A Convergence Analysis for Pose Graph Optimization via Gauss-Newton Methods

Luca Carlone

***Abstract*— In this work we present a convergence analysis of the pose graph optimization problem, that arises in the context of mobile robots localization and mapping. The analysis is performed under some simplifying assumptions on the structure of the measurement covariance matrix and provides non trivial results on the aspects affecting convergence in nonlinear optimization based on Gauss-Newton methods. We also provide estimates for the basin of attraction of the maximum likelihood solution and results on the uniqueness of the global optimum. The results confirm observations of related work and explain why common Simultaneous Localization and Mapping (SLAM) instances are so well-behaved in terms of convergence. Moreover, as a by-product of the derivation, we present different techniques that can enlarge the convergence radius *a-priori* (i.e., during robot operation) or *a-posteriori* (i.e., given the data). We validate the theoretical derivation with extensive experiments on standard benchmarking datasets.**

I. Introduction

A *pose graph* is a model used in probabilistic robotics to formalize the Simultaneous Localization and Mapping (SLAM) problem. In the pose graph, each node represents a pose assumed by a mobile robot, whereas an edge exists between two nodes if a relative measurement (inter-nodal constraint) is available between the corresponding poses. Then, the objective of the SLAM problem is to estimate nodes' poses (*pose graph configuration*) that maximize the likelihood of inter-nodal measurements. The maximum likelihood formulation naturally leads to a nonlinear non-convex optimization problem, whose (possibly non unique) minimum corresponds to the estimate of the pose graph configuration.

Related work. There exists an extensive literature dealing with the pose graph optimization problem. After the pioneering work [14], and the comprehensive formulation of Thrun *et al.* [20], several authors provided relevant contributions to exploit the structure of the optimization problem, for the purpose of speeding up the computation. Konolige investigated a reduction scheme, for shrinking the dimension of the optimization variable [12]; Frese *et al.* proposed a multilevel relaxation approach [6]; Olson *et al.* [17] and Grisetti *et al.* [8] showed the effectiveness of stochastic gradient descent for solving the optimization problem. More recently, Kaess *et al.* [10] presented a *variable ordering* technique for solving in incremental fashion the SLAM problem, while Carlone *et al.* [3] presented an approximation for pose graph optimization in a planar setup. In most of the previous papers, it was remarked that common problem instances are particularly well-behaved: although pose graph optimization is a non-convex problem, the solution can be find efficiently and the risk of being trapped in local minima can be ward off in most cases. Such observations pushed researchers to investigate the structure of the optimization problem, going beyond the computational aspects. In [18], Rizzini presented an interesting closed-form solution and showed relevant insights on the problem, under the hypothesis that the measurement covariances are identity matrices. In [7] Grisetti *et al.* investigated the formulation of the problem on manifolds, for correctly addressing the treatment of the angular information. In [9], Huang *et al.* discussed the convexity properties of SLAM and they drew conclusions about the importance of the orientation measurements, that are also confirmed by the results presented in this paper.

Contribution. In this paper we formally analyze the convergence properties of pose graph optimization, for the case in which a Gauss-Newton approach is used for computing a minimum of the cost function. Our attempt is to answer to the following questions: (i) why are common problem instances so well-behaved in terms of convergence to a global minimum? (ii) what aspects do affect the convergence properties of the optimization problem? (iii) what kind of actions can be taken *a-priori* (i.e., during robot operation) or *a-posteriori* (i.e., after all measurements are acquired) to improve global convergence? A partial answers to question (i) can be found in [9], where it is recognized the role of the orientation measurements in defining the convexity properties of the optimization problem. In this context, we will identify several non trivial aspects influencing convergence, hence answering to questions (i) and (ii). In particular, we show that a special role is played by three factors: the accuracy of the orientation measurements with respect to the Cartesian measurements, the inter-nodal distances, and the connectivity of the graph. We also provide an estimate for the basin of attraction of the maximum likelihood solution and results on the uniqueness of the global optimum. However, we consider the latter as minor results since experimental tests show that such estimates are conservative. As a by-product of the derivation, we present different techniques that can enhance global convergence, providing our answer for the issue (iii); in some cases these techniques provide suggestions for the design of the pose graph, i.e., what actions the robot can take during operation, in order to create a well-behaved problem instance. In other cases, the techniques show how, for given input data, the convergence radius can be increases *arbitrarily*, with slight modifications of the problem instance.

Remark. For sake of simplicity we tailor the formulation to the single robot batch SLAM problem. However, the results also apply to the case in which the different poses in the pose-graph correspond to different agents: this is the case in which the pose graph models multi robot localization [1] or multi robot SLAM [2] problems.

Notation and Preliminaries. A *directed* graph $\mathcal{G}$ is a pair $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a finite set of elements, called *vertices* or *nodes*, and $\mathcal{E}$ is a set containing ordered pairs of nodes. A generic element $e \in \mathcal{E}$, referred to as *edge*, is in the form

This work was partially funded by Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) under MEMONET National Research Project.

L. Carlone is with the Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy. luca.carlone@polito.it

$e = (i, j)$, meaning that edge e, incident on nodes i and j, leaves i (*tail*) and is directed towards node j (*head*) [11]. The number of nodes and edges are denoted with $n+1$ and m, respectively, i.e., $|\mathcal{V}| = n + 1$ and $|\mathcal{E}| = m$ ($|\cdot|$ denotes the cardinality of a set). We denote with $\mathcal{N}_{\text{out}}(i)$ the set of outgoing neighbors of node i, i.e., $\mathcal{N}_{\text{out}}(i) = \{j : (i, j) \in \mathcal{E}\}$. The *incidence matrix* $\mathcal{A}$ of a directed graph is a matrix in $\mathbb{R}^{(n+1)\times m}$ in which each column contains the information of an edge in $\mathcal{E}$; in particular the column corresponding to the edge $e = (i, j)$, has the i-th element equal to -1, the j-th element equal to $+1$ and all the others equal to zero. The *Laplacian matrix* $\mathcal{L}$ of a graph is a matrix such that the i-th element of the main diagonal is equal to the degree of node i (number of edges incident on i), and the off-diagonal entry in position (i, j) (with $i \neq j$) is equal to -1 if nodes i and j are adjacent (otherwise the entry is zero). The Laplacian matrix can be computed from the incident matrix as $\mathcal{L} = \mathcal{A}\mathcal{A}^\top$.

$M_{n\times m}$ denotes a matrix with n rows and m columns and $\otimes$ denotes the Kronecker product. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, $\mathbf{0}_n$ denotes a (column) vector of all zeros of dimension n. $\|M\|$ denotes the spectral norm (maximum singular value) of M, or the standard Euclidean norm, in case of vectors. Therefore, for a rectangular matrix M and for a positive semi-definite symmetric square matrix N, it holds:

- $\|M\| = \sqrt{\lambda_{max}\{MM^\top\}} = \sqrt{\lambda_{max}\{M^\top M\}}$ ($\lambda_{max}\{\cdot\}$ is the largest eigenvalue of the matrix);
- $\|N\| = \lambda_{max}\{N\}$;
- if M is full column rank, $\|M^\dagger\| \doteq \|(M^\top M)^{-1}M^\top\| = \frac{1}{\sqrt{\lambda_{min}\{MM^\top\}}} = \frac{1}{\sqrt{\lambda_{min}\{M^\top M\}}}$ ($M^\dagger$ is the pseudoinverse of M, and $\lambda_{min}\{\cdot\}$ is the smallest nonzero eigenvalue);
- if N is invertible, $\|N^{-1}\| = \frac{1}{\lambda_{min}\{N\}}$.

II. Problem Formulation

The objective of pose graph optimization is to provide an estimate of the poses assumed by the robot, namely $x \doteq \{x_0, \ldots, x_n\}$, that maximizes measurements likelihood. x is called *configuration* of poses; the $n + 1$ poses are in the form $x_i = [p_i^\top \ \theta_i]^\top \in \text{SE(2)}$, where $p_i \in \mathbb{R}^2$ is the Cartesian position of the i-th pose, and θ_i is its orientation. Sensor measurements are usually preprocessed to obtain relative pose information between pose pairs and inter-nodal measurements are modelled as edges in the pose graph. Therefore, to each edge (i, j) in the edge set $\mathcal{E}$, we associate a measurement:

$$\bar{\xi}_{i,j} = \xi_{i,j} + \epsilon_{i,j} = \begin{bmatrix} R_i^\top(p_j - p_i) \\ \langle\theta_j - \theta_i\rangle_{2\pi} \end{bmatrix} + \begin{bmatrix} \epsilon_{i,j}^\Delta \\ \epsilon_{i,j}^\delta \end{bmatrix}, \ \forall\,(i, j) \in \mathcal{E}, \tag{1}$$

where $\xi_{i,j}$ is the true (unknown) relative pose, $\epsilon_{i,j} \in \mathbb{R}^3$ is the *measurement noise*[1], $R_i \in \mathbb{R}^{2\times2}$ is a planar rotation matrix of an angle θ_i, $\langle\cdot\rangle_{2\pi}$ is a modulo-(2π) operator that forces angular measurements in the manifold SO(2), and $\epsilon_{i,j}^\Delta$ and $\epsilon_{i,j}^\delta$ are the (possibly correlated) *Cartesian* and *orientation noises*. According to related literature, we assume $\epsilon_{i,j}$ to be zero mean Gaussian noise, i.e., $\epsilon_{i,j} \sim \mathcal{N}(\mathbf{0}_3, P_{i,j})$, being $P_{i,j}$ a 3 by 3 covariance matrix. $\xi_{i,j}$ describes the relative transformation that leads pose i to overlap with pose j. By convention, if an edge is directed from node i to node j, the corresponding relative measurement is expressed in the reference frame of node i. We denote with m the number of available measurements, i.e., $|\mathcal{E}| = m$. Notice that we can rewrite each measurement as $\bar{\xi}_{ij} = [(\bar{\Delta}_{i,j}^l)^\top \ \check{\delta}_{i,j}]^\top$, where $\bar{\Delta}_{i,j}^l \in \mathbb{R}^2$ denotes the *relative position* measurement, and $\check{\delta}_{i,j} \in \text{SO(2)}$ denotes the *relative orientation* measurement. The superscript l in $\bar{\Delta}_{i,j}^l$ remarks that the relative position vector is expressed in a local frame.

In [3] the authors showed that the measurement model associated to the relative orientation measurements can be made linear by adding a suitable multiple of 2π, i.e, $\langle\theta_j - \theta_i\rangle_{2\pi} = \theta_j - \theta_i + 2k_{i,j}\pi$, with $\theta_i, \theta_j \in \mathbb{R}$, being $2k_{i,j}\pi$ ($k_{i,j} \in \mathbb{Z}$) a *regularization term*. In this paper we assume that the regularization terms have been correctly computed, using the technique presented in [3], and we call $\bar{\delta}_{i,j}$ the regularized measurements, i.e., we define $\bar{\delta}_{i,j} = \check{\delta}_{i,j} - 2k_{i,j}\pi$. Then the measurement model becomes:

$$\begin{bmatrix} \bar{\Delta}_{i,j}^l \\ \bar{\delta}_{i,j} \end{bmatrix} = \begin{bmatrix} R_i^\top(p_j - p_i) \\ \theta_j - \theta_i \end{bmatrix} + \begin{bmatrix} \epsilon_{i,j}^\Delta \\ \epsilon_{i,j}^\delta \end{bmatrix}, \ \forall\,(i, j) \in \mathcal{E}. \tag{2}$$

Let us number the available measurements from 1 to m; let us stack all the relative position measurements in the vector $\bar{\Delta}^l = [(\bar{\Delta}_1^l)^\top \ (\bar{\Delta}_2^l)^\top \ \ldots \ (\bar{\Delta}_m^l)^\top]^\top$, and all the relative orientation measurements in the vector $\bar{\delta} = [\bar{\delta}_1 \ \bar{\delta}_2 \ \ldots \ \bar{\delta}_m]^\top$. Let us stack similarly the orientation and the position noises, obtaining ϵ^Δ and ϵ^δ, respectively. Repeating the same procedure for the unknown robot positions and orientations we get the *nodes' position* $p = [p_1^\top \ \ldots \ p_n^\top]^\top$ and the *nodes' orientation* $\theta = [\theta_1 \ \ldots \ \theta_n]^\top$ (the first pose x_0 is conventionally set to the origin of the reference frame, then it does not take part in the optimization problem). According to [3], we can write in compact form the measurement model as:

$$\begin{bmatrix} \bar{\Delta}^l \\ \bar{\delta} \end{bmatrix} = \begin{bmatrix} R^\top A_2^\top p \\ A^\top\theta \end{bmatrix} + \begin{bmatrix} \epsilon^\Delta \\ \epsilon^\delta \end{bmatrix}, \tag{3}$$

where A is the *reduced* incidence matrix of the pose graph, $A_2 = A \otimes \mathbf{I}_2$, and $R = R(\theta)$ is a block diagonal matrix containing the rotation matrices that transform the corresponding local measurements in the global frame, i.e., the nonzero entries of R are in positions $(2k-1, 2k-1), (2k-1, 2k), (2k, 2k-1), (2k, 2k),\ k = 1, \ldots, m$, and the k-th diagonal block contains the rotation matrix converting the k-th relative position measurement in the global frame.

Then the maximum likelihood estimate of the network configuration $x = \{p, \theta\}$ attains the minimum of the following cost function (see [3] and the references therein):

$$f(x) = \begin{bmatrix} R^\top A_2^\top p - \bar{\Delta}^l \\ A^\top\theta - \bar{\delta} \end{bmatrix}^\top \Omega \begin{bmatrix} R^\top A_2^\top p - \bar{\Delta}^l \\ A^\top\theta - \bar{\delta} \end{bmatrix}, \tag{4}$$

where Ω is the *information matrix* of the noise vector $[(\epsilon^\Delta)^\top \ (\epsilon^\delta)^\top]^\top$. Therefore, the pose-based SLAM problem reduces to find a global minimum of $f(x)$, i.e., $x^{**} = \arg\min f(x)$ (*pose graph optimization problem*).

III. Convergence Analysis

In this section we provide some results on the convergence of a Gauss-Newton method, applied to the optimization problem at hand. Our derivation is based on a recent result

[1] A more formal definition of the uncertainty is $\epsilon_{i,j} = \Psi^{-1}(\bar{\xi}_{i,j} \ominus \xi_{i,j})$, where Ψ is the exponential map from a vector of the Lie algebra to an element of the Lie group. For sake of clarity, we use a less formal definition, with the understanding that the manifold locally behaves as a vector space.

on convergence for Gauss-Newton techniques [5]. For sake of clarity we here recall Theorem 18 from [5].

Theorem 1 (Theorem 18, [5]): Let us consider an optimization problem in the form:

$$\min_{\eta} \|r(\eta)\|^2, \tag{5}$$

where $\eta \in X_\eta \subseteq \mathbb{R}^{n_\eta}$ is the *optimization variable*, and $r(\cdot) : X_\eta \to \mathbb{R}^{m_\eta}$, is a continuously differentiable mapping (*residual error function*), with $m_\eta > n_\eta$. Let us call η^{**} a solution of (5) and define:

$$\psi \doteq \|r(\eta^{**})\| \tag{6}$$

$$\phi \doteq \|r'(\eta^{**})^\dagger\| \tag{7}$$

$$\kappa \doteq \sup\{\varepsilon \geq 0 : \mathcal{B}(\eta^{**}, \varepsilon) \subset X_\eta\}, \tag{8}$$

where $r'(\eta^{**}) = \frac{\partial r'(\eta)}{\partial \eta}\big|_{\eta^{**}}$ (Jacobian of $r(\cdot)$ with respect to η, evaluated in η^{**}), $r'(\eta^{**})^\dagger$ denotes the pseudoinverse of $r'(\eta^{**})$, and $\mathcal{B}(\eta^{**}, \varepsilon)$ is the Euclidean Ball with radius ε and center η^{**}. If the following conditions are satisfied:

- **Condition 1**: $r'(\eta^{**})$ is injective;
- **Condition 2**: there exists K such that:

$$\begin{cases} \textbf{(2.1)} & \|r'(\eta^a) - r'(\eta^b)\| \leq K\|\eta^a - \eta^b\|, \ \ \forall\, \eta^a, \eta^b \in \mathcal{B}(\eta^{**}, \kappa) \\ \textbf{(2.2)} & \sqrt{2}\psi\phi^2 K < 1 \end{cases}; \tag{9}$$

then, starting from any initial guess in the ball $\mathcal{B}(\eta^{**}, \gamma)$ with $\gamma = \min\left(\kappa, \frac{2-2\sqrt{2}K\phi^2\psi}{3K\phi}\right)$, the Gauss-Newton method, applied to problem (5), is well defined and converges to η^{**}. Moreover, defining $\phi_0 = \left\|[r'(\eta^{**})^\top r'(\eta^{**})]^{-1}\right\|$, if $2\psi\phi_0 K < 1$, then the solution η^{**} is unique in $\mathcal{B}(\eta^{**}, \gamma_0)$, with $\gamma_0 = \frac{2-2K\phi_0\psi}{K\phi}$.

□

We study the problem under the simplifying assumption that the measurement information matrix has the following structure:

$$\Omega = \begin{bmatrix} w_\Delta^2 \mathbf{I}_{2m} & \mathbf{0}_{2m,m} \\ \mathbf{0}_{m,2m} & w_\delta^2 \mathbf{I}_m \end{bmatrix}.$$

The previous assumption requires that the noises on the relative position measurements and on the relative orientation measurements have spherical covariance matrices. Accordingly, the cost function (4) can be rewritten as:

$$f(x) = \begin{bmatrix} R^\top A_2^\top p - \bar{\Delta}^l \\ A^\top\theta - \bar{\delta} \end{bmatrix}^\top \begin{bmatrix} w_\Delta^2 \mathbf{I}_{2m} & \mathbf{0}_{2m,m} \\ \mathbf{0}_{m,2m} & w_\delta^2 \mathbf{I}_m \end{bmatrix} \begin{bmatrix} R^\top A_2^\top p - \bar{\Delta}^l \\ A^\top\theta - \bar{\delta} \end{bmatrix} = \left\| \begin{bmatrix} w_\Delta(A_2^\top p - R\bar{\Delta}^l) \\ w_\delta(A^\top\theta - \bar{\delta}) \end{bmatrix} \right\|^2. \tag{10}$$

Now we can start by observing that a minimum of (10) has to annihilate the gradient of the cost function. Let us compute the gradient with respect to nodes' position p, and let us impose it to be zero:

$$\begin{aligned} \nabla_p(f) = 2w_\Delta^2 A_2(A_2^\top p - R(\theta)\bar{\Delta}^l) = \mathbf{0}_{2n} \Longrightarrow \\ \Longrightarrow p = (A_2 A_2^\top)^{-1} A_2 R(\theta)\bar{\Delta}^l. \end{aligned} \tag{11}$$

Therefore, for any given choice of nodes' orientation θ, choosing p as in equation (11) assures that the objective is minimized with respect to p. Hence, we can substitute the expression (11) in the cost function (10), obtaining an optimization problem in the sole variable θ:

$$\min_\theta \left\| \begin{bmatrix} w_\Delta[A_2^\top (A_2 A_2^\top)^{-1} A_2 R(\theta)\bar{\Delta}^l - R(\theta)\bar{\Delta}^l] \\ w_\delta(A^\top\theta - \bar{\delta}) \end{bmatrix} \right\|^2. \tag{12}$$

The following result allows to give more structure to the previous cost function.

Theorem 2 (Duality of the estimation in cycle space): The matrix $A_2^\top (A_2 A_2^\top)^{-1} A_2$ satisfies $A_2^\top (A_2 A_2^\top)^{-1} A_2 + C_2^\top (C_2 C_2^\top)^{-1} C_2 = \mathbf{I}_{2m}$, where $C_2 = C \otimes \mathbf{I}_2$ and C is a *cycle basis matrix* of the pose graph. Moreover $\|A_2^\top (A_2 A_2^\top)^{-1} A_2\| = \|C_2^\top (C_2 C_2^\top)^{-1} C_2\| = 1$.

Proof. We first observe that $A_2^\top (A_2 A_2^\top)^{-1} A_2$ and $C_2^\top (C_2 C_2^\top)^{-1} C_2$ are orthogonal projection matrices, since they are idempotent and symmetric [15]. Therefore it holds the result on the unit norm, according to [15], equation (5.13.10). Moreover, $A_2^\top$ is a basis for a subspace $\mathcal{S}_2^A$ of dimension $2n$ of $\mathbb{R}^{2m}$, whereas $C_2^\top$ is a basis of the null-space of $\mathcal{S}_2^A$, therefore according to [15], equation (5.13.6), it holds $A_2^\top (A_2 A_2^\top)^{-1} A_2 + C_2^\top (C_2 C_2^\top)^{-1} C_2 = \mathbf{I}_{2m}$. □

According to Theorem 2, we can rewrite the optimization problem (12) as:

$$\min_\theta \left\| \begin{bmatrix} -w_\Delta D_2 R(\theta)\bar{\Delta}^l \\ w_\delta(A^\top\theta - \bar{\delta}) \end{bmatrix} \right\|^2, \tag{13}$$

with $D_2 = C_2^\top (C_2 C_2^\top)^{-1} C_2$. We can now apply the result of Theorem 1, considering θ as optimization variable. Since the application of Theorem 1 is not straightforward, we discuss some intermediate results, before presenting the key result in Theorem 3. We structure the rest of this section as follows:

- Lemma 1 assesses Condition 1 in Theorem 1;
- Lemma 2 discusses Condition 2.1 in Theorem 1;
- Lemma 3 provides an upper bound for the value of $\phi \doteq \|r'(\theta^{**})^\dagger\|$ in (7);
- Theorem 3 takes advantage of the previous lemmas to discuss the convergence properties of a Gauss-Newton technique, applied to pose graph optimization.

Let us start by recalling the expression of the residual error in the cost function (13):

$$r(\theta) = \begin{bmatrix} -w_\Delta D_2 R(\theta)\bar{\Delta}^l \\ w_\delta(A^\top\theta - \bar{\delta}) \end{bmatrix}. \tag{14}$$

The norm of the residual error at a minimum is called $\psi \doteq \|r(\theta^{**})\|$, as in equation (6). The Jacobian of $r(\theta)$ can be computed explicitly and written in compact form as:

$$r'(\theta) = \begin{bmatrix} \frac{\partial r_1}{\partial \theta_1} & \cdots & \frac{\partial r_1}{\partial \theta_n} \\ \cdots & \ddots & \vdots \\ \frac{\partial r_{3m}}{\partial \theta_1} & \cdots & \frac{\partial r_{3m}}{\partial \theta_n} \end{bmatrix} = \begin{bmatrix} -w_\Delta D_2 R(\theta + \pi/2)\bar{\Lambda}^l \\ w_\delta A^\top \end{bmatrix} \tag{15}$$

where $R(\theta + \pi/2) \in \mathbb{R}^{2m,2m}$ is a block diagonal matrix, whose nonzero entries are in positions $(2k-1, 2k-1), (2k-1, 2k), (2k, 2k-1), (2k, 2k)$, $k = 1, \ldots, m$, and the k-th diagonal block, corresponding to the k-th relative measurement, say (i, j), contains a rotation matrix of an angle $\theta_i + \pi/2$; the matrix $\bar{\Lambda}^l \in \mathbb{R}^{2m,n}$ has the following structure: the rows $2k-1$ and $2k$ correspond to the relative measurement $k = (i, j)$, and the entries on such rows are zero everywhere, except for the i-th column that contains $\bar{\Delta}_k^l$, i.e.

$$\bar{\Lambda}^l(2k-1 : 2k, i) = \begin{cases} \bar{\Delta}_k^l, & \text{if } k = (i, j) \\ \mathbf{0}_2, & \text{otherwise} \end{cases} \tag{16}$$

where $\bar{\Lambda}^l(2k-1:2k,i)$ denotes the two entries of $\bar{\Lambda}^l$ having row indices $[2k-1,2k]$ and column index i.

Now we notice that, according to [3], angular variables are treated as vectors in $\mathbb{R}^n$, since the regularization procedure allows to frame the estimation from the manifold SO(2) into a linear vector space. Therefore, we can consider the whole $\mathbb{R}^n$ as domain of the optimization variable θ. Accordingly, κ in definition (8) (radius of the largest open ball contained in the domain of the optimization variable) can be chosen arbitrarily large. The following result assures the satisfaction of Condition 1 in Theorem 1.

Lemma 1: The linear mapping $r'(\theta)$ in (15) is injective for any $\theta \in \mathbb{R}^n$.

Proof. We first remark that the lemma does not discuss the properties of the mapping $r'(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{3m,n}$, but it refers to the linear map $r'(\theta) : \mathbb{R}^n \rightarrow \mathbb{R}^{3m}$, for a given $\theta \in \mathbb{R}^n$. Then, we recall that a linear mapping is injective if it is left invertible [15]. Now it is easy to see that a left inverse of $r'(\theta)$ is $[\mathbf{0}_{n,2m} \quad \frac{1}{w_\delta}(AA^\top)^{-1}A]$, and this holds for any $\theta \in \mathbb{R}^n$. □

Condition 2.1 in Theorem 1 is, instead, less straightforward to assess, and it is discussed in the following lemma.

Lemma 2: For a generic node i, let us define $\mathbf{dist}^2_{\text{out,i}} = \sum_{j\in\mathcal{N}_{\text{out}}(i)} \|\bar{\Delta}^l_{ij}\|^2$ (sum of the squared distances from node i to all its outgoing neighbors). Call $\mathbf{dist}_{\text{out,max}} = \max_{i\in\{1,\dots,n\}} \mathbf{dist}_{\text{out,i}}$. Then, the Jacobian $r'(\cdot)$ in (15) satisfies the Lipschitz condition:

$$\|r'(\theta^a) - r'(\theta^b)\| \leq K\|\theta^a - \theta^b\|, \;\; \forall\, \theta^a, \theta^b \in \mathbb{R}^n$$

with $K = w_\Delta \mathbf{dist}_{\text{out,max}}$.

Proof. See Appendix I. □

The next result, Lemma 3, provides an upper bound for the quantity in (7), and it will be used for assessing Condition 2.2 and for stating Theorem 3.

Lemma 3: Let us define the quantity $\mathbf{dist}_{\text{out,max}}$ as in Lemma 2 and the pseudoinverse $(A^\top)^\dagger = (AA^\top)^{-1}A$; considering a problem instance whose input data satisfy the following condition:

$$w_\delta > w_\Delta \|(A^\top)^\dagger\| \mathbf{dist}_{\text{out,max}} \;\; , \tag{17}$$

then the following bound holds for the pseudoinverse of $r'(\theta)$:

$$\|r'(\theta)^\dagger\| \leq \frac{\|(A^\top)^\dagger\|}{w_\delta - w_\Delta\|(A^\top)^\dagger\|\mathbf{dist}_{\text{out,max}}} \;\; \forall \;\; \theta \in \mathbb{R}^n. \tag{18}$$

Proof. See Appendix II. □

Using all the previous results, we can finally state the following theorem.

Theorem 3: Let us consider the optimization problem (13) in the optimization variable θ. Let us call θ^{**} a solution of (13), define $\psi \doteq \|r(\theta^{**})\|$, and define $\mathbf{dist}_{\text{out,max}}$ as in Lemma 2. If the following inequalities are satisfied:

Condition 3.1: $\beta_1 \doteq \frac{w_\Delta}{w_\delta}\|(A^\top)^\dagger\|\mathbf{dist}_{\text{out,max}} < 1 \;\; ,$

Condition 3.2: $\beta_2 \doteq \sqrt{2}\psi\phi^2 K = \frac{\sqrt{2}\psi w_\Delta \|(A^\top)^\dagger\|^2 \mathbf{dist}_{\text{out,max}}}{\left[w_\delta - w_\Delta\|(A^\top)^\dagger\|\mathbf{dist}_{\text{out,max}}\right]^2} < 1 \;\; ,$

then, the Gauss-Newton method applied to problem (13) is well defined and converges to θ^{**}, starting from any initial guess in the ball $\mathcal{B}(\theta^{**},\gamma)$ with:

$$\gamma = \frac{2\left[w_\delta - w_\Delta\|(A^\top)^\dagger\|\mathbf{dist}_{\text{out,max}}\right]}{3w_\Delta\|(A^\top)^\dagger\|\mathbf{dist}_{\text{out,max}}} - \frac{2\sqrt{2}\psi\|(A^\top)^\dagger\|}{3[w_\delta - w_\Delta\|(A^\top)^\dagger\|\mathbf{dist}_{\text{out,max}}]} \;\; .$$

Moreover, if $\beta_2 < \frac{\sqrt{2}}{2}$, then the solution θ^{**} is unique in the ball $\mathcal{B}(\theta^{**}, 3\gamma)$.

Proof. The theorem is an application of Theorem 1, hence the proof can be directly inferred from [5] and from Lemmas 1-3. We note that conditions 3.1 and 3.2 correspond to (17) and Condition 2.2 in Theorem 1, respectively. The only claim that needs to be proven is the claim of uniqueness of the solution in $\mathcal{B}(\theta^{**}, 3\gamma)$. For this purpose we observe that, according to Theorem 1, the radius in which the solution is unique is (i) $\gamma_0 = \frac{2-2\psi\phi_0 K}{\phi K}$ (with $\phi_0 = \left\|[r'(\theta^{**})^\top r'(\theta^{**})]^{-1}\right\|$), as long as (ii) $2\psi\phi_0 K < 1$. Now we start by observing that $\phi_0 = \left\|[r'(\theta^{**})^\top r'(\theta^{**})]^{-1}\right\| = \frac{1}{\lambda_{min}\{[r'(\theta^{**})^\top r'(\theta^{**})]\}}$ (see "Notation and preliminaries" section). Moreover, $\phi = \left\|[r'(\theta^{**})]^\dagger\right\| = \frac{1}{\sqrt{\lambda_{min}\{[r'(\theta^{**})^\top r'(\theta^{**})]\}}}$ (see "Notation and preliminaries"). Then, $\phi_0 = \phi^2$ and condition (ii) becomes $2\psi\phi^2 K < 1$ which is equivalent to $\sqrt{2}\beta_2 < 1$, i.e., $\beta_2 < \frac{\sqrt{2}}{2}$. Finally we observe that, using the equality $\phi_0 = \phi^2$, the uniqueness radius γ_0 in (i) becomes $\gamma_0 = \frac{2-2\psi\phi_0 K}{\phi K} = \frac{2-2\psi\phi^2 K}{\phi K} \geq \frac{2-2\sqrt{2}\psi\phi^2 K}{\phi K} = 3\gamma$; If $\gamma_0 \geq 3\gamma$, then the solution, which from Theorem 1 is unique in $\mathcal{B}(\theta^{**}, \gamma_0)$, is also unique in $\mathcal{B}(\theta^{**}, 3\gamma)$. □

IV. Relevance in Practical Applications

We begin this section with some useful observations on the results of Theorem 3, summarized in the following remark.

Remark 1: For a fixed ψ, the convergence radius γ is monotonically increasing in w_δ, whereas it is monotonically decreasing in w_Δ, $\mathbf{dist}_{\text{out,max}}$, and $\|(A^\top)^\dagger\|$. Moreover, β_1 and β_2 in conditions 3.1 and 3.2 are monotonically decreasing in w_δ, and monotonically increasing in w_Δ, $\mathbf{dist}_{\text{out,max}}$, and $\|(A^\top)^\dagger\|$ (i.e., the desirable situation, in which conditions 3.1 and 3.2 hold and the convergence radius is large, is attained for large w_δ and small values of w_Δ, $\mathbf{dist}_{\text{out,max}}$, and $\|(A^\top)^\dagger\|$).

A. Influence of measurement uncertainty (w_δ, w_Δ)

Probably the only parameter that was well known to be crucial in the convergence properties of the pose graph optimization was the orientation uncertainty. Our analysis confirms this evidence: w_δ, by definition, is the orientation measurement information (inverse of the standard deviation of the orientation measurements). Now we want to point out that what truly matters is not w_δ itself, but the ratio $w = \frac{w_\delta}{w_\Delta}$; for this purpose we observe that it is easy to rewrite conditions 3.1 as function of w (i.e., without w_δ and w_Δ appearing explicitly). Defining $\bar{\psi} = \frac{\psi}{w_\Delta} = \sqrt{\|D_2 R(\theta)\bar{\Delta}^l\|^2 + \|w(A^\top\theta - \bar{\delta})\|^2}$, also condition 3.2 and the convergence radius γ can be rewritten in function of w. Therefore the convergence radius of the optimization problem will be large for large values of w (orientation measurement much more accurate than relative position measurements). Luckily, this is exactly the case for common pose graph optimization instances, where the orientation measurements are accurate if compared with the position

measurements. The result is also in agreement with the limit case in which the information of relative position measurements approaches zero. For $w_\Delta = 0$ the term associated with the Cartesian information in the objective function disappears and the problem becomes convex: consistently, our estimate of the radius of convergence tends to infinite (in convex problems there is no risk of being trapped in local minima).

Practical use. Beside the obvious suggestion of increasing the orientation accuracy for improving convergence, we now want to show that it is possible to use our result to devise *a-posteriori* solutions for enhancing convergence, i.e., techniques that, starting from the same input data, can improve convergence performance. Let us define a positive scalar $\tilde{w}_\Delta$. Let us rewrite the pose graph optimization problem (13), substituting w_Δ with $\tilde{w}_\Delta$:

$$f(\theta, \tilde{w}_\Delta) = \left\| \begin{bmatrix} -\tilde{w}_\Delta D_2 R(\theta) \bar{\Delta}^l \\ w_\delta (A^\top \theta - \bar{\delta}) \end{bmatrix} \right\|^2. \tag{19}$$

According to the considerations in this section, for (i) $\tilde{w}_\Delta = 0$ problem (19) is convex, whereas for (ii) $\tilde{w}_\Delta = w_\Delta$ we recover the original problem. Moreover, for $\tilde{w}_\Delta$ passing from the case (i) to the case (ii) the convergence radius decreases: therefore, we can apply the following approach for improving convergence in the problem at hand:

Algorithm 1 (Pose graph Continuation):

1: ***Input:*** A, $\bar{\delta}$, w_δ, $\bar{\Delta}^l$, w_Δ, number of outer iterations T, and an initial guess $\theta^{(-1)}$;
2: ***Output:*** Configuration estimate θ^*
3: ***for*** $t = \{0, \dots, T\}$
$\tilde{w}_\Delta = \frac{t}{T} w_\Delta$;
$\theta^{(t)} = \mathtt{GN}(f(\theta, \tilde{w}_\Delta), \theta^{(t-1)})$;
4: ***endfor***
5: $\theta^* = \theta^{(t)}$.

In Algorithm 1, $\mathtt{GN}(f(\theta, \tilde{w}_\Delta), \theta^{(t-1)})$ denotes the results of the application of a Gauss-Newton approach to the cost function $f(\theta, \tilde{w}_\Delta)$, parametrized in $\tilde{w}_\Delta$, starting from the initial guess $\theta^{(t-1)}$. The previous algorithm resembles *continuation methods*, that are often used in distance-based localization problems to mitigate the problem of local minima [16]. We notice that, for the case $\tilde{w}_\Delta = 0$, the problem is quadratic and the orientation solution can be computed in closed-form [3]:

$$\hat{\theta} = \arg\min f(\theta, \tilde{w}_\Delta = 0) = (AA^\top)^{-1} A \bar{\delta}, \tag{20}$$

then the initial guess, given to Algorithm 1, is irrelevant in practice. We notice that the continuation approach not necessarily requires a large number of outer iterations: the approach proposed in [3] corresponds to Algorithm 1 with $T = 1$, $\tilde{w}_\Delta^{(0)} = 0$ and $\tilde{w}_\Delta^{(1)} = w_\Delta$ (a single Gauss-Newton iterations is considered in the second outer iteration). Therefore, the presented results also explains the effectiveness of the approximation [3] from the perspective of convergence.

B. Influence of inter-nodal distances ($\boldsymbol{dist}_{out,max}$)

According to Theorem 3, small inter-nodal distances correspond to better convergence properties (i.e., larger convergence domains, as well as better chances to satisfy conditions 3.1 and 3.2). Also in this case, pose graph optimization problems are particularly well-behaved, in the sense that common problem instances are characterized by small relative position measurements, since the available measured poses are naturally limited by the sensor range.

Practical use. This section suggests a surprising practical result: scaling the relative position measurements by a scalar smaller than 1 enlarges the basin of convergence of a Gauss-Newton approach, applied to the problem at hand. As a consequence, if the original problem has "bad" convergence properties, we can solve the problem using $s\Delta^l$ (where, $s < 1$ is a scalar), instead of Δ^l, obtaining a scaled configuration of the pose graph, which topologically resembles the original configuration, although the Cartesian components are scaled by a factor s. This property may appear counter-intuitive at first sight, but it will be further confirmed by experimental results presented in Section V.

C. Influence of graph structure ($\|(A^\top)^\dagger\|$)

The study of the influence of the graph structure on the convergence domain is not straightforward and needs some more insight on $\|(A^\top)^\dagger\|$. We recall that, for a fixed ψ, the convergence radius increases when $\|(A^\top)^\dagger\|$ decreases (i.e., we want $\|(A^\top)^\dagger\|$ as small as possible). The following result allows to connect the norm of $(A^\top)^\dagger$ with the graph structure.

Theorem 4: If we call $\breve{\mathcal{G}}$ the graph obtained by adding edges to a connected graph $\mathcal{G}$, then, $\|(\breve{A}^\top)^\dagger\| \leq \|(A^\top)^\dagger\|$, where A and $\breve{A}$ denote the reduced incidence matrix of the original and of the augmented graph, respectively.

Proof. For proving the claim we notice that adding an edge to the graph means adding a row to the matrix $A^\top$. Let us call the augmented incidence matrix $\breve{A}^\top = [A \; a]^\top$, where we added the rows $a^\top$ to $A^\top$. Then, the theorem claims that (i) $\|(\breve{A}^\top)^\dagger\| \leq \|(A^\top)^\dagger\|$. Assume by contradiction that (ii) $\|(\breve{A}^\top)^\dagger\| > \|(A^\top)^\dagger\|$. The contradiction assumption (ii) implies that $\lambda_{min}\{\breve{A}\breve{A}^\top\} < \lambda_{min}\{AA^\top\}$ (see "Notation and preliminaries" section and recall that, for a connected graph, $A^\top$ is full column rank [3]). According to the definition of $\breve{A}^\top$, we obtain $\breve{A}\breve{A}^\top = [A \; a][A \; a]^\top = AA^\top + aa^\top$. Now, for the Weyl inequality, we have that $\lambda_{min}\{AA^\top + aa^\top\} \geq \lambda_{min}\{AA^\top\} + \lambda_{min}\{aa^\top\}$. Since, $aa^\top$ is positive semi-definite by construction, all its eigenvalues are non-negative; then $\lambda_{min}\{\breve{A}\breve{A}^\top\} \geq \lambda_{min}\{AA^\top\} + \lambda_{min}\{aa^\top\} \geq \lambda_{min}\{AA^\top\}$, but this contradicts (ii), proving the claim. □

Practical use. The result of this section guarantees that, given a pose graph configuration, it is possible to reduce the quantity $\|(A^\top)^\dagger\|$, by adding edges to the pose graph. Further insight is provided by the following corollary.

Corollary 1: The quantity $\min_i \lambda_i\{L\} \doteq \min_i \lambda_i\{AA^\top\}$ is a measure of connectivity: it is zero when the graph $\mathcal{G}$ is not connected and it does not decrease when adding edges to $\mathcal{G}$.

Proof. We already demonstrated that, for a connected graph, $\|(A^\top)^\dagger\|$ is non increasing when adding edges, then $\lambda_{min}\{AA^\top\} = \frac{1}{\|(A^\top)^\dagger\|^2}$ is non decreasing when adding edges. It remains to demonstrate that $\min_i \lambda_i\{AA^\top\} = 0$ when the graph $\mathcal{G}$ is not connected. For this purpose we recall that the number of zero eigenvalues of the Laplacian matrix is equal to the number of connected components of the graph [4]: if $\mathcal{G}$ is not connected, it contains at least two connected components, then the two smallest eigenvalues of $\mathcal{L}$ are $\lambda_1\{\mathcal{A}\mathcal{A}^\top\} = \lambda_2\{\mathcal{A}\mathcal{A}^\top\} = 0$. For the Cauchy interlacing theorem, the smallest eigenvalues of the reduced Laplacian

$L = AA^\top$, namely $\min_i \lambda_i\{AA^\top\}$, satisfies $\lambda_1\{\mathcal{A}\mathcal{A}^\top\} \leq \min_i \lambda_i\{AA^\top\} \leq \lambda_2\{\mathcal{A}\mathcal{A}^\top\}$, then it has to be equal to zero, proving the claim. □

In principle this suggests that the robot, in full autonomy, can improve the convergence properties, simply acquiring more measurements. However, we have to notice that adding edges may also increase the quantity $\mathbf{dist}_{\text{out,max}}$ (Section IV-B).

V. Numerical Experiments

In this section we show that empirical evidence confirms the convergence results presented in the theoretical derivation. We study the convergence properties when (i) varying $\omega = \omega_\delta/\omega_\Delta$, (ii) scaling the Cartesian measurements by a factor s, (iii) reducing graph connectivity (by choosing a suitable incidence matrix $\breve{A}$). In particular, for the convergence study, we consider different radii $Q = \{10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4\}$. For each radius Q_i, we generate 500 random orientation guesses $\theta^{(0)}$ on the hypersphere $\|\theta^{(0)} - \theta^{**}\| = Q_i$ and, for each guess, we run a standard Gauss-Newton approach, starting from $\theta^{(0)}$. The global optimum θ^{**} is retrieved in a preliminary controlled experiment. After running the tests, we record the percentage of experiments in which the value of the attained objective is equal (up to numerical approximations) to the objective value at θ^{**}. Notice that we do not evaluate convergence on the basis of the proximity to θ^{**} in order not to rule out the possibility of different equivalent global optima. For sake of repeatability, we consider a standard SLAM dataset, the Intel Research Lab in Seattle, using the relative orientation and position measurements available online [13]. The results are summarized in Figures 1-3. One can imagine to consider a parametrized optimization problem $f(\theta, \omega, s, \breve{A})$: each test evaluates how the percentage of converging experiments changes in function of the parameters ω, s, and $\breve{A}$. We also report the distance of the estimate $\hat{\theta}$, see equation (20), from the global optimum. In Figure 1, we report the percentage of converging tests for $\omega = \{0.5, 1, 10^2\}$ with $s = 1$ (no scaling), and considering the original incidence matrix ($\breve{A} = A$). For smaller values of ω, the problem of local minima becomes so severe that also in the controlled experiment is challenging to obtain a global solution.

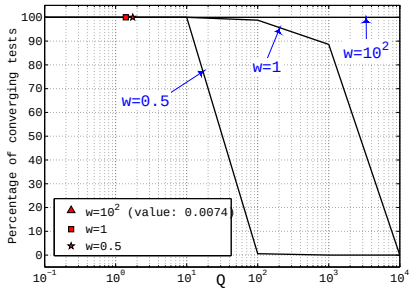

Fig. 1. Percentage of converging tests for different values of ω. We kept constant $s = 1$ and $\breve{A} = A$, and the different curves correspond to $\omega = \{0.5, 1, 10^2\}$. The markers denote the distance of $\hat{\theta}$ from θ^{**}, for different values of ω. For the case $\omega = 10^2$ the marker is not visible in the plot since $\|\hat{\theta} - \theta^{**}\| = 0.0074$.

Figure 2 considers the case in which s varies in the set $s = \{0.1, 0.5, 1, 5\}$, with $\omega = 1$ and $\breve{A} = A$. The scaling s, as the parameter ω, has a large influence on the convergence properties: this confirms the results described in Section IV-B; moreover, this result rises a warning for the cases in which pose graph optimization in used for large scale mapping. Figure 3 reports the convergence results for $\omega = 1$ and $s = 1$, for the case in which $\breve{A}$ is obtained from A by removing 100 and 200 edges, respectively. The figure shows that the convergence properties are less sensitive to changes in the incidence matrix; however, dropping edges may remarkably worsen the quality of the estimate $\hat{\theta}$. We can see that, in most tests, the estimate $\hat{\theta}$ falls within the (experimental) convergence basin. This means that the continuation approach of Section IV-A is able to assure global convergence, with only two outer iterations T.

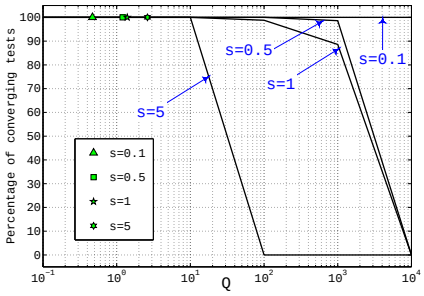

Fig. 2. Percentage of converging tests for different values of s. We kept constant $w = 1$ and $\breve{A} = A$, and the curves correspond to $s = \{0.1, 0.5, 1, 5\}$. The markers denote the distance of $\hat{\theta}$ from θ^{**}, for different values of s.

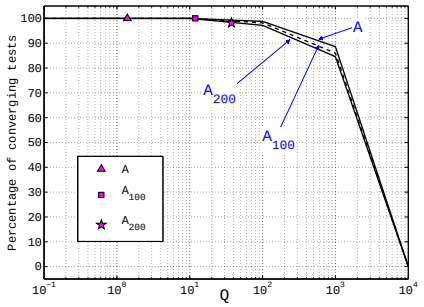

Fig. 3. Percentage of converging tests for different incidence matrices. We kept constant $w = 1$ and $s = 1$, and we considered the original incidence matrix A, and two incidence matrices obtained by removing 100 and 200 edges from A, respectively. The latter are called A_{100} and A_{200}. The curve corresponding to A_{100} is plotted as a dashed line for clarity. The markers denote the distance of $\hat{\theta}$ from θ^{**}.

	$s = 1$	$s = 0.1$	$s = 0.01$
$\omega = 10^4$	$\beta_1 = 1.2 \cdot 10^{-1}$	$\beta_1 = 1.2 \cdot 10^{-2}$	$\beta_1 = 1.2 \cdot 10^{-3}$
	$\beta_2 = 15.5$	$\beta_2 = 1.2$	$\beta_2 = 1.2 \cdot 10^{-1}$
	$\gamma = -$	$\gamma = -$	$\gamma = 495.6$
$\omega = 10^5$	$\beta_1 = 1.2 \cdot 10^{-2}$	$\beta_1 = 1.2 \cdot 10^{-3}$	$\beta_1 = 1.2 \cdot 10^{-4}$
	$\beta_2 = 1.2$	$\beta_2 = 1.2 \cdot 10^{-1}$	$\beta_2 = 1.2 \cdot 10^{-2}$
	$\gamma = -$	$\gamma = 495$	$\gamma = 5577$

TABLE I

Conditions 3.1 and 3.2 and convergence radius γ for different problem instances ($\breve{A} = A$).

Table I shows the values of β_1 and β_2 (conditions 3.1 and 3.2) and the convergence radius γ (when the conditions are satisfied) for different problem instances. It is worth noticing that the estimates are conservative (experimentally, the convergence radius is already bigger than 10^4 for $\omega = 10^2$ and $s = 1$, see Figure 1). For concluding the experimental section we show that, for the dataset under analysis, the convergence properties are likely to be worst when performing the optimization on the manifold SO(2) (i.e., normalizing the angular information in $[-\pi, \pi[$ at each iteration). In Figure 4 we show the convergence results for a problem instance with $\omega = 1$, $s = 1$, and $\breve{A} = A$, comparing a Gauss-Newton approach on manifold with the approach discussed in this paper.

VI. Conclusion

In this work we show that, when using a Gauss-Newton approach, the convergence properties of pose graph optimization are influenced by three elements: ratio between information content of relative orientation and relative position measurements, inter-nodal distances, and structure of the underlying graph. Moreover, we provide an estimate of the convergence domain and of the region in which the minimum is unique. The outcome of our study provides a theoretical motivation for the empirical evidence that common pose-based SLAM instances are well-behaved in terms of convergence. We also discuss possible ways to improve convergence during graph design or a-posteriori, given the data. Future work includes finding tighter convergence and uniqueness bounds and extending the experimental evaluation presented in this paper.

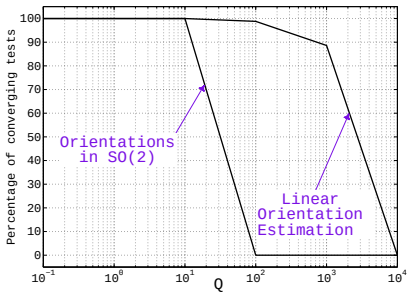

Fig. 4. Percentage of converging tests for a problem instance with $\omega = 1$, $s = 1$, $\breve{A} = A$, comparing a Gauss-Newton approach on manifold with the approach discussed in this paper.

Appendix I

In this appendix we demonstrate that Lemma 2 holds, i.e., that there exists a K such that the Lipschitz condition $\|r'(\theta^a) - r'(\theta^b)\| \leq K\|\theta^a - \theta^b\|$ is verified for any $\theta^a, \theta^b \in \mathbb{R}^n$. Our objective is to suitably bound the following quantity:

$$\|r'(\theta^a) - r'(\theta^b)\| = \left\| \begin{bmatrix} -w_\Delta D_2[R(\theta^a+\pi/2) - R(\theta^b+\pi/2)]\bar{\Lambda}^l \\ \mathbf{0}_{m,n} \end{bmatrix} \right\| =$$
$$= \|w_\Delta D_2[R(\theta^a + \pi/2) - R(\theta^b + \pi/2)]\bar{\Lambda}^l\| \leq$$
$$\leq \|w_\Delta\|\|D_2\|\|[R(\theta^a + \pi/2) - R(\theta^b + \pi/2)]\bar{\Lambda}^l\|.$$

According to Theorem 2, D_2 is a projection matrix and satisfies $\|D_2\| = 1$; moreover, w_Δ is a positive scalar by definition. Therefore the previous inequality becomes:

$$\|r'(\theta^a) - r'(\theta^b)\| \leq w_\Delta\|[R(\theta^a + \pi/2) - R(\theta^b + \pi/2)]\bar{\Lambda}^l\|. \tag{21}$$

Let us call $\bar{M} = [R(\theta^a + \pi/2) - R(\theta^b + \pi/2)]\bar{\Lambda}^l \in \mathbb{R}^{2m,n}$. We now need to compute $\|[R(\theta^a+\pi/2) - R(\theta^b+\pi/2)]\bar{\Lambda}^l\| = \|\bar{M}\| = \sqrt{\lambda_{max}\{\bar{M}^\top \bar{M}\}}$. Recalling the structure of $\bar{\Lambda}^l$, (16), and $R(\cdot)$ we notice that $\bar{M}$ has the following structure:

$$\bar{M}(2k-1:2k, i) = \begin{cases} (R_{\theta_i^a+\pi/2} - R_{\theta_i^b+\pi/2})\bar{\Delta}_k^l, & \text{if } k = (i,j) \\ \mathbf{0}_2 & \text{otherwise} \end{cases}, \tag{22}$$

where $\bar{M}(2k-1:2k, i)$ denotes the two entries of $\bar{M}$ having row indices $[2k-1, 2k]$ and column index i, and where $R_{\theta_i^a+\pi/2}$ and $R_{\theta_i^b+\pi/2}$ are 2 by 2 planar rotation matrices of angles $\theta_i^a + \pi/2$ and $\theta_i^b + \pi/2$, respectively. Now, if we call $\bar{N} \doteq \bar{M}^\top \bar{M}$, we can write the generic element $\bar{N}(i,j)$, with $i, j = 1, 2, \ldots, n$, as $\bar{N}(i,j) = \sum_{k=1}^{m} \bar{M}(2k-1:2k, i)^\top \bar{M}(2k-1:2k, j)$. However, according to the structure of $\bar{M}$, (22), the only case in which the previous sum is different from zero is the case $i = j$; therefore $\bar{N} = \bar{M}^\top \bar{M}$ is a diagonal matrix with diagonal entries

$$\bar{N}(i,i) = \sum_{j \in \mathcal{N}_{\text{out}}(i)} \|(R_{\theta_i^a+\pi/2} - R_{\theta_i^b+\pi/2})\bar{\Delta}_{ij}^l\|^2, \quad i = 1, 2, \ldots, n. \tag{23}$$

It is easy to see that a generic term in the sum (23) satisfies the following Lipschitz condition: $\|(R_{\theta_i^a+\pi/2} - R_{\theta_i^b+\pi/2})\bar{\Delta}_{ij}^l\|^2 \leq \|\bar{\Delta}_{ij}^l\|^2\|\theta_i^a - \theta_i^b\|^2$ (this can be seen, for instance, by computing the norm of the Jacobian of $R_{\theta_i+\pi/2}\bar{\Delta}_{ij}^l$ with respect to θ_i, and observing that it is bounded, with upper bound $\|\bar{\Delta}_{ij}^l\|$). Therefore the generic element $\bar{N}(i,i)$ satisfies $\bar{N}(i,i) \leq \sum_{j \in \mathcal{N}_{\text{out}}(i)} \|\bar{\Delta}_{ij}^l\|^2\|\theta_i^a - \theta_i^b\|^2$. Defining $\mathbf{dist}_{\text{out,max}}^2 = \max_{i \in \{1,\ldots,n\}} \sum_{j \in \mathcal{N}_{\text{out}}(i)} \|\bar{\Delta}_{ij}^l\|^2$, we obtain the following bound: $\bar{N}(i,i) \leq \mathbf{dist}_{\text{out,max}}^2 \|\theta_i^a - \theta_i^b\|^2$.

Since $\bar{N}$ is diagonal, the eigenvalues are simply the diagonal elements and the maximum eigenvalue is simply the infinity norm of the vector that composes the diagonal of $\bar{N}$, i.e.,

$$\lambda_{\max}\{\bar{N}\} = \left\| \begin{bmatrix} \bar{N}(1,1) \\ \bar{N}(2,2) \\ \vdots \\ \bar{N}(n,n) \end{bmatrix} \right\|_\infty.$$

Therefore the norm of $\bar{M}$ is:

$$\|\bar{M}\| = \sqrt{\lambda_{\max}\{\bar{N}\}} = \sqrt{\left\| \begin{bmatrix} \bar{N}(1,1) \\ \bar{N}(2,2) \\ \vdots \\ \bar{N}(n,n) \end{bmatrix} \right\|_\infty} \overset{(i)}{=} \left\| \begin{bmatrix} \sqrt{\bar{N}(1,1)} \\ \sqrt{\bar{N}(2,2)} \\ \vdots \\ \sqrt{\bar{N}(n,n)} \end{bmatrix} \right\|_\infty \leq$$
$$\overset{(ii)}{\leq} \left\| \begin{bmatrix} \sqrt{\bar{N}(1,1)} \\ \sqrt{\bar{N}(2,2)} \\ \vdots \\ \sqrt{\bar{N}(n,n)} \end{bmatrix} \right\| \overset{(iii)}{\leq} \mathbf{dist}_{\text{out,max}} \left\| \begin{bmatrix} \|\theta_1^a - \theta_1^b\| \\ \|\theta_2^a - \theta_2^b\| \\ \vdots \\ \|\theta_n^a - \theta_n^b\| \end{bmatrix} \right\| =$$
$$\overset{(iv)}{=} \mathbf{dist}_{\text{out,max}} \|\theta^a - \theta^b\|,$$

where the relations are explained as follows: in (i) the square root of the largest element of the vector is substituted with the largest square root (all the elements are nonnegative by definition); (ii) is a standard vector-norm inequality between the infinity and the Euclidean norm; (iii) is based on the upper bound on $\bar{N}(i,i)$ that we found few lines before; finally, equality (iv) is simply obtained applying the definition of the Euclidean vector norm. Summarizing, we obtained the following bound:

$$\|[R(\theta^a + \pi/2) - R(\theta^b + \pi/2)]\bar{\Lambda}^l\| \leq \mathbf{dist}_{\text{out,max}} \|\theta^a - \theta^b\|.$$

Restarting from equation (21) we can finally conclude:

$$\|r'(\theta^a) - r'(\theta^b)\| \leq w_\Delta \mathbf{dist}_{\text{out,max}} \|\theta^a - \theta^b\|,$$

proving Lemma 2, with $K = w_\Delta \mathbf{dist}_{\text{out,max}}$. □

APPENDIX II

In this appendix we prove Lemma 3, providing an upper bound for the quantity $\|r'(\theta)^\dagger\|$. According to [19], given two full column rank matrices V and V_2, such that there exists a third matrix V_1, with $V = V_1 + V_2$ and $\|V_1\|\|V_2^\dagger\| \leq 1$, then:

$$\|V^\dagger\| \leq \frac{\|V_2^\dagger\|}{1 - \|V_1\|\|V_2^\dagger\|} \tag{24}$$

We are now going to apply this result with:

$$V \doteq r'(\theta) = \begin{bmatrix} -w_\Delta D_2 R(\theta + \pi/2)\bar{\Lambda}^l \\ w_\delta A^\top \end{bmatrix}$$

$$V_1 \doteq \begin{bmatrix} -w_\Delta D_2 R(\theta + \pi/2)\bar{\Lambda}^l \\ \mathbf{0}_{m,n} \end{bmatrix} \quad V_2 \doteq \begin{bmatrix} \mathbf{0}_{2m,n} \\ w_\delta A^\top \end{bmatrix}$$

Let us compute $\|V_2^\dagger\|$ and $\|V_1\|$. The former norm is $\|V_2^\dagger\| = \|(w_\delta A^\top)^\dagger\| = \frac{1}{w_\delta}\|(A^\top)^\dagger\|$ Regarding $\|V_1\|$, we can see that:

$$\|V_1\| = \left\| \begin{bmatrix} -w_\Delta D_2 R(\theta + \pi/2)\bar{\Lambda}^l \\ \mathbf{0}_{m,n} \end{bmatrix} \right\| =$$

$$= \left\|w_\Delta D_2 R(\theta + \pi/2)\bar{\Lambda}^l\right\| \leq \|w_\Delta\|\|D_2\|\|R(\theta + \pi/2)\bar{\Lambda}^l\| \overset{(i)}{\leq}$$

$$\leq w_\Delta \|R(\theta + \pi/2)\bar{\Lambda}^l\|$$

where the inequality (i) uses the fact that $D_2 \doteq C_2^\top (C_2 C_2^\top)^{-1} C_2$ has unit norm (Theorem 2). Moreover, repeating the considerations in Appendix I we notice that, due to the structure of matrix $\bar{M} \doteq R(\theta + \pi/2)\bar{\Lambda}^l$, $\bar{N} \doteq \bar{M}^\top \bar{M}$ is a diagonal matrix, whose eigenvalues are on the main diagonal. Furthermore, the diagonal terms are in the form $\bar{N}(i,i) = \sum_{j \in \mathcal{N}_{\text{out}}(i)} \|\bar{\Delta}_{ij}^l\|^2$, $i = 1, \ldots, n$. Therefore, $\|\bar{M}\| = \sqrt{\lambda_{\max}\{\bar{N}\}} = \sqrt{\max_{i \in \{1,\ldots,n\}} \bar{N}(i,i)} = \sqrt{\max_{i \in \{1,\ldots,n\}} \sum_{j \in \mathcal{N}_{\text{out}}(i)} \|\bar{\Delta}_{ij}^l\|^2}$. We already defined the quantity $\mathbf{dist}_{\text{out,max}}^2 \doteq \max_{i \in \{1,\ldots,n\}} \sum_{j \in \mathcal{N}_{\text{out}}(i)} \|\bar{\Delta}_{ij}^l\|^2$ (see Lemma 2), therefore:

$$\|V_1\| \leq w_\Delta \mathbf{dist}_{\text{out,max}}$$

Finally, applying (24), we get that as long as:

$$\|V_1\|\|V_2^\dagger\| \leq \frac{w_\Delta}{w_\delta}\|(A^\top)^\dagger\|\mathbf{dist}_{\text{out,max}} < 1 \tag{25}$$

the following inequality holds:

$$\|V^\dagger\| = \|r'(\theta)^\dagger\| \leq \frac{\|V_2^\dagger\|}{1 - \|V_1\|\|V_2^\dagger\|} \leq$$
$$\leq \frac{\|(A^\top)^\dagger\|}{w_\delta - w_\Delta\|(A^\top)^\dagger\|\mathbf{dist}_{\text{out,max}}} . \tag{26}$$

It is now easy to see that conditions (25) and (26) correspond to (17) and (18), respectively, hence proving the thesis.

□

REFERENCES

[1] R. Aragues, L. Carlone, G.C. Calafiore, and C. Sagues. Multi-agent localization form noisy relative pose measurement. In *Proc. of the IEEE lnternational Conf. on Robotics and Automation*, pages 364–369, 2011.

[2] K. Been, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller. Multiple relative pose graphs for robust cooperative mapping. In *Proc. of the IEEE lnternational Conf. on Robotics and Automation*, pages 3185–3192, 2011.

[3] L. Carlone, R. Aragues, J. Castellanos, and B. Bona. A linear approximation for graph-based simultaneous localization and mapping. In *Proc. of Robotics: Science and Systems, in press*, 2011.

[4] F.R.K. Chung. *Spectral Graph Theory*. Providence, RI: Amer. Math. Soc., 1997.

[5] O.P. Ferreira, M.L.N. Gonçalves, and P.R. Oliveira. Local convergence analysis of the Gauss-Newton method under a majorant condition. *Journal of Complexity*, 27(1):111–125, 2011.

[6] U. Frese, P. Larsson, and T. Duckett. A multilevel relaxation algorithm for simultaneous localization and mapping. *IEEE Trans. on Robotics*, 21(2):196–207, 2005.

[7] G. Grisetti, R. Kuemmerle, C. Stachniss, U. Frese, and C. Hertzberg. A hierarchical optimization on manifolds for online 2D and 3D mapping. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2010.

[8] G. Grisetti, C. Stachniss, and W. Burgard. Non-linear constraint network optimization for efficient map learning. *IEEE Trans. on Intelligent Transportation Systems*, 10(3):428–439, 2009.

[9] S. Huang, Y. Lai, U. Frese, and G. Dissanayake. How far is SLAM from a linear least squares problem? In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2010.

[10] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert. iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2011.

[11] T. Kavitha, C. Liebchen, K. Mehlhorn, D. Michail, R. Rizzi, T. Ueckerdt, and K. Zweig. Cycle bases in graphs: Characterization, algorithms, complexity, and applications. *Computer Science Rev.*, 3(4):199–243, 2009.

[12] K. Konolige. Large-scale map-making. In *Proc. of the AAAI National Conf. on Artificial Intelligence*, 2004.

[13] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, and A. Kleiner. Slam benchmarking webpage, 2009.

[14] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4:333–349, 1997.

[15] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.

[16] J.J. Moré and Z. Wu. Global continuation for distance geometry problems. *SIAM J. on Optimization*, 7(3):814–836, 1997.

[17] E. Olson, J.J. Leonard, and S. Teller. Fast iterative optimization of pose graphs with poor initial estimates. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 2262–2269, 2006.

[18] D. Lodi Rizzini. A closed-form constraint networks solver for maximum likelihood mapping. In *Proc. of European Conference on Mobile Robots*, 2010.

[19] G.W. Stewart. On the continuity of the generalized inverse. *SIAM J. Appl. Math.*, 17(1):33–45, 1969.

[20] S. Thrun and M. Montemerlo. The GraphSLAM algorithm with applications to large-scale mapping of urban structures. *Int. J. Robot. Res.*, 25:403–429, 2006.