SIAM J. OPTIM.
Vol. 22, No. 1, pp. 108–134

ON THE SOLUTION OF THE GPS LOCALIZATION AND CIRCLE FITTING PROBLEMS*

AMIR BECK† AND DROR PAN†

Abstract. We consider the problem of locating a user's position from a set of noisy pseudoranges to a group of satellites. We consider both the nonlinear *least squares* formulation of the problem, which is nonconvex and nonsmooth, and the nonlinear *squared least squares* variant, in which the objective function is smooth, but still nonconvex. We show that the squared least squares problem can be reformulated as a generalized trust region subproblem and as such can be solved efficiently. Conditions for attainment of the optimal solutions of both problems are derived. The nonlinear least squares problem is shown to have tight connections to the well-known geometric circle fitting and orthogonal regression problems. Finally, a fixed point method for the nonlinear least squares formulation is derived and analyzed.

Key words. GPS localization, nonconvex optimization, nonlinear least squares, generalized trust region subproblem, existence of optimal solutions

AMS subject classifications. 90C26, 90C90

DOI. 10.1137/100809908

1. Introduction.

This paper is concerned with a problem arising in global positioning system (GPS) localization in which measurements are processed to determine a user's position $\mathbf{x} \in \mathbb{R}^n$ (in applications, n is either 2 or 3). GPS consists of a set of satellites transmitting time-stamped signals; the distance measurements are formed by multiplying the difference between the user clock and the transmission time (according to GPS time) with the propagation velocity usually assumed to be the speed of light. The user's clock cannot be assumed to be accurate, but the error, or *bias*, introduced by this clock is the same for all measurements. Distances that are contaminated by the same error caused by the clock bias are also called *pseudoranges*. Mathematically speaking, the pseudoranges $d_1, \ldots, d_m$ are given by the following approximate equations:

$$d_i \approx \|\mathbf{x} - \mathbf{a}_i\| - r, \quad i = 1, \ldots, m, \tag{1.1}$$

where r is the bias caused by the unknown user clock bias (see also [1]) and $\mathbf{a}_1, \ldots, \mathbf{a}_m$ are m vectors in $\mathbb{R}^n$ (in some applications, the satellites' locations). The above equations are only approximate in applications due to errors in measurements that are not related to the "shared" error r; that is, the exact model is in fact

$$d_i = \|\mathbf{x} - \mathbf{a}_i\| - r + \varepsilon_i, \quad i = 1, \ldots, m, \tag{1.2}$$

where for every $i = 1, \ldots, m$, ε_i is an unknown noise corresponding to the ith pseudorange. The problem of estimating $\mathbf{x}$ and r from the set of pseudoranges $d_1, \ldots, d_m$ is the *GPS localization problem.* Usually, in applications, n is set to 3, and in this setting 4 measurements (i.e., satellites) should be sufficient to estimate the position of the user; however, additional measurements can increase the estimation accuracy. A

*Received by the editors September 27, 2010; accepted for publication (in revised form) November 7, 2011; published electronically January 26, 2012.
http://www.siam.org/journals/siopt/22-1/80990.html

†Department of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel (becka@ie.technion.ac.il, drorpan@tx.technion.ac.il).

popular approach is to take a reference satellite, say j, and subtract the jth equation from all other equations. This way, the bias r is eliminated. The arising problem of estimating $\mathbf{x}$ from difference measurements has attracted a lot of attention in the more general context of source localization estimation from range-difference measurements; see, e.g., [2, 11] and the excellent review paper [18]. This indirect approach has several disadvantages: it is not clear how to choose the reference satellite, and the obtained solution might be sensitive to this choice; in addition, subtraction of equations introduces dependencies between the noise components, which are complicated to handle. Another approach, which was considered in [20], is to consider the differences between *all* pairs of equations. The estimate that was devised in [20] is based on a semidefinite relaxation of a maximum likelihood problem involving all these pair-differences.

Surprisingly, the GPS localization problem is closely related to another important and seemingly unrelated problem: circle fitting. Indeed, when $d_i = 0$ for all $i = 1, 2, \ldots, m$ the approximate equations amount to

$$r \approx \|\mathbf{x} - \mathbf{a}_i\|, \quad i = 1, \ldots, m,$$

which is the same as saying that the points $\mathbf{a}_1, \ldots, \mathbf{a}_m$ approximately reside on the circle[1] with center $\mathbf{x}$ and radius r. Circle fitting is an important problem with applications in many areas such as archeology [19], computer graphics [16], coordinate metrology [6], petroleum engineering [10], quality inspection for mechanical parts [13], and statistics [14]. More difficult geometric fitting problems, which are not discussed in this paper, are ellipse fitting problems; see, e.g., [9, 7] and the references therein.

In this paper we are primarily concerned with solving the least squares (LS) problem associated with the approximate equations (1.1). The precise formulation of this nonlinear LS problem is given in section 2, where we also show that when $d_i = 0$ for all $i = 1, \ldots, m$, the problem coincides with the circle geometric fit problem. The resulting problem is nonconvex and nonsmooth and as such is in principal difficult to solve. In particular, iterative methods designed to solve the problem are not guaranteed to converge to a global minimum. This is why finding a good approximate solution is an important task as it can serve as a good starting point of an iterative method. The approximate solution that we consider in section 3 is the solution of a "squared least squares" (SLS) problem associated with the approximate equations (1.1). This problem, although also nonconvex, can be cast as a generalized trust region subproblem [15] for which a global optimal solution can be efficiently obtained under suitable mild conditions, which also guarantee the attainment of the optimal solution. The SLS problem in the circle fitting setting (i.e., $d_i = 0$ for all i) is even simpler; it is shown to be equivalent to a linear LS problem—a generalization of a known result for the two-dimensional case; see [12]. Then, in section 4, we return to the LS problem and provide conditions under which the minimum of the problem is attained, showing also the connection to the so-called orthogonal regression problem. Finally, in section 5, we develop and analyze a fixed point–type method for solving the LS problem, and we conclude with some numerical results illustrating the advantage of the LS approach over the SLS approach.

2. Problem formulation: GPS navigation. We will assume without loss of generality that $\alpha \equiv \min_{i=1,\ldots,m} d_i = 0$. This is not a restrictive assumption since it is always possible to make the change of variables $\tilde{r} = r + \alpha$, define the perturbed

[1] In this paper the term "circle" represents the boundary of an n-dimensional ball (n is not necessarily 2).

measurements $\tilde{d}_i = d_i - \alpha$, and obtain the equivalent model $\tilde{d}_i = \|\mathbf{x} - \mathbf{a}_i\| - \tilde{r} + \varepsilon_i$, $i = 1, \ldots, m$, in which $\min_{i=1,\ldots,m} \tilde{d}_i = 0$.

In this paper we consider the LS approach in which $(\mathbf{x}, r)$ is chosen as the optimal solution of the problem

$$\min_{\mathbf{x},r} \left\{ \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - d_i - r)^2 : r \geq 0 \right\}. \tag{2.1}$$

The constraint $r \geq 0$ is added due to the following reason: in applications, for every i, the magnitude of the error ε_i is significantly smaller than the distance $\|\mathbf{x} - \mathbf{a}_i\|$, which is usually measured in hundreds of kilometers, and it is thus safe to assume that $r + d_i = \|\mathbf{x} - \mathbf{a}_i\| + \varepsilon_i$ is positive. Hence, $r + d_i \geq 0$ for every $i = 1, \ldots, m$ leading to the inequality $r \geq -\min_{i=1,\ldots,m} d_i = 0$. In addition, in the circle fitting problem, r stands for the radius, which is obviously nonnegative. The LS formulation (2.1) also has a statistical meaning: if the noise components ε_i are identical normally distributed zero-mean Gaussian variables, then an optimal solution of (2.1) is a maximum-likelihood estimator of $(\mathbf{x}, r)$.

It is possible to eliminate the decision variable r by fixing $\mathbf{x}$ and minimizing with respect to r. This leads to the following unconstrained and reduced formulation of the problem:

$$\min_{\mathbf{x}} \left\{ \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - d_i - r(\mathbf{x}))^2 \right\}, \tag{GPS$_{\text{ls}}$}$$

where

$$r(\mathbf{x}) := \left[\frac{1}{m} \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - d_i) \right]_+ \tag{2.2}$$

and where for a real number x, $[x]_+ \equiv \max\{x, 0\}$ denotes the nonnegative part of x. Problem (GPS$_{\text{ls}}$) is a nonsmooth nonconvex problem and as such is in principal a difficult problem to solve. A typical randomly generated two-dimensional example illustrating the nonconvexity and nonsmoothness of the problem is given in Figure 1.

An underlying assumption that will be made throughout the paper is that the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ do not reside in a lower-dimensional affine space (that is, a line

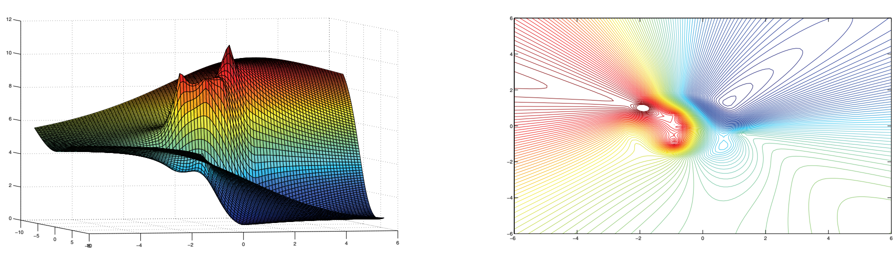

FIG. 1. *Surface and contour plots of problem* (GPS$_{\text{ls}}$) *with* $m = 5$, $n = 2$ *and* $\mathbf{a}_1 = (-0.89, 1.00)^T$, $\mathbf{a}_2 = (-0.62, -0.04)^T$, $\mathbf{a}_3 = (-0.87, 0.63)^T$, $\mathbf{a}_4 = (1.21, -0.42)^T$, $\mathbf{a}_5 = (-1.86, 1.00)^T$, $d_1 = 2.90$, $d_2 = 1.70$, $d_3 = 1.76$, $d_4 = 1.77$, $d_5 = 2.71$.

when $n = 2$ and a plane when $n = 3$). Mathematically, this condition can be written as follows.

Assumption A. The matrix $\tilde{\mathbf{A}}$ defined by

$$\tilde{\mathbf{A}} = \begin{pmatrix} 2\mathbf{a}_1^T & -1 \\ 2\mathbf{a}_2^T & -1 \\ \vdots & \vdots \\ 2\mathbf{a}_m^T & -1 \end{pmatrix} \tag{2.3}$$

has full column rank.

A direct consequence of this assumption is that $m \geq n + 1$.

2.1. The circle fitting problem. When $d_i = 0$ for all $i = 1, \ldots, m$, the problem in its unreduced form is given by

$$\min_{\mathbf{x}, r} \left\{ \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - r)^2 : r \geq 0 \right\},$$

and in its unreduced form by

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) := \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - r(\mathbf{x}))^2 \right\}, \tag{CF$_{\text{ls}}$}$$

where

$$r(\mathbf{x}) := \frac{1}{m} \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|.$$

Note that there is no need to take the nonnegative part of the above expression as it is already nonnegative. Problem (CF$_{\text{ls}}$) is also called the circle "geometric" fit problem (see, e.g., [9]) since it consists of finding a circle that minimizes the sum of squared distances to a certain set of points (here denoted by $\mathbf{a}_1, \ldots, \mathbf{a}_m$):

$$\min_{\mathbf{x}, \mathbf{r}} \sum_{i=1}^{m} d(\mathbf{a}_i, C(\mathbf{x}, r))^2, \tag{2.4}$$

where

$$C(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\| = r\},$$

and $d(\mathbf{y}, S)$ denotes the distance between a point $\mathbf{y}$ and a set S. The equivalence between (2.4) and (CF$_{\text{ls}}$) is made clear by the fact that $d(\mathbf{z}, C(\mathbf{x}, r))^2 = (\|\mathbf{z} - \mathbf{x}\| - r)^2$. An illustration of a geometric fit is given in Figure 2.

3. The SLS localization problem. One of the main objectives of this paper is to analyze problem (GPS$_{\text{ls}}$) and to devise an iterative method for solving it. However, as illustrated, for example, in Figure 1, one of the major difficulties in solving (GPS$_{\text{ls}}$) is its nonconvexity. This is why it is imperative to find a good quality starting point for the devised iterative method. We will therefore proceed now to analyze a different optimization problem, which provides an approximation of the (GPS$_{\text{ls}}$) problem. The main feature of this new problem is that it is tractable; that is, its global optimal solution can be found efficiently.

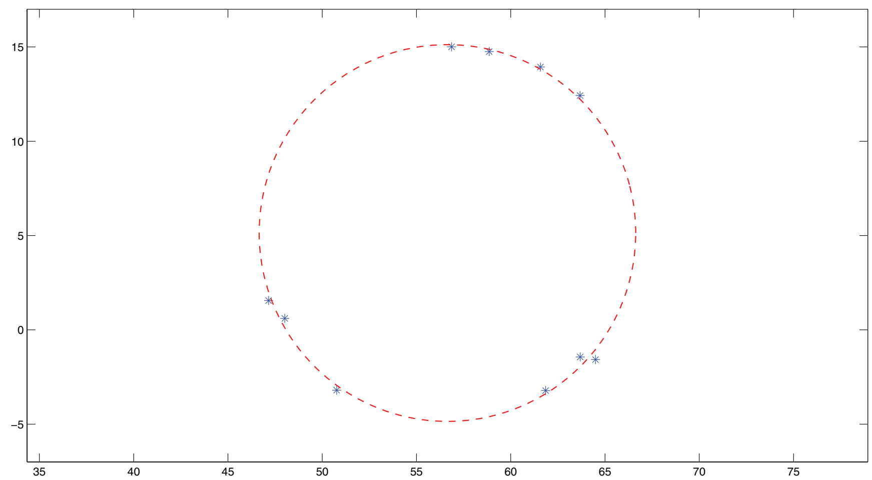

FIG. 2. *The best circle fit (the optimal solution of* $(\mathrm{CF}_{\mathrm{ls}})$*) of* 10 *points denoted by asterisks.*

3.1. Problem formulation. Recall that the model considered in the GPS localization problem consists of solving approximate equations of the form

$$\|\mathbf{x} - \mathbf{a}_i\| \approx r + d_i.$$

Instead of looking directly at an LS approach, let us first square both sides of the approximate equation,

$$\|\mathbf{x} - \mathbf{a}_i\|^2 \approx (r + d_i)^2,$$

and then choose $\mathbf{x}$ and r to minimize the sum of squares of differences between the two sides:

$$(\mathrm{GPS}_{\mathrm{sls}}) : \min \left\{ \sum_{i=1}^{m} \left(\|\mathbf{x} - \mathbf{a}_i\|^2 - (r + d_i)^2 \right)^2 \right\}. \tag{3.1}$$

Note that here we discarded the inequality $r \geq 0$ since the aim is to define a tractable problem, and the sign constraint introduces further difficulties into the problem. From the modeling perspective, this problem possesses several disadvantages in comparison to the LS problem $(\mathrm{GPS}_{\mathrm{ls}})$: (i) $(\mathrm{GPS}_{\mathrm{sls}})$ does not have the same statistical meaning as problem $(\mathrm{GPS}_{\mathrm{ls}})$ and as a result might produce a solution which is worse—in terms of the distance to the "true" solution—than the one provided by $(\mathrm{GPS}_{\mathrm{ls}})$. (ii) In the context of a circle fitting (i.e., $d_i = 0$ for all i) problem, $(\mathrm{GPS}_{\mathrm{sls}})$ is not the *geometric* fitting problem, which in some applications (in particular, those emanating in image processing) is in fact the problem of interest.

The main advantage of the "squared least squares" (SLS) problem $(\mathrm{GPS}_{\mathrm{sls}})$ is that, as opposed to problem $(\mathrm{GPS}_{\mathrm{ls}})$, it is tractable and, as illustrated by the numerical Example 4.1 of section 4.2, often provides a good approximation for the solution of problem $(\mathrm{GPS}_{\mathrm{ls}})$. As such, it is a good candidate for a starting point of iterative methods devised to solve $(\mathrm{GPS}_{\mathrm{ls}})$.

3.2. Formulation of $(\mathrm{GPS_{sls}})$ as a tractable problem. To show that $(\mathrm{GPS_{sls}})$ is indeed tractable, note that it is equivalent to

$$\min_{\mathbf{x},r}\left\{\sum_{i=1}^{m}\left(-2\mathbf{a}_i^T\mathbf{x}-2d_ir+\|\mathbf{x}\|^2-r^2+\|\mathbf{a}_i\|^2-d_i^2\right)^2\right\},$$

which can be rewritten as

$$\min_{\mathbf{x},r,\alpha}\left\{\sum_{i=1}^{m}\left(-2\mathbf{a}_i^T\mathbf{x}-2d_ir+\alpha+\|\mathbf{a}_i\|^2-d_i^2\right)^2 : \alpha=\|\mathbf{x}\|^2-r^2\right\}. \tag{3.2}$$

By denoting

$$\mathbf{B}=\begin{pmatrix}2\mathbf{a}_1^T & -1 & 2d_1\\ 2\mathbf{a}_2^T & -1 & 2d_2\\ \vdots & \vdots & \vdots\\ 2\mathbf{a}_m^T & -1 & 2d_m\end{pmatrix},\qquad \mathbf{b}=\begin{pmatrix}\|\mathbf{a}_1\|^2-d_1^2\\ \|\mathbf{a}_2\|^2-d_2^2\\ \vdots\\ \|\mathbf{a}_m\|^2-d_m^2\end{pmatrix} \tag{3.3}$$

and letting $\mathbf{y}=(\mathbf{x}^T,\alpha,r)^T$, we obtain that an equivalent formulation of (3.2) is

$$\min_{\mathbf{y}\in\mathbb{R}^{n+2}}\left\{\|\mathbf{B}\mathbf{y}-\mathbf{b}\|^2 : \mathbf{y}^T\mathbf{D}\mathbf{y}-2\mathbf{g}^T\mathbf{y}=0\right\}, \tag{3.4}$$

where

$$\mathbf{D}=\begin{pmatrix}\mathbf{I}_n & \mathbf{0}_n & \mathbf{0}_n\\ \mathbf{0}_n^T & 0 & 0\\ \mathbf{0}_n^T & 0 & -1\end{pmatrix},\qquad \mathbf{g}=\frac{1}{2}\begin{pmatrix}\mathbf{0}_n\\ 1\\ 0\end{pmatrix}. \tag{3.5}$$

Problem (3.4) is a *generalized trust region subproblem* (GTRS) [15, 8], that is, a problem consisting of minimizing a quadratic function subject to a single quadratic constraint. Recall that despite its nonconvexity, a GTRS of the form

$$\min\{\mathbf{x}^T\mathbf{A}_1\mathbf{x}+2\mathbf{b}_1^T\mathbf{x}+c_1 : \mathbf{x}^T\mathbf{A}_2\mathbf{x}+2\mathbf{b}_2^T\mathbf{x}+c_2=0\}, \tag{GTRS}$$

where $\mathbf{A}_1=\mathbf{A}_1^T,\mathbf{A}_2=\mathbf{A}_2^T\in\mathbb{R}^{n\times n},\mathbf{b}_1,\mathbf{b}_2\in\mathbb{R}^n,c_1,c_2\in\mathbb{R}$, possesses necessary and sufficient optimality conditions as recalled in the following theorem.

THEOREM 3.1 (see [15, Theorem 3.2]). *Suppose that $\mathbf{A}_2\neq\mathbf{0}$. Then $\mathbf{x}$ is an optimal solution of* (GTRS) *if and only if there exists $\lambda\in\mathbb{R}$ such that*

$$\begin{aligned}(\mathbf{A}_1+\lambda\mathbf{A}_2)\mathbf{x}+(\mathbf{b}_1+\lambda\mathbf{b}_2)&=\mathbf{0},\\ \mathbf{A}_1+\lambda\mathbf{A}_2&\succeq\mathbf{0},\\ \mathbf{x}^T\mathbf{A}_2\mathbf{x}+2\mathbf{b}_2^T\mathbf{x}+c_2&=0.\end{aligned}$$

As a consequence of Theorem 3.1, $\mathbf{y}^*$ is an optimal solution of (3.4) if and only if there exists $\lambda\in\mathbb{R}$ such that

$$(\mathbf{B}^T\mathbf{B}+\lambda\mathbf{D})\mathbf{y}^*=\mathbf{B}^T\mathbf{b}+\lambda\mathbf{g}, \tag{3.6}$$

$$\mathbf{B}^T\mathbf{B}+\lambda\mathbf{D}\succeq\mathbf{0}, \tag{3.7}$$

$$(\mathbf{y}^*)^T\mathbf{D}\mathbf{y}^*-2\mathbf{g}^T\mathbf{y}^*=0. \tag{3.8}$$

Note that the above conditions do not guarantee that the optimal value of (3.4) is attained. We will now derive conditions that guarantee not only the attainment of an optimal solution, but also the existence of a method for efficiently computing the global optimal solution of (3.4).

3.3. Attainment of the solution of (3.4). A known sufficient condition for the attainability of the minimum is the existence of $\hat{\lambda} \in \mathbb{R}$ for which

$$\mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D} \succ \mathbf{0}. \tag{3.9}$$

We are therefore led to discuss sufficient conditions under which (3.9) is satisfied.

THEOREM 3.2. *Let* $\mathbf{d} = (d_1, d_2, \ldots, d_m)^T$ *and let* $\tilde{\mathbf{A}}$ *be given in* (2.3). *The minimum of the GTRS* (3.4) *is attained if at least one of the following conditions is satisfied:*

(i) $\mathbf{d} \notin$ *Range*$(\tilde{\mathbf{A}})$.

(ii) $\mathbf{d} \in$ *Range*$(\tilde{\mathbf{A}})$ *and* $\|\mathbf{w}\| \neq \frac{1}{2}$, *where* $\mathbf{w}$ *consists of the first* n *components of the vector* $(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{d}$.

In addition, if (i) *is satisfied, then* (3.9) *holds true with* $\hat{\lambda} = 0$. *If* (ii) *is satisfied, then* $\|\mathbf{w}\| < \frac{1}{2}$ *implies that there exists an* $\varepsilon_1 > 0$ *for which* (3.9) *holds true for all* $\hat{\lambda} \in (-\varepsilon_1, 0)$, *and if* $\|\mathbf{w}\| > \frac{1}{2}$, *then there exists an* $\varepsilon_2 > 0$ *for which* (3.9) *holds true for all* $\hat{\lambda} \in (0, \varepsilon_2)$.

Proof. Note that

$$\mathbf{B} = (\tilde{\mathbf{A}}, 2\mathbf{d}).$$

By the latter relation, and by Assumption A, which states that the columns of $\tilde{\mathbf{A}}$ are linearly independent, it follows that $\mathbf{B}$ is of full column rank if and only if $\mathbf{d} \notin$ Range$(\tilde{\mathbf{A}})$. Therefore, if (i) is satisfied, then $\mathbf{B}$ is of full column rank and hence (3.9) is satisfied with $\hat{\lambda} = 0$. Now suppose that (ii) is satisfied. Note that

$$\mathbf{D} = \begin{pmatrix} \mathbf{E} & \mathbf{0}_{n+1} \\ \mathbf{0}_{n+1}^T & -1 \end{pmatrix},$$

where $\mathbf{E} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n^T & 0 \end{pmatrix}$. Therefore,

$$\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D} = \begin{pmatrix} \tilde{\mathbf{A}}^T\tilde{\mathbf{A}} + \lambda\mathbf{E} & 2\tilde{\mathbf{A}}^T\mathbf{d} \\ 2\mathbf{d}^T\tilde{\mathbf{A}} & 4\|\mathbf{d}\|^2 - \lambda \end{pmatrix}. \tag{3.10}$$

Recall that by Assumption A the matrix $\tilde{\mathbf{A}}$ has full column rank, and therefore there exists an $\varepsilon > 0$ such that $\tilde{\mathbf{A}}^T\tilde{\mathbf{A}} + \lambda\mathbf{E}$ is positive definite for $\lambda \in I$, where $I = (-\varepsilon, \infty)$. Thus, if $\lambda \in I$, then it follows by Schur complement (see, e.g., [4, Appendix A.5]) that the matrix given by (3.10) is positive definite if and only if

$$g(\lambda) := 4\|\mathbf{d}\|^2 - \lambda - 4\mathbf{d}^T\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}} + \lambda\mathbf{E})^{-1}\tilde{\mathbf{A}}^T\mathbf{d} > 0.$$

Note that in the setting of case (ii) we have

$$g(0) = 4\|\mathbf{d}\|^2 - 4\mathbf{d}^T\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{d} = 0. \tag{3.11}$$

This is due to the fact that the relation $\mathbf{d} \in$ Range$(\tilde{\mathbf{A}})$ implies that

$$\min_{\mathbf{x}} \|\tilde{\mathbf{A}}\mathbf{x} - \mathbf{d}\|^2 = 0, \tag{3.12}$$

but on the other hand, the optimum of (3.12) is attained at $\mathbf{x} = (\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{d}$, implying that the optimal value is also equal to

$$\|\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{d} - \mathbf{d}\|^2 = \|\mathbf{d}\|^2 - 4\mathbf{d}^T\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{d},$$

thus establishing the validity of (3.11). Since $g(0) = 0$, it is sufficient to prove that $g'(0) \neq 0$ to guarantee the existence of an interval of either the form $(-\varepsilon_1, 0)$ or the form $(0, \varepsilon_2)$ (where $\varepsilon_1, \varepsilon_2 > 0$) on which g is positive. A rather tedious technical computation shows that

$$g'(0) = -1 + 4\mathbf{d}^T\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\mathbf{E}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{d},$$

and therefore we conclude that if $\beta \equiv \|\mathbf{E}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{d}\| \neq \frac{1}{2}$, then $g'(0) \neq 0$. Note that the vector $\mathbf{w}$ defined in the premise of the theorem is equal to $\mathbf{E}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{d}$ so that $\beta = \|\mathbf{w}\|$. We can thus conclude that if $\beta < \frac{1}{2}$, then there exists $\varepsilon_1 > 0$ such that (3.9) is satisfied for $\hat{\lambda} \in (-\varepsilon_1, 0)$, and if $\beta > \frac{1}{2}$, then there exists $\varepsilon_2 > 0$ for which (3.9) is satisfied for $\hat{\lambda} \in (0, \varepsilon_2)$. ◻

The fact that problem ($\mathrm{GPS_{sls}}$) can be recast as the GTRS (3.4) immediately implies the following corollary.

COROLLARY 3.3 (attainment of the optimal solution of ($\mathrm{GPS_{sls}}$)). *If condition* (i) *or* (ii) *of Theorem* 3.2 *is satisfied, then the minimum of problem* ($\mathrm{GPS_{sls}}$) *given in* (3.1) *is attained.*

Note that in a sense the conditions of Theorem 3.2 are rather mild, although it is of course always possible to tailor special examples in which these conditions are not satisfied. Indeed, recall that the relation $m \geq n+1$ holds. If $m \geq n+2$, then $\tilde{\mathbf{A}}$ is a tall matrix (i.e., the number of rows is greater than the number of columns) and, at least for randomly generated instances, the relation $\mathbf{d} \notin \mathrm{Range}(\tilde{\mathbf{A}})$ is likely to be satisfied. If $m = n+1$, then $\tilde{\mathbf{A}}$ is square, which, combined with the fact that it has a full column rank, implies that it is nonsingular. Therefore, $\mathbf{d} \in \mathrm{Range}(\tilde{\mathbf{A}})$. However, in this case, condition (ii) is expected to be satisfied since there is no particular reason, at least in applications, why $\|\mathbf{w}\|$ will be *exactly* equal to $\frac{1}{2}$.

If $\mathbf{d} \notin \mathrm{Range}(\tilde{\mathbf{A}})$, then $\hat{\lambda} = 0$ satisfies (3.9). If $\mathbf{d} \in \mathrm{Range}(\tilde{\mathbf{A}})$ and $\|\mathbf{w}\| \neq \frac{1}{2}$, then the following line search procedure will find a $\hat{\lambda}$ satisfying (3.9).

A line search procedure for finding $\hat{\boldsymbol{\lambda}}$ satisfying (3.9).

Initialization: Set $\hat{\lambda} = \mathrm{sgn}\,(\|\mathbf{w}\| - \frac{1}{2})$.
Step 1: Compute $\mathbf{G} = \mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D}$.
Step 2: If $\mathbf{G} \succ \mathbf{0}$, then STOP. Otherwise, set $\hat{\lambda} = \frac{\hat{\lambda}}{2}$ and return to Step 1.

3.4. Finding a global optimal solution of ($\mathbf{GPS_{sls}}$). By using the necessary and sufficient optimality conditions for GTRS problems, it is possible to construct a dual-based method for finding the global optimal solution of problem ($\mathrm{GPS_{sls}}$). To do so, we first write explicitly the interval in which λ satisfies the semidefinite constraint (3.7). We use the following notation: for an $N \times N$ symmetric matrix $\mathbf{A}$ and an $N \times N$ positive definite matrix $\mathbf{B}$, the generalized eigenvalues of the matrix pair $\mathbf{A}, \mathbf{B}$ are

$$\lambda_1(\mathbf{A}; \mathbf{B}) \geq \lambda_2(\mathbf{A}; \mathbf{B}) \geq \cdots \geq \lambda_N(\mathbf{A}; \mathbf{B}).$$

LEMMA 3.1. *Let $\hat{\lambda}$ satisfy* (3.9). *Then $\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D} \succeq \mathbf{0}$ if and only if*

$$\mu_1 \leq \lambda \leq \mu_2, \tag{3.13}$$

where

$$\mu_1 = \hat{\lambda} - \frac{1}{\lambda_1(\mathbf{D}; \mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D})},$$
$$\mu_2 = \hat{\lambda} - \frac{1}{\lambda_{n+2}(\mathbf{D}; \mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D})}.$$

Proof. The relation $\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D} \succeq \mathbf{0}$ holds if and only if

$$\mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D} + (\lambda - \hat{\lambda})\mathbf{D} \succeq \mathbf{0}.$$

After multiplication by $(\mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D})^{-1/2}$ from the left and the right, the latter linear matrix inequality reads

$$\mathbf{I} + (\lambda - \hat{\lambda})(\mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D})^{-1/2}\mathbf{D}(\mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D})^{-1/2} \succeq \mathbf{0},$$

which is equivalent to

$$1 + (\lambda - \hat{\lambda})\lambda_i(\mathbf{D}; \mathbf{B}^T\mathbf{B} + \hat{\lambda}\mathbf{D}) \geq 0, \quad i = 1, \dots, n+2.$$

The latter inequality is the same as (3.13). ∎

We will assume that the so-called hard case does not occur; that is, we will assume that

$$\mathbf{B}^T\mathbf{b} + \mu_i\mathbf{g} \notin \text{Range}(\mathbf{B}^T\mathbf{B} + \mu_i\mathbf{D}), \quad i = 1, 2. \tag{3.14}$$

If (3.14) is indeed satisfied, then condition (3.6) will not be satisfied for $\lambda = \mu_1$ or $\lambda = \mu_2$. Therefore, the optimal λ will reside in the open interval (μ_1, μ_2) in which $\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D} \succ \mathbf{0}$. The optimal solution of the GTRS (3.4), $\mathbf{y}^*$, will be of the form $\mathbf{y}^* = \mathbf{y}(\lambda^*)$, where

$$\mathbf{y}(\lambda) \equiv (\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D})^{-1}(\mathbf{B}^T\mathbf{b} + \lambda\mathbf{g}),$$

and where λ^* is the unique root of the function

$$\phi(\lambda) \equiv \mathbf{y}(\lambda)^T\mathbf{D}\mathbf{y}(\lambda) - 2\mathbf{g}^T\mathbf{y}(\lambda), \tag{3.15}$$

which is strictly decreasing function over I (a proof of this property can be found in [15]).

The above discussion is summarized in the following theorem.

THEOREM 3.4. *Let* $\mathbf{B}, \mathbf{D}, \mathbf{b}, \mathbf{g}$ *be as defined in* (3.3) *and* (3.5) *and assume that* (3.14) *is satisfied. Then the global optimal solution of problem* (GPS$_{\text{sls}}$) *is composed of the first n components of the vector*

$$(\mathbf{B}^T\mathbf{B} + \lambda^*\mathbf{D})^{-1}(\mathbf{B}^T\mathbf{b} + \lambda^*\mathbf{g}),$$

where λ^* is the unique root over (μ_1, μ_2) of the strictly decreasing function ϕ defined in (3.15).

Thus, in order to solve the GTRS (3.4), it is required to invoke some kind of a root-finding procedure, such as bisection, in order to locate the optimal dual variable, and thus also the optimal solution of problem (GPS$_{\text{sls}}$).

3.5. The circle fitting SLS problem. When $d_i = 0$ for all i, problem (GPS$_{\text{sls}}$) reduces to the circle fitting SLS problem given by

$$\text{(CF}_{\text{sls}}\text{)} \qquad \min_{\mathbf{x},r} \left\{ \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\|^2 - r^2)^2 : \mathbf{x} \in \mathbb{R}^n, r \in \mathbb{R} \right\}.$$

As a special case of problem (GPS$_{\text{sls}}$), it can be cast as a GTRS and solved using the method described in the previous section. However, it is known that in the two-dimensional case ($n = 2$), this problem [12, 5] is in fact equivalent to a *linear LS problem*. In this section we generalize this result and show that problem (CF$_{\text{sls}}$) can be recast as a linear LS problem for a general n. We begin by noting that problem (CF$_{\text{sls}}$) is the same as

$$\min_{\mathbf{x},r} \left\{ \sum_{i=1}^{m} (-2\mathbf{a}_i^T \mathbf{x} + \|\mathbf{x}\|^2 - r^2 + \|\mathbf{a}_i\|^2)^2 : \mathbf{x} \in \mathbb{R}^n, r \in \mathbb{R} \right\}. \tag{3.16}$$

Making the change of variables $R = \|\mathbf{x}\|^2 - r^2$, the above problem reduces to

$$\min_{\mathbf{x},R} \left\{ \sum_{i=1}^{m} (-2\mathbf{a}_i^T \mathbf{x} + R + \|\mathbf{a}_i\|^2)^2 : \|\mathbf{x}\|^2 \geq R \right\}. \tag{3.17}$$

Note that the change of variables imposes an additional relation between the variables that is given by the constraint $\|\mathbf{x}\|^2 \geq R$. We will show that in fact this constraint can be dropped; that is, problem (3.17) is equivalent to the linear LS problem

$$\min_{\mathbf{x},R} \left\{ \sum_{i=1}^{m} (-2\mathbf{a}_i^T \mathbf{x} + R + \|\mathbf{a}_i\|^2)^2 \right\}. \tag{3.18}$$

Indeed, any optimal solution $(\hat{\mathbf{x}}, \hat{R})$ of (3.18) automatically satisfies $\|\hat{\mathbf{x}}\|^2 \geq \hat{R}$ since otherwise, if $\|\hat{\mathbf{x}}\|^2 < \hat{R}$, then we have

$$-2\mathbf{a}_i^T \hat{\mathbf{x}} + \hat{R} + \|\mathbf{a}_i\|^2 > -2\mathbf{a}_i^T \hat{\mathbf{x}} + \|\hat{\mathbf{x}}\|^2 + \|\mathbf{a}_i\|^2 = \|\hat{\mathbf{x}} - \mathbf{a}_i\|^2 \geq 0, \quad i = 1, \ldots, m.$$

Squaring both sides of the first inequality in the above and summing over i yields

$$\sum_{i=1}^{m} \left(-2\mathbf{a}_i^T \hat{\mathbf{x}} + \hat{R} + \|\mathbf{a}_i\|^2\right)^2 > \sum_{i=1}^{m} \left(-2\mathbf{a}_i^T \hat{\mathbf{x}} + \|\hat{\mathbf{x}}\|^2 + \|\mathbf{a}_i\|^2\right)^2,$$

showing that $(\hat{\mathbf{x}}, \|\hat{\mathbf{x}}\|^2)$ gives a lower function value than $(\hat{\mathbf{x}}, \hat{R})$ in contradiction to the optimality of $(\hat{\mathbf{x}}, R)$. To conclude, problem (CF$_{\text{sls}}$) is equivalent to the LS problem (3.18) that can also be written as

$$\min_{\mathbf{y} \in \mathbb{R}^{n+1}} \|\tilde{\mathbf{A}}\mathbf{y} - \mathbf{b}\|^2, \tag{3.19}$$

where $\mathbf{y} = (\mathbf{x}^T, R)^T$, $\tilde{\mathbf{A}}$ is given in (2.3), and $\mathbf{b} = (\|\mathbf{a}_1\|^2, \ldots, \|\mathbf{a}_m\|^2)^T$. Since, by Assumption A, $\tilde{\mathbf{A}}$ is of full column rank, it follows that the unique optimal solution is

$$\mathbf{y} = (\tilde{\mathbf{A}}^T \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^T \mathbf{b}.$$

The optimal $\mathbf{x}$ is given by the first n components of $\mathbf{y}$, and the radius r is given by

$$r = \sqrt{\|\mathbf{x}\|^2 - R},$$

where R is the last (i.e., $(n+1)$th) component of $\mathbf{y}$. We summarize the above discussion in the following theorem.

THEOREM 3.5. *Let* $\mathbf{y} = (\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T\mathbf{b}$, *where* $\tilde{\mathbf{A}}$ *is given in* (2.3). *Then the optimal solution of problem* ($\mathrm{CF}_{\mathrm{sls}}$) *is given by* $(\hat{\mathbf{x}}, \hat{r})$, *where* $\hat{\mathbf{x}}$ *consists of the first* n *components of* $\mathbf{y}$ *and* $r = \sqrt{\|\hat{\mathbf{x}}\|^2 - y_{n+1}}$.

4. Attainment of the optimal solution of problem ($\mathbf{GPS_{ls}}$). Problem ($\mathrm{GPS}_{\mathrm{ls}}$) is more difficult than problem ($\mathrm{GPS}_{\mathrm{sls}}$) in the sense that it is also nonsmooth in addition to being nonconvex, and indeed it does not seem to be possible, as it was for problem ($\mathrm{GPS}_{\mathrm{sls}}$), to find an equivalent formulation of ($\mathrm{GPS}_{\mathrm{ls}}$) as a tractable problem. The objective function of ($\mathrm{GPS}_{\mathrm{ls}}$) is not coercive, and it might be that the optimal solution is not attained. We therefore begin with establishing conditions under which attainment of the optimal solution is guaranteed.

4.1. The liminf of the objective function of ($\mathbf{GPS_{ls}}$). The following lemma, bounding $r(\mathbf{x})$ (see (2.2)), will be useful in what follows.

LEMMA 4.1. *Let*

$$N = \frac{1}{m}\sum_{i=1}^{m}(\|\mathbf{a}_i\| + d_i). \tag{4.1}$$

Then for every $\mathbf{x} \in \mathbb{R}^n$ *satisfying* $\|\mathbf{x}\| \geq N$ *the inequality*

$$\big|\|\mathbf{x}\| - r(\mathbf{x})\big| \leq N \tag{4.2}$$

is satisfied.

Proof. Let $\mathbf{x} \in \mathbb{R}^n$ satisfy $\|\mathbf{x}\| \geq N$. Then

$$\sum_{i=1}^{m}(\|\mathbf{x}-\mathbf{a}_i\|-d_i) \geq \sum_{i=1}^{m}(\|\mathbf{x}\|-\|\mathbf{a}_i\|-d_i) = m\|\mathbf{x}\|-\sum_{i=1}^{m}(\|\mathbf{a}_i\|+d_i) \geq mN-\sum_{i=1}^{m}(\|\mathbf{a}_i\|+d_i) = 0.$$

Therefore,

$$r(\mathbf{x}) = \frac{1}{m}\sum_{i=1}^{m}(\|\mathbf{x} - \mathbf{a}_i\| - d_i) \geq 0.$$

Now, on the one hand,

$$r(\mathbf{x}) = \frac{1}{m}\sum_{i=1}^{m}(\|\mathbf{x} - \mathbf{a}_i\| - d_i) \leq \frac{1}{m}\sum_{i=1}^{m}(\|\mathbf{x}\| + \|\mathbf{a}_i\| - d_i),$$

showing that

$$r(\mathbf{x}) - \|\mathbf{x}\| \leq \frac{1}{m}\sum_{i=1}^{m}(\|\mathbf{a}_i\| - d_i) \leq \frac{1}{m}\sum_{i=1}^{m}(\|\mathbf{a}_i\| + d_i). \tag{4.3}$$

On the other hand,

$$r(\mathbf{x}) = \frac{1}{m}\sum_{i=1}^{m}(\|\mathbf{x} - \mathbf{a}_i\| - d_i) \geq \frac{1}{m}\sum_{i=1}^{m}(\|\mathbf{x}\| - \|\mathbf{a}_i\| - d_i),$$

implying the inequality

$$r(\mathbf{x}) - \|\mathbf{x}\| \geq \frac{1}{m}\sum_{i=1}^{m}(-\|\mathbf{a}_i\| - d_i). \tag{4.4}$$

Combining (4.3) and (4.4), inequality (4.2) follows. □

A direct result of (4.2) is the following (soon-to-be) useful limit:

$$\lim_{\|\mathbf{x}\|\to\infty} \frac{r(\mathbf{x})}{\|\mathbf{x}\|} = 1. \tag{4.5}$$

The main result, which will be used in order to establish conditions for attainment of the optimal solution, is that the liminf of the objective function as $\|\mathbf{x}\| \to \infty$ of problem (GPS_{ls}) can be computed. In particular, if we denote the objective function of (GPS_{ls}) by

$$f(\mathbf{x}) \equiv \sum_{i=1}^{m}(\|\mathbf{x} - \mathbf{a}_i\| - d_i - r(\mathbf{x}))^2, \tag{4.6}$$

then we will prove that

$$\liminf_{\|\mathbf{x}\|\to\infty} f(\mathbf{x}) = f_{\text{liminf}}, \tag{4.7}$$

where

$$f_{\text{liminf}} \equiv \min_{\mathbf{z}} \left\{ (\mathbf{A}\mathbf{z} + \mathbf{d})^T \left(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T \right) (\mathbf{A}\mathbf{z} + \mathbf{d}) : \|\mathbf{z}\| = 1 \right\}, \tag{4.8}$$

and where $\mathbf{1}_m$ denotes a column vector of m ones and $\mathbf{A}$ is the matrix defined by

$$\mathbf{A} := \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_m^T \end{pmatrix}. \tag{4.9}$$

In this notation, the connection between $\mathbf{A}$ and the matrix $\tilde{\mathbf{A}}$ defined in (2.3) is given by

$$\tilde{\mathbf{A}} = (2\mathbf{A}, -\mathbf{1}_m).$$

Lemma 4.2 below shows that

$$\liminf_{\|\mathbf{x}\|\to\infty} f(\mathbf{x}) \leq f_{\text{liminf}}$$

by detecting a sequence $\{\mathbf{x}_k\}$ for which $f(\mathbf{x}_k)$ converges to f_{liminf}.

LEMMA 4.2. *Let $\mathbf{z}$ be an optimal solution of the problem* (4.8)*. Then the sequence defined by*

$$\mathbf{x}_k = k\mathbf{z}, \quad k = 1, 2, \ldots,$$

satisfies $\|\mathbf{x}_k\| \to \infty$ as $k \to \infty$ and

$$\lim_{k\to\infty} f(\mathbf{x}_k) = f_{\text{liminf}}. \tag{4.10}$$

Proof. Let $k > N$ (where N is given in (4.1)). Then $r(\mathbf{x}_k) = \frac{1}{m}\sum_{i=1}^m (\|\mathbf{x}_k - \mathbf{a}_i\| - d_i)$ and thus for every $i = 1, \ldots, m$ it holds that

$$\|\mathbf{x}_k - \mathbf{a}_i\| - d_i - r(\mathbf{x}_k) = \|k\mathbf{z} - \mathbf{a}_i\| - d_i - r(k\mathbf{z})$$

$$= \frac{\|\mathbf{z} - \frac{1}{k}\mathbf{a}_i\| - \frac{d_i}{k} - \frac{1}{m}\sum_{j=1}^m \left(\|\mathbf{z} - \frac{1}{k}\mathbf{a}_j\| - \frac{d_j}{k}\right)}{\frac{1}{k}} \tag{4.11}$$

$$= g_i\left(\frac{1}{k}\right), \tag{4.12}$$

where

$$g_i(\alpha) \equiv \frac{\|\mathbf{z} - \alpha\mathbf{a}_i\| - d_i\alpha - \frac{1}{m}\sum_{j=1}^m \|\mathbf{z} - \alpha\mathbf{a}_j\| + \bar{d}\alpha}{\alpha}$$

and

$$\bar{d} = \frac{1}{m}\sum_{j=1}^m d_j.$$

Since the denominator and the nominator in $g(\alpha)$ both converge to 0 as $\alpha \to 0^+$, it follows by l'Hôpital's rule that

$$\lim_{\alpha\to 0^+} g_i(\alpha) = \lim_{\alpha\to 0^+}\left[\frac{-\mathbf{a}_i^T[\mathbf{z} - \alpha\mathbf{a}_i] - d_i + \bar{d}}{\|\mathbf{z} - \alpha\mathbf{a}_i\|} - \frac{1}{m}\sum_{j=1}^m \frac{-\mathbf{a}_j^T[\mathbf{z} - \alpha\mathbf{a}_j]}{\|\mathbf{z} - \alpha\mathbf{a}_j\|}\right]$$
$$= -\left[\mathbf{a}_i^T\mathbf{z} + d_i - \frac{1}{m}\sum_{j=1}^m (\mathbf{a}_j^T\mathbf{z} + d_j)\right].$$

Therefore, by plugging $\alpha = \frac{1}{k}$ into the above limit and using (4.12), it follows that

$$\lim_{k\to\infty} \|\mathbf{x}_k - \mathbf{a}_i\| - d_i - r(\mathbf{x}_k) = -\left(\mathbf{a}_i^T\mathbf{z} + d_i - \frac{1}{m}\sum_{j=1}^m (\mathbf{a}_j^T\mathbf{z} + d_j)\right),$$

and thus

$$\lim_{k\to\infty} f(\mathbf{x}_k) = \sum_{i=1}^m \left(\mathbf{a}_i^T\mathbf{z} + d_i - \frac{1}{m}\sum_{j=1}^m (\mathbf{a}_j^T\mathbf{z} + d_j)\right)^2$$
$$= \left[(\mathbf{A}\mathbf{z} + \mathbf{d})^T\left(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T\right)(\mathbf{A}\mathbf{z} + \mathbf{d})\right].$$

Finally, since $\mathbf{z}$ is an optimal solution of (4.8), the result (4.10) follows. ▯

We have thus shown that $\liminf_{\|\mathbf{x}\|\to\infty} f(\mathbf{x}) \le f_{\text{liminf}}$. To show the reverse inequality, we will require the following technical lemma.

LEMMA 4.3. *There exist continuous functions $A(\mathbf{x})$ and $C(\mathbf{x})$ satisfying*

$$\lim_{\|\mathbf{x}\|\to\infty} A(\mathbf{x}) = 1, \qquad \lim_{\|\mathbf{x}\|\to\infty} C(\mathbf{x}) = 0$$

for which

$$f(\mathbf{x}) \ge A(\mathbf{x}) f_{\text{liminf}} + C(\mathbf{x})$$

for every $\|\mathbf{x}\| > \max\{1, N\}$, where N is given by (4.1).

Proof. First note that

$$
\begin{aligned}
f(\mathbf{x}) &= \sum_{i=1}^{m}(\|\mathbf{x}-\mathbf{a}_i\| - r(\mathbf{x}) - d_i)^2 \\
&= \sum_{i=1}^{m}\left(\frac{\|\mathbf{x}-\mathbf{a}_i\|^2 - r^2(\mathbf{x}) - 2d_i r(\mathbf{x}) - d_i^2}{\|\mathbf{x}-\mathbf{a}_i\| + r(\mathbf{x}) + d_i}\right)^2 \\
&= \sum_{i=1}^{m}\left(\frac{-2\mathbf{a}_i^T\mathbf{x} + \|\mathbf{x}\|^2 - r^2(\mathbf{x}) - 2d_i r(\mathbf{x}) - d_i^2 + \|\mathbf{a}_i\|^2}{\|\mathbf{x}-\mathbf{a}_i\| + r(\mathbf{x}) + d_i}\right)^2 \\
&= \sum_{i=1}^{m}\left(\frac{-2\mathbf{a}_i^T\mathbf{x} + \|\mathbf{x}\|^2 - r^2(\mathbf{x}) - 2d_i r(\mathbf{x}) - d_i^2 + \|\mathbf{a}_i\|^2}{2\|\mathbf{x}\|}\right)^2 \left(\frac{2\|\mathbf{x}\|}{\|\mathbf{x}-\mathbf{a}_i\| + r(\mathbf{x}) + d_i}\right)^2 \\
&\geq A(\mathbf{x})B(\mathbf{x}),
\end{aligned}
$$

where

$$
A(\mathbf{x}) \equiv \min_{i=1,\ldots,m}\left\{\left(\frac{2\|\mathbf{x}\|}{\|\mathbf{x}-\mathbf{a}_i\| + r(\mathbf{x}) + d_i}\right)^2\right\}, \tag{4.13}
$$

$$
B(\mathbf{x}) \equiv \sum_{i=1}^{m}\left(\frac{-2\mathbf{a}_i^T\mathbf{x} + \|\mathbf{x}\|^2 - r^2(\mathbf{x}) - 2d_i r(\mathbf{x}) - d_i^2 + \|\mathbf{a}_i\|^2}{2\|\mathbf{x}\|}\right)^2. \tag{4.14}
$$

The denominator in (4.13) is positive since by Lemma 4.1 the inequality $\|\mathbf{x}\| > N$ implies that $r(\mathbf{x}) > 0$. By (4.5) it follows that

$$
\lim_{\|\mathbf{x}\|\to\infty} A(\mathbf{x}) = 1.
$$

In addition, $B(\mathbf{x})$ can be rewritten as

$$
B(\mathbf{x}) = \sum_{i=1}^{m}\left(\mathbf{a}_i^T\frac{\mathbf{x}}{\|\mathbf{x}\|} + d_i - \frac{\|\mathbf{x}\|^2 - r^2(\mathbf{x})}{2\|\mathbf{x}\|} + \alpha_i(\mathbf{x})\right)^2,
$$

where

$$
\alpha_i(\mathbf{x}) \equiv \frac{2d_i(r(\mathbf{x}) - \|\mathbf{x}\|) + d_i^2 - \|\mathbf{a}_i\|^2}{2\|\mathbf{x}\|}.
$$

Now,

$$
\begin{aligned}
B(\mathbf{x}) = &\overbrace{\sum_{i=1}^{m}\left(\mathbf{a}_i^T\frac{\mathbf{x}}{\|\mathbf{x}\|} + d_i - \frac{\|\mathbf{x}\|^2 - r^2(\mathbf{x})}{2\|\mathbf{x}\|}\right)^2}^{B_1(\mathbf{x})} + \overbrace{\sum_{i=1}^{m}\alpha_i^2(\mathbf{x})}^{B_2(\mathbf{x})} \\
&\underbrace{+\,2\sum_{i=1}^{m}\left(\mathbf{a}_i^T\frac{\mathbf{x}}{\|\mathbf{x}\|} + d_i - \frac{\|\mathbf{x}\|^2 - r^2(\mathbf{x})}{2\|\mathbf{x}\|}\right)\alpha_i(\mathbf{x})}_{B_3(\mathbf{x})}.
\end{aligned}
$$

The first term in the above summation can be bounded below as follows:

$$
B_1(\mathbf{x}) \geq \min_y\left\{\sum_{i=1}^{m}\left(\mathbf{a}_i^T\mathbf{z} + d_i - y\right)^2\right\},
$$

where $\mathbf{z} = \mathbf{x}/\|\mathbf{x}\|$. The optimal solution of the above optimization problem is $y = \frac{1}{m}\sum_{i=1}^m(\mathbf{a}_i^T\mathbf{z} + d_i)$, which corresponds to the optimal value

$$(\mathbf{Az}+\mathbf{d})^T\left(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T\right)(\mathbf{Az}+\mathbf{d}).$$

Therefore, since $\|\mathbf{z}\| = \|\mathbf{x}/\|\mathbf{x}\|\| = 1$, it follows that

$$B_1(\mathbf{x}) \geq \min_{\mathbf{z}}\left\{(\mathbf{Az}+\mathbf{d})^T\left(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T\right)(\mathbf{Az}+\mathbf{d}) : \|\mathbf{z}\| = 1\right\} = f_{\text{liminf}}.$$

Since $\|\mathbf{x}\| \geq N$, we have by Lemma 4.1 that (4.2) holds, implying that $|\alpha_i(\mathbf{x})| \leq \frac{2d_iN+d_i^2+\|\mathbf{a}_i\|^2}{2\|\mathbf{x}\|}$, which in turn implies that

$$\alpha_i(\mathbf{x}) \to 0 \quad \text{as } \|\mathbf{x}\| \to \infty \tag{4.15}$$

and hence

$$\lim_{\|\mathbf{x}\|\to\infty} B_2(\mathbf{x}) = 0.$$

Finally, note that by (4.2) it follows that

$$\left|\frac{\|\mathbf{x}\|^2 - r^2(\mathbf{x})}{\|\mathbf{x}\|}\right| = \frac{|\|\mathbf{x}\| - r(\mathbf{x})| \cdot |\|\mathbf{x}\| + r(\mathbf{x})|}{\|\mathbf{x}\|} \leq N\frac{\|\mathbf{x}\| + r(\mathbf{x})}{\|\mathbf{x}\|} \leq N\frac{2\|\mathbf{x}\|+N}{\|\mathbf{x}\|} \leq 2N+N^2,$$

where the last inequality uses the fact that $\|\mathbf{x}\| \geq 1$. We can now conclude that

$$\begin{aligned}|B_3(\mathbf{x})| &\leq 2\sum_{i=1}^m\left|\left(\mathbf{a}_i^T\frac{\mathbf{x}}{\|\mathbf{x}\|} + d_i - \frac{\|\mathbf{x}\|^2 - r^2(\mathbf{x})}{2\|\mathbf{x}\|}\right)\right| \cdot |\alpha_i(\mathbf{x})| \\ &= 2\sum_{i=1}^m\left(\|\mathbf{a}_i\| + |d_i| + N + \frac{N^2}{2}\right)|\alpha_i(\mathbf{x})|,\end{aligned}$$

which combined with (4.15) implies that $B_3(\mathbf{x}) \to 0$ as $\|\mathbf{x}\| \to \infty$. The result then follows with $A(\mathbf{x})$ defined in (4.13) and $C(\mathbf{x}) \equiv A(\mathbf{x})(B_2(\mathbf{x}) + B_3(\mathbf{x}))$. ☐

The direct result of Lemmas 4.2 and 4.3 is the following proposition stating that f_{liminf} is indeed the liminf of $f(\mathbf{x})$ as $\|\mathbf{x}\| \to \infty$.

PROPOSITION 4.1. $\liminf_{\|\mathbf{x}\|\to\infty} f(\mathbf{x}) = f_{\text{liminf}}$.

Proof. By Lemma 4.2 it follows that $\liminf_{\|\mathbf{x}\|\to\infty} f(\mathbf{x}) \leq f_{\text{liminf}}$. To prove the reverse inequality, take a sequence $\{\mathbf{x}_k\}$ for which $\|\mathbf{x}_k\| \to \infty$, satisfying that the sequence $\{f(\mathbf{x}_k)\}$ converges to a finite value. Let $A(\cdot)$ and $C(\cdot)$ be as defined in Lemma 4.3. Then $A(\mathbf{x}_k) \to 1$ and $C(\mathbf{x}_k) \to 0$ as $k \to \infty$ and thus

$$\lim_{k\to\infty} f(\mathbf{x}_k) \geq \lim_{k\to\infty}\{A(\mathbf{x}_k)f_{\text{liminf}} + C(\mathbf{x}_k)\} = f_{\text{liminf}},$$

establishing the fact that $\liminf_{\|\mathbf{x}\|\to\infty} f(\mathbf{x}) \geq f_{\text{liminf}}$; hence the result follows. ☐

Note that the value f_{liminf} is computable as it is the optimal value of a GTRS, which, as already mentioned, can be efficiently solved.

4.2. Sufficient conditions for attainability. The direct consequence of Proposition 4.1 is that the condition

$$\text{[SC1]: there exists } \tilde{\mathbf{x}} \in \mathbb{R}^n \text{ such that } f(\tilde{\mathbf{x}}) < f_{\text{liminf}}$$

is sufficient for the attainability of the optimal solution of (GPS_{ls}). In order to construct a verifiable condition, we can take $\tilde{\mathbf{x}}$ as a "good" and computable approximation of the optimal solution of (GPS_{ls}). A natural candidate for such a point is the optimal solution of problem $(\text{GPS}_{\text{sls}})$, which, as mentioned in section 3.4, can be efficiently computed. Therefore, a computable sufficient condition is the following:

$$\text{[SC2]: } f(\mathbf{x}_{\text{sls}}) < f_{\text{liminf}},$$

where $\mathbf{x}_{\text{sls}}$ is the optimal solution of $(\text{GPS}_{\text{sls}})$.

Example 4.1. We now show empirically that condition [SC2] is likely to be satisfied in practical situations. Suppose that $m = 6$, $n = 2$. We performed Monte Carlo runs where in each run the locations $\mathbf{a}_j$, $j = 1, \dots, 6$, and the true source location $\mathbf{x}$ were randomly generated from a uniform distribution over the square $[-10, 10] \times [-10, 10]$. The observed distances d_j are given by (1.2), with ε_j being generated from a normal distribution with zero-mean and standard deviation σ and r is generated from a normal distribution with zero-mean and standard deviation 10. The results of the runs are summarized in Table 1 below. For each value of σ, 1000 realizations were generated. In our experiments σ takes on four different values: $10^{-2}, 10^{-1}, 1$, and 10. For each σ, N_σ denotes the number of runs for which condition [SC2] is satisfied.

TABLE 1
Number of runs (out of 1000*) for which condition* [SC2] *is satisfied.*

σ	10^{-2}	10^{-1}	1	10
N_σ	1000	1000	986	513

Clearly, for the smaller values of σ (10^{-2} and 10^{-1}), condition [SC2] was always satisfied, and it was almost always satisfied when $\sigma = 1$. For the largest value of σ, that is, $\sigma = 10$, [SC2] was satisfied for approximately half of the runs. It should be noted that $\sigma = 10$ is a huge standard deviation compared to the exact distances, meaning that the observed pseudoranges are essentially random and have only a mild connection to the "true" pseudoranges, and therefore this large σ cannot represent a practical scenario.

4.3. The connection between circle fitting and orthogonal regression. In the circle fitting problem (CF_{ls}), that is, when $\mathbf{d} = \mathbf{0}$, the liminf is given by

$$f_{\text{liminf}} = \min_{\mathbf{z}} \left\{ \mathbf{z}^T \mathbf{A}^T \left(\mathbf{I}_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^T \right) \mathbf{A}\mathbf{z} : \|\mathbf{z}\| = 1 \right\}.$$

The above expression can also be written as

$$f_{\text{liminf}} = \lambda_{\min} \left[\mathbf{A}^T \left(\mathbf{I}_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^T \right) \mathbf{A} \right].$$

In this case f_{liminf} has a nice geometric interpretation in the context of *orthogonal regression*, which we now recall. Consider the set of points $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$. For a given

$\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R}$, we define the hyperplane:

$$H_{\mathbf{x},y} := \left\{ \mathbf{a} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{a} = y \right\}.$$

In the orthogonal LS problem we seek to find a nonzero $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ such that the sum of squared Euclidean distances between the points $\mathbf{a}_1, \ldots, \mathbf{a}_m$ and $H_{\mathbf{x},y}$ is minimal; that is, the problem is given by

$$\min_{\mathbf{x},y} \left\{ \sum_{i=1}^m d(\mathbf{a}_i, H_{\mathbf{x},y})^2 : \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R} \right\}. \tag{4.16}$$

An illustration of the solution to the orthogonal regression problem is given in Figure 3 below.

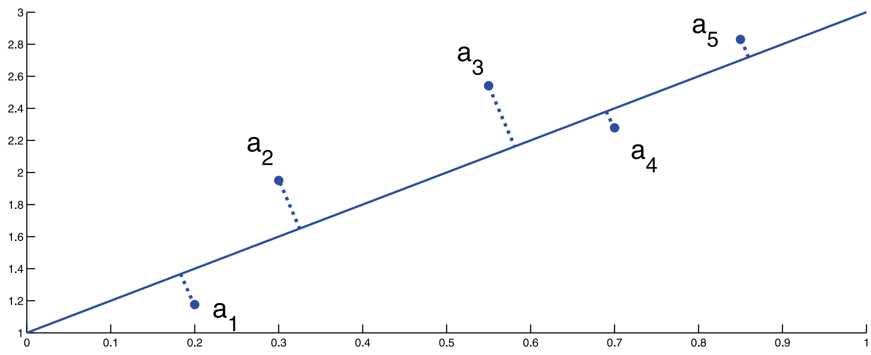

FIG. 3. *A two-dimensional example: given 5 points $\mathbf{a}_1, \ldots, \mathbf{a}_5$ in the plane, the orthogonal regression problems seeks to find the line for which the sum of squared norms of the dashed lines is minimal.*

The optimal solution of the orthogonal regression problem is described in the next well-known result (see, e.g., [17]), whose proof is given here for the sake of completeness.

PROPOSITION 4.2. *Let $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ and let $\mathbf{A}$ be the matrix given in* (4.9)*. Then an optimal solution of problem* (4.16) *is given by $\mathbf{x}$ that is an eigenvector of the matrix $\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T)\mathbf{A}$ associated with the minimum eigenvalue and $y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^T\mathbf{x}$. The optimal function value is $\lambda_{\min}\left[\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T)\mathbf{A}\right]$.*

Proof. Since the squared Euclidean distance between the point $\mathbf{a}_i$ and $H_{\mathbf{x},y}$ is given by

$$d(\mathbf{a}_i, H_{\mathbf{x},y})^2 = \frac{(\mathbf{a}_i^T\mathbf{x} - y)^2}{\|\mathbf{x}\|^2}, \quad i = 1, \ldots, m,$$

it follows that (4.16) is the same as

$$\min \left\{ \sum_{i=1}^m \frac{(\mathbf{a}_i^T\mathbf{x} - y)^2}{\|\mathbf{x}\|^2} : \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R} \right\}. \tag{4.17}$$

Fixing $\mathbf{x}$ and minimizing first with respect to y, we obtain that the optimal y is given by

$$y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^T\mathbf{x}.$$

Plugging the above expression of y into (4.17) and using the definition of the matrix $\mathbf{A}$ (4.9), we arrive at the following reformulation of (4.16):

$$\min_{\mathbf{x}} \left\{ \frac{\|(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T)\mathbf{A}\mathbf{x}\|^2}{\|\mathbf{x}\|^2} : \mathbf{x} \neq \mathbf{0} \right\},$$

which is the same as

$$\min_{\mathbf{x}} \left\{ \frac{\mathbf{x}^T[\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T)\mathbf{A}]\mathbf{x}}{\|\mathbf{x}\|^2} : \mathbf{x} \neq \mathbf{0} \right\}.$$

Therefore, an optimal solution of the problem is an eigenvector of the matrix $\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T)\mathbf{A}$ corresponding to the minimum eigenvalue; the optimal function value is the minimum eigenvalue $\lambda_{\min}\left[\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T)\mathbf{A}\right]$. $\square$

We can thus conclude that the liminf of the objective function of the circle fitting problem is equal to the optimal value of the orthogonal regression problem. This result is in fact quite natural, as illustrated by the following example.

Example 4.2. Consider the following five points in the plane:

$$\mathbf{a}_1 = (10.21, 4.51), \quad \mathbf{a}_2 = (8.63, 9.56), \quad \mathbf{a}_3 = (3.89, 13.56),$$
$$\mathbf{a}_4 = (3.22, 13.09), \quad \mathbf{a}_5 = (9.29, 6.26),$$

which are denoted by asterisks in all four plots of Figure 4. We plotted in Figure 4 the four circles

$$C(k\mathbf{z}, r(k\mathbf{z})), \quad k = 1, 10, 100, 1000,$$

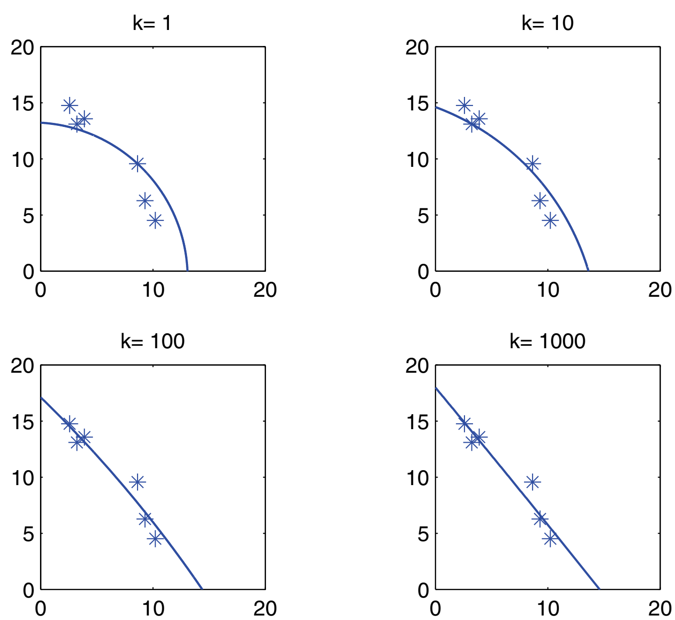

FIG. 4. *A sequence of circles for which the objective function of the circle fitting problem converges to the optimal value of the orthogonal regression problem.*

where $\mathbf{z}$ is an optimal solution of (4.8). Clearly, as k grows, the curvature of the circle in the surrounding of the five points goes to zero. As a result, the circle fitting objective function converges (as $k \to \infty$) to the optimal value of the orthogonal regression problem.

5. A fixed point method for solving ($\mathbf{GPS_{ls}}$). In this section we construct a simple fixed point method for solving problem ($\mathrm{GPS_{ls}}$). Since the problem is nonconvex, only convergence to stationary points will be established. We assume that condition [SC2] (see section 4) is satisfied, that is, that $f(\mathbf{x}_{\mathrm{sls}}) < f_{\mathrm{liminf}}$. First note that the objective function f of problem ($\mathrm{GPS_{ls}}$) can be rewritten as

$$f(\mathbf{x}) = \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - d_i)^2 - m r(\mathbf{x})^2. \tag{5.1}$$

Thus, problem ($\mathrm{GPS_{ls}}$) is closely related to the problem of estimating the location of a source from a set of distances to several anchors (see, e.g., [3] and the references therein) in which the objective function is given by

$$\tilde{f}(\mathbf{x}) = \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - d_i)^2.$$

In fact, the relation (5.1) means that f is a subtraction of $\tilde{f}$ and a convex function, and in that sense it is "less convex" than $\tilde{f}$, suggesting that problem ($\mathrm{GPS_{ls}}$) is more difficult than the source localization discussed in [3].

5.1. The method. Let

$$\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$$

be the set of points of nondifferentiability of the objective function f of ($\mathrm{GPS_{ls}}$). To derive the method, we begin by writing the optimality condition. For every $\mathbf{x} \notin \mathcal{A}$,

$$\nabla f(\mathbf{x}) = \mathbf{0}. \tag{5.2}$$

The equality (5.2) can be rewritten as

$$\begin{aligned} \frac{1}{2}\nabla f(\mathbf{x}) &= \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - d_i)\frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|} - r(\mathbf{x})\sum_{i=1}^{m} \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|} \\ &= m\mathbf{x} - \sum_{i=1}^{m} \mathbf{a}_i - \sum_{i=1}^{m} (r(\mathbf{x}) + d_i)\frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}. \end{aligned}$$

Therefore, a point $\mathbf{x} \notin \mathcal{A}$ is a stationary point if and only if

$$\mathbf{x} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{a}_i + \frac{1}{m}\sum_{i=1}^{m} (r(\mathbf{x}) + d_i)\frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}.$$

Let us denote the operator $T : \mathbb{R}^n \backslash \mathcal{A} \to \mathbb{R}^n$ by

$$T(\mathbf{x}) \equiv \frac{1}{m}\sum_{i=1}^{m} \mathbf{a}_i + \frac{1}{m}\sum_{i=1}^{m} (r(\mathbf{x}) + d_i)\frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}. \tag{5.3}$$

Then it is natural to define the following fixed point method:

A fixed point method for solving ($\mathbf{GPS_{ls}}$).

Initialization: Choose $\mathbf{x}_0 \in \mathbb{R}^n$.
General Step:

$$\mathbf{x}_{k+1} = T(\mathbf{x}_k), \quad k = 0, 1, 2, \ldots . \tag{5.4}$$

For the circle fitting problem the method takes the following form:

A fixed point method for solving ($\mathbf{CF_{ls}}$).

Initialization: Choose $\mathbf{x}_0 \in \mathbb{R}^n$.
General Step:

$$\mathbf{x}_{k+1} = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i + r(\mathbf{x}_k)\left[\frac{1}{m}\sum_{i=1}^m \frac{\mathbf{x}_k - \mathbf{a}_i}{\|\mathbf{x}_k - \mathbf{a}_i\|}\right], \quad k = 0, 1, 2, \ldots . \tag{5.5}$$

The method, as it is written, is not well defined since there might be a k for which $\mathbf{x}_k \in \mathcal{A}$, and in this case $\mathbf{x}_{k+1}$ cannot be computed via (5.4). In what follows, we will show how to avoid the points of nondifferentiability $\mathcal{A}$.

5.2. Convergence. The method (5.4) is very similar to the fixed point method that was devised for the source localization problem of [3]. Indeed, omitting the term $r(\mathbf{x})$ in the definition of $T(\mathbf{x})$ results with the exact same method introduced in [3]. The convergence analysis is also very similar, and we therefore present only the result and for completeness give the proof in the appendix.

THEOREM 5.1 (convergence of the fixed point method). *Let $\{\mathbf{x}^k\}$ be generated by (5.4) such that $\mathbf{x}^0$ satisfies*

$$f(\mathbf{x}^0) < \min\{f(\mathbf{a}_1), \ldots, f(\mathbf{a}_m), f_{\text{liminf}}\}. \tag{5.6}$$

Then the following hold:
(a) $\mathbf{x}^k \notin \mathcal{A}$ *for every* $k \geq 0$.
(b) *For every* $k \geq 0$, $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$ *and equality is satisfied if and only if* $\mathbf{x}^{k+1} = \mathbf{x}^k$.
(c) *The sequence of function values* $\{f(\mathbf{x}^k)\}$ *converges.*
(d) *The sequence* $\{\mathbf{x}^k\}$ *is bounded.*
(e) *Every convergent subsequence* $\{\mathbf{x}^{k_l}\}$ *satisfies* $\mathbf{x}^{k_l+1} - \mathbf{x}^{k_l} \to \mathbf{0}$.
(f) *Any limit point of* $\{\mathbf{x}^k\}$ *is a stationary point of* f.

Proof. See the appendix. ∎

A direct consequence of Theorem 5.1 is the following corollary.

COROLLARY 5.2. *Let $\{\mathbf{x}^k\}$ be the sequence generated by the fixed point algorithm satisfying* (5.6). *Then $f(\mathbf{x}^k) \to f^*$, where f^* is the function value at a stationary point of f.*

5.3. Initialization of the fixed point method. To make the fixed point method (5.4) well defined, it is crucial to find a point satisfying

$$f(\mathbf{x}_0) < \min\{f(\mathbf{a}_1), \ldots, f(\mathbf{a}_m), f_{\text{liminf}}\}. \tag{5.7}$$

If $f_{\text{liminf}} \le \min\{f(\mathbf{a}_1), \dots, f(\mathbf{a}_m)\}$, then by the validity of the sufficient condition [SC2] the choice $\mathbf{x}_0 = \mathbf{x}_{\text{sls}}$ will satisfy (5.7). Suppose that $f_{\text{liminf}} \ge \min\{f(\mathbf{a}_1), \dots, f(\mathbf{a}_m)\}$ and let

$$p \in \underset{i=1,\dots,m}{\operatorname{argmin}}\{f(\mathbf{a}_i)\}.$$

To find a point whose function value is smaller than $f(\mathbf{a}_p)$, we will seek a descent direction $\mathbf{v} \in \mathbb{R}^n$ satisfying $\|\mathbf{v}\| = 1$ for which $f'(\mathbf{a}_p; \mathbf{v}) < 0$ (that is, the directional derivative of the objective function at $\mathbf{a}_p$ in the direction $\mathbf{v}$ is negative). For such a direction there exists an $\varepsilon > 0$ satisfying $f(\mathbf{a}_p + t\mathbf{v}) < f(\mathbf{a}_p)$ for all $t \in (0, \varepsilon)$. The next lemma shows how to find such a direction under a mild condition.

LEMMA 5.1. *Let* $p \in \operatorname{argmin}_{i=1,\dots,m}\{f(\mathbf{a}_i)\}$ *and let*

$$g_j(\mathbf{x}) = \sum_{i=1, i\neq j}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - d_i)^2, \quad j = 1, \dots, m,$$

$$h_j(\mathbf{x}) = \frac{1}{m} \sum_{i=1, i\neq j}^{m} \|\mathbf{x} - \mathbf{a}_i\|, \quad j = 1, \dots, m.$$

Assume that the following three conditions do not hold simultaneously:

$$r(\mathbf{a}_p) = 0, \quad \nabla g_p(\mathbf{a}_p) = 0, \quad d_p = 0. \tag{5.8}$$

Then $f'(\mathbf{a}_p, \mathbf{v}) < 0$*, where the normalized vector* $\mathbf{v}$ *is given by*

$$\mathbf{v} = \begin{cases} \frac{\mathbf{z}_1}{\|\mathbf{z}_1\|} & \text{if } r(\mathbf{a}_p) > 0, \mathbf{z}_1 \neq \mathbf{0}, \\ \text{any normalized vector} & \text{if } r(\mathbf{a}_p) > 0, \mathbf{z}_1 = \mathbf{0}, \\ \text{any normalized vector} & \text{if } r(\mathbf{a}_p) = 0, \nabla g_p(\mathbf{a}_p) = \mathbf{0}, d_p > 0, \\ -\frac{\nabla g_p(\mathbf{a}_p)}{\|\nabla g_p(\mathbf{a}_p)\|} & \text{if } r(\mathbf{a}_p) = 0, \nabla g_p(\mathbf{a}_p) \neq \mathbf{0}, \end{cases}$$

and where

$$\mathbf{z}_1 = -\nabla g_p(\mathbf{a}_p) + 2mr(\mathbf{a}_p)\nabla h_p(\mathbf{a}_p). \tag{5.9}$$

Proof. Recall that

$$f(\mathbf{x}) = \sum_{i=1}^{m} (\|\mathbf{x} - \mathbf{a}_i\| - d_i)^2 - mr^2(\mathbf{x}).$$

Let us split the analysis into two cases.

Case 1. If $\sum_{i=1}^{m}(\|\mathbf{a}_p - \mathbf{a}_i\| - d_i) > 0$, then $r(\mathbf{a}_p) > 0$ and we have that

$$\begin{aligned} f'(\mathbf{a}_p; \mathbf{v}) &= \nabla g_p(\mathbf{a}_p)^T \mathbf{v} - 2d_p - 2mr(\mathbf{a}_p) \left(\nabla h_p(\mathbf{a}_p)^T \mathbf{v} + \frac{1}{m} \right) \\ &= -\mathbf{z}_1^T \mathbf{v} - 2d_p - 2r(\mathbf{a}_p). \end{aligned}$$

If $\mathbf{z}_1 = \mathbf{0}$, then any direction is a descent direction. Otherwise, by taking $\tilde{\mathbf{v}} = \frac{\mathbf{z}_1}{\|\mathbf{z}_1\|}$, we obtain that

$$f'(\mathbf{a}_p; \tilde{\mathbf{v}}) = -\|\mathbf{z}_1\| - 2d_p - r(\mathbf{a}_p),$$

which by the fact that $d_p \ge 0$ and $r(\mathbf{a}_p) > 0$ implies that $f'(\mathbf{a}_p; \tilde{\mathbf{v}}) < 0$.

Case 2. If $\sum_{i=1}^m (\|\mathbf{a}_p - \mathbf{a}_i\| - d_i) \le 0$, then $r(\mathbf{a}_p) = 0$. Consider the function $\tilde{f}(\mathbf{x}) = \sum_{i=1}^m (\|\mathbf{x} - \mathbf{a}_i\| - d_i)^2$. It is enough to show that there exists a normalized vector $\mathbf{v}$ for which $\tilde{f}'(\mathbf{a}_p; \mathbf{v}) < 0$. Now,

$$\tilde{f}'(\mathbf{a}_p; \mathbf{v}) = \nabla g_p(\mathbf{a}_p)^T \mathbf{v} - 2d_p.$$

If $\nabla g_p(\mathbf{a}_p) \neq 0$, then define $\tilde{\mathbf{v}} = -\frac{\nabla g_p(\mathbf{a}_p)}{\|\nabla g_p(\mathbf{a}_p)\|}$. It holds that

$$\tilde{f}'(\mathbf{a}_p; \tilde{\mathbf{v}}) = -\|\nabla g_p(\mathbf{a}_p)\| - 2d_p,$$

which by the fact that $d_p \ge 0$ implies that $\tilde{f}'(\mathbf{a}_p; \tilde{\mathbf{v}}) < 0$. If $\nabla g_p(\mathbf{a}_p) = 0$ and $d_p > 0$, then $\tilde{f}'(\mathbf{a}_p; \mathbf{v}) = -2d_p < 0$ for any $\mathbf{v} \in \mathbb{R}^n$, implying that *any* direction is a descent direction. $\square$

Note that the condition which states that (5.8) will not be satisfied is mild in the sense that it is very unlikely for "true" random data that the vector $\nabla g_p(\mathbf{a}_p)$ will be equal to the zeros vector. In fact, for the circle fitting problem, (5.8) will surely not be satisfied unless all the points $\mathbf{a}_1, \ldots, \mathbf{a}_m$ are equal to $\mathbf{a}_p$, which is certainly not an interesting case, and is also not possible under the underlying Assumption A. Under this condition, and based on Lemma 5.1, we can define the following procedure for finding an initial vector $\mathbf{x}_0$ for which the fixed point method (5.4) is well defined.

Procedure for finding an $\mathbf{x}_0 \in \mathbb{R}^n$ satisfying $f(\mathbf{x}_0) < \min\{f(\mathbf{a}_1), \ldots, f(\mathbf{a}_m), f_{\text{liminf}}\}$.

a. If $f_{\text{liminf}} \le \min\{f(\mathbf{a}_1), \ldots, f(\mathbf{a}_m)\}$, then choose $\mathbf{x}_0 = \mathbf{x}_{\text{sls}}$ and STOP.

b. Let $p \in \operatorname{argmin}_{i=1,\ldots,m} \{f(\mathbf{a}_i)\}$. Choose a descent direction of f at $\mathbf{a}_p$ according to the following cases:
   - If $r(\mathbf{a}_p) > 0$, then there are two options: if $\mathbf{z}_1 = -\nabla g_p(\mathbf{a}_p) + 2mr(\mathbf{a}_p)\nabla h_p(\mathbf{a}_p) \neq \mathbf{0}$, then take $\tilde{\mathbf{v}} = \frac{1}{\|\mathbf{z}_1\|}\mathbf{z}_1$; otherwise, take $\tilde{\mathbf{v}}$ to be any normalized vector and go to step c.
   - If $r(\mathbf{a}_p) = 0$, then either $\mathbf{z} = \nabla g_p(\mathbf{a}_p) \neq \mathbf{0}$, and in that case $\tilde{\mathbf{v}} = -\frac{1}{\|\mathbf{z}\|}\mathbf{z}$, or $\nabla g_p(\mathbf{a}_p) = \mathbf{0}$, and in that case $\tilde{\mathbf{v}}$ can be chosen as any normalized vector.

c. Set $s = 1$.

d. If $f(\mathbf{a}_p + s\tilde{\mathbf{v}}) < f(\mathbf{a}_p)$, then STOP. The output is $\mathbf{x}_0 = \mathbf{a}_p + s\tilde{\mathbf{v}}$. Otherwise, go to step e.

e. Set $s \leftarrow \frac{s}{2}$. Go back to step d.

5.4. Numerical examples. The following example illustrates the advantage of the LS solution (an optimal solution of problem (GPS_{ls})) over the SLS solution (an optimal solution of $(\text{GPS}_{\text{sls}})$). All the experiments were performed in MATLAB.

Example 5.1. The setting in this example is the same as the one used in Example 4.1, with the exception that we do not consider the value $\sigma = 10$. For each value of σ, 1000 realizations were generated. For each value of σ, the second column is the average over the 1000 runs of the relative error of the SLS solution $\frac{\|\mathbf{x}_{\text{sls}} - \mathbf{x}_{\text{true}}\|}{\|\mathbf{x}_{\text{true}}\|}$, while the third column is the average of the relative error of the LS solution $\frac{\|\mathbf{x}_{\text{ls}} - \mathbf{x}_{\text{true}}\|}{\|\mathbf{x}_{\text{true}}\|}$. The LS solution was obtained by using the fixed point method (5.4) with the SLS solution as an initial point. The fourth column contains the number of runs in

TABLE 2

The average of the relative errors of the LS and SLS solutions (second and third columns) and the number of runs in which the LS solution provided a better solution than the SLS solution.

σ	Relative error SLS	Relative error LS	I_σ
10^{-2}	0.0033	0.0028	651
10^{-1}	0.0355	0.0279	634
1	0.3193	0.2846	605

which the LS solution was closer than the SLS solution to the true solution (that is, $\|\mathbf{x}_{\text{ls}} - \mathbf{x}_{\text{true}}\| < \|\mathbf{x}_{\text{sls}} - \mathbf{x}_{\text{true}}\|$).

Obviously, the results summarized in Table 2 suggest that on average the LS estimate gives better results than the SLS estimate. At the same time, it seems that the SLS solution is a rather good approximation of the LS solution in the sense that it has the same order of magnitude of relative error and in many cases (approximately 35–40 percent) gives a better approximation than the LS solution.

Example 5.2. In this example we further illustrate the advantage of the LS solution over the SLS solution by considering one realization of the GPS localization problem in which

$$\mathbf{a}_1 = \begin{pmatrix} -29 \\ -18 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 7 \\ -24 \end{pmatrix}, \mathbf{a}_3 = \begin{pmatrix} -19 \\ -27 \end{pmatrix}, \mathbf{a}_4 = \begin{pmatrix} 10 \\ -27 \end{pmatrix}, \mathbf{a}_5 = \begin{pmatrix} -9 \\ 3 \end{pmatrix}, \mathbf{a}_6 = \begin{pmatrix} -33 \\ -34 \end{pmatrix},$$

and with the "true" source location chosen as $\mathbf{x}_{\text{true}} = (-8, -2)^T$. We then generated 1000 realizations of the noise components ε_i with standard deviation $\sigma = 1$, and computed the LS and SLS solutions (as in the previous example). The left image in Figure 5 describes the histogram of the errors of the LS solution ($\|\mathbf{x}_{\text{ls}} - \mathbf{x}_{\text{true}}\|$), while the right image describes the corresponding histogram of the errors of the SLS solution ($\|\mathbf{x}_{\text{sls}} - \mathbf{x}_{\text{true}}\|$).

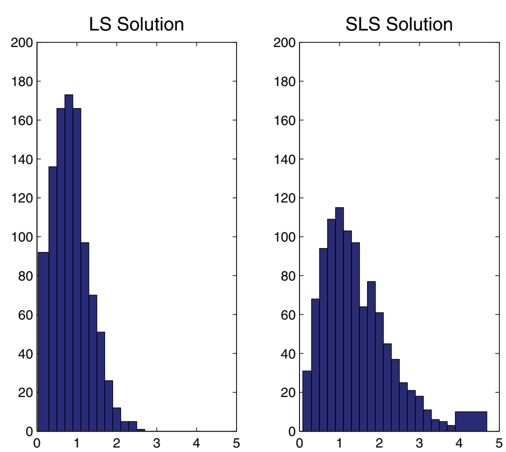

FIG. 5. *Histograms of the errors of the LS (left) and SLS (right) solutions.*

The histogram corresponding to the LS estimate is shifted closer to zero compared to the histogram of the SLS estimate and also has a smaller variance. This clearly indicates that overall the LS solution provides more accurate solutions than the SLS solution.

Example 5.3. In this example we demonstrate that the LS circle fitting problem ($\mathrm{CF_{ls}}$) can result with a more geometrical sense than the SLS circle fitting problem ($\mathrm{CF_{sls}}$). Consider the 6 points

$$\mathbf{a}_1 = \begin{pmatrix}1\\9\end{pmatrix}, \ \mathbf{a}_2 = \begin{pmatrix}2\\7\end{pmatrix}, \ \mathbf{a}_3 = \begin{pmatrix}5\\8\end{pmatrix}, \ \mathbf{a}_4 = \begin{pmatrix}7\\7\end{pmatrix}, \ \mathbf{a}_5 = \begin{pmatrix}9\\5\end{pmatrix}, \ \mathbf{a}_6 = \begin{pmatrix}3\\7\end{pmatrix}.$$

Then the solutions to ($\mathrm{CF_{ls}}$) and ($\mathrm{CF_{sls}}$) are given in Figure 6. Obviously the LS solution is much more reasonable from a geometrical point of view. Further evidence for this observation is that the sum of squares of the distances of the points to the circle obtained by the circle fitting LS problem ($\mathrm{CF_{ls}}$) is 3.1724, while the sum of distances of the points to the circle produced by the SLS circle fitting problem ($\mathrm{CF_{sls}}$) is 7.2081—more than twice as much.

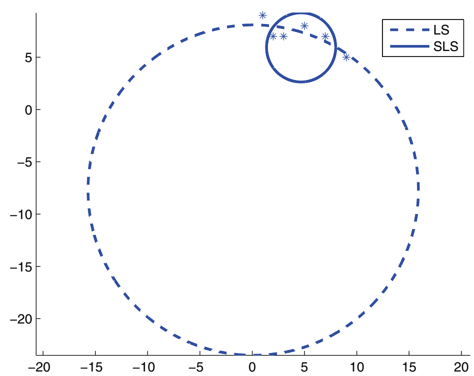

Fig. 6. *The dashed circle is the optimal solution to* ($\mathrm{CF_{ls}}$)*, while the circle plotted with a solid line is the optimal solution to* ($\mathrm{CF_{sls}}$)*.*

6. Concluding remarks. This paper considered two possible formulations of the GPS localization problem: the squared least squares (SLS) formulation ($\mathrm{GPS_{sls}}$), which is nonconvex and smooth, and the least squares (LS) variant ($\mathrm{GPS_{ls}}$), which is nonconvex and nonsmooth. Both problems are generalizations of circle fitting problems. The disadvantage of the SLS formulation is that it lacks the statistical and geometrical meaning of the LS problem. However, the SLS solution was shown (empirically) to be a good starting point for a fixed point method devised to solve the LS problem. It is still an open question whether an efficient method for finding the global optimal solution of the LS problem can be devised. Another interesting line of analysis would be to generalize the obtained results to the more complicated problems arising in the area of sensor network localization.

Appendix. Proof of Theorem 5.1. Before proving Theorem 5.1, we need to define the following auxiliary function:

$$h(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i - (r(\mathbf{y}) + d_i)h_i(\mathbf{y})\|^2, \tag{A.1}$$

where

$$h_i(\mathbf{y}) \equiv \frac{\mathbf{y} - \mathbf{a}_i}{\|\mathbf{y} - \mathbf{a}_i\|}, \quad i = 1, \ldots, m. \tag{A.2}$$

Then it holds that $\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} h(\mathbf{x}, \mathbf{x}_k)$. The following lemma states several key properties of the auxiliary function h.

LEMMA A.1.

(a) *$h(\mathbf{x}, \mathbf{x}) = f(\mathbf{x})$ for every $\mathbf{x} \notin \mathcal{A}$.*

(b) *$h(\mathbf{x}, \mathbf{y}) \geq f(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \notin \mathcal{A}$.*

(c) *If $\mathbf{y} \notin \mathcal{A}$, then*

$$T(\mathbf{y}) = \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n} h(\mathbf{x}, \mathbf{y}).$$

(d) *For every $\mathbf{y} \notin \mathcal{A}$*

$$f(T(\mathbf{y})) \leq f(\mathbf{y})$$

and equality holds if and only if $T(\mathbf{y}) = \mathbf{y}$.

Proof. (a) By the definition of the auxiliary function h and $h_1, \ldots, h_m$, it follows that

$$\begin{aligned} h(\mathbf{x}, \mathbf{x}) &= \sum_{i=1}^m \|\mathbf{x} - \mathbf{a}_i - (r(\mathbf{x}) + d_i)h_i(\mathbf{x})\|^2 \\ &= \sum_{i=1}^m \left\{\|\mathbf{x} - \mathbf{a}_i\|^2 - 2(r(\mathbf{x}) + d_i)h_i(\mathbf{x})^T(\mathbf{x} - \mathbf{a}_i) + (r(\mathbf{x}) + d_i)^2\|h_i(\mathbf{x})\|^2\right\} \\ &= \sum_{i=1}^m \left\{\|\mathbf{x} - \mathbf{a}_i\|^2 - 2(r(\mathbf{x}) + d_i)\|\mathbf{x} - \mathbf{a}_i\| + (r(\mathbf{x}) + d_i)^2\right\} \\ &= \sum_{i=1}^m (\|\mathbf{x} - \mathbf{a}_i\| - r(\mathbf{x}) - d_i)^2 = f(\mathbf{x}). \end{aligned}$$

(b) First, note that

$$\begin{aligned} h(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^m \|\mathbf{x} - \mathbf{a}_i - (r(\mathbf{y}) + d_i)h_i(\mathbf{y})\|^2 \\ &= \sum_{i=1}^m \left\{\|\mathbf{x} - \mathbf{a}_i\|^2 - 2(r(\mathbf{y}) + d_i)h_i(\mathbf{y})^T(\mathbf{x} - \mathbf{a}_i) + (r(\mathbf{y}) + d_i)^2\|h_i(\mathbf{y})\|^2\right\} \\ &= \sum_{i=1}^m \left\{\|\mathbf{x} - \mathbf{a}_i\|^2 - 2(r(\mathbf{y}) + d_i)h_i(\mathbf{y})^T(\mathbf{x} - \mathbf{a}_i) + (r(\mathbf{y}) + d_i)^2\right\}. \end{aligned} \tag{A.3}$$

By the Cauchy–Schwarz inequality it follows that

$$h_i(\mathbf{y})^T(\mathbf{x} - \mathbf{a}_i) = \frac{(\mathbf{y} - \mathbf{a}_i)^T(\mathbf{x} - \mathbf{a}_i)}{\|\mathbf{y} - \mathbf{a}_i\|} \leq \frac{\|\mathbf{y} - \mathbf{a}_i\| \cdot \|\mathbf{x} - \mathbf{a}_i\|}{\|\mathbf{y} - \mathbf{a}_i\|} = \|\mathbf{x} - \mathbf{a}_i\|,$$

which combined with the expression (A.3) for $h(\mathbf{x}, \mathbf{y})$ and the fact that $r(\mathbf{y}) \geq 0$ implies that

$$h(\mathbf{x}, \mathbf{y}) \geq \sum_{i=1}^m \left\{\|\mathbf{x}-\mathbf{a}_i\|^2 - 2(r(\mathbf{y})+d_i)\|\mathbf{x}-\mathbf{a}_i\| + (r(\mathbf{y})+d_i)^2\right\} = \sum_{i=1}^m (\|\mathbf{x}-\mathbf{a}_i\| - r(\mathbf{y}) - d_i)^2.$$

Finally, by the definition of $r(\mathbf{x})$ as the minimizer of $\sum_{i=1}^m (\|\mathbf{x}-\mathbf{a}_i\| - r - d_i)^2$ over all nonnegative r, it follows that

$$\sum_{i=1}^m (\|\mathbf{x}-\mathbf{a}_i\| - r(\mathbf{y}) - d_i)^2 \geq \sum_{i=1}^m (\|\mathbf{x}-\mathbf{a}_i\| - r(\mathbf{x}) - d_i)^2 = f(\mathbf{x}),$$

proving that $h(\mathbf{x},\mathbf{y}) \geq f(\mathbf{x})$.

(d) By (5.3) and the strict convexity of the function $\mathbf{x} \mapsto h(\mathbf{x},\mathbf{y})$, one has

$$h(T(\mathbf{y}),\mathbf{y}) < h(\mathbf{x},\mathbf{y}) \text{ for every } \mathbf{x} \neq T(\mathbf{y}).$$

In particular, if $T(\mathbf{y}) \neq \mathbf{y}$, then

$$h(T(\mathbf{y}),\mathbf{y}) < h(\mathbf{y},\mathbf{y}) = f(\mathbf{y}), \tag{A.4}$$

where the last equality follows from part (a). By part (b), $h(T(\mathbf{y}),\mathbf{y}) \geq f(T(\mathbf{y}))$, which combined with (A.4) establishes the desired strict monotonicity. $\square$

Proof of Theorem 5.1. (a) and (b) follow by induction on k using Lemma A.1(d).

(c) readily follows from the monotonicity and lower boundedness (by zero) of the sequence $\{f(\mathbf{x}^k)\}$.

(d) By part (b) all the iterates $\mathbf{x}^k$ are in the level set $Lev(f, f(\mathbf{x}^0))$, which, by the fact that $f(\mathbf{x}^0) < f_{\text{liminf}}$, establishes the boundedness of the sequence $\{\mathbf{x}^k\}$.

(e) and (f) Let $\{\mathbf{x}^{k_l}\}$ be a convergent subsequence of $\{\mathbf{x}^k\}$ with limit point $\mathbf{x}^*$. Since $f(\mathbf{x}^{k_l}) \leq f(\mathbf{x}^0) < \min_{j=1,\ldots,m} f(\mathbf{a}_j)$, it follows by the continuity of f that $f(\mathbf{x}^*) \leq f(\mathbf{x}^0) < \min_{j=1,\ldots,m} f(\mathbf{a}_j)$, proving that $\mathbf{x}^* \notin \mathcal{A}$. By (5.4),

$$\mathbf{x}^{k_l+1} = T(\mathbf{x}^{k_l}). \tag{A.5}$$

Therefore, since the subsequence $\{\mathbf{x}^{k_l}\}$ and its limit point $\mathbf{x}^*$ are not in $\mathcal{A}$, by the continuity of ∇f on $\mathbb{R}^n \setminus \mathcal{A}$, we conclude that the subsequence $\{\mathbf{x}^{k_l+1}\}$ converges to a vector $\bar{\mathbf{x}}$ satisfying

$$\bar{\mathbf{x}} = T(\mathbf{x}^*). \tag{A.6}$$

To prove (e), we need to show that $\bar{\mathbf{x}} = \mathbf{x}^*$. Since both $\mathbf{x}^*$ and $\bar{\mathbf{x}}$ are limit points of $\{\mathbf{x}^k\}$ and since the sequence of function values converges (by part (c)), then the continuity of f over $\mathbb{R}^n$ implies $f(\mathbf{x}^*) = f(\bar{\mathbf{x}})$. Invoking Lemma A.1 for $\mathbf{y} = \mathbf{x}^*$, we conclude that $\bar{\mathbf{x}} = \mathbf{x}^*$, proving claim (e). Part (f) follows from the observation that the equality $\mathbf{x}^* = T(\mathbf{x}^*)$ is equivalent (by the definition of T) to $\nabla f(\mathbf{x}^*) = \mathbf{0}$. $\square$

REFERENCES

[1] J. S. Abel, *A variable projection method for additive components with application to GPS*, IEEE Trans. Aerospace Electron. Systems, 30 (1994), pp. 928–930.

[2] A. Beck, P. Stoica, and J. Li, *Exact and approximate solutions of source localization problems*, IEEE Trans. Signal Process., 56 (2008), pp. 1770–1778.

[3] A. Beck, M. Teboulle, and Z. Chikishev, *Iterative minimization schemes for solving the single source localization problem*, SIAM J. Optim., 19 (2008), pp. 1397–1416.

[4] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[5] I. V. L. Clarkson and E. E. Zelniker, *A statistical analysis of the Delogne–Kasa method for fitting circles*, Digital Signal Process., 16 (2006), pp. 498–522.

[6] M. G. Cox and A. B. Forbes, *Strategies for Testing Form Assessment Software*, Technical report, NPL Report DITC 211/92, National Physical Laboratory, Middlesex, UK, 1992.

[7] A. Fitzgibbon, M. Pilu, and R. B. Fisher, *Direct least square fitting of ellipses*, IEEE Trans. Pattern Anal. Machine Intelligence, 21 (1999), pp. 476–480.

[8] C. Fortin and H. Wolkowicz, *The trust region subproblem and semidefinite programming*, Optim. Methods Softw., 19 (2004), pp. 41–67.

[9] W. Gander, G. H. Golub, and R. Strebel, *Least-squares fitting of circles and ellipses*, BIT, 34 (1994), pp. 24–40.

[10] M. Heidari and P. C. Heigold, *Determination of hydraulic conductivity tensor using a nonlinear least squares estimator*, Water Resources Bull., (1993), pp. 415–424.

[11] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, *Real-time passive source localization: A practical linear correction least-squares approach*, IEEE Trans. Speech Audio Process., 9 (2002), pp. 943–956.

[12] I. Kasa, *A circle fitting procedure and its error analysis*, IEEE Trans. Instrum. Meas., 25 (1976), pp. 8–14.

[13] U. M. Landau, *Estimation of a circular arc center and its radius*, Comput. Vis. Graph. Image Process., 38.

[14] Y. V. Linnik, *Method of Least Squares and Principles of the Theory of Observations*, Pergamon Press, New York, 1961.

[15] J. J. Moré, *Generalizations of the trust region subproblem*, Optim. Methods Softw., 2 (1993), pp. 189–209.

[16] V. Pratt, *Direct least squares fitting of algebraic surfaces*, ACM J. Computer Graphics, 21 (1987), pp. 145–152.

[17] L. S. Shapiro and M. Brady, *Rejecting outliers and estimating errors in orthogonal-regression framework*, Philos. Trans. R. Soc. Lond. Ser. A, 350 (1995), pp. 407–439.

[18] P. Stoica and J. Li, *Source localization from range-difference measurements*, IEEE Signal Process. Mag., 23 (2006), pp. 63–65, 69.

[19] A. Thom, *A statistical examination of the megalithic sites in Britain*, J. Roy. Stat. Soc. Ser. A, 118 (1955), pp. 275–295.

[20] G. Wang, K. Yang, and Z. Q. Luo, *Efficient convex relaxation methods for robust target localization by a sensor network using time differences of arrivals*, IEEE Trans. Signal Process., 57 (2009), pp. 2775–2784.