Initialization of 3D Pose Graph Optimization using Lagrangian duality

Jesus Briales Javier Gonzalez-Jimenez

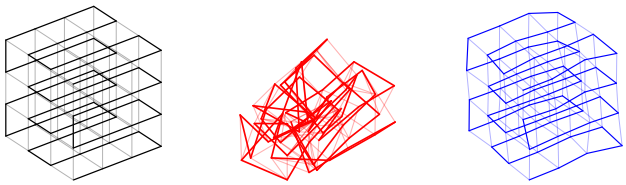
***Abstract*— Pose Graph Optimization (PGO) is the *de facto* choice to solve the trajectory of an agent in Simultaneous Localization and Mapping (SLAM). The Maximum Likelihood Estimation (MLE) for PGO is a non-convex problem for which no known technique is able to guarantee a globally optimal solution under general conditions. In recent years, Lagrangian duality has proved suitable to provide good, frequently *tight* relaxations of the hard PGO problem through *convex* Semidefinite Programming (SDP). In this work, we build from the state-of-the-art Lagrangian relaxation [1] and contribute a complete *recovery procedure* that, given the (tractable) optimal solution of the relaxation, provides either the *optimal* MLE solution if the relaxation is tight, or a remarkably good feasible guess if the relaxation is non-tight, which occurs in specially challenging PGO problems (very noisy observations, low graph connectivity, etc.). In the latter case, when used for initialization of local iterative methods, our approach outperforms other state-of-the-art approaches converging to better solutions. We support our claims with extensive experiments.**

I. Introduction

Pose Graph Optimization (PGO) is at present the most widespread formulation for *Simultaneous Localization and Mapping* (SLAM) in robotics, where it serves as back-end receiving the relative pose measurements produced by the front-end from the sensory data. PGO is also widely known as SE(d)-Synchronization, and it is a pervasive problem in many other fields, including computer vision and control, where closely related problems appear in the tasks of structure from motion [2], [3], extrinsic sensor calibration [4], [5], sensor network localization [6], etc.

The Pose Graph Optimization problem consists in finding the trajectory that best explains all the observed relative poses. This is formulated as a high-dimensional non-convex optimization problem, which makes finding the optimal solution to a PGO instance a *hard problem* in general. On the other hand, *local minima* in the context of PGO must be avoided at any cost as these may lie arbitrarily far from the optimal solution, rendering the estimate completely useless (see Fig. 1(b)). When applied at the core of real autonomous applications, which may involve safety issues (e.g. transportation of passengers, medical interventions, etc.), failing to converge to the *globally optimal solution* can thus drive to unnacceptable catastrophic failures [7].

An important venue of research that has shown remarkably successful for dealing with hard optimization problem like PGO is that of relaxation techniques. Relaxing an optimization problem is a modeling strategy and kind of an art, although plenty of references and examples exist in the literature for many various fields. In the case of PGO, whose *optimization domain* is highly non-convex, the *relaxation of the constraints* in the original problem yields a new optimization problem that is tractable in the sense that its *globally optimal solution* can be attained. The most important aspect about a *good* relaxation is that, when properly exploited, it may provide a suboptimal yet remarkably good estimate that serves as initialization to local iterative methods (which are the standard tool in the case of PGO [9]–[11]). A good initialization not only decreases the odds of converging to local minima, but it usually enables faster convergence. Last but not least, under certain circumstances, a relaxation may even provide a *certifiably globally optimal* solution for the original hard problem [12].



(a) groundtruth (b) chordal init. + GN (c) our init. + GN

Fig. 1. Example of a challenging PGO problem with non-tight (best) Lagrangian relaxation (see [$\mathcal{R}^\star$]). The groundtruth is corrupted with *severe rotation noise*. The state-of-the-art chordal initialization [8] drives to a local minimum, whereas the result from *our initialization* looks appealing when compared to the original groundtruth. The initialization was refined using Gauss-Newton (GN). This work exploits the Lagrangian relaxation of [1] to obtain better initializations for PGO under a larger range of conditions.

Our present contribution is depicted in Fig. 2 and adheres to this same line of research. The outcome of this work will be a well-founded *recovery procedure* that exploits a recently proposed Lagrangian relaxation [1] for the PGO problem and provides, depending on the case, either a very good initial guess or even the globally optimal solution for PGO. In both cases our approach outperforms the state-of-the-art alternatives, as supported by experiments in Section V.

The following section provides a more exhaustive review of relaxations in the context of PGO and places our contribution therein.

Several additional results and proofs are given as supplementary material (suppl. material) in [13]. Lastly, we make the code implementing the procedure available at `https://github.com/jbriales/PGO-LagInit`.

MAPIR-UMA Group. Dept. Ingenieria de Sistemas y Automatica. Instituto de Investigacion Biomedica de Malaga (IBIMA). Universidad de Malaga, 29071 Malaga, Spain. E-mail: jesusbriales@uma.es

This work has been supported by a University of Malaga travel grant, the Spanish grant program FPU14/06098 and the project PROMOVE (DPI2014-55826-R), funded by the Spanish Government and the "European Regional Development Fund".

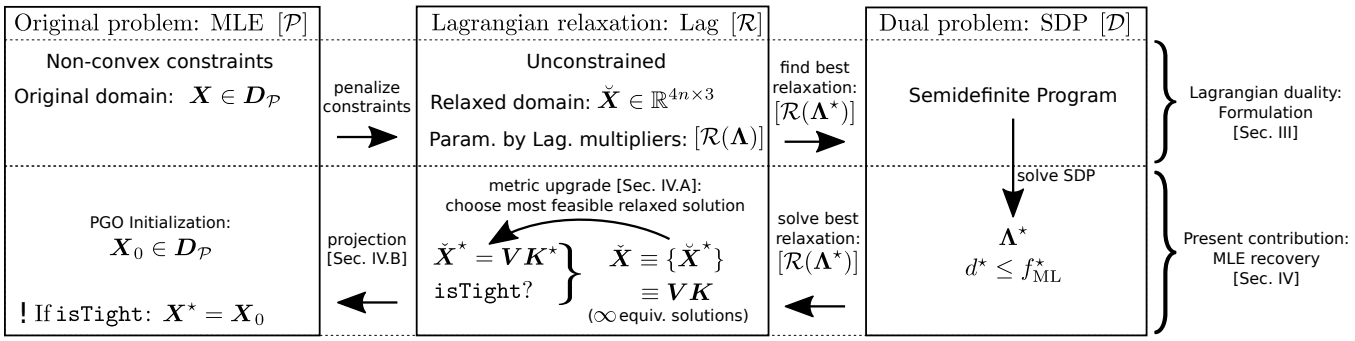


Fig. 2. Overview of the complete pipeline in this work. **Top row** (from left to right, relaxation in [1]): The original *hard* PGO problem $[\mathcal{P}]$ is relaxed to the unconstrained Lagrangian relaxation $[\mathcal{R}]$. The search for the best possible relaxation $[\mathcal{R}^\star]$ results in the dual problem $[\mathcal{D}]$. **Bottom row** (from right to left, our recovery procedure): Once the dual optimal solution $\boldsymbol{\Lambda}^\star$ is known, we obtain a parameterized family of solutions $\check{\boldsymbol{X}}$ for the best Lagrangian relaxation. We apply a *metric upgrade* (Sec. IV-A) to obtain the "most feasible" solution $\check{\boldsymbol{X}}^\star$ and finally *project* (Sec. IV-B) this onto the original domain.

II. Related work

An important and well-known example of relaxation is the *chordal relaxation* in the related (and simpler) problem of rotation synchronization, also known as linear or spectral relaxation in other fields [2], [8], [14], [15]. This relaxation basically drops the constraints in the original problem yielding a simpler linear problem (or an eigenvalue problem, depending on the variant) which is easy to solve. The chordal relaxation is tight only for trivial cases[1], but the suboptimal estimates obtained upon a simple recovery procedure are good enough to provide an initialization that works remarkably well for many cases in practice. The state-of-the-art initialization procedure for PGO [8] applies this relaxation on the rotational part of the problem to get a good initial estimate for the complete PGO problem.

A quite recent milestone in the context of PGO has been the effective exploitation of the *Lagrangian relaxation* [1], [17], [18]. In the Lagrangian relaxation the constraints are substituted by penalization terms. In order to find the *best* Lagrangian relaxation the dual problem, which consists in a Semidefinite Program (SDP) [19], needs to be solved. This results in a remarkably improved relaxation that empirically shows to be *tight* for common PGO instances with intermediate noise level.

In the 3D case, the Lagrangian relaxation was presented and used by Carlone *et al*. [18] both for performing global optimality verification and for recovering the globally optimal solution from the SDP solution, as long as the relaxation holds tight.

In our previous work [1], the Lagrangian relaxation is exploited again using a different formulation of the PGO problem. The result was a much smaller dual (SDP) problem with the same tightness properties for the relaxation. This novel relaxation was used in [1] for optimality verification purposes only, providing an important gain in efficiency and speed w.r.t. the reference method in [18].

In the present work we present an appropriate *recovery procedure* that allows us to get a feasible estimate for the original problem from the solution of the dual (SDP) problem presented in [1]. As for [18], the obtained estimate is globally optimal if the relaxation is *tight*.

More importantly, in the case the relaxation is *not tight*, our *recovery procedure* provides a good initialization even for very challenging problem instances (*e.g*. high rotational noise, low connectivity, etc.), surpassing in effectiveness the state-of-the-art chordal initialization [8]: our initialization tends to be much closer to the optimum rendering the convergence from our initial guess more robust and faster. Note that unlike the chordal initialization procedure [8], our initialization uses all the information available in the problem (both rotational and translational).

The key for the success of our procedure is a sensible heuristic founded on the theory underlying the Lagrangian relaxation. In [16], a similar heuristic was exploited in the context of 2D PGO, outperforming all current 2D alternatives in convergence success. Our *main contribution* involves the non-trivial development of the necessary framework for exploiting this same heuristic in the case of 3D PGO.

III. 3D PGO: MLE and Lagrangian relaxation

This section provides a brief review of the Maximum Likelihood Estimation formulation for 3D PGO and the corresponding Lagrangian relaxation presented in [1].

A. Quadratic MLE formulation

PGO is the problem of estimating a model consisting of n poses $(\boldsymbol{R}_i, \boldsymbol{t}_i)$ (the unknowns) from m relative measurements $(\bar{\boldsymbol{R}}_{ij}, \bar{\boldsymbol{t}}_{ij})$ (the data), both being entities in the special Euclidean group $\mathrm{SE}(3) \equiv \mathrm{SO}(3) \ltimes \mathbb{R}^3$ (for the 3D case). It is customary to associate this problem to a graph $G(V, E)$, associating the unknowns to the nodes $V = \{1, \ldots, n\}$ and the relative measurements to the edges $(i, j) \in E$.

The best model is obtained by maximizing the consistency of the modelled poses (the nodes) with the relative pose observations (the edges). Assuming the same (well-founded) *generative noise model* argued by Carlone *et al*. in [18], namely an isotropic[2] Gaussian distribution for $\bar{\boldsymbol{t}}_{ij}$ and an

[1] When the graph underlying the problem is balanced or a tree [16].

[2] Isotropy is an essential condition in the present formulation, non-isotropic observations can be approximated in a fundamented way by isotropic distributions as in [20].

isotropic[2] Langevin distribution [21] for $\bar{\boldsymbol{R}}_{ij}$, the Maximum Likelihood Estimate (MLE) problem takes the form

$$f^\star_{\text{ML}} = \min_{\{(\boldsymbol{R}_i,\boldsymbol{t}_i)\in \text{SE}(3)\}} \sum_{(i,j)\in E} \omega^2_{\bar{\boldsymbol{t}}_{ij}} \|\boldsymbol{t}_j - \boldsymbol{t}_i - \boldsymbol{R}_i \bar{\boldsymbol{t}}_{ij}\|_2^2 + \omega^2_{\bar{\boldsymbol{R}}_{ij}} \|\boldsymbol{R}_j - \boldsymbol{R}_i \bar{\boldsymbol{R}}_{ij}\|_F^2, \quad (1)$$

where the objective function $f(\cdot)$ will be referred to as the MLE (or *primal*) objective. The weights $\omega^2_{\bar{\boldsymbol{t}}_{ij}}$ and $\omega^2_{\bar{\boldsymbol{R}}_{ij}}$ are the scalar information and concentration parameter of the Gaussian and Langevin distribution, respectively. This MLE formulation is virtually equivalent to the more traditional formulations based on the absolute angular error (see [18] for details), but its *quadratic objective* makes the Lagrangian relaxation fairly simpler, yielding a Semidefinite Program (SDP) as the dual problem.

B. Compact MLE formulation

In our previous work [1] we obtained an equivalent *compact* matrix formulation for the MLE problem (1):

Problem $\mathcal{P}$ (Primal problem: Matrix form of MLE)**.**

$$f^\star_{ML} = \min_{\boldsymbol{X}} \frac{1}{2} \text{tr}(\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X}) + const. \quad (2)$$

$$s.t.\ \boldsymbol{X} = \left[\boldsymbol{R}_1, \dots, \boldsymbol{R}_n, \boldsymbol{t}_1, \dots, \boldsymbol{t}_n\right]^\top \in \boldsymbol{D}_{\mathcal{P}} \subset \mathbb{R}^{4n\times 3}. \quad (3)$$

This problem will be also referred to as the *primal* problem. Both const. ≥ 0 and a $4n \times 4n$ symmetric (positive semidefinite) matrix $\boldsymbol{M}$ gather all the data (observations $\bar{\boldsymbol{R}}_{ij}$ and $\bar{\boldsymbol{t}}_{ij}$) in the PGO problem [1, Sec. II.A]. We will refer to the set of feasible $\boldsymbol{X}$ points in the matrix reformulation (2) as the *original* or *primal domain* $\boldsymbol{D}_{\mathcal{P}}$, that is, $\boldsymbol{X} \in \boldsymbol{D}_{\mathcal{P}}$ only if the corresponding *rotation blocks* $\boldsymbol{X}_{[\boldsymbol{R}_i]}$ fulfill

$$\boldsymbol{X}^\top_{[\boldsymbol{R}_i]} \equiv \boldsymbol{R}_i \in \text{SO}(3) \iff \begin{cases} \boldsymbol{R}_i^\top \boldsymbol{R}_i = \boldsymbol{I}_3, \\ \det(\boldsymbol{R}_i) = +1. \end{cases} \quad (4)$$

No constraints apply on the translation blocks $\boldsymbol{X}_{[\boldsymbol{t}_i]}$ of $\boldsymbol{X}$, since $\boldsymbol{X}^\top_{[\boldsymbol{t}_i]} \equiv \boldsymbol{t}_i \in \mathbb{R}^3$.

C. Lagrangian relaxation and dual problem

The MLE (primal) problem [$\mathcal{P}$] is hard only because of the non-convexity in the rotational constraints, rendering the Lagrangian relaxation approach a perfect candidate to obtain a good relaxation of the problem. The Lagrangian relaxation of Problem [$\mathcal{P}$][3] provides a whole family of relaxations parameterized by the different weightings (Lagrange multipliers $\boldsymbol{\Lambda}_i$) employed. From here on, we use the notation $\mathbb{S}^n$ to denote the set of $n \times n$ symmetric matrices.

Problem $\mathcal{R}$ (Lagrangian relaxation of [$\mathcal{P}$])**.**

$$d(\boldsymbol{\Lambda}) = \min_{\breve{\boldsymbol{X}}\in\mathbb{R}^{4n\times 3}} \frac{1}{2} \text{tr}(\breve{\boldsymbol{X}}^\top \boldsymbol{M}_{\boldsymbol{\Lambda}} \breve{\boldsymbol{X}}) + \frac{1}{2} \text{tr}(\boldsymbol{\Lambda}) + const. \quad (5)$$

where $\boldsymbol{\Lambda} = \text{blkdiag}(\{\boldsymbol{\Lambda}_i\}_{i\in V}, \boldsymbol{0}_n)$, $\boldsymbol{\Lambda}_i \in \mathbb{S}^3$ *are the Lagrange multipliers for each orthogonality constraint* (4)*, and we define the* penalized cost matrix $\boldsymbol{M}_{\boldsymbol{\Lambda}} := \boldsymbol{M} - \boldsymbol{\Lambda}$*. We use* $\breve{\boldsymbol{X}}$ *to refer to a point in the* relaxed *domain* $\breve{\boldsymbol{X}} \in \mathbb{R}^{4n\times 3}$*, in comparison to points in the original domain* $\boldsymbol{X} \in \boldsymbol{D}_{\mathcal{P}}$*.*

Proof. This result is obtained in [1, Sec. III]. □

The optimal objective $d(\boldsymbol{\Lambda})$ of Problem [$\mathcal{R}$] is, by definition, a lower bound on $f^\star_{\text{ML}}$ for Problem [$\mathcal{P}$]: $d(\boldsymbol{\Lambda}) \leq f^\star_{\text{ML}}$. The search for the best possible relaxation [$\mathcal{R}$] is referred to as the dual problem or, more precisely in our case, the Lagrangian dual problem:

Problem $\mathcal{D}$ (Lagrangian dual problem of [$\mathcal{P}$])**.**

$$d^\star = const. + \max_{\boldsymbol{\Lambda}} \frac{1}{2} \text{tr}(\boldsymbol{\Lambda}), \quad s.t.\ \boldsymbol{M} - \boldsymbol{\Lambda} \succcurlyeq \boldsymbol{0}, \quad (6)$$

where $\succcurlyeq \boldsymbol{0}$ *denotes the matrix is Positive Semidefinite.*

Proof. This result is obtained in [1, Sec. III]. □

Upon close observation, this problem adopts the form of a (convex) Semidefinite Program (SDP) [19]. By definition of the dual problem it holds that $d^\star \leq d(\boldsymbol{\Lambda}) \leq f^\star_{\text{ML}}$. Note that the set of Lagrange multipliers $\boldsymbol{\Lambda}_i$ (also called *dual variables*) form the domain of the dual problem [$\mathcal{D}$].

The *best* Lagrangian relaxation [$\mathcal{R}$] is that obtained from the optimal solution $\boldsymbol{\Lambda}^\star$ of the dual problem [$\mathcal{D}$]:

Problem $\mathcal{R}^\star$ (Best Lagrangian relaxation of [$\mathcal{P}$])**.**

$$d^\star = \min_{\breve{\boldsymbol{X}}\in\mathbb{R}^{4n\times 3}} \frac{1}{2} \text{tr}(\breve{\boldsymbol{X}}^\top \boldsymbol{M}^\star_{\boldsymbol{\Lambda}} \breve{\boldsymbol{X}}) + \frac{1}{2} \text{tr}(\boldsymbol{\Lambda}^\star) + const. \quad (7)$$

where the optimal penalized cost matrix is $\boldsymbol{M}^\star_{\boldsymbol{\Lambda}} = \boldsymbol{M} - \boldsymbol{\Lambda}^\star$*.*

Note that since $d^\star = \frac{1}{2}\text{tr}(\boldsymbol{\Lambda}^\star) + \text{const.}$ (from [$\mathcal{D}$]), the optimal solution $\breve{\boldsymbol{X}}^\star$ for [$\mathcal{R}^\star$] fulfills $\text{tr}((\breve{\boldsymbol{X}}^\star)^\top \boldsymbol{M}^\star_{\boldsymbol{\Lambda}} \breve{\boldsymbol{X}}^\star) = 0$. $\boldsymbol{M}^\star_{\boldsymbol{\Lambda}} \succcurlyeq \boldsymbol{0}$ by the constraints of the dual problem [$\mathcal{D}$] so

$$\text{tr}((\breve{\boldsymbol{X}}^\star)^\top \boldsymbol{M}^\star_{\boldsymbol{\Lambda}} \breve{\boldsymbol{X}}^\star) = 0 \Rightarrow \boldsymbol{M}^\star_{\boldsymbol{\Lambda}} \breve{\boldsymbol{X}}^\star = \boldsymbol{0}_{4n\times 3}. \quad (8)$$

The relation (8) above states the equivalent condition that the optimal primal solution $\breve{\boldsymbol{X}}^\star$ belongs to the null space or kernel $\boldsymbol{\mathcal{N}} = \ker(\boldsymbol{M}^\star_{\boldsymbol{\Lambda}})$ of the optimal penalized cost matrix $\boldsymbol{M}^\star_{\boldsymbol{\Lambda}}$, which can be proved has $\text{rank}(\boldsymbol{\mathcal{N}}) > 3$.

An important property pointed in [1, Sec. IV] is:

Proposition 1 (Nullspace condition)**.** If $d^\star = f^\star_{ML}$*, that is,* if *the best Lagrangian relaxation [*$\mathcal{R}^\star$*] is* tight*, then*

$$\boldsymbol{M}^\star_{\boldsymbol{\Lambda}} \boldsymbol{X}^\star = \boldsymbol{0}_{4n\times 3} \Rightarrow \boldsymbol{X}^\star \in \ker(\boldsymbol{M} - \boldsymbol{\Lambda}^\star), \quad (9)$$

where $\boldsymbol{X}^\star$ *and* $\boldsymbol{\Lambda}^\star$ *solve [*$\mathcal{P}$*] and [*$\mathcal{D}$*], respectively.*

Proof. From duality theory [23, Sec. 5], if $d^\star = f^\star_{\text{ML}}$ then $\boldsymbol{X}^\star$, the optimal solution to [$\mathcal{P}$], is also a solution to the best Lagrangian relaxation [$\mathcal{R}^\star$]. The result (9) follows then from (8). □

The best Lagrangian relaxation [$\mathcal{R}^\star$] has the remarkable property of turning *tight* ($d^\star = f^\star_{\text{ML}}$) in a wide range of PGO problem instances, as empirically shown in [1], so the

[3] After dropping the virtually inactive determinant constraints [22].

result (9) holds in many practical instances. This will be the fundamental relation that we exploit in the subsequent section for recovering a good (frequently *globally optimal*) guess $\boldsymbol{X}_0 \in \boldsymbol{D}_{\mathcal{P}}$ for the MLE problem [$\mathcal{P}$] from the solution $\boldsymbol{\Lambda}^\star$ of the dual problem [$\mathcal{D}$].

IV. Recovery: From dual solution to MLE guess

This section focuses on the development of a *recovery procedure* $\texttt{RECOVER}(\boldsymbol{\Lambda}^\star) \mapsto \boldsymbol{X}_0$, that is, an approach that given the optimal solution $\boldsymbol{\Lambda}^\star$ of the dual problem [$\mathcal{D}$], returns a good (feasible) guess $\boldsymbol{X}_0$ for the original problem [$\mathcal{P}$]. In fact, if the best Lagrangian relaxation [$\mathcal{R}^\star$] is tight ($d^\star = f^\star_{\text{ML}}$), the obtained guess will directly be the optimal solution for [$\mathcal{P}$], that is, $\boldsymbol{X}^\star = \boldsymbol{X}_0$. This is done exploiting the *nullspace condition* (9).

In the general case where our relaxation is *not tight* ($d^\star \leq f^\star_{\text{ML}}$), the nullspace relation (9) does not hold and there is no way to recover the optimal solution of the original problem [$\mathcal{P}$] in a *certifiable* way[4]. However, to recover a good initial guess $\boldsymbol{X}_0 \in \boldsymbol{D}_{\mathcal{P}}$ for [$\mathcal{P}$] we apply a similar heuristic to Carlone *et al.* in [16, Sec. V.A.2] stemming from the nullspace condition (9): *The optimal solution $\boldsymbol{X}^\star$ should be "close" to the nullspace of $\boldsymbol{M}^\star_{\boldsymbol{\Lambda}}$*. According to this heuristic, the RECOVER procedure should seek for feasible candidates $\boldsymbol{X} \in \boldsymbol{D}_{\mathcal{P}}$ that are close to the nullspace of $\boldsymbol{M}^\star_{\boldsymbol{\Lambda}}$. We attain this in a two-step process: First, we look for the Lagrangian solution $\check{\boldsymbol{X}}^\star$ that is "most feasible" for the original problem. We will refer to this operation as a *metric upgrade*, for its resemblance with the metric upgrade proposed in [24]. Second, we project the chosen $\check{\boldsymbol{X}}^\star$ onto its "closest" feasible candidate $\boldsymbol{X}_0 \in \boldsymbol{D}_{\mathcal{P}}$. Both "close" or "most feasible" are notions that require some underlying metric to quantify the distance of a general matrix $\boldsymbol{X}$ to the feasible domain $\boldsymbol{D}_{\mathcal{P}}$ of the original problem [$\mathcal{P}$].

Feasibility metric: Given a point $\boldsymbol{X} \in \boldsymbol{D}_{\mathcal{P}}$, the constraints of the original problem apply only to the rotation blocks in $\boldsymbol{X}$: $\boldsymbol{X}_{[\boldsymbol{R}_i]}^\top \equiv \boldsymbol{R}_i \in \text{SO}(3)$. Note that

$$\boldsymbol{R}_i^\top \boldsymbol{R}_i = \boldsymbol{I}_3 \iff \|\boldsymbol{R}_i^\top \boldsymbol{R}_i - \boldsymbol{I}_3\|_F^2 = 0. \tag{10}$$

As a result a simple and convenient metric for the feasibility of a point $\boldsymbol{X}$ is[5]

$$\|\boldsymbol{X}\|_{\mathcal{P}}^2 = \frac{1}{2}\sum_{i=1}^{n} \|\boldsymbol{X}_{[\boldsymbol{R}_i]}\boldsymbol{X}_{[\boldsymbol{R}_i]}^\top - \boldsymbol{I}_3\|_F^2. \tag{11}$$

From here on, we will refer to $\|\cdot\|_{\mathcal{P}}$ as the *feasibility metric*. If this metric is zero, the point is feasible for [$\mathcal{P}$]: $\|\boldsymbol{X}\|_{\mathcal{P}} = 0 \Rightarrow \boldsymbol{X} \in \boldsymbol{D}_{\mathcal{P}}$.

Next, we briefly describe the two steps involved in the RECOVER procedure. A overview of the pipeline can be found in Algorithm 1.

[4]For details on the *optimality verification* process, see [1], [18].

[5]Note this metric does not account for the determinant constraint.

A. Metric upgrade of the nullspace

From here on, when we refer to the nullspace $\boldsymbol{\mathcal{N}}$ of $\boldsymbol{M}^\star_{\boldsymbol{\Lambda}}$, we will be considering any orthonormal basis $\boldsymbol{V} \in \mathbb{R}^{4n\times k}$ that spans this nullspace[6]. From here on, we will use $\check{\boldsymbol{X}} \equiv \check{\boldsymbol{X}}^\star$ to refer to any solution of the best Lagrangian relaxation [$\mathcal{R}^\star$]. According to (8), this family of solutions is characterized by all possible $4n\times 3$ matrices lying inside the nullspace $\boldsymbol{\mathcal{N}} = \ker(\boldsymbol{M}^\star_{\boldsymbol{\Lambda}})$, which can be parameterized as

$$\check{\boldsymbol{X}} \equiv \check{\boldsymbol{X}}(\boldsymbol{K}) = \boldsymbol{V}\boldsymbol{K} \equiv \{\check{\boldsymbol{X}}^\star\}, \quad \boldsymbol{K} \in \mathbb{R}^{k\times 3}. \tag{12}$$

Under this parameterization, the task of finding the "most feasible" point $\check{\boldsymbol{X}}$ in the nullspace for the chosen feasibility metric (11) yields the *metric upgrade problem* [24]:

$$\check{\boldsymbol{X}}^\star = \arg\min_{\check{\boldsymbol{X}}} \frac{1}{2}\|\check{\boldsymbol{X}}\|_{\mathcal{P}}^2, \quad \text{s.t. } \check{\boldsymbol{X}} = \boldsymbol{V}\boldsymbol{K}. \tag{13}$$

The rotation blocks in $\check{\boldsymbol{X}}$ can be written as $\check{\boldsymbol{X}}_{[\boldsymbol{R}_i]} = \boldsymbol{V}_{[\boldsymbol{R}_i]}\boldsymbol{K}$ where $\boldsymbol{V}_{[\boldsymbol{R}_i]}$ stands for the i-th $3\times k$ block in $\boldsymbol{V}$. For a given nullspace basis $\boldsymbol{V}$,

$$\begin{aligned}\frac{1}{2}\|\check{\boldsymbol{X}}\|_{\mathcal{P}}^2 &= \frac{1}{2}\|\boldsymbol{V}\boldsymbol{K}\|_{\mathcal{P}}^2 \\ &= \frac{1}{2}\sum_{i=1}^{n}\|\boldsymbol{V}_{[\boldsymbol{R}_i]}(\boldsymbol{K}\boldsymbol{K}^\top)\boldsymbol{V}_{[\boldsymbol{R}_i]}^\top - \boldsymbol{I}_3\|_F^2 \equiv g(\underbrace{\boldsymbol{K}\boldsymbol{K}^\top}_{k\times k}).\end{aligned} \tag{14}$$

The objective $g(\cdot)$ is referred to as the *upgrade objective*, and the equivalent metric upgrade problem we address is:

Problem Upg (Metric upgrade of $\check{\boldsymbol{X}}^\star$ in [$\mathcal{R}^\star$])**.**

$$g^\star = \min_{\boldsymbol{K}\in\mathbb{R}^{k\times 3}} g(\boldsymbol{K}\boldsymbol{K}^\top). \tag{15}$$

This is an unconstrained problem, but the upgrade objective $g(\boldsymbol{K}\boldsymbol{K}^\top)$ is a quartic polynomial[7] in the entries of $\boldsymbol{K}$. Instead, we introduce the auxiliar *lifted variable* $\boldsymbol{S} = \boldsymbol{K}\boldsymbol{K}^\top \in \mathbb{S}^k$ and address the equivalent formulation:

Problem UpgL (*Lifting* of metric upgrade problem [Upg])**.**

$$g^\star = \min_{\boldsymbol{S}\in\mathbb{S}^k} g(\boldsymbol{S}), \quad \textit{s.t. } \operatorname{rank}(\boldsymbol{S}) = 3, \quad \boldsymbol{S} \succcurlyeq \boldsymbol{0}. \tag{16}$$

The equivalent *lifted* metric upgrade problem [UpgL] is still not trivial due to the constraints affecting $\boldsymbol{S}$. However, its *linear relaxation*,

$$g^\star_{\mathcal{R}} = \min_{\boldsymbol{S}\in\mathbb{S}^k} g(\boldsymbol{S}), \tag{17}$$

is straightforward to solve as the minimization of a quadratic function[8]. If the solution $\boldsymbol{S}^\star_{\mathcal{R}}$ for this linear relaxation (17) has $3 < \operatorname{rank}(\boldsymbol{S}^\star_{\mathcal{R}}) \leq k$, $\boldsymbol{S}^\star_{\mathcal{R}}$ is not feasible for the lifted metric upgrade problem [UpgL]. In this case, we compute

[6]A basis for $\ker(\boldsymbol{M}^\star_{\boldsymbol{\Lambda}})$ is formed by the eigenvectors corresponding to the k smallest (*zero*) eigenvalues of the *positive semidefinite* matrix $\boldsymbol{M}^\star_{\boldsymbol{\Lambda}}$.

[7]The minimization of a quartic polynomial like this is a non-convex problem whose resolution is not trivial.

[8]The details about this linear problem are given in suppl. material [13].

Algorithm 1: Recovery procedure

Input: Dual solution $\boldsymbol{\Lambda}^\star$ of dual problem [6]
Output: Feasible guess $\boldsymbol{X}_0$, optimality flag `isOpt`

$\boldsymbol{V} \leftarrow \ker(\boldsymbol{M} - \boldsymbol{\Lambda}^\star)$; ▷ Compute nullspace
$\check{\boldsymbol{X}}^\star,\ g^\star \leftarrow$ `MetricUpgrade(`$\boldsymbol{V}$`)`; ▷ Sec. IV-A
$\boldsymbol{X}_0 \leftarrow$ `Project(`$\check{\boldsymbol{X}}^\star$`)`; ▷ Sec. IV-B
`isTight` $\leftarrow (g^\star == 0)$; ▷ Detect if optimal
return $\boldsymbol{X}_0$, `isOpt` $\leftarrow$ `isTight`

the rank-3 matrix $\boldsymbol{S}_0$ closest to $\boldsymbol{S}^\star_{\mathcal{R}}$ applying the matrix approximation lemma[9] and use this as an initial estimate to perform a local iterative search on the constrained problem (16). This operation can be readily performed using optimization approaches tailored to this kind of problem, *e.g.* the Manopt toolbox [25]. Further details are given in the suppl. material [13]. Note that performing local optimization on the metric upgrade problem might lead to local (suboptimal) solutions. Nevertheless, the empirical performance remains good in view of the experimental results.

Once we attain the desired optimal rank-3 estimate $\boldsymbol{S}^\star$, we decompose it into $\boldsymbol{S}^\star = \boldsymbol{K}^\star(\boldsymbol{K}^\star)^\top$ using the SVD decomposition of $\boldsymbol{S}^\star$. The product $\check{\boldsymbol{X}}^\star = \boldsymbol{V}\boldsymbol{K}^\star$ yields the seeked "most feasible" Lagrangian solution (13).

Exact MLE recovery[10]*:* An important case is that when $\operatorname{rank}(\boldsymbol{S}^\star_{\mathcal{R}}) = 3$ and $g^\star_{\mathcal{R}} = 0$, since this evidences that the Lagrangian relaxation is tight, $f^\star_{\text{ML}} = d^\star$. In this case, the upgrade output $\check{\boldsymbol{X}}^\star = \boldsymbol{V}\boldsymbol{K}^\star$ provides the globally optimal solution to the original problem [$\mathcal{P}$].

B. Projection to the feasible set

The "most feasible" candidate $\check{\boldsymbol{X}}^\star = \boldsymbol{V}\boldsymbol{K}^\star$ obtained through the metric upgrade is still not, in general, a feasible point for the original domain $\boldsymbol{D}_{\mathcal{P}}$. To fix this, we look for the point $\boldsymbol{X}_0 \in \boldsymbol{D}_{\mathcal{P}}$ closest to the chosen $\check{\boldsymbol{X}}^\star$ following the same approach as in [14, Sec. III.A]. Namely, we substitute each rotation block in $\check{\boldsymbol{X}}^\star$ by its closest rotation matrix. To account for possible issues with reflections in the solution, we also check the projection of $\check{\boldsymbol{X}}^\star\boldsymbol{J}$, where $\boldsymbol{J} = \operatorname{diag}([1, 1, -1])$, and keep the best projection as the initial guess $\boldsymbol{X}_0$.

V. Experiments

The objective of the experiments in this section is two-fold: First, we show that for low and moderate rotation noise the Lagrangian relaxation in [1] is *tight* and the proposed recovery procedure is able to recover the optimal MLE solution. Second, for PGO instances which are *non-tight* due to high rotation noise and other poor conditions of the problem we show that the initial estimate provided by our recovery procedure drives to better solutions (lower objective) than the state-of-the-art initialization in 3D PGO [8], especially in those problem regimes where the PGO problem becomes more challenging.

[9]Find details about the low-rank approximation in suppl. material [13].
[10]Find the proof for this result in suppl. material [13].

These claims above are supported by extensive Monte Carlo analysis on simulated datasets with various topologies. The conclusions were consistent regardless of the topology so we show here the results for the `grid` topology only. The results for the rest of topologies can be found in the suppl. material [13]. The `grid` scenario is equivalent to the simulation setup employed by Carlone *et al.* in [18] (see Fig. 1), so we refer interested readers to that work or our suppl. material [13] for further details on this synthetic dataset generation. The list of parameters parameterizing the generated instances are the noise level both in rotation measurements ($\sigma_{\boldsymbol{R}}$) and translation measurements ($\sigma_{\boldsymbol{t}}$), the level of connectivity in the graph given as the probability or ratio of loop closure P_c and, finally, the number of nodes in the graph n.

For each simulated PGO instance, the dual problem (6) was solved using `CVX` [26], providing the lower bound $d^\star \leq f^\star_{\text{ML}}$ on the optimal MLE objective and the [$\mathcal{D}$] solution $\boldsymbol{\Lambda}^\star$. We used the latter in our recovery procedure (Alg. 1) to obtain a feasible point $\boldsymbol{X}_0$ for the original problem [$\mathcal{P}$]. When necessary (for non-tight relaxations) we performed Gauss-Newton (GN) refinement on this initial guess. This approach is tagged as `A1`. For comparison we used also the state-of-the-art chordal initialization [2], [8] followed by local iterative refinement with GN as well. We tag this approach as `chordal`. When refining with GN, we iterated until convergence or up to a maximum of 100 iterations. The statistics were computed over 50 runs.

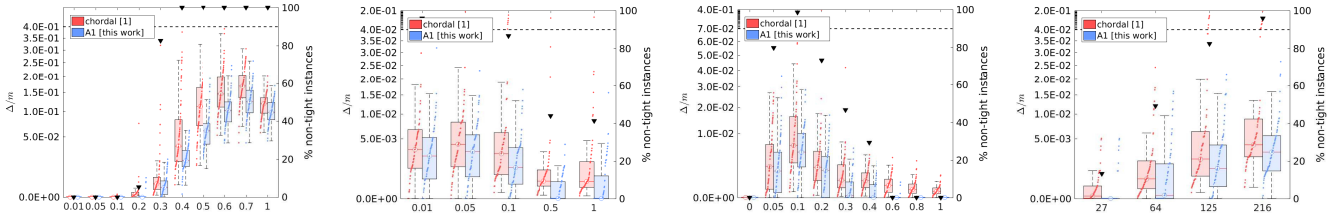
The performance of the methods is measured using the optimality gap $\Delta = \hat{f} - d^\star \geq 0$, where $\hat{f}$ is the final objective value attained by the approach. For tight problems, $f^\star = d^\star$ and $\Delta = 0$ certifies a solution is optimal. Otherwise, there is no way to certify optimality but $\Delta_a < \Delta_b$ still implies that "a" is a better estimate than "b".

The main results in Fig. 3 shows the results as a `boxplot` superimposed with a sorted display of the true underlying data. A custom square-root-based scale [27] is used to improve the visibility of the results. The upper portion of the left Y axis above the dashed line displays extreme outliers, which are significative here as a proof of occasional cases with very bad performance. The percentage of non-tight cases[11] is symbolized by ▼, in the right Y axis.

Results: First we will focus on the results under increasing rotation noise. Previous works has extensively pointed [1], [8], [16], [18] that the tightness of the Lagrangian relaxation (6) is sensitive to rotation noise above certain threshold. This trend appears in Fig. 3(a): the relaxation is tight for low noise, then tightness quickly deteriorates after $\sigma_{\boldsymbol{R}} = 0.2$ rad and from there on all the problems are just non-tight.

As for `chordal`, it reaches the optimal MLE solution for moderate noise levels, but as the problem becomes more difficult, `A1` reaches significantly lower Δ than `chordal`, especially for $\sigma_{\boldsymbol{R}}$ between 0.3 and 0.5 rad. This suggests the heuristic at the core of `A1` holds for a certain range of

[11]Optimality (in the tight case) is returned as a flag by Algorithm 1.



(a) rotation noise $\sigma_{\boldsymbol{R}}$ (b) translation noise $\sigma_{\boldsymbol{t}}$ (c) connectivity P_c (d) graph size n

Fig. 3. Optimality gap and average tightness vs. *grid* parameters. The default set of parameters is $\sigma_{\boldsymbol{R}} = 0.3$ rad, $\sigma_{\boldsymbol{t}} = 0.1$ m, $P_c = 0.2$, and $n = 125$. Left axis shows Δ after GN refinement. The points ▼ refer to the percentage of non-tight cases (right axis).

problems wich go from moderate to high rotation noise.

For the rest of analyzed parameters, we fixed the rotation noise at $\sigma_{\boldsymbol{R}} = 0.3$ rad. At this critical point, the results show that all the parameters affect the tightness, and again the performance for `A1` surpassed that of `chordal`.

Comparison with SDP relaxation in [18]: The recovery procedure for the Lagrangian relaxation in [18] does not make any special consideration about the non-tight casuistic, often driving to poor initialization. For tight cases, both recovery procedures provide the same optimal result. However, as shown in [1, Fig. 3], our dual problem [$\mathcal{D}$] is much faster to solve *by conventional interior point methods*.

Scalability issues: At the time of writing of this work, the solution of the dual problem [$\mathcal{D}$] by interior point solvers was the main computational burden in the pipeline. These present serious scalability issues, and thus prevented the application of our Algorithm 1 on large-scale datasets [8].

Luckily, *very fast* state-of-the-art solvers for the dual problem [$\mathcal{D}$] exploiting the low-rank structure for our problem have appeared since the original submission of this work that unties the potential of the proposed procedures for application in virtually any PGO instance regardless of its size [28]. Future work will include further evaluation on large-scale problem exploring the use of these solvers.

VI. Conclusions

This paper proposes an effective *recovery procedure* that exploits the state-of-the-art formulation of the Lagrangian relaxation for PGO [1] to return the globally optimal MLE solution if the relaxation is tight, or a remarkably good initialialization for non-tight relaxations. In both cases our results match or surpass those of state-of-the-art alternatives, specially in problem instances with severe noise regimes.

The performed experiments give some interesting hints about the relation between the PGO problem and different features of the graph data. We leave as future work exploring this aspect and exploiting better suited SDP solvers to apply these techniques in difficult large-scale real problems too.

References

[1] J. Briales and J. González-Jiménez, "Fast Global Optimality Verification in 3D SLAM," in *Int. Conf. on Intelligent Robots and Systems (IROS)*, IEEE/RSJ, 2016.

[2] D. Martinec and T. Pajdla, "Robust rotation and translation estimation in multiview reconstruction," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1–8, IEEE, 2007.

[3] M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri, "Global motion estimation from point matches," in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pp. 81–88, IEEE, 2012.

[4] R. Gomez-Ojeda, J. Briales, E. Fernandez-Moral, and J. Gonzalez-Jimenez, "Extrinsic calibration of a 2d laser-rangefinder and a camera based on scene corners," in *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 3611–3616, 2015.

[5] J. Briales and J. Gonzalez-Jimenez, "A minimal solution for the calibration of a 2D laser-rangefinder and a camera based on scene corners," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pp. 1891–1896, sep 2015.

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

[7] B. Vlasic and N. E. Boudette, "As U.S. investigates fatal Tesla crash, com- pany defends Autopilot system."

[8] 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 *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 4597–4604, IEEE, 2015.

[9] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," *2011 IEEE International Conference on Robotics and Automation*, pp. 3607–3613, may 2011.

[10] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The International Journal of Robotics Research*, 2011.

[11] D. M. Rosen, M. Kaess, and J. J. Leonard, "RISE: An incremental trust-region method for robust online sparse least-squares estimation," *Robotics, IEEE Transactions on*, vol. 30, no. 5, pp. 1091–1108, 2014.

[12] A. Bandeira, "A note on probably certifiably correct algorithms," *Comptes Rendus Mathematique*, 2016.

[13] J. Briales and J. Gonzalez, "Initialization of 3D Pose Graph Optimization using Lagrangian duality - Supplementary Material." Online. Available: `mapir.isa.uma.es/jbriales/ICRA17-suppl.pdf`.

[14] N. Boumal, A. Singer, and P.-A. Absil, "Robust estimation of rotations from relative measurements by maximum likelihood," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pp. 1156–1161, 2013.

[15] O. Ozyesil, N. Sharon, and A. Singer, "Synchronization over Cartan motion groups via contraction," *arXiv:1612.00059*, nov 2016.

[16] L. Carlone, G. C. Calafiore, C. Tommolillo, and F. Dellaert, "Planar Pose Graph Optimization: Duality, Optimal Solutions, and Verification," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 545–565, 2016.

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

[18] L. Carlone, D. M. Rosen, G. Calafiore, J. J. Leonard, and F. Dellaert, "Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pp. 125–132, sep 2015.

[19] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[20] D. M. Rosen, C. DuHadway, and J. J. Leonard, "A convex relaxation for approximate global optimization in simultaneous localization and mapping," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 5822–5829, IEEE, 2015.

[21] A. Chiuso, G. Picci, S. Soatto, and Others, "Wide-sense estimation on the special orthogonal group," *Communications in Information {&} Systems*, vol. 8, no. 3, pp. 185–200, 2008.

[22] R. Tron, D. M. Rosen, and L. Carlone, "On the Inclusion of Determinant Constraints in Lagrangian Duality for 3D SLAM," in *Robotics: Science and Systems, Workshop "The Problem of Mobile Sensors"*, 2015.

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

[24] R. Tron, X. Zhou, and K. Daniilidis, "A Survey on Rotation Optimization in Structure From Motion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 77–85, 2016.

[25] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1455–1459, 2014.

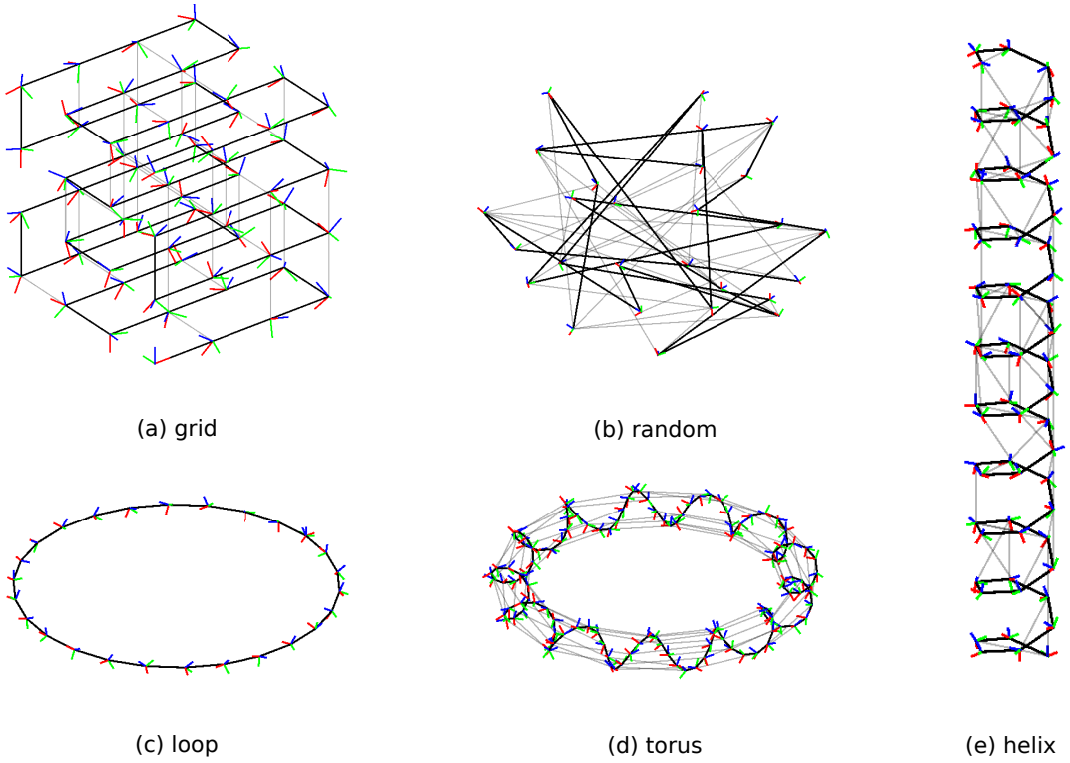
[26] M. Grant and S. Boyd, "{CVX}: Matlab Software for Disciplined Convex Programming, version 2.1." \url{http://cvxr.com/cvx}, 2014.

[27] C. Kleiber and A. Zeileis, "Visualizing Count Data Regressions Using Rootograms," may 2016.

[28] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "A Certifiably Correct Algorithm for Synchronization over the Special Euclidean Group," *International Workshop on the Algorithmic Foundations of Robotics*, nov 2016.

Initialization of 3D Pose Graph Optimization using Lagrangian duality Supplementary material

Jesus Briales Javier Gonzalez-Jimenez



(a) grid (b) random (c) loop (d) torus (e) helix

Fig. 1: Different topologies considered in the experiments.

I. Experiments: All topologies

In this section we show more experimental results supporting our claims *in the main document*, namely the superiority of our PGO initialization approach, tagged as `A1`, w.r.t. the state-of-the-art initialization in 3D PGO [1], tagged as `chordal`.

In order to have a more extensive testing framework, we considered various topologies for the Monte Carlo analysis: `grid`, `random`, `loop`, `helix` and `torus`. Next, each topology is described and we show the performance results considering different values for the parameters that affect the problem instances: Rotation noise $\sigma_{\boldsymbol{R}}$, translation noise $\sigma_{\boldsymbol{t}}$, level of connectivity[1] P_c and number of nodes n in the graph. Note that the default parameters chosen for each topology may vary (the specific values are given in the figure captions): we focus in the range of problems where strong duality begins to fail, since those are the cases of interest where common approaches break.

The statistics were computed over 50 runs.

[1]Connectivity was ignored for the `loop` scenario, as for that case connectivity is fixed to *a single* loop closure.

Figure format: The experimental results are shown in Figures 2 to 7. Each subfigure depicts the *mean* suboptimality gap (that is, the suboptimality gap Δ normalized by the number of measurements m) on the left Y axis and the percentage of non-tight cases in the right Y axis.

For the suboptimality gap values $\frac{\Delta}{m}$, we plot a simplified statistical description of the results by means of `boxplot`. We also display, for completeness, the whole set of values superimposed on the space of the corresponding box plot in a sorted fashion (in ascending order).

To improve the visibility of the results and to make comparisons easier we employed a non-linear square-root scale in the left Y axis (the square root of the results is shown). This scale gives more weight to values nearby 0 and shrinks values farther away from 0, and has been successfully used for visualization purposes *e.g.* in [2]. Furthermore, to present extreme outliers, which are significative here as a proof of occasional cases with very bad performance, we apply a much smaller scale the upper portion of the image above the dashed line.

A. `grid` scenario

The `grid` scenario consists of a 3D regular cubic grid (each grid cell with side 1m), which is traversed producing an odometric trajectory while loop closures are added with probability P_c between nearby nodes. See Fig. 1a.

The statistics for the `grid` scenario appear in Fig. 3. An example with a challenging case is shown in Fig. 8.

B. `random` scenario

To generate a `random` scenario we chose n random poses lying inside a $10\text{m} \times 10\text{m} \times 10\text{m}$ cube. Then a spanning path of the pose graph is taken as odometric trajectory, and further edges are added between any two nodes with a probability P_c to simulate loop closures. See Fig. 1b.

The statistics for the `random` scenario appear in Fig. 4.

Note that, in this scenario, any node provides a candidate loop closure with any other node in the graph. This makes the connectivity grow very quickly. Because of this, in Fig. 2 we focus in the range of cases where connectivity is bounded by a notably low P_c.

C. `loop` scenario

The `loop` scenario allocates poses uniformly spaced along a circumference, assigning random orientation to the poses. The circumference radius is chosen so that the distance between consecutive poses is 1m.

Then the odometric trajectory follows the expectable path in the loop, and a *single* loop closure edge is added between the last and first nodes. This makes the `loop` scenario a special case as it has the minimum possible connectivity producing a non-trivial problem. See Fig. 1c.

The statistics for the `loop` scenario appear in Fig. 5.

D. `helix` scenario

In the `helix` scenario the poses were uniformly spread along an helicoidal trajectory, with random orientation. The torus major radius was 4m and the minor radius 0.5m. was characterized by a radius of 1m and a slope of 15 deg.

The odometric trajectory follows the helicoidal path. Again, loop closure edges are added between nearby nodes only, with probability P_c. See Fig. 1e.

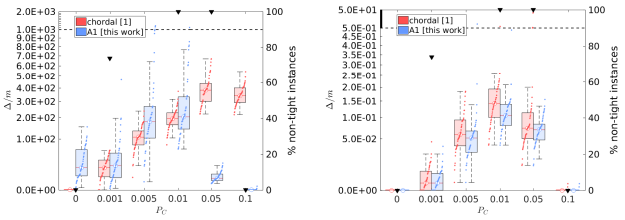
The statistics for the `helix` scenario appear in Fig. 6. An example with a challenging case is shown in Fig. 9.

E. `torus` scenario

Finally, in the `torus` scenario, the poses (with random orientation) were uniformly spread along a closed helicoidal trajectory along the torus surface.

The odometric trajectory was assigned to the helicoidal path, and random loop closures were added between nearby nodes, with probability P_c. See Fig. 1d.

The statistics for the `torus` scenario appear in Fig. 7. An example with a challenging case is shown in Fig. 10.



(a) init - connectivity (b) refined - connectivity

Fig. 2: Optimality gap and average tightness for the `random` scenario within the challenging range of connectivity values. The rest of parameters are $\sigma_{\boldsymbol{R}} = 0.6$ rad, $\sigma_{\boldsymbol{t}} = 0.1$ m and $n = 64$. The top row shows Δ_0 for the initial estimates. The bottom row shows Δ for the refined estimates. The points ▼ refer to the percentage of non-tight cases (right Y axis).

APPENDIX I
CALCULUS: METRIC UPGRADE IN MANOPT

The metric upgrade problem *in the main document* was finally reduced to a problem of unconstrained optimization in the manifold of rank-3, symmetric $k \times k$ positive semidefinite (PSD) matrices, $\mathbb{S}_+(k, 3)$:

$$g^\star = \min_{\boldsymbol{S} \in \mathbb{S}_+(k,3)} g(\boldsymbol{S}). \tag{1}$$

We use the Manopt toolbox [3] for the optimization, where the elements of $\boldsymbol{S} \in \mathbb{S}_+(k, 3)$ are parameterized in the `symfixedrankYYfactory` factory by full column rank matrices $\boldsymbol{Y} \in \mathbb{R}_*^{k \times 3}$ as $\boldsymbol{S} = \boldsymbol{Y}\boldsymbol{Y}^\top$.

The optimization objective is

$$g(\boldsymbol{S}) = \frac{1}{2}\boldsymbol{s}^\top \boldsymbol{A}\boldsymbol{s} - \boldsymbol{b}^\top \boldsymbol{s} + \frac{3n}{2}, \quad \boldsymbol{s} = \operatorname{vecs}(\boldsymbol{S}). \tag{2}$$

We use the second-order Riemannian trust-region method implemented in Manopt. This requires the gradient of the objective function $h(\boldsymbol{Y}) = g(\boldsymbol{Y}\boldsymbol{Y}^\top)$, which is obtained by the chain rule:

$$\frac{\partial h(\boldsymbol{Y})}{\partial \boldsymbol{Y}} = \frac{\partial g(\boldsymbol{s})}{\partial \boldsymbol{s}} \cdot \frac{\partial \operatorname{vecs}(\boldsymbol{S})}{\partial \operatorname{vec}(\boldsymbol{S})} \cdot \frac{\partial \operatorname{vec}(\boldsymbol{S})}{\partial \boldsymbol{Y}}. \tag{3}$$

The first derivative comes from the optimization objective,

$$\frac{\partial g(\boldsymbol{s})}{\partial \boldsymbol{s}} = (\boldsymbol{A}\boldsymbol{s} - \boldsymbol{b})^\top. \tag{4}$$

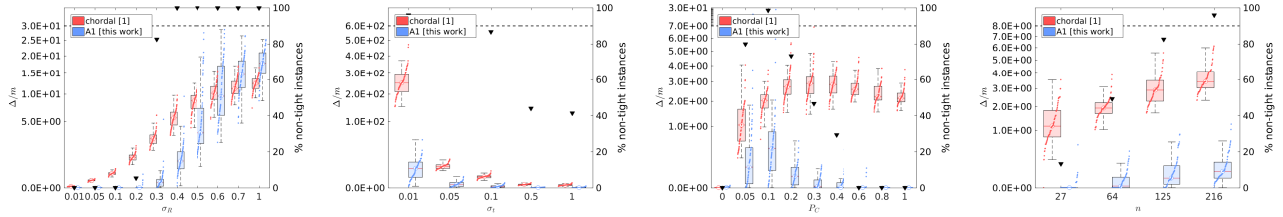
The second is the projection matrix $\boldsymbol{P} \in \mathbb{R}^{\frac{1}{2}k(k+1) \times k^2}$ between the vectorization operators, defined in [4], that fulfills

$$\operatorname{vecs}(\boldsymbol{S}) = \boldsymbol{P} \operatorname{vec}(\boldsymbol{S}). \tag{5}$$

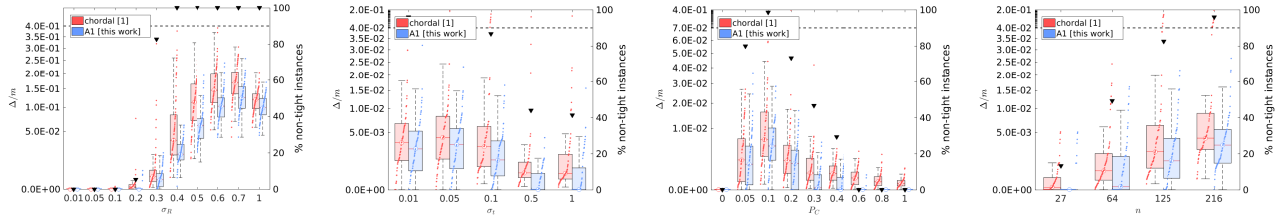
Finally, we apply the properties for vectorization and its derivatives available in [5] to get the third derivative:

$$\frac{\partial \operatorname{vec}(\boldsymbol{Y}\boldsymbol{Y}^\top)}{\partial \operatorname{vec}(\boldsymbol{Y})} = (\boldsymbol{I}_{k^2} + \boldsymbol{T}_{k,k})(\boldsymbol{Y} \otimes \boldsymbol{I}_k). \tag{6}$$

Here $\boldsymbol{T}_{k,k}$ is the $k^2 \times k^2$ permutation matrix that fulfills $\boldsymbol{T}_{k,k} \operatorname{vec} \boldsymbol{A} = \operatorname{vec} \boldsymbol{A}^\top$.

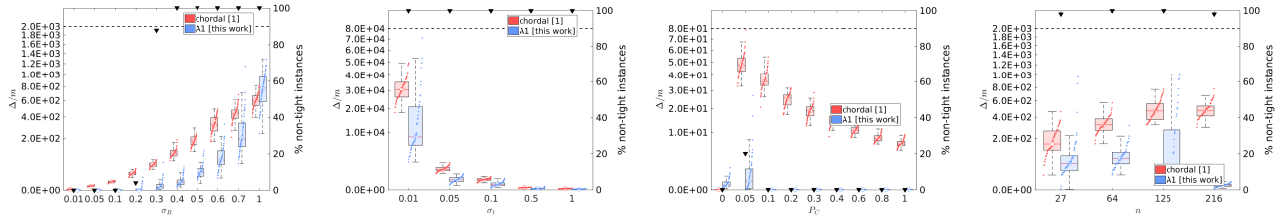


(a) init - rotation noise (b) init - translation noise (c) init - connectivity (d) init - graph size

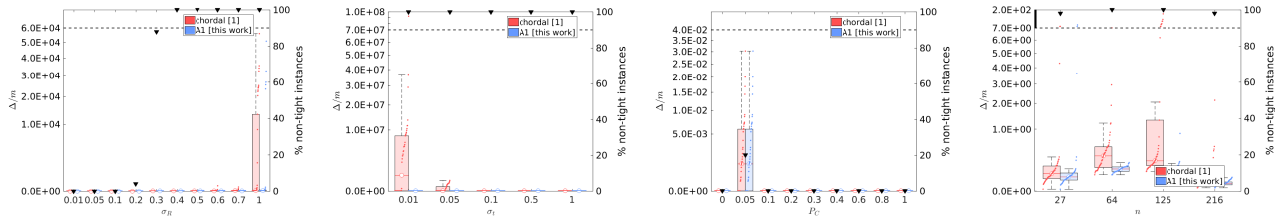


(e) refined - rotation noise (f) refined - translation noise (g) refined - connectivity (h) refined - graph size

Fig. 3: Optimality gap and average tightness vs. *grid* parameters. The default set of parameters is $\sigma_{\boldsymbol{R}} = 0.3$ rad, $\sigma_{\boldsymbol{t}} = 0.1$ m, $P_c = 0.2$, and $n = 64$. The top row shows Δ_0/m for the initial estimates. The bottom row shows Δ/m for the refined estimates. The points ▼ refer to the percentage of non-tight cases (right Y axis).



(a) init - rotation noise (b) init - translation noise (c) init - connectivity (d) init - graph size



(e) refined - rotation noise (f) refined - translation noise (g) refined - connectivity (h) refined - graph size

Fig. 4: Optimality gap and average tightness vs. *random* parameters. The default set of parameters is $\sigma_{\boldsymbol{R}} = 0.6$ rad, $\sigma_{\boldsymbol{t}} = 0.1$ m, $P_c = 0.02$, and $n = 64$. The top row shows Δ_0/m for the initial estimates. The bottom row shows Δ/m for the refined estimates. The points ▼ refer to the percentage of non-tight cases (right Y axis).

REFERENCES

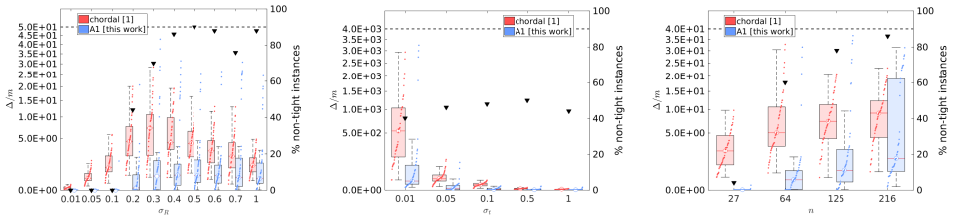
[1] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert, "Initialization techniques for 3D SLAM: a survey on rotation estimation and its use in pose graph optimization," in *Robot. Autom. (ICRA), 2015 IEEE Int. Conf.*, pp. 4597–4604, IEEE, 2015.

[2] C. Kleiber and A. Zeileis, "Visualizing Count Data Regressions Using Rootograms," may 2016.

[3] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab toolbox for optimization on manifolds," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1455–1459, 2014.

[4] K. Schäcke, "On the kronecker product," 2013.

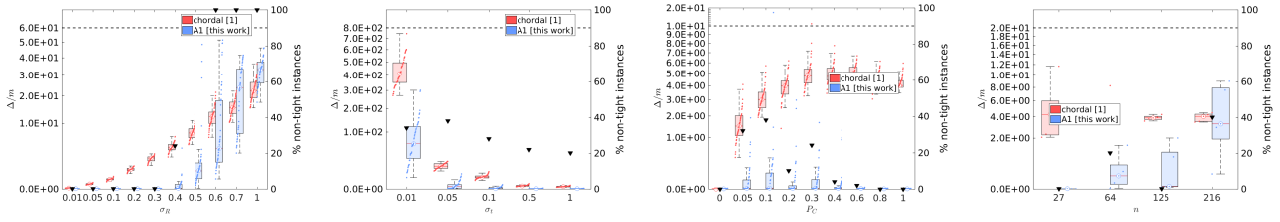
[5] P. L. Fackler, "Notes on Matrix Calculus," pp. 1–14, 2005.



(a) init - rotation noise (b) init - translation noise (c) init - graph size

(d) refined - rotation noise (e) refined - translation noise (f) refined - graph size

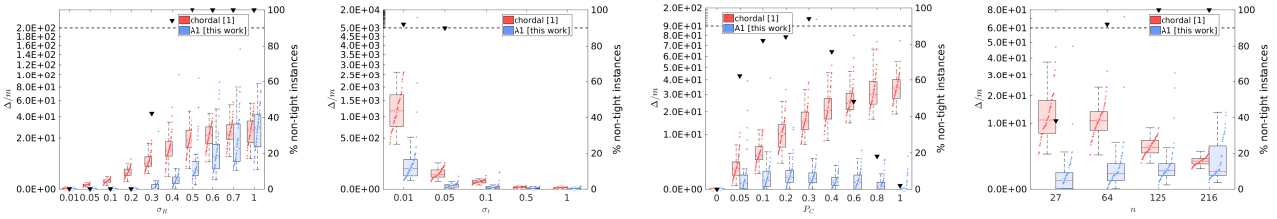
Fig. 5: Optimality gap and average tightness vs. *loop* parameters. The default set of parameters is $\sigma_{\boldsymbol{R}} = 0.2$ rad, $\sigma_{\boldsymbol{t}} = 0.1$ m and $n = 64$. The top row shows Δ_0 for the initial estimates. The bottom row shows Δ for the refined estimates. The points ▼ refer to the percentage of non-tight cases (right Y axis).



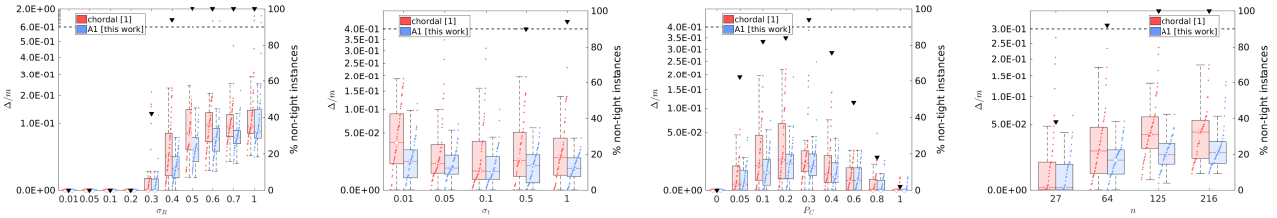
(a) init - rotation noise (b) init - translation noise (c) init - connectivity (d) init - graph size

(e) refined - rotation noise (f) refined - translation noise (g) refined - connectivity (h) refined - graph size

Fig. 6: Optimality gap and average tightness vs. *helix* parameters. The default set of parameters is $\sigma_{\boldsymbol{R}} = 0.4$ rad, $\sigma_{\boldsymbol{t}} = 0.1$ m, $P_c = 0.2$, and $n = 64$. The top row shows Δ_0/m for the initial estimates. The bottom row shows Δ/m for the refined estimates. The points ▼ refer to the percentage of non-tight cases (right Y axis).

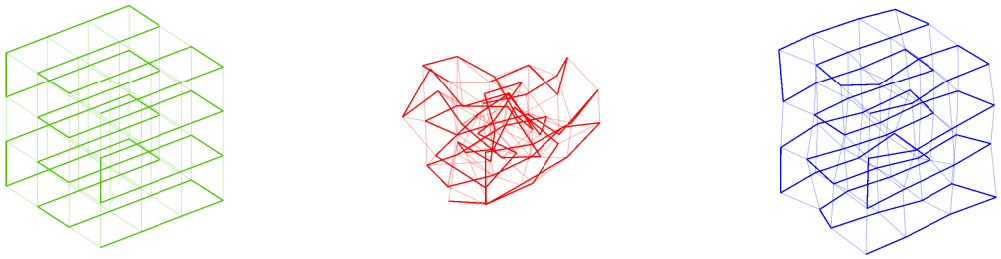


(a) init - rotation noise (b) init - translation noise (c) init - connectivity (d) init - graph size



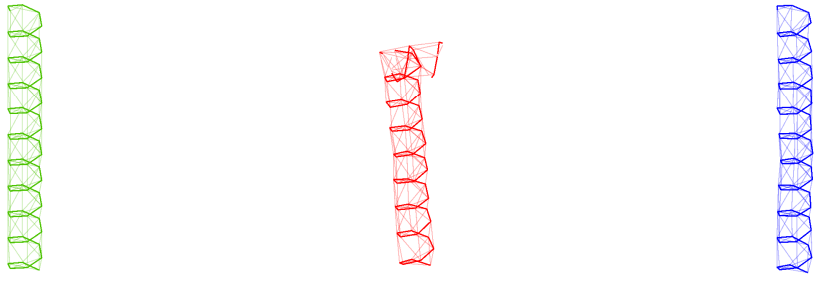
(e) refined - rotation noise (f) refined - translation noise (g) refined - connectivity (h) refined - graph size

Fig. 7: Optimality gap and average tightness vs. *torus* parameters. The default set of parameters is $\sigma_{\boldsymbol{R}} = 0.4$ rad, $\sigma_{\boldsymbol{t}} = 0.1$ m, $P_c = 0.2$, and $n = 64$. The top row shows Δ_0/m for the initial estimates. The bottom row shows Δ/m for the refined estimates. The points ▼ refer to the percentage of non-tight cases (right Y axis).



(a) Groundtruth (b) chordal + GN (c) ours + GN

Fig. 8: Challenging `grid` instance. The refinement from `chordal` initialization (b) clearly failed whereas from our `A1` initialization (c) we got visually appealing results.



(a) Groundtruth (b) chordal + GN (c) ours + GN

Fig. 9: Challenging `helix` instance. The refinement from `chordal` initialization (b) clearly failed whereas from our `A1` initialization (c) we got visually appealing results.

(a) Groundtruth (b) chordal + GN (c) ours + GN

Fig. 10: Challenging `torus` instance. The refinement from `chordal` initialization (b) clearly failed whereas from our `A1` initialization (c) we got visually appealing results.