Guaranteeing Practical Convergence in Algorithms for Sensor and Source Localization

Barış Fidan, *Member, IEEE*, Soura Dasgupta, *Fellow, IEEE*, and Brian D. O. Anderson, *Life Fellow, IEEE*

***Abstract*—This paper considers localization of a source or a sensor from distance measurements. We argue that linear algorithms proposed for this purpose are susceptible to poor noise performance. Instead given a set of sensors/anchors of known positions and measured distances of the source/sensor to be localized from them we propose a potentially nonconvex weighted cost function whose global minimum estimates the location of the source/sensor one seeks. The contribution of this paper is to provide nontrivial ellipsoidal and polytopic regions surrounding these sensors/anchors of known positions, such that if the object to be localized is in this region, localization occurs by globally exponentially convergent gradient descent in the noise free case. Exponential convergence in the noise free case represents practical convergence as it ensures graceful performance degradation in the presence of noise. These results guide the deployment of sensors/anchors so that small subsets can be made responsible for practical localization in geographical areas determined by our approach.**

***Index Terms*—Global convergence, gradient descent, localization, optimization, sensors.**

I. Introduction

OVER the last few years the problem of source/sensor localization has assumed increasing significance (see [1] for application scope). Specifically *source localization* refers to a set of sensors estimating the precise location of a source using information related to their relative position to the source. In *sensor localization* a sensor estimates its own position using similar information relative to several nodes of known positions called *anchors*. This information can be distance, bearing, power level (indirectly related to distance), and time difference of arrival (TDOA). *In this paper, we will focus on distances only*. In abstract terms, our goal is the following. Given known 2-D or 3-D vectors $x_1, \cdots, x_N$ ($N > 2$ and $N > 3$ in 2-D and 3-D, respectively) and an unknown vector y^*, estimate the value of y^*, from the measured distances $d_i = \|y^* - x_i\|$. Here, as elsewhere in this paper, unless otherwise noted, all vector norms are 2-norms. In the source localization problem, y^* represents the position of the unknown source, and the x_i the positions of the sensors seeking to estimate its location. In the sensor localization problem, the x_i are the positions of the anchors, and y^* the position of the sensor estimating its own position.

We are particularly interested in issues of anchor/sensor placement. For reasons to be explained later, we believe that it is important to deploy anchors/sensors in a way so that small subsets can be responsible for localization within easily determined geographical regions. To aide such deployment, we wish to characterize regions surrounding small groups of anchors/sensors such that sensors/sources lying in these regions can be localized in a *practical* and effective way by using their distances from the members of these groups. The definition and motivation of "practical localization" is provided in the sequel.

We note that distances can be estimated through various means. For example if a source emits a signal, the signal intensity and the characteristics of the medium provides a distance estimate. In this case with A the source signal strength, and η the power loss coefficient, the received signal strength (RSS) at a distance d from the source is given by

$$s = A/d^{\eta}. \tag{1.1}$$

Thus, A, s, and η provide d. Alternatively, a sensor may transmit signals of its own, and estimate the distance by measuring the time it takes for this signal reflected off the target to return. Another means of estimating distances is when a group of sensors collaboratively use TDOA information.

In 2-D, localization from distance measurements generically requires that distances from y^* of at least three noncollinearly situated x_i be available. To be precise, with just two distances, the position can be determined to within a binary ambiguity. Occasionally, *a priori* information may be available which will resolve that ambiguity. Otherwise, a third distance is needed. In three dimensions, one generically needs at least four noncoplanar x_i.

Prior work on anchor deployment has mostly concerned placing anchors in a way so that every sensor can measure its distance from a sufficient number of anchors [2], [3]. Indeed, as explained in Section II if the d_i uniquely specify y^*, y^* can be estimated using linear algorithms [4], [5]. However, in practice, as also explained in Section II, such a linear algorithm with certain geometries may deliver highly inaccurate estimates with noisy measurements of the distances, even when the noise is small.

Manuscript received December 4, 2006; revised January 22, 2008. Published August 13, 2008 (projected). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhi Tian. Supported by NSF grants ECS-0622017 and CCF-072902, and by National ICT Australia, which is funded by the Australian Government's Department of Communications, Information Tehcnology and the Arts and the Australian Research Council through the Backing Australia's Ability initiative and the ICT Centre of Excellence Program.

B. Fidan and B. D. O. Anderson are with Australian National University and National ICT Australia, Limited, Canberra ACT 2601, Australia (e-mail: Baris.Fidan@nicta.com.au; Brian.Anderson@nicta.com.au).

S. Dasgupta is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA (e-mail: dasgupta@engineering.uiowa.edu).

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

Digital Object Identifier 10.1109/TSP.2008.924138

On the other hand, several papers adopt a nonlinear estimation approach [6]–[10]. Many of these directly work with (1.1), with known A and η. Thus, rather than assuming that the d_i are directly available, they work with the RSS at several sensors and choose y^*, to be the y that minimizes

$$\sum_{i=1}^{N}\left(s_i - \frac{A}{||x_i - y||^{\eta}}\right)^2. \tag{1.2}$$

In (1.2) there are N sensors, the ith located at x_i, and s_i is the RSS at the ith sensor. It should be noted that unless one makes the unrealistic assumption that the noise perturbing the RSS is Gaussian, (1.2) *does not provide the maximum-likelihood (ML) estimate of* y^*. A more realistic assumption on the noise perturbing s_i is that it is log normal, from which an ML-algorithm can be developed.

Cost functions such as (1.2) and indeed the ML-cost function, are inevitably nonconvex, and their minimization manifested with locally attractive false optima. This is true, for example, for [7] and [8] which conduct searches *in two dimensions* that, as noted in [9], are sensitive to spurious stationary points. While it is easy to detect convergence of search procedures to false minima, one would have to reinitialize the search process, potentially multiple times, wasting precious power, and impairing time critical localization.

As partial amelioration, [9] provides search alternatives involving the so-called Projection on Convex Sets (POCS) approach also in two dimensions, with $\eta = 2$. It has however, the unique solution of y^* in the noise free case if and only if y^* is in the convex hull of the x_i. Convergence fails if y^* lies outside the convex hull.

As an extension [10] proposes the so-called hyperbolic POCS, that does sometimes converge even if y^* is outside the convex hull of the x_i. However, no characterization of conditions for convergence is given. This is also true for the hybrid of the hyperbolic and circular POCS, also advocated in [10].

A. The Approach

The foregoing indicates that sensor/anchor placement that guarantee availability of sufficient numbers of distance measurements will not suffice for practical convergence. One must also take into account the convergence behavior of any nonlinear algorithm one employs. In particular, to avoid repeated reinitializations that sap energy and induce delays, it is important to characterize regions surrounding *small numbers of sensors/anchors* such that if a source/sensor lies in them, the resulting cost function can be minimized using a gradient descent law that is *globally exponentially convergent* in the noise free case. We assert that such a characterization facilitates *practical localization in three respects*.

First, there are several results in the nonlinear stability literature (e.g., [13, Ch. 5]) that show that global exponential convergence in the noise free case guarantees graceful performance degradation in the presence of noise. These include results where convergence in distribution, [12], with variance increasing with increasing noise variance is proven. Second, assigning just a small number of sensors/anchors the responsibility of localizing a given region also has practical benefits: Collaboration involving large numbers of sensors strains the systems resources both in terms of computational and communication costs. Third, such regions remove the need for repeated initializations which deplete resources and are inimical to time critical localization. Characterization of such regions for cost functions such as (1.2), which in any case does not lead to a ML estimate, is difficult as the corresponding gradient descent update kernel is rational for integer η and irrational in general. The corresponding cost function for ML estimation under a more realistic log normal noise assumption is even more problematic, due to the presence of logarithms in the update kernel. Consequently, any benefits accruing from obtaining an ML estimate are outweighed by the lack of guaranteed convergence.

Instead our starting point is that some how distance measurements have been obtained without reference to the underlying signal model except possibly, as in [10], in its use in obtaining distance estimates. Localization is accomplished by obtaining y^* as the y minimizing the cost function defined below for certain designer selected weights $\lambda_i > 0$

$$J(y) = 0.5\sum_{i=1}^{N}\lambda_i\left(||x_i - y||^2 - d_i^2\right)^2. \tag{1.3}$$

In the cost function (1.3), each addend term $\lambda_i(||x_i - y||^2 - d_i^2)^2$, for $i = 1,\ldots,N$, penalizes the difference between the calculated distance of x_i to the source/sensor position estimate y and the measured distance of x_i to y^*. The coefficients λ_i $(i = 1,\cdots,N)$ are weighting terms that can be chosen based on the any additional *a priori* information that maybe available. For example, if it is known that certain d_i estimates are more reliable than others, then one may choose the corresponding λ_i to be larger. On the other hand, see Section IV, these weight selections may also facilitate practical convergence in the sense adopted in this paper.

Of course, other indexes can be contemplated. For example one could work with $||x_i - y|| - d_i$, rather than the difference of the squares, and one could work with the $2m$th power of the difference, rather than the second power. Working with the difference $||x_i - y|| - d_i$, rather than the difference of the squares would be much harder to treat by the methods of this paper, as the derivative of the index is more awkward analytically. On the other hand, working with a power $2m$ rather than 2 may be a more tractable extension.

Section II provides more background and examples showing the non-convexity of (1.3). In Section III, given $\lambda_i > 0$ and x_i, we characterize a nontrivial *ellipsoidal* set, members of which are guaranteed to be exponentially localized through the gradient descent minimization of (1.3) in the noise free case. In Section IV, given x_i, we quantify a nontrivial polytopic set for which there exist $\lambda_i > 0$ such that members of this set can be similarly localized. Section V demonstrates the performance of gradient descent minimization under noisy distance measurements. It is after all an important aim of this paper to present an algorithm initially derived to cope with noiseless measurements, but able to cope with noisy ones in a graceful manner. Section VI is the conclusion.

II. Background and Preliminaries

In this section, we discuss linear algorithms and examples demonstrating the nonconvexity of (1.3).

A. Linear Algorithms

We discuss now the practical ramifications of linear localization algorithms. Consider three noncollinear x_i in 2-D and equations

$$\|x_i - y^*\|^2 = d_i^2, \quad \text{for } i \in \{1,2,3\}. \tag{2.1}$$

Subtracting the first equation from the remaining two one obtains

$$2\begin{bmatrix}(x_1-x_2)^T\\(x_1-x_3)^T\end{bmatrix}y^* = \begin{bmatrix}\|x_1\|^2 - \|x_2\|^2 + d_2^2 - d_1^2\\ \|x_1\|^2 - \|x_3\|^2 + d_3^2 - d_1^2\end{bmatrix}. \tag{2.2}$$

For noncollinear x_i, $\det([(x_1-x_2),(x_1-x_3)]^T) \neq 0$, i.e., y^* can be solved uniquely. But, the solution is invariant if for any α the d_i^2 are replaced by $d_i^2+\alpha$, suggesting and verified by example in [11], that such linear algorithms may have poor noise performance.

Indeed consider in two dimensions $x_1 = 0$, $x_2 = [43,7]^T$, $x_3 = [47,0]^T$ and $y^* = [17.9719, -29.3227]^T$. Its distances from the x_i are $d_1 = 34.392$, $d_2 = 44.1106$, and $d_3 = 41.2608$. Now suppose the measured distances from x_1, x_2 and x_3 are 35, 42, and 43, respectively. These estimated distances are inconsistent in that no single y can simultaneously meet these distance constraints. The linear algorithm provides an estimate of y^* that is $[16.8617, -6.5076]^T$, i.e., relatively small errors in distance measurements translate to very substantial localization error. *On the other hand, the gradient descent optimization of* (1.3), *using* $\lambda_i = 1$, *and the algorithm in* (2.4) *below, initialized with the estimate provided by the linear algorithm, converges to* $[18.2150, -29.2443]^T$ *a point that is much closer to* y^*.

B. Preliminaries of (1.3)

Our standing assumption below ensures that $J(y) = 0$ if and only if $y = y^*$.

Assumption 2.1: In two dimensions, $N > 2$ and the x_i, $i \in \{1, \cdots, N\}$, are noncollinear. In three dimensions, $N > 3$, and they are noncoplanar.

As an initial point we will seek to find conditions under which

$$\partial J(y)/\partial y = \sum_{i=1}^{N} \lambda_i \left(\|y - x_i\|^2 - d_i^2\right)(y - x_i) = 0 \text{ iff } y = y^*. \tag{2.3}$$

The fact that $y = y^*$ guarantees that $\partial J(y)/\partial y = 0$ is trivially seen by noting that by definition $\|y^* - x_i\|^2 = d_i^2$ for all i. Examples presented in this section show that in fact the reverse implication does not always hold.

Consider the iterative gradient descent algorithm

$$y[k+1] = y[k] - \mu \sum_{i=1}^{N} \lambda_i \left(\|y[k] - x_i\|^2 - d_i^2\right)(y[k] - x_i) \tag{2.4}$$

where $y[k]$, for $k = 1, 2, \ldots$, denotes the estimate of y^* at the kth iteration, and $y[0]$ is the initial estimate to be chosen. Under (2.3), it is well known that given an arbitrary constant $C > 0$, there exists μ^* dependent on C such that for all $\mu \leq \mu^*$, the algorithm (2.4) is globally uniformly asymptotically convergent to $y = y^*$ for every $\|y[0]\| \leq C$. As a matter of fact in keeping with our goal of ensuring practical localization, the conditions we obtain are in fact stronger, in that they guarantee exponential rather than just the uniform asymptotic convergence guaranteed by (2.3).

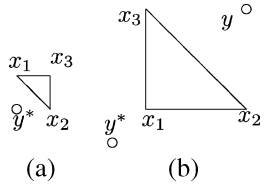

Fig. 1. (a) False unstable stationary point at x_3. (b) False stable stationary point at y.

We provide two examples where (2.3) fails. The first leads to a situation where the resulting false stationary point is locally unstable.

Example 2.1: Thus, consider the 2-D case where $\lambda_i = 1$, $x_1 = [-1,0]^T$, $x_2 = [0,-1]^T$, $x_3 = 0$, and $y^* = [-1,-1]^T$ depicted in Fig. 1(a). In this case, $d_1^2 = d_2^2 = 1$ and $d_3^2 = 2$. Observe in this case, with $y = [y_1, y_2]^T$ the first element of $\partial J/\partial y$ is given by

$$\left((y_1+1)^2 + y_2^2 - 1\right)(y_1+1) + \left((y_2+1)^2 + y_1^2 - 1\right)y_1 + \left(y_1^2 + y_2^2 - 2\right)y_1$$

just as the second element is provided by

$$\left((y_1+1)^2 + y_2^2 - 1\right)y_2 + \left((y_2+1)^2 + y_1^2 - 1\right)(y_2+1) + \left(y_1^2 + y_2^2 - 2\right)y_2.$$

The underlying symmetry of the two expressions ensures that they are simultaneously zero only if

$$y_1 = y_2 = \delta$$

and δ must obey:

$$\begin{aligned}\left((\delta+1)^2 + \delta^2 - 1\right)(2\delta+1) + (2\delta^2 - 2)\delta &= 0\\ \Leftrightarrow 2\delta\left((\delta+1)(2\delta+1) + \delta^2 - 1\right) &= 0\\ \Leftrightarrow 6\delta^2(\delta+1) &= 0.\end{aligned}$$

Thus the only two stationary points are $y = y^*$ and $y = 0$.

In (2.4) $y_1[k+1] - y_1[k]$ equals

$$-\mu\left[2y_1^2[k] + (y_1[k] + y_2[k])^2 + 3y_1[k]\left(y_1^2[k] + y_2^2[k]\right)\right].$$

For sufficiently small $\|y[k]\|$, the first two terms dominate. Thus, if $y_1[k] < 0$, and $y[k]$ is in the vicinity of the origin, $y_1[k+1] < y_1[k]$, exhibiting the local instability of $y = 0$. In practical terms, as is well known (see, e.g., [13, Ch. 2]), the local instability of this stationary point will make it unattainable in that trajectories will rarely stick to it. Nonetheless this stationary point is inconsistent with the requirements of global exponential convergence.

On the other hand, there are examples, e.g., Example 2.2 below where the spurious stationary points are locally stable. This occurs, [14] if the Hessian

$$\mathcal{H} = \frac{\partial}{\partial y}\left[\sum_{i=1}^{N}\lambda_i\left(\|y-x_i\|^2 - d_i^2\right)(y-x_i)\right]$$
$$= 2\sum_{i=1}^{N}\lambda_i(y-x_i)(y-x_i)^T - \sum_{i=1}^{N}\lambda_i\left(\|y-x_i\|^2 - d_i^2\right)I$$

is positive definite at such a stationary point.

Example 2.2: Choose $\lambda_i = 1$, $x_1 = [1,1]^T$, $x_2 = [1,3]^T$, $x_3 = [3,1]^T$, and the true $y^* = 0$ depicted in Fig. 1(b). In this case $d_3^2 = d_2^2 = 10$ and $d_1^2 = 2$. By direct verification it can be seen that $y = [3,3]^T$ is a spurious stationary point. At this point

$$\mathcal{H} = 2\begin{bmatrix}8 & 4\\ 4 & 8\end{bmatrix} - 6I = \begin{bmatrix}10 & 8\\ 8 & 10\end{bmatrix}$$

which is positive definite. Thus, $y = [3,3]^T$ is in fact a local minimum.

It is thus important to consider conditions under which (2.3) holds.

III. Guaranteed Convergence for Given Weights

The last section demonstrates that cost functions of the form of (1.3) may well have local minima. In this section, we provide a sufficient condition under which (2.3) holds for *fixed* $\lambda_i > 0$. There are various equivalent statements of the sufficient condition. One (see Theorem 3.3) is that y^* lies in a certain ellipsoid (or ellipse in two dimensions).

The sufficient condition provided here in fact has broader implications. Not only does it guarantee convergence of gradient descent minimization, but in fact induces exponentially fast convergence.

First some notation: We define the N-vector

$$u_N = [1,\cdots,1]^T, \tag{3.1}$$

the $3\times N$ or $4\times N$ matrix in 2-D and 3-D, respectively,

$$\mathcal{X} = \left[[x_1,\cdots,x_N]^T, u_N\right]^T, \tag{3.2}$$

the 3×1 or 4×1 vector in 2-D and 3-D, respectively,

$$\bar{y} = [y^{*T}, 1]^T \tag{3.3}$$

and $\mathrm{Co}(x_1,\cdots,x_N)$ as the convex hull of $\{x_1,\cdots,x_N\}$. We prove an initial result.

Lemma 3.1: Consider x_i in 2-D or 3-D with Assumption 2.1 holding. Then for every y^* there exist scalar β_i obeying

$$\sum_{i=1}^{N}\beta_i = 1 \tag{3.4}$$

for which

$$\sum_{i=1}^{N}\beta_i x_i = y^*. \tag{3.5}$$

Proof: Observe because of (3.2)–(3.5)

$$\mathcal{X}\beta = \bar{y} \tag{3.6}$$

where

$$\beta = [\beta_1,\cdots,\beta_N]^T. \tag{3.7}$$

Thus, we need only show that

$$\mathrm{rank}[\mathcal{X}] = \dim(\bar{y}). \tag{3.8}$$

If the contrary were true then, with γ a scalar there exists $\theta = [\bar{\theta}^T, \gamma]^T \neq 0$ for which

$$\theta^T\mathcal{X} = 0$$

i.e., for all $i \in \{1,\ldots,N\}$

$$\bar{\theta}^T x_i = -\gamma.$$

Further $\bar{\theta} \neq 0$. This is certainly true when $\gamma \neq 0$. It is also true when $\gamma = 0$, as $\theta \neq 0$. Then in 2-D, the x_i are collinear and in 3-D coplanar, violating Assumption 2.1. ■

If $\beta_i \geq 0$ for all i, then $y^* \in \mathrm{Co}\{x_1,\cdots,x_N\}$. Further, for $N > 3$ in 2-D and $N > 4$ in 3-D, the β_i that obey (3.4), (3.5) are in general non-unique. We now develop a condition (contained in Lemma 3.2 below) involving β_i and λ_i to ensure (2.3).

Define

$$\tilde{x}_i = x_i - y^* \text{ and } \tilde{y} = y - y^*. \tag{3.9}$$

Then (2.3) holds if and only if

$$\sum_{i=1}^{N}\lambda_i\left(\|\tilde{y}-\tilde{x}_i\|^2 - \|\tilde{x}_i\|^2\right)(\tilde{y}-\tilde{x}_i) = 0 \Leftrightarrow \tilde{y} = 0. \tag{3.10}$$

Further because of (3.4) and (3.5)

$$\sum_{i=1}^{N}\beta_i\tilde{x}_i = 0. \tag{3.11}$$

With

$$\Lambda = \mathrm{diag}\{\lambda_1,\cdots,\lambda_N\} \tag{3.12}$$
$$e_i(y) = \tilde{y} - \tilde{x}_i \text{ and } E(y) = [e_1(y),\cdots,e_N(y)] \tag{3.13}$$

as distances are invariant under a coordinate translation

$$\partial J/\partial y = \sum_{i=1}^{N}\lambda_i\left(\|\tilde{y}-\tilde{x}_i\|^2 - \|\tilde{x}_i\|^2\right)(\tilde{y}-\tilde{x}_i)$$
$$= \sum_{i=1}^{N}\lambda_i\tilde{y}^T(\tilde{y}-2\tilde{x}_i)(\tilde{y}-\tilde{x}_i)$$
$$= \sum_{i=1}^{N}\lambda_i\left(2(\tilde{y}-\tilde{x}_i)(\tilde{y}-\tilde{x}_i)^T - (\tilde{y}-\tilde{x}_i)\tilde{y}^T\right)\tilde{y}$$
$$= 2E(y)\Lambda E^T(y)\tilde{y} - E(y)\Lambda u_N\tilde{y}^T\tilde{y}. \tag{3.14}$$

Because of (3.11), and (3.4)

$$E(y)\beta = \tilde{y}. \tag{3.15}$$

Define

$$P = \Lambda(2I - u_N\beta^T) \tag{3.16}$$

and observe from (3.14) and (3.15) that

$$\begin{aligned}\partial J/\partial y &= 2E(y)\Lambda E(y)^T\tilde{y} - E(y)\Lambda u_N\beta^T E(y)^T\tilde{y}\\ &= E(y)PE^T(y)\tilde{y}. \end{aligned} \tag{3.17}$$

Then we provide an intermediate result that concerns a sufficient condition on P for assuring (2.3).

Lemma 3.2: Consider y^*, $x_1,\cdots x_N$ with Assumption 2.1 in place. Suppose $\beta = [\beta_1,\cdots,\beta_N]^T$, obeys (3.4) and (3.5), $\lambda_i > 0$, E is as in (3.9) and (3.13), and P is defined through (3.1), (3.12), and (3.16). Suppose also that

$$P + P^T > 0. \tag{3.18}$$

Then, we have the following:

A) there exists $\alpha > 0$, such that for all y,

$$E(y)(P + P^T)E^T(y) \geq \alpha I; \tag{3.19}$$

B) Condition (2.3) holds.

Proof: Because of (3.17), A) will prove B). In more detail, suppose $\partial J/\partial y = 0$. Then, $\tilde{y}^T\partial J/\partial y = 0$, i.e., $\tilde{y}^T E(y)(P + P^T)E^T(y)\tilde{y} = 0$. Then, (3.19) implies $\tilde{y} = 0$.

Observe now that Assumption 2.1 implies that there exists a $\rho > 0$ such that for all $||\theta|| = 1$, and all y

$$\begin{aligned}\left\|E^T(y)\theta\right\| &= \left\|\begin{bmatrix} y^T\theta - x_1^T\theta \\ \vdots \\ y^T\theta - x_N^T\theta \end{bmatrix}\right\|\\ &\geq \rho,\end{aligned}$$

as otherwise for every $\epsilon > 0$, one can choose a unit θ for which $x_i^T\theta$ are all within ϵ of each other, violating Assumption 2.1. Thus, for all y and θ

$$\left\|E^T(y)\theta\right\| \geq \rho||\theta||.$$

Then A) follows directly from (3.18). ∎

The Lemma provides only a sufficient condition for (2.3). Even if (3.18) is violated, (2.3) will hold unless some $\tilde{y} \neq 0$ is in the null space of $E(y)PE^T(y)$ which itself depends on $\tilde{y}$. Yet, we show below that the condition of the Lemma does quantify a nontrivial domain where (2.3) holds. Second, as we show in Theorem 3.1 under this condition not only does convergence occur but it does so at an exponential rate.

Theorem 3.1: Consider the algorithm update (2.4) and the various quantities defined in Lemma 3.2. Suppose (3.18) holds. Then, for every $M > 0$, there exists a $\mu^*(M)$ such that $y[k] - y^*$ converges exponentially to zero whenever

$$||y[0] - y^*|| \leq M \tag{3.20}$$

and

$$0 < \mu \leq \mu^*(M). \tag{3.21}$$

Proof: Define

$$\tilde{y}[k] = y[k] - y^*.$$

Then, because of (3.17), with

$$Q\left(y[k]\right) = E\left(y[k]\right)PE^T\left(y[k]\right)$$

(2.4) becomes

$$\tilde{y}[k+1] = \tilde{y}[k] - \mu Q\left(y[k]\right)\tilde{y}[k]. \tag{3.22}$$

Thus

$$\begin{aligned}||\tilde{y}[k+1]||^2 = ||\tilde{y}[k]||^2 &- 2\mu\tilde{y}^T[k]Q\left(y[k]\right)\tilde{y}[k]\\ &+\mu^2\tilde{y}^T[k]Q^T\left(y[k]\right)Q\left(y[k]\right)\tilde{y}[k].\end{aligned} \tag{3.23}$$

Because of A) of Lemma 3.2

$$2\tilde{y}^T[k]Q\left(y[k]\right)\tilde{y}[k] \geq \alpha\left||\tilde{y}[k]\right||^2.$$

Also as $Q^T(y[k])Q(y[k])$ is a polynomial function of $\tilde{y}[k]$, for every M, there exists a C such that

$$\left\|Q^T\left(y[k]\right)Q\left(y[k]\right)\right\| \leq C$$

where $\|\cdot\|$ for matrices denote the induced 2-norm, whenever $||\tilde{y}[k]|| \leq M$. Choose

$$\mu^*(M) < \frac{\alpha}{C}. \tag{3.24}$$

Then if $||\tilde{y}[k]|| \leq M$, from (3.23)

$$\begin{aligned}||\tilde{y}[k+1]||^2 &\leq \left(1 - \mu(\alpha - \mu C)\right)||\tilde{y}[k]||^2\\ &\leq \left(1 - \mu\left(\alpha - \mu^*(M)C\right)\right)||\tilde{y}[k]||^2\\ &= \left(1 - \mu\gamma\right)||\tilde{y}[k]||^2\end{aligned} \tag{3.25}$$

where because of (3.24)

$$\gamma = \alpha - \mu^*(M)C > 0.$$

Thus

$$||\tilde{y}[k+1]|| \leq ||\tilde{y}[k]|| \leq M$$

and because of (3.25) exponential convergence obtains. ∎

Observe from (3.16) that P depends only on the λ_i and β_i. Thus, our next goal is to characterize conditions on λ_i and β_i, for which (3.18) holds. To this end, we first present an intermediate lemma.

Lemma 3.3: Consider two $N \times 1$ vectors a, b. Then

$$4I - ab^T - ba^T > 0 \tag{3.26}$$

if and only if

$$\sqrt{(b^Tb)(a^Ta)} < 4 - a^Tb. \tag{3.27}$$

Proof: First, we assert that the characteristic polynomial of $ab^T + ba^T$ i.e., $\det(sI - (ab^T + ba^T))$ is given by

$$s^{N-2}\left(s^2 - 2(b^Ta)s + (a^Tb)^2 - (a^Ta)(b^Tb)\right). \tag{3.28}$$

To see this observe that as $ab^T + ba^T$ has rank 2 for suitable α_i,

$$\det\left(sI - (ab^T + ba^T)\right) = s^{N-2}(s^2 + \alpha_1 s + \alpha_2).$$

It is well known [15, App. 7] that

$$\alpha_1 = -\text{trace}(ab^T + ba^T) = -2\text{trace}(b^T a) = -2b^T a.$$

Also from [15, App. 7]

$$\begin{aligned}\alpha_2 &= -\frac{1}{2}\text{trace}\left((ab^T + ba^T)(ab^T + ba^T + \alpha_1 I)\right) \\ &= -\frac{1}{2}\text{trace}\left((ab^T + ba^T)^2\right) - \frac{\alpha_1}{2}\text{trace}(ab^T + ba^T).\end{aligned}$$

Then (3.28) holds because

$$(ab^T + ba^T)^2 = (b^T a)(ab^T + ba^T) + (a^T a)(bb^T) + (b^T b)(aa^T).$$

From (3.28), it follows that $N-2$ eigenvalues of $ab^T + ba^T$ are at zero, and the remaining two are at $b^T a \pm \sqrt{(a^T a)(b^T b)}$. Thus, the smallest eigenvalue of $4I - ab^T - ba^T$ is $4 - b^T a - \sqrt{(a^T a)(b^T b)}$, whence the result. ∎

Lemmas 3.3 and 3.2 and (3.4) help characterize a region where (2.3) holds in terms of the β_i.

Theorem 3.2: Consider in 2-D or 3-D, x_i, obeying Assumption 2.1, $\lambda_i > 0$, Λ as in (3.12), P defined through (3.1), (3.12), (3.16) and any y^* and β obeying (3.7), (3.4), and (3.5). Then (3.18) holds if and only if

$$(\beta^T \Lambda^{-1} \beta)\left(u_N^T \Lambda u_N\right) = \left(\sum_{i=1}^N \beta_i^2/\lambda_i\right)\left(\sum_{i=1}^N \lambda_i\right) < 9. \quad (3.29)$$

Further (2.3) holds if (3.29) is true.

Proof: Observe that (3.18) is true if and only if

$$4I - \Lambda^{1/2} u_N \beta^T \Lambda^{-1/2} - (\Lambda^{1/2} u_N \beta^T \Lambda^{-1/2})^T > 0.$$

Then Lemma 3.3 proves that (3.29) is necessary and sufficient for (3.18) to hold. [One must use the fact that $\beta^T \Lambda^{-1/2} \Lambda^{1/2} u_n = 1$ by (3.4).] Then Lemma 3.2 proves the result. ∎

Consider now the special case of unity weights, i.e., $\lambda_i = 1$, and $N < 9$. Then we argue that $\text{Co}(x_1, \cdots, x_N)$ is a *proper subset* of the region characterized by Theorem 3.2. Indeed if $y^* \in \text{Co}(x_1, \cdots, x_N)$ then in (3.4) and (3.5) $\beta_i \geq 0$, and

$$\sqrt{\beta^T \beta} \leq \beta^T u_N = 1.$$

Thus as $\Lambda = I$

$$(\beta^T \Lambda^{-1} \beta)\left(u_N^T \Lambda u_N\right) \leq 8.$$

Recalling that in 2-D and 3-D, it suffices to have $N = 3$ and $N = 4$, respectively, for given x_i satisfying Assumption 2.1, the set characterized by Theorem 3.2 can be chosen to be significantly larger than their convex hull. This means that in the sensor localization problem, just a few anchors will achieve substantial geographical coverage, just as in source localization just a few sensors will achieve a large coverage. Of course, there is a benefit to having a large number of sensors as more data can be used. This must be balanced against the fact that when sensors collaboratively localize in a decentralized fashion, larger numbers increase the communications cost in effecting that collaboration. Further larger number of sensors also often increase the number of false minima, defeating a basic premise of practical convergence.

Theorem 3.2 characterizes the set for which (3.18) holds in terms of β, but not directly in terms of y^*. Of course, the β_i themselves have a relationship to y^*. The next theorem exploits this relationship to characterize this set directly in terms of y^*, and in fact shows that the set of y^* satisfying (3.4), (3.5) and (3.29) is an *ellipsoid* (an ellipse in two dimension).

Theorem 3.3: For every $\lambda_i > 0$, and x_i, obeying Assumption 2.1, the set of y^* for which scalar β_i satisfying (3.4), (3.5), and (3.29) exist, is a nonempty ellipsoid, determined entirely by x_i and λ_i.

Proof: Note, (3.29) holds for some choice of λ_i if and only if it holds with any $\delta > 0$, for $\delta\lambda_i$. Thus, without loss of generality, we may assume the λ_i are such that $u_N^T \Lambda u_N = 1$. Then, $\beta_i = \lambda_i$ satisfies (3.4) and (3.29). Thus, the set of y^* is nonempty.

Recall (3.4) and (3.5) combine to give (3.6) and that Assumption 2.1 ensures (3.8).

Consider then for unitary matrices U and V, and positive-definite real diagonal D, the singular value decomposition (SVD) of $\mathcal{X}$, i.e.,

$$\mathcal{X} = U[D, 0]V^T.$$

Here, in 2-D, U, V, and D are 3×3, $N \times N$, and 3×3, respectively. In 3-D, U, V, and D are 4×4, $N \times N$, and 4×4, respectively. Then, (3.6) becomes:

$$U[D, 0]V^T \beta = \bar{y}.$$

Consequently with an *arbitrary* z as a $(N-3)$-vector in 2-D, and as a $(N-4)$-vector in 3-D, β is completely characterized by

$$\beta = V \begin{bmatrix} D^{-1} U^T \bar{y} \\ z \end{bmatrix}.$$

Thus, with the nonsingular matrix

$$F = D^{-1} U^T$$

and for arbitrary z as above, β satisfying (3.4) and (3.5) is completely characterized by

$$\beta = V \begin{bmatrix} F\bar{y} \\ z \end{bmatrix} \quad (3.30)$$

where the nonsingular matrices F and V are determined entirely by x_i. Thus, with A a $N \times N$, positive definite symmetric matrix given by

$$A = \left(u_N^T \Lambda u_N\right) V^T \Lambda^{-1} V$$

i.e., determined by Λ, and the x_i, (3.29) becomes

$$\min_z \left([\bar{y}^T F^T, z^T] A \begin{bmatrix} F\bar{y} \\ z \end{bmatrix} \right) < 9. \quad (3.31)$$

Partition A as

$$A = \begin{bmatrix} A_{11} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix}$$

where in 2-D and 3-D, A_{11} is 3×3 and 4×4, respectively. Observe as $A = A^T > 0$ the Schur's complement obeys

$$B = \left[A_{11} - A_{21}^T A_{22}^{-1} A_{21}\right] > 0. \tag{3.32}$$

Now consider the minimization of

$$G(z) = [\bar{y}^T F^T, z^T] A \begin{bmatrix} F\bar{y} \\ z \end{bmatrix}$$

with respect to z. Observe the minimizing $z = z^*$ must obey

$$\left.\frac{\partial G(z)}{\partial z}\right|_{z=z^*} = 0$$

i.e.,

$$[\bar{y}^T F^T, z^{*T}] A \begin{bmatrix} 0 \\ I \end{bmatrix} = 0$$

leading to

$$z^* = -A_{22}^{-1} A_{21} F\bar{y}.$$

Thus, (3.31) becomes

$$\begin{aligned}[\bar{y}^T F^T, z^{*T}] A \begin{bmatrix} F\bar{y} \\ z^* \end{bmatrix} &= \bar{y}^T F^T [I, -A_{21}^T A_{22}^{-1}] \\ &\quad \times \begin{bmatrix} A_{11} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I \\ -A_{22}^{-1} A_{21} \end{bmatrix} F\bar{y} \\ &= \bar{y}^T F^T B F \bar{y} < 9.\end{aligned}$$

Thus, the set of $\bar{y} = [y^{*T}, 1]^T$ is an ellipsoid. The matrix F is entirely defined by $\mathcal{X}$, i.e., the x_i, and the matrix B is defined using the λ_i and x_i. Hence, since the set of y^* is nonempty it is an ellipsoid determined entirely by x_i and λ_i. ■

IV. Choosing the Weights

In the previous section, we took the λ_i as given, and characterized the y^* and β_i for which (2.3) holds. Now the λ_i are effectively user-chosen parameters. One way of choosing them is to use *a priori* knowledge of the type described in the introduction. However, more in keeping with our objectives of sensor/anchor placement for practical convergence, in this section we quantify the set of y^* for which there exists a set of λ_i so that (2.3) holds. The first question we address is the following. By changing the λ_i, can one alter false stationary points should they exist? An affirmative answer to this question permits one to detect convergence to a false stationary point by changing the λ_i by a nontrivial amount. The Theorem below shows that barring a highly non-generic situation, the answer is indeed yes. This nongeneric situation arises when one of the x_i has exactly the same distance from each of the other x_j as does y^*. In this case x_i is a false stationary point. Example 2.1 exemplifies this situation.

Theorem 4.1: Consider distinct x_i, obeying Assumption 2.1. Suppose for some $y \neq y^*$ and all nonnegative λ_i,

$$\sum_{i=1}^{N} \lambda_i \left(\|y - x_i\|^2 - d_i^2\right)(y - x_i) = 0. \tag{4.1}$$

Then, there exists $K \subset \{1, \cdots, N\}$, $|K| = N - 1$, such that for all $i \in K$ and $j \notin K$, $\|x_i - x_j\| = d_i$. Further, $y = x_j$.

Proof: Since (4.1) holds for the same y and all $\lambda_i \geq 0$, for each $i \in \{1, \cdots, N\}$ either

$$\|y - x_i\|^2 = d_i^2 \tag{4.2}$$

or

$$y = x_i. \tag{4.3}$$

Since $y \neq y^*$ Assumption 2.1 ensures that (4.2) cannot hold for all i. Thus, for some $i = j$ (4.3) holds. Since the x_i are distinct (4.3) holds for only one element of $\{x_1, \cdots, x_N\}$. ■

Since generically the choice of λ_i affects the location and existence of false stationary points, we now characterize conditions under which λ_i exist for the sufficiency condition (3.18) to hold thereby guaranteeing practical convergence for the algorithm. A related question is: Given the existence of such a set of λ_i, how is one to select them? To understand the underlying intuition on this last question, suppose y^* is much closer to x_1 than the other x_i. Intuition suggests that one should emphasize d_1 more than the other d_i, by choosing λ_1 to be relatively larger. This is not just because the distance estimates at a closer location will be more reliable. Rather it is inherent in the underlying geometry of the problem. For example, in an extreme case, if a $d_i = 0$, then this measurement alone suffices for localization, i.e., the other weights can in the limit be selected as zero. The results of this section should be viewed in this context. We first present the following theorem that characterizes regions in question in terms of the β_i.

Theorem 4.2: Under the hypotheses of Theorem 3.2, there exist $\lambda_i > 0$ for which (3.18) holds if and only if

$$\sum_{i=1}^{N} |\beta_i| < 3. \tag{4.4}$$

Further under (4.4)

$$\lambda_i = |\beta_i| \tag{4.5}$$

always guarantees (3.18).

Proof: First observe that should (4.4) hold then under (4.5),

$$\left(\sum_{i=1}^{N} \beta_i^2/\lambda_i\right)\left(\sum_{i=1}^{N} \lambda_i\right) = \left(\sum_{i=1}^{N} |\beta_i|\right)^2 < 9.$$

Thus, from Theorem 3.2, (3.18) holds. Further, from the Cauchy–Schwarz inequality

$$\left(\sum_{i=1}^{N} |\beta_i|\right)^2 = \left(\sum_{i=1}^{N} \frac{|\beta_i|}{\sqrt{\lambda_i}} \sqrt{\lambda_i}\right)^2 \leq \left(\sum_{i=1}^{N} \beta_i^2/\lambda_i\right)\left(\sum_{i=1}^{N} \lambda_i\right).$$

Thus, the violation of (4.4) implies the violation of (3.29) and hence (3.18). ■

Observe (4.5) in particular is in accord with the intuition we have foreshadowed. In particular, if y^* is very close to x_1 then there exists at least one choice of the β_i, such that β_1 has a much

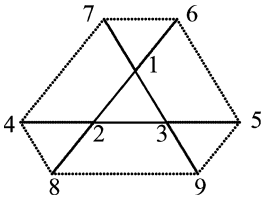

Fig. 2. Illustration of the polytope suggested by Theorem 4.2.

larger magnitude than the remaining β_i. For such a choice, (4.5) forces λ_1 to be much larger than the remaining λ_i.

Nonetheless, this result is in terms of β_i and thus only indirectly characterizes the spread of the geographical area that guarantees (3.18). Theorem 4.3 below provides a direct characterization.

Theorem 4.3: With x_i, obeying Assumption 2.1, the largest set of y^* for each member of which one can find a choice of nonnegative λ_i that guarantee (3.18), has the following properties:

a) it is a convex polytope'
b) it has $\mathrm{Co}\{x_1, \cdots, x_N\}$ as a proper subset;
c) it can be quantified by a linear program that is entirely determined by the x_i.

Proof: Clearly, this set of y^* is polytopic as it satisfies (4.4) and (3.4). Convexity is easily verified. Thus, because of (3.5), it is also polytopic in y^*. Further as under (3.4), $\beta_i \geq 0$ for all i guarantees (4.4), $\mathrm{Co}\{x_1, \cdots, x_N\}$ is a proper subset of this polytope. To prove c), observe from the proof of Theorem 3.3, for a given y^* the set of β_i that satisfy (3.4) and (3.5) is given by (3.30), with F and V determined entirely from the x_i and z a completely free vector. Thus, the set we seek is characterized by y^* that obey, under (3.3),

$$\min_z \left\| V \begin{bmatrix} F\bar{y} \\ z \end{bmatrix} \right\|_1 < 3. \tag{4.6}$$

where $\|\cdot\|_1$ denotes the vector 1-norm, which of course can be obtained by solving a linear program determined entirely from the x_i. ∎

The fact that a linear program can determine this polytope is of course attractive from a computational view point. While this polytope contains $\mathrm{Co}\{x_1, \cdots, x_N\}$, it is in fact much larger than just $\mathrm{Co}\{x_1, \cdots, x_N\}$. Indeed consider the 2-D example depicted in Fig. 2 where 1, 2, and 3 represent the x_i locations. Choose $\beta_1 = 0$. Then y^* satisfying (3.4) and (4.4) are in the interval (4,5). Here [2,3] is a closed subinterval of (4,5), and the lengths of the segments joining 4 and 2, 3 and 5, and 2 and 3 are all equal. By similarly extending [1,3] and [1,2], one could come up with a hexagon 6, 7, 4, 8, 9, 5 that defines the desired polytope.

Note also that though in this set (4.5) provides a choice of the λ_i, these are not the only choice one can make. The more one enters the interior of this polygon, the more the available choices of λ_i, and indeed the larger the region where a common set of λ_i guarantees (2.3). This in particular has implications to the positioning of the x_i. Thus, with a potentially rough estimate of the position of a source, groups of sensors can collaborate to determine whether they can settle upon a λ_i which ensures (2.3). This provides guidance on how to deploy fewer sensors to achieve greater coverage.

Similar conclusions will follow for two-dimensional examples with $N > 3$, and for three-dimensional examples.

At the same time it should be noted that when the source is located near the boundaries of these polytopes, not only is the choice of available λ_i limited, but the convergence becomes less robust, and performance is degraded. This performance degradation also occurs when the source is close to the boundaries of the ellipses characterized in the previous section. Thus, while the polytopes and ellipses characterized in this paper serve as theoretical benchmarks, in practical terms the regions for practical convergence should have sufficient cushions.

V. Simulations

Section II-A already contains a simulation example in $\Re^2$, where a noisy set of measurements led to a vastly inaccurate estimate using a linear algorithm, but produced much improved localization using (2.4). In this section we provide two sets of simulation results. In both cases we consider a signal model based on the RSS distance measurement mechanism described by (1.1) that is affected by a log-normal shadowing term ω_s. Assuming that other measurement noises in the RSS mechanism are insignificant in comparison to the log-normal shadowing, the shadowing effect can be included in the signal model (2.4) as follows:

$$s = \omega_s A/d^{\eta} \tag{5.1}$$

where $\omega_s[\mathrm{dB}] \triangleq 10\log\omega_s$ is a zero-mean Gaussian noise, i.e., $\omega_s[\mathrm{dB}] \sim \mathcal{N}(0, \sigma_s^2)$. The model (5.1) is often expressed, considering all signal measurements in decibels, as

$$s[\mathrm{dB}] = A[dB] - 10\eta\log(d) + \omega_s[\mathrm{dB}]. \tag{5.2}$$

where $s[\mathrm{dB}] \triangleq 10\log s$, $A[\mathrm{dB}] \triangleq 10\log A$. $\omega_s[\mathrm{dB}] \triangleq 10\log\omega_s$.

The first case considers comparison between our algorithm and the ML estimate. The second set provides a comparison with the POCS algorithms of [9] and [10]. Given our underlying motivation of sensor/anchor placement in a manner that a *small number of sensors/anchors* are responsible for practical localization in the regions determined in the previous sections, in both cases we consider a small number of well-placed sensors/anchors.

It should be emphasized that the use of the RSS model is purely for illustration purposes. Our algorithm does not require such a model but simply assumes that distances have somehow been acquired. As such, similar to [10], we assume here that, each single distance measurement/estimate $\hat{d}$ corresponding to the sensor-source distance d above is obtained using maximum likelihood estimation as

$$\hat{d} = (A/s)^{1/\eta}. \tag{5.3}$$

In the simulations, the values of the source signal strength and the power loss coefficient are taken as $A[\mathrm{dB}] = 30$ dB and either $\eta = 3$ or $\eta = 4$.

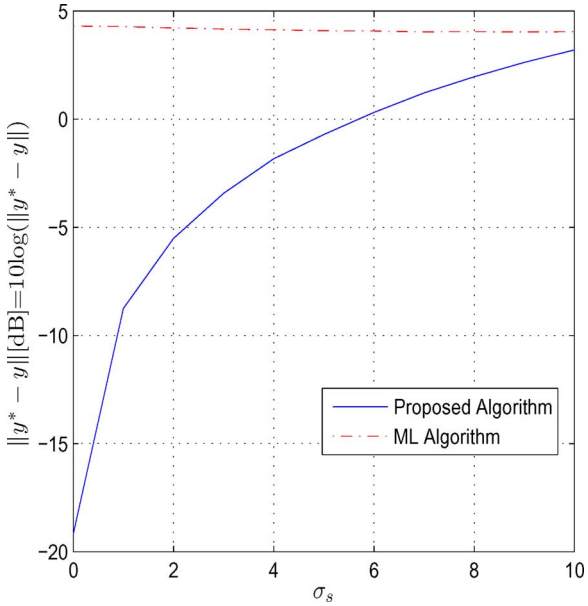

Fig. 3. Example of poor performance of ML localization: y^* source location and y its estimate.

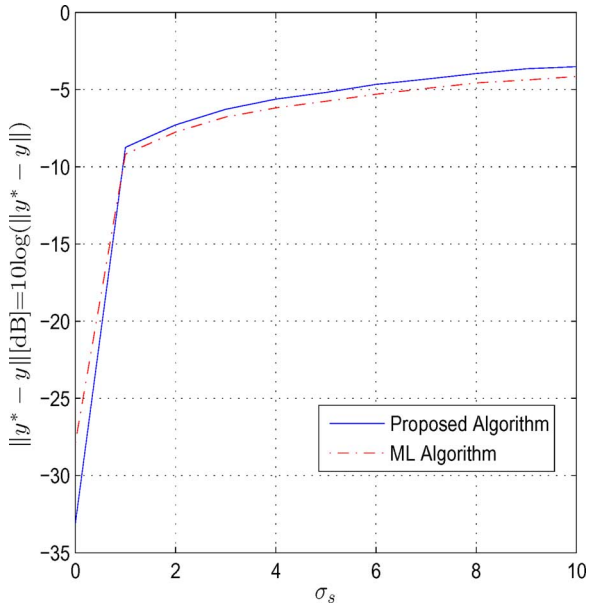

Fig. 4. Example of good performance of ML localization.

A. Comparison With ML Estimation

It is readily seen that the ML-estimate of the source position can be obtained by minimizing the following cost function:

$$J_{\mathrm{ML}}(y) = \sum_{i=1}^{N} \left[\log s_i - \log \left(\frac{A}{\|x_i - y\|^{\eta}} \right) \right]^2. \qquad (5.4)$$

In this section, we compare the gradient descent minimization of (5.4) with that of (1.3). We choose $\eta = 3$, four sensors at $[-2,-1]'$, $[-1,-3]'$, $[-1,1]'$, and $[1,0]'$, and a source located at $[0,-1]'$. Our goal is to illustrate the sensitivity of the ML-algorithm to the presence of false minima. In this example, (5.4) has a local minimum in the vicinity of $[-2.5,.3]'$. Thus, in the first example depicted in Fig. 3, we choose 50 randomly selected initial conditions lying in the square $[-4,-2]\times[0,1]$. Specifically for each value of $\sigma_s \in [0,10]$, we obtained the mean square localization error by averaging over each of these 50 initial conditions over 100 runs per initial condition. Each individual run uses a fixed value of noise, and has $\mu = 0.01$. For the algorithm of this paper, $\lambda_i = 1$. One observes that while the proposed algorithm performs well the performance of the ML algorithm is poor. An interesting fact is that for large values of σ_s, ML localization actually improves. This is so because for large noise the probability that the estimates are outside the basin of attraction of the false minimum is higher.

The second example, depicted in Fig. 4, uses precisely the same parameters as above but with gradient descent randomly initialized from the square $[-0.5,0.5]\times[-1.5,-0.5]$. Observe now the ML-based estimation performs much better than previously. Nonetheless, the performance of the proposed algorithm comparable even though it is, is outperformed at high noise variances. The fact that the ML-based algorithm is outperformed at small values of σ_s can be attributed to the fact that in this non-Gaussian noise setting the ML estimate is not the MMSE estimate.

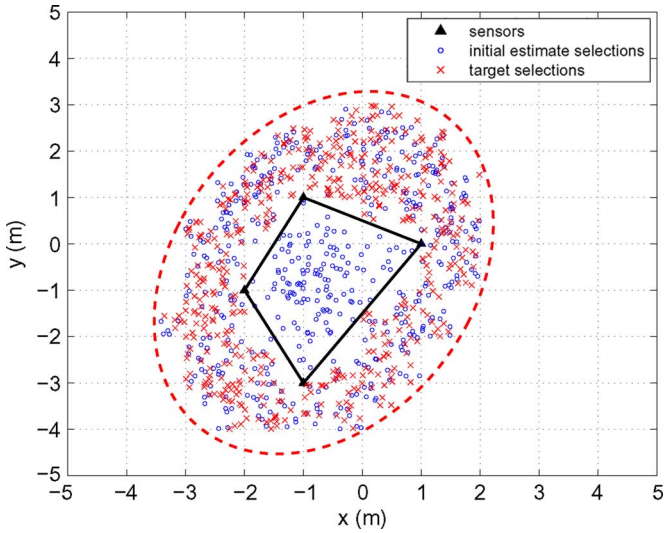

Fig. 5. Ellipse defined in Theorem 3.3 for the four sensors shown is plotted in red, with out-of-hull sources and random initial conditions.

B. Relative Performance With POCS

We consider settings where the λ_i, A, η and the sensor locations are identical to those in the previous section. The goal is to compare the performance of our algorithm with the circular POCS algorithm of [9] and the circular/hyperbolic hybrid POCS algorithm of [10].

Two sets of randomly selected pairs of source locations and initial conditions are used. Each comprises 500 pairs and is depicted in Figs. 5 and 6. In Fig. 5 depicting the first set, the random sources are outside the convex hull of the sensors but inside the corresponding ellipse defined in Theorem 3.3, depicted as the red dotted line contour. Fig. 6 depicts the second set where the sources are all inside the convex hull of the sensors. In both cases the red crosses are the sources and the blue circles are the initial conditions.

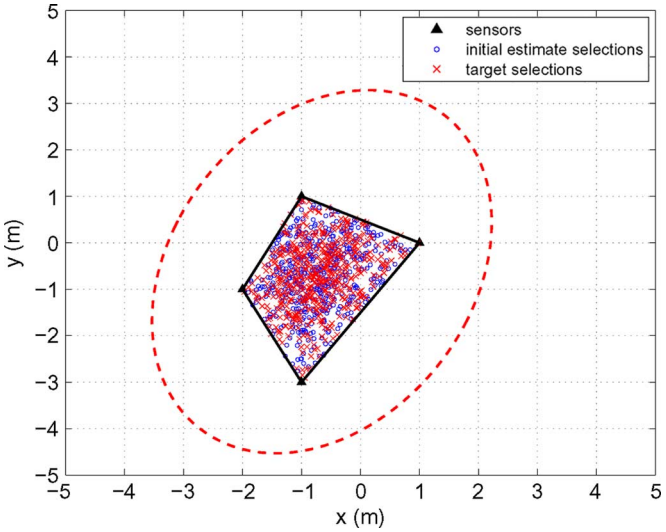

Fig. 6. Ellipse defined in Theorem 3.3 for the four sensors shown is plotted in red, with sources inside the convex hull and random initial conditions.

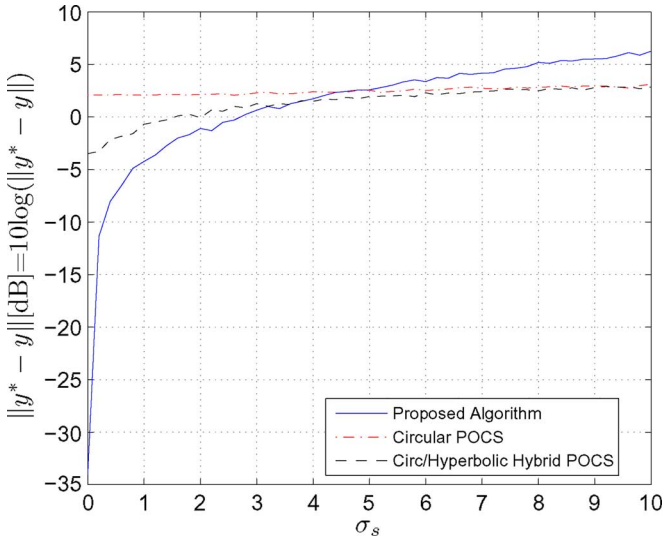

Fig. 8. Average estimation error versus standard deviation of the shadowing noise, with out-of-hull sources.

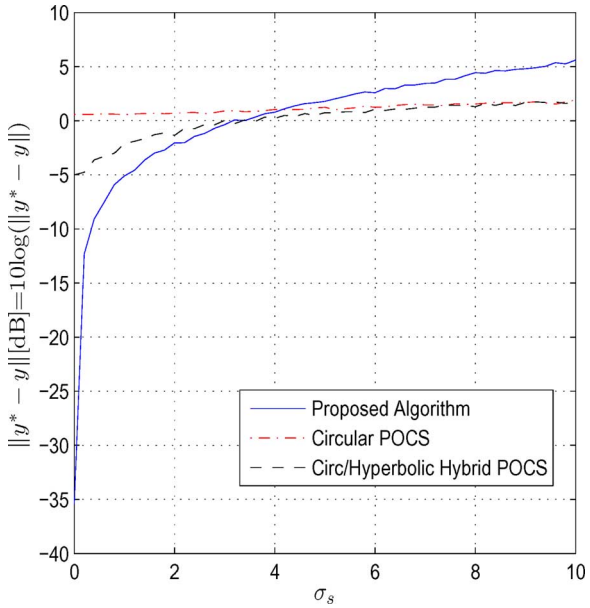

Fig. 7. Comparison with the POCS algorithms.

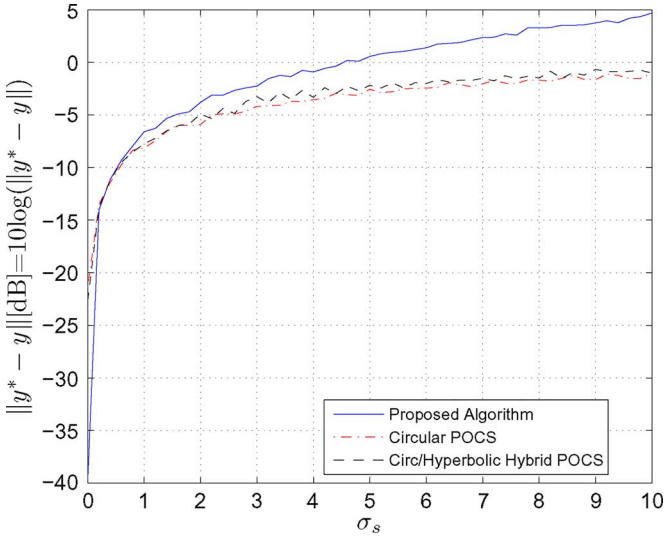

Fig. 9. Average estimation error versus standard deviation of the shadowing noise, with in-hull sources.

Observe that the ellipse characterized by Theorem 3.3 is much larger than the convex hull of the sensors.

Fig. 7 illustrates the square of the estimation error averaged over 1000 runs, with each run corresponding to one pair of the source location and initial condition from the two sets described above, as a function of the standard deviation σ_s of the shadowing noise. The measurement set generated for each particular run is kept the same for all the 1000 steps of this particular run.

The algorithm parameters used in these simulations are as follows: For the algorithm (2.4), $\mu = 0.001$. For the circular POCS (according to the notation in [9]), $\lambda_k = 1$ for $k \leq 320$ and $\lambda_k = 4/(k-316)$ for $k > 320$. The circular/hyperbolic hybrid POCS is implemented as 500 steps of hyperbolic POCS followed by 500 steps of circular POCS. For the hyperbolic POCS part (according to the notation in [10]), $K = 0.1$, $\lambda_k = 1$ for $k \leq 150$ and $\lambda_k = 3/(k-147)$ for $k > 150$. For the circular POCS part (again according to the notation in [9]), $\lambda_k = 1$ for $k \leq 160$ and $\lambda_k = 2/(k-158)$ for $k > 160$.

Observe that even though the area of the complement of the convex hull inside the ellipse is larger than that of the convex hull, equal numbers of source locations have been selected from each set. The performance of the proposed algorithm is better than that of circular and hybrid POCS until about $\sigma_s/\eta = 4/3$, and the situation reverses after that. *A similar reversal was noted for $\eta = 4$, where the point of reversal was roughly* 6/4.

To get a better insight into the relative performance with sources located inside and outside the convex hull of the sensors, consider Figs. 8 and 9, that respectively depict the relative performance in these two cases. While for the out of hull case the relative performance mirrors that above, in the inside the hull case the POCS algorithms outperform the proposed algorithm earlier. *This however, buttresses the thesis of this paper that this algorithm is desirable when convergence of competing algorithms is in doubt, particularly as the area*

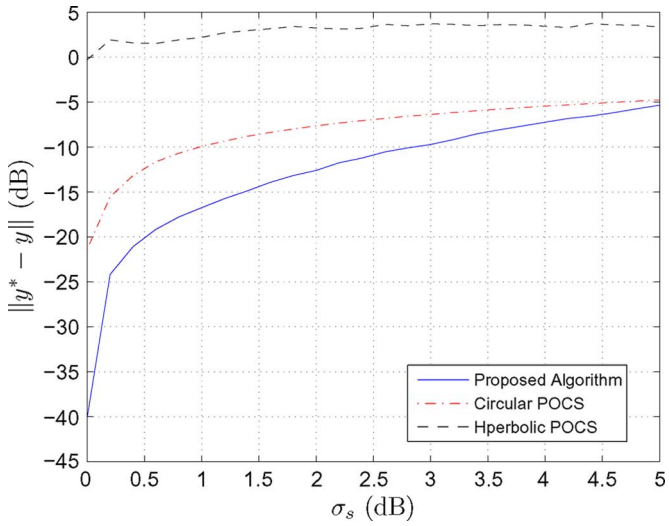

Fig. 10. Average estimation error versus standard deviation of the shadowing noise, with continuous source emission, and in-hull sources.

inside the convex hull is much smaller than that outside it, but inside the ellipse.

The foregoing assumed that only one distance measurement is acquired in any single localization setting. However, when the source emission is continuous, a new distance estimate can be acquired at each iteration of the cognizant algorithms. The relative in-hull performance is depicted in Fig. 10, with $\eta = 4$. It is seen that the proposed algorithm outperforms its POCS counterparts for a longer interval of σ_s.

VI. Conclusion

We have studied conditions under which a localization algorithm involves a globally exponentially convergent gradient descent minimization problem. In particular given a set of nodes with known positions (e.g., sensors), this algorithm seeks to localize an object (e.g., a source) whose distances from these nodes are available. Given a set of such sensors and a set of weights we characterize a nontrivial ellipsoidal geographic region surrounding these sensors such that a source lying in this region can be localized through a minimization described above. We also characterize a polytopic region for which there exist weights that permit similar localization. These characterizations provide guidance for placing sensors/anchors to achieve a desired level of geographical coverage.

Even though the exponential convergence in the noise free case demonstrated by us, guarantees graceful performance degradation in the presence of noise, an explicit convergence analysis in the noisy case represents a useful line of future research.

References

[1] *IEEE Signal Process. Mag.*, vol. 22, no. 4, Jul. 2005.

[2] J. Aspnes, T. Eren, D. Goldenberg, A. S. Morse, W. Whiteley, Y. Yang, B. D. O. Anderson, and P. Belhumeur, "A theory of network localization," *IEEE Trans. Mobile Comput.*, vol. 5, no. 12, pp. 1663–1678, Dec. 2006.

[3] G. Mao, B. Fidan, and B. D. O. Anderson, "Localization," in *Sensor Network and Configuration*, N. P. Mahalik, Ed. New York: Springer-Verlag, 2006, ch. 13, pp. 281–315.

[4] A. Tarighat and A. H. Sayed, "Network based wireless localization," *IEEE Signal Process. Mag.*, pp. 24–40, Jul. 2005.

[5] S. Dandach, B. Fidan, S. Dasgupta, and B. D. O. Anderson, "Adaptive source localization with mobile agents," in *Proc. CDC*, San Diego, CA, Dec. 2006, pp. 2045–2050.

[6] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.

[7] M. G. Rabbat and R. D. Nowak, "Decentralized source localization and tracking," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montreal, QC, Canada, May 2004, vol. III, pp. III-921–III-924.

[8] M. G. Rabbat and R. D. Nowak, "Distributed optimization in sensor networks," in *Proc. 3rd Int. Symp. Information Processing in Sensor Networks*, Berkeley, CA, Apr. 2004, pp. 20–27.

[9] D. Blatt and A. O. Hero, "Energy-based sensor network source localization via projection onto convex sets," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3614–3619, Sep. 2006.

[10] M. Rydstrom, E. G. Strom, and A. Svensson, "Robust sensor network positioning based on projection onto circular and hyperbolic convex sets (POCS)," presented at the SPAWC, Cannes, France, Jul. 2006.

[11] M. Cao, B. D. O. Anderson, and A. S. Morse, "Sensor network localization with imprecise distance measurements," *Syst. Control Lett.*, vol. 55, pp. 87–93, 2006.

[12] R. R. Bitmead, "Convergence in distribution of LMS-type adaptive parameter estimates," *IEEE Trans. Autom. Control*, vol. 28, no. 1, pp. 54–60, Jan. 1983.

[13] H. K. Khalil, *Nonlinear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[14] M. Vidyasagar, "Lyapunov stability," in *Nonlinear Systems Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1993, ch. 5.

[15] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

Barış Fidan (M'03) received the B.S. degrees in electrical engineering and mathematics from Middle East Technical University, Turkey, in 1996, the M.S. degree in electrical engineering from Bilkent University, Turkey, in 1998, and the Ph.D. degree in electrical engineering at the University of Southern California, Los Angeles, in 2003.

He has been with National ICT Australia and the Research School of Information Sciences and Engineering of the Australian National University, Canberra, Australia, since 2005, where he is currently a Senior Researcher. His research interests include autonomous formations, sensor networks, cooperative localization, adaptive and nonlinear control, switching and hybrid systems, mechatronics, and various control applications.

Soura Dasgupta (F'98) was born in Calcutta, India, in 1959. He received the B.E. degree in electrical engineering from the University of Queensland, Australia, in 1980, and the Ph.D. in systems engineering from the Australian National University, Canberra, in 1985.

He is currently Professor of electrical and computer engineering at the University of Iowa, Iowa City. In 1981, he was a Junior Research Fellow in the Electronics and Communications Sciences Unit at the Indian Statistical Institute, Calcutta. He has held visiting appointments at the University of Notre Dame, University of Iowa, Université Catholique de Louvain-La-Neuve, Belgium, and the Australian National University. His research interests are in controls, signal processing, and communications.

Between 1988 and 1991, and 2004 and 2007, he respectively served as an Associate Editor of the IEEE Transactions on Automatic Control and the IEEE Transactions on Circuits and Systems-II. He is a corecipient of the Gullimen Cauer Award for the best paper published in the IEEE Transactions on Circuits and Systems in the calendar years of 1990 and 1991, a past Presidential Faculty Fellow, an Associate Editor for the IEEE Control Systems Society Conference Editorial Board, a subject editor for the *International Journal of Adaptive Control and Signal Processing*, and a member of the editorial board of the *EURASIP Journal of Wireless Communications*.

Brian D. O. Anderson (LF'06) was born in Sydney, Australia, and received the undergraduate degrees in pure mathematics in 1962 and electrical engineering in 1964 from the University of Sydney, Australia, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1966.

Following completion of his education, he worked in industry in Silicon Valley and served as a faculty member in the Department of Electrical Engineering at Stanford University. He was Professor of electrical engineering at the University of Newcastle, Australia, from 1967 until 1981 and is now a Distinguished Professor at the Australian National University and Distinguished Researcher in National ICT Australia, Ltd. His interests are in control and signal processing.

Dr. Anderson is a Fellow of the Royal Society London, the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, the Honorary Fellow of the Institution of Engineers, Australia, and the Foreign Associate of the U.S. National Academy of Engineering. He holds doctorates (honoris causa) from the Universit Catholique de Louvain, Belgium, Swiss Federal Institute of Technology, Zürich, and the University of Sydney, the University of Melbourne, the University of New South Wales, and the University of Newcastle. He served a term as President of the International Federation of Automatic Control from 1990 to 1993 and as President of the Australian Academy of Science between 1998 and 2002. His awards include the IEEE Control Systems Award of 1997, the 2001 IEEE James H. Mulligan, Jr. Education Medal, and the Guillemin–Cauer Award, IEEE Circuits and Systems Society in 1992 and 2001, the Bode Prize of the IEEE Control System Society in 1992, and the Senior Prize of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING in 1986.