

Sparsity-Aware Multi-Source TDOA Localization

Hadi Jamali-Rad, *Student Member, IEEE*, and Geert Leus, *Fellow, IEEE*

Abstract—The problem of source localization from time-difference-of-arrival (TDOA) measurements is in general a non-convex and complex problem due to its hyperbolic nature. This problem becomes even more complicated for the case of *multi-source* localization where TDOAs should be assigned to their respective sources. We simplify this problem to an ℓ_1 -norm minimization by introducing a novel TDOA fingerprinting and grid design model for a multi-source scenario. Moreover, we propose an *innovative trick* to enhance the performance of our proposed fingerprinting model in terms of the number of identifiable sources. An interesting by-product of this enhanced model is that under some conditions we can convert the given underdetermined problem to an overdetermined one that could be solved using classical least squares (LS). Finally, we also tackle the problem of off-grid source localization as a case of grid mismatch. Our extensive simulation results illustrate a good performance for the introduced TDOA fingerprinting paradigm as well as a significant detection gain for the enhanced model.

Index Terms—Multi-source localization, TDOA fingerprinting, sparse reconstruction.

I. INTRODUCTION

DETERMINING the position of multiple sources in a two-dimensional or three-dimensional (2-D or 3-D) space is a fundamental problem which has received an upsurge of attention recently [1]. Many different approaches have been proposed in literature to recover the source locations based on time-of-arrival (ToA), time-difference-of-arrival (TDOA) or received-signal-strength (RSS) measurements between the source nodes (SNs) and some fixed receivers or access points (APs). A traditional wisdom in RSS-based localization tries to extract distance information from the RSS measurements. However, this approach fails to provide accurate location estimates due to the complexity and unpredictability of the wireless channel. This has motivated another category of RSS-based positioning, the so-called location fingerprinting, which discretizes the physical 2-D or 3-D space into grid points (GPs) and creates a map representing the space by assigning to every GP a set of location-dependent RSS parameters, one for every AP. The location of the source is then estimated by

comparing real-time measurements with the fingerprinting map at APs, for instance using K-nearest neighbors (KNN) [2] or Bayesian classification (BC) [3].

A closer look at the grid-based fingerprinting localization problem reveals that the source location is unique in the spatial domain, and can thus be represented by a 1-sparse vector. This motivated the use of compressive sampling (CS) [4] to recover the location of the source using only a few measurements by solving an ℓ_1 -norm minimization problem. This idea (for RSS measurements) illustrated promising results for the first time in [5], [6] as well as in the subsequent works [7]–[9]. Existing RSS-based sparse localization algorithms only make use of the signal/RSS readings at different receivers (or APs) separately. However, there is potential information in the cross-correlations of these received signals at different APs which has not been exploited in the aforementioned works. In [10], we have proposed to reformulate the sparse localization problem so that we can make use of the cross-correlations of the signal readings at different APs, which leads to a considerable improvement in terms of the number of identifiable sources. Notably, all the aforementioned studies consider on-grid target(s) or source(s).

On the other hand, the problem of TDOA-based localization for a single (multiple) source(s) has been investigated from different perspectives in literature, for instance in the speech and acoustic domain [11]–[15]. In speech processing, algorithms often rely on the speech non-stationarity (TDOAs can be assigned to different sources using this assumption) which does not hold in our context. That is why some of these studies consider disjoint sources such as [12] and in many others linear array receivers are assumed and thus the problem basically boils down to direction of arrival (DOA) estimation [15]. In a big line of research, the conversion of phase to TDOA leads to aliasing effects at high frequencies for large receiver spacings [13], [15]. In [13], for instance, a blind source separation (BSS) signal model is considered and a beamforming procedure is proposed to produce an acoustic map of the covered area. To obtain such a map, distance information (between source(s) and receivers) is required which becomes computationally demanding for a near-field assumption. In [14], a fingerprinting-like approach is proposed and the area is discretized into a set of GPs for which an acoustic map function is defined. Through a proper processing of the acoustic map and de-emphasizing the effect of the dominant source, they illustrate a good performance in localizing two sources, but in some situations their performance drops if the number of targets is larger than three. Interestingly, none of the aforementioned studies exploits CS or sparse reconstruction ideas and surprisingly, not much work can be found on TDOA-based source localization within a sparse representation framework. In [16], a single-source TDOA-based localization

Manuscript received August 24, 2012; revised January 15, 2013, April 11, 2013, and June 15, 2013; accepted June 26, 2013. Date of publication July 04, 2013; date of current version September 03, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andrzej Cichocki. This work was supported by NWO-STW under the VICI program (Project 10382). Part of this work is presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2013.

The authors are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2826 CD Delft, The Netherlands (e-mail: h.jamali-rad@tudelft.nl; g.j.t.leus@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2013.2272288

TABLE I
DESCRIPTION OF THE SYMBOLS

Symbol	Description
$\mathbf{s}_k(t)$	k -th SN's signal
$x_i(t), n_i(t)$	Received signal and noise at the i -th AP
$r_i(\Delta)$	Cross-correlation w.r.t. (AP ₁ , AP _{i}) pair
$\Delta_i^{(k)}$	k -th TDOA peak in $r_i(\Delta)$
$\mathbf{y}^{(k)}$	Measurement vector containing $\Delta_i^{(k)}$'s
$\Delta_{i,k}$	TDOA of the k -th SN in $r_i(\Delta)$
\mathbf{y}_k	Measurement vector containing $\Delta_{i,k}$'s
$\Delta_{i,n}^g$	TDOA of n -th GP w.r.t. (AP ₁ , AP _{i}) pair
\mathbf{y}_n^g	Measurement vector containing $\Delta_{i,n}^g$'s
hyp _{i,k}	Hyperbola of k -th SN w.r.t. (AP ₁ , AP _{i}) pair
hyp _{i,n} ^g	Hyperbola of n -th GP w.r.t. (AP ₁ , AP _{i}) pair

is proposed wherein the sparsity of the multipath channel is exploited for time delay estimation but we are basically interested in spatial source sparsity, i.e., we want to exploit the fact that the sources are sparse in the 2-D or 3-D space. On the other hand, in [17], the spatial source sparsity is exploited to simplify the hyperbolic source localization problem into an ℓ_1 -norm minimization. However, the algorithm in [17] treats different sources separately, i.e., it is in principle a single-source localization approach. Besides, the problem of off-grid source localization is not really tackled in [17]. A conference pre-cursor of the current work is presented in [18].

The contribution of this paper is four-fold. Firstly, we formulate the problem of sparsity-aware multi-source localization by defining a novel TDOA fingerprinting paradigm to simplify the complexity and non-convexity of the multi-source TDOA localization problem. The proposed paradigm solves the problem of the TDOA assignment and multi-source localization in a joint fashion. Second, we present an appropriate grid design for our fingerprinting model. Further, we propose a *novel trick* to enhance our proposed fingerprinting paradigm in terms of the number of identifiable sources, which leads to a significant detection gain. And finally, we extend our ideas by tackling the problem of off-grid source localization. To this aim, we propose two algorithms inspired by the grid mismatch concept as well as the sparse total least squares (STLS) method proposed in [19]. It is worth pointing out that the proposed algorithms can be applied in outdoor environments where location-based services are of interest. Therefore, there is no limitation to employ the proposed ideas in wireless local area networks (WLANs) or wireless sensor networks (WSNs) operating in a centralized fashion. A notation summary of the symbols used in the following sections is given in Table I.

The rest of the paper is organized as follows. In Section II, the TDOA network model as well as our measurement model are explained. Section III introduces our novel sparse multi-source TDOA localization idea as well as the proposed grid design. Section IV presents an innovative approach to enhance the performance of our proposed multi-source algorithm. The problem of off-grid source localization is investigated in Section V. Extensive simulations in Section VI corroborate our analytical

claims in several scenarios. Finally, the paper is wrapped up in Section VII with brief concluding remarks.

II. TDOA NETWORK MODEL

Consider that we have M APs distributed over a 2-D or 3-D area which is discretized into N GPs. Note that the APs can be located anywhere, not necessarily on the GPs. We consider K SNs which are randomly located either on any of these GPs (“*on-grid*”) or possibly “*off-grid*”. We assume that the APs are connected to each other in a wireless or wired fashion so that they can cooperate by exchanging their signal readings. Now, if the k -th source broadcasts a time domain signal $s_k(t)$, the received signal at the i -th AP can be expressed by

$$x_i(t) = \sum_{k=1}^K h_{i,k} s_k(t - \tau_{i,k}) + n_i(t), \quad (1)$$

where in general $h_{i,k}$ is the channel coefficient and $\tau_{i,k}$ is the time delay from the k -th source to the i -th AP and $n_i(t)$ represents additive white noise. Here, for the sake of simplicity, we have considered a single-tap flat fading channel. We only consider a single-path scenario here, since it might be more suited to an outdoor environment and since it simplifies the setting in order to have a better focus on the core idea of this paper.

In this work, we choose a set of $M - 1$ TDOA measurements (the so-called non-redundant set) by always considering the first AP as the reference. Since we consider a passive source localization scenario, taking cross-correlations of the received signals is the optimal approach for extracting the TDOAs [20]. The signals $s_k(t)$ and $n_i(t)$ are assumed to be ergodic, mutually uncorrelated white sequences, i.e.,

$$\int_t s_k(t) s_{k'}(t - \Delta) dt = \begin{cases} 0, & k \neq k' \\ \delta(\Delta), & k = k', \end{cases} \quad (2a)$$

$$\int_t n_i(t) n_j(t - \Delta) dt = \begin{cases} 0, & i \neq j \\ \delta(\Delta), & i = j, \end{cases} \quad (2b)$$

$$\int_t s_k(t) n_i(t) dt = 0. \quad (2c)$$

where $\delta(\cdot)$ stands for the unit impulse function. Therefore, by considering (2), the cross-correlation between the received signal at the i -th AP and the reference AP is given by

$$\begin{aligned} r_i(\Delta) &= \int_t \left(\sum_{k=1}^K h_{i,k} s_k(t - \tau_{i,k}) + n_i(t) \right) \\ &\quad \times \left(\sum_{k'=1}^K h_{1,k'} s_{k'}(t - \Delta - \tau_{1,k'}) + n_1(t - \Delta) \right) dt \\ &= \sum_{k=1}^K \sum_{k'=1}^K \int_t (h_{i,k} s_k(t - \tau_{i,k}) + n_i(t)) \\ &\quad \times (h_{1,k'} s_{k'}(t - \Delta - \tau_{1,k'}) + n_1(t - \Delta)) dt \\ &= \sum_{k=1}^K \int_t (h_{i,k} s_k(t - \tau_{i,k}) + n_i(t)) \\ &\quad \times (h_{1,k} s_k(t - \Delta - \tau_{1,k}) + n_1(t - \Delta)) dt \\ &= \sum_{k=1}^K h_{i,k} h_{1,k} \delta(\Delta - \Delta_{i,k}), \end{aligned} \quad (3)$$

where $\Delta_{i,k} = \tau_{1,k} - \tau_{i,k}$ is the TDOA of the k -th source w.r.t. the AP pair (AP_1, AP_i) . As is shown by (3), for a single-tap channel as considered here, the K dominant peaks of $r_i(\Delta)$ return the TDOA values $\{\Delta_{i,k}\}_k$ related to the K sources. Note that in this work we assume that K is known even though target counting algorithms (such as a modified version of [9]) can be applied to estimate K in advance.

The main problem with (3) is that even though we can estimate the set of TDOAs $\{\Delta_{i,k}\}_k$, we do not know the source indices of the TDOAs. This leads to an assignment problem to relate the TDOAs to the sources. To make it more clear, as shown in Fig. 1, we define the $\Delta_i^{(k)}$'s which denote the TDOAs in an increasing order ($\Delta_i^{(1)} \leq \dots \leq \Delta_i^{(K)}$). These $\Delta_i^{(k)}$'s can be measured for $i = 2, \dots, M$ and they are stacked in the measurement vectors $\mathbf{y}^{(k)} = [\Delta_2^{(k)}, \dots, \Delta_M^{(k)}]^T$. Note the difference with the $\Delta_{i,k}$'s, which denote the TDOA values ordered according to the source indices leading to the vectors $\mathbf{y}_k = [\Delta_{2,k}, \dots, \Delta_{M,k}]^T$. It is worth mentioning that while the $\mathbf{y}^{(k)}$ vectors are perfectly known, the \mathbf{y}_k vectors are not. Now, the problem considered herein can be stated as follows. *How can we assign the TDOAs to the different sources and simultaneously localize them?* We would like to emphasize that we tackle the problem of *passive* multi-source localization where we have no knowledge about the signals transmitted by the sources except for the common assumption that they are mutually uncorrelated white sequences; otherwise, any sort of information about the signal (such as an identification label, the occupied bandwidth, the time slot in which they are transmitted, etc.) can help to dissect the problem into K separate localization problems that can be solved disjointly. We start our solution development by considering on-grid sources and then we extend it to the case of off-grid sources.

III. SPARSITY-AWARE TDOA LOCALIZATION

In order to assign the TDOAs to the different sources and simultaneously localize them, we propose a fingerprinting procedure. We start this procedure with an *initialization* phase where the fingerprinting map is determined. Then, in the *runtime* phase, this map is used together with the measured TDOAs to determine the location of the SNs.

A. Initialization Phase

In the initialization phase, we basically discretize the physical space into GPs and create a map (the so-called fingerprinting map) representing the space by assigning to every GP a set of location-dependent parameters. For the TDOA setup under consideration, the location-dependent parameter set will consist of the TDOA measurements from the APs. For every GP, we determine the $M - 1$ TDOAs at the different APs w.r.t. the first AP. Next, by concatenating the measurements from N GPs we construct a fingerprinting map Ψ of size $(M - 1) \times N$ of the form

$$\Psi = \begin{bmatrix} \Delta_{2,1}^g & \cdots & \Delta_{2,N}^g \\ \vdots & \ddots & \vdots \\ \Delta_{M,1}^g & \cdots & \Delta_{M,N}^g \end{bmatrix}, \quad (4)$$

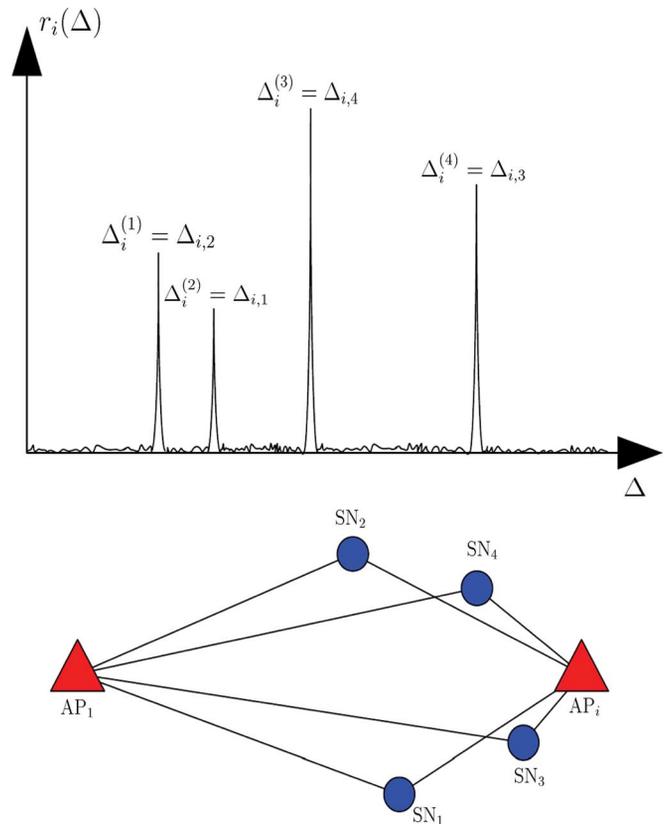


Fig. 1. Artificial setup for assignment problem; definition of $\Delta_i^{(k)}$ and $\Delta_{i,k}$. Note that SN_2 produces the smallest TDOA while SN_3 produces the largest one.

where $\Delta_{i,n}^g$ represents the TDOA of the received signal at the i -th AP and the reference AP from a source located at the n -th GP. Note the difference with $\Delta_{i,k}$ which is the measured TDOA from the k -th source w.r.t. the (AP_1, AP_i) pair. To determine the $\Delta_{i,n}^g$'s, we can simply use the known geometric configuration of the APs and GPs. This is highly desirable as we can avoid *exhaustive* classical training procedures.

B. Runtime Phase

For the runtime phase, we make a distinction between a single-source and multi-source scenario as explained in the following.

1) *Single-Source Scenario*: In the single-source case, the location of the source is estimated by comparing the runtime phase TDOA measurements $\mathbf{y} = \mathbf{y}^{(1)} = [\Delta_2^{(1)}, \dots, \Delta_M^{(1)}]^T$ with the fingerprinting map, at a central unit connected to the APs. One way to carry out this comparison is by exploiting the source sparsity and considering that the source can only be located at a single GP. This way, the single-source localization problem can be cast into a sparse representation framework given by $\mathbf{y} = \Psi\boldsymbol{\theta} + \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon}$ an $(M - 1) \times 1$ vector containing the additive noise on the TDOAs, and $\boldsymbol{\theta}$ an $N \times 1$ vector with all elements equal to zero except for one element equal to one corresponding to the index of the GP where the source is located.

¹Note that only for a single-source scenario $\mathbf{y}^{(1)} = \mathbf{y}_1$, but this cannot be generalized to a multi-source scenario, i.e., in that case we generally have $\mathbf{y}^{(k)} \neq \mathbf{y}_k$.

Thus, \mathbf{y} will be a 1-sparse TDOA vector characterized by the sparsity basis Ψ and the ultimate goal is to recover θ only by determining the index of its non-zero element.

Solving $\mathbf{y} = \Psi\theta + \epsilon$ with classical LS produces an incorrect estimate due to the underdetermined nature of the problem ($M - 1 \ll N$). Instead, sparse reconstruction techniques (or CS) aim to reconstruct θ from \mathbf{y} , by taking the source sparsity concept into account. It is worth mentioning that here we have a natural compression in the problem in the sense that the number of measurements is limited to $M - 1$ which in many practical scenarios is much less than the number of GPs N . Therefore, we will estimate θ by solving the following ℓ_1 -norm minimization problem (similar to [17]) $\min_{\theta} \|\mathbf{y} - \Psi\theta\|_2^2 + \lambda\|\theta\|_1$ where λ is a regularization parameter that controls the trade-off between sparsity and reconstruction fidelity of the estimated θ . It is worth mentioning that for a single-source scenario some other simpler methods, like matching pursuit [21], can also be used to recover the location of the source.

2) *Multi-Source Scenario*: Having explained the single-source TDOA localization within a sparse framework, now, the question is how we can extend this single-source localization scheme to a multi-source one. Before explaining the idea, we would like to remind the reader of a natural phenomenon in RSS fingerprinting. Different from TDOA measurements, the RSSs of the source signals will sum up at the APs [7], [10]. On the other hand, TDOA measurements do not simply follow this pattern. Nevertheless, this motivated us to sum up the measured $\Delta_i^{(k)}$ values for different sources at the APs, i.e., $\mathbf{y} = \sum_k \mathbf{y}^{(k)}$. Note that this vector is equal to $\mathbf{y} = \sum_k \mathbf{y}_k$ and thus automatically leads to a similar formulation as for the single-source case

$$\mathbf{y} = \Psi\theta + \epsilon, \quad (5)$$

where here θ is a K -sparse vector (containing all zeros except for K ones) to accommodate the K sources. We would like to emphasize again that in practice we can only measure the $\mathbf{y}^{(k)}$ vectors because it is still unknown to which source they belong, i.e., the \mathbf{y}_k vectors cannot be separately calculated. However, the *beauty* of the proposed sparsity-aware multi-source TDOA localization (SMTL) framework is that since we work with $\mathbf{y} = \sum_k \mathbf{y}^{(k)} = \sum_k \mathbf{y}_k$, it does not really require such assignment information. Therefore, similar to the single-source scenario, (5) can also be solved using an ℓ_1 -norm minimization

$$\hat{\theta}_{\text{SMTL}} = \arg \min_{\theta} \|\mathbf{y} - \Psi\theta\|_2^2 + \lambda\|\theta\|_1, \quad (6)$$

where λ is defined as earlier. Notably, outliers in the measured TDOAs \mathbf{y} can be handled within our sparsity-aware framework by exploiting the ideas proposed in [22].

Remark 1 (Identifiability of SMTL): To elaborate on the identifiability of localization using SMTL, it is worth mentioning that in classical (2-D) TDOA localization, as long as there are $M > 3$ APs (not lying on a straight line) associated with a source, that source can be uniquely identified and localized. In a multi-source case, however, all possible assignments between TDOAs and sources have to be checked. On the other hand, the sparse reconstruction-based nature of SMTL imposes an extra

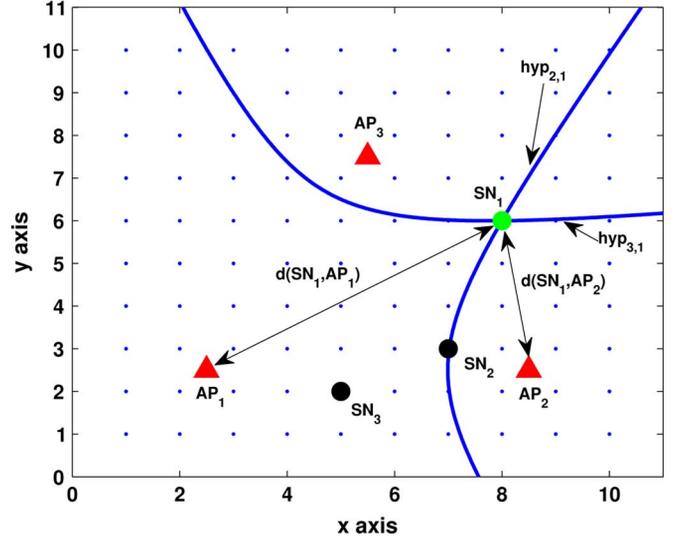


Fig. 2. Coincident Δ 's in a uniform GP configuration.

constraint $M - 1 \geq 2K$ ($M > 3$ should also be satisfied) because for a perfect reconstruction we require every $2K$ -column subset of Ψ to be full column rank so that we can reconstruct a K -sparse θ . All in all, this leads to $M > \max(2K, 3)$ as a necessary condition for identifiability and reconstruction. ■

C. Grid Design

In the earlier proposed TDOA formulation an unintentional grid problem shows up. Consider that we have three APs (AP_1 to AP_3) and three source nodes (SN_1 to SN_3) as in Fig. 2. Now, assume that SN_1 is located on (8, 6) as shown in the figure. The set of points $\mathbf{x} = [x, y]^T$ that represents a constant TDOA w.r.t. AP_1 and AP_i ($\Delta_{i,1}$ is constant) defines a hyperbola given by

$$\begin{aligned} \text{hyp}_{i,1} : \quad & \frac{1}{\nu}(d(\text{SN}_1, \text{AP}_1) - d(\text{SN}_1, \text{AP}_i)) \\ & = \frac{1}{\nu}(d(\mathbf{x}, \text{AP}_1) - d(\mathbf{x}, \text{AP}_i)), \quad (7) \end{aligned}$$

where $d(A, B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$ is the Euclidean distance between points A and B and ν denotes the speed of the signal propagation. For $i = 2, 3$ this results in the two hyperbolas ($\text{hyp}_{2,1}$ and $\text{hyp}_{3,1}$) plotted with solid blue lines in Fig. 2. In general, $\text{hyp}_{i,k}$ denotes the hyperbola related to the TDOA of the source SN_k w.r.t. the (AP_1, AP_i) pair. Right now, if any other source falls on either one of these two hyperbolas, that source will have a similar TDOA as SN_1 w.r.t. either the (AP_1, AP_2) or (AP_1, AP_3) pair. Because in Fig. 2 SN_2 lies on $\text{hyp}_{2,1}$ and SN_3 does not, the output of the cross-correlation related to the (AP_1, AP_2) pair will contain only two dominant peaks instead of three peaks. Obviously, in such a case this coincidence cannot be resolved based on the amplitude of the peaks because the signals arrive at the APs with different amplitudes depending on the fading channel. It is worth mentioning that with the uniform GP configuration as shown in Fig. 2, the probability of obtaining such (approximately) equal Δ values in each row of (4) is not low and this probability increases with the number of GPs N . Next,

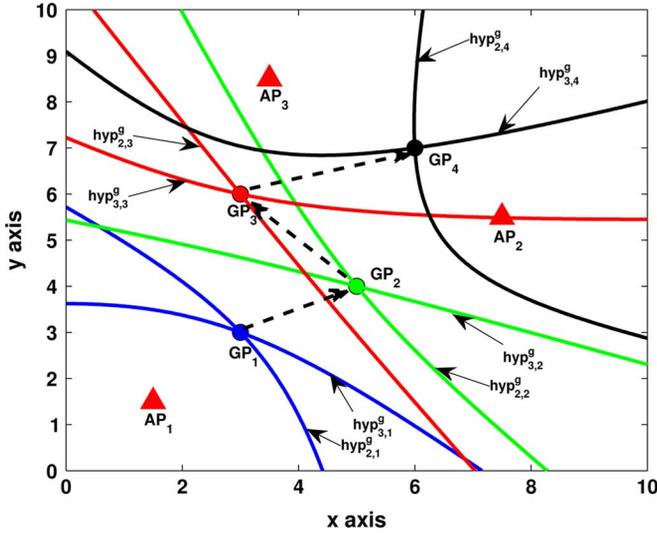


Fig. 3. Proposed sequential GP placement.

we propose a new grid configuration to avoid this issue, if the sources are on-grid. Note that in many practical situations, the APs are part of the existing infrastructure and we do not have the privilege neither to change their number nor their location. This basically motivates the following grid design based on a fixed AP configuration.

For a given AP configuration, we propose a sequential GP placement so that none of the sources will have a similar TDOA w.r.t. any of the AP pairs, i.e., (AP_1, AP_i) , $i = 2, \dots, M$. Let us consider the simple scenario shown in Fig. 3 where again only three APs exist. We start by choosing a desired location for the first GP (GP_1). Note that we have no restriction on the location of GP_1 . Now, GP_1 defines $M - 1 = 2$ hyperbolas ($hyp_{2,1}^g$ and $hyp_{3,1}^g$) with $hyp_{i,n}^g$ defined similar to (7) but for the GPs as

$$\begin{aligned} hyp_{i,n}^g &: \frac{1}{\nu} (d(GP_n, AP_1) - d(GP_n, AP_i)) \\ &= \frac{1}{\nu} (d(\mathbf{x}, AP_1) - d(\mathbf{x}, AP_i)), \quad i = 2, 3, n = 1, \dots, N, \end{aligned} \quad (8)$$

with AP_1 chosen as the reference. Each of these hyperbolas excludes a curve from the 2-D plane of the covered area and leaves the remaining part of the plane as a possible option to place the next GP. Therefore, if we place GP_2 on either $hyp_{2,1}^g$ or $hyp_{3,1}^g$ there will be one overlapping peak in the output of the cross-correlation corresponding to the pair (AP_1, AP_2) or (AP_1, AP_3) , respectively. After placing GP_2 , two more hyperbolas should be excluded from the 2-D plane for the next GP. This means, we should not place GP_3 on any of $hyp_{2,1}^g$, $hyp_{3,1}^g$, $hyp_{2,2}^g$ and $hyp_{3,2}^g$, as is also illustrated in Fig. 3. The following GPs are placed in a similar fashion and this procedure can be continued until we find N GPs.

Remark 2 (Backward Checking): It is important to observe that $hyp_{i,n}^g$ of GP_n can never cross $hyp_{i,n'}$ of $GP_{n'}$. This is because if they could cross, then at the crossing point we would have $\Delta_{i,n} = \Delta_{i,n'}$ and considering (8) this would mean that the two hyperbolas should coincide everywhere and thus GP_n and $GP_{n'}$ would be located on the same hyperbola. This is impossible according to our grid design. As a result, a hyperbola related to a GP can never cross a previously deployed GP, which

means that our proposed sequential GP placement procedure does not require a backward checking modification when we place the GPs. ■

IV. ENHANCED SPARSITY-AWARE MULTI-SOURCE LOCALIZATION (ESMTL)

The proposed sparsity-aware multi-source algorithm of Section III has a limited source detection capability which comes from the fact that we sum the measured TDOAs at the APs, thereby losing a significant amount of information. This basically limits the number of detectable sources (K) through the number of measurements (see Remark 1). The question is how this problem can be solved without taking additional TDOA measurements. The *innovative trick* we use here is to consider not just the sum of the TDOAs as $\mathbf{y} = \sum_k \mathbf{y}^{(k)} = \sum_k \mathbf{y}_k$, but the sum of any function of the TDOAs as

$$\mathbf{y}_{f_l} = \sum_k f_l(\mathbf{y}^{(k)}) = \sum_k f_l(\mathbf{y}_k), \quad (9)$$

where

$$f_l(\mathbf{y}^{(k)}) = [f_{l,1}(\Delta_2^{(k)}), \dots, f_{l,M-1}(\Delta_M^{(k)})]^T, \quad (10)$$

with $f_{l,i}(\cdot)$ being any possible *measurement function*. If we combine a set of L such sums, i.e.,

$$\bar{\mathbf{y}} = [\mathbf{y}_{f_1}^T, \mathbf{y}_{f_2}^T, \dots, \mathbf{y}_{f_L}^T]^T, \quad (11)$$

this newly defined measurement vector $\bar{\mathbf{y}}$ calls for a new fingerprinting map $\bar{\Psi}$ which can accordingly be defined as

$$\bar{\Psi} = [f_1(\Psi)^T, \dots, f_L(\Psi)^T]^T, \quad (12)$$

where

$$f_l(\Psi) = \begin{bmatrix} f_{l,1}(\Delta_{2,1}^g) & \cdots & f_{l,1}(\Delta_{2,N}^g) \\ \vdots & \ddots & \vdots \\ f_{l,M-1}(\Delta_{M,1}^g) & \cdots & f_{l,M-1}(\Delta_{M,N}^g) \end{bmatrix}, \quad (13)$$

and thus the model (5) can be extended to

$$\bar{\mathbf{y}} = \bar{\Psi}\boldsymbol{\theta} + \bar{\boldsymbol{\epsilon}}. \quad (14)$$

The new $\bar{\Psi}$ has $L(M - 1)$ rows instead of only $M - 1$ rows, i.e., it is capable of detecting more sources simultaneously, if the measurement functions $f_{l,i}(\cdot)$ own certain properties. First of all, they should be nonlinear in general since linear functions generate dependent rows in $\bar{\Psi}$ which in principle does not increase the number of independent equations in (14). Moreover, these functions should not impair the restricted isometry property (RIP) [23] of $\bar{\Psi}$ required for a high quality reconstruction. Having this issue in mind, an orthonormalization procedure on the resulting $\bar{\Psi}$ can help to improve the RIP, as we also show numerically later on.

Remark 3 (Identifiability of ESMTL): For the enhanced model, the expected necessary identifiability condition (as explained in Remark 1) will be $L(M - 1) \geq 2K$ and $M > 3$ which results in $M > \max(\lceil (2K + 1)/L \rceil + 1, 3)$, where $\lceil \cdot \rceil$ denotes the ceiling operator. A detailed analysis of the

TABLE II
RIP TEST

Matrix	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8	δ_9	δ_{10}
$\mathbf{N}_{5 \times 25}$	0.000	0.942	1.775	2.497	3.152	3.651	4.112	4.506	4.840	5.162
$\mathbf{R} \Psi$	0.000	0.951	1.821	2.664	3.350	4.084	4.398	4.729	5.061	5.389
$\mathbf{N}_{25 \times 25}$	0.000	0.608	0.959	1.240	1.475	1.709	1.897	2.058	2.186	2.279
$\tilde{\Psi} = \mathbf{R} \tilde{\Psi}$	0.000	0.517	0.725	0.851	0.946	0.974	0.982	0.989	0.994	0.996

dependence of the measurement functions on the identifiability is a complicated mathematical exercise which is outside the scope of this paper and is left for future work. ■

In principle, the measurement functions $f_{l,i}(\cdot)$ can be any nonlinear function. We could for instance consider a base set of L non-linear functions denoted as $\{g_l(\cdot)\}_{l=1}^L$ (the $g_l(\cdot)$ functions could for example be monomials, i.e., $g_l(\cdot) = (\cdot)^l$) and take $f_{l,i}(\cdot) = g_l(\cdot)$. In addition, to improve the RIP we could further apply the operator \mathbf{R} of size $L(M-1) \times L(M-1)$ to the measurements, i.e., $\tilde{\mathbf{y}} = \mathbf{R}\bar{\mathbf{y}}$, leading to the new map $\tilde{\Psi} = \mathbf{R}\Psi$. One option to design \mathbf{R} could be to force the *columns* of $\mathbf{R}\tilde{\Psi}$ to be as close as possible to orthonormal by solving

$$\min_{\mathbf{R}} \|(\mathbf{R}\tilde{\Psi})^T(\mathbf{R}\tilde{\Psi}) - \mathbf{I}_N\|_F^2. \quad (15)$$

Based on a detailed derivation in Appendix A, if $L(M-1) \leq N$ this results in the following solution

$$\mathbf{R} = \Sigma^\dagger(1 : L(M-1), :) \mathbf{U}^T, \quad (16)$$

while if $L(M-1) > N$ it leads to

$$\mathbf{R} = \begin{bmatrix} \Sigma^\dagger \\ \mathbf{0}_{(L(M-1)-N) \times L(M-1)} \end{bmatrix} \mathbf{U}^T, \quad (17)$$

where \mathbf{U} and Σ come from the singular value decomposition (SVD) of $\tilde{\Psi}$, i.e., $\tilde{\Psi} = \mathbf{U}\Sigma\mathbf{V}^T$. Surprisingly, this corresponds to orthonormalizing the rows of $\tilde{\Psi}$ (see also Appendix A), which has indeed been shown to improve the RIP [7]. Having said that, by employing the operator \mathbf{R} , (14) should be modified to

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{R}\bar{\mathbf{y}} = \mathbf{R}\tilde{\Psi}\boldsymbol{\theta} + \mathbf{R}\bar{\boldsymbol{\epsilon}} \\ &= \tilde{\Psi}\boldsymbol{\theta} + \tilde{\boldsymbol{\epsilon}}. \end{aligned} \quad (18)$$

Finally, (18) can be solved by

$$\hat{\boldsymbol{\theta}}_{\text{ESMTL}} = \arg \min_{\boldsymbol{\theta}} \|\tilde{\mathbf{y}} - \tilde{\Psi}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1, \quad (19)$$

where λ is defined as earlier.

A. RIP Investigation

As we explained earlier, Ψ and $\tilde{\Psi}$ are proved to be the sparsifying bases for the SMTL and the ESMTL. Having satisfied the sparsity property, the only issue that should be assessed is the mutual incoherence between the columns of Ψ and $\tilde{\Psi}$ or alternatively the RIP. In this subsection, we try to numerically investigate the RIP property of the proposed fingerprinting maps to illustrate that the reconstruction will indeed have a high quality. As we discussed earlier, to improve the ℓ_1 -norm reconstruction problem we apply the orthonormalization operator \mathbf{R} to $\tilde{\Psi}$ (and

similarly to Ψ) and that is why we only investigate the RIP of the resulting matrices. As is well documented in literature [23], for $K = 1, 2, \dots$ the RIP constant δ_K of a matrix \mathbf{A} (with normalized columns) is the smallest number for which

$$-\delta_K \leq \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} - 1 \leq \delta_K, \quad (20)$$

for all K -sparse $\mathbf{x} \in \mathbb{R}^N$. Roughly speaking, as long as $0 < \delta_K < 1$ the RIP holds. However, the fact that we need to know all the combinations $\binom{N}{K}$ for $K = 1, 2, \dots$ makes the problem NP-hard. For the sake of computational complexity, we use the definition in [23] where δ_K is defined as the maximum distance from 1 of all the eigenvalues of the $\binom{N}{K}$ submatrices, $\mathbf{A}_\Lambda^T \mathbf{A}_\Lambda$, derived from \mathbf{A} , where Λ is a set of indices with cardinality K which selects those columns of \mathbf{A} indexed by Λ . It means that for each K , the RIP constant is given by

$$\delta_K = \max_{\Lambda} (|\lambda_{\max}(\mathbf{A}_\Lambda^T \mathbf{A}_\Lambda) - 1|, |\lambda_{\min}(\mathbf{A}_\Lambda^T \mathbf{A}_\Lambda) - 1|). \quad (21)$$

For the sake of feasibility of the computations, we consider the case where $M-1 = 5$, $N = 25$, and $L = 4$ (for the ESMTL), which is also the setup considered in one of our simulation scenarios in Section VI. For such a case, we have computed the δ_K with $K = 1, \dots, 10$ for $\mathbf{R}\Psi$ and $\tilde{\Psi} = \mathbf{R}\tilde{\Psi}$ as well as for matrices with the same size containing elements drawn from a random normal distribution, i.e., $\mathbf{N}_{5 \times 25}$ and $\mathbf{N}_{25 \times 25}$. Note that such random matrices are proved to be a good choice in terms of the RIP and that is why we use them as a benchmark. The results are presented in Table II.

As is clear from the table, our proposed fingerprinting map for the SMTL ($\mathbf{R}\Psi$) is almost similar to $\mathbf{N}_{5 \times 25}$ and loosely satisfies the RIP up to $K = 2$. However, for $K > 2$, δ_K starts increasing. Interestingly, we see that by the aid of the added rows using our innovative $\tilde{\Psi} = \mathbf{R}\tilde{\Psi}$, the RIP is met for K up to 10, which is even better than for $\mathbf{N}_{25 \times 25}$. It is also worth stressing that this way we could demonstrate that the proposed innovative trick indeed improves the RIP of $\tilde{\Psi} = \mathbf{R}\tilde{\Psi}$ over $\mathbf{R}\Psi$.

B. Advantages of ESMTL

Besides the enhanced source detection capability, there are a number of other advantages in using the ESMTL approach as explained in the following.

First of all, an important advantage of this idea is that the recently added elements of $\bar{\mathbf{y}}$ and thus $\tilde{\mathbf{y}}$ are simply generated based on the existing TDOA measurements and *no extra measurements* are required in the runtime phase. The same holds for the new rows of $\tilde{\Psi}$ and thus $\tilde{\Psi}$ which can be computed from the

rows of $\tilde{\Psi}$. This important characteristic of the proposed TDOA fingerprinting avoids imposing extra *cost-prohibitive* measurements on the central unit.

Another important corollary of this new $\tilde{\Psi}$ is healing the case of coincident Δ peaks in the output of the cross-correlations. Now that we can have several extra equations, a simple solution to heal the issue of a uniform GP configuration (explained in Section III) is that when computing cross-correlations, say for the (AP_1, AP_i) pair, if we notice that some peaks are overlapping (number of dominant peaks is less than K), we can ignore the corresponding elements in $\tilde{\mathbf{y}}$ and correspondingly the rows in $\tilde{\Psi}$. This means that instead of (19), we solve

$$\hat{\boldsymbol{\theta}}_{\text{ESMTL}} = \min_{\boldsymbol{\theta}} \|\tilde{\mathbf{y}}' - \tilde{\Psi}'\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1, \quad (22)$$

with $\tilde{\mathbf{y}}' = \mathbf{R}\mathbf{T}\tilde{\mathbf{y}}$ and $\tilde{\Psi}' = \mathbf{R}\mathbf{T}\tilde{\Psi}$ where \mathbf{T} is a selection matrix which removes the elements and rows corresponding to the measurements with coincident peaks from $\tilde{\mathbf{y}}$ and $\tilde{\Psi}$, respectively, and with \mathbf{R} computed based on $\mathbf{T}\tilde{\Psi}$ instead of $\tilde{\Psi}$. Note that this way we are actually removing some APs; however, we can live with the uniform grid configuration until we violate the necessary identifiability condition $M > \max(\lceil(2K+1)/L\rceil + 1, 3)$.

Moreover, it is noteworthy that by finding appropriate measurement functions we can keep on increasing the number of rows so that we can attain a full column rank $\tilde{\Psi}$ matrix. In such a case, we can drop the sparsity-awareness when complexity is an issue or K is very large and recover $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}}_{\text{LS}} = \tilde{\Psi}^\dagger \tilde{\mathbf{y}}. \quad (23)$$

Further, if we are given the statistics of $\tilde{\boldsymbol{\epsilon}}$ (e.g., the mean $\mathbf{m}_{\tilde{\boldsymbol{\epsilon}}}$ and the covariance matrix $\mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}$), we obtain $\mathbb{E}\{\hat{\boldsymbol{\theta}}_{\text{LS}}\} = \tilde{\Psi}^\dagger \mathbb{E}\{\tilde{\mathbf{y}}\} = \tilde{\Psi}^\dagger(\tilde{\Psi}\boldsymbol{\theta} + \mathbf{m}_{\tilde{\boldsymbol{\epsilon}}}) = \boldsymbol{\theta} + \tilde{\Psi}^\dagger \mathbf{m}_{\tilde{\boldsymbol{\epsilon}}}$, and $\text{MSE}(\hat{\boldsymbol{\theta}}_{\text{LS}}) = \mathbb{E}\{\|\hat{\boldsymbol{\theta}}_{\text{LS}} - \boldsymbol{\theta}\|_2^2\} = \text{tr}\{\tilde{\Psi}^\dagger \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}} (\tilde{\Psi}^\dagger)^T\} + \mathbf{m}_{\tilde{\boldsymbol{\epsilon}}}^T (\tilde{\Psi}^\dagger)^T \tilde{\Psi}^\dagger \mathbf{m}_{\tilde{\boldsymbol{\epsilon}}}$, where $\mathbb{E}\{\cdot\}$ stands for the statistical expectation and $\text{tr}(\cdot)$ denotes the trace operator. This information can also be employed to solve the problem using weighted LS (WLS) as

$$\hat{\boldsymbol{\theta}}_{\text{WLS}} = (\tilde{\Psi}^T \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}} \tilde{\Psi})^{-1} \tilde{\Psi}^T \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}} [\tilde{\mathbf{y}} - \mathbf{m}_{\tilde{\boldsymbol{\epsilon}}}], \quad (24)$$

A detailed analysis of the mean and the covariance of the error on TDOA estimation using cross-correlations can be found in [24].

V. TACKLING GRID MISMATCH FOR OFF-GRID SOURCES

The classical idea of TDOA fingerprinting as well as our proposed multi-source localization ideas (SMTL and ESMTL) are based on the assumption that the sources are located on the GPs. However, as we will show in Section VI, the considered models defined by (5), (14) and (18) return inaccurate estimates if the sources are not located on their postulated GPs. This motivated us to tackle this problem for the case of multi-source TDOA localization. One generic possibility to deal with off-grid sources

is to employ the adaptive grid refinement in [25], but this requires several steps of refinement. Hence, we try to interpret this phenomenon in the form of grid or map mismatch where the measurements of the sources in \mathbf{y} , instead of (5), follow a perturbed model as

$$\mathbf{y} = [\Psi + \mathbf{E}]\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad (25)$$

which means that \mathbf{y} is now K -sparse within the sparsity basis $\Psi + \mathbf{E}$. To develop our idea of mismatch recovery, we start by analyzing the relation between the measurements from off-grid sources and \mathbf{E} for a noiseless case. For our TDOA model, Fig. 4 illustrates the case of an off-grid source in a simple setup consisting of a source SN_k and two APs (AP_1 and AP_i). As can be seen, every GP defines a so-called cell where the GP forms the center of the cell. Here, we consider that the source SN_k lies in the cell related to the n -th GP, denoted as GP_n , with $n = z(k)$ where $z(\cdot)$ indicates the mapping between sources and GPs. Assuming that the variations of $\Delta_{i,k}$ are small within the cell related to $\text{GP}_{z(k)}$, we propose to estimate the value of the perturbed TDOA ($\Delta_{i,k}$) by considering a first-order Taylor expansion as (assuming a noiseless case)

$$\Delta_{i,k} = \Delta_{i,z(k)}^g + \left[\frac{\partial \Delta_i(\mathbf{x})}{\partial x} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} \quad \frac{\partial \Delta_i(\mathbf{x})}{\partial y} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} \right] \times \begin{bmatrix} x_k - x_{z(k)}^g \\ y_k - y_{z(k)}^g \end{bmatrix}, \quad (26)$$

where $\mathbf{x}_{z(k)}^g = [x_{z(k)}^g, y_{z(k)}^g]^T$ denotes the location of $\text{GP}_{z(k)}$, $\mathbf{x}_k = [x_k, y_k]^T$ denotes the location of SN_k , and $\Delta_i(\mathbf{x})$ is the TDOA at the location $\mathbf{x} = [x, y]^T$ w.r.t. the $(\text{AP}_1, \text{AP}_i)$ pair given by

$$\begin{aligned} \Delta_i(\mathbf{x}) &= \frac{1}{\nu} (d(\text{AP}_1, \mathbf{x}) - d(\text{AP}_i, \mathbf{x})) \\ &= \frac{1}{\nu} \left(\sqrt{(x - x_{\text{AP}_1})^2 + (y - y_{\text{AP}_1})^2} \right. \\ &\quad \left. - \sqrt{(x - x_{\text{AP}_i})^2 + (y - y_{\text{AP}_i})^2} \right), \end{aligned} \quad (27)$$

and thus its partial derivatives will be

$$\frac{\partial \Delta_i(\mathbf{x})}{\partial x} = \frac{1}{\nu} \left(\frac{x - x_{\text{AP}_1}}{\sqrt{(x - x_{\text{AP}_1})^2 + (y - y_{\text{AP}_1})^2}} - \frac{x - x_{\text{AP}_i}}{\sqrt{(x - x_{\text{AP}_i})^2 + (y - y_{\text{AP}_i})^2}} \right), \quad (28a)$$

$$\frac{\partial \Delta_i(\mathbf{x})}{\partial y} = \frac{1}{\nu} \left(\frac{y - y_{\text{AP}_1}}{\sqrt{(x - x_{\text{AP}_1})^2 + (y - y_{\text{AP}_1})^2}} - \frac{y - y_{\text{AP}_i}}{\sqrt{(x - x_{\text{AP}_i})^2 + (y - y_{\text{AP}_i})^2}} \right). \quad (28b)$$

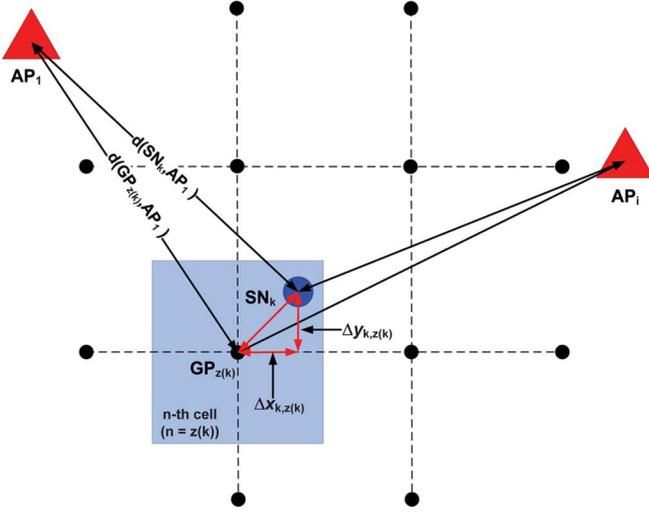


Fig. 4. Grid mismatch.

In order to fit this into our network model, we can extend (26) for the case of M APs again by considering AP_1 to be the reference AP as

$$\begin{aligned}
 \underbrace{\begin{bmatrix} \Delta_{2,k} \\ \vdots \\ \Delta_{M,k} \end{bmatrix}}_{\mathbf{y}_k} &= \underbrace{\begin{bmatrix} \Delta_{2,z(k)}^g \\ \vdots \\ \Delta_{M,z(k)}^g \end{bmatrix}}_{\mathbf{y}_{z(k)}^g} \\
 &+ \underbrace{\begin{bmatrix} \frac{\partial \Delta_2(\mathbf{x})}{\partial x} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} & \frac{\partial \Delta_2(\mathbf{x})}{\partial y} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} \\ \vdots & \vdots \\ \frac{\partial \Delta_M(\mathbf{x})}{\partial x} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} & \frac{\partial \Delta_M(\mathbf{x})}{\partial y} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} \end{bmatrix}}_{\Delta \Psi_{z(k)}} \\
 &\times \underbrace{\begin{bmatrix} x_k - x_{z(k)}^g \\ y_k - y_{z(k)}^g \end{bmatrix}}_{\Delta \mathbf{x}_{k,z(k)}}. \quad (29)
 \end{aligned}$$

It is notable that the first term on the right-hand-side of (29) is $\mathbf{y}_{z(k)}^g$ and corresponds to the measurements received from an on-grid source. Clearly, in order to be able to compute the grid mismatch for SN_k ($\Delta \mathbf{x}_{k,z(k)}$), we first have to find the closest GP corresponding to that source given by the mapping $z(k)$. The closer this GP is to the real source location, the better the first-order Taylor expansion will work. We will come back to this problem after extending (29) for a multi-source scenario.

In a multi-source scenario what happens is that we receive $\sum_{k=1}^K \mathbf{y}_k$ instead of \mathbf{y}_k which explicitly means that (29) should be solved for all the sources simultaneously as modeled by

$$\sum_{k=1}^K \mathbf{y}_k = \sum_{k=1}^K \mathbf{y}_{z(k)}^g + \sum_{k=1}^K \Delta \Psi_{z(k)} \Delta \mathbf{x}_{k,z(k)}, \quad (30)$$

where assuming that the sources (through the mapping $z(\cdot)$) are related to different GPs, we have

$$\sum_{k=1}^K \mathbf{y}_k = \sum_{k=1}^K \mathbf{y}_{z(k)}^g + \Delta \Psi \Delta \mathbf{X} \boldsymbol{\theta}, \quad (31)$$

with $\Delta \Psi = [\Delta \Psi_1, \dots, \Delta \Psi_N]$ and $\Delta \mathbf{X} = \text{diag}(\Delta \mathbf{x}_1, \dots, \Delta \mathbf{x}_N)$, which defines a block-diagonal matrix with $\Delta \mathbf{x}_1$ to $\Delta \mathbf{x}_N$ as its blocks where

$$\Delta \mathbf{x}_n = \begin{cases} \Delta \mathbf{x}_{k,z(k)}, & \exists k : n = z(k) \\ \text{don't care}, & \text{otherwise} \end{cases}. \quad (32)$$

By exploiting $\mathbf{y} = \sum_{k=1}^K \mathbf{y}_k$ and $\Psi \boldsymbol{\theta} = \sum_{k=1}^K \mathbf{y}_{z(k)}^g$, (31) can be rewritten as

$$\mathbf{y} - \Psi \boldsymbol{\theta} = \Delta \Psi \Delta \mathbf{X} \boldsymbol{\theta}, \quad (33)$$

and this can fit into the mismatch model (25) by taking $\mathbf{E} = \Delta \Psi \Delta \mathbf{X}$ which immediately gives an insight about the *structure* of the mismatch in our model.

In order to recover the mismatch, we now propose two approaches, both relying on the idea that if we know the indices of the closest GPs to the sources (i.e., the set $\{z(k) | k = 1, \dots, K\}$) given by $\boldsymbol{\theta}$, (33) is overdetermined and can efficiently be solved using classical LS. More specifically, since we can derive that

$$\begin{aligned}
 \Delta \Psi \Delta \mathbf{X} \boldsymbol{\theta} &= \sum_{n=1}^N \Delta \Psi_n \Delta \mathbf{x}_n [\boldsymbol{\theta}]_n \\
 &= \sum_{n=1}^N \Delta \Psi_n \text{diag}([\boldsymbol{\theta}]_n \otimes \mathbf{1}_2) \Delta \mathbf{x}_n \\
 &= \Delta \Psi \text{diag}(\boldsymbol{\theta} \otimes \mathbf{1}_2) \Delta \mathbf{x}, \quad (34)
 \end{aligned}$$

where $[\boldsymbol{\theta}]_n$ stands for the n -th element of $\boldsymbol{\theta}$ and $\Delta \mathbf{x} = [\Delta \mathbf{x}_1^T, \dots, \Delta \mathbf{x}_N^T]^T$, we obtain

$$\Delta \mathbf{x} = [\Delta \Psi \text{diag}(\boldsymbol{\theta} \otimes \mathbf{1}_2)]^\dagger [\mathbf{y} - \Psi \boldsymbol{\theta}]. \quad (35)$$

In order to solve (35), we have to find $\boldsymbol{\theta}$ under a grid mismatch. One way to do this is to solve the following sparse total least squares (STLS) problem for the enhanced model (18)

$$\min_{\tilde{\mathbf{E}}, \tilde{\boldsymbol{\epsilon}}, \tilde{\boldsymbol{\theta}}} \|[\tilde{\mathbf{E}}, \tilde{\boldsymbol{\epsilon}}]\|_F^2 + \lambda \|\tilde{\boldsymbol{\theta}}\|_1, \quad (36a)$$

$$\text{s.t. } \tilde{\mathbf{y}} = [\tilde{\Psi} + \tilde{\mathbf{E}}] \tilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\epsilon}}, \quad (36b)$$

using the coordinate descent (CD) algorithm in [19]. Note that $[\tilde{\mathbf{E}}, \tilde{\boldsymbol{\epsilon}}]$ denotes the augmented matrix composed of $\tilde{\mathbf{E}}$ and $\tilde{\boldsymbol{\epsilon}}$. It is worth pointing out that $\tilde{\mathbf{E}}$ is different from \mathbf{E} and similarly can be written as $\tilde{\mathbf{E}} = \Delta \tilde{\Psi} \Delta \mathbf{X}$. Therefore, instead of $\Delta \Psi$ we have to compute $\Delta \tilde{\Psi}$ which instead of $\partial \Delta_i(\mathbf{x}) / \partial x$ and $\partial \Delta_i(\mathbf{x}) / \partial y$ evaluated at $\mathbf{x} = \mathbf{x}_{z(k)}^g$ would contain

$$\begin{aligned}
 &\frac{\partial f_{l,i}(\Delta_i(\mathbf{x}))}{\partial x} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} \\
 &= \frac{\partial f_{l,i}(\Delta_i(\mathbf{x}))}{\partial \Delta_i(\mathbf{x})} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} \frac{\partial \Delta_i(\mathbf{x})}{\partial x} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g}, \quad (37a)
 \end{aligned}$$

$$\begin{aligned}
 &\frac{\partial f_{l,i}(\Delta_i(\mathbf{x}))}{\partial y} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} \\
 &= \frac{\partial f_{l,i}(\Delta_i(\mathbf{x}))}{\partial \Delta_i(\mathbf{x})} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g} \frac{\partial \Delta_i(\mathbf{x})}{\partial y} \Big|_{\mathbf{x}=\mathbf{x}_{z(k)}^g}. \quad (37b)
 \end{aligned}$$

As can be seen from (37), the elements of $\Delta \tilde{\Psi}$ are scaled by a multiplicative term $\partial f_{l,i}(\Delta_i(\mathbf{x}))/\partial \Delta_i(\mathbf{x})$.

Now, if we do not exploit the explored structure of the perturbations in $\tilde{\mathbf{E}}$, the STLS problem of (36) can be solved by an iterative block CD algorithm yielding successive estimates of $\boldsymbol{\theta}$ with $\tilde{\mathbf{E}}$ fixed, and alternately of $\tilde{\mathbf{E}}$ with $\boldsymbol{\theta}$ fixed. Given $\tilde{\mathbf{E}}(m)$ the cost in (36) has the form of a LASSO problem

$$\hat{\boldsymbol{\theta}}_{\text{STLS}}(m) = \arg \min_{\boldsymbol{\theta}} \|\tilde{\mathbf{y}} - [\tilde{\Psi} + \tilde{\mathbf{E}}(m)]\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1, \quad (38)$$

while given $\boldsymbol{\theta}(m)$ it reduces to a quadratic form with optimal $\tilde{\mathbf{E}}(m+1)$ given by

$$\hat{\mathbf{E}}_{\text{STLS}}(m+1) = (1 + \|\boldsymbol{\theta}(m)\|_2^2)^{-1} [\tilde{\mathbf{y}} - \tilde{\Psi}\boldsymbol{\theta}(m)]\boldsymbol{\theta}(m)^T, \quad (39)$$

where (m) denotes the m -th iteration. As explained in [19], the CD algorithm tries to find the values of $\tilde{\mathbf{E}}$ as well as $\boldsymbol{\theta}$ which has only K non-zero values corresponding to the K sources. Therefore, we propose to use a two-step algorithm to solve the problem of grid mismatch called STLS-LS. First, we use the CD algorithm to end up with $\boldsymbol{\theta}$. Next, using the indices of the non-zero elements of the recovered $\boldsymbol{\theta}$, i.e., the set $\{z(k)|k = 1, \dots, K\}$, as the location of the GPs, we run the grid mismatch recovery proposed in (35). It is worth mentioning that the convergence of the CD algorithm is investigated in [19]. The overall STLS-LS algorithm is summarized in Algorithm 1.

Algorithm 1: Mismatch recovery using STLS-LS

- 1: Run the iterative STLS CD algorithm given by (38) and (39).
 - 2: Find the indices of the GPs corresponding to the sources.
 - 3: Compute $\Delta \tilde{\Psi}$ and solve (35) to recover the off-grid locations.
-

More accurate results for $\boldsymbol{\theta}$ can be acquired if we estimate the location of the closest GPs by considering the structure of the perturbations. This becomes even more precise if we have knowledge about the covariance matrix of $\Delta \mathbf{x}$ ($\mathbf{C}_{\Delta \mathbf{x}}$) and $\tilde{\boldsymbol{\epsilon}}$ ($\mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}$), assuming that $\mathbf{m}_{\tilde{\boldsymbol{\epsilon}}} = \mathbf{0}$ as is shown in [24]. Note that such information about the statistics of $\Delta \mathbf{x}$ can be computed by considering the fact that the elements of $\Delta \mathbf{x}_{k,z(k)}$ should lie within a cell around the GP with a uniform distribution $\mathcal{U}(-a/2, a/2)$ in each dimension, where a is the length of a square cell. This in turn yields $\mathbf{C}_{\Delta \mathbf{x}} = a^2 \mathbf{I}_{2N}/12$. Next, we solve the following weighted structured STLS (WSSTLS) problem for the enhanced model

$$\min_{\Delta \mathbf{x}, \tilde{\boldsymbol{\epsilon}}, \boldsymbol{\theta}} \Delta \mathbf{x}^T \mathbf{C}_{\Delta \mathbf{x}}^{-1} \Delta \mathbf{x} + \tilde{\boldsymbol{\epsilon}}^T \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}^{-1} \tilde{\boldsymbol{\epsilon}} + \lambda \|\boldsymbol{\theta}\|_1, \quad (40a)$$

$$\text{s.t. } \tilde{\mathbf{y}} - \tilde{\Psi}\boldsymbol{\theta} = [\Delta \tilde{\Psi} \Delta \mathbf{X}] \boldsymbol{\theta} + \tilde{\boldsymbol{\epsilon}}. \quad (40b)$$

By taking the structure of $\tilde{\mathbf{E}}$ into account, we again use (34), which helps us to rewrite (40) as

$$\min_{\Delta \mathbf{x}, \tilde{\boldsymbol{\epsilon}}, \boldsymbol{\theta}} \Delta \mathbf{x}^T \mathbf{C}_{\Delta \mathbf{x}}^{-1} \Delta \mathbf{x} + \tilde{\boldsymbol{\epsilon}}^T \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}^{-1} \tilde{\boldsymbol{\epsilon}} + \lambda \|\boldsymbol{\theta}\|_1, \quad (41a)$$

$$\text{s.t. } \tilde{\mathbf{y}} - \tilde{\Psi}\boldsymbol{\theta} = \Delta \tilde{\Psi} \text{diag}(\boldsymbol{\theta} \otimes \mathbf{1}_2) \Delta \mathbf{x} + \tilde{\boldsymbol{\epsilon}}. \quad (41b)$$

Let us start with $\Delta \mathbf{x}(m)$ (similarly $\Delta \mathbf{X}(m)$) known, which results in

$$\min_{\tilde{\boldsymbol{\epsilon}}, \boldsymbol{\theta}} \tilde{\boldsymbol{\epsilon}}^T \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}^{-1} \tilde{\boldsymbol{\epsilon}} + \lambda \|\boldsymbol{\theta}\|_1, \quad (42a)$$

$$\text{s.t. } \tilde{\mathbf{y}} = [\tilde{\Psi} + \Delta \tilde{\Psi} \Delta \mathbf{X}(m)] \boldsymbol{\theta} + \tilde{\boldsymbol{\epsilon}}, \quad (42b)$$

which by substituting $\tilde{\boldsymbol{\epsilon}}$ from (42b) in (42a) is equivalent to solving the following convex problem (quadratic form regularized by ℓ_1 -norm as in LASSO)

$$\hat{\boldsymbol{\theta}}_{\text{WSSTLS}}(m) = \arg \min_{\boldsymbol{\theta}} \left(\tilde{\mathbf{y}} - [\tilde{\Psi} + \Delta \tilde{\Psi} \Delta \mathbf{X}(m)] \boldsymbol{\theta} \right)^T \times \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}^{-1} \left(\tilde{\mathbf{y}} - [\tilde{\Psi} + \Delta \tilde{\Psi} \Delta \mathbf{X}(m)] \boldsymbol{\theta} \right) + \lambda \|\boldsymbol{\theta}\|_1. \quad (43)$$

Having $\boldsymbol{\theta}(m)$ in hand, the next step is to solve

$$\min_{\Delta \mathbf{x}, \tilde{\boldsymbol{\epsilon}}} \Delta \mathbf{x}^T \mathbf{C}_{\Delta \mathbf{x}}^{-1} \Delta \mathbf{x} + \tilde{\boldsymbol{\epsilon}}^T \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}^{-1} \tilde{\boldsymbol{\epsilon}}, \quad (44a)$$

$$\text{s.t. } \tilde{\mathbf{y}} - \tilde{\Psi}\boldsymbol{\theta}(m) = \Delta \tilde{\Psi} \text{diag}(\boldsymbol{\theta}(m) \otimes \mathbf{1}_2) \Delta \mathbf{x} + \tilde{\boldsymbol{\epsilon}}, \quad (44b)$$

which is quadratic in $\Delta \mathbf{x}$ and results in

$$\Delta \hat{\mathbf{x}}_{\text{WSSTLS}}(m+1) = [\mathbf{C}_{\Delta \mathbf{x}}^{-1} + \mathbf{S}^T \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}^{-1} \mathbf{S}]^\dagger \mathbf{S}^T \mathbf{C}_{\tilde{\boldsymbol{\epsilon}}}^{-1} \mathbf{q}, \quad (45)$$

where $\mathbf{S} = \Delta \tilde{\Psi} \text{diag}(\boldsymbol{\theta}(m) \otimes \mathbf{1}_2)$ and $\mathbf{q} = \tilde{\mathbf{y}} - \tilde{\Psi}\boldsymbol{\theta}(m)$. The detailed derivation of (45) is explained in Appendix B. All in all, the two-step mismatch recovery procedure by using WSSTLS (called WSSTLS-LS) is summarized in Algorithm 2.

Algorithm 2: Mismatch recovery using WSSTLS-LS

- 1: Run the proposed WSSTLS CD algorithm given by (43) and (45).
 - 2: Find the indices of the GPs corresponding to the sources.
 - 3: Compute $\Delta \tilde{\Psi}$ and solve (35) to recover the off-grid locations.
-

VI. SIMULATION RESULTS

In this section, we investigate the performance of our proposed sparsity-aware multi-source localization algorithms (SMTL and ESMTL) in terms of the localization accuracy and the number of identifiable sources. To this aim, we consider a wireless network of size $10 \times 10 \text{ m}^2$ divided into $N = 100$ GPs and we consider M APs covering the whole area and K SNs to be simultaneously localized in our simulations. Instead of taking infinite integrals (as in (3)), in practice we work with discrete-time signals of limited length and hence the computations of the autocorrelations as well as the cross-correlations will not be ideal as in the derivations of Section II. As a result, the noise terms $n_i(t)$ will not be completely eliminated and will affect our performance through ϵ and $\tilde{\boldsymbol{\epsilon}}$. Here, we consider a baseband signal (satisfying the properties mentioned in Section II) sampled at $T_s = 1 \text{ ms}$ and compute the autocorrelations and cross-correlations during a time-slot of length $T = 1 \text{ s}$. This is equal to recording $N_s = T/T_s = 1000$ samples for our computations. The speed of signal propagation is $\nu = 340 \text{ m/s}$. Meanwhile, we assume that none of the received signals is so weak that it will be considered as noise in $r_i(\Delta)$ and cannot

be detected. We define the signal to noise ratio (SNR) at the i -th AP as the ratio of the received signal power to the noise power. Notably, we consider a distance-independent noise on the received signals at the different APs which according to [24], [26] results in an ϵ on the TDOA measurements specified by its covariance matrix

$$[\mathbf{C}_\epsilon]_{i,j} \approx \begin{cases} \frac{3 T_s^2 (1+2\text{SNR})}{\pi^2 N_s \text{SNR}^2}, & i = j \\ \frac{3 T_s^2}{\pi^2 N_s \text{SNR}}, & i \neq j \end{cases}. \quad (46)$$

In order to be able to quantitatively compare the performances of the algorithms under consideration, we consider the positioning root mean squared error (PRMSE) defined by $\text{PRMSE} = \sqrt{\sum_{p=1}^P \sum_{k=1}^K e_{k,p}^2 / (PK)}$, where $e_{k,p}$ represents the distance between the real location of the k -th source and its estimated location at the p -th Monte Carlo (MC) trial. All simulations are averaged over $P = 100$ MC runs where in each run the sources are deployed on different random locations. In the following simulations, we consider both the uniform grid structure as well as our proposed grid design, where for the former case if it happens that we encounter coincident Δ values, we use the solution proposed in Subsection IV-B, i.e., we remove the effect of the corresponding APs from the measurement vector and the map. For the next simulations, whenever we talk about ESMTL, we consider $L = 2, \dots, 5$ monomial base functions, i.e., $f_{l,i}(\cdot) = g_l(\cdot) = (\cdot)^l$, $l = 1, \dots, L$ to enhance the proposed SMTL by introducing new rows in $\tilde{\Psi}$. Further, we use the explained orthonormalization technique (using \mathbf{R}) to compute $\tilde{\Psi}$. This way, $\tilde{\Psi}$ will be of size $5(M-1) \times N = 45 \times 100$, while Ψ is of size $(M-1) \times N = 9 \times 100$. For all reconstruction problems, we try to find the best λ by cross-validation [27].

For the purpose of comparison, we also simulate the conventional TDOA positioning method proposed in [28] (called TDOA), as well as an optimal constrained weighted least squares method (called TDOA-CWLS) [29]. Notably, both algorithms localize the sources disjointly which gives them an edge over the proposed algorithms but of course this requires that the TDOAs can be exactly assigned to the correct sources. We would like to point out that we do not compare our results with the KNN, the BC, or even semi-definite relaxation (SDR)-based algorithms because the superiority of the ℓ_1 -norm minimization approach compared to KNN, BC and SDR-based algorithms for similar contexts (e.g., RSS-based localization) is respectively illustrated in [7] and [17]. Instead, motivated by the consideration of the aforementioned disjoint conventional methods, as a *benchmark*, we compute the Cramér-Rao lower bound (CRLB) [30] for the location of a single source, but averaged over the positions of the multiple sources. The corresponding fisher information matrix (FIM) associated with SN_k can be given by

$$\mathbf{I}_k = \begin{bmatrix} \left(\frac{\partial \mathbf{y}_k}{\partial x_k} \right)^T \mathbf{C}_\epsilon^{-1} \left(\frac{\partial \mathbf{y}_k}{\partial x_k} \right) & \left(\frac{\partial \mathbf{y}_k}{\partial x_k} \right)^T \mathbf{C}_\epsilon^{-1} \left(\frac{\partial \mathbf{y}_k}{\partial y_k} \right) \\ \left(\frac{\partial \mathbf{y}_k}{\partial y_k} \right)^T \mathbf{C}_\epsilon^{-1} \left(\frac{\partial \mathbf{y}_k}{\partial x_k} \right) & \left(\frac{\partial \mathbf{y}_k}{\partial y_k} \right)^T \mathbf{C}_\epsilon^{-1} \left(\frac{\partial \mathbf{y}_k}{\partial y_k} \right) \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \text{tr} \left[\mathbf{C}_\epsilon^{-1} \frac{\partial \mathbf{C}_\epsilon}{\partial x_k} \mathbf{C}_\epsilon^{-1} \frac{\partial \mathbf{C}_\epsilon}{\partial x_k} \right] & \text{tr} \left[\mathbf{C}_\epsilon^{-1} \frac{\partial \mathbf{C}_\epsilon}{\partial x_k} \mathbf{C}_\epsilon^{-1} \frac{\partial \mathbf{C}_\epsilon}{\partial y_k} \right] \\ \text{tr} \left[\mathbf{C}_\epsilon^{-1} \frac{\partial \mathbf{C}_\epsilon}{\partial y_k} \mathbf{C}_\epsilon^{-1} \frac{\partial \mathbf{C}_\epsilon}{\partial x_k} \right] & \text{tr} \left[\mathbf{C}_\epsilon^{-1} \frac{\partial \mathbf{C}_\epsilon}{\partial y_k} \mathbf{C}_\epsilon^{-1} \frac{\partial \mathbf{C}_\epsilon}{\partial y_k} \right] \end{bmatrix},$$

where $\text{tr}(\cdot)$ stands for the trace operator and the elements of \mathbf{y}_k as well as their derivatives are defined earlier using (27)-(29).

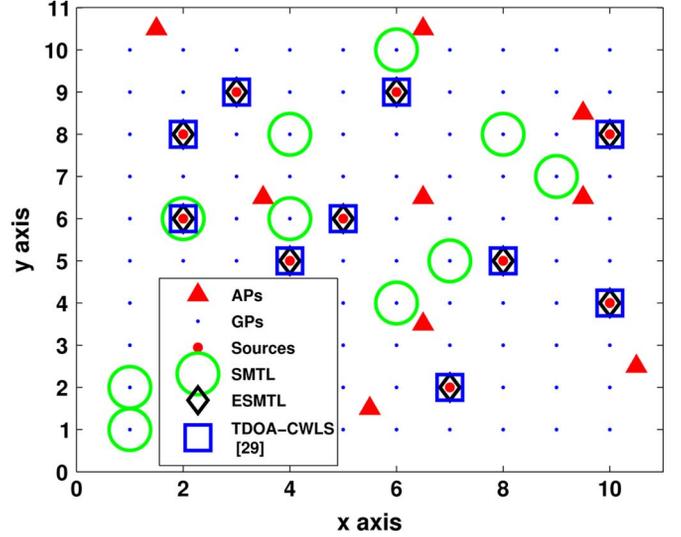


Fig. 5. Multi-source ($K = 10$) localization with $M = 10$ APs.

From (46), \mathbf{C}_ϵ is independent of the location of the source and hence the second term on the right-hand-side of (47) will be equal to zero. Therefore, corresponding to the PRMSE, the total root-CRLB (RCRLB) of the K sources is given by

$$\text{RCRLB} = \sqrt{\frac{\sum_{k=1}^K \text{tr}(\mathbf{I}_k^{-1})}{K}}. \quad (47)$$

A. Localization of On-Grid Sources

We start by investigating the performance of the proposed algorithms for the case of on-grid sources. In the first simulation, as shown by Fig. 5, we consider $K = 10$ sources randomly deployed over the covered area and $M = 10$ APs which are deployed uniformly at random. The SNR is assumed to be 20 dB for all the APs. We recover θ using both SMTL and ESMTL algorithms and we expect ESMTL to be able to locate more sources simultaneously. This is shown in Fig. 5 where SMTL can only localize a single source with minimum error. However, by using the ESMTL algorithm we can locate all the $K = 10$ sources and this clearly illustrates the enhanced performance of ESMTL compared to SMTL. As can be seen, the disjoint TDOA-CWLS is capable of reaching a high accuracy, as well. This highlights the fact that our ESMTL can perform as good as a disjoint algorithm, which is assisted with signal assignment information and treats the sources separately.

In order to further investigate this improvement in terms of the number of identifiable sources, in Fig. 6, we illustrate the PRMSE of localization versus the number of sources increasing up to $K = 10$. We have $M = 10$ APs and the simulation results of the ESMTL are presented for $L = 2, \dots, 5$. The SNR is again set to 20 dB. As can be seen from the figure, by increasing K , the PRMSE of localization for SMTL increases sharply while ESMTL (with $L = 5$) can handle almost all the sources simultaneously with minimum error. A notable (and expected) observation is that by increasing L from 2 to 5 the potential capability of ESMTL gradually increases from $K = 2$ sources being localized to $K = 10$. The figure also illustrates the considerable improvement of TDOA-CWLS over TDOA which helps it to almost attain the CRLB. Note that we do not plot the results for

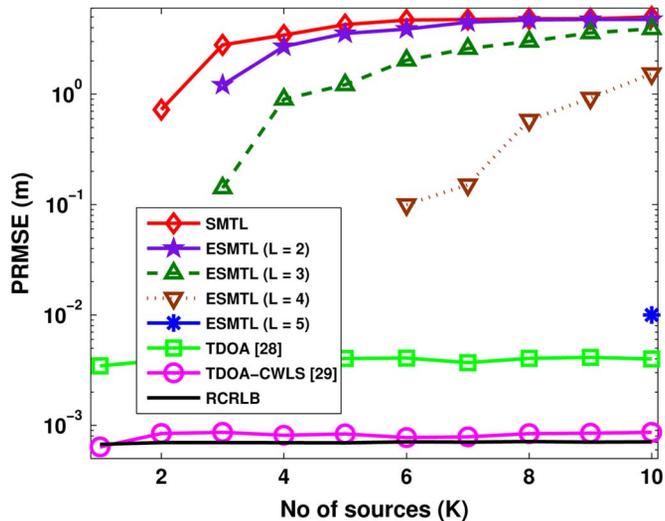


Fig. 6. PRMSE versus K for $L = 2, \dots, 5$ and $\text{SNR} = 20$ dB. The unplotted data points correspond to zero error in logarithmic scale.

$K > 10$ sources since for those cases θ is not really sparse, i.e., we do not have $K \ll N$.

In order to investigate the localization accuracy, we also plot the PRMSE versus SNR for the same previous setup but with $K = 5$ SNs in Fig. 7. As can be seen, increasing the SNR leads to a gradual improvement in the performance of the ESMTL so that for $\text{SNR} > 5$ dB we attain zero error. However, SMTL is in principle incapable of localizing $K = 5$ sources simultaneously, as it was also shown in Fig. 6, and that is why its performance does not improve with SNR. It is worth mentioning that the performance of the TDOA and TDOA-CWLS schemes is better than the one of ESMTL for lower SNRs. One reason for this is that the conventional approaches are disjoint, i.e., they treat the sources separately. Therefore, the measurement noise does not have any effect on the disambiguation of the sources. However, in the ESMTL, the measurement noise affects the values of the TDOAs (from the cross-correlations) as well as the disambiguation which is solved using ℓ_1 -norm minimization. Therefore, the disambiguation (assignment problem) can be badly affected by noise for low SNRs, which can in turn lead to a large error. Another important point worthy of being mentioned is that we attain zero error for $\text{SNR} > 5$ dB, which means that we go below the benchmark CRLB. This can be justified by the fact that we consider the on-grid scenario and have a limited number of candidates for the locations of the sources, i.e., the GPs. This feature helps the ℓ_1 -norm minimization to exactly locate the sources, as long as the noise is not too strong. More specifically, since within the region of a cell, there is only one possible point for the location of a source, the ℓ_1 -norm minimization becomes robust against small noise values.

In the next simulation, we investigate the performance of the ESMTL solved with classical LS, which means we have to make sure that Ψ has full column rank. Hence, we prefer to keep the generated rows instead of removing them for coincident Δ 's and use our proposed grid design of Subsection III-C. To simplify our simulations and reach a full rank with less complexity, we consider $N = 25$ GPs and only $M = 6$ APs, i.e., we require only $L = 5$ functions as defined earlier (in that case $L(M-1) = 25$). The results are shown in Fig. 8, where we consider $K = 10$ and 20 sources. As is clear from the figure, even though with

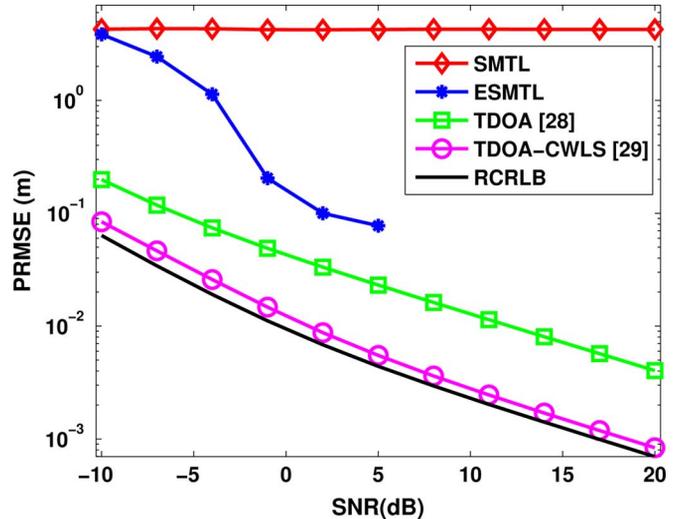


Fig. 7. PRMSE versus SNR for $K = 5$. The unplotted data points correspond to zero error in logarithmic scale.

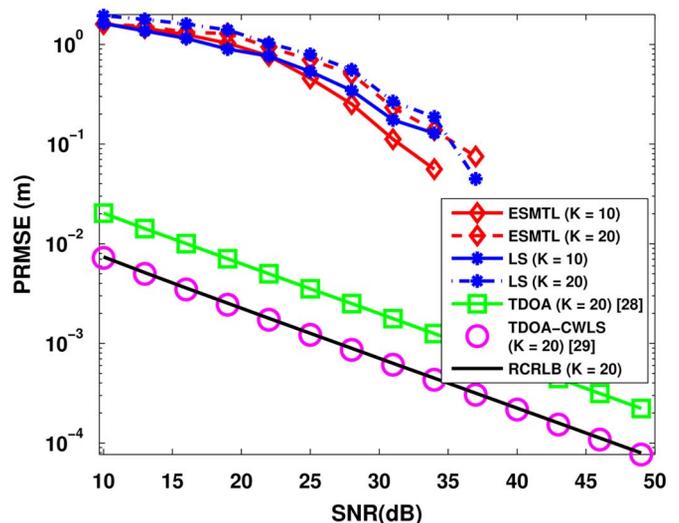
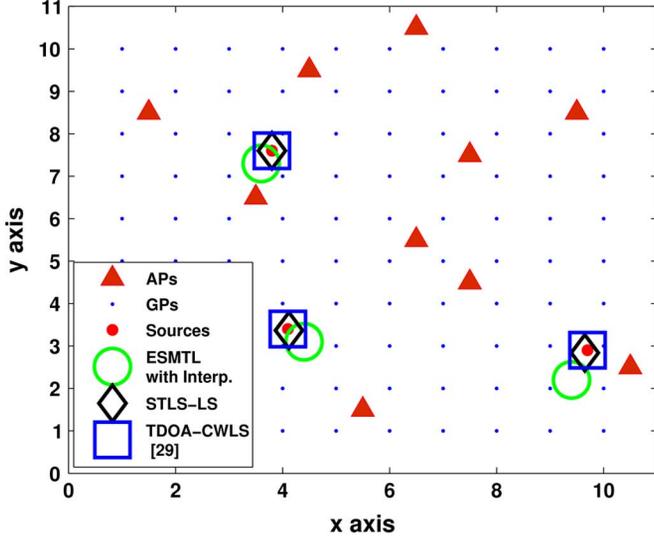
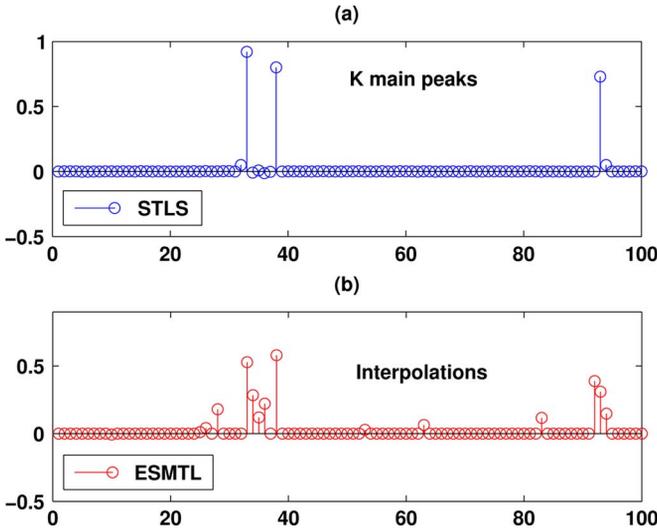


Fig. 8. PRMSE versus SNR for $K = 10$ and 20. The unplotted data points correspond to zero error in logarithmic scale.

$K/N = 10/25$ (or even further with $K/N = 20/25$) θ is not sparse anymore, the ESMTL (solved with LS) is capable of localizing the sources with minimum error for $\text{SNR} > 35$ dB. However, increasing K increases the probability of wrong Δ computations for a limited bin length N_s and thus leads to a performance degradation for $K = 20$ compared to $K = 10$. As can be seen, the ESMTL (no LS) will still work here but no gain is expected over LS as the problem is not sparse. Obviously, SMTL fails to operate here and is omitted for the sake of clarity. Notably, we observe that for specific AP configurations, it might happen that the newly generated rows with monomials do not necessarily lead to fully independent columns. As there is no restriction on the type of measurement functions, this can be healed to some extent by using different types of nonlinear functions.

B. Localization of Off-Grid Sources

The following simulations are devoted to the performance evaluation for the case of off-grid sources, i.e., tackling the grid mismatch problem. In Figs. 9 and 10, the setup is almost the

Fig. 9. Multi-source ($K = 3$) localization with grid mismatch.Fig. 10. Recovered θ with ESMTL and STLS.

same as in the previous subsection ($M = 10$ APs), except that here we consider a different AP configuration. As is clear from Fig. 9, the sources are randomly placed within the cells. The first solution, along the lines of existing literature, consists of using ESMTL (or SMTL) and interpolating the peaks in the recovered θ as is also used in [19] for a single off-grid source. For the multi-source scenario under consideration, to avoid overlapping peaks in the recovered θ , we have considered less sources (only $K = 3$) and we keep them distant from each other. We expect that if the sources are located far enough from each other, as in this case, we would have 4 peaks in the recovered θ corresponding to each off-grid source (altogether 12 peaks in the recovered θ for $K = 3$ sources) and then based on those peaks (shown in Fig. 10(b)) we can conduct a linear interpolation to locate each source (ESMTL-Interp.).

On the other hand, in order to locate the off-grid sources, we use the first proposed approach of Section V using STLS-LS summarized in Algorithm 1. In the first step, the CD algorithm

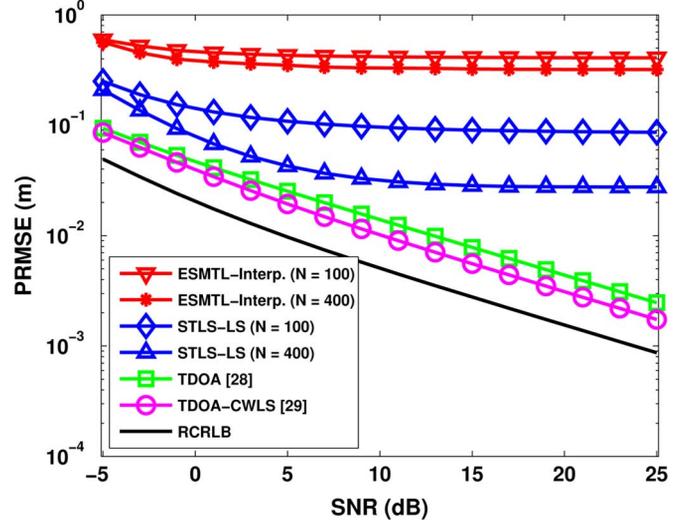


Fig. 11. Mismatch recovery with ESMTL and STLS.

is used to recover a θ which satisfies (25). The recovered θ is depicted in Fig. 10(a) and as can be seen, the main $K = 3$ peaks correspond to the closest GPs to the sources, i.e., $z(k) = 33, 38, 93$ located on (4,3), (4,8) and (10,3). In the second step, knowing the closest GPs, we compute $\Delta\Psi$ and estimate the mismatch. As is clear from Fig. 9, our proposed mismatch recovery algorithm is successful to locate the off-grid sources with a reasonable accuracy and much better than the ESMTL-Interp. A notable observation is that we still face difficulties to resolve two sources located in one cell.

Finally, Fig. 11 illustrates the PRMSE performance versus SNR for ESMTL-Interp. as well as for the proposed mismatch recovery algorithm STLS-LS when there exist $K = 3$ off-grid sources randomly deployed over the covered area. As can be seen from the figure, while STLS-LS is capable of locating the off-grid sources with a PRMSE of about 9cm for a large span of SNRs, ESMTL-Interp. cannot attain an accuracy better than 45cm for high SNRs. This stresses the fact that in order to obtain centimeter accuracy, the ESMTL should be modified with the proposed mismatch recovery process for the case of *multiple off-grid* sources. Notably, the conventional disjoint TDOA algorithms (TDOA and TDOA-CWLS) outperform both ESMTL-Interp. and STLS-LS because they are provided with the signal assignment information and they are independent of the GPs and hence indifferent w.r.t. the off-grid effect. We highlight that for more accurate results, the second proposed approach based on WSSTLS-LS (summarized in Algorithm 2) can be used, but it is more demanding in terms of computational cost. We would also like to comment on the attainable accuracy of the STLS-LS for large SNRs. As is clear from the figure, the attainable accuracy does not considerably improve with SNR for large SNRs. This effect originates from the 1st-order Taylor expansion. Obviously, the larger the size of the cells, the larger the variations of the TDOA in the cell and hence the worse a 1st-order Taylor expansion will work. This effect can be healed to some extent by decreasing the cell size as is confirmed by the simulation results for $N = 400$ where a PRMSE of 3cm (three times better than $N = 100$) is attained by STLS-LS for large SNRs.

VII. CONCLUSION

This paper tackles the problem of multi-source TDOA localization. We have proposed to simplify the involved issues (i.e., solving hyperbolic equations and multi-source disambiguation) by introducing a novel TDOA fingerprinting and grid design paradigm to convert this non-convex problem to a convex ℓ_1 -norm minimization. Moreover, we have proposed a novel trick to enhance the proposed model to be capable of localizing more sources. As a result, we even become able to convert the problem to an overdetermined one which can be efficiently solved using classical LS, if wanted. Finally, in order to extend our ideas, we have proposed two algorithms to handle off-grid sources. Our extensive simulation results corroborate the efficiency of the proposed algorithms in terms of localization accuracy as well as detection capability.

APPENDIX A THE OPTIMAL OPERATOR \mathbf{R}

Let us start by rewriting (15) as

$$\min_{\mathbf{R}} \|(\mathbf{R}\bar{\Psi})^T(\mathbf{R}\bar{\Psi}) - \mathbf{I}_N\|_F^2 = \min_{\mathbf{\Gamma}} \|\bar{\Psi}^T \mathbf{\Gamma} \bar{\Psi} - \mathbf{I}_N\|_F^2, \quad (48)$$

which is due to the fact that the solution of the right-hand side is always symmetric and allows for a decomposition as $\mathbf{\Gamma} = \mathbf{R}^T \mathbf{R}$. By applying $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$, with \otimes denoting the Kronecker product, we can further write

$$\text{vec}(\bar{\Psi}^T \mathbf{\Gamma} \bar{\Psi}) = (\bar{\Psi}^T \otimes \bar{\Psi}^T) \boldsymbol{\gamma}, \quad (49)$$

where $\boldsymbol{\gamma} = \text{vec}(\mathbf{\Gamma})$ and $\text{vec}(\cdot)$ denotes the standard vectorization operator. Therefore, (15) can be rewritten as the following LS problem

$$\min_{\boldsymbol{\gamma}} \|(\bar{\Psi}^T \otimes \bar{\Psi}^T) \boldsymbol{\gamma} - \text{vec}(\mathbf{I}_N)\|_2^2,$$

with its solution given by

$$\boldsymbol{\gamma} = [\bar{\Psi}^T \otimes \bar{\Psi}^T]^\dagger \text{vec}(\mathbf{I}_N). \quad (50)$$

Now, using $(\mathbf{A} \otimes \mathbf{B})^\dagger = \mathbf{A}^\dagger \otimes \mathbf{B}^\dagger$ and $(\mathbf{B}^T)^\dagger = (\mathbf{B}^\dagger)^T$, we can further simplify (50) as

$$\boldsymbol{\gamma} = [(\bar{\Psi}^\dagger)^T \otimes (\bar{\Psi}^\dagger)^T] \text{vec}(\mathbf{I}_N).$$

Next, we have

$$\begin{aligned} \mathbf{\Gamma} &= \text{ivec}(\boldsymbol{\gamma}) = \text{ivec}([(\bar{\Psi}^\dagger)^T \otimes (\bar{\Psi}^\dagger)^T] \text{vec}(\mathbf{I}_N)) \\ &= (\bar{\Psi}^\dagger)^T \mathbf{I}_N \bar{\Psi}^\dagger = (\bar{\Psi}^\dagger)^T \bar{\Psi}^\dagger, \end{aligned} \quad (51)$$

with $\text{ivec}(\cdot)$ denoting the inverse $\text{vec}(\cdot)$ operation, which is indeed a symmetric matrix as claimed earlier. Note that we are now looking for an \mathbf{R} of size $L(M-1) \times L(M-1)$ such that $\mathbf{\Gamma} = \mathbf{R}^T \mathbf{R}$; therefore, the solution is not $\bar{\Psi}^\dagger$. We need to employ the singular value decomposition (SVD) to decompose $\bar{\Psi}$

as $\bar{\Psi} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, and thus $\bar{\Psi}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^\dagger\mathbf{U}^T$, which allows us to rewrite (51) as

$$\begin{aligned} \mathbf{\Gamma} &= \mathbf{U}(\boldsymbol{\Sigma}^\dagger)^T \mathbf{V}^T \mathbf{V} \boldsymbol{\Sigma}^\dagger \mathbf{U}^T \\ &= \mathbf{U}(\boldsymbol{\Sigma}^\dagger)^T \boldsymbol{\Sigma}^\dagger \mathbf{U}^T = \mathbf{R}^T \mathbf{R}. \end{aligned}$$

Therefore, the desired operator \mathbf{R} of size $L(M-1) \times L(M-1)$ if $L(M-1) \leq N$ is given by

$$\mathbf{R} = \boldsymbol{\Sigma}^\dagger(1 : L(M-1), :) \mathbf{U}^T, \quad (52)$$

while if $L(M-1) > N$, it is given by

$$\mathbf{R} = \begin{bmatrix} \boldsymbol{\Sigma}^\dagger \\ \mathbf{0}_{(L(M-1)-N) \times L(M-1)} \end{bmatrix} \mathbf{U}^T. \quad (53)$$

This means that $\mathbf{R}\bar{\Psi} = \mathbf{R}\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ is given by $\mathbf{V}^T(1 : L(M-1), :)$ if $L(M-1) \leq N$ or $[\mathbf{V}, \mathbf{0}_{N \times (L(M-1)-N)}]^T$ if $L(M-1) > N$, which surprisingly means that this is equal to row orthonormalization as proposed in [7]. ■

APPENDIX B COMPUTATION OF THE OPTIMAL $\Delta \mathbf{x}$

Substituting $\tilde{\epsilon}$ from (44b) into (44a) while using $\mathbf{q} = \tilde{\mathbf{y}} - \tilde{\Psi}\boldsymbol{\theta}(m)$ and $\mathbf{S} = \Delta \tilde{\Psi} \text{diag}(\boldsymbol{\theta}(m) \otimes \mathbf{1}_2)$ leads to minimizing

$$\begin{aligned} J &= \Delta \mathbf{x}^T \mathbf{C}_{\Delta \mathbf{x}}^{-1} \Delta \mathbf{x} + [\mathbf{q} - \mathbf{S} \Delta \mathbf{x}]^T \mathbf{C}_{\tilde{\epsilon}}^{-1} [\mathbf{q} - \mathbf{S} \Delta \mathbf{x}] \\ &= \Delta \mathbf{x}^T \mathbf{C}_{\Delta \mathbf{x}}^{-1} \Delta \mathbf{x} + \mathbf{q}^T \mathbf{C}_{\tilde{\epsilon}}^{-1} \mathbf{q} - \mathbf{q}^T \mathbf{C}_{\tilde{\epsilon}}^{-1} \mathbf{S} \Delta \mathbf{x} \\ &\quad + \Delta \mathbf{x}^T \mathbf{S}^T \mathbf{C}_{\tilde{\epsilon}}^{-1} \mathbf{S} \Delta \mathbf{x} - \Delta \mathbf{x}^T \mathbf{S}^T \mathbf{C}_{\tilde{\epsilon}}^{-1} \mathbf{q}. \end{aligned}$$

By taking the partial derivative of J w.r.t. $\Delta \mathbf{x}$ and setting it equal to zero we obtain

$$\frac{\partial J}{\partial \Delta \mathbf{x}} = 2\mathbf{C}_{\Delta \mathbf{x}}^{-1} \Delta \mathbf{x} - 2\mathbf{S}^T \mathbf{C}_{\tilde{\epsilon}}^{-1} \mathbf{q} + 2\mathbf{S}^T \mathbf{C}_{\tilde{\epsilon}}^{-1} \mathbf{S} \Delta \mathbf{x} = \mathbf{0},$$

which results in

$$\Delta \hat{\mathbf{x}}_{\text{WSSTLS}}(m+1) = [\mathbf{C}_{\Delta \mathbf{x}}^{-1} + \mathbf{S}^T \mathbf{C}_{\tilde{\epsilon}}^{-1} \mathbf{S}]^\dagger \mathbf{S}^T \mathbf{C}_{\tilde{\epsilon}}^{-1} \mathbf{q}. \quad \blacksquare$$

REFERENCES

- [1] G. Mao, B. Fidan, and B. D. O. Anderson, "Wireless sensor network localization techniques," *Comput. Netw.*, vol. 51, no. 10, pp. 2529–2553, Jul. 2009.
- [2] B. Li, J. Salter, A. G. Dempster, and C. Rizos, "Indoor positioning techniques based on wireless LAN," in *Proc. Aust. Wireless Conf. (AuWireless)*, Mar. 2006, pp. 13–16.
- [3] D. Madigan, E. Elnahrawy, R. R. Matrin, W. Ju, P. Krishnan, P. Krishnan, and A. Krishnakumar, "Bayesian indoor positioning systems," in *Proc. IEEE Intl. Conf. Comput. Commun. (INFOCOM)*, Mar. 2005, pp. 1217–1227.
- [4] D. L. Donoho, "Compressive sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Sep. 2006.
- [5] V. Cevher and R. G. Baraniuk, "Compressive sensing for sensor calibration," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jul. 2008, pp. 175–178.
- [6] V. Cevher, M. F. Durate, and R. G. Baraniuk, "Distributed target localization via spatial sparsity," in *Proc. Eur. Signal Process. Conf. (EU-SIPCO)*, Lausanne, Switzerland, Aug. 2008.

- [7] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Received signal strength based indoor positioning using compressive sensing," *IEEE Trans. Mobile Comput.*, vol. 11, no. 12, pp. 1983–1993, Dec. 2012.
- [8] S. Nikitaki and P. Tsakalides, "Localization in wireless networks based on jointly compressed sensing," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug.–Sep. 2011, pp. 1809–1813.
- [9] B. Zhang, X. Cheng, N. Zhang, Y. Cui, Y. Li, and Q. Liang, "Sparse target counting and localization in sensor networks based on compressive sensing," in *Proc. IEEE Intl. Conf. Comput. Commun. (INFOCOM)*, Apr. 2011, pp. 2255–2263.
- [10] H. Jamali-Rad, H. Ramezani, and G. Leus, "Sparse multi-target localization using cooperative access points," in *Proc. IEEE Sensor Array Multichannel Process. Workshop (SAM)*, Jun. 2012, pp. 353–356.
- [11] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1479–1489, Nov. 2008.
- [12] L. Yang and K. C. Ho, "An approximately efficient TDOA localization algorithm in closed-form for locating multiple disjoint sources with erroneous sensor positions," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4598–4615, Dec. 2009.
- [13] A. Lombard, T. Rosenkranz, H. Buchner, and W. Kellermann, "Multidimensional localization of multiple sound sources using averaged directivity patterns of blind source separation systems," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2009, pp. 233–236.
- [14] A. Brutti, M. Omologo, and P. Svaizer, "Multiple source localization based on acoustic map de-emphasis," *EURASIP J. Audio, Speech, Music Process.*, vol. 2010, no. 147495, Dec. 2010.
- [15] C. Blandin, A. Ozerov, and E. Vincent, "Multi-source TDOA estimation in reverberant audio using angular spectra and clustering," *Signal Process.*, vol. 92, no. 8, pp. 1950–1960, Aug. 2012.
- [16] C. R. Comsa, A. M. Haimovich, S. C. Schwartz, Y. H. Dobyns, and J. A. Dabin, "Source localization using time difference of arrival within a sparse representation framework," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2011, pp. 2872–2875.
- [17] C. R. Comsa, A. M. Haimovich, S. C. Schwartz, Y. H. Dobyns, and J. A. Dabin, "Time difference of arrival based source localization within a sparse representation framework," in *Proc. Conf. Inf. Syst. (CISS)*, Mar. 2011, pp. 1–6.
- [18] H. Jamali-Rad and G. Leus, "Sparsity-aware TDOA localization of multiple sources," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2013, pp. 4021–4025.
- [19] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.
- [20] K. C. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
- [21] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, July 2011.
- [22] G. B. Giannakis, G. Mateos, S. Farahmand, V. Kekatos, and H. Zhu, "USPACOR: Universal sparsity-controlling outlier rejection," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2011, pp. 1952–1955.
- [23] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?," *SIAM Rev.*, vol. 53, no. 1, pp. 105–125, 2011.
- [24] H. C. So, Y. T. Chan, and F. K. W. Chan, "Closed form formulae for time difference of arrival estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2614–2620, Jun. 2008.
- [25] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
- [26] Y. Wang and G. Leus, "Reference-free time-based localization for an asynchronous target," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 101186, Jan. 2012.
- [27] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009.
- [28] B. Friedlander, "A passive localization algorithm and its accuracy analysis," *IEEE J. Ocean. Eng.*, vol. 12, no. 1, pp. 234–245, Jan. 1987.
- [29] K. Cheung, H. C. So, W.-K. Ma, and Y. T. Chan, "A constrained least squares approach to mobile positioning: Algorithms and optimality," *EURASIP J. Adv. Signal Process.*, vol. 2006, no. 20858, Dec. 2006.
- [30] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.



Hadi Jamali-Rad (S'07) received the B.Sc. degree in Electrical Engineering from the Iran University of Science and Technology (IUST), Tehran, Iran, in 2007 and the M.Sc. degree (Honors) in telecommunications engineering from the IUST, in 2010. He is currently working towards the Ph.D. degree at the department of Electrical Engineering of the Delft University of Technology (TU Delft), Delft, The Netherlands. In 2012, he was a visiting researcher at the Signals, Identification, System Theory and Automation (SISTA) division of the Katholieke

Universiteit Leuven (KU Leuven), Leuven, Belgium. In 2013, he was an intern with Shell Global Solutions, Rijswijk, The Netherlands, working on seismic signal processing. His general interests lie in the area of signal processing for communications, wireless networking, and cooperative communications. Mr. Jamali-Rad has served as a reviewer for several IEEE journals and major conferences.



Geert Leus (F'12) received the electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in 1996 and 2000, respectively. Currently, he is an Antoni van Leeuwenhoek Full Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, Delft, The Netherlands. His research interests are in the area of signal processing for communications. Dr. Leus received a 2002 IEEE Signal Processing Society Young Author Best Paper Award

and a 2005 IEEE Signal processing Society Best Paper Award. He was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE SIGNAL PROCESSING LETTERS. Currently, he is a member of the IEEE Sensor Array and Multichannel Technical Committee and serves as the Editor-in-Chief of the *EURASIP Journal on Advances in Signal Processing*.