Local Strong Convexity of Maximum-Likelihood TDOA-Based Source Localization and Its Algorithmic Implications

Huikang Liu, Yuen-Man Pun, and Anthony Man-Cho So
Dept. of Syst. Eng. & Eng. Manag., The Chinese Univ. of Hong Kong, Hong Kong SAR, China
Email: {hkliu, ympun, manchoso}@se.cuhk.edu.hk

***Abstract*—We consider the problem of single source localization using time-difference-of-arrival (TDOA) measurements. By analyzing the maximum-likelihood (ML) formulation of the problem, we show that under certain mild assumptions on the measurement noise, the estimation errors of both the closed-form least-squares estimate proposed in [1] and the ML estimate, as measured by their distances to the true source location, are of the same order. We then use this to establish the curious result that the objective function of the ML estimation problem is actually locally strongly convex at an optimal solution. This implies that some lightweight solution methods, such as the gradient descent (GD) and Levenberg-Marquardt (LM) methods, will converge to an optimal solution to the ML estimation problem when properly initialized, and the convergence rates can be determined by standard arguments. To the best of our knowledge, these results are new and contribute to the growing literature on the effectiveness of lightweight solution methods for structured non-convex optimization problems. Lastly, we demonstrate via simulations that the GD and LM methods can indeed produce more accurate estimates of the source location than some existing methods, including the widely used semidefinite relaxation-based methods.**

I. Introduction

Source localization has received considerable attention in the signal processing community for its many applications in everyday life, such as self-located map services, person and asset tracking, wireless network security, and advanced location-based services [2], [3]. In practice, time-difference-of-arrival (TDOA) measurements are often used to estimate the source localization, as they have less stringent synchronization requirements on the sensors [3], [4]. Over the years, many methods have been proposed to tackle the TDOA-based source localization problem. For instance, an equation-error formulation approach was proposed in [1], which involves solving a certain system of linear equations to produce a closed-form least-squares estimate of the source location. This approach is simple but generally suffers from poor accuracy. More recently, a popular approach to tackling the TDOA-based source localization problem is to use the semidefinite relaxation (SDR) technique [5]. Such an approach has been studied in [6] and [7]. The SDR approach is computationally demanding and its performance could sometimes be far from the Cramér-Rao lower bound (CRLB); see, e.g., [7]. The above discussion motivates us to ask whether there are lightweight methods that can yield accurate estimates of the source location using TDOA measurements. In this paper, we consider two candidate methods, namely, the gradient descent (GD) method and the Levenberg-Marquardt (LM) method. Both of these methods have been used to tackle the ML formulation of the TDOA-based source localization problem; see, e.g., [8] and the references therein. However, since the ML formulation is non-convex in general, it is not clear whether these methods will get trapped at local minima of the objective function. To address this issue, we conduct a theoretical analysis of the ML formulation of the TDOA-based source localization problem. Our contribution is twofold. First, we show that under some mild assumptions on the measurement noise, the estimation errors of both the least-squares estimate of [1] and the ML estimate are of the same order; see Theorem 1 and Corollary 1. Then, we use this result to show that the objective function of the ML estimation problem is locally strongly convex at an optimal solution. This implies that with a proper initialization, the GD (resp. LM) method will converge *linearly* (resp. *quadratically*) to an optimal solution to the ML estimation problem. To the best of our knowledge, these results are new and contribute to the growing literature (see, e.g., [9], [10], [11] and the references therein) on the effectiveness of lightweight solution methods for structured non-convex optimization problems.

The paper is organized as follows. In Section II, we introduce the ML formulation of the TDOA-based source localization problem. Then, in Section III, we review several lightweight solution methods for tackling the ML formulation. Next, we provide a theoretical analysis of the ML formulation in Section IV and verify our theoretical results via simulations in Section V. Lastly, we conclude in Section VI.

II. Problem Formulation

We are interested in the n-dimensional TDOA-based single source localization problem, in which the goal is to estimate the source location $\boldsymbol{x}^\star \in \mathbb{R}^n$ based on a given set of noisy range-difference measurements of the form

$$d_i = \|\boldsymbol{x}^\star - \boldsymbol{a}_i\|_2 - \|\boldsymbol{x}^\star - \boldsymbol{a}_1\|_2 + n_i \quad \text{for } i = 2, \ldots, m, \tag{1}$$

where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$ are the known locations of m given sensors and $n_2, \ldots, n_m$ are the measurement noise. Here, the first sensor is designated as the reference sensor and we assume that the measurement noise takes the form $n_i = g_i - g_1$, where $g_1, \ldots, g_m$ are independent and identically distributed Gaussian random variables with mean zero and variance $\sigma^2 > 0$. Under this setting, the covariance matrix of the noise vector $\boldsymbol{n} = [n_2 \quad n_3 \quad \cdots \quad n_m]^T$ is given by $\sigma^2 \boldsymbol{Q}$, where $\boldsymbol{Q} = \boldsymbol{1}_{m-1}\boldsymbol{1}_{m-1}^T + \boldsymbol{I}_{m-1} \in \mathbb{R}^{(m-1)\times(m-1)}$ with $\boldsymbol{1}_{m-1}$ and $\boldsymbol{I}_{m-1}$ being the $(m-1)$-dimensional all-one vector and the $(m-1)\times(m-1)$ identity matrix, respectively. The maximum-likelihood (ML) estimator $\hat{\boldsymbol{x}}$ of $\boldsymbol{x}^\star$ can then be found via

$$\hat{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ f(\boldsymbol{x}) := (\boldsymbol{d}-\boldsymbol{U}\boldsymbol{h})^T \boldsymbol{Q}^{-1}(\boldsymbol{d}-\boldsymbol{U}\boldsymbol{h}) \right\}, \quad (2)$$

where

$$\begin{aligned} \tau_i &= \|\boldsymbol{x}-\boldsymbol{a_i}\|_2 \quad \text{for } i=1,\dots,m, \\ \boldsymbol{h} &= [\tau_1 \quad \tau_2 \quad \cdots \quad \tau_m]^T, \\ \boldsymbol{d} &= [d_2 \quad d_3 \quad \cdots \quad d_m]^T, \\ \boldsymbol{U} &= [-\mathbf{1}_{m-1} \quad \boldsymbol{I}_{m-1}]. \end{aligned} \quad (3)$$

III. LIGHTWEIGHT SOLUTION METHODS

Although the ML estimation problem (2) is non-convex, it can be tackled by the GD method and the LM method. Let us now briefly review these two lightweight methods.

A. Gradient Descent Method

By defining

$$\boldsymbol{\Phi}(\boldsymbol{x}) = \begin{bmatrix} \left(\frac{\boldsymbol{x}-\boldsymbol{a}_1}{\|\boldsymbol{x}-\boldsymbol{a}_1\|_2} - \frac{\boldsymbol{x}-\boldsymbol{a}_2}{\|\boldsymbol{x}-\boldsymbol{a}_2\|_2}\right)^T \\ \vdots \\ \left(\frac{\boldsymbol{x}-\boldsymbol{a}_1}{\|\boldsymbol{x}-\boldsymbol{a}_1\|_2} - \frac{\boldsymbol{x}-\boldsymbol{a}_m}{\|\boldsymbol{x}-\boldsymbol{a}_m\|_2}\right)^T \end{bmatrix}, \quad (4)$$

$$\boldsymbol{d}(\boldsymbol{x}) = \begin{bmatrix} \|\boldsymbol{x}-\boldsymbol{a}_2\|_2 - \|\boldsymbol{x}-\boldsymbol{a}_1\|_2 \\ \|\boldsymbol{x}-\boldsymbol{a}_3\|_2 - \|\boldsymbol{x}-\boldsymbol{a}_1\|_2 \\ \vdots \\ \|\boldsymbol{x}-\boldsymbol{a}_m\|_2 - \|\boldsymbol{x}-\boldsymbol{a}_1\|_2 \end{bmatrix},$$

we can express the gradient of f as

$$\nabla f(\boldsymbol{x}) = \boldsymbol{\Phi}(\boldsymbol{x})^T \boldsymbol{Q}^{-1}(\boldsymbol{d}-\boldsymbol{d}(\boldsymbol{x})). \quad (5)$$

The GD method proceeds by using the steepest descent direction to update the source location estimate; i.e.,

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha_k \nabla f(\boldsymbol{x}^k),$$

where $\alpha_k > 0$ is the step size. One can either use a constant step size or perform a line search to obtain the step size that can yield the largest decrease in the objective value in each iteration. In Section V, we will discuss the choice of the step sizes in greater detail.

B. Levenberg-Marquardt Method

In [8], the authors proposed to use the Levenberg-Marquardt (LM) method to tackle (2). The LM update is based on a damped Gauss-Newton (GN) procedure and is given by

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \left(\boldsymbol{A}^k + \lambda_k \boldsymbol{I}\right)^{-1} \nabla f(\boldsymbol{x}^k),$$

where

$$\boldsymbol{A}^k = \boldsymbol{\Phi}(\boldsymbol{x}^k)^T \boldsymbol{Q}^{-1} \boldsymbol{\Phi}(\boldsymbol{x}^k) \quad (6)$$

and λ_k is a damping parameter that can be set using, e.g., the line-search criterion in [8]. Although the per-iteration complexity of the LM method is higher than that of the GD method, it requires fewer iterations than the GD method to converge, espcially when the source is outside the convex hull of the fixed set of sensors.

C. Initialization

Since the ML estimation problem (2) is non-convex, it is natural to expect that the quality of the solutions produced by the GD and LM methods will depend on the initialization. As it turns out, by using the initialization obtained from the equation-error formulation approach in [1], it is possible to analyze the performance of the GD and LM methods. Let us now briefly describe the equation-error formulation approach.

Consider the auxiliary parameter $R = \|\boldsymbol{x}\|_2$, which is unknown at the beginning. Suppose that $\boldsymbol{a}_1 = \mathbf{0}$. Let $R_i = \|\boldsymbol{a}_i\|_2$. Then, we can rewrite (1) as

$$\boldsymbol{\delta} - 2R\boldsymbol{d} - 2\boldsymbol{S}\boldsymbol{x} = \boldsymbol{\epsilon},$$

where

$$\boldsymbol{\delta} = \begin{bmatrix} R_2^2 - d_2^2 \\ R_3^2 - d_3^2 \\ \vdots \\ R_m^2 - d_m^2 \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{a}_2^T \\ \boldsymbol{a}_3^T \\ \vdots \\ \boldsymbol{a}_m^T \end{bmatrix}, \quad (7)$$

and $\boldsymbol{\epsilon}$ is the residue due to the measurement noise $\boldsymbol{n}$. Given R, the least-squares solution $\boldsymbol{x} = \boldsymbol{x}(R)$ is given by

$$\boldsymbol{x}(R) = \frac{1}{2}\bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}(\boldsymbol{\delta} - 2R\boldsymbol{d}), \quad (8)$$

where $\bar{\boldsymbol{S}} = (\boldsymbol{S}^T\boldsymbol{Q}^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{Q}^{-\frac{1}{2}}$; see [1]. Since $R^2 = \boldsymbol{x}^T\boldsymbol{x}$, equation (8) leads to a quadratic equation in R, which we can solve to get

$$R = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad (9)$$

where

$$a = 4 - 4\boldsymbol{d}^T\bar{\boldsymbol{S}}^T\bar{\boldsymbol{S}}\boldsymbol{d}, \; b = 4\boldsymbol{d}^T\bar{\boldsymbol{S}}^T\bar{\boldsymbol{S}}\boldsymbol{\delta}, \; c = -\boldsymbol{\delta}^T\bar{\boldsymbol{S}}^T\bar{\boldsymbol{S}}\boldsymbol{\delta} \quad (10)$$

(we take the positive root since $R \geq 0$ by definition). We then substitute R into (8) to obtain the initial source location estimate $\boldsymbol{x}^0 = \boldsymbol{x}(R)$.

IV. THEORETICAL ANALYSIS

With the preparations in the previous section, we are now ready to study two issues related to the ML estimation problem (2). First, observe that the ML estimate $\hat{\boldsymbol{x}}$, which is an optimal solution to (2), needs not equal to the true source location $\boldsymbol{x}^\star$ in general. Thus, it is natural to ask whether we can bound the estimation error $\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\|_2$. Second, we are interested in the local growth behavior of the objective function of (2) around an optimal solution. This could then shed light on the local convergence behavior of the GD and LM methods.

A. Estimation Error of an ML Estimator

Let us begin by addressing the first question. Observe that if we replace the noisy range-difference vector $\boldsymbol{d}$ in (8) by the true range-difference vector $\boldsymbol{d}^\star = \boldsymbol{d} - \boldsymbol{n}$, then we have

$$\boldsymbol{x}^\star = \frac{1}{2}\bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}(\boldsymbol{\delta}^\star - 2R^\star\boldsymbol{d}^\star) \quad (11)$$

where $\boldsymbol{\delta}^\star$ and $R^\star$ are defined by (7), (9), and (10) with $\boldsymbol{d}$ replaced by $\boldsymbol{d}^\star$. Using (11) and the definition of $\boldsymbol{x}^0$, we get

$$\begin{aligned}
&\|\boldsymbol{x}^\star - \boldsymbol{x}^0\| \\
&= \left\| \frac{1}{2}\bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}(\boldsymbol{\delta}^\star - \boldsymbol{\delta}) - R^\star \bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}(\boldsymbol{d}^\star - \boldsymbol{d}) \right. \\
&\quad \left. - (R^\star - R)\bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{d} \right\|_2 \\
&\le \left\| \frac{1}{2}\bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}(\boldsymbol{\delta}^\star - \boldsymbol{\delta}) \right\|_2 + \left\| R^\star \bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{n} \right\|_2 \\
&\quad + |R^\star - R| \cdot \left\| \bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{d} \right\|_2 \\
&\le \left\| \frac{1}{2}\bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}\left(2\mathrm{Diag}(\boldsymbol{d}^\star)\boldsymbol{n} + \mathrm{Diag}(\boldsymbol{n})\boldsymbol{n}\right) \right\|_2 \\
&\quad + R^\star \|\bar{\boldsymbol{S}}\|_{\mathrm{op}} \left\| \boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{n} \right\|_2 + |R^\star - R| \cdot \left\| \bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{d} \right\|_2 \\
&\le \left\| \bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}\mathrm{Diag}(\boldsymbol{d}^\star)\boldsymbol{Q}^{\frac{1}{2}} \right\|_{\mathrm{op}} \left\| \boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{n} \right\|_2 + R^\star \|\bar{\boldsymbol{S}}\|_{\mathrm{op}} \left\| \boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{n} \right\|_2 \\
&\quad + \left\| \frac{1}{2}\bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}\mathrm{Diag}(\boldsymbol{n})\boldsymbol{n} \right\|_2 + |R^\star - R| \cdot \left\| \bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{d} \right\|. \quad (12)
\end{aligned}$$

By assumption, the noise vector $\boldsymbol{n}$ is a Gaussian random vector with mean zero and covariance matrix $\sigma^2\boldsymbol{Q}$. Hence, by standard concentration arguments, we have $\|\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{n}\|_2 \le 3\sqrt{m}\sigma$ with high probability. This allows us to bound the first two terms in (12). The third term is bounded by $O(\sigma^2)$. Now, it remains to bound the term $|R^\star - R|$. Since $R^\star$ and R are defined using (9), it suffices to give upper bounds on the terms $|a - a^\star|$, $|b - b^\star|$, and $|c - c^\star|$, which can be obtained by mimicking the derivation of (12). Putting the pieces together, we obtain the following result:

Theorem 1: Let $\boldsymbol{x}^\star$ be the true source location and $\boldsymbol{x}^0$ be defined in Section III-C. Suppose that $\|\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{n}\|_2 \le 3\sqrt{m}\sigma$. Then, there exist a constant $L > 0$, which is determined by $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$, $\boldsymbol{x}^\star$, and $\boldsymbol{Q}$, such that

$$\|\boldsymbol{x}^0 - \boldsymbol{x}^\star\|_2 \le L\sigma.$$

Using Theorem 1, we can bound the estimation error $\|\boldsymbol{x}^\star - \hat{\boldsymbol{x}}\|_2$ of any ML estimator $\hat{\boldsymbol{x}}$.

Corollary 1: Under the same conditions as Theorem 1, we have

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\|_2 \le 2L\sigma.$$

Proof: Let $\hat{\boldsymbol{n}} = \boldsymbol{d} - \boldsymbol{U}\hat{\boldsymbol{h}}$ and $\hat{\boldsymbol{d}} = \boldsymbol{d} - \hat{\boldsymbol{n}}$, where $\hat{\boldsymbol{h}}$ is defined by (3) with the entries $\|\boldsymbol{x} - \boldsymbol{a}_i\|_2$ replaced by $\|\hat{\boldsymbol{x}} - \boldsymbol{a}_i\|_2$. Then, we have

$$\hat{\boldsymbol{x}} = \frac{1}{2}\bar{\boldsymbol{S}}\boldsymbol{Q}^{-\frac{1}{2}}(\hat{\boldsymbol{\delta}} - 2\hat{R}\hat{\boldsymbol{d}}),$$

where $\hat{\boldsymbol{\delta}}$ and $\hat{R}$ are defined by (7), (9), and (10) with $\boldsymbol{d}$ replaced by $\hat{\boldsymbol{d}}$. Since $\hat{\boldsymbol{x}}$ is an optimal solution to (2), we have

$$f(\hat{\boldsymbol{x}}) = \|\boldsymbol{Q}^{-\frac{1}{2}}\hat{\boldsymbol{n}}\|_2^2 \le f(\boldsymbol{x}^\star) = \|\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{n}\|_2^2 \le 9m\sigma^2$$

with high probability. Thus, by using the same argument in the proof of Theorem 1, we get $\|\hat{\boldsymbol{x}} - \boldsymbol{x}^0\|_2 \le L\sigma$, which implies that $\|\hat{\boldsymbol{x}} - \boldsymbol{x}^\star\|_2 \le 2L\sigma$, as desired. ■

B. Local Strong Convexity of the Objective Function

Although the ML estimation problem (2) is non-convex, under suitable assumptions on the noise vector $\boldsymbol{n}$, it can be shown that the objective function is strongly convex around a neighborhood of an optimal solution. To establish this curious result, recall that the gradient of the objective function f is given in (5). Then, the Hessian of f can be computed as

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{\Phi}(\boldsymbol{x})^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{x}) + \boldsymbol{E}(\boldsymbol{x}), \quad (13)$$

where $\boldsymbol{E}(\boldsymbol{x})$ satisfies

$$\|\boldsymbol{E}(\boldsymbol{x})\|_{\mathrm{op}} \le M\|\boldsymbol{Q}^{-1}(\boldsymbol{d} - \boldsymbol{d}(\boldsymbol{x}))\|_2$$

for some $M > 0$ whenever $\boldsymbol{x}$ lies in a bounded set and $\min_{i\in\{1,\dots,m\}} \|\boldsymbol{x} - \boldsymbol{a}_i\|_2 > 0$. In particular, the second term in (13) is small when the noise power σ^2 is small and $\boldsymbol{x}$ close to $\boldsymbol{x}^\star$. This justifies the approximation of the Hessian matrix $\nabla^2 f(\boldsymbol{x})$ by the matrix in (6) in the LM method. Now, observe that

$$\nabla^2 f(\hat{\boldsymbol{x}}) = \boldsymbol{\Phi}(\hat{\boldsymbol{x}})^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\hat{\boldsymbol{x}}) + \boldsymbol{E}(\hat{\boldsymbol{x}})$$

with $\|\boldsymbol{E}(\hat{\boldsymbol{x}})\|_{\mathrm{op}} \le M\|\boldsymbol{Q}^{-\mathbf{1}}\hat{\boldsymbol{n}}\|_2$. Moreover, a simple computation shows that

$$\|\boldsymbol{\Phi}(\hat{\boldsymbol{x}}) - \boldsymbol{\Phi}(\boldsymbol{x}^\star)\|_F \le 2m\|\boldsymbol{x}^\star - \hat{\boldsymbol{x}}\|_2,$$

which implies that

$$\|\boldsymbol{\Phi}(\boldsymbol{x}^\star)^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{x}^\star) - \boldsymbol{\Phi}(\hat{\boldsymbol{x}})^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\hat{\boldsymbol{x}})\|_{\mathrm{op}} \le N\|\boldsymbol{x}^\star - \hat{\boldsymbol{x}}\|_2$$

for some constant $N > 0$. Therefore, we have

$$\begin{aligned}
&\|\nabla^2 f(\hat{\boldsymbol{x}}) - \boldsymbol{\Phi}(\boldsymbol{x}^\star)^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{x}^\star)\|_{\mathrm{op}} \\
\le\quad & M\|\boldsymbol{Q}^{-1}\hat{\boldsymbol{n}}\|_2 + N\|\boldsymbol{x}^\star - \hat{\boldsymbol{x}}\|_2.
\end{aligned}$$

This, together with Corollary 1, gives the following theorem:

Theorem 2: Under the same condition as Theorem 1, we have

$$\|\nabla^2 f(\hat{\boldsymbol{x}}) - \boldsymbol{\Phi}(\boldsymbol{x}^\star)^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{x}^\star)\|_{\mathrm{op}} \le (3M\sqrt{m} + 2NL)\sigma.$$

In particular, we have $\nabla^2 f(\hat{\boldsymbol{x}}) \succ \mathbf{0}$ whenever

$$\sigma < \frac{1}{3M\sqrt{m} + 2NL}\lambda_{\min}(\boldsymbol{\Phi}(\boldsymbol{x}^\star)^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{x}^\star)). \quad (14)$$

Theorem 2 shows that when the noise power is sufficiently small, the Hessian of f at an optimal solution $\hat{\boldsymbol{x}}$ to (2) is positive definite, which implies that f is locally strongly convex at $\hat{\boldsymbol{x}}$. Thus, with proper initialization, the GD method will converge linearly and the LM method will converge quadratically to an optimal solution to (2). This explains in part the numerical observations in [8]. We summarize the above discussion in the following corollary:

Corollary 2: Under the same condition as Theorem 1, suppose that the noise power σ^2 satisfies (14) and the GD and LM methods are initialized by the point $\boldsymbol{x}^0$ defined in Section III-C. Then, the sequence of iterates $\{\boldsymbol{x}^k\}_{k\ge 0}$ generated by the GD (resp. LM) method will converge linearly (resp. quadratically) to an optimal solution $\hat{\boldsymbol{x}}$ to the ML estimation problem (2).

V. Simulations

To verify our theoretical results, we examine in this section the performance of closed-form least-squares estimation in [1] before and after postprocessing by the GD and LM methods. Our goal is to demonstrate that initializing the GD and LM methods using the approach in [1] will yield a good estimate of the source location.

In each simulation, we perform $N = 1000$ Monte Carlo runs in each iteration and compute the mean-squared error (MSE) $(1/N)\sum_{i=1}^{N}\|\boldsymbol{x}^\star - \hat{\boldsymbol{x}}^i\|_2^2$, where $\boldsymbol{x}^\star$ and $\hat{\boldsymbol{x}}^i$ denote the true source position and estimated source position in the i-th run, respectively. Sensors are set along the sides of the square $[-40, 40] \times [-40, 40]$ with

$$\boldsymbol{a}_1 = \begin{bmatrix}40\\0\end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix}0\\40\end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix}-40\\0\end{bmatrix}, \quad \boldsymbol{a}_4 = \begin{bmatrix}0\\-40\end{bmatrix},$$
$$\boldsymbol{a}_5 = \begin{bmatrix}40\\40\end{bmatrix}, \quad \boldsymbol{a}_6 = \begin{bmatrix}40\\-40\end{bmatrix}, \quad \boldsymbol{a}_7 = \begin{bmatrix}-40\\40\end{bmatrix}, \quad \boldsymbol{a}_8 = \begin{bmatrix}-40\\-40\end{bmatrix}.$$

In our experiments, we use the constant step size $\alpha_k = \alpha = \sigma^2/m$ for the GD method. For the LM method, we use the parameters suggested in [8]. We terminate our methods when the norm of the gradient at the current iterate is less than 10^{-4}.

We compare our results with those produced by the SDR method with penalty on the true range-difference matrix [7] and the SDR method with tighter constraints on the geometric structure inside the convex hull of the sensors [6]. The penalty coefficient β in the former method is set to $\beta = 10^{-5}$. We also compare the performance of our method with the CRLB

$$\mathsf{CRLB}(\boldsymbol{x}) = \text{trace}(\boldsymbol{\Phi}(\boldsymbol{x})^T \boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{x}))^{-1},$$

which serves as a statistical lower bound on the performance of any TDOA algorithm [12].

Example 1. We assess the performance of the algorithms when the source lies inside the convex hull of the fixed set of sensors, where the source is located at $\boldsymbol{x}^\star = [30 \quad 10]^T$. In our experiments, the GD and LM methods give similar results. Thus, we shall only present the results for the GD method in this example. Figure 1 shows the MSE performance versus the noise power σ^2, where σ varies from $10^{-1.9}$ to 1. It is observed that the GD method can indeed improve the closed-form least-squares estimator. Moreover, it outperforms the two SDR methods. The performance of the GD method is also close to the CRLB. This demonstrates the strength of the GD method.

Example 2. We assess the performance of the algorithms when the source lies outside the convex hull of the fixed set of sensors, where the source is located at $\boldsymbol{x}^\star = [120 \quad 150]^T$. Figure 2 shows the MSE performance verses the noise power σ^2, where σ varies from $10^{-2.9}$ to 10^{-1}. In this example, we present the results for the LM method due to its faster convergence. As can be seen from the figure, both the closed-form least-squares estimator and the SDR estimators have large MSEs, but the estimator produced by the LM method has a good MSE performance.

Our choice of σ varies from $10^{-2.9}$ to 10^{-1}, which is much smaller than that in the previous example. This is owing to the fact that outside-convex-hull measurements lead to poor conditioning of the matrix $\boldsymbol{\Phi}(\boldsymbol{x})$ in (4), which means that the term $\lambda_{\min}(\boldsymbol{\Phi}(\boldsymbol{x}^\star)^T\boldsymbol{Q}^{-1}\boldsymbol{\Phi}(\boldsymbol{x}^\star))$ becomes much smaller. Thus, we need a smaller σ to guarantee convergence.

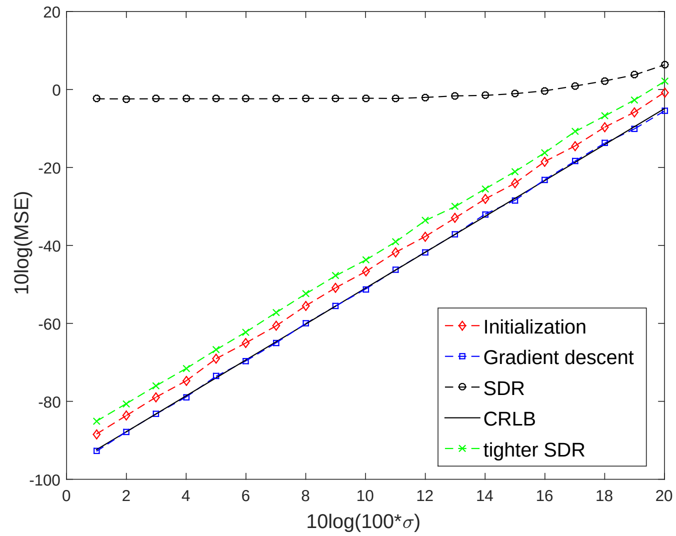

Fig. 1. Performance of the GD method: Source inside the convex hull of sensors

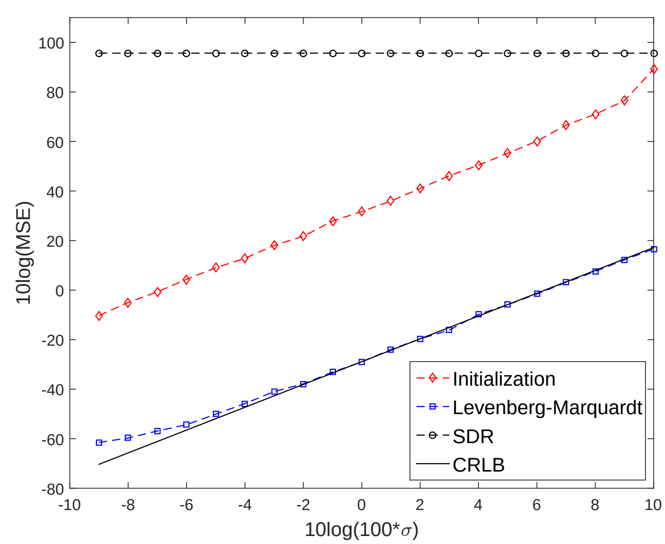

Fig. 2. Performance of the LM method: Source outside the convex hull of sensors

VI. Conclusion

In the paper, we conducted a theoretical analysis of the ML formulation of the TDOA-based source localization problem. Under certain assumptions on the measurement noise, we showed that the GD (resp. LM) method will converge linearly (resp. quadratically) to an optimal solution to the ML estimation problem when properly initialized. We then demonstrated via simulations that the GD and LM methods can indeed produce more accurate estimates of the source location than some existing methods.

References

[1] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 12, pp. 1661–1669, 1987.

[2] A. Bensky, *Wireless Positioning: Technologies and Applications*. Norwood, MA: Artech House, Inc., 2008.

[3] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 41–53, 2005.

[4] 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.

[5] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[6] 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.

[7] Y. Yan, H. Wang, X. Shen, K. He, and X. Zhong, "TDOA-based source collaborative localization via semidefinite relaxation in sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2015, p. Article ID 248970, 2015.

[8] C. Mensing and S. Plass, "Positioning algorithms for cellular networks using TDOA," in *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006)*, 2006, pp. IV–513–IV–516.

[9] J. Sun, Q. Qu, and J. Wright, "When Are Nonconvex Problems Not Scary?" 2015, manuscript, available at http://arxiv.org/abs/1510.06096.

[10] H. Liu, M.-C. Yue, and A. M.-C. So, "On the Statistical Performance and Convergence Rate of the Generalized Power Method for Angular Synchronization," 2017, accepted for publication in *SIAM Journal on Optimization*.

[11] H. Liu, M.-C. Yue, A. M.-C. So, and W.-K. Ma, "A Discrete First-Order Method for Large-Scale MIMO Detection with Provable Guarantees," in *Proceedings of the 18th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2017)*, 2017, pp. 669–673.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice Hall PTR, 1993.