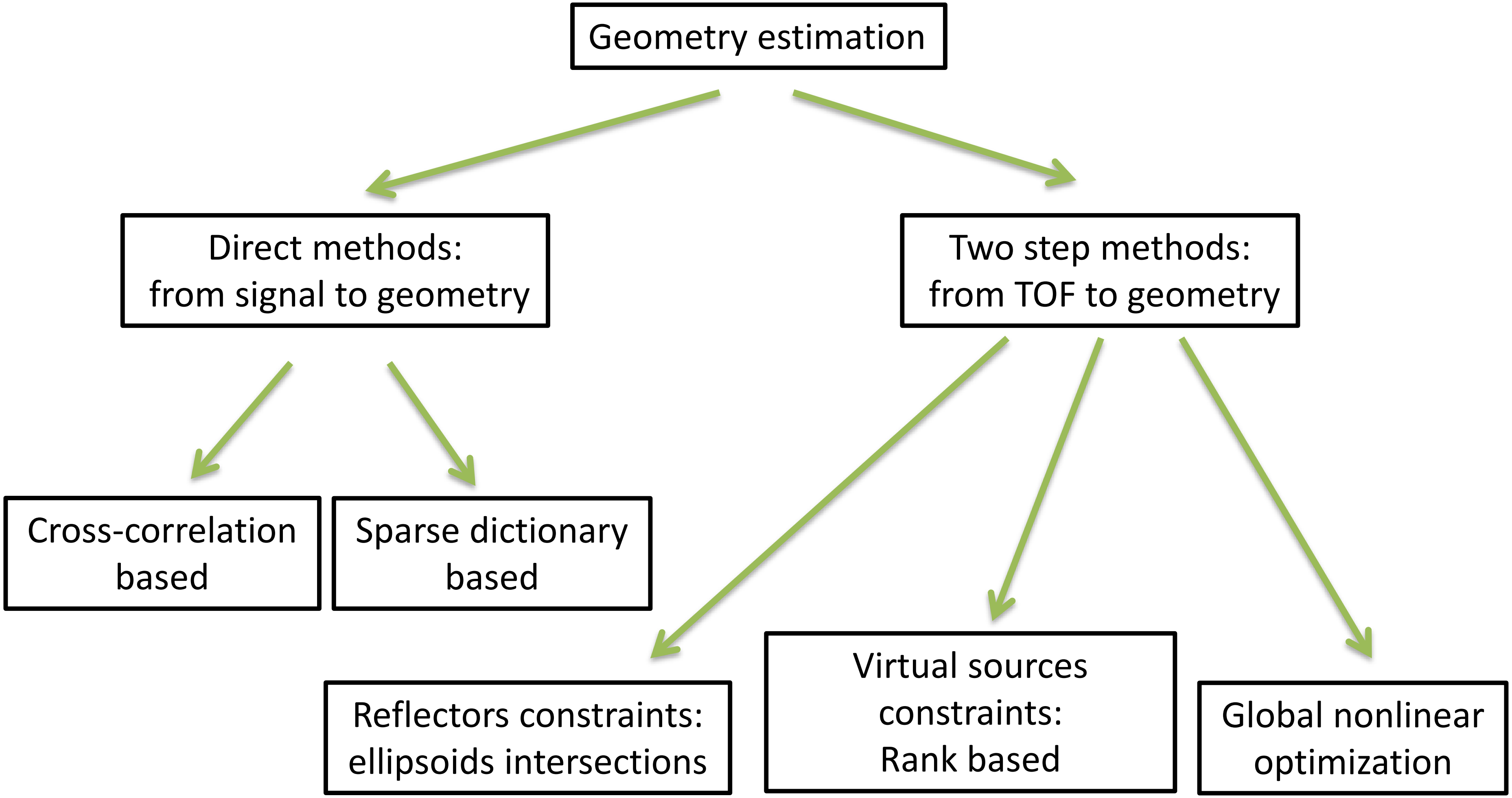
Room geometry estimation

Taxonomy



Geometry estimation

Direct methods: from signal to geometry

Two step methods: from TOF to geometry

Cross-correlation based

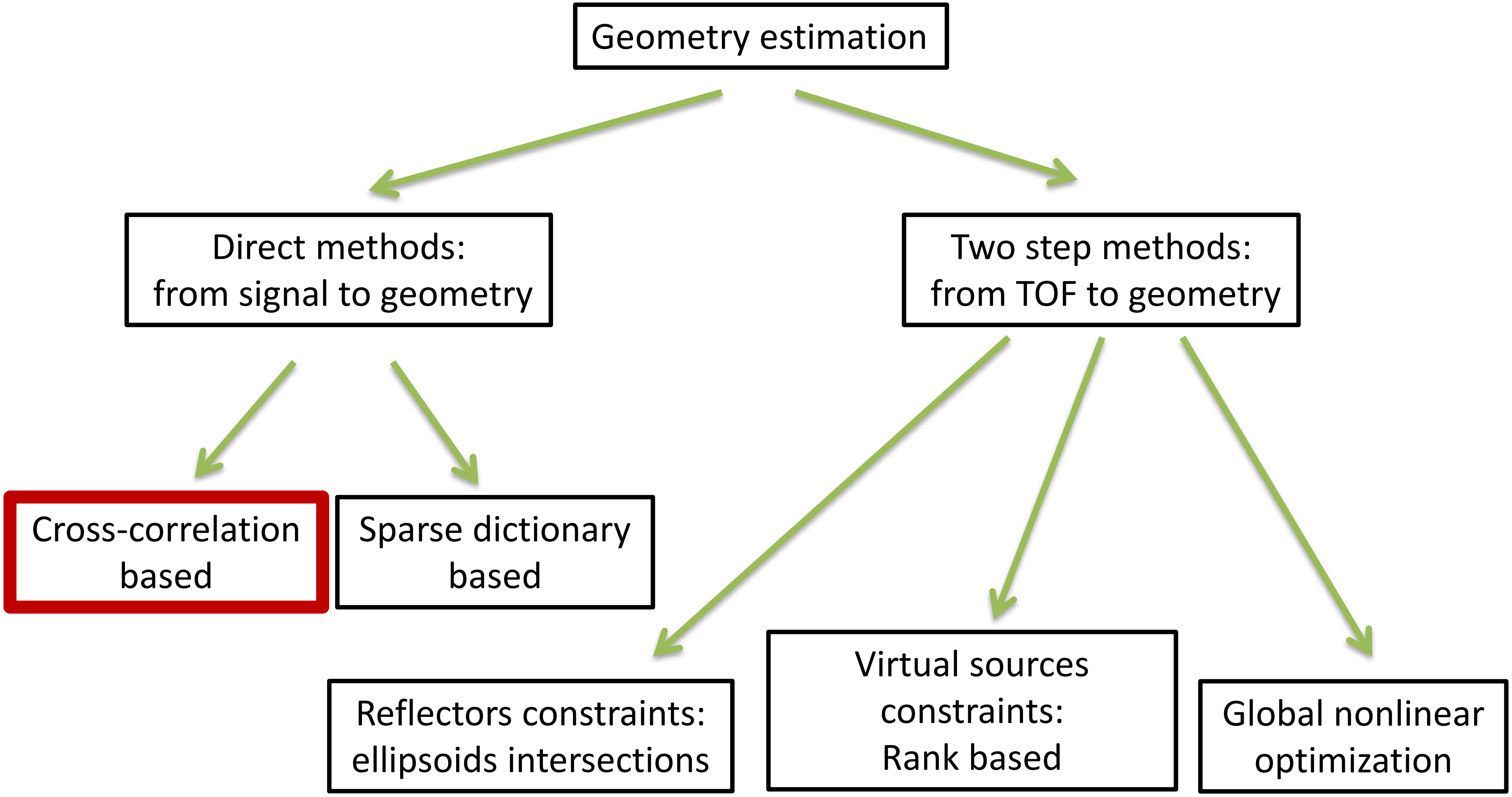
Sparse dictionary based

Reflectors constraints: ellipsoids intersections

Virtual sources constraints: Rank based

Global nonlinear optimization

Taxonomy



Geometry estimation

Direct methods: from signal to geometry

Two step methods: from TOF to geometry

Cross-correlation based

Sparse dictionary based

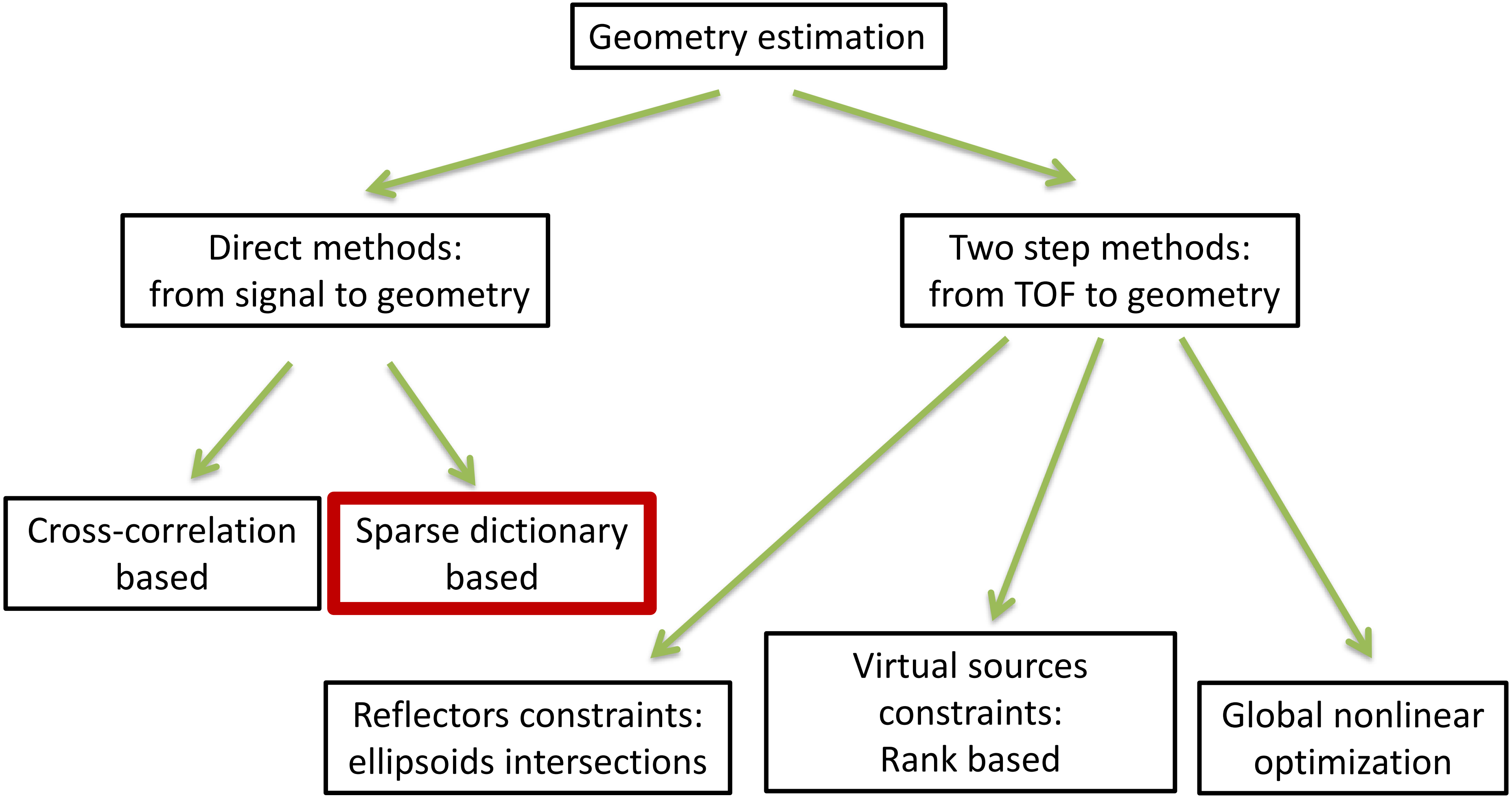
Reflectors constraints: ellipsoids intersections

Virtual sources constraints: Rank based

Global nonlinear optimization

Tervo & Korhonen, “Estimation of reflective surfaces from continuous signals.” ICASSP 2010 .

Taxonomy



Geometry estimation

Direct methods: from signal to geometry

Two step methods: from TOF to geometry

Cross-correlation based

Sparse dictionary based

Reflectors constraints: ellipsoids intersections

Virtual sources constraints: Rank based

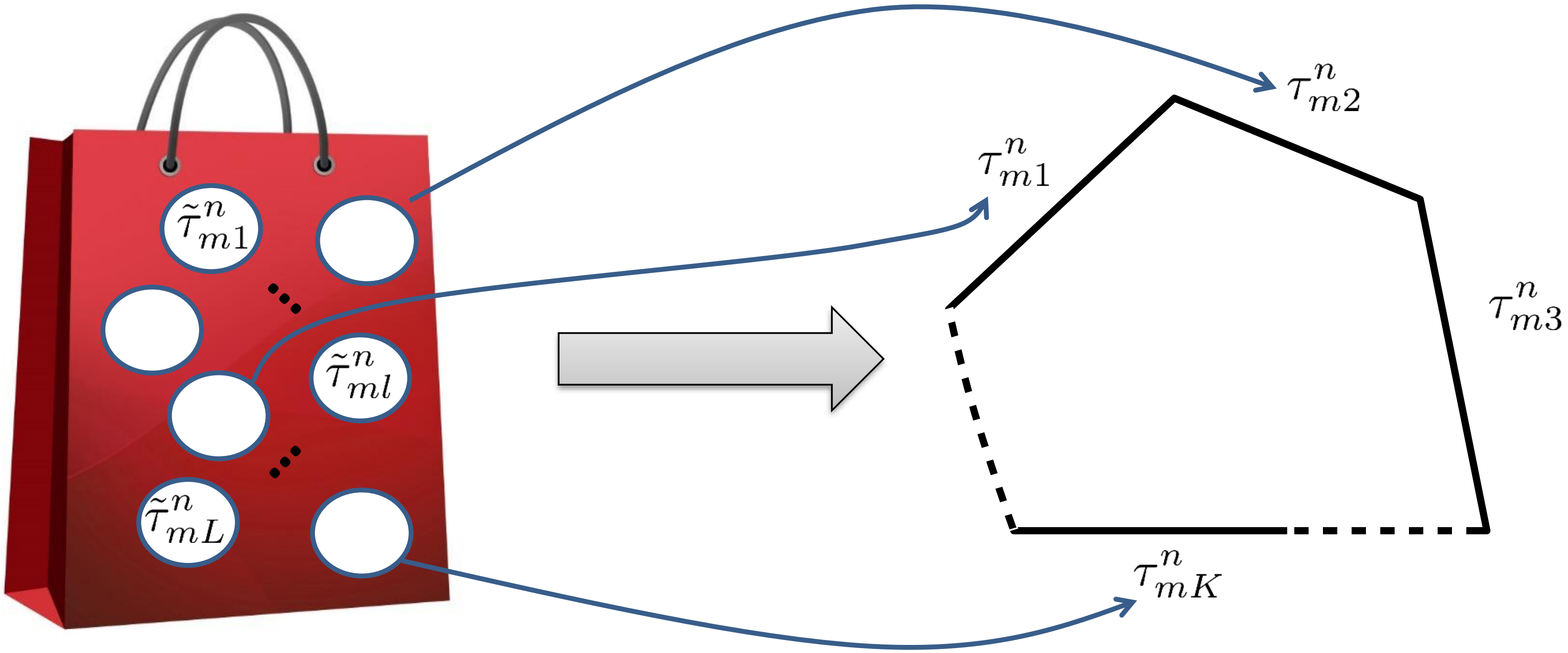
Global nonlinear optimization

Ribeiro et al., "Geometrically Constrained Room Modeling With Compact Microphone Arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, 2012.

Challenges

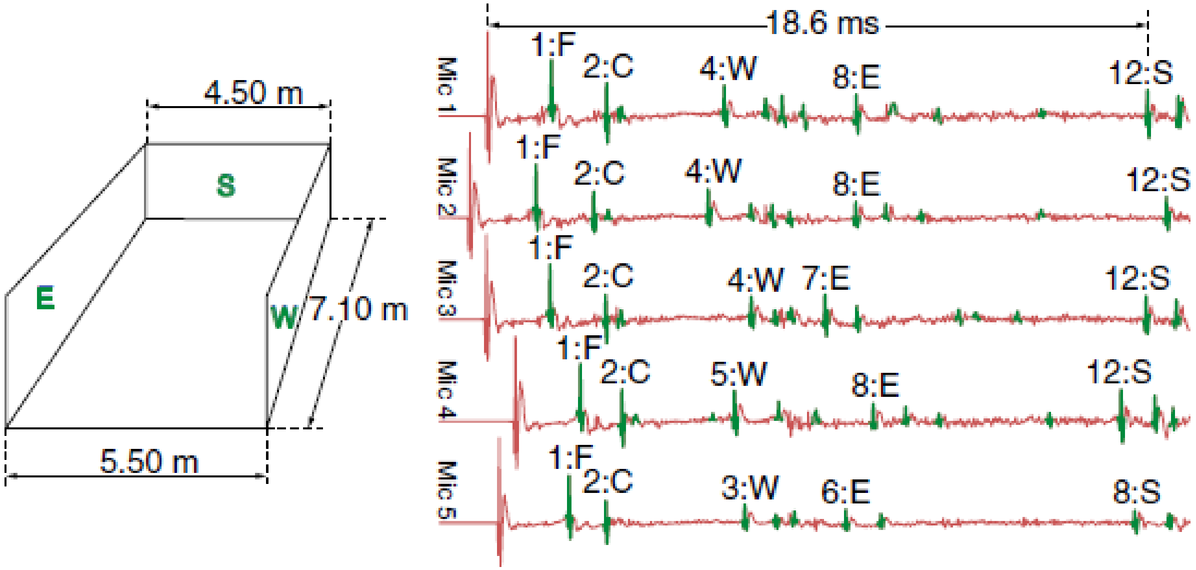
- Noise on measured delays;
- Spurious delays due to reflections from objects, diffraction effects, etc.;
- Missing delays due to low SNR measurements;
- **Unlabeled delays:** matching between planar reflectors and delays at each microphone is in general unknown.

Unlabeled delays



- Matching between delays and walls (echo sorting) has to be estimated.
- NP-hard permutation problem

Disambiguation of delays: a real case

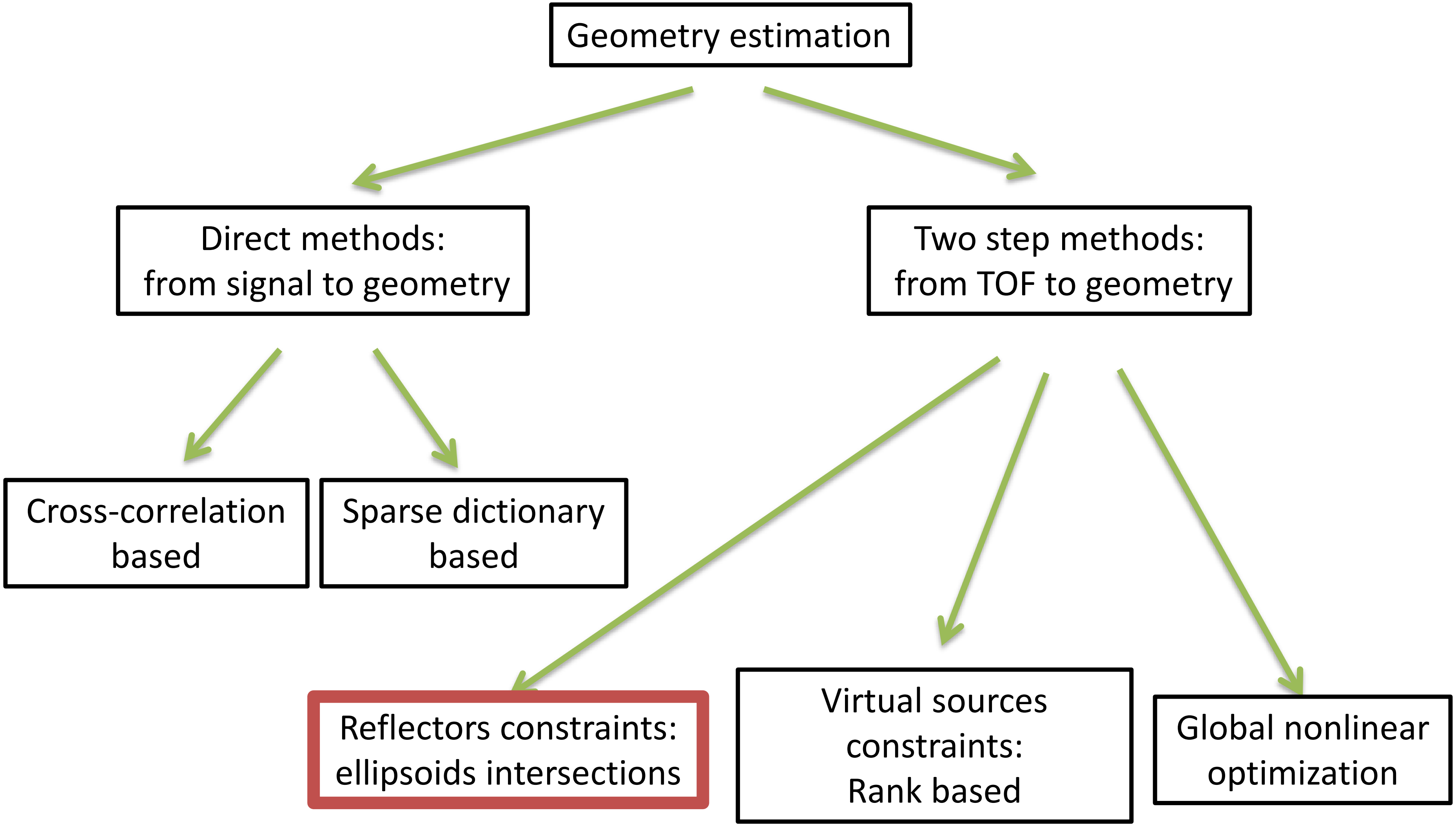


Courtesy of [Dokmanic et al. 2013]

After detecting delays in the real signals, the echo sorting problem assigns delays to ceiling (C), floor (F) and walls (E, S, W) of the room.

Second order delays may complicate further the problem with very spread configurations of microphones.

Taxonomy



Geometry estimation

Direct methods: from signal to geometry

Two step methods: from TOF to geometry

Cross-correlation based

Sparse dictionary based

Reflectors constraints: ellipsoids intersections

Virtual sources constraints: Rank based

Global nonlinear optimization

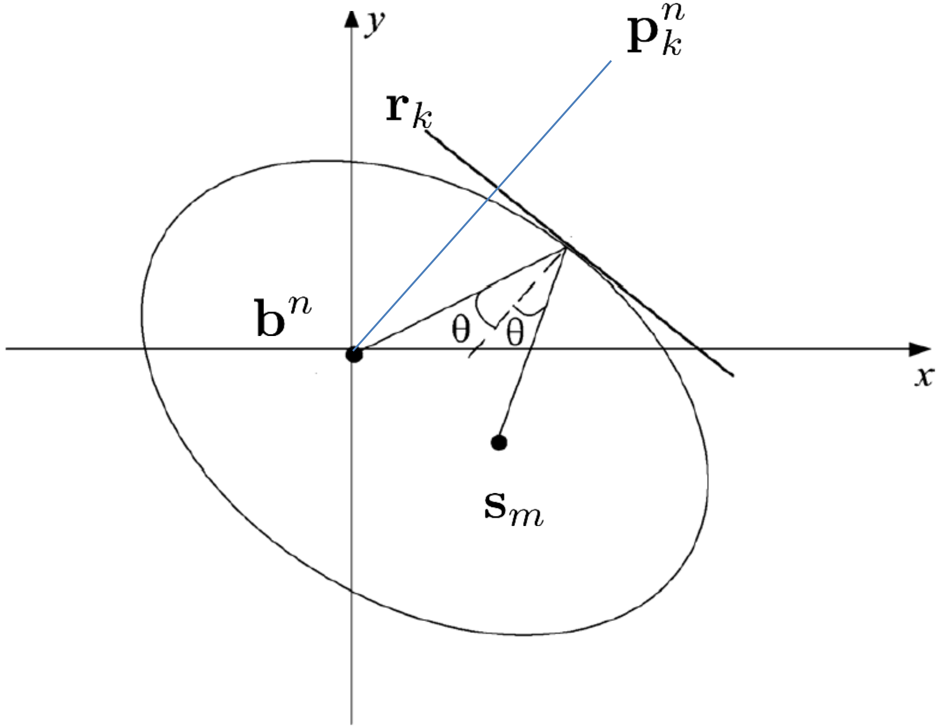
Reflectors constraints: ellipsoids intersections

Requirements:

- Sources and Microphones positions known;
- Tx signal known;
- Each source is placed very close to a planar reflector.

Reflector constraint (2D)

Given a TOF τ_{mk}^{n} extracted from a signal from microphone $\mathbf{s}_m$ and a source $\mathbf{b}^n$, the corresponding reflector *k* is tangent to an ellipse of major diameter $d_{mk}^{n} = c\tau_{mk}^{n}$ and foci equal to $\mathbf{b}^n$ and $\mathbf{s}_m$.



$$C(d_{mk}^{n}, \mathbf{b}^n, \mathbf{s}_m) = \{\mathbf{x} \;:\; \|\mathbf{x} - \mathbf{b}^n\|_2^2 + \|\mathbf{x} - \mathbf{s}_m\|_2^2 = (d_{mk}^{n})^2\}$$

Antonacci et al. "Inference of room geometry from acoustic impulse responses." ***IEEE Trans. On Audio, Speech, and Lang. Proc., 2012.***

Conics in dual forms

Move to homogeneous coordinates: $\bar{\mathbf{x}}^{\top} = [\mathbf{x} \;\; 1]^{\top}$

Conics (ellipses) in homogeneous coordinates can be expressed as a 3 x 3 symmetric matrix:

$$\mathsf{C}^{n}_{mk} = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

such that: $C(d^{n}_{mk}, \mathbf{s}_m, \mathbf{b}^{n}) = \bar{\mathbf{x}}^{\top} \mathsf{C}^{n}_{mk} \bar{\mathbf{x}}$

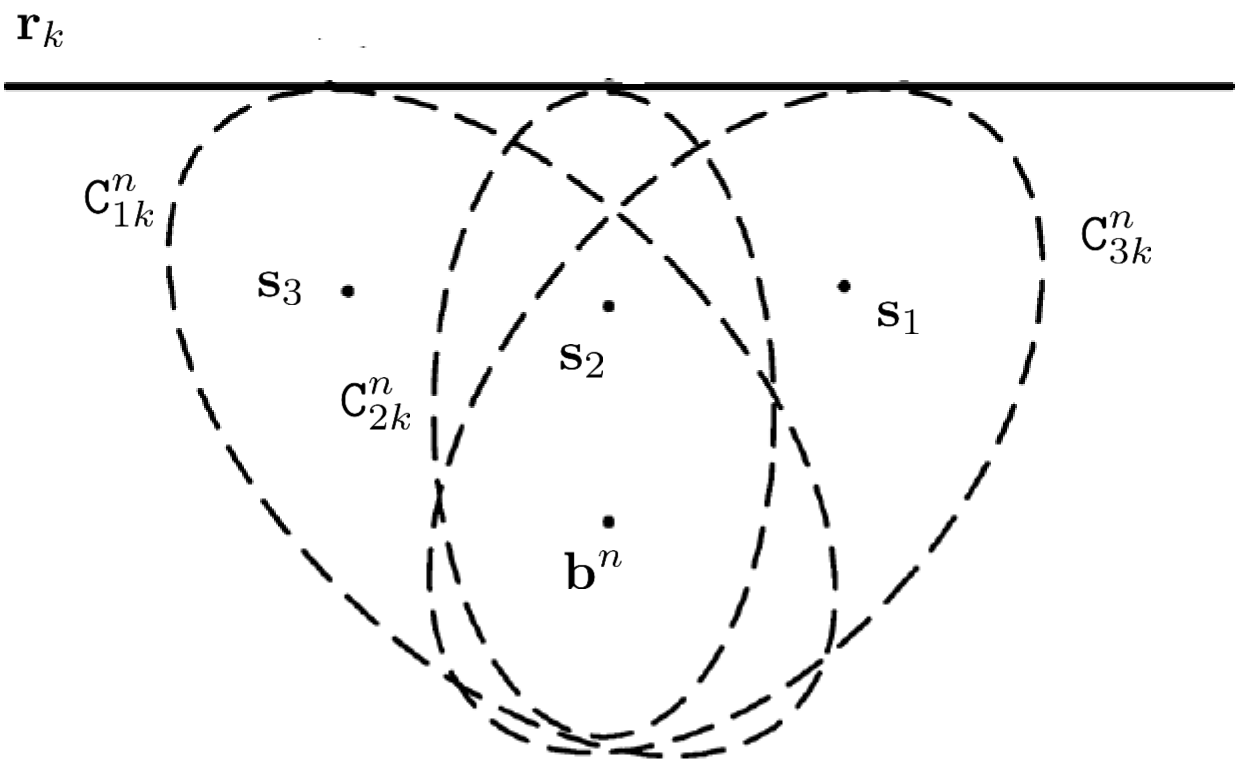
Dual form of a conic is parametrised by tangent lines to the conics rather than points. This gives the following matrix form:

$$\tilde{\mathsf{C}}^{n}_{mk} = {\mathsf{C}^{n}_{mk}}^{-1}$$

with a constraint on the reflector line such that: $\bar{\mathbf{r}}_k^{\top} \; \tilde{\mathsf{C}}^{n}_{mk} \; \bar{\mathbf{r}}_k = 0$

where $\bar{\mathbf{r}}_k$ is the homogeneous form of the vector normal to the line, whose modulus is equal to the line distance from the origin: $\bar{\mathbf{r}}_k = [\mathbf{r}_k^{\top} \;\; 1]^{\top}$

Multiple microphones



Multiple ellipses from delays related to different microphones define a unique common tangent reflector. This requires the solution of the labelling problem!

Antonacci et al. "Inference of room geometry from acoustic impulse responses." ***IEEE Trans. On Audio, Speech, and Lang. Proc., 2012.***

Cost function of quadrics

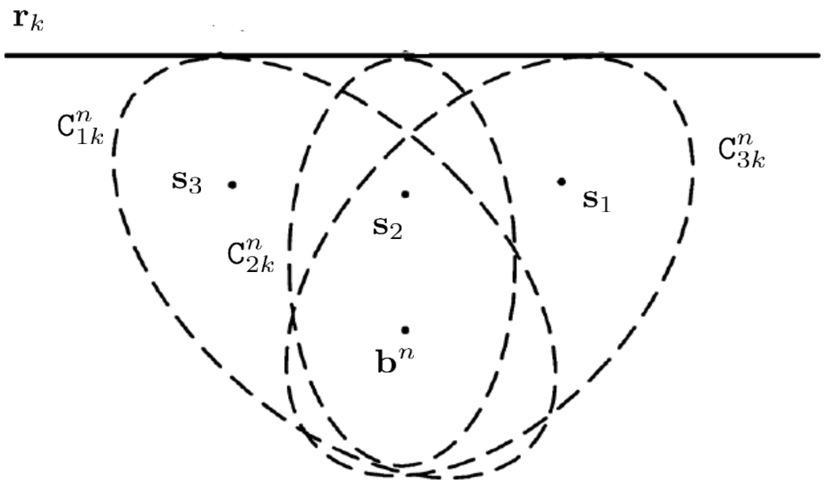
In real cases the relation:

$$\bar{\mathbf{r}}_k^\top \ \tilde{\mathtt{C}}_{mk}^n \ \bar{\mathbf{r}}_k = 0$$

holds only approximately

Minimize the nonlinear cost function:

$$H(\bar{\mathbf{r}}) = \sum_{m=1}^{M} \|\bar{\mathbf{r}}^\top \tilde{\mathtt{C}}_{mk}^n \bar{\mathbf{r}}^\top\|_2^2$$



- One of the homogeneous coordinates fixed in order to avoid the trivial solution.

- Taking the gradient = 0, one obtains a fourth order polynomial equations systems with a finite number of solutions.

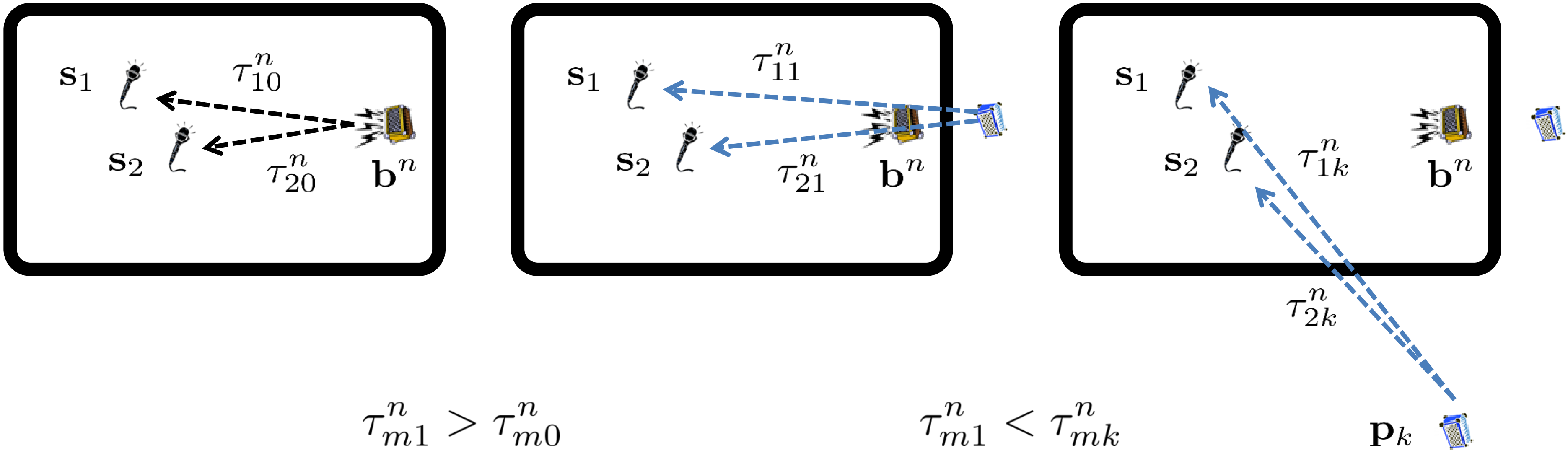
- Keep the solution giving the lowest value of the cost function.

Labeling problem: proposed solution

- For each reflector put the source as close as possible to it. In this way the first delay after the direct path should belong to the same reflector for all the microphones.

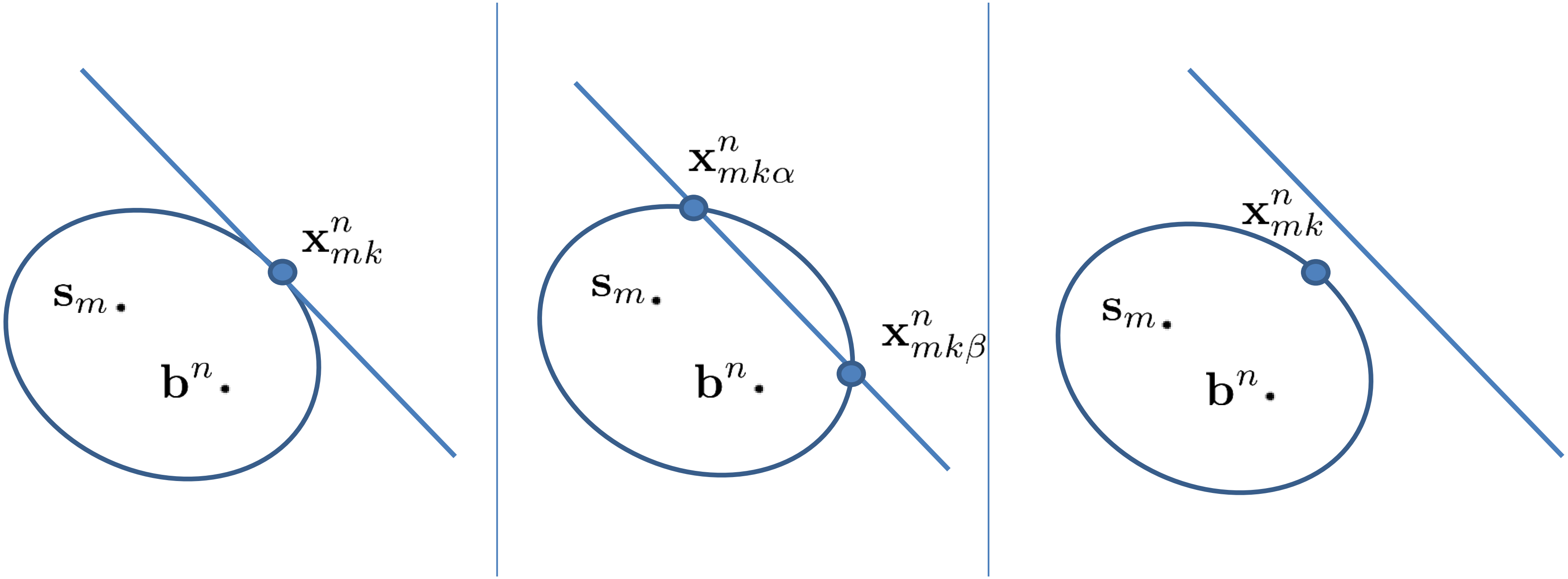
- Estimate the line reflector from the collected set of delays from the above procedure.

- Repeat the procedure for each reflector.



Solution Refinement (1)

Given an estimated reflector *k*, evaluate for each ellipse *nmk* the intersecting, tangent, or most close point to the reflector.

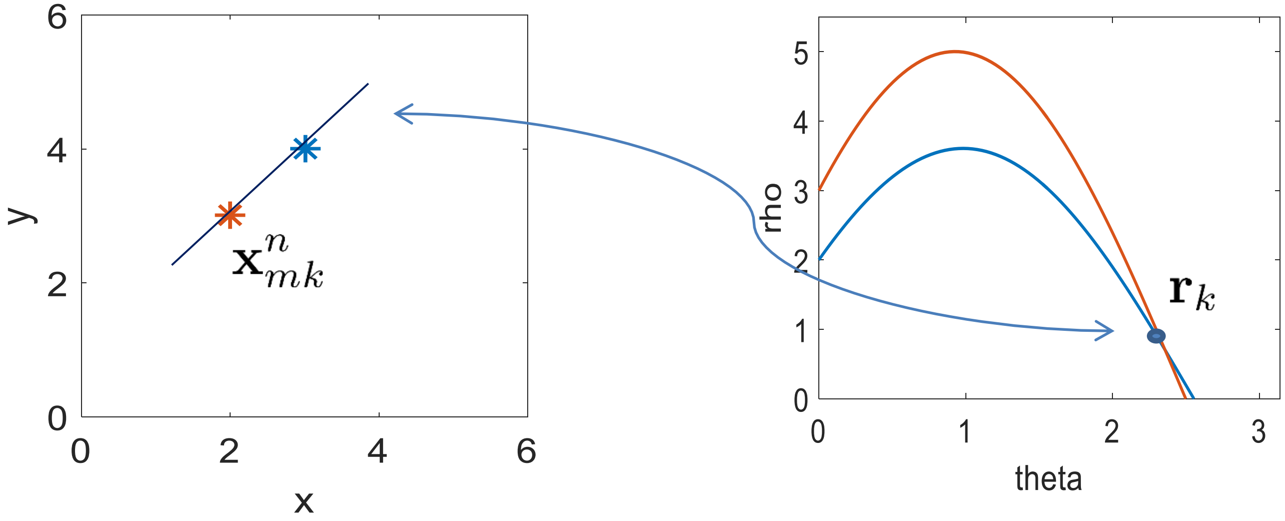


Antonacci et al. “Inference of room geometry from acoustic impulse responses.” *IEEE Trans. On Audio, Speech, and Lang. Proc., 2012.*

Refinement procedure (2): Hough Transform

Hough transform maps points on curves in (ρ, θ) space. If a set of points belongs to the same line, the corresponding curves will intersect in a single point.

$$\rho = x\ cos(\theta) + y\ sin(\theta)$$

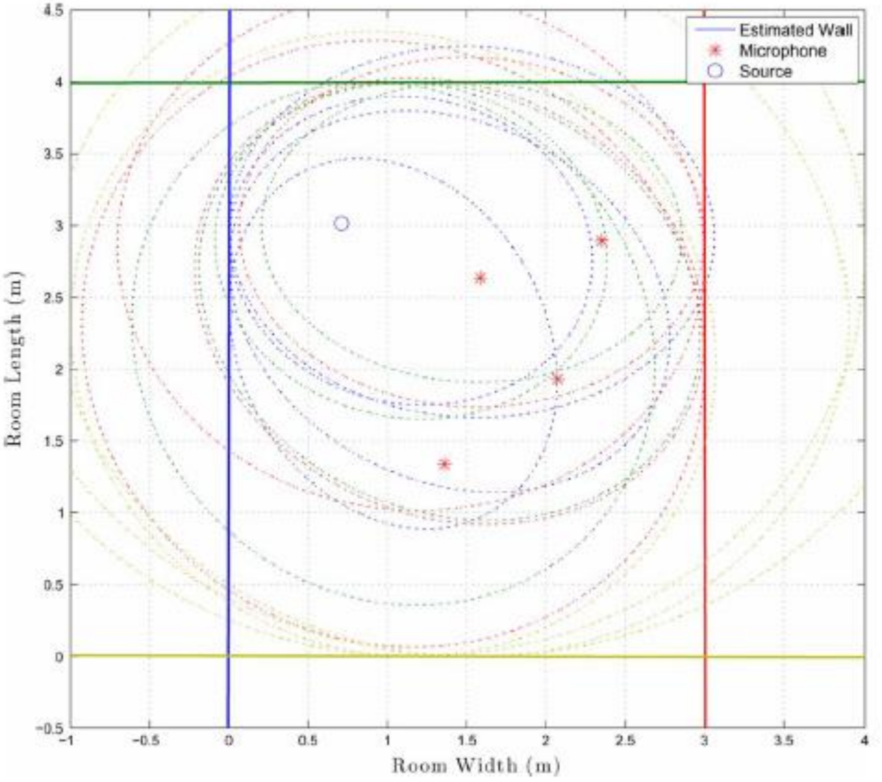


Refinement procedure (2)

Collect all the points $\mathbf{x}_{mk}^{n}$ and perform Hough transform discretizing the curves into a grid in (theta, rho) space.

Increment a counter for each grid point crossed by a curve in the transformed space.

Take the position of the *K* largest maxima in (theta , rho) space as the refined positions for the *K* reflectors.

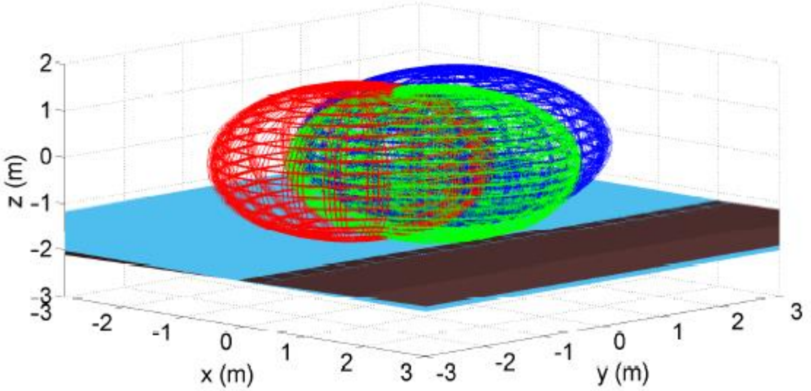


Antonacci et al., “Inference of room geometry from acoustic impulse responses.” *IEEE Trans. On Audio, Speech, and Lang. Proc.,* 2012.

Extension to 3D: ellipsoids

In 3D space ellipses becomes ellipsoids, and linear reflectors become planar reflectors

$$\mathsf{Q}^n_{mk} = \begin{bmatrix} a & b & c & d \\ b & e & f & g \\ c & f & h & i \\ d & g & i & l \end{bmatrix}$$



$$\tilde{\mathsf{Q}}^n_{mk} = {\mathsf{Q}^n_{mk}}^{-1}$$

$$\bar{\mathbf{r}}_k^\top \; \tilde{\mathsf{Q}}^n_{mk} \; \bar{\mathbf{r}}_k = 0$$

$$H(\bar{\mathbf{r}}) = \sum_{m=1}^{M} \|\bar{\mathbf{r}}^\top \tilde{\mathsf{Q}}^n_{mk} \bar{\mathbf{r}}^\top\|_2^2$$

Remaggi et al. "A 3D model for room boundary estimation." ICASSP 2015.

Pros and Cons

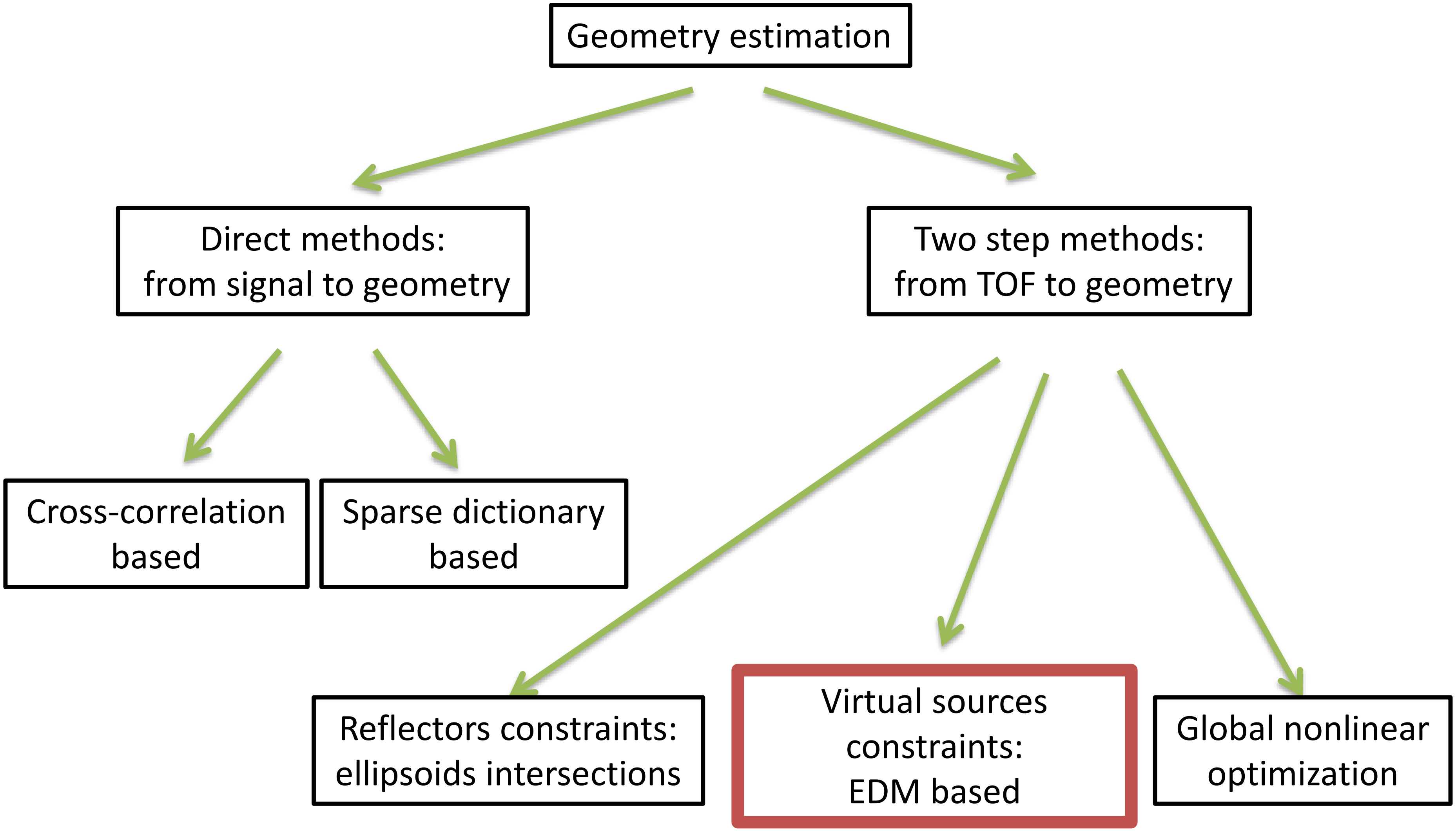
Pros

- Robustness to outliers (spurious delays) and echo labeling problem solved by local maxima search in the Hough space.

Cons

- Requires knowledge of both microphone and source positions.
- Source must be placed very close to each planar reflector.

Taxonomy



Geometry estimation

Direct methods:
from signal to geometry

Two step methods:
from TOF to geometry

Cross-correlation based

Sparse dictionary based

Reflectors constraints:
ellipsoids intersections

Virtual sources constraints:
EDM based

Global nonlinear optimization

Virtual sources constraints: EDM based

Requirements:

- One source, at least 4 microphones;
- Relative position among microphones assumed to be known.

Euclidean Distance Matrices $\mathbb{EDM}$

Given a set of 3D points $\mathbf{x}_i, \quad i = 1, \cdots, I$ an Euclidean Distance Matrix $\mathbb{EDM}$ is defined as a matrix $\mathtt{D}$ of pairwise squared Euclidean distances between points:

$$\mathtt{D}[i,j] = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$$

$\mathbb{EDM}$ has rank at most equal to 5:

$$\mathtt{D} = \underbrace{\mathbf{1} diag(\mathtt{X}^\top \mathtt{X})^\top}_{\text{Rank = 1}} - \underbrace{2\mathtt{X}^\top \mathtt{X}}_{\text{Rank = 3}} + \underbrace{diag(\mathtt{X}^\top \mathtt{X})\mathbf{1}^\top}_{\text{Rank = 1}}$$

$$\mathtt{X} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_I \end{bmatrix} \in \mathcal{R}^{I\times 3} \qquad \mathbf{1} \in \mathcal{R}^{I\times 1}$$

Dokmanić et al., "Acoustic echoes reveal room shape.“, PNAS 2013

Augmented $\mathbb{EDM}$

Build the $\mathbb{EDM}$ from the pairwise (known) microphone-microphone distances

$$\mathtt{D}[m_1, m_2] = \|\mathbf{s}_{m_1} - \mathbf{s}_{m_2}\|_2^2$$

Now, suppose to know the delay labeling and build a vector of distances related to the same reflector *k*:

$$\mathbf{d}_k = [d_{1k}^2,\ d_{2k}^2, \cdots,\ d_{Mk}^2]$$

Build an augmented matrix $\mathtt{D}_{aug}(\mathbf{d}_k)$ with the distances between the microphones and the *k*-virtual source .

$$\mathtt{D}_{aug} = \begin{bmatrix} \mathtt{D} & \mathbf{d}_k \\ \mathbf{d}_k^\top & 0 \end{bmatrix}$$

In absence of measurement errors on distances the matrix rank = 5 if and only if all the distances in vector $\mathbf{d}_k$ belong to the same virtual source *k.*

Dokmanić et al., "Acoustic echoes reveal room shape.“, PNAS 2013

Rank based approach (1)

Build a distance vector picking one TOF for each microphone:

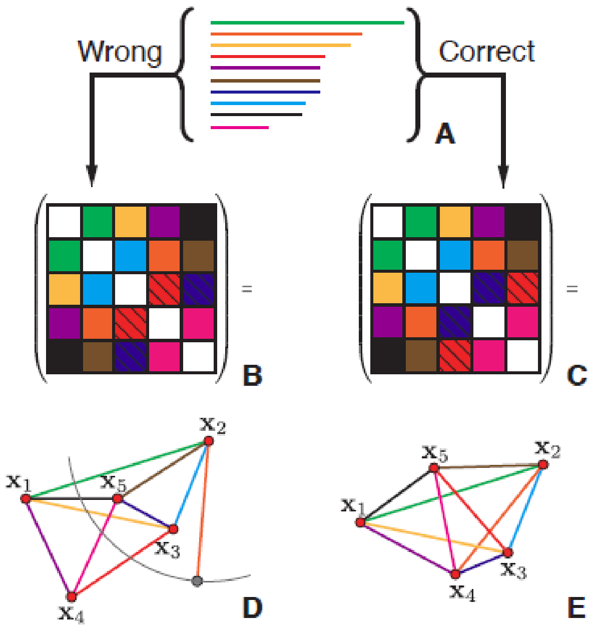
$$\mathbf{d}_{(i_1,\cdots,i_M)} = [d^2_{1i_1},\ d^2_{2i_2},\cdots,\ d^2_{Mi_M}]$$

d_{mi_m}: distance related to the TOF i_m from microphone *m*

Build an augmented distance matrix:

$$\mathtt{D}_{aug}(i_1,\cdots,i_M) = \begin{bmatrix} \mathtt{D} & \mathbf{d}_{(i_1,\cdots,i_M)} \\ \mathbf{d}^\top_{(i_1,\cdots,i_M)} & 0 \end{bmatrix}$$

Check the matrix rank of $\mathtt{D}_{aug}(i_1,\cdots,i_M)$ and retain the index set giving rank $\mathtt{D}_{aug}(i_1,\cdots,i_M)$< 6



If just 4 microphones are available, a modified augmented $\mathbb{EDM}$ with rank < 5 can be exploited, by subtracting a common row or column from $\mathtt{D}_{aug}$.

Dokmanić et al., "Acoustic echoes reveal room shape.“, PNAS 2013

S-stress approach

Rank checking is not robust to errors and noise in the TOF estimation

rely on S-stress measure:

$$s(i_1, \cdots, i_M) = \min_{\mathtt{E}} \|\mathtt{D}(i_1, \cdots, i_M) - \mathtt{E}\|_2^2, s.t. \;\; \mathtt{E} \in \mathbb{EDM}$$

$$\mathtt{E}[i, j] = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$$

S-stress measure of how close the matrix of measured squared distances is to an $\mathbb{EDM}$

The method ranks in ascending order the s-stress cost given by the above problem and it keeps the index sets yelding the best results.

Testing for all the index sets may be cumbersome, need to rely on heuristics:
Distance between TOFs from different microphones related to the same planar cannot be higher than the maximum distance between pairs of microphones.

Dokmanić et al., "Acoustic echoes reveal room shape.“, PNAS 2013

Solve for virtual sources

For each selected index set, pick the corresponding distances and solve for the virtual source positions

$$d_{mk}^2 = \|\mathbf{s}_m - \mathbf{p}_k\|_2^2 = \|\mathbf{s}_m\|_2^2 - 2\mathbf{s}_m^\top \mathbf{p}_k + \|\mathbf{p}_k\|_2^2$$

$$\tilde{d}_{mk}^2 = -\frac{1}{2}\left(d_{mk}^2 - \|\mathbf{s}_m\|_2^2\right) = \mathbf{s}_m^\top \mathbf{p}_k - \frac{1}{2}\|\mathbf{p}_k\|_2^2$$

$$\begin{bmatrix} \tilde{d}_{1k}^2 \\ \tilde{d}_{2k}^2 \\ \vdots \\ \tilde{d}_{Mk}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{s}_1^\top & -\frac{1}{2} \\ \mathbf{s}_2^\top & -\frac{1}{2} \\ \vdots & \\ \mathbf{s}_M^\top & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \mathbf{p}_k \\ \|\mathbf{p}_k\|_2^2 \end{bmatrix} \quad \Longrightarrow \quad \tilde{\mathbf{d}}_k = \mathsf{S} \begin{bmatrix} \mathbf{p}_k \\ \|\mathbf{p}_k\|_2^2 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{p}_k \\ \|\mathbf{p}_k\|_2^2 \end{bmatrix} = \mathsf{S}^\dagger \tilde{\mathbf{d}}_k$$

Given the virtual sources it is straightforward to recover corresponding reflectors $\mathbf{r}_k$

Pros and Cons

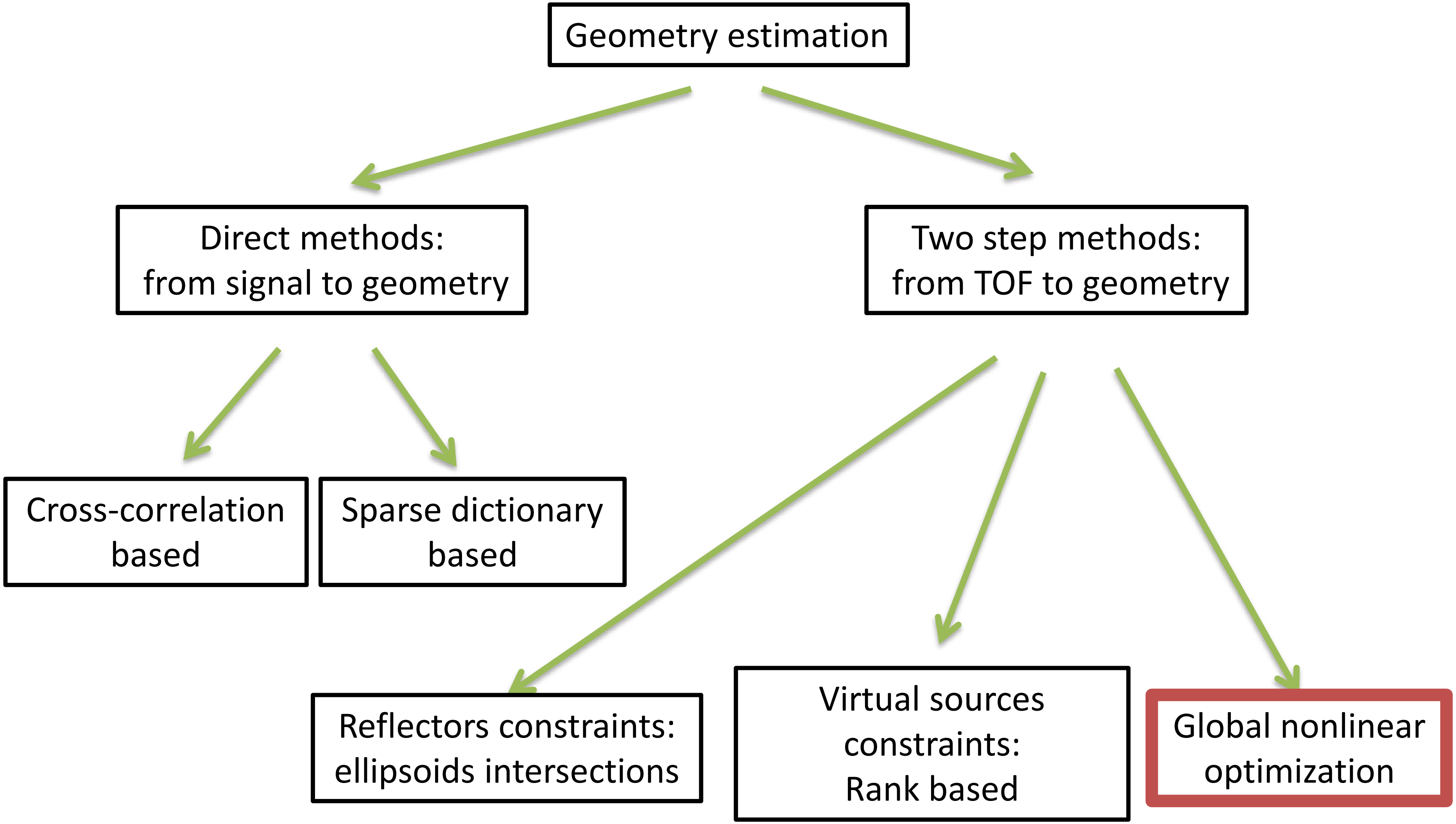
Pros

- Arbitrary (known) microphone displacements;
- Needs only first order reflections.

Cons

- Knowledge of relative microphone positions
- Evaluation of all TOF combinations may be cumbersome: heuristics needed to prune the selected sets.

Taxonomy



Geometry estimation

Direct methods: from signal to geometry

Two step methods: from TOF to geometry

Cross-correlation based

Sparse dictionary based

Reflectors constraints: ellipsoids intersections

Virtual sources constraints: Rank based

Global nonlinear optimization

Global nonlinear optimization

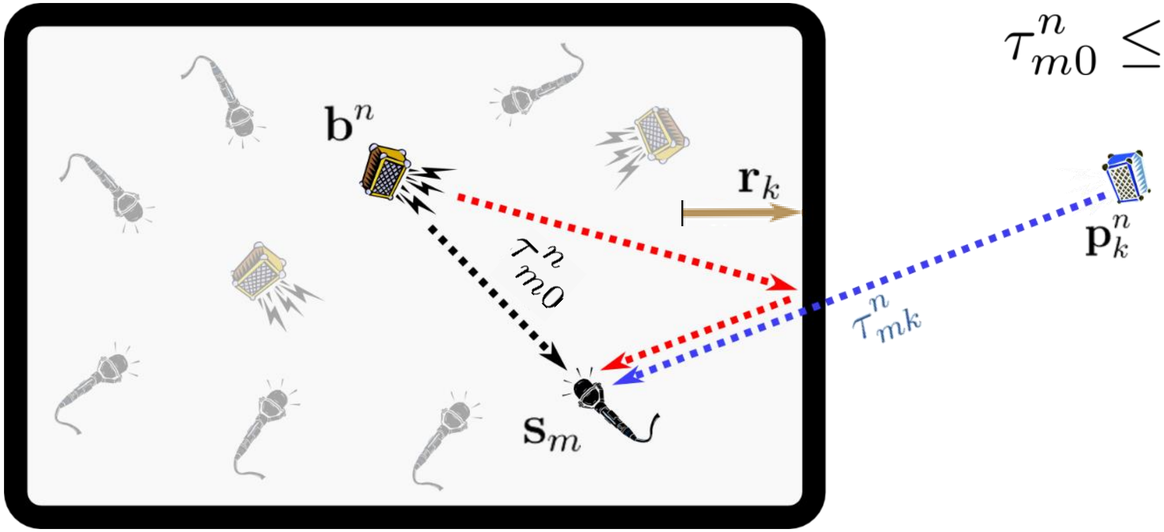
Requirements:

- RX signals at multiple microphones from multiple sources.
- Number of planar reflectors known.

- **Not needed:** knowledge of microphone and sources position and emission and offset times (**fully uncalibrated method**).

Exploit direct path properties

For each mic and source (m,n) the smallest delay of arrival corresponds to the direct path

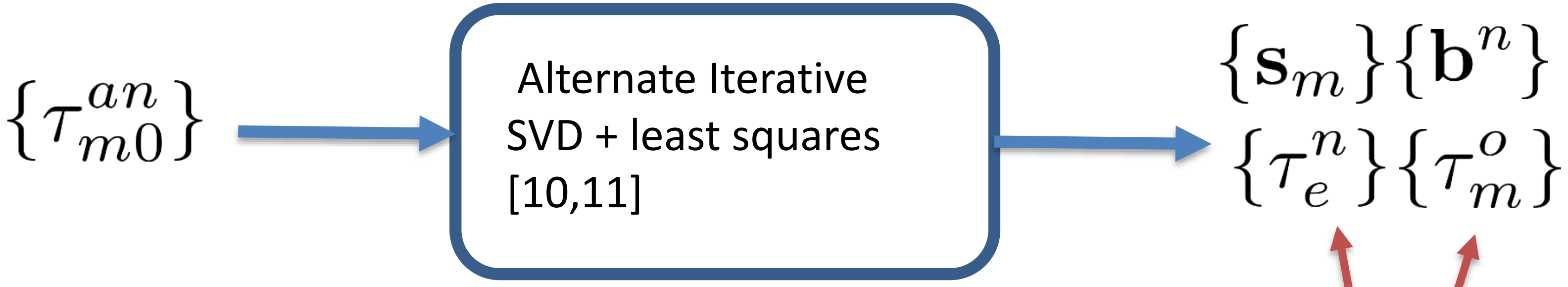


$$\tau^n_{m0} \leq \tau^n_{mk}$$

Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

Sources and mics positions from direct path delays

For each couple *n,m* sort the *K+1* delays in ascending order and take the first one.



$\{\tau^{an}_{m0}\}$

Alternate Iterative SVD + least squares [10,11]

$\{\mathbf{s}_m\}\{\mathbf{b}^n\}$

$\{\tau^n_e\}\{\tau^o_m\}$

Bilinear factorization problem (requires knowledge of emission and offset times)

Grounded on [11] : estimates also signal emission and offset times

[11] M. Crocco, A. Del Bue, and V. Murino, “A bilinear approach to the position self-calibration of multiple sensors,” *IEEE Transactions on Signal Processing*, vol. 60, pp. 660–673, 2012.

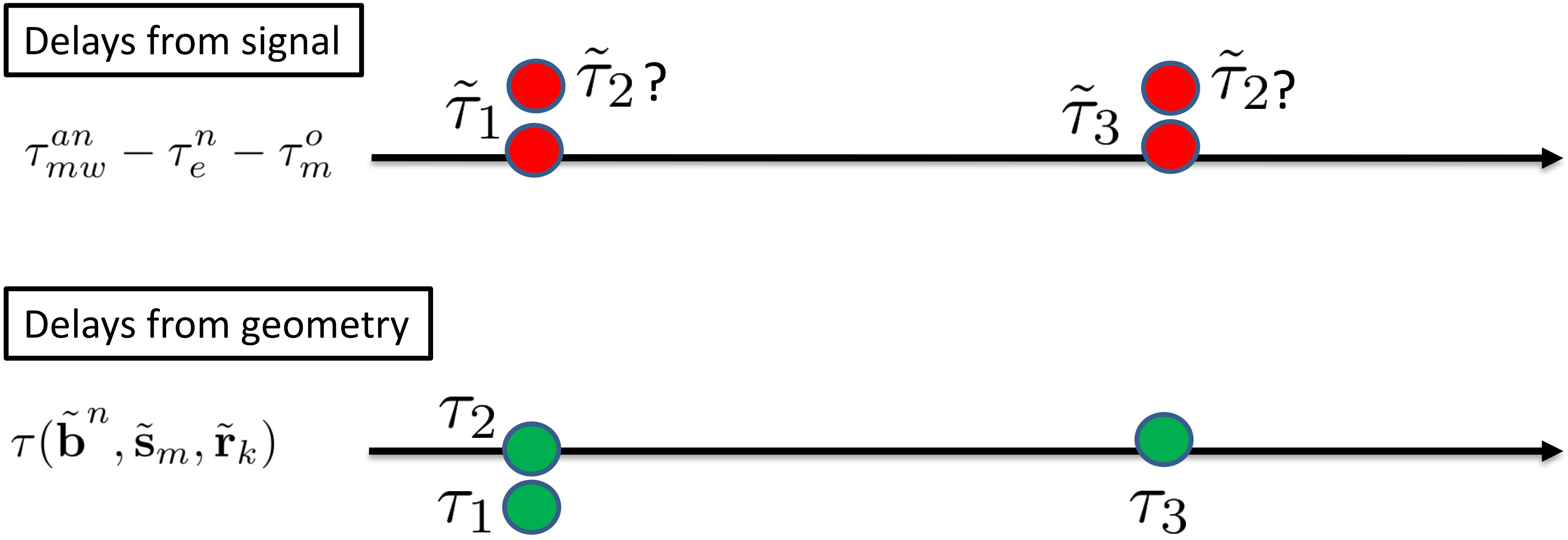
[10] Nikolay D. Gaubitch, W.Bastiaan Kleijn, and Richard Heusdens, “Auto-localization in ad-hoc microphone arrays,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 106–110.

Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

Ill-conditioning in presence of close delays

- If two or more delays are close to each other the problem is ill-conditioned.
- Same cost function minimum for a wrong delay reconstruction

Example of wrong delay reconstruction:



Delays from signal

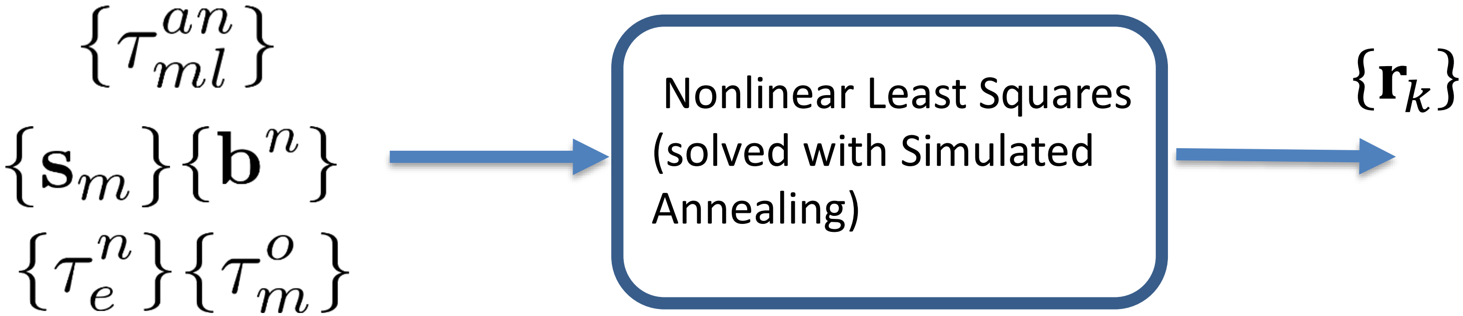
$\tau_{mw}^{an} - \tau_{e}^{n} - \tau_{m}^{o}$

Delays from geometry

$\tau(\tilde{\mathbf{b}}^{n}, \tilde{\mathbf{s}}_m, \tilde{\mathbf{r}}_k)$

Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

First guess estimation of walls coordinates



$\{\tau^{an}_{ml}\}$

$\{\mathbf{s}_m\}\{\mathbf{b}^n\}$

$\{\tau^n_e\}\{\tau^o_m\}$

Nonlinear Least Squares (solved with Simulated Annealing)

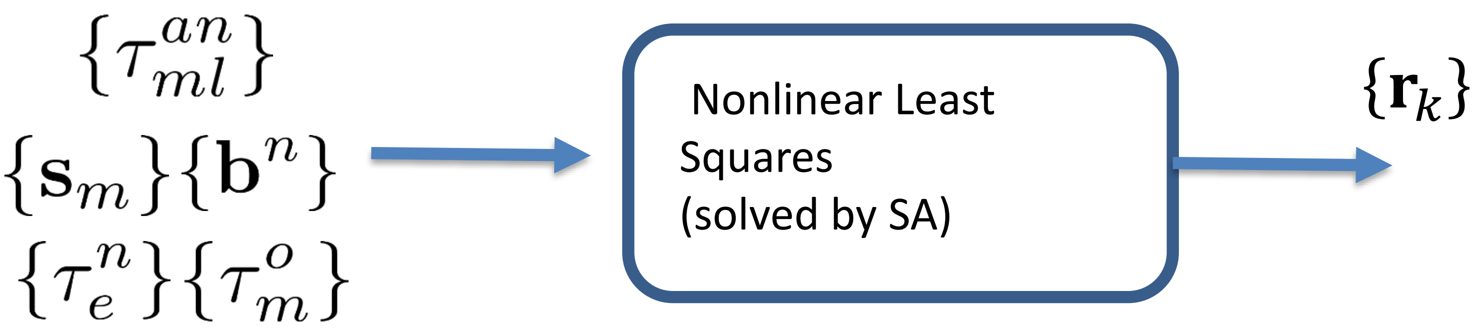
$\{\mathbf{r}_k\}$

$$\min_{\mathbf{r}_k} \sum_{nm} I(n,m) \sum_k \left(\tau^{an}_{mk} - \tau^n_e - \tau^o_m - \tau(\tilde{\mathbf{b}}^n, \tilde{\mathbf{s}}_m, \mathbf{r}_{h(k)})\right)$$

Index function sorting the set of delays in ascending order.
Matching problem between walls and delays is bypassed.

Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

First guess estimation of walls coordinates



$\{\tau^{an}_{ml}\}$

$\{\mathbf{s}_m\}\{\mathbf{b}^n\}$

$\{\tau^n_e\}\{\tau^o_m\}$

Nonlinear Least Squares
(solved by SA)

$\{\mathbf{r}_k\}$

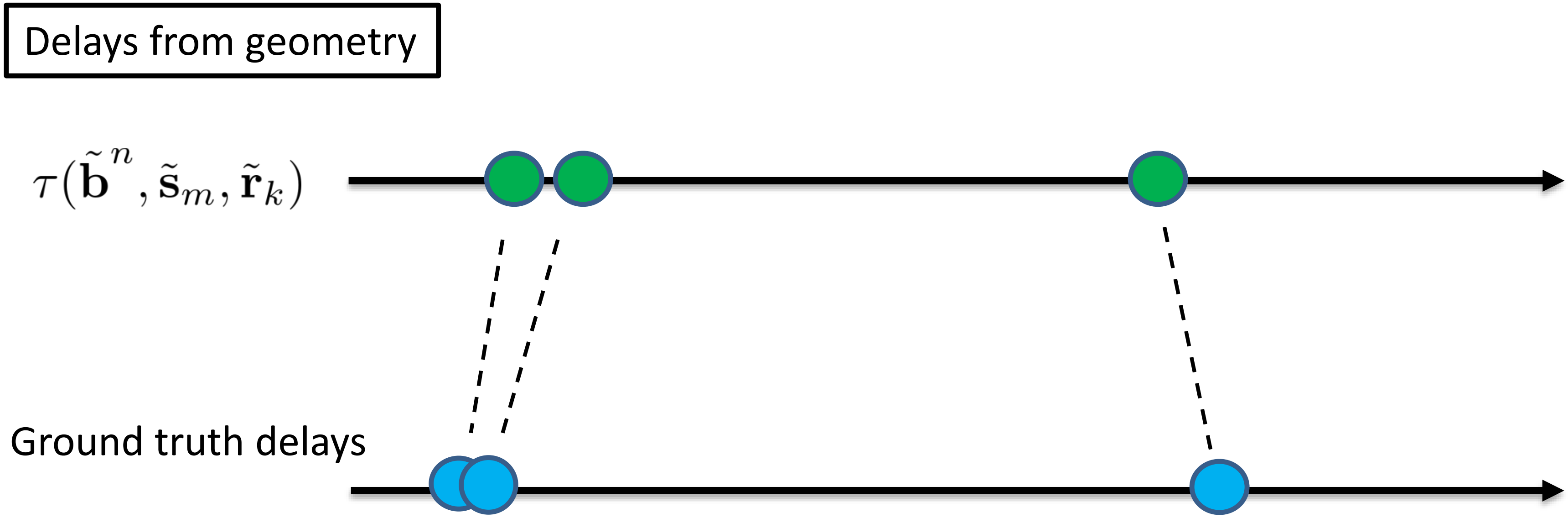
$$\min_{\mathbf{r}_k} \sum_{nm} I(n,m) \sum_k \left(\tau^{an}_{mk} - \tau^n_e - \tau^o_m - \tau(\tilde{\mathbf{b}}^n, \tilde{\mathbf{s}}_m, \mathbf{r}_{h(k)})\right)$$

Pruning strategy: if $\tau(\tilde{\mathbf{b}}^n, \tilde{\mathbf{s}}_m, \mathbf{r}_{h(k_1)}) \approx \tau(\tilde{\mathbf{b}}^n, \tilde{\mathbf{s}}_m, \mathbf{r}_{h(k_2)})$ for some $k_1, k_2 :\ I(n,m) = 0$.
Terms containing possible wrong delays estimations are pruned out.

Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

Fetching back ambiguous delays (1)

Delays estimated from geometry are not precise due to error accumulation along the procedure but are not subject to ambiguities.



Delays from geometry

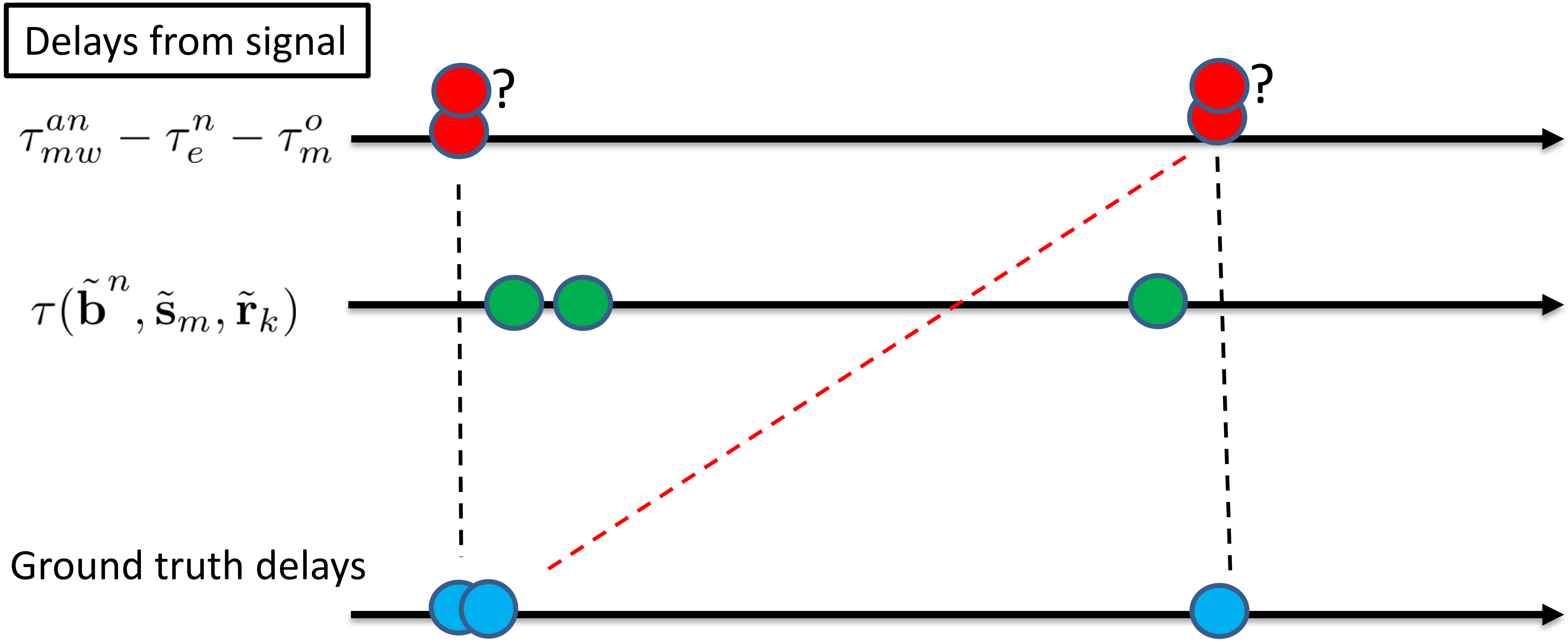
$\tau(\tilde{\mathbf{b}}^n, \tilde{\mathbf{s}}_m, \tilde{\mathbf{r}}_k)$

Ground truth delays

Crocco et al., "Towards Fully Uncalibrated Room Reconstruction with Sound", EUSIPCO 2014

Fetching back ambiguous delays (2)

Delays estimated from the signals are more precise but are subject to ambiguities.



Delays from signal

$\tau^{an}_{mw} - \tau^{n}_{e} - \tau^{o}_{m}$

$\tau(\tilde{\mathbf{b}}^{n}, \tilde{\mathbf{s}}_m, \tilde{\mathbf{r}}_k)$

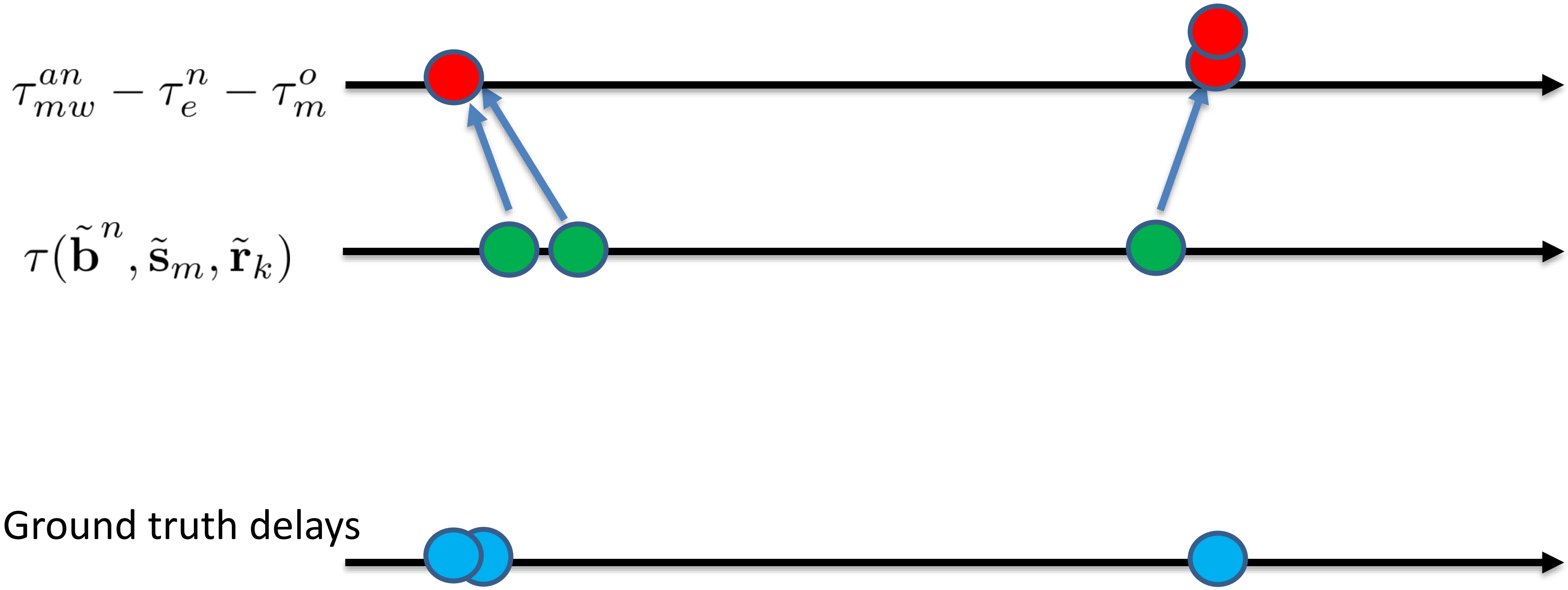
Ground truth delays

Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

Fetching back ambiguous delays (3)

Nearest neigbour procedure

$$\tilde{w}_k = \arg\min_{w} \left(\tau_{mw}^{an} - \tau_{e}^{n} - \tau_{m}^{o} - \tau(\tilde{\mathbf{b}}^{n}, \tilde{\mathbf{s}}_m, \tilde{\mathbf{r}}_k) \right)$$

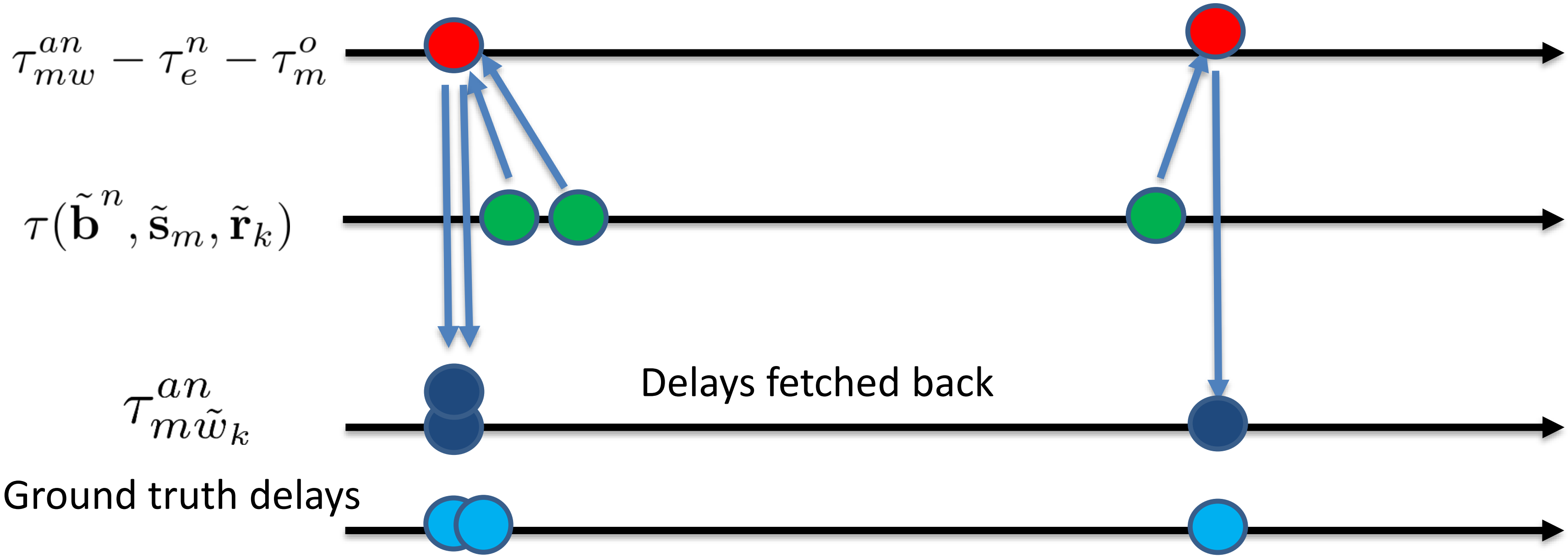

$\tau_{mw}^{an} - \tau_{e}^{n} - \tau_{m}^{o}$
$\tau(\tilde{\mathbf{b}}^{n}, \tilde{\mathbf{s}}_m, \tilde{\mathbf{r}}_k)$
Ground truth delays

Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

Fetching back ambiguous delays (4)

$$\tilde{w}_k = \arg\min_{w} \left(\tau^{an}_{mw} - \tau^{n}_{e} - \tau^{o}_{m} - \tau(\tilde{\mathbf{b}}^{n}, \tilde{\mathbf{s}}_m, \tilde{\mathbf{r}}_k) \right)$$

Final delays recovered back are precise and with no ambiguities



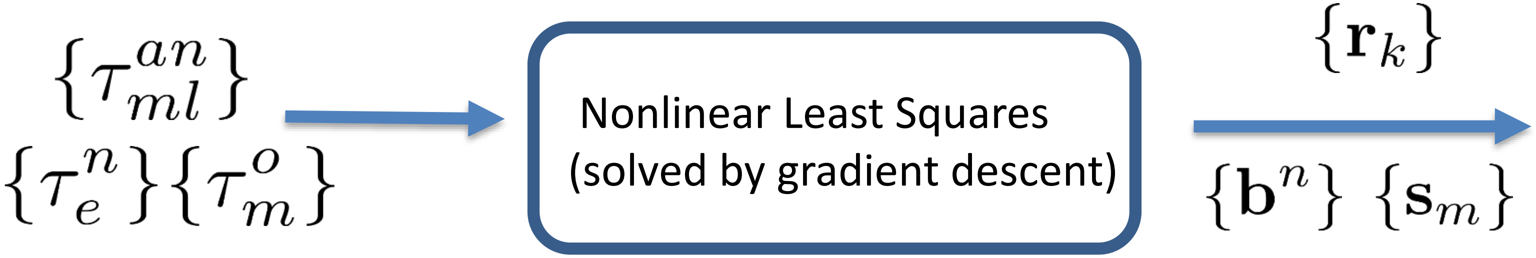
$\tau^{an}_{mw} - \tau^{n}_{e} - \tau^{o}_{m}$
$\tau(\tilde{\mathbf{b}}^{n}, \tilde{\mathbf{s}}_m, \tilde{\mathbf{r}}_k)$
$\tau^{an}_{m\tilde{w}_k}$
Delays fetched back
Ground truth delays

Crocco et al., "Towards Fully Uncalibrated Room Reconstruction with Sound", EUSIPCO 2014

Final geometric optimization

Jointly estimate **walls, sources and microphones** positions by using all the set of delays

$$\min_{\mathbf{b}^n,\mathbf{s}_m,\mathbf{r}_k} \sum_{nmk} \left(\tau^{an}_{m\tilde{w}_k} - \tau^{n}_{e} - \tau^{o}_{m} - \tau(\mathbf{b}^n, \mathbf{s}_m, \mathbf{r}_k)\right)^2$$



$\{\tau^{an}_{ml}\}$

$\{\tau^{n}_{e}\}\{\tau^{o}_{m}\}$

Nonlinear Least Squares
(solved by gradient descent)

$\{\mathbf{r}_k\}$

$\{\mathbf{b}^n\}$ $\{\mathbf{s}_m\}$

Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

Pros and Cons

Pros

- Fully uncalibrated method: no required knowledge of microphone and sources positions, as well as TX emission and RX offset times.
- Delay labeling problem bypassed by sorting delays at each iteraton of simulated annealing.

Cons

- Nonlinear, non-convex cost function involved, no guarantee to find the global maximum.
- Computationally demanding due to simulated annealing procedure.
- Limited robustness to missing or spurious delays (the problem might be alleviated adopting different loss functions like Huber).

NEXT: Dataset & evaluation