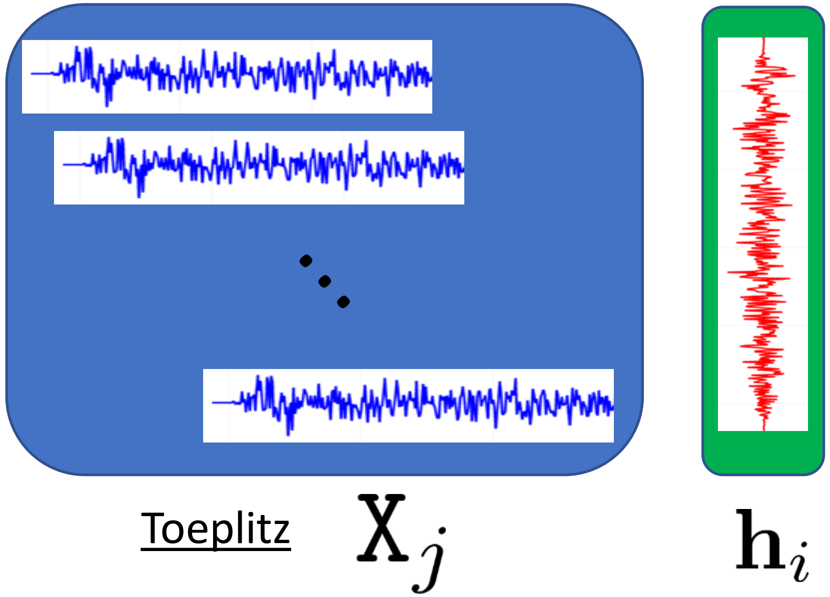


Toeplitz $\mathbf{X}_j$ $\mathbf{h}_i$

The Cross – Relation Identity (2)

$$x_j(k) * h_i(k) = \mathbf{X}_j \, \mathbf{h}_i$$

Shift to matrix form



In absence of noise

$$\mathbf{X}_i \mathbf{h}_j = \mathbf{X}_j \mathbf{h}_i$$

Quadratic cost function

$$J(\mathbf{h}) = \sum_{i \neq j} \|\mathtt{X}_i \mathbf{h}_j - \mathtt{X}_j \mathbf{h}_i\|_2^2 = \mathbf{h}^\top \mathtt{Q} \mathbf{h}$$

$$\underset{\mathbf{h}}{min}\ J(\mathbf{h}) \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1$$

Quadratic cost function

$$J(\mathbf{h}) = \sum_{i \neq j} \|\mathtt{X}_i \mathbf{h}_j - \mathtt{X}_j \mathbf{h}_i\|_2^2 = \mathbf{h}^\top \mathtt{Q} \mathbf{h}$$

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

To avoid the trivial solution

Singular value problem [Tong et al., 1994]:
solution given by the singular vector corresponding to the smallest singular value of $\mathtt{Q}$

[Tong et al., 1994]: L. Tong, G. Xu and T. Kailath. «Blind identification and equalization based on second order statistics: a time domain approach», IEEE Trans. On Information Theory, vol. 40, no. 2, pp.340-349, Mar. 1994..

Quadratic cost function

$$J(\mathbf{h}) = \sum_{i \neq j} \|\mathtt{X}_i \mathbf{h}_j - \mathtt{X}_j \mathbf{h}_i\|_2^2 = \mathbf{h}^\top \mathtt{Q} \mathbf{h}$$

$$\underset{\mathbf{h}}{min}\ J(\mathbf{h}) \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1$$

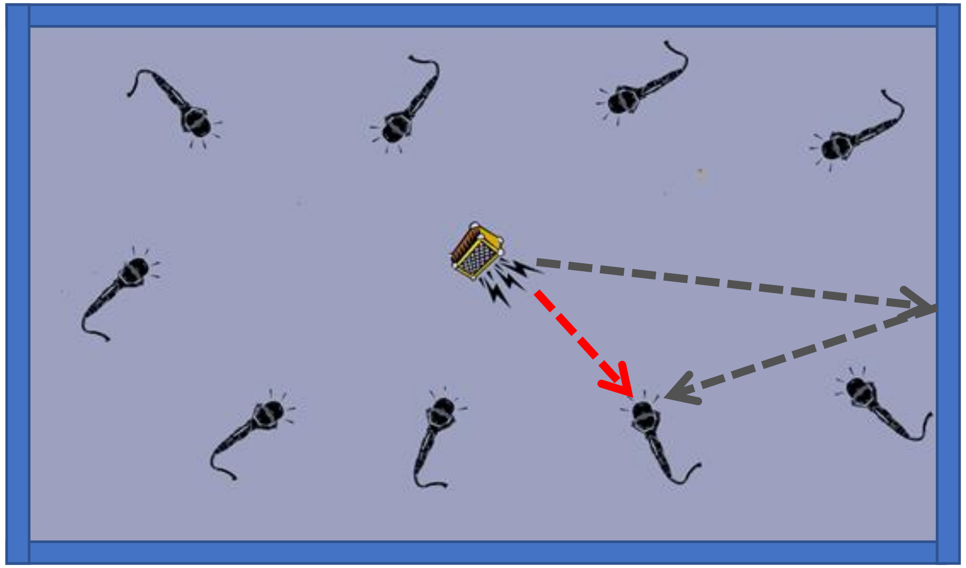
To avoid the trivial solution

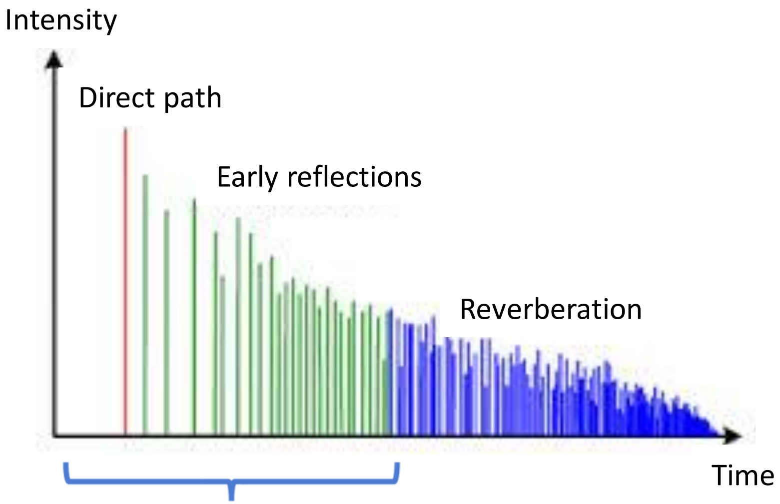
Singular value problem [Tong et al., 1994]:
solution given by the singular vector corresponding to the smallest singular value of $\mathtt{Q}$

Drawbacks:
- Channels must be co-prime
- RIR length must be known
- Sensitivity to «holes» in the signal spectrum

[Tong et al., 1994]: L. Tong, G. Xu and T. Kailath. «Blind identification and equalization based on second order statistics: a time domain approach», IEEE Trans. On Information Theory, vol. 40, no. 2, pp.340-349, Mar. 1994..

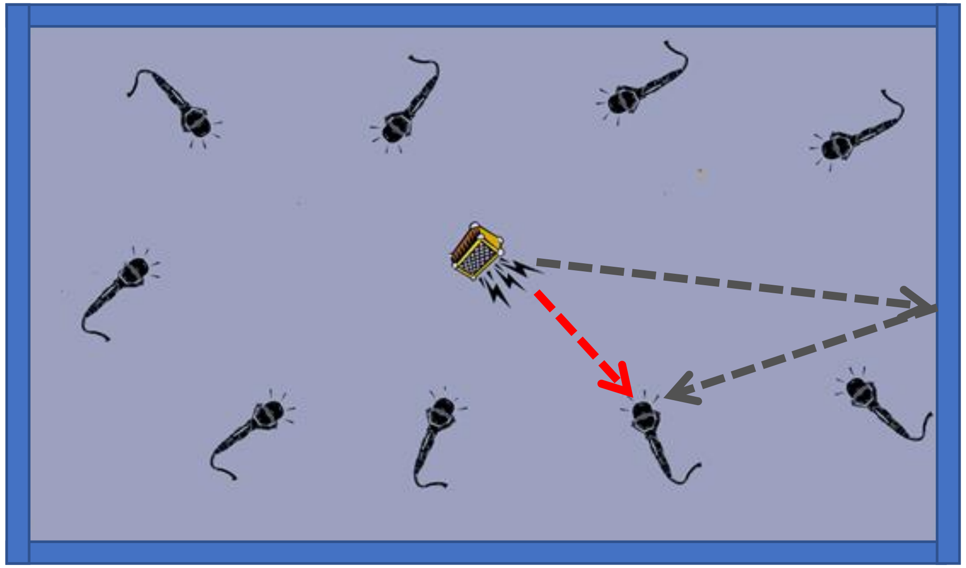
RIR sparsity prior

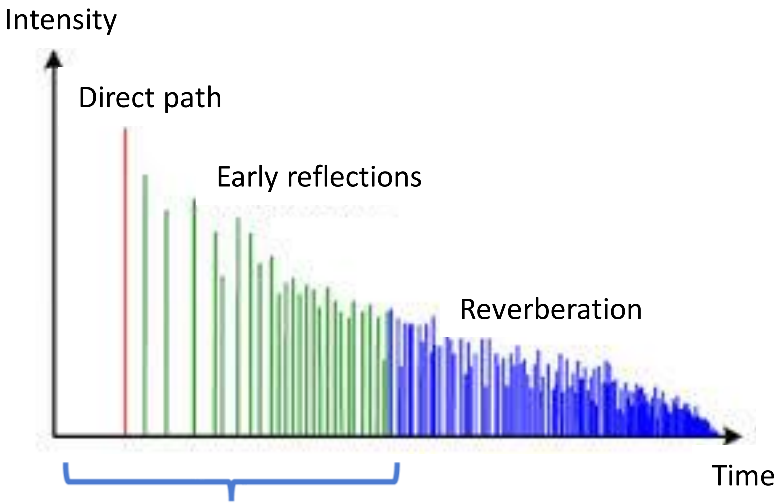




Sparsity holds for direct path and early reflections

RIR sparsity prior





Sparsity holds for direct path and early reflections

Reconstruction of <u>RIR direct path and early reflections</u> is sufficient for room geometry reconstruction and has proven to work also in speech enhancement [Yu et al 2012] and dereverberation [Lin et al, 2007].

[Yu et al 2012]: Yu, Meng, et al. "Multi-Channel Regularized Convex Speech Enhancement Model and Fast Computation by the Split Bregman Method." *Audio, Speech, and Language Processing, IEEE Transactions on* 20.2 (2012): 661-675.

[Lin et al, 2007]: Lin, Yuanqing, et al. "Blind channel identification for speech dereverberation using l1-norm sparse learning." *Advances in Neural Information Processing Systems.* 2007.

Penalty inducing sparsity

$$\underset{\mathbf{h}}{min}\ J(\mathbf{h}) + \lambda \|\mathbf{h}\|_0 \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1$$

NP- hard

Penalty inducing sparsity

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda\|\mathbf{h}\|_0 \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1$$

NP- hard

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda\|\mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1$$

[Kowalczyk et al, 2013]

[Kowalczyk et al, 2013]: Kowalczyk, Konrad, et al. "Blind System Identification Using Sparse Learning for TDOA Estimation of Room Reflections." *Sig. Proc. Letters, IEEE* 20.7 (2013): 653-656.

Penalty inducing sparsity

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_0 \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1$$

NP- hard

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1$$

[Kowalczyk et al, 2013]

Drawbacks:
- Non-convex problem due to the quadratic equality constraint
- L_1 norm penalizes larger coefficients more: the solution is not in general the same of L_0 norm.

[Kowalczyk et al, 2013]: Kowalczyk, Konrad, et al. "Blind System Identification Using Sparse Learning for TDOA Estimation of Room Reflections." *Sig. Proc. Letters, IEEE* 20.7 (2013): 653-656.

Anchor constraint and non-negativity constraint

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad h_1(a) = 1, \quad \mathbf{h} \geq 0$$

[Lin *et al* 2007]

Convex formulation

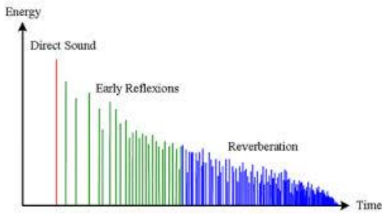
[Lin et al 2007]: Lin, Yuanqing, et al. "Blind sparse-nonnegative (BSN) channel identification for acoustic time-difference-of-arrival estimation." *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*. IEEE, 2007.

Anchor constraint and non-negativity constraint

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \; h_1(a) = 1, \quad \mathbf{h} \geq 0$$

[Lin *et al* 2007]

Convex formulation



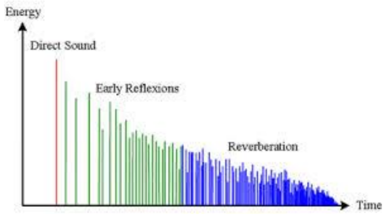
[Lin et al 2007]: Lin, Yuanqing, et al. "Blind sparse-nonnegative (BSN) channel identification for acoustic time-difference-of-arrival estimation." *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*. IEEE, 2007.

Anchor constraint and non-negativity constraint

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad h_1(a) = 1, \quad \mathbf{h} \geq 0$$

[Lin *et al* 2007]

Convex formulation



Drawbacks:
- amplitude distortion, peak of the anchor overly enhanced
- does not solve the L_1 penalty limitations

[Lin et al 2007]: Lin, Yuanqing, et al. "Blind sparse-nonnegative (BSN) channel identification for acoustic time-difference-of-arrival estimation." *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*. IEEE, 2007.

New approach

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_1 = 1, \quad \mathbf{h} \geq 0.$$

L_1 equality constraint

Convex solution, no distortion due to anchor

New approach

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_1 = 1, \quad \mathbf{h} \geq 0.$$

L_1 equality constraint

Convex solution, no distortion due to anchor

<u>BUT</u> L_1 norm appears both as a constraint and a penalty:

New approach

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_1 = 1, \quad \mathbf{h} \geq 0.$$

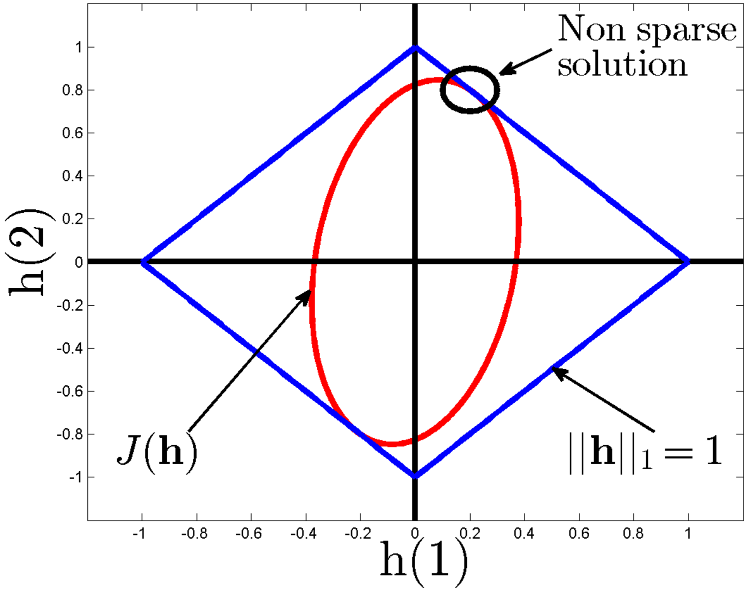
L_1 equality constraint

Convex solution, no distortion due to anchor

BUT L_1 norm appears both as a constraint and a penalty:

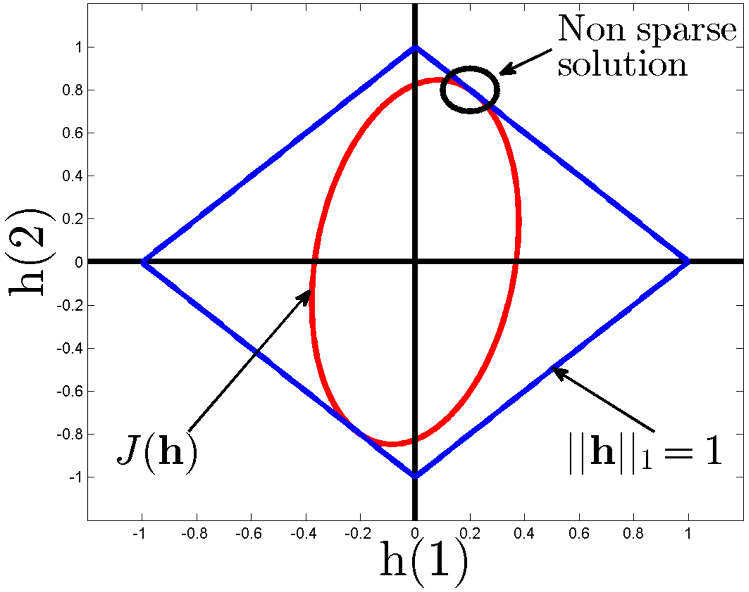
no more sparsity - inducing effect !!

Toy problem in two dimensions (1)

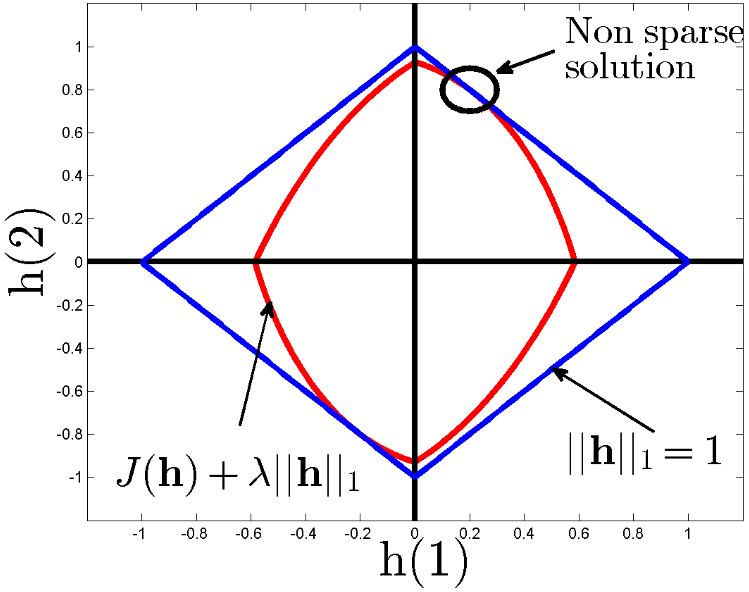


Quadratic cost function
without penalties

Toy problem in two dimensions (1)



Quadratic cost function
without penalties



Quadratic cost function
without L_1 penalty

Iterative reweighted L_1 penalty

Solve a sequence of sub - problems for z = 1, ... Z :

$$\hat{\mathbf{h}}^{(z)} = \min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{p}^{(z)} \odot \mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_1 = 1, \quad \mathbf{h} \geq 0$$

Iterative reweighted L_1 penalty

Solve a sequence of sub - problems for z = 1, ... Z :

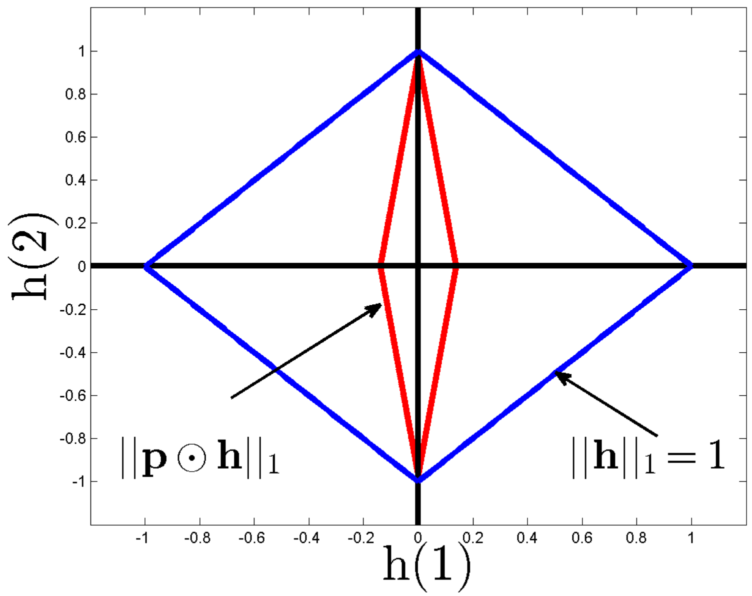
$$\hat{\mathbf{h}}^{(z)} = \min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{p}^{(z)} \odot \mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_1 = 1, \quad \mathbf{h} \geq 0$$

Weight update rule

$$p(l)^{(z)} = \frac{1}{\hat{h}(l)^{(z-1)} + \epsilon} \quad for \;\; l = 1, \cdots, L$$

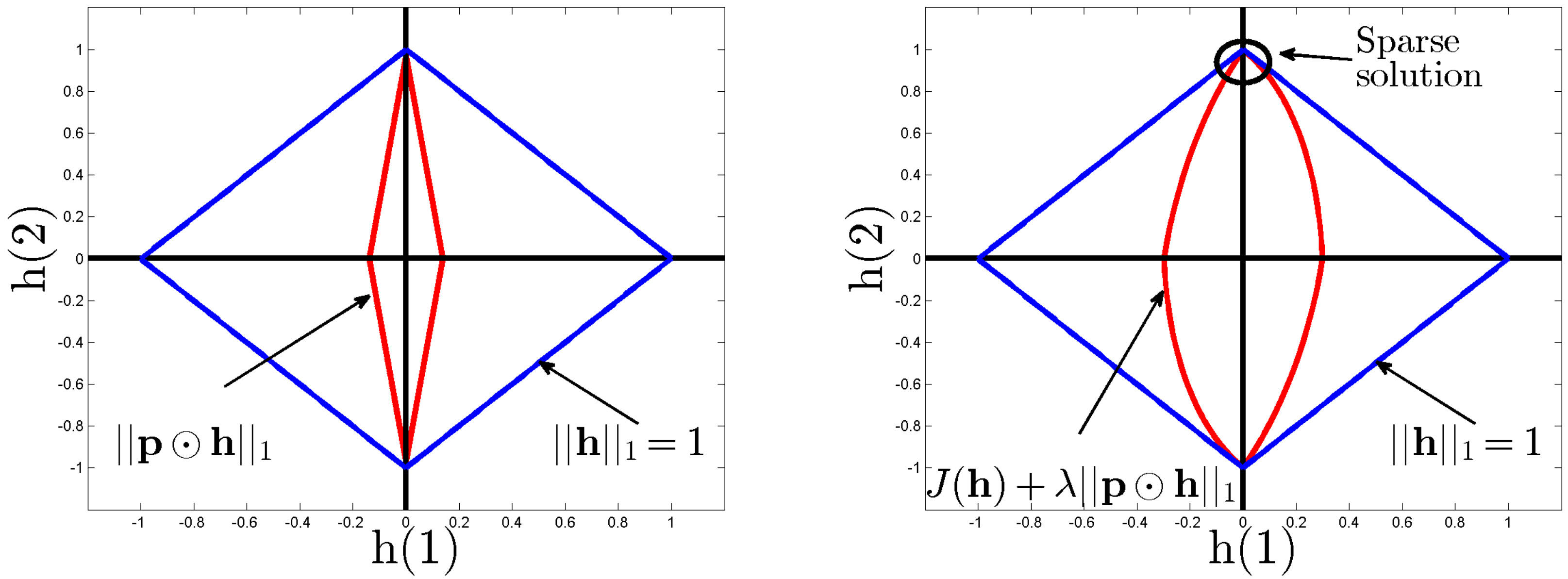
Smaller elements of vector $\mathbf{h}$ are more penalized than bigger ones.

Toy problem in two dimensions (2)



Weighted L_1 penalty $p(1) > p(2)$

Toy problem in two dimensions (2)



Weighted L1 penalty $p(1) > p(2)$

Quadratic cost function with weighted L_1 penalty

Initialization and convergence

How to initialize $\mathbf{p}^{(1)}$?

<u>Use the solution of the anchor – constrained problem</u>

$$\hat{\mathbf{h}}^{(1)} = \underset{\mathbf{h}}{argmin} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad h_1(a) = 1, \quad \mathbf{h} \geq 0$$

Initialization and convergence

How to initialize $\mathbf{p}^{(1)}$?

Use the solution of the anchor – constrained problem

$$\hat{\mathbf{h}}^{(1)} = \underset{\mathbf{h}}{argmin} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad h_1(a) = 1, \quad \mathbf{h} \geq 0$$

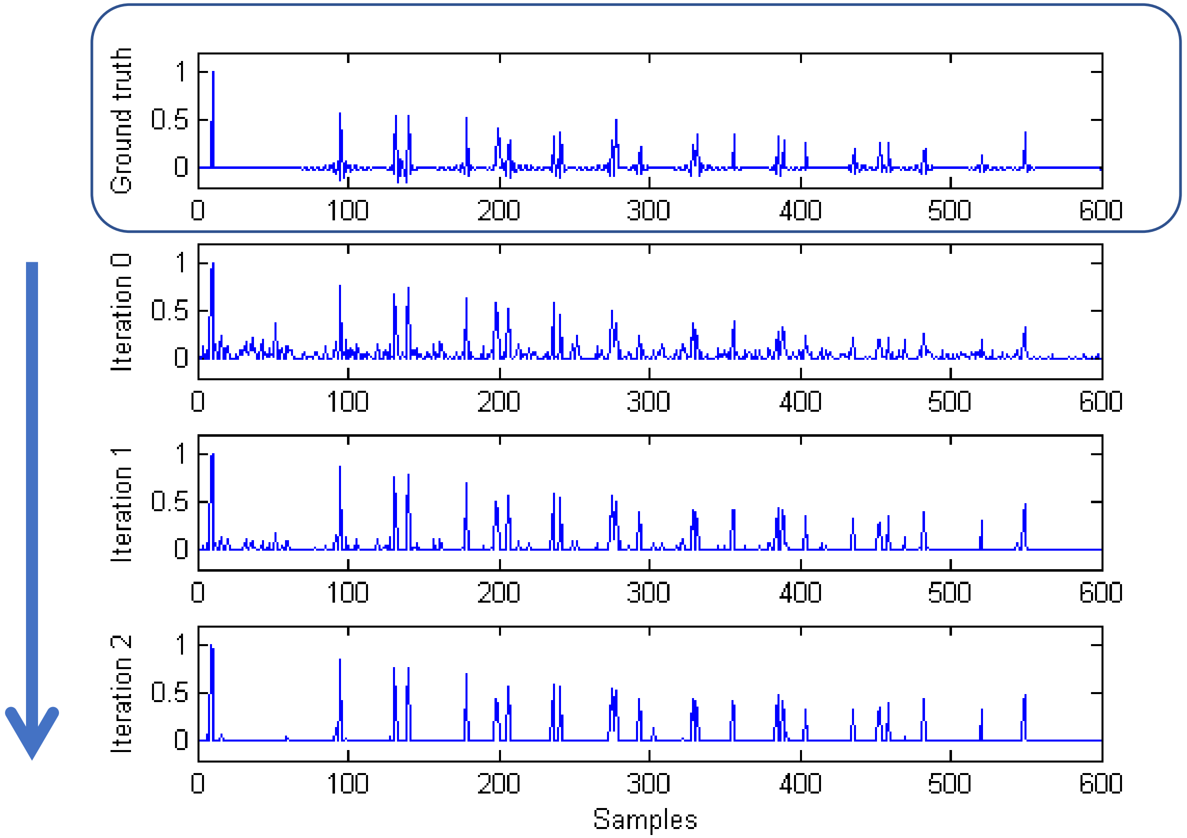
At convergence $\hat{\mathbf{h}}^{(z-1)} \approx \hat{\mathbf{h}}^{(z)}$ therefore

$$p(l)^{(z)}\hat{h}(l)^{(z)} \approx 1 \;\; for \;\; \hat{h}(l)^{(z)} \neq 0$$

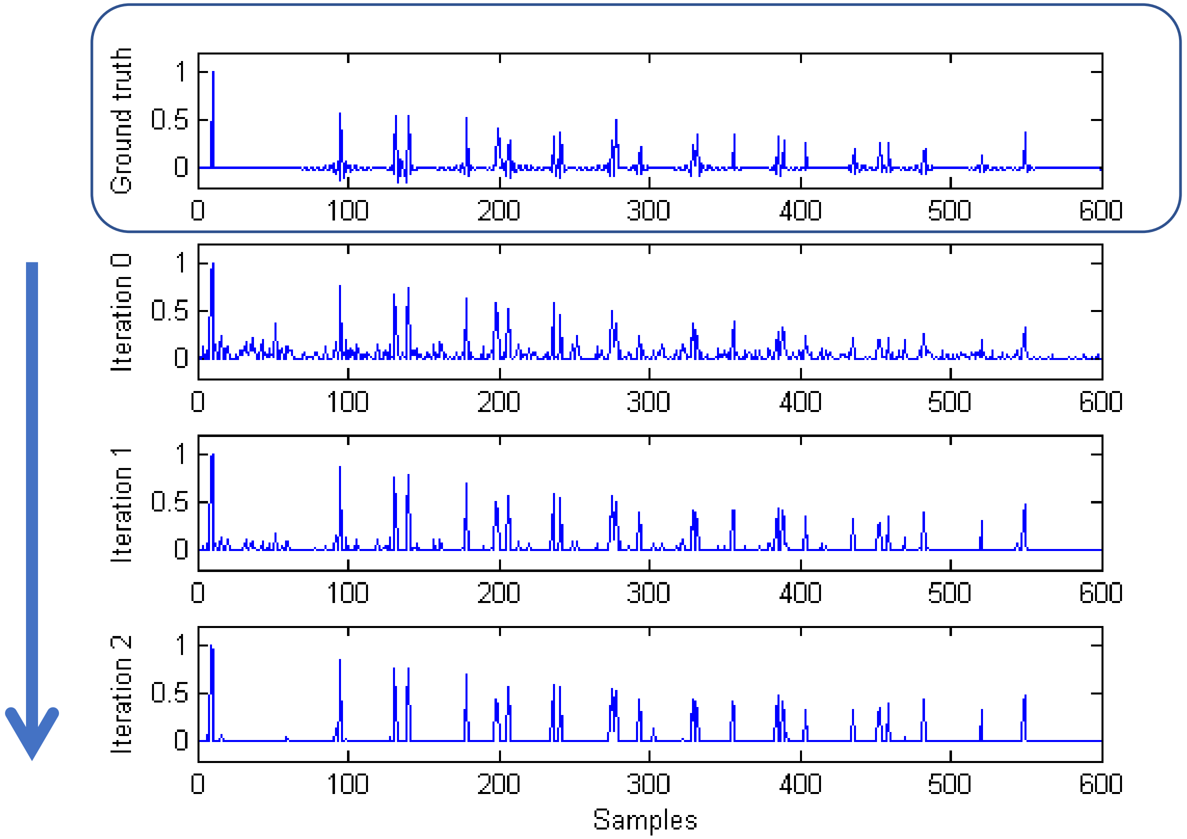
$$p(l)^{(z)}\hat{h}(l)^{(z)} \approx 0 \;\; for \;\; \hat{h}(l)^{(z)} = 0$$

→ Weighted L_1 norm is equivalent to L_0 norm

Example of the iterative procedure



Example of the iterative procedure

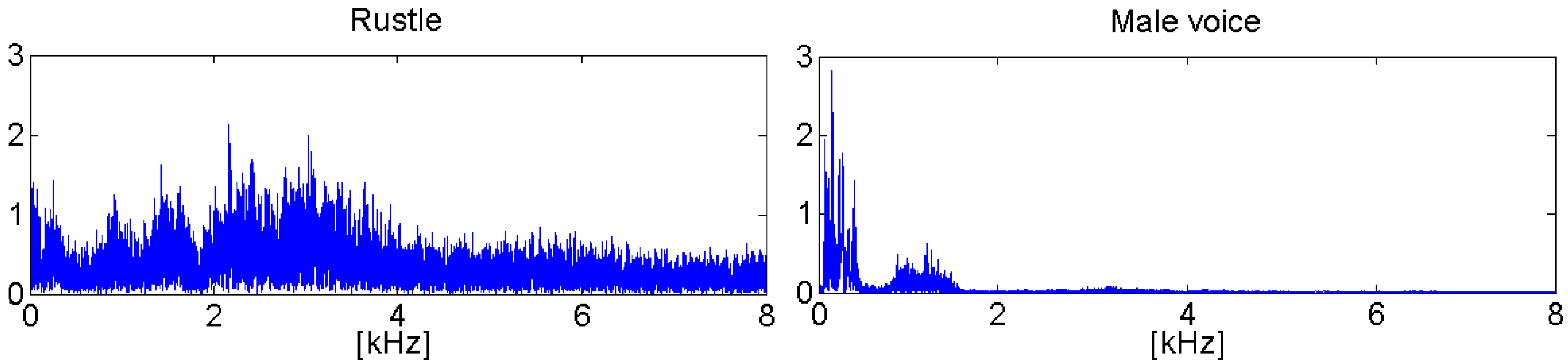


<u>Spurious peaks</u> are gradually eliminated from the estimated RIR

Even if non–negativity is not perfect in the GT RIR, the estimated RIR well preserves <u>peaks position and energy</u>

Experiments

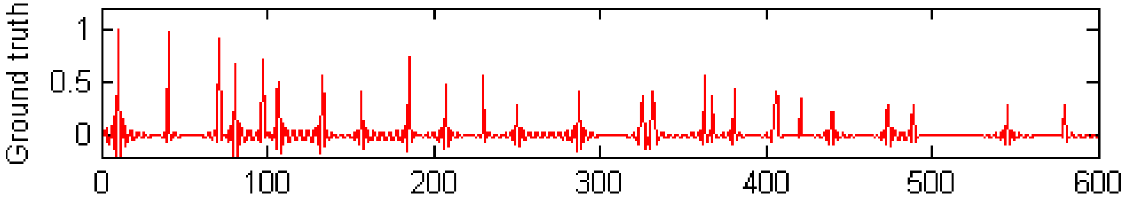
- Simulated room of 5 x 4 x 6 m
- 2 microphones and a source randomly placed
- RIRs simulated with the image method [Allen & Berkley, 1979]
- Synthetic and real signals: white noise, rustle, male voice
- Variable SNR: 0, 6, 14, 20, 40 dB
- 50 Monte Carlo simulations for each SNR



[Allen & Berkley, 1979]: Allen, Jont B., and David A. Berkley. "Image method for efficiently simulating small-room acoustics." *The Journal of the Acoustical Society of America* 65.4 (1979): 943-950.

Example of RIR estimation

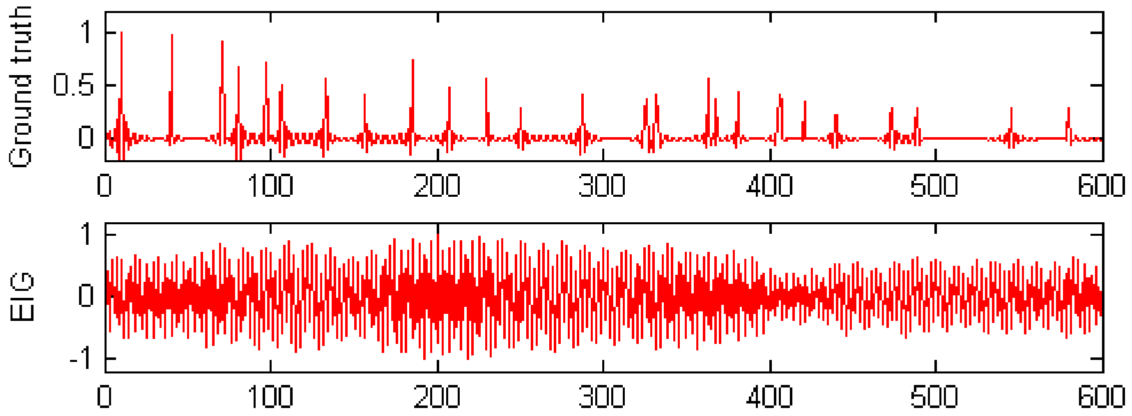
<u>Ground Truth</u>



Example of RIR estimation

Ground Truth

EIG: eigenvalue problem
[Tong et al. 1994]

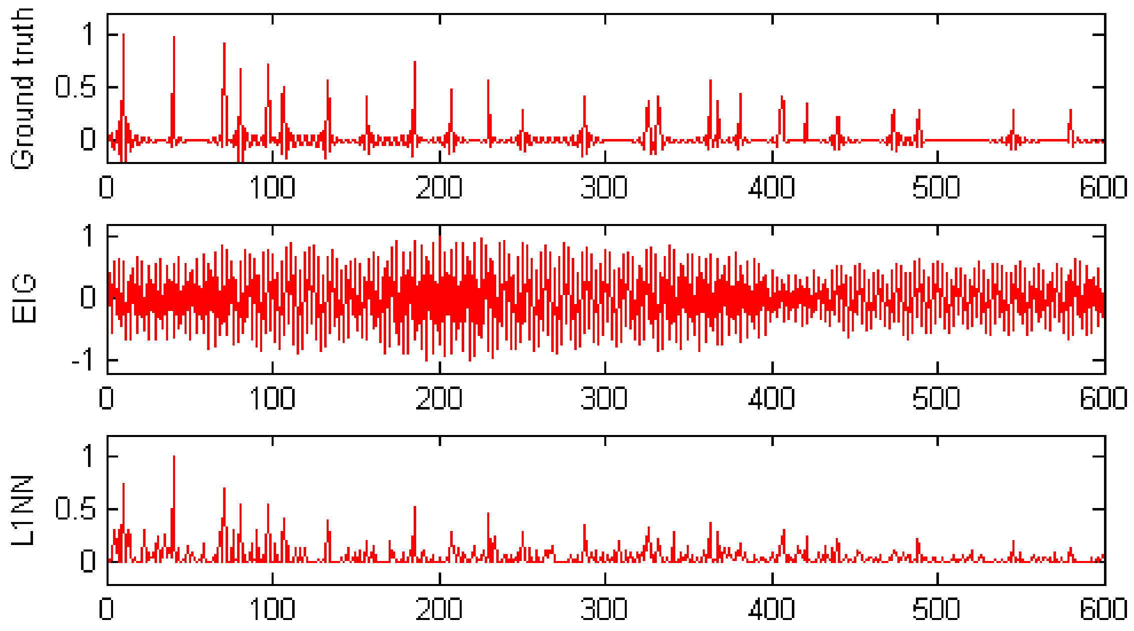


Example of RIR estimation

<u>Ground Truth</u>

EIG: eigenvalue problem
[Tong et al. 1994]

L1NN: anchor constraint
(initialization) [Lin et al. 2007]



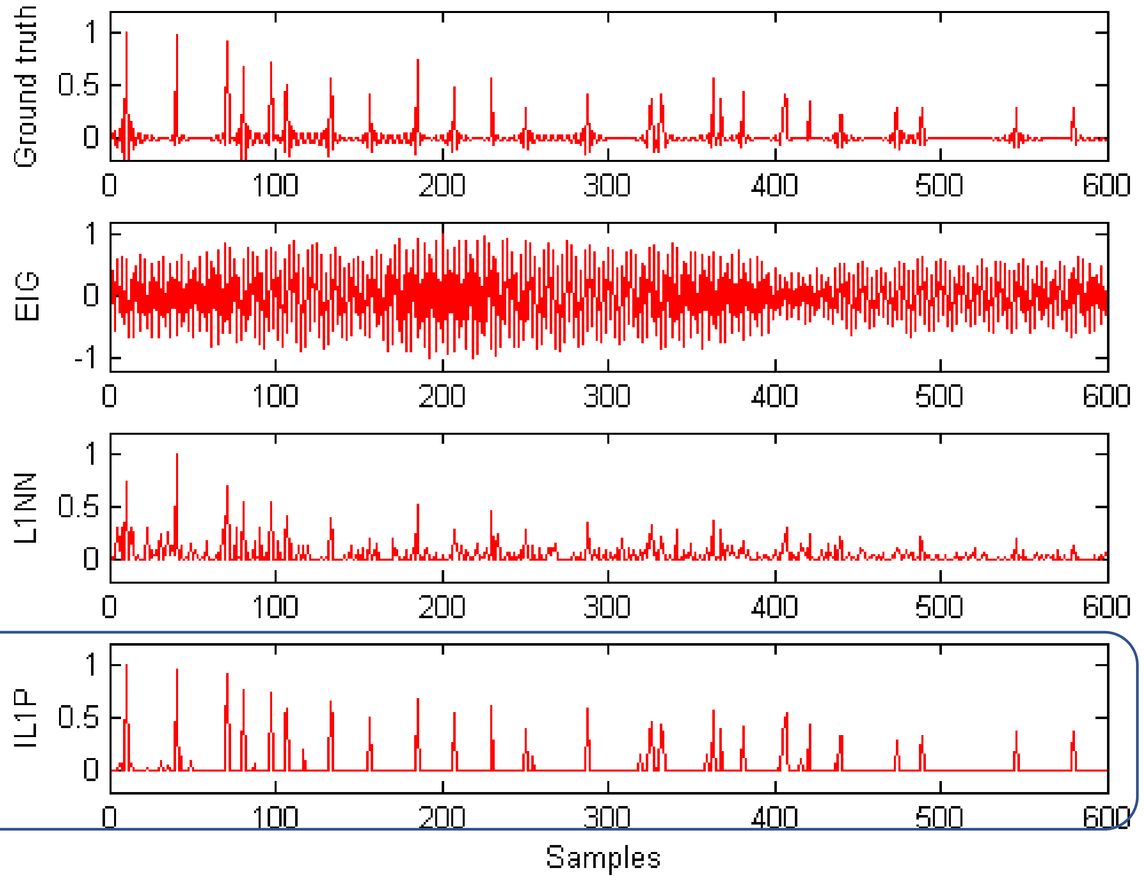
Example of RIR estimation

Ground Truth

EIG: eigenvalue problem [Tong et al. 1994]

L1NN: anchor constraint (initialization) [Lin et al. 2007]

IL1P: Proposed method



Metrics of evaluation

Average peak position mismatch

$$A_{PPM} = \sum_{i=1}^{50} \sum_{p=1}^{P_i} \frac{|\tau_{p,i}^{gt} - \tau_{p,i}^{e}|}{50 P_i}$$

Peak position accuracy over the <u>inliers</u>

Metrics of evaluation

Average peak position mismatch

$$A_{PPM} = \sum_{i=1}^{50} \sum_{p=1}^{P_i} \frac{|\tau_{p,i}^{gt} - \tau_{p,i}^{e}|}{50 P_i}$$

Peak position accuracy over the <u>inliers</u>

Average percentage of unmatched peaks

$$A_{PUP} = \sum_{i=1}^{50} \frac{K - P_i}{50K}$$

Percentage of <u>outliers</u> (> 20 samples)

<u>Metrics evaluated on the first (sparse) part of the RIR</u>

Quantitative results

Synthetic:
white noise

SNR	EIG	L1NN	IL1P
0 dB	4.51 [0.34]	1.96 [0.16]	**1.17 [0.11]**
6 dB	4.24 [0.34]	0.77 [0.10]	**0.40 [0.04]**
14 dB	3.86 [0.31]	0.34 [0.03]	**0.28 [0.01]**
20 dB	3.42 [0.30]	0.28 [0.03]	**0.27 [0.01]**
40 dB	3.20 [0.28]	**0.26 [0.00]**	0.27 **[0.00]**

$A_{PPM}\ [A_{PUP}]$

Quantitative results

Synthetic: white noise

SNR	EIG	L1NN	IL1P
0 dB	4.51 [0.34]	1.96 [0.16]	**1.17 [0.11]**
6 dB	4.24 [0.34]	0.77 [0.10]	**0.40 [0.04]**
14 dB	3.86 [0.31]	0.34 [0.03]	**0.28 [0.01]**
20 dB	3.42 [0.30]	0.28 [0.03]	**0.27 [0.01]**
40 dB	3.20 [0.28]	**0.26 [0.00]**	0.27 **[0.00]**

Real: rustle

SNR	EIG	L1NN	IL1P
0 dB	4.53 [0.39]	2.46 [0.18]	**2.23 [0.14]**
6 dB	4.50 [0.37]	1.29 [0.12]	**0.54 [0.02]**
14 dB	3.82 [0.38]	0.39 [0.02]	**0.28 [0.00]**
20 dB	3.65 [0.35]	0.29 [0.01]	**0.28 [0.00]**
40 dB	3.31 [0.29]	**0.28 [0.00]**	**0.28 [0.00]**

A_{PPM} $[A_{PUP}]$

Quantitative results

	SNR	EIG	L1NN	IL1P
Synthetic: white noise	0 dB	4.51 [0.34]	1.96 [0.16]	**1.17 [0.11]**
	6 dB	4.24 [0.34]	0.77 [0.10]	**0.40 [0.04]**
	14 dB	3.86 [0.31]	0.34 [0.03]	**0.28 [0.01]**
	20 dB	3.42 [0.30]	0.28 [0.03]	**0.27 [0.01]**
	40 dB	3.20 [0.28]	**0.26 [0.00]**	0.27 **[0.00]**
Real: rustle	0 dB	4.53 [0.39]	2.46 [0.18]	**2.23 [0.14]**
	6 dB	4.50 [0.37]	1.29 [0.12]	**0.54 [0.02]**
	14 dB	3.82 [0.38]	0.39 [0.02]	**0.28 [0.00]**
	20 dB	3.65 [0.35]	0.29 [0.01]	**0.28 [0.00]**
	40 dB	3.31 [0.29]	**0.28 [0.00]**	**0.28 [0.00]**
Real: male voice	0 dB	4.66 [0.38]	**2.87** [0.23]	3.31 **[0.16]**
	6 dB	4.91 [0.37]	1.96 [0.19]	**1.36 [0.08]**
	14 dB	4.44 [0.40]	0.98 [0.10]	**0.58 [0.01]**
	20 dB	4.30 [0.35]	0.50 [0.04]	**0.39 [0.01]**
	40 dB	3.64 [0.35]	0.32 [0.01]	**0.29 [0.00]**

$A_{PPM}\ [A_{PUP}]$

Conclusions

- The proposed method yields a convex, computationally efficient formulation for the blind sparse non-negative RIR estimation problem.
- The iterative weighted L_1 penalty fairly approximates the optimal L_0 penalty.
- The proposed method outperforms current approachs in almost all SNR conditions and for different sound sources, both synthetic and real.
- Future works will assess the method effectiveness on a set of applications like speech enhancement and room geometry reconstruction.