Lagrangian Duality in 3D SLAM:
Verification Techniques and Optimal Solutions

Luca Carlone,   David M. Rosen,   Giuseppe Calafiore,   John J. Leonard,   Frank Dellaert

Abstract—State-of-the-art techniques for simultaneous localization and mapping (SLAM) employ iterative nonlinear optimization methods to compute an estimate for robot poses. While these techniques often work well in practice, they do not provide guarantees on the quality of the estimate. This paper shows that *Lagrangian duality* is a powerful tool to assess the quality of a given candidate solution. Our contribution is threefold. First, we discuss a revised formulation of the SLAM inference problem. We show that this formulation is probabilistically grounded and has the advantage of leading to an optimization problem with quadratic objective. The second contribution is the derivation of the corresponding *Lagrangian dual problem*. The SLAM dual problem is a (convex) *semidefinite program*, which can be solved reliably and globally by off-the-shelf solvers. The third contribution is to discuss the relation between the original SLAM problem and its dual. We show that from the dual problem, one can evaluate the quality (i.e., the suboptimality gap) of a candidate SLAM solution, and ultimately provide a certificate of optimality. Moreover, when the duality gap is zero, one can compute a guaranteed optimal SLAM solution from the dual problem, circumventing non-convex optimization. We present extensive (real and simulated) experiments supporting our claims and discuss practical relevance and open problems.

I. INTRODUCTION

Simultaneous localization and mapping (SLAM) is an enabling technology for many applications, including service and industrial robotics, autonomous driving, search and rescue, planetary exploration, and augmented reality.

The last decade has witnessed several groundbreaking results in SLAM, and state-of-the-art approaches are now transitioning from academic research to industrial applications. Standard techniques compute an estimate (e.g., for robot poses) by minimizing a nonlinear cost function, whose global minimum is the *maximum likelihood estimate* (or *maximum a posteriori* estimate in presence of priors). The optimization problem underlying SLAM is commonly solved using iterative nonlinear optimization methods, e.g. the Gauss-Newton method [1], [2], [3], the gradient method [4], [5], trust region methods [6], or ad-hoc approximations [7], [8].

Despite the success of state-of-the-art techniques, some practical and theoretical problems remain open. While iterative approaches are observed to work well in many problem instances, they cannot guarantee the correctness (global optimality) of the estimates that they compute. This is due to the fact that the optimization problem is non-convex, hence iterative

L. Carlone and F. Dellaert are with the Georgia Institute of Technology, Atlanta, GA, USA, luca.carlone@gatech.edu, frank@cc.gatech.edu.

D.M. Rosen and J.J. Leonard are with the Massachusetts Institute of Technology, Cambridge, MA, USA {dmrosen,jleonard}@mit.edu.

G. Calafiore is with the Politecnico di Torino, Torino, Italy, giuseppe.calafiore@polito.it.
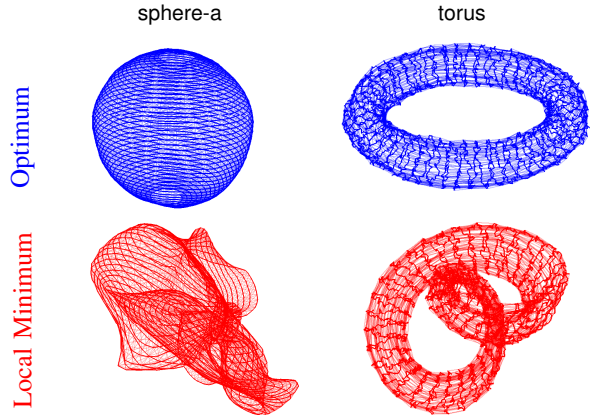
Fig. 1.   This paper exploits Lagrangian duality to evaluate the quality of a SLAM solution. For instance, our verification techniques are able to certify the optimality of the solutions in the first row of the figure, while they reject as suboptimal the estimates shown in the bottom row.

techniques may be trapped in local minima[1] (Fig. 1), which correspond to wrong estimates. Recent work [9], [10] shows that iterative techniques fail to converge to a correct estimate even in fairly simple (real and simulated) 3D problems. Recent research efforts have addressed the issue of global convergence from several angles. Olson *et al.* [4], Grisetti *et al.* [5], Rosen *et al.* [6], and Tron *et al.* [11] study iterative techniques with larger basins of convergence. Carlone *et al.* [8], [12], and Rosen *et al.* [10] propose initialization techniques to bootstrap iterative optimization. Huang *et al.* [13], Wang *et al.* [14], Carlone [15], and Khosoussi *et al.* [16] investigate the factors influencing local convergence and the quality of the optimal SLAM solution. While these techniques provide remarkable insights into the problem, and working solutions to improve convergence, none of them can guarantee the recovery of a globally optimal solution to SLAM.

The motivation behind this work is that the transition of SLAM from research topic to industrial technology requires techniques with *guaranteed performance*. In autonomous vehicle applications, failure to produce a correct SLAM solution may put passengers' lives at risk. In other applications, SLAM failures can possibly cascade into path planning failures (if the plan is computed using a wrong map), and this may prevent the reliable operation of mobile robots.

Therefore, in this paper we address the following question:
Verification Problem: *Given a candidate SLAM estimate (e.g., returned by a state-of-the-art iterative solver), is it possible to evaluate the quality of this estimate (e.g., its suboptimality gap), possibly certifying its optimality?*

To address this problem, we introduce a powerful tool,

---

[1]We use the term "local minimum" to denote a stationary point of the cost which does not attain the optimal objective.

Lagrangian duality, borrowing the corresponding theory from the optimization community. Duality was first applied to 2D SLAM by Carlone and Dellaert [9]. In this work, we provide a nontrivial extension of [9] to 3D SLAM.

This paper contains three main contributions. The first contribution is a revised formulation of the SLAM inference problem. Our formulation is a probabilistically grounded maximum likelihood (ML) estimator, and has the advantage of leading to an optimization problem with quadratic objective, which facilitates the derivation of the dual problem. Our revised SLAM formulation is presented in Section II.

The second contribution is the derivation of the *Lagrangian dual problem*. A key step towards this goal consists of rewriting the 3D SLAM problem as a quadratic optimization problem with quadratic equality constraints. Intuitively, the constraints impose that the pose estimates are members of SE(3), while the objective minimizes the mismatch w.r.t. the measurements. The dual SLAM problem, introduced in Section III, is a (convex) *semidefinite program* (SDP), and can be solved globally by off-the-shelf solvers.

The third contribution is to provide *verification techniques* to assess the quality of a given SLAM solution, leveraging the relation between the standard SLAM problem and its dual. We show that solving the dual problem allows us to bound the sub-optimality gap of a given candidate solution, hence we are able to quantify how far the candidate solution is from being optimal. Since current SDP solvers do not scale well to large problems, we also propose a second verification technique that does not require solving the SDP. As a by-product of our derivation, we show that, when the duality gap is zero, we can compute an optimal SLAM solution directly from the dual problem. Our verification techniques, presented in Section IV, can be seamlessly integrated in standard SLAM pipelines. Experimental evidence (Section V) confirms that these techniques enable the certification of globally optimal solutions in both real and simulated experiments. Extra results and visualizations are given in the supplemental material [17].

II. 3D Pose Graph Optimization Revisited

We consider the *pose graph optimization* (PGO) formulation of the SLAM problem. PGO computes the *maximum likelihood* estimate for n poses $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, given m relative pose measurements $\bar{\boldsymbol{x}}_{ij}$ between pairs of poses i and j. In a 3D setup, both the unknown poses and the measurements are quantities in $\mathrm{SE}(3) \doteq \{(\boldsymbol{R}, \boldsymbol{t}) : \boldsymbol{R} \in \mathrm{SO}(3), \boldsymbol{t} \in \mathbb{R}^3\}$. We use the notation $\boldsymbol{x}_i = (\boldsymbol{R}_i, \boldsymbol{t}_i)$ and $\bar{\boldsymbol{x}}_{ij} = (\bar{\boldsymbol{R}}_{ij}, \bar{\boldsymbol{t}}_{ij})$ to make explicit the rotation and the translation of each pose. PGO can be visualized as a directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, in which we associate a node $i \in \mathcal{V} = \{1, \ldots, n\}$ to each pose $\boldsymbol{x}_i$ and an edge $(i, j) \in \mathcal{E}$ to each relative measurement $\bar{\boldsymbol{x}}_{ij}$.

In this section we propose a revised PGO formulation. The key difference w.r.t. related work is the use of the *chordal distance* to quantify the rotation errors (more details in Section II-B). To lay the groundwork for this formulation, we begin with a generative model for our measurements, and then derive the corresponding ML estimator.

A. Generative Noise Model

We assume the following generative model for the relative pose measurements $(\bar{\boldsymbol{R}}_{ij}, \bar{\boldsymbol{t}}_{ij})^2$:

$$
\begin{aligned}
\bar{\boldsymbol{t}}_{ij} &= \boldsymbol{R}_i^\mathsf{T}(\boldsymbol{t}_j - \boldsymbol{t}_i) + \boldsymbol{t}_\epsilon & \boldsymbol{t}_\epsilon &\sim \text{gaussian}(\boldsymbol{0}_3, \omega_t^2 \mathbf{I}_3) \\
\bar{\boldsymbol{R}}_{ij} &= \boldsymbol{R}_i^\mathsf{T} \boldsymbol{R}_j \boldsymbol{R}_\epsilon & \boldsymbol{R}_\epsilon &\sim \text{vonMises}(\mathbf{I}_3, \omega_R^2)
\end{aligned} \tag{1}
$$

where "gaussian$(\boldsymbol{\mu}, \boldsymbol{\Omega})$" denotes a Gaussian distribution with mean $\boldsymbol{\mu}$ and *information* matrix $\boldsymbol{\Omega}$, while "vonMises$(\boldsymbol{S}, \kappa)$" denotes the isotropic von Mises-Fisher distribution, with mean $\boldsymbol{S} \in \mathrm{SO}(3)$ and *concentration parameter* κ. The key difference w.r.t. to measurement models in other PGO formulations lies in the use of the von Mises-Fisher distribution as the model for the rotational measurements errors $\boldsymbol{R}_\epsilon \in \mathrm{SO}(3)$.

The isotropic *von Mises-Fisher* (or *Langevin*) [18] distribution on $\mathrm{SO}(n)$ with mean $\boldsymbol{S} \in \mathrm{SO}(n)$ and concentration parameter $\kappa \geq 0$ can be written explicitly as:

$$
\mathbb{P}(\boldsymbol{R}_\epsilon) = \frac{1}{c_n(\kappa)} \exp\left(\kappa \operatorname{tr}\left(\boldsymbol{S}^\mathsf{T} \boldsymbol{R}_\epsilon\right)\right), \tag{2}
$$

where $\operatorname{tr}(\cdot)$ is the matrix trace and $c_n(\kappa)$ is a normalization term. Closed-form expressions for $c_n(\kappa)$ are given in [18]; these are inconsequential for our derivation. For $\kappa \to 0$, the distribution tends to the uniform distribution over $\mathrm{SO}(n)$. For $\kappa \to \infty$, $\boldsymbol{R}_\epsilon = \boldsymbol{S}$ with probability one. Roughly speaking, one may think at κ in terms of *information* content.

We are now ready to introduce the maximum likelihood estimator for the poses, given the measurement model (1).

B. Maximum Likelihood Estimator

The ML estimate corresponds to the set of poses maximizing the likelihood of the measurements, or, equivalently, minimizing the negative log-likelihood:

$$
f_{\mathrm{ML}}^\star = \min_{\{\boldsymbol{x}_i \in \mathrm{SE}(3)\}} \sum_{(i,j) \in \mathcal{E}} -\log \mathcal{L}(\bar{\boldsymbol{R}}_{ij} | \boldsymbol{x}) - \log \mathcal{L}(\bar{\boldsymbol{t}}_{ij} | \boldsymbol{x}). \tag{3}
$$

The negative log-likelihood of the Cartesian measurements can be easily computed from the Gaussian distribution:

$$
-\log \mathcal{L}(\bar{\boldsymbol{t}}_{ij} | \boldsymbol{x}) = \omega_t^2 \|\boldsymbol{t}_j - \boldsymbol{t}_i - \boldsymbol{R}_i \bar{\boldsymbol{t}}_{ij}\|^2 + \text{const.} \tag{4}
$$

Using (2), the negative log-likelihood for $\bar{\boldsymbol{R}}_{ij}$ is:

$$
\begin{aligned}
-\log \mathcal{L}(\bar{\boldsymbol{R}}_{ij} | \boldsymbol{x}) &= -\omega_R^2 \operatorname{tr}\left(\boldsymbol{R}_j^\mathsf{T} \boldsymbol{R}_i \bar{\boldsymbol{R}}_{ij}\right) + \text{const.} \\
&= \frac{\omega_R^2}{2} \|\boldsymbol{R}_j - \boldsymbol{R}_i \bar{\boldsymbol{R}}_{ij}\|_\mathsf{F}^2 + \text{const.}
\end{aligned} \tag{5}
$$

where $\|\cdot\|_\mathsf{F}^2$ is the Frobenius matrix norm (sum of the squares of the entries), and we used $\|\boldsymbol{S} - \boldsymbol{R}\|_\mathsf{F}^2 = \operatorname{tr}\left((\boldsymbol{S} - \boldsymbol{R})(\boldsymbol{S} - \boldsymbol{R})^\mathsf{T}\right)$. The norm $\|\boldsymbol{S} - \boldsymbol{R}\|_\mathsf{F}^2$ is usually referred to as the *chordal distance* between two rotations $\boldsymbol{S}$ and $\boldsymbol{R}$ [19].

Plugging (4) and (5) back into (3) we obtain our ML estimator:

$$
\begin{aligned}
f_{\mathrm{ML}}^\star = \min_{\substack{\{\boldsymbol{t}_i \in \mathbb{R}^3\} \\ \{\boldsymbol{R}_i \in \mathrm{SO}(3)\}}} \sum_{(i,j) \in \mathcal{E}} &\omega_t^2 \|\boldsymbol{t}_j - \boldsymbol{t}_i - \boldsymbol{R}_i \bar{\boldsymbol{t}}_{ij}\|^2 \quad \text{(PGO)} \\
&+ \frac{\omega_R^2}{2} \|\boldsymbol{R}_j - \boldsymbol{R}_i \bar{\boldsymbol{R}}_{ij}\|_\mathsf{F}^2
\end{aligned} \tag{6}
$$

---

^{2}To keep notation simple, we consider measurements with the same distribution. The extension to heterogeneous ω_R^2 and ω_t^2 is trivial.

The main difference between (6) and formulations in related work is the use of the chordal distance (related work instead uses the *geodesic distance* $\left\|\mathrm{Log}\left(\bar{\boldsymbol{R}}_{ij}^{\mathsf{T}}\boldsymbol{R}_i^{\mathsf{T}}\boldsymbol{R}_j\right)\right\|^2$). References [9], [19] show that for small residual errors $\frac{1}{2}\left\|\boldsymbol{R}_j - \boldsymbol{R}_i\bar{\boldsymbol{R}}_{ij}\right\|_{\mathrm{F}}^2 \approx \left\|\mathrm{Log}\left(\bar{\boldsymbol{R}}_{ij}^{\mathsf{T}}\boldsymbol{R}_i^{\mathsf{T}}\boldsymbol{R}_j\right)\right\|^2$, making the formulations equivalent from a practical standpoint.

The advantage of the formulation (6) is that it has a quadratic objective function. This facilitates the derivation of the Lagrangian dual problem, as shown in the next section.

III. Lagrangian Duality in 3D PGO

The main goal of this paper is to provide tools to check if a candidate SLAM solution $\hat{\boldsymbol{x}}$ is globally optimal. If we knew the optimal cost $f_{\mathrm{ML}}^{\star}$ this would be easy: calling $f_{\mathrm{ML}}(\cdot)$ the objective function of (6), if $f_{\mathrm{ML}}(\hat{\boldsymbol{x}}) = f_{\mathrm{ML}}^{\star}$ then $\hat{\boldsymbol{x}}$ is optimal. Unfortunately, $f_{\mathrm{ML}}^{\star}$ is unknown. Our contribution is to show that we can compute close proxies of $f_{\mathrm{ML}}^{\star}$ using duality theory. To make the derivation easier, we first rewrite the problem as a quadratic problem with equality constraints (Section III-A), and then derive the dual (Section III-B).

A. Quadratic Problem with Quadratic Equality Constrains

In this section, we rewrite (6) in order to (i) have vector variables (the rotations $\boldsymbol{R}_i$ are matrices), and (ii) formulate the constraints $\boldsymbol{R}_i \in \mathrm{SO}(3)$ as quadratic equality constraints.

We define $\boldsymbol{r}_i \in \mathbb{R}^9$ as the vectorized version of $\boldsymbol{R}_i$: $\boldsymbol{r}_i \doteq [\boldsymbol{R}_i^{(1)} \ \boldsymbol{R}_i^{(2)} \ \boldsymbol{R}_i^{(3)}]^{\mathsf{T}}$, where $\boldsymbol{R}_i^{(k)}$ is the kth *row* of $\boldsymbol{R}_i$. We use the shorthand $\boldsymbol{r}_i = \mathrm{rows}(\boldsymbol{R}_i)$ to obtain the vector representation $\boldsymbol{r}_i$ of a 3×3 matrix $\boldsymbol{R}_i$. Using this parametrization, each summand in the objective in (6) becomes (using that $\|\boldsymbol{R}\|_{\mathrm{F}} = \|\boldsymbol{R}^{\mathsf{T}}\|_{\mathrm{F}}$ in the first expression):

$$\omega_t^2 \left\|\boldsymbol{t}_j - \boldsymbol{t}_i - \boldsymbol{R}_i\bar{\boldsymbol{t}}_{ij}\right\|^2 + \frac{\omega_R^2}{2}\left\|\boldsymbol{R}_j^{\mathsf{T}} - \bar{\boldsymbol{R}}_{ij}^{\mathsf{T}}\boldsymbol{R}_i^{\mathsf{T}}\right\|_{\mathrm{F}}^2$$

$$= \omega_t^2 \left\|\boldsymbol{t}_j - \boldsymbol{t}_i - \boldsymbol{T}_{ij}\boldsymbol{r}_i\right\|^2 + \frac{\omega_R^2}{2}\left\|\boldsymbol{r}_j - \boldsymbol{Q}_{ij}\boldsymbol{r}_i\right\|^2 \quad (7)$$

where $\boldsymbol{T}_{ij} \doteq \mathbf{I}_3 \otimes \bar{\boldsymbol{t}}_{ij}^{\mathsf{T}} \in \mathbb{R}^{3\times 9}$, $\boldsymbol{Q}_{ij} \doteq \mathbf{I}_3 \otimes \bar{\boldsymbol{R}}_{ij}^{\mathsf{T}} \in \mathbb{R}^{9\times 9}$, and $\otimes$ is the Kronecker product.

We cannot choose arbitrary vectors $\boldsymbol{r}_i \in \mathbb{R}^9$, but have to limit ourself to choices of $\boldsymbol{r}_i$ that produce meaningful rows of a rotation matrix $\boldsymbol{R}_i \in \mathrm{SO}(3)$. The rotation group $\mathrm{SO}(3)$ is defined as $\mathrm{SO}(3) \doteq \{\boldsymbol{R} \in \mathbb{R}^{3\times 3} : \boldsymbol{R}^{\mathsf{T}}\boldsymbol{R} = \mathbf{I}_3, \det(\boldsymbol{R}) = 1\}$, which, written in terms of the rows of $\boldsymbol{R}_i$, becomes:

$$\boldsymbol{R}_i^{\mathsf{T}}\boldsymbol{R}_i = \mathbf{I}_3 \Leftrightarrow (\boldsymbol{R}_i^{(u)})^{\mathsf{T}}\boldsymbol{R}_i^{(v)} = \begin{cases} 1 & \text{if } u = v, \quad u, v = \\ 0 & \text{if } u \neq v, \quad 1, 2, 3 \end{cases}$$

$$\det(\boldsymbol{R}_i) = 1 \Leftrightarrow \boldsymbol{R}_i^{(1)} \times \boldsymbol{R}_i^{(2)} = \boldsymbol{R}_i^{(3)} \quad (8)$$

where $\times$ is the cross product. In other words, the rows of a rotation matrix have to be orthonormal, and have to satisfy the right-hand rule. To derive the dual problem we relax the second condition ($\det(\boldsymbol{R}) = 1$), which amounts to performing estimation in $\mathrm{O}(3)$ rather than $\mathrm{SO}(3)$ (i.e., resulting matrices can have determinant $\det(\boldsymbol{R}) = \pm 1$). Then in Proposition 4 we show how to reconcile our verification techniques to work directly on the original PGO problem (6).

Using (7) and (8) and relaxing the determinant constraints, we rewrite the PGO problem (6) as:

$$f^{\star} = \min_{\{\boldsymbol{r}_i, \boldsymbol{t}_i\}} \sum_{(i,j)\in\mathcal{E}} \omega_t^2 \left\|\boldsymbol{t}_j - \boldsymbol{t}_i - \boldsymbol{T}_{ij}\boldsymbol{r}_i\right\|^2 + \frac{\omega_R^2}{2}\left\|\boldsymbol{r}_j - \boldsymbol{Q}_{ij}\boldsymbol{r}_i\right\|^2$$

$$\text{subject to} \quad \begin{array}{ll} \boldsymbol{r}_i^{\mathsf{T}}\boldsymbol{E}_{uv}\boldsymbol{r}_i = 1, & u = v \\ \boldsymbol{r}_i^{\mathsf{T}}\boldsymbol{E}_{uv}\boldsymbol{r}_i = 0, & u \neq v \end{array} \left.\begin{array}{l} \\ \end{array}\right\} \begin{array}{l} u, v = 1, 2, 3 \\ i = 1, \ldots, n \end{array} \quad (9)$$

where $\boldsymbol{E}_{uv}$ is a 9×9 selection matrix composed of 3×3 blocks that are zero everywhere except the 3×3 block in position (u, v), which is the identity matrix. The matrices $\boldsymbol{E}_{uv}$ are built such that $\boldsymbol{r}_i^{\mathsf{T}}\boldsymbol{E}_{uv}\boldsymbol{r}_i = (\boldsymbol{R}_i^{(u)})^{\mathsf{T}}\boldsymbol{R}_i^{(v)}$, hence the constraints in (9) correspond to the orthonormality constraints in (8).

In order to write (9) is a more compact matrix notation, we define the vector $\breve{\boldsymbol{x}} = [\boldsymbol{t}_1^{\mathsf{T}}, \ldots, \boldsymbol{t}_n^{\mathsf{T}}, \boldsymbol{r}_1^{\mathsf{T}}, \ldots, \boldsymbol{r}_n^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{R}^{12n}$. Using this notation, (9) becomes:

$$f^{\star} = \min_{\breve{\boldsymbol{x}}} \quad \|\breve{\boldsymbol{A}}\breve{\boldsymbol{x}}\|^2 \quad (10)$$

$$\text{subject to} \quad \begin{array}{ll} \breve{\boldsymbol{x}}^{\mathsf{T}}\breve{\boldsymbol{E}}_{iuv}\breve{\boldsymbol{x}} = 1, & u = v \\ \breve{\boldsymbol{x}}^{\mathsf{T}}\breve{\boldsymbol{E}}_{iuv}\breve{\boldsymbol{x}} = 0, & u \neq v \end{array} \left.\begin{array}{l} \\ \end{array}\right\} \begin{array}{l} u, v = 1, 2, 3 \\ i = 1, \ldots, n \end{array}$$

where the matrices $\breve{\boldsymbol{A}}$ and $\breve{\boldsymbol{E}}_{iuv}$ are obtained by stacking the coefficient matrices in (9), with suitable zero blocks for padding; ω_t^2 and ω_R^2 are included in the definition of $\breve{\boldsymbol{A}}$.

Finally, since absolute poses are not observable from relative measurements, we fix a pose to be our reference frame. Without loss of generality we fix the pose of the first node to the identity pose ($\boldsymbol{t}_1 = \boldsymbol{0}_3$ and $\boldsymbol{R}_1 = \mathbf{I}_3$, or, equivalently $\boldsymbol{r}_1 = \mathrm{rows}(\mathbf{I}_3)$). This process is usually called *anchoring*. Fixing the first pose modifies (10) as follows:

$$f^{\star} = \min_{\boldsymbol{x}} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \quad (11)$$

$$\text{subject to} \quad \begin{array}{ll} \boldsymbol{x}^{\mathsf{T}}\boldsymbol{E}_{iuv}\boldsymbol{x} = 1, & u = v \\ \boldsymbol{x}^{\mathsf{T}}\boldsymbol{E}_{iuv}\boldsymbol{x} = 0, & u \neq v \end{array} \left.\begin{array}{l} \\ \end{array}\right\} \begin{array}{l} u, v = 1, 2, 3 \\ i = 1, \ldots, n - 1 \end{array}$$

where $\boldsymbol{x} \in \mathbb{R}^{12(n-1)}$ is obtained by removing the first pose from $\breve{\boldsymbol{x}}$, $\boldsymbol{A}$ is obtained by removing from $\breve{\boldsymbol{A}}$ the columns corresponding to the first pose, and $\boldsymbol{b}$ is the known right-hand-side arising from anchoring; $\boldsymbol{E}_{iuv}$ are the same as $\breve{\boldsymbol{E}}_{iuv}$ but without the rows and columns corresponding to the first pose.

We conclude this section by transforming (11) into an equivalent problem with *homogeneous* objective (i.e., without constant terms in the squared cost). For this purpose, we note that solving (11) is the same as solving:

$$f^{\star} = \min_{\boldsymbol{x}, y} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}y\|^2 \quad \text{(primal problem)}$$

$$\text{subject to} \quad \begin{array}{ll} \boldsymbol{x}^{\mathsf{T}}\boldsymbol{E}_{iuv}\boldsymbol{x} = 1, & u = v \\ \boldsymbol{x}^{\mathsf{T}}\boldsymbol{E}_{iuv}\boldsymbol{x} = 0, & u \neq v \end{array} \left.\begin{array}{l} \\ \end{array}\right\} \begin{array}{l} u, v = 1, 2, 3 \\ i = 1, \ldots, n - 1 \end{array}$$

$$y^2 = 1 \quad (12)$$

meaning that the two problems have the same optimal objective, and the corresponding solutions can be mapped to each other. Intuitively, if the solution of (12) is $[\boldsymbol{x}_H^{\star} \ 1]$, then $\boldsymbol{x}^{\star} = \boldsymbol{x}_H^{\star}$ is also optimal for (11), while if the solution is $[\boldsymbol{x}_H^{\star} \ {-1}]$, then $\boldsymbol{x}^{\star} = -\boldsymbol{x}_H^{\star}$ will be optimal for (11). The inclusion of the *slack* variable y is often referred to as *homogenization*. We refer to (12) as the *primal* problem.

B. The dual problem

In this section we apply Lagrangian duality to the primal problem (12), borrowing the corresponding theory from the optimization community [20], [21]. We begin by recalling basic properties and notions about duality theory and then we tailor these concepts to our SLAM problem.

The key insight of duality is that for every constrained optimization problem of the form:

$$f^\star = \min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) \tag{13}$$
$$\text{subject to} \quad c_i(\boldsymbol{x}) = 0 \quad \forall i \in \mathcal{C}$$

(where $\mathcal{C}$ is a set indexing the constraints $c_i(\boldsymbol{x})$), there is an associated *unconstrained* optimization problem:

$$d^\star = \max_{\boldsymbol{\lambda}} \left(\overbrace{\inf_{\boldsymbol{x}} f(\boldsymbol{x}) + \sum_{i \in \mathcal{C}} \lambda_i c_i(\boldsymbol{x})}^{d(\boldsymbol{\lambda})} \right) \tag{14}$$

called the *dual problem*. The scalar variables λ_i appearing in (14) are called *Lagrange multipliers* or *dual variables*, and the function $d(\boldsymbol{\lambda})$ is called the *dual function*; $\boldsymbol{\lambda}$ is a vector stacking all dual variables. With reference to the dual problem (14), problem (13) is referred to as the *primal* problem. Intuitively, the minimization ("inf") in $d(\boldsymbol{\lambda})$ can be understood as a relaxation of the original problem (13) in which the constraints are transformed into penalty terms in the objective, whose "importance" is controlled by $\boldsymbol{\lambda}$; hence, the maximization (w.r.t., $\boldsymbol{\lambda}$) tries to make this relaxation as tight as possible.

The dual problem (14) has two important properties. First, since the dual function $d(\boldsymbol{\lambda})$ is the pointwise infimum of a family of affine functions of $\boldsymbol{\lambda}$, it is *always* concave, and therefore the dual *maximization* problem (14) is a convex program [20, Sec. 5.2]. Its convexity guarantees that the dual problem can always be solved *globally optimally* using *local* search techniques. Second, given any feasible $\boldsymbol{x}$ for (13) (for which $c_i(\boldsymbol{x}) = 0$), the definition of $d(\boldsymbol{\lambda})$ in (14) shows that $d(\boldsymbol{\lambda}) \leq f(\boldsymbol{x})$ for any choice of $\boldsymbol{\lambda}$. In particular, this must also hold at the optima $\boldsymbol{x}^\star$ and $\boldsymbol{\lambda}^\star$ in (13) and (14), so that:

$$d^\star \leq f^\star. \tag{15}$$

The inequality (15) is referred to as *weak (Lagrangian) duality*, and it enables us to lower-bound the optimal value $f^\star$ of the (possibly very difficult, nonconvex) primal problem (13) using the optimal value $d^\star$ of the (convex) dual problem (14). For some problems, the inequality (15) is tight (i.e. $d^\star = f^\star$), for which we say that *strong (Lagrangian) duality* holds. The quantity $f^\star - d^\star \geq 0$ is called the *duality gap*.

Using weak duality, it is easy to show that, given a primal feasible point $\hat{\boldsymbol{x}}$ and a dual point $\hat{\boldsymbol{\lambda}}$, the following chain of inequality holds:

$$d(\hat{\boldsymbol{\lambda}}) \leq d(\boldsymbol{\lambda}^\star) \doteq d^\star \leq f^\star \doteq f(\boldsymbol{x}^\star) \leq f(\hat{\boldsymbol{x}}) \tag{16}$$

where the first inequality stems from the fact that $\boldsymbol{\lambda}^\star$ attains the maximum over all $\boldsymbol{\lambda}$, and the last follows from the optimality of $f^\star$ (which is the global minimum among all feasible $\boldsymbol{x}$).

Therefore, in this work we exploit a simple idea: given a candidate solution $\hat{\boldsymbol{x}}$, if we are able to find a $\hat{\boldsymbol{\lambda}}$, for which

$d(\hat{\boldsymbol{\lambda}}) = f(\hat{\boldsymbol{x}})$, then the chain of inequalities (16) becomes tight ($d(\hat{\boldsymbol{\lambda}}) = d^\star = f^\star = f(\hat{\boldsymbol{x}})$), which implies that $\hat{\boldsymbol{x}}$ is an optimal solution. Equation (16) thus provides a means of *certifying* the global optimality of a candidate solution $\hat{\boldsymbol{x}}$ for (13) and enables our derivation of *algorithmic* approaches for certifying the correctness of SLAM solutions.

We are now ready to apply duality to our primal problem (12). From (12) and (14), the dual function is:

$$d(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x},y} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}y\|^2 + \sum_{i=1}^{n-1} \Bigg[\sum_{u=1,2,3} \lambda_{iuu}(1 - \boldsymbol{x}^\mathsf{T} \boldsymbol{E}_{iuu}\boldsymbol{x})$$
$$+ \sum_{\substack{u,v=1,2,3 \\ u \neq v}} \lambda_{iuv}(-\boldsymbol{x}^\mathsf{T} \boldsymbol{E}_{iuv}\boldsymbol{x}) \Bigg] + \lambda_y(1 - y^2), \tag{17}$$

where $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers λ_{iuv} and λ_y, associated with the orthonormality and homogeneity constraints in (12), respectively. We observe that the quadratic terms in (17) can be written more compactly as:

$$\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}y\|^2 - \boldsymbol{x}^\mathsf{T} \Bigg[\sum_{i=1}^{n-1} \sum_{u,v=1}^{3} \lambda_{iuv} \boldsymbol{E}_{iuv} \Bigg] \boldsymbol{x} - \lambda_y \, y^2$$
$$= \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix}^\mathsf{T} \begin{bmatrix} \boldsymbol{H}(\boldsymbol{\lambda}) & -\boldsymbol{A}^\mathsf{T}\boldsymbol{b} \\ -\boldsymbol{b}^\mathsf{T}\boldsymbol{A} & \boldsymbol{b}^\mathsf{T}\boldsymbol{b} - \lambda_y \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix}, \tag{18}$$

where

$$\boldsymbol{H}(\boldsymbol{\lambda}) \doteq \boldsymbol{A}^\mathsf{T}\boldsymbol{A} - \sum_{i=1}^{n-1} \sum_{u,v=1}^{3} \lambda_{iuv} \boldsymbol{E}_{iuv}. \tag{19}$$

Calling $\boldsymbol{M}(\boldsymbol{\lambda})$ the matrix in (18), the dual function (17) can thus be written as:

$$d(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x},y} \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix}^\mathsf{T} \boldsymbol{M}(\boldsymbol{\lambda}) \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix} + \sum_{\substack{i=1,\dots,n-1 \\ u=1,2,3}} \lambda_{iuu} + \lambda_y. \tag{20}$$

Now in the dual problem (14) we try to maximize $d(\boldsymbol{\lambda})$; however, from (20) we see that $d(\boldsymbol{\lambda}) = -\infty$ if $\boldsymbol{M}(\boldsymbol{\lambda})$ has a negative eigenvalue (by letting $[\boldsymbol{x}, y]$ lie in the corresponding eigenspace). Consequently, we can safely restrict our search to the vectors $\boldsymbol{\lambda}$ that preserve positive semi-definiteness of $\boldsymbol{M}(\boldsymbol{\lambda})$ [20, Sec. 5.1.5]. Moreover:

$$\boldsymbol{M}(\boldsymbol{\lambda}) \succeq 0 \quad \Rightarrow \quad \inf_{\boldsymbol{x},y} \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix}^\mathsf{T} \boldsymbol{M}(\boldsymbol{\lambda}) \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix} = 0, \tag{21}$$

as the minimization over $[\boldsymbol{x} \; y]$ in the homogenized problem (21) is unconstrained. The dual problem (14) thus becomes:

$$d^\star = \max_{\boldsymbol{\lambda}} \quad \sum_{\substack{i=1,\dots,n \\ u=1,2,3}} \lambda_{iuu} + \lambda_y \quad \text{(dual problem)}$$
$$\text{subject to} \quad \boldsymbol{M}(\boldsymbol{\lambda}) \succeq 0. \tag{22}$$

The dual SLAM problem turns out to be a *semidefinite program* (SDP), for which specialized solvers exist [22]. In the following section we discuss the relations between the primal and the dual problem, and elucidate on its practical use.

IV. Relation between the Primal and the Dual Problem and Practical Use

In this section we present two powerful applications of the dual problem (22). Section IV-A deals with the case in which one is given a candidate PGO solution, and wants to evaluate its quality, possibly certifying its optimality. Section IV-B shows that in particular cases (when the duality gap is zero) one can obtain an optimal solution of the primal problem from the solution $\boldsymbol{\lambda}^\star$ of the dual.

In both sections, we use the following property.

Lemma 1 (Primal optimal solution and zero duality gap): If the duality gap is zero ($d^\star = f^\star$), then any primal optimal solution $[\boldsymbol{x}^\star\ 1]$ of (12) is in the null space of the matrix $\boldsymbol{M}(\boldsymbol{\lambda}^\star)$, where $\boldsymbol{\lambda}^\star$ is the solution of the dual problem (22).

Proof: When the duality gap is zero, any minimizer of the primal problem is also a minimizer for the infimum in the dual function (14) [20, Sec. 5.5.5]. Consider a primal optimal solution $[\boldsymbol{x}^\star\ 1]$. We already observed than any such minimizer annihilates the quadratic term in (20), and therefore it holds that $[\boldsymbol{x}^\star\ 1]^\mathsf{T}\boldsymbol{M}(\boldsymbol{\lambda}^\star)[\boldsymbol{x}^\star\ 1] = 0$, which implies that $[\boldsymbol{x}^\star\ 1]$ is in the null space of $\boldsymbol{M}(\boldsymbol{\lambda}^\star)$, proving the claim. ∎

We give an alternative proof, which does not require prior knowledge on duality, in the supplemental material [17].

A. Verification

In this section we consider the case in which we are given a candidate solution $\hat{\boldsymbol{x}}$ for the primal problem, and we want to evaluate the quality of this solution. For brevity, we denote with $f(\hat{\boldsymbol{x}})$ the objective evaluated at $\hat{\boldsymbol{x}}$, for both (11) and its homogeneous form (12) (for the latter we imply $y = 1$).

We begin with the following proposition, whose proof easily follows from (16) and the discussion in Section III-B.

Proposition 2 (Verification of Primal Optimal Objective): Given a candidate solution $\hat{\boldsymbol{x}}$ for the primal problem (12), if $f(\hat{\boldsymbol{x}}) = d^\star$, then the duality gap is zero and $\hat{\boldsymbol{x}}$ is an optimal solution of (12). Moreover, even if the duality gap is nonzero, $f(\hat{\boldsymbol{x}}) - d^\star \geq f(\hat{\boldsymbol{x}}) - f^\star$, meaning that $f(\hat{\boldsymbol{x}}) - d^\star$ is an upper-bound for the sub-optimality gap of $\hat{\boldsymbol{x}}$.

Proposition 2 ensures that the candidate $\hat{\boldsymbol{x}}$ is optimal when $f(\hat{\boldsymbol{x}}) = d^\star$. Moreover, even in the case in which we get $f(\hat{\boldsymbol{x}}) > d^\star$, the quantity $f(\hat{\boldsymbol{x}}) - d^\star$ can be used as an indicator of how far $\hat{\boldsymbol{x}}$ is from the global optimum.

While Proposition 2 already provides means of verifying a candidate solution, it requires solving the dual problem, to compute $d^\star$. The following proposition provides a technique to verify the optimality of $\hat{\boldsymbol{x}}$ without solving the SDP.

Proposition 3 (Verification of Primal Optimal Solution): Given a candidate solution $\hat{\boldsymbol{x}}$ for the primal problem (12), if the solution $\hat{\boldsymbol{\lambda}}$ of the linear system

$$\boldsymbol{M}(\hat{\boldsymbol{\lambda}}) \begin{bmatrix} \hat{\boldsymbol{x}} \\ 1 \end{bmatrix} = \boldsymbol{0} \qquad \text{(to be solved w.r.t. } \hat{\boldsymbol{\lambda}}) \qquad (23)$$

is such that $\boldsymbol{M}(\hat{\boldsymbol{\lambda}}) \succeq 0$ and $d(\hat{\boldsymbol{\lambda}}) = f(\hat{\boldsymbol{x}})$, then the duality gap is zero and $\hat{\boldsymbol{x}}$ is a primal optimal solution.

Proof: From Lemma 1, we know that when the duality gap is zero, it must hold $\boldsymbol{M}(\boldsymbol{\lambda}^\star)[\boldsymbol{x}^\star\ 1] = \boldsymbol{0}$. Therefore,

in Proposition 3 we solve the linear system (23), trying to obtain $\boldsymbol{\lambda}^\star$. When $\boldsymbol{M}(\hat{\boldsymbol{\lambda}}) \succeq 0$, the solution $\hat{\boldsymbol{\lambda}}$ of (23) is such that $d(\hat{\boldsymbol{\lambda}}) \leq d^\star$ (recall that $d^\star$ is the maximum over $\boldsymbol{\lambda}$). Therefore, it holds that (i) $d(\hat{\boldsymbol{\lambda}}) \leq d^\star \leq f^\star \leq f(\hat{\boldsymbol{x}})$ (by weak duality and optimality of $f^\star$). However, if $d(\hat{\boldsymbol{\lambda}}) = f(\hat{\boldsymbol{x}})$, the chain of inequalities (i) becomes tight, $d(\hat{\boldsymbol{\lambda}}) = d^\star = f^\star = f(\hat{\boldsymbol{x}})$, implying that $\hat{\boldsymbol{x}}$ attains the optimal objective $f^\star$. ∎

In practice, iterative SLAM solvers optimize (6), rather than the primal problem (12). A natural question is then how to use the results in Propositions 2-3 (which relate $f^\star$ and $d^\star$) to verify the solution of (6), whose optimal value is $f_{\text{ML}}^\star \geq f^\star$. This extension is given by the following proposition, which essentially states that Propositions 2-3 can be applied directly to check the solution of (6).

Proposition 4 (Verification techniques for PGO): The following statements hold true:

(V1) Given a candidate solution $\hat{\boldsymbol{x}}$ for the PGO problem (6), if $f_{\text{ML}}(\hat{\boldsymbol{x}}) = d^\star$, then $\hat{\boldsymbol{x}}$ is an optimal solution of (6). Moreover, $f_{\text{ML}}(\hat{\boldsymbol{x}}) - d^\star \geq f_{\text{ML}}(\hat{\boldsymbol{x}}) - f_{\text{ML}}^\star$, i.e., $f_{\text{ML}}(\hat{\boldsymbol{x}}) - d^\star$ is an upper-bound for the sub-optimality gap of $\hat{\boldsymbol{x}}$.

(V2) Given a candidate solution $\hat{\boldsymbol{x}}$ for the PGO problem (6), if the solution $\hat{\boldsymbol{\lambda}}$ of the linear system (23) is such that $\boldsymbol{M}(\hat{\boldsymbol{\lambda}}) \succeq 0$ and $d(\hat{\boldsymbol{\lambda}}) = f_{\text{ML}}(\hat{\boldsymbol{x}})$, then the duality gap is zero and $\hat{\boldsymbol{x}}$ is an optimal solution of (6).

Proof: The first claim can be proven by observing that the following chain of inequalities holds (i) $d^\star \leq f^\star \leq f_{\text{ML}}^\star \leq f_{\text{ML}}(\hat{\boldsymbol{x}})$, hence $f_{\text{ML}}(\hat{\boldsymbol{x}}) = d^\star$ implies that $d^\star = f^\star = f_{\text{ML}}^\star = f_{\text{ML}}(\hat{\boldsymbol{x}})$, which implies that $\hat{\boldsymbol{x}}$ is optimal. The inequality $f_{\text{ML}}(\hat{\boldsymbol{x}}) - d^\star \geq f_{\text{ML}}(\hat{\boldsymbol{x}}) - f_{\text{ML}}^\star$ easily follows from (i). The second claim can be proven in the same way, noting that (ii) $d(\hat{\boldsymbol{\lambda}}) \leq d^\star \leq f^\star \leq f_{\text{ML}}^\star \leq f_{\text{ML}}(\hat{\boldsymbol{x}})$: if we are able to compute a $\hat{\boldsymbol{\lambda}}$ that is dual feasible ($\boldsymbol{M}(\hat{\boldsymbol{\lambda}}) \succeq 0$) and such that $d(\hat{\boldsymbol{\lambda}}) = f_{\text{ML}}(\hat{\boldsymbol{x}})$, then it must hold $d(\hat{\boldsymbol{\lambda}}) = d^\star = f^\star = f_{\text{ML}}^\star = f_{\text{ML}}(\hat{\boldsymbol{x}})$, which implies that $f_{\text{ML}}(\hat{\boldsymbol{x}})$ is optimal. ∎

B. Primal Optimal Solutions

In this section, rather than verifying the quality of a given candidate solution, we show how to use duality to compute a primal optimal solution directly. We focus on the particular case in which the duality gap is zero. This case is of interest as we observe that the duality gap is often zero in practice.

From Lemma 1 we know that an optimal solution must be in the null space of $\boldsymbol{M}(\boldsymbol{\lambda}^\star)$. This motivates the following proposition, which provides a way to compute a primal optimal solution directly from the solution $\boldsymbol{\lambda}^\star$ of (22).

Proposition 5: If the duality gap is zero and $\boldsymbol{\lambda}^\star$ is an optimal solution of (22), then an optimal solution $\boldsymbol{x}^\star$ of (12) can be computed by solving the following linear system:

$$\begin{bmatrix} \boldsymbol{H}(\boldsymbol{\lambda}^\star) \\ \boldsymbol{b}^\mathsf{T}\boldsymbol{A} \end{bmatrix} \boldsymbol{x}^\star = \begin{bmatrix} \boldsymbol{A}^\mathsf{T}\boldsymbol{b} \\ \lambda_y^\star - \boldsymbol{b}^\mathsf{T}\boldsymbol{b} \end{bmatrix} \quad \text{(to be solved w.r.t. } \boldsymbol{x}^\star)$$
$$(24)$$

Proof: From Lemma 1, we know that when the duality gap is zero, it holds $\boldsymbol{M}(\boldsymbol{\lambda}^\star)[\boldsymbol{x}^\star\ 1] = \boldsymbol{0}$. Then, recalling the structure of $\boldsymbol{M}(\boldsymbol{\lambda}^\star)$ from (18), it's easy to see that (24) only rewrites the condition $\boldsymbol{M}(\boldsymbol{\lambda}^\star)[\boldsymbol{x}^\star\ 1] = \boldsymbol{0}$, moving the constant terms to the right-hand side. ∎
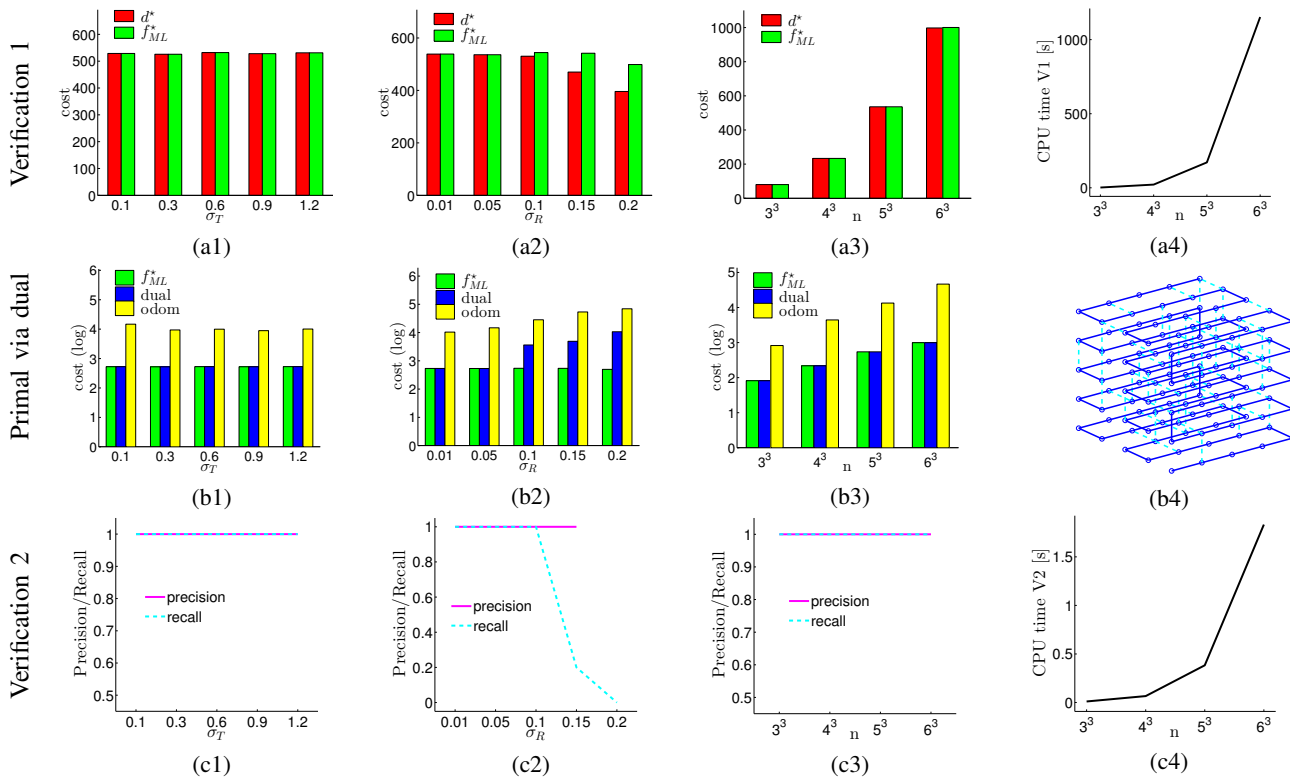
Fig. 2. Statistics regarding the verification technique V1 (first row), the computation of the primal solution via the dual (second row), and the verification technique V2 (third row). Results are shown for different levels of translation noise σ_T (first column), rotation noise σ_R (second column), and size of the problem (third column). The rightmost column shows the CPU time required by V1 and V2 and the datasets used for the Monte Carlo simulations.

Note that Lemma 1 ensures that the linear system (24) admits a solution. Proposition 5 allows finding an optimal solution to the primal problem (12). If this satisfies the determinant constraints in (6), it follows that this solution is also optimal for the original PGO problem (6). While currently we cannot prove that the determinant constraints are always satisfied, this was always the case in our experiments.

V. EXPERIMENTS

In this section we show that the first verification technique in Proposition 4 (referred to as V1) enables accurate quantification of the sub-optimality gap of a candidate SLAM solution, but that current SDP solvers do not scale well with increasing problem size. The second verification technique in Proposition 4 (referred to as V2) provides a convenient alternative for large-scale problems: it is reliable and less computationally demanding. Finally, we provide empirical evidence that when the duality gap is zero we can compute an optimal solution from the solution of the dual problem, as suggested by Proposition 5. The SDP (22) is solved using SDPA [22].

Effectiveness of V1. Proposition 4(V1) ensures that for any candidate solution $\hat{x}$, $f_{\mathrm{ML}}(\hat{x}) - d^\star$ represents an upper-bound on the sub-optimality gap $f_{\mathrm{ML}}(\hat{x}) - f_{\mathrm{ML}}^\star$ of $\hat{x}$. However, it is possible that this bound is very loose, in which case it would be of little practical utility. In this section, we show that $d^\star$ is close to $f_{\mathrm{ML}}^\star$ in practice, hence $f_{\mathrm{ML}}(\hat{x}) - d^\star$ is a very good measure of the sub-optimality gap of the candidate $\hat{x}$.

In our experiments we compute the "optimal" solution $f_{\mathrm{ML}}^\star$ of (6) by refining the *chordal* initialization of [23], [24] with 10

Gauss-Newton (GN) iterations. While one cannot guarantee *a priori* that this approach always produces the optimal estimate, using the results of this paper we will be able to check optimality a posteriori.

We evaluated how close $d^\star$ is to $f_{\mathrm{ML}}^\star$ on the cube dataset of Fig. 2(b4). In this dataset, the odometric trajectory is simulated as the robot travels on a 3D grid world, and random loop closures are added between nearby nodes, with probability 0.3. Relative pose measurements are obtained by contaminating the true relative poses with zero-mean Gaussian noise, with standard deviation σ_T and σ_R for the translational and rotational noise, respectively. Statistics are computed over 10 runs: for each run we create a cube with random connectivity and random measurement noise. We consider an example with $n = 5^3$ poses and varying noise levels σ_T and σ_R.

Fig. 2(a1) shows $d^\star$ and $f_{\mathrm{ML}}^\star$ for different translational noise levels, fixing $\sigma_R = 0.05$rad. The figure shows that $d^\star = f_{\mathrm{ML}}^\star$ (zero duality gap) independently on the translational noise level, hence $d^\star$ is a very good proxy of $f_{\mathrm{ML}}^\star$.

Fig. 2(a2) shows $d^\star$ and $f_{\mathrm{ML}}^\star$ for different rotational noise, fixing $\sigma_T = 0.1$m. In this case the duality gap $f_{\mathrm{ML}}^\star - d^\star$ is more sensitive to the noise level, and for large rotational noise $d^\star$ becomes smaller than $f_{\mathrm{ML}}^\star$. However, the gap $f_{\mathrm{ML}}^\star - d^\star$ remains small, and is within 20% of $f_{\mathrm{ML}}^\star$ in all cases.

Fig. 2(a3) shows $d^\star$ and $f_{\mathrm{ML}}^\star$ for different sizes of the cube dataset, fixing $\sigma_R = 0.05$rad and $\sigma_T = 0.1$m. Again in this case $d^\star = f_{\mathrm{ML}}^\star$ (zero duality gap) independently of the size of the dataset. However, we observe that the SDP (22) becomes intractable for larger problem sizes, as shown in Fig. 2(a4).

Primal optimal solution via the dual. Here we demonstrate experimentally that one can recover a primal optimal solution $\boldsymbol{x}^\star$ for the SLAM problem from the dual optimal solution $\boldsymbol{\lambda}^\star$ of (22) whenever strong duality holds.

For each of the experimental trials described previously, we also computed an estimate $\hat{\boldsymbol{x}}$ for the SLAM solution directly from the optimal solution $\boldsymbol{\lambda}^\star$ for the dual program (22) using equation (24). Figs. 2(b1)-(b3) compare the value $f(\hat{\boldsymbol{x}})$ of this estimate against the value $f_{\texttt{ML}}^\star$ of the solution obtained by solving the SLAM problem directly using the chordal initialization, and against the cost of the odometric guess.

We can see that for those experimental conditions in which strong duality holds (those experiments in Figs. 2(a1)-(a3) for which the green and red bars are the same height), the estimate $\hat{\boldsymbol{x}}$ in fact achieves the (certified) globally optimal cost $f_{\texttt{ML}}^\star$ and is therefore a globally optimal solution for the SLAM problem, as guaranteed by Proposition 5. More interestingly, we also find that even in those cases when strong duality does *not* hold (in which case Proposition 5 no longer guarantees that $\hat{\boldsymbol{x}}$ is a primal optimal solution), the quality of the candidate $\hat{\boldsymbol{x}}$ degrades gracefully (i.e. the gap $f(\hat{\boldsymbol{x}}) - f_{\texttt{ML}}^\star$ increases gradually) with increasing noise levels. In particular, we find that $\hat{\boldsymbol{x}}$ outperforms the odometric guess in all tested cases.

These experiments confirm that we can extract a primal optimal solution $\boldsymbol{x}^\star$ for the SLAM problem *directly* from the solution $\boldsymbol{\lambda}^\star$ of the convex dual problem whenever strong duality holds. However, this approach is unfortunately not currently practical as a general-purpose SLAM technique, due to the high computational cost of solving large scale SDPs.

Effectiveness of $\texttt{V2}$**.** In this section we show that $\texttt{V2}$ is a computationally tractable verification approach, and it preserves the desirable properties of $\texttt{V1}$, i.e., it is able to discern optimal estimates from suboptimal ones. In contrast to $\texttt{V1}$, $\texttt{V2}$ does not quantify the sub-optimality gap, but can only give a binary answer: either it certifies that the estimate $\hat{\boldsymbol{x}}$ is optimal (by producing a dual certificate $\hat{\boldsymbol{\lambda}}$), or it is inconclusive. Before moving to the real tests, we consider the cube scenario discussed in the previous section.

We perform the following test: for each realization of the cube scenario, we compute the optimal solution $\boldsymbol{x}^\star$ (attaining $f_{\texttt{ML}}^\star$) as in the previous tests, and a sub-optimal solution $\boldsymbol{x}^\dagger$ (attaining $f^\dagger > f^\star$) by bootstrapping the Gauss-Newton method with a random initial guess. Then, we apply the verification technique to both $\boldsymbol{x}^\star$ and $\boldsymbol{x}^\dagger$ and see if they pass the optimality test. Using the results we can compute the *precision* and *recall* of our classification:

$$\text{precision} = \frac{|\checkmark\ \boldsymbol{x}^\star|}{|\checkmark\ \boldsymbol{x}^\star| + |\checkmark\ \boldsymbol{x}^\dagger|} \quad \text{recall} = \frac{|\checkmark\ \boldsymbol{x}^\star|}{|\checkmark\ \boldsymbol{x}^\star| + |\text{X}\ \boldsymbol{x}^\star|} \quad (25)$$

where $|\checkmark\ \boldsymbol{x}^\star|$ denotes the number of tests in which an optimal solution $\boldsymbol{x}^\star$ was accepted as optimal by $\texttt{V2}$, $|\checkmark\ \boldsymbol{x}^\dagger|$ is the number of tests in which a suboptimal solution was accepted as optimal, and $|\text{X}\ \boldsymbol{x}^\star|$ is the number of tests in which $\texttt{V2}$ was not able to certify the optimality of an optimal solution.

We point out that Proposition 4($\texttt{V2}$) guarantees that our certification approach always has precision equal to 1; however, recall may be less than 1, and will be for cases in which strong duality does not hold. We plot precision/recall for different

		$f_{\texttt{ML}}(\hat{\boldsymbol{x}})$	$d(\hat{\boldsymbol{\lambda}})$	μ	Time	V2
sphere	Init.	$5.7595\cdot10^2$	$5.75\cdot10^2$	$-1.80\cdot10^{-4}$	609	✓
$n{=}2500$ $m{=}4949$	Odom.	$5.8019\cdot10^2$	$4.38\cdot10^2$	$-1.45\cdot10^{-4}$	597	X
sphere-a	Init.	$1.2485\cdot10^6$	$1.25\cdot10^6$	$-9.64\cdot10^{-3}$	332	✓
$n{=}2200$ $m{=}8647$	Odom.	$3.0413\cdot10^6$	$3.04\cdot10^6$	$-1.01\cdot10^2$	2.5	X
torus	Init.	$1.2114\cdot10^4$	$1.21\cdot10^4$	$-7.85\cdot10^{-2}$	39	✓
$n{=}5000$ $m{=}9048$	Odom.	$2.7666\cdot10^4$	$2.76\cdot10^4$	$-1.04\cdot10^2$	3.7	X
cube	Init.	$4.216\cdot10^4$	$4.22\cdot10^4$	$-1.24\cdot10^{-2}$	1646	✓
$n{=}8000$ $m{=}22236$	Odom.	$2.7465\cdot10^5$	$2.74\cdot10^5$	$-9.98\cdot10^1$	68.2	X
garage	Init.	$6.2994\cdot10^{-1}$	$6.11\cdot10^{-1}$	$-6.78\cdot10^{-2}$	12.3	✓
$n{=}1661$ $m{=}6275$	Odom.	$6.2997\cdot10^{-1}$	$3.53\cdot10^{-1}$	$-6.75\cdot10^{-2}$	11.2	X
cubicle	Init.	$6.2481\cdot10^2$	$6.25\cdot10^2$	$-4.76\cdot10^{-1}$	16.8	✓
$n{=}5750$ $m{=}16869$	Odom.	$6.2484\cdot10^2$	$6.16\cdot10^2$	$-4.75\cdot10^{-1}$	15.5	✓
rim	Init.	$1.235\cdot10^4$	$1.23\cdot10^4$	$-9.77\cdot10^1$	7.7	X
$n{=}10195$ $m{=}29743$	Odom.	$1.6985\cdot10^4$	$-2.8\cdot10^4$	$-9.38\cdot10^1$	7.1	X

TABLE I
VERIFICATION TECHNIQUE $\texttt{V2}$ ON LARGE-SCALE SLAM DATASETS.

levels of translational and rotational noise, and for different sizes of the dataset, as shown in Figs. 2(c1)-(c2)-(c3). We observe that only for larger rotational noise Fig. 2(c2) the recall decreases; these are exactly the cases in which the duality gap is nonzero. (For $\sigma_R = 0.2$rad, $\texttt{V2}$ was not able to certify optimality in any case, which means that the precision becomes undefined ($\frac{0}{0}$) and for this reason we do not show the corresponding data point in Fig. 2(c2).) Finally, Fig. 2(d2) shows the CPU time required by $\texttt{V2}$. This is the time required to solve the linear system (23), and check if $\boldsymbol{M}(\hat{\boldsymbol{\lambda}}) \succeq 0$. $\texttt{V2}$ is computationally cheap, as it does not require solving the SDP.

We conclude the experimental part of this paper by testing the performance of the verification technique $\texttt{V2}$ on large-scale SLAM datasets. We consider the same datasets as [23]: the sphere, sphere-a, torus and cube are simulated datasets, while the garage, cubicle and rim are real datasets. For the results in this paper we substituted the covariances in the datasets of the scenarios sphere, sphere-a, garage, cubicle and rim with isotropic ones, as required by the PGO formulation (6).

Table I shows the results of the application of $\texttt{V2}$ to the SLAM datasets. The first column shows the cost obtained by applying a $\texttt{GN}$ method starting from the initialization [23] (rows "Init." in the table) and from the odometric guess (rows "Odom."). These are the candidate solutions that we want to check, using $\texttt{V2}$. The columns "$d(\hat{\boldsymbol{\lambda}})$" and "$\mu$" show intermediate results of $\texttt{V2}$. In particular, $d(\hat{\boldsymbol{\lambda}})$ is the same one described in Proposition 4($\texttt{V2}$), while μ is the smallest eigenvalue of $\boldsymbol{M}(\hat{\boldsymbol{\lambda}})$. Recall that $\texttt{V2}$ certifies optimality when $f_{\texttt{ML}}(\hat{\boldsymbol{x}}) = d(\hat{\boldsymbol{\lambda}})$ and $\mu \geq 0$ (if the smallest eigenvalue is non-negative, then $\boldsymbol{M}(\hat{\boldsymbol{\lambda}}) \succeq 0$). We specify a tolerance in these tests, since the $\texttt{GN}$ estimate $\hat{\boldsymbol{x}}$ will not attain the optimal solution $\boldsymbol{x}^\star$ exactly. Consequently, the solution $\hat{\boldsymbol{\lambda}}$ of (23) is not exact, so we consider that $f_{\texttt{ML}}(\hat{\boldsymbol{x}}) = d(\hat{\boldsymbol{\lambda}})$ if $|f_{\texttt{ML}}(\hat{\boldsymbol{x}}) - d(\hat{\boldsymbol{\lambda}})|/f_{\texttt{ML}}(\hat{\boldsymbol{x}}) < 20\%$. Similarly, we add some tolerance to the condition $\mu \geq 0$, and accept $\mu \geq -1$. Recall that μ is expected to be slightly negative, as the objective in (22) tends to push $\boldsymbol{M}(\boldsymbol{\lambda})$ towards the boundary of the positive definite cone, and the SDP is solved numerically.

Let us start our analysis from the sphere dataset. When using the initialization ("Init." row), the $\texttt{GN}$ method attains an objective $5.7595 \cdot 10^2$. In this case, the two conditions for $\texttt{V2}$ are satisfied and we can certify the optimality of the resulting estimate (green check-mark in the rightmost column).

When using the odometric initialization the resulting cost is larger (red entry in the column $f_{\mathtt{ML}}(\hat{x})$), hence the estimate is suboptimal. V2 is able to identify the suboptimality, since $d(\hat{\boldsymbol{\lambda}})$ becomes much smaller than $f_{\mathtt{ML}}(\hat{x})$. Hence V2 correctly decides not to certify the optimality of the odometric estimate (red "X" in the last column). Similar considerations hold for the second scenario, the challenging sphere-a: in this case the initialized estimate (shown in Fig. 1, top left) is accepted as optimal. The odometric estimate, instead, is trapped in a local minimum (Fig. 1, bottom left), and the optimality test V2 correctly rejects the estimate since it leads to a very negative μ. Similar considerations hold for the torus, cube, and the cubicle datasets: our technique is able to discern optimal solutions from suboptimal estimates in all cases.

For the garage dataset, the initialized estimate is classified as optimal, while the odometric estimate, which has a similar cost ($6.2997 \cdot 10^{-1}$ vs $6.2994 \cdot 10^{-1}$), is rejected as suboptimal. We observed the corresponding trajectory estimates (see [17]) and, while they are are both visually correct and have very similar costs, they do not overlap. This may indicate the presence of regions in the cost functions that are nearly flat, i.e., for which different estimates can have similar cost. Allowing for extra GN iterations, the odometric estimate converges to the same cost as the initialized one, and our technique is able to certify its optimality. Empirically, we observed that estimates that are suboptimal because they need extra iterations to converge tend to fail the check on $d(\hat{\boldsymbol{\lambda}})$ (compare with sphere and garage), while estimates that converge to a wrong minimum tend to fail the check on μ (compare with sphere-a, torus, and cube).

For the rim dataset, both estimates ("init" and "odom") are rejected by V2. While the initialized estimate is visually correct, currently, we cannot conclude anything as it might be that there exists a better estimate (i.e., GN failed), or our technique failed because the duality gap was nonzero.

Finally, in the column "Time", we report the CPU time (in seconds) required to perform the second verification technique. Most of the time here is spent computing the smallest eigenvalue μ of $M(\hat{\boldsymbol{\lambda}})$, which was obtained using Matlab's eigs, specifying -100 as guess for the eigenvalue. The CPU time depends on the size of the problem, but also depends on the distance between our guess (-100) and the closest eigenvalue. We leave a more thorough investigation of these computational aspects for future work.

VI. Conclusion

We show that Lagrangian duality is an effective tool to assess the quality of a given SLAM estimate. We propose two techniques to judge if an estimate is globally optimal (i.e., it is the ML estimate), and we show that when the duality gap is zero one can compute an optimal SLAM solution from the dual problem. The performance of our verification techniques is extensively tested in real and simulated datasets, including large-scale benchmarks. Many theoretical questions remain open, e.g., under which conditions the duality gap is zero? is it possible to derive bounds on the duality gap? We also leave as future work more practical problems, such as the one of exploring more efficient solvers for large SDPs, possibly exploiting problem structure.

References

[1] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous Robots*, pp. 333–349, Apr 1997.

[2] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, Shanghai, China, May 2011.

[3] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Intl. J. of Robotics Research*, vol. 31, pp. 217–236, Feb 2012.

[4] E. Olson, J. Leonard, and S. Teller, "Fast iterative alignment of pose graphs with poor initial estimates," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2006, pp. 2262–2269.

[5] G. Grisetti, C. Stachniss, and W. Burgard, "Non-linear constraint network optimization for efficient map learning," *Trans. on Intelligent Transportation systems*, vol. 10, no. 3, pp. 428–439, 2009.

[6] D. Rosen, M. Kaess, and J. Leonard, "RISE: An incremental trust-region method for robust online sparse least-squares estimation," *IEEE Trans. Robotics*, 2014.

[7] F. Dellaert and A. Stroupe, "Linear 2D localization and mapping for single and multiple robots," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2002.

[8] L. Carlone, R. Aragues, J. Castellanos, and B. Bona, "A fast and accurate approximation for planar pose graph optimization," *Intl. J. of Robotics Research*, 2014.

[9] L. Carlone and F. Dellaert, "Duality-based verification techniques for 2D SLAM," in *Intl. Conf. on Robotics and Automation (ICRA), accepted*, 2015.

[10] D. Rosen, C. DuHadway, and J. Leonard, "A convex relaxation for approximate global optimization in simultaneous localization and mapping," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2015.

[11] R. Tron, B. Afsari, and R. Vidal, "Intrinsic consensus on SO(3) with almost global convergence," in *IEEE Conf. on Decision and Control*, 2012.

[12] L. Carlone and A. Censi, "From angular manifolds to the integer lattice: Guaranteed orientation estimation with application to pose graph optimization," *IEEE Trans. Robotics*, 2014.

[13] S. Huang, Y. Lai, U. Frese, and G. Dissanayake, "How far is SLAM from a linear least squares problem?" in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2010, pp. 3011–3016.

[14] H. Wang, G. Hu, S. Huang, and G. Dissanayake, "On the structure of nonlinearities in pose graph SLAM," in *Robotics: Science and Systems (RSS)*, 2012.

[15] L. Carlone, "Convergence analysis of pose graph optimization via Gauss-Newton methods," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 965–972.

[16] K. Khosoussi, S. Huang, and G. Dissanayake, "Novel insights into the impact of graph structure on SLAM," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2014.

[17] L. Carlone, D. Rosen, G. Calafiore, J. Leonard, and F. Dellaert, "Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions. supplemental material." [Online]. Available: www.lucacarlone.com/index.php/resources/research/duality3DPGO

[18] N. Boumal, A. Singer, and P. Absil, "Cramer-rao bounds for synchronization of rotations," *Information and Inference*, vol. 3, no. 1, pp. 1–39, 2014.

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

[20] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[21] G. Calafiore and L. E. Ghaoui, *Optimization Models*. Cambridge University Press, 2014.

[22] M. Yamashita, K. Fujisawa, and M. Kojima, "Implementation and evaluation of SDPA 6.0 (semidefinite programming algorithm 6.0)," *Optimization Methods and Software*, vol. 18, no. 4, pp. 491–505, Aug. 2003.

[23] 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 *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2015.

[24] D. Martinec and T. Pajdla, "Robust rotation and translation estimation in multiview reconstruction," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.