ESTIMATION OF TDOA FOR ROOM REFLECTIONS BY ITERATIVE WEIGHTED L1 CONSTRAINT

ISTITUTO ITALIANO DI TECNOLOGIA

Marco Crocco and Alessio Del Bue

Visual Geometry and Modelling (VGM) Lab

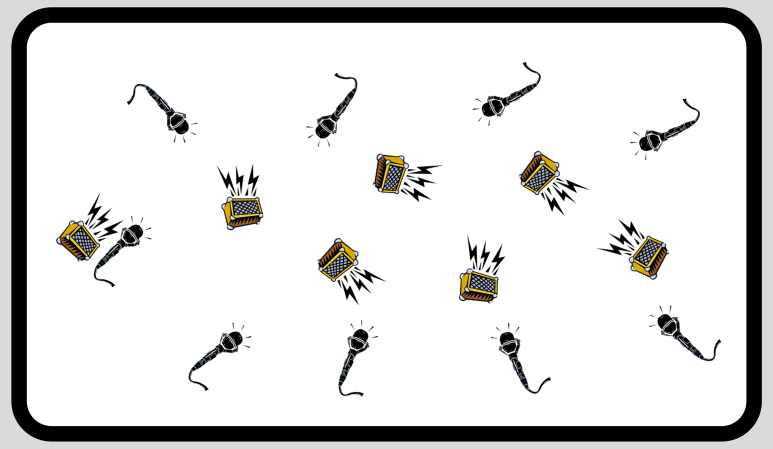
Pattern Analysis and Computer Vision (PAVIS) Department

Istituto Italiano di Tecnologia (IIT), Genova, Italy

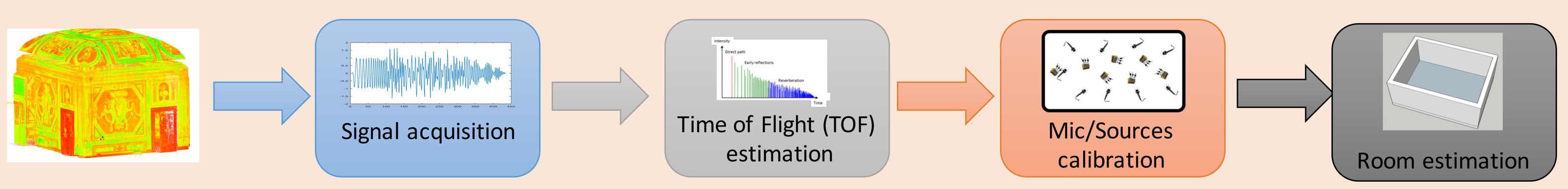
Signal and Information Processing : The Heartbeat of a Smart Society
Shanghai, China
ICASSP 2016
The 41st IEEE International Conference on Acoustics, Speech and Signal Processing, 20-25 March 2016

A method to extract Time Difference of Arrivals (TDOAs) for room reflections:

- Estimation of TDOAs with **no a priori** information (statistic of signals, walls/reflectors positions)
- **Few parameters** to tune.
- Fast, iterative solution that promotes **sparsity and non-negativity.**
- Tested with **real signals** (speech and non-speech).



Room Reconstruction Problem



The estimation of reliable TDOAs is a necessary step for room reconstruction to recover the 3D position and orientation of reflectors and the position of sound sources/microphones.

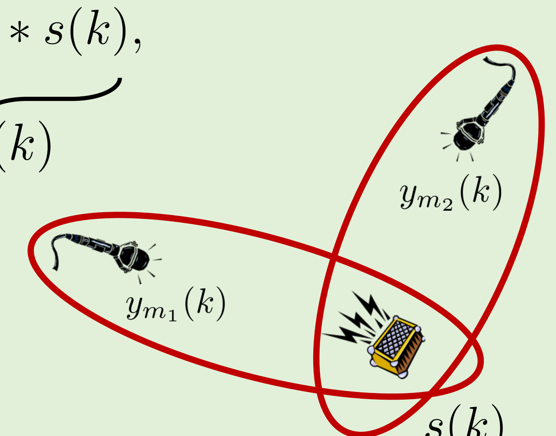
The cross-relation property of a Multi-Output system:

$h_{m_i}(k)$ is the Room Impulse Response (RIR) between source location and microphone m_i location. For every couple of microphones we have that:

$$h_{m_1}(k) * \underbrace{h_{m_2}(k) * s(k)}_{y_{m_2}(k)} = h_{m_2}(k) * \underbrace{h_{m_1}(k) * s(k)}_{y_{m_1}(k)},$$

Thus giving:

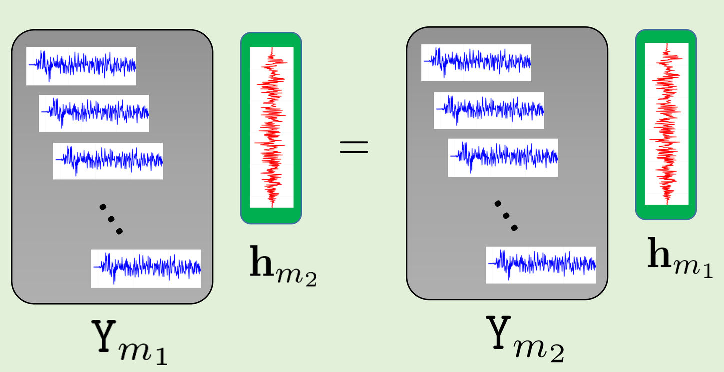
$$h_{m_1}(k) * y_{m_2}(k) = h_{m_2}(k) * y_{m_1}(k)$$



$$y_{m1}(k) * h_{m_2}(k) = \mathtt{Y}_{m_1}\mathbf{h}_{m_2}$$

Shift to matrix form

$$\mathtt{Y}_{m_1}\mathbf{h}_{m_2} = \mathtt{Y}_{m_2}\mathbf{h}_{m_1}$$



Leading to: $J(\mathbf{h}) = \sum_{m_1 \neq m_2} \|\mathtt{Y}_{m_1}\mathbf{h}_{m_2} - \mathtt{Y}_{m_2}\mathbf{h}_{m_1}\|_2^2 = \mathbf{h}^\top \mathtt{Q}\mathbf{h}$

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, 1994.

Solution of the quadratic cost function $J(\mathbf{h})$:

Include penalty terms in the cost function to avoid trivial solutions and to promote properties of the RIR: **Sparsity and non-negativity.**

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

NP- hard

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

Drawbacks:

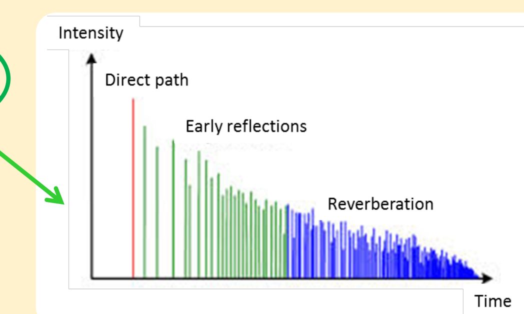
- Non-convex problem due to the quadratic equality constraint;
- L_1 norm penalizes larger coefficients more: the solution might be not equal to the L_0 norm.

Kowalczyk et al. "Blind System Identification Using Sparse Learning for TDOA Estimation of Room Reflections." *Sig. Proc. Letters,* 2013.

Non-negative:

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

Convex formulation



Drawbacks:

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

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

Iterative weighted L_1 constraint (IL1C)

IL1C promotes sparsity and non-negativity by removing the previous drawbacks. It is an iterative method that at **each iteration re-weights to improve sparsity**.

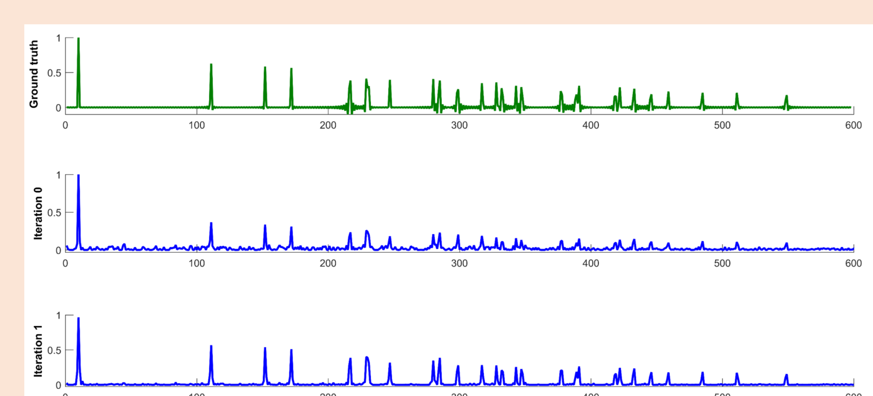
Solve a sequence of sub-problems for $z = 1, \ldots, Z$:

$$\mathbf{h}^z = arg\,min J(\mathbf{h}) + \|\mathbf{h}\|_1, \quad s.t. \; \mathbf{p}^{z\top}\mathbf{h} = 1, \quad \mathbf{h} > 0$$

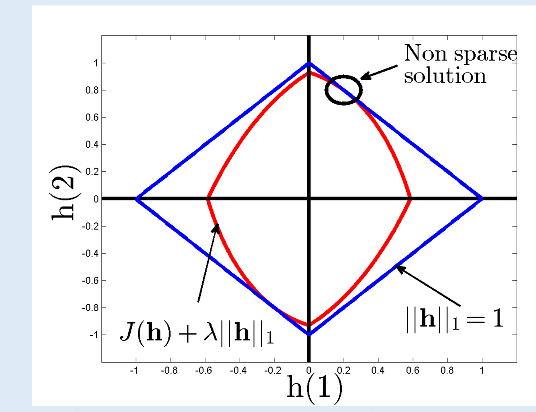
Weight update rule $\mathbf{p}^z = \mathbf{h}^{z-1}$

No anchor constraint and penalizing more the small components

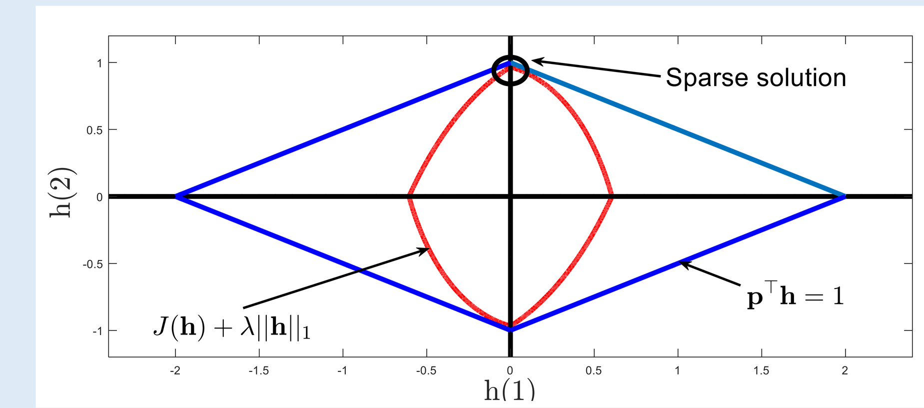
Each iteration consists of the minimization of a **quadratic cost function + linear constraints: convex problem** easily solved by standard optimization methods.



Geometrical interpretation



Quadratic cost function with L_1 penalty

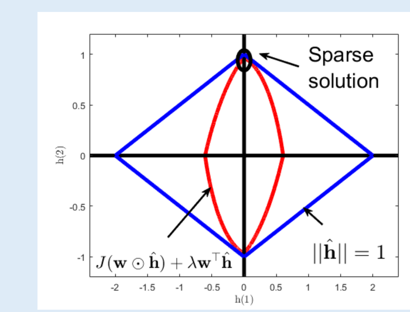


Quadratic cost function with L_1 penalty and weighted L_1 **constraint**

Let us make a variable change: $\mathbf{h} = \mathbf{w} \odot \hat{\mathbf{h}}$

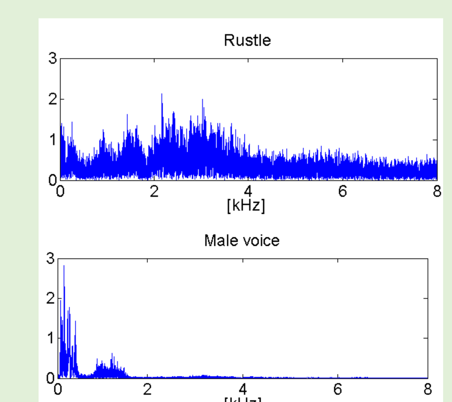
with: $w(k) = \dfrac{1}{p(k)}$

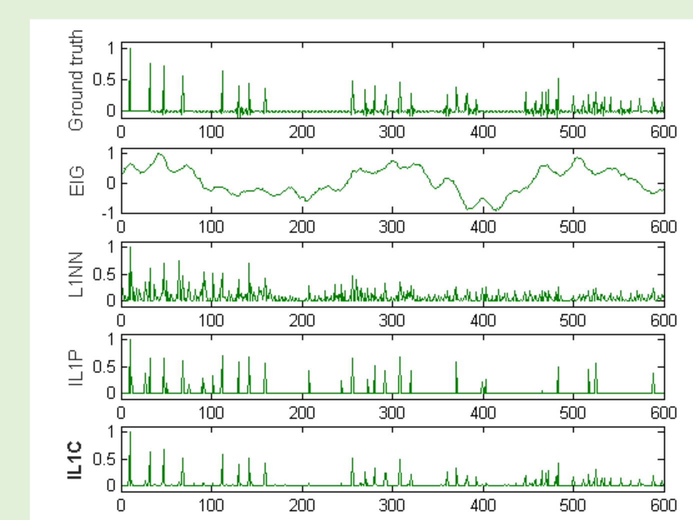
$$J(\mathbf{w} \odot \hat{\mathbf{h}}) + \mathbf{w}^\top\hat{\mathbf{h}}, \quad s.t. \; \|\hat{\mathbf{h}}\|_1 = 1, \quad \mathbf{h} > 0$$



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





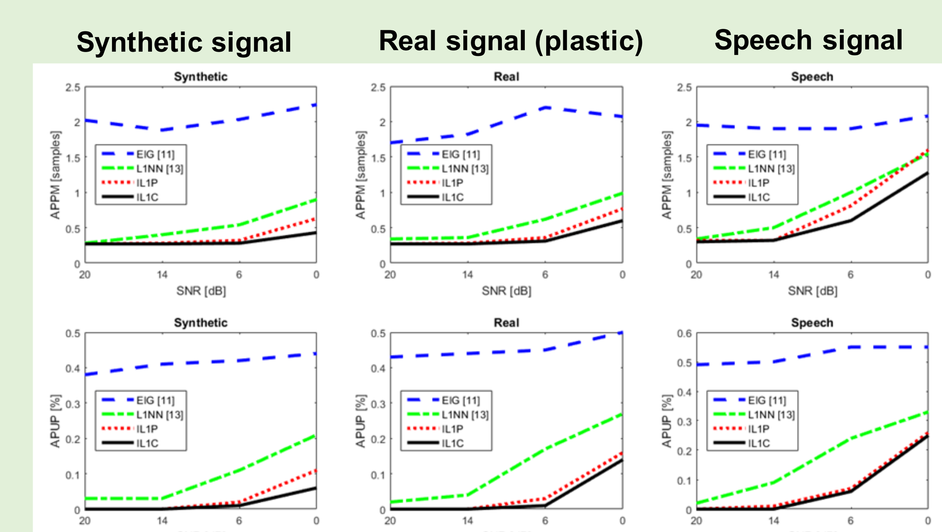
EIG: eigenvalue problem [Tong et al. 1994]

L1NN: anchor constraint [Lin et al. 2007]

IL1P : iterative L1 penalty [Crocco and Del Bue, 2015]

IL1C : iterative L1 constraint [Crocco and Del Bue. 2016]

Results show that SVD based solution fails while anchor constraints do not achieve desired sparsity. IL1P solution is a penalty based solution over the L1 regulariser while IL1C achieves the desired level of sparsity



Average peak position mismatch

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

Peak position accuracy over the inliers

Average percentage of unmatched peaks

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

Percentage of outliers (> 20 samples)