Global Strong Convexity and Characterization of Critical Points of Time-of-Arrival-Based Source Localization

Yuen-Man Pun[a,1], Anthony Man-Cho So[b]

[a] *CIICADA Lab, School of Engineering, Australian National University, Canberra, ACT 2601, Australia*

[b] *Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong SAR, China*

Abstract

In this work, we study a least-squares formulation of the source localization problem given time-of-arrival measurements. We show that the formulation, albeit non-convex in general, is globally strongly convex under certain condition on the geometric configuration of the anchors and the source and on the measurement noise. Next, we derive a characterization of the critical points of the least-squares formulation, leading to a bound on the maximum number of critical points under a very mild assumption on the measurement noise. In particular, the result provides a sufficient condition for the critical points of the least-squares formulation to be isolated. Prior to our work, the isolation of the critical points is treated as an assumption without any justification in the localization literature. The said characterization also leads to an algorithm that can find a global optimum of the least-squares formulation by searching through all critical points. We then establish an upper bound of the estimation error of the least-squares estimator. Finally, our numerical results corroborate the theoretical findings and show that our proposed algorithm can obtain a global solution regardless of the geometric configuration of the anchors and the source.

Keywords: source localization, time of arrival (TOA), global optimization,

[1] The first author is supported in part by the Australian Research Council under the Discovery Project DP210102454. Most of the work of the first author was done when she was a Ph.D. student in the Department of Systems Engineering and Engineering Management of The Chinese University of Hong Kong.

global strong convexity, critical points

1. Introduction

The source localization problem has been extensively studied for decades. From emerging applications in mobile phones [7, 6, 1] to asset tracking, intelligent transportation system, and rescue and surveillance [21, 11, 19], localization is ubiquitous in our daily life. There are two main categories of wireless location technologies, namely, mobile-based and network-based [19]. Mobile-based location technology determines the location of a source from signals received from a set of emitters, while network-based location technology determines the location of a source by measuring its signal parameters received at the anchors of the network.

In this work, we consider the latter setting and are interested in locating a source via a collection of noisy measurements from m known anchors. Specifically, let $\boldsymbol{x}^\star \in \mathbb{R}^d$ (with $d = 2, 3$) be the unknown source position. We set m anchors located at $\boldsymbol{a}_i \in \mathbb{R}^d$, for $i = 1, \ldots, m$, to measure the signal propagation time between the source and each anchor. Let $\boldsymbol{c} \coloneqq \frac{1}{m}\sum_{i=1}^m \boldsymbol{a}_i$ be the centroid of the anchors. Suppose that $\{\boldsymbol{a}_i - \boldsymbol{c}\}_{i=1}^m$ spans $\mathbb{R}^d$. For $i = 1, \ldots, m$, the time-of-arrival measurement obtained from the i-th anchor is given by

$$t_i = \frac{1}{c}\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + n_i,$$

where c denotes the speed of the signal and $\boldsymbol{n} = [n_1, \ldots, n_m]^T$ denotes the measurement noise. To simplify notation, we consider the range measurements

$$r_i = \|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i \text{ for } i = 1, \ldots, m,$$

which can be obtained easily from the time-of-arrival measurement, with measurement noise $\boldsymbol{w} = [w_1, \ldots, w_m]^T = c\boldsymbol{n}$. With the above setup, one can then consider a least-squares formulation of the source localization problem in the squared domain [2, 3, 4, 20]:

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \coloneqq \sum_{i=1}^m (\|\boldsymbol{x} - \boldsymbol{a}_i\|^2 - r_i^2)^2. \tag{LS}$$

Although the algorithmic aspect of localization problems has been studied extensively, the geometric landscape of their optimization formulations is largely unexplored. Until recently, some papers show that the maximum-likelihood (ML) functions of some localization problems given range measurements with Gaussian measurement noise possess local strong convexity at their global optima; see, [10, 18]. However, due to the non-smoothness of these ML functions, iterative methods often get stuck at their local minima unless given a good initialization; see the discussion in [3, Example 4.1]. On the contrary, despite the non-convexity of (LS), it is smooth and is known that a global solution can be computed efficiently under some assumptions [2, 3]. This motivates us to study the geometric landscape of problem (LS), which can then be exploited to design algorithms for solving (LS) with theoretical guarantees.

1.1. Related Works

Cheung et al. [4] considered the source localization problem on $\mathbb{R}^2$ and reformulated it as a quadratic minimization problem with a quadratic constraint. The authors solved the problem via the Karush-Kuhn-Tucker (KKT) condition and determined the Lagrange multiplier by a five-root equation, which can be solved via any root-finding algorithm. However, it is not clear whether the result can be extended to the source localization problem on $\mathbb{R}^3$. From an algorithmic point of view, the number of real roots is unknown before solving the equation, thus root-finding algorithms could become less effective.

Later, Beck et al. [2] viewed the same reformulation as in [4] as a *generalized trust region subproblem (GTRS)*, the necessary and sufficient optimality conditions of which and efficient algorithms for which have been extensively studied in the literature; see, e.g., [8, 12, 23]. Specifically, the problem reduces to that of finding a root of a strictly decreasing function over a certain interval, which can be solved via a bisection algorithm. Despite the monotonicity of the function over the interval, every iteration of the bisection algorithm requires finding a matrix inverse, which can be computationally costly. Moreover, this method makes an assumption that depends on the solution and therefore cannot be checked a priori. The method also assumes that f admits isolated stationary points [3]. Although this assumption is considered standard in the literature, it is not clear when it holds.

Another line of work is to solve the localization problem by taking the *sum-of-squares (SOS)* approach; see, e.g., [20, 14]. In [20], Shames et al. reformulated the problem based on SOS relaxation, which can be solved by semidefinite programming techniques. Moreover, it is shown that the

SOS relaxation is exact. However, the approach is computationally heavy in practice and is more of theoretical interest. Nevertheless, the result suggests that (LS) might possess some favorable problem structure for analysis.

1.2. Our Contributions

Existing works design algorithms for problem (LS) without considering the underlying landscape of its objective function f. Although various algorithms can tackle the problem efficiently in practice, not much is known about the reason behind their success, as the problem is non-convex in general. Our work aims to fill this gap in our understanding by studying problem (LS) from the following perspectives:

- *Geometric Perspective:* We show that, under certain assumptions on the geometric configuration of the anchors and the source, problem (LS), albeit non-convex in general, exhibits global strong convexity (see Theorem 1). In particular, if the source lies within the convex hull of regularly located anchors on $\mathbb{R}^2$ and the measurement noise is sufficiently small, then the function f is strongly convex on the whole space (see Example 1). This provides an alternative perspective on why localization inside the convex hull of anchors is favorable from an algorithmic point of view. Besides, we characterize the gradient and Hessian of f via the centroid of anchors (see Proposition 1). As a result, we bound the maximum number of critical points of f under a very mild assumption on the measurement noise (see Propositions 2 and 3). This provides the first sufficient condition for the critical points of f to be isolated. Prior to our work, such a property is assumed without any justification; see, e.g., [3, 17].

- *Algorithmic Perspective:* Based on the characterization of critical points of f, we propose an algorithm, called *critical point finding algorithm (CPFA)*, that finds a global minimum of problem (LS) by searching through all of its critical points via Newton's method (see Algorithm 1). The key observation here is that the negative eigenvalues of the Hessian $\nabla^2 f(\boldsymbol{c})$ reveal the number of critical points of f and the information that facilitates the root finding procedure. If the measurement noise is sufficiently small, we further show that problem (LS) has a unique global minimum, which is also a first in the localization literature (see Proposition 4).

- *Estimation Perspective:* We establish an upper bound on the estimation error of a global minimum of problem (LS), that is, the distance between the global minimum and the true source location $\boldsymbol{x}^\star$ (see Theorem 2). Although similar results have been developed in the localization literature (see, e.g., [10, 18]), our bound is new for problem (LS).

1.3. Notation

The notation in this paper is mostly standard. We use $\|\cdot\|$ and $\|\cdot\|_{\mathrm{op}}$ to denote the Euclidean norm and the operator norm, respectively. We use $\boldsymbol{I}$ to denote the identity matrix. Given a vector $\bar{\boldsymbol{x}} \in \mathbb{R}^d$ and a scalar $\rho \in \mathbb{R}$, we use $B(\bar{\boldsymbol{x}}, \rho) = \{\boldsymbol{x} \in \mathbb{R}^d \colon \|\boldsymbol{x} - \bar{\boldsymbol{x}}\| \leq \rho\}$ to denote the ball centered at $\bar{\boldsymbol{x}}$ with radius ρ. Given a symmetric matrix $\boldsymbol{A}$, we use $\lambda(\boldsymbol{A})$ to denote its eigenvalues. In particular, $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ denote the largest and smallest eigenvalues of $\boldsymbol{A}$, respectively. We use $\boldsymbol{A} \succ \boldsymbol{0}$ to indicate that $\boldsymbol{A}$ is positive definite.

2. Optimization Landscape

Recent works [10, 18] show that some non-convex localization problems are locally strongly convex around their global optima. This suggests that, despite its non-convexity, problem (LS) might possess some desirable global optimization landscape that can be exploited in algorithm design. Surprisingly, under certain assumption on the configuration of the anchors and the source and on the measurement noise, the function f is *globally* strongly convex on $\mathbb{R}^d$. The following theorem presents a necessary and sufficient condition for the strong convexity of f to hold.

Theorem 1. *Define*

$$R^2 := \|\boldsymbol{x}^\star - \boldsymbol{c}\|^2 + \frac{1}{m}\sum_{i=1}^{m}(2w_i\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i^2). \tag{1}$$

We have $\nabla^2 f(\boldsymbol{x}) \succ \boldsymbol{0}$ for all $\boldsymbol{x} \in \mathbb{R}^d$ if and only if

$$R^2 < \frac{2}{m}\lambda_{\min}\left(\sum_{i=1}^{m}(\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T\right). \tag{2}$$

Proof. First, the Hessian of f in (LS) can be computed as

$$\nabla^2 f(\boldsymbol{x}) = \sum_{i=1}^{m}\left[4(\|\boldsymbol{x} - \boldsymbol{a}_i\|^2 - r_i^2)\boldsymbol{I} + 8(\boldsymbol{x} - \boldsymbol{a}_i)(\boldsymbol{x} - \boldsymbol{a}_i)^T\right]. \tag{3}$$

Expanding the first term and recalling that $\boldsymbol{c} = \frac{1}{m}\sum_{i=1}^m \boldsymbol{a}_i$, we get

$$\begin{aligned}\sum_{i=1}^m (\|\boldsymbol{x} - \boldsymbol{a}_i\|^2 - r_i^2) &= \sum_{i=1}^m (\|\boldsymbol{x}\|^2 - 2\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle + \|\boldsymbol{a}_i\|^2 - r_i^2) \\ &= m\left(\|\boldsymbol{x}\|^2 - 2\left\langle \frac{1}{m}\sum_{i=1}^m \boldsymbol{a}_i, \boldsymbol{x}\right\rangle\right) + \sum_{i=1}^m \left(\|\boldsymbol{a}_i\|^2 - r_i^2\right) \\ &= m\left(\|\boldsymbol{x} - \boldsymbol{c}\|^2\right) - m\left(\frac{1}{m}\sum_{i=1}^m (r_i^2 - \|\boldsymbol{a}_i\|^2) + \|\boldsymbol{c}\|^2\right).\end{aligned} \tag{4}$$

Since $r_i^2 = \|\boldsymbol{x}^\star - \boldsymbol{a}_i\|^2 + 2w_i\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i^2$ for $i = 1, \ldots, m$, by completing squares for the second term above, we have

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^m (r_i^2 - \|\boldsymbol{a}_i\|^2) + \|\boldsymbol{c}\|^2 \\ &= \frac{1}{m}\sum_{i=1}^m (\|\boldsymbol{x}^\star - \boldsymbol{a}_i\|^2 - \|\boldsymbol{a}_i\|^2) + \frac{1}{m}\sum_{i=1}^m (2w_i\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i^2) + \|\boldsymbol{c}\|^2 \\ &= \|\boldsymbol{x}^\star\|^2 - 2\left\langle \frac{1}{m}\sum_{i=1}^m \boldsymbol{a}_i, \boldsymbol{x}^\star\right\rangle + \|\boldsymbol{c}\|^2 + \frac{1}{m}\sum_{i=1}^m (2w_i\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i^2) \\ &= \|\boldsymbol{x}^\star - \boldsymbol{c}\|^2 + \frac{1}{m}\sum_{i=1}^m (2w_i\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i^2) = R^2.\end{aligned} \tag{5}$$

Substituting (5) into (4), we obtain

$$4\sum_{i=1}^m (\|\boldsymbol{x} - \boldsymbol{a}_i\|^2 - r_i^2) = 4m\left(\|\boldsymbol{x} - \boldsymbol{c}\|^2 - R^2\right). \tag{6}$$

Putting it into (3), the Hessian of f can then be written as

$$\nabla^2 f(\boldsymbol{x}) = 4m(\|\boldsymbol{x} - \boldsymbol{c}\|^2 - R^2)\boldsymbol{I} + 8\left(\sum_{i=1}^m (\boldsymbol{x} - \boldsymbol{a}_i)(\boldsymbol{x} - \boldsymbol{a}_i)^T\right). \tag{7}$$

This implies that the smallest eigenvalue of $\nabla^2 f(\boldsymbol{x})$ is given by

$$\lambda_{\min}(\nabla^2 f(\boldsymbol{x})) = 4m(\|\boldsymbol{x} - \boldsymbol{c}\|^2 - R^2) + 8\lambda_{\min}\left(\sum_{i=1}^m (\boldsymbol{x} - \boldsymbol{a}_i)(\boldsymbol{x} - \boldsymbol{a}_i)^T\right). \tag{8}$$

Hence, the "only if" part can be trivially obtained by putting $\boldsymbol{x} = \boldsymbol{c}$ in (8) and noting that $\lambda_{\min}(\nabla^2 f(\boldsymbol{c})) > 0$.

Conversely, note that

$$\begin{aligned}
&\lambda_{\min}\left(\sum_{i=1}^{m}(\boldsymbol{x}-\boldsymbol{a}_i)(\boldsymbol{x}-\boldsymbol{a}_i)^T\right)\\
&= \min_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^T\left(\sum_{i=1}^{m}(\boldsymbol{x}-\boldsymbol{a}_i)(\boldsymbol{x}-\boldsymbol{a}_i)^T\right)\boldsymbol{v} = \min_{\|\boldsymbol{v}\|=1}\sum_{i=1}^{m}(\boldsymbol{v}^T(\boldsymbol{x}-\boldsymbol{a}_i))^2.
\end{aligned}$$

Putting this back to (8), we obtain

$$\lambda_{\min}(\nabla^2 f(\boldsymbol{x})) = \min_{\|\boldsymbol{v}\|=1} 4m\left(\|\boldsymbol{x}-\boldsymbol{c}\|^2 - R^2 + \frac{2}{m}\sum_{i=1}^{m}(\boldsymbol{v}^T(\boldsymbol{x}-\boldsymbol{a}_i))^2\right). \tag{9}$$

Now, for any $\boldsymbol{v} \in \mathbb{R}^d$ satisfying $\|\boldsymbol{v}\| = 1$, define

$$g_{\boldsymbol{v}}(\boldsymbol{x}) \coloneqq 4m\left(\|\boldsymbol{x}-\boldsymbol{c}\|^2 - R^2 + \frac{2}{m}\sum_{i=1}^{m}(\boldsymbol{v}^T(\boldsymbol{x}-\boldsymbol{a}_i))^2\right).$$

Then, the gradient and Hessian of $g_{\boldsymbol{v}}$ can be computed as

$$\begin{aligned}
\nabla g_{\boldsymbol{v}}(\boldsymbol{x}) =& 4m\left(2(\boldsymbol{x}-\boldsymbol{c}) + \frac{2}{m}\sum_{i=1}^{m}2(\boldsymbol{v}^T(\boldsymbol{x}-\boldsymbol{a}_i))\boldsymbol{v}\right)\\
=& 4m\left(2(\boldsymbol{x}-\boldsymbol{c}) + 4\left(\boldsymbol{v}^T(\boldsymbol{x}-\boldsymbol{c})\right)\boldsymbol{v}\right)
\end{aligned}$$

and

$$\nabla^2 g_{\boldsymbol{v}}(\boldsymbol{x}) = 4m(2\boldsymbol{I} + 4\boldsymbol{v}\boldsymbol{v}^T).$$

In particular, we see that

$$\nabla^2 g_{\boldsymbol{v}}(\boldsymbol{x}) \succ \boldsymbol{0} \quad \text{and} \quad \nabla g_{\boldsymbol{v}}(\boldsymbol{c}) = \boldsymbol{0}.$$

This implies that for any $\boldsymbol{v} \in \mathbb{R}^d$ satisfying $\|\boldsymbol{v}\| = 1$, the global minimum of $g_{\boldsymbol{v}}$ is attained at $\boldsymbol{c}$. By putting $\boldsymbol{x} = \boldsymbol{c}$ in (9) and requiring the resulting expression to be positive, we obtain condition (2), as desired. □

Theorem 1 is interesting because the objective function in (LS), albeit non-convex in general, could exhibit strong convexity on the whole space $\mathbb{R}^d$

under certain condition on the geometric configuration of the source and the anchors and on the measurement noise. Here, let us interpret condition (2) by considering its left-hand side (lhs) and right-hand side (rhs) separately.

On the lhs, if we assume that $w_i \ll \|\boldsymbol{x}^\star - \boldsymbol{a}_i\|$ for $i = 1, \ldots, m$,[2] then the quantity R^2 roughly measures the squared distance between the source and the centroid of anchors. On the rhs, $\frac{2}{m}\lambda_{\min}(\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T) = \frac{2}{m}\sum_{i=1}^m (\tilde{\boldsymbol{v}}^T(\boldsymbol{a}_i - \boldsymbol{c}))^2$ is proportional to the average of the squared magnitude of $\boldsymbol{a}_i - \boldsymbol{c}$ projected onto the vector $\tilde{\boldsymbol{v}}$, where $\tilde{\boldsymbol{v}}$ is an eigenvector corresponding to the smallest eigenvalue of $\nabla^2 f(\boldsymbol{c})$. Such a value can be viewed as a measure of degeneracy of the anchors. For example, when $\{\boldsymbol{a}_i - \boldsymbol{c}\}_{i=1}^m$ do not span $\mathbb{R}^d$, the smallest eigenvalue is zero, thereby the rhs of (2) vanishes. Geometrically, condition (2) is equivalent to $B(\boldsymbol{c}, R)$ lying within $B(\boldsymbol{c}, \rho)$, where R is defined in (1) and ρ is defined as $\rho = \sqrt{\frac{2}{m}\lambda_{\min}(\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T)}$.

In other words, the non-degeneracy of the anchors and the nearness of the source $\boldsymbol{x}^\star$ to the centroid $\boldsymbol{c}$ of anchors are sufficient to guarantee the strong convexity of f on the whole space $\mathbb{R}^d$. Contrary to the common belief in the literature that localization inside the convex hull of anchors offers better performance in a statistical sense (see, e.g., [16]), Theorem 1 provides an alternative perspective of what is a desirable configuration in a source localization problem in an algorithmic sense. When the anchors are regularly located, the convex hull of anchors roughly approximates the region depicted in condition (2). On the other hand, when the vectors $\{\boldsymbol{a}_i\}_{i=1}^m$ do not spread evenly on $\mathbb{R}^d$, the region that the source $\boldsymbol{x}^\star$ can be located for f to enjoy a desirable geometry is much smaller. To understand condition (2) better, we now look into some examples.

Example 1. *Consider the source localization on $\mathbb{R}^2$. Suppose that there are m anchors located at $(\ell \cos \frac{2\pi i}{m}, \ell \sin \frac{2\pi i}{m})$, for $i = 1, \ldots, m$; i.e., the anchors are distributed uniformly on a circle centered at $(0, 0)$ with radius ℓ. It can*

[2] Such an assumption is standard in the localization literature; see, e.g., [24, 18].

be easily computed that $\boldsymbol{c} = \boldsymbol{0}$ *and*

$$\begin{aligned}\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T &= \sum_{i=1}^{m} \begin{pmatrix} \ell \cos \frac{2\pi i}{m} \\ \ell \sin \frac{2\pi i}{m} \end{pmatrix} \begin{pmatrix} \ell \cos \frac{2\pi i}{m} & \ell \sin \frac{2\pi i}{m} \end{pmatrix} \\ &= \sum_{i=1}^{m} \begin{pmatrix} \ell^2 \cos^2 \frac{2\pi i}{m} & \ell^2 \cos \frac{2\pi i}{m} \sin \frac{2\pi i}{m} \\ \ell^2 \cos \frac{2\pi i}{m} \sin \frac{2\pi i}{m} & \ell^2 \sin^2 \frac{2\pi i}{m} \end{pmatrix} \\ &= \ell^2 \begin{pmatrix} \frac{m}{2} & 0 \\ 0 & \frac{m}{2} \end{pmatrix},\end{aligned}$$

where the last equality is given by standard trigonometric identities. This implies that condition (2) *is satisfied if the source lies within the ball* $B(\boldsymbol{0}, \ell)$ *given sufficiently small noise. Here, the ball* $B(\boldsymbol{0}, \ell)$ *fully covers the convex hull of the anchors* $\{\boldsymbol{a}_i\}_{i=1}^m$*. Figure 1a illustrates the case where* $m = 4$*.*

Example 2. *Consider the source localization on* $\mathbb{R}^3$*. Suppose that there are 8 anchors located at* $\boldsymbol{a}_1 = (\ell, \ell, \ell)$*,* $\boldsymbol{a}_2 = (\ell, \ell, -\ell)$*,* $\boldsymbol{a}_3 = (\ell, -\ell, \ell)$*,* $\boldsymbol{a}_4 = (-\ell, \ell, \ell)$*,* $\boldsymbol{a}_5 = (-\ell, -\ell, \ell)$*,* $\boldsymbol{a}_6 = (-\ell, \ell, -\ell)$*,* $\boldsymbol{a}_7 = (\ell, -\ell, -\ell)$*, and* $\boldsymbol{a}_8 = (-\ell, -\ell, -\ell)$*; i.e., the anchors are at the vertices of a cube centered at* $(0, 0, 0)$ *with side length* 2ℓ*. It can be computed that* $\boldsymbol{c} = \boldsymbol{0}$ *and*

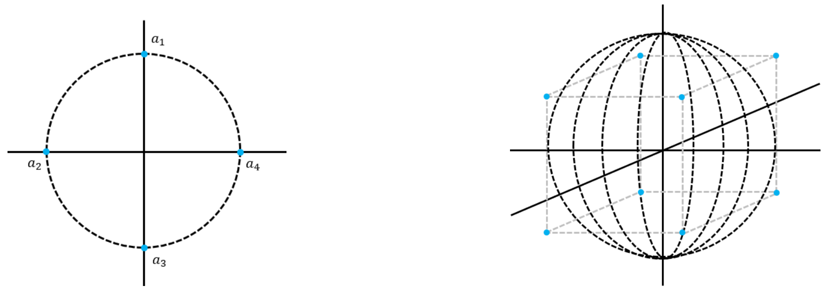
$$\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T = 8 \begin{pmatrix} \ell^2 & 0 & 0 \\ 0 & \ell^2 & 0 \\ 0 & 0 & \ell^2 \end{pmatrix}.$$

Therefore, if the source lies within the ball with radius $\sqrt{2}\ell$ *(as illustrated in Figure 1b), condition (2.7) is satisfied. This shows that the ball* $B(\boldsymbol{0}, \sqrt{2}\ell)$ *roughly covers the convex hull of the anchors* $\{\boldsymbol{a}_i\}_{i=1}^m$*.*

The above two examples correspond to the scenario where the anchors are regularly located, which allows the vectors $\{\boldsymbol{a}_i - \boldsymbol{c}\}_{i=1}^m$ to spread uniformly over $\mathbb{R}^d$. Next, let us consider another two examples in which the anchors are less regularly located.

Example 3. *Consider the source localization problem on* $\mathbb{R}^2$*. Suppose that the anchors are located at the corners of a rectangle; i.e.,* $\boldsymbol{a}_1 = (w, h)$*,* $\boldsymbol{a}_2 = (-w, h)$*,* $\boldsymbol{a}_3 = (-w, -h)$*, and* $\boldsymbol{a}_4 = (w, -h)$*, for some* $w, h > 0$*. Similar to the previous examples, we have* $\boldsymbol{c} = \boldsymbol{0}$ *and*

$$\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T = 4 \begin{pmatrix} w^2 & 0 \\ 0 & h^2 \end{pmatrix}.$$



(a) Example 1.

(b) Example 2.

Figure 1: The geometry of anchors (denoted by the blue dots) and the ball $B(\boldsymbol{c}, \rho)$ (denoted by the black dashed lines) in Examples 1 and 2.

Condition (2) *is equivalent to* $R^2 < 2\min\{w^2, h^2\}$*. The ball* $B(\mathbf{0}, \sqrt{2}\min\{w, h\})$ *is shown by the black dashed line in Figure 2a. As the rectangle gets flatter,* $B(\mathbf{0}, \sqrt{2}\min\{w, h\})$ *gets much smaller than the convex hull of the anchors.*

Example 4. *Consider the source localization problem on* $\mathbb{R}^2$*. Suppose that there are 9 anchors located at* $\boldsymbol{a}_1 = (w, h)$*,* $\boldsymbol{a}_2 = (-w, h)$*,* $\boldsymbol{a}_3 = (-w, -h)$*,* $\boldsymbol{a}_4 = (w, -h)$*,* $\boldsymbol{a}_5 = (\frac{w}{n}, \frac{h}{n})$*,* $\boldsymbol{a}_6 = (-\frac{w}{n}, \frac{h}{n})$*,* $\boldsymbol{a}_7 = (-\frac{w}{n}, -\frac{h}{n})$*,* $\boldsymbol{a}_8 = (\frac{w}{n}, -\frac{h}{n})$*, and* $\boldsymbol{a}_9 = (0, 0)$*, for some* $w, h, n > 0$*; see Figure 2b. It can be computed that* $\boldsymbol{c} = \mathbf{0}$ *and*

$$\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T = 4\left(1 + \frac{1}{n^2}\right)\begin{pmatrix} w^2 & 0 \\ 0 & h^2 \end{pmatrix}.$$

According to condition (2)*, if*

$$R^2 < \frac{2}{m}\lambda_{\min}\left(\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T\right) = \frac{8}{9}\left(1 + \frac{1}{n^2}\right)\min\{w^2, h^2\} = \rho^2,$$

then $\nabla^2 f(\boldsymbol{x}) \succ \mathbf{0}$ *for all* $\boldsymbol{x} \in \mathbb{R}^d$*. Given that the first four anchors* $\{\boldsymbol{a}_i\}_{i=1}^4$ *the same as those in Example 3, it is surprising to see that the ball* $B(\mathbf{0}, \rho)$ *becomes smaller after adding five more anchors* $\{\boldsymbol{a}_i\}_{i=5}^9$ *inside their convex hull. The ball becomes even smaller when* n *gets larger; i.e., the anchors are getting closer to the centroid. This demonstrates that adding anchors can destroy the strong convexity of the function* f*.*

Next, we present one example with an asymmetric set of anchors and another example with a very small region $B(\boldsymbol{c}, \rho)$.

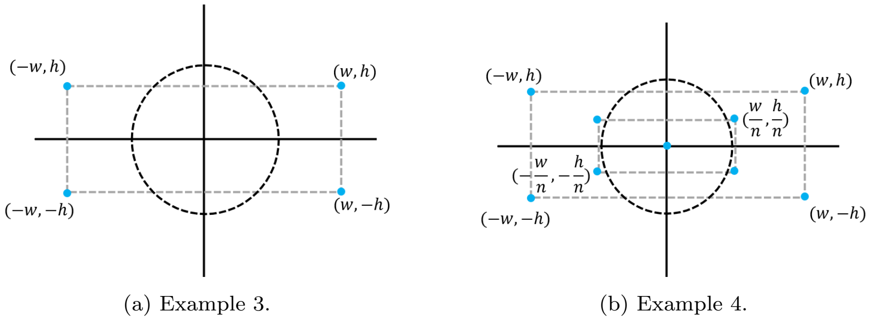


Figure 2: The geometry of anchors (denoted by the blue dots) and the ball $B(\boldsymbol{c}, \rho)$ (denoted by the black dashed lines) in Examples 3 and 4.

Example 5. *Consider the source localization problem on* $\mathbb{R}^2$*. Suppose that there are 4 anchors located at* $\boldsymbol{a}_1 = (\ell, \ell)$*,* $\boldsymbol{a}_2 = (\ell/2, -\ell)$*,* $\boldsymbol{a}_3 = (-\ell, 2\ell)$*, and* $\boldsymbol{a}_4 = (-\ell/2, -2\ell)$*, for some* $\ell > 0$*; see Figure 3a. It can be computed that* $\boldsymbol{c} = \boldsymbol{0}$ *and*

$$\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T = \ell^2 \begin{pmatrix} 2.5 & -0.5 \\ -0.5 & 10 \end{pmatrix}.$$

According to condition (2)*, if*

$$R^2 < \frac{2}{m} \lambda_{\min} \left(\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T \right) = 1.2334\ell^2 = \rho^2,$$

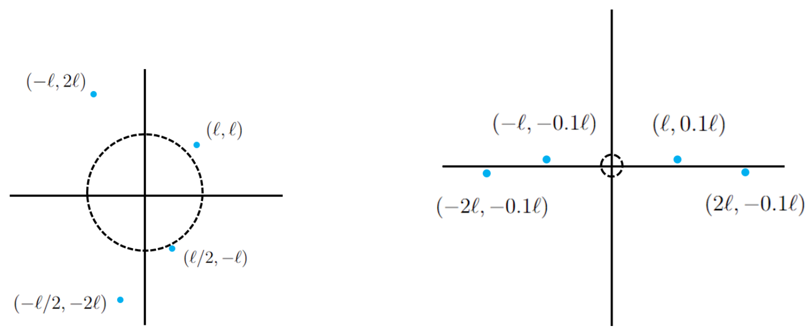
then $\nabla^2 f(\boldsymbol{x}) \succ \boldsymbol{0}$ *for all* $\boldsymbol{x} \in \mathbb{R}^d$*. Although the anchors are located asymmetrically, we see that the region* $B(\boldsymbol{c}, \rho)$ *is of reasonable size. This is because the vectors* $\{\boldsymbol{a}_i - \boldsymbol{c}\}_{i=1}^m$ *span the space* $\mathbb{R}^d$ *rather evenly.*

Example 6. *Consider the source localization problem on* $\mathbb{R}^2$*. Suppose that there are 4 anchors located at* $\boldsymbol{a}_1 = (-\ell, -0.1\ell)$*,* $\boldsymbol{a}_2 = (\ell, 0.1\ell)$*,* $\boldsymbol{a}_3 = (-2\ell, -0.1\ell)$*, and* $\boldsymbol{a}_4 = (2\ell, -0.1\ell)$*, for some* $\ell > 0$*; see Figure 3b. It can be computed that* $\boldsymbol{c} = \boldsymbol{0}$ *and*

$$\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T = \begin{pmatrix} 10\ell^2 & 0 \\ 0 & 0.04^2 \end{pmatrix}.$$

According to condition (2)*, if*

$$R^2 < \frac{2}{m} \lambda_{\min} \left(\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T \right) = 0.1\sqrt{2}\ell = \rho^2,$$



(a) Example 5.

(b) Example 6.

Figure 3: The geometry of anchors (denoted by the blue dots) and the ball $B(\boldsymbol{c}, \rho)$ (denoted by the black dashed lines) in Examples 5 and 6.

then $\nabla^2 f(\boldsymbol{x}) \succ \mathbf{0}$ *for all* $\boldsymbol{x} \in \mathbb{R}^d$. *In this example, the trapezium formed by the four anchors is very flat, implying that the minimum eigenvalue of* $\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T$ *is very small. As a result, the region* $B(\boldsymbol{c}, \rho)$ *is also very small.*

Strong convexity of a function is powerful in that it favors the fast convergence of many numerical algorithms. Here, we record a corollary regarding the convergence rate of gradient descent when applied to a strongly convex objective function.

Corollary 1. *If condition* (2) *holds, then gradient descent converges linearly to the global minimum of problem* (LS) *regardless of the initialization.*

Proof. Follows directly from, e.g., [13, Theorem 2.1.15]. □

3. Critical Points

In the previous section, we have shown that condition (2) is both necessary and sufficient for the objective function f in (LS) to be strongly convex. Such a condition involves not only the geometric configuration of the source and anchors but also the measurement noise. In this section, we consider the general setting where condition (2) need not hold and study the critical points of f. Let us begin by expressing the gradient and Hessian of f in terms of the centroid of anchors.

Proposition 1. *For any* $\boldsymbol{v} \in \mathbb{R}^d$, *we have*

$$\nabla f(\boldsymbol{c} + \boldsymbol{v}) = 4m\|\boldsymbol{v}\|^2 \boldsymbol{v} + \nabla^2 f(\boldsymbol{c})\boldsymbol{v} + \nabla f(\boldsymbol{c}) \tag{10}$$

and

$$\nabla^2 f(\boldsymbol{c}+\boldsymbol{v}) = \nabla^2 f(\boldsymbol{c}) + 4m\|\boldsymbol{v}\|^2\boldsymbol{I} + 8m\boldsymbol{v}\boldsymbol{v}^T. \tag{11}$$

Proof. Every $\boldsymbol{x} \in \mathbb{R}^d$ can be written as $\boldsymbol{x} = \boldsymbol{c}+\boldsymbol{v}$ for some $\boldsymbol{v} \in \mathbb{R}^d$. Plugging this into (7), we have

$$\begin{aligned}\nabla^2 f(\boldsymbol{c}+\boldsymbol{v}) =& 4m(\|\boldsymbol{v}\|^2 - R^2)\boldsymbol{I} + 8\left(\sum_{i=1}^m (\boldsymbol{c}+\boldsymbol{v}-\boldsymbol{a}_i)(\boldsymbol{c}+\boldsymbol{v}-\boldsymbol{a}_i)^T\right)\\ =& 4m(\|\boldsymbol{v}\|^2 - R^2)\boldsymbol{I} + 8\left(\sum_{i=1}^m \left((\boldsymbol{c}-\boldsymbol{a}_i)(\boldsymbol{c}-\boldsymbol{a}_i)^T + \boldsymbol{v}\boldsymbol{v}^T\right)\right)\\ =& \nabla^2 f(\boldsymbol{c}) + 4m\|\boldsymbol{v}\|^2\boldsymbol{I} + 8m\boldsymbol{v}\boldsymbol{v}^T.\end{aligned}$$

By the fundamental theorem of calculus, we can then write the gradient as

$$\begin{aligned}\nabla f(\boldsymbol{c}+\boldsymbol{v}) - \nabla f(\boldsymbol{c}) =& \int_0^1 \nabla^2 f(\boldsymbol{c}+t\boldsymbol{v})\boldsymbol{v}dt\\ =& \int_0^1 (\nabla^2 f(\boldsymbol{c}) + 12mt^2\|\boldsymbol{v}\|^2\boldsymbol{I})\boldsymbol{v}dt\\ =& (\nabla^2 f(\boldsymbol{c}) + 4m\|\boldsymbol{v}\|^2\boldsymbol{I})\boldsymbol{v}.\end{aligned}$$

Rearranging the terms yields the desired result. □

Proposition 1 is significant, as it elucidates the geometry of the objective function f in (LS). For one thing, it shows that as a point $\boldsymbol{x}$ gets further away from the centroid $\boldsymbol{c}$ of the anchors, all eigenvalues of $\nabla^2 f(\boldsymbol{x})$ get larger. In particular, this implies that f is locally strongly convex at any $\boldsymbol{x}$ that is sufficiently far from $\boldsymbol{c}$. For another thing, it provides a more direct path to understanding the critical points of f. Indeed, instead of characterizing the critical points of f as those vectors $\boldsymbol{x}$ that satisfy

$$\nabla f(\boldsymbol{x}) = \sum_{i=1}^m 4(\|\boldsymbol{x}-\boldsymbol{a}_i\|^2 - r_i^2)(\boldsymbol{x}-\boldsymbol{a}_i) = \boldsymbol{0},$$

we can characterize them as those vectors $\boldsymbol{v}$ that satisfy

$$\nabla f(\boldsymbol{c}+\boldsymbol{v}) = 4m\|\boldsymbol{v}\|^2\boldsymbol{v} + \nabla^2 f(\boldsymbol{c})\boldsymbol{v} + \nabla f(\boldsymbol{c}) = \boldsymbol{0}. \tag{12}$$

The latter is more amenable to analysis. Based on the characterization (12), we can deduce the maximum number of critical points of f.

Proposition 2. *Suppose that for $i = 1, \ldots, m$, the measurement noise w_i follows a probability distribution that is absolutely continuous with respect to the Lebesgue measure. Let $\gamma \leq d$ be the number of negative eigenvalues of $\nabla^2 f(\boldsymbol{c})$. Then, the number of critical points of f is at most $2\gamma + 1$ almost surely.*

Proof. Suppose that $\nabla^2 f(\boldsymbol{c}) = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ for some orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{d\times d}$ and diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}$. Write $\boldsymbol{U} = \begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_d\end{bmatrix}$ and $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_d)$. Every vector $\boldsymbol{x} \in \mathbb{R}^d$ can be written as $\boldsymbol{x} = \boldsymbol{c} + \boldsymbol{v}$ for some vector $\boldsymbol{v} \in \mathbb{R}^d$. Since the vectors $\{\boldsymbol{u}_i\}_{i=1}^d$ form a basis of $\mathbb{R}^d$, we can write $\nabla f(\boldsymbol{c}) = \sum_{i=1}^d b_i \boldsymbol{u}_i$ and $\boldsymbol{v} = \sum_{i=1}^d p_i \boldsymbol{u}_i$. Denote

$$s = \sum_{i=1}^{d} p_i^2. \tag{13}$$

Then, (12) can be written as

$$4msp_i + \lambda_i p_i + b_i = 0, \quad i = 1, \ldots, d. \tag{14}$$

To proceed, we need the following two lemmas.

Lemma 1. *The matrices $\nabla^2 f(\boldsymbol{c})$ and $\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T$ are simultaneously diagonalizable.*

Lemma 2. *Let $\boldsymbol{z} \in \mathbb{R}^d$ be any non-zero vector. Suppose that for $i = 1, \ldots, m$, the measurement noise w_i follows a probability distribution that is absolutely continuous with respect to the Lebesgue measure. Then, the event $\boldsymbol{z}^T \nabla f(\boldsymbol{c}) = 0$ is of measure zero.*

Although $\nabla^2 f(\boldsymbol{c})$ involves the measurement noise $\{w_i\}_{i=1}^m$, Lemma 1 shows that the eigenbasis $\boldsymbol{U}$ of $\nabla^2 f(\boldsymbol{c})$ is deterministic and is determined by the geometric configuration of the anchors $\{\boldsymbol{a}_i\}_{i=1}^m$. Since we have $\boldsymbol{b} = \boldsymbol{U}^T \nabla f(\boldsymbol{c})$ by definition, Lemma 2 implies that $b_i \neq 0$ for $i = 1, \ldots, d$ almost surely. This, together with (14), implies that $4ms + \lambda_i \neq 0$ and $p_i \neq 0$ for $i = 1, \ldots, m$. Upon combining (13) and (14), we see that s satisfies

$$\sum_{i=1}^{d} \left(-\frac{b_i}{4ms + \lambda_i} \right)^2 = s. \tag{15}$$

In other words, s satisfies the system of equations

$$\begin{cases} y = \sum_{i=1}^{d} \left(\frac{b_i}{4ms + \lambda_i} \right)^2 =: h(s) \\ y = s. \end{cases} \tag{16}$$

It is clear that for $i = 1, \ldots, d$, the function h is discontinuous at $s_i = -\frac{\lambda_i}{4m}$ if $\lambda_i < 0$. Without loss of generality, let us assume that the eigenvalues are given in ascending order; i.e., $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$. Let $\gamma \leq d$ be the number of negative eigenvalues of $\nabla^2 f(\boldsymbol{c})$. Then, the negative eigenvalues of $\nabla^2 f(\boldsymbol{c})$ are given by $0 > \lambda_\gamma \geq \cdots \geq \lambda_1$. Consequently, the function h possesses γ discontinuity points, which break $\mathbb{R}_+$ into $\gamma + 1$ intervals. Note that h is well-defined on each of these intervals. Now, observe that the second derivative of h is given by

$$h''(s) = 96m^2 \sum_{i=1}^{d} \frac{b_i^2}{(4ms + \lambda_i)^4}. \tag{17}$$

We see that $h''(s) > 0$ for all s except at those discontinuity points and $h''(s) \geq C$ for some constant $C > 0$ when s is bounded. This implies that h is locally strongly convex on the first γ intervals and strictly convex on the last interval. Therefore, there are at most two intersection points between the graph $y = h(s)$ and the line $y = s$ on $\left(-\frac{\lambda_{j+1}}{4m}, -\frac{\lambda_j}{4m}\right)$ for $j = 1, \ldots, \gamma$, with $\lambda_{\gamma+1}$ being defined as zero.

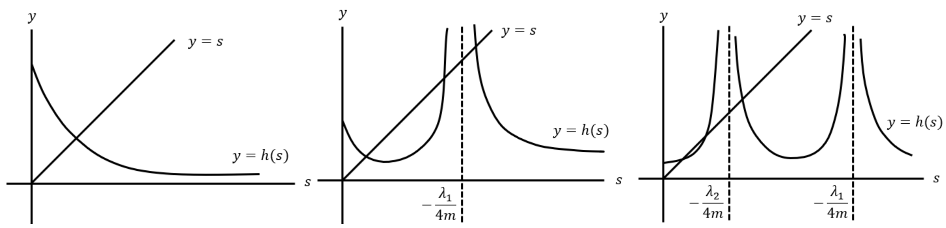
Moreover, since the first derivative of h is given by

$$h'(s) = -8m \sum_{i=1}^{d} \frac{b_i^2}{(4ms + \lambda_i)^3},$$

the function h is monotonically decreasing and tends to zero as $s \to \infty$ on the interval $\left(\max\left\{0, -\frac{\lambda_1}{4m}\right\}, \infty\right)$. Thus, the graph $y = h(s)$ intersects with the line $y = s$ at most once on this interval. It follows that the total number of solutions to (15) is at most $2\gamma + 1$. $\square$

Proof of Lemma 1. Plugging $\boldsymbol{x} = \boldsymbol{c}$ into (7), the Hessian of f at $\boldsymbol{c}$ is given by

$$\nabla^2 f(\boldsymbol{c}) = -4mR^2 \boldsymbol{I} + 8 \left(\sum_{i=1}^{m} (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T \right),$$



(a) $0 \leq \lambda_1 \leq \lambda_2$. (b) $\lambda_1 < 0 \leq \lambda_2$. (c) $\lambda_1 \leq \lambda_2 < 0$.

Figure 4: Sketch of system (16) in different localization scenarios on $\mathbb{R}^2$.

where R is defined in (1). Suppose that $\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^T$ for some orthogonal matrix $\boldsymbol{V}$ and diagonal matrix $\boldsymbol{\Sigma}$. Then,

$$\nabla^2 f(\boldsymbol{c}) = -4mR^2\boldsymbol{I} + 8\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{V}\left(8\boldsymbol{\Sigma} - 4mR^2\boldsymbol{I}\right)\boldsymbol{V}^T,$$

where $8\boldsymbol{\Sigma} - 4mR^2\boldsymbol{I}$ is also a diagonal matrix. Therefore, both $\nabla^2 f(\boldsymbol{c})$ and $\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T$ share the same eigenvectors, which establishes the lemma. □

Proof of Lemma 2. Recall that $r_i = \|\boldsymbol{x}^\star - \boldsymbol{a}_i\|_2 + w_i$ for $i = 1, \ldots, m$. The gradient of f at $\boldsymbol{c}$ is given by

$$\begin{aligned}\nabla f(\boldsymbol{c}) = & \sum_{i=1}^m 4\left(\|\boldsymbol{c} - \boldsymbol{a}_i\|^2 - \|\boldsymbol{x}^\star - \boldsymbol{a}_i\|^2\right)(\boldsymbol{c} - \boldsymbol{a}_i) \\ & - \sum_{i=1}^m 8w_i\|\boldsymbol{x}^\star - \boldsymbol{a}_i\|(\boldsymbol{c} - \boldsymbol{a}_i) + \sum_{i=1}^m 4w_i^2(\boldsymbol{c} - \boldsymbol{a}_i).\end{aligned}$$

Therefore, the equation $\boldsymbol{z}^T\nabla f(\boldsymbol{c}) = 0$ is quadratic in $\boldsymbol{w}$. Upon applying the result in [15], the desired result is obtained. □

Under a very mild assumption on the measurement noise, Proposition 2 reveals that the number of critical points of f is closely related to the number of negative eigenvalues of the Hessian of f at $\boldsymbol{c}$. Note that using (7), the eigenvalues of $\nabla^2 f(\boldsymbol{c})$ are given by

$$\lambda(\nabla^2 f(\boldsymbol{c})) = -4mR^2 + 8\lambda\left(\sum_{i=1}^m (\boldsymbol{c} - \boldsymbol{a}_i)(\boldsymbol{c} - \boldsymbol{a}_i)^T\right),$$

where R (defined in (1)) roughly measures the distance between the source and the centroid of anchors. Hence, it is expected that the number of critical points of f is smaller when the anchors are more regularly located and the source is closer to the centroid of anchors, and vice versa. This is the first result in the literature that shows the relationship between the geometry of the source localization problem and the maximum number of critical points. Furthermore, the result implies that, as long as the measurement noise follows a probability distribution that is absolutely continuous with respect to the Lebesgue measure, the critical points of f are isolated. For the first time, this provides a sufficient condition for the critical points of a least-squares formulation of the source localization problem to be isolated and justifies a common assumption in the literature (see, e.g., [3, 17]). It is worth noting that our sufficient condition is very mild, because by the Radon–Nikodym theorem, it essentially requires the measurement noise follows a probability distribution that has a density with respect to the Lebesgue measure. Such a requirement is satisfied by many continuous probability distributions, such as the uniform distribution, Gaussian distribution, and Cauchy distribution. Thus, Proposition 2 can be applied to a variety of noise distributions.

As a matter of fact, when considering the source localization problem on $\mathbb{R}^2$, it can be shown that the critical points of f are isolated, regardless of the probability distribution of the measurement noise. Specifically, we have the following result:

Proposition 3. *For the source localization problem on $\mathbb{R}^2$, the number of critical points of f is at most 5.*

Proof. Using the same notation as in the proof of Proposition 2, it remains to prove the statement for the case where $b_1 = 0$ or $b_2 = 0$. To begin, let us consider the case where $b_1 = b_2 = 0$. Using the definition of $\{b_i\}_{i=1}^2$, it is easy to see that $b_1 = b_2 = 0$ implies $\nabla f(\boldsymbol{c}) = 0$. Suppose there exists a non-zero vector $\boldsymbol{v} \in \mathbb{R}^d$ such that $\nabla f(\boldsymbol{c} + \boldsymbol{v}) = 0$. Writing $\boldsymbol{v} = \sum_{i=1}^2 p_i \boldsymbol{u}_i$, we can, without loss of generality, assume that $p_1 \neq 0$ and $\lambda_1 = -4ms$ using (14). Then, we have $\nabla f(\boldsymbol{c} + t\boldsymbol{u}_1) = 0$ for all $t \in \mathbb{R}$, implying that f is constant along the line $\{\boldsymbol{c} + t\boldsymbol{u}_1 : t \in \mathbb{R}\}$. However, when $m \geq 1$, the function f is coercive; i.e., for any sequence $\{\boldsymbol{x}_i\}_{i \geq 0}$ satisfying $\|\boldsymbol{x}_i\| \to \infty$, we have $f(\boldsymbol{x}_i) \to \infty$. Hence, when $b_1 = b_2 = 0$, there exists only one critical point of f, which is $\boldsymbol{c}$.

Now, let us consider the case where there is exactly one $i \in \{1, 2\}$ such that $b_i = 0$ (in other words, $s > 0$). Without loss of generality, we assume that $b_1 = 0$ and $b_2 \neq 0$. We consider the following cases:

Case 1: $\lambda_1 \geq 0$. Since $s > 0$, we see that $4ms + \lambda_1 \neq 0$. This implies that $p_1 = 0$. Then, by direct substitution into (15), we have

$$s = 0 + \left(-\frac{b_2}{4ms + \lambda_2}\right)^2,$$

which in total yields at most 3 solutions.

Case 2: $\lambda_1 < 0$. From (14), there are two possibilities, namely,

$$4ms + \lambda_1 = 0 \quad \text{or} \quad p_1 = 0.$$

For $p_1 = 0$, following the result in Case 1, there are at most 3 solutions. For $4ms + \lambda_1 = 0$ (i.e., $s = -\frac{\lambda_1}{4m}$), we first observe that $\lambda_1 \neq \lambda_2$. Otherwise, by (14), we have $b_2 = 0$, which contradicts with our assumption. Plugging this into (15) leads to

$$\left(-\frac{\lambda_1}{4m}\right) = p_1^2 + \left(\frac{-b_2}{4m\left(\frac{-\lambda_1}{4m}\right) + \lambda_2}\right)^2,$$

which yields at most 2 solutions. Combining the above two cases with the case where $b_1, b_2 \neq 0$ in Proposition 2, we conclude that there are at most 5 solutions to (15). $\square$

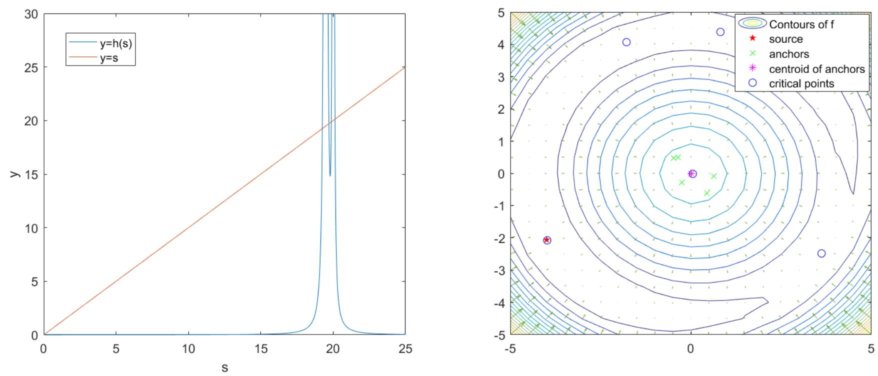
Now, let us show that the bounds on the number of critical points of f given in Propositions 2 and 3 are tight. Specifically, we construct an instance of the source localization problem on $\mathbb{R}^2$ (i.e., $d = 2$) in which the number of critical points is exactly $2d + 1 = 5$.

Example 7. *Consider the source localization problem on $\mathbb{R}^2$. Suppose that the source is located at $\boldsymbol{x}^\star = [-3.9939, -2.0593]^T$ and the five anchors are located at*

$$\boldsymbol{a}_1 = [-0.4818, 0.4816]^T, \quad \boldsymbol{a}_2 = [-0.2438, -0.2857]^T, \quad \boldsymbol{a}_3 = [0.4484, -0.6043]^T,$$
$$\boldsymbol{a}_4 = [0.6287, -0.0933]^T, \quad \boldsymbol{a}_5 = [-0.3515, 0.5016]^T.$$

Suppose that the measurement noise vector is given by

$$\boldsymbol{w} = [-0.0071, 0.0006, -0.0185, -0.0040, -0.0054]^T.$$



(a) Plot of system (16). (b) Contour lines and gradient flow of f.

Figure 5: Computation visualization and geometric landscape of the source localization problem in Example 7.

Then, as shown in Figure 5, there are in total five critical points, namely

$$\boldsymbol{\xi}_1 = [0.0536, -0.0151]^T, \quad \boldsymbol{\xi}_2 = [-3.6174, -2.4827]^T, \quad \boldsymbol{\xi}_3 = [-1.7820, 4.0657]^T,$$
$$\boldsymbol{\xi}_4 = [0.8177, 4.3817]^T, \quad \boldsymbol{\xi}_5 = [-3.9774, -2.0759]^T.$$

As can be verified using the Hessian of f in (3)*, the first one is a local maximum, the second and third are strict saddles, and the last two are local minima. Figure 5a plots the system* (16)*, while Figure 5b shows the contours of f and its gradient-descent vector field.*

The careful reader may notice that neither Proposition 2 nor Proposition 3 applies to the source localization problem on $\mathbb{R}^3$ in which the noise distribution is not absolutely continuous with respect to the Lebesgue measure. Indeed, one can construct an instance of the source localization problem on $\mathbb{R}^3$ under the noiseless setting in which the critical points are not even isolated; i.e., there is an infinite number of critical points.

Example 8. *Suppose that the source is located at $\boldsymbol{x}^\star = [2, 0, 0]^T$ and the eight anchors (i.e., $m = 8$) are located at*

$$\boldsymbol{a}_1 = [0.5, 1, 1]^T, \quad \boldsymbol{a}_2 = [-0.5, 1, 1]^T, \quad \boldsymbol{a}_3 = [0.5, 1, -1]^T, \quad \boldsymbol{a}_4 = [-0.5, 1, -1]^T$$
$$\boldsymbol{a}_5 = [0.5, -1, 1]^T, \quad \boldsymbol{a}_6 = [-0.5, -1, 1], \quad \boldsymbol{a}_7 = [0.5, -1, -1]^T, \quad \boldsymbol{a}_8 = [-0.5, -1, -1].$$

Suppose that the measurements are noiseless; i.e., $w_i = 0$ *for* $i = 1, \ldots, m$. *Then, the gradient and Hessian of* f *are given by*

$$\nabla f(\boldsymbol{c}) = \begin{pmatrix} -32 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(\boldsymbol{c}) = \begin{pmatrix} -112 & 0 & 0 \\ 0 & -64 & 0 \\ 0 & 0 & -64 \end{pmatrix},$$

respectively. Decompose $\nabla^2 f(\boldsymbol{c}) = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ *and write* $\nabla f(\boldsymbol{c}) = \boldsymbol{U}\boldsymbol{b}$. *We then have* $\boldsymbol{U} = \boldsymbol{I}$, $\lambda_1 = -112,\ \lambda_2 = -64,\ \lambda_3 = -64,$ *and* $b_1 = -32,\ b_2 = 0,\ b_3 = 0$. *The system* (14) *implies that*

$$(4ms + \lambda_2)p_2 = 0.$$

Consider the case where $4ms + \lambda_2 = 0$ *(i.e.,* $s = -\frac{-64}{4\times 8} = 2$*). We can compute from* (14) *that*

$$p_1 = \frac{-b_1}{4ms + \lambda_1} = \frac{-32}{4(8)(2) + (-112)} = \frac{2}{3}.$$

Putting this into (13), *we have*

$$2 = \left(\frac{2}{3}\right)^2 + p_2^2 + p_3^2 \quad \Longleftrightarrow \quad p_2^2 + p_3^2 = \frac{14}{9}.$$

This shows that f *possesses a circle of critical points, which are not isolated.*

4. Algorithm

The previous section characterizes the critical points of f in (12) and bounds their maximum number in Propositions 2 and 3. The results suggest that finding all the critical points of f is as easy as finding the roots of a univariate equation. In this section, we present and analyze an algorithm for finding all the critical points and hence also a global minimum of f.

Let us collect the ingredients from the previous section and briefly outline the algorithm here. Recall from (12) that every critical point of f satisfies

$$\nabla f(\boldsymbol{c} + \boldsymbol{v}) = 4m\|\boldsymbol{v}\|^2\boldsymbol{v} + \nabla^2 f(\boldsymbol{c})\boldsymbol{v} + \nabla f(\boldsymbol{c}) = \boldsymbol{0}$$

for some $\boldsymbol{v} \in \mathbb{R}^d$. Performing eigendecomposition on the Hessian $\nabla^2 f(\boldsymbol{c}) = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ and writing $\nabla f(\boldsymbol{c}) = \boldsymbol{U}\boldsymbol{b}$, $\boldsymbol{v} = \boldsymbol{U}\boldsymbol{p}$ for some vectors $\boldsymbol{b} \in \mathbb{R}^d$ and

$\boldsymbol{p} \in \mathbb{R}^d$, the critical points of f can be found by solving the following system for $\{p_i\}_{i=1}^d$ and $s \in \mathbb{R}_+$:

$$\begin{cases} s = \sum_{i=1}^{d} p_i^2; \\ 4msp_i + \lambda_i p_i + b_i = 0 \quad \text{for } i = 1, \ldots, d. \end{cases} \tag{18}$$

Suppose that each measurement noise w_i, for $i = 1, \ldots, m$, follows some absolutely continuous probability distribution with respect to the Lebesgue measure. Then, because of the result in Lemma 2, we can solve for the roots of the equation

$$g(s) := \sum_{i=1}^{d} \left(\frac{b_i}{4ms + \lambda_i} \right)^2 - s = 0 \tag{19}$$

via Newton's method. Specifically, at every step, we perform an update on s by

$$s_{k+1} = s_k - g(s_k)/g'(s_k) \tag{20}$$

with

$$g'(s_k) = \sum_{i=1}^{d} -8mb_i^2(4ms + \lambda_i)^{-3} - 1. \tag{21}$$

Substituting the solutions s into (18) and converting the solution vectors $\boldsymbol{p} \in \mathbb{R}^d$ back to the standard basis of $\mathbb{R}^d$, we obtain all the critical points of f via

$$\boldsymbol{\xi} = \boldsymbol{c} + \boldsymbol{U}\boldsymbol{p}. \tag{22}$$

Eventually, it is a trivial task to determine a global minimum of f by looking for a critical point with the minimum objective value. Algorithm 1 outlines the implementation details of the above discussion.

In the existing literature, several notable works, such as [4] and [2], also solve problem (LS) by finding the roots of certain equations. Yet, as discussed in Section 1.1, our algorithm is the first that utilizes the geometry of the objective function f and is thus more reliable and efficient.

Algorithm 1, composing of a diagonalization step and Newton's method, has a complexity of $\mathcal{O}(d^3 + \log(d)\text{eval}(f))$ for obtaining an ε-approximate solution. It is worth pointing out that the diagonalization of $\nabla^2 f(\boldsymbol{c})$ plays a key role in the algorithm. In particular:

Algorithm 1 Critical Point Finding Algorithm (CPFA)

Input: Anchors' location $\{\boldsymbol{a}_i\}_{i=1}^m$, and range measurements $\{r_i\}_{i=1}^m$.
Output: Global optimum of f.
Compute $\boldsymbol{c}$, $\nabla f(\boldsymbol{c})$ and $\nabla^2 f(\boldsymbol{c})$.
Diagonalize $\nabla^2 f(\boldsymbol{c}) = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, with $\lambda_1 \leq \cdots \leq \lambda_d$.
Compute $\boldsymbol{b} = \boldsymbol{U}^T \nabla f(\boldsymbol{c})$.
Compute all possible solutions s via Algorithm 2.
For each s obtained, compute $p_i = -\frac{b_i}{4ms+\lambda_i}$ for $i = 1, \ldots, d$.
Compute all critical points of f by (22).
Return the critical point with the minimum objective value.

Algorithm 2 Root-Finding Algorithm

Input: γ = number of negative eigenvalues of $\nabla^2 f(\boldsymbol{c})$, diagonalization of $\nabla^2 f(\boldsymbol{c}) = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, $\boldsymbol{b} = \boldsymbol{U}^T \nabla f(\boldsymbol{c})$, number of measurements m, perturbation ε.
Output: $\{s^1, \ldots, s^{2\gamma+1}\}$.
if $\lambda_i \geq 0$, for $i = 1, \ldots, d$, **then**
    $s_0 = \varepsilon$.
    while stopping criterion is not met **do**
        $s_{k+1} = s_k - g(s_k)/g'(s_k)$.
else
    $\lambda_{\gamma+1} \leftarrow 0$.
    for $i = 1, \ldots, \gamma$ **do**
        Initialize $s_0^{2i-1} = -\lambda_{i+1}/(4m) + \varepsilon$ and $s_0^{2i} = -\lambda_i/(4m) - \varepsilon$.
    Initialize $s_0^{2\gamma+1} = -\lambda_1/(4m) + \varepsilon$.
    for $j = 1, \ldots, 2\gamma + 1$ **do**
        while stopping criterion is not met **do**
            $s_{k+1}^j = s_k^j - g(s_k^j)/g'(s_k^j)$.

- *Elucidate the geometry:* The eigenvalues of $\nabla^2 f(\boldsymbol{c})$ reveals (i) whether the objective function f possesses global strong convexity and (ii) the number of critical points of f. The information allows Newton's method to search for roots over certain intervals.

- *Avoid computing matrix inverses repeatedly:* Without the diagonalization of $\nabla^2 f(\boldsymbol{c})$, one has to solve for the roots of

  $$\|(4ms\boldsymbol{I} + \nabla^2 f(\boldsymbol{c}))^{-1}\nabla f(\boldsymbol{c})\|^2 - s = 0$$

  instead of (19). Although the complexities of computing the diagonalization of $\nabla^2 f(\boldsymbol{c})$ and the matrix inverse $(4ms\boldsymbol{I} + \nabla^2 f(\boldsymbol{c}))^{-1}$ are on the same order, this approach would incur extra computational cost since a matrix inverse has to be computed in every Newton iteration for different s, while the diagonalization of $\nabla^2 f(\boldsymbol{c})$ has to be done only once in CPFA.

- *Facilitate localizing different sources given the same set of anchors:* According to the result in Lemma 1, the diagonalization of the Hessian $\nabla^2 f(\boldsymbol{c})$ is solely determined by the geometric configuration of the anchors. Therefore, as long as the locations of the anchors are fixed, the diagonalization of $\nabla^2 f(\boldsymbol{c})$ can be pre-computed for recovering different sources, resulting in a cheaper computational cost of the algorithm.

In Algorithm 1, we need to perform Newton's method over several intervals (see Line 10 in Algorithm 2) to look for all solutions of s and obtain a global minimum by comparing the objective values of all critical points derived from different solutions of s. Therefore, if we have some knowledge about which interval the best s (i.e., the one that corresponds to a global minimum of problem (LS)) is lying, we can focus our search on that interval and thus reduce the computational time of the algorithm. As it turns out, when the noise is sufficiently small, it can be shown that the best s lies on $\left(-\frac{\lambda_{\min}(\nabla^2 f(\boldsymbol{c}))}{4m}, \infty\right)$.

Proposition 4. *Let $\hat{s} := \|\hat{\boldsymbol{x}} - \boldsymbol{c}\|^2$, where $\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$. If the measurement noise w_i is sufficiently small for $i = 1, \ldots, m$, then $\hat{s}$ lies on $\left(-\frac{\lambda_{\min}(\nabla^2 f(\boldsymbol{c}))}{4m}, \infty\right)$. In particular, the global optimum of f is unique.*

The proof of the proposition relies on the following bound on the estimation error of $\hat{\boldsymbol{x}}$, which is defined as the distance between $\hat{\boldsymbol{x}}$ and the true source location $\boldsymbol{x}^\star$ and will be proved in Section 5.

Fact 1. *The estimation error of $\hat{\boldsymbol{x}}$ satisfies*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\| \leq 4JK\sqrt{m}\|\boldsymbol{w}\| + 2K\|\boldsymbol{w}\|^2$$

for some constants J and K defined in (28) *and* (29), *respectively.*

Proof. It remains to show when

$$\hat{s} = \|\hat{\boldsymbol{x}} - \boldsymbol{c}\|^2 > -\frac{\lambda_{\min}(\nabla^2 f(\boldsymbol{c}))}{4m} \tag{23}$$

holds. Note that $\hat{s}$ can be upper bounded by

$$\begin{aligned}\hat{s} = \|\hat{\boldsymbol{x}} - \boldsymbol{c}\|^2 &= \|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\|^2 + 2\langle \hat{\boldsymbol{x}} - \boldsymbol{x}^\star, \boldsymbol{x}^\star - \boldsymbol{c}\rangle + \|\boldsymbol{x}^\star - \boldsymbol{c}\|^2 \\ &\geq -2\|\boldsymbol{x}^\star - \boldsymbol{c}\|(4JK\sqrt{m}\|\boldsymbol{w}\| + 2K\|\boldsymbol{w}\|^2) + \|\boldsymbol{x}^\star - \boldsymbol{c}\|^2. \end{aligned} \tag{24}$$

Also, applying the results in (8) and (1), we obtain

$$\begin{aligned} &-\frac{\lambda_{\min}(\nabla^2 f(\boldsymbol{c}))}{4m} \\ &\overset{(8)}{=} R^2 - \frac{2}{m}\lambda_{\min}\left(\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T\right) \\ &\overset{(1)}{=} \|\boldsymbol{x}^\star - \boldsymbol{c}\|^2 + \frac{1}{m}\sum_{i=1}^m (2w_i\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i^2) - \frac{2}{m}\lambda_{\min}\left(\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T\right). \end{aligned} \tag{25}$$

Combining these two inequalities then yields a sufficient condition for (23) to hold:

$$\begin{aligned} &-2\|\boldsymbol{x}^\star - \boldsymbol{c}\|(4JK\sqrt{m}\|\boldsymbol{w}\| + 2K\|\boldsymbol{w}\|^2) - \frac{1}{m}\sum_{i=1}^m (2w_i\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i^2) \\ &> -\frac{\lambda_{\min}\left(\sum_{i=1}^m (\boldsymbol{a}_i - \boldsymbol{c})(\boldsymbol{a}_i - \boldsymbol{c})^T\right)}{4m}. \end{aligned} \tag{26}$$

Under the assumption that $\{\boldsymbol{a}_i - \boldsymbol{c}\}_{i=1}^m$ spans $\mathbb{R}^d$, the rhs of (26) is strictly negative. Therefore, when the measurement noise is sufficiently small, condition (26) could be satisfied. Applying the argument in the proof of Proposition 2, we can then see that the global optimum of (LS) is unique. □

Therefore, if the measurement noise is sufficiently small, then it suffices to perform Newton's method on the interval $\left(-\frac{\lambda_{\min}(\nabla^2 f(\boldsymbol{c}))}{4m}, \infty\right)$, regardless of the number of critical points of f. Moreover, the function f is known to be monotonically decreasing on the interval, which also reduces the cost of searching for the root. Besides, Proposition 4 provides a sufficient condition for the uniqueness of the global optimum of the least-squares formulation (LS) to hold without the strong convexity of the objective function f, which is new in the literature; cf. [10, 18].

As a closing remark for this subsection, without any assumption on the measurement noise, a similar algorithm can be developed for source localization problem on $\mathbb{R}^2$ based on the proof of Proposition 3. For the sake of brevity, we omit the algorithm here.

5. Estimation Performance

In the previous sections, we study the geometric properties of problem (LS) and their algorithmic consequences. In this section, we shift our focus to the estimation quality of a global optimum $\hat{\boldsymbol{x}}$ of problem (LS) (also referred to as a least-squares estimator). In particular, we are interested in the estimation error of $\hat{\boldsymbol{x}}$, which is defined as the distance between $\hat{\boldsymbol{x}}$ and the true source location $\boldsymbol{x}^\star$. Intuitively, when the measurement noise is small, the estimation error of $\hat{\boldsymbol{x}}$ should also be small. The following result formalizes this intuition.

Theorem 2. *Let $\hat{\boldsymbol{x}}$ be a global optimum of problem* (LS). *Then, the estimation error of $\hat{\boldsymbol{x}}$ satisfies*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\| \leq 4JK\sqrt{m}\|\boldsymbol{w}\| + 2K\|\boldsymbol{w}\|^2, \tag{27}$$

where

$$J = \max_{i \in \{1,\ldots,m\}} (\|\boldsymbol{x}^\star - \boldsymbol{a}_i\|) \quad \text{and} \tag{28}$$

$$K = \left(\lambda_{\min}\left(\sum_{i=1}^{m} (\boldsymbol{a}_{i+1} - \boldsymbol{a}_i)(\boldsymbol{a}_{i+1} - \boldsymbol{a}_i)^T\right)\right)^{-1/2}. \tag{29}$$

Before we proceed, several comments are in order. First, although estimation error bounds for other localization problems have been established in [10, 18], the result in Theorem 2 is new for (LS). Second, Theorem 2 does not impose any assumption on the distribution of the measurement noise, so one can apply results from probability to obtain estimation error bounds for different noise distributions. For instance, when the measurement noise $\boldsymbol{w}$ follows a Gaussian distribution with mean zero and variance $\sigma^2 \boldsymbol{I}$, then, with probability at least $1 - (2\exp(-\mu^2/2))^m$ for some constant $\mu > 0$, we have

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\| \leq 4\mu J K m \sigma + 2\mu^2 K m \sigma^2 \tag{30}$$

with J and K defined in (28) and (29), respectively. As a result, the estimation error is controlled by the standard deviation σ of the noise with high probability. Third, unlike existing estimation error bounds in the localization literature [10, 18], we make explicit how the constants in the bound depend on the locations of the anchors and the number of measurements.

The proof of Theorem 2 utilizes the fact that the solution of (LS) can be approximated by a so-called ordinary least-squares estimator, which is computable in closed-form; see, e.g., [19]. Specifically, each measurement r_i roughly measures the distance between the source and the i-th anchor; i.e.

$$\|\boldsymbol{x}^\star - \boldsymbol{a}_i\|^2 \approx r_i^2, \quad i = 1, \ldots, m.$$

Therefore, by subtracting the i-th equation from the $(i+1)$-st equation, we obtain the system of linear equations

$$2(\boldsymbol{a}_{i+1} - \boldsymbol{a}_i)^T \boldsymbol{x}^\star \approx \|\boldsymbol{a}_{i+1}\|^2 - \|\boldsymbol{a}_i\|^2 + r_i^2 - r_{i+1}^2, \quad i = 1, \ldots, m.$$

The above suggests that the source location $\boldsymbol{x}^\star$ can be estimated by the solution of the following minimization problem:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \|\boldsymbol{B}\boldsymbol{x} - \boldsymbol{g}\|^2. \tag{31}$$

Here, we have

$$\boldsymbol{B} := \begin{pmatrix} (\boldsymbol{a}_2 - \boldsymbol{a}_1)^T \\ \vdots \\ (\boldsymbol{a}_m - \boldsymbol{a}_{m-1})^T \end{pmatrix} \quad \text{and} \quad \boldsymbol{g} := \frac{1}{2} \begin{pmatrix} \|\boldsymbol{a}_2\|^2 - \|\boldsymbol{a}_1\|^2 + r_1^2 - r_2^2 \\ \vdots \\ \|\boldsymbol{a}_m\|^2 - \|\boldsymbol{a}_{m-1}\|^2 + r_{m-1}^2 - r_m^2 \end{pmatrix}. \tag{32}$$

The assumption that $\{\boldsymbol{a}_i - \boldsymbol{c}\}_{i=1}^m$ spans $\mathbb{R}^d$ implies that $\{\boldsymbol{a}_{i+1} - \boldsymbol{a}_i\}_{i=1}^{m-1}$ spans $\mathbb{R}^d$. Hence, the solution of (31), which we call an ordinary least-squares estimator, is given by

$$\boldsymbol{x}_{\mathrm{OLS}} = (\boldsymbol{B}^T \boldsymbol{B})^{-1} \boldsymbol{B}^T \boldsymbol{g}. \tag{33}$$

Proof of Theorem 2. Let $\hat{r}_i = \|\hat{\boldsymbol{x}} - \boldsymbol{a}_i\|$ be the distance between an optimal solution of (LS) and the i-th anchor for $i = 1, \ldots, m$, and $\hat{\boldsymbol{g}}$ be defined similarly as $\boldsymbol{g}$ in (32) but with $\{r_i\}_{i=1}^m$ replaced by $\{\hat{r}_i\}_{i=1}^m$. It is easy to see that

$$\|\boldsymbol{x}_{\mathrm{OLS}} - \hat{\boldsymbol{x}}\| \leq \|(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\|_{\mathrm{op}}\|\boldsymbol{g} - \hat{\boldsymbol{g}}\|. \tag{34}$$

On one hand, we see that

$$\begin{aligned}
\|(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\|_{\mathrm{op}} &= \sqrt{\lambda_{\max}((\boldsymbol{B}^T\boldsymbol{B})^{-1})} \\
&= \left(\lambda_{\min}\left(\sum_{i=1}^m (\boldsymbol{a}_{i+1} - \boldsymbol{a}_i)(\boldsymbol{a}_{i+1} - \boldsymbol{a}_i)^T\right)\right)^{-1/2} = K;
\end{aligned} \tag{35}$$

on the other hand, defining $\boldsymbol{M} \coloneqq \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(m-1)\times m}$ and using the optimality of $\hat{\boldsymbol{x}}$, we have

$$\|\boldsymbol{g} - \hat{\boldsymbol{g}}\| \leq \frac{1}{2}\|\boldsymbol{M}\|_{\mathrm{op}}\sqrt{f(\hat{\boldsymbol{x}})} \leq \frac{1}{2}\|\boldsymbol{M}\|_{\mathrm{op}}\sqrt{f(\boldsymbol{x}^\star)}. \tag{36}$$

Moreover, the result in [9, Theorem 2.2] implies that

$$\|\boldsymbol{M}\|_{\mathrm{op}} = \sqrt{\lambda_{\max}(\boldsymbol{M}\boldsymbol{M}^T)} \leq 2. \tag{37}$$

Now, upon writing $r_i^\star = \|\boldsymbol{x}^\star - \boldsymbol{a}_i\|$ as the distance between the true source location and the i-th anchor for $i = 1, \ldots, m$ and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\sqrt{f(\boldsymbol{x}^\star)} &= \sqrt{\sum_{i=1}^m (r_i^\star - r_i^2)^2} \leq \sum_{i=1}^m \left(2r_i^\star|w_i| + w_i^2\right) \\
&\leq 2\max_{i\in\{1,\ldots,m\}}(r_i^\star)\sqrt{m}\|\boldsymbol{w}\| + \|\boldsymbol{w}\|^2.
\end{aligned} \tag{38}$$

Hence,

$$\|\boldsymbol{x}_{\mathrm{OLS}} - \hat{\boldsymbol{x}}\| \leq 2K\max_{i\in\{1,\ldots,m\}}(r_i^\star)\sqrt{m}\|\boldsymbol{w}\| + K\|\boldsymbol{w}\|^2. \tag{39}$$

Besides, by defining $\boldsymbol{g}^\star$ similarly as $\boldsymbol{g}$ but with $\{r_i\}_{i=1}^m$ replaced by $\{r_i^\star\}_{i=1}^m$, a similar computation shows that

$$\begin{aligned}\|\boldsymbol{x}_{\mathrm{OLS}} - \boldsymbol{x}^\star\| \leq & \|(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\|_{\mathrm{op}}\|\boldsymbol{g} - \boldsymbol{g}^\star\| \\ \leq & \|(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T\|_{\mathrm{op}} \cdot \frac{1}{2}\|\boldsymbol{M}\|_{\mathrm{op}}\sqrt{f(\boldsymbol{x}^\star)} \\ \leq & 2K \max_{i\in\{1,\ldots,m\}}(r_i^\star)\sqrt{m}\|\boldsymbol{w}\| + K\|\boldsymbol{w}\|^2. \end{aligned} \tag{40}$$

Combining (39) and (40), the desired result follows from triangle inequality. □

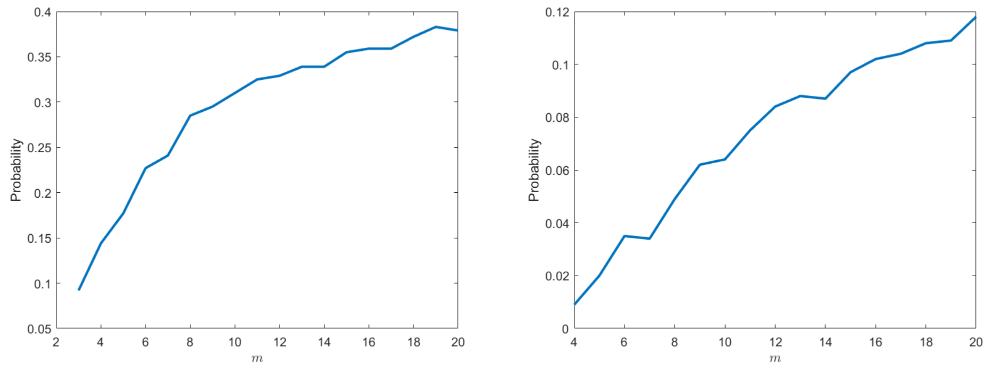
6. Numerical Simulation

In this section, we conduct numerical experiments to demonstrate our theoretical findings and examine the performance of our proposed algorithm CPFA. All numerical experiments were conducted on a PC running Windows 10 with an Intel® Core™ i7-7700 3.60GHz CPU and 16GB memory.

6.1. Global Strong Convexity

In this subsection, we examine how likely problem (LS) is globally strongly convex when a source and m anchors are generated randomly from a uniform distribution. Specifically, consider the source localization problem on $\mathbb{R}^2$, where we randomly generate m anchors (for $m = 3, \ldots, 20$) and a source on $[-5, 5] \times [-5, 5]$. Each measurement is contaminated by a noise that follows a Gaussian distribution with mean zero and variance $\sigma^2 = 0.01$. We conduct 1000 Monte Carlo simulations and evaluate the empirical probability that condition (2) (or equivalently, global strong convexity) is satisfied. Figure 6a plots the empirical probability that problem (LS) is globally strongly convex versus the number of anchors m from $m = 3$ to $m = 20$. As can be seen, the empirical probability increases from 0.092 to 0.379 when the number of anchors increases from $m = 3$ to $m = 20$.

Similarly, consider the source localization problem on $\mathbb{R}^3$, where we randomly generate m anchors (for $m = 4, \ldots, 20$) and a source on $[-5, 5] \times [-5, 5] \times [-5, 5]$. The measurement noise again follows a Gaussian distribution with mean zero and variance $\sigma^2 = 0.01$. Figure 6b plots the empirical probability that problem (LS) is globally strongly convex versus the number of anchors m from $m = 4$ to $m = 20$. The empirical probability increases from 0.009 to 0.118 when the number of anchors increases from $m = 4$ to



(a) On $\mathbb{R}^2$.

(b) On $\mathbb{R}^3$.

Figure 6: Empirical probability that problem (LS) is globally strongly convex.

$m = 20$. It can be seen that with the same number of anchors, the empirical probability that problem (LS) is globally strongly convex is generally smaller as the dimension of the source localization problem increases.

6.2. Performance of CPFA

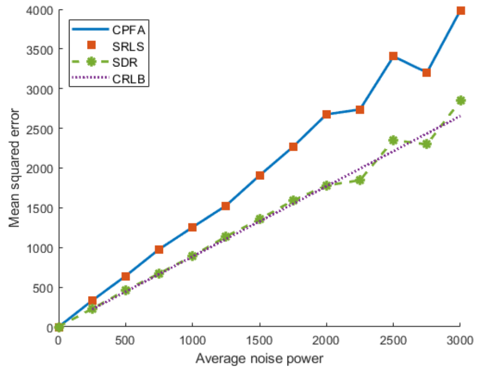
6.2.1. Particular Configuration

In this subsection, we consider the source localization problem on $\mathbb{R}^2$, where the anchors are set at [4, Section 5]

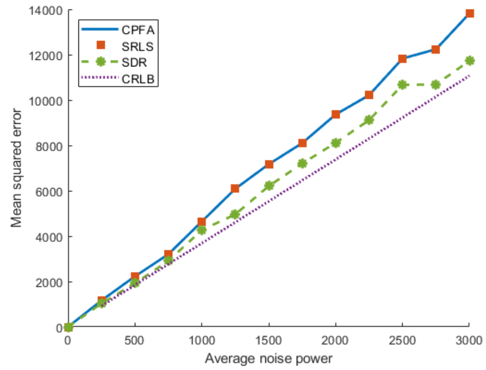
$$\boldsymbol{a}_1 = (0,0)^T, \boldsymbol{a}_2 = (3000\sqrt{3}, 3000)^T, \boldsymbol{a}_3 = (0, 6000)^T,$$
$$\boldsymbol{a}_4 = (-3000\sqrt{3}, 3000)^T, \boldsymbol{a}_5 = (-3000\sqrt{3}, -3000)^T$$

and the source is located either at $\boldsymbol{x}^\star = (1000, 2000)^T$ (which is within the convex hull of the anchors) or at $\boldsymbol{x}^\star = (-8000, -6000)^T$ (which is outside the convex hull of the anchors). Each measurement is independently contaminated by a zero-mean Gaussian noise with average power ranging from 0 to 3000 at an interval of 250. For each of the source locations, we run 1000 Monte Carlo simulations and take the average.

Figures 7a and 7b plot the mean squared error versus average noise power from 0 to 3000 when the source lies at $\boldsymbol{x}^\star = (1000, 2000)^T$ and $\boldsymbol{x}^\star = (-8000, -6000)^T$, respectively. We compare the performance of CPFA with that of SRLS [2] and SDR [5], as well as with the Cramér-Rao lower bound (CRLB). The perturbation parameter of Newton's method in CPFA is set to be $\varepsilon = 10^{-4}$. As can be seen in Figure 7a, when the source is located



(a) $\boldsymbol{x}^\star = (1000, 2000)^T$.



(b) $\boldsymbol{x}^\star = (-8000, -6000)^T$.

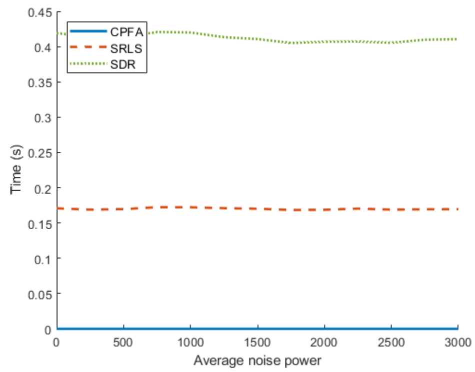
Figure 7: Mean squared error versus average noise power.

at $\boldsymbol{x}^\star = (1000, 2000)^T$, CPFA and SRLS obtain the same mean squared error under different noise levels. SDR performs better than both methods with its mean squared error close to CRLB, particularly when the average noise power is large. Figure 7b shows a similar result when the source is located at $\boldsymbol{x}^\star = (-8000, -6000)^T$. The mean squared errors of CPFA and SRLS overlap, while SDR performs slightly better than both methods. This is because the SDR solution is an approximation of the maximum likelihood estimator, while the CPFA and SRLS solutions are not.

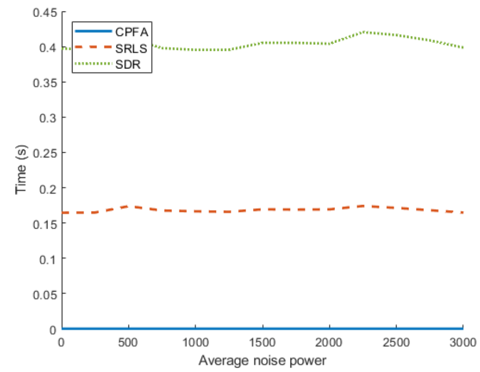
Next, Figures 8a and 8b plot the computational time of CPFA, SRLS, and SDR under different noise levels when the source lies at $\boldsymbol{x}^\star = (1000, 2000)^T$ and $\boldsymbol{x}^\star = (-8000, -6000)^T$, respectively. As can be seen in Figures 8a and 8b, the methods have similar computational time with different source locations and under different noise levels. In both configurations, CPFA is the fastest, while SDR is the slowest. This demonstrates the efficiency of CPFA compared with other state-of-the-art localization algorithms.

6.2.2. Random Configurations

In this example, we consider two scenarios on $\mathbb{R}^2$ with five anchors: (i) the anchors and the source are randomly distributed on $[-10, 10] \times [-10, 10]$; (ii) the anchors are randomly distributed on $[-5, 5] \times [-5, 5]$ while the source is randomly distributed on $[-10, 10] \times [-10, 10]$; cf. [2, Example 2]. For both scenarios, we evaluate the performance of CPFA, SRLS, and SDR under different noise levels, namely, $\sigma = 0.01$, $\sigma = 0.1$, and $\sigma = 1$. The perturbation parameter of Newton's method in CPFA is set to be $\varepsilon = 10^{-4}$. We run 1000



(a) $\boldsymbol{x}^{\star} = (1000, 2000)^T$.



(b) $\boldsymbol{x}^{\star} = (-8000, -6000)^T$.

Figure 8: Time (s) versus average noise power.

Monte Carlo simulations and record the mean squared errors in scenarios (i) and (ii) in Tables 1a and 1b, respectively. We see from Table 1a that CPFA and SRLS always obtain the same mean squared error over 1000 Monte Carlo simulations regardless of measurement noise levels. Moreover, both methods perform better than SDR when the noise power is small (i.e., when $\sigma = 0.01$ and $\sigma = 0.1$) but perform worse than SDR when the noise power is large (i.e., when $\sigma = 1$). Similar phenomenon can be observed in scenario (ii). We see from Table 1b that, when $\sigma = 0.01$, CPFA, SRLS, and SDR obtain the same mean squared error; when $\sigma = 0.1$, CPFA and SRLS perform equally good and outperform SDR; when $\sigma = 1$, SDR outperforms CPFA and SRLS. It is expected that when the geometric configuration of the source and anchors is randomly generated, the SDR solution might not be tight enough, resulting in a poor estimate of the source location. However, since problem (LS) is suboptimal in the maximum likelihood sense, when the noise level is large, CPFA and SRLS cannot estimate the source location well. Therefore, SDR outperforms both methods in this case.

Tables 2a and 2b show the average computational time of CPFA, SRLS, and SDR in scenarios (i) and (ii), respectively. As can be seen in both tables, CPFA has the shortest computational time among the three algorithms. SRLS is about 4-10 times slower than CPFA, while SDR is about 700-4000 times slower than CPFA. This demonstrates that CPFA is a fast algorithm, regardless of the noise level and the geometric configuration of the source and anchors.

σ	CPFA	SRLS	SDR
0.01	**0.0002**	**0.0002**	0.0004
0.1	**0.0238**	**0.0238**	0.0322
1	3.0414	3.0414	**2.9246**

(a) Source and anchors are randomly distributed on $[-10, 10] \times [-10, 10]$.

σ	CPFA	SRLS	SDR
0.01	**0.0002**	**0.0002**	**0.0002**
0.1	**0.0162**	**0.0162**	0.0189
1	2.0769	2.0768	**1.6825**

(b) Source and anchors are randomly distributed on $[-10, 10] \times [-10, 10]$ and $[-5, 5] \times [-5, 5]$, respectively.

Table 1: Mean squared errors of CPFA, SRLS and SDR under different noise levels.

σ	CPFA	SRLS	SDR
0.01	**0.0005**	0.0050	0.3840
0.1	**0.0013**	0.0050	0.3747
1	**0.0007**	0.0049	0.3729

(a) Source and anchors are randomly distributed on $[-10, 10] \times [-10, 10]$.

σ	CPFA	SRLS	SDR
0.01	**0.0004**	0.0049	0.3817
0.1	**0.0007**	0.0050	0.3797
1	**0.0001**	0.0054	0.4073

(b) Source and anchors are randomly distributed on $[-10, 10] \times [-10, 10]$ and $[-5, 5] \times [-5, 5]$, respectively.

Table 2: Computational time of CPFA, SRLS and SDR under different noise levels.

7. Conclusion

In this work, we established the global strong convexity of the squared-range least-squares formulation of the time-of-arrival-based source localization problem under certain assumption on the configuration of the anchors and the source and on the measurement noise. We also presented several instances that satisfy this assumption when the measurement noise is negligible. Next, we characterized the critical points of the least-squares formulation via the gradient and Hessian of the objective function at the centroid of anchors. As a result, we obtained a finite upper bound on the maximum number of critical points (implying that the critical points are isolated) under a very mild assumption on the measurement noise. The characterization also leads to an algorithm that solves the least-squares formulation globally by searching through all critical points. Moreover, we established an upper bound on the estimation error of the least-squares estimator. Our numerical results support the theoretical findings and show that our proposed algorithm can consistently obtain a global minimum of the least-squares formulation regardless of the geometric configuration of the source and the sensors. An interesting future direction is to see whether a similar analysis can be performed on other localization problems (such as the sensor network localization problem; see, e.g., [22]) or structured sum-of-squares problems.

Acknowledgement

We would like to express our thanks to the anonymous reviewers for their helpful and insightful suggestions on the paper.

References

[1] Madeline Balaam, Rob Comber, Ed Jenkins, Selina Sutton, and Andrew Garbett. Feedfinder: A location-mapping mobile application for breast-feeding women. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1709–1718, 2015.

[2] Amir Beck, Petre Stoica, and Jian Li. Exact and approximate solutions of source localization problems. *IEEE Transactions on Signal Processing*, 56(5):1770–1778, 2008.

[3] Amir Beck, Marc Teboulle, and Zahar Chikishev. Iterative minimization schemes for solving the single source localization problem. *SIAM Journal on Optimization*, 19(3):1397–1416, 2008.

[4] Ka Wai Cheung, Hing-Cheung So, Wing-Kin Ma, and Yiu-Tong Chan. Least squares algorithms for time-of-arrival-based mobile location. *IEEE Transactions on Signal Processing*, 52(4):1121–1130, 2004.

[5] K.W. Cheung, W.K. Ma, and H.C. So. Accurate approximation algorithm for TOA-based maximum likelihood mobile location using semidefinite programming. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages ii–145, 2004.

[6] Maria Luisa Damiani. Location privacy models in mobile applications: Conceptual view and research directions. *GeoInformatica*, 18:819–842, 2014.

[7] Javier JM Diaz, Rodrigo de A Maues, Rodrigo B Soares, Eduardo F Nakamura, and Carlos MS Figueiredo. Bluepass: An indoor bluetooth-based localization system for mobile applications. In *The IEEE symposium on Computers and Communications*, pages 778–783. IEEE, 2010.

[8] Charles Fortin and Henry Wolkowicz. The trust region subproblem and semidefinite programming. *Optimization Methods and Software*, 19(1):41–67, 2004.

[9] Devadatta Kulkarni, Darrell Schmidt, and Sze-Kai Tsui. Eigenvalues of tridiagonal pseudo-Toeplitz matrices. *Linear Algebra and its Applications*, 297:63–80, 1999.

[10] Huikang Liu, Yuen-Man Pun, and Anthony Man-Cho So. Local strong convexity of maximum-likelihood TDOA-based source localization and its algorithmic implications. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5. IEEE, 2017.

[11] Balmukund Mishra, Deepak Garg, Pratik Narang, and Vipul Mishra. Drone-surveillance for search and rescue in natural disaster. *Computer Communications*, 156:1–10, 2020.

[12] Jorge J Moré. Generalizations of the trust region problem. *Optimization Methods and Software*, 2(3-4):189–209, 1993.

[13] Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Boston, 2004.

[14] Jiawang Nie. Sum of squares method for sensor network localization. *Computational Optimization and Applications*, 43(2):151–179, 2009.

[15] Masashi Okamoto. Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Annals of Statistics*, 1(4):763–765, 1973.

[16] Christopher E O'Lone and R Michael Buehrer. An analysis of the convex hull's impact on localization performance. In *Proceedings of IEEE/ION PLANS 2016*, pages 519–526, 2016.

[17] Michael Pauley and Jonathan H Manton. Global optimisation for time of arrival-based localisation. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, pages 791–795. IEEE, 2018.

[18] Yuen-Man Pun and Anthony Man-Cho So. Local strong convexity of source localization and error bound for target tracking under time-of-arrival measurements. *IEEE Transactions on Signal Processing*, 70:190–201, 2022.

[19] Ali H Sayed, Alireza Tarighat, and Nima Khajehnouri. Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Processing Magazine*, 22(4):24–40, 2005.

[20] Iman Shames, Brian DO Anderson, and Barış Fidan. On the use of convex optimization in sensor network localization and synchronization. *IFAC Proceedings Volumes*, 42(20):228–233, 2009.

[21] Rathin Chandra Shit, Suraj Sharma, Deepak Puthal, Philip James, Biswajeet Pradhan, Aad van Moorsel, Albert Y. Zomaya, and Rajiv Ranjan. Ubiquitous localization (ubiloc): A survey and taxonomy on device free localization for smart world. *IEEE Communications Surveys & Tutorials*, 21(4):3532–3564, 2019.

[22] Anthony Man-Cho So and Yinyu Ye. Theory of semidefinite programming for sensor network localization. *Mathematical Programming, Series B*, 109(2):367–384, 2007.

[23] Alex L Wang and Fatma Kılınç-Karzan. The generalized trust region subproblem: Solution complexity and convex hull results. *Mathematical Programming*, 191(2):445–486, 2022.

[24] Gang Wang, Anthony Man-Cho So, and Youmin Li. Robust convex approximation methods for TDOA-based localization under NLOS conditions. *IEEE Transactions on Signal Processing*, 64(13):3281–3296, 2016.