Implementing Robust M-Estimators with Certifiable Factor Graph Optimization

Zhexin Xu*, Hanna Jiamei Zhang, Helena Calatrava, Pau Closas, and David M. Rosen

***Abstract*— Parameter estimation in robotics and computer vision faces formidable challenges from both outlier contamination and nonconvex optimization landscapes. While M-estimation addresses the problem of outliers through robust loss functions, it creates severely nonconvex problems that are difficult to solve globally. *Adaptive reweighting* schemes provide one particularly appealing strategy for implementing M-estimation in practice: these methods solve a sequence of simpler weighted least squares (WLS) subproblems, enabling both the use of standard least squares solvers and the recovery of higher-quality estimates than simple local search. However, adaptive reweighting still crucially relies upon solving the inner WLS problems effectively, a task that remains challenging in many robotics applications due to the intrinsic nonconvexity of many common parameter spaces (e.g. rotations and poses).**

In this paper, we show how one can easily implement adaptively-reweighted M-estimators with *certifiably correct* inner WLS solvers using only fast *local* optimization over smooth manifolds. Our approach exploits recent work on *certifiable factor graph optimization* to provide global optimality certificates for the inner WLS subproblems while seamlessly integrating into existing factor graph-based software libraries and workflows. Experimental evaluation on pose-graph optimization and landmark SLAM tasks demonstrates that our adaptively reweighted certifiable estimation approach provides higher-quality estimates than alternative local search-based methods, while scaling tractably to realistic problem sizes.

I. Introduction

Current state-of-the-art approaches to parameter estimation in robotics and computer vision typically formalize and solve this task using maximum likelihood estimation (MLE) [1], [2]. This approach is appealing for both its conceptual simplicity, and for the strong statistical performance guarantees that MLE affords [3]. However, MLE faces two critical challenges. First, it is typically implemented using *local* optimization methods that are highly sensitive to initialization when applied to nonconvex problems [4]. Second, basic MLE estimators have *zero breakdown point*, meaning that even a vanishingly small fraction of outlier contamination in the data can produce arbitrarily poor MLE estimates [5].

M-estimation [5] addresses the outlier problem by replacing the standard negative log-likelihood loss used in basic MLE with *robust* alternatives that are designed to attenuate the ill effects of outlier measurements on the resulting estimates. While this approach can dramatically improve robustness to outliers, achieving this effective attenuation typically requires the use of *nonconvex* loss functions that substantially exacerbate the nonconvexity (and hence sensitivity to initialization) already present in MLE.

All authors are with the Institute for Experiential Robotics, Northeastern University, 360 Huntington Ave, Boston, MA 02115, USA. Corresponding author: Zhexin Xu (xu.zhex@northeastern.edu).

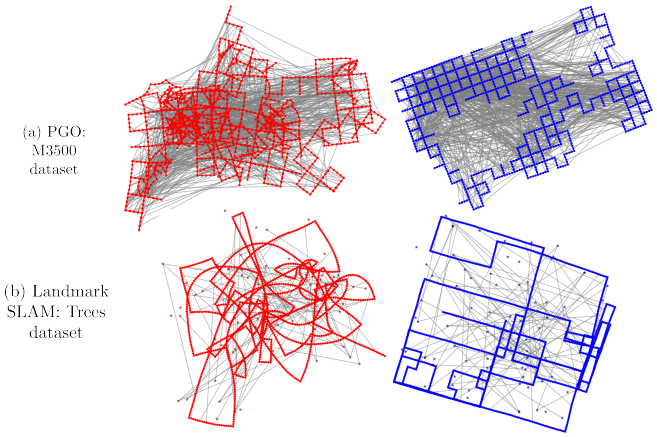

Fig. 1: Examples of solutions obtained with `GNC-Local` (left) and the proposed `Certi-GNC` framework (right), in the presence of 30 % outlier loop closures (shown in grey) for a) pose graph optimization (PGO) and b) landmark simultaneous localization and mapping (SLAM).

Implementing M-estimation in practice thus frequently entails a tradeoff between performance guarantees and computational tractability. On one hand, global optimization methods (such as branch-and-bound or global polynomial optimization techniques) [6], [7] can guarantee the recovery of correct (i.e. *globally optimal*) M-estimates, but are typically intractably expensive to apply to large-scale problems. Conversely, *local* optimization remains computationally efficient, but determining a suitable high-quality initialization becomes even more challenging in the presence of potential outlier contamination [8].

Adaptive reweighting schemes provide one particularly appealing strategy for implementing M-estimation. These methods proceed by solving a *sequence* of weighted least squares (WLS) problems, which is advantageous because the latter are frequently easier to optimize than robust formulations (and in fact often take the form of a standard MLE for which a fast solver is already available). Adaptive reweighting thus provides a computationally efficient approach to implementing M-estimation using standard solvers, while also typically recovering substantially better estimates than direct application of simple local search. However, this strategy crucially relies upon solving the *inner* WLS, a task that is itself still challenging in many robotics applications, due to the inherent nonconvexity of the underlying parameter spaces (e.g. rotation and pose manifolds) [9]. While some recent work has demonstrated the use of convex relaxation-based *certifiably correct* methods to perform this inner WLS

optimization [10], at present certifiable methods are typically hand-designed for specific problem classes (e.g. rotation averaging or PGO), and thus this strategy – while highly effective – is limited to those specific applications for which an existing certifiable optimizer is already available [9].

In this paper, we show how one can easily implement a wide range of adaptively-reweighted M-estimators with certifiably-correct inner WLS solvers using only fast *local* optimization over smooth manifolds. Our approach exploits the *certifiable factor graph* optimization framework proposed in [11], which provides a simple procedure for designing and deploying a wide range of (non-robust) certifiable estimators for robot perception problems using the same factor graph modeling and local optimization paradigm already employed ubiquitously throughout robotics and computer vision [1]. Our proposed M-estimation approach thus seamlessly integrates into existing factor graph-based software libraries and workflows, without requiring the use of external, hand-designed, problem-specific certifiable solvers [9]. Experimental evaluation on a variety of pose-graph and pose-and-landmark SLAM tasks demonstrates that our adaptively-reweighted certifiable estimation approach outperforms alternative M-estimation schemes based upon simple *local* search in terms of solution quality, while scaling tractably to realistic estimation tasks.

II. Review of Robust Estimation

This section reviews fundamentals of robust estimation, focusing on the use of adaptive reweighting schemes to solve the M-estimation problem [12].

M-estimation computes a state estimate $\boldsymbol{x}^*$ according to:

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \sum_i \rho(r_i(\boldsymbol{x})), \tag{1}$$

where $\rho(\cdot)$ is a *robust* loss function that grows sub-quadratically for large errors. Ensuring hard rejection of very large residuals requires a loss whose gradient *vanishes* for large errors, which is necessarily nonconvex; this additional nonconvexity makes robust M-estimation harder than its non-robust counterpart.

One can address this challenge by directly applying *global* optimization to solve (1); however, this problem is NP-hard in general, and thus global methods quickly become intractable for large-scale SLAM or computer vision problems [9], [12]. Consequently, practitioners often rely on local nonlinear methods, which are initialization-sensitive and prone to poor local minima under outlier contamination [8].

A. Reformulating M-estimation as Outlier Processes

Rather than relying solely on global or purely local optimization to directly solve the robust problem (1), a practical alternative is to *robustify* an existing non-robust estimator by solving a sequence of simpler WLS subproblems via adaptive reweighting schemes. The Black-Rangarajan (BR) duality [13] formalizes this strategy by describing an equivalence between solving the M-estimation problem by applying *(i)* a robust loss to the measurement residuals and *(ii)* a weighted, non-robust joint formulation over $(\boldsymbol{x}, \boldsymbol{w})$, where $\boldsymbol{w}$ are a set of auxiliary weights multiplying the quadratic residual terms. This equivalence is presented in the following theorem:

Algorithm 1: GNC for Robust M-estimation.

Input: Initial estimate $\boldsymbol{x}^{(0)}$, initial weights $\boldsymbol{w}^{(0)}$, surrogate loss family $\rho_\mu(\cdot)$
Output: Final estimate $\hat{\boldsymbol{x}}$

Initialize: $\boldsymbol{x} \leftarrow \boldsymbol{x}^{(0)}$, $\boldsymbol{w} \leftarrow \boldsymbol{w}^{(0)}$, $\mu \leftarrow \mu_0$
repeat
  repeat
    $\boldsymbol{x} \leftarrow \arg\min_{\boldsymbol{x}} \sum_{i=1}^{n} w_i\, r_i^2(\boldsymbol{x})$
    for $i \in [n]$ **do**
      $w_i \leftarrow \arg\min_{w_i \in [0,1]} \Phi_{\rho_\mu}(w_i) + w_i r_i^2(\boldsymbol{x})$
    end
  until ConvergenceWLS() *(Sec. III-D)*
  $\mu \leftarrow$ UpdateSchedule(μ)
until ConvergenceGNC() *(Sec. III-D)*
return $\hat{\boldsymbol{x}} \leftarrow \boldsymbol{x}$

Theorem 1: (Black–Rangarajan Duality [13]) *Given a robust loss $\rho(\cdot)$, define $\phi(z) := \rho(\sqrt{z})$. If $\phi(z)$ satisfies $\lim_{z\to 0} \phi'(z) = 1$, $\lim_{z\to\infty} \phi'(z) = 0$, and $\phi''(z) < 0$, then the M-estimation problem in (1) is equivalent to*

$$\min_{\substack{\boldsymbol{x} \\ \boldsymbol{w}\in[0,1]^n}} \sum_{i=1}^{n} \left[w_i r_i^2(\boldsymbol{x}) + \Phi_\rho(w_i) \right], \tag{2}$$

where $w_i \in [0,1]$ are auxiliary weights, and $\Phi_\rho(w_i)$ is an outlier process induced by $\rho(\cdot)$.

The conditions on $\rho(\cdot)$ are satisfied by most common robust loss functions [13]. One nice feature of this approach is that the optimal auxiliary weights can be interpreted as *soft inlier indicators*. Note that this probabilistic interpretation is a byproduct of the BR reformulation and is not available when minimizing a robust loss alone.

B. IRLS via Alternating Minimization

Directly minimizing *jointly* over $(\boldsymbol{x}, \boldsymbol{w})$ in (2) is frequently difficult. However, two *partial* minimizations are often much more straightforward: *(i)* for fixed $\boldsymbol{w}$, minimizing over $\boldsymbol{x}$ reduces to an ordinary WLS problem; and *(ii)* for fixed $\boldsymbol{x}$, the minimization over the weights decouples into n independent one-dimensional problems on $[0,1]$, which for many losses admit closed-form solutions. Note that while this WLS problem is easier to implement with standard NLS methods (Gauss–Newton, Levenberg-Marquardt (LM)) than the original problem, it remains challenging for most robotics applications. Applying alternating minimization to the BR formulation (Theorem 1) simply alternates these two steps and yields the classical iteratively reweighted least squares (IRLS) algorithm. Because each partial minimization cannot increase the objective, IRLS produces a monotonically non-increasing cost sequence. Nevertheless, the nonconvexity of

robust redescending losses renders the convergence of IRLS highly sensitive to initialization [12].

C. GNC as a Continuation Strategy

Graduated non-convexity (GNC) is a continuation scheme that aims to reduce sensitivity to initialization by introducing a homotopy from an "easy" convex surrogate to the "hard" target robust loss, with each stage solved by WLS. Compared with IRLS [14], GNC introduces a control parameter μ that defines surrogate losses $\rho_\mu(\cdot)$, which are initially convex and gradually recover the original robust loss as μ is varied. This yields a sequence of WLS subproblems with progressively increasing robustness (see Fig. 2). The BR duality extends naturally to $\rho_\mu(\cdot)$, resulting in the corresponding outlier processes $\Phi_{\rho_\mu}(w_i)$ (see for instance [12] for some common explicit forms). This leads to the three steps shown in Algorithm 1.

III. Robust Certifiable Factor Graph Optimization

In the previous section, we showed that the robust estimation problem can be reduced to a sequence of non-robust problems. However, obtaining *global* solutions to the resulting non-robust subproblems remains challenging for many tasks of interest in robotics and computer vision.

The central question of this work is: within robust adaptive weighting schemes, *how can we solve the inner WLS problem (see Algorithm 1) so that it is both certifiably correct and broadly applicable?* We address this by solving this inner problem through `Certi-FGO` [11], a framework that makes certifiable estimation accessible in factor graph libraries (e.g., GTSAM [15]) by exploiting the structural correspondence between factor graphs and block-separable quadratically constrained quadratic programs (QCQPs).

In the remainder of this section, we introduce Shor's convex relaxation of these block-separable QCQPs, yielding a semidefinite program (SDP) that serves as our certifiable surrogate, and show how to implement and solve the resulting relaxations efficiently by only using standard local factor graph optimization (FGO) routines. We therefore preserve the practicality of FGO and equip it with certifiable guarantees.

A. Shor's Relaxation of QCQPs

A growing body of recent work has shown that hard nonconvex MLE problems can be solved by expressing them as a QCQP, relaxing them to SDP, and then applying Burer-Monteiro (BM) factorization [16] to handle large problem instances efficiently [9], [17].

Let us assume that we have an estimation problem in the form of a QCQP:

$$f^*_{\text{QCQP}} = \min_{\mathbf{X}\in\mathbb{R}^{n\times d}} \langle \mathbf{Q}, \mathbf{X}\mathbf{X}^\top\rangle \quad \text{s.t.} \langle \mathbf{A}_i, \mathbf{X}\mathbf{X}^\top\rangle = b_i,\ \forall i \in [m], \tag{3}$$

where $\mathbf{Q} \in \mathbb{S}^n$, $\mathbf{A}_i \in \mathbb{S}^n$ for all $i \in [m]$, and $\mathbf{b} \in \mathbb{R}^m$.

To address NP-hardness of (3), Shor's relaxation [18] replaces $\mathbf{X}\mathbf{X}^\top$ with a generic positive semidefinite (PSD) matrix $\mathbf{Z} \in \mathbb{S}^n_+$, removing the implicit rank-d constraint to yield the convex SDP relaxation:

$$f^*_{\text{SDP}} = \min_{\mathbf{Z}\in\mathbb{S}^n_+} \langle \mathbf{Q}, \mathbf{Z}\rangle \quad \text{s.t.} \langle \mathbf{A}_i, \mathbf{Z}\rangle = b_i,\ \forall i \in [m]. \tag{4}$$

Note that (4) is convex, and can thus be solved to global optimality using e.g. interior-point methods. Furthermore, its optimal value always *lower bounds* the optimal value of (3): $f^*_{\text{SDP}} \leq f^*_{\text{QCQP}}$. Finally, note that if a minimizer $\mathbf{Z}^*$ of (4) happens to have a rank-d factorization of the form $\mathbf{Z}^* = \mathbf{X}^*\mathbf{X}^{*\top}$, then $\mathbf{X}^*$ is in fact a *global* minimizer of (3).

Remark 1: Crucially, a large body of recent work has shown that this exact solution recovery often occurs when the measurement noise is small [9], [19].

B. Efficiently Solving the SDP Relaxation via Low-Rank Factorization and Riemannian Staircase

Rather than solving (4) directly, we can exploit its low-rank structure to reduce computational complexity for large-scale problems. We assume a factorization of the form $\mathbf{Z} = \mathbf{Y}\mathbf{Y}^\top$ for some $\mathbf{Y} \in \mathbb{R}^{n\times p}$ and $d < p \ll n$, yielding the *lifted rank-p* BM factorization [16]:

$$\min_{\mathbf{Y}\in\mathbb{R}^{n\times p}} \langle \mathbf{Q}, \mathbf{Y}\mathbf{Y}^\top\rangle \quad \text{s.t.} \quad \langle \mathbf{A}_i, \mathbf{Y}\mathbf{Y}^\top\rangle = b_i,\ \forall i \in [m], \tag{5}$$

which automatically enforces positive semidefiniteness and reduces the decision space from roughly n^2 to np variables (where $p \ll n$), at the cost of reintroducing nonconvexity. Note that setting $p = d$ recovers the original QCQP (3). Moreover, a candidate $\mathbf{Z} = \mathbf{Y}\mathbf{Y}^\top$ for a solution of (4) constructed from a *local* minimizer $\mathbf{Y}$ of (5) can still be checked for *global* optimality using (4)'s Karush-Kuhn-Tucker (KKT) conditions.

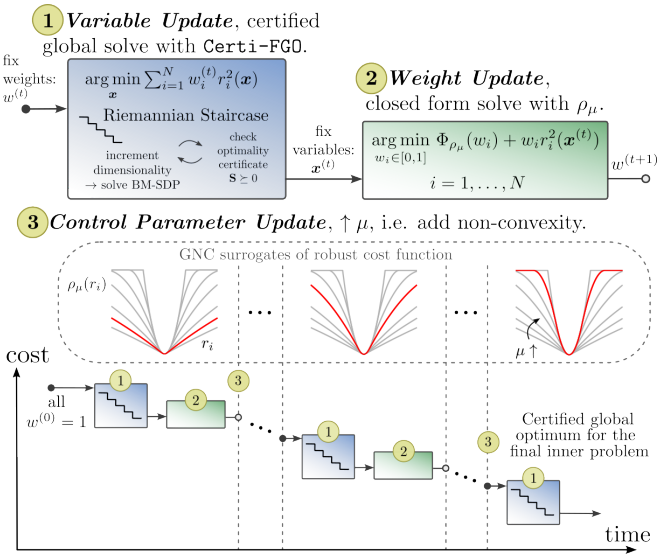

Fig. 2: Overview of our `Certi-GNC` framework: `Certi-FGO` [11] provides 1) certifiable global inner solves (i.e. non-linear weighted least squares (WLS) problem) within a graduated non-convexity (GNC) framework. The control parameter μ defines convex surrogates $\rho_\mu(r_i)$ of the target non-convex robust loss, yielding a sequence of WLS problems with 2) weights $w_i \in [0, 1]$ updated via. closed form solve from residuals r_i. 3) A truncated least squares (TLS) robust loss function is used here: μ is increased to gradually recover non-convexity, with $\mu \to \infty$ recovering hard truncation.

Remark 2: This low-rank factorization approach is justified by the fact that (4) is known to admit low-rank solutions for many problems of interest under mild conditions [9], [19].

Through comparison of the KKT conditions of the BM and SDP problems, one can show that a BM solution can be certified as globally optimal by checking the positive semidefiniteness of the *certificate matrix* [20]:

$$\mathbf{S} \triangleq \mathbf{Q} + \mathcal{A}^*(\boldsymbol{\lambda}) = \mathbf{Q} + \sum_{i=1}^{m} \lambda_i \mathbf{A}_i, \tag{6}$$

where $\mathcal{A}^* : \mathbb{R}^m \to \mathbb{S}^n$ denotes the adjoint operator. For a KKT point $\mathbf{Y}$ of the rank-p BM factorization with multipliers $\boldsymbol{\lambda}$, the matrix $\mathbf{S}$ in (6) serves as a certificate of optimality: if $\mathbf{S} \succeq 0$, then $\mathbf{Z} = \mathbf{Y}\mathbf{Y}^\top$ is globally optimal for the SDP (and, when $p = d$, also solves the QCQP); otherwise, the eigenvector associated with the smallest eigenvalue of $\mathbf{S}$ gives a direction of negative curvature. The *Riemannian Staircase* [21], [22] wraps this certify-or-lift step by solving a sequence of low-rank problems: parameterize $\mathbf{Z} = \mathbf{Y}\mathbf{Y}^\top$ at rank p, solve the rank-p manifold subproblem, recover multipliers to form $\mathbf{S}$, and either certify or increase the rank to $p+1$ by augmenting $\mathbf{Y}$ along the minimum-eigenvector direction. In practice, one or two staircase steps typically suffice [17].

C. Certifiable Estimation for Factor Graphs

We begin by considering the following general MLE problem: jointly minimizing a sum of *data fitting* terms with *sparse* dependencies over a collection of variables,

$$\min_{\boldsymbol{x}_i \in \mathcal{M}_i} \quad \sum_{k=1}^{u} l_k(\boldsymbol{x}_{S_k}), \tag{7}$$

where $\mathcal{M}_i$ is the domain of $\boldsymbol{x}_i$ and each summand l_k depends *only* upon the subset of variables $\boldsymbol{x}_{S_k}$ indexed by $S_k \subseteq [n]$.

This sparse dependency structure of (7) admits a natural graphical representation using *factor graphs* [1]. Formally the factor graph associated with (7) is the bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{F}, \mathcal{E})$ in which:

1) *variable nodes* $\mathcal{V} \triangleq \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ are the model parameters to be estimated;
2) *factor nodes* $\mathcal{F} \triangleq \{l_1, \ldots, l_u\}$ consist of the individual factors l_k;
3) *edge set* $\mathcal{E} \triangleq \{(\boldsymbol{x}_i, l_k) \in \mathcal{V} \times \mathcal{F} \,|\, \boldsymbol{x}_i \in \boldsymbol{x}_{S_k}\}$, i.e. variable $\boldsymbol{x}_i$ and factor l_k are joined by an edge in $\mathcal{G}$ if and only if $\boldsymbol{x}_i$ is an argument of l_k.

See Fig. 3 for a representative example.

Factor graphs serve two primary functions. First, the edge set $\mathcal{E}$ in $\mathcal{G}$ directly encodes the sparsity structure of Prob. (7). Second, factor graph models provide a convenient modular modeling language for constructing high dimensional optimization problems by composing simple elementary constituent parts (i.e. individual variables and factors). Consequently, many current state-of-the-art software libraries for state estimation in robotics and computer vision employ factor graph-based abstractions for instantiating and solving MLE problems of the form (7) [15], [23].

Now let us additionally suppose that the generic maximum likelihood estimation (7) takes the form of the QCQP (3). Observe that the sparsity pattern captured in (7) places algebraic restrictions on the QCQP data matrices: S_k captures the *block sparsity pattern* of the k-th objective matrix $\mathbf{Q}_k$, while the Cartesian product structure of the full domain implies constraint matrices $\mathbf{A}_i$ are *block-diagonal* with one nonzero block, i.e., they are *block separable*. `Certi-FGO` [11] exploits this correspondence between sparse factor graphs and QCQPs with block-structure by *lifting* each variable to a higher-dimensional domain corresponding to the BM-factored relaxation. Crucially, this lifting preserves the sparsity structure of (7). Assuming that the lifted variable domains are smooth manifolds, the resulting *lifted factor graph* can thus be optimized on the product manifold using Riemannian optimization [24], where the Riemannian Staircase [17], [22] (Algorithm 2) automatically manages rank increases and solution verification. Thus, `Certi-FGO` provides certifiable global optimality and competitive scalability without requiring problem-specific implementations.

Remark 3: The use of this method *presupposes* that the estimation problem (7) is a QCQP. This is the case for many robotics and computer vision estimation problems [9] [7].

In the following, we show how factor graph structure is preserved through the problem transformations (QCQP→SDP→BM) described in Secs. III-A and III-B. Likewise, the verification and saddle-escape procedures III-B can be efficiently performed block wise by exploiting the same factor graph structure.

1) QCQP and Shor's relaxation over Factor Graphs: We begin by explicitly writing the factor graph MLE problem (7) as a QCQP (3), revealing the block separable structure of the objective and constraints induced by the factor graph formulation. The decision variable $\mathbf{X} \in \mathbb{R}^{n \times d}$ can be partitioned into K *block-row variables* $\mathbf{X}_i \in \mathbb{R}^{d_i \times d}$ corresponding to individual factor graph variables $\boldsymbol{x}_i$ in (7), such that $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_K \end{bmatrix}^\top$ with $n = \sum_{i=1}^{K} d_i$.

The decomposition (7) places strong restrictions on the data matrices parameterizing the corresponding QCQP. First, recall that each factor $l_k(\cdot)$ only depends upon the subset of variables indexed by S_k. This fact constrains the sparsity patterns of the data matrices $\mathbf{Q}_k \in \mathbb{S}^n$ parameterizing the quadratic summands $l_k(\cdot)$ in (7): note that we must have $(\mathbf{Q}_k)_{i,j} = 0$ if $(i,j) \notin S_k \times S_k$. Second, the feasible set of (7) is a *Cartesian product* of the individual variable domains $\mathcal{M} = \mathcal{M}_1 \times \cdots \times \mathcal{M}_K$, i.e., each variable can be varied independently of all others. In order to be consistent with this product structure, it follows that each of the quadratic constraints appearing in (3) can involve only *one* of the variables $\mathbf{X}_i$. Consequently, we may partition the index set $[m]$ of the constraints into subsets, where $L_i \subseteq [m]$ contains the indices of those constraints associated with variable $\mathbf{X}_i$. This implies that if $\ell \in L_i$, then A_ℓ is block diagonal, with a single nonzero block in the (i,i)-th position. Thus, the factor graph decomposition (7) implies the following block decomposition for the corresponding QCQP:

$$\begin{aligned} \min_{\mathbf{X}\in\mathbb{R}^{n\times d}} \quad & \sum_{k=1}^{u}\sum_{i,j=1}^{K} \langle (\mathbf{Q}_k)_{i,j}, \mathbf{X}_i\mathbf{X}_j^\top\rangle \\ \text{s.t.} \quad & \langle (\mathbf{A}_\ell)_{i,i}, \mathbf{X}_i\,\mathbf{X}_i^\top\rangle = b_\ell,\ \forall \ell\in L_i,\ i\in[K], \end{aligned} \tag{8}$$

2) Burer-Monteiro Factorization over Factor Graphs: For scalability, we apply the low-rank BM factorization $\mathbf{Z} = \mathbf{Y}\mathbf{Y}^\top$ to the SDP obtained with Shor's relaxation, partitioning $\mathbf{Y} \in \mathbb{R}^{n\times p}$ into K variable blocks $\mathbf{Y}_i \in \mathbb{R}^{d_i\times p}$ to arrive at

$$\begin{aligned} \min_{\mathbf{Y}\in\mathbb{R}^{n\times p}} & \sum_{k=1}^{u}\sum_{i,j=1}^{K} \langle (\mathbf{Q}_k)_{i,j}, \mathbf{Y}_i\mathbf{Y}_j^\top\rangle \rangle \\ \text{s.t} \quad & \langle (\mathbf{A}_\ell)_{i,i}, \mathbf{Y}_i\mathbf{Y}_i^\top\rangle = b_\ell, \forall \ell\in L_i, i\in[K]. \end{aligned} \tag{9}$$

Note that the sparsity and block-separability structure of the BM factorization (9) *exactly matches* the one assumed in the factor graph decomposition (7) for the initial QCQP (8).

Rather than use traditional non-linear programming solvers to solve the constrained nonconvex problem (9), we leverage *intrinsic optimization*—reformulating it as unconstrained optimization on manifolds—for substantial efficiency gains. The constraint set of Prob. (9) inherits the same block-diagonal and block-separability properties as that of the original QCQP (8). It follows that the feasible set for our desired intrinsic reformulation must also be a Cartesian product, and that the individual factors comprising this product are determined by:

$$\begin{aligned} \mathcal{M}_i^{(p)} := & \big\{\mathbf{Y}_i \in \mathbb{R}^{d_i\times p} : \langle (\mathbf{A}_\ell)_{i,i}, \mathbf{Y}_i\mathbf{Y}_i^\top\rangle = b_\ell\big\}, \\ & \forall \ell\in L_i, i\in[K], \end{aligned} \tag{10}$$

with the overall feasible set $\mathcal{M}^{(p)} := \mathcal{M}_1^{(p)} \times \cdots \times \mathcal{M}_K^{(p)}$. The lifted objective (9) inherits the same sparse dependency structure from the original data matrices $\mathbf{Q}_k$ (8), ensuring the reformulated problem maintains a factor graph representation with variables and factors in *one-to-one* correspondence with the original MLE (7). This leads to the intrinsic formulation:

$$\min_{\mathbf{Y}\in\mathcal{M}^{(p)}} \big\langle \mathbf{Q},\, \mathbf{Y}\,\mathbf{Y}^\top\big\rangle. \tag{11}$$

Since the constraints are absorbed into the manifold geometry $\mathbf{Y} \in \mathcal{M}^{(p)}$, this becomes unconstrained optimization over a product of smooth manifolds[1], enabling standard Riemannian methods from factor graph optimization libraries.

3) Optimality Verification and Saddle-Escape in the Factor Graph Setting: The Lagrange multipliers used in the optimality certificate of (6) can be obtained by solving a least-squares problem. Because each constraint matrix $\mathbf{A}_\ell$ is block-diagonal with a single nonzero block $(\mathbf{A}_\ell)_{i,i}$ acting on variable $\mathbf{Y}_i$ for all $\ell \in L_i$, the operator $\mathcal{A}^*(\boldsymbol{\lambda})$ is also block-diagonal with blocks aligned to the factor graph variables. Consequently, the linear system characterizing Lagrange multipliers for the BM factorization (5) actually decomposes into K *independent* linear systems involving the K blocks on the diagonal of $\mathcal{A}^*(\boldsymbol{\lambda})$. It follows that we can solve for the Lagrange multipliers λ_i (and hence the nonzero blocks of the certificate matrix $\mathbf{S}$) by performing independent operations on the individual manifolds $\mathcal{M}_i^{(p)}$. This blockwise structure enables efficient optimality verification and the construction of negative-curvature directions for saddle escape that directly reflect the sparsity of the underlying factor graph.

[1]This is possible only when the constraints acting on individual variables are smooth manifolds, which is a mild assumption in practice as rotations and poses induce smooth manifold structure, specifically the Stiefel manifold [24].

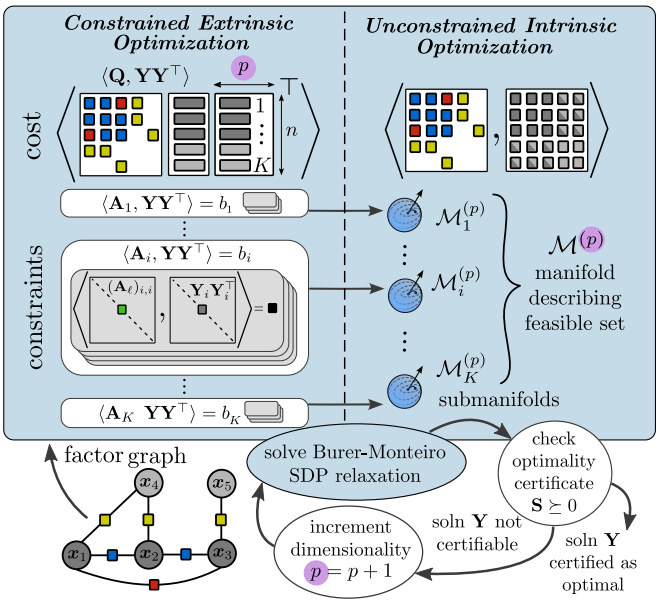

Fig. 3: Illustration of the `Certi-FGO` framework showing the transformation from constrained to unconstrained optimization leveraging underlying sparse factor graph structure. The BM variable $\mathbf{Y}$ is partitioned into K block rows $\mathbf{Y}_i$ corresponding to variables $\boldsymbol{x}_i$ in the factor graph. The preserved block structure is apparent in this BM form of the factor graph MLE (9), in that each constraint block $(\mathbf{A}_\ell)_{i,i}$ acts locally on a single block $\mathbf{Y}_i$, defining individual submanifolds $\mathcal{M}_i^{(p)}$. Sparse separable structure is also present in the objective, though not explicitly visualized. The constrained optimization problem (Eq. (9), left) is reformulated as unconstrained Riemannian optimization over the product manifold $\mathcal{M}^{(p)} = \mathcal{M}_1^{(p)} \times \cdots \times \mathcal{M}_K^{(p)}$ (Eq. (11), right), enabling efficient solution using standard manifold optimization techniques.

Such efficient optimality verification and saddle escape are carried out using the Riemannian Staircase algorithm [17], [22] adapted to the factor graph setting, described in Algorithm 2. At each rank p, local Riemannian optimization of the BM problem is performed over the product manifold $\mathcal{M}^{(p)}$, the certificate matrix $\mathbf{S}$ is computed from the blockwise multipliers, and global optimality is either certified or the rank is increased along a negative-curvature direction to continue the search.

In practice, this block-separable formulation makes the Riemannian optimization problem almost automatic: the product manifold $\mathcal{M}^{(p)}$ can be assembled directly from the individual submanifolds $\mathcal{M}_i^{(p)}$ defined in (10), and the GNC scheme already implemented in GTSAM can be applied to the lifted factors without the need for problem-specific code. See [11] for details on constructing these lifted factors. As a result, practitioners can construct an efficient, certifiably robust solver that efficiently exploits problem structure simply by lifting their original factor graph, rather than hand-deriving problem-specific data matrices or manifold models.

Algorithm 2: `Certi-FGO`, Certifiable Estimation in Factor Graphs [11]

Input: Initial values $\mathbf{Y} = \{\mathbf{Y}_i\}_{i=1}^n$, factor graph $\mathcal{G}$, initial rank p.
Output: A feasible estimate $\hat{\mathbf{X}}$ and lower bound f^*_{SDP} on its optimal value.

function CERTIFIABLEFGO($\mathbf{Y}$, $\mathcal{G}$, p)**:**
  while *true* **do**
    // Lift variables to rank-p
    $\mathbf{Y}_p \leftarrow \text{LIFT}(\mathbf{Y})$
    // Construct lifted factors for rank-p lift
    $\mathcal{G}_p \leftarrow \text{CONSTRUCTLIFTEDFACTORS}(\mathcal{G}, p)$
    $(\mathbf{Y}^*_p) \leftarrow \text{LOCALOPTIMIZATION}(\mathcal{G}_p, \mathbf{Y}_p)$
    // Construct certificate and verify by computing minimum eigenvalue
    $(\lambda_{\min}, v_{\min}) \leftarrow \text{VERIFICATION}(\mathcal{G}_p, \mathbf{Y}_p)$
    if $\lambda_{\min} > 0$ **then**
      return $\{\hat{\mathbf{Y}}, f^*_{\text{SDP}}\}$
    else
      // Increase rank, use current solution as initialization in next iterate.
      $p \leftarrow p + 1$
      $\mathbf{Y}_p \leftarrow \text{SADDLEESCAPE}(\mathbf{Y}^*_p, v_{\min})$
    end
  end
  $\hat{\mathbf{X}} \leftarrow \text{ROUNDSOLUTION}(\mathbf{Y}_p)$
  return $\{\hat{\mathbf{X}}, f^*_{\text{SDP}}\}$
end

D. Robustifying Certifiable FGO

In Sec. II-A, we showed that the BR duality reformulates M-estimation as iterative variable and weight updates (Algorithm 1). The variable update (**??**) is a WLS problem—a type of QCQP—which `Certi-FGO` can solve (see Remark 3). Therefore, `Certi-FGO` naturally fits as the inner solver for adaptive reweighting schemes.

We emphasize how easy it is to *implement* `Certi-FGO` in practice within existing factor graph-based adaptive reweighting robust estimation frameworks. Unlike purpose-built methods like SE-Sync—which require hand-crafting the objective and constraint matrices for Problem (5), manually specifying the manifold structure, and implementing custom solver subroutines (ex. gradient and Hessian computations)—`Certi-FGO` operates directly on factors. This seamless integration with GTSAM's mature solver library and GNC implementation, provides certifiable robust estimation for factor graphs without manual problem reformulation and minimal implementation efforts. In this paper we present one realization of this pipeline which we refer to as `Certi-GNC`, consisting of Algorithm 1 with the inner WLS variable update solved with Algorithm 2.

At each rank-p lift of `Certi-FGO` Algorithm 2, optimization on the lifted manifold $\mathcal{M}^{(p)}$ as per (11) is performed using GTSAM's native Riemannian LM solver, which handles the manifold geometry through tangent space computations and retractions without requiring any custom implementation. In our experiments, `GNC-Local`—our non-certifiable GNC baseline—uses this same LM solver for its inner solves, ensuring observed performance differences stem from our use of certifiable inner WLS solves, rather than from details of the underlying factor graph optimizers.

In practice, the inner loop of Algorithm 1 (lines 3-8) is truncated to a single iteration. When homotopy parameter updates are sufficiently small, the minimizer for the next step lies close to the current solution $\boldsymbol{x}^*$, allowing this warm-started single iteration to adequately approximate the solution. For `GNC-Local`, this means performing one local solve per μ value. For `Certi-GNC`, each inner solve achieves global optimality, inherently eliminating any need for iterative refinement.

Algorithm Termination. To ensure fair comparison, `Certi-GNC` and `GNC-Local` share the same GNC *outer loop* termination criteria. The loop terminates when any of the following occur between iterations: *(i)* weights w_i converge, i.e. $\max_{1\leq i\leq n}\left| w_i - \text{round}(w_i)\right| \leq \varepsilon$, *(ii)* cost converges, i.e. $\Delta\langle \mathbf{Q}, \mathbf{Y}\mathbf{Y}^\top\rangle \leq c_{tol} = 10^{-6}$, *(iii)* the max number of GNC iterations reached. With small μ steps, a single warm-started WLS per stage often suffices and the inner loop may be omitted. Similarly, the exit criterion for the *inner loop* WLS solves are (`Certi-FGO` or LM solve of `GNC-Local`): *(i)* cost converges, i.e. $\Delta\langle \mathbf{Q}, \mathbf{Y}\mathbf{Y}^\top\rangle \leq c_{tol} = 10^{-5}$, *(ii)* $p_{max} = 30$, maximum iterations exceeded, *(iii)* $S + \eta\mathbf{I} \succcurlyeq 0$, i.e. solution certification achieved. Note that criterion *(iii)* only applies to the Riemannian Staircase of `Certi-FGO` in Algorithm 2.

Remark 4: `Certi-FGO` performance depends on parameters η and p_{max}, which can be tuned to balance certificate accuracy and runtime. While smaller η values give stricter certificates, in practice using larger η values for approximate certificates can, in some instances, still achieve near-optimal solutions. While the maximum rank p_{max} affects solver speed, certificates are usually obtained at lower ranks than the specified maximum.

IV. FGO Applications and Experiments

We benchmark `Certi-GNC`, our iterative framework with globally optimal subproblem solves that requires no good initialization (see Fig. 2), against a purely local one `GNC-Local` under two initializations: random sampling on the feasible set (rand init) and a favorable (outlier-free) odometry-based initialization (odom init). We evaluate two representative factor graph problems: PGO and landmark-based SLAM. To the best of our knowledge, these experiments report the first *robust* certifiable solver to appear in the literature for landmark-based SLAM (although [26], [27] present fast *non-robust* certifiable methods).

A. Experimental Setup

Problem Definition: The standard PGO problem estimates poses $\{(\mathbf{R}_i, \mathbf{t}_i)\}_{i\in[K]}$, with $\mathbf{R}_i \in \text{SO}(d)$ and $\mathbf{t}_i \in \mathbb{R}^d$, from noisy relative measurements $\{(\tilde{\mathbf{R}}_{ij}, \tilde{\mathbf{t}}_{ij})\}_{(i,j)\in\mathcal{E}}$ by

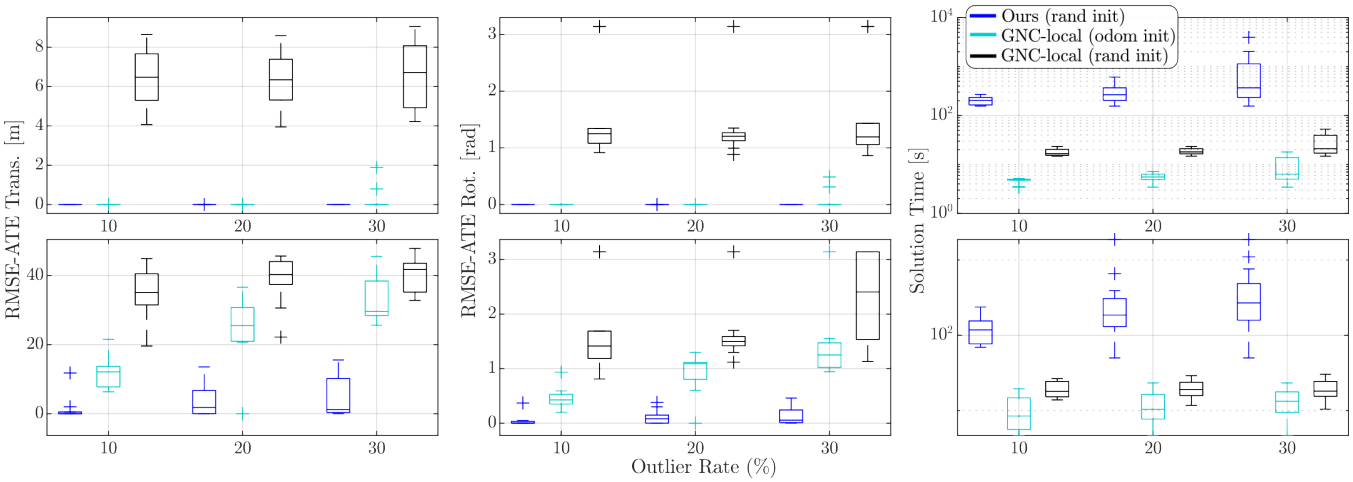

Fig. 4: Performance of our `Certi-GNC` framework with random initialization compared with the non-certifiable baseline `GNC-Local` given a good initialization from odometry and a random one. Columns (left to right): RMSE-ATE (translation), RMSE-ATE (rotation), and solution time. Rows (top to bottom): a) PGO problems from Intel dataset and b) landmark SLAM from Trees dataset [25].

minimizing $\sum_{(i,j)\in\mathcal{E}} \kappa_{ij}\|\mathbf{R}_j - \mathbf{R}_i\tilde{\mathbf{R}}_{ij}\|_F^2 + \tau_{ij}\|\mathbf{t}_j - \mathbf{t}_i - \mathbf{R}_i\tilde{\mathbf{t}}_{ij}\|_2^2$, where κ_{ij} and τ_{ij} are the measurement precisions. Landmark-based SLAM extends PGO by estimating landmark positions $\{\mathbf{l}_k\}_{k\in[L]}$ from noisy pose–landmark measurements $\{\tilde{\mathbf{l}}_{ik}\}_{(i,k)\in\mathcal{E}_{\text{lm}}}$, with $\mathbf{l}_k \in \mathbb{R}^d$, and adds the residuals $\sum_{(i,k)\in\mathcal{E}_{\text{lm}}} \tau_{ik}\|\mathbf{l}_k - \mathbf{t}_i - \mathbf{R}_i\tilde{\mathbf{l}}_{ik}\|_2^2$, where τ_{ik} is the measurement precision.

Metrics: We report runtime (in seconds) and estimation accuracy, the latter measured by the root-mean-square absolute trajectory error (RMSE-ATE) [28].

Implementation Details: For outlier generation, we corrupt only loop-closure (PGO) or pose–landmark (landmark SLAM) edges at rates (10%, 20%, 30%), mimicking data association failures; odometry measurements are left unperturbed and thus remain inliers. Each problem is evaluated over 10 Monte Carlo trials with independently generated outlier realizations. `Certi-GNC` and `GNC-Local` use the same GNC parameters as introduced in the previous section. The GTSAM LM solver is used as the local solver for both `Certi-GNC` and `GNC-Local` `relative_error` 10^{-8}, `absolute_error` 10^{-8}, and `max_iteration` 100. All experiments use the certifiable FGO implementation of [11] within the GNC scheme from GTSAM [10], [15], and were run on a laptop equipped with an Intel Core i7-11800H CPU and 32 GB RAM under Ubuntu 22.04. Full dataset details appear in our code release[1]; for TREES (landmark SLAM), we evaluate a reduced variant, extracting the first 1600 poses and the corresponding landmark observations for those poses.

B. Accuracy and Runtime Results

We assess whether `Certi-GNC` provides practical benefits over local methods for state estimation problems of varying complexity in the presence of outliers. The results in Fig. 4 reveal two key findings.

In terms of translation and rotation error, and for the easier estimation problem of PGO, `GNC-Local` with good odometry initialization and `Certi-GNC` provide similar results, though `GNC-Local` exhibits higher variability, indicating less reliable performance. For the more challenging landmark SLAM problem, `Certi-GNC` consistently outperforms `GNC-Local` across all conditions, showing that global optimization guarantees become essential as problem complexity increases due to higher dimensionality and more opportunities for local methods to fail.

While `Certi-GNC` exhibits longer convergence times than local alternatives, this is expected given the additional computational complexity of solving the underlying global optimization problem and the certification framework's computational overhead [29] (see Remark 4). This is justified by the improved solution reliability.

C. Tightness Properties Under Outlier Contamination

Under *bounded noise*, certifiable estimators relying on SDP relaxations are generically tight, yielding exact, low-rank solutions (see Remarks 1, 2). These properties enable the use of BM factorization to achieve computational tractability. However, in robust estimation the presence of outliers violates this bounded-noise assumption. Thus, we cannot generically expect FGO problems with outliers to exhibit the low-rank structure and exactness that `Certi-FGO` is designed to exploit. On the other hand, if the iterative reweighting procedure used in GNC succeeds in effectively separating inliers from outliers, we should hope that the *reweighted* inner least-squares estimation problems produced at the end of the homotopy will recover these tightness properties.

For a set of representative Monte Carlo PGO samples subject to varying outlier contamination, the termination level and relative suboptimality gap $\frac{f_{\text{QCQP}} - f_{\text{SDP}}}{f_{\text{SDP}}}$ of the `Certi-FGO` solution, embedded within each GNC iteration, are plotted in Fig. 5. Across all iterates the termination level (Fig. 5, top) remains relatively small and converges to a low value, demonstrating that the requisite low-rank structure is indeed present and that `Certi-FGO` successfully exploits this. Since the termination level directly corresponds

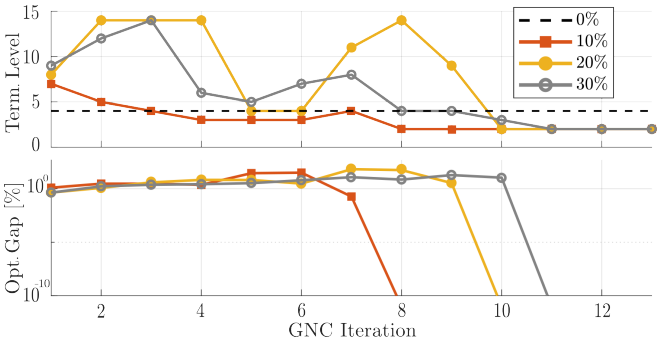

Fig. 5: A single Monte Carlo trial on the Intel PGO dataset with outlier rates of 0%, 10%, 20%, and 30%, showing (top) the Riemannian-staircase termination rank (a proxy for computational effort) and (bottom) the stagewise optimality gap of the `Certi-FGO` solve at each GNC iteration. In the 0% outlier case, GNC weights converge in the first iteration and the optimality gap is 10^{-12} as per Remark 1, the terminal level is visualized with a dashed line. All trials start at level 2 with random initialization; the trials with outlier terminated at GNC iteration 13.

to computational effort, the maximum level of $p = 15$ indicates that BM factorization provides scalability even when used as an inner solver in outlier-contaminated settings. (For comparison, in this example there are $K = 1728$ block variables consisting of $2d$ poses, so the corresponding matrix decision variable $\mathbf{Z}$ is of order $n = K \times 3 = 5184$.) Similarly, the relative suboptimality gap approaching 0 (Fig. 5, bottom) indicates that our non-convex SDP relaxations become asymptotically tight at the termination of GNC (i.e., after finding suitable assignments to the weights w_i), confirming the recovery of the low-rank structure and exactness that our method is designed to exploit.

V. Conclusion

In this paper, we show how to implement *robustified* certifiable estimators using only fast *local* optimization over smooth manifolds, thus enabling practitioners to easily design and deploy these state-of-the-art methods using standard factor graph-based software libraries and workflows. We embed the recently proposed non-robust certifiable factor graph estimation framework in [11] as the inner solver within an robust adaptive reweighting scheme, GNC, providing stagewise global optimality certificates for the inner subproblems.

In terms of estimation accuracy, our framework exceeds the performance of M-estimation strategies based on local search. While our method is more expensive to run than local baselines, this is expected because we are performing *global* rather than *local* optimization for the inner WLS solves. This trade-off is worthwhile for the improved reliability our approach affords. Finally, by requiring no problem-specific relaxations and leveraging standard facto graph software (e.g. GTSAM), our approach facilitates broad deployment across diverse estimation tasks, thereby democratizing access to the powerful machinery of robust certifiable estimation.

References

[1] F. Dellaert, M. Kaess *et al.*, "Factor graphs for robot perception," *Foundations and Trends in Robotics*, vol. 6, no. 1-2, pp. 1–139, 2017.

[2] T. D. Barfoot, *State Estimation for Robotics*. Cambridge University Press, 2017.

[3] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

[4] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 1999.

[5] E. M. Ronchetti and P. J. Huber, *Robust statistics*. John Wiley & Sons Hoboken, NJ, USA, 2009.

[6] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. Optim.*, vol. 11, no. 3, pp. 796–817, 2001. [Online]. Available: https://doi.org/10.1137/S1052623400366802

[7] H. Yang and L. Carlone, "One ring to rule them all: Certifiably robust geometric perception with outliers," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 18 846–18 859, 2020.

[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 *IEEE Intl. Conf. on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 4597–4604.

[9] D. M. Rosen, K. Khosoussi, C. Holmes, G. Dissanayake, T. Barfoot, and L. Carlone, "Certifiably optimal solvers and theoretical properties of SLAM," in *SLAM Handbook. From Localization and Mapping to Spatial Intelligence*, L. Carlone, A. Kim, T. Barfoot, D. Cremers, and F. Dellaert, Eds. Cambridge University Press.

[10] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *Robotics and Automation Letters*, vol. 5, no. 2, pp. 1127–1134, 2020.

[11] Z. Xu, N. R. Sanderson, and D. M. Rosen, "Simplifying certifiable estimation: A factor graph optimization approach," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2025, Workshop: Robotics in the Wild. [Online]. Available: https://dartmouthrobotics.github.io/icra-2025-robots-wild/spotlight-papers/icra-2025-robots-wild-16.pdf

[12] H. Yang, J. Mangelson, Y. Chang, J. Shi, and L. Carlone, "Robustness to incorrect data association and outliers," in *SLAM Handbook. From Localization and Mapping to Spatial Intelligence*, L. Carlone, A. Kim, T. Barfoot, D. Cremers, and F. Dellaert, Eds. Cambridge University Press.

[13] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Intl. J. of Computer Vision*, vol. 19, no. 1, pp. 57–91, 1996.

[14] A. Blake and A. Zisserman, *Visual reconstruction*. MIT press, 1987.

[15] F. Dellaert and G. Contributors, "borglab/gtsam," May 2022. [Online]. Available: https://github.com/borglab/gtsam

[16] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Mathematical Programming*, vol. 95, no. 2, pp. 329–357, 2003.

[17] A. Papalia, Y. Tian, D. M. Rosen, J. P. How, and J. J. Leonard, "An overview of the Burer-Monteiro method for certifiable robot perception," *arXiv preprint arXiv:2410.00117*, 2024.

[18] N. Z. Shor, "Quadratic optimization problems," *Soviet Journal of Computer and Systems Sciences*, vol. 25, pp. 1–11, 1987.

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

[20] D. M. Rosen, "Scalable low-rank semidefinite programming for certifiably correct machine perception." Springer, 2020, pp. 551–566.

[21] N. Boumal, V. Voroninski, and A. Bandeira, "The non-convex Burer-Monteiro approach works on smooth semidefinite programs," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 29, 2016.

[22] N. Boumal, "A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints," *arXiv preprint arXiv:1506.00575*, 2015.

[23] S. Agarwal, K. Mierle, T. Team *et al.*, "Ceres solver," 2012.

[24] N. Boumal, *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023.

[25] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the bayes tree," *Intl. J. of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.

[26] T. Fan, H. Wang, M. Rubenstein, and T. Murphey, "CPL-SLAM: Efficient and certifiably correct planar graph-based SLAM using the complex number representation," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1719–1737, 2020.

[27] C. Holmes and T. D. Barfoot, "An efficient global optimality certificate for landmark-based SLAM," *Robotics and Automation Letters*, no. 3, pp. 1539–1546, 2023.

[28] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 7244–7251.
[29] D. M. Rosen, "Accelerating certifiable estimation with preconditioned eigensolvers," *Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 507–12 514, 2022.