2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Distributed Joint Localization and Clock Synchronization in TOA-based Sensor Networks Using Joint Rigidity Theory

Ruixin Wen, Eric Schoof and Airlie Chapman

Abstract— **This paper studies the joint localization and clock synchronization problem in a time-of-arrival-based sensor network employing joint rigidity theory. First, we study the topological conditions and prior information requirements for uniquely determining the network position and clock parameters. Second, we propose a distributed algorithm for the joint localization and clock synchronization problem and discuss the approaches for algorithm initialization. Finally, we analyze the poor conditioning of the problem and propose two possible methods that result in better conditioning as well as maintaining the distributed manner.**

I. INTRODUCTION

Localization and clock synchronization are two critical aspects of self-organizing location-aware wireless sensor networks. The network localization problem determines the position of sensors in the network, which is highly desirable in a variety of applications relating to environmental monitoring and surveillance [1]. The network clock synchronization problem calibrates the drift between the independent clock of each sensor in the network, enabling communication coordination and distributed data fusion [2]. Distributed methods for these two problems offer advantages in terms of fault tolerance and network scalability, which have been separately studied in the literature [3]–[5]. The distributed localization algorithm estimates the sensor positions by using knowledge of the absolute positions of a limited number of sensors (called anchors) and distance or bearing measurements between network connected sensors [3], [4]. The distributed clock synchronization algorithm estimates the clock parameters using a consensus-based method [5].

The position and clock parameters are tightly coupled in time-of-arrival (TOA) based ranging methods, where the distance is calculated from the time-of-flight as the signal travels with the speed of light. This coupled relationship has attracted research interest in the joint localization and clock synchronization problem [6]–[11]. This problem aims to estimate the position and clock parameters in the network simultaneously. Two-step methods are traditionally used, performing clock synchronization first to provide inter-sensor distance estimates for the subsequent localization [6], [7]. Such two-step methods have reduced accuracy as strict coupling between the position and clock parameters is not enforced. Centralized methods have been proposed for simultaneously solving the problems in one step [8], [9], which only locate and synchronize one node in the network. Distributed methods have been studied based on the probabilistic model between sensor node pairs [10], [11], where the network topologies are not fully leveraged. Joint rigidity theory proposed in [12] argues that, as long as specific topological conditions are satisfied, the position and clock parameters in a wireless sensor network can be solved up to some trivial variations based on only one round of TOA measurements. Joint rigidity theory elaborates on the impact of network topology in determining the position and clock parameters in the network, providing an opportunity to apply existing rigidity-related tools in TOA-based joint localization and clock synchronization.

In this paper, the joint localization and clock synchronization problem is formulated as an optimization problem. First, we study the condition for a unique local solution to the optimization problem based on joint rigidity theory. Such a condition indicates the requirements for uniquely determining the sensor position and clock parameters in the network. Second, we propose a distributed gradient-descent algorithm that exhibits local convergence. An initialization method is proposed to overcome the difficulty in determining the initial estimate of clock parameters. Finally, we analyze the poor conditioning of the optimization problem and discuss two possible methods that result in better conditioning while maintaining the distributed property of the algorithm.

The paper is structured as follows. Section II reviews joint rigidity theory. Section III formulates the localization and clock synchronization problem and studies the condition for a unique solution. Section IV proposes a distributed algorithm and an initialization method for local convergence. Section V analyzes the ill-conditioning of the problem and proposes two possible solutions. Simulation examples are also provided. Section VI concludes the paper and discusses the future work.

Notation: The vectors $\mathbf{0}$ and $\mathbf{1}_n$ denote the vector of all zero entries and the $n \times 1$ vector of all ones, respectively. The identity matrix is denoted as $I_n \in \mathbb{R}^{n \times n}$. Let $\|\cdot\|$ be the Euclidean norm of a vector or the 2-norm of a matrix. The closed ball of radius $r > 0$ centered at a vector x is denoted by $\mathcal{B}_r[x]$. Consider matrices $A, B \in \mathbb{R}^{p \times q}$. The rank and null space of A are denoted by $\text{rank}(A)$ and $\text{Null}(A)$, respectively. The minimum and maximum eigenvalue of a symmetric matrix A is denoted as $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. The Kronecker product of A and B is written as $A \otimes B$, and $\text{diag}\{A, B\}$ denotes a diagonal block matrix whose diagonal blocks are A and B. An elemental rotation in d-dimensional space is a rotation about the $(d-2)$-dimensional subspace

This work is supported by the University of Melbourne through Melbourne Research Scholarship. R. Wen is with the Department of Mechanical Engineering, The University of Melbourne, Victoria 3010, Australia ruixinw1@student.unimelb.edu.au E. Schoof is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010, Australia eschoof@unimelb.edu.au A. Chapman is with the Department of Mechanical Engineering, The University of Melbourne, Victoria 3010, Australia airlie.chapman@unimelb.edu.au

containing a set of $(d-2)$ vectors in the standard basis. Matrix J_d^i denotes the infinitesimal generator of the ith elemental rotation in d-dimensional space, where $i \in \{1, 2, ..., d(d-1)/2\}$. For example, for $d = 2$ and $d = 3$,

$$J_2^1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$
$$J_3^1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, J_3^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, J_3^3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

II. Joint Rigidity Theory

Joint rigidity theory [12] studies the topological conditions under which a wireless sensor network is localizable and clock synchronizable up to some trivial variations based on one round of TOA timestamps. In this section, we review the concept of infinitesimal joint rigidity and how it is determined in wireless sensor networks.

A wireless sensor network is termed as a TOA-based sensor network if its sensors can measure high-precision timestamps for signal transmission and reception. One example is the ultrawideband sensor network [13], which is widely considered in joint localization and clock synchronization problems. Consider a TOA-based sensor network where every sensor node has its own position and an independent clock. For a signal sending from node i to node j, the sending and receiving timestamps at their local time frames are denoted by $T^i_{(i,j)}$ and $T^j_{(i,j)}$, respectively. The position of node i is denoted by $p_i \in \mathbb{R}^d$. The clock of node i is characterized by two constant clock parameters, clock skew w_i and clock offset ϕ_i. The clock offset and clock skew correspond to the node's time difference and frequency difference, respectively, compared to a global clock. The clock parameters follow the first-order clock model, $t_i = w_i t + \phi_i$, where t is the global reference time and t_i is the local time at node i. For a sensor network with n nodes deployed in d-dimensional space, we denote a position configuration $p = [p_1^T, ..., p_n^T]^T \in \mathbb{R}^{dn}$ and a clock configuration $\varphi = [\varphi_1^T, ..., \varphi_n^T]^T \in \mathbb{R}^{2n}$ with $\varphi_i = [\alpha_i, \beta_i]^T$, where $\alpha_i = w_i^{-1} \in \mathbb{R}^+$ and $\beta_i = -w_i^{-1}\phi_i \in \mathbb{R}$. Thus, $t = \alpha_i t_i + \beta_i$. For brevity, we refer to α_i and β_i as simply the clock skew and clock offset, respectively.

The topology of a TOA-based sensor network can be represented by a directed graph $\mathcal{D} = (\mathcal{V}, \mathcal{E})$, where the vertex set $\mathcal{V}$ represent the sensor nodes and the directed edge set $\mathcal{E}$ represents a signal transmission from its tail node (where the arrow starts) to its head node. Suppose that the cardinality $|\mathcal{V}| = n$ and $|\mathcal{E}| = m$. A TOA-based sensor network can be represented by a position-clock framework in $\mathbb{R}^{d+2}$, denoted by $(\mathcal{D}, \sigma)$, where $\sigma = [p^T, \varphi^T]^T \in \mathbb{R}^{n(d+2)}$ is the position-clock configuration of the network. Consider a signal transmitted from one node to the other, the distance between the node pair equals the product of the speed of light and the time of flight. Thus, a position-clock framework satisfies the following constraint for every edge $(i, j) \in \mathcal{E}$,

$$\|p_i - p_j\| - c(\alpha_j T^j_{(i,j)} + \beta_j - \alpha_i T^i_{(i,j)} - \beta_i) = 0, \quad (1)$$

where c is the speed of light. Note that the timestamps are measured in the given framework $(\mathcal{D}, \sigma)$, where the sending timestamps $T^i_{(i,j)}$ are typically user-defined.

Given a position-clock framework $(\mathcal{D}, \sigma)$ and its TOA timestamp measurements, we consider a variable $\hat{\sigma} = [\hat{p}^T, \hat{\varphi}^T]^T$ and define the following function based on (1),

$$f_k(\hat{\sigma}) = \|\hat{p}_i - \hat{p}_j\| - c(\hat{\alpha}_j T^j_{(i,j)} + \hat{\beta}_j - \hat{\alpha}_i T^i_{(i,j)} - \hat{\beta}_i), \quad (2)$$

where k represents the node pair $(i, j) \in \mathcal{E}$ for some edge ordering $k \in \{1, ..., m\}$, denoted as $k \sim (i, j) \in \mathcal{E}$. Define the **TOA constraint function** $F(\hat{\sigma}) = [f_1(\hat{\sigma}), ..., f_m(\hat{\sigma})]^T \in \mathbb{R}^m$. Thus, we have $F(\sigma) = \mathbf{0}$.

The **joint rigidity matrix** is defined as

$$R(\hat{\sigma}) = \frac{\partial F(\hat{\sigma})}{\partial \hat{\sigma}} \in \mathbb{R}^{m \times n(d+2)}. \quad (3)$$

Due to the term $\|\hat{p}_i - \hat{p}_j\|$ in (2), the domain of $R(\hat{\sigma})$ does not include the points with $\hat{p}_i = \hat{p}_j$ for all $(i, j) \in \mathcal{E}$. Such a case corresponds to collocated sensors, which is *non-generic* in sensor networks. The definition here slightly differs from the definition in [12] due to the variation of the constraint (1). When $\hat{\sigma} = \sigma$, $R(\sigma)$ is a scaled version of the joint rigidity matrix defined in [12], hence all the properties and theorems provided in [12] still hold for the current $R(\sigma)$. The null space of $R(\sigma)$ characterizes the variations that preserve the TOA constraints, termed infinitesimal variations. Given TOA timestamp measurements, a position-clock framework is called **infinitesimally joint rigid** if all its infinitesimal variations are trivial, i.e., one of the following variations or a combination of them: a position translation, a position rotation, a clock offset translation, and a position-clock scaling. The following theorem states the corresponding rank condition on $R(\sigma)$, where J_d^i is defined in the notation subsection.

Theorem 1 ([12]): Given TOA timestamp measurements, for a position-clock framework $(\mathcal{D}, \sigma)$ in $\mathbb{R}^{d+2}$, the following conditions are equivalent:

1) $(\mathcal{D}, \sigma)$ is infinitesimally joint rigid;
2) $\text{rank}(R(\sigma)) = (d+2)n - d(d+1)/2 - 2$;
3) $\text{Null}(R(\sigma)) = \text{span}\{[\mathbf{1}_n^T \otimes I_d, \mathbf{0}^T]^T, [\mathbf{0}^T, \mathbf{1}_n^T \otimes [0, 1]]^T, \sigma, [p^T(I_n \otimes J_d^1)^T, \mathbf{0}^T]^T, ..., [p^T(I_n \otimes J_d^{\frac{d(d-1)}{2}})^T, \mathbf{0}^T]^T\}$.

Note that the basis of $\text{Null}(R(\sigma))$ in 3) represents the trivial infinitesimal variations mentioned above. As these trivial variations exist in all TOA-based sensor networks, statement 2) presents the maximum rank of $R(\sigma)$, which is achieved when the framework is infinitesimally joint rigid. Following the existing research on rigidity theory [14], [15], the infinitesimal joint rigidity largely depends on the graph topology rather than the configuration. A graph is called **generically joint rigid** if its framework is infinitesimally joint rigid for almost all position-clock configurations and TOA timestamps. Thus, when *generic* configurations and timestamps are considered, an infinitesimally joint rigid framework can be determined or constructed by satisfying the generically joint rigid graph condition. The corresponding test and construction method refer to [12]. Examples of generically joint rigid graphs are given in Fig. 1.

III. Problem Formulation

The joint localization and clock synchronization problem aims to uniquely determine the node positions and clock parameters in a TOA-based sensor network. In this section,

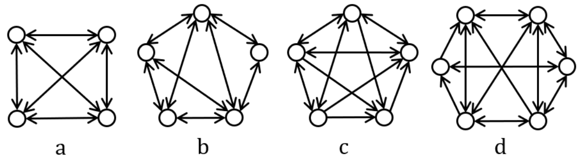


Fig. 1. Generically joint rigid graph examples in $\mathbb{R}^{2+2}$. The double-headed arrow indicates the bidirectional TOA measurements between node pairs.

we formulate the problem as an optimization problem and present the conditions on the prior knowledge of sensor nodes to guarantee that the true configuration of the sensor network is an isolated local minimizer of a position-clock cost function.

Joint rigidity theory provides the topological conditions that ensure localizability and clock synchronizability up to trivial variations. To uniquely localize and synchronize the network, prior knowledge of some nodes' positions and clock parameters is required to restrict these trivial variations. The nodes with known positions/clock skews/clock offsets are termed position/skew/offset anchors, respectively.

Consider a TOA-based sensor network, represented by the position-clock framework $(\mathcal{D}, \sigma)$. The joint localization and clock synchronization problem can be formulated as the following minimization problem.

$$\begin{aligned} \min_{\hat{\sigma}} \quad & V(\hat{\sigma}) = \sum_{k\sim(i,j)\in\mathcal{E}} f_k(\hat{\sigma})^2 \\ \text{s.t.} \quad & \hat{p}_i = p_i, \quad i \in \mathcal{V}_p, \\ & \hat{\alpha}_i = \alpha_i, \quad i \in \mathcal{V}_\alpha, \quad \hat{\beta}_i = \beta_i, \quad i \in \mathcal{V}_\beta \end{aligned} \tag{4}$$

where $\hat{\sigma}$ is the optimization variable and f_k is defined in (2). The sets $\mathcal{V}_p$, $\mathcal{V}_\alpha$, and $\mathcal{V}_\beta$ denote the node set of position/skew/offset anchors, respectively.

The subvectors of $\hat{\sigma}$, corresponding to the unknown non-anchor and known anchor states, are denoted by $\hat{\sigma}_r \in \mathbb{R}^{n_r}$ and $\hat{\sigma}_0 \in \mathbb{R}^{n_0}$, respectively. Thus, there exist two selection matrices $P_r \in \mathbb{R}^{n_r \times n(d+2)}$ and $P_0 \in \mathbb{R}^{n_0 \times n(d+2)}$ such that $\hat{\sigma}_r = P_r\hat{\sigma}$ and $\hat{\sigma}_0 = P_0\hat{\sigma}$. We also term P_0 the anchor selection matrix. Note that according to the structure of P_r and P_0, the matrix $P = [P_r^T, P_0^T]^T$ is a permutation matrix. Therefore, we have $P_rP_r^T = I_{n_r}$, $P_0P_0^T = I_{n_0}$, $P_rP_0^T = \mathbf{0}$ and $\hat{\sigma} = P_r^T\hat{\sigma}_r + P_0^T\hat{\sigma}_0$. By substituting $\hat{\sigma}_0 = P_0\sigma$ into $V(\hat{\sigma})$, the constrained optimization problem (4) can be converted to an unconstrained problem

$$\min_{\hat{\sigma}_r} \quad \widetilde{V}(\hat{\sigma}_r) = \sum_{k\sim(i,j)\in\mathcal{E}} \widetilde{f}_k(\hat{\sigma}_r)^2, \tag{5}$$

where $\widetilde{f}_k(\hat{\sigma}_r) = f_k(P_r^T\hat{\sigma}_r + P_0^TP_0\sigma)$. Since $\widetilde{f}_k(\sigma_r) = f_k(\sigma) = 0$ for all $k\sim(i,j) \in \mathcal{E}$, the true configuration $\sigma_r = P_r\sigma$ is a global minimizer of (5). The function $\widetilde{V}(\hat{\sigma}_r)$ is continuous and twice differentiable in the neighborhood of any generic σ_r. We define the **reduced joint rigidity matrix** $R_r(\hat{\sigma}) = R(\hat{\sigma})P_r^T$, which is the partial derivative of $F(\hat{\sigma})$ with respect to $\hat{\sigma}_r$. The following lemma indicates the condition under which σ_r is an isolated local minimizer.

Lemma 2: The true configuration σ_r is an isolated local minimizer of (5) if $R_r(\sigma)$ is full column rank.

Proof: The gradient and Hessian of $\widetilde{V}$ with respect to $\hat{\sigma}_r$ are

$$\nabla\widetilde{V} = 2\sum_{k=1}^{m} \widetilde{f}_k \cdot \nabla\widetilde{f}_k, \tag{6}$$

$$\nabla^2\widetilde{V} = 2\sum_{k=1}^{m} (\nabla\widetilde{f}_k(\nabla\widetilde{f}_k)^T + \widetilde{f}_k \cdot \nabla^2\widetilde{f}_k). \tag{7}$$

Evaluated at σ, $\widetilde{f}_k(\sigma_r) = 0$. Thus, $\nabla\widetilde{V} = 0$. Since $\nabla\widetilde{f}_k(\hat{\sigma}_r) = \nabla f_k(P_r^T\hat{\sigma}_r + P_0^TP_0\sigma)$, we have $\nabla^2\widetilde{V} = 2\sum_{k=1}^{m}\nabla\widetilde{f}_k(\nabla\widetilde{f}_k)^T = 2R_r(\sigma)^TR_r(\sigma)$ by the definition. If $R_r(\sigma)$ is full column rank, $\nabla^2\widetilde{V}(\sigma_r)$ is positive definite, therefore σ_r is an isolated local minimizer. ■

The following lemma builds the relationship between the full column rank of $R_r(\sigma)$ and the anchor selection matrix.

Lemma 3: Given a position-clock framework $(\mathcal{D}, \sigma)$ and an anchor selection matrix P_0, the reduced joint rigidity matrix $R_r(\sigma)$ is full column rank if P_0M is full column rank where M is a matrix whose column vectors form a basis of Null$(R(\sigma))$.

Proof: Suppose a vector $v_r \in \mathbb{R}^{n_r}$ satisfying $R_rv_r = \mathbf{0}$. By $R_r = RP_r^T$, we have $P_r^Tv_r \in$ Null(R), i.e., there exists a vector w such that $Mw = P_r^Tv_r$. By $P_0P_r^T = \mathbf{0}$, $P_0Mw = \mathbf{0}$. If P_0M is full column rank, then $w = 0$. Consequently, $v_r = P_rMw = \mathbf{0}$, so $R_r(\sigma)$ is full column rank. ■

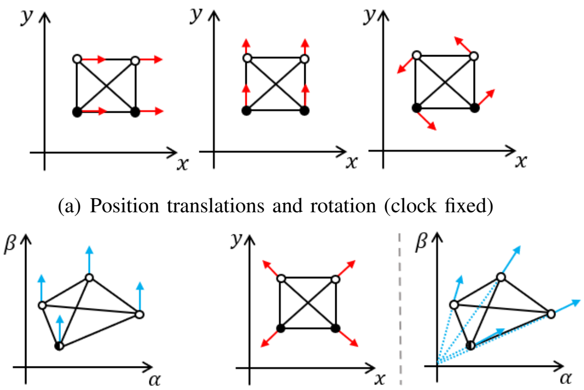
Lemma 3 indicates that the full column rank property of $R_r(\sigma)$, which is relevant to the minimizer property, can be determined by examining the anchor selection matrix and the null space of the joint rigidity matrix. It also indicates that the minimum number of anchors required for (5), i.e., the row number of P_0, is relevant to the minimum column rank of M, i.e., $\dim(\text{Null}(R(\sigma)))$. The minimum of $\dim(\text{Null}(R(\sigma)))$ occurs when the framework is infinitesimally joint rigid, as discussed in Section II. The following theorem states the anchor requirement under infinitesimal joint rigidity. The matrix J_d^i is defined in the notation subsection.

Theorem 4: Consider a position-clock framework $(\mathcal{D}, \sigma)$ in $\mathbb{R}^{d+2}$ consisting of K_p position anchors with position configuration $p_0 \in \mathbb{R}^{dK_p}$, K_α skew anchors with clock skew configuration $\alpha_0 \in \mathbb{R}^{K_\alpha}$ and K_β offset anchors with clock offset configuration $\beta_0 \in \mathbb{R}^{K_\beta}$. Define a matrix

$$M_0 = \begin{bmatrix} \mathbf{1}_{K_p} \otimes I_d & \mathbf{0} & p_0 & Q(p_0) \\ \mathbf{0} & \mathbf{0} & \alpha_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{K_\beta} & \beta_0 & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{K \times (\frac{d(d+1)}{2}+2)},$$

where $Q(p_0) = [(I_{K_p} \otimes J_d^1)p_0, (I_{K_p} \otimes J_d^2)p_0, ..., (I_n \otimes J_d^{d(d-1)/2})p_0] \in \mathbb{R}^{dK_p \times d(d-1)/2}$ and $K = dK_p + K_\alpha + K_\beta$. If $(\mathcal{D}, \sigma)$ is infinitesimally joint rigid and the matrix M_0 is full column rank, then the true configuration σ_r is an isolated local minimizer of (5).

Proof: By Theorem 1, if $(\mathcal{D}, \sigma)$ is infinitesimally joint rigid, the null space of $R(\sigma)$ is known. Thus, a matrix M, whose columns are formed by a basis of Null$(R(\sigma))$, holds $M = [[\mathbf{1}_n^T \otimes I_d, \mathbf{0}^T]^T, [\mathbf{0}^T, \mathbf{1}_n^T \otimes [0, 1]]^T, \sigma, [p^T(I_n \otimes J_d^1)^T, \mathbf{0}^T]^T, ..., [p^T(I_n \otimes J_d^{\frac{d(d-1)}{2}})^T, \mathbf{0}^T]^T]$. Therefore, $M_0 = P_0M$, where P_0 is the anchor selection matrix. By Lemma 2 and 3, if M_0 is full column rank, then $R_r(\sigma)$ is full column



(a) Position translations and rotation (clock fixed)

(b) Clock offset translation (position fixed)

(c) Position-clock scaling

Fig. 2. Basic trivial infinitesimal variations in $\mathbb{R}^{2+2}$ and how they are restricted by the chosen anchors. The framework is infinitesimally joint rigid where the undirected graph is a simplified presentation of Fig. 1(a). Position variations (red arrows) and the position anchors (black) are illustrated in the xy-coordinate plane. Clock variations (blue arrows) and the offset anchor (half-black) are illustrated in the $\alpha\beta$-coordinate plane.

rank, consequently the true configuration σ_r is an isolated local minimizer of (5). ■

Theorem 4 provides an approach to examine whether the chosen anchors provide enough information to eliminate all possible variations around the true configuration. An anchor selection candidate, satisfying the condition in Theorem 4, must have at least one offset anchor and d position anchors. However, a skew anchor is not necessary. For $d = 2$ as an example, two position anchors and one offset anchor guarantee that M_0 is full column rank provided position anchors are not collocated. For $d = 3$, three position anchors and one offset anchor guarantee that M_0 is full column rank provided position anchors are not collinear. Note that the position/skew/offset anchors can be the same or different nodes. Fig. 2 illustrates an example in $\mathbb{R}^{2+2}$, where the framework is infinitesimally joint rigid. The chosen anchors restrict all the trivial infinitesimal variations, indicating the position and clock parameters can be uniquely determined in the neighborhood of the true configuration.

Setting up anchors requires an accurate measurement of the corresponding information. Typically, position information is easier to obtain by manual measurements or existing positioning systems compared to clock parameters. As indicated by Theorem 4, the minimum requirement on known clock parameters is a single offset anchor provided position anchors are properly chosen. Since a global clock is assumed to be a common reference rather than an actual clock, we can simply set an arbitrary node's clock offset as zero to have a single offset anchor. It is equivalent to assuming the global clock starts at the same time as this node's clock.

IV. Distributed Gradient Descent Algorithm

In this section, we propose a distributed gradient-descent algorithm for solving (5). We also discuss the requirements for the initial configuration estimate and the method for determining an appropriate initial estimate.

Let $\hat{\sigma}_r^{(\tau)}$ be the configuration estimate computed by the algorithm at iteration τ for $\tau \geq 1$ and $\hat{\sigma}_r^{(0)}$ be the initial estimate. The minimization of (5) can be obtained by the gradient descent method. At each iteration, the configuration estimate is updated by

$$\begin{aligned}\hat{\sigma}_r^{(\tau+1)} &= \hat{\sigma}_r^{(\tau)} - \gamma_\tau \nabla \widetilde{V}(\hat{\sigma}_r^{(\tau)}) \\ &= \hat{\sigma}_r^{(\tau)} - \gamma_\tau R_r(\widetilde{\sigma}^{(\tau)})^T F(\widetilde{\sigma}^{(\tau)}),\end{aligned} \tag{8}$$

where $\widetilde{\sigma}^{(\tau)} = P_r^T \hat{\sigma}_r^{(\tau)} + P_0^T P_0 \sigma$ and the step size $\gamma_\tau > 0$. The update at the ith node is

$$\hat{\sigma}_{r,i}^{(\tau+1)} = \hat{\sigma}_{r,i}^{(\tau)} - \gamma_\tau \sum_{k \in \mathcal{E}_i} \widetilde{f}_k(\hat{\sigma}_r^{(\tau)}) \cdot \nabla \widetilde{f}_k(\hat{\sigma}_r^{(\tau)}), \tag{9}$$

where $\mathcal{E}_i$ contains all edges connected to node i. The update for each node only requires the information from its adjacent nodes, hence the algorithm is distributed.

Suppose that the framework is infinitesimally joint rigid and the given anchor selection follows the condition in Theorem 4, then $R_r(\sigma)$ is full column rank and σ_r is an isolated local minimizer. Therefore, there exist some $\epsilon > 0$ such that for any initial estimate $\hat{\sigma}_r^{(0)} \in \mathcal{B}_\epsilon[\sigma_r]$, i.e., $\|\hat{\sigma}_r^{(0)} - \sigma_r\| \leq \epsilon$, and appropriate step size [16], the gradient-descent algorithm is guaranteed to converge to σ_r.

Due to the local convergence, the initial estimate needs to be in the neighborhood of the true configuration, therefore, careful considerations need to be applied when initializing the estimate in practice. Let the true configuration $\sigma_r = [p_r^T, \varphi_r^T]^T$ and the estimate $\hat{\sigma}_r^{(\tau)} = [(\hat{p}_r^{(\tau)})^T, (\hat{\varphi}_r^{(\tau)})^T]^T$. In practice, the position $\hat{p}_r^{(0)}$ can be initialized by observation or a less accurate positioning system so that $\hat{p}_r^{(0)} \in \mathcal{B}_\delta[p_r]$, where $\delta > 0$ depends on the accuracy of the positioning system. However, it is usually harder to initialize the clock $\hat{\varphi}_r^{(0)}$ close to φ_r such that $\hat{\sigma}_r^{(0)} \in \mathcal{B}_\epsilon[\sigma_r]$. To determine a good initial estimate, we propose the following initialization procedure. Denote $\hat{\sigma}_r = [\hat{p}_r^T, \hat{\varphi}_r^T]^T$. By substituting $\hat{p}_r = \hat{p}_r^{(0)}$ into (5), $\hat{\varphi}_r^{(0)}$ is initialized as the minimizer of (5) with respect to only the clock configuration, i.e.,

$$\hat{\varphi}_r^{(0)} = \arg\min_{\hat{\varphi}_r} V_1(\hat{\varphi}_r) := \widetilde{V}(\begin{bmatrix} \hat{p}_r^{(0)} \\ \hat{\varphi}_r \end{bmatrix}). \tag{10}$$

The structure of (1) enables rewriting $\widetilde{V}$ as $\widetilde{V}(\hat{\sigma}_r) = \|Y_r \hat{\varphi}_r - b(\hat{p}_r)\|^2$, where Y_r is a constant matrix. The function $b(\hat{p}_r) = [b_1, ..., b_m]^T$ where $b_k = \|\hat{p}_i - \hat{p}_j\|$ for $k \sim (i,j) \in \mathcal{E}$ subject to $\hat{p}_i = p_i$ for all $i \in \mathcal{V}_p$. Note that Y_r is also the partial derivative of $F(\hat{\sigma})$ with respect to $\hat{\varphi}_r$. The reduced joint rigidity matrix $R_r(\hat{\sigma})$ can be written as

$$R_r(\hat{\sigma}) = \begin{bmatrix} R_{dr}(\hat{p}) & Y_r \end{bmatrix}, \tag{11}$$

where $R_{dr}(\hat{p}) = \partial F(\hat{\sigma}) / \partial \hat{p}_r$. When the framework is infinitesimally joint rigid and the given anchor selection follows the condition in Theorem 4, $R_r(\sigma)$ is full column rank, hence Y_r is full column rank. Thus, $V_1(\hat{\varphi}_r) = \|Y_r \hat{\varphi}_r - b(\hat{p}_r^{(0)})\|^2$ has a unique minimizer, i.e., $\hat{\varphi}_r^{(0)} = (Y_r^T Y_r)^{-1} Y_r^T b(\hat{p}_r^{(0)})$. Since the true configuration σ_r satisfies $\widetilde{V}(\sigma_r) = 0$, we have $Y_r \varphi_r = b(p_r)$. As the function b is locally L-Lipschitz, there exists $\delta' > 0$ such that for $\hat{p}_r^{(0)} \in \mathcal{B}_{\delta'}[p_r]$,

$$\|\hat{\varphi}_r^{(0)} - \varphi_r\| \leq a \|b(\hat{p}_r^{(0)}) - b(p_r)\| \leq aL \|\hat{p}_r^{(0)} - p_r\|, \tag{12}$$

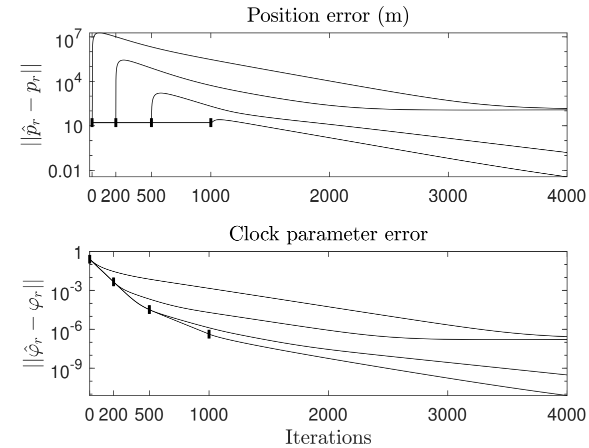


Fig. 3. The position and clock estimate error under the diagonally scaled algorithm (15) with different numbers of iterations in the initialization (10). Four curves from top to bottom correspond to 0, 200, 500, and 1000 iterations of initialization (10) followed by (15) after the indicated vertical bars. The network topology follows Fig. 1(a) and the anchors satisfy the condition in Theorem 4.

where $a = \|(Y_r^T Y_r)^{-1} Y_r^T\|$. The initialization in (10) ensures that the error of the initialized clocks is bounded by the error of the initial positions. When the initial position estimates satisfy $\hat{p}_r^{(0)} \in \mathcal{B}_\delta[p_r] \subseteq \mathcal{B}_{\delta'}[p_r]$, by (10), the initialized position-clock estimates satisfy $\|\hat{\sigma}_r^{(0)} - \sigma_r\| \leq \|\hat{p}_r^{(0)} - p_r\| + \|\hat{\varphi}_r^{(0)} - \varphi_r\| \leq (1 + aL)\delta$. Therefore, we can always find a δ so that $\|\hat{\sigma}_r^{(0)} - \sigma_r\| \leq \epsilon$, i.e., $\hat{\sigma}_r^{(0)} \in \mathcal{B}_\epsilon[\sigma_r]$. In other words, when $\hat{\sigma}_r^{(0)}$ is not in the convergence domain of the minimizer σ_r, it can be simply reinitialized by improving the accuracy of the initial position estimate.

Similar to (8), the minimization in (10) also can be realized by a distributed gradient descent algorithm, which is globally convergent since V_1 is quadratic. This approach addresses the difficulty in initializing clock configuration estimates, allowing the algorithm to be applied in a wider range of scenarios. A simulation example is given in Fig. 3.

V. Diagonal scaling and 2-block Gauss-Seidel method

Ill-conditioning is an unfavorable property in optimization problems and results in a slow convergence rate and high noise sensitivity. Unfortunately, the TOA-based joint localization and clock synchronization problem formulated in (5) is ill-conditioned due to the poor relative scaling between position and clock variables. In this section, we discuss two possible methods for better conditioning while maintaining the distributed property of the algorithm.

According to the TOA-based relationship between position and clock variables in (1), the clock variables are scaled by the speed of light when contributing to the objective function $\widetilde{V}(\hat{\sigma}_r)$ in (5), while position variables are not scaled. The disproportionate variable effects cause the ill-conditioning of (5). Considering the quadratic approximation of $\widetilde{V}(\hat{\sigma}_r)$ near σ_r, we next show the ill-conditioning by examining the condition number of the Hessian $\nabla^2 \widetilde{V}(\sigma_r)$. By (7), we have $\nabla^2 \widetilde{V}(\sigma_r) = R_r^T R_r$, where R_r is the reduced joint rigidity matrix evaluated at the true configuration σ. According to the partition of R_r given in (11),

$$R_r^T R_r = \begin{bmatrix} R_{dr}(p)^T R_{dr}(p) & R_{dr}(p)^T Y_r \\ Y_r^T R_{dr}(p) & Y_r^T Y_r \end{bmatrix}. \qquad (13)$$

Thus, we have the minimum eigenvalue $\lambda_{\min}(R_r^T R_r) \leq \lambda_{\min}(R_{dr}(p)^T R_{dr}(p))$ [17, Theorem 4.3.28] and the maximum eigenvalue $\lambda_{\max}(R_r^T R_r) = \|R_r\|^2$. The 2-norm $\|R_r\| \geq \max_{i,j} |r_{ij}|$, where r_{ij} is the entry in the ith row and jth column of R_r [18]. Since R_r is the partial derivative of F with respect to σ_r, according to the function expression in (2), we have $r_{ij} = c$ for some i, j. Consequently, $\lambda_{\max}(R_r^T R_r) \geq c^2$ and the condition number of the Hessian $\nabla^2 \widetilde{V}(\sigma_r)$ is

$$\kappa = \frac{\lambda_{\max}(R_r^T R_r)}{\lambda_{\min}(R_r^T R_r)} \geq \frac{c^2}{\lambda_{\min}(R_{dr}(p_r)^T R_{dr}(p_r))}. \qquad (14)$$

With generic position configurations, $\lambda_{\min}(R_{dr}(p_r)^T R_{dr}(p_r))$ is typically small. Due to the magnitude of the speed of light c, the condition number is overly large, indicating that the problem is ill-conditioned.

A scaled version of the gradient descent method is generally considered in ill-conditioned problems. The update is

$$\hat{\sigma}_r^{(\tau+1)} = \hat{\sigma}_r^{(\tau)} - \gamma_\tau D_\tau \nabla \widetilde{V}(\hat{\sigma}_r^{(\tau)}), \qquad (15)$$

where D_τ is positive definite and symmetric. This iteration is equivalent to the steepest descent applied in a new coordinate system of a vector $\hat{x}^{(\tau)} = (D_\tau)^{1/2} \hat{\sigma}_r^{(\tau)}$. The Hessian of $\widetilde{V}$ with respect to $\hat{x}^{(\tau)}$ is $(D_\tau)^{1/2} \nabla^2 \widetilde{V}(\hat{\sigma}_r^{(\tau)}) (D_\tau)^{1/2}$. By choosing an appropriate D_τ to achieve a smaller condition number, the problem (5) can be better conditioned. For distributed implementation and low computation complexity, we consider a constant diagonal matrix $D_\tau = \mathrm{diag}\{I_{n_p}, c^{-2} I_{n_\varphi}\}$, where n_p and n_φ correspond to the dimension of p_r and φ_r, respectively. This matrix rescales the clock parameters to eliminate the impact of the speed of light, which significantly improves the conditioning. A simulation example is given in Fig. 3, where the condition number of the Hessian matrix at the minimizer is decreased from 3.6×10^{19} to 535.8 by applying the diagonally scaled algorithm.

Another method for alleviating the ill-conditioning of (5) is to decompose the position and clock variables in the minimization of the objective function. It can be realized by the 2-block Gauss-Seidel method [19], also referred to as the block coordinate descent method, formulated as follows. Given a configuration estimate $\hat{\sigma}_r^{(\tau)} = [(\hat{p}_r^{(\tau)})^T, (\hat{\varphi}_r^{(\tau)})^T]^T$ at iteration τ, the estimate at the next iteration $\hat{\sigma}_r^{(\tau+1)} = [(\hat{p}_r^{(\tau+1)})^T, (\hat{\varphi}_r^{(\tau+1)})^T]^T$ where

$$\hat{\varphi}_r^{(\tau+1)} = \arg\min_{\hat{\varphi}_r} \widetilde{V}(\begin{bmatrix} \hat{p}_r^{(\tau)} \\ \hat{\varphi}_r \end{bmatrix}), \qquad (16)$$

$$\hat{p}_r^{(\tau+1)} = \arg\min_{\hat{p}_r} \widetilde{V}(\begin{bmatrix} \hat{p}_r \\ \hat{\varphi}_r^{(\tau+1)} \end{bmatrix}). \qquad (17)$$

As the function $\widetilde{V}$ is continuously differentiable in the neighborhood of true configuration and bounded below, subproblems (16) and (17) have optimal solutions following a similar analysis for (5), hence the 2-block Gauss-Seidel method is well-defined and converges to a local minimizer

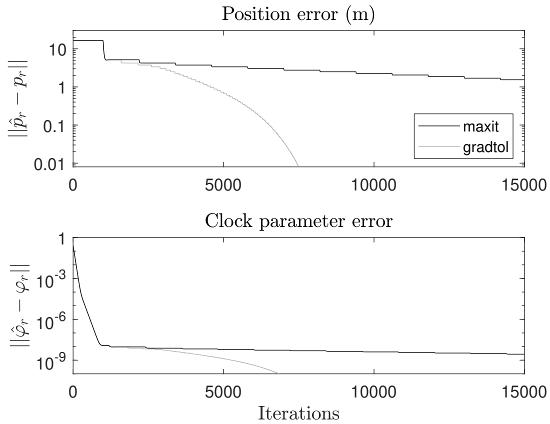


Fig. 4. Position and clock estimate error using the 2-block Gauss-Seidel method with different termination criteria. For the black curve (maxit), the convergence of each block is terminated after a predetermined number of iterations. For the gray curve (gradtol), the convergence of each block is terminated when the 2-norm of the node-based gradient component is less than a predetermined value for all nodes[1]. The example follows the same network setup as in Fig. 3.

of (5) [19, Corollary 2]. Subproblems (16) and (17), which involve clock and position variables, respectively, both have better conditioning compared to the original problem as the impact of the speed of light is eliminated. The conditioning properties of (16) and (17) are characterized by the relatively small condition number of $Y_r^T Y_r$ and $R_{dr}(\hat{p})^T R_{dr}(\hat{p})$, respectively. The minimization of subproblems can also be solved distributedly by a gradient descent algorithm. A simulation example is given in Fig. 4, where the condition number of the Hessian matrix of (16) and (17) at the minimizer are 137.8 and 12.8, respectively.

Better conditioning of the subproblems in the 2-block Gauss-Seidel method improves the speed of convergence and the robustness to numerical errors. It generally requires more iterations compared to the diagonally scaled algorithm but has lower computational complexity per iteration due to the decomposition of the optimization variables. Furthermore, by first updating the clock variable block, the subproblem (16) plays the same role as the initialization approach (10) in Section IV. Therefore, no additional initialization procedure for clock configuration is needed when applying the 2-block Gauss-Seidel method.

VI. Conclusions

In this paper, we studied the joint localization and clock synchronization problem in TOA-based sensor networks by formulating it as an optimization problem. Based on joint rigidity theory, we presented the topological condition and anchor requirements for uniquely determining the position and clock parameters. A distributed algorithm with local convergence was proposed along with an efficient method for initialization. We also studied the ill-conditioning of the problem and proposed two methods that alleviate the ill-conditioning and maintain the distributed manner.

[1]Detecting whether all nodes satisfy the termination condition requires additional protocol, e.g., termination message passing through a *spanning tree* [20].

The analysis in the paper is based on accurate TOA timestamp measurements. Characterizing and reducing the impact of noisy measurements on the estimation result is a direction of future work. For an extension where multiple timestamp measurements are sequentially acquired across the network, a stochastic gradient descent method can be considered. The construction of a joint rigid network relies on the bidirectional communication assumption, see [12], which restricts the communication structure in the network. Constructing a joint rigid network without bidirectional communication assumption is another future direction.

References

[1] G. Mao, B. Fidan, and B. D. Anderson, “Wireless sensor network localization techniques,” *Computer Networks*, vol. 51, no. 10, pp. 2529–2553, 2007.
[2] Y.-C. Wu, Q. Chaudhari, and E. Serpedin, “Clock synchronization of wireless sensor networks,” *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 124–138, 2010.
[3] J. Aspnes, T. Eren, D. K. Goldenberg, A. S. Morse, W. Whiteley, Y. R. Yang, B. D. O. Anderson, and P. N. Belhumeur, “A theory of network localization,” *IEEE Transactions on Mobile Computing*, vol. 5, no. 12, pp. 1663–1678, Dec 2006.
[4] S. Zhao and D. Zelazo, “Localizability and distributed protocols for bearing-based network localization in arbitrary dimensions,” *Automatica*, vol. 69, pp. 334–341, 2016.
[5] L. Schenato and F. Fiorentin, “Average timesynch: A consensus-based protocol for clock synchronization in wireless sensor networks,” *Automatica*, vol. 47, no. 9, pp. 1878–1886, 2011.
[6] B. Denis, J.-B. Pierrot, and C. Abou-Rjeily, “Joint distributed synchronization and positioning in UWB ad hoc networks using TOA,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 4, pp. 1896–1911, 2006.
[7] M. Hamer and R. D’Andrea, “Self-calibrating ultra-wideband network supporting multi-robot localization,” *IEEE Access*, vol. 6, pp. 22 292–22 304, 2018.
[8] J. Zheng and Y.-C. Wu, “Joint time synchronization and localization of an unknown node in wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1309–1320, 2009.
[9] S. Zhu and Z. Ding, “Joint synchronization and localization using TOAs: A linearization based WLS solution,” *IEEE Journal On Selected Areas in Communications*, vol. 28, no. 7, pp. 1017–1025, 2010.
[10] F. Meyer, B. Etzlinger, F. Hlawatsch, and A. Springer, “A distributed particle-based belief propagation algorithm for cooperative simultaneous localization and synchronization,” in *2013 Asilomar Conference on Signals, Systems and Computers*, 2013, pp. 527–531.
[11] W. Yuan, N. Wu, B. Etzlinger, H. Wang, and J. Kuang, “Cooperative joint localization and clock synchronization based on gaussian message passing in asynchronous wireless networks,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 9, pp. 7258–7273, 2016.
[12] R. Wen, E. Schoof, and A. Chapman, “Clock rigidity and joint position-clock estimation in ultrawideband sensor networks,” *IEEE Transactions on Control of Network Systems*, vol. 10, no. 3, pp. 1209–1221, 2023.
[13] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, “Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks,” *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, 2005.
[14] B. D. O. Anderson, C. Yu, B. Fidan, and J. M. Hendrickx, “Rigid graph control architectures for autonomous formations,” *IEEE Control Systems Magazine*, vol. 28, no. 6, pp. 48–63, Dec 2008.
[15] S. Zhao and D. Zelazo, “Bearing rigidity and almost global bearing-only formation stabilization,” *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1255–1268, May 2016.
[16] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 1999.
[17] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.
[18] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013.
[19] L. Grippo and M. Sciandrone, “On the convergence of the block nonlinear gauss–seidel method under convex constraints,” *Operations Research Letters*, vol. 26, no. 3, pp. 127–136, 2000.
[20] D. Bertsekas and J. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Athena Scientific, 2015.