Multidim Syst Sign Process
DOI 10.1007/s11045-013-0233-8

TDOA-based acoustic source localization in the space–range reference frame

P. Bestagini · M. Compagnoni · F. Antonacci · A. Sarti · S. Tubaro

Received: 8 October 2012 / Revised: 8 February 2013 / Accepted: 11 March 2013

Abstract In this paper we consider a class of geometric methods for acoustic source localization based on range differences (or time differences of arrival), and we offer a new and unifying perspective on such methods based on the adoption of a multidimensional reference frame obtained by adding the range difference coordinate to the spatial coordinates of the source. In this extended coordinate system the working principles of a wide range of source localization methods becomes clear and immediate. The space–range reference frame, however, has a practical purpose as well, as it can be used for gaining insight on why some configurations of microphones lead to better localization performance than others and it suggests methods for improving existing localization techniques. In particular, we derive a closed-form solution of the constrained least-squares localization problem for linear arrays of microphones.

Keywords Acoustic source localization · Microphone array · Range difference · Time difference of arrival

P. Bestagini · F. Antonacci (✉)· A. Sarti · S. Tubaro
Dipartimento di Elettronica ed Informazione, Politecnico di Milano,
piazza Leonardo da Vinci, 20133 Milano, Italy
e-mail: antonacc@elet.polimi.it

P. Bestagini
e-mail: bestagini@elet.polimi.it

A. Sarti
e-mail: sarti@elet.polimi.it

S. Tubaro
e-mail: tubaro@elet.polimi.it

M. Compagnoni
Dipartimento di Matematica "Francesco Brioschi", Politecnico di Milano,
piazza Leonardo da Vinci, 20133 Milano, Italy
e-mail: marco.compagnoni@polimi.it

Published online: 29 March 2013

1 Introduction

Acoustic source localization has been a leading research topic in the audio and acoustics communities for quite a few decades. In the past two decades, however, this area of research has raised particular attention as applications to teleconferencing and hands-free communications, as well as distributed sensor networks have become to proliferate. Among the methodologies that are available in the literature, those based on measurements of Time Differences of Arrivals (TDOAs), taken with microphone arrays, have proved robust against reverberations and environmental noise. Relevant examples are Huang et al. (2000, 2001, 2004), Gillette and Silverman (2008), Smith and Abel (1987a,b), Abel and Smith (1987), Schau and Robinson (1987), Huang and Benesty (2004). In this manuscript we focus on this class of solutions, and we propose a novel representation for the related acoustic constraints, which sheds new light on acoustic source localization problems, as it offers a unifying perspective and additional insight.

Two important categories of solutions are those based on Maximum Likelihood (ML) and least squares (LS) criteria. In ML-based methods TDOA measurements are assumed to be affected by gaussian i.i.d. error and a nonlinear cost function is defined, whose minimum corresponds to the searched source location. Due to its nonlinearity, minimizing the ML cost function turns out to be a difficult task, therefore several simplifications have been proposed based on the LS criterion. More specifically, LS techniques localize the source by finding the minimum of a cost function that depends on the source location itself and on the distance between the reference microphone and the source. As the range depends on the source location itself, techniques based on a Constrained Least-Squares (CLS) have been proposed in the literature, which constrain the range to be equal to the distance between source and reference microphone.

Other works in the literature have given an insight into state-of-the-art acoustic source localization techniques. A relevant example is Huang and Benesty (2004), where an useful and interesting categorization is given. In this manuscript we keep this categorization, but we enrich this description with a geometric interpretation in the space–range reference frame. Due to the role played by the range in the localization procedure, in this manuscript we propose to visualize microphones and measurements in a space–range reference frame, obtained by adding the range difference coordinate (i.e. the distance of the point from the reference microphone) to the spatial coordinates of the source. In this multidimensional reference frame the working principle behind different localization techniques takes on new intuitive interpretation. This reference frame, in fact, was already adopted in Compagnoni et al. (2012) for localization purposes. In this manuscript we show how that multidimensional coordinate system can be use for understanding a wide range of source localization algorithms and, in some cases, how to improve them.

The rest of this manuscript is structured as follows: in Sect. 2 the space–range reference frame is introduced. Sections 3 and 4 offer a visual re-interpretation of various localization techniques in the 3D space, and discusses their limitations. Section 5 describes how to select configurations of microphones that lead to good localization results. This discussion is backed with some simulations.

2 The space–range coordinate system

Let us consider a microphone array whose sensors are in $\mathbf{m}_i = [x_i, y_i]^T, i = 0, \ldots, N$, as shown in Fig. 1. The 0th sensor is the reference microphone, which means that the Time

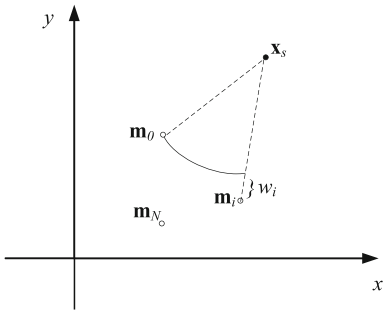


Fig. 1 Geometric setup of the localization problem

Differences Of Arrival (TDOAs) are all computed with respect to it. Without loss of generality we also assume that the origin of the reference frame is placed on the reference sensor, i. e. $\mathbf{m}_0 = [0, 0]^T$. The acoustic source is located at $\mathbf{x}_S = [x_S, y_S]^T$. The Time Of Arrival (TOA) of the ith sensor is the time of flight of the signal from the acoustic source to the sensor itself, and is therefore given by $\tau_i = ||\mathbf{x}_S - \mathbf{m}_i||/c$, c being the sound speed. However, as there is no synchronization between source and microphones, only TDOAs can be measured. In particular, the TDOA τ_{i0} is defined as the difference between the time of flights from the source to the ith microphone and from the source to the reference microphone

$$\tau_{i0} = (||\mathbf{x}_S - \mathbf{m}_i|| - ||\mathbf{x}_S||)/c \ .$$

The range difference w_i is defined as the difference between the distances from the source to the ith sensor and from the source to the reference sensor, therefore range difference and TDOA will be proportional to each other as

$$w_i = c\tau_{i0} = ||\mathbf{x}_S - \mathbf{m}_i|| - ||\mathbf{x}_S|| \ . \tag{1}$$

Because of the sampling and the noise in the acquisition chain the measurements of the TDOA $\hat{\tau}_{i0}$ will be affected by an error and will therefore produce noisy estimations $\hat{w}_i = c\hat{\tau}_{i0}$ of the range difference.

The goal of acoustic source localization from TDOA measurements is to produce an estimate $\hat{\mathbf{x}}_S$ of the true source location $\mathbf{x}_S$ through the analysis of the measured range differences $\hat{w}_i, \ i = 1, \ldots, N$.

As anticipated in the Introduction, we want to define a new reference frame based on spatial coordinates and range difference, in order to visualize and re-interpret the problem of source localization. Each point $\mathbf{x} = [x, y]^T$ on the space plane is mapped onto the 3D space–range $[\mathbf{x}^T, w]^T$, where

$$w = ||\mathbf{x}_S - \mathbf{x}|| - ||\mathbf{x}_S|| \ . \tag{2}$$

Using this definition, the w coordinate for the ith microphone is w_i defined in Eq.(1), and for the source is

$$w_S = -||\mathbf{x}_S|| \ . \tag{3}$$

The work in Compagnoni et al. (2012) considers a similar reference frame, but in that case $w_S = 0$. Even if not developed with this reference frame in mind, also state-of-the-art localization methodologies have a clear and intuitive representation in the space–range reference frame. In the following sections we revisit the theory of some methodologies, at least one for each class of localization techniques, with the goal of deriving their interpretation in the space–range reference frame.

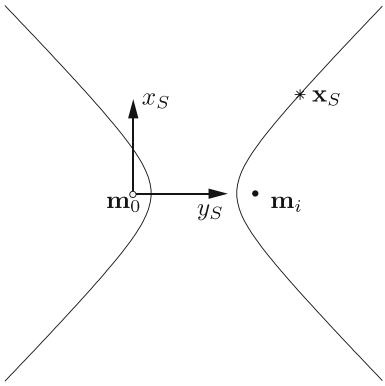


Fig. 2 The hyperbola described by a TDOA measurement. The *dots* are the foci of the hyperbola, where the microphones are. The *white dot* is the reference microphone. The source (marked by an *asterisk* $*$) is bound to lie on one of the two branches of the hyperbola

3 Maximum-likelihood TDOA localization

The hyperbolic Least Squares (LS) error on the ith microphone is defined in Huang and Benesty (2004) as

$$\begin{aligned} e_{h,i}(\mathbf{x}_S) &= \hat{w}_i - w_i(\mathbf{x}_S) \\ &= \hat{w}_i - (\|\mathbf{x}_S - \mathbf{m}_i\| - \|\mathbf{x}_S\|), \end{aligned} \tag{4}$$

therefore the corresponding hyperbolic LS cost function is

$$J_h(\mathbf{x}_S) = \sum_{i=1}^{N} e_{h,i}(\mathbf{x}_S)^2 \ . \tag{5}$$

If the measurements are affected by an additive white gaussian noise $\hat{w}_i$, the cost function (5) is proportional to the Maximum Likelihood (ML) cost function

$$J_{ML}(\mathbf{x}_S) = \sum_{i=1}^{N} \frac{[\hat{w}_i - w_i(\mathbf{x}_S)]^2}{\sigma^2} = \frac{J_h(\mathbf{x}_S)}{\sigma^2} \ ,$$

and the estimated source position is

$$\hat{\mathbf{x}}_{ML} = \arg\min_{\mathbf{x}_S}(J_h(\mathbf{x}_S)) \ .$$

A well-known geometric interpretation of the $e_{h,i}$ and J_h is in terms of hyperbolic curves in the geometric space of source locations. If $\hat{w}_i \ , i = 1, \ldots, N$ are noiseless we have the set of equations

$$e_{h,i}(\mathbf{x}_S) = \hat{w}_i - w_i = 0, \ i = 1, \ldots, N \tag{6}$$

in the variable $\mathbf{x}_S$, which define a branch of the hyperbola with foci $(\mathbf{m}_0, \mathbf{m}_i)$ and major axis length $|\hat{w}_i|$ (see Fig. 2).

If the measurements were noiseless, the source $\mathbf{x}_S$ would lie on the intersection of hyperbola. Because of measurement noise, however, the hyperbola of Eq. (6) will not simultaneously pass through a single point for $i = 1, \ldots, N$. The cost function (5) measures how well all all such equations are satisfied for a single value of $\mathbf{x}_S$.

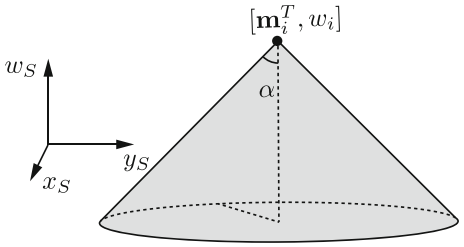


Fig. 3 The negative half-cone related to the ith microphone

In the $3D$ space–range reference frame, we can derive a second geometric interpretation of the hyperbolic errors and cost function. Using the coordinate w_S, Eq. (6) becomes

$$w_S - \hat{w}_i = -\|\mathbf{x}_S - \mathbf{m}_i\|. \tag{7}$$

Equation (7) in $(\mathbf{x}_S, w_S)$ defines a negative right circular half-cone, with apex $[\mathbf{m}_i^T, \hat{w}_i]^T$ and vertex angle $\alpha = \pi/4$, as shown in Fig. 3.

As for hyperbolas, with noiseless measurements the source $\mathbf{x}_S$ should fall in the intersection of all such half-cones, but because of the measurement noise this does not happen. Notice that the distance between $[\mathbf{x}_S^T, w_S]^T$ and the surface of the cone is

$$d = \frac{\sqrt{2}}{2}\left|\|\mathbf{x}_S - \mathbf{m}_i\| - (\hat{w}_i - w_S)\right| = \frac{\sqrt{2}}{2}\left|e_{h,i}(\mathbf{x}_S)\right|,$$

therefore the cost function (5) turns out to be proportional to the sum of the squared distances between the source and each cone and $\hat{\mathbf{x}}_{ML}$ is the point that minimizes the cost function. However the ML method minimizes the cost function $J_h(\mathbf{x}_S, w_S)$ with respect to the variables $\mathbf{x}_S$ only, implicitly accounting for the fact that w_S depends on the spatial coordinates. In fact the ML minimization could also be rewritten as

$$\hat{\mathbf{x}}_{ML} = \underset{\mathbf{x}_S, w_S}{\arg\min}(J_h(\mathbf{x}_S, w_S)) \quad \text{s.t.} \quad w_S = -\|\mathbf{x}_S\|,$$

which is a constrained optimization problem in the space–range coordinate system. Let us define the reference half-cone as

$$w_S = -\|\mathbf{x}_S\|. \tag{8}$$

This is the negative right circular half-cone with apex $[\mathbf{m}_0^T, 0]^T = [0, 0, 0]^T$ and vertex angle $\pi/4$. The ML estimated source position $\hat{\mathbf{x}}_{ML}$, as shown in Fig. 4, is therefore the point on the reference half-cone (8) that lies the closest to the remaining N half-cones (7).

The ML estimator is one of the most popular approaches to TDOA-based localization thanks to its well-established advantage of asymptotic efficiency for a wide sample space. However, the cost function (5) is highly nonlinear and in general its minimization is a difficult task. In order to overcome this problem, many methods based on different algebraic error definitions (statistically sub-optimal) have been developed Huang et al. (2000, 2001, 2004), Gillette and Silverman (2008), Smith and Abel (1987a,b), Abel and Smith (1987), Schau and Robinson (1987), Huang and Benesty (2004). In the next section we show the relation between the referenced techniques and ML, which becomes very easy to interpret in the space–range reference frame.

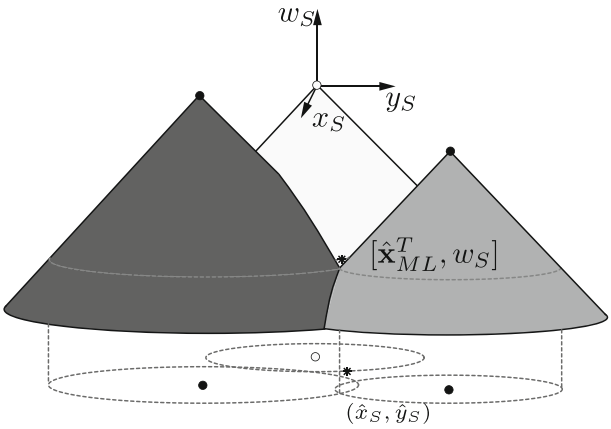


Fig. 4 ML localization technique. The reference microphone is marked by *open circle*, the other microphones by *filled circle*, the source is marked by *asterisk*. A projection of the intersecting cones on the horizontal plane is shown too. When the measurements are noisy, cones do not intersect and the ML optimization procedure localizes the source as the point on the reference half-cone that minimizes the cost-function

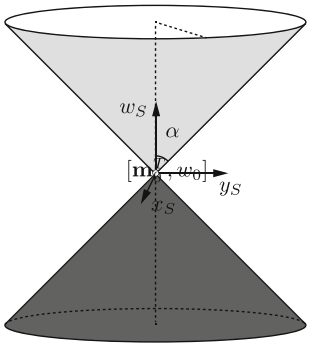


Fig. 5 The reference half-cone (*dark*) and reference double-cone (*dark* and *light*)

4 Towards a closed-form solution

In the previous section we showed the geometric interpretation of the ML estimator of the source position $\mathbf{x}_S$ as the constrained least square estimation of the closest point $\hat{\mathbf{x}}_S$ to the half-cone surfaces (7). In order to obtain algebraic cost functions, Eq. (7) is squared to obtain

$$\begin{aligned}(w_S - \hat{w}_i)^2 &= \|\mathbf{x}_S - \mathbf{m}_i\|^2 \\ &\Leftrightarrow (x_S - x_i)^2 + (y_S - y_i)^2 - (w_S - \hat{w}_i)^2 = 0,\end{aligned} \tag{9}$$

which defines a double (positive and negative) cone with the apex in $[\mathbf{m}_i^T, w_i]^T$.

In particular, the reference cone (8) becomes

$$w_S^2 = \|\mathbf{x}_S\|^2 \quad \Leftrightarrow \quad x_S^2 + y_S^2 - w_S^2 = 0, \tag{10}$$

as shown in Fig. 5.

These equations are the starting point to obtain all the main known least squares localization methods, as it offers a neat geometric interpretation of global validity.

4.1 Cone equation

As explained in Sect. 3, if the measurements $\hat{w}_i$ are noiseless, the source coincides with the intersection of all the cones. However, in a real-world scenario the measurements are noisy and (9,10) are not simultaneously satisfied. Defining the cone LS errors as

$$e_{c,0}(\mathbf{x}_S) = x_S^2 + y_S^2 - w_S^2, \tag{11}$$

$$e_{c,i}(\mathbf{x}_S) = (x_S - x_i)^2 + (y_S - y_i)^2 - (w_S - \hat{w}_i)^2, \tag{12}$$

the cone-based LS cost function

$$J_c(\mathbf{x}_S, w_S) = \sum_{i=0}^{N} e_{c,i}^2 \tag{13}$$

measures how cone equations are simultaneously satisfied.

A source localization method based on the cost function (13), although with a different derivation and geometrical interpretation, is proposed in Compagnoni et al. (2012). Here the authors exchange the roles of (x_S, y_S, w_S) and $(x_i, y_i, \hat{w}_i)$ in eqs. (9,10), interpreting them as $N+1$ conditions in the space–range reference frame on the points $[x_i, y_i, \hat{w}_i]^T$ to lie on a unique (propagation) cone with apex $[x_S, y_S, w_S]^T$. In the presence of noisy measurements the points do not perfectly fit this cone. In order to solve the localization problem, authors find the best fitting cone by minimizing the cost function (13), and estimate the source position $[\hat{\mathbf{x}}_C^T, \hat{w}_C]^T$ as the propagation cone vertex.

4.2 Plane equation

The cost function (13) is a polynomial of fourth degree in (x_S, y_S, w_S), therefore the minimization problem is again a non-trivial task. In order to simplify the problem we can manipulate the polynomial system that defines the cones. If we expand Eq. (9) and use Eq. (10), we obtain

$$x_i x_S + y_i y_S - \hat{w}_i w_S - \frac{1}{2}(x_i^2 + y_i^2 - \hat{w}_i^2) = 0. \tag{14}$$

These are linear equations in (x_S, y_S, w_S) for $i = 1, \ldots, N$, which define N planes Π_i in the space–range reference frame, with normal vector $\mathbf{n}_i = [x_i, y_i, -\hat{w}_i]^T$. Each plane Π_i contains the curve (conic section whose projection on the space plane is the hyperbola of Fig. 2) obtained as the intersection between the ith cone (9) and the reference cone (10).

As usual, with noiseless measurements the N planes intersect exactly in $\mathbf{x}_S$. In a real scenario we define the errors

$$e_{s,i}(\mathbf{x}_S) = x_i x_S + y_i y_S - \hat{w}_i w_S - \frac{1}{2}\left(x_i^2 + y_i^2 - \hat{w}_i^2\right).$$

These are exactly the spherical LS errors given in Huang and Benesty (2004). The spherical LS cost function

$$J_s(\mathbf{x}_S, w_S) = \sum_{i=1}^{N} e_{s,i}^2 \tag{15}$$

is a quadratic polynomial in (x_S, y_S, w_S) and gives a measure of how well the plane equations are satisfied. Notice that the spherical function (15) coincides with the cone function (13) if one assumes that the error $e_{c,0}$ on the reference microphones be equal to zero, i.e. if

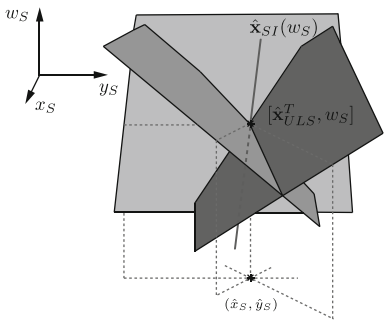


Fig. 6 Intersecting planes. The estimated location of the source and its projection on the horizontal plane are denoted by *asterisk*. The solution given by SI is the same

the reference cone equation (10) is satisfied. Geometrically speaking, this means that the restriction on the reference cone of the two functions is the same,

$$J_s(\mathbf{x}_S, w_S)|_{w_S^2=\|\mathbf{x}_S\|^2} = J_c(\mathbf{x}_S, w_S)|_{w_S^2=\|\mathbf{x}_S\|^2} \quad .$$

Notice that the difference between the two functions grows as we move away from the reference cone surface.

Because of its simplicity, many authors proposed localization methods based on the minimization of the spherical cost function. They differ exclusively on how the constraint (10) is taken into account.

4.2.1 Unconstrained least squares (ULS)

In the simplest approach to source localization based on the cost function (15), one simply discards the reference cone constraint. What remains to be solved is just the unconstrained minimization problem for the quadratic function (15)

$$\hat{\mathbf{x}}_{ULS} = \underset{(\mathbf{x}_S, w_S)}{\arg\min}(J_s(\mathbf{x}_S, w_S)).$$

Looking at the problem in the space–range reference frame, the ULS method searches for the point that best fits the equations of the planes (Fig. 6).

A first method for solving the ULS problem is discussed in Huang et al. (2000), Gillette and Silverman (2008) and it is known simply as the LS algorithm. Let us rewrite the cost function (15) in matrix form as

$$J_s(\mathbf{s}) = \|\mathbf{A}\mathbf{s} - \mathbf{b}\|^2,$$

where

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & -\hat{w}_1 \\ x_2 & y_2 & -\hat{w}_2 \\ \vdots & \vdots & \vdots \\ x_N & y_N & -\hat{w}_N \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} x_S \\ y_S \\ w_S \end{bmatrix},$$
$$\mathbf{b} = \frac{1}{2} \begin{bmatrix} x_1^2 + y_1^2 - \hat{w}_1^2 \\ x_2^2 + y_2^2 - \hat{w}_2^2 \\ \vdots \\ x_N^2 + y_N^2 - \hat{w}_N^2 \end{bmatrix}. \tag{16}$$

The gradient equation is

$$\frac{\partial J_s(\mathbf{s})}{\partial \mathbf{s}} = 2\mathbf{s}^T\mathbf{A}^T\mathbf{A} - 2\mathbf{b}^T\mathbf{A} = \mathbf{0}. \tag{17}$$

The estimated source position is therefore

$$[\hat{\mathbf{x}}_{LS}^T, \hat{w}_{LS}]^T = \hat{\mathbf{s}} = \arg\min_{\mathbf{s}}(J_s(\mathbf{s})) = \mathbf{A}^\dagger\,\mathbf{b}. \tag{18}$$

$\hat{\mathbf{s}}$ is the global minimum of the spherical cost function $J_s(\mathbf{s})$ in the space–range reference frame.

In Smith and Abel (1987a,b), Abel and Smith (1987) a different solution to the ULS problem has been proposed: the Spherical Interpolation (SI) algorithm. It is a two-step method. First the SI method solves the gradient equation (17) with respect to the $\mathbf{x}_S$ variable only, finding the estimated source position $\hat{\mathbf{x}}_{SI}$ as a linear function of w_S,

$$\hat{\mathbf{x}}_{SI}(w_S) = \mathbf{S}^\dagger\,(\mathbf{b} + w_S\hat{\mathbf{w}}), \tag{19}$$

where

$$\mathbf{S} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix},\quad \hat{w} = \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \vdots \\ \hat{w}_N \end{bmatrix}.$$

In the second step the algorithm substitutes the function $\hat{\mathbf{x}}_{SI}(w_S)$ into the linear system

$$\mathbf{As} - \mathbf{b} = \mathbf{0}\,. \tag{20}$$

If we define

$$\mathbf{P} := \mathbf{I} - \mathbf{SS}^\dagger,$$

the system (20) takes on the simpler form

$$\mathbf{P}(w_S\hat{\mathbf{w}} + \mathbf{b}) = \mathbf{0}\,. \tag{21}$$

This is an overdetermined linear system in the remaining variable w_S, which can be solved in a least-squares sense leading to

$$\hat{w}_{SI} = -\frac{\hat{\mathbf{w}}^T\mathbf{Pb}}{\hat{\mathbf{w}}^T\mathbf{P}\hat{w}} \quad\Rightarrow\quad \hat{\mathbf{x}}_{SI} = \mathbf{S}^\dagger\left[\mathbf{I} - \frac{\hat{\mathbf{w}}\hat{\mathbf{w}}^T\mathbf{P}}{\hat{\mathbf{w}}^T\mathbf{P}\hat{\mathbf{w}}}\right]\mathbf{b}. \tag{22}$$

Theorem 4.1 *LS and SI methods give the same solution to the ULS problem*

$$[\hat{\mathbf{x}}_{LS}^T, \hat{w}_{LS}]^T = [\hat{\mathbf{x}}_{SI}^T, \hat{w}_{SI}]^T.$$

Proof An algebraic proof of the theorem has been given in Huang and Benesty (2004). Here we offer an alternative interpretation by means of the space–range reference frame.

For each real value of $\overline{w}$, the function $\hat{\mathbf{x}}_{SI}(\overline{w})$ is bound to give the point lying on the plane $w_S = \overline{w}$ that minimizes the spherical cost function $J_s(\mathbf{x}_S, \overline{w})$

$$\hat{\mathbf{x}}_{SI}(\overline{w}) = \arg\min_{(\mathbf{x}_S, w_S)}(J_s(\mathbf{x}_S, w_S)) \quad \text{s.t.} \quad w_S = \overline{w}\,.$$

Geometrically speaking, $\hat{\mathbf{x}}_{SI}(\overline{w})$ is the point on the plane $w_S = \overline{w}$ that best fits the plane equations (14) and the function $\hat{\mathbf{x}}_{SI}(w_S)$ is a parametric description of the line in the $3D$

space–range coordinate system that contains all these points. Notice that the global minimum $\hat{\mathbf{s}}$ of the spherical cost function is bound to lie on this line. As a consequence, searching for the least square solution of the system (21) is completely equivalent to minimizing the spherical cost function $J_s(\mathbf{x}_S, w_S)$ along the line $\hat{\mathbf{x}}_{SI}(w_S)$, i.e.

$$\hat{w}_{SI} = \arg\min_{w_S}(\|\mathbf{P}(\mathbf{b} + w_S\hat{\mathbf{w}})\|^2) \tag{23}$$

$$= \arg\min_{(\mathbf{x}_S, w_S)}(J_s(\mathbf{x}_S, w_S)) \quad \text{s.t.} \quad \mathbf{x}_S = \hat{\mathbf{x}}_{SI}(w_S). \tag{24}$$

As $\hat{\mathbf{s}}$ lies on the line, the solution of the constrained minimization in (23) is $\hat{w}_{LS}$ and $\hat{\mathbf{x}}_{SI} = \hat{\mathbf{x}}_{SI}(\hat{w}_{LS}) = \hat{\mathbf{x}}_{LS}$. □

As shown above, while the LS algorithm finds the global minimum $\hat{\mathbf{s}}$ of the spherical cost function solving directly the gradient equation (17), the SI algorithm first solves the two spatial components of the gradient equation (17) and then minimizes the cost function along the third direction $\hat{\mathbf{x}}_{SI}(w_S)$.

Equations (18) and (22) give equivalent closed-form solutions of the ULS problem. However, since the constraint (8) has been completely discarded, in many situations the ULS yields poor localization accuracy, to the point of often becoming physically meaningless.

4.2.2 Constrained LS (CLS)

Many authors (Schau and Robinson 1987; Huang et al. 2001; Beck et al. 2008) suggest the reintroduction of the reference cone constraint (8) in order to increase the localization accuracy. This way we study the constrained minimization of the spherical cost function (15)

$$\hat{\mathbf{x}}_{CLS} = \arg\min_{(\mathbf{x}_S, w_S)}(J_s(\mathbf{x}_S, w_S)) \quad \text{s.t.} \quad w_S = -\|\mathbf{x}_S\|.$$

From the geometric standpoint, CLS searches for the point lying on the reference half-cone (8) that best fits the equations of the planes (see Fig. 7). In the literature there is no closed-form and exact solution. However, both closed-form approximate methods (Schau and Robinson 1987) and iterative ones (Huang et al. 2001) and, more recently, the exact iterative algorithm (Beck et al. 2008) have been developed. In the following, due to its relation to the SI algorithm and its immediate geometric interpretation, we will offer an in-depth discussion on the Spherical Intersection algorithm (SX) (Schau and Robinson 1987).

The SX algorithm is a modified version of the SI and is organized in two steps. Similarly to the SI method, the SX algorithm first finds the line (19)

$$\hat{\mathbf{x}}_{SI}(w_S) = \mathbf{S}^{\dagger}(\mathbf{b} + w_S\hat{\mathbf{w}}) ,$$

then it estimates the range $\hat{w}_{SX}$ by replacing $\hat{\mathbf{x}}_{SI}(w_S)$ into the reference (double) cone constraint (10), and by taking the negative solution of the resulting degree-two polynomial equation in w_S, i.e.

$$\|\hat{\mathbf{x}}_{SI}(w_S)\|^2 = w_S^2 ,$$

which finally leads to

$$\hat{\mathbf{x}}_{SX} = \hat{\mathbf{x}}_{SI}(\hat{w}_{SX}) .$$

Geometrically speaking, the SX method estimates the source position $[\hat{\mathbf{x}}_{SX}^T, \hat{w}_{SX}]^T$ as the intersection point between the line (19) and the reference half-cone (8) (see Fig. 8).

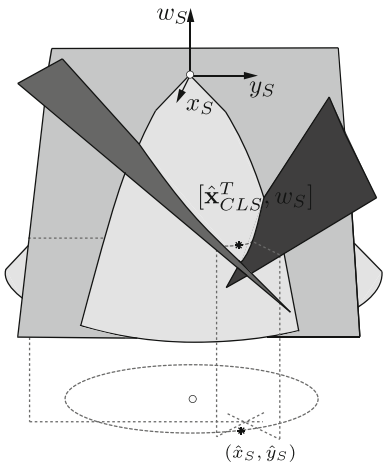


Fig. 7 The solution to the CLS problem is the point lying on the cone surface, which best fits the planes equations. A projection of the cone and planes on the horizontal plane is also shown

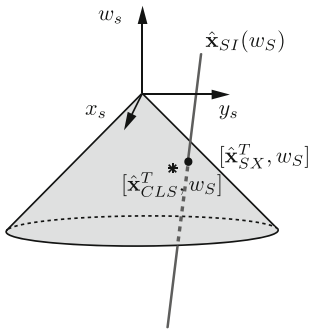


Fig. 8 The source location estimate $\hat{\mathbf{x}}_{SX}$ is the intersection between the line $\hat{\mathbf{x}}_{SI}(w_S)$ and the reference half-cone. The result is an approximation of the CLS solution $\hat{\mathbf{x}}_{CLS}$

The SX method is a closed-form estimator, but it suffers from some limitations. First, it fails if the line does not intersect the negative reference half–cone. Second, it only offers an approximate solution to the CLS problem. In fact $[\hat{\mathbf{x}}_{SX}^T, \hat{w}_{SX}]^T$ is the minimum of the spherical cost function (15) along the intersection of the reference cone with the plane $w_S = \hat{w}_{SX}$. There is no guarantee, however, that this is also the minimum of the cost function on the whole reference half-cone.

5 Discussion

In Sects. 3 and 4 we have given a unified geometric interpretation of the main localization methods. The 3D space–range reference frame is also a useful tool for understanding and interpreting the behavior of localization techniques. However, a complete geometric analysis of the source localization problem is beyond the scope of this manuscript. In order to show the potential of our geometric approach, in this section we study the ULS problem for two particular configurations of microphones in the case of noiseless measurements. Even in this advantageous condition, in fact, there are situations in which the ULS is unable to localize the source. Moreover, as we will see in the last section, this simplified analysis provides the starting point to predict some of the critical configurations of ULS in more realistic scenarios.

Since measurements are noiseless, we have

$$\hat{w}_i = w_i = \|\mathbf{x}_S - \mathbf{m}_i\| - \|\mathbf{x}_S\|.$$

Preliminarly, we present some simple results, which turn out to be useful in the next paragraphs.

Lemma 5.1 *Let* Π_i *be the planes defined by* Eq. (14) *for* $i = 1, \ldots, N$. *Then*

$$\bigcap_{i=1}^{N} \Pi_i \neq \varnothing.$$

Proof In a noiseless measurements scenario, the source $[\mathbf{x}_S^T, w_S]^T$ satisfies all the equations (14), thus

$$[\mathbf{x}_S^T, w_S]^T \in \bigcap_{i=1}^{N} \Pi_i \neq \varnothing.$$

□

Corollary 5.2 *The system* $\mathbf{As} = \mathbf{b}$ *admits solutions, therefore*

$$\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}|\mathbf{b}).$$

Corollary 5.3 *The solution of the ULS minimization is given by the intersection of the planes* (14)

$$\arg\min_{(\mathbf{x}_S, w_S)} (J_s(\mathbf{x}_S, w_S)) = \bigcap_{i=1}^{N} \Pi_i.$$

The above results are used in the next paragraphs.

5.1 Aligned microphones

In this paragraph we consider the case of aligned microphones, i.e. exists a line $r \subset \mathbb{R}^2$ containing all the microphones $\mathbf{m}_i$, $i = 0, \ldots, N$ (Fig. 9). Without loss of generality, we set

$$r: \; y = 0 \quad \Rightarrow \quad \mathbf{m}_i = [x_i, 0]^T, \quad i = 1, \ldots, N,$$

therefore

$$\mathbf{A} = \begin{bmatrix} x_1 & 0 & -w_1 \\ x_2 & 0 & -w_2 \\ \vdots & \vdots & \vdots \\ x_N & 0 & -w_N \end{bmatrix}.$$

We set r^0 as the smallest segment that contains all the microphones and $r^c := r \setminus r^0$ as its complement in r.

Theorem 5.4 *In the aligned scenario, the ULS does not have an unique solution. More specifically:*

- *if* $\mathbf{x}_S \notin r^c$, *we have*

$$\bigcap_{i=1}^{N} \Pi_i = l \subset \mathbb{R}^3,$$

where l *is a line with direction versor* $\mathbf{v} = [0, 1, 0]^T$;

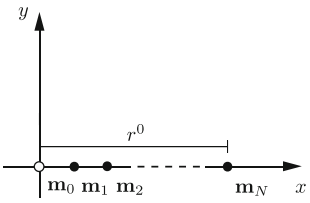


Fig. 9 Linear microphone array. *Dots* represent microphones, the *white* one being the reference one. In this example r^0 is the segment from $\mathbf{m}_0$ to $\mathbf{m}_N$

- *if* $\mathbf{x}_S \in r^c$, *we have*

$$\bigcap_{i=1}^{N} \Pi_i = \Pi \subset \mathbb{R}^3,$$

where Π *is a plane containing the origin and with normal versor* $\mathbf{n} = \frac{1}{\sqrt{2}}[\pm 1, 0, 1]^T$.

Proof We have $1 \leq \text{rank}(\mathbf{A}) \leq 2$, thus, by Lemma 5.1, there are only two possible cases.

- If $\text{rank}(\mathbf{A}) = 2$ the intersection of all the planes defines a linear subspace of dimension 1 in the space–range reference frame. In particular, planes Π_i determine a pencil of planes all passing through the line $l \subset \mathbb{R}^3$. Since the vector orthogonal to the ith plane is $\mathbf{n}_i = [x_i, 0, -w_i]^T$, $i = 1, \ldots, N$, the line l is parallel to the vector $\mathbf{v} = [0, 1, 0]^T$.
- If $\text{rank}(\mathbf{A}) = 1$,the planes Π_i are coincident.

We need, therefore, to determine the loci of points where $\text{rank}(\mathbf{A}) = 2$ and $\text{rank}(\mathbf{A}) = 1$, which depends on the configuration of sources and microphones in the acoustic scene. First, notice that $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}|\mathbf{b}) = 1$ if, and only if,

$$\text{rank} \begin{bmatrix} x_i & -w_i & b_i \\ x_j & -w_j & b_j \end{bmatrix} = 1 \quad \text{for each} \quad 1 \leq i < j \leq N.$$

We consider here the case $i = 1$, $j = 2$. We should therefore study the equation system

$$\begin{cases} x_1 w_2 = x_2 w_1 \\ x_1 b_2 = x_2 b_1 \\ w_1 b_2 = w_2 b_1 \end{cases}. \tag{25}$$

Let us assume that $\mathbf{x}_S \notin r$. From Eq. (16), the second line of (25) is

$$x_1(x_2^2 - w_2^2) = x_2(x_1^2 - w_1^2).$$

By replacing the first line of (25) and simplifying, we get

$$w_1 w_2 = x_1 x_2.$$

In the case $\mathbf{x}_S \notin r$ the triangular inequality implies that

$$|w_i| = |\|\mathbf{x}_S - \mathbf{m}_i\| - \|\mathbf{x}_S\|| < |x_i| \quad \Rightarrow \quad |w_1 w_2| < |x_1 x_2|,$$

which means that $\text{rank}(\mathbf{A}) = 2$ for each $\mathbf{x}_S \notin r$. In this case, the intersection of the planes is a line l passing through the source position $[\mathbf{x}_S^T, w_S]^T$.
Conversely, if $\mathbf{x}_S = [x_S, 0]^T \in r$ we have

$$w_i = |x_S - x_i| - |x_S|.$$

It is then straightforward to verify that

$$\det \begin{bmatrix} x_i & -w_i \\ x_j & -w_j \end{bmatrix} = -x_i w_j + x_j w_i = 0, \quad 1 \leq i < j \leq N, \tag{26}$$

are satisfied if, and only if, $\mathbf{x}_S \in r^c$. Moreover, in this case we have $b_i = 0$ and $\mathbf{n}_i = \frac{1}{\sqrt{2}}[\pm 1, 0, 1]^T$, for each $i = 1, \ldots, N$. □

The above theorem confirms that both LS and SI algorithms fail to localize the source, because do not carry any information on the coordinate y_S of the source, no matter what the noise level of the measurements is. However, the situation changes if we consider the reference half-cone constraint (8). In the aligned scenario the above analysis suggests us the way to obtain the closed-form and exact resolution of the CLS problem, working even in the presence of noisy $\hat{w}_i$. However, we observe that, in the general noisy case, the planes Π_i could have trivial intersection. Anyway, with the exception of the special situations where $\text{rank}(\mathbf{A}) = 1$, the solution of the ULS minimization is still a line $l \subset \mathbb{R}^3$ with direction versor $\mathbf{v} = [0, 1, 0]^T$, because the spherical cost function does not depend on y_S. The ULS solution for (x_S, w_S) is given by

$$[\hat{x}_{ULS}, \hat{w}_{ULS}]^T = \mathbf{A}_{\mathbf{R}}^{\dagger} \mathbf{b},$$

where

$$\mathbf{A}_{\mathbf{R}} = \begin{bmatrix} x_1 & -\hat{w}_1 \\ x_2 & -\hat{w}_2 \\ \vdots & \vdots \\ x_N & -\hat{w}_N \end{bmatrix}.$$

Before stating the theorem 5.6 on the CLS, we give some preliminary properties about ULS and the plane Π_{xw} defined by $y_S = 0$. If $\text{rank}(A) = 2$, the restriction $J_s|_{\Pi_{xw}}$ is a positive quadratic form, with the unique global minimum $[\hat{x}_{ULS}, 0, \hat{w}_{ULS}]^T$, which is the projection of l on Π_{xw}. Moreover, let us define the open region

$$U := \{[x_S, 0, w_S]^T \mid x_S < -w_S, \ x_S > w_S\} \subset \Pi_{xw}. \tag{27}$$

Its closure $\bar{U}$ is a convex closed set and it is the projection of the negative reference half-cone (8) on Π_{xw}. In fact, we observe that the generatrices $x_S = \pm w_S$ divide Π_{xw} into four regions and it is straightforward to verify that any point on the negative half-cone region (e.g. $[0, 0, -1]^T$) satisfies the inequalities (27) (see Fig. 10).

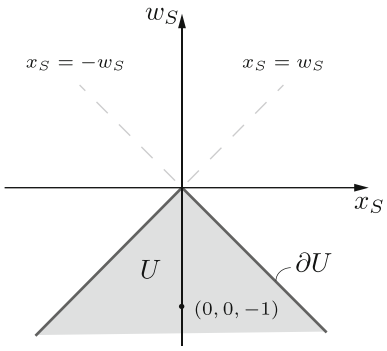


Fig. 10 The U region highlighted in *gray*

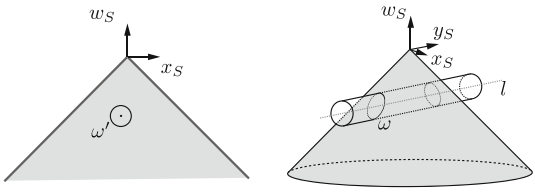


Fig. 11 Reference cone intersected by l. The open neighborhood ω and its projection into the $\bar{U}$ region are shown

Finaly, the boundary of U is

$$\partial U := \{[x_S, 0, w_S]^T \mid x_S = \pm w_S, w_S \leq 0\}.$$

Lemma 5.5 *The point* $[\hat{\mathbf{x}}_{CLS}^T, \hat{w}_{CLS}]^T$ *on the reference cone is a local minimum of CLS if, and only if, its projection* $[\hat{x}_{CLS}, 0, \hat{w}_{CLS}]^T \in \Pi_{xw}$ *is a local minimum of* $J_s|_{\Pi_{xw}}$ *on* $\bar{U}$.

Proof Let us assume that $[\hat{\mathbf{x}}_{CLS}^T, \hat{w}_{CLS}]^T$ be a CLS minimum point, then there exists an open neighborhood ω of the point on the reference cone surface, such that

$$J_s(\mathbf{x}_S, w_S) \geq J_s(\hat{\mathbf{x}}_{CLS}, \hat{w}_{CLS}), \tag{28}$$

for each $[\mathbf{x}_S^T, w_S]^T \in \omega$. Let us define $\omega' \subset \bar{U}$ as the projection on Π_{xw} of the set ω. It is an open neighborhood of $[\hat{x}_{CLS}, 0, \hat{w}_{CLS}]^T$ in $\bar{U}$ (Fig. 11).

Since J_s does not depend on y_S, we have $J_s(x_S, 0, w_S) \equiv J_s(\mathbf{x}_S, w_S)$, hence inequality (28) implies

$$J_s(x_S, 0, w_S) \geq J_s(\hat{x}_{CLS}, 0, \hat{w}_{CLS}), \tag{29}$$

for each $[x_S, 0, w_S]^T \in \omega'$. This proves one direction of the lemma. The converse is similar. □

Theorem 5.6 *Assume that the microphones be aligned and* $\mathrm{rank}(A) = 2$. *Then:*

- *if* $[\hat{x}_{ULS}, 0, \hat{w}_{ULS}]^T \in U$, *the solutions of the CLS are two points*

$$\hat{\mathbf{x}}_{CLS}^{\pm} = \left[\hat{x}_{ULS}, \pm\sqrt{-\hat{x}_{ULS}^2 + \hat{w}_{ULS}^2}\right]^T; \tag{30}$$

- *if* $[\hat{x}_{ULS}, 0, \hat{w}_{ULS}]^T \notin U$, *the solution of the CLS is exactly one of the following three points:*

$$\hat{\mathbf{x}}_{CLS}^{+} = \left[\frac{\sum(x_i - \hat{w}_i)^2(x_i + \hat{w}_i)}{\sum(x_i - \hat{w}_i)^2}, 0\right]^T, \tag{31}$$

$$\hat{\mathbf{x}}_{CLS}^{-} = \left[\frac{\sum(x_i + \hat{w}_i)^2(x_i - \hat{w}_i)}{\sum(x_i + \hat{w}_i)^2}, 0\right]^T, \tag{32}$$

$$\hat{\mathbf{x}}_{CLS}^{0} = [0, 0]^T. \tag{33}$$

Proof First, we study the minimization problem of $J_s|_{\Pi_{xw}}$ on $\bar{U}$. Since the restriction $J_s|_{\Pi_{xw}}$ is a positive quadratic form and $\bar{U}$ is a convex closed set, there exists one, and only one, global minimum $[\hat{x}_{CLS}, 0, \hat{w}_{CLS}]^T$. By Lemma 5.5, this point is the projection on the plane Π_{xw} of the solutions $[\hat{\mathbf{x}}_{CLS}^T, \hat{w}_{CLS}]^T$ of CLS.
If $[\hat{x}_{ULS}, 0, \hat{w}_{ULS}]^T \in U$, then it has to match with $[\hat{x}_{CLS}, 0, \hat{w}_{CLS}]^T$. In this case the line l passes through $[\hat{x}_{CLS}, 0, \hat{w}_{CLS}]^T$ and intersects the negative reference half-cone (8). Hence,

the solutions $\hat{\mathbf{x}}^{\pm}_{CLS}$ of the CLS are the projections of the intersection points on the horizontal plane $w_S = 0$. If we replace the equation of l into Eq. (10) we get the solution in (30).

If $[\hat{x}_{ULS}, 0, \hat{w}_{ULS}]^T \notin U$, the constrained minimum of $J_s|_{\Pi_{xw}}$ is located on ∂U, which means that J_s, in turn, can have the constrained minimum only at

$$\hat{\mathbf{x}}^{\pm}_{CLS} = \underset{(\mathbf{x}_S, w_S)}{\arg\min}(J_s(\mathbf{x}_S, w_S)) \quad \text{s.t.} \quad y_s = 0, \; w_S = \pm x_S \tag{34}$$

or at $\hat{\mathbf{x}}^{0}_{CLS}$, that is the unique irregular point of ∂U. If we keep into account of the constraint in (34), we get the result in (31)–(33). □

Summarizing, Theorem (5.6) gives us the prescription to obtain the exact solution of the CLS minimization problem, which, by Lemma 5.5, is essentially equivalent to the constrained minimization problem of $J_s|_{\Pi_{xw}}$ on $\bar{U}$. Notice that

- if $[\hat{x}_{ULS}, 0, \hat{w}_{ULS}]^T \in U$ we have two solutions $\hat{\mathbf{x}}^{\pm}_{CLS}$, which correspond to the two intersections of l with the reference half-cone and it is not possible to discriminate between them;
- if $[\hat{x}_{ULS}, 0, \hat{w}_{ULS}]^T \notin U$, it is first necessary to verify if $\hat{\mathbf{x}}^{\pm}_{CLS}$ lies on the negative reference half-cone. The estimated source position is then the admissible point where the spherical function J_s assumes the minimum value.

5.2 Microphones at the vertices of a square

A necessary condition for the ULS methods to be able to localize the source is that the microphones are not collinear. On the other hand, one might wonder if this is also a sufficient condition. In this Subsection we prove that other configurations could cause problems in the localization. Let us assume a four microphones setup, one at each vertex of a unit square. Without loss of generality, we set

$$\mathbf{m}_0 = [0, 0]^T, \quad \mathbf{m}_1 = [1, 0]^T, \quad \mathbf{m}_2 = [1, 1]^T, \quad \mathbf{m}_3 = [0, 1]^T.$$

Therefore, we have

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & -w_1 \\ 1 & 1 & -w_2 \\ 0 & 1 & -w_3 \end{bmatrix},$$

where

$$\begin{aligned} w_1 &= \sqrt{(x_S - 1)^2 + y_S^2} - \sqrt{x_S^2 + y_S^2}, \\ w_2 &= \sqrt{(x_S - 1)^2 + (y_S - 1)^2} - \sqrt{x_S^2 + y_S^2}, \\ w_3 &= \sqrt{x_S^2 + (y_S - 1)^2} - \sqrt{x_S^2 + y_S^2}. \end{aligned} \tag{35}$$

Theorem 5.7 *In the above scenario, ULS does not have a unique solution if, and only if, the source* $\mathbf{x}_S$ *lies on one of the two straight lines*

$$r \subset \mathbb{R}^2 : \; x = \frac{1}{2} \quad \textit{and} \quad r' \subset \mathbb{R}^2 : \; y = \frac{1}{2}.$$

In these cases, in fact, we have

$$\bigcap_{i=1}^{N} \Pi_i = l,$$

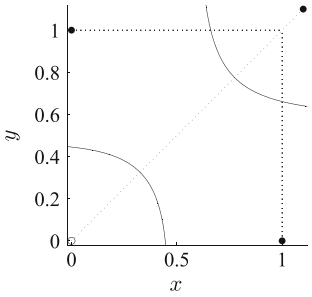


Fig. 12 The curve $|\mathbf{A}| = 0$ when microphones (*dots*) are located at $\mathbf{m}_0 = [0, 0]^T$, $\mathbf{m}_1 = [1, 0]^T$, $\mathbf{m}_2 = [1, 1]^T$, $\mathbf{m}_3 = [1.1, 1.1]^T$

which is a line in $\mathbb{R}^3$ *contained, respectively, in the planes* $x = \frac{1}{2}$ *or* $y = \frac{1}{2}$. *In particular, if* $\mathbf{x}_S \neq [\frac{1}{2}, \frac{1}{2}]^T$, *the projection of* l *on the horizontal plane* $w_S = 0$ *is, respectively, the line* r *or* r'.

Proof We have $\text{rank}(\mathbf{A}) \geq 2$. Using Lemma 5.1, the planes Π_i, $i = 1, 2, 3$ intersect along a straight line $l \subset \mathbb{R}^3$ if, and only if, $\text{rank}(\mathbf{A}) = 2$. Let us study

$$|\mathbf{A}| = -w_1 + w_2 - w_3 = 0,$$

with respect to $\mathbf{x}_S$. Using (35) and simplifying, the determinant equation becomes

$$\sqrt{(x_S - 1)^2 + (y_S - 1)^2} + \sqrt{x_S^2 + y_S^2} = \sqrt{(x_S - 1)^2 + y_S^2} + \sqrt{x_S^2 + (y_S - 1)^2}.$$

Squaring both sides of the equation and simplifying, we obtain

$$\sqrt{((x_S - 1)^2 + (y_S - 1)^2)(x_S^2 + y_S^2)} = \sqrt{((x_S - 1)^2 + y_S^2)(x_S^2 + (y_S - 1)^2)},$$

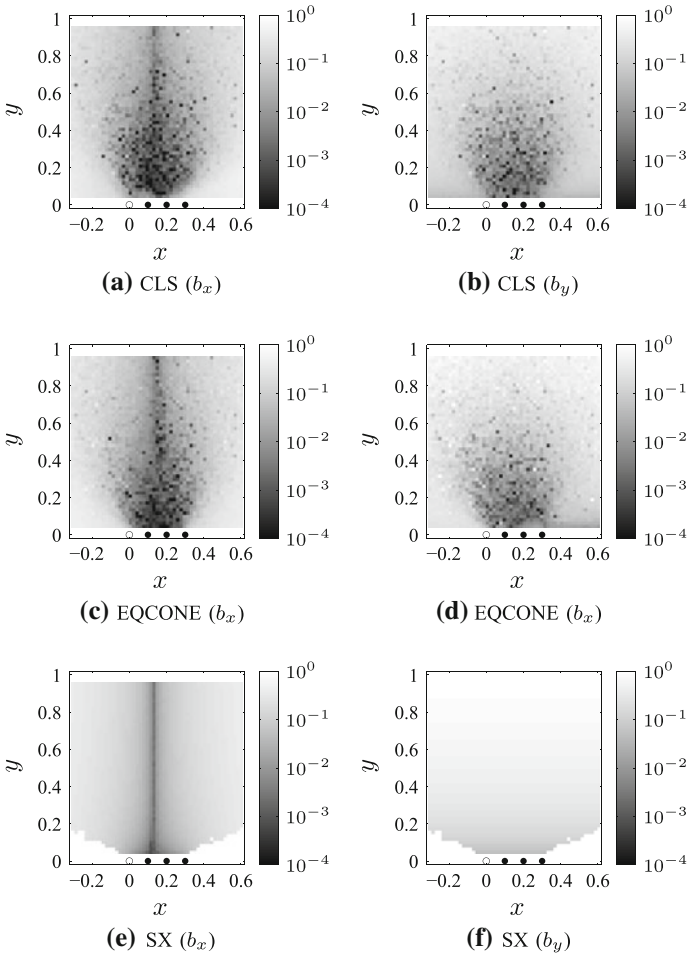
then, squaring once more, yields

$$(-2x_S + 1)(-2y_S + 1) = 0.$$

If $\mathbf{x}_S \in r$, we have $w_1 = 0$ and $w_2 = w_3$. Hence, the vectors orthogonal to the planes Π_i are, respectively, $\mathbf{n}_1 = [1, 0, 0]^T$, $\mathbf{n}_2 = [1, 1, -w_2]^T$ and $\mathbf{n}_3 = [0, 1, -w_2]^T$, which implies that the vector parallel to the line l is $\mathbf{v} = [0, w_2, 1]^T$. If $\mathbf{x}_S \neq [\frac{1}{2}, \frac{1}{2}]^T$, we have $w_2 \neq 0$, thus the projection of l on the plane $w_S = 0$ is exactly r. The case $\mathbf{x}_S \in r'$ is analogous. □

In the configurations of Theorem (5.7), the ULS fails to localize the source. In the special case $\mathbf{x}_S = [\frac{1}{2}, \frac{1}{2}]^T$, the line l is perpendicular to the plane $w_S = 0$, therefore $\hat{\mathbf{x}}_{ULS} = [\frac{1}{2}, \frac{1}{2}]^T$. Anyway, also in this situation both the LS and the SI algorithms fail. In fact, since $\mathbf{w} = \mathbf{0}$, one cannot define $\mathbf{A}^\dagger$ and Eqs. (22) become meaningless.

Notice also that even in irregular configurations of microphones it could happen that $|\mathbf{A}| = 0$ and, in this case, the ULS still retains non-isolated minima for $\mathbf{x}_S$ lying on some non-trivial curve. Figure 12 shows a possible configuration of microphones that can cause problems. In particular, $|\mathbf{A}| = 0$ when the source lies on the depicted curve.



(a) CLS (b_x) **(b)** CLS (b_y)

(c) EQCONE (b_x) **(d)** EQCONE (b_x)

(e) SX (b_x) **(f)** SX (b_y)

Fig. 13 Bias on x and y using the linear setup expressed in [m]. *Open circle* represents the reference microphone, *filled circle* the other ones

5.3 Experimental results

So far we have presented a theoretical analysis considering only noiseless measurements. We now consider what happens in the more common scenario where measurements are, in fact, affected by noise. In order to answer this question, in this Section we verify the theoretical results and that they hold even in the case of noisy measurements.

For every simulation, localization algorithms have been tested for every source position on a regular grid of 51 $\times$ 51 points spaced by 18mm. For every source position, 100 realizations of range differences corrupted with a zero-mean and 1cm standard deviation gaussian noise have been considered. The metric used for comparison is the average bias on the x and y coordinates of the localized source. For the x coordinate it is computed as

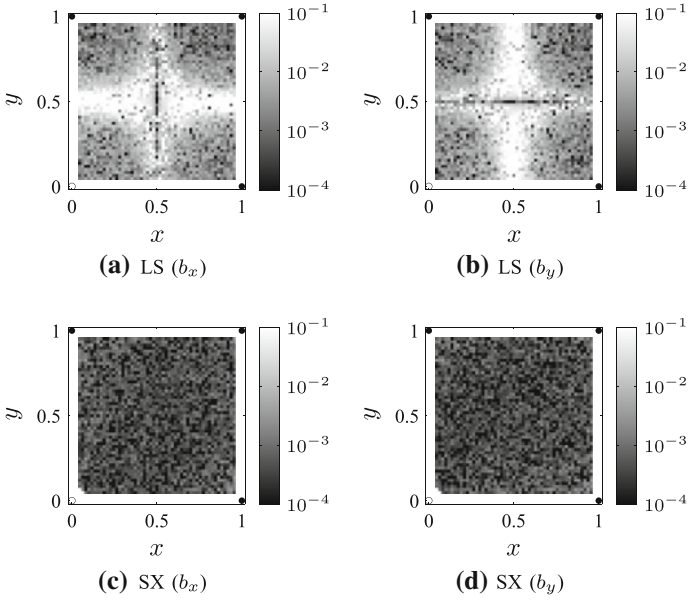


Fig. 14 Bias on x and y using the square setup expressed in $[m]$. *Open circle* represents the reference microphone, *filled circle* the other ones

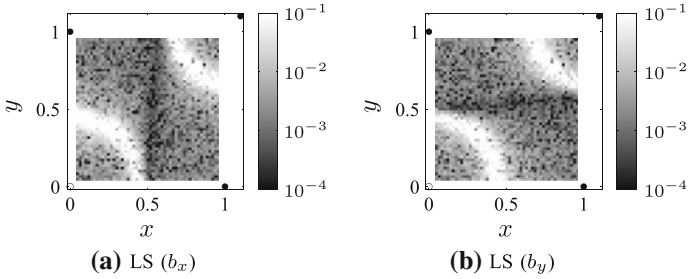


Fig. 15 Bias on x and y expressed in $[m]$ with a microphone on the square diagonal. *Open circle* represents the reference microphone, *filled circle* the other ones

$$b_x = \left| \frac{1}{n} \sum_{i=1}^{n} (x_S - \hat{x}_{S,i}) \right|,$$

where n is the number of noisy measurements tested for each source location, x_S is the x coordinate of the source, and $\hat{x}_{S,i}$ is the estimation based on the ith realization. The definition of the bias b_y on the y coordinate is straightforward.

Let us first consider the case of a linear array. We have simulated an array of four microphones spaced by $10cm$ and sources are placed in front of the array. Fig. 13 shows b_x and b_y for the CLS algorithm presented in Sect. 5.1, SX, and one of the methods presented in Compagnoni et al. (2012) (EQCONE). As the localization suffers of ambiguities for sources that are placed in front and behind the array, we preliminarily informed the algorithm that

sources are placed in the half-space $y > 0$. Notice that, for this configuration, the proposed CLS algorithm outperforms SX and its accuracy is at least comparable with that of EQCONE. It is worth noticing that the choice of these algorithms is due to the fact that, as explained before, LS and SI do not work with this configuration. Moreover, notice that also the algorithm in Beck et al. (2008) cannot be applied, since it requires a full rank matrix $\mathbf{A}$.

We consider now the scenario of four microphones located at the vertices of a unit square. Sources are bound to lie inside this square. Figure 14 shows b_x and b_y for LS and SX. Notice that LS fails to estimate the x or y coordinate of the source when it lies respectively on $r' : y = 1/2$ or $r : x = 1/2$. On the other hand, SX works properly, confirming that the localization problems are related to the LS algorithm, as from theorem (5.7) for the noiseless case. However in a noisy case, planes Π_i will be perturbed into $\hat{\Pi}_i = \Pi_i + \epsilon_i$, where ϵ_i is a noise term. We have therefore

$$\bigcap_{i=1}^{3} \hat{\Pi}_i = \mathbf{p},$$

which is a point in $\mathbb{R}^3$, such that the distance from the line l is small for small values of ϵ_i. This means that a source placed near r' or r will be estimated by ULS at a point not far from l (e.g. sources with $y_S \simeq \frac{1}{2}$ are localized close to the line $y = \frac{1}{2}$, leading to a correct y_S estimation, and a wrong x_S one). This explains why there are areas surrounding r and r' where the source is not well localized in a noisy case.

In the same way for other setups, if we find positions where the source cannot be localized in a noiseless scenario, we expect to find areas around those positions where the localization accuracy is poor when noise is present. Indeed Fig. 15 shows b_x and b_y for LS, when three microphones are on the vertices of a square, and the fourth one lies on the square diagonal. As mentioned at the end of par. 5.2, in fact, there could exists positions of the sources such that $|\mathbf{A}| = 0$, also for irregular configurations of the microphones. In the noiseless case the source LS algorithm cannot localize the source when it lies on the curve in Fig. 12, while in the noisy case we can clearly see two areas of low accuracy around this curve.

6 Conclusions

In this manuscript we have presented a framework that provides a unifying perspective on TDOA-based techniques for acoustic source localization. More specifically, we defined a space–range reference frame and used it to visualize and visually interpret numerous source localization techniques. In particular, localizing a source corresponds to minimizing a cost function, possibly constrained, in the space–range reference frame. This theory, however, goes beyond the mere tutorial intent. In fact, the space–range reference frame can also be used for predicting configurations of microphones and sources that could cause localization problems. We also presented a closed-form solution of the constrained least squares technique for the case of aligned microphones, which is derived from analysis in the space–range reference frame.

Acknowledgments The authors thank Roberto Notari for his useful suggestions during the writing of the paper. Marco Compagnoni and Paolo Bestagini equally contributed to the work presented in this manuscript and should be considered as first co-authors.

References

Abel, J. & Smith, J. (1987). The spherical interpolation method for closed-form passive source localization using range difference measurements. In *Acoustics, speech, and signal processing, IEEE international conference on ICASSP '87* (vol. 12, apr pp. 471–474).

Beck, A., Stoica, P., & Li, J. (2008). Exact and approximate solutions of source localization problems. *IEEE Transactions on Signal Processing*, *56*(5), 1770–1778.

Compagnoni, M., Bestagini, P., Antonacci, F., Sarti, A., & Tubaro, S. (2012). Localization of acoustic sources through the fitting of propagation cones using multiple independent arrays. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(7), 1964–1975.

Gillette, M., & Silverman, H. (2008). A linear closed-form algorithm for source localization from time-differences of arrival. *IEEE Signal Processing Letters*, *15*, 1–4.

Huang, Y., Benesty, J., & Elko, G. (2000). Passive acoustic source localization for video camera steering. In *Acoustics, speech, and signal processing, 2000. ICASSP '00. Proceedings. 2000 IEEE international conference on* (vol. 2, pp. II909–II912).

Huang, Y., Benesty, J., Elko, G., & Mersereati, R. (2001). Real-time passive source localization: a practical linear-correction least-squares approach. *IEEE Transactions on Speech and Audio Processing*, *9*(8), 943–956.

Huang, Y., & Benesty, J. (2004). *Audio signal processing for next generation multimedia communication systems*. Dordrecht: Kluwer Academic Publishers.

Huang, Y., Benesty, J., & Elko, G. (2004). *Source localization ch. 9*. Dordrecht: Kluwer Academic Publishers.

Schau, H., & Robinson, A. (1987). Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *35*(8), 1223–1225.

Smith, J., & Abel, J. S. (1987a). Closed-form least-squares source location estimation from range-difference measurements. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-35*, 1661–1669.

Smith, J., & Abel, J. (1987b). The spherical interpolation method of source localization. *IEEE Journal of Oceanic Engineering*, *12*, 246–252.

Author Biographies

Paolo Bestagini was born in Novara (Italy) on February 22, 1986. He received the M.Sc. degree cum laude in 2010 in Telecommunication Engineering at Politecnico di Milano, Italy. He is currently a Ph.D. student at Image and Sound Processing Group in Dipartimento di Elettronica ed Informazione at Politecnico di Milano, Italy. He initially worked on audio signal-processing for source localization. His recent research focuses on reverse engineering of multi media data for forensics analyses.

Marco Compagnoni was born in Tirano (Italy) on April 25, 1980. He received the Laurea cum laude degree in Theoretical Physics (Universitá di Milano Bicocca, 2004) and the Ph.D. degree in Applied Mathematics (Politecnico di Milano, 2009). He is currently a post-doc researcher at the Geometry—Algebra In Application group (GAIA) in Dipartimento di Matematica at Politecnico di Milano. His research interests are applied math, especially in the application of geometry in mathematical physics and in statistics, in image processing and pattern recognition and in space-time processing of audio signals.

Fabio Antonacci was born in Bari (Italy) on July 26, 1979. He received Laurea cum laude degree in 2004 and Ph.D. in 2008, both at Politecnico di Milano, Italy. He is currently a post-doc researcher at Image and Sound Processing Group in Dipartimento di Elettronica ed Informazione at Politecnico di Milano, Italy. His research focuses on space-time processing of audio signals, for both speaker and microphone arrays (source localization, acoustic scene analysis, rendering of spatial sound) and on modeling of acoustic propagation (visibility-based beam tracing). He published approximately 40 articles in proceedings of international conferences and on peer-reviewed journals.

Augusto Sarti (M'04) received the M.S. and the Ph.D. degrees in electronic engineering, both from the University of Padua, Italy, in 1988 and 1993, respectively, with research on nonlinear communication systems. His graduate studies included a joint graduate program with the University of California at Berkeley, where he worked on nonlinear system theory. In 1993, he joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, Milan, Italy, where he is currently an Associate Professor. His research interests are in the area of digital signal processing, with particular focus on sound analysis, processing and synthesis; space-time audio processing; geometrical acoustics; music information retrieval. He also worked on problems of multidimensional signal processing, vision-based 3D scene reconstruction; camera calibration; image analysis; motion planning and nonlinear system theory. He coauthored over 180 scientific publications on international journals and congresses as well as numerous patents in the multimedia signal processing area. He coordinates the Sound and Music Computing Lab of the Image and Sound Processing Group of the Politecnico di Milano. He promoted and coordinated or contributed to numerous (20+) EC-funded project. He is a member of the IEEE Technical Committee on Audio and Acoustics Signal Processing, and Associate Editor of IEEE Signal Processing Letters. He has served as guest editor for numerous special issues of international journals. He was co-chairman of the 2005 Edition of the IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS); Chairman of 2009 edition of the Digital Audio Effects conference, (DAFx); and in the organizing committees of numerous other conferences in the area of signal processing.

Stefano Tubaro (M'01) was born in Novara, Italy, in 1957. He received his Electronic Engineering degree at the Politecnico di Milano, Milano, Italy, in 1982. He then joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, first as a researcher of the National Research Council; then as an Associate Professor (1991) and finally as a Full Professor (2004). He initially worked on problems related to speech analysis; motion estimation/compensation for video analysis/coding; and vector quantization applied to hybrid video coding. In the past few years, his research interests have focused on image and video analysis for the geometric and radiometric modeling of 3-D scenes; and advanced algorithms for video coding and sound processing. He has authored over 150 scientific publications on international journals and congresses. He co-authored two books on digital processing of video sequences. He also co-authored several patents on image processing techniques. He coordinates the research activities of the Image and Sound Processing Group (ISPG) at the Dipartimento di Elettronica e Informazione of the Politecnico di Milano; which is involved in several research programs funded by industrial partners, the Italian Government, and by the European Commission. He has been involved in the IEEE Technical Committee of Multimedia Signal Processing (2005–2009), and is currently involved in that of Image Video and Multidimensional Signal Processing.