Novel Parameterization for Gauss–Newton Methods in 3-D Pose Graph Optimization

Seyed-Mahdi Nasiri, *Member, IEEE*, Reshad Hosseini, and Hadi Moradi, *Senior Member, IEEE*

***Abstract*—Pose graph optimization (PGO), or equivalently pose synchronization, that is synchronizing rotations and positions, is the state-of-the-art formulation for simultaneous localization and mapping in robotics. In this article, we first present a new manifold Gauss–Newton method for solving the rotation synchronization problem. In this article, we derive an efficient implementation of the method and develop a convergence theory thereof. A structural parameter appears in the proof, which has a significant influence on the convergence basin. In this article, we show that this structural parameter is the norm of the inverse of the reduced graph Laplacian and obtains the explicit relation of this parameter for special graph structures. We also present another method that directly solves the pose synchronization using both relative rotation and translation observations. Experimental results show that our rotation synchronization method can be successfully used to initialize iterative PGO solvers. Furthermore, we show that our pose synchronization algorithm outperforms state-of-the-art solvers in high-noise cases.**

***Index Terms*—Gauss–Newton, initialization, pose graph optimization (PGO), 3-D simultaneous localization and mapping (SLAM).**

I. Introduction

Pose graph optimization (PGO) is the problem of estimating a set of unknown poses[1] so that relative poses best match a collection of pairwise relative measurements. It is the most applied formulation of *simultaneous localization and mapping* (SLAM) [1] in robotics and also known as SE(d)-synchronization or pose synchronization (PS) in many other fields. Similar formulations arise in sensor network localization [2] and structure from motion [3].

The PGO problem is a nonconvex optimization problem and hard to solve in general. Well-known iterative numerical methods were used to solve the PGO problem, such as Gauss–Newton [4]–[6], Levenberg–Marquardt [7], and trust region [8], [9]. The nonconvexity of the PGO problem and the convergence behavior of iterative methods were investigated in the literature [10]–[13], and it was shown that depending on the pose graph structure and the noise level of measurements, the problem may have different local minima that are far from the global minimum. Grisetti *et al.* [14] proposed a method that has larger basins of convergence. But Carlone *et al.* [15], [16] and Rosen *et al.* [17] showed that state-of-the-art algorithms are entangled in local minima and fail to converge to the global optimum in some cases.

Various attempts have been made to solve this problem in the literature. Convex relaxation was adopted by a line of research [18], [19]. The main idea is to find the global optimum of a semidefinite programming relaxation of the PGO (or its dual) problem. It has been shown experimentally that the convex relaxation, in many practical problems in which the amount of noise is limited, is exact and the duality gap is zero [19]. Similar ideas were used to verify the optimality of a given solution [20]–[22].

Aloise and Grisetti [23] proposed an alternative error function based on chordal distance introduced in [17] and [21] for the 3-D PGO problem to reduce the nonlinearity of the problem. They consequently obtained a larger basin of convergence and better numerical stability for their proposed algorithm.

Another set of efforts aimed at finding good initial points for iterative solvers for improving the convergence speed and reducing the risk of convergence to a local minimum. Many initialization methods took advantage of the fact that one part of the PGO cost function related to rotation measurements is independent of the positions, and can be solved separately [3], [15], [16], [24]. Then, the orientations are fixed and the positions are calculated from the other part of the cost function giving a simple least squares problem for the positions. This solution is then used as the initializer to bootstrap iterative methods.

In our earlier work [25], we proposed a method to find an initial estimation of the rotations. In this article, it is referred to as *rotation synchronization* (RS) algorithm. It was shown that this method outperforms the well-known chordal initialization method and can be successfully used to initialize iterative PGO solvers. The current work goes beyond our previous work in several important aspects. The major additions are as follows.

1) We give a convergence analysis of RS which was originally introduced in ICRA2018 [25]. We show that close enough to the global optimum solution, our algorithm gets closer to the optimum up to a certain bound.
2) We propose a more powerful method, PS, which takes into account both rotations and positions.

Manuscript received April 11, 2020; accepted September 28, 2020. This article was recommended for publication by Associate Editor N. Gans and Editor F. Chaumette upon evaluation of the reviewers' comments. *(Corresponding authors: Reshad Hosseini; Hadi Moradi.)*

Seyed-Mahdi Nasiri and Hadi Moradi are with the School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran (e-mail: s.m.nasiri@gmail.com; moradih@ut.ac.ir).

Reshad Hosseini is with the School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran, and is also with the School of Computer Science, Institute of Research in Fundamental Sciences (IPM), Tehran 19395-5746, Iran (e-mail: reshad.hosseini@ut.ac.ir).

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

Digital Object Identifier 10.1109/TRO.2020.3034021

[1]A pose comprises a position and an orientation.

3) We introduce a structural parameter that affects the basin of convergence. We give its analytical expression and its relation to the reduced Laplacian matrix for complete pose graphs.

The simulation results for evaluating the RS initialization method go beyond those of our conference paper [25]. These results reaffirm the convergence analysis and show that our RS algorithm has a wider convergence basin in comparison to the existing iterative solvers. We show in the experiments that by using our RS algorithm as an initializer and our PS algorithm as a solver, we obtain a method that outperforms the state-of-the-art SE-Sync method in high-noise datasets.

In the following, we will first provide an overview of the related literature in Section II. Then, we explain the basic mathematic notations and preliminaries in Section III. The method of our conference paper [25] is reviewed in Section IV. The convergence analysis is then given for noise-free (see Section V-A) and noisy measurements (Section V-B). In Section VI, the structural parameter is calculated for complete graphs[2] and the obtained convergence bound is compared to an existing bound in the literature. The second algorithm, PS, which uses both rotations and positions, is presented in Section VII. Finally, in Section VIII, through several experiments on a benchmark dataset and noisy versions of the same dataset, the performance of the proposed algorithms against the current state-of-the-art methods is investigated.

II. Related Works

In the efficient and robust implementations of Gauss–Newton and Levenberg–Marquardt methods, available in G^{2}o [26] and GTSAM [27], a minimal representation for increments in local Euclidean space was proposed and an update method was defined that accumulates the increments in a non-Euclidean manifold geometry. These methods can get stuck in the local minima of the cost function, and it has been shown that a proper initialization is important to obtain a good accuracy in these methods [28].

Carlone *et al.* [28] reviewed and compared several initialization methods [3], [29]–[32] and showed that using the chordal method as an initializer and GTSAM as a solver is the best combination for iteratively solving the problem. All of these initialization methods give a rough estimate for the rotation part of the PGO problem. Nasiri *et al.* [25] showed that RS, which solves the rotation part of the PGO problem, outperforms the chordal initialization method and can help iterative solvers to attain a more reliable solution.

In several recent works, convex relaxations of the PGO problem were proposed to overcome the nonconvexity of the problem and dependence of available iterative solvers on the initialization [17], [33], [34]. Some of these methods employ Lagrangian duality to obtain and verify the global optimality of the solution in planar [18], [21] and 3-D SLAM [19], [20], [22]. These methods can retrieval the global optimal solution of the original problem if the duality gap is zero. Rosen *et al.* [19] stated that there exists an upper bound for the noise level under which their method, SE-Sync, guarantees the optimality of the solution. The bound was not calculated theoretically, but the authors showed empirically that the bound is greater than typical noise levels encountered in robotic tasks. We show in the experiments that for high-noise levels, our proposed algorithm, PS, outperforms the SE-Sync method.

All of the abovementioned certifiable algorithms use a specific form of the cost function in which the rotation and translation measurements are considered to be independent and also have the same uncertainty in all directions. However, in a recent work [35], the planar PGO problem is formulated as a convex polynomial optimization program for a more general form of the measurement noise relaxing the assumption of the same uncertainty in all directions. The method of [35] can find the global optimal solution regardless of the noise level.

The rotation part of the PGO problem is known as rotation averaging [30] or attitude synchronization [36] in computer vision. Various algorithms were presented to solve this problem in recent years [37]–[39]. In a recent work, Eriksson *et al.* [40] utilized the spectral graph theory to present a theoretical analysis of the Lagrangian duality in the rotation averaging. They analytically specified an error bound for the noise levels that ensures there is no duality gap, and consequently their proposed algorithm converges to the optimal solution. Section VI gives a comparison between the maximum error bound needed for the convergence of our RS algorithm and the bound obtained in [40]. We see experimentally that the convergence behavior of other solvers also depend on the pose graph structure.

The role of graph structure on the convergence of various solvers and the number of local minima were investigated in the literature [10], [41], [42]. Khosoussi *et al.* [43] investigated the role of graph structure in the accuracy of maximum-likelihood (ML) estimation in SLAM. They theoretically analyzed the impact of average vertex degree[3] on the reliability of the ML estimation, the impact that was observed experimentally in [41]. They also explored the relations between the Cramer–Rao lower bound and some structural features of the pose graph for two special classes of 2-D SLAM, that is compass-SLAM and linear sensor network, and extended their analysis to planar SLAM in [44] and [45].

III. Notations and Preliminaries

A. PGO and Graph Topology

A set of poses in 3-D space is denoted by $n+1$ positions $\mathbf{p}_0, ..., \mathbf{p}_n \in \mathbb{R}^3$ and $n+1$ rotation matrices $R_0, ..., R_n \in \mathrm{SO}(3)$. The relative noisy displacement and rotation measurements from pose i to pose j are denoted by $\mathbf{d}_{ij}$ and Z_i^j, respectively. PGO is the problem of estimating poses, i.e., $\mathbf{p}_i$s and R_is, from m relative noisy observations between poses. We can consider poses and relative observations between poses as a directed graph $G = (V, E)$, where $V = \{v_0, v_1, ..., v_n\}$ is a set of $n+1$ vertices representing poses and $E = \{e_1, ..., e_m\}$ is a set of edges representing relative measurements between

[2] A graph in which every pair of distinct vertices is connected.

[3] The number of edges incident to the vertex.

poses. The relative position and rotation measurements between vertices i and j are represented by

$$Z_i^j = R_i^j R_{\text{noise}} = R_j^T R_i R_{\text{noise}} \tag{1}$$

$$\mathbf{d}_{ij} = R_i^T(\mathbf{p}_j - \mathbf{p}_i) + \mathbf{d}_{\text{noise}}. \tag{2}$$

The direction of the edge $e_k = (i, j), k \in \{1, \ldots, m\}$, indicates that the relative observation is from vertex i to vertex j.

The incidence matrix $\bar{A}$ of the aforementioned directed graph is an $m \times (n+1)$ matrix in which the kth row is related to the edge $e_k = (i, j)$. There are two nonzero entries in the kth row of $\bar{A}$, a "+1" in the jth column, and a "−1" in the ith column. Since all measurements in PGO are relative, without loss of generality, we can assume that v_0 is the origin, i.e., $\mathbf{p}_0 = [0, 0, 0]^T$ and $R_0 = I_3$. This operation is usually known as *anchoring* in the literature [20], [22]. Anchored reduced incidence matrix A is formed by eliminating the first column of $\bar{A}$ [19], [40], [46]. $A \in \mathbb{R}^{m \times n}$ is a full column rank matrix for connected graphs [47].

B. Operators

In this article, $\text{Sym}(X)$ extracts the symmetric part of matrix $X \in \mathbb{R}^{3\times3}$ and $\text{Skew}(X)$ extracts the skew-symmetric part of matrix X. Let $\mathbf{a}^T = (a_x, a_y, a_z)$ be a 3-vector. $[\mathbf{a}]_\times$ denotes the skew-symmetric matrix generated by $\mathbf{a}$

$$[\mathbf{a}]_\times = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix}. \tag{3}$$

Assume that vector $\mathbf{a}$ can represent the skew-symmetric part of matrix A, i.e., $\text{Skew}(A) = [\mathbf{a}]_\times$. The operator $\mathbf{s} : \mathbb{R}^{3\times3} \to \mathbb{R}^3$ receives matrix A and returns the 3-vector representing the skew-symmetric part of the matrix

$$\mathbf{s}(A) = \mathbf{a}. \tag{4}$$

C. Rodrigues' Formula

Rodrigues' rotation formula is a method for computing a rotation matrix from a rotation axis and a rotation angle. Assume that $\mathbf{a}$ is the normalized vector of the rotation axis and θ is the rotation angle. According to the Rodrigues' rotation formula, the rotation matrix is obtained by

$$R = I + J\sin\theta + J^2(1 - \cos\theta) \tag{5}$$

where I is a 3×3 identity matrix and $J = [\mathbf{a}]_\times$.

We define the rotation vector as

$$\boldsymbol{\delta} = \mathbf{a}\sin\theta \tag{6}$$

corresponding to the rotation matrix R, then the Rodrigues' formula for the rotation matrix can be re-expressed by

$$R = I + [\boldsymbol{\delta}]_\times + \beta(\|\boldsymbol{\delta}\|)[\boldsymbol{\delta}]_\times^2 \tag{7}$$

where

$$\beta(\|\delta\|) = \left(\frac{1 - \sqrt{1 - \|\boldsymbol{\delta}\|^2}}{\|\boldsymbol{\delta}\|^2}\right) = \left(\frac{1 - \cos\theta}{\sin^2\theta}\right). \tag{8}$$

Since $[\boldsymbol{\delta}]_\times$ is a skew-symmetric matrix, $[\boldsymbol{\delta}]_\times^2$ is symmetric and therefore

$$R^{-1} = R^T = I - [\boldsymbol{\delta}]_\times + \beta(\|\boldsymbol{\delta}\|)[\boldsymbol{\delta}]_\times^2. \tag{9}$$

D. Distance Metrics in SO(3)

Given two rotation matrices, R_1 and R_2, the following two metrics are used in this article.

1) Chordal distance: The Frobenius norm of the matrix $R_1 - R_2$, that is $d_{\text{chord}}(R_1, R_2) = \|R_1 - R_2\|_F$.
2) Angular distance: The rotation angle of $R_2^T R_1$, that is $d_{\text{ang}}(R_1, R_2) = \sin^{-1}\|\boldsymbol{\delta}\| = |\theta|$, wherein δ is the rotation vector ()orresponding to the rotation matrix $R_2^T R_1$.

E. PGO Formulation

Assume noisy measurements, $\mathbf{d}_{ij}$ (2), and Z_i^j (1), of a directed pose graph $G(V, E)$, are given, PGO is the problem of finding poses from these noisy measurements [5]. For some measurement noises commonly assumed by various works in the literature,[4] the ML estimation of the graph vertices, i.e., $\mathbf{p}_i, R_i, i \in \{1, ..., n\}$, becomes the solution of the following optimization problem [19], [22]:

$$\min_{\substack{R_i \in \text{SO}(3) \\ \mathbf{p}_i \in \mathbb{R}^3}} \sum_{e_k = \{i,j\} \in E} \lambda_k \left\|\mathbf{d}_{ij} - R_i^T(\mathbf{p}_j - \mathbf{p}_i)\right\|^2 + \omega_k \left\|Z_i^j - R_j^T R_i\right\|_F^2. \tag{10}$$

Let R_i^*, $i = \{1, ..., n\}$, be the optimal rotations given by the solution of (10). Then, the optimal positions are the solution of the translation part of the problem (10) given by [15]

$$\{\mathbf{p}_i^*\} = \underset{\mathbf{p}_i \in \mathbb{R}^3}{\text{argmin}} \sum_{e_k = \{i,j\} \in E} \lambda_k \left\|\mathbf{d}_{ij} - R_i^{*T}(\mathbf{p}_j - \mathbf{p}_i)\right\|^2. \tag{11}$$

The second part of (10) as given below, called the RS problem, is independent of the positions, and can be minimized, separately

$$\{R_i^\star\} = \underset{R_i \in \text{SO}(3)}{\text{argmin}} \sum_{e_k = \{i,j\} \in E} \omega_k \left\|Z_i^j - R_j^T R_i\right\|_F^2. \tag{12}$$

It is clear that the solutions of (12) are not exactly the optimal rotations of (10), but they can be used as initialization for gradient solvers as is the case for the famous chordal initialization method [3].

The problem (11) is a linear least squares problem and has the following closed form solution:

$$P^* = (A^T \Lambda A)^{-1} A^T \Lambda D \tag{13}$$

[4]The assumptions are the independence of the rotation and position measurements, zero-mean isotropic Gaussian distributed noises for the positions, and isotropic Langevin distributed noises with the identity mode for the rotations. These assumptions are not essential for developing our methods, but we make such assumptions in this article to simplify the formulas, to develop theoretical convergence results, and to make it possible to compare our method with a method, such as SE-Sync, which needs such assumptions. The development of our method, which can be used with other cost functions with a relaxed isotropic assumption, is briefly described in Appendix A.

where $P^*_{n\times 3} = [\mathbf{p}^*_1 \dots \mathbf{p}^*_n]^T$ is the matrix of positions, $A_{m\times n}$ is the reduced incidence matrix, $\Lambda_{m\times m} = \text{diag}([\lambda_1 \dots \lambda_m])$ is the diagonal matrix of weights, and $D_{m\times 3}$ is the matrix of all relative displacement measurements in the reference coordinate system, i.e., the kth row of D is $d^T_{ij}R^T_i$ where $e_k = (i,j) \in E$.

IV. RS Algorithm

In this section, we propose our RS algorithm for solving the RS problem (12). The proposed algorithm is a manifold Gauss–Newton method that starts from some initial rotations and iteratively refines the estimates of rotations. Suppose $\hat{R}_i(t)$, $i \in \{1, ..., n\}$, are estimates for the rotations at time step t. Assume the following change of variables for R_is in the RS problem (12):

$$R_i = \Psi_i(t)\hat{R}_i(t) \tag{14}$$

for $\Psi_i(t)$, $i \in \{1, ..., n\}$. To avoid clutter, we omit the time step t in the rest of this section. Using Rodrigues' formula, Ψ_i can be written as

$$\Psi_i = I + [\boldsymbol{\delta}_i]_\times + \beta_i(\|\boldsymbol{\delta}_i\|)[\boldsymbol{\delta}_i]^2_\times. \tag{15}$$

Substituting (14) and (15) in the RS problem (12), the arguments of the cost function are changed to $\boldsymbol{\delta}_i$, $i \in \{1, ..., n\}$. Thus, the problem can be rewritten as follows:

$$\begin{aligned}
&\underset{R_1,\dots,R_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| Z^j_i - R^T_j R_i \right\|^2_F \\
&\equiv \underset{\Psi_1\dots,\Psi_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| Z^j_i - \hat{R}^T_j \Psi^T_j \Psi_i \hat{R}_i \right\|^2_F \\
&\equiv \underset{\Psi_1,\dots,\Psi_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| \hat{R}_j Z^j_i \hat{R}^T_i - \Psi^T_j \Psi_i \right\|^2_F \\
&\equiv \underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \Big\| \hat{R}_j Z^j_i \hat{R}^T_i - I + [\boldsymbol{\delta}_j]_\times - [\boldsymbol{\delta}_i]_\times \\
&\quad + S_2(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) \Big\|^2_F
\end{aligned} \tag{16}$$

where $S_2(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j)$ contains all terms with multiplication of at least two skew-symmetric terms, $[\boldsymbol{\delta}_i]_\times$ or $[\boldsymbol{\delta}_j]_\times$, which makes it negligible compared to the other terms in (16) for small δ_i and δ_j

$$\begin{aligned}
S_2(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) = {} & [\boldsymbol{\delta}_j]_\times[\boldsymbol{\delta}_i]_\times + [\boldsymbol{\delta}_j]_\times\beta_i[\boldsymbol{\delta}_i]^2_\times - \beta_j[\boldsymbol{\delta}_j]^2_\times[\boldsymbol{\delta}_i]_\times \\
& - \beta_j[\boldsymbol{\delta}_j]^2_\times\beta_i[\boldsymbol{\delta}_i]^2_\times - \beta_j[\boldsymbol{\delta}_j]^2_\times - \beta_i[\boldsymbol{\delta}_i]^2_\times.
\end{aligned} \tag{17}$$

By removing $S_2(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j)$ from (16), the RS problem can be approximated by

$$\underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| [\boldsymbol{\delta}_j]_\times - [\boldsymbol{\delta}_i]_\times - B_k \right\|^2_F \tag{18}$$

where $B_k \triangleq I - \hat{R}_j Z^j_i \hat{R}^T_i$. We use the following trivial proposition to further simplify the cost function of (18).

Proposition 1: Let $A_{n\times n}$ be an arbitrary squared matrix, then

$$\|A\|^2_F = \|\text{Skew}(A)\|^2_F + \|\text{Sym}(A)\|^2_F.$$

(Proved in [25].)

According to Proposition 1, we have

$$\begin{aligned}
&\underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| [\boldsymbol{\delta}_j]_\times - [\boldsymbol{\delta}_i]_\times - B_k \right\|^2_F \\
&= \underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| \text{Skew}\left([\boldsymbol{\delta}_j]_\times - [\boldsymbol{\delta}_i]_\times - B_k\right) \right\|^2_F \\
&\qquad + \omega_k \left\| \text{Sym}\left([\boldsymbol{\delta}_j]_\times - [\boldsymbol{\delta}_i]_\times - B_k\right) \right\|^2_F \\
&= \underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| \text{Skew}\left([\boldsymbol{\delta}_j]_\times - [\boldsymbol{\delta}_i]_\times - B_k\right) \right\|^2_F \\
&\qquad + \omega_k \left\| \text{Sym}\left(-B_k\right) \right\|^2_F \\
&= \underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| \text{Skew}\left([\boldsymbol{\delta}_j]_\times - [\boldsymbol{\delta}_i]_\times - B_k\right) \right\|^2_F \\
&= \underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \left\| \boldsymbol{\delta}_j - \boldsymbol{\delta}_i - \mathbf{s}(B_k) \right\|^2.
\end{aligned}$$

Therefore, we get the following optimization problem for each time step t:

$$\hat{\Delta}(t) = \underset{\Delta}{\text{argmin}} \left\{ \text{tr}\left((A\Delta - B(t))^T \Omega (A\Delta - B(t)) \right) \right\} \tag{19}$$

where $A_{m\times n}$ is the reduced incidence matrix, Ω is a fixed matrix $\Omega = \text{diag}\Big(\begin{bmatrix} \omega_1 & \dots & \omega_m \end{bmatrix} \Big)$, and

$$B(t) = \begin{bmatrix} \mathbf{s}(B_1(t))^T \\ \vdots \\ \mathbf{s}(B_m(t))^T \end{bmatrix}, \hat{\Delta}(t) = \begin{bmatrix} \hat{\boldsymbol{\delta}}^T_1(t) \\ \vdots \\ \hat{\boldsymbol{\delta}}^T_n(t) \end{bmatrix}. \tag{20}$$

The problem (19) has the following closed-form solution:

$$\hat{\Delta}(t) = A^\ddagger B(t) \tag{21}$$

where

$$A^\ddagger = \left(A^T \Omega A\right)^{-1} A^T \Omega. \tag{22}$$

The ith row of $\hat{\Delta}(t)$ is a rotation vector $\hat{\delta}_i(t)$ and its corresponding rotation matrix is $\hat{\Psi}_i(t)$. According to (6), the 3-vector $\hat{\delta}_i(t)$ has the constraint $\|\hat{\delta}_i(t)\| \leq 1$, so one way to fulfill this condition is projecting $\hat{\delta}_i(t)$ into the unit sphere if it violates the constraint.

Equation (21) takes a Gauss–Newton step for the problem (16). To update the estimated rotations, $\hat{R}_i(t)$s, a mapping is needed to find the rotation matrices on the SO(3) manifold from $\hat{\delta}_i(t)$s, which are on the tangent space. To this aim, we propose to use (14) to implement the retraction[5] as follows:

$$\begin{aligned}
\hat{R}_i(t+1) &\leftarrow \hat{\Psi}_i(t)\hat{R}_i(t) \\
&= \left(I + [\hat{\delta}_i(t)]_\times + \beta_i(\|\hat{\delta}_i(t)\|)[\hat{\delta}_i(t)]^2_\times \right) \hat{R}_i(t).
\end{aligned} \tag{23}$$

This gives an iterative algorithm for finding the vertices rotations summarized in Algorithm 1. The algorithm needs initial

[5]Retraction is a mapping from the tangent space to the manifold space. See [48] for more information.

Algorithm 1: RS: Rotation Synchronization Algorithm.

Input: edges of pose graph $(Z_i^j, \mathbf{d}_{ij}, \forall e_k = \{i,j\} \in E)$, stopping criteria δ_{th} and $\text{itr}_{\max}$, and matrix of weights Ω from (20)
Output: estimated poses of vertices $(R_i, \mathbf{p}_i,\ i \in \{1,\dots,n\})$

$t \leftarrow 1$
$\hat{R}_i(t) \leftarrow$ chordal initialization
repeat
  $\mathbf{b}_k \leftarrow \mathbf{s}(I - \hat{R}_j(t) Z_i^j \hat{R}_i^T(t)), \forall e_k = \{i,j\} \in E$
  $B(t) \leftarrow [\mathbf{b}_1, \dots, \mathbf{b}_m]^T$
  $\hat{\Delta}(t) \leftarrow (A^T \Omega A)^{-1} A^T \Omega B(t)$
  for $i \in \{1,\dots,n\}$ **do**
    $\hat{\delta}_i(t)^T \leftarrow i$th row of $\hat{\Delta}(t)$
    if $\|\hat{\delta}_i(t)\| > 1$ **then** $\delta_i(t) \leftarrow \delta_i(t)/\|\delta_i(t)\|$
    $\hat{\Psi}_i(t) \leftarrow I + [\hat{\delta}_i(t)]_\times + \beta_i(\|\hat{\delta}_i(t)\|)[\hat{\delta}_i(t)]_\times^2$
    $\hat{R}_i(t+1) \leftarrow \hat{\Psi}_i(t) \hat{R}_i(t)$
  end for
  $t \leftarrow t+1$;
until $\max_i \|\hat{\delta}_i\| > \delta_{\text{th}}$ & $t < \text{itr}_{\max}$
$R_i \leftarrow \hat{R}_i(t),\ \forall i \in \{1,\dots,n\}$
return $R_i,\ \forall i \in \{1,\dots,n\}$

estimates of the rotations that is calculated by the chordal method of Martinec and Pajdla [3]. It is worth mentioning that with our proposed parameterization δ in (6), the problem (18) is more tighter approximation of (16) in comparison to the common method of using the canonical rotation angle vector, i.e., $\delta = \mathbf{a}\theta$. We observed experimentally that using the canonical rotation angle vector has slower convergence and is more entangled in local minima.

V. Convergence Analysis

To make the convergence analysis more understandable, we first analyze the noise-free cases and prove the optimality of the solution. Then, the analysis is presented for noisy cases. Note that, the convergence proof of the method in noise-free cases is not important by itself. But, it can clarify the proof for noisy cases and makes this article easy to follow.

A. Noise-Free Cases

Assume $R_i^\star, i \in \{1,\dots,n\}$, are the optimal solutions of RS (12). Matrices $\Psi_i^\star(t)$ align the estimated rotation matrices $\hat{R}_i(t)$ to the optimal solutions $R_i^\star$

$$R_i^\star = \Psi_i^\star(t) \hat{R}_i(t). \tag{24}$$

Suppose that $\delta_i^\star(t)$ is the rotation vector (6) corresponding to the matrix $\Psi_i^\star(t)$. In order to prove the convergence of the algorithm to the optimal solution, i.e., $\hat{R}_i(t) \to R_i^\star,\ i \in 1, ..., n$, it should be shown that in each step, $\Psi_i^\star(t)$ comes closer to the identity matrix or equally $\|\delta_i^\star(t)\| \to 0$. We define the following key parameter appearing in the convergence proof.

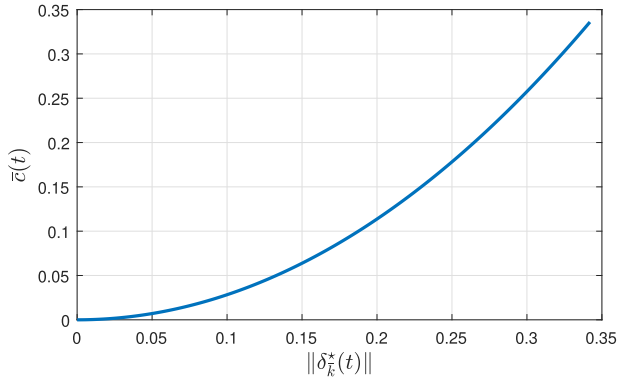

Fig. 1. Value of $\bar{c}(t)$ (27) with respect to the norm of maximum optimal rotation vector of the RS problem $\|\delta_{\bar{k}}^\star(t)\|$.

Definition 1: The structural parameter $\bar{a}$ is the maximum magnitude of $\mathbf{a}_k^\ddagger$

$$\bar{a} = \max_k \left\| \mathbf{a}_k^\ddagger \right\| \tag{25}$$

in which $\mathbf{a}_k^\ddagger$ represents the kth row of $A^\ddagger$ given in (22).

It is worth noting that the magnitude of $\mathbf{a}_k^\ddagger,\ k \in 1, ..., n$, depends on the pose graph topology and is usually small (less than 2 for torus, sphere-a, and cube in [28] datasets).

The following lemma gives a bound between the rotation vectors $\delta_i^\star(t)$ and rotation vectors that are the solutions of the synchronization problem (18). This bound is used in the convergence proof of Theorem 1.

Lemma 1: The magnitude of difference between the rotation vector $\delta_k^\star(t)$, the optimal solution of (16) for the kth rotation, and $\hat{\delta}_k(t)$, the kth row of $\hat{\Delta}(t)$ in RS, is bounded by $\bar{a}\bar{c}(t)$

$$\left\| \delta_k^\star(t) - \hat{\delta}_k(t) \right\| \le \bar{a}\bar{c}(t) \tag{26}$$

where

$$\bar{c}(t) = \left\| \mathbf{s}\left(-S_2(\delta_{\bar{k}}^\star(t), -\delta_{\bar{k}}^\star(t)) \right) \right\| \tag{27}$$

wherein $\bar{k} = \text{argmax}_k \|\delta_k^\star(t)\|$. The value of $\bar{c}(t)$ with respect to $\delta_{\bar{k}}^\star$ is depicted in Fig. 1.

Proof: According to the definition, $\delta_i^\star$s are the minimizers of the cost function (16)

$$\begin{aligned} &\{\delta_1^\star(t), \dots, \delta_n^\star(t)\} \\ &\quad = \underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \Big\| \hat{R}_j(t) Z_i^j \hat{R}_i^T(t) - I \\ &\qquad + [\delta_j]_\times - [\delta_i]_\times + S_2(\delta_i, \delta_j) \Big\|_F^2. \end{aligned} \tag{28}$$

Due to the noise-free assumption, $Z_i^j = R_i^j$, therefore, the minimum value of the cost function is zero. So, a new relation for $\delta_i^\star(t)$s can be given by the following equation:

$$\begin{aligned} &\{\delta_1^\star(t), \dots, \delta_n^\star(t)\} \\ &\quad = \underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\text{argmin}} \sum_{e_k=\{i,j\}\in E} \omega_k \Big\| \hat{R}_j(t) R_i^j \hat{R}_i^T(t) - I \\ &\qquad + [\delta_j]_\times - [\delta_i]_\times + \underline{S_2(\delta_i^\star(t), \delta_j^\star(t))} \Big\|_F^2. \end{aligned} \tag{29}$$

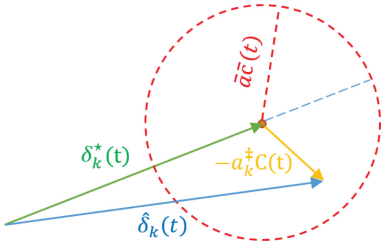

Fig. 2. Magnitude of difference between the kth optimal rotation vector $\delta_k^\star(t)$ and the solution of the approximated problem by removing higher order terms $\hat{\delta}_k(t)$ is smaller or equal to $\bar{a}\bar{c}(t)$, where $\bar{a}$ is the structural parameter and $\bar{c}(t)$ is an upper bound of $\|\mathbf{c}_k(t)\|$ described in (27).

Using the same strategy for deriving (19), the solution of (29) can be obtained as follows:

$$\Delta^\star(t) = \underset{\Delta}{\operatorname{argmin}} \left\{ \operatorname{tr}\left((A\Delta - (B(t) + C(t)))^T \Omega \times (A\Delta - (B(t) + C(t))) \right) \right\} \tag{30}$$

where

$$\Delta^\star(t)^T = [\delta_1^\star(t), \ldots, \delta_n^\star(t)] \tag{31}$$

$$C(t)^T = [\mathbf{c}_1(t), \ldots, \mathbf{c}_m(t)] \tag{32}$$

and for each $e_k = \{i, j\} \in E$

$$\mathbf{c}_k(t) = \mathbf{s}\left(-S_2\left(\delta_i^\star(t), \delta_j^\star(t)\right)\right). \tag{33}$$

Problems (19) and (30) are standard linear least squares problems. Hence, their solutions are as follows:

$$\hat{\Delta}(t) = A^\ddagger(B(t)) \tag{34}$$

$$\Delta^\star(t) = A^\ddagger(B(t) + C(t)). \tag{35}$$

Therefore, the difference between the solutions is

$$\Delta^\star(t) - \hat{\Delta}(t) = A^\ddagger C(t) \tag{36}$$

and, therefore

$$\delta_k^\star(t) - \hat{\delta}_k(t) = \mathbf{a}_k^\ddagger C(t), \quad k \in \{1, \ldots, n\} \tag{37}$$

where $\mathbf{a}_k^\ddagger$ represents the kth row of $A^\ddagger$.

It can be shown that the magnitude of $\mathbf{c}_k(t)$ is bounded to $\bar{c}(t)$. See Appendix B for more details. Therefore

$$\|\mathbf{c}_k(t)\| \leq \bar{c}(t) \quad \forall k \in \{1, \ldots, n\}. \tag{38}$$

According to (37) and (38) and the definition of $\bar{a}$ in (25), the following bound can be obtained:

$$\left\| \delta_k^\star(t) - \hat{\delta}_k(t) \right\| \leq \left\| \mathbf{a}_k^\ddagger \right\| \max_{k'} \|\mathbf{c}_{k'}(t)\| \leq \bar{a}\bar{c}(t). \tag{39}$$

Fig. 2 gives a geometrical view of the bound given in (39). ■

The following theorem proves a convergence basin of the RS algorithm in terms of the structural parameter.

Theorem 1: If the maximum angular distance between the initial and optimal rotations is less than a threshold for the given structural parameter $\bar{a}$ shown in Fig. 3, then RS converges to the optimal solution.

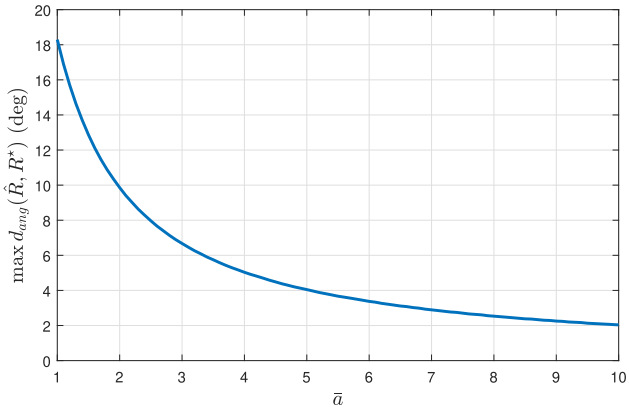

Fig. 3. Convergence basin, i.e., the maximum angular distance between the initial and optimal rotations, as a function of the structural parameter $\bar{a}$.

Proof: We first investigate the convergence behavior for $\bar{a} = 1$. Then, the effect of $\bar{a}$ on the obtained convergence area is investigated. By substituting (23) into (24), we obtain

$$R_k^\star = \Psi_k^\star(t)\hat{\Psi}_k^T(t)\hat{R}_k(t+1). \tag{40}$$

Thus, based on (24) and (40), it is easy to see

$$\Psi_k^\star(t+1) = \Psi_k^\star(t)\hat{\Psi}_k^T(t). \tag{41}$$

Given $\hat{\delta}_k(t)$ and $\delta_k^\star(t)$, the vector $\delta_k^\star(t+1)$ can be obtained by (41). The maximum magnitude of $\delta_k^\star(t+1)$ occurs when $\delta_k^\star(t)$ and $\hat{\delta}_k(t)$ are in the same direction and $\hat{\delta}_k(t)$ takes its maximum value in the bound of (39) (see Appendix C for detailed proof of this). By increasing $\delta_k^\star(t)$, the upper bound is increased, therefore $\delta_k^\star(t+1)$ is bounded by

$$\|\delta_{\bar{k}}^\star(t+1)\| \geq \|\delta_k^\star(t+1)\| \quad \forall k \in \{1, \ldots, n\} \tag{42}$$

where $\bar{k} = \operatorname{argmax}_k \|\delta_k^\star(t)\|$. Hence, the maximum bound of $\|\delta_k^\star(t+1)\|$ can be obtained as a function of the maximum value of $\|\delta_k^\star(t)\|$, which is shown in Fig. 4. The function is below the bisector of the first quadrant up to a certain value. It means that, if $\|\delta_k^\star(t)\|$ is less than a certain value, $\|\delta_k^\star(t+1)\|$ will always be smaller than $\|\delta_k^\star(t)\|$, and therefore the algorithm converges to the optimal solution.

Different values of $\bar{a}$ change the radius of the red circle in Fig. 2. The bigger $\bar{a}$, the larger radius for the geometric location of $\hat{\delta}_k(t)$, consequently results in a larger $\delta_k^\star(t+1)$. Finally, it results in a smaller convergence area. Fig. 3 shows the effect of $\bar{a}$ on the convergence area. In this figure, the vertical axis is the maximum initial estimation error that ensures the convergence. The estimation error is the angle of rotation in degrees. Each point on this curve corresponds to a case in which $\max \|\delta_k^\star(t)\| = \max \|\delta_k^\star(t+1)\|$ for different values of $\bar{a}$. Since these points are calculated through (15), (17), (27), (39), and (41), there is no clear closed-form solution. Thus, the analytical expression is excluded and the relation is shown as a curve in Fig. 3. ■

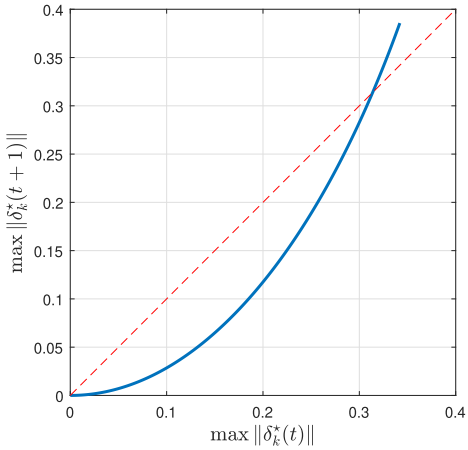

Fig. 4. Bound for the maximum magnitude of $\delta_k^\star(t+1)$ with respect to the maximum magnitude of $\delta_k^\star(t)$ assuming $\bar{a}=1$ (blue solid line), and the bisector of the first quadrant (red dashed line). Being on the curve below the bisector line means that the maximum bound of $\|\delta_k^\star(t+1)\|$ will be smaller than the maximum bound of $\|\delta_k^\star(t)\|$ and the algorithm will converge. Therefore, the convergence basin will be up to a point where $\max\|\delta_k^\star(t+1)\| = \max\|\delta_k^\star(t)\|$.

B. Noisy Cases

Similar to the noiseless case, we assume that $\Psi_k^\star(t)$ rotates the current rotation estimation $\hat{R}_k(t)$ to the real rotation matrix $R_k^\star$. Again, we first prove a lemma for bounding the difference between the rotation vectors $\delta_k^\star(t)$ and $\hat{\delta}_k(t)$. Because of noise, an extra term appears in the bound depending on the amount of noise in the measurements.

Lemma 2: The magnitude of the difference between the rotation vectors $\delta_k^\star(t)$ and $\hat{\delta}_k(t)$ is bounded by $\bar{a}(\bar{c}(t)+\bar{e})$

$$\left\|\delta_k^\star(t) - \hat{\delta}_k(t)\right\| \leq \bar{a}\,(\bar{c}(t)+\bar{e}) \tag{43}$$

where $\bar{c}$ is defined in (27) and

$$\bar{e} = \max_{\{i,j\}\in E} \left\|\mathbf{s}(\Phi_i^j)\right\| \tag{44}$$

wherein Φ_i^j is the difference between the real relative rotation R_i^j and the noisy measurement Z_i^j, i.e.,

$$\Phi_i^j \triangleq R_i^j - Z_i^j. \tag{45}$$

The proof of Lemma 2 is similar to the proof of Lemma 1 and is explained in Appendix D.

Equation (43) says that the geometric location of $\hat{\delta}_k(t)$ falls within the circle with the radius of $\bar{a}\bar{c}(t)+\bar{a}\bar{e}$ as illustrated in Fig. 5. Fig. 6 shows the magnitude of $\mathbf{e}_k$ with respect to the relative rotation measurement noise. In this figure, the measurement noise is represented by the angle of rotation in degrees. Now, we are ready to prove the main convergence theorem below. The proof obtains the convergence basin as a function of the key parameter $\bar{a}$ and the rotation measurement noise. This theorem also gives the least accuracy of RS in noisy cases.

Theorem 2: If the initial error (the maximum angular distance between the initial and optimal rotations) is less than a threshold for a given structural parameter $\bar{a}$ and given measurement noise

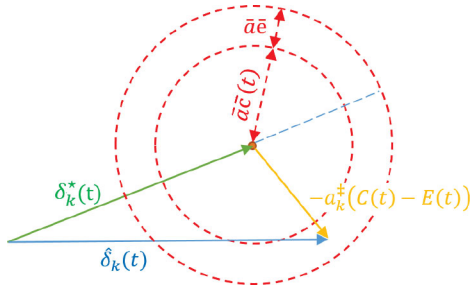

Fig. 5. Radius of geometric location of $\hat{\delta}_k(t)$ as in Fig. 2. Here, we have larger radius because of noise manifested as extra $\bar{a}\bar{e}$ term on the radius.

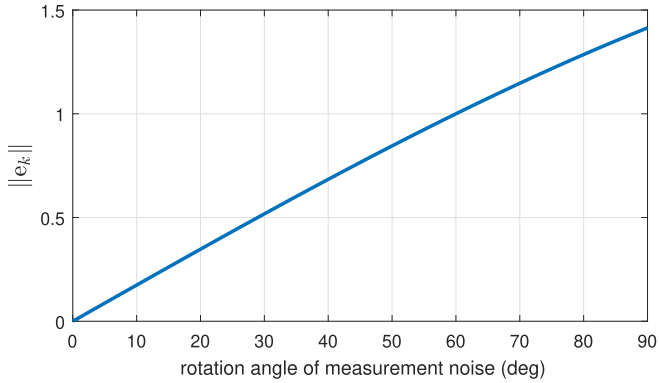

Fig. 6. Assuming rotation measurement noise of (1), the magnitude of e_k will only be a function of noise rotation angle.

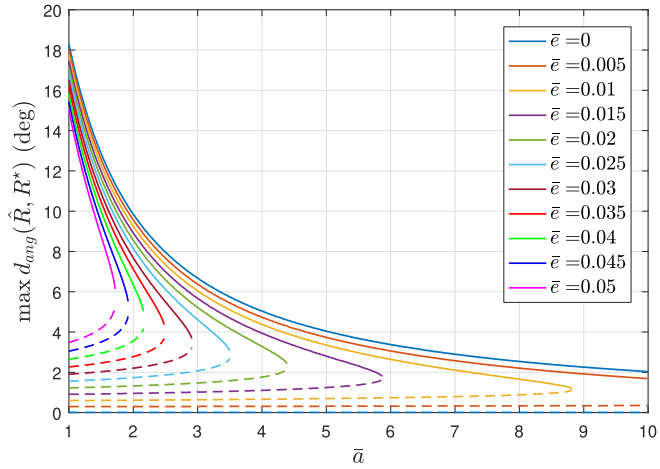

Fig. 7. Convergence basin (solid lines), i.e., the maximum initial rotation estimation error, and the least accuracy of RS (dashed lines), as a function of $\bar{e}$ (the maximum value of $\|\mathbf{e}_k\|$ shown in Fig. 6) and $\bar{a}$ (the structural parameter).

upper bound $\bar{e}$, then each iteration of RS reduces the errors up to the bound depicted in Fig. 7.

Proof: The magnitude of $\delta_k^\star(t+1)$ is obtained by (41). Like the noise-free condition, the largest value for the magnitude of $\delta_k^\star(t+1)$ is obtained when $k = \text{argmax}_l \|\delta_l^\star(t)\|$ and $\delta_k^\star(t)$ and $\hat{\delta}_k(t)$ are in the same direction. In noisy cases, the radius of geometric location of $\hat{\delta}_k(t)$ is larger by the extra value $\bar{a}\bar{e}$.

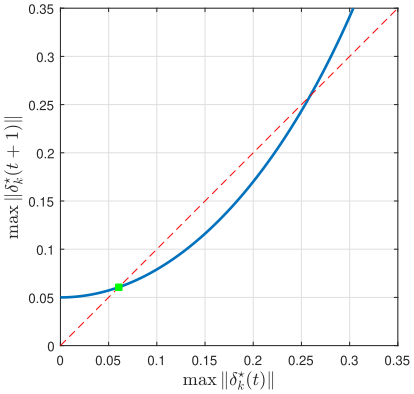

Fig. 8. Bound for the maximum magnitude of $\delta_k^\star(t+1)$ with respect to the maximum magnitude of $\delta_k^\star(t)$ (blue solid line), assuming $\bar{a}=1$ and $\bar{e}=0.05$, and the bisector of the first quadrant (red dashed line).

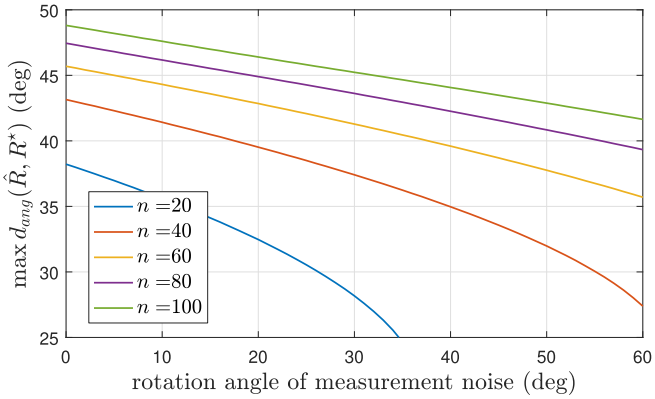

Fig. 9. Convergence basin as a function of the maximum rotation angle of the measurement noise, for a complete graph with n vertices.

Therefore, $\|\hat{\delta}_k(t)\|$ can take larger values, which results in larger $\|\delta_k^\star(t+1)\|$. Similar to noise-free condition, $\max_k \|\delta_k^\star(t+1)\|$ is bounded by a function of $\max_k \|\delta_k^\star(t)\|$ for a given $\bar{a}$ and $\bar{e}$. As an example, see Fig. 8 for a value of $\bar{e}$. Again, if $\max_k \|\delta_k^\star(t)\|$ is less than a certain value, the function will be under the bisector of the first quadrant and consequently $\max_k \|\delta_k^\star(t+1)\|$ will be less than $\max_k \|\delta_k^\star(t)\|$.

The effect of $\bar{e}$ and $\bar{a}$ on the maximum initial error and the final solution accuracy, i.e., the difference between the estimated values and true values, are shown in Fig. 7. ■

VI. Structural Parameter and Bounds

In this section, we first show that the structural parameter is actually the norm of the inverse reduced graph Laplacian. Inverse reduced graph Laplacian is an important entity for the PGO problem and was shown to be proportional to the Cramer–Rao bound for 2-D SLAM [43]. Then, we give an explicit relation for the convergence bound of complete graphs, the case where the structural parameter can be computed analytically.

A. Structural Parameter and Graph Laplacian

According to Definition 1, the structural parameter $\bar{a}$ can be equivalently defined as

$$\bar{a} = \sqrt{\max\left[\mathrm{diag}(A^\ddagger A^{\ddagger T})\right]}. \tag{46}$$

Assuming equal noise variance for all measurements and subsequently equal weights for all rotation measurements, i.e., $\Omega = \omega I$, we have

$$\begin{aligned} A^\ddagger A^{\ddagger T} &= (A^T \omega A)^{-1} A^T \omega^2 A (A^T \omega A)^{-T} \\ &= (A^T A)^{-1} A^T A (A^T A)^{-T} \\ &= (A^T A)^{-1} = \mathcal{L}^{-1} \end{aligned} \tag{47}$$

where $\mathcal{L}$ is the reduced graph Laplacian matrix [40], [46].

Remark 1: The [inverse] reduced Laplacian matrix of a connected graph is a positive definite matrix.

Proof: Any connected graph has got at least one spanning tree, and by Kirchhoff's theorem [46], any cofactor of its Laplacian matrix is positive. Therefore, $\det(\mathcal{L}) > 0$, and subsequently $\det(\mathcal{L}^{-1}) > 0$. ■

Proposition 2: The maximum value on the diagonal entries of a positive definite matrix is also its max norm. (Proved in Appendix E.)

According to the above-mentioned remarks, (46) and (47), the structural parameter is the max norm of the inverse reduced Laplacian matrix.

B. Structural Parameter of a Complete Graph

For a complete graph with n vertices, the reduced Laplacian matrix has the following simple structure:

$$\mathcal{L}_{n\times n} = n I_{n\times n} - 1_{n\times n}. \tag{48}$$

Its inverse is

$$\mathcal{L}^{-1} = \frac{1}{n+1}\left(I_{n\times n} + 1_{n\times n}\right). \tag{49}$$

Therefore, for a complete graph and for the aforementioned equal noise variance assumption, the structural parameter is obtained by

$$\bar{a} = \sqrt{\frac{2}{(n+1)}}. \tag{50}$$

It is easy to show that removing an edge from the graph increases the structural parameter. The mathematical developments and proofs are available in Appendix F.

Fig. 9 depicts the convergence basin obtained in Section V for a complete graph with different number of vertices. As it can be seen in the figure, for a complete graph with 40 vertices and initial estimation error, i.e., angular distance between initial estimations and real rotations, less than 35°, RS converges if the maximum rotation angle of measurement noise is less than 40°. The convergence basin is increased by increasing the number of vertices.

In a recent related work Eriksson *et al.* [40], obtained a bound for the maximum angular residual, i.e., angular distance between

optimal relative rotations R_i^j and observations Z_i^j, for which there is no duality gap. No duality gap implies that the solution of Lagrangian dual formulation is also an optimal solution of the primal problem. According to their analysis, for a complete graph with n vertices, the solution is optimal when the maximum angular residual is less than $\alpha_{\max}$, which is given as follows:

$$\alpha_{\max} = 2\arcsin\left(\sqrt{\frac{1}{4} - \frac{n}{2(n-1)}} - \frac{1}{2}\right). \tag{51}$$

This implies that the maximum angular residual for a complete graph with $n = 40$ is about 43.8°. The maximum angular residual is decreased by increasing the number of vertices. Note that the maximum angular residual can be larger than the measurement noise angle. For instance, in a complete graph with $n = 40$ vertices and 40° rotation angle of the measurement noise, the maximum angular residual can be more than 58°. We saw experimentally (see Appendix G) that the convergence basin of our method is not only much wider than what is suggested by the theory, but it is also wider than the convergence basin obtained by the Lagrange duality method [40] given in (51).

VII. PS Algorithm

The RS algorithm uses only relative rotation observations to estimate vertices rotations. In this part, we propose the PS algorithm that uses both relative rotation and position observations to estimate the rotations. Similar to RS, Rodrigues rotation formula is employed. Substituting (14) and (15) into the PGO problem (10) and removing the second-order terms, we obtain

$$\begin{aligned}
&\min_{\delta_i,\mathbf{p}_i} \sum_{e_k\in E} \lambda_{ij} \left\|\mathbf{p}_j - \mathbf{p}_i - \hat{R}_i\mathbf{d}_{ij} - [\boldsymbol{\delta_i}]_\times \hat{R}_i\mathbf{d}_{ij}\right\|^2 \\
&\qquad + \omega_{ij}\left\|\hat{R}_j Z_i^j \hat{R}_i^T - I + [\delta_j]_\times - [\delta_i]_\times\right\|_F^2 \\
&\equiv \min_{\delta_i,\mathbf{p}_i} \sum_{e_k\in E} \lambda_{ij} \left\|\mathbf{p}_j - \mathbf{p}_i - \hat{R}_i\mathbf{d}_{ij} + \left[\hat{R}_i\mathbf{d}_{ij}\right]_\times \boldsymbol{\delta_i}\right\|^2 \\
&\qquad + 2\omega_{ij}\left\|\delta_j - \delta_i - \mathbf{s}\left(I - \hat{R}_j Z_i^j \hat{R}_i^T\right)\right\|^2.
\end{aligned} \tag{52}$$

The optimization problem can be rewritten in the following matrix form:

$$\min_{\delta_i,\mathbf{p}_i} \left\| \begin{bmatrix} (\sqrt{2\Omega}A)\otimes I & O \\ M & (\sqrt{\Lambda}A)\otimes I \end{bmatrix} \begin{bmatrix} \delta_T \\ \mathbf{p}_T \end{bmatrix} - \begin{bmatrix} Z_\Omega \\ D_\lambda \end{bmatrix} \right\|^2 \tag{53}$$

in which $\otimes$ represents the Kronecker product and

$$\delta_T = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_n \end{bmatrix}, \ \mathbf{p}_T = \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_n \end{bmatrix}, \ Z_\Omega = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_m \end{bmatrix}, \ D_\lambda = \begin{bmatrix} \mathbf{l}_1 \\ \vdots \\ \mathbf{l}_m \end{bmatrix} \tag{54}$$

and for each $e_k = \{i, j\} \in E$

$$\mathbf{z}_k = \sqrt{2\omega_{ij}}\mathbf{s}\left(I - \hat{R}_j Z_i^j \hat{R}_i^T\right), \quad \mathbf{l}_k = \sqrt{\lambda_{ij}}\hat{R}_i\mathbf{d}_{ij}. \tag{55}$$

Algorithm 2: PS: Pose Synchronization Algorithm.

Input: edges of pose graph $(Z_i^j, \mathbf{d}_{ij},\ \forall e_k = \{i,j\} \in E)$, stopping criteria δ_{th}, cost_{th}, and $\text{itr}_{\max}$, and weights $\Omega = \text{diag}([\omega_1 \dots \omega_m])$ and $\Lambda_{m\times m} = \text{diag}([\lambda_1 \dots \lambda_m])$
Output: estimated poses of vertices $(R_i, \mathbf{p}_i,\ i \in \{1,\dots,n\})$

$t \leftarrow 1$
$\hat{R}_i(t) \leftarrow$ chordal initialization
repeat
  for $e_k = \{i,j\} \in E$ **do**
    $M_k \leftarrow [0_{3\times3}, \dots, 0_{3\times3}, \overbrace{[\hat{R}_i(t)\mathbf{d}_{ij}]_\times}^{i\text{th column-block}}, 0_{3\times3}, \dots, 0_{3\times3}]$
    $\mathbf{z}_k \leftarrow \sqrt{2\omega_{ij}}\mathbf{s}(I - \hat{R}_j Z_i^j \hat{R}_i^T)$
    $\mathbf{l}_k \leftarrow \sqrt{\lambda_{ij}}\hat{R}_i\mathbf{d}_{ij}$
  end for
  $M(t) \leftarrow [M_1^T, \dots, M_m^T]^T$
  $Z_\Omega \leftarrow [\mathbf{z}_1^T, \dots, \mathbf{z}_m^T]^T$
  $D_\lambda \leftarrow [\mathbf{l}_1^T, \dots, \mathbf{l}_m^T]^T$
  $G(t) \leftarrow \begin{bmatrix} (\sqrt{2\Omega}A)\otimes I & O \\ M & (\sqrt{\Lambda}A)\otimes I \end{bmatrix}$
  Solve for $\delta_T(t)$ and $p_T(t)$ in $G\begin{bmatrix}\delta_T(t)\\ \mathbf{p}_T(t)\end{bmatrix} = \begin{bmatrix} Z_\Omega \\ D_\lambda\end{bmatrix}$
  $f = \left\| G\begin{bmatrix}\delta_T(t)\\ \mathbf{p}_T(t)\end{bmatrix} - \begin{bmatrix} Z_\Omega \\ D_\lambda\end{bmatrix}\right\|^2$
  for $i \in \{1,\dots,n\}$ **do**
    $\hat{\delta}_i(t) \leftarrow (3i)$th to $(3i+3)$th row of $\delta_T(t)$
    if $\|\hat{\delta}_i(t)\| > 1$ **then** $\delta_i(t) \leftarrow \delta_i(t)/\|\delta_i(t)\|$
      $\hat{\Psi}_i(t) \leftarrow I + [\hat{\delta}_i(t)]_\times + \beta_i(\|\hat{\delta}_i(t)\|)[\hat{\delta}_i(t)]_\times^2$
      $\hat{R}_i(t+1) \leftarrow \hat{\Psi}_i(t)\hat{R}_i(t)$
    end for
    $t \leftarrow t+1$
until $\max_i \|\hat{\delta}_i\| > \delta_{\text{th}}$ & $t < \text{itr}_{\max}$ & $\Delta f/f < \text{cost}_{\text{th}}$
$R_i \leftarrow \hat{R}_i(t),\ \forall i \in \{1,\dots,n\}$
Calculate P from (13)
$\mathbf{p}_i^T \leftarrow i$th row of P, $\forall i \in \{1,\dots,n\}$
return $\mathbf{p}_i, R_i,\ \forall i \in \{1,\dots,n\}$

M is a matrix with m rows and n columns of 3×3 blocks in which, for each $e_k = \{i,j\} \in E$, the ith column block in the kth row block is equal to $\sqrt{\lambda_{ij}}[\hat{R}_i\mathbf{d}_{ij}]_\times$ and other terms are zero.

The solution of (53) is obtained by

$$\begin{bmatrix}\delta_T^*\\ \mathbf{p}_T^*\end{bmatrix} = G^\ddagger \begin{bmatrix} Z_\Omega \\ D_\lambda\end{bmatrix}. \tag{56}$$

The matrix G is a sparse matrix, therefore, sparse solvers can be used to solve (53). Equation (56) is computationally more expensive than (21) but leads to more accurate results.

The PS algorithm, which is summarized in Algorithm 2, uses just the expressed formulations iteratively to solve the PGO problem.

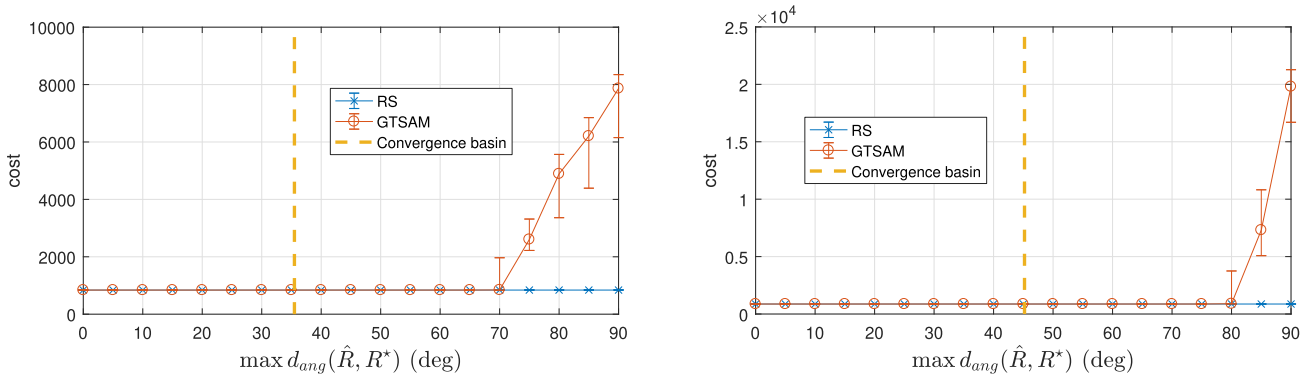

Fig. 10. Comparison of RS with an iterative RS solver of GTSAM for different amounts of noise in the initial guess for complete graphs with (a) 60 and (b) 100 number of vertices. The measurements are contaminated by random rotation noises with random axes and random rotation angles up to (a) 50° and (b) 30°. The figures show the median and positive and negative standard deviations of the final costs of RS and GTSAM (vertical axes) for 100 runs, starting from different initial guess with different "maximum angular error" (horizontal axes).

VIII. Evaluation

In this section, the performance of the proposed algorithms is evaluated through several experiments. The simulations are twofold. In the first part, the proposed RS algorithm is compared to state-of-the-art RS solvers. In the next part, the PS algorithm is compared to the state-of-the-art SE-Sync method for real and synthetic datasets.

In all experiments, $\text{itr}_{\max}$ and δ_{th} are 10^2 and 10^{-7}, respectively, for both RS and PS, and cost_{th} is set to 10^{-7}. All experiments are tested using MATLAB 9.1 on a computer with an Intel Core i7-7700HQ at 2.8 GHz, with 12 GB of memory.[6]

A. RS Problem

In this part, an experimental study is presented to support our claim that the proposed RS method for RS has a large convergence basin and is better than the Gauss–Newton method of GTSAM. The algorithms are applied to a set of complete graphs. RS convergence basin for complete graphs is obtained in Section VI (see Fig. 9). Fig. 10 shows the costs obtained by RS and GTSAM in complete graphs with (a) 60 and (b) 100 number of vertices for different amounts of noise in initial guesses. As it can be seen in Fig. 10, the RS convergence basin is larger than the GTSAM convergence basin. It is also larger than what is guaranteed by our theoretical analysis. More experiments are presented in Appendix H, which compare RS and GTSAM on different datasets and using Langevin distribution rotation noise.

B. PGO Problem

The evaluations of this part are performed on the common benchmark datasets [28] and noisy versions of them. To generate more realistic noises with relatively similar properties of the noise in the original data, we try to keep original shapes of noise histograms. This is simply done by multiplying the noise of the original data by the factor $\eta > 1$. To be more concrete, let us assume that $R_i, \mathbf{p}_i, i \in \{1, \dots, n\}$, are the optimal solutions and $\mathbf{d}_k, Z_k$ for all $e_k = \{i, j\}, k \in \{1, \dots, m\}$, are the original observations of a dataset. Using Rodrigues' formula for the noise term in (1), we have

$$Z_k = R_j^T R_i \left(I + J_k \sin\theta_k + J_k^2 (1 - \cos\theta_k)\right). \tag{57}$$

The noisy observations are generated as follows:

$$\mathbf{d}'_k = R_i^T(\mathbf{p}_j - \mathbf{p}_i) + \eta\left(\mathbf{d}_k - R_i^T(\mathbf{p}_j - \mathbf{p}_i)\right) \tag{58}$$

$$Z'_k = R_j^T R_i \left(I + J_k \sin\eta\theta_k + J_k^2 (1 - \cos\eta\theta_k)\right). \tag{59}$$

TABLE I
Mean and Standard Deviation of Observation Noise in Different Datasets

Dataset	sphere-a	torus	cube	garage	cubicle	rim
mean($\|d_{noise}\|$)	0.906	0.105	0.125	0.011	0.015	0.035
std($\|d_{noise}\|$)	0.495	0.045	0.053	0.009	0.023	0.086
mean(θ_k) (deg)	15.51	3.087	3.692	0.084	0.365	1.084
std(θ_k) (deg)	6.649	1.316	1.567	0.064	0.848	3.566

Table I gives the mean and standard deviation of $\|d_{\text{noise}}\|$ and θ_k for all original datasets. This way of generating noise makes the experiments and the results reproducible by potential readers. It is worth mentioning that we tried other (nonmultiplicative) types of noise and observed the same performance gains.

1) Comparing With SE-Sync on Original Datasets: Table II compares the results of RS+PS[7] and the state-of-the-art method SE-Sync [19] on the common datasets [28]. In this table, the number of vertices (n), the number of edges (m), the odometry cost (f_{odometry}), the optimal cost (f^*), the number of RS and PS iterations (itrRS + itrPS), the number of outer and inner SE-Sync iterations (outer(inner)), and the execution time of the algorithms (time) and the chordal initialization are presented. As shown in the table, in most experiments, RS+PS is faster than SE-Sync. It is worth noting that for sphere-a, garage, cubicle, and rim, the cost attained by RS+PS is slightly better than that by SE-Sync (due to small tolerance chosen for RS+PS stopping

[6] Efficient implementation of the proposed algorithms in MATLAB is available via http://visionlab.ut.ac.ir/resources/pgorls.zip.

[7] In this article, the phrase method1 + method2 means using the method1 as initializer and the method2 as the solver. The phrase "chord" means the chordal initialization.

TABLE II
COST, NUMBER OF ITERATIONS, AND CPU TIME OF RS+PS AND SE-SYNC METHODS ON DIFFERENT LOW-NOISE DATA

Dataset	n	m	$f_{odometry}$	f^*	chord	RS+PS		SE-Sync	
					$time[s]$	$itrRS{+}itrPS$	$time[s]$	$outer(inner)$	$time[s]$
sphere_a	2200	8647	2.269×10^8	2.962×10^6	0.14	10+6	0.83	10(29)	1.66(3.2)
torus	5000	9047	1.345×10^7	2.423×10^4	0.24	8+27	3.99	4(35)	2.26(5.15)
cube	8000	22236	1.705×10^8	8.432×10^4	0.92	7+5	9.17	4(36)	8.26(15.43)
garage	1661	6275	1.930×10^5	1.263×10^0	0.05	4+3	0.18	4(383)	3.84(4.03)
cubicle	5750	16869	5.041×10^6	7.171×10^2	0.26	6+15	3.05	4(82)	4.35(5.78)
rim	10195	29743	8.075×10^7	5.461×10^3	0.50	20+23	10.1	5(265)	25.88(26.16)

The CPU time of SE-Sync is computed without (with) the time spent for optimality check.

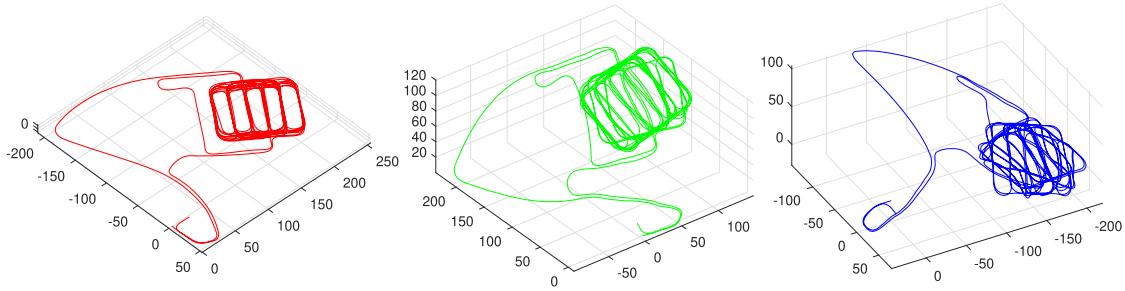

Fig. 11. Optimal solution of the garage original dataset (left), the result of RS+PS (center), and the result of SE-Sync initialized by the chordal method (right) in garage with noise factor $\eta = 80$.

TABLE III
RESULTS OF DIFFERENT METHODS IN DIFFERENT DATASETS WITH DIFFERENT NOISE FACTORS

	η	40	60	80
garage	odometry	1.948×10^5	1.972×10^5	2.007×10^5
	chord	2.282×10^3	5.152×10^3	9.189×10^3
	RS	2.268×10^3	5.108×10^3	9.082×10^3
	RS+PS	2.020×10^3	4.542×10^3	8.070×10^3
	chord+SE-Sync	2.020×10^3	4.542×10^3	8.982×10^3
	η	5	7	10
cubicle	odometry	5.084×10^6	5.093×10^6	5.108×10^6
	chord	1.879×10^4	3.643×10^4	7.545×10^4
	RS	1.828×10^4	3.521×10^4	6.979×10^4
	RS+PS	1.618×10^4	3.042×10^4	5.679×10^4
	chord+SE-Sync	1.618×10^4	3.212×10^4	6.162×10^4
	η	5	7	10
rim	odometry	8.122×10^7	8.122×10^7	8.122×10^8
	chord	1.729×10^5	3.072×10^5	5.752×10^5
	RS	1.659×10^5	2.968×10^5	5.064×10^5
	RS+PS	8.525×10^4	1.793×10^5	2.314×10^5
	chord+SE-Sync	1.027×10^5	1.927×10^5	3.615×10^5
	η	5	7	10
sphere_a	odometry	2.458×10^8	2.558×10^8	2.622×10^8
	chord	6.696×10^6	6.787×10^6	6.843×10^6
	RS	4.661×10^6	4.814×10^6	4.875×10^6
	RS+PS	4.658×10^6	4.813×10^6	4.873×10^6
	chord+SE-Sync	4.650×10^6	4.821×10^6	4.870×10^6
	η	5	7	10
torus	odometry	1.370×10^7	1.405×10^7	1.469×10^7
	chord	6.894×10^5	1.480×10^6	3.470×10^6
	RS	6.101×10^5	1.211×10^6	2.497×10^6
	RS+PS	6.015×10^5	1.187×10^6	2.430×10^6
	chord+SE-Sync	6.015×10^5	1.248×10^6	2.785×10^6
	η	5	10	15
cube	odometry	1.713×10^8	1.744×10^8	1.797×10^8
	chord	2.276×10^6	1.060×10^7	2.300×10^7
	RS	2.152×10^6	8.404×10^6	1.832×10^7
	RS+PS	2.088×10^6	8.100×10^6	1.666×10^6
	chord+SE-Sync	2.088×10^6	8.100×10^6	1.941×10^6

criterion). The difference between the costs for torus and cube is smaller than 10^{-3}.

2) Comparing With SE-Sync in Challenging Datasets: To compare the methods in more challenging datasets, we added noise to relative rotation and translation observations in all benchmark datasets. Table III compares the results of different methods for different noise factors. In almost all datasets, RS+PS outperforms chord+SE-Sync when the noise factors increases. Fig. 11 illustrates the results of RS+PS and chord+SE-Sync in garage with the noise factor $\eta = 80$. As it can be seen in Table I, the garage original dataset has a very low observation noise and the noise factor $\eta = 80$ gives an amount of noise almost equal to the sphere-a original observation noise. More experiments can be found in Appendixes I and J.

IX. CONCLUSION

In this article, new manifold Gauss–Newton methods, RS and PS, were introduced to solve the RS and PGO problems, respectively. It was shown in the experiments that RS outperformed the chordal initialization method and offered better starting point for iterative techniques resulting in escaping from local minima. We also compared our RS as an initializer and PS as a solver to the state-of-the-art method that used a chordal initialization and SE-Sync as a solver. The experiments showed that our algorithm was faster than SE-Sync in low-noise datasets and resulted in better solution in high-noise cases. A convergence analysis was developed for RS and a bound was given in which each iteration of our algorithm get closer to the global minimum. It was observed in the experiments that the convergence area was much larger than the area proved in Section V. A future direction of research is proving a better basin of convergence.

It was displayed that the basin of convergence depended on the structural parameter $\bar{a}$. We showed its relation to the reduced Laplacian matrix and derived its analytical expression for complete graphs. Khosoussi *et al.* [43] computed a closed-from expression for Fisher information matrix (FIM) in 2-D SLAM. Following their results, it is easy to see that FIM for a 2-D RS

problem is equal to $\sigma_\theta^2 L$, where σ_θ^2 is an angular measurement noise variance. Hence, the max norm of the Cramer–Rao lower bound[8] is equal to

$$\|\mathcal{I}^{-1}\|_{\max} = \sigma_\theta^2 \bar{a}^2. \tag{60}$$

Equation (60) reveals the effect of the measurement noise and the graph structure on the max norm of the covariance matrix lower bound in 2-D SLAM. Our next step is to find such relations for 3-D pose graphs.

In the experiments of Appendix I, our proposed method attained lower costs in almost all datasets in comparison to SE-Sync, and it also attained better or equal lower localization error measured as ATE in the most of datasets. But, interestingly, we observed that on few highly noisy datasets, SE-Sync attain lower ATE in comparison with our proposed PS method. It is remained for a future work to investigate the reason of this behavior. We think that it is possible to combine the power of SE-Sync and our proposed method, to reach a method with a better localization accuracy than any of these methods.

Appendix A
Proposed Methods for the General Case of the Cost Function

The general form of the 3-D PGO problem that covers colored and nonisotropic measurements noise has the form

$$\min \sum \mathbf{e}_{ij}^T \mathcal{I}_{ij} \mathbf{e}_{ij} \tag{61}$$

where $\mathcal{I}_{ij}$s $\in \mathbb{R}^{6\times 6}$ are information matrices and $\mathbf{e}_{ij}$s $\in \mathbb{R}^6$ are the error vectors that measure how the graph parameters, i.e., position and rotation of the graph vertices, match the measurements. Each $\mathbf{e}_{ij}$ contains a 3-vector related to the position error $\mathbf{e}_{ij}^t$ and a 3-vector related to the rotation error $\mathbf{e}_{ij}^r$.

We define the error vectors as follows:

$$\mathbf{e}_{ij}^t = (\mathbf{p}_j - \mathbf{p}_i) - R_i \mathbf{d}_{ij}$$
$$\mathbf{e}_{ij}^r = \mathbf{s}(R_j Z_i^j R_i^T).$$

$R_j Z_i^j R_i^T$ is a rotation matrix and can be expressed in Rodregues format (7)

$$R_j Z_i^j R_i^T = I + [\delta_{ij}]_\times + \beta(\|\delta_{ij}\|)[\delta_{ij}]_\times^2.$$

It is easy to show that $\mathbf{s}(R_j Z_i^j R_i^T)$ is equal to δ_{ij}.

In order to obtain the least squares of the problem at each step in the Gauss–Newton method, similar to what it can be seen in this article, we substitute (14) and (15) in the error vectors and remove the second-order terms. Thus, we have

$$\begin{aligned}\mathbf{e}_{ij}^t &\simeq \mathbf{p}_j - \mathbf{p}_i - \hat{R}_i \mathbf{d}_{ij} - [\delta_i]_\times \hat{R}_i \mathbf{d}_{ij} \\ &= \mathbf{p}_j - \mathbf{p}_i - \hat{R}_i \mathbf{d}_{ij} + \left[\hat{R}_i \mathbf{d}_{ij}\right]_\times \delta_i \\ \mathbf{e}_{ij}^r &\simeq \delta_j - \delta_i + \hat{\delta}_{ij} = \delta_j - \delta_i - \mathbf{s}\left(I - \hat{R}_j Z_i^j \hat{R}_i^T\right) \\ &= \delta_j - \delta_i - \mathbf{s}\left(I - \hat{R}_j Z_i^j \hat{R}_i^T\right).\end{aligned}$$

[8]Well-known Cramer–Rao bound states that a covariance matrix of any unbiased estimator of rotations satisfies, i.e., $\mathbb{C}\text{ov}_\theta \succeq \mathcal{I}^{-1}$, where $\mathcal{I}$ is FIM.

These error vectors are the same error vectors used in (52). The error vectors can be reexpressed in a matrix form as follows:

$$\mathbf{e}_{ij} = \begin{bmatrix}\mathbf{e}_{ij}^t \\ \mathbf{e}_{ij}^r\end{bmatrix} \simeq \begin{bmatrix}-I & I & \left[\hat{R}_i \mathbf{d}_{ij}\right]_\times & O \\ O & O & -I & I\end{bmatrix} \begin{bmatrix}\mathbf{p}_i \\ \mathbf{p}_j \\ \delta_i \\ \delta_j\end{bmatrix} - \begin{bmatrix}\hat{R}_i \mathbf{d}_{ij} \\ -\hat{\delta}_{ij}\end{bmatrix}. \tag{62}$$

Substituting (62) in (61), we obtain

$$\min_{\delta_i, \mathbf{p}_i} \sum_{e_k \in E} \left\| \mathcal{I}_{ij}^{1/2} \begin{bmatrix}-I & I & \left[\hat{R}_i \mathbf{d}_{ij}\right]_\times & O \\ O & O & -I & I\end{bmatrix} \begin{bmatrix}\mathbf{p}_i \\ \mathbf{p}_j \\ \delta_i \\ \delta_j\end{bmatrix} - \mathcal{I}_{ij}^{1/2} \begin{bmatrix}\hat{R}_i \mathbf{d}_{ij} \\ -\hat{\delta}_{ij}\end{bmatrix} \right\|^2.$$

Solving the above-mentioned least squares problem, we obtain the Gauss–Newton step. Now that we have the Gauss–Newton step, we can update the parameters using the proposed retraction (lines 15–20 of Algorithm 2). It is evident that (52) and (53) are special cases of (61) for $\mathcal{I}_{ij} = \text{diag}([\lambda_{ij}, \lambda_{ij}, \lambda_{ij}, 2\omega_{ij}, 2\omega_{ij}, 2\omega_{ij}])$.

Appendix B
Maximum Magnitude of $\mathbf{c}_k(t)$

In this part, we obtain the maximum magnitude of $\mathbf{c}_k(t)$ and prove the inequality (38). The magnitude of $\mathbf{c}_k(t)$ is obtained as

$$\|\mathbf{c}_k(t)\| = \|\mathbf{s}\left(-S_2\left(\delta_i^\star(t), \delta_j^\star(t)\right)\right)\|$$

where

$$\begin{aligned}S_2(\delta_i^\star, \delta_j^\star) &= [\delta_j^\star]_\times [\delta_i^\star]_\times + [\delta_j^\star]_\times \beta_i [\delta_i^\star]_\times^2 \\ &\quad - \beta_j [\delta_j^\star]_\times^2 [\delta_i^\star]_\times - \beta_j [\delta_j^\star]_\times^2 \beta_i [\delta_i^\star]_\times^2 - \beta_j [\delta_j^\star]_\times^2 - \beta_i [\delta_i^\star]_\times^2.\end{aligned}$$

The following propositions reveal the fact that rotating the vectors $\delta_i(t)$ and $\delta_j(t)$ by equal amounts has no effect on the magnitude of $\mathbf{c}_k(t)$.

Proposition 3: Let $\mathbf{a}$ be an arbitrary 3-vector and R be an arbitrary rotation matrix, then

$$[R\mathbf{a}]_\times = R[\mathbf{a}]_\times R^T.$$

Proposition 4: Let M be an arbitrary 3×3 matrix and R be an arbitrary rotation matrix, then

$$\|s(RMR^T)\| = \|s(M)\|.$$

According to Propositions 3 and 4, for an arbitrary R, we have

$$\begin{aligned}\|\mathbf{c}_k(t)\| &= \|\mathbf{s}\left(-S_2\left(\delta_i^\star(t), \delta_j^\star(t)\right)\right)\| \\ &= \|\mathbf{s}\left(-S_2\left(R\delta_i^\star(t), R\delta_j^\star(t)\right)\right)\|.\end{aligned}$$

The rotation matrix R is selected so that the vector $\delta_i^\star$ is aligned to the global x-axis and $\delta_j^\star$ is places in the global xy-plane. Therefore

$$R\delta_i^\star = [\delta_i, 0, 0]^T$$
$$R\delta_j^\star = [\delta_j \cos\alpha_{ij}, \delta_j \sin\alpha_{ij}, 0]^T$$

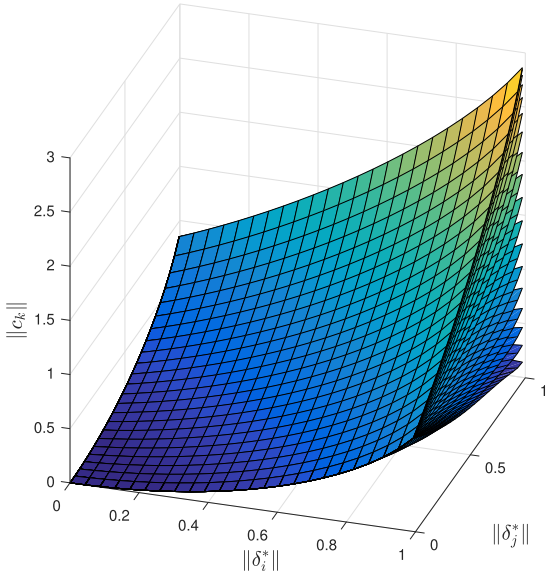

Fig. 12. Magnitude of $\mathbf{c}_k(t)$ as a function of $\|\delta_i^\star\|$, $\|\delta_j^\star\|$, and α_{ij}. Each surface corresponds to a specific α_{ij} (the angle between $\delta_i^\star$ and $\delta_j^\star$). The lowest surface corresponds to $\alpha_{ij}=0$ and the highest surface corresponds to $\alpha_{ij}=\pi$.

where $\delta_i=\|\delta_i^\star\|$, $\delta_j=\|\delta_j^\star\|$, and α_{ij} is the angle between $\delta_i^\star$ and $\delta_j^\star$. Hence, the magnitude of $\mathbf{c}_k(t)$ is a function of three scalars δ_i, δ_j, and α_{ij}, i.e., $\|\mathbf{c}_k(t)\|=f(\delta_i,\delta_j,\alpha_{ij})$. Analytic derivative of f with respect to these parameters reveals that the maximum magnitude of $\|\mathbf{c}_k(t)\|$ for given $\|\delta_i^\star\|$ and $\|\delta_j^\star\|$ occurs when $\alpha_{ij}=\pi$. Also, $f(\delta_i,\delta_j,\pi)$ is an increasing function with respect to δ_i and δ_j. Therefore

$$\begin{aligned}\|\mathbf{c}_k(t)\| &= \|\mathbf{s}\left(-S_2\left(\delta_i^\star(t),\delta_j^\star(t)\right)\right)\| \\ &= \|\mathbf{s}\left(-S_2\left([\delta_i,0,0]^T,[\delta_j\cos(\alpha_{ij}),\delta_j\sin(\alpha_{ij}),0]^T\right)\right)\| \\ &\le \|\mathbf{s}\left(-S_2\left([\delta_i,0,0]^T,-[\delta_j,0,0]^T\right)\right)\| \\ &\le \left\|\mathbf{s}\left(-S_2(\delta_{\bar{k}}^\star(t),-\delta_{\bar{k}}^\star(t))\right)\right\| = \bar{c}(t). \end{aligned} \tag{63}$$

Fig. 12 illustrates the magnitude of $\mathbf{c}_k(t)$ with respect to $\|\delta_i^\star\|$, $\|\delta_j^\star\|$, and α_{ij}.

Appendix C
Theorem 1 Proof Details

Equation (41) implies that

$$\begin{aligned}\Psi_k^\star(t+1) &= \Psi_k^\star(t)\hat{\Psi}_k^T(t) \\ &= \left(I+[\delta_k^\star(t)]_\times+\beta_k(\|\delta_k^\star(t)\|)[\delta_k^\star(t)]_\times^2\right) \\ &\quad\times\left(I-[\hat{\delta_k}(t)]_\times+\beta_k(\|\hat{\delta_k}(t)\|)[\hat{\delta_k}(t)]_\times^2\right).\end{aligned}$$

According to Proposition 3, for an arbitrary rotation matrix R, we have

$$\begin{aligned}R\Psi_k^\star(t+1)R^T &= R\Psi_k^\star(t)\hat{\Psi}_k^T(t)R^T \\ &= \left(I+R[\delta_k^\star(t)]_\times R^T+R\beta_k(\|\delta_k^\star(t)\|)[\delta_k^\star(t)]_\times^2R^T\right) \\ &\quad\left(I-R[\hat{\delta_k}(t)]_\times R^T+R\beta_k(\|\hat{\delta_k}(t)\|)[\hat{\delta_k}(t)]_\times^2R^T\right)\end{aligned}$$

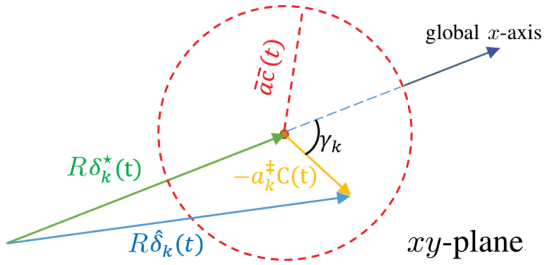

Fig. 13. Vectors $R\delta_k^\star(t)$ and $R\hat{\delta}_k(t)$ in the global coordinate frame.

$$\begin{aligned}&= \left(I+[R\delta_k^\star(t)]_\times+\beta_k(\|R\delta_k^\star(t)\|)[R\delta_k^\star(t)]_\times^2\right) \\ &\quad\left(I-[R\hat{\delta_k}(t)]_\times+\beta_k(\|R\hat{\delta_k}(t)\|)[R\hat{\delta_k}(t)]_\times^2\right).\end{aligned}$$

This indicates that equal rotation of $\delta_k^\star(t)$ and $\hat{\delta}_k(t)$ will not change the magnitude of $\delta_k^\star(t+1)$. The rotation matrix R can be selected in a way that the vector $\delta_k^\star(t)$ is aligned to the global x-axis and the vector $\hat{\delta}_k(t)$ is placed in the global xy-plane. Therefore, the vector $(-a_k^\ddagger C)=\hat{\delta}_k(t)-\delta_k^\star(t)$ also lies in the global xy-plane (see Fig. 13). Therefore

$$\begin{aligned}R\delta_k^\star &= [\delta_k^\star,0,0]^T \\ R\hat{\delta_k} &= [\delta_k^\star,0,0]^T+\zeta_k[\cos\gamma_k,\sin\gamma_k,0]^T\end{aligned}$$

where $\delta_k^\star=\|\delta_k^\star(t)\|$, γ_k is the angle between vectors $\delta_k^\star$ and $(-a_k^\ddagger C)$, and ζ_k is the magnitude of $(-a_k^\ddagger C)$, and $0\le\zeta_k\le\bar{a}\bar{c}$. Therefore, the magnitude of $\delta_k^\star(t+1)$ is a function of three scalars $\delta_k^\star$, $\hat{\delta_k}$, and γ_k, i.e., $\|\delta_k^\star(t+1)\|=f(\delta_k^\star,\zeta_k,\gamma_k)$. Analytic derivative of f with respect to γ_k reveals that for given $\delta_k^\star$ and ζ_k, the maximum magnitude of $\delta_k^\star(t+1)$ occurs when $\gamma_k=0$. Also, $f(\delta_k^\star,\zeta_k,0)$ is an increasing function with respect to $\delta_k^\star$ and ζ_k. Hence, the maximum magnitude of $\delta_k^\star(t+1)$ occurs when $\delta_k^\star(t)$ and $\hat{\delta}_k(t)$ are in the same direction and $\hat{\delta}_k(t)$ takes its maximum value. It can be also concluded that $\delta_{\bar{k}}^\star(t+1)\ge\delta_k^\star(t+1)\tilde{}\forall k\in\{1,\dots,n\}$.

Appendix D
Proof of Lemma 2

Proof: In each step of the RS algorithm, $\hat{\delta}_i(t)$s are calculated by the following optimization problem:

$$\begin{aligned}&\{\hat{\delta}_1(t),\dots,\hat{\delta}_n(t)\} \\ &= \underset{\boldsymbol{\delta}_1\dots\boldsymbol{\delta}_n}{\operatorname{argmin}}\sum_{e_k=\{i,j\}\in E}\omega_k\Big\|\hat{R}_j(t)Z_i^j\hat{R}_i^T(t) \\ &\qquad - I+[\delta_j]_\times-[\delta_i]_\times\Big\|_F^2.\end{aligned} \tag{64}$$

According to (45), (64) can be written as follows:

$$\begin{aligned}&\{\hat{\delta}_1(t),\dots,\hat{\delta}_n(t)\} \\ &= \underset{\boldsymbol{\delta}_1,\dots,\boldsymbol{\delta}_n}{\operatorname{argmin}}\sum_{e_k=\{i,j\}\in E}\omega_k\Big\|\hat{R}_j(t)R_i^j\hat{R}_i^T(t) \\ &\qquad - \hat{R}_j(t)\Phi_i^j\hat{R}_i^T(t)-I+[\delta_j]_\times-[\delta_i]_\times\Big\|_F^2.\end{aligned} \tag{65}$$

The solution to (65) is

$$\hat{\Delta}(t) = A^{\ddagger}(B(t) + E(t)) \tag{66}$$

where $B(t)$ was given in (20) and

$$E(t)^T = [\mathbf{e}_1(t), \ldots, \mathbf{e}_m(t)] \tag{67}$$

in which

$$\mathbf{e}_k = \mathbf{s}\left(\hat{R}_j(t)\Phi_i^j\hat{R}_i^T(t)\right). \tag{68}$$

It is easy to see that $\hat{R}_i(t)$ and $\hat{R}_j(t)$ do not change the magnitude of $\mathbf{e}_k$, only relative rotation measurement noise, i.e., Φ_i^j, affects its magnitude

$$\|\mathbf{e}_k\| = \left\|\mathbf{s}\left(\Phi_i^j\right)\right\|. \tag{69}$$

According to (35) and (66), the difference between solutions is obtained by

$$\Delta^{\star}(t) - \hat{\Delta}(t) = A^{\ddagger}(C(t) - E(t)) \tag{70}$$

and therefore

$$\delta_k^{\star}(t) - \hat{\delta}_k(t) = \mathbf{a}_k^{\ddagger}(t)(C(t) - E(t)) \tag{71}$$

where $\mathbf{a}_k^{\ddagger}(t)$ represents the kth row of $A^{\ddagger}$.

Hence, the magnitude of the difference between solutions satisfy the following inequality:

$$\begin{aligned}\left\|\delta_k^{\star}(t) - \hat{\delta}_k(t)\right\| &\leq \left\|\mathbf{a}_k^{\ddagger}\right\|\left(\max_{\bar{k}}\|\mathbf{c}_{\bar{k}}(t)\| + \max_{k'}\|\mathbf{e}_{k'}(t)\|\right)\\ &\leq \bar{a}(\bar{c}(t) + \bar{e}).\end{aligned} \tag{72}$$

■

Appendix E
Proof of Proposition 2

Proof: Let a_{kk} be the largest diagonal element of a positive definite matrix A. Every principal submatrix of A is also positive definite. In particular, for arbitrary i and j where $i \neq j$, we have

$$\begin{aligned}&\det\begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj}\end{bmatrix} > 0\\ &\Rightarrow a_{ii}a_{jj} - a_{ji}a_{ij} > 0\\ &\overset{a_{ij}=a_{ji}}{\Rightarrow} a_{ii}a_{jj} - a_{ij}^2 > 0\\ &\overset{a_{kk}\text{ is the largest diagonal element}}{\Rightarrow} a_{kk}^2 - a_{ij}^2 > 0 \Rightarrow |a_{kk}| > |a_{ij}|\\ &\Rightarrow \|A\|_{\max} = a_{kk}.\end{aligned}$$

Appendix F
Effect of Removing an Edge of a Graph on Its Structural Parameter

The structural parameter is the maximum diagonal entries of inverse reduced Laplacian matrix in equal noise variance assumption. To show the effect of removing an edge of a graph on its structural parameter, we investigate its effect on diagonal entries of inverse reduced Laplacian matrix.

Assume that $\mathcal{L}$ is the reduced Laplacian matrix of a graph. The diagonal entry $\mathcal{L}(i,i)$ is equal to the degree of the ith vertex of the graph, i.e., the number of edges incident with the ith vertex. The nondiagonal entry $\mathcal{L}(i,j)$ is -1 if vertices i and j are adjacent and is 0 if they are not.

Suppose that $\mathcal{L}$ is the reduced Laplacian matrix of a graph in which vertices 1 and 2 are adjacent. Therefore, $\mathcal{L}$ can be partitioned as follows:

$$\mathcal{L}_{n\times n} = \begin{bmatrix} A & B \\ B^T & D\end{bmatrix}$$

where $A \in \mathbb{R}^{2\times 2}, B \in \mathbb{R}^{2\times(n-2)}, D \in \mathbb{R}^{(n-2)\times(n-2)}$, and

$$A = \begin{bmatrix} d_1 & -1 \\ -1 & d_2\end{bmatrix}, d_1 > 1, d_2 > 1.$$

By eliminating the edge between the vertices 1 and 2, a new graph is formed with a new reduced Laplacian matrix $\mathcal{L}'$

$$\mathcal{L}'_{n\times n} = \begin{bmatrix} A' & B \\ B^T & D\end{bmatrix}$$

where

$$A' = \begin{bmatrix} d_1 - 1 & 0 \\ 0 & d_2 - 1\end{bmatrix}, d_1 > 1, d_2 > 1.$$

The matrix $\mathcal{L}$ can be inverted blockwise

$$\mathcal{L}^{-1} = \begin{bmatrix} A & B \\ B^T & D\end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}B^T)^{-1} & -(A - BD^{-1}B^T)^{-1}BD^{-1} \\ -D^{-1}B^T(A - BD^{-1}B^T)^{-1} & D^{-1} + D^{-1}B^T(A - BD^{-1}B^T)^{-1}BD^{-1}\end{bmatrix}.$$

Assume that

$$BD^{-1}B^T = \begin{bmatrix} e & g \\ g & f\end{bmatrix}.$$

Hence

$$\begin{aligned}A - BD^{-1}B^T &= \begin{bmatrix} d_1 & -1 \\ -1 & d_2\end{bmatrix} - \begin{bmatrix} e & g \\ g & f\end{bmatrix}\\ &= \begin{bmatrix} d_1 - e & -1 - g \\ -1 - g & d_2 - f\end{bmatrix}\end{aligned}$$

$$\Rightarrow (A - BD^{-1}B^T)^{-1} = \frac{\begin{bmatrix} d_2 - f & g + 1 \\ g + 1 & d_1 - e\end{bmatrix}}{(d_1 - e)(d_2 - f) - g^2 - 2g - 1}.$$

The matrix $\mathcal{L}'$ can be also inverted blockwise

$$\mathcal{L}'^{-1} = \begin{bmatrix} A' & B \\ B^T & D\end{bmatrix}^{-1} = \begin{bmatrix} (A' - BD^{-1}B^T)^{-1} & -(A' - BD^{-1}B^T)^{-1}BD^{-1} \\ -D^{-1}B^T(A' - BD^{-1}B^T)^{-1} & D^{-1} + D^{-1}B^T(A' - BD^{-1}B^T)^{-1}BD^{-1}\end{bmatrix}.$$

Hence

$$A' - BD^{-1}B^T = \begin{bmatrix} d_1 - 1 & 0 \\ 0 & d_2 - 1 \end{bmatrix} - \begin{bmatrix} e & g \\ g & f \end{bmatrix} = \begin{bmatrix} d_1 - e - 1 & -g \\ -g & d_2 - f - 1 \end{bmatrix}$$

$$\Rightarrow (A' - BD^{-1}B^T)^{-1} = \frac{\begin{bmatrix} d_2 - f - 1 & g \\ g & d_1 - e - 1 \end{bmatrix}}{(d_1 - e)(d_2 - f) - (d_1 - e) - (d_2 - f) - g^2 + 1}.$$

We first compare the first diagonal entries of $\mathcal{L}^{-1}$ and $\mathcal{L}'^{-1}$ here. Let $a = d_1 - e$ and $b = d_2 - f$, then[9]

$$\mathcal{L}^{-1}(1,1) < \mathcal{L}'^{-1}(1,1)$$

$$\Leftarrow \frac{b}{ab - g^2 - 2g - 1} < \frac{b-1}{ab - a - b - g^2 + 1} \overset{\text{determinants are positive}}{\Leftarrow}$$

$$ab^2 - ab - b^2 - bg^2 + b$$

$$< ab^2 - g^2 b - 2gb - b - ab + g^2 + 2g + 1$$

$$\Leftarrow 0 < -2gb + g^2 + 2g + 1 - 2b - b^2 \Leftarrow 0 < (b - g - 1)^2.$$

Therefore, the first diagonal entry of reduced Laplacian matrix is increased by removing the edge between vertices 1 and 2. The second diagonal entry is increased by the same amount. For other diagonal entries, the diagonal entries of $D^{-1}B^T(A - BD^{-1}B^T)^{-1}BD^{-1}$ and $D^{-1}B^T(A' - BD^{-1}B^T)^{-1}BD^{-1}$ are compared. Let $S = [s_{ij}] = BD^{-1}$. Here, we compare the $(k+2)$th elements of $\mathcal{L}$ and $\mathcal{L}'$

$$\mathcal{L}^{-1}(k+2, k+2) < \mathcal{L}'^{-1}(k+2, k+2)$$

$$\Leftarrow \frac{bs_{1k}^2 + 2gs_{1k}s_{2k} + 2s_{1k}s_{2k} + as_{2k}^2}{ab - g^2 - 2g - 1}$$

$$< \frac{bs_{1k}^2 - s_{1k}^2 + 2gs_{1k}s_{2k} + as_{2k}^2 - s_{2k}^2}{ab - a - b - g^2 + 1} \overset{\text{determinants are positive}}{\Leftarrow}$$

$$\big(b^2 s_{1k}^2 - bs_{1k}^2 a - b^2 s_{1k}^2 - bs_{1k}^2 g^2 + bs_{1k}^2 + 2gs_{1k}s_{2k}ab$$

$$- 2gs_{1k}s_{2k}a - 2gs_{1k}s_{2k}b - 2g^3 s_{1k}s_{2k} + 2gs_{1k}s_{2k}$$

$$+ 2s_{1k}s_{2k}ab - 2s_{1k}s_{2k}a - 2s_{1k}s_{2k}b - 2s_{1k}s_{2k}g^2 + 2s_{1k}s_{2k}$$

$$+ a^2 s_{2k}^2 b - a^2 s_{2k}^2 - abs_{2k}^2 - as_{2k}^2 g^2 + as_{2k}^2\big)$$

$$< \big(ab^2 s_{1k}^2 - abs_{1k}^2 + 2gs_{1k}s_{2k}ab + a^2 bs_{2k}^2 - abs_{2k}^2$$

$$- g^2 bs_{1k}^2 + g^2 s_{1k}^2 - 2g^3 s_{1k}s_{2k} - g^2 as_{2k}^2 + g^2 s_{2k}^2 - 2gbs_{1k}^2$$

$$+ 2gs_{1k}^2 - 4g^2 s_{1k}s_{2k} - 2gas_{2k}^2 + 2gs_{2k}^2 - bs_{1k}^2 + s_{1k}^2$$

$$-2gs_{1k}s_{2k} - as_{2k}^2 + s_{2k}^2\big)$$

$$\Leftarrow 0 < (s_{1k} - s_{2k} + s_{1k}g - s_{2k}g + as_{2k} - bs_{1k})^2.$$

This implies that all diagonal entries of inverse reduced Laplacian matrix are increased by removing the edge between vertices 1 and 2. The same result is achieved by removing any edge using a simple permutation.

[9]To prove the inequality $\mathcal{L}^{-1}(1,1) < \mathcal{L}'^{-1}(1,1)$, we simplify the equations and reach an obvious true inequality. It is clear that the correct way to prove the inequality is to go backward, in accordance with the arrows direction.

TABLE IV
RESULTS OF RS AND SE-SYNC ON A COMPLETE GRAPH WITH $n = \{10, 20, 30\}$ VERTICES

n	rotation angle of measurement noise (deg)	maximum angular residual (deg)	Number of experiments in which $f_{RS}^\star \square f_{SE}^\star$ $=$	$<$	$>$
10	65	124	1000	0	0
	70	149	1000	0	0
	75	155	986	13	2
	80	178	674	312	14
20	70	123	1000	0	0
	75	145	1000	0	0
	80	167	996	4	0
	85	177	934	58	8
30	75	132	1000	0	0
	80	149	1000	0	0
	85	154	1000	0	0
	90	174	921	67	12

The number of 1000 experiments were done for each measurement noise angle. The maximum angular residuals are computed using verified optimal solutions.

TABLE V
RESULTS OF RS AND GTSAM FOR ISOTROPIC LANGEVIN DISTRIBUTION NOISE IN ROTATION MEASUREMENTS OF TORUS AND CUBICLE DATASETS

dataset	κ	RMSE of the rotation noise (deg)	chord	GTSAM	RS
torus	200	2.86 ± 0.02	4049 ± 78	4043 ± 78	4043 ± 78
	50	5.73 ± 0.05	4082 ± 69	4048 ± 67	4048 ± 67
	20	9.05 ± 0.06	4135 ± 79	4040 ± 74	4040 ± 74
	5	18.6 ± 0.08	4544 ± 112	4178 ± 80	4177 ± 80
	2	31.3 ± 0.29	6020 ± 318	4829 ± 371	4652 ± 129
	1.5	37.8 ± 0.38	7645 ± 391	6631 ± 436	5066 ± 144
cubicle	200	2.87 ± 0.02	11157 ± 106	11148 ± 105	11148 ± 105
	50	5.73 ± 0.02	11158 ± 103	11115 ± 103	11115 ± 103
	20	9.06 ± 0.06	11181 ± 118	11075 ± 116	11075 ± 116
	5	18.6 ± 0.09	12068 ± 150	11588 ± 130	11583 ± 130
	2	31.4 ± 0.15	14133 ± 243	12979 ± 495	12552 ± 172
	1.5	37.8 ± 0.24	16228 ± 388	16186 ± 403	13449 ± 117

The table lists concentration parameter κ of Langevin distribution, RMSE of the angle of rotation noise, and the mean and std of the final RS costs (12) obtained by chord, RS, and GTSAM in 20 runs.

APPENDIX G
CONVERGENCE BASIN FOR COMPLETE GRAPHS

To support our claim that "the convergence basin is more wider than the bound obtained in this article" and also compare our RS solution to the solution of the Lagrange dual problem, the relative rotation observations of a set of complete graphs with 10, 20, and 30 vertices are contaminated by a random rotation noise with random axes and different rotation angles. Then, the solution of RS problem, using RS, and the solution of its dual problem, using SE-Sync, are obtained. The initial rotations are chosen randomly. The experiments were repeated 1000 times for each noise angle and for each number of vertices.

Table IV illustrates the number of experiments in which the RS solutions are the same as, better than, or worse than the solutions of the dual problems. The tolerance for accepting the costs to be equal is 10^{-3}. As it is shown in the table, RS and SE-Sync attained the same solution in all experiments with

TABLE VI
FINAL COST VALUES OF DIFFERENT METHODS IN DIFFERENT DATASETS WITH DIFFERENT NOISE LEVELS

Dataset	RMSE of the rotation noise (deg)	10	20	30	40	50	60	70
sphere_a	RS+PS	25776±235	25910±227	26285±209	26657±240	26958±217	26702±213	25596±188
	chord+SE-Sync	25776±235	25910±227	26285±209	26657±240	26983±220	27301±360	28861±424
torus	RS+PS	15976±153	15819±194	15680±168	15789±205	15625±150	14280±117	12204±85
	chord+SE-Sync	15976±153	15819±194	15680±162	16727±421	19345±385	18566±212	16490±180
cube	RS+PS	55822±327	55882±326	55889±309	56301±327	57055±314	57381±357	55780±326
	chord+SE-Sync	55822±327	55882±326	55889±309	56301±327	57055±316	57776±360	59787±645
garage	RS+PS	16990±179	16920±175	16822±187	16534±178	16075±176	15618±357	15375±221
	chord+SE-Sync	16990±179	16920±175	16822±187	16534±178	16075±176	16452±3871	16785±1649
cubicle	RS+PS	43280±278	43651±310	44252±293	45117±308	44968±303	43408±264	40527±253
	chord+SE-Sync	43280±278	43651±310	44252±294	45117±308	44965±302	43408±264	41387±584
rim	RS+PS	76653±344	77712±403	78645±365	78992±360	78036±366	75617±320	69310±342
	chord+SE-Sync	76653±344	77712±403	78645±366	78992±359	78036±366	75614±320	69323±361

TABLE VII
ATE VALUES OF DIFFERENT METHODS IN DIFFERENT DATASETS WITH DIFFERENT NOISE LEVELS

Dataset	RMSE of the rotation noise (deg)	10	20	30	40	50	60	70
sphere_a	RS+PS	485±21	1009±49	1477±66	1794±84	2243±114	2809±203	3955±438
	chord+SE-Sync	485±21	1010±49	1479±67	1797±85	2248±117	2835±154	4518±458
torus	RS+PS	43±1	87±3	131±10	191±28	301±38	401±23	445±21
	chord+SE-Sync	43±1	87±3	131±6	184±11	310±31	405±23	462±26
cube	RS+PS	31±0	57±1	63±1	96±2	129±3	169±41	219±53
	chord+SE-Sync	31±0	57±1	63±1	96±2	129±3	159±5	201±6
garage	RS+PS	5±1	19±4	47±9	104±26	225±51	1223±476	2470±297
	chord+SE-Sync	5±1	19±4	47±9	104±25	215±46	1145±452	2365±297
cubicle	RS+PS	14±2	28±4	38±5	50±6	61±20	80±16	124±50
	chord+SE-Sync	14±2	28±4	38±5	50±6	61±8	80±16	121±29
rim	RS+PS	21±2	42±4	53±6	71±6	94±12	126±20	203±27
	chord+SE-Sync	21±2	42±4	53±6	71±7	96±13	130±24	227±41

up to 70° measurement noise angle. The optimality of these solutions are certified by the optimality verification of SE-Sync. This implies that the algorithm converges in datasets with much more rotation measurement noise than the bound obtained in this article. Also, the proposed algorithm usually gives a better solution for high-noise datasets in which SE-Sync does not give an optimal solution. The maximum angular residual is also given in the table, which shows that it can be much larger than the maximum rotation angle of measurement noise.

APPENDIX H
RS VERSUS GTSAM (LANGEVIN DISTRIBUTION NOISE)

Table V depicts the mean and std of the final RS costs attained by chord, RS, and GTSAM for the torus and the cubicle datasets at different amounts of Langevin noise. RS and GTSAM show equal performance for up to about 19° RMSE of the angle of rotation noise (concentration parameter of 5). For higher amount of noise, RS outperforms GTSAM.

APPENDIX I
RS+PS VERSUS CHORD+SE-SYNC (LANGEVIN DISTRIBUTION ROTATION NOISE AND GAUSSIAN TRANSLATION NOISE)

In this part, the performance of RS+PS and chord+SE-Sync are evaluated on datasets with different noise levels. The rotation observations were contaminated by random rotations with random axes and Langevin distributed random angles. The translation observations were also contaminated by a random Gaussian noise. The standard deviations of translation noises were chosen proportional to the concentration parameters of rotation noises. The weights, i.e., λ_ks and ω_ks, were chosen according to the noise levels, i.e., proportional to the concentration parameters of rotation noise and the standard deviations of translation noise.

Each experiment was repeated 100 times. Table VI shows the mean and standard deviation of cost attained by RS+PS and chord+SE-Sync in different datasets with different amounts of noise. In almost all scenarios, RS+PS attains equal or better cost than that by chord+SE-Sync. The absolute trajectory errors (ATEs) are also reported in Table VII. This localization error metrics measures the distance between all nodes and their ground truth trajectory. In all the experiments, the solution of noiseless data is assumed as the ground truth. In most of the experiments, RS+PS converged to a solution with lower ATE. There are few cases where ATEs attained by chord+SE-Sync are better than those by RS+PS, although their cost function are worse. We observed that in these cases, the added amount of noise is so high such that it deforms the trajectory and therefore the optimal solution of the noisy data takes distance from the noiseless trajectory. In these cases, the optimal minimum of the cost function may have more ATE in comparison to suboptimal solutions.

APPENDIX J
COMPARING RS+PS WITH CHORD+GTSAM

Carlone *et al.* [28] compared different initializers and solvers for the PGO problem and showed that chord+GTSAM has the best performance among gradient-based methods. In this section, we see how our proposed methods can improve this baseline solver.

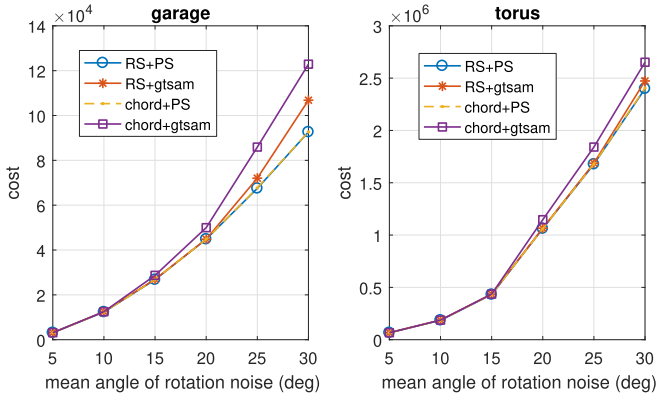

Fig. 14. Comparison of RS and PS, with chord and GTSAM in garage and torus datasets with different amounts of noise.

Fig. 14 compares the cost (10) attained by four combinations of initializers and solvers, initializers being RS and chord and solvers being PS and GTSAM, for garage and torus datasets with different amounts of noise. The noise is created similar to what we explained in Section VIII-B. As it can be seen in the figure, the methods RS+PS and chord+PS attain the best results in all experiments. We can see that our RS algorithm outperforms the chord initialization method and can help GTSAM to reach more accurate solutions. It is clear from the figure that our PS algorithm outperforms the GTSAM algorithm in the high-noise regimes.

ACKNOWLEDGMENT

The authors appreciate the anonymous reviewers' consideration and guidance in all rounds of the review process. Their comments really helped improve the quality and contribution of this work.

REFERENCES

[1] G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based SLAM," *IEEE Intell. Transp. Syst. Mag.*, vol. 2, no. 4, pp. 31–43, Winter 2010.

[2] G. Piovan, I. Shames, B. Fidan, F. Bullo, and B. D. Anderson, "On frame and orientation localization for relative sensing networks," *Automatica*, vol. 49, no. 1, pp. 206–213, 2013.

[3] D. Martinec and T. Pajdla, "Robust rotation and translation estimation in multiview reconstruction," in *Proc. IEEE Conf. Comput. Vision. Pattern Recognit.*, 2007, pp. 1–8.

[4] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Auton. Robots*, vol. 4, no. 4, pp. 333–349, 1997.

[5] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G^2o: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 3607–3613.

[6] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. Robot.*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.

[7] E. Olson, J. Leonard, and S. Teller, "Fast iterative alignment of pose graphs with poor initial estimates," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2006, pp. 2262–2269.

[8] D. M. Rosen, M. Kaess, and J. J. Leonard, "An incremental trust-region method for robust online sparse least-squares estimation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 1262–1269.

[9] D. M. Rosen, M. Kaess, and J. J. Leonard, "Rise: An incremental trust-region method for robust online sparse least-squares estimation," *IEEE Trans. Robot.*, vol. 30, no. 5, pp. 1091–1108, Oct. 2014.

[10] L. Carlone, "A convergence analysis for pose graph optimization via Gauss–Newton methods," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2013, pp. 965–972.

[11] S. Huang, Y. Lai, U. Frese, and G. Dissanayake, "How far is slam from a linear least squares problem?," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2010, pp. 3011–3016.

[12] S. Huang, H. Wang, U. Frese, and G. Dissanayake, "On the number of local minima to the point feature based slam problem," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 2074–2079.

[13] H. Wang, S. Huang, U. Frese, and G. Dissanayake, "The nonlinearity structure of point feature SLAM problems with spherical covariance matrices," *Automatica*, vol. 49, no. 10, pp. 3112–3119, 2013.

[14] G. Grisetti, C. Stachniss, and W. Burgard, "Nonlinear constraint network optimization for efficient map learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 428–439, Sep. 2009.

[15] L. Carlone, R. Aragues, J. A. Castellanos, and B. Bona, "A fast and accurate approximation for planar pose graph optimization," *Int. J. Robot. Res.*, vol. 33, no. 7, pp. 965–987, 2014.

[16] L. Carlone and A. Censi, "From angular manifolds to the integer lattice: Guaranteed orientation estimation with application to pose graph optimization," *IEEE Trans. Robot.*, vol. 30, no. 2, pp. 475–492, Apr. 2014.

[17] D. M. Rosen, C. DuHadway, and J. J. Leonard, "A convex relaxation for approximate global optimization in simultaneous localization and mapping," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 5822–5829.

[18] L. Carlone, G. C. Calafiore, C. Tommolillo, and F. Dellaert, "Planar pose graph optimization: Duality, optimal solutions, and verification," *IEEE Trans. Robot.*, vol. 32, no. 3, pp. 545–565, Jun. 2016.

[19] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group," *Int. J. Robot. Res.*, vol. 38, no. 2/3, pp. 95–125, 2019.

[20] J. Briales and J. Gonzalez-Jimenez, "Fast global optimality verification in 3D SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 4630–4636.

[21] L. Carlone and F. Dellaert, "Duality-based verification techniques for 2D SLAM," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 4589–4596.

[22] L. Carlone, D. M. Rosen, G. Calafiore, J. J. Leonard, and F. Dellaert, "Lagrangian duality in 3d slam: Verification techniques and optimal solutions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 125–132.

[23] I. Aloise and G. Grisetti, "Chordal based error function for 3-D pose-graph optimization," *IEEE Robot. Autom. Lett.*, vol. 5, no. 1, pp. 274–281, Jan. 2020.

[24] L. Carlone, R. Aragues, J. A. Castellanos, and B. Bona, "A linear approximation for graph-based simultaneous localization and mapping," in *Robotics: Science and Systems*. Cambridge, MA, USA: MIT Press, 2011, pp. 41–48.

[25] S.-M. Nasiri, H. Moradi, and R. Hosseini, "A linear least square initialization method for 3d pose graph optimization problem," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2018, pp. 2474–2479.

[26] R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G^2o: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 3607–3613.

[27] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Inst. Robot. Intell. Mach., Georgia Inst. Technol., Atlanta, GA, USA, Tech. Rep. GT-RIM-CP&R-2012-002, 2012.

[28] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert, "Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 4597–4604.

[29] J. R. Peters, D. Borra, B. Paden, and F. Bullo, "Sensor network localization on the group of three-dimensional displacements," *SIAM J. Control Optim.*, vol. 53, no. 6, pp. 3534–3561, 2015.

[30] R. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation averaging," *Int. J. Comput. Vision*, vol. 103, no. 3, pp. 267–305, 2013.

[31] J. Fredriksson and C. Olsson, "Simultaneous multiple rotation averaging using Lagrangian duality," in *Proc. Asian Conf. Comput. Vision*, 2012, pp. 245–258.

[32] R. Tron and R. Vidal, "Distributed 3-D localization of camera sensor networks from 2-D image measurements," *IEEE Trans. Autom. Control*, vol. 59, no. 12, pp. 3325–3340, Dec. 2014.

[33] O. Ozyesil, A. Singer, and R. Basri, "Stable camera motion estimation using convex programming," *SIAM J. Imag. Sci.*, vol. 8, no. 2, pp. 1220–1262, 2015.

[34] A. S. Bandeira, A. Singer, and D. A. Spielman, "A Cheeger inequality for the graph connection Laplacian," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 4, pp. 1611–1630, 2013.
[35] J. G. Mangelson, J. Liu, R. M. Eustice, and R. Vasudevan, "Guaranteed globally optimal planar pose graph and landmark slam via sparse-bounded sums-of-squares programming," in *Proc. Int. Conf. Robot. Autom.*, 2019, pp. 9306–9312.
[36] L. Wang and A. Singer, "Exact and stable recovery of rotations for robust synchronization," *Inf. Inference, J. IMA*, vol. 2, no. 2, pp. 145–193, 2013.
[37] R. Hartley, K. Aftab, and J. Trumpf, "L1 rotation averaging using the Weiszfeld algorithm," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2011, pp. 3041–3048.
[38] R. Tron and R. Vidal, "Distributed image-based 3-D localization of camera sensor networks," in *Proc. IEEE Conf. Decis. Control.*, 2009, pp. 901–908.
[39] F. Arrigoni, L. Magri, B. Rossi, P. Fragneto, and A. Fusiello, "Robust absolute rotation estimation via low-rank and sparse matrix decomposition," in *Proc. Int. Conf. 3D Vision*, 2014, pp. 491–498.
[40] A. Eriksson, C. Olsson, F. Kahl, and T.-J. Chin, "Rotation averaging and strong duality," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 127–135.
[41] E. Olson and M. Kaess, "Evaluating the performance of map optimization algorithms," in *Proc. Workshop Good Exp. Methodology Robot.*, 2009.
[42] H. Wang, G. Hu, S. Huang, and G. Dissanayake, "On the structure of nonlinearities in pose graph SLAM," in *Robotics: Science and Systems*. Cambridge, MA, USA: MIT Press, 2013, pp. 425–432.
[43] K. Khosoussi, S. Huang, and G. Dissanayake, "Novel insights into the impact of graph structure on slam," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 2707–2714.
[44] K. Khosoussi, S. Huang, and G. Dissanayake, "Good, bad and ugly graphs for SLAM," in *Proc. Workshop Problem Mobile Sensors*, 2015.
[45] K. Khosoussi, S. Huang, and G. Dissanayake, "Tree-connectivity: Evaluating the graphical structure of SLAM," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2016, pp. 1316–1322.
[46] C. Godsil and G. Royle, *Algebraic Graph Theory* (Graduate Texts in Mathematics). New York, NY, USA: Springer, 2001.
[47] R. B. Bapat, *Graphs and Matrices*. New York, NY, USA: Springer, 2010.
[48] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

Seyed-Mahdi Nasiri (Member, IEEE) received the B.Sc. degree in electrical engineering from Shahed University, Tehran, Iran, in 2010, and the M.Sc. degree in control engineering from the Department of Electrical Engineering, Sharif University of Technology, Tehran, in 2012. He is currently working toward the Ph.D. degree in machine intelligence and robotics with the School of Electrical and Computer Engineering, University of Tehran, Tehran.

From 2008 to 2012, he was a member of Omid Robotics Team and received several international competitions rewards.

Reshad Hosseini received the B.Sc. degree in electrical engineering from the School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran, in 2004, and the Ph.D. degree in computer science from the Faculty of Elektrotechnik und Informatik, TU Berlin, Berlin, Germany, in 2012.

He is currently an Assistant Professor with the School of Electrical and Computer Engineering, University of Tehran. His current research interests include manifold optimization, large-scale mixture models, three-dimensional reconstruction, neural system identification, and deep learning for recognition.

Hadi Moradi (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran, in 1988, and the Ph.D. degree in computer engineering from the University of Southern California, Los Angeles, CA, USA, in 2012.

He is currently an Associate Professor with the School of Electrical and Computer Engineering, University of Tehran. His current research interests include robotics, cognitive games, intelligent autism screening, and rehabilitation.