Selecting Good Measurements via ℓ_1 Relaxation: a Convex Approach for Robust Estimation over Graphs

Luca Carlone, Andrea Censi, Frank Dellaert

***Abstract*— Pose graph optimization is an elegant and efficient formulation for robot localization and mapping. Experimental evidence suggests that, in real problems, the set of measurements used to estimate robot poses is prone to contain *outliers*, due to perceptual aliasing and incorrect data association. While several related works deal with the rejection of outliers during pose estimation, the goal of this paper is to propose a grounded strategy for *measurements selection*, i.e., the output of our approach is a set of "reliable" measurements, rather than pose estimates. Because the classification in *inliers*/outliers is not observable in general, we pose the problem as finding the maximal subset of the measurements that is internally coherent. In the linear case, we show that the selection of the maximal coherent set can be (conservatively) relaxed to obtain a linear programming problem with ℓ_1 objective. We show that this approach can be extended to (nonlinear) planar pose graph optimization using similar ideas as our previous work on linear approaches to pose graph optimization. We evaluate our method on standard datasets, and we show that it is robust to a large number of outliers and different outlier generation models, while entailing the advantages of linear programming (fast computation, scalability).**

I. Introduction

A *pose graph* is a model for single and multi robot localization and mapping. The standard setup in the literature considers a graph where each node is associated to a pose (or configuration) $\boldsymbol{x}_i$, and the edges $\mathcal{E}$ are associated to the available measurements. For each edge $(i,j) \in \mathcal{E}$, we have available the measurement

$$\boldsymbol{z}_{ij} = h(\boldsymbol{x}_i, \boldsymbol{x}_j) + \boldsymbol{\epsilon}_{ij}, \tag{1}$$

where the function h returns the pose difference (in the Lie group sense), and $\boldsymbol{\epsilon}_{ij}$ is additive zero-mean Gaussian noise.

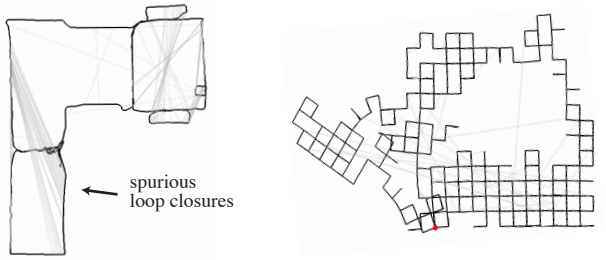
The objective of *pose graph optimization* is to compute an estimate of the nodes configuration $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ that maximizes the likelihood of the measurements. This is a nonconvex optimization problem, which has been extensively studied with various approaches (see, e.g., [1–5]). Some ready-to-use implementations are available online [6, 7].

Optimization "back-ends" (methods that solve the pose graph optimization problems) should be robust to mistakes of the "front-end" (methods that create the pose graph from the robot's sensory data), that may introduce spurious measurements that do not follow the nominal model (1).

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

A. Censi is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA, censi@mit.edu.

This work was partially funded by the National Science Foundation Award 11115678 "RI: Small: Ultra-Sparsifiers for Fast and Scalable Mapping and 3D Reconstruction on Mobile Robots".



(a) Bicocca dataset (b) Coherent outlier in M3500

Fig. 1. Pose graph optimization must be robust to outliers created by perceptual aliasing (to the left, visually similar corridors create spurious loop closures). However, in general, it is unobservable whether a measurement is an outlier or an inlier (to the right, an example of outlier that is coherent with the measurements). In this paper we abandon the notion of "inlier" or "outlier" and we seek for the maximal "*coherent*" set of measurements. We describe a fast and accurate convex relaxation of the combinatorial problem of selecting a maximal set of coherent measurements.

One particularly troubling source of spurious data arises from perceptual aliasing: if a robot visits two places that look similar, the front-end might generate spurious loop closure constraints; moreover, in ambiguous environments we might find groups of spurious loop closures that are locally coherent (Fig. 1a).

Several techniques have been proposed to identify and discard spurious constraints [8–10]; these are all based on iterative nonconvex optimization, which means that the robustness of the solution depends on the quality of the initial guess, and the computational cost does not scale well.

Contribution: We describe an approach to outlier rejection in planar pose graph optimization based on linear techniques, which allows to have a fast and *global* solution; in practice, the approach is as accurate as state-of-the-art techniques, while not relying on the availability of an initial guess.

In Section II we introduce our formulation, which slightly differs from related work. We do not focus on deciding whether a measurement is an "outlier" or "inlier", because this distinction is in general unobservable; rather, we look for the maximal set of measurements that are internally "coherent".

In Section III we consider the linear case. The problem of finding the largest coherent set can be written as an ℓ_0 problem with linear constraints, thanks to the "big M" trick. Moreover, we show that the ℓ_0 problem can be relaxed to a linear program (LP) minimizing an ℓ_1-norm objective, which is easily solvable with mature techniques, and scales well with problem size.

In Section IV we show the application to planar pose graph optimization. Here, we build on our previous work on linear approaches to planar pose graph optimization [5, 11, 12]. We

solve the problem in two stages: a first stage considering only the angular constraints, and a second phase considering a joint problem of Cartesian and angular constraints.

The experiments in Section V show that this purely linear formulation is able to obtain results as accurate as the state-of-the-art, while being much faster, and not depending on an accurate initial guess.

Classical approaches to outlier rejection in estimation: The lack of robusteness in estimators such as least squares was early recognized [13] and prompted the development of robustified alternativates, such as M-estimators and the *Least Trimmed Squares* estimator, proposed by Rousseeuw and Leroy [14]. More modern alternatives, such as the *Penalized Trimmed Squares* estimator [15], are based on minimization of a cost including squared errors and penalty for discarding measurements. This approach is sensitive to the penalty for measurement removal and requires solving a quadratic mixed-integer problem (NP-hard). Zioutas *et al.* [16] combine new penalty functions with the idea of ϵ-insensitive loss function from support vector machine, to improve efficiency.

Outlier rejection in SLAM and related fields: Early work entrusted outlier rejection to the front-end [17, 18]: measurements are tested against the current belief and unlikely observations are rejected as outliers. These techniques can usually reject a large portion of the outliers; the problem is that even a single outlier can compromise the estimate.

Sünderhauf and Protzel [8] propose to augment the optimization problem with latent variables that are responsible for deactivating outliers. Agarwal *et al.* [19] avoid the introduction of slack optimization variables by acting directly on measurement covariances. Olson and Agarwal [9] discuss the use of a max-mixture distribution to accommodate multiple hypotheses on the noise distribution of a measurement. Latif *et al.* [10] recognize that, rather then solving a joint problem, a good outlier rejection strategy should look for "internally consistent" constraints. The two concepts of *inter* and *intra*-cluster consistency guide the selection of a good subset of constraints; however, testing each subset requires solving a nonconvex optimization problem. Sünderhauf and Protzel [20] present a comprehensive comparative study of existing approaches [8–10].

Casafranca *et al.* [21] propose to minimize the ℓ_1-norm of the residual errors in pose graph optimization, and show that this formulation is less sensitive to the presence of spurious measurements. Lee *et al.* [22] discuss an expectation maximization (EM) approach and use a Cauchy function for the choice of the weights. Indelman *et al.* [23] apply EM in multi robot scenarios. We also remark that robust kernels are implemented in state-of-the-art SLAM libraries [7]. All these approaches use nonconvex optimization at their core. In practice this means that the output heavily depends on the initial guess, hence no guarantees are available.

Outlier rejection has been addressed in other fields with problems similar to pose graph optimization, such as multiview geometry problems. Li [24] formulates ℓ_∞ triangulation as an LP-type problem (a generalization of linear programs) and proposes a technique to discard κ outliers from the measurement set. The same work also discusses limitations of standard algorithms like RANSAC, which hold for pose graph optimization as well. Other approaches to outlier rejection based on nonlinear optimization or heuristics include Dalalyan and Keriven [25] and Day *et al.* [26]. Shames *et al.* [27] apply robust regression to fault detection on graphs.

Our approach shares some ideas with these works. For example, the idea of not penalizing measurements with small residual error is similar to Zioutas *et al.* [16]; however, the final formulation is radically different as we do not mix the two terms (error minimization and penalty for measurement removal) in the cost function and we obtain an LP problem instead of a mixed-integer problem. The idea of using the ℓ_1-norm as a *sparsifier* is at the basis of the *Dantzig selector* [28] and appears in related work [27]. Our approach also shares some similarities with RRR [10]. The method in [10] can be interpreted as a brute force approach to deal with the combinatorial complexity of selecting a subset of "good" measurements: measurements are clustered and each combination is tested for consistency. We show that we can avoid the combinatorial complexity using a (conservative) ℓ_1 relaxation, making measurement selection reliable and fast.

II. From "outlier" Rejection to Selection of Coherent Measurement Sets

The problem of outlier rejection is hard because, in general, it is unobservable whether a measurement is an outlier or an inlier. In this section we describe this fundamental problem, and we present an alternative formulation that focuses on finding subsets of measurements that are "internally coherent".

A. The distinction between inlier and outlier is unobservable

Consider the ideal case in which we know the true configuration $\boldsymbol{x}^\circ$. Then, to test whether a measurement $\boldsymbol{z}_{ij}$ could be generated by the nominal model (1), we could test if it satisfies

$$\left\| \boldsymbol{z}_{ij} - h(\boldsymbol{x}_i^\circ, \boldsymbol{x}_j^\circ) \right\| \leq \beta_{ij}, \tag{2}$$

where β_{ij} is a suitable bound, e.g. a 3σ confidence interval. We are not specifying the nature of the norm $\|\cdot\|$ in (2) as the discussion can be easily generalized to different choices, the most common being the Mahalanobis distance.

If the test fails, we can conclude that the measurement is an outlier, with a confidence given by the choice of β_{ij}. However, if the test succeeds, we cannot conclude anything, because the measurement could be generated by a model different from (1), but can still satisfy the bound (2). Note that a measurement that succeeds at the inlier test is still "dangerous": in fact, an estimator that includes spurious measurements will in general lose the property of consistency (it is more confident than it should be) and, depending on how the spurious measurements are generated, other properties such as unbiasedness. Thus, we cannot reliably distinguish between inliers and outliers, even if we knew the true configuration.

B. Selection of coherent measurements

In this paper we do not use the concept of "inlier" and "outlier"; rather, we use the concept of "coherent" measurement set. A measurement subset is coherent if there exists a configuration that explains those measurements. We know that the set of "inliers" is a coherent subset. However, there are,

in general, other subsets of measurements that are coherent but include some "outliers".

Definition 1 (Coherent Measurement Set): Given the measurement model (1) and an upper bound β_{ij} on the tolerance for each edge, a measurement subset $\mathcal{S} \subseteq \mathcal{E}$ is *internally coherent* if there exists a configuration $\boldsymbol{x} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ such that

$$\|\boldsymbol{z}_{ij} - h(\boldsymbol{x}_i, \boldsymbol{x}_j)\| \leq \beta_{ij}, \quad \text{for all } (i,j) \in \mathcal{S}. \tag{3}$$

Our definition of coherent set depends on the choice of an upper bound β_{ij}. In the UBB (Unknown-But-Bounded noise [29]) interpretation this is the bound on the magnitude of the noise. In a probabilistic interpretation, this is a notion of risk of accepting an "inlier" as incoherent.

If a set is coherent, all its subsets are coherent as well. Therefore, it makes sense to ask for the *maximal* subsets. $\mathcal{S}$ is maximal if there does not exist another internally coherent subset $\mathcal{S}'$ such that $\mathcal{S} \subset \mathcal{S}'$. Many small subsets of measurements are coherent but not interesting. For example, the empty set is coherent (though not maximal), and the set composed by one outlier is coherent as well. Therefore, we impose the requirements that the sets are large enough to guarantee observability of the configuration.

Definition 2 (Observable Subset): A measurement subset $\mathcal{S} \subset \mathcal{E}$ is *observable* if the configuration $\boldsymbol{x}$ is observable given the measurements $\boldsymbol{z}_{ij}$, $(i,j) \in \mathcal{S}$.

Finally, we put in the formulation the possibility that some of the measurements are trusted. For example, in pose graph optimization, we know that the edges corresponding to odometry measurements are not corrupted by outliers. Therefore, we partition the edges $\mathcal{E}$ in the subsets of "trusted" measurements $\mathcal{E}^{\circ}$ (in our case, odometric edges) and the subset of "untrusted" measurements $\mathcal{E}^{\mathsf{L}}$ (loop closures).

Thus the most general way to pose the outlier rejection problem is to ask for all possible subsets that are *coherent*, *observable*, and *maximal*.

Problem 1 (Coherent Measurement Sets Detection): Given:

- the measurement model (1);
- a partition of the edges $\mathcal{E} = \mathcal{E}^{\circ} \cup \mathcal{E}^{\mathsf{L}}$;
- the measurements $\boldsymbol{z}_{ij}$ for all edges $(i,j) \in \mathcal{E}$;
- an upper bound β_{ij} on each measurement;

find all subsets of $\mathcal{E}$ that satisfy the following:

1) They contain all trusted edges $\mathcal{E}^{\circ}$;
2) They are coherent (Definition 1);
3) They are observable (Definition 2);
4) They are maximal with respect to $\subset$.

Note that this is a very hard problem to solve, being combinatorial in nature; however, it is the only proper way to approach the problem without making more assumptions on how outliers are generated. The key innovation of this paper is to focus on the problem of selecting coherent subsets of measurements. Rather than trying to solve an optimization problem that jointly solves for the removal of the outliers and for an estimate of the poses, our objective is to select a set of coherent measurements by solving feasibility problems rather than optimization problems.

III. Coherent Set Detection in the Linear Model

This section describes an approach to detect coherent measurement sets for the linear case, using convex relaxation.

The algorithm we propose builds upon a few ideas:

1) It is natural to look for the largest coherent set, as it is the one that most likely contains the least outliers.
2) This combinatorial problem can be written as an optimization problem in the ℓ_0 norm, using the "big M" trick.
3) The ℓ_0 problem can be (conservatively) relaxed to obtain an ℓ_1 problem that can be efficiently solved.

In the end we obtain a linear program that minimizes an ℓ_1 proxy for the number of measurements to exclude.

A. Linear measurement model on graphs

Consider the case in which each node configuration x_i is a real number and the measurements are of the form $h(x_i, x_j) = x_j - x_i$, thus giving the measurement model

$$z_{ij} = (x_j - x_i) + \epsilon_{ij}, \qquad \text{for } (i,j) \in \mathcal{E}. \tag{4}$$

All results could be extended to the vector case ($\boldsymbol{x}_i \in \mathbb{R}^d$, $d > 1$), but for simplicity we continue in the scalar case.

The model (4) can be written compactly using the vector notation. We stack the m measurements z_{ij} in a vector $\boldsymbol{z} \in \mathbb{R}^m$ ($|\mathcal{E}| = m$), and the configuration in the vector $\boldsymbol{x} \in \mathbb{R}^n$.

We then define the incidence matrix $\boldsymbol{A}$ of the graph [30]. The incidence matrix $\boldsymbol{A}$ is an $n \times m$ matrix, with elements taking values in the set $\{-1, 0, +1\}$. The element A_{ik} is $+1$ if node i is the head of edge k, -1 if node i is the tail of edge k, and it is 0 otherwise. Using this notation, (4) becomes

$$\boldsymbol{z} = \boldsymbol{A}^{\mathsf{T}} \boldsymbol{x} + \boldsymbol{\epsilon}. \tag{5}$$

B. Checking coherence of a measurement set

If the measurements follow model (5), then checking whether the set $\mathcal{E}$ is coherent according to Definition 1 corresponds to solving a linear feasibility program.

To see this, note that, according to Definition 1, the set $\mathcal{E}$ is coherent if $|z_{ij} - (x_j - x_i)| \leq \beta_{ij}$ for all $(i,j) \in \mathcal{E}$ and for some choice of $\boldsymbol{x}$. We write this in vector form by stacking all bounds β_{ij} in a vector $\boldsymbol{\beta}$, obtaining the feasibility problem

$$\text{find } \boldsymbol{x}, \qquad \text{subject to } \left|\boldsymbol{z} - \boldsymbol{A}^{\mathsf{T}} \boldsymbol{x}\right| \leq \boldsymbol{\beta}. \tag{6}$$

We use "$\leq$" to denote component-wise inequality.

While solving this (linear) feasibility problem can be used as a check on the original set of measurements $\mathcal{E}$, in the following we discuss the case in which $\mathcal{E}$ contains spurious measurements and we want to detect a maximal coherent subset of $\mathcal{E}$.

C. Indicator variables and the "big M" trick

We use an indicator variable $b_{ij} \in \{0, 1\}$ for each measurement $(i,j) \in \mathcal{E}^{\mathsf{L}}$ that is not trusted. If b_{ij} is 0 then the measurement is selected, and if b_{ij} is 1 it is excluded.

Given such indicator variables, there are multiple ways to incorporate them in the problem. For example, Sünderhauf and Protzel [8] use indicator variables in the cost function to multiply the penalty for each untrusted edge. In our formalization we do not have a cost function, as we focus on

solving a feasibility problem, but we could use the indicator variables to deactivate the inequality, by writing the untrusted constraints as

$$(1 - b_{ij}) \left| z_{ij} - (x_j - x_i) \right| \leq \beta_{ij}, \quad \text{for all } (i,j) \in \mathcal{E}^{\mathsf{L}}.$$

If b_{ij} is 1, the constraint is automatically satisfied. The drawback of this approach is that it destroys the linearity of the problem.

An alternative is using the "big M" trick, which is used in the optimization literature to show the equivalence of variable selection problems to integer programming [31]. In this approach we add to the right-hand side of the inequality a large constant M that multiplies the indicator variable:

$$\left| z_{ij} - (x_j - x_i) \right| \leq \beta_{ij} + M\, b_{ij}, \quad \text{for all } (i,j) \in \mathcal{E}^{\mathsf{L}}. \quad (7)$$

If b_{ij} is 0, then we obtain the original constraint. If b_{ij} is 1, then the constraint is irrelevant, provided that M is large enough. Therefore, indicator variables control the activation of constraints without losing the linearity property.

D. The ℓ_0 formulation

In this section we introduce a first relaxation of Problem 1. While Problem 1 requires to compute *all* maximal subsets that are coherent and observable, we now only look for the *largest* subset of measurements that is coherent and observable.

Just like the edges $\mathcal{E}$ are split in the subsets $\mathcal{E}^{\circ}$ (odometric edges) and $\mathcal{E}^{\mathsf{L}}$ (loop closures), we split the measurements $\boldsymbol{z}$ in $\boldsymbol{z}^{\circ}$ and $\boldsymbol{z}^{\mathsf{L}}$, and the bounds $\boldsymbol{\beta}$ in $\boldsymbol{\beta}^{\circ}$ and $\boldsymbol{\beta}^{\mathsf{L}}$. Similarly, we split the matrix $\boldsymbol{A}$ in two submatrices: $\boldsymbol{A} = [\boldsymbol{A}_{\circ}\ \boldsymbol{A}_{\mathsf{L}}]$.

For the trusted edges in $\mathcal{E}^{\circ}$, the bounds are written in the original form: $\left| \boldsymbol{z}^{\circ} - \boldsymbol{A}_{\circ}^{\mathsf{T}} \boldsymbol{x} \right| \leq \boldsymbol{\beta}^{\circ}$.

For the untrusted edges in $\mathcal{E}^{\mathsf{L}}$ we use the "big M" trick. We stack the binary variables b_{ij}, $(i,j) \in \mathcal{E}^{\mathsf{L}}$ in a single vector $\boldsymbol{b} \in \{0,1\}^{\ell}$, where $\ell = |\mathcal{E}^{\mathsf{L}}| = m - n + 1$ is the number of cycles in the graph. Writing (7) in vector form gives

$$\left| \boldsymbol{z}^{\mathsf{L}} - \boldsymbol{A}_{\mathsf{L}}^{\mathsf{T}} \boldsymbol{x} \right| \leq \boldsymbol{\beta}^{\mathsf{L}} + M\, \boldsymbol{b}.$$

With this notation we can state the following proposition.

Proposition 3 (ℓ_0 Coherent Measurement Set Detection): A maximum-cardinality coherent and observable measurement set $\mathcal{S}^{\star} \subseteq \mathcal{E}$ can be obtained from a solution $(\boldsymbol{x}^{\star}, \boldsymbol{b}^{\star})$ of the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{x},\, \boldsymbol{b}} \quad & \|\boldsymbol{b}\|_0, \qquad \boldsymbol{b} \in \{0,1\}^{\ell}, \qquad (8) \\ \text{subject to} \quad & \left| \boldsymbol{z}^{\circ} - \boldsymbol{A}_{\circ}^{\mathsf{T}} \boldsymbol{x} \right| \leq \boldsymbol{\beta}^{\circ}, \\ & \left| \boldsymbol{z}^{\mathsf{L}} - \boldsymbol{A}_{\mathsf{L}}^{\mathsf{T}} \boldsymbol{x} \right| \leq \boldsymbol{\beta}^{\mathsf{L}} + M\, \boldsymbol{b}. \end{aligned}$$

by setting $\mathcal{S}^{\star} = \mathcal{E}^{\circ} \cup \{(i,j) : b_{ij}^{\star} = 0\}$.

Proof: For every $\boldsymbol{b}^{\star}$ that solves (8), $\boldsymbol{x}^{\star}$ is an admissible configuration, hence the resulting set is *coherent*. The objective minimizes the number of deactivated constraints, hence $\boldsymbol{b}^{\star}$ describes a *maximum-cardinality* subset. Regarding observability, the set of trusted (odometric) edges defines a spanning path, which guarantees observability of $\boldsymbol{x}$ [11]. ■

E. ℓ_1 relaxation

Problem (8) is nonconvex and combinatorial in nature, so one cannot expect to solve large problem instances in small time. We formulate a relaxed version using an ℓ_1-norm objective, which is known to encourage sparsity as it is a convex proxy of the ℓ_0-norm [32].

Proposition 4 (ℓ_1 Coherent Measurement Set Detection): A coherent and observable measurement set $\mathcal{S}^{\star} \subseteq \mathcal{E}$ can be obtained from a solution $(\boldsymbol{x}^{\star}, \boldsymbol{b}^{\star})$ of the optimization problem

$$\begin{aligned} \min_{\boldsymbol{x},\, \boldsymbol{b}} \quad & \|\boldsymbol{b}\|_1, \qquad \boldsymbol{b} \in \mathbb{R}^{\ell}, \qquad (9) \\ \text{subject to} \quad & \left| \boldsymbol{z}^{\circ} - \boldsymbol{A}_{\circ}^{\mathsf{T}} \boldsymbol{x} \right| \leq \boldsymbol{\beta}^{\circ}, \\ & \left| \boldsymbol{z}^{\mathsf{L}} - \boldsymbol{A}_{\mathsf{L}}^{\mathsf{T}} \boldsymbol{x} \right| \leq \boldsymbol{\beta}^{\mathsf{L}} + M\, \boldsymbol{b}. \end{aligned}$$

by setting $\mathcal{S}^{\star} = \mathcal{E}^{\circ} \cup \{(i,j) : b_{ij}^{\star} = 0\}$.

We omit a formal proof as it is clear that when $b_{ij}^{\star} = 0$ the corresponding constraint in (9) is satisfied and the corresponding measurement is in the coherent set. The observability requirement proof is the same as the ℓ_0 case.

After the ℓ_1 relaxation, the elements of $\boldsymbol{b}^{\star}$ quantify how much a given constraint needs to be enlarged to be feasible. We expect the ℓ_1 norm to give small values for measurements in the coherent set and large values to the outliers (Fig. 2).

The advantage is that Problem (9) is now convex; indeed, it is a *linear program* (LP). LPs can be solved efficiently and scale well to very large problems; moreover, mature state-of-the-art solvers are available [33, 34]. While we preserved the guarantees (observability, coherence), the drawback is that with the ℓ_1 relaxation we cannot guarantee the maximum-cardinality property. Extensive experimental evidence shows that the ℓ_1-norm encourages sparsity [32], but in general we cannot guarantee that the solution sets of (8) and (9) coincide.

F. Simultaneous decisions over dependent constraints

We further extend the formulation to account for dependent constraints. For example, in pose graph optimization, we want the same decision variable to activate both cartesian and angular constraints of a given relative pose measurement. Suppose that there are only $s < \ell$ decisions to take, and let $\boldsymbol{b} \in \mathbb{R}^s$. We can let entry of the decision vector $\boldsymbol{b}$ control the activation of multiple constraints by substituting the term $M\boldsymbol{b}$ in (9) with a term $\boldsymbol{M}\boldsymbol{b}$, where $\boldsymbol{M}$ is a $\ell \times s$ matrix, and M_{hk} is nonzero if the k-th decision variable controls the h-th constraint. If different constraints use different units (such as meters and radians), the nonzero values of $\boldsymbol{M}$ can normalize all of them to the same unit.

With this extension, Problem (9) becomes:

$$\begin{aligned} \min_{\boldsymbol{x},\, \boldsymbol{b}} \quad & \|\boldsymbol{b}\|_1, \qquad \boldsymbol{b} \in \mathbb{R}^s, \qquad (10) \\ \text{subject to} \quad & \left| \hat{\boldsymbol{z}}^{\circ} - \boldsymbol{B}_{\circ}^{\mathsf{T}}\, \boldsymbol{x} \right| \leq \boldsymbol{\beta}^{\circ}, \\ & \left| \hat{\boldsymbol{z}}^{\mathsf{L}} - \boldsymbol{B}_{\mathsf{L}}^{\mathsf{T}}\, \boldsymbol{x} \right| \leq \boldsymbol{\beta}^{\mathsf{L}} + \boldsymbol{M}\boldsymbol{b}. \end{aligned}$$

IV. Coherent Set Detection in Planar Pose Graph Optimization

In this section we discuss the case of planar pose optimization, where the node configuration lies in $\mathrm{SE}(2)$ and the

observations are the relative poses between nodes. We show that the problem of outlier detection in this nonlinear setting can be reduced to solving linear problems, just like it has been shown in previous work in the outlier-free setting [5, 11, 12].

A. The planar pose graph optimization setting

We parametrize the node configuration $\boldsymbol{x}_i \in \mathrm{SE}(2)$ using the node orientation $\theta_i \in (-\pi, +\pi]$ and the position $\boldsymbol{p}_i \in \mathbb{R}^2$:

$$\boldsymbol{x}_i = (\boldsymbol{p}_i, \theta_i)\,.$$

For each edge $(i,j) \in \mathcal{E}$ we are given noisy observations of the relative pose $\boldsymbol{x}_i^{-1}\boldsymbol{x}_j$. We parametrize these relative observations using the relative angular measurements $\delta_{ij} \in (-\pi, +\pi]$ and the relative position measurements $\boldsymbol{d}_{ij} \in \mathbb{R}^2$.

The nominal model for the relative angular measurements $\delta_{ij} \in (-\pi, +\pi]$ is

$$\delta_{ij} = \langle \theta_j - \theta_i + \epsilon_{ij}^{\theta} \rangle_{2\pi}, \tag{11}$$

where $\langle\cdot\rangle_{2\pi}$ is the 2π-modulo operator that normalizes the angular measurements in the interval $(-\pi, +\pi]$. The noise ϵ_{ij}^{θ} is Gaussian with variance σ_{ij}^{θ}. The nominal model for the relative position measurements $\boldsymbol{d}_{ij} \in \mathbb{R}^2$ is

$$\boldsymbol{d}_{ij} = \boldsymbol{R}(\theta_i)^{\mathsf{T}}(\boldsymbol{p}_j - \boldsymbol{p}_i) + \boldsymbol{\epsilon}_{ij}^{p}, \tag{12}$$

where $\boldsymbol{R}(\theta_i)$ is a planar rotation of an angle θ_i and $\boldsymbol{\epsilon}_{ij}^{p}$ is Gaussian noise with covariance $\boldsymbol{P}_{ij}^{d}$. As in [11] we assume the noises ϵ_{ij}^{θ} and $\boldsymbol{\epsilon}_{ij}^{p}$ to be independent.

B. Linear approaches to planar pose graph optimization

Previous work has shown that the problem of planar pose graph optimization, without outliers, has an efficient approximated solution using linear estimation [5, 11] and has global solutions using mixed-integer programming [12].

We summarize the basic ideas that we re-use in this paper:

1) The estimation problem can be decoupled in angular and translational components because estimating the orientations $\boldsymbol{\theta}$ given only angular constraints gives a very good estimate; in other words, the posterior $p(\boldsymbol{\theta}|\boldsymbol{z}_\theta)$ is already peaked.

2) If the orientations $\boldsymbol{\theta}$ are known, estimating translations is a linear problem, as can be seen from (12).

3) The problem of estimating the orientations alone is "almost" linear. The observation model when written in the form (11) is nonlinear because of $\langle\cdot\rangle_{2\pi}$; however, if we introduce extra integer variables $k_{ij} \in \mathbb{Z}$ in the formalization, it can be written as a purely *linear* observation model:

$$\delta_{ij} = (\theta_j - \theta_i) + 2\pi\, k_{ij} + \epsilon_{ij}. \tag{13}$$

4) If the angular noise is low, then the values of k_{ij} can be computed by setting them to zero for edges belonging to the odometric spanning tree, and by rounding errors accumulated along cycles to the closest multiple of 2π for loop closures. For a loop closure edge $u = (i,j)$, consider the fundamental circuit composed by that edge and a subset of edges from the odometric spanning tree. Define a vector $\boldsymbol{c}^u \in \mathbb{R}^\ell$, such that $\boldsymbol{c}_v^u$ is 0, if the v-th edge does not appear in the fundamental circuit associated to edge u; and it is 1 (respectively, -1) if it is traversed forward (respectively, backward) in the circuit. The approximation for k_u is

$$\hat{k}_u = \begin{cases} 0, & \text{if } u \in \mathcal{E}^{\circ}, \\ \mathrm{round}\left(\frac{1}{2\pi}\sum_v \boldsymbol{c}_v^u\, \delta_v\right), & \text{if } u \in \mathcal{E}^{\mathsf{L}}. \end{cases} \tag{14}$$

The rounding returns the correct value of k_{ij} when the orientation noise is small [11], which is the case for most SLAM datasets, but it fails when the noise is large [12].

5) In the outlier-free case, given the integers k_{ij}, an estimate for $\boldsymbol{\theta}$ can be computed via linear least squares.

C. Coherent set detection in planar pose optimization

In the following we propose an approach to detect a coherent measurement set in a planar optimization problem.

The method works in two stages. In the first stage (Section IV-D) we consider only the relative rotation measurements. We use the regularization terms given in (14) to obtain a purely linear problem for the orientations $\boldsymbol{\theta}$. Therefore, we can directly apply the result in Proposition 4 that returns a set of observable and coherent measurements. This set can be then used to estimate the nodes orientations in $\mathrm{SO}(2)$.

In the second stage (Section IV-E), we use the angular estimates from the first stage to reformulate the relative pose measurements using a linear model. Then we use (10) to take simultaneous decisions over angular and translational components of a relative pose measurement. The output of the second stage is a pair $(\boldsymbol{x}^\star, \boldsymbol{b}^\star)$. The vector $\boldsymbol{b}^\star$ represents the (relaxed) selection of the coherent measurement set. The configuration $\boldsymbol{x}^\star$ is merely a by-product in our formulation, and can be disregarded, used as pose estimate, or used to bootstrap nonlinear estimation using the coherent measurement set.

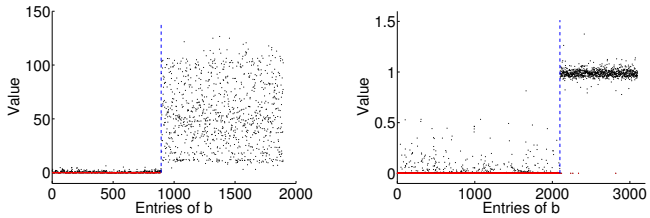
D. First Stage: selection of coherent set of relative rotations

We use (14) to obtain the regularization terms k_{ij}. Given these regularization factors we rewrite the observation model (11) as in (13). Model (13) is linear, therefore we can use the approach presented in Section III for the detection of a coherent measurement set.

Call $\hat{\boldsymbol{z}}_\theta$ the vector obtained by stacking all the *regularized* orientation measurements, i.e., $\delta_{ij} - 2\pi k_{ij}$. Using $\hat{\boldsymbol{z}}_\theta$ we can write the (nominal) measurement model (13) as

$$\hat{\boldsymbol{z}}_\theta = \boldsymbol{A}^{\mathsf{T}}\boldsymbol{\theta} + \boldsymbol{\epsilon}_\theta, \tag{15}$$

which is identical to the linear model assumed in (5). Therefore, it is straightforward to use Proposition 4 to compute an observable and coherent set of angular measurements.



(a) (b)

Fig. 2. Example of solution $\boldsymbol{b}^\star$ of the ℓ_1 problem. Measurements with index on the right of the blue vertical line are outliers. (a) INTEL with $\kappa = 1000$ outliers and with weight matrix $\boldsymbol{M}$ as in (20); (b) M3500 with $\kappa = 1000$ outliers and weight matrix $\boldsymbol{M}$ as in (21).

We can then use this set of measurements to compute an estimate of the nodes' orientations: using standard linear estimation over the coherent set we obtain an estimate of the rotations $\hat{\boldsymbol{\theta}}$ and the corresponding covariance $\boldsymbol{P}_{\boldsymbol{\theta}}$.

1) Impact of Outliers on Regularization: Note that so far we are somehow assuming a correct computation of the regularization terms k_{ij}. Since the integers k_{ij} are computed from the measurements, and those are possibly outliers, this point deserves a remark. According to (14), the regularization terms for the odometric edges will be all zeros. This is a first advantage, as the choice of k_{ij} cannot influence the odometric edges (assumed to be inliers), making them untrusted. Moreover, the regularization term for the loop closure (i, j) only depends on the loop closure itself and on a set of odometric edges; this implies that if the loop closure is not an outlier, the computation of the regularization term will be correct. If the loop closure is an outlier, an incorrect computation of the regularization term will only offset the measurement of a multiple of 2π, while our approach will be in charge of deciding on his coherence with respect to the other measurements.

E. Second Stage: Selection of Coherent Set of Relative Poses

In the second stage we assume to have an estimate of the orientations $\hat{\boldsymbol{\theta}}$ and their covariance matrix $\boldsymbol{P}_{\boldsymbol{\theta}}$.

1) Linear model for translations: Let us rewrite the nominal model (12) by using the approximation $\theta_i = \hat{\theta}_i$ as to obtain

$$\boldsymbol{d}_{ij} = \boldsymbol{R}(\hat{\theta}_i)^{\mathsf{T}}(\boldsymbol{p}_j - \boldsymbol{p}_i) + \boldsymbol{\epsilon}^p_{ij}. \tag{16}$$

Multiplying both sides by $\boldsymbol{R}(\hat{\theta}_i)$ we obtain

$$\boldsymbol{R}(\hat{\theta}_i)\, \boldsymbol{d}_{ij} = (\boldsymbol{p}_j - \boldsymbol{p}_i) + \boldsymbol{R}(\hat{\theta}_i)\, \boldsymbol{\epsilon}^p_{ij}. \tag{17}$$

The noise $\hat{\boldsymbol{\epsilon}}^p_{ij} \doteq \boldsymbol{R}(\hat{\theta}_i)\, \boldsymbol{\epsilon}^p_{ij}$ is exactly Gaussian only for a fixed θ_i. Nevertheless, we can obtain a conservative approximation as a Gaussian variable by computing first and second-order moments, using the uncertainty in $\hat{\boldsymbol{\theta}}$, described by the covariance $\boldsymbol{P}_{\boldsymbol{\theta}}$, and the original distribution of $\boldsymbol{\epsilon}^p_{ij}$.

Rewrite the left hand-side as $\hat{\boldsymbol{d}}_{ij} \doteq \boldsymbol{R}(\hat{\theta}_i)\boldsymbol{d}_{ij}$. We have obtained a purely linear measurement model for the translations:

$$\hat{\boldsymbol{d}}_{ij} = (\boldsymbol{p}_j - \boldsymbol{p}_i) + \hat{\boldsymbol{\epsilon}}^p_{ij}. \tag{18}$$

2) Joint measurement model: We can now combine the linear models (5) and (18) to define the joint (nominal) linear model involving both orientation and position measurements. To do so we need to define a configuration $\boldsymbol{x} \in \mathbb{R}^{3n}$ that contains both angular and positional variables, a vector of observations $\hat{\boldsymbol{z}} \in \mathbb{R}^{3m}$ that contains the relative angular and translation measurements, and a joint noise vector $\hat{\boldsymbol{\epsilon}}$ that contains the noise terms. Finally, we define a matrix $\boldsymbol{B}$ with entries in $\{-1, 0, +1\}$ derived from the incidence matrix $\boldsymbol{A}$ and write the joint measurement model as

$$\hat{\boldsymbol{z}} = \boldsymbol{B}\, \boldsymbol{x} + \hat{\boldsymbol{\epsilon}}. \tag{19}$$

As before, we divide the measurements in trusted and untrusted, and distinguish quantities related to odometric edges ($\hat{\boldsymbol{z}}^{\circ}$, $\boldsymbol{B}_{\circ}$, $\boldsymbol{\beta}^{\circ}$) from those related to loop closures ($\hat{\boldsymbol{z}}^{\mathsf{L}}$, $\boldsymbol{B}_{\mathsf{L}}$, $\boldsymbol{\beta}^{\mathsf{L}}$).

3) Joint coherent measurement detection: We have obtained a joint linear measurement model for both translations and rotations. Note, however, that the measurements are now dependent, because each edge corresponds to three measurements (one angular, two cartesian). Hence we use the extension described in Section III-F to let one decision variable control three constraints. Thus we write (10) with a $3\ell \times \ell$ matrix $\boldsymbol{M}$, as we have 3ℓ constraints for ℓ decision variables. The three rows in $\boldsymbol{M}$ corresponding to the k-th relative pose measurement are zero everywhere, except for the k-th column.

There are many possibilities to choose the nonzero entries in $\boldsymbol{M}$, as they act similarly to weights in the ℓ_1 objective. Candes *et al.* [32] have shown that varying the weights enhances the effectiveness of the ℓ_1-norm to enforce sparsity.

In our formulation we need to weight the elements of each block to normalize Cartesian and angular measurements; but we can choose the relative importance for each constraint.

The naive choice is to simply choose the elements to be be proportional to the standard deviations. The only nonzero elements in $\boldsymbol{M}$ are those corresponding to entries $(3k + a, k)$ (using zero-based indices), and take the value

$$\boldsymbol{M}_{3k+a,k} = \hat{\sigma}_{3k+a}. \tag{20}$$

Here, k ranges from 0 to $\ell - 1$, and a ranges in $\{0, 1, 2\}$.

A smarter strategy exploits graph structure. As each untrusted measurement corresponds to a loop closure we use the absolute value of the error accumulated along the corresponding loop in the corresponding row of $\boldsymbol{M}$. Using the notation for the fundamental circuits introduced in Section IV-B the elements of $\boldsymbol{M}$ can be written as

$$\boldsymbol{M}_{3k+a,k} = \left|\textstyle\sum_{k'} \boldsymbol{c}^k_{k'}\, \hat{z}_{3k'+a}\right|. \tag{21}$$

Note that in the noiseless case, measurements sum to zero along circuits [11]. Intuitively, in the noisy case, (21) promotes nonzero entries in $\boldsymbol{b}^{\star}$ for the loops that accumulate large errors. Examples of $\boldsymbol{b}^{\star}$ for the two choices of $\boldsymbol{M}$ are shown in Fig. 2

F. Summary of guarantees and approximations

The application of Proposition 4 to the linear model (10), allows to draw the following conclusion: the zero entries of the optimal solution $\boldsymbol{b}^{\star}$, computed from (9), describe a coherent set of relative pose measurements. Because of the ℓ_1 relaxation we cannot conclude that the set is maximal and we discuss in the experimental section the practical implications of this fact.

We remark two underlying assumptions that ensure validity of (10). First, the estimator $\hat{\boldsymbol{\theta}}$ is Normally distributed, with covariance $\boldsymbol{P}_{\boldsymbol{\theta}}$; this assumption can be challenged by the fact that we may have outliers in the coherent set computed in the first stage: this suggests to set strict bounds for the first stage, so to select only measurements generated by the nominal model. Second, we are assuming that the noise on the relative measurements (17) can be approximated as Gaussian; this assumption is mild, as in our approach the covariances only guide the selection of the bounds $\boldsymbol{\beta}$. Experiments show that when we choose $\boldsymbol{\beta}$ close to nominal bounds (3σ), we may obtain "open-minded" results, like in Fig. 1b. When we choose $\boldsymbol{\beta}$ to be more conservative (e.g., 2σ), we obtain results

qualitatively similar to methods in the literature, in the sense that the "outliers" are rejected, and we can robustly estimate trajectories close to the optimal outlier-free estimate.

Finally, the choice of $\boldsymbol{\beta}$ is complicated by the fact that the covariance information in datasets is typically unreliable.

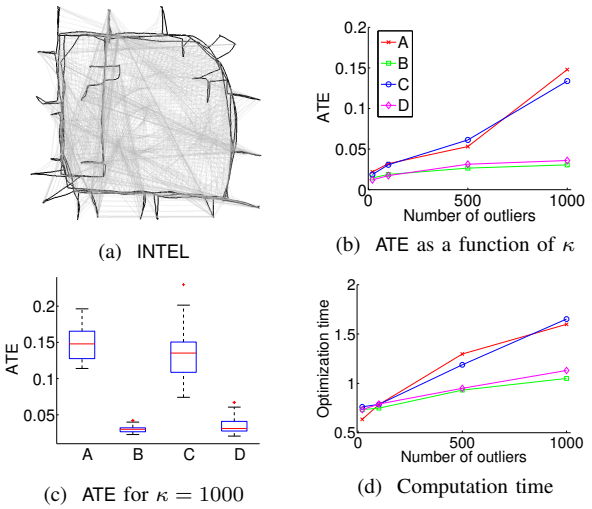
V. Experiments on Robust Pose Graph Optimization

1) Datasets: The Bicocca datasets ($n = 8358$, m varies between 8380 and 8804 depending on the problem instance) [35] contains 41 problem instances with different loop closing edges due to different thresholds for detection of place similarity [10]. Outliers are due to sensory aliasing (Fig. 1a). Ground truth is provided by differential GPS [36]. The INTEL dataset ($n = 943$, $m = 1837$) is a real dataset without outliers. The M3500 dataset ($n = 3500$, $m = 5598$) is a synthetic dataset without outliers. For INTEL and M3500 we used the version published on the Vertigo website [37] and we take as ground truth the optimized trajectory of the outlier-free data.

2) Outlier generation model: We use the outliers generation models described by Sünderhauf & Protzel [8] and their software Vertigo [37] to add outliers to the datasets INTEL and M3500. The four models we used are: A (random), B (local), C (randomly grouped, in groups of 20), D (locally grouped). We vary the number of outliers κ as 20, 100, 500, 1000.

3) Implementation notes: Our approach is implemented in Matlab and uses CVX-MOSEK [34, 38] as parser/solver for the ℓ_1 problems. After solving the ℓ_1 problem we select the constraints with $b^\star_{ij} = 0$ to create a new outlier-free dataset, which is then optimized using g2o. The choice of the back-end once outliers are rejected is not critical, and we use g2o for having a direct comparison with the results in [20]. Experiments are run on a Desktop computer with Intel i7 processor (3.4Ghz).

For Bicocca, which has $n = 8358$ nodes, we neglect the uncertainty $\boldsymbol{P_\theta}$ in the orientation estimate from the first stage to avoid inverting a large matrix; in that dataset this is a good approximation because angular measurements are very accurate.



(a) INTEL (b) ATE as a function of κ (c) ATE for $\kappa = 1000$ (d) Computation time

Fig. 3. INTEL dataset: (a) Estimated trajectory for an instance with $\kappa = 1000$ randomly grouped outliers; (b) Mean ATE for $\{20, 100, 500, 1000\}$ injected outliers and the four generative models (Section V-2). (c) Bar plot of ATE for $\kappa = 1000$. (e) Average computation time in seconds.

4) Performance measures: We use the performance measures used by [20]. The ATE (Absolute Trajectory Error) performance measure is the mean of the norm of the Cartesian errors with respect to the ground truth. For the cases in which we have a ground truth for the set of outliers (INTEL, M3500) we compute the ATE with respect to the optimal trajectory computed from the outlier-free data.

5) Results: For the INTEL dataset we show the results using 1σ bounds in the first stage, 2σ bounds in the second (see comments in Section IV-F), and the simple choice (20) for M. The ATE is generally low for all outlier generation models, with the maximum equal to 0.25 m (Fig. 3b). In general, we obtain better results for models B (local) and D (locally grouped). For models A (random) and C (randomly grouped) the method tends to discard many inliers; this is a lack of precision rather than catastrophic failure. The ATE appears to have linear relation with the number of outliers, but it increases more sharply for models A and C (Fig. 3c). Computation time appears to increase linearly with the number of outliers and is between 0.5 to 1.5 seconds for 1000 outliers (Fig. 3d).

For the Bicocca dataset we show the results using 1σ and 2σ bounds in the two stages. The ATE averaged over the 41 runs was 1.65m, while the maximum recorded value was 3m. These are practically the same results obtained by RRR [35] which has the best results over other state-of-the-art techniques as reported by Sünderhauf & Protzel ([20], Table III). Our method is faster than RRR on comparable machines (1.2 s vs 2.29 s).

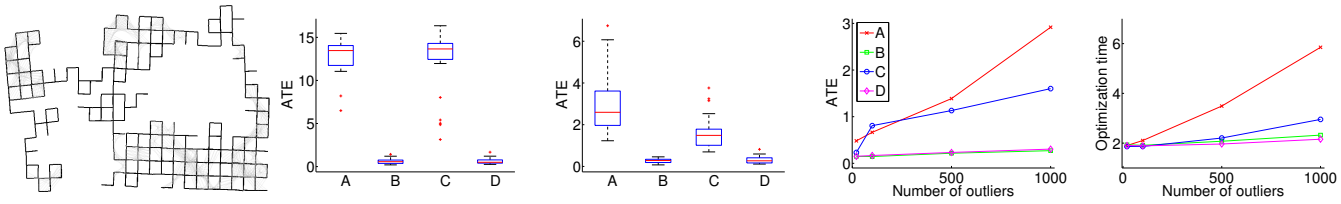
The M3500 dataset is more challenging, as it is less constrained and thus outliers are more likely to be coherent (Fig. 1b). For this dataset we observed that the declared covariances, in the version published on the Vertigo website [37], are considerably larger than what can be identified from the data and the ground truth (actual standard deviation are 10 times smaller than the declared ones). Thus we use 0.1σ for the first stage, and 0.2σ for the second stage using the declared σ; this corresponds to use 1σ and 2σ of the actual covariances.

If we use the naive choice of $\boldsymbol{M}$ (20), we notice the same trends observed for the INTEL dataset, where models B and D are handled much better than A and C. However, the ATE is much larger for this dataset (Fig. 4b) and it is unacceptable.

When we use the weighted $\boldsymbol{M}$ (21), we notice a dramatic improvement in the ATE (Fig. 4c). (For the INTEL dataset there is not much difference between using the two strategies because it is a well constrained dataset.) Intuitively, the second strategy makes easier to disable loop closures that accumulate large errors along the corresponding fundamental cycle; from the optimization perspective, eq. (21) is a good rescaling of the entries of $\boldsymbol{b}$, making the ℓ_1 relaxation a better proxy of the ℓ_0-norm. We notice that the ATE is absolutely insensitive to the number of outliers for cases B and D (Fig. 4d). Optimization time is in the range 2 to 6 s for this dataset (Fig. 4e).

VI. Conclusion

If one does not assume a generative model for faulty measurements, outliers are not observable. As a consequence,



(a) (b) (c) (d) (e)

Fig. 4. M3500 dataset: (a) Estimated trajectory for an instance with $\kappa = 1000$ locally grouped outliers; (b) Bar plot of ATE for different generative models of outliers ($\kappa = 1000$), using the naive choice of $\boldsymbol{M}$ (20); (c) Bar plot of ATE for different generative models of outliers ($\kappa = 1000$), using $\boldsymbol{M}$ as in (21); (d) Mean ATE for $\{20, 100, 500, 1000\}$ injected outliers, using (21), and for generative model A (red), B (green), C (blue), D (magenta); (e) Average optimization time for $\{20, 100, 500, 1000\}$ injected outliers, using (21), and for generative model A (red), B (green), C (blue), D (magenta).

in this work, rather than looking for inliers, we define and provide practical strategies to identify a set of "internally coherent" measurements. In the linear case, we propose an ℓ_0-norm formulation that returns the largest set of coherent measurements. As ℓ_0-norm minimization is hard, we consider a convex relaxation of the problem. Our final formulation is a linear program in which an ℓ_1-norm objective function encourages large (but non necessarily maximal) sets of inliers. We apply our formulation to planar pose graph optimization and we show that it allows robust and efficient solution of large problem instances with many outliers. We leave as future work a thorough investigation of the impact of the choice of weights on the ℓ_1-norm relaxation.

REFERENCES

[1] E. Olson, J. Leonard, and S. Teller. "Fast iterative alignment of pose graphs with poor estimates". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2006.

[2] G. Grisetti, C. Stachniss, and W. Burgard. "Non-linear Constraint Network Optimization for Efficient Map Learning". In: *IEEE Trans. on Intelligent Transportation Systems* 10.3 (2009).

[3] M. Kaess, A. Ranganathan, and F. Dellaert. "iSAM:Incremental Smoothing and Mapping". In: *IEEE Trans. on Robotics* 24.6 (2008).

[4] M. Kaess et al. "iSAM2:Incremental Smoothing and Mapping Using the Bayes Tree". In: *Int. J. Robot. Res.* 31 (2012).

[5] L. Carlone et al. "A Fast and Accurate Approximation for Pose Graph Optimization". In: *Int. J. Robot. Res.* (2012). accepted, available at: http://www.lucacarlone.com/index.php/publications.

[6] C. Stachniss, U. Frese, and G. Grisetti. *OpenSLAM*. URL: `http://www.openslam.org/`.

[7] F. Dellaert. *GTSAM*. URL: `https://borg.cc.gatech.edu`.

[8] N. Sünderhauf and P. Protzel. "Switchable Constraints for Robust Pose Graph SLAM". In: *Proc. of the IEEE/RSJ Int. Conf. on Intelligent RObots and Systems (IROS)*. 2012.

[9] E. Olson and P. Agarwal. "Inference on networks of mixtures for robust robot mapping". In: *Robotics: Science and Systems (RSS)*. 2012.

[10] Y. Latif, C. Cadena, and J. Neira. "Robust loop closing over time for pose graph SLAM". In: *Int. J. Robot. Res.* 32.14 (2013).

[11] L. Carlone et al. "A Linear Approximation for Graph-based Simultaneous Localization and Mapping". In: *Robotics: Science and Systems (RSS)*. 2011.

[12] L. Carlone and A. Censi. "From Angular Manifolds to the Integer Lattice: Guaranteed Orientation Estimation with Application to Pose Graph Optimization". In: IEEE Trans. on Robotics, accepted, ArXiv preprint: http://arxiv.org/abs/1211.3063, 2012. 2014.

[13] P. Huber. *Robust Statistics*. John Wiley and Sons, 2009.

[14] P. Rousseeuw and A. Leroy. *Robust regression and outlier detection*. John Wiley and Sons, 1987.

[15] G. Zioutas and A. Avramidis. "Deleting Outliers in Robust Regression with Mixed Integer Programming". In: *Acta Mathematicae Applicatae Sinica* 21.2 (2005).

[16] G. Zioutas, L. Pitsoulis, and A. Avramidis. "Quadratic mixed integer programming and support vectors for deleting outliers in robust regression". In: *Annals of Operations Research* 166.1 (2009).

[17] J. Neira and J. Tardós. "Data association in stochastic mapping using the joint compatibility test". In: *IEEE Trans. Robot. Automat.* 17.6 (2002).

[18] J. Leonard et al. "Towards Robust Data Association and Feature Modeling for Concurrent Mapping and Localization". In: *Springer Tracts in Advanced Robotics* 6 (2003).

[19] P. Agarwal et al. "Robust map optimization using dynamic covariance scaling". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2013.

[20] N. Sünderhauf and P. Protzel. "Switchable constraints vs. max-mixture models vs. RRR - A comparison of three approaches to robust pose graph SLAM". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2013.

[21] J. Casafranca, L. Paz, and P. Piniès. "A back-end L1 norm based solution for Factor Graph SLAM". In: *Proc. of the IEEE/RSJ Int. Conf. on Intelligent RObots and Systems (IROS)*. 2013.

[22] G. Lee, F. Fraundorfer, and M. Pollefeys. "Robust pose-graph loop-closures with expectation-maximizatio". In: *Proc. of the IEEE/RSJ Int. Conf. on Intelligent RObots and Systems (IROS)*. 2013.

[23] V. Indelman et al. "Multi-Robot Pose Graph Localization and Data Association from Unknown Initial Relative Poses via Expectation Maximization". In: *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2014.

[24] H. Li. "A practical algorithm for L∞ triangulation with outliers". In: *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2007.

[25] A. Dalalyan and R. Keriven. "L1-penalized robust estimation for a class of inverse problems arising in multiview geometry". In: *In Annual Conference on Neural Information Processing Systems*. 2009.

[26] Y. Dai, H. Mingyi, and L. Hongdong. "Two Efficient Algorithms for Outlier Removal in Multi-view Geometry Using L∞ Norm". In: *Fifth Intl. Conf. on Image and Graphics*. 2009.

[27] I. Shames et al. "Fault Detection and Mitigation in Kirchhoff Networks". In: *IEEE Signal Processing Letters* 19.11 (2012).

[28] E. Candès and T. Tao. "The Dantzig selector: Statistical estimation when p is much larger than n". In: *Ann. Statist.* 35.6 (2007).

[29] L. E. Ghaoui and G. Calafiore. "Robust Filtering for Discrete-Time Systems with Bounded Noise and Parametric Uncertainty". In: *IEEE Trans. on Automatic Control* 46 (2001).

[30] W. Chen. *Graph Theory and Its Engineering Applications*. Advanced Series in Electrical and Computer Engineering, 1997.

[31] H. Taha. *Operations research: an introduction*. Volume 8. NJ: Prentice Hall Upper Saddle River, 1997.

[32] E. Candès, M. Wakin, and S. Boyd. "Enhancing Sparsity by Reweighted L1 Minimization". In: *J. Fourier Anal. Appl.* 14.6 (2008).

[33] I. Corporation. *IBM ILOG CPLEX Optimization Studio*. URL: `http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/`.

[34] E. Andersen and K. Andersen. *MOSEK*. URL: `http://www.mosek.com/`.

[35] Y. Latif. *RRR*. URL: `github.com/ylatif/rrr.git`.

[36] A. Bonarini et al. "RAWSEEDS: Robotics Advancement through Web-publishing of Sensorial and Elaborated Extensive Data Sets, http://www.rawseeds.org/". In: *Proc. of the IROS'06 Workshop on Benchmarks in Robotics Research*. 2006.

[37] N. Sünderhauf. *Vertigo: Versatile Extensions for Robust Inference using Graph Optimization*. URL: `http://openslam.org/vertigo.html`.

[38] M. Grant and S. Boyd. *CVX: Matlab software for disciplined convex programming*. URL: `http://cvxr.com/cvx`.