Generalized Proximal Methods for Pose Graph Optimization

Taosha Fan[1] and Todd Murphey[1]

Northwestern University, 2145 Sheridan Rd, Evanston, IL 60208,
taosha.fan@u.northwestern.edu, t-murphey@northwestern.edu

Abstract. In this paper, we generalize proximal methods that were originally designed for convex optimization on normed vector space to non-convex pose graph optimization (PGO) on special Euclidean groups, and show that our proposed generalized proximal methods for PGO converge to first-order critical points. Furthermore, we propose methods that significantly accelerate the rates of convergence almost without loss of any theoretical guarantees. In addition, our proposed methods can be easily distributed and parallelized with no compromise of efficiency. The efficacy of this work is validated through implementation on simultaneous localization and mapping (SLAM) and distributed 3D sensor network localization, which indicate that our proposed methods are a lot faster than existing techniques to converge to sufficient accuracy for practical use.

1 Introduction

Pose graph optimization (PGO) estimates a number of unknown poses from noisy relative measurements, in which we associate each pose with a vertex and each measurement with an edge of a graph. PGO has important applications in a number of areas, for example, simultaneous localization and mapping (SLAM) in robotics [1], structural analysis of biological macromolecules in cryo-electron microscopy [2], sensor network localization in distributed sensing [3], etc.

In the last twenty years, a number of PGO methods have been developed, which are either first-order optimization methods [3–5] or second-order optimization methods [6–9]. In general, first-order PGO methods typically converge slowly when close to critical points, and thus, second-order PGO methods are preferable in most applications. In spite of this, second-order PGO methods have to continuously solve linear systems to evaluate descent directions, which is difficult to distribute and parallelize, and can be time-consuming for large-scale optimization problems [6–10].

In optimization and applied mathematics, there are a number of algorithms to accelerate the rates of convergence of first-order optimization methods [11–14]. Nevertheless, most of existing accelerated first-order optimization methods [11–14] rely on proximal methods [15] and need a proximal operator that is also an upper bound of the objective function, and such a proximal operator, though exists, it is usually unclear for PGO. In addition, it is common in first-order PGO methods to formulate PGO as optimization on special Euclidean groups, and update pose estimates using Riemannian instead of Euclidean gradients [3–5], whereas in general accelerated first-order optimization methods [11–14] only apply to optimization on normed vector space and are inapplicable for optimization using Riemannian gradients. As a result, it is in strictly limited to accelerate existing first-order PGO methods [3–5] with [11–14].

In this paper, we generalize proximal methods [15] that were originally designed for convex optimization on normed vector space to non-convex PGO on special Euclidean groups, and show that our proposed methods converge to first-order critical points. Different from existing first-order PGO methods [3–5], our proposed methods do not rely on Riemannian gradients to update pose estimates and there is no need to perform line search to guarantee convergence. Instead, our proposed methods update pose estimates by solving optimization sub-problems in closed form. Furthermore, we present methods that significantly accelerate the rates of convergence using [11, 12] with no loss of theoretical guarantees. To our knowledge, neither proximal methods nor accelerated first-order methods for PGO have been presented before. In addition, our proposed methods can be easily distributed and parallelized without compromise of efficiency. In spite of being first-order PGO methods, our proposed methods are empirically several times faster than second-order PGO methods to converge to modest accuracy that is sufficient for practical use. In cases when higher accuracy is required, our proposed methods can be combined with second-order PGO methods [6–10] to improve the overall performance.

The rest of this paper is organized as follows. Section 2 introduces notations that are used throughout this paper. Section 3 reformulates proximal methods in a more general way that is used in this paper to solve PGO. Section 4 formulates and simplifies PGO. Section 5 proposes a generalized proximal operator that is also an upper bound of PGO, which is fundamental to our proposed methods. Sections 6 and 7 present unaccelerated and accelerated generalized proximal methods for PGO, respectively, which is the major contribution of this paper. Section 8 implements our proposed methods on SLAM and distributed 3D sensor network localization, and makes comparisons with $\mathsf{SE-Sync}$ [9]. The conclusions are made in Section 9

2 Notation

$\mathbb{R}$ denotes the sets of real numbers; $\mathbb{R}^{m\times n}$ and $\mathbb{R}^n$ denote the sets of $m\times n$ matrices and $n\times 1$ vectors, respectively; and $SO(d)$ and $SE(d)$ denote the sets of special orthogonal groups and special Euclidean groups, respectively. For a matrix $X\in\mathbb{R}^{m\times n}$, the notation $[X]_{ij}$ denotes the (i,j)-th entry or (i,j)-th block of X. The notation $\|\cdot\|$ denotes the Frobenius norm of matrices and vectors. For symmetric matrices $Y, Z\in\mathbb{R}^{n\times n}$, $Y\succeq Z$ (or $Z\preceq Y$) and $Y\succ Z$ (or $Z\prec Y$) mean that $Y-Z$ is positive semidefinite and positive definite, respectively. If $F:\mathbb{R}^{m\times n}\to\mathbb{R}$ is a function, $\mathcal{M}\subset\mathbb{R}^{m\times n}$ is a manifold and $X\in\mathcal{M}$, the notation $\nabla F(X)$ and $\operatorname{grad} F(X)$ denote the Euclidean and Riemannian gradients, respectively.

3 Generalized Proximal Methods

For an optimization problem

$$\min_{X\in\mathcal{X}} F(X),$$

in which $\mathcal{X}$ is a closed set and $F : \mathcal{X} \to \mathbb{R}$ is a function with Lipschitz smooth gradient $\nabla F(X)$ for a scalar $L > 0$ such that

$$F(Y) \leq F(X) + \nabla F(X)^\top (Y - X) + \frac{L}{2}\|Y - X\|^2, \tag{1}$$

then the proximal operator of the first-order approximation at $X^{(k)} \in \mathcal{X}$ is defined to be [15]

$$X^{(k+1)} = \arg\min_{X \in \mathcal{X}} F(X^{(k)}) + \nabla F(X^{(k)})^\top \big(X - X^{(k)}\big) + \frac{L}{2}\|X - X^{(k)}\|^2, \tag{2}$$

from which it can be concluded that $F(X^{(k+1)}) \leq F(X^{(k)})$. An optimization algorithm using Eq. (2) to generate iterates $\{X^{(k)}\}$ is called the proximal method. Though originally designed for convex optimization [11, 12, 15], proximal methods have been used to solve non-convex optimization problems and get quite good results [13, 14].

From Eq. (2), a prerequisite of proximal methods is that there exists a positive scalar L w.r.t. which $F(X)$ is Lipschitz smooth. In most cases, L is unknown, and finding such a scalar L can be time-consuming. Instead, if there exists a positive definite matrix Ω such that

$$F(Y) \leq F(X) + \nabla F(X)^\top (Y - X) + \frac{1}{2}(Y - X)^\top \Omega (Y - X), \tag{3}$$

we obtain an first-order approximation that is also an upper bound of $F(X)$ as

$$X^{(k+1)} = \arg\min_{X \in \mathcal{X}} F(X^{(k)}) + \nabla F(X^{(k)})^\top \big(X - X^{(k)}\big) + \\ \frac{1}{2}(X - X^{(k)})^\top \Omega (X - X^{(k)}). \tag{4}$$

We term Eq. (4) as the generalized proximal operator and an optimization algorithm using the equation above to generate iterates $\{X^{(k)}\}$ as the generalized proximal method. For a number of optimization problems, finding a matrix Ω satisfying Eq. (3) is much easier than finding a scalar L satisfying Eq. (1). Even though it is possible to determine a scalar $L > 0$ satisfying Eq. (1) as the greatest eigenvalue of Ω, it is still expected that Eq. (4) results in a better approximation and a tighter upper bound than Eq. (2).

In the following sections, we will propose generalized proximal methods using Eq. (4) to solve PGO.

4 Problem Formulation

In this section, we review PGO that can be formulated as a least-square optimization problem and be simplified to a compact quadratic form. It should be noted that both the least-square and quadratic formulations of PGO have been well addressed by Rosen *et. al* in [9], and due to space limitations, we only present the main results and interested readers can refer to [9] for a detailed introduction.

PGO estimates n unknown poses $g_i \triangleq (R_i, t_i) \in SE(d)$ with m noisy measurements of relative poses $g_{ij} \triangleq g_i^{-1} g_j \triangleq (R_{ij}, t_{ij}) \in SE(d)$. In PGO, the n poses g_i and

m relative measurements g_{ij} are described through a directed graph $\overrightarrow{G} \triangleq (\mathcal{V}, \overrightarrow{\mathcal{E}})$ in which $\mathcal{V} \triangleq \{1, \cdots, n\}$ and each index i is associated with g_i, and $(i, j) \in \overrightarrow{\mathcal{E}} \subset \mathcal{V}\times\mathcal{V}$ if and only if g_{ij} exists. If we ignore the orientation of edges in $\mathcal{E}$, an undirected graph $G \triangleq (\mathcal{V}, \mathcal{E})$ is obtained. In the rest of this paper, it is assumed that $\overrightarrow{G}$ is weakly connected and G is connected. Following [9], we also assume that the m measurements (R_{ij}, t_{ij}) are random variables:

$$\tilde{t}_{ij} = \underline{t}_{ij} + t_{ij}^{\epsilon} \qquad t_{ij}^{\epsilon} \sim \mathcal{N}(\mathbf{0}, \tau_{ij}^{-1}\mathbf{I}), \tag{5a}$$

$$\widetilde{R}_{ij} = \underline{R}_{ij} R_{ij}^{\epsilon} \qquad R_{ij}^{\epsilon} \sim \mathcal{L}(\mathbf{I}, \kappa_{ij}), \tag{5b}$$

in which $\underline{g}_{ij} \triangleq (\underline{R}_{ij}, \underline{t}_{ij}) \in SE(d)$ is the true (latent) value of g_{ij}, and $\mathcal{N}(\mu_t, \Sigma_t)$ denotes the normal distribution with mean $\mu_t \in \mathbb{R}^d$ and covariance $0 \preceq \Sigma_t \in \mathbb{R}^{d\times d}$, and $\mathcal{L}(\mu_R, \kappa_R)$ denotes the isotropic Langevin distribution with mode $\mu_R \in SO(d)$ and concentration parameter $\kappa_R \geq 0$.

From the perspective of maximum likelihood estimation [9], PGO can be formulated as a least square optimization problem on $SE(d)^n$

$$\min_{\substack{R_i\in SO(d),\, t_i \in R^d,\\ i=1,\cdots,n}} \sum_{(i,j)\in\overrightarrow{\mathcal{E}}} \frac{1}{2}\left(\kappa_{ij}\cdot\|R_i\widetilde{R}_{ij} - R_j\|^2 + \tau_{ij}\cdot\|R_i\tilde{t}_{ij} + t_i - t_j\|^2\right), \tag{6}$$

in which $R_i \in SO(d)$ and $t_i \in \mathbb{R}^d$. A straightforward derivation further simplifies Eq. (6) to

$$\min_{X\in\mathbb{R}^{d\times n}\times SO(d)^n} F(X) \triangleq \frac{1}{2}\operatorname{trace}(X\widetilde{M}X^\top) \tag{7}$$

in which $F(X)$ is a quadratic function and $X \triangleq \begin{bmatrix} t_1 \cdots t_n \; R_1 \cdots R_n\end{bmatrix} \in \mathbb{R}^{d\times n} \times SO(d)^n \subset \mathbb{R}^{d\times(d+1)n}$. For $F(X)$ of Eq. (7), $\widetilde{M} \in \mathbb{R}^{(d+1)n\times(d+1)n}$ is a positive-semidefinite matrix

$$\widetilde{M} \triangleq \begin{bmatrix} L(W^\tau) & \widetilde{V} \\ \widetilde{V}^\top & L(\widetilde{G}^\rho) + \widetilde{\Sigma}\end{bmatrix} \in \mathbb{R}^{(d+1)n\times(d+1)n}, \tag{8}$$

in which $L(W^\tau) \in \mathbb{R}^{n\times n}$, $L(\widetilde{G}^\rho) \in \mathbb{R}^{dn\times dn}$, $\widetilde{V} \in \mathbb{R}^{n\times dn}$ and $\widetilde{\Sigma} = \operatorname{diag}\{\widetilde{\Sigma}_i, \cdots, \widetilde{\Sigma}_n\} \in \mathbb{R}^{dn\times dn}$ are sparse matrices defined as Eqs. (13) to (16) in [9, Section 4].

In the next section, we will propose a generalized proximal operator of Eqs. (6) and (7) whose minimization is n independent optimization problems on $SE(d)$, and thus can be efficiently solved, which is fundamental to our proposed methods for PGO.

5 The Generalized Proximal Operator for PGO

In this section, we propose an upper bound of PGO, and show that the resulting upper bound is a generalized proximal operator of the first-order approximation of PGO.

For any matrices A and B of the same size, it is known that

$$\frac{1}{2}\|A-B\|^2=\min_{P\in\mathbb{R}^{m\times n}}\|A-P\|^2+\|B-P\|^2, \tag{9}$$

the unique optimal solution to which is $P=\frac{1}{2}A+\frac{1}{2}B$. As a result of Eq. (9), if we introduce m pairs of extra variables $P_{ij}\in\mathbb{R}^{d\times d}$ and $p_{ij}\in\mathbb{R}^d$ for each $(i,j)\in\overrightarrow{\mathcal{E}}$ to Eq. (6), an upper bound of PGO is obtained as

$$\begin{aligned}\min_{\substack{R_i\in SO(d),\,t_i\in\mathbb{R}^d,\\ i=1,\cdots,n}}\sum_{(i,j)\in\overrightarrow{\mathcal{E}}}\Big(&\kappa_{ij}\cdot\|R_i\widetilde{R}_{ij}-P_{ij}\|^2+\tau_{ij}\cdot\|R_i\tilde{t}_{ij}+t_i-p_{ij}\|^2+\\ &\kappa_{ij}\cdot\|R_j-P_{ij}\|^2+\tau_{ij}\cdot\|t_j-p_{ij}\|^2\Big).\end{aligned} \tag{10}$$

If P_{ij} and p_{ij} are chosen as

$$P_{ij}=\frac{1}{2}R_i^{(k)}\widetilde{R}_{ij}+\frac{1}{2}R_j^{(k)},\qquad p_{ij}=\frac{1}{2}R_i^{(k)}\tilde{t}_{ij}+\frac{1}{2}t_i^{(k)}+\frac{1}{2}t_j^{(k)}, \tag{11}$$

with $(R_i^{(k)},\,t_i^{(k)})\in SE(d)$ and $i=1,\cdots,n$, then Eq. (6) and Eq. (10) attain the same objective value at $R_i^{(k)}$ and $t_i^{(k)}$. As a matter of fact, Eq. (10) results in a generalized proximal operator of the first-order approximation of PGO as stated in Theorem 1.

Theorem 1. *Let P_{ij} and p_{ij} are chosen as Eq. (11) with $(R_i^{(k)},\,t_i^{(k)})\in SE(d)$ and $i=1,\cdots,n$. Then, there exists a constant matrix $0\preceq\widetilde{\Omega}\in\mathbb{R}^{(d+1)n\times(d+1)n}$ such that Eq. (10) is equivalent to*

$$\begin{aligned}\min_{X\in\mathbb{R}^{d\times n}\times SO(d)^n}F(X^{(k)})&+\operatorname{trace}\Big((X-X^{(k)})^\top\nabla F(X^{(k)})\Big)+\\ &\frac{1}{2}\operatorname{trace}\Big((X-X^{(k)})\widetilde{\Omega}(X-X^{(k)})^\top\Big)\end{aligned} \tag{12}$$

in which $X^{(k)}=\begin{bmatrix}t_1^{(k)}\cdots t_n^{(k)}\ R_1^{(k)}\cdots R_n^{(k)}\end{bmatrix}\in\mathbb{R}^{d\times n}\times SO(d)^n$, $F(X^{(k)})=\frac{1}{2}\operatorname{trace}(X^{(k)}\widetilde{M}X^{(k)\top})$ and $\nabla F(X^{(k)})=X^{(k)}\widetilde{M}$.

Proof. See [16, Appendix B.1]. □

It should be noted that Eq. (12) is an upper bound of PGO as well as Eqs. (6) and (7), in which $\widetilde{M}$ and $\widetilde{\Omega}$ are closely related as follows.

Theorem 2. *Let $\widetilde{M}$ and $\widetilde{\Omega}$ be defined as Eqs. (7) and (12), respectively. Then,*
(1) $\widetilde{\Omega}\succeq\widetilde{M}$;
(2) for any $c\in\mathbb{R}$, $c\cdot\mathbf{I}\succeq\widetilde{\Omega}$ if $\frac{c}{2}\cdot\mathbf{I}\succeq\widetilde{M}$.

Proof. See [16, Appendix B.2]. □

As a result of Theorems 1 and 2, Eq. (12) is a generalized proximal operator and an upper bound of PGO, which suggests the possibility of generalized proximal methods to solve PGO. It should be noted that only the Euclidean gradient $\nabla F(X^{(k)})$ is involved in Eq. (12), and as a result, we might also accelerate generalized proximal methods using [11, 12]. In Section 6, we will propose generalized proximal methods for PGO, and in Section 7, we will further accelerate our proposed generalized proximal methods for PGO, which is the major contribution of this paper.

6 Generalized Proximal Methods for PGO

In this section, we propose two generalized proximal methods to solve PGO and show that our proposed methods for PGO converge to first-order critical points.

6.1 The $\mathsf{GPM-PGO}$ Method

Algorithm 1 The $\mathsf{GPM-PGO}$ Method

1: **Input**: An initial iterate $X^{(0)} = \begin{bmatrix} t_1^{(0)} & \cdots & t_n^{(0)} & R_1^{(0)} & \cdots & R_n^{(0)} \end{bmatrix} \in \mathbb{R}_d^n \times SO(d)^n$, and the maximum number of iterations N.
2: **Output**: A sequence of iterates $\{X^{(k)}\}$.
3: **function** $\mathsf{GPM-PGO}(X^{(0)}, N)$
4:   **for** $k = 0 \rightarrow N-1$ **do**
5:     $\begin{bmatrix} \theta_1^{(k)} & \cdots & \theta_n^{(k)} \end{bmatrix} \leftarrow X^{(k)}\Theta$
6:     **for** $i = 1 \rightarrow n$ **do**
7:       $R_i^{(k+1)} \leftarrow \arg\max\limits_{R_i \in SO(d)} \mathrm{trace}(R_i^\top \theta_i^{(k)})$
8:     **end for**
9:     $R^{(k+1)} \leftarrow \begin{bmatrix} R_1^{(k+1)} & \cdots & R_n^{(k+1)} \end{bmatrix}$
10:     $t^{(k+1)} \leftarrow R^{(k+1)}\Xi + X^{(k)}\Psi$
11:     $X^{(k+1)} \leftarrow \begin{bmatrix} t^{(k+1)} & R^{(k+1)} \end{bmatrix}$
12:   **end for**
13:   **return** $\{X^{(k)}\}$
14: **end function**

According to Theorem 2, it is known that $\widetilde{\Omega} \succeq \widetilde{M}$, and thus, we obtain a series of upper bounds of PGO:

$$\min_{X \in \mathbb{R}^{d\times n}\times SO(d)^n} G(X|X^{(k)}) \triangleq F(X^{(k)}) + \mathrm{trace}\left((X - X^{(k)})^\top \nabla F(X^{(k)})\right) + \frac{1}{2}\mathrm{trace}\left((X - X^{(k)})\widetilde{\Gamma}(X - X^{(k)})^\top\right), \tag{13}$$

in which $\widetilde{\Gamma} = \widetilde{\Omega} + \alpha \cdot \mathbf{I} \succeq \widetilde{M}$ and $\alpha \geq 0$. From Eqs. (10) and (12), a straightforward mathematical manipulation indicates that $\widetilde{\Gamma}$ in Eq. (13) takes the form as $\widetilde{\Gamma} \triangleq \begin{bmatrix} \Gamma^\tau & \widetilde{\Gamma}^{\nu\top} \\ \widetilde{\Gamma}^\nu & \widetilde{\Gamma}^\rho \end{bmatrix}$ with $\Gamma^\tau = \mathrm{diag}\{\Gamma_1^\tau, \cdots, \Gamma_n^\tau\} \in \mathbb{R}^{n\times n}$, $\widetilde{\Gamma}^\rho = \mathrm{diag}\{\widetilde{\Gamma}_1^\rho, \cdots, \widetilde{\Gamma}_n^\rho\} \in \mathbb{R}^{dn\times dn}$ and $\widetilde{\Gamma}^\nu = \mathrm{diag}\{\widetilde{\Gamma}_1^\nu, \cdots, \widetilde{\Gamma}_n^\nu\} \in \mathbb{R}^{dn\times n}$, in which

$$\Gamma_i^\tau = \alpha + \sum_{(i,j)\in\mathcal{E}} 2\cdot\tau_{ij} \in \mathbb{R}, \tag{14a}$$

$$\widetilde{\Gamma}_i^\rho = \alpha\cdot\mathbf{I} + \sum_{(i,j)\in\mathcal{E}} 2\cdot\kappa_{ij}\cdot\mathbf{I} + \sum_{(i,j)\in\overrightarrow{\mathcal{E}}} 2\cdot\tau_{ij}\cdot\tilde{t}_{ij}\tilde{t}_{ij}^\top \in \mathbb{R}^{d\times d}, \tag{14b}$$

$$\widetilde{\Gamma}_i^\nu = \sum_{(i,j)\in\overrightarrow{\mathcal{E}}} 2\cdot\tau_{ij}\cdot\tilde{t}_{ij} \in \mathbb{R}^d. \tag{14c}$$

For simplicity and clarity, we rewrite $\nabla F(X^{(k)})$ in Eq. (12) as $\nabla F(X^{(k)}) = \left[\overline{\gamma}_1^\tau \cdots \overline{\gamma}_n^\tau\ \overline{\gamma}_1^\rho \cdots \overline{\gamma}_n^\rho\right] \in \mathbb{R}^{d\times(d+1)n}$, in which $\overline{\gamma}_i^\tau \in \mathbb{R}^d$ and $\overline{\gamma}_i^\rho \in \mathbb{R}^{d\times d}$ are Euclidean gradients w.r.t. t_i and R_i, respectively. Substituting Eqs. (14a) to (14c) into Eq. (12) and simplifying the resulting equation, we obtain

$$\begin{aligned}\min_{\substack{R_i\in SO(d),\, t_i\in\mathbb{R}^d,\\ i=1,\cdots,n}} F(X^{(k)}) + \sum_{i\in\mathcal{V}}\Big(\frac{1}{2}\operatorname{trace}\left((R_i - R_i^{(k)})\widetilde{\Gamma}_i^\rho(R_i - R_i^{(k)})^\top)\right) +\\ \widetilde{\Gamma}_i^{\nu\top}(R_i - R_i^{(k)})^\top(t_i - t_i^{(k)}) + \frac{1}{2}(t_i - t_i^{(k)})^\top\Gamma_i^\tau(t_i - t_i^{(k)}) +\\ \operatorname{trace}\left(\overline{\gamma}_i^{\rho\top}(R_i - R_i^{(k)})\right) + \overline{\gamma}_i^{\tau\top}(t_i - t_i^{(k)})\Big),\end{aligned} \tag{15}$$

which is equivalent to n independent optimization problems on $(R_i,\, t_i) \in SE(d)$:

$$\begin{aligned}\min_{R_i\in SO(d),\, t_i\in\mathbb{R}^d} \frac{1}{2}\operatorname{trace}\left((R_i - R_i^{(k)})\widetilde{\Gamma}_i^\rho(R_i - R_i^{(k)})^\top)\right) +\\ \widetilde{\Gamma}_i^{\nu\top}(R_i - R_i^{(k)})^\top(t_i - t_i^{(k)}) + \frac{1}{2}\Gamma_i^\tau(t_i - t_i^{(k)})^\top(t_i - t_i^{(k)}) +\\ \operatorname{trace}\left(\overline{\gamma}_i^{\rho\top}(R_i - R_i^{(k)})\right) + \overline{\gamma}_i^{\tau\top}(t_i - t_i^{(k)}).\end{aligned} \tag{16}$$

Furthermore, if $R_i \in SO(d)$ is given, $t_i \in \mathbb{R}^d$ can be recovered as

$$t_i = t_i^{(k)} - R_i\widetilde{\Gamma}_i^\nu\Gamma_i^{\tau-1} + \left(R_i^{(k)}\widetilde{\Gamma}_i^\nu - \overline{\gamma}_i^\tau\right)\Gamma_i^{\tau-1}. \tag{17}$$

Substituting Eq. (17) into Eq. (16) to cancel out t_i and applying $R_iR_i^\top = \mathbf{I}$ to simplify the resulting equation, we obtain

$$R_i^{(k+1)} = \arg\max_{R_i\in SO(d)} \operatorname{trace}(R_i^\top\theta_i^{(k)}), \tag{18}$$

in which

$$\theta_i^{(k)} = R_i^{(k)}(\widetilde{\Gamma}_i^\rho - \widetilde{\Gamma}_i^\nu\Gamma_i^{\tau-1}\widetilde{\Gamma}_i^{\nu\top}) + \overline{\gamma}_i^\tau\Gamma_i^{\tau-1}\widetilde{\Gamma}_i^{\nu\top} - \overline{\gamma}_i^\rho \in \mathbb{R}^{d\times d}. \tag{19}$$

From $\nabla F(X^{(k)}) = X^{(k)}\widetilde{M}$ and Eqs. (8) and (13), we might matricize Eq. (19) as

$$\theta^{(k)} = X^{(k)}\Theta,$$

in which $\theta^{(k)} = \left[\theta_1^{(k)} \cdots \theta_n^{(k)}\right] \in \mathbb{R}^{d\times dn}$, and $\Theta \in \mathbb{R}^{(d+1)n\times dn}$ is a sparse matrix

$$\Theta = \begin{bmatrix} L(W^\tau) \\ \widetilde{V}^\top \end{bmatrix}\Gamma^{\tau-1}\widetilde{\Gamma}^{\nu\top} + \begin{bmatrix} \mathbf{0} \\ \widetilde{\Gamma}^\rho - \widetilde{\Gamma}^\nu\Gamma^{\tau-1}\widetilde{\Gamma}^{\nu\top} \end{bmatrix} - \begin{bmatrix} \widetilde{V} \\ L(\widetilde{G}^\rho) + \widetilde{\Sigma} \end{bmatrix}. \tag{20}$$

Similarly, as a result of Eqs. (8), (13) and (17), we obtain

$$t^{(k+1)} = R^{(k+1)}\Xi + X^{(k)}\Psi, \tag{21}$$

in which $t^{(k+1)} = \begin{bmatrix} t_1^{(k+1)} \cdots t_n^{(k+1)} \end{bmatrix} \in \mathbb{R}^{d\times n}$, $R^{(k+1)} = \begin{bmatrix} R_1^{(k+1)} \cdots R_n^{(k+1)} \end{bmatrix} \in SO(d)^n \subset \mathbb{R}^{d\times dn}$, and $\Xi \in \mathbb{R}^{dn\times n}$ and $\Psi \in \mathbb{R}^{(d+1)n\times n}$ are sparse matrices with $\Xi = -\widetilde{\Gamma}^{\nu}\Gamma^{\tau-1}$ and $\Psi = \begin{bmatrix} \Gamma^{\tau} - L(W^{\tau}) \\ \widetilde{\Gamma}^{\nu} - \widetilde{V}^{\top} \end{bmatrix} \Gamma^{\tau-1}$, respectively.

It is obvious that Eq. (16) is simplified to Eq. (18). From [17], if $\theta_i^{(k)} \in \mathbb{R}^{d\times d}$ admits a singular value decomposition $\theta_i^{(k)} = U_i \Sigma_i V_i^{\top}$ in which U_i and $V_i \in O(d)$ are orthogonal (but not necessarily special orthogonal) matrices, and $\Sigma_i = \mathrm{diag}\{\sigma_1, \sigma_2, \cdots, \sigma_d\} \in \mathbb{R}^{d\times d}$ is a diagonal matrix, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d \geq 0$ are singular values of θ_i, then the optimal solution to Eq. (18) is

$$R_i = \begin{cases} U_i \Sigma^{+} V_i^{\top}, & \det(U_i V_i^{\top}) > 0, \\ U_i \Sigma^{-} V_i^{\top}, & \det(U_i V_i^{\top}) < 0, \end{cases} \tag{22}$$

in which $\Sigma^{+} = \mathrm{diag}\{1, 1, \cdots, 1\}$ and $\Sigma^{-} = \mathrm{diag}\{1, 1, \cdots, -1\}$. If $d = 2$, the equation above is equivalent to the polar decomposition of 2×2 matrices, and if $d = 3$, there are fast algorithms for singular value decomposition of 3×3 matrices [18]. In both cases of $d = 2$ and $d = 3$, Eq. (18) can be efficiently solved. As long as $R_i \in SO(d)$ is known, we can further recover $t_i \in \mathbb{R}^d$ using Eq. (21) so that a solution $(t_i, R_i) \in \mathbb{R}^d \times SO(d)$ to Eq. (16) is obtained with which Eq. (15) is also solved.

Therefore, Eq. (15) only involves n singular value decomposition to solve Eq. (18) on $R_i \in SO(d)$, and a matrix-vector multiplication to retrieve $t_i \in \mathbb{R}^d$ using Eq. (21), which suggests the $\mathsf{GPM} - \mathsf{PGO}$ method (Algorithm 1).

6.2 The $\mathsf{GPM} - \mathsf{PGO}^*$ Method

It is according to Eq. (7) that PGO can be reformulated as

$$\min_{t\in\mathbb{R}^{d\times n},\, R\in SO(d)^n} \frac{1}{2}\,\mathrm{trace}\left(tL(W^{\tau})t^{\top}\right) + \mathrm{trace}\left(\tilde{V}R^{\top}t\right) + \frac{1}{2}\,\mathrm{trace}(RL(\widetilde{G}^{\rho})R^{\top}) + \frac{1}{2}\,\mathrm{trace}(R\widetilde{\Sigma}^{\rho}R^{\top}), \tag{23}$$

in which $t = \begin{bmatrix} t_1 \cdots t_n \end{bmatrix} \in \mathbb{R}^{d\times n}$ and $R = \begin{bmatrix} R_1 \cdots R_n \end{bmatrix} \in SO(d)^n$. Following a similar procedure in [9], if the rotation $R = R^{(k+1)}$ in the equation above is given, we can optimally recover the corresponding translation $t = t^{(k+1)}$ as $t^{(k+1)} = -R^{(k+1)}\widetilde{V}^{\top}L(W^{\tau})^{\dagger}$, which is further simplified to

$$t^{(k+1)} = -R^{(k+1)}\widetilde{T}^{\top}\Omega A^{\top}(A\Omega A^{\top})^{\dagger}. \tag{24}$$

In Eq. (24), $A \in \mathbb{R}^{n\times m}$, $\Omega \in \mathbb{R}^{m\times m}$ and $\widetilde{T} \in \mathbb{R}^{m\times dn}$ are sparse matrices that are defined as Eqs.(7), (22) and (23) in [9, Sections 3 and 4], respectively.

As a result, instead of computing each $t_i^{(k+1)} \in \mathbb{R}^d$ sub-optimally using Eqs. (17) and (21), we might use Eq. (24) to optimally recover $t^{(k+1)} \in \mathbb{R}^{d\times n}$ w.r.t. $R^{(k+1)} \in SO(d)^n$ as a whole. Furthermore, if $t^{(k+1)}$ is recovered by Eq. (24), we have $\overline{\gamma}_i^{\tau} = \mathbf{0}$ for all $i = 1, \cdots, n$ in Eq. (16), and thus, only the Euclidean gradient $\overline{\gamma}^{\rho} = \begin{bmatrix} \overline{\gamma}_1^{\rho} \cdots \overline{\gamma}_n^{\rho} \end{bmatrix} = \nabla_R F(X^{(k)})$ w.r.t. $R^{(k)}$ needs to be computed, and then $\theta_i^{(k)}$ is simplified to

Algorithm 2 The GPM − PGO* Method

Input: An initial iterate $x^{(0)} = \begin{bmatrix} t^{(0)} & R^{(0)} \end{bmatrix} \in \mathbb{R}^{d\times n} \times SO(d)^n$ in which $t^{(0)} = -R^{(0)}\widetilde{T}^\top \Omega A^\top (A\Omega A^\top)^\dagger$, and the maximum number of iterations N.
Output: A sequence of iterates $\{X^{(k)}\}$.
function GPM − PGO*$(X^{(0)}, N)$
  for $k = 0 \to N-1$ **do**
    $\begin{bmatrix} \theta_1^{(k)} & \cdots & \theta_n^{(k)} \end{bmatrix} \leftarrow X^{(k)}\Phi$
    for $i = 1 \to n$ **do**
      $R_i^{(k+1)} \leftarrow \arg\max\limits_{R_i \in SO(d)} \mathrm{trace}(R_i^\top \theta_i^{(k)})$
    end for
    $R^{(k+1)} \leftarrow \begin{bmatrix} R_1^{(k+1)} & \cdots & R_n^{(k+1)} \end{bmatrix}$
    $t^{(k+1)} \leftarrow -R^{(k+1)}\widetilde{T}^\top \Omega A^\top (A\Omega A^\top)^\dagger$
    $X^{(k+1)} \leftarrow \begin{bmatrix} t^{(k+1)} & R^{(k+1)} \end{bmatrix}$
  end for
  return $\{X^{(k)}\}$
end function

$$\theta_i^{(k)} = R_i^{(k)}(\widetilde{\Gamma}_i^\rho - \widetilde{\Gamma}_i^\nu \Gamma_i^{\tau-1} \widetilde{\Gamma}_i^{\nu\top}) - \overline{\gamma}_i^\rho.$$

Following a similar procedure to Eq. (20), we obtain

$$\theta^{(k)} = X^{(k)}\Phi,$$

in which $\theta^{(k)} = \begin{bmatrix} \theta_1^{(k)} & \cdots & \theta_n^{(k)} \end{bmatrix} \in \mathbb{R}^{d\times dn}$, and $\Phi \in \mathbb{R}^{(d+1)n\times dn}$ is a sparse matrix

$$\Phi = \begin{bmatrix} \mathbf{0} \\ \widetilde{\Gamma}^\rho - \widetilde{\Gamma}^\nu \Gamma^{\tau-1} \widetilde{\Gamma}^{\nu\top} \end{bmatrix} - \begin{bmatrix} \widetilde{V} \\ L(\widetilde{G}^\rho) + \widetilde{\Sigma} \end{bmatrix}. \tag{25}$$

From Eqs. (24) and (25), we obtain GPO − PGO* (Algorithm 2), which always recovers the translation t optimally w.r.t. R and thus is expected to outperform Algorithm 1.

It is important to establish whether GPM − PGO and GPM − PGO* solve PGO. Empirically, we observe in the experiments that our proposed methods always converge to the global optima if the noise magnitudes are below a certain threshold. Theoretically, we can provide guarantees that GPM − PGO and GPM − PGO* converge to first-order critical points. Note that existing first- and second-order PGO methods in general need to choose stepsize carefully to guarantee the convergence to first-order critical points, whereas there is no stepsize tuning involved in either GPM − PGO or GPM − PGO*.

Theorem 3. *Let $\{X^{(k)}\}$ be a sequence of iterates that is generated by either* GPM − PGO *or* GPM − PGO*. *Then,*

(1) $F(X^{(k)})$ is non-increasing;
(2) $F(X^{(k)}) \to F^\infty$ as $k \to \infty$;
(3) $\|X^{(k+1)} - X^{(k)}\| \to 0$ as $k \to \infty$ if $\widetilde{\Gamma} \succ \widetilde{M}$;
(4) $\mathrm{grad}\, F(X^{(k)}) \to \mathbf{0}$ as $k \to \infty$ if $\widetilde{\Gamma} \succ \widetilde{M}$;
(5) $\|X^{(k+1)} - X^{(k)}\| \to 0$ as $k \to \infty$ if $\alpha > 0$;
(6) $\mathrm{grad}\, F(X^{(k)}) \to \mathbf{0}$ as $k \to \infty$ if $\alpha > 0$.

Proof. See [16, Appendix B.3]. □

Algorithm 3 The $\mathsf{NAG-PGO}^*$ Method

Input: An initial iterate $X^{(0)} = \begin{bmatrix} t^{(0)} & R^{(0)} \end{bmatrix} \in \mathbb{R}^{d\times n} \times SO(d)^n$ and $X^{(-1)} = \begin{bmatrix} t^{(-1)} & R^{(-1)} \end{bmatrix} \in \mathbb{R}^{d\times n} \times SO(d)^n$ in which $t^{(0)} = -R^{(0)}\widetilde{T}^\top \Omega A^\top (A\Omega A^\top)^\dagger$ and $t^{(-1)} = -R^{(-1)}\widetilde{T}^\top \Omega A^\top (A\Omega A^\top)^\dagger$, $s^{(0)} \in [1, +\infty)$, and the maximum number of iterations N.
Output: A sequence of iterates $\{X^{(k)},\, s^{(k)}\}$.
function $\mathsf{NAG-PGO}^*(X^{(0)},\, X^{(-1)},\, s^{(0)},\, N)$
  for $k = 0 \to N-1$ **do**
    $s^{(k+1)} \leftarrow \dfrac{\sqrt{4s^{(k)2}+1}+1}{2}, \quad Y^{(k)} \leftarrow X^{(k)} + \dfrac{s^{(k)}-1}{s^{(k+1)}}\left(X^{(k)} - X^{(k-1)}\right)$
    $\begin{bmatrix} \theta_1^{(k)} & \cdots & \theta_n^{(k)} \end{bmatrix} \leftarrow Y^{(k)}\Phi$
    for $i = 1 \to n$ **do**
      $R_i^{(k+1)} \leftarrow \arg\max_{R_i \in SO(d)} \mathrm{trace}(R_i^\top \theta_i^{(k)})$
    end for
    $R^{(k+1)} \leftarrow \begin{bmatrix} R_1^{(k+1)} & \cdots & R_n^{(k+1)} \end{bmatrix}$
    $t^{(k+1)} \leftarrow -R^{(k+1)}\widetilde{T}^\top \Omega A^\top (A\Omega A^\top)^\dagger$
    $X^{(k+1)} \leftarrow \begin{bmatrix} t^{(k+1)} & R^{(k+1)} \end{bmatrix}$
  end for
  return $\{X^{(k)},\, s^{(k)}\}$
end function

7 Accelerated Generalized Proximal Methods for PGO

$\mathsf{GPM-PGO}$ and $\mathsf{GPM-PGO}^*$ generalize proximal methods that use Euclidean gradients to update pose estimates, which is different from existing first-order PGO methods [3–5] using Riemannian gradients, and thus, it is possible to accelerate $\mathsf{GPM-PGO}$ and $\mathsf{GPM-PGO}^*$ using [11–14].

Following Nesterov's accelerated proximal method [11, 12], we might extend $\mathsf{GPM-PGO}^*$ to $\mathsf{NAG-PGO}^*$ (Algorithm 3). $\mathsf{NAG-PGO}^*$ is almost the same as $\mathsf{GPM-PGO}^*$ at the beginning when $s^{(k)}$ is small but then more governed by the momentum term $X^{(k)} - X^{(k-1)}$ as k increases. If we relax the constraints of $R_i \in SO(d)$ to any closed convex sets and choose initial iterate $X^{(0)} = X^{(-1)}$ and $s^{(0)} = 1$, $\mathsf{NAG-PGO}^*$ would converge to the global optima within $O(1/N^2)$ time, whereas theoretically $\mathsf{GPM-PGO}^*$ can not have a rate of convergence better than $O(1/N)$ [11, 12]. Even though PGO is a non-convex optimization problem, $\mathsf{NAG-PGO}^*$ is expected to inherit the characteristics of Nesterov's accelerated proximal method and outperform $\mathsf{GPM-PGO}^*$, and empirically, $\mathsf{NAG-PGO}^*$ is indeed much faster than $\mathsf{GPM-PGO}^*$.

Different from $\mathsf{GPM-PGO}^*$, $\mathsf{NAG-PGO}^*$ is not a descent algorithm, and might have "Nesterov ripples" due to high momentum term as k increases [19]. Moreover, even though $\mathsf{NAG-PGO}^*$ is empirically much faster than $\mathsf{GPM-PGO}^*$ to converge to first-order critical points, it seems difficult to have any theoretical guarantees of convergence for $\mathsf{NAG-PGO}^*$. In order to address these theoretical and practical drawbacks of $\mathsf{NAG-PGO}^*$, we propose $\mathsf{AGPM-PGO}^*$ (Algorithm 4) that is an extension of $\mathsf{NAG-PGO}^*$ with adaptive restart – a restart scheme is commonly used to im-

Algorithm 4 The $\mathsf{AGPM-PGO}^*$ Method

1: **Input**: An initial iterate $X^{(0)} = \begin{bmatrix} t^{(0)} & R^{(0)} \end{bmatrix} \in \mathbb{R}^{d\times n} \times SO(d)^n$ in which $t^{(0)} = -R^{(0)}\widetilde{T}^\top \Omega A^\top (A\Omega A^\top)^\dagger$, the maximum number of outer iterations N, the maximum number of inner iterations N_0, $\eta \in (0,\, 1]$, and $\delta \in [0,\, \infty)$.
2: **Output**: A sequence of iterates $\{X^{(k)}\}$.
3: **function** $\mathsf{AGPM-PGO}^*(X^{(0)},\, N,\, N_0,\, \delta)$
4: $\quad a^{(0)} \leftarrow 1, \quad T^{(0)} \leftarrow X^{(0)}, \quad f^{(0)} \leftarrow F(X^{(0)})$
5: $\quad$ **for** $k = 0 \to N-1$ **do**
6: $\quad\quad \{V^{(i)},\, s^{(i)}\} \leftarrow \mathsf{NAG-PGO}^*(X^{(k)},\, T^{(k)},\, a^{(k)},\, N_0)$
7: $\quad\quad$ **if** $F(V^{(N_0)}) \leq f^{(k)} - \delta \cdot \|V^{(N_0)} - X^{(k)}\|^2$ **then**
8: $\quad\quad\quad a^{(k+1)} \leftarrow s^{(N_0)}, \quad X^{(k+1)} \leftarrow V^{(N_0)}, \quad T^{(k+1)} \leftarrow V^{(N_0-1)}$
9: $\quad\quad$ **else**
10: $\quad\quad\quad \{Z^{(i)}\} \leftarrow \mathsf{GPM-PGO}^*(X^{(k)},\, N_0)$
11: $\quad\quad\quad a^{(k+1)} \leftarrow 1, \quad X^{(k+1)} \leftarrow Z^{(N_0)}, \quad T^{(k+1)} \leftarrow Z^{(N_0)}$
12: $\quad\quad$ **end if**
13: $\quad\quad f^{(k+1)} \leftarrow (1-\eta)\cdot f^{(k)} + \eta \cdot F(X^{(k+1)})$
14: $\quad$ **end for**
15: $\quad$ **return** $\{X^{(k)}\}$
16: **end function**

prove the convergence of accelerated proximal methods in convex optimization [19]. In $\mathsf{AGPM-PGO}^*$, we implement $\mathsf{GPM-PGO}^*$ for several iterations and then restart $\mathsf{NAG-PGO}^*$ whenever the momentum term seems to take us in a bad direction, and as is shown later, though not a descent algorithm, $\mathsf{AGPM-PGO}^*$ is guaranteed to converge to first-order critical points under mild conditions. Since $\mathsf{NAG-PGO}^*$ is usually preferred than $\mathsf{GPM-PGO}^*$, it is recommended to choose a small δ in line 7 of Algorithm 4. Besides acceleration, $\mathsf{NAG-PGO}^*$ and $\mathsf{AGPM-PGO}^*$ are expected to escape saddle points faster than $\mathsf{GPM-PGO}^*$ with some additional simple strategies adopted, for which interested readers can refer to [14] for more details. Similar to $\mathsf{NAG-PGO}^*$ and $\mathsf{AGPM-PGO}^*$, we might also extend $\mathsf{GPM-PGO}$ to obtain $\mathsf{NAG-PGO}$ and $\mathsf{AGPM-PGO}$, which are shown in [16, Appendix A].

In the experiments, we observe that $\mathsf{AGPM-PGO}$ and $\mathsf{AGPM-PGO}^*$ converge to the global optima as long as $\mathsf{GPM-PGO}$ and $\mathsf{GPM-PGO}^*$ converge to the global optima, however, $\mathsf{AGPM-PGO}$ and $\mathsf{AGPM-PGO}^*$ are a lot faster than $\mathsf{GPM-PGO}$ and $\mathsf{GPM-PGO}^*$. Even though $\mathsf{AGPM-PGO}$ and $\mathsf{AGPM-PGO}^*$ are not descent algorithms if $\eta < 1$, we still prove that $\mathsf{AGPM-PGO}$ and $\mathsf{AGPM-PGO}^*$ converge to first-order critical points under mild conditions.

Theorem 4. *Let $\{X^{(k)}\}$ be a sequence of iterates that is generated by either $\mathsf{AGPM-PGO}$ or $\mathsf{AGPM-PGO}^*$. Then,*

(1) $F(X^{(k)}) \to F^\infty$ *as* $k \to \infty$*;*
(2) $\|X^{(k+1)} - X^{(k)}\| \to 0$ *as* $k \to \infty$ *if* $\widetilde{\Gamma} \succ \widetilde{M}$ *and* $\delta > 0$*;*
(3) $\operatorname{grad} F(X^{(k)}) \to \mathbf{0}$ *as* $k \to \infty$ *if* $\widetilde{\Gamma} \succ \widetilde{M}$, $\delta > 0$ *and* $N_0 = 1$*;*
(4) $\|X^{(k+1)} - X^{(k)}\| \to 0$ *as* $k \to \infty$ *if* $\alpha > 0$ *and* $\delta > 0$*;*
(5) $\operatorname{grad} F(X^{(k)}) \to \mathbf{0}$ *as* $k \to \infty$ *if* $\alpha > 0$, $\delta > 0$ *and* $N_0 = 1$.

Proof. See [16, Appendix B.4]. □

Note that even though (3) and (5) of Theorem 4 require the maximum number of inner iterations $N_0 = 1$ to guarantee $\operatorname{grad} F(X^{(k)}) \to 0$, we still observe in the experiments that $\mathsf{AGPM-PGO}$ and $\mathsf{AGPM-PGO}^*$ always converge to first-order critical points for any $N_0 > 1$. As a result, it should be empirically all right to specify $N_0 > 1$ so that the number of objective function evaluation is reduced and the overall efficiency is improved.

8 Experiments

In this section, we evaluate the performance of generalized proximal methods for PGO that are proposed in Sections 6 and 7 on SLAM and distributed 3D sensor network localization, and make comparisons with existing techniques. All the tests have been performed on a Thinkpad P51 laptop with a 3.1GHz Intel Core Xeon that runs Ubuntu 18.04 and uses g++ 7.4 as C++ compiler.

8.1 SLAM Benchmark Datasets

In the first set of experiments, we implement $\mathsf{AGPM-PGO}^*$ (Algorithm 4) on a variety of popular 2D and 3D SLAM datasets and compare the results with $\mathsf{SE-Sync}$ [9], which is one of the fastest PGO methods.

For each of the dataset, we choose $\alpha = 0$, $N_0 = 10$, $\eta = 1$ and $\delta = 1\times 10^{-5}$ for $\mathsf{AGPM-PGO}^*$, and $\mathsf{AGPM-PGO}^*$ terminates once the relative improvement of the objective function is less than $\epsilon = 0.002$, i.e., $F(X^{(k)}) \leq (1+\epsilon)F(X^{(k+1)})$. For $\mathsf{SE-Sync}$, we use the default settings except the stopping criteria. In default, $\mathsf{SE-Sync}$ does not stop until attaining a local optimum, whereas in our experiments, for a fair comparison, we terminate $\mathsf{SE-Sync}$ once it achieves an equivalent accuracy as $\mathsf{AGPM-PGO}^*$, which takes less time than the default settings. For all the datasets, we use the chordal initialization [20] for both $\mathsf{AGPM-PGO}^*$ and $\mathsf{SE-Sync}$.

The results for these experiments are shown in Tables 1 and 2 and Figs. 1 to 3. In Tables 1 and 2, n is the number of unknown poses, m is the number of edges, f^* is the globally optimal objective value that can be obtained using $\mathsf{SE-Sync}$, and f is the objective value attained by $\mathsf{AGPM-PGO}^*$ and $\mathsf{SE-Sync}$. In Figs. 1 and 2, we present the speed-up v.s. $\mathsf{SE-Sync}$ and the relative objective error $(f-f^*)/f^*$ of $\mathsf{AGPM-PGO}^*$. In all the experiments, $\mathsf{AGPM-PGO}^*$ is several times faster than $\mathsf{SE-Sync}$ to achieve modest accurate solutions with an average speed-up of 5.14x for 2D SLAM datasets and 9.14x for 3D SLAM datasets. In addition, the average relative objective errors of $\mathsf{AGPM-PGO}^*$ for 2D and 3D SLAM datasets are 0.25% and 0.075%, respectively, and such an accuracy is generally sufficient for practical use in SLAM. Furthermore, though not presented in this paper, $\mathsf{AGPM-PGO}^*$ converge to the global optima in all the experiments if enough computational time is given.

We also compare the convergence of $\mathsf{GPM-PGO}$, $\mathsf{GPM-PGO}^*$, $\mathsf{AGPM-PGO}$ and $\mathsf{AGPM-PGO}^*$ with $\mathsf{SE-Sync}$ on 2D and 3D SLAM datasets, whose results are shown in [16, Appendix C].

Dataset	n	m	f^*	SE − Sync [9]		AGPM − PGO* [ours]	
				f	Time (s)	f	Time (s)
ais2klin	15115	16727	1.885×10^2	1.885×10^2	4.83×10^{-1}	1.901×10^2	6.19×10^{-2}
city	10000	20687	6.386×10^2	6.387×10^2	3.19×10^{-1}	6.388×10^2	4.02×10^{-2}
CSAIL	1045	1172	3.170×10^1	3.170×10^1	4.02×10^{-3}	3.171×10^1	7.81×10^{-4}
manhatta	3500	5453	6.432×10^3	6.434×10^3	9.02×10^{-3}	6.435×10^3	3.26×10^{-3}
intel	1728	2512	5.235×10^1	5.235×10^1	1.05×10^{-2}	5.248×10^1	5.10×10^{-3}

Table 1: Results of the 2D SLAM datasets

Dataset	n	m	f^*	SE − Sync [9]		AGPM − PGO* [ours]	
				f	Time (s)	f	Time (s)
cubicle	5750	16869	7.171×10^2	7.174×10^2	1.99×10^{-1}	7.178×10^2	6.64×10^{-2}
garage	1661	6275	1.263×10^0	1.263×10^0	4.38×10^{-1}	1.264×10^0	1.21×10^{-2}
grid	8000	22236	8.432×10^4	8.433×10^4	1.10×10^0	8.433×10^4	1.54×10^{-1}
rim	10195	29743	5.461×10^3	5.463×10^3	1.15×10^0	5.465×10^3	3.03×10^{-1}
sphere	2500	4949	1.687×10^3	1.687×10^3	1.74×10^{-1}	1.687×10^3	2.66×10^{-2}
torus	5000	9048	2.423×10^4	2.423×10^4	1.89×10^{-1}	2.425×10^4	5.50×10^{-2}
sphere-a	2200	8827	2.962×10^6	2.963×10^6	1.56×10^{-1}	2.963×10^6	4.06×10^{-2}

Table 2: Results of the 3D SLAM datasets

8.2 Distributed 3D Sensor Network Localization

In the experiments of distributed 3D sensor network localization, it is assumed that the 3D sensor network is static and connected, and each node in the network can only communicate with its neighbours and can only measure the relative pose w.r.t. its neighbours, and as a result, we need to solve it distributedly to estimate poses of each node. As mentioned before, GPM − PGO* and AGPM − PGO* always optimally recover the translation t, and thus, expect a faster convergence than GPM − PGO and AGPM − PGO. However, in distributed PGO, similar to second-order PGO methods [6–9] solving linear systems to evaluate the descent direction, GPM − PGO* and AGPM − PGO* have to use iterative solvers to solve Eq. (24) to recover the translation t, which usually reduces the efficiency of optimization. In contrast, GPM − PGO and AGPM − PGO only need the estimated poses of itself and its neighbours in optimization and do not have to solve linear systems as GPM − PGO* and AGPM − PGO* and second-order PGO methods [6–9]. Therefore, GPM − PGO and AGPM − PGO are well suited for the distributed PGO without inducing extra efforts. Furthermore, there is no loss of theoretical guarantees for GPM − PGO and AGPM − PGO in distributed PGO.

In the experiments, we simulate a distributed 3D sensor network on an ellipsoid with $n = 200$ vertices (nodes) and $m = 600$ edges. The noisy measurements $\tilde{g}_{ij} = (\widetilde{R}_{ij}, \tilde{t}_{ij}) \in SE(3)$ are generated according to the model $\widetilde{R}_{ij} = \underline{R}_{ij} \exp(\xi_{ij}^R)$ and

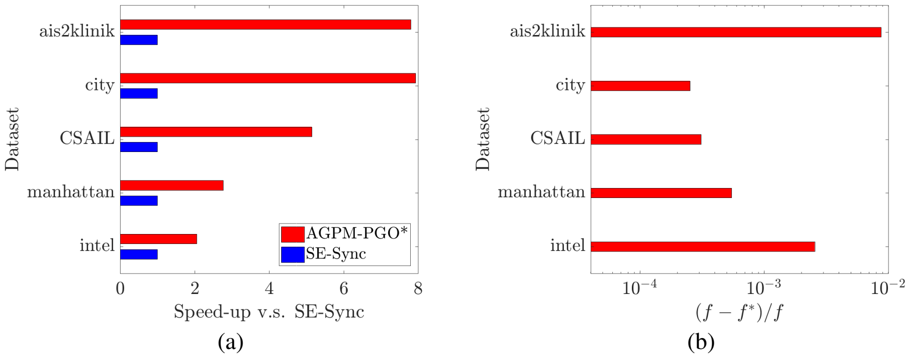


Fig. 1: The results of $\mathsf{AGPM-PGO}^*$ for 2D SLAM datasets. The results are (a) speed-up v.s. $\mathsf{SE-Sync}$ [9] and (b) the relative objective error $(f - f^*)/f^*$ of $\mathsf{AGPM-PGO}^*$. For 2D SLAM datasets, the average speed up is 5.14x and the average relative objective error is 0.25%.

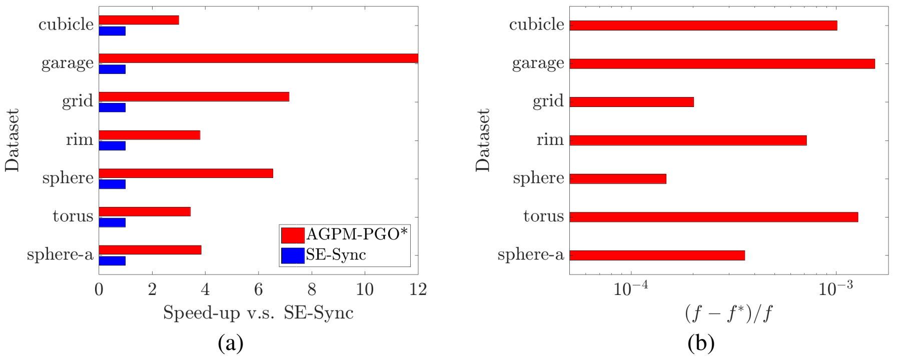


Fig. 2: The results of $\mathsf{AGPM-PGO}^*$ for 3D SLAM datasets. The results are (a) speed-up v.s. $\mathsf{SE-Sync}$ [9] and (b) the relative objective error $(f - f^*)/f^*$ of $\mathsf{AGPM-PGO}^*$. For 3D SLAM datasets, the average speed up is 9.14x and the average relative objective error is 0.075%.

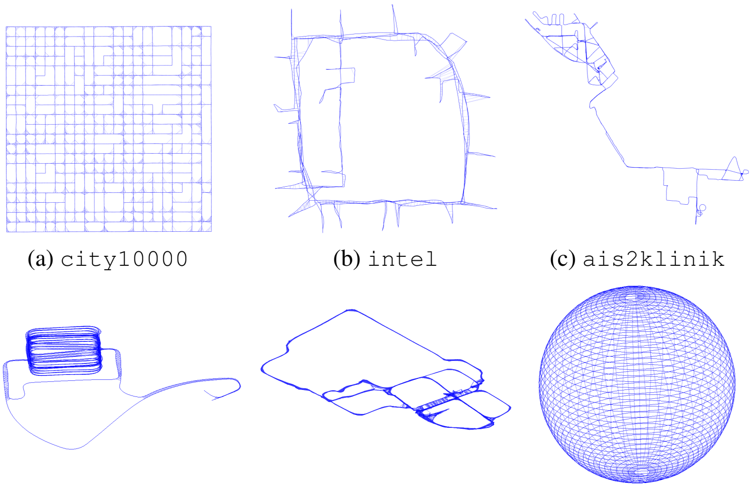
$\tilde{t}_{ij} = \underline{t}_{ij} + \xi^t_{ij}$, in which $\xi^R_{ij} \sim \mathcal{N}(0,\, \sigma_R \cdot \mathbf{I})$ with $\delta_R = 0.05\,\mathrm{rad}$ with $\xi^t_{ij} \sim \mathcal{N}(0,\, \sigma_t \cdot \mathbf{I})$ and $\delta_t = 0.05\,\mathrm{m}$. We compute the statistics over 30 runs and use the chordal initialization[1] for all the runs. The results are shown in Fig. 4. In all the 30 runs, $\mathsf{AGPM-PGO}$ converge to the global optima with an average rotation error of 0.0253 rad and an average relative translation error of 1.60%.

9 Conclusions

In this paper, we have proposed generalized proximal methods for PGO and proved that our proposed methods converge to first-order critical points. In addition, we have accelerated the rates of convergence without loss of any theoretical guarantees. Our

[1] The chordal initialization can be distributedly solved by relaxing PGO as convex quadratic programming.

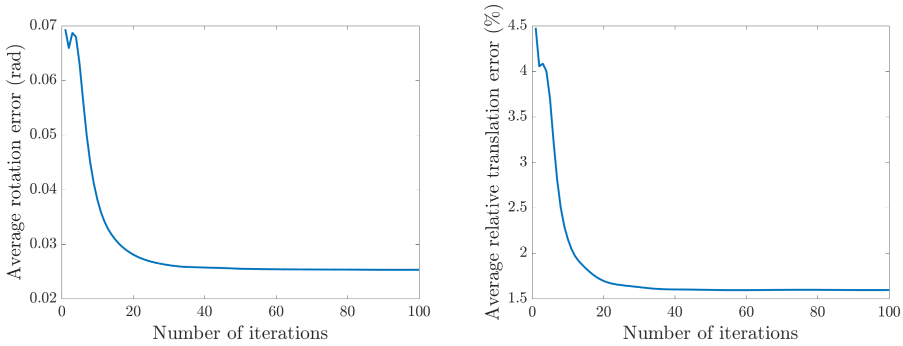
proposed methods can be distributed and parallelized with minimal efforts and with no compromise of efficiency. In the experiments, our proposed methods are much faster than existing techniques to converge to modest accuracy that is sufficient for practical use. Though not presented in this paper, our proposed methods can also be extended for incremental smoothing [21].



(a) `city10000` (b) `intel` (c) `ais2klinik`

(d) `garage` (e) `rim` (f) `sphere`

Fig. 3: The results of $\mathsf{AGPM-PGO}^*$ on some 2D and 3D SLAM benchmark datasets.



(a) (b)

Fig. 4: The results of distributed sensor network localization using $\mathsf{AGPM-PGO}$ over 30 runs. The results are (a) average rotation error and (b) average relative translation error. In all the 30 runs, $\mathsf{AGPM-PGO}$ converge to global optima.

References

1. Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*, 2016.

2. Amit Singer and Yoel Shkolnisky. Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming. *SIAM journal on imaging sciences*, 4(2):543–572, 2011.
3. Roberto Tron and René Vidal. Distributed image-based 3-D localization of camera sensor networks. In *IEEE Conference on Decision and Control (CDC)*, 2009.
4. Edwin Olson, John Leonard, and Seth Teller. Fast iterative alignment of pose graphs with poor initial estimates. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2006.
5. Giorgio Grisetti, Cyrill Stachniss, and Wolfram Burgard. Nonlinear constraint network optimization for efficient map learning. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):428–439, 2009.
6. Michael Kaess, Hordur Johannsson, Richard Roberts, Viorela Ila, John J Leonard, and Frank Dellaert. iSAM2: Incremental smoothing and mapping using the bayes tree. *The International Journal of Robotics Research*, 2012.
7. David Rosen, Michael Kaess, and John Leonard. RISE: An incremental trust-region method for robust online sparse least-squares estimation. *IEEE Transactions on Robotics*, 2014.
8. R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A general framework for graph optimization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3607–3613, Shanghai, China, May 2011.
9. David Rosen, Luca Carlone, Afonso Bandeira, and John Leonard. SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. *arXiv preprint arXiv:1612.07386*, 2016.
10. Taosha Fan, Hanlin Wang, Michael Rubenstein, and Todd Murphey. Efficient and guaranteed planar pose graph optimization using the complex number representation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.
11. Yurii Nesterov. A method for unconstrained convex minimization problem with the rate of convergence O (1/kˆ 2). In *Doklady AN USSR*, volume 269, pages 543–547, 1983.
12. Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.
13. Saeed Ghadimi and Guanghui Lan. Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Mathematical Programming*, 156(1-2):59–99, 2016.
14. Chi Jin, Praneeth Netrapalli, and Michael I Jordan. Accelerated gradient descent escapes saddle points faster than gradient descent. In *Conference On Learning Theory*, 2018.
15. Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.
16. Taosha Fan and Todd Murphey. Generalized proximal methods for pose graph optimization. [Online]. Available: https://northwestern.box.com/s/goaan0th6t6uteejvy02k1qksf9m6q59.
17. Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (4):376–380, 1991.
18. Aleka McAdams, Andrew Selle, Rasmus Tamstorf, Joseph Teran, and Eftychios Sifakis. Computing the singular value decomposition of 3x3 matrices with minimal branching and elementary floating point operations. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2011.
19. Brendan O'donoghue and Emmanuel Candes. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, 15(3):715–732, 2015.
20. Luca Carlone, Roberto Tron, Kostas Daniilidis, and Frank Dellaert. Initialization techniques for 3D SLAM: a survey on rotation estimation and its use in pose graph optimization. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.
21. Taosha Fan and Todd D Murphey. Fast incremental smoothing using generalized proximal methods. In preparation.

22. Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.
23. P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

A The $\mathsf{NAG-PGO}$ and $\mathsf{AGPM-PGO}$ Methods

A.1 The $\mathsf{NAG-PGO}$ Method

Algorithm 5 The $\mathsf{NAG-PGO}$ Method

Input: An initial iterate $X^{(0)} = \begin{bmatrix} t^{(0)} & R^{(0)} \end{bmatrix} \in \mathbb{R}^{d\times n} \times SO(d)^n$ and $X^{(-1)} = \begin{bmatrix} t^{(-1)} & R^{(-1)} \end{bmatrix} \in \mathbb{R}^{d\times n} \times SO(d)^n$, $s^{(0)} \in [1, +\infty)$, and the maximum number of iterations N.

Output: A sequence of iterates $\{X^{(k)},\, s^{(k)}\}$.

function $\mathsf{NAG-PGO}\,(X^{(0)},\, X^{(-1)},\, s^{(0)},\, N)$

    for $k = 0 \to N-1$ **do**

        $s^{(k+1)} \leftarrow \dfrac{\sqrt{4s^{(k)2}+1}+1}{2}, \quad Y^{(k)} \leftarrow X^{(k)} + \dfrac{s^{(k)}-1}{s^{(k+1)}}\left(X^{(k)} - X^{(k-1)}\right)$

        $\begin{bmatrix} \theta_1^{(k)} & \cdots & \theta_n^{(k)} \end{bmatrix} \leftarrow Y^{(k)}\Phi$

        for $i = 1 \to n$ **do**

            $R_i^{(k+1)} \leftarrow \arg\max\limits_{R_i \in SO(d)} \mathrm{trace}(R_i^\top \theta_i^{(k)})$

        end for

        $R^{(k+1)} \leftarrow \begin{bmatrix} R_1^{(k+1)} & \cdots & R_n^{(k+1)} \end{bmatrix}$

        $t^{(k+1)} \leftarrow R^{(k+1)}\Xi + X^{(k)}\Psi$

        $X^{(k+1)} \leftarrow \begin{bmatrix} t^{(k+1)} & R^{(k+1)} \end{bmatrix}$

    end for

    return $\{X^{(k)},\, s^{(k)}\}$

end function

A.2 The $\mathsf{AGPM-PGO}$ Method

Algorithm 6 The $\mathsf{AGPM-PGO}$ Method

Input: An initial iterate $X^{(0)} = \begin{bmatrix} t^{(0)} & R^{(0)} \end{bmatrix} \in \mathbb{R}^{d\times n} \times SO(d)^n$, the maximum number of outer iterations N, the maximum number of inner iterations N_0, $\eta \in (0,\, 1]$, and $\delta \in [0,\, \infty)$.

Output: A sequence of iterates $\{X^{(k)}\}$.

function $\mathsf{AGPM-PGO}(X^{(0)},\, N,\, N_0,\, \delta)$

    $a^{(0)} \leftarrow 1, \quad T^{(0)} \leftarrow X^{(0)}, \quad f^{(0)} \leftarrow F(X^{(0)})$

    for $k = 0 \to N-1$ **do**

        $\{V^{(i)},\, s^{(i)}\} \leftarrow \mathsf{NAG-PGO}^*(X^{(k)},\, T^{(k)},\, a^{(k)},\, N_0)$

        if $F(V^{(N_0)}) \leq f^{(k)} - \delta \cdot \|V^{(N_0)} - X^{(k)}\|^2$ **then**

            $a^{(k+1)} \leftarrow s^{(N_0)}, \quad X^{(k+1)} \leftarrow V^{(N_0)}, \quad T^{(k+1)} \leftarrow V^{(N_0-1)}$

        else

            $\{Z^{(i)}\} \leftarrow \mathsf{GPM-PGO}^*(X^{(k)},\, N_0)$

            $a^{(k+1)} \leftarrow 1, \quad X^{(k+1)} \leftarrow Z^{(N_0)}, \quad T^{(k+1)} \leftarrow Z^{(N_0)}$

        end if

        $f^{(k+1)} \leftarrow (1-\eta)\cdot f^{(k)} + \eta \cdot F(X^{(k+1)})$

    end for

    return $\{X^{(k)}\}$

end function

B Proofs

B.1 Proof of Theorem 1

If we let

$$F_{ij}^{R}(X)=\frac{1}{2}\|R_i\widetilde{R}_{ij}-R_j\|^2 \tag{B.1}$$

and

$$F_{ij}^{t}(X)=\frac{1}{2}\|R_i\tilde{t}_{ij}+t_i-t_j\|^2, \tag{B.2}$$

then we obtain

$$\nabla_{R_i}F_{ij}^{R}(X)=R_i-R_j\widetilde{R}_{ij}^{\top}, \tag{B.3a}$$

$$\nabla_{R_j}F_{ij}^{R}(X)=R_j-R_i\widetilde{R}_{ij}, \tag{B.3b}$$

and

$$\nabla_{t_i}F_{ij}^{t}(X)=R_i\tilde{t}_{ij}+t_i-t_j, \tag{B.4a}$$

$$\nabla_{R_i}F_{ij}^{t}(X)=\big(R_i\tilde{t}_{ij}+t_i-t_j\big)\tilde{t}_{ij}^{\top}, \tag{B.4b}$$

$$\nabla_{t_j}F_{ij}^{t}(X)=t_j-R_i\tilde{t}_{ij}-t_i. \tag{B.4c}$$

Note that $F_{ij}^{R}(X)$ and $F_{ij}^{t}(X)$ only depend on t_i, R_i, t_j and R_j, and as a result, $\nabla F_{ij}^{R}(X)$ and $\nabla F_{ij}^{t}(X)$ are well defined by Eqs. (B.3) and (B.4), respectively. Then, from Eqs. (B.1) to (B.4), it is straightforward to show that

$$\begin{aligned}
&\|R_i\widetilde{R}_{ij}-\frac{1}{2}R_i^{(k)}\widetilde{R}_{ij}-\frac{1}{2}R_j^{(k)}\|^2+\|R_j-\frac{1}{2}R_i^{(k)}\widetilde{R}_{ij}-\frac{1}{2}R_j^{(k)}\|^2\\
=&\|R_i-R_i^{(k)}\|^2+\|R_j-R_j^{(k)}\|^2+\mathrm{trace}\left(\left(R_i-R_i^{(k)}\right)^{\top}\nabla_{R_i}F_{ij}^{R}(X^{(k)})\right)+\\
&\mathrm{trace}\left(\left(R_j-R_j^{(k)}\right)^{\top}\nabla_{R_j}F_{ij}^{R}(X^{(k)})\right)+F_{ij}^{R}(X^{(k)})\\
=&\|R_i-R_i^{(k)}\|^2+\|R_j-R_j^{(k)}\|^2+\\
&\mathrm{trace}\left(\left(X-X_i^{(k)}\right)^{\top}\nabla F_{ij}^{R}(X^{(k)})\right)+F_{ij}^{R}(X^{(k)}),
\end{aligned} \tag{B.5}$$

and

$$\begin{aligned}
&\|R_i\tilde{t}_{ij}+t_i-\frac{1}{2}R_i^{(k)}\tilde{t}_{ij}-\frac{1}{2}t_i^{(k)}-\frac{1}{2}t_j^{(k)}\|^2+\|t_j-\frac{1}{2}R_i^{(k)}\tilde{t}_{ij}-\frac{1}{2}t_i^{(k)}-\frac{1}{2}t_j^{(k)}\|^2\\
=&\|(R_i-R_i^{(k)})\tilde{t}_{ij}+t_i-t_i^{(k)}\|^2+\|t_j-t_j^{(k)}\|^2+\left(t_i-t_i^{(k)}\right)^{\top}\nabla_{t_i}F_{ij}^{t}(X^{(k)})+\\
&\mathrm{trace}\left(\left(R_i-R_i^{(k)}\right)^{\top}\nabla_{R_i}F_{ij}^{t}(X^{(k)})\right)+\left(t_j-t_j^{(k)}\right)^{\top}\nabla_{t_j}F_{ij}^{t}(X^{(k)})+F_{ij}^{t}(X^{(k)})\\
=&\|(R_i-R_i^{(k)})\tilde{t}_{ij}+t_i-t_i^{(k)}\|^2+\|t_j-t_j^{(k)}\|^2+\\
&\mathrm{trace}\left(\left(X-X_i^{(k)}\right)^{\top}\nabla F_{ij}^{t}(X^{(k)})\right)+F_{ij}^{t}(X^{(k)}),
\end{aligned} \tag{B.6}$$

It is by definition that

$$F(X) = \sum_{(i,j)\in\overrightarrow{\mathcal{E}}} \left(\kappa_{ij}\cdot F_{ij}^R(X) + \tau_{ij}\cdot F_{ij}^t(X)\right), \tag{B.7}$$

and

$$\nabla F(X) = \sum_{(i,j)\in\overrightarrow{\mathcal{E}}} \left(\kappa_{ij}\cdot \nabla F_{ij}^R(X) + \tau_{ij}\cdot \nabla F_{ij}^t(X)\right). \tag{B.8}$$

Substitute Eqs. (B.5) and (B.6) into Eq. (10) and simplify the resulting equation with Eqs. (B.7) and (B.8), the result is

$$F(X^{(k)}) + \operatorname{trace}\left((X - X^{(k)})^\top \nabla F(X^{(k)})\right) + \\ \frac{1}{2}\operatorname{trace}\left((X - X^{(k)})\widetilde{\Omega}(X - X^{(k)})^\top\right), \tag{B.9}$$

in which $\widetilde{\Omega} \triangleq \begin{bmatrix} \Omega^\tau & \widetilde{\Omega}^{\nu\top} \\ \widetilde{\Omega}^\nu & \widetilde{\Omega}^\rho \end{bmatrix}$ with $\widetilde{\Omega}^\tau = \operatorname{diag}\{\widetilde{\Omega}_1^\tau, \cdots, \widetilde{\Omega}_n^\tau\} \in \mathbb{R}^{n\times n}$, $\widetilde{\Omega}^\rho = \operatorname{diag}\{\widetilde{\Omega}_1^\rho, \cdots, \widetilde{\Omega}_n^\rho\} \in \mathbb{R}^{dn\times dn}$ and $\widetilde{\Omega}^\nu = \operatorname{diag}\{\widetilde{\Omega}_1^\nu, \cdots, \widetilde{\Omega}_n^\nu\} \in \mathbb{R}^{dn\times n}$, in which

$$\Omega_i^\tau = \sum_{(i,j)\in\mathcal{E}} 2\cdot\tau_{ij} \in \mathbb{R}, \tag{B.10a}$$

$$\widetilde{\Omega}_i^\rho = \sum_{(i,j)\in\mathcal{E}} 2\cdot\kappa_{ij}\cdot\mathbf{I} + \sum_{(i,j)\in\overrightarrow{\mathcal{E}}} 2\cdot\tau_{ij}\cdot\tilde{t}_{ij}\tilde{t}_{ij}^\top \in \mathbb{R}^{d\times d}, \tag{B.10b}$$

$$\widetilde{\Omega}_i^\nu = \sum_{(i,j)\in\overrightarrow{\mathcal{E}}} 2\cdot\tau_{ij}\cdot\tilde{t}_{ij} \in \mathbb{R}^d. \tag{B.10c}$$

The proof is completed.

B.2 Proof of Theorem 2

Proof of (1) It is without loss of any generality to reformulate Eq. (7) as

$$F(X) = F(X^{(k)}) + \operatorname{trace}\left((X - X^{(k)})^\top \nabla F(X^{(k)})\right) + \\ \frac{1}{2}\operatorname{trace}\left((X - X^{(k)})\widetilde{M}\left(X - X^{(k)}\right)^\top\right). \tag{B.11}$$

Note that (B.9) is an upper bound of $F(X)$ and Eq. (B.11), and as a result, we obtain $\widetilde{\Omega} \succeq \widetilde{M}$, which completes the proof of (1).

Proof of (2) From Eqs. (B.10a) to (B.10c), if we reorder $X = \begin{bmatrix} t_1 \cdots t_n \ R_1 \cdots R_n \end{bmatrix}$ to $X' = \begin{bmatrix} t_1 \ R_1 \ t_2 \ R_2 \cdots t_n \ R_n \end{bmatrix}$, then $\widetilde{\Omega}$ is accordingly reordered to a block diagonal $\widetilde{\Omega}' \triangleq \mathrm{diag}\{\widetilde{\Omega}'_1, \cdots, \widetilde{\Omega}'_n\} \in \mathbb{R}^{(d+1)n\times(d+1)n}$, in which $\widetilde{\Omega}'_i \in \mathbb{R}^{(d+1)\times(d+1)}$ are the principal minors of $2\cdot\widetilde{M}$. Let $\lambda_{\max}(\widetilde{\Omega}'_i)$, $\lambda_{\max}(\widetilde{\Omega})$, $\lambda_{\max}(\widetilde{M})$, etc., be the greatest eigenvalue of corresponding matrices. As a result of Courant-Fischer theorem [22, Theorem 4.2.6], it is straightforward to show $\lambda_{\max}(\widetilde{\Omega}'_i) \leq 2\cdot\lambda_{\max}(\widetilde{M})$, from which we further obtain $\lambda_{\max}(\widetilde{\Omega}) = \lambda_{\max}(\widetilde{\Omega}') \leq 2\cdot\lambda_{\max}(\widetilde{M})$. Then, for any $c\in\mathbb{R}$, if $\frac{c}{2}\cdot\mathbf{I}\succeq\widetilde{M}$, we obtain $c \geq 2\cdot\lambda_{\max}(\widetilde{M})$, and thus, $c \geq \lambda_{\max}(\widetilde{\Omega})$ and $c\cdot\mathbf{I}\succeq\widetilde{\Omega}$, which completes the proof of (2).

B.3 Proof of Theorem 3

Proof of (1) For $\mathsf{GPM}-\mathsf{PGO}$, it should be noted that we define $G(X|X^{(k)})$ as

$$G(X|X^{(k)}) = F(X^{(k)}) + \mathrm{trace}\left((X-X^{(k)})^\top \nabla F(X^{(k)})\right) + \frac{1}{2}\,\mathrm{trace}\left((X-X^{(k)})\widetilde{\Gamma}(X-X^{(k)})^\top\right) \quad \text{(B.12)}$$

in Eq. (13), with which Eq. (13) is equivalent to

$$X^{(k+1)} = \arg\min_{X\in\mathbb{R}^{d\times n}\times SO(d)^n} G(X|X^{(k)}). \quad \text{(B.13)}$$

Since $G(X|X^{(k)})$ is an upper bound of $F(X)$ that attains the same value with $F(X)$ at $X^{(k)}$ and $X^{(k+1)}$ minimizes $G(X|X^{(k)})$, it can be concluded that

$$F(X^{(k+1)}) \leq G(X^{(k+1)}|X^{(k)}) \leq G(X^{(k)}|X^{(k)}) = F(X^{(k)}), \quad \text{(B.14)}$$

which suggests that $\mathsf{GPM}-\mathsf{PGO}$ is non-increasing.

For $\mathsf{GPM}-\mathsf{PGO}^*$, if we substitute

$$t = -R\widetilde{V}^\top L(W^\tau)^\dagger \quad \text{(B.15)}$$

into Eq. (7) and simplify the resulting equation, we obtain

$$\min_{R\in SO(d)^n} F'(R) \triangleq \frac{1}{2}\,\mathrm{trace}(R\widetilde{M}_R R^\top), \quad \text{(B.16)}$$

in which

$$\widetilde{M}_R = L(\widetilde{G}^\rho) + \widetilde{\Sigma}^\rho - \widetilde{V}^\top L(W^\tau)^\dagger \widetilde{V}. \quad \text{(B.17)}$$

For each iterate in $\mathsf{GPM}-\mathsf{PGO}^*$, from Eq. (24), it is straight to show that $F'(R^{(k)}) = F(X^{(k)})$, $\nabla F'(R^{(k)}) = \nabla_R F(X^{(k)})$ and $\nabla_t F(X^{(k)}) = \mathbf{0}$, and as a result, $G(X|X^{(k)})$ can be simplified to

$$G(X|X^{(k)}) = F'(R^{(k)}) + \mathrm{trace}\left(\nabla F'(R^{(k)})^\top (R-R^{(k)})\right) + \frac{1}{2}\,\mathrm{trace}\left((X-X^{(k)})\widetilde{\Gamma}(X-X^{(k)})^\top\right). \quad \text{(B.18)}$$

If we substitute Eq. (17) into Eq. (B.18) and marginalize out $t \in \mathbb{R}^{d\times n}$, we obtain

$$G'(R|R^{(k)}) = F'(R^{(k)}) + \mathrm{trace}\left(\nabla F'(R^{(k)})^\top (R - R^{(k)})\right) + \\ \frac{1}{2}\,\mathrm{trace}\left(\left(R - R^{(k)}\right)\widetilde{\Gamma}'\left(R - R^{(k)}\right)^\top\right), \tag{B.19}$$

in which

$$\widetilde{\Gamma}' = \widetilde{\Gamma}^\rho - \widetilde{\Gamma}^{\nu\top}\Gamma^{\tau-1}\widetilde{\Gamma}^\nu$$

is positive semidefinte, and it should be noted that

$$G'(R^{(k)}|R^{(k)}) = F'(R^{(k)}) = F(X^{(k)}).$$

Furthermore, Eq. (18) is equivalent to

$$R^{(k+1)} = \arg\min_{R\in SO(d)^n} G'(R|R^{(k)}), \tag{B.20}$$

and it is by definition that

$$F(X^{(k+1)}) = F'(R^{(k+1)}) \leq G'(R^{(k+1)}|R^{(k)}) \leq F'(R^{(k)}) = F(X^{(k)}), \tag{B.21}$$

which suggests that $\mathsf{GPM}-\mathsf{PGO}^*$ is non-increasing.

From Eqs. (B.14) and (B.21), it can be concluded that $F(X^{(k)})$ is non-increasing, which completes the proof of (1).

Proof of (2) From (1) of Theorem 3, it is known that $F(X^{(k)})$ is non-increasing. Furthermore, $F(X)$ is bounded below, and as a result, there exists $F^\infty \in \mathbb{R}$ such that $F(X^{(k)}) \to F^\infty$, which completes the proof of (2).

Proof of (3) For $\mathsf{GPM}-\mathsf{PGO}$, it is from Eqs. (B.12) and (B.13) that

$$F(X^{(k)}) + \mathrm{trace}\left(\left(X^{(k+1)} - X^{(k)}\right)^\top \nabla F(X^{(k)})\right) + \\ \frac{1}{2}\,\mathrm{trace}\left(\left(X^{(k+1)} - X^{(k)}\right)\widetilde{\Gamma}\left(X^{(k+1)} - X^{(k)}\right)^\top\right) \leq F(X^{(k)}),$$

and from Eq. (B.11), $F(X^{(k+1)})$ is equivalent to

$$F(X^{(k+1)}) = F(X^{(k)}) + \mathrm{trace}\left(\left(X^{(k+1)} - X^{(k)}\right)^\top \nabla F(X^{(k)})\right) + \\ \frac{1}{2}\,\mathrm{trace}\left(\left(X^{(k+1)} - X^{(k)}\right)\widetilde{M}\left(X^{(k+1)} - X^{(k)}\right)^\top\right),$$

which implies

$$F(X^{(k+1)}) - F(X^{(k)}) \leq \\ \frac{1}{2}\,\mathrm{trace}\left(\left(X^{(k+1)} - X^{(k)}\right)(\widetilde{M} - \widetilde{\Gamma})\left(X^{(k+1)} - X^{(k)}\right)^\top\right). \tag{B.22}$$

If $\widetilde{\Gamma} \succ \widetilde{M}$, there exists $\epsilon > 0$ such that $\widetilde{\Gamma} \succeq \widetilde{M} + \epsilon \cdot \mathbf{I}$. Then, from Eq. (B.22), it can be concluded that

$$F(X^{(k+1)}) - F(X^{(k)}) \leq -\epsilon \cdot \|X^{(k+1)} - X^{(k)}\|^2. \tag{B.23}$$

For $\mathsf{GPM-PGO}^*$, it is from Eqs. (B.19) and (B.20) that

$$F'(R^{(k)}) + \operatorname{trace}\left(\left(R^{(k+1)} - R^{(k)}\right)^\top \nabla F'(X^{(k)})\right) + \frac{1}{2}\operatorname{trace}\left(\left(R^{(k+1)} - R^{(k)}\right)\widetilde{\Gamma}'\left(R^{(k+1)} - R^{(k)}\right)^\top\right) \leq F'(R^{(k)}),$$

and from Eq. (B.16), $F(R^{(k+1)})$ is equivalent to

$$F'(R^{(k+1)}) = F'(R^{(k)}) + \operatorname{trace}\left(\left(R^{(k+1)} - R^{(k)}\right)^\top \nabla F'(R^{(k)})\right) + \frac{1}{2}\operatorname{trace}\left(\left(R^{(k+1)} - R^{(k)}\right)\widetilde{M}'\left(R^{(k+1)} - R^{(k)}\right)^\top\right),$$

which implies

$$F'(R^{(k+1)}) - F'(R^{(k)}) \leq \frac{1}{2}\operatorname{trace}\left(\left(R^{(k+1)} - R^{(k)}\right)(\widetilde{M}' - \widetilde{\Gamma}')\left(R^{(k+1)} - R^{(k)}\right)^\top\right). \tag{B.24}$$

From $\widetilde{\Gamma} \succ \widetilde{M} \succeq 0$, it is straightforward to show that $\widetilde{\Gamma}' \succ \widetilde{M}'$ and there exists $\epsilon_R > 0$ such that $\widetilde{\Gamma}' \succeq \widetilde{M}' + \epsilon_R \cdot \mathbf{I}$, and as a result, we obtain

$$F'(R^{(k+1)}) - F'(R^{(k)}) \leq -\epsilon_R \cdot \|R^{(k+1)} - R^{(k)}\|^2. \tag{B.25}$$

Furthermore, from Eqs. (B.25) and (24), it can be shown that there exists $\epsilon > 0$ such that

$$\epsilon_R \cdot \|R^{(k+1)} - R^{(k)}\|^2 \geq \epsilon \cdot \|X^{(k+1)} - X^{(k)}\|^2. \tag{B.26}$$

As a result of $F(X^{(k)}) = F'(R^{(k)})$ and $F(X^{(k+1)}) = F'(R^{(k+1)})$ and Eqs. (B.25) and (B.26), we obtain

$$F(X^{(k+1)}) - F(X^{(k)}) \leq -\epsilon \cdot \|X^{(k+1)} - X^{(k)}\|^2. \tag{B.27}$$

From (2) of Theorem 3, we obtain $F(X^{(k)}) \rightarrow F^\infty$. Then, as a result of Eqs. (B.23) and (B.27), it can be shown that for $\mathsf{GPM-PGO}$ and $\mathsf{GPM-PGO}^*$, $\|X^{(k+1)} - X^{(k)}\| \rightarrow 0$ as $k \rightarrow \infty$, which completes the proof of (3).

Proof of (4) For $\mathsf{GPM-PGO}$, from Riemannian optimization [9, 23], if we assume that the Euclidean gradient is $\nabla F(X) = \begin{bmatrix}\nabla_t F(X) & \nabla_R F(X)\end{bmatrix}$, in which $\nabla_t F(X) \in \mathbb{R}^{d\times n}$ and $\nabla_R F(X) \in \mathbb{R}^{d\times dn}$ correspond to the translation $t = \begin{bmatrix}t_1 & \cdots & t_n\end{bmatrix} \in \mathbb{R}^{d\times n}$ and the rotation $R = \begin{bmatrix}R_1 & \cdots & R_n\end{bmatrix} \in SO(d)^n$, respectively, then the Riemannian gradient $\operatorname{grad} F(X)$ can be computed as

$$\operatorname{grad} F(X) = \begin{bmatrix}\operatorname{grad}_t F(X) & \operatorname{grad}_R F(X)\end{bmatrix}, \tag{B.28}$$

in which

$$\text{grad}\,_t F(X) = \nabla_t F(X) \tag{B.29}$$

corresponds to the translation t, and

$$\text{grad}\,_R F(X) = \nabla_R F(X) - R\ \text{SymBlockDiag}_d(R^\top \nabla_R F(X)) \tag{B.30}$$

corresponds to the rotation R. In Eq. (B.30), $\text{SymBlockDiag}_d : \mathbb{R}^{dn\times dn} \to \mathbb{R}^{dn\times dn}$ is a linear operator

$$\text{SymBlockDiag}_d(Z) \triangleq \frac{1}{2}\text{BlockDiag}_d(Z + Z^\top),$$

in which $\text{BlockDiag}_d : \mathbb{R}^{dn\times dn} \to \mathbb{R}^{dn\times dn}$ is also a a linear operator that extracts the $(d \times d)$-block diagonals of a matrix, i.e.,

$$\text{BlockDiag}_d(Z) \triangleq \begin{bmatrix} Z_{11} & & \\ & \ddots & \\ & & Z_{nn} \end{bmatrix}.$$

As a result, Eqs. (B.28) to (B.30) result in a linear operator $\mathcal{Q}_X : \mathbb{R}^{d\times(d+1)n} \to \mathbb{R}^{d\times(d+1)n}$ that depends on X such that the Riemannian gradient $\text{grad}\, F(X)$ and the Euclidean gradient $\nabla F(X)$ are related as

$$\text{grad}\, F(X) = \mathcal{Q}_X\big(\nabla F(X)\big). \tag{B.31}$$

Similarly, for $G(X|X^{(k)})$ in Eq. (B.12), the Riemannian gradient $\text{grad}\, G(X|X^{(k)})$ is

$$\begin{aligned}
\text{grad}\, G(X|X^{(k)}) =& \mathcal{Q}_X\big(\nabla F(X^{(k)})\big) + \mathcal{Q}_X\big((X - X^{(k)})\widetilde{\Gamma}\big)\\
=& \mathcal{Q}_X\big(\nabla F(X)\big) + \mathcal{Q}_X\big(\nabla F(X^{(k)}) - \nabla F(X)\big) + \\
& \mathcal{Q}_X\big((X - X^{(k)})\widetilde{\Gamma}\big)\\
=& \text{grad}\, F(X) + \mathcal{Q}_X\big(\nabla F(X^{(k)}) - \nabla F(X)\big) + \\
& \mathcal{Q}_X\big((X - X^{(k)})\widetilde{\Gamma}\big),
\end{aligned}$$

and thus, we obtain

$$\begin{aligned}
\text{grad}\, G(X^{(k+1)}|X^{(k)}) =& \text{grad}\, F(X^{(k+1)}) + \mathcal{Q}_{X^{(k+1)}}\big(\nabla F(X^{(k)}) - \nabla F(X^{(k+1)})\big) + \\
& \mathcal{Q}_{X^{(k+1)}}\big((X^{(k+1)} - X^{(k)})\widetilde{\Gamma}\big).
\end{aligned} \tag{B.32}$$

It is known that $\nabla F(X)$ is Lipschitz continuous, then there exists $L > 0$ such that

$$\|\mathcal{Q}_{X^{(k+1)}}\big(\nabla F(X^{(k)}) - \nabla F(X^{(k+1)})\big)\| \leq L \cdot \|\mathcal{Q}_{X^{(k+1)}}\|_2 \cdot \|X^{(k)} - X^{(k+1)}\|, \tag{B.33}$$

in which $\|\cdot\|_2$ denotes the induced 2-norm of linear operators. From Eqs. (B.28) to (B.30), note that $\mathcal{Q}_X$ only depends on $R \in SO(d)^n$ and $SO(d)^n$ is a compact manifold, and thus, $\|\mathcal{Q}_X\|_2$ is bounded. Then, from (3) of Theorem 3 that $\|X^{(k+1)} - X^{(k)}\| \to 0$ if $\widetilde{\Gamma} \succ \widetilde{M}$ and Eq. (B.33), we obtain

$$\mathcal{Q}_{X^{(k+1)}}\big(\nabla F(X^{(k)}) - \nabla F(X^{(k+1)})\big) \to \mathbf{0}. \tag{B.34}$$

Similarly, since $\widetilde{\Gamma}$ is also bounded, we obtain

$$\mathcal{Q}_{X^{(k+1)}}\big((X^{(k+1)} - X^{(k)})\widetilde{\Gamma}\big) \rightarrow \mathbf{0} \tag{B.35}$$

From Eqs. (B.32), (B.34) and (B.35), it can be concluded that

$$\operatorname{grad} F(X^{(k+1)}) \rightarrow \operatorname{grad} G(X^{(k+1)}|X^{(k)}). \tag{B.36}$$

Note that $X^{(k+1)}$ minimizes $G(X|X^{(k)})$, and thus, $\operatorname{grad} G(X^{(k+1)}) = \mathbf{0}$ always holds, then from Eq. (B.36), we obtain

$$\operatorname{grad} F(X^{(k+1)}) \rightarrow \mathbf{0}$$

as $k \rightarrow \infty$, which completes the proof of (4).

For $\mathsf{GPM-PGO}^*$, the Riemannian gradient of $F'(R)$ in Eq. (B.16) is

$$\operatorname{grad} F'(R) = \nabla F'(R) - R\ \mathrm{SymBlockDiag}_d(R^\top \nabla F'(R)). \tag{B.37}$$

From Eq. (B.19), we obtain

$$\begin{aligned}\operatorname{grad} G'(R|R^{(k)}) =& \mathcal{Q}_R\big(\nabla F'(R^{(k)})\big) + \mathcal{Q}_R\big((R - R^{(k)})\widetilde{\Gamma}'\big)\\ =& \mathcal{Q}_R\big(\nabla F'(R)\big) + \mathcal{Q}_R\big(\nabla F'(R^{(k)}) - \nabla F'(R)\big) +\\ & \mathcal{Q}_R\big((R - R^{(k)})\widetilde{\Gamma}'\big)\\ =& \operatorname{grad} F'(R) + \mathcal{Q}_R\big(\nabla F'(R^{(k)}) - \nabla F'(R)\big) +\\ & \mathcal{Q}_R\big((R - R^{(k)})\widetilde{\Gamma}'\big),\end{aligned}$$

in which $\mathcal{Q}_R : \mathbb{R}^{d\times dn} \rightarrow \mathbb{R}^{d\times dn}$ is the linear operator defined by Eq. (B.37), and we obtain

$$\begin{aligned}\operatorname{grad} G'(R^{(k+1)}|R^{(k)}) =& \operatorname{grad} F'(R^{(k+1)}) + \mathcal{Q}_{R^{(k+1)}}\big(\nabla F'(R^{(k)}) - \nabla F'(R^{(k+1)})\big) +\\ & \mathcal{Q}_{R^{(k+1)}}\big((R^{(k+1)} - R^{(k)})\widetilde{\Gamma}'\big).\end{aligned} \tag{B.38}$$

Similar to $\mathsf{GPM-PGO}$, it can be shown that

$$\operatorname{grad} F'(R^{(k+1)}) \rightarrow \operatorname{grad} G'(R^{(k+1)}|R^{(k)}), \tag{B.39}$$

and since it is by definition that $\operatorname{grad} G'(R^{(k+1)}|R^{(k)}) \rightarrow \mathbf{0}$, we can conclude that $\operatorname{grad} F'(R^{(k+1)}) \rightarrow \mathbf{0}$. Moreover, as mentioned in the proof of (1), we have $\nabla F'(R^{(k)}) = \nabla_R F(X^{(k)})$ and $\nabla_t F(X^{(k)}) = \mathbf{0}$, which further indicates that $\operatorname{grad} F(X^{(k)}) \rightarrow \mathbf{0}$.

Proof of (5) and (6) Note that $\widetilde{\Gamma} = \alpha \cdot \mathbf{I} + \widetilde{\Omega} \succ \widetilde{M}$ if $\alpha > 0$. Then, from (3) and (4) of Theorem 3, it can be concluded that (5) and (6) hold as long as $\alpha > 0$, which completes the proof of (5) and (6).

B.4 Proof of Theorem 4

Proof of (1) Even though Algorithm 4 is not necessarily a descent algorithm, we can still prove $f^{(k+1)} \leq f^{(k)}$ and $F(X^{(k)}) \leq f^{(k)}$ by induction. Note that $f^{(0)} = F(X^{(0)})$, and thus, $F(X^{(0)}) \leq f^{(0)}$. If $X^{(k+1)}$ is generated from $\mathsf{NAG-PGO}$ or $\mathsf{NAG-PGO}^*$, we obtain $F(X^{(k+1)}) \leq f^{(k)}$; otherwise, $X^{(k+1)}$ is generated from $\mathsf{GPM-PGO}$ or $\mathsf{GPM-PGO}^*$, and according to Theorem 3, we obtain $F(X^{(k+1)}) \leq F(X^{(k)})$, from which it can be further shown that $F(X^{(k+1)}) \leq f^{(k)}$ as long as $F(X^{(k)}) \leq f^{(k)}$. If $F(X^{(k+1)}) \leq f^{(k)}$, we obtain $f^{(k+1)} = (1-\eta)\cdot f^{(k)} + \eta \cdot F(X^{(k+1)}) \leq f^{(k)}$ and $F(X^{(k+1)}) \leq f^{(k+1)}$ for any $\eta \in (0, 1]$. As a result, it can be concluded that $f^{(k+1)} \leq f^{(k)}$ and $F(X^{(k)}) \leq f^{(k)}$. Furthermore, $f^{(k)}$ is a actually convex combination of $F(X^{(0)}), \cdots, F(X^{(k)})$, and $F(X)$ is bounded below, and thus, $f^{(k)}$ is also bounded below and there exists F^∞ such that $f^{(k)} \to F^\infty$. Since $f^{(k)} \to F^\infty$, we obtain $(1-\eta)\cdot f^{(k)} + \eta \cdot F(X^{(k)}) \to F^\infty$ as well, then it can be concluded that $F(X^{(k)}) \to F^\infty$ as $k \to \infty$, which completes the proof of (1).

Proof of (2) If $\widetilde{\Gamma} \succ \widetilde{M}$, there exists $\epsilon > 0$ such that $\widetilde{\Gamma} \succeq \widetilde{M} + \epsilon \cdot \mathbf{I}$. From Algorithms 4 and 6, it can be concluded that

$$F(X^{(k+1)}) \leq f^{(k)} - \delta \cdot \|X^{(k+1)} - X^{(k)}\|^2$$

if $X^{(k+1)}$ is from $\mathsf{NAG-PGO}$ or $\mathsf{NAG-PGO}^*$, or

$$F(X^{(k+1)}) \leq f^{(k)} - \epsilon \cdot \|X^{(k+1)} - X^{(k)}\|^2$$

if $X^{(k+1)}$ is from $\mathsf{GPM-PGO}$ or $\mathsf{GPM-PGO}^*$. As a result, we obtain

$$F(X^{(k+1)}) \leq f^{(k)} - \phi \cdot \|X^{(k+1)} - X^{(k)}\|^2, \tag{B.40}$$

in which $\phi = \min\{\delta, \epsilon\}$. From Eq. (B.40) and $f^{(k+1)} = (1-\eta)\cdot f^{(k)} + \eta\cdot F(X^{(k)})$, we obtain

$$f^{(k+1)} \leq f^{(k)} - \eta \cdot \phi \cdot \|X^{(k+1)} - X^{(k)}\|^2.$$

Since $f^{(k)} \to F^\infty$ and $\eta, \phi > 0$, it can concluded that $\|X^{(k+1)} - X^{(k)}\| \to 0$ as $k \to \infty$, which completes the proof of (2).

Proof of (3) For $\mathsf{AGPM-PGO}$, if the inner number of iterations $N_0 = 1$, then for $\mathsf{NAG-PGO}$, $X^{(k+1)}$ is actually evaluated as

$$X^{(k+1)} = \arg \min_{X \in \mathbb{R}^{d\times n} \times SO(d)^n} G(X|Y^{(k)}), \tag{B.41}$$

in which

$$Y^{(k)} = X^{(k)} + \frac{s^{(k)} - 1}{s^{(k+1)}}\left(X^{(k)} - X^{(k-1)}\right). \tag{B.42}$$

From Eq. (B.32), we obtain

$$\begin{aligned}\operatorname{grad} G(X^{(k+1)}|Y^{(k)}) =& \operatorname{grad} F(X^{(k+1)}) + \mathcal{Q}_{X^{(k+1)}}\big(\nabla F(Y^{(k)}) - \nabla F(X^{(k+1)})\big) + \\ & \mathcal{Q}_{X^{(k+1)}}\big((X^{(k+1)} - Y^{(k)})\widetilde{\Gamma}\big).\end{aligned} \tag{B.43}$$

From Algorithms 3 and 5, note that $s^{(k)} \geq 1$, and

$$0 \leq \frac{s^{(k)}-1}{s^{k+1}} = \frac{2s^{(k)}-2}{\sqrt{4s^{(k)2}+1}+1} \leq \frac{2s^{(k)}-2}{2s^{(k)}} \leq 1. \tag{B.44}$$

From Eqs. (B.42) and (B.44), we obtain

$$\|X^{(k)} - Y^{(k)}\| = \frac{s^{(k)}-1}{s^{(k+1)}}\|X^{(k)} - X^{(k-1)}\| \leq \|X^{(k)} - X^{(k-1)}\|. \tag{B.45}$$

From (2) of Theorem 4 and Eq. (B.45), we obtain $\|X^{(k)} - Y^{(k)}\| \to 0$. In addition, from the triangular inequality of the Frobenius norm, it can be further concluded that

$$\|X^{(k+1)} - Y^{(k)}\| \to 0 \tag{B.46}$$

Similar to the proof of (4) of Theorem 3, from Eq. (B.46), we obtain

$$\mathcal{Q}_{X^{(k+1)}}\big(\nabla F(Y^{(k)}) - \nabla F(X^{(k+1)})\big) \to \mathbf{0}, \tag{B.47}$$

and

$$\mathcal{Q}\big((X^{(k+1)} - Y^{(k)})\widetilde{\Gamma}\big) \to \mathbf{0}. \tag{B.48}$$

From Eqs. (B.43), (B.47) and (B.48), it can be concluded that

$$\operatorname{grad} F(X^{(k+1)}) \to \operatorname{grad} G(X^{(k+1)}|Y^{(k)}).$$

In addition, from Eq. (B.41), we obtain $\operatorname{grad} G(X^{(k+1)}|Y^{(k)}) = \mathbf{0}$, which implies

$$\operatorname{grad} F(X^{(k+1)}) \to \mathbf{0}. \tag{B.49}$$

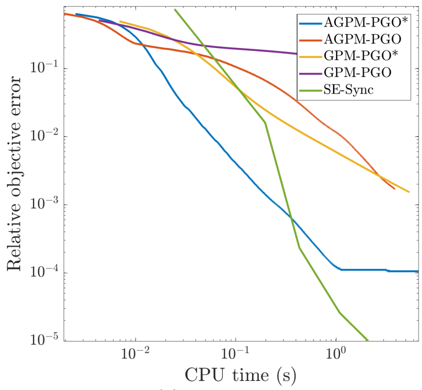
Thus, it can be concluded that $\operatorname{grad} F(X^{(k+1)}) \to \mathbf{0}$ if $X^{(k+1)}$ is generated from $\mathsf{NAG-PGO}$. Following the same proof of (4) of Theorem 3, we can also prove that $\operatorname{grad} F(X^{(k+1)}) \to \mathbf{0}$ if $X^{(k+1)}$ is generated from $\mathsf{GPM-PGO}$. As a result, $\operatorname{grad} F(X^{(k)}) \to \mathbf{0}$ as $k \to \infty$, which completes the proof of (3).

For $\mathsf{AGPM-PGO}^*$, the proof is similar to that of $\mathsf{AGPM-PGO}$, which is omitted due to space limitation.

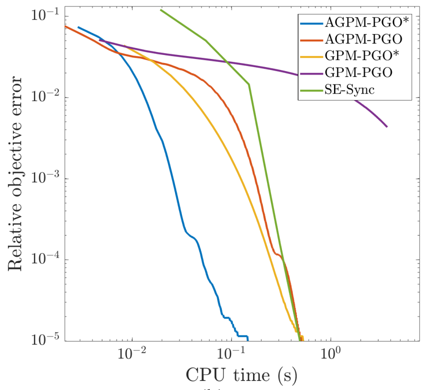
Proof of (4) and (5) Note that $\widetilde{\Gamma} = \alpha \cdot \mathbf{I} + \widetilde{\Omega} \succ \widetilde{M}$ if $\alpha > 0$. Then, the proofs of (4) and (5) are implementation of (2) and (3) of Theorem 4, respectively.

C Experiment Results

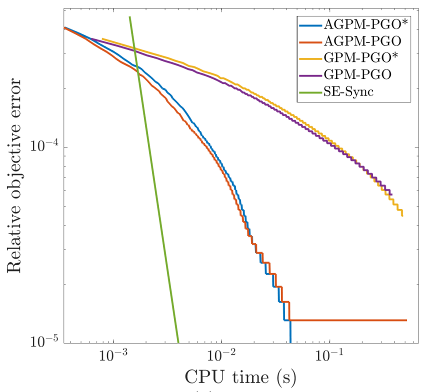
C.1 The Comparisons of Convergence on SLAM Datasets



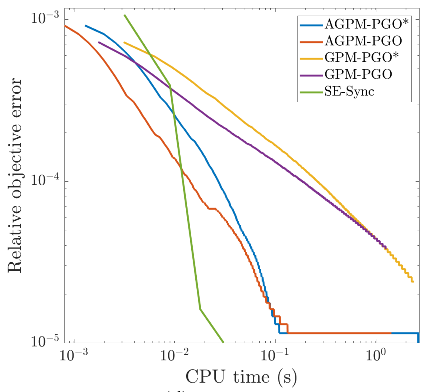
(a) `ais2klinik`



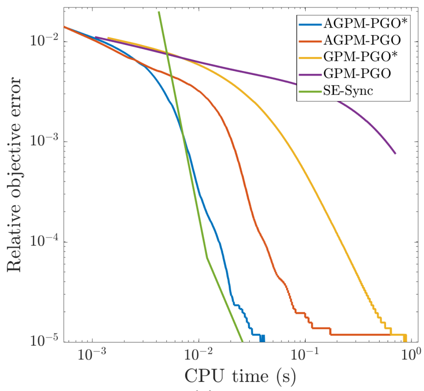
(b) `city`



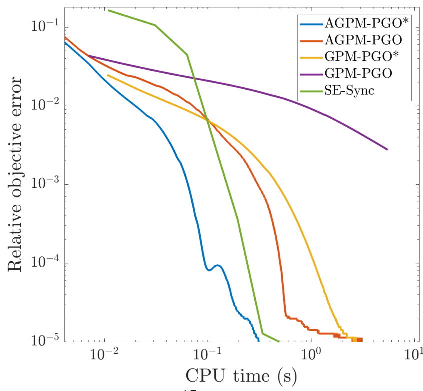
(c) `CSAIL`



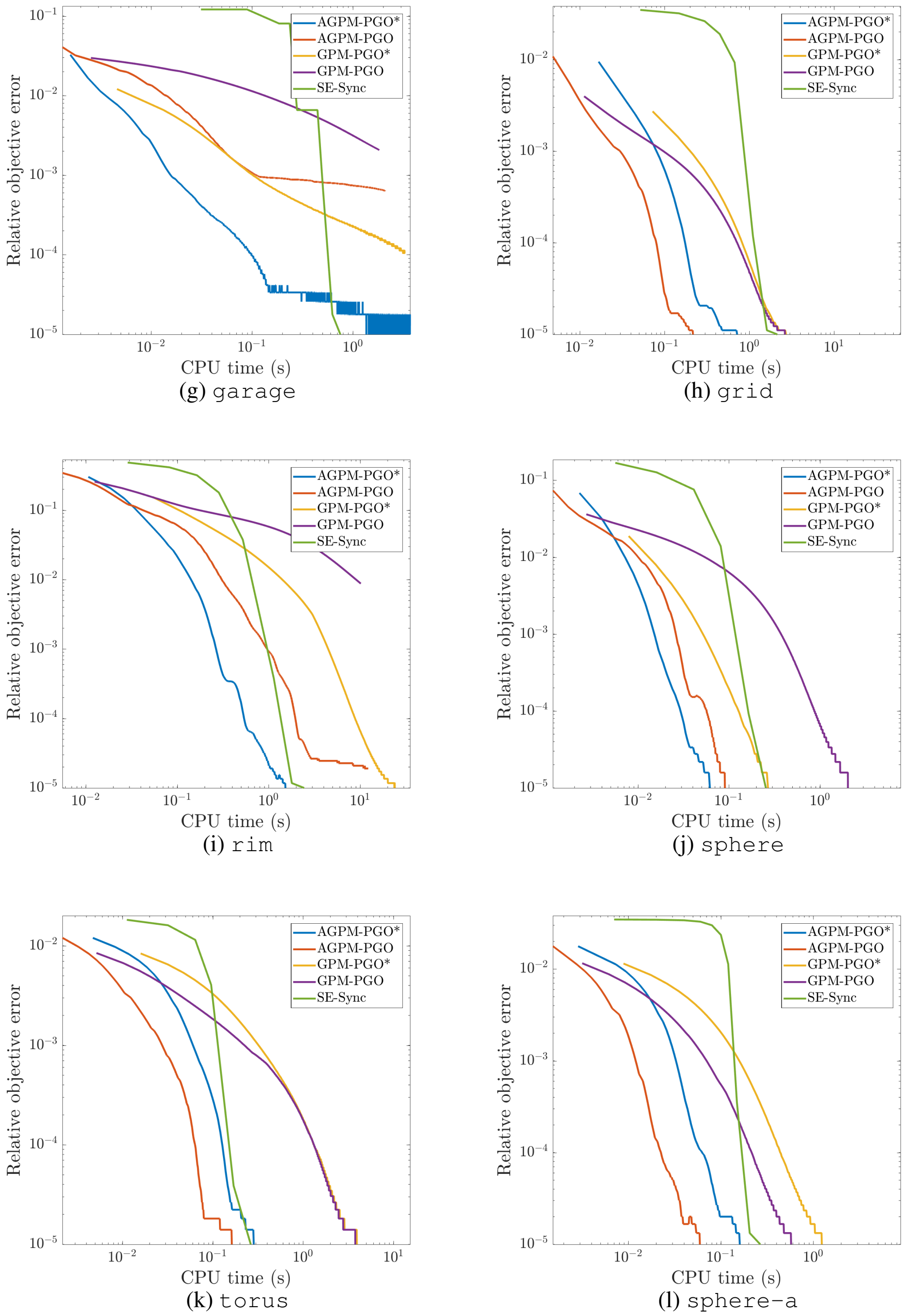
(d) `manhattan`



(e) `intel`



(f) `cubicle`



(g) `garage` (h) `grid`

(i) `rim` (j) `sphere`

(k) `torus` (l) `sphere-a`

Fig. 5: The convergence comparison of AGPM − PGO*, AGPM − PGO, GPM − PGO*, GPM − PGO and SE − Sync on 2D and 3D SLAM datasets.