Local Strong Convexity of Source Localization and Error Bound for Target Tracking under Time-of-Arrival Measurements

Yuen-Man Pun and Anthony Man-Cho So, *Senior Member, IEEE*

Abstract—In this paper, we consider a time-varying optimization approach to the problem of tracking a moving target using noisy time-of-arrival (TOA) measurements. Specifically, we formulate the problem as that of sequential TOA-based source localization and apply online gradient descent (OGD) to it to generate the position estimates of the target. To analyze the tracking performance of OGD, we first revisit the classic least-squares formulation of the (static) TOA-based source localization problem and elucidate its estimation and geometric properties. In particular, under standard assumptions on the TOA measurement model, we establish a bound on the distance between an optimal solution to the least-squares formulation and the true target position. Using this bound, we show that the loss function in the formulation, albeit non-convex in general, is locally strongly convex at its global minima. To the best of our knowledge, these results are new and can be of independent interest. By combining them with existing techniques from online strongly convex optimization, we then establish the first non-trivial bound on the cumulative target tracking error of OGD. Our numerical results corroborate the theoretical findings and show that OGD can effectively track the target at different noise levels.

Index Terms—source localization, target tracking, time of arrival (TOA), online gradient descent, tracking error bound

I. INTRODUCTION

Target tracking [2], [3] is a key enabling technology in many applications of multi-agent systems and wireless sensor networks, such as motion planning [4], [5] and surveillance [6], [7]. In one of its basic forms, the tracking problem aims to maintain position estimates of a moving, signal-emitting target over time using noisy measurements of the emitted signal collected by stationary sensors. Such a sequential localization formulation has been extensively studied in the control and signal processing communities, and various approaches for tackling it have been proposed. When a model on the target dynamics and noise statistics is available, a classic approach is to employ Kalman filtering techniques to perform the tracking; see, e.g., [8]–[11] and the references therein. In recent years, however, there have been increasing efforts in developing tracking techniques that require only minimal assumptions on the target trajectory and/or noise distribution. One approach is to view the sequential localization formulation through the lens of time-varying optimization [12], [13]. Specifically, at each time step, the position estimate of the target is given by a minimizer of a loss function that depends on the noisy signal measurements collected at that time step. However, since the time interval between successive measurements is often very short and the sensors have limited computational power, it is impractical to solve the loss minimization problem at each time step exactly. This motivates the use of online optimization techniques to tackle the target tracking problem. To evaluate the performance of an online method, various metrics are available; see [13]. These metrics differ in how they measure the discrepancy between the solutions generated by the method at different time steps and the optimal solutions at the corresponding time steps. When the loss function is convex at every time step, it has been shown that many online methods enjoy strong performance guarantees under different metrics; see, e.g., [14]–[20] and the references therein.

Although the results just mentioned cover a wide variety of target tracking scenarios, they do not apply to those where the loss function of interest is *non-convex*. One such scenario is *time-of-arrival (TOA)-based tracking*, in which sensors collect TOA measurements of the target signal and the tracking is achieved by minimizing a non-convex least-squares loss function associated with the measurements collected at each time step [21]–[23]. In this scenario, the tracking problem can be viewed as a sequential version of the well-studied TOA-based source localization problem; see, e.g., [24]–[30]. As far as we know, the TOA-based tracking problem has barely been investigated from the time-varying or online optimization perspective in the literature. Recently, there have been some works that study time-varying optimization problems with general non-convex loss functions. However, the results are not entirely satisfactory when specialized to the TOA-based tracking problem. For instance, the work [23] proposes an online Newton's method (ONM) and establishes a bound on its *dynamic regret* (i.e., the difference between the cumulative loss incurred by the sequence of solutions generated by the method and that incurred by the sequence of optimal solutions; see [13]) by assuming, among other things, that the Hessian of the loss function at each time step satisfies a non-degeneracy condition. It also demonstrates the numerical performance of ONM on the TOA-based tracking problem. Nevertheless, since ONM needs to compute the inverse of the Hessian of the loss function at each time step, it can be computationally expensive. In addition, the work does not shed any light on whether the TOA-based tracking problem satisfies the assumptions

underlying the dynamic regret analysis of ONM. As such, the theoretical performance of ONM for TOA-based tracking remains unclear. On the other hand, the work [31] develops a dual averaging method and obtains a bound on its dynamic regret under relatively mild assumptions on the loss functions. However, the method is mainly of theoretical interest, as it needs to compute a distribution over the feasible solutions and sample a solution from this distribution at each time step, and neither of these is straightforward to implement for the TOA-based tracking problem.

Motivated by the above discussion, we are interested in developing a low-complexity online method for TOA-based tracking and establishing theoretical guarantee on its performance. One method that naturally suggests itself is online gradient descent (OGD). The method only needs to perform a single gradient descent update at each time step, thus making it well-suited for the target tracking task. However, there has been no performance analysis of OGD for our problem setting so far. Not surprisingly, a major difficulty is that the least-squares loss function associated with the TOA measurements is non-convex. The main contribution of this work is the development of the first non-trivial performance bound for OGD when it is applied to the TOA-based tracking problem. The performance metric we adopt is the *cumulative target tracking error* (CTTE), which is defined as the sum of the distances between the estimated target position and the *true target position* at different time steps. Our bound makes explicit the dependence of the CTTE of OGD on the path length of the target trajectory and the noise power of the TOA measurements. It is important to note that there is a subtle yet fundamental difference in nature between the CTTE metric and most other metrics used in the time-varying or online optimization literature. The former measures the performance relative to the *true values of the parameter* we wish to estimate (viz. the true positions of the target at different time steps), while the latter (such as the dynamic regret or the usual tracking error) measure the performance relative to the *optimal solutions* to the loss minimization problems at different time steps. In the context of the TOA-based tracking problem, it is clear that the CTTE defined above is a more relevant performance metric, as ultimately we are interested in how well the online method tracks the true target positions rather than the optimal solutions to the time-varying loss minimization problem. Nevertheless, the use of the true target positions in the definition of CTTE makes it a more challenging metric to analyze.

To establish the said CTTE bound, we proceed in two steps. First, we revisit the classic least-squares formulation of the (static) TOA-based source localization problem and elucidate its estimation and geometric properties. Specifically, under standard assumptions on the TOA measurement model, we establish a bound on the estimation error of any least-squares estimate of the true target position and use this bound to show that the loss function, albeit non-convex in general, is locally strongly convex at its global minima. Moreover, we give an explicit estimate of the size of the strong convexity region. We remark that similar results have previously been established for a *time-difference-of-arrival* (TDOA)-based least-squares

loss function [32]. However, to the best of our knowledge, our results for the TOA-based least-squares loss function are new and can be of independent interest. In particular, it provides further theoretical justification for the good empirical performance of gradient-based schemes observed in [27] when solving the TOA-based source localization problem. Second, we extend our local strong convexity result from the static localization setting to the dynamic target tracking setting. Specifically, we show that as long as the aforementioned assumptions on the TOA measurement model are satisfied and the distance between the true positions of the target at consecutive time steps is sufficiently small, the position estimate of the target at the current time step will lie in the strong convexity region of the loss function at the next time step. This allows us to utilize techniques from online strongly convex optimization to establish the advertised CTTE bound for OGD.

The notation in this paper is mostly standard. We use $\| \cdot \|_1$ and $\| \cdot \|$ to denote the ℓ_1-norm and Euclidean norm, respectively. Given a vector $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ and a scalar $r > 0$, we use $B(\bar{\boldsymbol{x}}, r) := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \bar{\boldsymbol{x}}\| \leq r\}$ to denote the closed Euclidean ball with center $\bar{\boldsymbol{x}}$ and radius r. Given a symmetric matrix $\boldsymbol{A}$, we use $\lambda_{\min}(\boldsymbol{A})$ to denote its smallest eigenvalue and $\boldsymbol{A} \succ \boldsymbol{0}$ to indicate that it is positive definite.

The rest of the paper is organized as follows. In Section II, we present a time-varying optimization formulation of the TOA-based tracking problem and describe how it can be tackled by OGD. In Section III, we study the estimation error and local strong convexity property of the static TOA-based source localization problem. Using these results, we establish our bound on the CTTE of OGD for the TOA-based tracking problem in Section IV. In Section V, we present numerical results to demonstrate the efficacy of OGD for TOA-based tracking and illustrate our theoretical findings. We then end with some closing remarks in Section VI.

II. PROBLEM FORMULATION AND PRELIMINARIES

We begin by describing the setup for TOA-based tracking. Let $\boldsymbol{x}_t^\star \in \mathbb{R}^n$ be the unknown true position of the moving target at time t, where $t = 1, \ldots, T$ and T is the time horizon of interest. Furthermore, let $\boldsymbol{a}_i \in \mathbb{R}^n$ $(i = 1, \ldots, m)$ be the known position of the ith sensor and suppose that the vectors $\{\boldsymbol{a}_i - \boldsymbol{a}_1\}_{i=2}^m$ span $\mathbb{R}^n$ (in particular, we have $m \geq n+1$). We consider the following model for TOA-based range measurements:

$$r_i^t = \|\boldsymbol{x}_t^\star - \boldsymbol{a}_i\| + w_i^t, \quad i = 1, \ldots, m; \, t = 1, \ldots, T. \quad (1)$$

Here, w_i^t is the measurement noise and r_i^t is the noisy TOA-based range measurement between the target and the ith sensor at time t. We assume that w_i^t is a random variable with mean zero, variance bounded above by σ_t^2 and is independent of the noise at other sensors and at other time steps. We also assume that $|w_i^t| \ll \|\boldsymbol{x}_t^\star - \boldsymbol{a}_i\|$ for $i = 1, \ldots, m$ and $t = 1, \ldots, T$. It is worth noting that similar assumptions have appeared in the localization literature; see, e.g., [33].

To estimate the target position at time t, a natural approach is to consider the following non-convex least-squares formulation:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ f_t(\boldsymbol{x}) := \sum_{i=1}^{m} (\|\boldsymbol{x} - \boldsymbol{a}_i\| - r_i^t)^2. \qquad (2)$$

Such a formulation is motivated by the fact that when the measurement noise vector $\boldsymbol{w}^t = (w_1^t, \dots, w_m^t)$ is Gaussian, every optimal solution to Problem (2) is a maximum-likelihood estimate of the true target position $\boldsymbol{x}_t^\star$; see, e.g., [24]. Henceforth, we shall use $\hat{\boldsymbol{x}}_t$ to denote an optimal solution to (2) (i.e., $\hat{\boldsymbol{x}}_t \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} f_t(\boldsymbol{x})$) and refer to it as a *least-squares estimate* of the true target position $\boldsymbol{x}_t^\star$. In this paper, we propose to apply OGD to tackle the time-varying optimization formulation (2), as it may not be computationally feasible to find an (approximately) optimal solution to (2) at every time step. Specifically, given an estimate $\boldsymbol{x}_{t-1}$ of the target position at time $t-1$ and the noisy range measurements $\{r_i^t\}_{i=1}^m$ at time t, we generate an estimate $\boldsymbol{x}_t$ of the target position at time t via the one-step gradient descent update

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} - \eta_t \nabla f_t(\boldsymbol{x}_{t-1}), \quad t = 1, \dots, T, \qquad (3)$$

where $\eta_t > 0$ is the step size. We remark that the update (3) should be interpreted in a formal sense at this point, as the function f_t is non-differentiable at $\boldsymbol{x} \in \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_m\}$. We shall justify the validity of (3) in the following sections.

Naturally, we are interested in evaluating the performance of the sequence of position estimates $\{\boldsymbol{x}_t\}_{t=1}^T$. For that purpose, we employ the notion of CTTE, which is defined as

$$\text{CTTE}\left(\{\boldsymbol{x}_t\}_{t=1}^T\right) := \sum_{t=1}^{T} \|\boldsymbol{x}_t - \boldsymbol{x}_t^\star\|.$$

Note that the definition of CTTE involves the sequence of *true target positions* $\{\boldsymbol{x}_t^\star\}_{t=1}^T$, not the sequence of *optimal solutions* $\{\hat{\boldsymbol{x}}_t\}_{t=1}^T$ to Problem (2), as it is the former that we are interested in tracking. Indeed, a small CTTE implies that the estimate $\boldsymbol{x}_t$ is close to the true target position $\boldsymbol{x}_t^\star$ at every time step t. Our goal is to bound the CTTE in terms of the variations in the target trajectory $\{\|\boldsymbol{x}_{t+1}^\star - \boldsymbol{x}_t^\star\|\}_{t=1}^{T-1}$ and the noise power $\{\sigma_t^2\}_{t=1}^T$ and to derive conditions that can guarantee a sublinear CTTE bound (i.e., $\frac{1}{T}\text{CTTE}\left(\{\boldsymbol{x}_t\}_{t=1}^T\right) \to 0$) on the tracking performance of OGD. We remark that a sublinear CTTE bound is a desirable property for a tracking algorithm to have, as it implies that the target tracking error of the algorithm—i.e., the distance between the target position estimate produced by the algorithm and the true target position—vanishes asymptotically. In the next section, we will develop two results that are key to achieving this goal. Specifically, under the assumption that the power of the measurement noise σ_t^2 is sufficiently small, we will first establish a bound on the estimation error $\|\hat{\boldsymbol{x}}_t - \boldsymbol{x}_t^\star\|$ and then use this bound to show that the loss function f_t is locally strongly convex at the least-squares estimate $\hat{\boldsymbol{x}}_t$.[1]

---

[1] A function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be *locally strongly convex at* $\bar{\boldsymbol{x}}$ if there exists an $r > 0$ such that g is strongly convex on the ball $B(\bar{\boldsymbol{x}}, r)$ [34].

III. LOCAL STRONG CONVEXITY OF TOA-BASED SOURCE LOCALIZATION

Consider a fixed time t. Then, Problem (2) reduces to the classic TOA-based source localization problem (see, e.g., [30]), in which the target is considered static. For notational simplicity, we drop the index t and write Problem (2) as

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \ f(\boldsymbol{x}) := \sum_{i=1}^{m} (\|\boldsymbol{x} - \boldsymbol{a}_i\| - r_i)^2 \qquad (4)$$

with $r_i = \|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + w_i$. As before, we assume that w_i is a random variable with mean zero, variance bounded above by σ^2 and satisfies $|w_i| \ll \|\boldsymbol{x}^\star - \boldsymbol{a}_i\|$. Let $\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$ denote a least-squares estimate of the true target position $\boldsymbol{x}^\star$. The following proposition, which plays a crucial role in our subsequent development, shows that $\hat{\boldsymbol{x}}$ and $\boldsymbol{x}^\star$ are close when the power of the measurement noise vector $\boldsymbol{w} = (w_1, \dots, w_m)$ is small.

Proposition 1 (Estimation Error of Least-Squares Estimator). *Suppose that $\|\boldsymbol{w}\| \leq c_0 \sqrt{m}\sigma$ for some constant $c_0 > 0$. Then, there exist constants $K_1, K_2 > 0$, which are determined by $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ and $\boldsymbol{x}^\star$, such that*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\| \leq K_1 \sqrt{m}\sigma + K_2 m\sigma^2.$$

The proof of Proposition 1 can be found in Appendix A.

The assumption on $\|\boldsymbol{w}\|$ in Proposition 1 is rather mild, as it can be satisfied with high probability when, e.g., $w_1, \dots, w_m$ are sub-Gaussian random variables [35, Chapter 3]. Now, using Proposition 1, we can prove the following theorem, which establishes the local strong convexity of f at $\hat{\boldsymbol{x}}$ and provides an explicit estimate on the size of the strong convexity region around $\hat{\boldsymbol{x}}$. This constitutes our first main result in this paper.

Theorem 1 (Local Strong Convexity of TOA-Based Source Localization). *Consider the setting of Proposition 1. Suppose that for some given $\delta > 0$, the noise power σ^2 satisfies*

$$\|\boldsymbol{x}^\star - \boldsymbol{a}_i\| > K_1 \sqrt{m}\sigma + K_2 m\sigma^2 + \delta, \quad i = 1, \dots, m \quad (5)$$

and

$$\kappa := \frac{\delta}{10m} \cdot \Lambda - (K_1 \sqrt{m}\sigma + K_2 m\sigma^2) - \frac{4c_0\sigma}{5} > 0, \quad (6)$$

where

$$\Lambda := \lambda_{\min}\left(\sum_{i=1}^{m} \left(\frac{\boldsymbol{x}^\star - \boldsymbol{a}_i}{\|\boldsymbol{x}^\star - \boldsymbol{a}_i\|}\right)\left(\frac{\boldsymbol{x}^\star - \boldsymbol{a}_i}{\|\boldsymbol{x}^\star - \boldsymbol{a}_i\|}\right)^T\right).$$

Then, we have $\nabla^2 f(\hat{\boldsymbol{x}} + \boldsymbol{\epsilon}) \succ \boldsymbol{0}$ for all $\boldsymbol{\epsilon} \in \mathbb{R}^n$ satisfying $\|\boldsymbol{\epsilon}\| \leq \kappa$; i.e., f is strongly convex over $B(\hat{\boldsymbol{x}}, \kappa)$.

The proof of Theorem 1 can be found in Appendix B. Here, let us elaborate on the assumptions of the theorem.

1) Condition (5) stipulates that the target should be sufficiently far from the sensors, which is not very restrictive in practice. Moreover, when combined with Proposition 1, the condition implies that $\|\hat{\boldsymbol{x}} - \boldsymbol{a}_i\| > \delta$ for $i = 1, \dots, m$, which shows that the loss function f is smooth around the least-squares estimate $\hat{\boldsymbol{x}}$. This allows

us to use the Hessian $\nabla^2 f$ to characterize the local strong convexity of f at $\hat{\boldsymbol{x}}$.

2) Since the vectors $\{\boldsymbol{a}_i - \boldsymbol{a}_1\}_{i=2}^m$ span $\mathbb{R}^n$ by assumption, it can be shown that the vectors $\{\boldsymbol{x}^\star - \boldsymbol{a}_i\}_{i=1}^m$ also span $\mathbb{R}^n$. This implies that $\Lambda > 0$. Thus, condition (6) can be satisfied when $\sigma > 0$ is sufficiently small (incidentally, condition (5) also becomes easier to satisfy as σ becomes smaller). An important insight drawn from (6) is that the landscape of the loss function f around the least-squares estimate $\hat{\boldsymbol{x}}$ depends on the noise power level and the geometric configuration of the target and sensors.

We remark that although the TOA-based source localization problem has been extensively studied in the literature, Theorem 1 is, to the best of our knowledge, the first result that elicits the local strong convexity property of the non-convex least-squares formulation (4). Now, since the strong convexity region $B(\hat{\boldsymbol{x}}, \kappa)$ around $\hat{\boldsymbol{x}}$ is compact and $\nabla^2 f$ is continuous over $B(\hat{\boldsymbol{x}}, \kappa)$, we see that ∇f is Lipschitz continuous over $B(\hat{\boldsymbol{x}}, \kappa)$. Thus, Theorem 1 implies that when applying the gradient descent method to tackle Problem (4), the resulting sequence of iterates will converge to the optimal solution $\hat{\boldsymbol{x}}$ at a linear rate, provided that the initial point lies in the strong convexity region around $\hat{\boldsymbol{x}}$. This can be deduced using the following well-known result.

Fact 1 (Linear Convergence of Gradient Descent for Strongly Convex Minimization; cf. [36, Theorem 2.1.15]). *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a function that is smooth, μ-strongly convex, and L-gradient Lipschitz continuous on an open convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Suppose that g has a global minimizer $\hat{\boldsymbol{x}}$ over $\mathcal{X}$. Then, the sequence $\{\boldsymbol{x}_k\}_{k\geq 0}$ generated by the gradient descent method*

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \eta \nabla g(\boldsymbol{x}_k)$$

with initial point $\boldsymbol{x}_0 \in \mathcal{X}$ and step size $\eta \in (0, 2/(\mu + L)]$ satisfies

$$\|\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}\|^2 \leq \left(1 - \frac{2\eta\mu L}{\mu + L}\right) \|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|^2.$$

In particular, Theorem 1 provides a means to justify the good empirical performance of gradient-based schemes observed in [27] when solving the TOA-based source localization problem.

IV. CTTE OF OGD FOR TOA-BASED TRACKING

Let us now address the main goal of this paper—namely, to establish a bound on the CTTE of OGD for TOA-based tracking. The results in Section III suggest that if the iterate generated by OGD at time t lies in the strong convexity region of the loss function at time $t+1$ for $t = 0, 1, \ldots, T-1$, then the tracking problem is essentially reduced to that of minimizing a time-varying strongly convex function. This opens up the possibility of using techniques from online strongly convex optimization to bound the CTTE of OGD for TOA-based tracking.

To realize the above idea, we need to first introduce some additional preliminaries and collect some consequences of the

results in Section III. Observe that the constants K_1, K_2, Λ in Theorem 1 involve the target position $\boldsymbol{x}^\star$. Since the target is moving in the tracking setting, it will simplify our subsequent analysis if we can find uniform bounds on these constants. Towards that end, we further assume that the target stays within a fixed compact region $\mathcal{T} \subseteq \mathbb{R}^n$ throughout the tracking task. Such an assumption is rather mild in practice. Moreover, since K_1, K_2, Λ depend continuously on $\boldsymbol{x}^\star$, it implies the existence of finite upper bounds on K_1, K_2 and a positive lower bound on Λ that hold for all $t \geq 0$. As a slight abuse of notation, we shall use K_1, K_2, Λ to denote these uniform bounds in the sequel.

Following the setting of Theorem 1, let $\hat{\boldsymbol{x}}_t \in \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} f_t(\boldsymbol{x})$ denote a least-squares estimate of the true target position $\boldsymbol{x}_t^\star$ at time t and $c_0 > 0$ be a constant such that $\|\boldsymbol{w}^t\| \leq c_0 \sqrt{m} \sigma_t$ for $t = 1, \ldots, T$. Furthermore, suppose that for some given $\delta > 0$, the maximum noise power $\sigma^2 := \max_{t \in \{1,\ldots,T\}} \sigma_t^2$ satisfies

$$\|\boldsymbol{x}_t^\star - \boldsymbol{a}_i\| > K_1\sqrt{m}\sigma + K_2 m\sigma^2 + \delta,$$
$$i = 1, \ldots, m; \ t = 1, \ldots, T \quad (7)$$

and

$$\kappa := \frac{\delta}{10m} \cdot \Lambda - (K_1\sqrt{m}\sigma + K_2 m\sigma^2) - \frac{4c_0\sigma}{5} > 0 \quad (8)$$

(recall from the discussion in the preceding paragraph that K_1, K_2, Λ are now uniform in t and hence κ is also uniform in t). Then, using Theorem 1, the expressions for $\nabla f_t, \nabla^2 f_t$, and the assumption that the target stays within the compact region $\mathcal{T}$, we deduce the existence of constants $\mu, L > 0$ such that for $t = 1, \ldots, T$,

1) f_t is μ-strongly convex over $B(\hat{\boldsymbol{x}}_t, \kappa)$—i.e., for any $\boldsymbol{x}, \boldsymbol{y} \in B(\hat{\boldsymbol{x}}_t, \kappa)$,

$$f_t(\boldsymbol{x}) \geq f_t(\boldsymbol{y}) + \nabla f_t(\boldsymbol{y})^T (\boldsymbol{x} - \boldsymbol{y}) + \frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2; \quad (9)$$

2) ∇f_t is L-Lipschitz continuous over $B(\hat{\boldsymbol{x}}_t, \kappa)$—i.e., for any $\boldsymbol{x}, \boldsymbol{y} \in B(\hat{\boldsymbol{x}}_t, \kappa)$,

$$\|\nabla f_t(\boldsymbol{y}) - \nabla f_t(\boldsymbol{x})\| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|; \quad (10)$$

Now, let $\{\boldsymbol{x}_t\}_{t=1}^T$ be the sequence of iterates generated by the OGD update (3) with initial point $\boldsymbol{x}_0$ and step size $\eta_t \equiv \eta \in (0, 2/(\mu + L)]$ for $t = 1, \ldots, T$. In addition, let $v_t := \|\boldsymbol{x}_{t+1}^\star - \boldsymbol{x}_t^\star\|$ ($t = 1, \ldots, T-1$) denote the variation in the true target position between time t and $t+1$ and $v := \max_{t \in \{1,\ldots,T-1\}} v_t$ denote the maximum variation in the true target position between successive time steps. The following proposition shows that under suitable conditions, OGD maintains the invariant that the iterate generated at the current time step lies in the strong convexity region of the loss function at the next time step.

Proposition 2 (Invariant of OGD). *Suppose that in addition to (7) and (8), the maximum noise power σ^2 and maximum variation v satisfy*

$$\kappa \geq \frac{2(K_1\sqrt{m}\sigma + K_2 m\sigma^2) + v}{1 - \rho}, \quad (11)$$

where $\rho := \left(1 - \frac{2\eta\mu L}{\mu+L}\right)^{1/2} \in (0,1)$ with μ, L given by (9), (10), respectively, and $\kappa > 0$ is the radius of the strong convexity region of the loss function f_t around the least-squares estimate $\hat{x}_t$ for $t = 1,\ldots,T$. Furthermore, suppose that the initial point x_0 satisfies $\|x_0 - x_1^\star\| \le K_1\sqrt{m}\sigma + K_2 m\sigma^2$. Then, for $t = 0, 1, \ldots, T-1$, the iterate x_t lies in the strong convexity region $B(\hat{x}_{t+1}, \kappa)$ of the loss function f_{t+1}.

Proof. We proceed by induction on t. For $t = 0$, we have

$$\begin{aligned}
\|x_0 - \hat{x}_1\| &\le \|x_0 - x_1^\star\| + \|x_1^\star - \hat{x}_1\| \\
&\le 2(K_1\sqrt{m}\sigma + K_2 m\sigma^2) \qquad (12) \\
&\le \kappa,
\end{aligned}$$

where the second inequality follows from our assumption on x_0 and Proposition 1 and the last follows from our choice of κ in (11). This establishes the base case. Now, for $t \ge 1$, we have

$$\begin{aligned}
\|x_t - \hat{x}_{t+1}\| &\le \|x_t - \hat{x}_t\| + \|\hat{x}_t - \hat{x}_{t+1}\| \\
&\le \rho\|x_{t-1} - \hat{x}_t\| + \|\hat{x}_t - x_t^\star\| + \|x_{t+1}^\star - \hat{x}_{t+1}\| \\
&\quad + \|x_t^\star - x_{t+1}^\star\| \\
&\le \rho\kappa + 2(K_1\sqrt{m}\sigma + K_2 m\sigma^2) + v_t \\
&\le \kappa,
\end{aligned}$$

where the second inequality follows from the OGD update (3), the inductive hypothesis (i.e., x_{t-1} lies in the strong convexity region of f_t), and Fact 1; the third follows from the inductive hypothesis and Proposition 1; the last follows from our choice of κ in (11). This completes the inductive step and also the proof of Proposition 2. □

We remark that since the loss functions $\{f_t\}_{t=1}^T$ are non-convex, some conditions on the maximum noise power, maximum variation, and quality of the initial point are to be expected in the CTTE analysis of OGD for tackling the TOA-based tracking problem (2). In fact, the performance analysis of ONM for general time-varying non-convex optimization in [23], though focusing on the dynamic regret metric, makes use of similar conditions on the maximum variation and quality of the initial point as those in Proposition 2.

Armed with Proposition 2, we can prove the following theorem, which establishes a CTTE bound for OGD when it is applied to the TOA-based tracking problem. This constitutes our second main result in this paper.

Theorem 2 (CTTE of OGD for TOA-Based Tracking). *Under the setting of Proposition 2, the sequence of iterates $\{x_t\}_{t=1}^T$ satisfies*

$$\text{CTTE}\left(\{x_t\}_{t=1}^T\right) = \mathcal{O}(1 + V(T) + N_1(T) + N_2(T)),$$

where $V(T) := \sum_{t=1}^{T-1}\|x_{t+1}^\star - x_t^\star\| = \sum_{t=1}^{T-1} v_t$ denotes the path length of the target trajectory, $N_1(T) := \sum_{t=1}^T \sigma_t$ denotes the cumulative noise standard deviation, and $N_2(T) := \sum_{t=1}^T \sigma_t^2$ denotes the cumulative noise variance.

Proof. Using the definition of CTTE and the triangle inequality, we have

$$\begin{aligned}
\text{CTTE}\left(\{x_t\}_{t=1}^T\right) &= \sum_{t=1}^T \|x_t - x_t^\star\| \\
&\le \sum_{t=1}^T \|x_t - \hat{x}_t\| + \sum_{t=1}^T \|\hat{x}_t - x_t^\star\|. \quad (13)
\end{aligned}$$

Let us now bound the two terms in (13) separately.

For the first term, we begin by adapting the argument used in the proof of [17, Theorem 1] to our time-varying optimization setting and bound

$$\begin{aligned}
\sum_{t=1}^T \|x_t - \hat{x}_t\| &\le \rho \sum_{t=1}^T \|x_{t-1} - \hat{x}_t\| \\
&\le \rho\|x_0 - \hat{x}_1\| + \rho\sum_{t=2}^T \|x_{t-1} - \hat{x}_{t-1}\| + \rho\sum_{t=2}^T \|\hat{x}_{t-1} - \hat{x}_t\| \\
&= \rho\left(\|x_0 - \hat{x}_1\| - \|x_T - \hat{x}_T\|\right) + \rho\sum_{t=1}^T \|x_t - \hat{x}_t\| \\
&\quad + \rho\sum_{t=1}^{T-1} \|\hat{x}_t - \hat{x}_{t+1}\|,
\end{aligned}$$

where the first inequality follows from the OGD update (3), Proposition 2, and Fact 1. It follows that

$$\sum_{t=1}^T \|x_t - \hat{x}_t\| \le \frac{\rho}{1-\rho}\left(\|x_0 - \hat{x}_1\| + \sum_{t=1}^{T-1} \|\hat{x}_t - \hat{x}_{t+1}\|\right). \quad (14)$$

Now, using (12), we get

$$\|x_0 - \hat{x}_1\| \le 2(K_1\sqrt{m}\sigma + K_2 m\sigma^2).$$

Furthermore, we have

$$\begin{aligned}
&\sum_{t=1}^{T-1} \|\hat{x}_t - \hat{x}_{t+1}\| \\
&\le \sum_{t=1}^{T-1} \left(\|\hat{x}_t - x_t^\star\| + \|x_t^\star - x_{t+1}^\star\| + \|x_{t+1}^\star - \hat{x}_{t+1}\|\right) \\
&\le \sum_{t=1}^{T-1} \left(K_1\sqrt{m}(\sigma_t + \sigma_{t+1}) + K_2 m(\sigma_t^2 + \sigma_{t+1}^2) + v_t\right) \\
&= \sum_{t=1}^{T-1} v_t + 2K_1\sqrt{m}\sum_{t=1}^T \sigma_t + 2K_2 m\sum_{t=1}^T \sigma_t^2,
\end{aligned}$$

where the second inequality follows from Proposition 1. Substituting the above into (14) yields

$$\sum_{t=1}^T \|x_t - \hat{x}_t\| = \mathcal{O}(1 + V(T) + N_1(T) + N_2(T)).$$

For the second term, we simply invoke Proposition 1 to get

$$\begin{aligned}
\sum_{t=1}^T \|\hat{x}_t - x_t^\star\| &\le K_1\sqrt{m}\sum_{t=1}^T \sigma_t + K_2 m\sum_{t=1}^T \sigma_t^2 \\
&= \mathcal{O}(N_1(T) + N_2(T)).
\end{aligned}$$

The desired result now follows by substituting the above into (13). □

Theorem 2 reveals that OGD can achieve sublinear CTTE when both the path length $V(T)$ and the cumulative noise power $N_2(T)$ grow sublinearly (note that the latter, together with the fact that $N_1(T) \leq \sqrt{T \cdot N_2(T)}$, implies the sublinear growth of the cumulative noise standard deviation $N_1(T)$). Roughly speaking, this means that if the target is not moving too fast and the noise power decays at a sufficiently fast rate over time, then the target tracking error of OGD will vanish asymptotically. It is important to note that our CTTE bound is expressed in terms of the path length of the *target trajectory* (i.e., $V(T) = \sum_{t=1}^{T-1} \|x_{t+1}^\star - x_t^\star\|$), not the path length of the *optimal solution trajectory* of the time-varying loss function (i.e., $V'(T) := \sum_{t=1}^{T-1} \|\hat{x}_{t+1} - \hat{x}_t\|$). Although the latter is commonly used in existing performance analyses of online methods (see, e.g., [17], [20], [23]), the former captures the actual variations in the target trajectory and is thus more relevant to the tracking problem considered in this paper. It is also worth noting that our CTTE bound shows explicitly how the TOA measurement noise affects the tracking performance of OGD through the terms $N_1(T)$ and $N_2(T)$.

V. NUMERICAL SIMULATIONS

In this section, we present numerical results to demonstrate the efficacy of OGD for the TOA-based tracking problem and illustrate our theoretical findings. Specifically, we apply both OGD and ONM—the latter has previously been used in [23] to tackle the TOA-based tracking problem—to various test instances and compare their tracking performance. In all the considered instances, there are $m = 3$ sensors, which are located at $a_1 = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}^T$, $a_2 = \begin{bmatrix} 0 & 0.5 \end{bmatrix}^T$, and $a_3 = \begin{bmatrix} 0.5 & 0 \end{bmatrix}^T$. Given the time horizon of interest T and the target trajectory $\{x_t^\star\}_{t=1}^T$, the measurement noise w_i^t in (1) is generated according to the Gaussian distribution with mean zero and variance σ_t^2 for $i = 1, \ldots, m; t = 1, \ldots, T$, and the TOA-based range measurements $\{r_i^t : i = 1, \ldots, m; t = 1, \ldots, T\}$ are then obtained using (1). We consider two initialization strategies for OGD and ONM. One is *exact initialization*, which assumes that the true initial target position $x_1^\star$ is known and takes $x_0 = x_1^\star$ as the initial point. The other is *ordinary least-squares (OLS) initialization*, which takes

$$x_0 = (A^T A)^{-1} A^T b_1 \quad (15)$$

with

$$A := \begin{bmatrix} (a_2 - a_1)^T \\ \vdots \\ (a_m - a_{m-1})^T \end{bmatrix}, \quad (16)$$

$$b_1 := \frac{1}{2} \begin{bmatrix} \|a_2\|^2 - \|a_1\|^2 + (r_1^1)^2 - (r_2^1)^2 \\ \vdots \\ \|a_m\|^2 - \|a_{m-1}\|^2 + (r_{m-1}^1)^2 - (r_m^1)^2 \end{bmatrix} \quad (17)$$

as the initial point; see [37]. The OLS estimate in (15) can be obtained as follows: Observe that any x satisfying

$$\|x - a_i\|^2 \approx (r_i^1)^2, \quad i = 1, \ldots, m$$

can serve as an estimate of the true initial target position $x_1^\star$. Upon subtracting the ith equation from the $(i+1)$st, where $i = 1, \ldots, m-1$, we get

$$2(a_{i+1} - a_i)^T x \approx \|a_{i+1}\|^2 - \|a_i\|^2 + (r_i^1)^2 - (r_{i+1}^1)^2.$$

In particular, we can obtain an estimate of $x_1^\star$ by solving

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2, \quad (18)$$

where A and b are given by (16) and (17), respectively. Since the vectors $\{a_i - a_1\}_{i=2}^m$ span $\mathbb{R}^n$ by assumption, the solution to (18) is readily given by (15). It is worth noting that the OLS estimate in (15) can be computed simply by using the sensor positions $\{a_i\}_{i=1}^m$ and noisy range measurements $\{r_i^1\}_{i=1}^m$. Thus, it is an attractive choice for initializing OGD and ONM. We use the step size $\eta_t = 0.1$ for $t = 1, \ldots, T$ in OGD. Then, OGD generates the position estimates of the target via (3), while ONM generates those via

$$x_t = x_{t-1} - \left(\nabla^2 f_t(x_{t-1})\right)^{-1} \nabla f_t(x_{t-1}), \quad t = 1, \ldots, T.$$

All computations were carried out in MATLAB on an Intel(R) Core(TM) i5-8600 CPU 3.10GHz CPU machine. The CTTE shown in the figures are averaged over 1000 Monte Carlo runs.

A. Small Noise Level and Path Variation

To begin, we construct the following set of test instances (cf. [23, Section IV]): The time horizon of interest T is set to 500. The target's initial position is set to $x_1^\star = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$ and its positions at subsequent time steps are given by

$$x_{t+1}^\star = x_t^\star + \frac{0.005}{\sqrt{2(t+1)}} u_t, \quad t = 1, \ldots, T-1, \quad (19)$$

where $u_1, \ldots, u_{T-1} \in \mathbb{R}^2$ are independently and uniformly distributed on the unit circle centered at the origin. We consider three scenarios, which correspond to three different noise levels: (i) $\sigma_t = 0.0001$ for $t = 1, \ldots, T$; (ii) $\sigma_t = 0.01$ for $t = 1, \ldots, T$; (iii) $\sigma_t = \frac{0.01}{\sqrt{t}}$ for $t = 1, \ldots, T$. Figures 1a–1c show the CTTE of OGD and ONM with exact and OLS initialization at these three noise levels. Figures 1d–1f show the tracking trajectories generated by OGD and ONM for particular instances at those noise levels with OLS initialization. We also include the trajectories of the least-squares estimates $\{\hat{x}_t\}_{t=1}^T$ in the figures for reference. These trajectories are generated using gradient descent (GD) at each time step. Specifically, at time t, we use the true target position $x_t^\star$ as the initial point and perform the GD updates using the constant step size $1/m$ until either the norm of the gradient is smaller than 10^{-8} or the number of iterations reaches 5000. We then declare the last iterate to be $\hat{x}_t$.

In the first scenario, the noise level is small compared to the path variation (i.e., $\sigma_{t+1} = 0.0001$ vs. $v_t = \frac{0.005}{\sqrt{2(t+1)}}$ for $t = 1, \ldots, T-1$ with $T = 500$). We see from Figure 1a that ONM has a smaller CTTE than OGD. This can be explained as follows: First, Proposition 1 implies that the least-squares estimate $\hat{x}_t$ is close to the true target position $x_t^\star$ for $t = 1, \ldots, T$. Second, since ONM uses both first- and second-order information of the loss function f_t, the point it generates
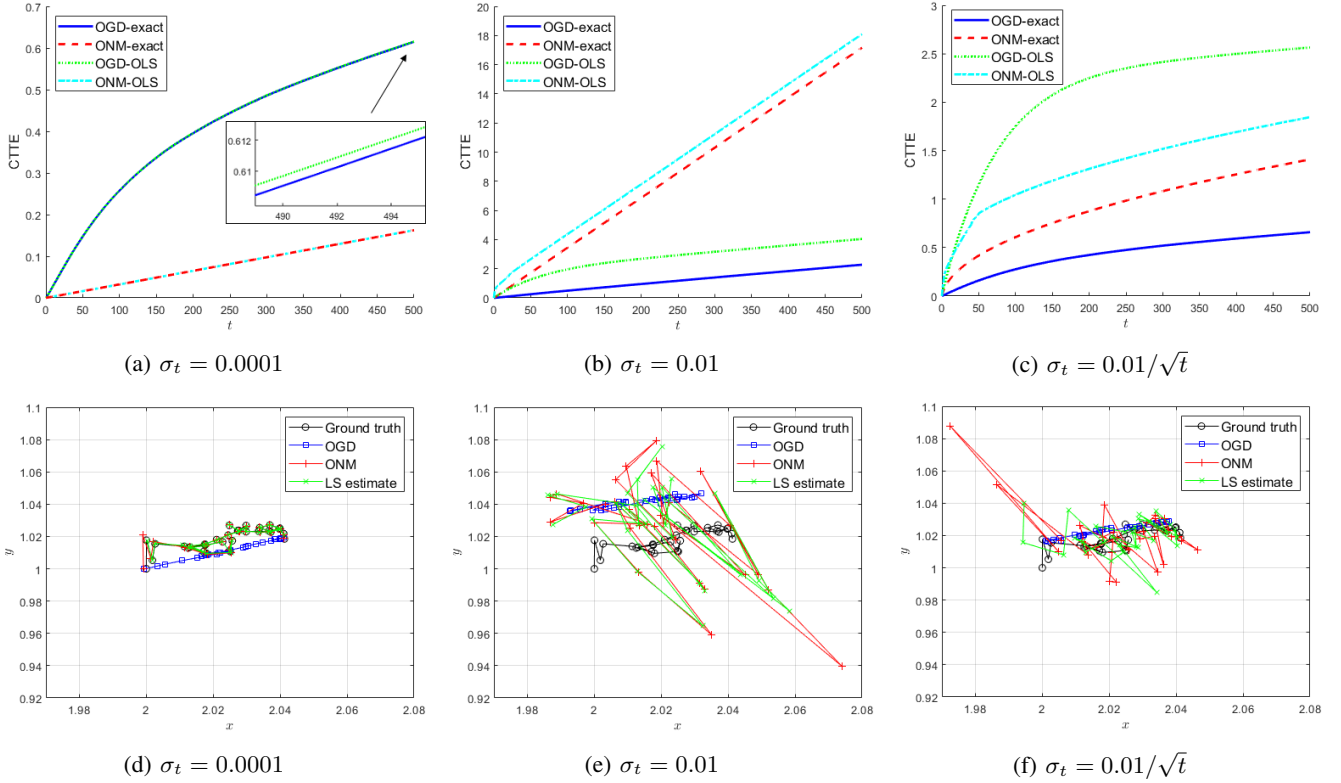
| (a) $\sigma_t = 0.0001$ | (b) $\sigma_t = 0.01$ | (c) $\sigma_t = 0.01/\sqrt{t}$ |

| (d) $\sigma_t = 0.0001$ | (e) $\sigma_t = 0.01$ | (f) $\sigma_t = 0.01/\sqrt{t}$ |

Fig. 1: CTTE (top row) and tracking trajectories (bottom row) of OGD and ONM at different noise levels.

is closer to $\hat{\boldsymbol{x}}_t$ than that generated by OGD. This suggests that ONM is better at tracking the least-squares estimates than OGD. In fact, these two claims are corroborated by our numerical results; see Figure 1d.

In the second scenario, the noise level increases relative to the path variation (i.e., $\sigma_{t+1} = 0.01$ vs. $v_t = \frac{0.005}{\sqrt{2(t+1)}}$ for $t = 1, \ldots, T-1$ with $T = 500$). Here, the ability of ONM to track the least-squares estimates closely becomes a liability, because Proposition 1 suggests that the true target position and the least-squares estimate will be further apart. Indeed, as shown in Figure 1b, ONM has a larger CTTE than OGD, and the gap widens as time goes by. We see from Figure 1e that ONM is much better at tracking the least-squares estimates than OGD. However, the least-squares estimates are quite far from the true target positions, and OGD is better at tracking the latter.

We note that in the above two scenarios, the noise level is constant, and the CTTE of OGD eventually grows linearly (see Figures 1a and 1b). This is consistent with the result in Theorem 2, as $N_1(T) = \Theta(T)$ and $N_2(T) = \Theta(T)$ and both terms dominate $V(T) = \Theta(\sqrt{T})$.

In the third scenario, the noise level diminishes as time goes by, but the relative magnitude between noise level and path variation stays roughly constant (i.e., $\sigma_{t+1} = \frac{0.01}{\sqrt{t}}$ vs. $v_t = \frac{0.005}{\sqrt{2(t+1)}}$ for $t = 1, \ldots, T-1$ with $T = 500$). From Figure 1c, we see that with exact initialization, OGD has a smaller CTTE than ONM. This suggests that the high initial noise level, which causes the least-squares estimate to deviate

from the true target position, throws off ONM and degrades its subsequent tracking performace even though the noise level is diminishing. Moreover, given the high initial noise level, the OLS initialization strategy tends to produce an inaccurate estimate of the true initial target position. Consequently, with OLS initialization, the CTTE of both OGD and ONM grow rapidly in the beginning, though the former is more affected by the quality of the OLS estimate than the latter. Nevertheless, we observe that the CTTE gap between OGD and ONM narrows as time goes by. This supports our earlier claim that OGD is better at tracking the true target positions than ONM; see also Figure 1f. Lastly, we note that the CTTE of OGD grows sublinearly. This is consistent with the result in Theorem 2, as we have $V(T) = \Theta(\sqrt{T})$, $N_1(T) = \Theta(\sqrt{T})$, and $N_2(T) = \Theta(\log T)$.

We also compare the per-iteration CPU time of OGD and ONM. As can be seen in Table I, OGD is about 2-3 times faster than ONM. The higher runtime of the latter can be attributed to the computation of the inverse of the Hessian of the loss function.

Noise Level	OGD	ONM
$\sigma_t = 0.0001$	5.23×10^{-6}s	1.36×10^{-5}s
$\sigma_t = 0.01$	5.11×10^{-6}s	1.31×10^{-5}s
$\sigma_t = 0.01/\sqrt{t}$	5.05×10^{-6}s	1.29×10^{-5}s

TABLE I: Per-iteration CPU time of OGD and ONM.

To better understand the effect of the relative magnitude between noise level and path variation on the tracking per-
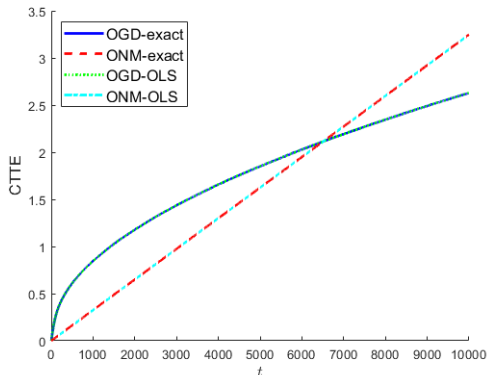
Fig. 2: CTTE of OGD and ONM at noise level $\sigma_t = 0.0001$, $T = 10000$.

formance of OGD and ONM, let us plot Figure 1a again but with the longer time horizon $T = 10000$. The result is shown in Figure 2. Although the CTTE of OGD is higher than that of ONM in the beginning, the latter eventually overtakes the former as t increases. This is consistent with our earlier observation that ONM is better at tracking the least-squares estimates than OGD. Indeed, when t is sufficiently large, the noise level $\sigma_{t+1} = 0.0001$ is larger than the path variation $v_t = \frac{0.005}{\sqrt{2(t+1)}}$. Thus, as time goes by, the true target position and the least-squares estimate become further apart (see Proposition 1), and ONM starts to incur a higher target tracking error at each time step. This suggests that the performance of ONM is rather sensitive to the noise level, while that of OGD is quite stable.

As a further illustration, we construct another set of test instances with $T = 10000$, the same initial target position $\boldsymbol{x}_1^\star = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$ and target trajectory (19) as before, and the following two different noise levels: (i) $\sigma_t = \frac{0.005}{\sqrt{2t}}$ for $t = 1, \ldots, T$; (ii) $\sigma_t = \frac{0.008}{\sqrt{2t}}$ for $t = 1, \ldots, T$. For $t = 1, \ldots, T - 1$, the ratios of noise level σ_{t+1} to path variation v_t in these two cases are 1 and 1.6, respectively. Figures 3a–3b show the CTTE of OGD and ONM with exact and OLS initialization at these two noise levels.

When the noise level to path variation ratio is 1, Figure 3a shows that ONM performs better than OGD, regardless of whether exact or OLS initialization is used. However, when the ratio increases to 1.6, Figure 3b shows that OGD eventually performs better than ONM, regardless of whether exact or OLS initialization is used. These results corroborate our earlier account that OGD is better at tracking the true target positions, while ONM is better at tracking the least-squares estimates.

B. Large Noise Level and Path Variation

Next, we study the CTTE of OGD and ONM when the two methods are applied to test instances that violate one or more of the conditions (7), (8), and (11). In particular, there is no guarantee that the iterate generated by OGD at the current time step lies in the strong convexity region of the loss function at the next time step.

We first construct a test instance that has large noise level and path variation but the ratio between them is small. The time horizon of interest is set to $T = 500$. The target's initial position is set to $\boldsymbol{x}_1^\star = \begin{bmatrix} 2 & 1 \end{bmatrix}^T$ and its subsequent positions are given by

$$\boldsymbol{x}_{t+1}^\star = \boldsymbol{x}_t^\star + \frac{0.1}{\sqrt{2(t+1)}} \boldsymbol{u}_t, \quad t = 1, \ldots, T-1.$$

Here, as before, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{T-1} \in \mathbb{R}^2$ are independently and uniformly distributed on the unit circle centered at the origin. The noise levels are given by $\sigma_t = \frac{0.1}{\sqrt{2t}}$ for $t = 1, \ldots, T$. Figure 4a shows the CTTE of OGD and ONM. We observe that the CTTE of OGD is much lower than that of ONM with both exact and OLS initialization. One possible explanation is that the good performance of ONM relies heavily on the local strong convexity of the loss function, and the lack of such a property seriously affects its performance.

Now, let us construct a test instance that has a small noise level but large path variation, so that the ratio between them is small. The time horizon of interest and the target's initial position are the same as before. The target trajectory is given by

$$\boldsymbol{x}_{t+1}^\star = \boldsymbol{x}_t^\star + \frac{0.5}{\sqrt{2(t+1)}} \boldsymbol{u}_t, \quad t = 1, \ldots, T-1,$$

while the noise levels are given by $\sigma_t = \frac{0.001}{\sqrt{2t}}$ for $t = 1, \ldots, T$. Figure 4b shows the CTTE of OGD and ONM. We see that the CTTE of OGD is much lower than that of ONM. In fact, when the iterates are no longer guaranteed to lie in the strong convexity regions of the loss functions, ONM becomes rather unstable regardless of the noise level to path variation ratio. This supports our earlier explanation that the local strong convexity of the loss function is crucial to the good performance of ONM. By contrast, OGD is much more robust and can better track the true target positions even when the conditions for local strong convexity are violated.

VI. CONCLUSION

In this paper, we established the first non-trivial performance bound for OGD when it is applied to a time-varying non-convex least-squares formulation of the TOA-based tracking problem. The performance metric we adopted is the CTTE, which measures the cumulative discrepancy between the trajectory of position estimates and that of the true target. To establish the said performance bound, we developed new results on the estimation and geometric properties of the classic static TOA-based source localization problem, which can be of independent interest. Our numerical results corroborate the theoretical findings and show that OGD can effectively track the target at different noise levels.

A possible future direction is to design and analyze online methods for TDOA-based tracking, which corresponds to a sequential version of the TDOA-based source localization problem (see, e.g., [38] and the references therein). One possible approach is to combine the results in [32] with the techniques developed in this paper. Another future direction is to study the performance of different online methods for solving the TOA-based tracking problem.
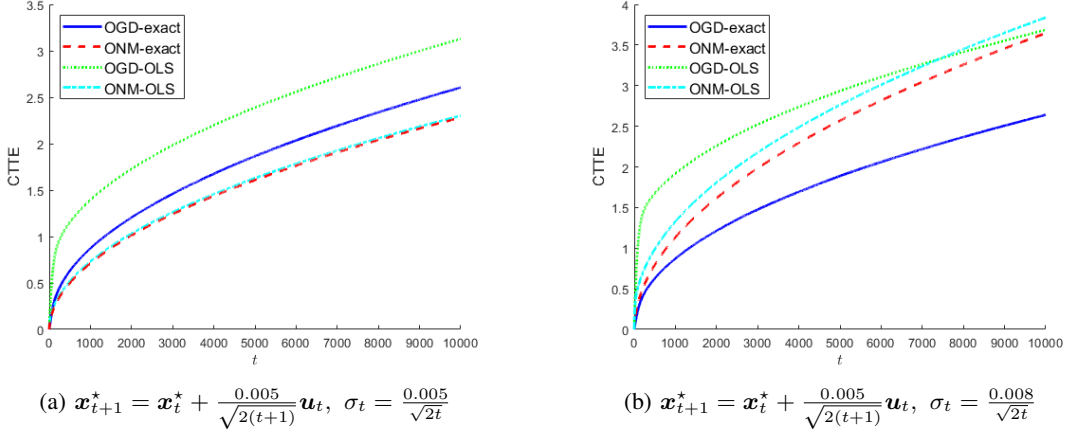
(a) $\boldsymbol{x}_{t+1}^\star = \boldsymbol{x}_t^\star + \frac{0.005}{\sqrt{2(t+1)}}\boldsymbol{u}_t,\ \sigma_t = \frac{0.005}{\sqrt{2t}}$

(b) $\boldsymbol{x}_{t+1}^\star = \boldsymbol{x}_t^\star + \frac{0.005}{\sqrt{2(t+1)}}\boldsymbol{u}_t,\ \sigma_t = \frac{0.008}{\sqrt{2t}}$

Fig. 3: CTTE of OGD and ONM when applied to different target trajectories and noise levels.

(a) $\boldsymbol{x}_{t+1}^\star = \boldsymbol{x}_t^\star + \frac{0.1}{\sqrt{2(t+1)}}\boldsymbol{u}_t,\ \sigma_t = \frac{0.1}{\sqrt{2t}}$

(b) $\boldsymbol{x}_{t+1}^\star = \boldsymbol{x}_t^\star + \frac{0.5}{\sqrt{2(t+1)}}\boldsymbol{u}_t,\ \sigma_t = \frac{0.001}{\sqrt{2t}}$
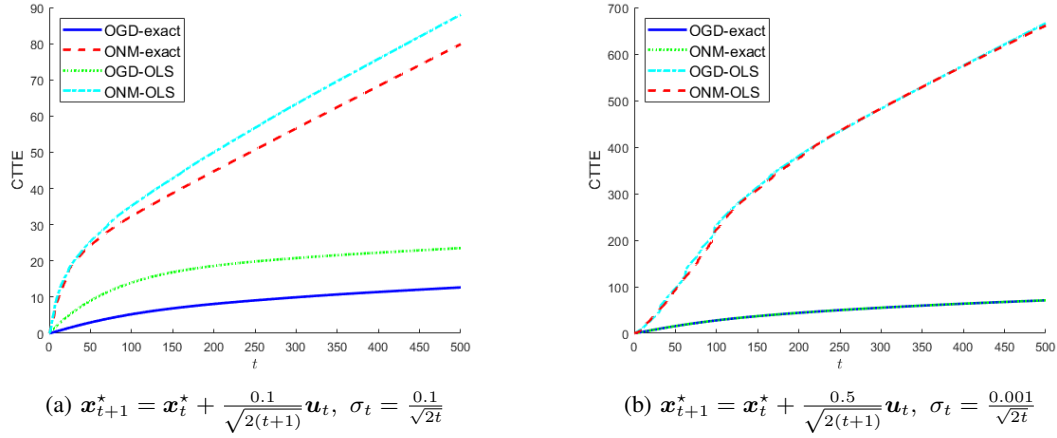
Fig. 4: CTTE of OGD and ONM with large noise level and/or path variation.

APPENDIX

A. Proof of Proposition 1

Following the development in Section V, the OLS estimate of the true target position $\boldsymbol{x}^\star$ is given by

$$\boldsymbol{x}_{\mathrm{OLS}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b},$$

where $\boldsymbol{A}$ is given in (16) and

$$\boldsymbol{b} := \frac{1}{2}\begin{bmatrix} \|\boldsymbol{a}_2\|^2 - \|\boldsymbol{a}_1\|^2 + r_1^2 - r_2^2 \\ \vdots \\ \|\boldsymbol{a}_m\|^2 - \|\boldsymbol{a}_{m-1}\|^2 + r_{m-1}^2 - r_m^2 \end{bmatrix}. \quad (20)$$

Now, let $r_i^\star = \|\boldsymbol{x}^\star - \boldsymbol{a}_i\|$ for $i = 1,\ldots,m$ and let $\boldsymbol{b}^\star$ be the vector obtained by replacing r_i with $r_i^\star$ in (20). Then, the derivation in Section V shows that $\boldsymbol{x}^\star$ satisfies $\boldsymbol{x}^\star = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}^\star$. This implies that

$$\|\boldsymbol{x}_{\mathrm{OLS}} - \boldsymbol{x}^\star\| = \|(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{b}^\star)\|.$$

Using the fact that $r_i = r_i^\star + w_i$ for $i = 1,\ldots,m$, we get $r_i^2 - (r_i^\star)^2 = 2r_i^\star w_i + w_i^2$ and hence

$$\|\boldsymbol{b} - \boldsymbol{b}^\star\| \le C_0\|\boldsymbol{w}\| + \frac{1}{2}\|\tilde{\boldsymbol{w}}\|$$

for some constant $C_0 > 0$, where

$$\tilde{\boldsymbol{w}} := \begin{bmatrix} w_1^2 - w_2^2 \\ w_2^2 - w_3^2 \\ \vdots \\ w_{m-1}^2 - w_m^2 \end{bmatrix}.$$

Since $\|\tilde{\boldsymbol{w}}\| \le \|\tilde{\boldsymbol{w}}\|_1 \le 2\|\boldsymbol{w}\|^2$, our assumption on $\|\boldsymbol{w}\|$ yields

$$\|\boldsymbol{x}_{\mathrm{OLS}} - \boldsymbol{x}^\star\| \le C_1\sqrt{m}\sigma + C_2 m\sigma^2 \quad (21)$$

for some constants $C_1, C_2 > 0$.

Next, let $\hat{r}_i = \|\hat{\boldsymbol{x}} - \boldsymbol{a}_i\|$ for $i = 1,\ldots,m$ and let $\hat{\boldsymbol{b}}$ be the vector obtained by replacing r_i with $\hat{r}_i$ in (20). By repeating the same argument as above and noting that

$$r_i^2 - \hat{r}_i^2 = 2r_i(r_i - \hat{r}_i) - (r_i - \hat{r}_i)^2$$
$$= 2(r_i^\star + w_i)(r_i - \hat{r}_i) - (r_i - \hat{r}_i)^2,$$

$$f(\hat{\boldsymbol{x}}) = \sum_{i=1}^m (\hat{r}_i - r_i)^2 \le f(\boldsymbol{x}^\star) = \|\boldsymbol{w}\|^2 \le c_0^2 m\sigma^2,$$

$$|w_i| \ll r_i^\star,$$

we have

$$\|\boldsymbol{b} - \hat{\boldsymbol{b}}\| \le C_3\sqrt{f(\hat{\boldsymbol{x}})} + \frac{1}{2}\|\tilde{\boldsymbol{r}}\| \le C_3\sqrt{f(\hat{\boldsymbol{x}})} + C_4 f(\hat{\boldsymbol{x}})$$
$$\le C_5\sqrt{m}\sigma + C_6 m\sigma^2$$

for some constants $C_3, C_4, C_5, C_6 > 0$, where

$$\tilde{\boldsymbol{r}} := \begin{bmatrix} (r_2 - \hat{r}_2)^2 - (r_1 - \hat{r}_1)^2 \\ (r_3 - \hat{r}_3)^2 - (r_2 - \hat{r}_2)^2 \\ \vdots \\ (r_m - \hat{r}_m)^2 - (r_{m-1} - \hat{r}_{m-1})^2 \end{bmatrix}.$$

This gives

$$\|\boldsymbol{x}_{\mathrm{OLS}} - \hat{\boldsymbol{x}}\| \le C_7\sqrt{m}\sigma + C_8 m\sigma^2 \tag{22}$$

for some constants $C_7, C_8 > 0$. The desired result then follows by applying the triangle inequality to (21) and (22).

B. Proof of Theorem 1

We begin with two technical lemmas.

Lemma 1. *Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ be two linearly independent vectors. Then,*

$$\lambda_{\min}(\boldsymbol{u}\boldsymbol{u}^T - \boldsymbol{v}\boldsymbol{v}^T) = \frac{\|\boldsymbol{u}\|^2 - \|\boldsymbol{v}\|^2 - \|\boldsymbol{u} - \boldsymbol{v}\|\|\boldsymbol{u} + \boldsymbol{v}\|}{2}. \tag{23}$$

Proof. Let $(\lambda, \boldsymbol{w}) \in \mathbb{R} \times \mathbb{R}^n$ be an eigenpair of $\boldsymbol{u}\boldsymbol{u}^T - \boldsymbol{v}\boldsymbol{v}^T$; i.e., $\|\boldsymbol{w}\| = 1$ and

$$(\boldsymbol{u}\boldsymbol{u}^T - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{w} = \lambda\boldsymbol{w}.$$

If $\boldsymbol{w} \in \mathrm{span}\{\boldsymbol{u}, \boldsymbol{v}\}^\perp$, then $\lambda = 0$. Otherwise, we can write $\boldsymbol{w} = a\boldsymbol{u} + b\boldsymbol{v}$ for some $a, b \in \mathbb{R}$ and compute

$$(\boldsymbol{u}\boldsymbol{u}^T - \boldsymbol{v}\boldsymbol{v}^T)(a\boldsymbol{u} + b\boldsymbol{v})$$
$$= (a\|\boldsymbol{u}\|^2 + b(\boldsymbol{u}^T\boldsymbol{v}))\boldsymbol{u} - (a(\boldsymbol{u}^T\boldsymbol{v}) + b\|\boldsymbol{v}\|^2)\boldsymbol{v}$$
$$= \lambda a\boldsymbol{u} + \lambda b\boldsymbol{v}. \tag{24}$$

Consider the following cases:

Case 1: $\boldsymbol{u}^T\boldsymbol{v} = 0$.
It is immediate that $\lambda_{\min}(\boldsymbol{u}\boldsymbol{u}^T - \boldsymbol{v}\boldsymbol{v}^T) = -\|\boldsymbol{v}\|^2$ and the corresponding eigenvector is $\boldsymbol{w} = \boldsymbol{v}/\|\boldsymbol{v}\|$. Since $\|\boldsymbol{u} - \boldsymbol{v}\|^2 = \|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$, we obtain (23).

Case 2: $\boldsymbol{u}^T\boldsymbol{v} \neq 0$.
We claim that both a and b must be non-zero. Indeed, suppose to the contrary that $a = 0$. Since $\boldsymbol{w} \neq \boldsymbol{0}$, we have $b \neq 0$. It follows from (24) that

$$(\boldsymbol{u}^T\boldsymbol{v})\boldsymbol{u} = (\lambda + \|\boldsymbol{v}\|^2)\boldsymbol{v}.$$

As $\boldsymbol{u}^T\boldsymbol{v} \neq 0$, we also have $\lambda \neq -\|\boldsymbol{v}\|^2$. However, the above relation contradicts the linear independence of $\boldsymbol{u}$ and $\boldsymbol{v}$. Thus, we conclude that $a \neq 0$. A similar argument shows that $b \neq 0$.

Now, by equating terms in (24), we have

$$\lambda = \frac{a\|\boldsymbol{u}\|^2 + b(\boldsymbol{u}^T\boldsymbol{v})}{a} = \frac{-a(\boldsymbol{u}^T\boldsymbol{v}) - b\|\boldsymbol{v}\|^2}{b}. \tag{25}$$

Solving the quadratic equation

$$(\boldsymbol{u}^T\boldsymbol{v})a^2 + (\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2)ab + (\boldsymbol{u}^T\boldsymbol{v})b^2 = 0,$$

we obtain a relationship between a and b. Plugging this relationship into (25) yields

$$\lambda = \frac{(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2) \pm \sqrt{(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2)^2 - 4(\boldsymbol{u}^T\boldsymbol{v})^2}}{2} - \|\boldsymbol{v}\|^2$$
$$= \frac{\|\boldsymbol{u}\|^2 - \|\boldsymbol{v}\|^2 \pm \|\boldsymbol{u} - \boldsymbol{v}\|\|\boldsymbol{u} + \boldsymbol{v}\|}{2}.$$

The negative root gives a non-positive eigenvalue by the Cauchy-Schwarz inequality. This establishes (23). $\square$

Lemma 2. *For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$, we have*

$$\left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} - \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \right\| \le \frac{\|\boldsymbol{x} - \boldsymbol{y}\|}{\min\{\|\boldsymbol{x}\|, \|\boldsymbol{y}\|\}}.$$

Proof. If $\|\boldsymbol{x}\| = \|\boldsymbol{y}\|$, then the inequality trivially holds as equality. Hence, we may assume without loss of generality that $\|\boldsymbol{x}\| < \|\boldsymbol{y}\|$. Consider

$$\left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} - \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \right\|^2 = \frac{1}{\|\boldsymbol{x}\|^2} \left\| \boldsymbol{x} - \frac{\|\boldsymbol{x}\|}{\|\boldsymbol{y}\|}\boldsymbol{y} \right\|^2.$$

Writing $\alpha = \|\boldsymbol{x}\|/\|\boldsymbol{y}\|$, where $0 < \alpha < 1$ because $\|\boldsymbol{x}\| < \|\boldsymbol{y}\|$, we see that

$$\|\boldsymbol{x} - \alpha\boldsymbol{y}\|^2$$
$$= \|\boldsymbol{x}\|^2 - 2\boldsymbol{x}^T\boldsymbol{y} + \|\boldsymbol{y}\|^2 + 2(1-\alpha)\boldsymbol{x}^T\boldsymbol{y} + (\alpha^2 - 1)\|\boldsymbol{y}\|^2$$
$$= \|\boldsymbol{x} - \boldsymbol{y}\|^2 + (1 - \alpha)(2\boldsymbol{x}^T\boldsymbol{y} - (\alpha + 1)\|\boldsymbol{y}\|^2)$$

and

$$2\boldsymbol{x}^T\boldsymbol{y} - (\alpha + 1)\|\boldsymbol{y}\|^2 \le 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| - (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)\|\boldsymbol{y}\|$$
$$= \|\boldsymbol{y}\|(\|\boldsymbol{x}\| - \|\boldsymbol{y}\|) < 0.$$

This completes the proof. $\square$

Armed with Lemmas 1 and 2, we are now ready to prove Theorem 1. By Proposition 1 and the assumption of Theorem 1, we have

$$\|\hat{\boldsymbol{x}} - \boldsymbol{a}_i\| \ge \|\boldsymbol{x}^\star - \boldsymbol{a}_i\| - \|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\| > \delta, \quad i = 1, \ldots, m. \tag{26}$$

Thus, the loss function f is twice continuously differentiable at $\hat{\boldsymbol{x}}$ with

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

Our goal is to prove that $\nabla^2 f(\hat{\boldsymbol{x}} + \boldsymbol{\epsilon}) \succ \boldsymbol{0}$ for all $\boldsymbol{\epsilon}$ within some ball. In particular, this would imply that $\nabla^2 f(\hat{\boldsymbol{x}}) \succ \boldsymbol{0}$. To begin, consider a fixed $i \in \{1, \ldots, m\}$. Since $\Lambda/m \le 1$, we have $\kappa < \delta/10$. This, together with Proposition 1 and the assumption that $\|\boldsymbol{\epsilon}\| \le \kappa$, gives

$$\|\hat{\boldsymbol{x}} + \boldsymbol{\epsilon} - \boldsymbol{a}_i\| \le \|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\| + \|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + \|\boldsymbol{\epsilon}\|$$
$$\le \|\boldsymbol{x}^\star - \boldsymbol{a}_i\| + K_1\sqrt{m}\sigma + K_2 m\sigma^2 + \frac{\delta}{10}.$$

Moreover, since $|w_i| \ll \|x^\star - a_i\|$, we may take $r_i \geq \|x^\star - a_i\|/2$. Putting these together and using the assumption of Theorem 1, we obtain

$$\frac{r_i}{\|\hat{x} + \epsilon - a_i\|}$$
$$\geq \frac{1/2}{1 + \|x^\star - a_i\|^{-1}\left(K_1\sqrt{m}\sigma + K_2 m\sigma^2 + \delta/10\right)}$$
$$\geq \frac{1}{4}.$$

Hence, we can bound

$$\lambda_{\min}(\nabla^2 f(\hat{x} + \epsilon))$$
$$\geq 2 \cdot \lambda_{\min}\left(\sum_{i=1}^m \frac{r_i}{\|\hat{x} + \epsilon - a_i\|^3}(\hat{x} + \epsilon - a_i)(\hat{x} + \epsilon - a_i)^T\right)$$
$$+ 2\sum_{i=1}^m \left(1 - \frac{r_i}{\|\hat{x} + \epsilon - a_i\|}\right)$$
$$\geq \frac{1}{2} \cdot \lambda_{\min}\left(\sum_{i=1}^m \left(\left(\frac{\hat{x} + \epsilon - a_i}{\|\hat{x} + \epsilon - a_i\|}\right)\left(\frac{\hat{x} + \epsilon - a_i}{\|\hat{x} + \epsilon - a_i\|}\right)^T\right.\right.$$
$$\left.\left. - \left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)\left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)^T\right)\right)$$
$$+ \frac{1}{2} \cdot \lambda_{\min}\left(\sum_{i=1}^m \left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)\left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)^T\right)$$
$$+ 2\sum_{i=1}^m \frac{\|\hat{x} + \epsilon - a_i\| - r_i}{\|\hat{x} + \epsilon - a_i\|}. \tag{27}$$

Now, let us bound the first and last terms in (27) separately. For the first term, we have

$$\lambda_{\min}\left(\sum_{i=1}^m \left(\left(\frac{\hat{x} + \epsilon - a_i}{\|\hat{x} + \epsilon - a_i\|}\right)\left(\frac{\hat{x} + \epsilon - a_i}{\|\hat{x} + \epsilon - a_i\|}\right)^T\right.\right.$$
$$\left.\left. - \left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)\left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)^T\right)\right)$$
$$\geq -\frac{1}{2}\sum_{i=1}^m \left\|\frac{\hat{x} + \epsilon - a_i}{\|\hat{x} + \epsilon - a_i\|} - \frac{x^\star - a_i}{\|x^\star - a_i\|}\right\| \cdot$$
$$\left\|\frac{\hat{x} + \epsilon - a_i}{\|\hat{x} + \epsilon - a_i\|} + \frac{x^\star - a_i}{\|x^\star - a_i\|}\right\|$$
$$\geq -\sum_{i=1}^m \left\|\frac{\hat{x} + \epsilon - a_i}{\|\hat{x} + \epsilon - a_i\|} - \frac{x^\star - a_i}{\|x^\star - a_i\|}\right\|$$
$$\geq -\sum_{i=1}^m \frac{\|\hat{x} + \epsilon - x^\star\|}{\min\{\|\hat{x} + \epsilon - a_i\|, \|x^\star - a_i\|\}}$$
$$\geq -\frac{2m}{\delta} \cdot \|\hat{x} + \epsilon - x^\star\|, \tag{28}$$

where the first inequality follows from Lemma 1, the second follows by applying triangle inequality to the term $\left\|\frac{\hat{x}+\epsilon-a_i}{\|\hat{x}+\epsilon-a_i\|} + \frac{x^\star-a_i}{\|x^\star-a_i\|}\right\|$, the third follows from Lemma 2, and the last is due to

$$\min\{\|\hat{x} + \epsilon - a_i\|, \|x^\star - a_i\|\} \geq \frac{\delta}{2},$$

which follows from (26) and the assumption that $\|\epsilon\| \leq \kappa < \delta/2$. For the last term in (27), we use

$$\left|\|\hat{x} + \epsilon - a_i\| - r_i\right| \leq \|\hat{x} + \epsilon - x^\star\| + |w_i|,$$

(26), and the assumption that $\|\epsilon\| \leq \kappa < \delta/2$ to bound

$$\sum_{i=1}^m \frac{\|\hat{x} + \epsilon - a_i\| - r_i}{\|\hat{x} + \epsilon - a_i\|} \geq -\sum_{i=1}^m \frac{|w_i| + \|\hat{x} + \epsilon - x^\star\|}{\|\hat{x} + \epsilon - a_i\|}$$
$$\geq -\frac{2}{\delta}\sum_{i=1}^m \left(|w_i| + \|\hat{x} + \epsilon - x^\star\|\right). \tag{29}$$

Upon substituting (28) and (29) into (27) and noting that $\sum_{i=1}^m |w_i| \leq \sqrt{m}\|w\| \leq c_0 m\sigma$ by assumption, we obtain

$$\lambda_{\min}(\nabla^2 f(\hat{x} + \epsilon))$$
$$\geq \frac{1}{2} \cdot \lambda_{\min}\left(\sum_{i=1}^m \left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)\left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)^T\right)$$
$$- \frac{5m}{\delta}(\|\hat{x} - x^\star\| + \|\epsilon\|) - \frac{4c_0 m\sigma}{\delta}. \tag{30}$$

Using Proposition 1, we see that the right-hand side of (30) is positive whenever

$$\|\epsilon\| < \frac{\delta}{10m} \cdot \lambda_{\min}\left(\sum_{i=1}^m \left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)\left(\frac{x^\star - a_i}{\|x^\star - a_i\|}\right)^T\right)$$
$$- (K_1\sqrt{m}\sigma + K_2 m\sigma^2) - \frac{4c_0\sigma}{5}. \tag{31}$$

Since σ satisfies (6), the right-hand side of (31) is positive. This completes the proof.

REFERENCES

[1] Y.-M. Pun and A. M.-C. So, "Dynamic regret bound for moving target tracking based on online time-of-arrival measurements," in *Proceedings of the 59th IEEE Conference on Decision and Control (CDC 2020)*, 2020, pp. 5968–5973.

[2] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17–29, 2002.

[3] É. L. Souza, E. F. Nakamura, and R. W. Pazzi, "Target tracking for sensor networks: A survey," *ACM Computing Surveys*, vol. 49, no. 2, p. Article 30, 2016.

[4] C. Chen, H. Wang, N. T. Chew, J. Ibañez-Guzmán, S. Jian, and C. C. Wah, "Target-tracking and path planning for vehicle following in jungle environment," in *Proceedings of the 8th International Conference on Control, Automation, Robotics and Vision (ICARCV 2004)*, vol. 1, 2004, pp. 455–460.

[5] J. Derenick, J. Spletzer, and A. Hsieh, "An optimal approach to collaborative target tracking with performance guarantees," *Journal of Intelligent and Robotic Systems*, vol. 56, no. 1–2, pp. 47–67, 2009.

[6] P. Chakravarty and R. Jarvis, "Multiple target tracking for surveillance: A particle filter approach," in *Proceedings of the 2005 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, 2005, pp. 181–186.

[7] B. Pannetier, J. Dezert, and G. Sella, "Multiple target tracking with wireless sensor network for ground battlefield surveillance," in *Proceedings of the 17th International Conference on Information Fusion (FUSION 2014)*, 2014, pp. 1–8.

[8] T. Li, A. Ekpenyong, and Y.-F. Huang, "Source localization and tracking using distributed asynchronous sensors," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3991–4003, 2006.

[9] P.-H. Tseng, K.-T. Feng, Y.-C. Lin, and C.-L. Chen, "Wireless location tracking algorithms for environments with insufficient signal sources," *IEEE Transactions on Mobile Computing*, vol. 8, no. 12, pp. 1676–1689, 2009.

[10] Y. Wang, G. Leus, and X. Ma, "Tracking a mobile node by asynchronous networks," in *Proceedings of the 12th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2011)*, 2011, pp. 161–165.

[11] L. Yi, S. G. Razul, Z. Lin, and C. M. See, "Target tracking in mixed LOS/NLOS environments based on individual measurement estimation and LOS detection," *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 99–111, 2014.

[12] E. Dall'Anese, A. Simonetto, S. Becker, and L. Madden, "Optimization and learning with information streams: Time-varying algorithms and applications," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 71–83, 2020.

[13] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, "Time-varying convex optimization: Time-structured algorithms and applications," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2032–2048, 2020.

[14] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proceedings of the 20th International Conference on International Conference on Machine Learning (ICML 2003)*, 2003, pp. 928–935.

[15] E. C. Hall and R. M. Willett, "Online convex optimization in dynamic environments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 647–662, 2015.

[16] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, "Online optimization : Competing with dynamic comparators," in *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS 2015)*, 2015, pp. 398–406.

[17] A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro, "Online optimization in dynamic environments: Improved regret rates for strongly convex problems," in *Proceedings of the 55th IEEE Conference on Decision and Control (CDC 2016)*, 2016, pp. 7195–7201.

[18] A. Simonetto, A. Mokhtari, A. Koppel, G. Leus, and A. Ribeiro, "A class of prediction-correction methods for time-varying convex optimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4576–4591, 2016.

[19] S. Shahrampour and A. Jadbabaie, "An online optimization approach for multi-agent tracking of dynamic parameters in the presence of adversarial noise," in *Proceedings of the 2017 American Control Conference (ACC)*, 2017, pp. 3306–3311.

[20] A. S. Bedi, P. Sarma, and K. Rajawat, "Tracking moving agents via inexact online gradient descent algorithm," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 202–217, 2018.

[21] D. Zhang, F. Xia, Z. Yang, L. Yao, and W. Zhao, "Localization technologies for indoor human tracking," in *Proceedings of the 5th International Conference on Future Information Technology*, 2010, pp. 1–6.

[22] E. Xu, Z. Ding, and S. Dasgupta, "Target tracking and mobile sensor navigation in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 177–186, 2013.

[23] A. Lesage-Landry, J. A. Taylor, and I. Shames, "Second-order online nonconvex optimization," 2020, accepted for publication in *IEEE Transactions on Automatic Control*.

[24] K. W. Cheung, W.-K. Ma, and H. C. So, "Accurate approximation algorithm for TOA-based maximum likelihood mobile location using semidefinite programming," pp. II–145–II–148, 2004.

[25] A. M.-C. So and Y. Ye, "Theory of semidefinite programming for sensor network localization," *Mathematical Programming, Series B*, vol. 109, no. 2, pp. 367–384, 2007.

[26] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770–1778, 2008.

[27] A. Beck, M. Teboulle, and Z. Chikishev, "Iterative minimization schemes for solving the single source localization problem," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1397–1416, 2008.

[28] E. Xu, Z. Ding, and S. Dasgupta, "Source localization in wireless sensor networks from signal time-of-arrival measurements," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2887–2897, 2011.

[29] S. Ji, K.-F. Sze, Z. Zhou, A. M.-C. So, and Y. Ye, "Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization," in *Proceedings of the 32nd IEEE International Conference on Computer Communications (INFOCOM 2013)*, 2013, pp. 2499–2507.

[30] H. C. So, "Source localization: Algorithms and analysis," in *Handbook of Position Location: Theory, Practice, and Advances*, 2nd ed., S. A. (Reza) Zekavat and R. M. Buehrer, Eds. New Jersey: John Wiley & Sons, Inc., 2019, pp. 59–106.

[31] A. Héliou, M. Martin, P. Mertikopoulos, and T. Rahier, "Online non-convex optimization with imperfect feedback," in *Advances in Neural Information Processing Systems 33: Proceedings of the 2020 Conference,* to appear, 2020.

[32] H. Liu, Y.-M. Pun, and A. M.-C. 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)*. IEEE, 2017, pp. 1–5.

[33] G. Wang, A. M.-C. So, and Y. Li, "Robust convex approximation methods for TDOA-based localization under NLOS conditions," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3281–3296, 2016.

[34] J.-P. Vial, "Strong convexity of sets and functions," *Journal of Mathematical Economics*, vol. 9, no. 1–2, pp. 187–205, 1982.

[35] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2018, vol. 47.

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

[37] A. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24–40, 2005.

[38] K. W. K. Lui, F. K. W. Chan, and H. C. So, "Semidefinite programming approach for range-difference based source localization," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1630–1633, 2009.

Yuen-Man Pun received the BSc degree in Mathematics and MPhil degree in Systems Engineering and Engineering Management (SEEM), both from The Chinese University of Hong Kong (CUHK). She is currently pursuing a PhD degree in SEEM at CUHK. Her research focuses on algorithmic design and analysis and its applications in data science and signal processing.
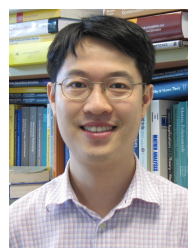
Anthony Man-Cho So (M'12-SM'17) received the BSE degree in Computer Science from Princeton University, Princeton, NJ, USA, with minors in Applied and Computational Mathematics, Engineering and Management Systems, and German Language and Culture. He then received the M.Sc. degree in Computer Science and the Ph.D. degree in Computer Science with a Ph.D. minor in Mathematics from Stanford University, Stanford, CA, USA.

Dr. So joined The Chinese University of Hong Kong (CUHK) in 2007. He is now the Associate Dean of Student Affairs in the Faculty of Engineering, Deputy Master of Morningside College, and Professor in the Department of Systems Engineering and Engineering Management. His research focuses on optimization theory and its applications in various areas of science and engineering, including computational geometry, machine learning, signal processing, and statistics.

Dr. So is appointed as an Outstanding Fellow of the Faculty of Engineering at CUHK in 2019. He has received a number of research and teaching awards, including the 2018 IEEE Signal Processing Society Best Paper Award, the 2015 IEEE Signal Processing Society Signal Processing Magazine Best Paper Award, the 2014 IEEE Communications Society Asia-Pacific Outstanding Paper Award, the 2013 CUHK Vice-Chancellor's Exemplary Teaching Award, and the 2010 Institute for Operations Research and the Management Sciences (INFORMS) Optimization Society Optimization Prize for Young Researchers. He currently serves on the editorial boards of Journal of Global Optimization, Optimization Methods and Software, and SIAM Journal on Optimization. He was also the Lead Guest Editor of IEEE SIGNAL PROCESSING MAGAZINE Special Issue on "Non-Convex Optimization for Signal Processing and Machine Learning".