Lifelong, Learning-Augmented Robot Navigation

by

Kevin J. Doherty
B.E., Stevens Institute of Technology (2017)
S.M., Massachusetts Institute of Technology (2019)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
and
WOODS HOLE OCEANOGRAPHIC INSTITUTION
February 2023

Author .....
Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
January 11, 2023

Certified by .....
Nicholas Roy
Professor of Aeronautics and Astronautics
Massachusetts Institute of Technology
Thesis Chair

Certified by .....
John J. Leonard
Samuel C. Collins Professor of Mechanical and Ocean Engineering
Massachusetts Institute of Technology
Thesis Supervisor

Certified by .....
Luca Carlone
Professor of Aeronautics and Astronautics
Massachusetts Institute of Technology

Certified by .....
Yogesh Girdhar
Research Scientist
Woods Hole Oceanographic Institution

Accepted by .....
Jonathan P. How
R. C. Maclaurin Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee
Massachusetts Institute of Technology

Accepted by .....
David Ralston
Chairman, Joint Committee for Applied Ocean Science & Engineering
Woods Hole Oceanographic Institution

Lifelong, Learning-Augmented Robot Navigation

by

Kevin J. Doherty

Submitted to the Department of Aeronautics and Astronautics
on January 11, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Simultaneous localization and mapping (SLAM) is the process by which a robot constructs a global model of an environment from local observations of it; this is a fundamental perceptual capability supporting planning, navigation, and control. We are interested in improving the expressiveness and operational longevity of SLAM systems. In particular, we are interested in leveraging state-of-the-art machine learning methods for object detection to augment the maps robots can build with *object-level* semantic information. To do so, a robot must combine continuous geometric information about its trajectory and object locations with discrete semantic information about object classes. This problem is complicated by the fact that object detection techniques are often unreliable in novel environments, introducing outliers and making it difficult to determine the correspondence between detected objects and mapped landmarks. For robust long-term navigation, a robot must contend with these discrete sources of ambiguity. Finally, even when measurements are not corrupted by outliers, long-term SLAM remains a challenging computational problem: typical solution methods rely on local optimization techniques that require a good "initial guess," and whose computational expense grows as measurements accumulate.

The first contribution of this thesis addresses the problem of inference for hybrid probabilistic models, i.e., models containing both discrete and continuous states we would like to estimate. These problems frequently arise when modeling e.g., outlier contamination (where binary variables indicate whether a measurement is corrupted), or when performing object-level mapping (where discrete variables may represent measurement-landmark correspondence or object categories). The former application is crucial for designing more robust perception systems. The latter application is especially important for enabling robots to construct semantic maps; that is, maps containing objects whose states are a mixture of continuous (geometric) information and (discrete) categorical information (such as class labels). The second contribution of this thesis is, a novel *spectral initialization* method which is efficient to compute, easy to implement, and *admits the first formal performance guarantees* for a SLAM initialization method. The final contribution of this thesis aims to curtail the growing computational expense of long-term SLAM. In particular, we propose an efficient algorithm for graph sparsification capable of reducing the computational burden of

SLAM methods without significantly degrading SLAM solution quality. Taken together, these contributions improve the *robustness* and *efficiency* of robot perception approaches in the lifelong setting.

Thesis Supervisor: John J. Leonard
Title: Samuel C. Collins Professor of Mechanical and Ocean Engineering
Massachusetts Institute of Technology

Acknowledgments

This thesis would not have been possible without the support of so many wonderful people. I would like to thank my advisor, John Leonard, who gave me the opportunity to join the Marine Robotics Group, and has been a constant source of positivity and encouragement throughout my time at MIT.

I am grateful to Nick Roy, who responded to one fateful email back in 2016, and ended up providing a home base for me in the Robust Robotics Group during my first year of grad school (and plenty of advice along the way). The RRG housed so many inspiring people during my time there, and I am so thankful for all of them.

Erin Fischell, Luca Carlone, and Michael Kaess have been stellar committee members. I appreciate their wisdom and patience as the ideas in this thesis came together. More importantly, all of them have been great role models and supporters of my development as a researcher.

I would like to thank the members of the Marine Robotics Group, past and present. I couldn't imagine a more positive and more supportive group of people to work with for the past several years. I especially want to thank Ziqi Lu, Kurran Singh, David Baxter, and Eddie Schneeweiss, who I had the opportunity to collaborate with on works that directly appeared in this thesis. However, all of the folks in the MRG, including Alan, Carol, Brendan, Chad, Tonio, Violet, Yihao, Dehann, and Pedro, have contributed to this work in one way or another. I'm grateful to Nira for keeping things running smoothly in the MRG and helping me navigate the administrative intricacies of MIT.

David Rosen was a spectacular collaborator and mentor on a number of projects appearing in this thesis. I am thankful for all of the time and energy he put into helping me improve the mathematical rigor of my work.

I have been incredibly fortunate to be a part of both the MIT-WHOI Joint Program and the MIT AeroAstro department during my time as a PhD student. Many people are involved in keeping things afloat in these programs, but I especially owe thanks to Beth, Kris, Andone, and Lea, who made it possible for me to be a part of

both of these fantastic programs simultaneously.

I owe a tremendous debt of gratitude to Brendan Englot and the members of his Robust Field Autonomy Lab at Stevens Institute of Technology. It was through Brendan and his lab that I found my way into robotics in the first place, and ultimately ended up pursuing my PhD. I have immensely enjoyed continued collaboration with the folks in the RFAL.

Many friends have made this place my home for the past several years. I've been lucky to have great roommates (Adrian and Alex), climbing/hiking buddies (Ellen, Jing, Fiona, Drew), and a fantastic joint program cohort (Sunfish, you know who you are!). At the same time, I'd like to thank all my friends from before grad school, especially Dave, Reeba, Julia, Jason, Colin, and Downey, for being amazing friends over all these years.

I want to thank my family, especially my parents, grandparents, and siblings, who have supported me every step of the way. Finally, I *need* to thank my partner, Ashley. She is the most patient person in the world, hearing more of my research ramblings and practice talks than perhaps anyone else over the past several years. I am so thankful to have had her love and support all along the way, and I could not have found a better person to navigate life with.

This work was generously supported by the NSF Graduate Research Fellowship Program (GRFP), ONR Neuro-Autonomy MURI grant N00014-19-1-2571, ONR grant N00014-18-1-2832, and the MIT-Portugal Program Flagship Project: Knowledge to Data.

Contents

1 Introduction **15**

1.1 Motivation 15

1.2 Contributions 17

1.2.1 Publications 19

1.3 Overview 19

2 Background and preliminaries **21**

2.1 Overview of Related Literature 21

2.1.1 Inference in hybrid probabilistic models 21

2.1.2 Data association and outlier rejection 24

2.1.3 Map representation and semantic SLAM 26

2.1.4 Performance guarantees and certifiable machine perception . . 28

2.1.5 Network design and information summarization 30

2.2 Notation and mathematical preliminaries 33

2.3 Factor graphs and probabilistic models 35

2.4 Probabilistic inference as nonlinear optimization 39

2.5 Common examples of SLAM problems 40

2.5.1 Landmark-based SLAM 40

2.5.2 Pose-graph SLAM 41

3 Discrete-continuous smoothing and mapping **43**

3.1 Problem formulation 45

3.2 Overview of the approach 47

3.2.1 Alternating minimization 48
3.2.2 Online, incremental inference 51
3.2.3 Recovering marginals 52
3.3 Example applications 53
3.3.1 Point-cloud registration 54
3.3.2 Robust pose graph optimization 56
3.4 Discussion 59
3.4.1 When is alternating minimization efficient? 59
3.4.2 When can we ensure accurate solutions? 59
3.5 Summary 60

4 Robust object-level semantic SLAM 63
4.1 Problem statement 65
4.1.1 Semantic SLAM with unknown data association 65
4.1.2 Problem Statement 66
4.2 Approach 68
4.2.1 From MAP inference to max-product elimination 69
4.2.2 From marginal MAP inference to sum-product elimination . . 71
4.2.3 Computing candidate association hypotheses 72
4.3 Experimental results 75
4.3.1 MIT RACECAR dataset 76
4.3.2 KITTI datasets 79
4.4 Summary 82

5 Performance guarantees for spectral initialization in rotation averaging and pose-graph SLAM 85
5.1 Problem formulation 87
5.2 Spectral methods for initialization 89
5.3 Main results 91
5.4 Experimental results 96
5.4.1 Evaluation on synthetic data 97

5.4.2 Evaluation on standard SLAM benchmark datasets 101
5.5 Summary 102

6 Spectral measurement sparsification for pose-graph SLAM 103
6.1 Problem formulation 105
6.2 Approach 106
6.2.1 Solving the relaxation 107
6.2.2 Post hoc suboptimality guarantees 111
6.3 Experimental results 113
6.4 Summary 118

7 Conclusion 119
7.1 Limitations and Future Work 120
7.1.1 Expressive models and robust inference 121
7.1.2 Self-supervised and unsupervised learning 122
7.1.3 Performance guarantees for robot perception 122
7.1.4 Efficient inference, compression, and hierarchy 123

A Proofs for Chapter 5 125
A.1 Structure of the data matrices 125
A.2 Analysis of the spectral relaxation 127
A.2.1 Recovering minimizers of Problem 6 as eigenvectors 127
A.2.2 Symmetric perturbations of symmetric matrices 128
A.3 Proof of the main results 130
A.3.1 An upper bound for the estimation error in Problem 6 130
A.3.2 An upper bound for the estimation error in Problem 5 131
A.3.3 An upper bound on $d_{\mathcal{S}}(R^{(0)}, R^*)$ 135
A.4 Relationship to the method of Moreira et al. [95] 137

B Proofs for Chapter 6 141
B.1 Subgradients of the Fiedler value 141
B.2 Solving the direction-finding subproblem 143

List of Figures

2-1 Examples of coupled semantic SLAM systems. 26
2-2 Illustration of landmark-based SLAM. 41
2-3 Illustration of pose-graph SLAM. 42

3-1 Examples of discrete-continuous factor graphs in robotics. 44
3-2 Overview of a single iteration of optimization with DC-SAM. 47
3-3 Example applications of DC-SAM. 54
3-4 Application of DC-SAM to robust pose graph optimization. 57

4-1 Conditional independence of data association variables in semantic SLAM. 68
4-2 MIT RACECAR platform. 77
4-3 MIT RACECAR dataset trajectory error results. 79
4-4 MIT RACECAR dataset semantic maps. 80
4-5 Qualitative object-level SLAM results for KITTI dataset 00. 83

5-1 Illustration comparing true, optimal, and initial rotation estimates. . 87
5-2 Qualitative results for initialization on benchmark pose-graph datasets. 96
5-3 Cube dataset overview. 98
5-4 Influence of dataset parameters on the performance bounds for the Cube experiments. 99

6-1 Qualitative results for pose-graph sparsification. 113
6-2 Quantitative results for pose-graph sparsification. 114

List of Tables

4.1 Quantitative results for semantic SLAM on the KITTI dataset 81

5.1 Quantitative pose graph initialization results for benchmark SLAM datasets. 102

6.1 Summary of the datasets used in our sparsification experiments. . . . 113

Chapter 1

Introduction

1.1 Motivation

Imagine a robot traveling deep under the ocean, or even on another planet. As it wanders, it may observe flora or fauna, or perhaps extraterrestrial geology. Most of the data it collects may not be scientifically interesting, but suppose after a few hours of deployment, it notices something surprising. Immediately, the robot communicates to scientists – maybe they are located topside on a boat, or thousands of miles away on Earth. Because of communication delays, it may be a few seconds or a few minutes before the data makes it to a scientific expert. By the time a scientist can send back a command to revisit the point where the observation was made, our robot could be far away! In order to return and sample the region (i.e. *reacquire* the target), the robot must store an internal map of its environment and maintain some notion of its own location within that map. Furthermore, we should expect that the quality of this map must be good enough that we can *guarantee* that using this map, our robot has a reasonable chance of making it back to the correct location.

Now, consider the same robot, but rather than operating for minutes or hours, it can operate for days, weeks, or months at a time without the need for direct intervention, instead only intermittently communicating interesting observations to a scientist elsewhere. Indeed, such *resident* systems represent a major component of the modern vision for the future of underwater autonomous vehicles. In this setting,

we ask: "what capabilities would this robot need in order to perform its primary task?" What does this robot's internal representation of the world "look like?" For example, how can we encode the types of semantically relevant concepts into our robot's representation that would be important to a scientific operator? Moreover, in this regime, it may no longer be straightforward to guarantee that our robot can return to locations flagged by an operator with high probability. Even in the situation where the phenomenon of interest is static (i.e. unchanging in time), it may take days for the robot to return to its previous location, and, without a good map, accumulating positional error in the process. The key to addressing this problem is understanding *which properties* of the measurements collected by a robot influence its positional error, and ensure that our robot takes actions that ensure high-quality localization.

We envision a scenario in which an underwater autonomous vehicle is tasked with exploring and persistently monitoring an *a priori* unknown environment. We assume the environment to be sparsely populated with a set of objects (natural or man-made), such that the semantic class or category of these objects (or some subset thereof) is known, and therefore that the system may be equipped with a detection model capable, at least in a noisy sense, of locating and classifying some of these objects. The system is tasked with keeping track of these semantic landmarks, and, due to bandwidth constraints, it is limited to performing onboard computations (in real-time) and, at best, transferring a compact summary of its observations to a topside vessel. The straightforward principal goal of this robot is to build an accurate map (and localize itself within this map) over an extended period of observation. To do so, it must reason online about errors in its detection model and misattribution of object measurements to previously mapped landmarks as well as *what to remember* and *what to forget* in order to ensure reasonably bounded time and memory during operation.

1.2 Contributions

This thesis addresses the problem of developing expressive *representations* and *algorithms* that enable robust, efficient long-term mapping and localization. From a representational standpoint, we propose new computational tools and algorithms for inference in *hybrid* probabilistic models; i.e. models where a subset of variables of interest are *continuous* and others are *discrete*, taking on values from a finite set. In turn, we develop new models, leveraging these techniques, for capturing both *uncertainty* in object class predictions of learned perception methods and *ambiguity* about measurement-object correspondence within a navigational framework. From an algorithmic standpoint, we consider two issues: First, we consider the issue of bounding the error of solutions to the specific problem of pose-graph SLAM. In so doing, we present the first *initialization technique* for pose-graph SLAM that provably achieves bounded error. Finally, we consider the issue of *graph sparsification*, presenting an algorithm based on *maximizing algebraic connectivity* capable of producing parsimonious graphs that retain the quality of SLAM estimates.

The key contributions of this thesis are four-fold:

1. We develop DC-SAM, a library permitting straightforward representation and local optimization of *hybrid factor graphs*. DC-SAM extends existing tools for nonlinear least-squares optimization in the setting of SLAM by allowing for optimization in hybrid, *discrete-continuous* models. This extension is crucial for representing and solving problems with both continuous and discrete states of interest, arising commonly in robot perception, controls, and planning applications. We develop a local optimization technique that leverages the conditional independence structure present in a hybrid factor graph model to perform efficient inference. We show experimentally in several examples motivated by robot perception applications that DC-SAM is *expressive* and *performant*, enabling representation of rich hybrid probabilistic models for data association and outlier rejection (taken as two particular relevant problem instances), and providing fast, accurate solutions in practice.

2. We develop a novel semantic SLAM approach, taking into consideration uncertainty and ambiguity. The hybrid factor graph representations supported by our work with DC-SAM allow us to develop a probabilistic model considering jointly the *uncertainty* in classifications provided by a learning-based perception model (in our case, an object detector), as well as the *ambiguity* in data association. Our model seamlessly couples semantic and geometric information in a coupled manner without additional specialized techniques. Inference is achieved, in real-time for many practical scenarios, by making use of the local optimization procedure in DC-SAM. We establish new connections between different representations for ambiguous measurement-landmark correspondence through the lens of variable elimination in factor graphs. We experimentally demonstrate the practical advantages of different representational choices for ambiguity in this setting, providing results on real data from the MIT RACECAR platform as well as through the use of benchmark stereo visual navigation data from the KITTI dataset.

3. We address the lack of performance guarantees for initialization techniques and solutions to the SLAM problem. We describe a spectral initialization technique which we show admits the *first formal performance guarantees* for a pose-graph SLAM initialization method. Our analysis links the performance of estimators for pose-graph SLAM to key spectral graph theoretic properties of pose graphs.

4. We develop MAC, an algorithm for graph sparsification based on *maximizing algebraic connectivity*. MAC is simple and computationally inexpensive, and admits formal post hoc performance guarantees on the quality of the solutions it provides. In experiments on benchmark pose-graph SLAM datasets, we show that our approach quickly produces high-quality sparsification results which retain the connectivity of the graph and, in turn, the quality of corresponding SLAM solutions, as compared to a baseline approach which does not consider graph connectivity.

1.2.1 Publications

The content of Chapter 3 originally appeared in: Kevin J Doherty, Ziqi Lu, Kurran Singh, and John J Leonard. Discrete-Continuous Smoothing and Mapping. *IEEE Robotics and Automation Letters*, October 2022 [47]. Chapter 4 is based on a refined mathematical treatment of the ideas and algorithms originally presented in: Kevin Doherty, David Baxter, Edward Schneeweiss, and John J. Leonard. Probabilistic data association via mixture models for robust semantic SLAM. *In IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2020 [46]. Chapter 5 was originally presented in: Kevin J Doherty, David M Rosen, and John J Leonard. Performance Guarantees for Spectral Initialization in Rotation Averaging and Pose-Graph SLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2022 [48]. Finally, Chapter 6 initially appeared in: Kevin J Doherty, David M Rosen, and John J Leonard. Spectral Measurement Sparsification for Pose-Graph SLAM. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022 [49].

1.3 Overview

The remainder of this thesis is organized as follows: Chapter 2 provides relevant preliminaries and background on two- and three-dimensional geometry, graph theory, probability and statistics (including maximum *a posteriori* estimation as it pertains to SLAM), and an overview of related work.

Chapter 3 introduces the DC-SAM library and optimization methods for inference in hybrid factor graphs. Building off of DC-SAM as a natural tool for synthesizing the output of learned object detection and classification models with geometric measurements for estimation, Chapter 4 develops our approach to robust object-level semantic SLAM which accounts for uncertainty in semantic predictions as well as *ambiguity* in measurement-landmark correspondences.

In Chapter 5, we focus attention on a restricted subset of SLAM problems, namely pose-graph SLAM, and consider the issue of bounding the estimation error of SLAM solutions. In particular, we ask whether it is possible to produce an *initial guess* in

some manner which achieves provably bounded estimation error and bounded deviation from the globally optimal estimate. We show that indeed it is possible to do so, and present an algorithm based on spectral decomposition that admits *formal performance guarantees* on solution quality. We also show that the *algebraic connectivity* or *Fiedler value* of the measurement graphs arising in pose-graph SLAM is a key parameter controlling estimation accuracy.

Chapter 6 develops the MAC algorithm for graph sparsification. We consider in particular the issue of long-term navigation, whereby controlling computational expense and memory requirements of a SLAM system entails *sparsifying* the measurement graph. Motivated by the insights from Chapter 5, MAC designs sparse pose-graphs by *maximizing algebraic connectivity*.

Chapter 2

Background and preliminaries

This chapter covers relevant background on the problems considered in this thesis. Section 2.1 covers discusses relevant prior work and the "research gap" this thesis aims to address. Section 2.2 gives a brief exposition of important notation and preliminaries. Section 2.3 provides more in depth background on *factor graphs* an important representational tool we use extensively in this thesis.

2.1 Overview of Related Literature

2.1.1 Inference in hybrid probabilistic models

The problem of inference in discrete-continuous (hybrid) graphical models arises in many domains and intersects a number of communities, even within the field of robotics. Our focus in this thesis will be on applications in robot perception, so we primarily discuss related works in these settings. The interested reader may refer to Dellaert [39] for a discussion of these models in broader robotics applications or Koller and Friedman [75, Ch. 14] for a discussion of computational hardness, inference techniques, and a detailed review of literature on the general problem of inference in hybrid models. Finally, while we discuss the *particular* optimization approach considered in Chapter 3 in relation to existing methods, it is important to note: the mere availability of a consistent framework in which these solutions could be implemented

(as we develop in Chapter 3) enables practitioners to compare different approaches without the need to develop the additional scaffolding usually required to adapt an existing method.

Multi-Hypothesis Methods. The class of approaches addressing hybrid estimation by enumerating and pruning solutions to the discrete states are referred to as *multi-hypothesis methods*. These methods appeared in classical detection and tracking problems [115] and early SLAM applications [36]. MH-iSAM2 [63] extends the capabilities of iSAM2 [70] to the case where measurements between continuous variables may have ambiguity, which can be represented by the introduction of discrete variables. MH-iSAM2 maintains a hypothesis tree, which can be constructed and updated in an incremental fashion, like iSAM2, making the solver efficient. The types of ambiguities they consider can all be represented as factors in a factor graph where the discrete variables are all conditionally independent. This limits application to scenarios where individual discrete variables can be *decoupled*. However, correlations between discrete variables may arise in problem settings as diverse as switching systems (Figure 3-1; see also [67]), outlier rejection,[1] and as we explore in Chapter 4, semantic SLAM. In order to retain computational efficiency, MH-iSAM2, like all multi-hypothesis methods, must prune hypotheses, which risks the deletion of hypotheses that would have later become high-probability modes. iMHS [67] takes a qualitatively similar approach to MH-iSAM2, but focuses on the problem of smoothing in dynamic hybrid models, exploiting the specific temporal structure of this problem setting. Their approach extends to the setting where correlations among discrete variables are present. Like MH-iSAM2, however, the efficiency of iMHS rests on the ability to prune incorrect modes.

Hybrid and Non-Gaussian Inference. Hybrid inference in graphical models has been considered previously in many settings (see [124] for a review). Prior solution methods focus on either *specific* models, such as conditional linear Gaussian models (e.g. [113]) or attempt to approximate more general models in a manner amenable to

[1] Though we do not explore the issue of outlier rejection problems with correlations, the interested reader may see Lajoie et al. [79] for a formulation in the setting of SLAM.

standard techniques (e.g. by discretizing continuous state spaces to form a discrete inference problem). Models encountered in robotics applications are typically high-dimensional (often with numbers of states in the thousands) and non-Gaussian [119], and solutions are often required quickly. This precludes *direct* application of these techniques to the problems we explore in Section 3.3.

Several approaches have been presented which consider non-Gaussian inference with application to robot perception; many of these methods can be viewed as adaptations of general hybrid inference techniques tailored toward the computational requirements and problem structure in specific robot perception problems. FastSLAM [94] is an approach to filtering in SLAM with non-Gaussian models based on particle filters. In particular, a set of particles representing the current state of a robot is retained, and each particle independently samples associations from a distribution over hypotheses. Multimodal iSAM (mm-iSAM) [53] and NF-iSAM [65] perform incremental non-Gaussian inference for continuous-valued variables using nonparametric belief propagation [132] and normalizing flows, respectively. In situations where discrete variables can be efficiently marginalized to produce a problem exclusively involving continuous states, they can approximate the posterior marginals over the remaining continuous variables.

In contrast, our work focuses on the task of MAP estimation from the perspective of local optimization. While we describe a mechanism for approximating marginals *given* an (approximate) MAP estimate, the uncertainties provided by non-Gaussian inference techniques can be substantially richer. However, considering this somewhat more restricted problem setting (and coarser marginal approximation) affords us considerable benefits in terms of computational expense. Prior works applying optimization techniques for MAP estimation in non-Gaussian models (e.g. [108, 120]) do so by marginalizing out discrete variables and using smooth local optimization techniques on the resulting *continuous-only* estimation problem. Consequently, they do not permit the explicit estimation of discrete states, as we consider in this thesis.

Existing Tools. Several existing solvers perform optimization with models that can be represented in terms of factor graphs. Ceres [4] and g2o [60] provide nonlinear

least-squares optimization tools suitable for robotics applications, but they are not suitable for inference in hybrid factor graphs, e.g. as in Figure 3-1. GTSAM [38] provides incremental nonlinear least-squares solvers, like iSAM2 [70], and tools for representing and solving discrete factor graphs; it is for these reasons that we choose to extend the capabilities of GTSAM to the setting of hybrid, discrete-continuous models. Finally, Caesar.jl [34] implements mm-iSAM [53], supporting approximate, incremental non-Gaussian inference over graphical models commonly encountered in SLAM, including discrete-continuous models in scenarios where discrete variables can be eliminated through marginalization to produce a problem exclusively involving continuous variables.

2.1.2 Data association and outlier rejection

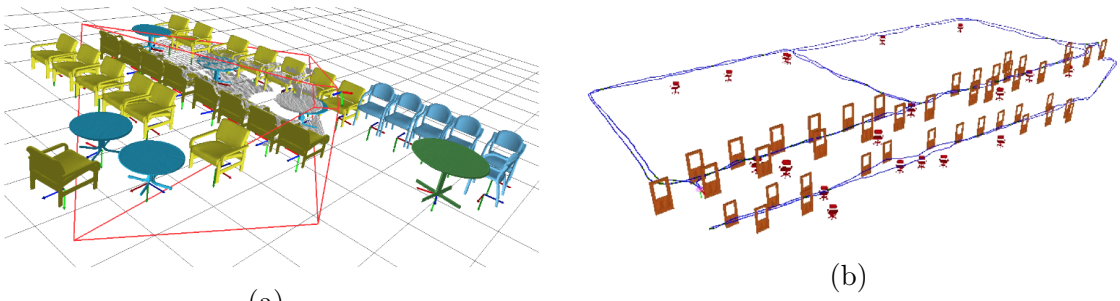
Chapters 3 and 4 deal with problems of *data association* and *outlier rejection*. Classical work on target tracking led to the introduction of approaches like probabilistic data association (PDA) [115] and multi-hypothesis tracking (MHT) [7] (as discussed in Section 2.1.1). These methods were, in turn, applied in the context of filtering-based state estimation [35, 36]. The joint compatibility heuristic for branch-and-bound search was later proposed to prune the large number of plausible hypotheses that arise in the context of multiple hypothesis tracking [100]. FastSLAM [93] introduced a particle filtering-based approach to the non-Gaussian inference problem of data association in which a data association sampler was introduced, serving as an alternative to explicit search over associations.

Many methods for data association and outlier rejection have been proposed in the context of smoothing-based graphical SLAM. The switchable constraints method [133] incorporated outlier rejection into the usual smoothing-based SLAM estimation process by introducing binary decision variables attached to each measurement. Max-mixtures [105] choose the minimum cost assignment to discrete hypothesis variables for a given estimate of the vehicle trajectory and map. The hybrid junction tree inference method of Segal et al. [127] iteratively updates discrete variables. Methods like pairwise-consistent measurement set maximization (PCM) [88] reformulate the

outlier rejection problem as one of find the largest pairwise-consistent set of measurements. Though we do not apply PCM in our setting, one possible enhancement to our approach would be to incorporate pairwise consistency constraints to rule out unlikely data associations. Yang et al. [145] propose a graduated nonconvexity (GNC) procedure for optimizing (typically nonconvex) outlier-robust cost functions in which they successively produce and optimize more well-behaved convex surrogate cost functions. In a series of papers, Pfeifer et al. [106–108] address a variety of robot perception applications by means of optimization over Gaussian mixture models. Their approach can be shown to be equivalent to the sum-product elimination approach we discuss, but rather than directly optimizing the marginal posterior, as in their work, we make use of expectation-maximization, which leads to a somewhat simpler implementation for optimizing the same objective.

The multi-hypothesis methods discussed in Section 2.1.1, like MH-iSAM2 [63], MH-JCBB [142], and iMHS [67] are the most recent incarnations of the multi-hypothesis tracking paradigm as applied to data association and outlier rejection for SLAM systems. MH-JCBB uses the joint compatibility heuristic [100] for robust data association, maintaining multiple "tracks" of data associations. All of these methods require online pruning of association hypotheses, which may discard correct combinations of hypotheses. Notably, the work presented in Chapters 3 and 4 allows us to exploit the specific conditional factorization of robot perception problems to compactly represent a large number of these hypotheses without the need to perform pruning.

Recent techniques consider robust estimation using convex relaxations [25, 27, 79, 144] to mitigate the effects of perceptual aliasing, often in the context of laser scan matching or appearance-based loop closure. Restricting consideration to certain measurement models (e.g. those encountered in pose-graph optimization or point cloud registration), these approaches allow computation of lower bounds on the objectives encountered in outlier robust estimation. Existing methods, however, are not adapted to the sorts of measurement models we consider here. Our approach allows for a much broader range of measurement models, while sacrificing convexity. An interesting area of future work would be to consider convex relaxations of the



(a) (b)

Figure 2-1: **Coupled semantic SLAM systems.** (a) SLAM++ [123] is one of the first coupled semantic SLAM systems and uses maximum-likelihood data association for semantic landmarks (with *a priori* known 3D object models). (b) The recent work of Bowman et al. [18] uses expectation-maximization to address the semantic SLAM problem with unknown data association. Figures adapted from [123] and [18], respectively.

formulation in our current presentation, even for a slightly restricted class of models.

Finally, random finite sets [98] are a related formalism useful for describing and performing inference involving the joint posterior distributions encountered when combining the SLAM and data association problems. In this work, we opt to formulate the data association and SLAM problems as a joint optimization procedure, seeking only a point estimate of robot and landmark states. However, random finite set approaches may present an interesting alternative in the setting where we aim to infer a *distribution* over the variables of interest rather than a point estimate (see, for example, [45] for prior work on the topic of semantic SLAM).

2.1.3 Map representation and semantic SLAM

Chapter 4 of this thesis deals with map representation; particularly *object-based, semantic map representations*, which may capture not just where environmental features or landmarks are, but also *what* they are. Broadly, we are concerned with navigation methods that incorporate *learning-augmented perception methods*, such as, but not exclusively, object detectors. Most commonly, semantic map representations are constructed in a *decoupled* fashion, i.e. assuming the availability of accurate poses from a SLAM or odometry system, e.g. [91, 121], and synthesizing semantic information

(be it from an object detector, or other means) in a *post hoc* manner. However, classifications or descriptors obtained from object detectors have recently been used to aid data association in the context of semantic SLAM, where inference of (typically discrete) semantic labels is *coupled* with data association. Most work on this *coupled* form of the semantic SLAM problem considers maximum-likelihood data association [92, 101, 102, 123, 146] (see Figure 2-1a), which is effective for small spatial scales or short-term operations where uncertainty growth can be more easily mitigated, a few approaches consider robust alternatives. Robust handling of measurements acquired via learned perception techniques is especially important in the long-term regime where these methods are likely to encounter data inconsistent with that encountered at training time (and we should not expect perfect precision and recall in the regime as operation time "goes to infinity").

A few recent approaches to semantic SLAM consider more robust alternatives. The nonparametric Bayesian approach of Mu et al. [97] alternates between sampling data associations and recomputing SLAM solutions. By retaining the association variables during optimization, this method offers improved robustness compared to maximum-likelihood data association, but the requirement to store and recompute discrete association variables is computationally demanding. Bowman et al. [18] describe an expectation-maximization procedure for optimizing the marginal posterior. In their setting, association variables are marginalized out and optimization is performed on the resulting factor graph by alternating between computing data association weights and improved assignments to robot and landmark states. In our prior work [45], we considered a treatment of the semantic SLAM problem where we marginalize out the data association variables and aim to approximate the non-Gaussian posterior using nonparametric belief propagation [53, 132], and later, using nonlinear least-squares for optimization [46]. In this work, we reveal that in fact prior work making use of marginalization of data association variables (i.e. [18, 45, 46]) can all be understood in terms of elimination operations on a factor graph. Moreover, this allows us to construct an elimination approach in such a way that the SLAM problem defined over the eliminated graph is *equivalent* to the original problem. Finally,

with these tools in hand, we show that we can represent the resulting association-ambiguous measurements in terms of *smart factors* [28]. In consequence, one can implement these models simply as factors within a hybrid factor graph optimization framework (e.g. DC-SAM [47] as described in Chapter 3), thereby allowing practitioners to make use of these methods without the need to develop custom optimization routines.

2.1.4 Performance guarantees and certifiable machine perception

In the interest of better understanding the performance of estimators used for robot perception, in Chapter 5, we discuss the problem of initialization for pose-graph SLAM and establish several performance guarantees for estimators. SLAM (as well as *multiple rotation averaging*, where the variables we aim to estimate are restricted to rotations) are often formulated as high-dimensional, nonconvex optimization problems, which are solved using local search techniques. Consequently, solving these problems requires efficient algorithms for producing an "initial guess." Historically, research on this topic has focused on developing cheap, but typically inexact, convex or linear relaxations of the SLAM (resp. rotation averaging) problems [30, 90]. While these techniques often work well in practice, the fact that they are obtained as heuristic approximations makes it difficult to ascertain *what specific features* of SLAM or RA problems determine their performance. Consequently, it is difficult to assess *under what conditions* these techniques can be reliably deployed.

A related line of research is the development of Cramér-Rao bounds for the pose-graph SLAM and rotation averaging problems [17, 33, 72]; these works provide *lower bounds* on the achievable estimation error *in expectation*. In this work, we derive a complementary set of *upper bounds* on the estimation error on a *per instance* basis. Interestingly, our estimation error *upper* bounds depend upon precisely the same spectral quantities as do the Cramér-Rao (*lower*) bounds, indicating that graph spectra are objects of central importance in understanding the statistical properties of SLAM

and RA estimators.

The spectral relaxation approach to initialization that we consider has previously appeared in other problem settings, particularly in the area of phase synchronization problems (cf. [15, 16, 80, 128]). In particular, Ling [80] describe error bounds that are qualitatively similar to those described in Chapter 5, though theirs are concerned specifically with *orthogonal* group synchronization problems. Liu et al. [81] take a similar approach to ours in order to derive error bounds for spectral estimators of synchronization problems defined over subgroups of the orthogonal group (including $\mathrm{SO}(d)$), but employ a different definition of the perturbation than the one we consider here. As we will show, our notion of perturbation has the advantage that it follows naturally from a generative model of SLAM and RA, and furthermore, directly reveals the spectral properties of the measurement network (specifically, a kind of generalized algebraic connectivity) as the key quantities controlling the worst-case performance of our spectral initialization method.

Recently, Moreira et al. [95] proposed a computationally-efficient Krylov-Schur decomposition approach for pose-graph SLAM. We show in Appendix A.4 that their method is formally equivalent to a *special case* of the one we present in Chapter 5 (namely, an *unweighted*, *rotation-only* variant of our spectral initialization procedure). However, our construction arises more naturally from spectral relaxation, and additionally allows for the incorporation of translational measurements, which we show in Section 5.4 can have a significant impact on estimation quality. Arrigoni et al. [5] also describe a spectral method for $\mathrm{SE}(d)$-synchronization. While an analysis similar to ours could likewise be carried out for their method, the form of the relaxation they consider would lead to more complicated bounds due to a dependence on the scale of the translational states. Finally, Boots and Gordon [14] consider spectral techniques for the range-only SLAM problem. Though their problem setting differs from the one considered here, extension of the techniques presented in this work to scenarios with different types of measurement models is an interesting area for future work.

Finally, *certifiably-correct* machine perception has emerged as a key area of interest to the robotics community, resulting in the development of algorithms capable of

directly computing *globally optimal* solutions of certain nonconvex estimation problems under moderate noise [20, 26, 29, 31, 41, 51, 118, 136]. Our analysis provides new bounds on the estimation error of the maximum likelihood estimators recovered by these techniques in terms of the magnitude of the measurement noise. Moreover, the bounds we present suggest that when these estimators, which are often based on large-scale semidefinite relaxations, do attain globally optimal solutions, the resulting estimates have error bounds that *match* (up to small constant factors) the error bounds we derive for our spectral initialization, which is easily implemented and computationally inexpensive.

2.1.5 Network design and information summarization

The work in Chapter 5 as well as prior work on performance guarantees for robot perception (see e.g. [119] for a review) identify spectral properties of measurement graphs as key properties controlling estimation error. In Chapter 6 we will describe approaches for leveraging these insights to develop algorithms for long-term SLAM. In particular, these properties motivate an answer to the question *what information should we keep, and what can safely be forgotten?* in the setting of lifelong navigation. Prior work on this topic can be divided into information summarization and sparsification, where we are interested in removal or compression of redundant information online or after measurements are acquired, and network design, where the principal objective is to decide where to add edges to a measurement graph or, more concretely, where a robot should go next to gather information. While the latter topic (active SLAM) is not a central focus of this thesis, many of the performance criteria for making these decisions are identical to (or substantially overlap) those encountered in the setting of information summarization or sparsification.

Importance of the algebraic connectivity in SLAM The importance of algebraic connectivity in general has been observed since at least 1973, with the seminal work of Fiedler [52]. In robot perception, the algebraic connectivity has appeared in the context of rotation averaging [17], linear SLAM problems and sensor network

localization [72, 73], and pose-graph SLAM [33] as a key quantity controlling estimation performance. In particular, Boumal et al. [17] observed that the inverse of the algebraic connectivity bounds (up to constants) the Cramér-Rao lower bound on the expected mean squared error for rotation averaging. In Chapter 5, we show that it appears as a key quantity controlling the *worst-case* error of estimators applied to measurement graphs in pose-graph SLAM and rotation averaging (where *larger* algebraic connectivity is associated with (statistically) lower error).

Maximizing the algebraic connectivity The problem of maximizing the algebraic connectivity subject to cardinality constraints has been considered previously for a number of related applications. Ghosh and Boyd [58] consider a semidefinite program relaxation of the same objective we consider. Alternatively, Nagarajan [99] considered a mixed-integer approach to optimize this objective. While our overall approach can make use of any solution to the relaxation we consider, neither of these methods scales to the types of problems we are considering. To the best of our knowledge, this is the first time an approach has been proposed for pose graph sparsification which makes use of *any* approach to solving a convex (or concave) relaxation of the algebraic connectivity maximization problem.

Network design and pose graph sparsification The theory of optimal experimental design (TOED) [111] gives several optimality criteria applicable to network design. Specifically, A-optimality, T-optimality, E-optimality, and D-optimality are common criteria, each of which corresponds to optimizing a different property of the information matrix describing the distribution of interest (in SLAM, this is typically the joint distribution over robot and landmark states). Briefly, A-optimal designs minimize the trace of the inverse of the information matrix, D-optimal designs maximize the determinant of the information matrix, E-optimal designs maximize the smallest eigenvalue of the information matrix, and T-optimal designs maximize the trace of the information matrix. Chen et al. [33] discuss the connections between Cramér-Rao bounds for pose-graph SLAM and the A-optimality and T-optimality

criteria. Historically, Cramér-Rao bounds and the optimal design metrics arising from them have been popular tools for network design and active SLAM [33, 72], including, e.g. in application to the planning of underwater inspection routes [74].

Khosoussi et al. [73] established many of the first results for optimal graph sparsification (i.e. measurement subset selection) in the setting of SLAM. The convex relaxation they consider is perhaps the closest existing work in the literature to ours. However, in contrast to the approach we present in Chapter 6, they consider the D-optimality criterion, while the results in Chapter 5 as well as previous work on Cramér-Rao bounds [17, 33] strongly suggest that the quantity of interest with regard to estimation performance is the algebraic connectivity, and therefore the E-optimality criterion.[2] More practically, the E-optimality criterion is both less computationally expensive to compute *and* to optimize.

Several methods have been proposed to reduce the *number of states* which need to be estimated in a SLAM problem (e.g. [22, 23, 64, 68]), typically by marginalizing out state variables. This procedure is usually followed by an edge pruning operation to mitigate the unwanted increase in graph density. Previously considered approaches rely on linearization of measurement models at a particular state estimate in order to compute approximate marginals and perform subsequent pruning. Consequently, little can be said concretely about the quality of the statistical estimates obtained from the sparsified graph compared to the original graph. In contrast, our approach does not require linearization, and provides explicit performance guarantees on the graph algebraic connectivity as compared to the globally optimal algebraic connectivity (which is itself linked to both the *best* and *worst* case performance of estimators applied to the SLAM problem).

[2]Of course, by maximizing the smallest eigenvalue, the E-optimality criterion also selects for information matrices with larger determinant.

2.2 Notation and mathematical preliminaries

Lie groups and matrix manifolds: We will make use of the matrix realizations of several Lie groups, most prominently the d-dimensional special Euclidean and special orthogonal groups, denoted $\mathrm{SE}(d)$ and $\mathrm{SO}(d)$, respectively. $\mathrm{SE}(d)$ can be realized as a matrix group according to:

$$\mathrm{SE}(d) \triangleq \left\{ \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{(d+1)\times(d+1)} \mid R \in \mathrm{SO}(d),\ t \in \mathbb{R}^d \right\}, \tag{2.1}$$

and the group $\mathrm{SO}(d)$ can be realized as:

$$\mathrm{SO}(d) \triangleq \left\{ R \in \mathbb{R}^{d\times d} \mid R^T R = I_d,\ \det(R) = 1 \right\}, \tag{2.2}$$

where I_d is the $(d \times d)$ identity matrix. The *Stiefel manifold* $\mathrm{St}(k, n)$ is the set of orthonormal k-frames in $\mathbb{R}^n$ $(k \leq n)$:

$$\mathrm{St}(k, n) \triangleq \left\{ V \in \mathbb{R}^{n\times k} \mid V^{\mathsf{T}} V = I_k \right\}. \tag{2.3}$$

Linear algebra: For a symmetric matrix S, $S \succeq 0$ denotes that S is positive-semidefinite. The eigenvalues of a symmetric matrix $S \in \mathbb{R}^{n\times n}$ are denoted $\lambda_1(S) \leq \lambda_2(S) \leq \ldots \leq \lambda_n(S)$. We will also consider several block-structured matrices, and make use of a few special operators acting on them. Following the notation of Rosen et al. [118], given square matrices $A_i \in \mathbb{R}^{d\times d}, i = 1, \ldots, n$, we let $\mathrm{Diag}(A_1, \ldots . A_n)$ denote the matrix direct sum (i.e., the block-diagonal matrix having $A_1, \ldots, A_n$ as its diagonal blocks). Furthermore, given a block-structured matrix B, let $\mathrm{BlockDiag}_d(B)$ denote the operator extracting a $d \times d$ block-diagonal matrix from B. Finally, let $\mathrm{SBD}(d, n)$ denote the set of $dn \times dn$ symmetric block-diagonal matrices with diagonal blocks of size $d \times d$, and $\mathrm{SymBlockDiag}_d(A)$ be the operator extracting the symmetrization of the $d \times d$ block-diagonal part of A.

Probability and statistics: We denote the multivariate Gaussian distribution with mean $\mu \in \mathbb{R}^d$ and covariance $\Sigma \in \mathbb{S}^d_+$ as $\mathcal{N}(\mu, \Sigma)$. We denote the isotropic Langevin distribution on $\mathrm{SO}(d)$ with mode $M \in \mathrm{SO}(d)$ and concentration parameter $\kappa \geq 0$ as Langevin(M, κ); this is the distribution whose probability density function is:

$$p(R; M, \kappa) = \frac{1}{c_d(\kappa)} \exp\left(\kappa \operatorname{tr}\left(M^T R\right)\right), \tag{2.4}$$

with respect to the Haar measure on $\mathrm{SO}(d)$, with $c_d(\kappa)$ a normalization constant.

For an unknown variable Z we aim to infer, we denote its true (latent) value by $\underline{Z}$ and a noisy measurement of Z by $\tilde{Z}$. We use the notation $p(X \mid Z)$ to denote the conditional distribution on a variable, X, given another, Z. When conditioning on a *particular* (fixed) assignment to Z, e.g. $Z = \tilde{Z}$, we will often write $p(X \mid \tilde{Z})$ when we mean $p(X \mid Z = \tilde{Z})$.

Gauge-invariant distance metrics: A key property of many of the geometric estimation problems we consider (particularly in Chapters 5 and 6) is that they admit infinitely many solutions due to *gauge symmetry*. We therefore define the following *orbit distances* in order to compare solutions to the rotation estimation problems encountered in Chapters 5 and 6 in a symmetry-aware manner:

$$d_{\mathcal{S}}(X, Y) \triangleq \min_{G \in \mathrm{SO}(d)} \|X - GY\|_F, \quad X, Y \in \mathrm{SO}(d)^n \tag{2.5a}$$

$$d_{\mathcal{O}}(X, Y) \triangleq \min_{G \in \mathrm{O}(d)} \|X - GY\|_F, \quad X, Y \in \mathrm{O}(d)^n. \tag{2.5b}$$

It will be convenient to "overload" the $\mathrm{O}(d)$ orbit distance to act on elements of the set $\mathcal{Y} \triangleq \left\{Y \in \mathbb{R}^{d \times dn} \mid YY^{\mathsf{T}} = nI_d\right\}$.[3] That is, for $X, Y \in \mathcal{Y}$:

$$d_{\mathcal{O}}(X, Y) \triangleq \min_{G \in \mathrm{O}(d)} \|X - GY\|_F. \tag{2.6}$$

[3]The elements of $\mathcal{Y}$ admit a straightforward interpretation as transposed and re-scaled elements of the Stiefel manifold $\mathrm{St}(d, dn)$ (see (2.3)).

Each of these distances can be computed in closed form by means of a singular value decomposition (see Rosen et al. [118, Theorem 5]).

2.3 Factor graphs and probabilistic models

A *factor graph* $\mathcal{G} \triangleq \{\mathcal{V}, \mathcal{F}, \mathcal{E}\}$ with factor nodes $f_k \in \mathcal{F}$, variable nodes $v_i \in \mathcal{V}$, and edges $\mathcal{E}$ is a graphical representation of a product factorization of a function:

$$\begin{aligned} f(\mathcal{V}) &= \prod_k f_k(\mathcal{V}_k) \\ \mathcal{V}_k &\triangleq \{v \in \mathcal{V} \mid (f_k, v) \in \mathcal{E}\}. \end{aligned} \tag{2.7}$$

From a representational standpoint, factor graphs are tremendously general. In particular, it's straightforward to verify that *any* function f can be decomposed in the form of (2.7); simply consider the trivial factor graph containing a single factor node adjacent to *all* variable nodes in $\mathcal{V}$. Second, it's clear that there can be many factor graphs $\mathcal{G}$ representing the same function. The key benefit of factor graphs as a modeling tool is only realized when we (as practitioners) are careful about the *particular* factor graph we use to model our function. We benefit from factor graphs when we specifically intend to explicate conditional independence relationships that we *know* exist in our model. Furthermore, there is an important "compositional" property of factor graphs that will be relevant in our applications to SLAM: as we will see, we will often construct the function $f(\mathcal{V})$ dynamically *online* by adding new factors f_k to $\mathcal{G}$ which depend only on a local subset of $\mathcal{V}$ (e.g. as a robot navigates through an environment). In this setting, the conditional independence relationships for the full model $f(\mathcal{V})$ need not be known *a priori*, knowledge of only the scope of individual factors (the subset of variables a factor relates) is sufficient to explicate conditional independence relations for the entire model. This is a property we will use extensively throughout this thesis, and instances of its practical application will appear specifically in Chapters 3, 4, and 5.

In this thesis (and broadly in SLAM applications), we are primarily interested in

factor graphs as they appear in probabilistic models. In particular, we will often be interested in determining the most probable value of some unobserved states Θ given a set of measurements $\tilde{Z} \sim p(Z \mid \underline{\Theta})$ which are presumed to be sampled from a (known) generative model $p(Z \mid \Theta)$ conditioned on the true (unknown) values of the states $\underline{\Theta}$. Both the measurements $\tilde{Z}$ and the unknown states Θ may be *continuous* (i.e. they may take on any value from an uncountable set) or *discrete* (taking on values from a countable set of outcomes). In practical robot perception applications, continuous states and measurements are usually elements of some (often, but not necessarily connected) subset of Euclidean space, e.g. rotations represented as elements of the special orthogonal groups SO(2) or SO(3), translations in $\mathbb{R}^2$ or $\mathbb{R}^3$, or rigid body transformations represented as elements of SE(2) or SE(3). Discrete sets encountered in application are essentially always *enumerable* (meaning they can be put in one-to-one correspondence with a finite subset of natural numbers). Commonly encountered examples include: binary variables in the set $\{0, 1\}$, used to indicate whether to "keep" a particular measurement or discard it as an outlier, or to indicate contact with the ground; association variables which may indicate multiple hypotheses about which environmental landmark was observed in a given measurement; and categorical variables, e.g. in the set $\{\text{cat}, \text{dog}, \text{house}, \text{chair}\}$ indicating object classes.

Given this context, the most fundamental problems encountered in SLAM are maximum *a posteriori* (MAP)inference and computing marginals. Specifically, the problem of MAP inference is as follows:

Problem 1 (MAP inference)**.** Given a set of measurements $\tilde{Z}$, a model $p(Z \mid \Theta)$ relating measurements to unobserved states Θ, and a prior distribution $p(\Theta)$ over unobserved states (which may be uniform), the problem of MAP inference is to determine an argument Θ^* maximizing the posterior probability $p(\Theta \mid \tilde{Z})$ over unknown

states given the realized measurements $\tilde{Z}$, i.e.:

$$\begin{aligned}\Theta^* &= \operatorname*{argmax}_{\Theta} p(\Theta \mid \tilde{Z}) \\ &= \operatorname*{argmax}_{\Theta} p(\Theta, Z = \tilde{Z}) \\ &= \operatorname*{argmax}_{\Theta} p(\tilde{Z} \mid \Theta) p(\Theta),\end{aligned} \tag{2.8}$$

where in the second line we have used the fact that $p(\Theta \mid \tilde{Z}) = p(\Theta, \tilde{Z})/p(\tilde{Z})$ and that the measurement values are constant with respect to the maximization. Note that there may not be a unique maximizing argument in equation 2.8; in such cases (in this thesis) we will be content with *any* maximizer.

The problem of marginal computation[4] can be described as follows:

Problem 2 (Marginal computation)**.** Given a set of measurements $\tilde{Z}$, a model $p(Z \mid \Theta)$ relating measurements to unobserved states Θ, and a prior distribution $p(\Theta)$ over unobserved states (which may be uniform), the problem of marginal computation for a subset of states $\Theta_i \subset \Theta$ is to compute the distribution:

$$\begin{aligned}p(\Theta_i \mid \tilde{Z}) &= \int_{\Theta \setminus \Theta_i} p(\Theta \mid \tilde{Z}) \\ &= \frac{1}{p(\tilde{Z})} \int_{\Theta \setminus \Theta_i} p(\tilde{Z} \mid \Theta) p(\Theta).\end{aligned} \tag{2.9}$$

In applications, integrals of the form encountered in equation (2.9) for typical forms of the model $p(Z \mid \Theta)$ will often be computationally intractable and we will settle for some approximation of the marginals of interest (for example the *Laplace approximation* [13, Sec. 4.4] as we will encounter in Chapters 3 and 4).

In many robot perception tasks (including SLAM), it is common that an individual measurement $\tilde{z}_k \in \tilde{Z}$ depends only on a small subset $\Theta_k \subset \Theta$ of states to be estimated. For example, a global positioning system (GPS) measurement of a robot's position

[4]Note that the marginals we consider here arise from integration or summation over the values of the marginalized variables. We will also discuss *max*-marginals in this thesis; these are functions obtained by replacing the integral in eq. (2.9) with a max over the same values. These operations (summation and maximization) correspond to sum-product and max-product variable elimination procedures for probabilistic graphical models (cf. [75] for a general reference).

depends on the location of the robot at the time the measurement was taken, but not where it was before or after the measurement. Similarly, the relative position of a landmark as measured by a moving observer depends only on the pose of the observer and the location of the landmark. These properties are reflected in the *conditional independence relationships* that appear in our model. In particular, we will assume that a measurement $\tilde{z}_k$ is independent of all other measurements $\tilde{Z} \setminus \{\tilde{z}_k\}$ and states $\Theta \setminus \Theta_k$ given the specific unknown states $\Theta_k \subset \Theta$ it relates. Formally, we say:

$$p(\tilde{z}_k \mid \Theta, \tilde{Z} \setminus \{\tilde{z}_k\}) = p(\tilde{z}_k \mid \Theta_k). \tag{2.10}$$

The straightforward consequence of this conditional independence structure is that the *joint likelihood* $p(\tilde{Z} \mid \Theta)$ admits a factorization as:

$$\begin{aligned} p(\tilde{Z} \mid \Theta) &= \prod_k p(\tilde{z}_k \mid \Theta_k) \\ \Theta_k &\subset \Theta. \end{aligned} \tag{2.11}$$

In turn, we recognize that the joint distribution $p(\Theta, Z = \tilde{Z})$ admits a very natural factor graph representation $\mathcal{G} = \{\mathcal{V}, \mathcal{F}, \mathcal{E}\}$ (of the form in eq. (2.7)) which explicates the same conditional independence relations appearing in eq. (2.11), i.e.

$$\begin{aligned} p(\Theta, \tilde{Z}) &= \prod_k f_k(\Theta_k) \\ \Theta_k &= \{\theta \in \Theta \mid (f_k, \theta) \in \mathcal{E}\}, \end{aligned} \tag{2.12}$$

where each factor f_k is in correspondence with a measurement likelihood term of the form $p(\tilde{z}_k \mid \Theta_k)$ or a prior term of the form $p(\Theta_k)$.

2.4 Probabilistic inference as nonlinear optimization

Given the decomposition of the joint in eq. (2.12), we may rewrite the MAP inference problem (Problem 1) in the following form:

$$\begin{aligned}\Theta^* &= \operatorname*{argmax}_{\Theta} p(\Theta \mid \tilde{Z}) \\ &= \operatorname*{argmax}_{\Theta} p(\Theta, \tilde{Z}) \\ &= \operatorname*{argmax}_{\Theta} \prod_k f_k(\Theta_k) \\ &= \operatorname*{argmin}_{\Theta} \sum_k -\log f_k(\Theta_k).\end{aligned} \tag{2.13}$$

That is, we may obtain a MAP estimate Θ^* by minimizing the negative logarithm of the joint (evaluated at $Z = \tilde{Z}$).

Everything we have done so far has been perfectly general. It will often be convenient, however, to restrict consideration to models where the optimization in eq. (2.13) is equivalent to a *nonlinear least-squares problem*. Formally, we say:

$$\begin{aligned}\Theta^* &= \operatorname*{argmax}_{\Theta} p(\Theta \mid \tilde{Z}) \\ &= \operatorname*{argmin}_{\Theta} \sum_k \|r_k(\Theta_k)\|_2^2,\end{aligned} \tag{2.14}$$

for some (typically nonlinear) function $r_k : \Omega \to \mathbb{V}$ mapping a subset of unknown states Θ_k to vectors in a subset $\mathbb{V}$ of $\mathbb{R}^n$. It turns out that the conditions under which the equivalence in eq. (2.14) hold are quite general (cf. Rosen et al. [120, Theorem 1]).[5]. If we further assume that $f_k \in C^1(\Omega)$ (by which we mean f_k is *continuously differentiable* on Ω, the space of values taken on by Θ_k), then a reasonable approach for approximate inference would be to apply gradient-based numerical optimization starting from an initial assignment to Θ. This is the approach taken by essentially every state-of-the-art technique for MAP inference in graphical models that has been applied to SLAM (cf. [21, 39, 119] for relevant reviews, and [60, 70] for specific

[5]In particular, it suffices to assume that factors f_k are positive and bounded above.

techniques).

2.5 Common examples of SLAM problems

The mathematical form of SLAM problems is essentially indistinguishable from the more general Bayesian inference problems of MAP inference (Problem 1) and marginal computation (Problem 2). There are three distinguishing considerations of SLAM specifically that are not present in the more general forms of these problems: (1) the particular (typically geometric) nature of the measurement models and states to be estimated, (2) the need to construct a model incrementally and online, and (3) the practical necessity of solving these problems (estimating unknown states) quickly (i.e. in real time), even if this means resorting to approximation. With respect to the first consideration, the following subsections provide exposition for the two most common formulations of SLAM problems: *landmark-based SLAM* in which a robot aims build a map of surrounding landmarks while localizing itself within that map, and *pose-graph SLAM* in which all measurements are relative transforms (e.g. elements of SE(3) for three-dimensional problems) between robot poses (also elements of SE(3)).

2.5.1 Landmark-based SLAM

In (three-dimensional) landmark-based SLAM (illustrated in Figure 2-2), we consider the problem of inferring jointly a *map* of environmental landmarks $L \triangleq \{\ell_1, \ldots, \ell_m\}$, $\ell_j \in \mathbb{R}^3$ and the trajectory of a robot $X \triangleq \{x_1, \ldots, x_n\}, x_i \in \mathrm{SE}(3)$ given a subset of measurements of their pairwise relationships (including, e.g. the relative position of a landmark ℓ_j in the frame of pose x_i, the relative SE(3) transform from one pose to another, among others), denoted $\tilde{Z}$.

A critical component of landmark-based SLAM is the determination of which (if any) mapped landmark was measured during the observation $\tilde{z}_k$. This is referred to as a *correspondence problem* or the *data association problem*, and it will be a key focus of study in Chapter 4.

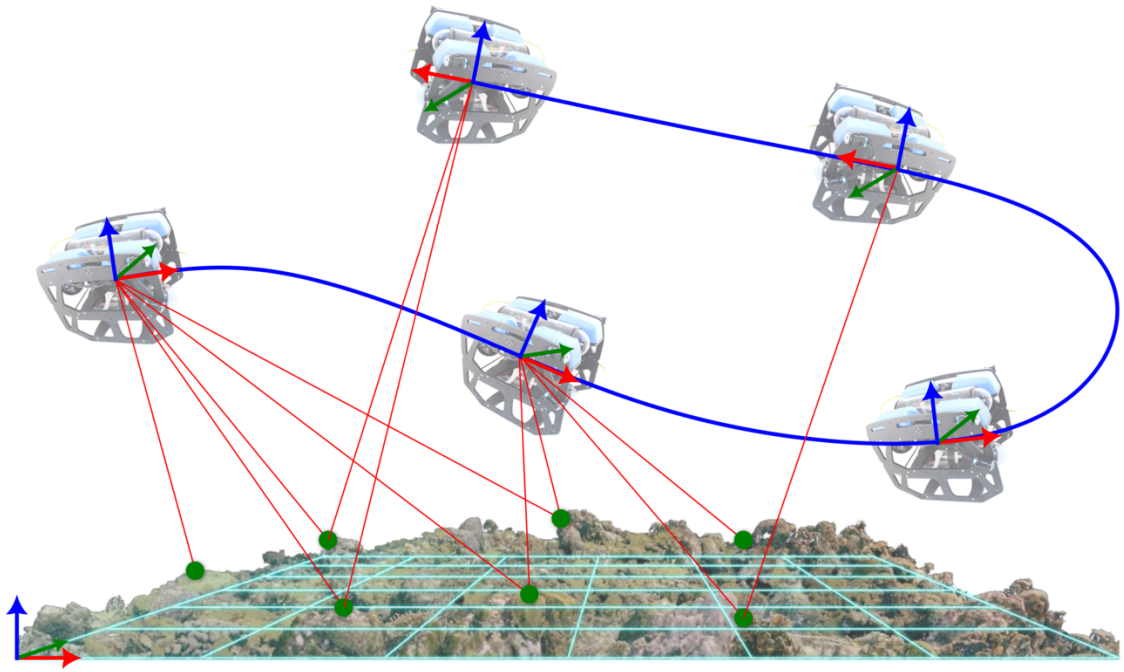


Figure 2-2: **Illustration of landmark-based SLAM.** The example above illustrates the problem of landmark-based SLAM where a robot simultaneously builds a map of environmental landmarks (here represented as points on a coral reef) and uses repeated observations of those landmarks to localize itself within a globally-consistent coordinate frame (here represented by the axes on the bottom-left of the figure).

2.5.2 Pose-graph SLAM

In contrast to landmark-based SLAM, in pose-graph SLAM we only aim to estimate the trajectory of a robot $X \triangleq \{x_1, \ldots, x_n\}$, $x_i \in \mathrm{SE}(3)$ given (potentially noisy) measurements $\tilde{x}_{ij} \in \mathrm{SE}(3)$ of a subset of their true pairwise relative transforms $\underline{x}_i^{-1}\underline{x}_j$. In pose-graph SLAM, a place recognition system, for determining when an observation is sufficiently "similar" to a previous one to suggest that our robot has returned to a previous location, replaces explicit long-term data association for the purposes loop closure (see Lowry et al. [82] for a review). Specific instances of pose-graph SLAM problems will appear in Chapters 3, 5, and 6.

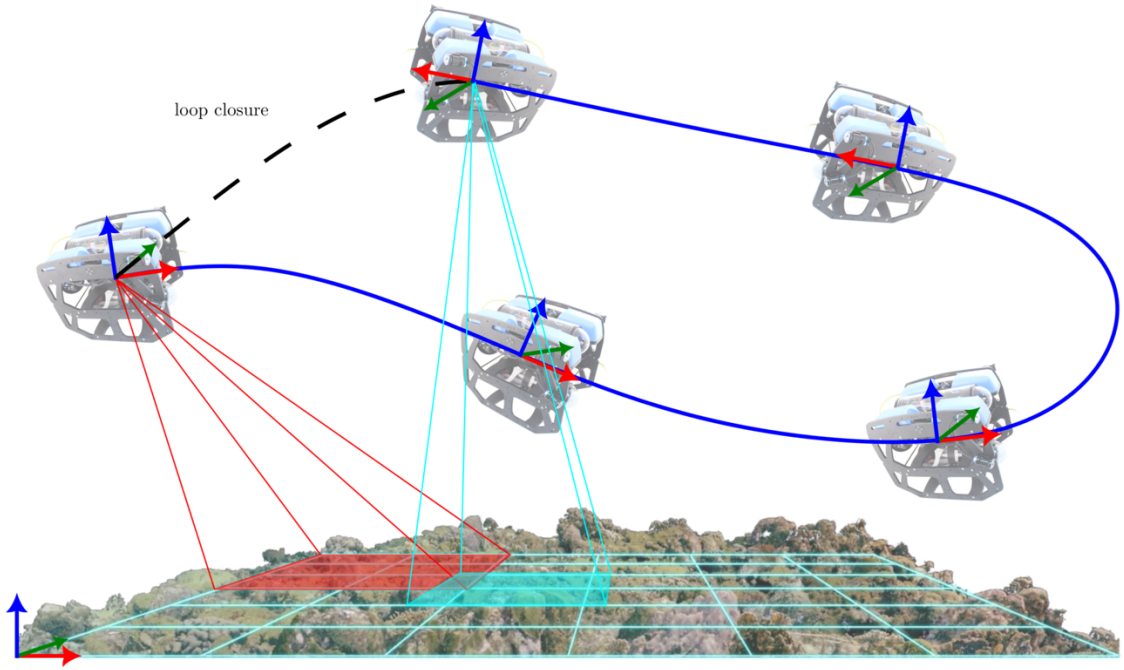


Figure 2-3: **Illustration of pose-graph SLAM.** The example above illustrates the problem of pose-graph SLAM. Here, a robot does not explicitly maintain a map of landmarks, but visually similar observations are used to determine loop closures in order to localize the robot within a globally-consistent coordinate frame (here represented by the axes on the bottom-left of the figure).

Chapter 3

Discrete-continuous smoothing and mapping

The previous sections considered the motivation for this thesis and discussed relevant preliminaries for understanding state-of-the-art techniques for inference in graphical models (and their application to localization and mapping problems). This chapter presents the first major contribution of this thesis.

Specifically, the probabilistic modeling approach described in the previous chapter has become the dominant representational paradigm in robot perception applications, appearing in a wide range of important estimation problems. This formalism has led to the development of numerous algorithms and software libraries, such as GTSAM [38], which provide flexible and modular languages for specifying and solving optimization problems defined by these models (typically in terms of factor graphs). Among the models relevant to robotics applications, *discrete-continuous graphical models* capture a great breadth of key problems arising in robot perception, task and motion planning [54, Sec 3.2], and navigation, including data association, outlier rejection, and semantic simultaneous localization and mapping (SLAM) [119] (see Figure 3-1). Despite the importance of these models, while *ad hoc* solutions have been proposed for *particular* problem instances, at present there is no off-the-shelf approach for hybrid problems that is either as general or as easy-to-use as similar methods for their continuous-only or discrete-only counterparts. Notably, the state-of-the-art gradient-

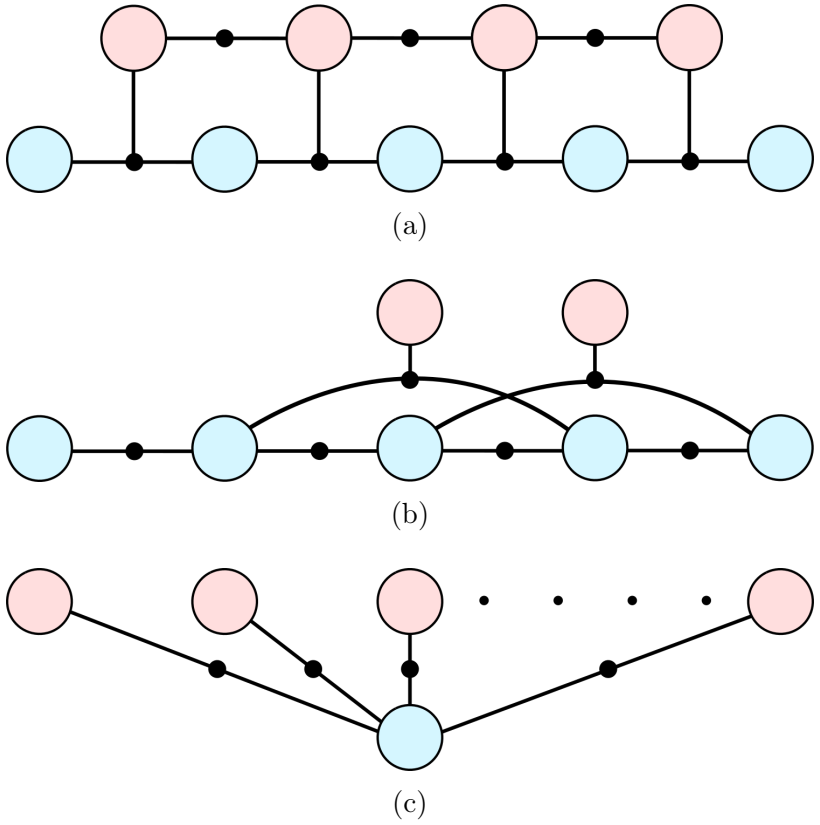


Figure 3-1: **Discrete-continuous factor graphs in robotics.** Factor graphs modeling several relevant discrete-continuous robot perception problems. Discrete variable nodes are colored red, continuous variable nodes are blue, and factor nodes are black. (a) Switching systems: discrete states control the evolution of a continuous process. (b) Outlier rejection: discrete inlier/outlier variables control whether a subset of untrusted measurements should be used in estimating continuous variables. (c) Point-cloud registration: discrete variables represent correspondences and the continuous variable is the relative transformation from a source to target point-cloud.

based approaches described in the previous chapters are not directly applicable in this case. This is the problem that we address in this chapter.

Our key insight is that in many instances, while maximum *a posteriori* (MAP) inference for graphical models containing both discrete and continuous variables is hard (see e.g. [75, Sec. 14.3.1]), if we fix either the discrete or continuous variables, local optimization of the other set is easy. Continuous optimization can be performed using smooth, gradient-based methods, while discrete optimization can be performed *exactly* for a fixed assignment to the continuous variables by means of standard max-

product variable elimination [75, Sec. 13.2.1]. In turn, our approach can be perform efficient inference in high-dimensional, nonlinear models commonly encountered in robotics. Moreover, this approach naturally extends many of the additional desired capabilities of an inference approach in robotics applications, such as incremental computation [70] and uncertainty estimation (cf. [69]) to the hybrid setting.

Our contributions are as follows: From a robotics science standpoint, we show that by leveraging the conditional independence structure of hybrid factor graphs commonly encountered in robotics problems, efficient local optimization can be performed using alternating optimization, which we prove guarantees monotonic improvement in the objective. Because our approach naturally respects the incremental structure of many such problems, it easily scales to thousands of discrete variables *without* the need to prune discrete assignments. From a systems standpoint, our discrete-continuous smoothing and mapping (DC-SAM) library[1] extends existing GTSAM tools by adding (1) explicit constructions for hybrid discrete-continuous factors, (2) a new solver capable of computing approximate solutions to the corresponding estimation problems, and (3) an approach for approximating uncertainties associated with solutions to these problems which does not depend on the solver we employ (and therefore is likely to be of independent interest). To the best of our knowledge, these are the first openly-available tools for general discrete-continuous factor graphs encountered in robotics applications. We demonstrate the application of our methods to point-cloud registration, robust pose graph optimization. In the next chapter, we will make use of the tools and ideas developed in this chapter in application to semantic SLAM.

3.1 Problem formulation

We are interested in determining the most probable assignment to a set of discrete variables D and continuous variables C given a set of measurements $\tilde{Z}$. Under the

[1]The DC-SAM library is currently available at `https://www.github.com/MarineRoboticsGroup/dcsam`.

assumption that each measurement $\tilde{z}_k$ is independent of all others given the subset of variables $\mathcal{V}_k \subseteq \mathcal{V}$ it relates, we can decompose the posterior $p(C, D \mid \tilde{Z})$ into a product of measurement factors f_k, each of which depends only on a subset of variables $\mathcal{V}_k$ (cf. eq. (2.11)):

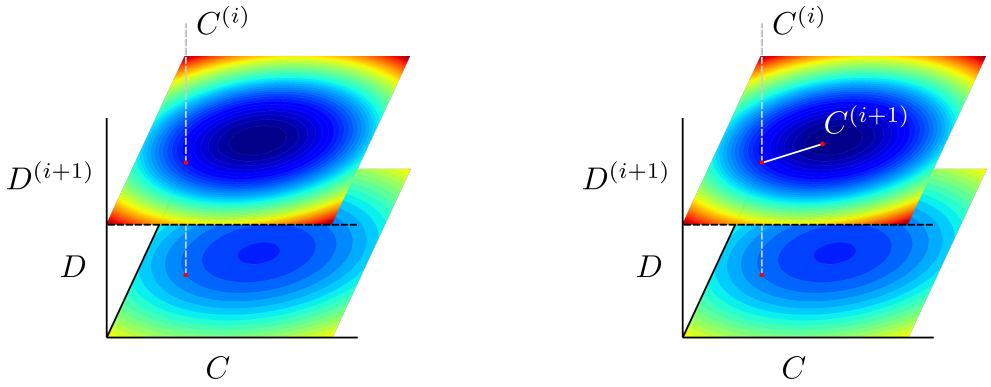
$$\begin{aligned} p(C, D \mid \tilde{Z}) &\propto \prod_k f_k(\mathcal{V}_k), \\ \mathcal{V}_k &\triangleq \{v \in \mathcal{V} \mid (f_k, v) \in \mathcal{E}\}, \end{aligned} \tag{3.1}$$

where each factor f_k is in correspondence with either a measurement likelihood of the form $p(\tilde{z}_k \mid \mathcal{V}_k)$ or a prior $p(\mathcal{V}_k)$. From (3.1), the posterior $p(C, D \mid \tilde{Z})$ can be decomposed into factors f_k of three possible types: *discrete* factors $f_k(D_k)$ where $D_k \subseteq D$, *continuous* factors $f_k(C_k)$, $C_k \subseteq C$, and *discrete-continuous* factors $f_k(C_k, D_k)$. In turn, the maximum *a posteriori* inference problem can be posed in terms of the following adaptation of Problem 1:

$$\begin{aligned} C^*, D^* &= \operatorname*{argmax}_{C,D} p(C, D \mid \tilde{Z}) \\ &= \operatorname*{argmax}_{C,D} \prod_k f_k(\mathcal{V}_k) \\ &= \operatorname*{argmin}_{C,D} \sum_k -\log f_k(\mathcal{V}_k). \end{aligned} \tag{3.2}$$

That is to say, we can maximize the posterior probability $p(C, D \mid \tilde{Z})$ by minimizing the negative log posterior, which in turn decomposes as a summation. Though the theoretical aspects of the methods we propose are quite general, in application we will primarily be concerned with factor graphs in which maximum likelihood estimation (or maximum *a posteriori* inference) can be represented in terms of a nonlinear least-squares problem, which permits the application of incremental nonlinear least-squares solvers like iSAM2 [70].[2] In particular, we consider discrete-continuous factors

[2] As we note in Section 2.3, this turns out not to be particularly restrictive, as any factor which is positive and bounded admits an equivalent representation in terms of a nonlinear least-squares cost function for the purposes of optimization.



(a) Discrete solve iteration. (b) Continuous solve iteration.

Figure 3-2: **Overview of a single iteration of optimization.** (a) First, given an initial iterate $C^{(i)}$ we solve exactly for the optimal assignment to the discrete variables using max-product elimination. (b) Next, given the latest assignment to the discrete variables, we update the continuous variables (e.g. using a trust-region method [117]). Color depicts the objective value of a solution, ranging from low cost (blue) to high cost (red).

$f_k(C_k, D_k)$ admitting a description as:

$$\begin{aligned} -\log f_k(C_k, D_k) &= \|r_k(C_k, D_k)\|_2^2, \\ C_k \subseteq C,\ D_k &\subseteq D, \end{aligned} \tag{3.3}$$

where the function $r_k : \Omega \times \mathbb{D} \to \mathbb{R}^m,\ \Omega \subseteq \mathbb{R}^d,\ \mathbb{D} \subseteq \mathbb{N}_0^{|D|}$ is first-order differentiable with respect to C. We consider factors involving only continuous variables admitting an analogous representation. We place no restriction on discrete factors.

3.2 Overview of the approach

The following subsections describe our approach to solving optimization problems of the form in (3.2). In Section 3.2.1, we outline our core alternating minimization procedure and prove that our approach guarantee monotonic descent guarantees. In Section 3.2.2, we describe how our approach can easily benefit from existing incremental optimization techniques to efficiently solve large-scale estimation problems. Finally, in Section 3.2.3, we consider the issue of estimating *uncertainties* for the solutions provided by our method.

3.2.1 Alternating minimization

In general, the MAP inference problem in (3.2) is computationally intractable [75, Sec. 13.1.1]. Indeed, even the purely continuous estimation problems arising in robot perception are typically NP-hard, including rotation averaging and pose-graph SLAM [119]. Despite this, smooth (local) optimization methods often perform quite well on such problems, both in their computational efficiency (owing to the fact that gradient computations are typically inexpensive) and quality of solutions when a good initialization can be supplied. However, even if we assume the ability to efficiently solve continuous estimation problems, the introduction of discrete variables complicates matters considerably: in the worst-case, solving for the joint MAP estimate globally requires that for each assignment to the discrete states we solve a continuous optimization subproblem, and discrete state spaces grow *exponentially* in the number of discrete variables under consideration. Consequently, efficient approximate solutions are needed.

Our key insight is that we can leverage the conditional independence structure of the factor graph model to develop an efficient local optimization method which we prove guarantees monotonic improvement in the posterior probability. To motivate our approach, we first observe that if we fix any assignment to the discrete states, the only variables remaining are continuous and approximate inference can be performed efficiently using smooth optimization techniques [70], [117]. In this sense, if we happened to know the assignment to the discrete variables, continuous optimization becomes "easy." On the other hand, if we fix an estimate for the continuous variables, we are left with an optimization problem defined over a discrete factor graph which can be solved to global optimality using max-product variable elimination [75, Sec. 13.2.1], but in the worst case may still require exploration of *exponentially many* discrete states. However, it turns out that for many commonly encountered problems, we can often do much better than the worst case.

For example, consider a partition of the discrete states into mutually exclusive

subsets $D_j \subseteq D$ which are *conditionally independent* given the continuous states:

$$p(D \mid C, \tilde{Z}) \propto \prod_j p(D_j \mid C, \tilde{Z}). \tag{3.4}$$

It is straightforward to verify from the mutual exclusivity of each set D_j that the problem of optimizing the conditional in (3.4) then breaks up into subproblems involving each D_j:

$$\max_D p(D \mid C, \tilde{Z}) \propto \prod_j \left[\max_{D_j} p(D_j \mid C, \tilde{Z}) \right]. \tag{3.5}$$

Critically, we have exchanged computation of the maximum of the product with the product of each maximum computed *independently*. In cases where the discrete states decompose into particularly small subsets ($|D_j| \ll |D|$), inference may be carried out efficiently. Many hybrid optimization problems encountered in robotics admit such advantageous conditional independence structures. For example, Figures 3-1b and 3-1c, depicting robust pose graph optimization and point-cloud registration, respectively, admit a decomposition of the form in equation (3.4) where each subset D_j contains only *a single discrete variable*. Moreover, some discrete factor graphs do not decompose quite so drastically after conditioning on continuous states, but may still permit efficient inference. For example, Figure 3-1a depicts a switching system in which, after conditioning on the continuous variables, the resulting discrete graph is a hidden Markov model, for which the most probable assignment to the discrete states can be computed in polynomial time using the Viterbi algorithm [140].

In turn, we will use these ideas to construct an algorithm for efficiently producing solutions to problems of the form in (3.2).[3] Consider the negative log posterior, defined as:

$$\mathcal{L}(C, D) \triangleq -\log p(C, D \mid \tilde{Z}). \tag{3.6}$$

[3]The approach we present does not *require* that a model admit a conditional factorization like the one in equation (3.4), though it improves computational efficiency considerably (see Section 3.4.1 for a discussion).

From (3.2), the joint optimization problem of interest can be formulated as:

$$C^*, D^* = \underset{C,D}{\operatorname{argmin}}\, \mathcal{L}(C, D). \tag{3.7}$$

Our alternating minimization approach (depicted in Figure 3-2) proceeds as follows: first, fix an initial iterate $C^{(i)}$. Then, we aim to solve the following subproblems:

$$D^{(i+1)} = \underset{D}{\operatorname{argmin}}\, \mathcal{L}(C^{(i)}, D) \tag{3.8a}$$

$$C^{(i+1)} = \underset{C}{\operatorname{argmin}}\, \mathcal{L}(C, D^{(i+1)}). \tag{3.8b}$$

We may then repeat (3.8a) and (3.8b) until the relative decrease in $\mathcal{L}(C, D)$ is sufficiently small or we have reached a maximum desired number of iterations. Finding minimizers for the subproblems (3.8a) and (3.8b) may still be challenging. Fortunately, one need not find a minimizer for the subproblems (3.8a) and (3.8b) in order for our approach to ensure monotonic improvements to the objective. In particular, we require only that at each iteration the following descent criteria hold:

$$\mathcal{L}(C^{(i)}, D^{(i+1)}) \leq \mathcal{L}(C^{(i)}, D^{(i)}) \tag{3.9a}$$

$$\mathcal{L}(C^{(i+1)}, D^{(i+1)}) \leq \mathcal{L}(C^{(i)}, D^{(i+1)}). \tag{3.9b}$$

There are many methods for updating the discrete and continuous states that satisfy (3.9a) and (3.9b), respectively. For the discrete states, the descent criterion in (3.9a) can be ensured by using the max-product algorithm to compute the *optimal solution* to the subproblem in (3.8a). For the continuous states, the descent criterion in (3.9b) can be guaranteed by, for instance, using a trust region method (e.g. [117]) to refine the continuous states with respect to the objective in (3.8b) . In turn, we obtain the following proposition:

Proposition 1. *Let $\mathcal{L}(C, D)$ be the objective to be minimized, with initial iterate $C^{(0)}, D^{(0)}$. Suppose that at each iteration, the discrete update satisfies the descent criterion in* (3.9a) *and likewise for the continuous update in* (3.9b)*. Then, the estimates*

$C^{(i)}, D^{(i)}$ obtained by alternating optimization satisfy:

$$\mathcal{L}(C^{(0)}, D^{(0)}) \geq \mathcal{L}(C^{(1)}, D^{(1)}) \geq \ldots \geq \mathcal{L}(C^{(T)}, D^{(T)}). \tag{3.10}$$

That is, this procedure ensures monotonic improvement in the objective.

Proof. Fix an initial iterate $(C^{(i)}, D^{(i)})$. By hypothesis, after a discrete update, we have $\mathcal{L}(C^{(i)}, D^{(i+1)}) \leq \mathcal{L}(C^{(i)}, D^{(i)})$ (from (3.9a)). Consequently, the updated assignment comprised of the pair $(C^{(i)}, D^{(i+1)})$ is *at least as good* as the previous assignment. By the same reasoning, performing a subsequent continuous update gives a pair $(C^{(i+1)}, D^{(i+1)})$ satisfying $\mathcal{L}(C^{(i+1)}, D^{(i+1)}) \leq \mathcal{L}(C^{(i)}, D^{(i+1)})$ (from (3.9b)). Combining these inequalities, we have:

$$\mathcal{L}(C^{(i+1)}, D^{(i+1)}) \leq \mathcal{L}(C^{(i)}, D^{(i+1)}) \leq \mathcal{L}(C^{(i)}, D^{(i)}). \tag{3.11}$$

The above chain of inequalities holds for all i, completing the proof. □

3.2.2 Online, incremental inference

Many robotics problems naturally admit *incremental* solutions wherein new information impacts only a small subset of the states we would like to estimate. Because our alternating minimization approach relies only upon the ability to provide an improvement in each of the *separate* discrete and continuous subproblem steps, we can rely on existing techniques to solve these problems in an incremental fashion. In particular, in the continuous optimization subproblem, we use iSAM2 [70] to refactor the graph containing continuous variables into a *Bayes tree*, permitting incremental inference of the continuous variables. Similarly, owing to the discrete factorization in (3.4), if, for example, we introduce new discrete variables which are conditionally independent of all previous discrete states given the current continuous state estimate, we are able to solve for the most probable assignment to these variables *without* the need to recompute solutions for previously estimated variables. In turn, we are able to efficiently solve online inference problems, as we will demonstrate in Section 4, in

which we produce solutions to online SLAM problems.

3.2.3 Recovering marginals

Uncertainty representation is important in many applications of robot perception. DC-SAM supports *post hoc* recovery of approximate marginal distributions for discrete and continuous variables from an estimate. For continuous variables, we use the *Laplace approximate* [13, Sec. 4.4] adopted by several nonlinear least-squares solvers (Ceres, g2o, and GTSAM). In particular, we fix a linearization point for the continuous variables (and a current estimate for discrete variables) and compute an approximate linear Gaussian distribution centered at this linearization point. For discrete variables, we fix an assignment to the continuous variables and compute the exact discrete marginals conditioned on this linearization point using sum-product variable elimination [75, Ch. 9-10]. The marginals we recover, then are:

$$p(D_j \mid \hat{C}, \tilde{Z}) = \sum_{D \setminus D_j} p(D \mid \hat{C}, \tilde{Z}), \quad D_j \subseteq D, \tag{3.12a}$$

$$p(C_j \mid \hat{D}, \tilde{Z}) = \int_{C \setminus C_j} p(C \mid \hat{D}, \tilde{Z}), \quad C_j \subseteq C. \tag{3.12b}$$

The reason for this approach is that in general, the number of posterior modes captured by a particular (discrete-continuous) factor graph can grow combinatorially. Computing exact marginals (i.e. determining exact solutions to Problem 2) can easily become intractable. In contrast, by making use of the conditional factorization in (3.4), solving for the discrete marginals in (3.12a) is often tractable.[4] Notably, our approach to marginal recovery does not require that one use the alternating minimization strategy outlined in Section 3.2.1; any method of providing an estimate $(\hat{C}, \hat{D})$ will suffice.

The continuous marginals in (3.12b) are estimated using the Laplace approximation [69]. In our derivation, it will be convenient to consider the continuous states as

[4]It is also interesting to note that the discrete marginals we recover are *exactly* the "weights" computed in the expectation step of the well-known *expectation-maximization* (EM) algorithm [42].

a vector $C \in \mathbb{R}^d$. Assume the point $(\hat{C}, \hat{D})$ is a critical point of the continuous subproblem (3.8b), i.e. $\nabla \mathcal{L}(C, \hat{D})|_{\hat{C}} = 0$. Consider a Taylor expansion of the objective $\mathcal{L}(C, \hat{D})$ about the point $\hat{C}$:

$$\mathcal{L}(C, \hat{D}) \approx \mathcal{L}(\hat{C}, \hat{D}) - \frac{1}{2} A \left(C - \hat{C} \right), \tag{3.13}$$

with the $d \times d$ Hessian matrix A defined as:

$$A \triangleq -\nabla^2 \mathcal{L}(C, \hat{D})|_{\hat{C}}. \tag{3.14}$$

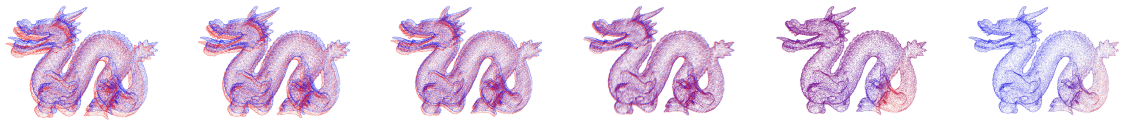
Exponentiating both sides of (3.13) and appropriately normalizing the result gives the linear Gaussian approximation:

$$p(C \mid \hat{D}, \tilde{Z}) \approx \frac{|A|^{1/2}}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} \| C - \hat{C} \|^2_{A^{-1}} \right\}, \tag{3.15}$$

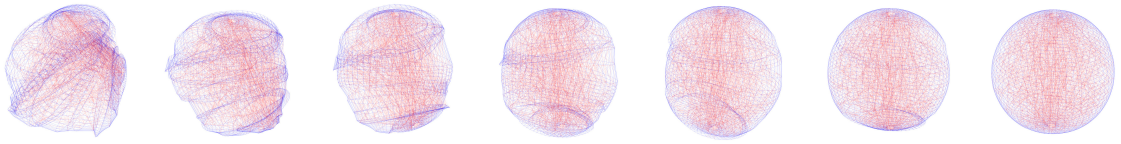
where $\|c\|_{A^{-1}}$ denotes the Mahalanobis norm $\sqrt{c^\mathsf{T} A c}$. When all factors involving continuous variables take the form in (3.3), the locally linear approximation of $\mathcal{L}$ about $\hat{C}$ admits a Hessian A which can be expressed in terms of the Jacobian of the measurement function r, and we have $A \succeq 0$ [40]. Additionally, the relevant components of the matrix A for estimating the marginals for a subset of variables C_j can be recovered from its square root, i.e. the *square-root information matrix* (cf. [69]).

3.3 Example applications

In the following sections we provide example applications motivated by typical robot perception problems. In Section 3.3.1 we demonstrate application of DC-SAM to the problem of point-cloud registration and show that it naturally generalizes the well-known *iterative closest point* (ICP) method [12, 32]. In Section 3.3.2, we consider the problem of *robust pose graph optimization*, where we aim to estimate a set of poses given only noisy measurements between a subset of them, and some fraction of those



(a) Point-cloud registration



(b) Robust pose graph optimization

Figure 3-3: **Example applications.** (a) Point-cloud registration using the *Stanford Dragon* dataset [37]. (b) Robust pose graph optimization using the *Sphere* dataset [70]. Each row displays the sequence of iterates for our method. In each case, we obtain high-quality solutions in just a few iterations.

measurements may be outliers. We implement a straightforward approach to solving this problem using DC-SAM and show that it produces competitive results.

3.3.1 Point-cloud registration

As a simple first example, we will consider the point-cloud registration problem. Consider a source point-cloud $\mathcal{P}_S = \{p_i^S \in \mathbb{R}^d,\ i = 1, \dots, n\}$ and target point-cloud $\mathcal{P}_T = \{p_j^T \in \mathbb{R}^d,\ j = 1, \dots, m\}$. Associate with each point in the source cloud p_i^S a discrete variable $d_i \in \{1, \dots, m\}$ determining the corresponding point in the target cloud. The goal of point-cloud registration is to identify the rigid-body transformation $T \in \mathrm{SE}(3)$ that minimizes the following objective:

$$\min_{T \in \mathrm{SE}(3)} \sum_{i=1}^{n} \|T p_i^S - p_{d_i}^T\|_2^2. \tag{3.16}$$

The key challenge encountered in this setting is that the correspondence variables d_i are unknown and unobserved. We might consider, then, introducing the correspondence variables into the optimization, to determine the *best* set of correspondence variables *and* the corresponding rigid-body transformation of the point-cloud, ob-

taining the following problem:

$$\min_{d_i \in 1:m, T \in \mathrm{SE}(3)} \sum_{i=1}^{n} \|T p_i^S - p_{d_i}^T\|_2^2. \tag{3.17}$$

Unfortunately, this problem is nonconvex and solving it to global optimality is, in general, NP-hard, requiring search over $\mathcal{O}(n^m)$ discrete state assignments.

A popular algorithm for solving the problem in equation (3.17) is to first posit an initial guess for the transformation T, determine the transformed locations of each of the points in the source cloud, then associate each point in the source cloud with the *nearest* point in the target cloud after the transformation. This is the *iterative closest point* (ICP) algorithm [12, 32]. Defining $r_i(T, d_i) = T p_i^S - p_{d_i}^T$, we can see that the problem in equation (3.17) is concisely described in terms of factors of the form (3.3). Moreover, the conditional independence structure of the graph corresponding to this problem (depicted in Figure 3-1) immediately motivates our alternating optimization approach, since each d_i in fact *decouples* when conditioned on T. Finally, one can verify that our alternating optimization procedure turns out to be identical to ICP (as described above) in this setting. To demonstrate this fact, we applied our method to point cloud registration using the *Stanford Dragon* dataset [37], the results of which are depicted in Figure 3-3a. Indeed, we observe that our approach produces qualitatively reasonable results in just a few iterations. Moreover, while implementing ICP typically requires that we explicitly write the (independent) correspondence updates and transform update, we need not encode this explicitly at all: the fact that the discrete (correspondence) update separates into independent subproblems is simply a consequence of the conditional independence structure of the factor graph model in Figure 3-1c. That said, our approach does not have knowledge about the particular *spatial* structure of the problem and therefore performs naïve search over discrete assignments. In contrast, a typical implementation of ICP would make use of efficient spatial data structures to speed up the solution to the discrete subproblem, see [122] (indeed, such optimizations for *particular* problems like this would make for interesting future applications of the DC-SAM library). However, unlike any *partic-*

ular ICP implementation, our solver can be readily extended (without modification) to more complex cost functions or models because the structure of the subproblems is dictated by the independence structure inherent in the graphical model.

3.3.2 Robust pose graph optimization

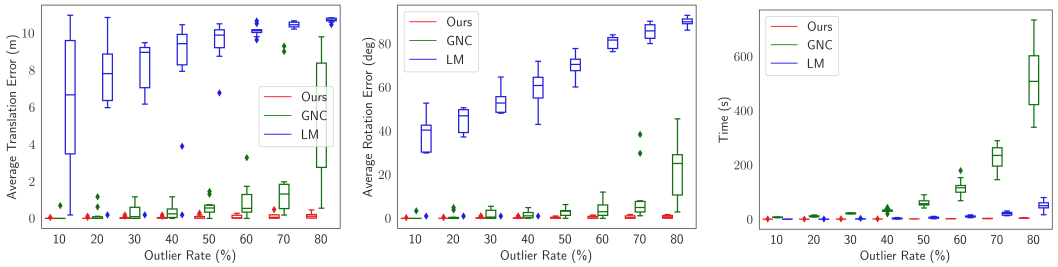
In this section we consider *robust* pose graph optimization. Recall (from Section 2.5.2) that in pose graph optimization we are interested in estimating a set of poses $x_1, \ldots, x_n \in \mathrm{SE}(3)$ from noisy measurements $\tilde{x}_{ij}$ of a subset of their (true) relative transforms $\underline{x}_{ij} = \underline{x}_i^{-1}\underline{x}_j$. This problem possesses a natural graphical structure $\mathcal{G} = \{\mathcal{V}, \vec{\mathcal{E}}\}$ where nodes correspond to the poses x_i to be estimated and edges correspond to the available noisy measurements between them. Pose graph optimization then aims to solve the following problem:

$$\min_{x_i \in \mathrm{SE}(3)} \sum_{\{i,j\} \in \vec{\mathcal{E}}} \big\| \underbrace{\log\left(\tilde{x}_{ij}^{-1} x_i^{-1} x_j\right)^\vee}_{r_{ij}(x_i, x_j)} \big\|_\Sigma^2, \tag{3.18}$$

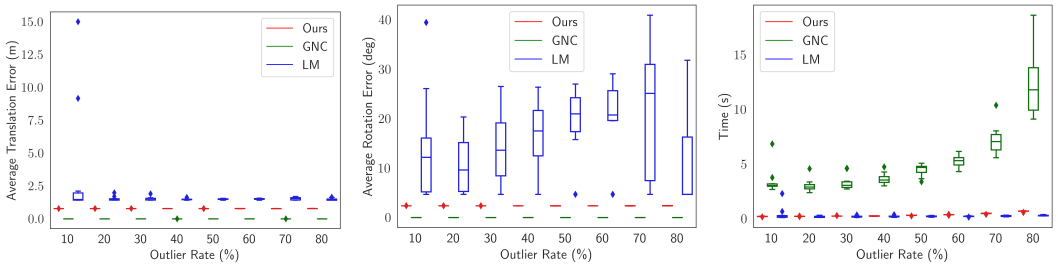
where $\log(\cdot)^\vee : \mathrm{SE}(3) \to \mathbb{R}^6$ takes an element of $\mathrm{SE}(3)$ to an element of the tangent space (cf. [8, Sec. 8.3.2]), and $\Sigma \in \mathbb{R}^{6\times 6}$ is a covariance matrix.

Suppose however, that some fraction of our measurements are corrupted by an unknown outlier process. We would like to determine the subset of outlier measurements and inlier measurements, as well as the corresponding optimal poses. It is typical to assume that the edges $\vec{\mathcal{E}}$ partition into a set of trusted odometry edges $\vec{\mathcal{E}}_\mathcal{O}$ and a set of untrusted loop closure edges $\vec{\mathcal{E}}_\mathcal{L}$. It is common to address this problem by introducing binary variables $d_{ij} \in \{0, 1\}$ for each of the untrusted edges (cf. [3, 105, 127, 133]), where $d_{ij} = 1$ indicates that the measurement $\tilde{x}_{ij}$ is drawn from the outlier process. Since the outlier distribution is unknown, it is common to assume that the outlier generating process is Gaussian with covariance $\tilde{\Sigma} \succ \Sigma$ much larger than the inlier model covariance. In turn, the problem of interest can be posed as follows:

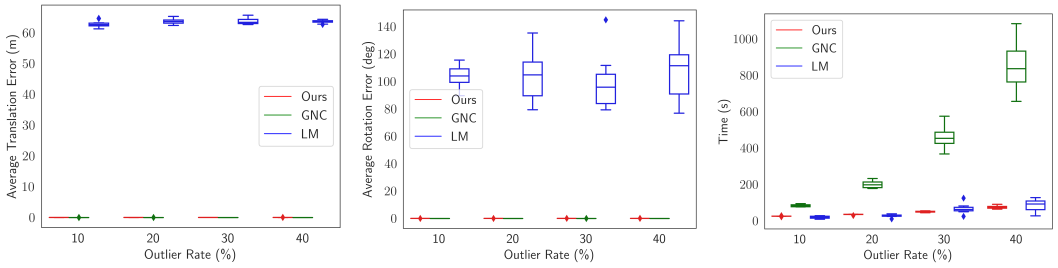
$$\min_{\substack{x_i \in \mathrm{SE}(3) \\ d_{ij} \in \{0,1\}}} \sum_{\{i,j\} \in \vec{\mathcal{E}}_\mathcal{O}} \|r_{ij}(x_i, x_j)\|_\Sigma^2 + \sum_{\{i,j\} \in \vec{\mathcal{E}}_\mathcal{L}} e_{ij}(x_i, x_j, d_{ij}), \tag{3.19}$$



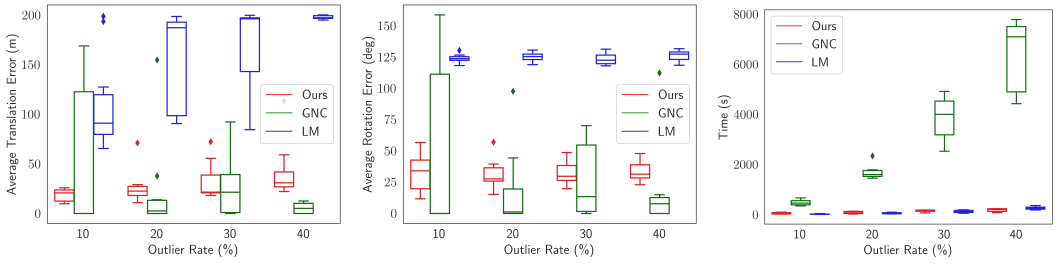
(a) *Intel*



(b) *CSAIL*



(c) *Sphere*



(d) *Garage*

Figure 3-4: **Robust pose graph optimization.** Average trajectory errors on (a) the *Intel* dataset, (b) the *CSAIL* dataset, (c) the *Sphere* dataset, and (d) the *Garage* dataset. Left to right: translation error, rotation error, and computation time. Statistics computed over 10 Monte Carlo trials. LM refers to the result obtained by running Levenberg-Marquardt on the corrupted graph.

where

$$e_{ij}(x_i, x_j, d_{ij}) \triangleq \begin{cases} -\log \omega_0 + \|r_{ij}(x_i, x_j)\|^2_\Sigma, \ d_{ij} = 0, \\ -\log \omega_1 + \|r_{ij}(x_i, x_j)\|^2_{\tilde{\Sigma}}, \ d_{ij} = 1, \end{cases} \tag{3.20}$$

and $\omega_0, \ \omega_1 \in [0, 1]$ are prior weights on the inlier and outlier hypotheses, respectively. Letting $|\vec{\mathcal{E}}_\mathcal{L}| = m$, there are $\mathcal{O}(2^m)$ possible assignments to the discrete variables in this problem. However, the above formulation can easily be represented in terms of discrete factors for the weights $\omega_0, \ \omega_1$ and discrete-continuous factors of the form in (3.3) to switch between the Gaussian inlier and outlier hypotheses. Moreover, once again, the discrete variables decouple from one another conveniently when we condition on an assignment to the continuous variables (Fig. 3-1b shows the corresponding graph).

In our experimental setup, we corrupt pose graphs with outliers generated between a random pair of (non-adjacent) poses with relative translation sampled uniformly from a cube of side-length 10 meters and rotation sampled from the uniform distribution over rotations (a similar process to the one described in [138, Section VI.C]). Based on the prior work of Olson and Agarwal [105], we made the outlier covariance model isotropic with variance 10^7 times larger than the inlier variance and set the weights $\omega_0, \ \omega_1$ to be the corresponding Gaussian normalizing constants. We provide two points of comparison: a Levenberg-Marquardt (LM) solver applied to the graph corrupted by outliers (as a "worst case") and the state-of-the-art graduated nonconvexity (GNC) solver [145].[5] Our results are summarized in Figure 3-4. In particular, we observe that in the cases that we are able to supply a high-quality initialization, optimization using our approach enables recovery of accurate SLAM solutions *significantly faster* than the GNC approach (and in some cases, faster than the non-robust baseline).[6] Our approach is susceptible to local optima (leading to

[5]We use the GNC approach implemented in GTSAM with the truncated least-squares cost. We use the default parameters from the GTSAM implementation, though the performance of GNC (in terms of computation time and solution quality) may be improved over the results shown here through further parameter tuning.

[6]The computation speed of our approach is primarily derived from two factors: first, we exploit efficient incremental optimization via iSAM2, and second, our optimization procedure is purely local, as opposed to GNC which requires solving re-weighted variants of the original pose graph optimization problem several times in an effort to improve robustness to initialization.

suboptimal performance on the CSAIL dataset). We will revisit this issue in Section 3.4.2.

3.4 Discussion

3.4.1 When is alternating minimization efficient?

The conditional factorization in equation (3.4) serves to give some intuition for when our optimization approach is computationally efficient. If the distribution over discrete variables conditioned on the continuous assignment admits a factorization into small subsets D_j, then the optimization problem in (3.8a) decouples into separate problems in direct correspondence with each set D_j. Since we perform exact inference on this distribution, solving for the most probable assignment is in the worst case exponential in the size of D_j [75]. Consequently, in graphs with densely connected discrete variables that are not decoupled by continuous variables, the per iteration complexity of alternating minimization can increase dramatically. That said, Proposition 1 ensures monotonic improvement in the objective so long as each optimization subproblem admits a solution no worse than the current iterate. Therefore, it is reasonable to consider extending this approach by allowing for *local* optimization in the discrete subproblem [126].

3.4.2 When can we ensure accurate solutions?

Though we are able to make some claims about when solutions to the discrete and continuous subproblems in our alternating minimization approach can be tractably computed, the question remains as to when one can ensure that these local search methods recover *high-quality* solutions. Since the alternating minimization approach is a *descent* method, we rely on the ability to provide a "good" initial guess from which purely going "downhill" in the cost landscape is enough to obtain a high-quality estimate. However, this is already a requirement of off-the-shelf tools for solving many robot perception problems, such as pose-graph SLAM, which (by virtue

of the nonconvexity of the optimization problems they attempt to solve) require high-quality initialization [119].[7] Nonetheless, the consideration of discrete variables *can* make initialization more challenging. The specifics of providing an initial guess will ultimately depend heavily on the application.

One can also attempt to reduce the initialization sensitivity of solutions obtained by our local optimization approach. A number of methods along these lines have been proposed. For example, graduated nonconvexity (GNC) [145] as discussed in Section 3.3.2, optimizes nonconvex functions by successively producing (and optimizing) a more well-behaved (typically convex) surrogate. Sampling methods and simulated annealing methods can improve convergence by allowing for the exploration of states that may *increase* cost or by initializing a descent method like our proposed approach from several starting points [86, 135]. Similarly, stochastic gradient descent is a classical approach for nonconvex optimization (and has appeared in the setting of robust pose-graph SLAM [104]), which could reasonably be adapted to our approach. Finally, heuristics have been considered which use consistency of measurements to filter out unlikely hypotheses [88] or to *re-initialize* estimates for factor graphs [84].

3.5 Summary

In this chapter we presented an approach to optimization in discrete-continuous graphical models based on *alternating minimization*. Our key insight is that the structure of the alternating optimization procedure allows us to leverage the conditional independence relations exposed by factor graphs to efficiently perform local search. We showed how the complexity of inference in this setting is related to structure of the graphical model itself. Critically, we observed that many important problems in robotics can be framed in terms of graphical models admitting particularly advantageous structures for application of our approach. We provided a method for addressing the issue of recovering uncertainties associated with estimates in the

[7] Moreover, even in these "simpler" problem instances, *verification* that a globally optimal solution has been found has only been demonstrated for certain special cases (see [119, Sec 2] for a review) and is otherwise itself an open problem.

discrete-continuous setting. Our solver and associated tools are implemented as part of our library, DC-SAM, which is, to the best of our knowledge, the first openly available library for addressing these *hybrid* discrete-continuous optimization problems. Finally, we demonstrate the application of our method to the key problems of robust pose graph optimization. In the next chapter, we will demonstrate another important application of these tools and ideas to semantic SLAM.

Chapter 4

Robust object-level semantic SLAM

As mentioned in Section 2.5.1, landmark-based SLAM hinges critically on *data association*, the ability to recognize previously mapped landmarks. Unfortunately, achieving consistent data association over long periods of operation, while vital for reliable robot navigation in the operational regime as "time goes to infinity," is far more difficult than short-term data association. In this regime a robot's pose uncertainty may grow large enough that many landmarks present reasonable loop closure candidates. Consequently, data association is a major failure mode of modern navigation systems. Indeed "regardless of the type of data association employed, it is highly likely that if an estimation technique fails, the blame can be squarely placed on bad data association" [8, p. 154]. For this reason, any mechanism by which we can associate landmarks *uniquely* is of interest.

Recently, advances in the capabilities of learned perception models, especially deep neural networks, motivate their use for the extraction of "higher-level" feature descriptors or object-level semantic landmarks for mapping [134]. Not only do these methods provide additional information that can be used to disambiguate environmental landmarks, but often this information carries with it a semantic interpretation, grounding the robot map representation in terms of objects. That said, no object detection or recognition model could be expected to offer perfect performance over the lifetime of a robot. In this chapter, we aim to build semantic map representations that support robot navigation in a way that does not depend on perfect detection

and classification of objects; rather we aim to construct estimation procedures that may leverage the strengths of these methods but are robust to their failures. Solving this problem requires joint inference of discrete (data associations and landmark classes) and continuous (robot poses and landmark locations) variables, which is fundamentally a computationally hard problem [39, 119]. In consequence, approximate inference methods which remain computationally tractable in application are needed.

To that end, this chapter makes several contributions. First, we show that under common assumptions on the factorization of the posterior over robot and landmark states, the (latent) discrete association variables are conditionally independent, making *exact elimination* possible [28]. Viewed through this lens, we show that prior work on approximating sum-mixtures of Gaussians in SLAM using *max-mixtures* [105] in fact arises directly from (exact) max-product elimination of discrete variables. Moreover, this suggests that the max-mixtures approach is best interpreted *not* simply as a computationally tractable approximation to a sum-mixture (in the setting of nonlinear least-squares optimization for SLAM), but rather as an exact representation of the factor graph model resulting from analytic elimination of association variables. Crucially, this perspective enables a natural generalization the max-mixtures approach to the setting where landmark states and measurements are jointly discrete and continuous (possessing a semantic class as well as a position), which is necessary for its application to semantic SLAM problems. Our prior work [43, 46] made use of similar ideas, but it was motivated as a computationally efficient approximation which was primarily heuristic in nature. This chapter gives an exposition from the perspective of inference in graphical models that makes the exactness of our approach obvious. Second, we consider sum-product elimination of discrete variables. We give a derivation of an expectation-maximization procedure for optimizing the resulting marginal posterior over the remaining variables. This approach is formally equivalent to that of Bowman et al. [18] and admits a convenient representation in terms a single factor definition. Both of these approaches are capable of incorporating association uncertainty in the setting of discrete-continuous landmark states in a computationally tractable manner. Finally, we observe that the resulting MAP inference problem of

interest is defined with respect to a *hybrid* model (including continuous states like robot poses and landmark locations, as well as discrete landmark class variables), and so to implement these ideas we employ the DC-SAM library and solver developed in the previous chapter. We provide experimental validation of these approaches using real data from a mobile robot and the KITTI dataset [56], demonstrating their practicality in real applications.

4.1 Problem statement

4.1.1 Semantic SLAM with unknown data association

Formally, we define the semantic SLAM problem in 3 dimensions as the inference of robot poses $X \triangleq \{x_i : x_i \in \mathrm{SE}(3),\ i = 1, \dots N\}$, and semantic landmarks $\ell_j \triangleq (\ell_j^\rho, \ell_j^c),\ j = 1, \dots, M$ comprising the set L, where each landmark is split into a (continuous) geometric component $\ell_j^\rho \in \mathbb{R}^3$ and a (discrete) semantic class component $\ell_j^c \in \mathcal{C}$ from a known set of class labels $\mathcal{C} \triangleq \{1, \dots, C\}$, given a set of measurements $\tilde{Z}$. This fits the general landmark-based SLAM paradigm outlined in Section 2.5.1. Moreover, this is a natural generalization of a purely geometric landmark-based SLAM approach; by simply taking $\mathbb{C} = 1$, we recover the usual geometric formulation. This corresponds to the following maximum *a posteriori* (MAP) inference problem:

$$X^*, L^* = \underset{X,L}{\operatorname{argmax}}\, p(X, L \mid \tilde{Z}). \tag{4.1}$$

When associations between measurements and map landmarks are unknown, they must also be inferred during navigation. In particular, suppose there are K landmark measurements with unknown associations. We introduce the discrete association variables $D \triangleq \{d_k : d_k \in \mathbb{N}_{\leq M},\ k = 1, \dots K\}$. Modifying the problem in (4.1) to accommodate these data association variables, we obtain:

$$X^*, L^*, D^* = \underset{X,L,D}{\operatorname{argmax}}\, p(X, L, D \mid \tilde{Z}). \tag{4.2}$$

Unfortunately, solving this problem to guaranteed global optimality is in general intractable, requiring search over all $\mathcal{O}(M^K)$ possible data association hypotheses.

In SLAM, we are primarily interested in the robot and landmark states, X and L; we are typically not concerned about the optimal assignments to the data association variables themselves. Instead, then, we can consider eliminating the data association variables. to form the following marginal MAP inference problem:

$$X^*, L^* = \underset{X,L}{\operatorname{argmax}} \left[\max_D p(X, L, D \mid Z) \right], \tag{4.3a}$$

$$X^+, L^+ = \underset{X,L}{\operatorname{argmax}} \underbrace{\sum_D p(X, L, D \mid Z)}_{p(X,L|Z)}, \tag{4.3b}$$

where here we are using the notation X^+, L^+ to distinguish the marginal MAP from the MAP estimate itself, X^*, L^*. It would seem that we have not gained anything computationally by eliminating data association variables. Indeed, it is straightforward to verify that (4.3a) is essentially a reorganization of terms in (4.2). Moreover, the marginal distribution $p(X, L \mid Z) = \sum_D p(X, L, D \mid Z)$ in (4.3b) is non-Gaussian, even when the measurement models themselves are corrupted by Gaussian noise. As we will show in Section 4.2, the key advantage of variable elimination in this setting derives from the conditional independence structure of the graphical SLAM problem.

4.1.2 Problem Statement

The semantic SLAM formulations described in Section 4.1.1 are broad enough to encompass a number of NP-hard problems (see, e.g. [75, 119]). In fact, the semantic SLAM process is made even more challenging by the issue of false positives, i.e. sensor measurements that occur with no landmark present, as well as the crucial fact that the number of landmarks M is not known beforehand; rather, it typically grows during navigation. Consequently, we assume access to a reasonable landmark *hypothesis set* $\mathcal{H}_k \subseteq L$ for a landmark measurement $\tilde{z}_k$. Beyond this assumption, the methods we present are not restricted to a particular class of measurement model.

In this chapter, however, our application of interest is semantic SLAM, and we will consider sensor configurations giving access to odometry measurements between subsequent keyframes, geometric landmark measurements, and semantic landmark measurements. All measurements of continuous variables are assumed to be corrupted by additive Gaussian noise. We divide measurements into two types, corresponding to odometry factors $\psi_{ij} \in \mathcal{F}_{\mathcal{O}}$ between poses x_i and x_j and (ambiguous) landmark measurements $\phi_{ik} \in \mathcal{F}_{\mathcal{L}}$ between a pose x_i, all landmarks in the hypothesis set $\mathcal{H}_k$, and a discrete association variable d_k, giving the MAP inference problem:

$$\max_{X,L,D} \prod_{\psi_{ij} \in \mathcal{F}_{\mathcal{O}}} \psi_{ij}(x_i, x_j) \prod_{\phi_{ik} \in \mathcal{F}_{\mathcal{L}}} \phi_{ik}(x_i, \mathcal{H}_k, d_k). \tag{4.4}$$

We will also consider the maximum marginal estimation problem from (4.3b), which can be written in terms of the above factorization as:

$$\max_{X,L} \sum_{D} \prod_{\psi_{ij} \in \mathcal{F}_{\mathcal{O}}} \psi_{ij}(x_i, x_j) \prod_{\phi_{ik} \in \mathcal{F}_{\mathcal{L}}} \phi_{ik}(x_i, \mathcal{H}_k, d_k). \tag{4.5}$$

Since our focus is primarily on the issue of data association, which involves only landmark measurements, it will be convenient throughout to abuse notation slightly by defining a single function summarizing the joint likelihood of all odometry measurements:

$$\psi(X) \triangleq \prod_{\psi_{ij} \in \mathcal{F}_{\mathcal{O}}} \psi_{ij}(x_i, x_j), \tag{4.6}$$

which do not depend on the associations. We assume that the landmark measurement model factors into conditionally independent geometric and semantic components given the association variable: i.e.:

$$\begin{aligned} \phi(x, \mathcal{H}, d) &\triangleq p(\tilde{z} \mid x, \mathcal{H}, d) \\ &= p(\tilde{z}^{\rho} \mid x, \ell_d^{\rho}) p(\tilde{z}^{c} \mid \ell_d^{c}), \end{aligned} \tag{4.7}$$

where $\tilde{z}^{\rho}$ and $\tilde{z}^{c}$ are the geometric and semantic components of the measurement, respectively. The geometric measurement likelihood $p(\tilde{z}^{\rho} \mid x, \ell_d^{\rho})$ is taken to be Gaus-

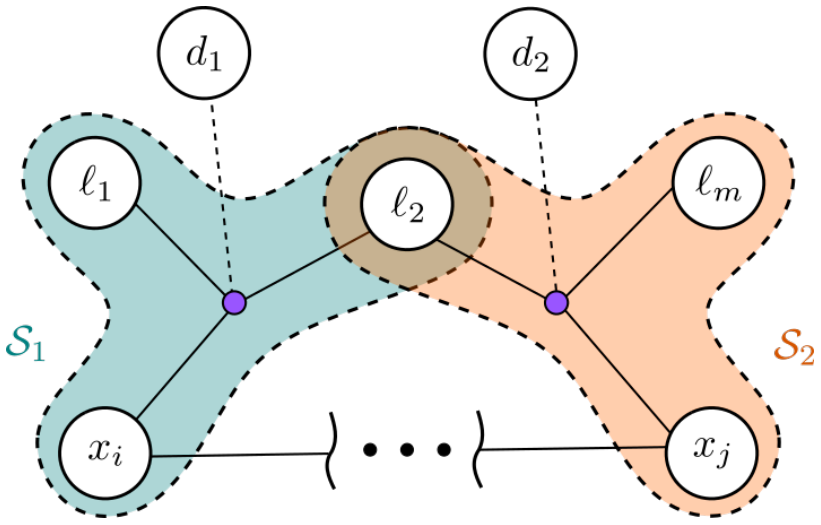


Figure 4-1: **Conditional independence of data associations.** In the setting of landmark-based SLAM, data associations which are conditionally independent can be analytically eliminated. The hypothesis sets $\mathcal{H}_k$ consist of sets of landmarks potentially associated with the k-th measurement. The associations d_1 and d_2 are conditionally independent given the values assigned to the variables in their separators sets $\mathcal{S}_k = \mathcal{H}_k \cup x_i$ (respectively x_j).

sian with mean $g(x, \ell_d^\rho)$ for some (typically nonlinear) measurement model g. In this work, we will consider object detections providing access to range and bearing to a 3D point (comprising the geometric measurement $\tilde{z}^\rho$), so g gives the range and bearing between a pose and a 3D point, and an object class label $\tilde{z}^c$. Finally, we assume that odometry measurements and geometric landmark measurements are corrupted by additive Gaussian noise, and we assume knowledge of the misclassification statistics of the detector, $p(\tilde{z}^c \mid \ell_d^c)$, i.e. the probability that the detector outputs $\tilde{z}^c$ given knowledge of the true class of the corresponding landmark, which could be expressed, for example, in terms of the $C \times C$ confusion matrix for an object classifier.

4.2 Approach

The following subsections discuss the application of two approaches to eliminate data association variables in the SLAM problem. Specifically, we describe *max-product* elimination to solve the MAP inference problem described in (4.3a), and *sum-product* elimination to address the marginal MAP inference problem in (4.3b). In the latter

case, we describe an inference approach based on the expectation-maximization algorithm (c.f. [18]). As we will show, each approach can conveniently be represented in terms of a factor amenable to optimization using standard tools (e.g. GTSAM [39]). Finally, we provide an approach for determining a reasonable hypothesis set in Section 4.2.3.

4.2.1 From MAP inference to max-product elimination

First, we consider directly solving the MAP inference problem in (4.4). It is clear that naïve search over all possible combinations association decisions would be intractable even for modestly sized problems. However, we can simplify the problem considerably. Since the association variables D do not appear in the odometry factors within the MAP inference problem (4.4), we can immediately rewrite the optimization as:

$$p^* = \max_{X,L} \psi(X) \left[\max_{D} \prod_{f_{ik} \in \mathcal{F}_{\mathcal{L}}} f_{ik}(x_i, \mathcal{H}_k, d_k) \right], \tag{4.8}$$

where p^* is the posterior probability attained by the MAP estimate. Now, each factor f_{ik} depends only on x_i, $\mathcal{H}_k$, and d_k. This is a direct consequence of the conditional independence structure exposed by the factor graph representation, and is depicted in Figure 4-1. Moreover, each association variable d_k is involved with *exactly one* factor f_{ik}. Therefore, the maximum over D of the product of all f_{ik} is attained simply as the product of the maxima taken separately over each d_k:

$$p^* = \max_{X,L} \psi(X) \prod_{f_{ik} \in \mathcal{F}_{\mathcal{L}}} \max_{d_k} f_{ik}(x_i, \mathcal{H}_k, d_k). \tag{4.9}$$

Consider the factor defined as:

$$f_{ik}^{mm}(x_i, \mathcal{H}_k) \triangleq \max_{d} f_{ik}(x_i, \mathcal{H}_k, d). \tag{4.10}$$

We immediately observe that this factor does not depend on any *particular* assignment to the association variable. Rather, for any assignment to a pose and the

set of landmarks in the hypothesis set, it determines the optimal association for a particular measurement. The quantity in equation 4.10 is commonly known as a "max-marginal" in the setting of probabilistic graphical models [75]. In the setting of pose-graph SLAM where the measurement models of interest are restricted to be relative measurements between robot poses, an analogous expression has appeared before as a max-mixture [105], where it was observed that when the component measurement models appearing in the inner maximization are Gaussian, the result can serve as an approximation of a sum-mixture of Gaussians. However, we now demonstrate the following remarkable fact: if we eliminate association variables according to (4.10), the resulting optimization problem over the remaining variables is *equivalent* to the original MAP inference problem. Substitution of (4.10) into (4.9) reveals:

$$p^* = \max_{X,L} \ \psi(X) \prod_{\phi_{ik}^{mm} \in \mathcal{F}_{\mathcal{L}}} \phi_{ik}^{mm}(x_i, \mathcal{H}_k). \tag{4.11}$$

Moreover, given solutions (X^*, L^*) to this problem, we can recover the corresponding association variables d_k^* as follows:

$$d_k^* = \operatorname*{argmax}_d f_{ik}(x_i^*, \mathcal{H}_k^*, d). \tag{4.12}$$

To see this, form the problem:

$$\begin{aligned} p^* &= \max_D \psi(X^*) \prod_{f_{ik} \in \mathcal{F}_{\mathcal{L}}} f_{ik}(x_i^*, \mathcal{H}_k^*, d_k) \\ &= \psi(X^*) \max_D \prod_{f_{ik} \in \mathcal{F}_{\mathcal{L}}} f_{ik}(x_i^*, \mathcal{H}_k^*, d_k) \\ &= \psi(X^*) \prod_{f_{ik} \in \mathcal{F}_{\mathcal{L}}} \max_{d_k} f_{ik}(x_i^*, \mathcal{H}_k^*, d_k), \end{aligned} \tag{4.13}$$

where in the last line we have used again the conditional independence of each d_k to factor the maximization into a set of subproblems in one-to-one correspondence with the landmark measurement factors. Finally, we observe that the maximum in (4.13) is specifically attained when each d_k takes on the optimal assignment for its

individual subproblem. By construction, d_k^* as defined in (4.12) is this assignment.

The equivalence of the problems in (4.4) and (4.11) is significant. By virtue of this equivalence, we have not skirted the issue of the computational hardness of rendering global solutions to the problem in (4.4). However, this change has the profound benefit of making the problem amenable to standard (computationally efficient) optimization techniques based on local search. Furthermore, since the geometric part of each component factor in (4.10) is assumed to be Gaussian with respect to a nonlinear measurement model, we can specifically make use of SLAM solvers tailored toward incremental nonlinear least-squares applications (e.g. iSAM2 [70]). Finally, from an optimizer X^*, L^*, of (4.11), we can recover the corresponding optimal associations (and, in the event that X^* and L^* happen to be *global* optimizers of (4.11), then the triple (X^*, L^*, D^*) will likewise be a globally optimal solution to (4.2)).

4.2.2 From marginal MAP inference to sum-product elimination

In this section we consider the marginal MAP inference problem in (4.5). The same reasoning we applied in the previous section can be used to show that the sum-marginal distribution on the right-hand side of equation (4.5) admits a similarly convenient decomposition whereby the max operator is replaced by a summation:

$$p^+ = \max_{X,L} \psi(X) \prod_{f_{ik} \in \mathcal{F}_\mathcal{L}} \sum_{d_k} f_{ik}(x_i, \mathcal{H}_k, d_k), \tag{4.14}$$

where p^+ is the optimal value for the sum marginal optimization problem. However, the factor representation analogous to the one in (4.10) (corresponding the summation terms appearing in (4.14)) is non-Gaussian, prohibiting the direct application of nonlinear least-squares optimization approaches to solve (4.5) (remedying this was the motivation for the original max-mixtures approach of Olson and Agarwal [105]). Rosen et al. [120] provide an algebraic reduction that can transform any positive, bounded factor into a representation suitable for nonlinear least-squares optimiza-

tion. Pfeifer et al. [108] subsequently applied this technique in the setting of point set registration. However, we consider an alternative approach based on expectation-maximization which admits a straightforward implementation and, as we will show, optimizes the same objective.

Originally applied to semantic SLAM by Bowman et al. [18], the expectation-maximization approach iteratively optimizes the marginal likelihood (or, in this case, the marginal posterior) in equation 4.3b in a two-step procedure [42]. In the expectation step, a distribution over the hidden (data association) variables given the remaining variables is computed, allowing the formation of a lower-bound on the marginal likelihood (or posterior). In the subsequent maximization step, the lower bound function is maximized with respect to the remaining variables (poses and landmarks). Given an iterate $\hat{X}, \hat{L}$, the factor representation for this approach can be expressed as:

$$f^{em}(x_i, \mathcal{H}_k) \triangleq \prod_{d \in \mathcal{H}_k} f(x_i, \mathcal{H}_k, d)^{w_d}, \tag{4.15}$$

where

$$w_d \triangleq \frac{f(\hat{x}, \hat{\mathcal{H}}, d)}{\sum_{d' \in \mathcal{H}} f(\hat{x}, \hat{\mathcal{H}}, d')}. \tag{4.16}$$

In our setting, the apparent benefit of this representation, as opposed to the exact formulation, is that the negative logarithm of this factor now takes the form of a weighted combination of the (negative logarithm of the) original factors. This makes implementation straightforward when one already has access to the component measurement models.

4.2.3 Computing candidate association hypotheses

Thus far, we have assumed knowledge of a collectively exhaustive set of data association hypotheses for each measurement. In practice, however, a reasonable hypothesis set must be determined as a robot collects new measurements. In this section, we suggest a method for determining candidate hypotheses based on the commonly employed *Laplace approximation* for the robot's belief state at a particular time (e.g.

as in [69, 129]). At a high-level, our approach is to first compute a linear Gaussian approximation to the true (non-Gaussian) posterior belief at a linearization point corresponding to our current state estimate.[1] We will use this approximation to estimate the marginal probability of a landmark measurement $\tilde{z}$ under each possible data association hypothesis (i.e. one hypothesis per mapped landmark), given all previous measurements, denoted $\tilde{Z}_{\setminus\{\tilde{z}\}}$. A threshold on the marginal measurement probability determines whether a measurement corresponds to a previously mapped landmark, or a new landmark. If the measurement is determined to be a new landmark, it is added to the map, otherwise all landmarks passing the gate threshold are considered as potential hypotheses and incorporated into the factor graph as a mixture factor.

Consider a single object detection measurement $\tilde{z} = (\tilde{z}^c, \tilde{z}^\rho)$ consisting jointly of geometric and semantic information obtained from a pose x. We will make the posterior marginal approximation $p(x, \ell \mid \tilde{Z}_{\setminus\{\tilde{z}\}}) \approx p(x, \ell^\rho \mid \tilde{Z}_{\setminus\{\tilde{z}\}}) p(\ell^c \mid \tilde{Z}_{\setminus\{\tilde{z}\}})$.[2] From this, and the factored measurement model, the likelihood of the form $p(\tilde{z} \mid d, \tilde{Z}_{\setminus\{\tilde{z}\}})$ can be broken into the product of separate semantic and geometric likelihoods:

$$p(\tilde{z}^c, \tilde{z}^\rho \mid d = j, \tilde{Z}_{\setminus\{\tilde{z}\}}) \approx p\left(\tilde{z}^c \mid \cdot\right) p\left(\tilde{z}^\rho \mid \cdot\right). \tag{4.17}$$

Each term on the right-hand side can be expanded as follows into the summation over landmark classes:

$$p(\tilde{z}^c \mid d = j, \tilde{Z}_{\setminus\{z\}}) = \sum_{\ell_j^c} p(\tilde{z}^c \mid \ell_j^c) p(\ell_j^c \mid \tilde{Z}_{\setminus\{\tilde{z}\}}), \tag{4.18}$$

and integral over robot pose and landmark location:

$$p(\tilde{z}^\rho \mid d = j, \tilde{Z}_{\setminus\{\tilde{z}\}}) = \int_{x, \ell_j^\rho} p\left(\tilde{z}^\rho \mid x, \ell_j^\rho\right) p(x, \ell_j^\rho \mid \tilde{Z}_{\setminus\{\tilde{z}\}}). \tag{4.19}$$

[1] As one might expect, a unimodal Gaussian distribution may quite poorly capture the true non-Gaussian belief. While this is computationally convenient, it is a challenge of applying these methods. Methods for global loop closure, e.g. incorporating finer-grained descriptors of objects, can improve the robustness of this process, but do not address the key difficulty in representing the non-Gaussian belief.

[2] The motivation for this approximation is that recovering the exact *hybrid* marginal will generally be intractable.

The marginal distribution $p(x, \ell_j^\rho \mid \tilde{Z}_{\setminus\{\tilde{z}\}})$ is generally non-Gaussian, and in turn this integral is typically intractable. Consequently, we approximate the marginal distribution via the Laplace approximation centered at the current estimate. Specifically, letting $\hat{\mathbf{x}}$ denote the state vector obtained by concatenating an estimate of the robot pose x and landmark position ℓ_j^ρ, we obtain an approximating distribution which is Gaussian with mean $\hat{\mathbf{x}}$ and covariance Σ:[3]

$$p(x, \ell_j^\rho \mid \tilde{Z}_{\setminus\{\tilde{z}\}}) \approx \mathcal{N}(\hat{\mathbf{x}}, \Sigma). \tag{4.20}$$

With this approximation, all of the terms in the integral are Gaussian and we obtain (as in [69]):

$$p(\tilde{z}^\rho \mid d = j, \tilde{Z}_{\setminus\{\tilde{z}\}}) \approx \frac{1}{\sqrt{|2\pi R_j|}} e^{-\frac{1}{2}\|g(\hat{\mathbf{x}}) - \tilde{z}^\rho\|^2_{R_j}}. \tag{4.21}$$

The covariance R_j is defined as:

$$R_j \triangleq \left.\frac{\partial g}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}} \Sigma \left.\frac{\partial g}{\partial \mathbf{x}}\right|^{\mathsf{T}}_{\hat{\mathbf{x}}} + \Gamma, \tag{4.22}$$

where Σ is the block joint covariance matrix between pose x_t and candidate landmark position ℓ_j^ρ, $\partial g / \partial \mathbf{x}$ is the Jacobian of the measurement function, and Γ is the covariance of the geometric measurement model. This result, combined with the expression in (4.18) gives the marginal likelihood in (4.17) that we normalize to compute data association probabilities. Those landmarks for which this approximate marginal computation is greater than a pre-determined threshold will be considered hypotheses. In particular, with the observation that the squared-error term $\|g(\hat{\mathbf{x}}) - \tilde{z}^\rho\|^2_{R_j}$ is a chi-squared distributed random variable, we base the acceptance criterion around a chi-squared test χ^2_α having a number of degrees of freedom equal to the measurement

[3] Approximate marginals of this form have the additional benefit that they are readily accessed using existing computational tools for SLAM like iSAM and iSAM2 [70] in libraries like GTSAM [38]. To recover the marginals for both the discrete and continuous states, we employ the approach described in Section 3.2.3 as implemented in DC-SAM.

dimension and with confidence level α, giving:

$$-2\log p(\tilde{z}^c \mid d = j, \tilde{Z}_{\setminus\{\tilde{z}\}}) + \|g(\hat{\mathbf{x}}) - \tilde{z}^\rho\|^2_{R_j} < \chi^2_\alpha, \tag{4.23}$$

where the factor of 2 appropriately scales the log probability of the semantic measurement component to match the geometric component.[4] The expression in equation (4.23) can be intuitively thought of as an adaptive chi-squared test on only the geometric measurement component. It is clear that if the semantic measurement and landmark class agree perfectly we have $p(\tilde{z}^c \mid \cdot\ ) = 1$, and therefore (4.23) reverts to a standard chi-squared test with statistic χ^2_α on the geometric measurement. In the usual event where the semantic measurement and landmark class disagree somewhat, we have: $p(\tilde{z}^c \mid \cdot\ ) < 1$ which makes the test in (4.23) equivalent to a *more stringent* chi-squared test on the geometric measurement. This satisfies expectations: if the class measurement and the landmark class disagree, we require a commensurate *increase* in the probability of the geometric measurement in order to accept an association.

Finally, in practice, we may also want to consider a "null-hypothesis" where the measurement is assumed to be a false-positive detection and has no correspondence with a known landmark. We can account for this possibility by placing a prior weight on a null-hypothesis component factor, which is implemented as a Gaussian component with large variance for the geometric part and a uniform distribution over landmark classes for the semantic part.

4.3 Experimental results

We compared variable elimination approaches for data association in two ways: First, in results presented in Section 4.3.1, we consider artificial semantic SLAM tasks using data obtained during indoor navigation with an MIT RACECAR vehicle[5] equipped

[4] More precisely, while a factor of $\frac{1}{2}$ appears in the logarithm of the probability of a geometric measurement in (4.21), the test statistic for geometric measurements appearing in (4.23) is twice this value, so we scale the log probability of the semantic measurement accordingly.

[5] `https://mit-racecar.github.io/`

only with a ZED stereo camera [1]. We simulated semantic SLAM tasks using AprilTag fiducials [87, 143], which allow us to generate synthetic object class measurements (for any number of classes up to the total number of AprilTags in the environment) based on the known unique ID number of each tag. Knowledge of the AprilTag ID also allows us to construct baseline solutions with known data association. We added odometry noise and random misclassifications at varying rates in order to assess the robustness of each approach, the results of which are summarized in Section 4.3.1. Second, in Section 4.3.2 we evaluated each approach on real stereo image data from the KITTI dataset [56, 57] using detections of cars for loop closures.

We implemented our approach in C++ using the discrete-continuous smoothing and mapping (DC-SAM) library [47], which makes use of iSAM2 [70] and the GTSAM [38] library for optimization and covariance recovery. Experiments were run on a single core of a 2.2 GHz Intel i7 CPU. We use evo [61] for trajectory evaluation.

In all of the experiments, we compare three approaches to data association: naïve maximum-likelihood association, denoted (ML), max-product elimination (or, max-mixtures), denoted (MM), and expectation-maximization, denoted (EM).[6] For all of the approaches other than maximum-likelihood, we also examine the addition of a null hypothesis decision (denoted +NH where applicable).

4.3.1 MIT RACECAR dataset

We collected roughly 25 minutes of data during indoor navigation with the MIT RACECAR mobile robot platform (depicted in Figure 4-2) over a roughly 1.08 km trajectory. We sampled AprilTag [87, 143] detection keyframes at a rate of 1 Hz resulting in 702 observations of 262 unique tags. The dataset itself consists of several repeated traversals of a single large loops with occasional 180 degree turns. The nature of this dataset makes the problem of data association a significant challenge, particularly for closing large loops and dealing with hairpin turns.

Odometry was obtained using the ZED stereo camera visual odometry [1]. The

[6]The expectation-maximization approach is equivalent to the approach of Bowman et al. [18] applied in our problem setting.

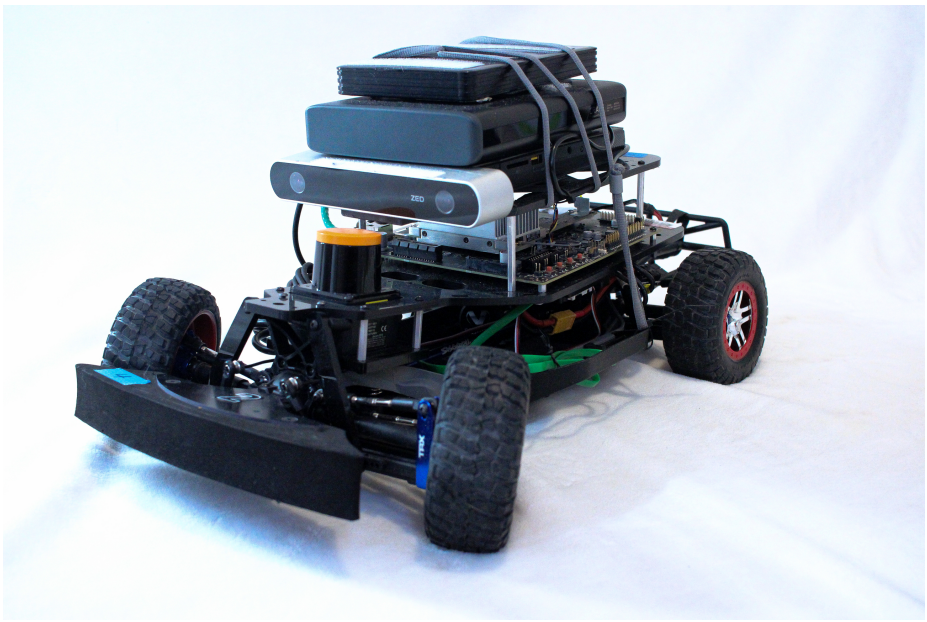
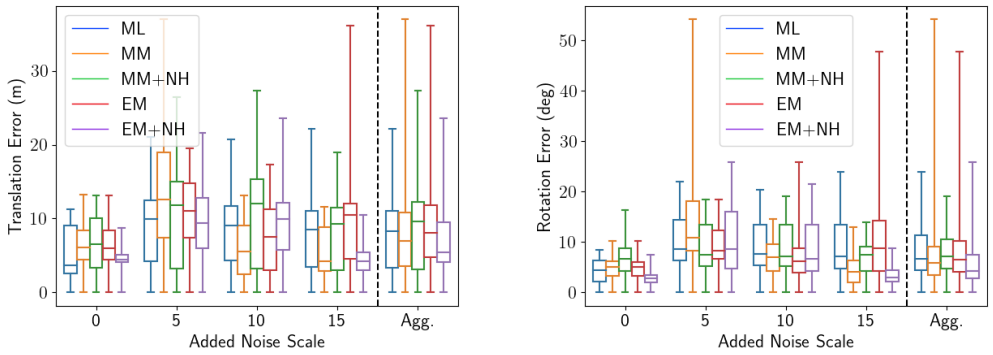


Figure 4-2: **MIT RACECAR platform.** The MIT RACECAR mobile robot platform used in the experiments. Figure adapted from [10].

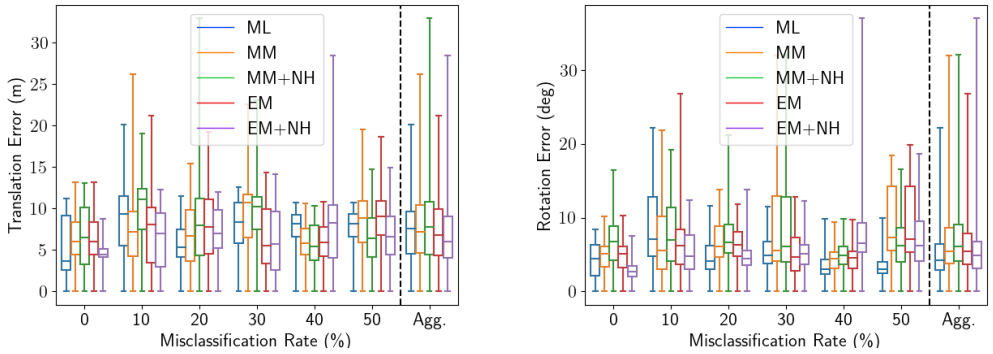
use of AprilTags allows us to obtain a baseline solution with known data associations. We assigned semantic labels to each AprilTag as the true tag ID modulo C for a C-class semantic SLAM problem. While AprilTags provide full rotation and translation estimates for tag poses in the camera frame, we do not expect this in general of neural network-based object detectors. Consequently, we consider only the range and bearing to AprilTags in our semantic SLAM system.

We examined the performance of each system across two parameters of interest: noise in odometry measurements and misclassification rate. To simulate added noise in odometry measurements, we perturb the existing measurements with additive Gaussian noise. We restrict this addition of noise to measurements of planar motion and sample from a distribution with standard deviations equal to 10^{-3} times the base model standard deviation. We examine the performance of each estimator while varying the scale applied to these models; that is, for an added noise "scale" of 10, the standard deviation of the augmented noise is 10^{-2} times that of the base model. To simulate misclassification, we randomly perturb classifications with a particular probability. That is, for a misclassification rate of 0.1, there is a 90% chance of observing the true landmark class and a 10% chance that it will be misclassified. In the case of

odometry noise and misclassification rate, we assume access to an accurate model of the noise characteristics. As a consequence, for a two-class classification problem (as we consider here), a 50% misclassification rate causes the system to discard semantic information entirely. Quantitative results for these experiments are summarized in Figure 4-3. The corresponding qualitative results are provided in Figure 4-4. The difficulty of this dataset is reflected by the relatively poor estimation accuracy of all of the methods, and is a consequence of the fact that accurate state estimation on this particular dataset requires reliably closing large loops. Since loop closure is determined by a measurement gate, this places the burden of robustness predominantly on the gating procedure. Since our approach does not address the issue of maintaining multiple hypotheses about the *existence* of a new landmark for a particular measurement (only the *identity* of a landmark correspondence once existence is established) we cannot recover from situations where the system creates spurious landmarks (this is true of all the methods we tested). Moreover, if the correct landmark hypothesis is not in the set of correspondences for a given measurement, the best we can hope for in the current system is that the measurement will be rejected as null. Furthermore, the specific gating procedure we employ here makes use of a unimodal Gaussian approximation of a posterior distribution that we do not expect to be truly unimodal. The quality of this approximation relies on the posterior distribution being concentrated around the current estimate of the map and robot trajectory. Of course, if the current estimate is wrong (e.g., as is common when information thus far is insufficient to distinguish between multiple landmark hypotheses) this can produce a misleading representation of the solution uncertainty. The performance of the expectation maximization approach, which effectively combines the impact of multiple hypotheses (including the null hypothesis) on the state uncertainty in a weighted average, seems to offer the most reliable performance using this gating procedure. Overall, our results suggest that the determination of new landmarks (i.e. the gating procedure) is an important challenge for future work on these systems. Better posterior approximations, such as the non-Gaussian formulation in [45] may help to some degree, but ultimately if we are to rely on a plausibly inaccurate state estimate for any decisions



(a)



(b)

Figure 4-3: **MIT RACECAR dataset trajectory error.** Error distribution for the MIT RACECAR dataset experiments. (a) Error as a function of added odometry noise. (b) Error as a function of misclassification rate.

whatsoever, it seems that we must have the ability to "revert" those decisions. Doing this in a computationally tractable manner is an open question, though heuristics have been produced in certain problem domains that may present initial steps in this direction (see, e.g., [84]).

4.3.2 KITTI datasets

We also evaluate our approach on stereo camera data from the KITTI dataset odometry sequences [56]. In our experiments, we use the YOLO object detector [114]. We threshold the confidence of the detector at 0.8, using detections of cars as landmarks. We use VISO2 stereo odometry for visual odometry [55]. We estimate the range and bearing to cars as the median range and bearing to all points that project into the

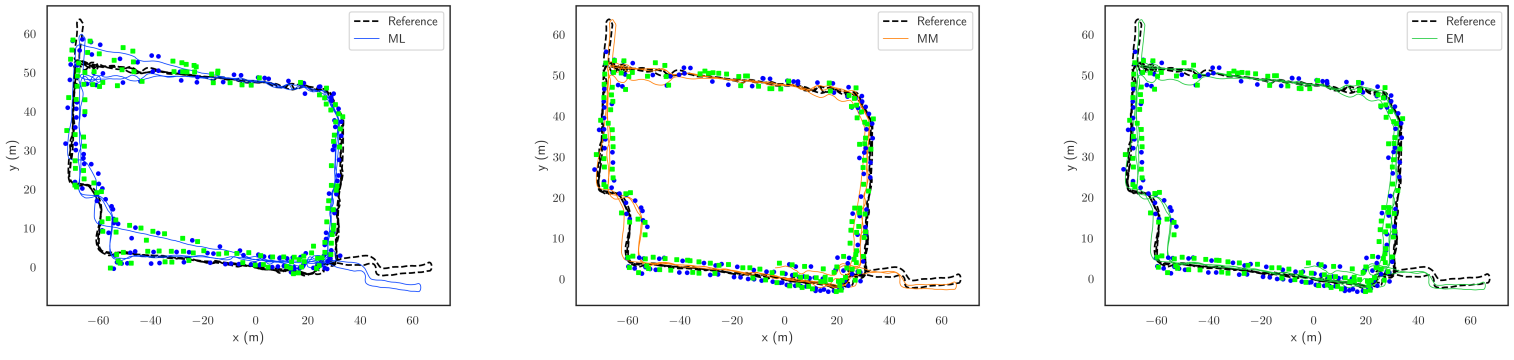


Figure 4-4: **MIT RACECAR dataset semantic maps.** Semantic maps produced by each method (no added noise). Trajectory colors consistent with Figure 4-3. Landmarks are colored green and blue depending on the semantic class.

		Translation Error (m)				Rotation Error (deg)			
Sequence	Method	Max	Mean	Median	RMSE	Max	Mean	Median	RMSE
00	ML	5.21	**2.52**	2.56	**2.76**	6.22	**1.31**	**1.22**	**1.49**
	MM	6.85	2.94	2.63	3.27	5.99	1.57	1.36	1.82
	EM	6.05	2.68	2.65	2.88	6.30	1.44	1.27	1.70
	MM+NH	5.26	2.60	**2.39**	2.78	6.00	1.40	1.29	1.59
	EM+NH	**4.86**	2.64	2.43	2.86	**5.68**	1.45	1.28	1.62
05	ML	4.53	2.21	2.09	2.49	2.78	1.12	0.99	1.31
	MM	**4.42**	2.18	2.06	**2.44**	2.73	**1.11**	**0.96**	**1.29**
	EM	4.45	**2.17**	**2.03**	**2.44**	**2.72**	**1.11**	**0.96**	**1.29**
	MM+NH	**4.42**	2.18	2.06	**2.44**	2.73	**1.11**	**0.96**	**1.29**
	EM+NH	4.45	**2.17**	**2.03**	**2.44**	**2.72**	**1.11**	**0.96**	**1.29**
07	ML	14.74	6.15	7.06	7.51	6.82	3.91	4.10	4.24
	MM	**5.59**	2.57	2.78	2.87	4.66	2.18	1.98	2.38
	EM	5.82	**2.54**	**2.72**	**2.84**	**4.06**	**2.16**	**1.96**	**2.34**
	MM+NH	**5.59**	2.57	2.78	2.87	4.66	2.18	1.98	2.38
	EM+NH	5.82	**2.54**	**2.72**	**2.84**	**4.06**	**2.16**	**1.96**	**2.34**

Table 4.1: **KITTI datasets.** Absolute translation and rotation errors (in meters and degrees, respectively) attained by semantic SLAM systems making use of each of the three data association methods: maximum-likelihood (ML), max-mixtures (MM), and expectation maximization (EM).

bounding box for a given car detection.

In this demonstration, we consider semantic SLAM using stereo camera data from the KITTI dataset [57]. We sample keyframes every two seconds, using VISO2 [55] to obtain stereo odometry measurements and YOLO [114] for noisy detections of cars. We estimate the range and bearing to an object's position as that of the median depth point projecting into a detected object's bounding box. Using DC-SAM, we are able to compute solutions to this problem online.[7] Table 4.1 gives a quantitative comparison of our approach with the odometric estimate from VISO2. Our approach substantially improves upon the translational errors of the odometric estimate and additionally enables the estimation of discrete landmark classes. Figure 4-5 gives a qualitative example demonstrating landmark class inference, in which we distinguish between cars and trucks as object-level landmarks.

4.4 Summary

In this chapter we presented an approach for semantic SLAM with unknown data association based on *exact elimination* of the association variables. We showed that for typical graphical models encountered in semantic SLAM, the association variables can be *analytically eliminated*. Inference over the eliminated graphs is equivalent to the original MAP (or marginal MAP) inference problem, but can be performed efficiently using local search techniques; in particular, incremental nonlinear least-squares solvers [70] as employed within our hybrid solver and library DC-SAM. Since our approach requires that we provide a suitable hypothesis set, we also gave a method based on a chi-squared hypothesis test to produce such a set. Our method for hypothesis set determination requires only the ability to compute approximate marginals over robot poses and landmark states, which are readily available using off-the-shelf tools like GTSAM [38]. Finally, we demonstrated our approach on semantic SLAM problems using real data from a mobile robot with simulated (noisy) object detections

[7] We run our solver on an Intel i7 2.6 GHz CPU and YOLO on an NVIDIA Quadro RTX 3000 GPU.

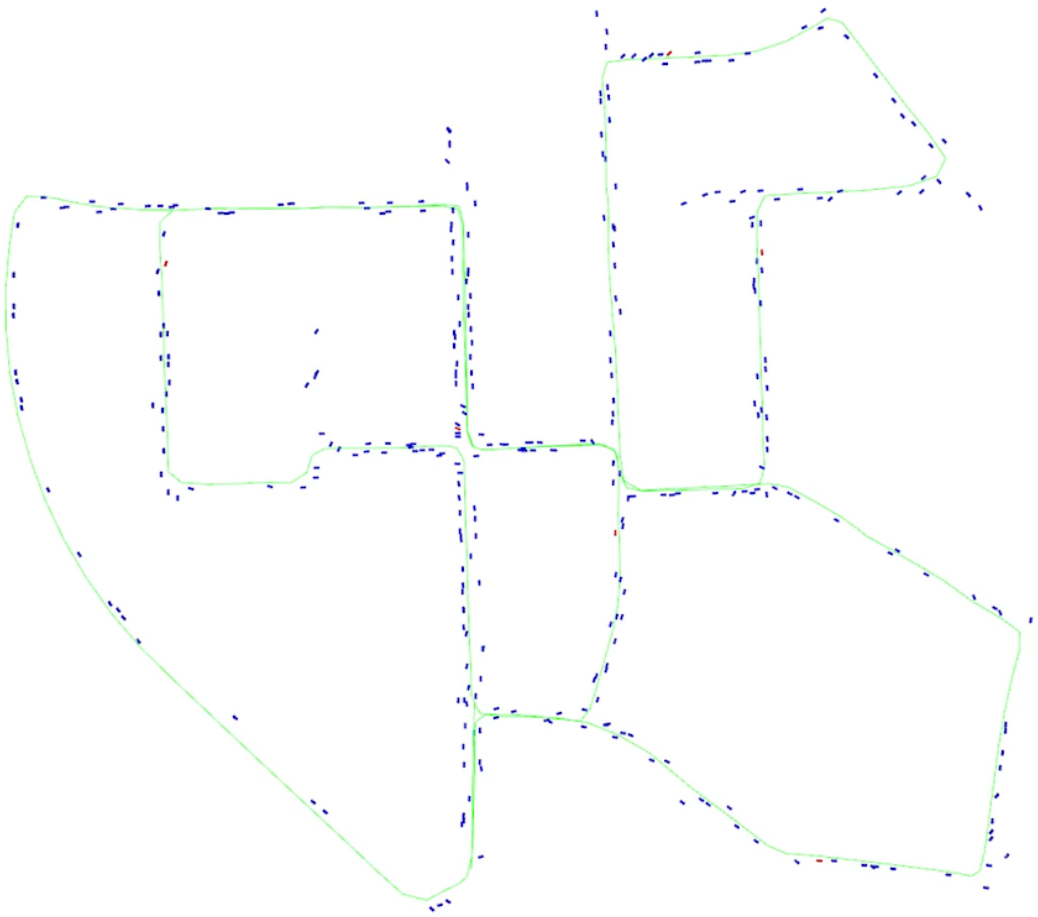


Figure 4-5: **Qualitative object-level SLAM results for KITTI dataset 00.** The output of our system (using the max-mixture model) is depicted above for a semantic SLAM problem on the KITTI dataset using observations of both cars and trucks. Cars are depicted in blue with trucks depicted in red. The estimated vehicle trajectory is shown in green.

obtained using fiducials as well as on the KITTI dataset using real object detections from YOLO [114].

With the recognition that even in cases where the cost functions to be optimized in the outlier free setting are convex, robust variants of these cost functions (e.g. the *truncated least-squares* (TLS) cost function) are nonconvex, methods have been proposed based on *graduated nonconvexity* (GNC), which initially solve a convex problem and aim to gradually recover the nonconvex, outlier-robust cost [145]. These approaches may improve the sensitivity of nonconvex robust estimation methods to initialization. To date, these methods have not been applied in the setting of landmark-based data association, though likewise it may be possible to consider similar techniques in the setting

Chapter 5

Performance guarantees for spectral initialization in rotation averaging and pose-graph SLAM

In Chapters 3 and 4 we considered problems of MAP inference in hybrid probabilistic models. The DC-SAM algorithm we developed in Chapter 3 is based on iterative improvement of an initial assignment to the states we aimed to estimate. However, we set aside the application-specific issue of actually obtaining that initial iterate. In this chapter, we examine the issue of initialization for robot perception problems (defined on continuous states): rotation averaging (RA), where all states and measurements are elements of the special orthogonal group; and pose-graph SLAM, where all states and measurements are elements of the special Euclidean group. The fact that the states in these problems are constrained to lie in sets which are nonconvex makes these estimation problems inherently *nonconvex*, with many bad local minima that can entrap the local optimization methods commonly applied to solve them. The performance of standard SLAM and RA algorithms thus crucially depends upon the quality of the estimates used to initialize the local search. In consequence, a great deal of prior work has been dedicated to the development of initialization techniques (see Carlone et al. [30] for a review). While many of these techniques often work well in practice, the fact that they are obtained as heuristic approximations makes it

difficult to ascertain *what specific features* of SLAM or RA problems determine their performance. As a result, it is difficult to say when, or under what conditions, these techniques can be *reliably* deployed.

We propose a simple *spectral initialization* method for pose-graph SLAM and rotation averaging that we prove enjoys *explicit performance guarantees*. To the best of our knowledge, these are the first concrete guarantees to appear in the literature for any initialization technique adapted to these applications. Our analysis gives direct control over the estimation error of a spectral initialization in terms of the spectral properties of the measurement network.[1] This allows us to control the distance from the spectral estimate to the global minimizer of the estimation problem; this is critical for ensuring that the initialization lies in the locally convex region around the global minimizer, and therefore that this minimizer can be recovered by a subsequent local refinement (see Figure 5-1). Our proof of this result relies on new estimation error bounds for the global minimizers (i.e. the *maximum likelihood* estimators) of SLAM and rotation averaging problems, which are likely to be of independent interest. Algorithmically, our approach only requires computing the first few eigenpairs of a symmetric matrix, which can be achieved using any off-the-shelf implementation of the Lanczos method (e.g. the MATLAB `eigs` command). Our empirical results on both synthetic data and standard pose-graph SLAM benchmarks demonstrate that the spectral estimator typically performs far better than our worst-case analysis suggests, achieving solution quality and computation times competitive with state-of-the-art approaches. Beyond its utility as an *initialization method* for , our results show that spectral relaxation provides an inexpensive method for rotation averaging and pose-graph optimization in its own right (i.e. *without* the need to perform subsequent nonconvex optimization or semidefinite relaxation) that attains an asymptotic error bound comparable to the (globally optimal) estimator, and provides near-optimal estimates in practice.

[1] Recent work has identified spectral properties of measurement networks as key quantities controlling the performance of estimators for these problems, though this connection (particularly in the context of SLAM) remains under-explored (see [119] for a recent review).

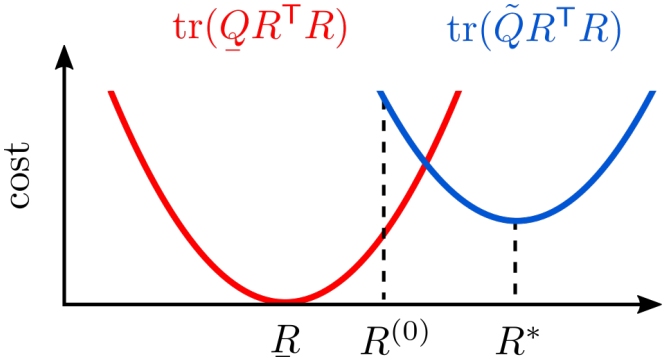


Figure 5-1: **Comparing true, optimal, and initial rotation estimates.** We are interested in bounds on the deviation of an initial estimate $R^{(0)}$ from the (latent) ground truth $\underline{R}$ and the globally optimal solution R^*.

5.1 Problem formulation

We consider the problem of synchronization over the $\mathrm{SO}(d)$ group: this is the problem of estimating n unknown values $R_1, \ldots, R_n \in \mathrm{SO}(d)$ given a set of noisy measurements $\tilde{R}_{ij}$ of a subset of their pairwise relative rotations $\underline{R}_{ij} \triangleq \underline{R}_i^{-1}\underline{R}_j$.[2] The problem of $\mathrm{SO}(d)$-synchronization captures, in particular, the problems of rotation averaging and, under common modeling assumptions, pose graph optimization (as we show in Problem 5 and equation (PGO)), where the variables of interest are the orientations of a robot (or more generally, a rigid body) at different points in time (see, for example Grisetti et al. [59]). This problem possesses a natural graphical structure $\mathcal{G} \triangleq (\mathcal{V}, \vec{\mathcal{E}})$, where nodes $\mathcal{V}$ correspond to latent variables $R_i \in \mathrm{SO}(d)$ and edges $(i,j) \in \vec{\mathcal{E}}$ correspond to (noisy) measured relative rotations $\tilde{R}_{ij}$ between R_i and R_j. In particular, for the problem of *rotation averaging*, we adopt the following standard generative model for rotation measurements: For each edge $(i,j) \in \vec{\mathcal{E}}$, we sample a noisy relative measurement $\tilde{R}_{ij}$ according to (cf. [41, 118]):

$$\tilde{R}_{ij} = \underline{R}_{ij} R_{ij}^{\epsilon}, \quad R_{ij}^{\epsilon} \sim \mathrm{Langevin}(I_d, \kappa_{ij}). \tag{5.1}$$

[2] As a brief notational remark: in the previous chapters we typically made use of d to indicate a discrete state we would like to estimate. Here, there are no discrete states, and d is simply used to denote the dimension of the problem under consideration (e.g. $d = 2$ for two-dimensional rotation synchronization, and so on).

Given a set of noisy pairwise relative rotations $\tilde{R}_{ij}$ sampled according to the generative model (5.1), a maximum likelihood estimate $R^* \in \mathrm{SO}(d)^n$ for the latent rotational states $R_1, \ldots, R_n$ is obtained as a minimizer of the following problem [41, 118]:

Problem 3 (Maximum likelihood estimation for rotation averaging)**.**

$$\min_{R_i \in \mathrm{SO}(d)} \sum_{(i,j)\in\vec{\mathcal{E}}} \kappa_{ij} \|R_j - R_i \tilde{R}_{ij}\|_F^2. \tag{5.2}$$

For pose-graph SLAM ($\mathrm{SE}(d)$-synchronization), we adopt the following generative model for rotation and translation measurements: For each edge $(i,j) \in \vec{\mathcal{E}}$, we sample a noisy relative measurement $\tilde{x}_{ij} = (\tilde{t}_{ij}, \tilde{R}_{ij}) \in \mathrm{SE}(d)$ according to:

$$\tilde{R}_{ij} = \underline{R}_{ij} R_{ij}^{\epsilon}, \quad R_{ij}^{\epsilon} \sim \mathrm{Langevin}(I_d, \kappa_{ij}) \tag{5.3a}$$

$$\tilde{t}_{ij} = \underline{t}_{ij} + t_{ij}^{\epsilon}, \quad t_{ij}^{\epsilon} \sim \mathcal{N}(0, \tau_{ij}^{-1} I_d), \tag{5.3b}$$

where $\underline{x}_{ij} = \underline{x}_i^{-1} \underline{x}_j = (\underline{t}_{ij}, \underline{R}_{ij})$ is the true relative transformation from x_i to x_j. Under this noise model, a maximum likelihood estimate $x^* \in \mathrm{SE}(d)^n$ for the latent states $x_1, \ldots, x_n$ is obtained as a minimizer of the following problem [118]:

Problem 4 (Maximum-likelihood estimation for $\mathrm{SE}(d)$ synchronization)**.**

$$\min_{\substack{t_i \in \mathbb{R}^d \\ R_i \in \mathrm{SO}(d)}} \sum_{(i,j)\in\vec{\mathcal{E}}} \kappa_{ij} \|R_j - R_i \tilde{R}_{ij}\|_F^2 + \tau_{ij} \|t_j - t_i - R_i \tilde{t}_{ij}\|_2^2. \tag{5.4}$$

Note that under these modeling assumptions, both pose-graph optimization and rotation averaging can be written as particular instances of the following general optimization problem:

Problem 5 (Quadratic minimization over $\mathrm{SO}(d)^n$)**.**

$$p^* = \min_{R \in \mathrm{SO}(d)^n} \mathrm{tr}(\tilde{Q} R^{\mathsf{T}} R), \tag{5.5}$$

where $\tilde{Q} \in \mathrm{Sym}(dn),\ \tilde{Q} \succeq 0$.

Algorithm 1 Spectral initialization procedure

Input: The data matrix $\tilde{Q}$ from (RA) or (PGO)
Output: A spectral initialization $R^{(0)}$

function SpectralInitialization($\tilde{Q}$)
  Compute orthogonal set of eigenvectors Φ corresponding to the d smallest eigenvalues
  of $\tilde{Q}$. ▷ Solve Problem 6.
  for $i = 1, \ldots, n$ **do**
    Set $R_i^{(0)} \leftarrow \Pi_{\mathcal{S}}(\Phi_i)$, where Φ_i is the i-th $(d \times d)$ block of Φ. ▷ Definition 2
  end for
  return $R^{(0)}$
end function

Specifically, the problems of rotation averaging (RA) and pose-graph optimization (PGO) in Problems 3 and 4, respectively, can be parameterized in terms of the following data matrices:

$$\tilde{Q} = L(\tilde{G}^{\rho}), \tag{RA}$$

$$\tilde{Q} = L(\tilde{G}^{\rho}) + \tilde{Q}^{\tau}, \tag{PGO}$$

where $L(\tilde{G}^{\rho})$ is the *rotation connection Laplacian* and $\tilde{Q}^{\tau}$ is a data matrix comprised of translation measurements. For the purposes of the approach presented in this chapter, the specific structure of $\tilde{Q}$ is not important; we require only that in the *noiseless case*, where $\tilde{Q} = \underline{Q}$, we have $\underline{R}^{\mathsf{T}} \in \ker(\underline{Q})$, where $\underline{R}$ is the set of (latent) ground-truth rotational states, and $L(\underline{G}^{\rho}) \succeq 0$ and $\underline{Q}^{\tau} \succeq 0$ (see [118, Appendix C.3] for a detailed analysis of the noiseless case). The interested reader may refer to Appendix A.1 for a complete description of these data matrices.

5.2 Spectral methods for initialization

The nonconvexity of the $\mathrm{SO}(d)$ constraint renders Problem 5 computationally hard to solve in general. However, we can generate a tractable *spectral relaxation* of Problem 5 by relaxing the $\mathrm{SO}(d)$ constraint as follows:

Problem 6 (Spectral Relaxation of Problem 5)**.**

$$\begin{aligned} p_{\mathrm{S}}^* = \min_{Y \in \mathbb{R}^{d \times dn}} \operatorname{tr}(\tilde{Q} Y^{\mathsf{T}} Y) \\ \text{s.t. } Y Y^{\mathsf{T}} = n I_d. \end{aligned} \tag{5.7}$$

Here, the $\mathrm{SO}(d)$ constraint on each $(d \times d)$ block of the variable Y has been replaced by the (weaker) constraint that $YY^{\mathsf{T}} = nI_d$, i.e. the matrix Y is comprised of d orthogonal rows of norm $\sqrt{n}$. While the relaxed constraints in (5.7) are still quadratic and nonconvex, in Appendix A.2.1 we prove that a feasible point Y is a (*global*) minimizer of Problem 6 if and only if its rows are comprised of d pairwise orthogonal (and appropriately scaled) eigenvectors corresponding to the minimum d eigenvalues of $\tilde{Q}$. Therefore, one can recover an optimizer Y^* of Problem 6 via a simple eigenvector computation.[3]

For the noiseless problem parameterized by $\underline{Q}$, the relaxation in Problem 6 is exact in the sense that $\underline{R} = GY^*$ for some $G \in \mathrm{O}(d)$.[4] This follows from the fact that, by construction, the ground truth rotations $\underline{R}^{\mathsf{T}}$ lie in $\ker(\underline{Q})$,[5] and $\underline{R}\underline{R}^{\mathsf{T}} = nI_d$ since $\underline{R} \in \mathrm{SO}(d)^n$. Likewise, since $\underline{R}$ is a minimizer of the relaxed problem *and* is in the feasible set for the Problem 5, it is also a minimizer for Problem 5. In general, however, we do not expect such a nice correspondence to hold. Indeed, a minimizer of Problem 6 need not even be *feasible* for Problem 5, since the former is obtained from the latter by relaxing constraints. Therefore, we must in general *round* the estimate provided by the spectral relaxation to obtain an approximate solution $R^{(0)} \in \mathrm{SO}(d)^n$ in the feasible set of Problem 5. The following definition makes this precise.

Definition 2 (Projection onto $\mathrm{SO}(d)$)**.** For $X \in \mathbb{R}^{d \times d}$, the projection $\Pi_{\mathcal{S}}(X)$ of X onto $\mathrm{SO}(d)$ is by definition a minimizer of the following:

$$\min_{G \in \mathrm{SO}(d)} \|X - G\|_F. \tag{5.8}$$

[3]This justifies our referring to Problem 6 as a "spectral" relaxation of Problem 5.

[4]The spectral relaxation in Problem 6, like Problem 5, admits infinitely many solutions: if Y^* is a minimizer of Problem 6, then any $GY^*, G \in \mathrm{O}(d)$ is also a minimizer.

[5]We refer the reader to [118, Appendix C.3] for detailed analysis of the noiseless case.

A minimizer for this problem is given in closed-form as [62, 139]:

$$\Pi_{\mathcal{S}}(X) = U\Xi V^{\mathsf{T}}. \tag{5.9}$$

where $X = U\Sigma V^{\mathsf{T}}$ is a singular value decomposition, and Ξ is the matrix:

$$\Xi = \operatorname{Diag}\left(1, 1, \det(UV^{\mathsf{T}})\right). \tag{5.10}$$

In the context of subsequent derivations, it will be convenient to "overload" this rounding operation to $Y \in \mathbb{R}^{d\times dn}$ as follows:

$$\Pi_{\mathcal{S}}(Y) = \left(\Pi_{\mathcal{S}}(Y_1), \ldots, \Pi_{\mathcal{S}}(Y_n)\right), \tag{5.11}$$

where $Y_i \in \mathbb{R}^{d\times d}$ are the n blocks of Y.

Therefore, we can obtain an approximate solution to Problem 5 from a minimizer Y^* of the relaxation in Problem 6 as $R^{(0)} \triangleq \Pi_{\mathcal{S}}(Y^*)$. Our overall spectral initialization procedure is summarized in Algorithm 1.

5.3 Main results

This section presents our main results, which are three-fold: First, we provide a bound on the error of our spectral initialization $R^{(0)}$ with respect to the ground-truth rotations $\underline{R}$. Second, we give a new bound on the error of *globally optimal* solutions R^* with respect to $\underline{R}$: this bound differs from prior work (e.g. Preskitt [110], Rosen et al. [118]) in that it is defined with respect to the orbit distance $d_{\mathcal{S}}$ on $\mathrm{SO}(d)^n$. Previous work used the orbit distance $d_{\mathcal{O}}$ on $\mathrm{O}(d)^n$ due to mathematical convenience; however, the estimation error one considers in application is actually over $\mathrm{SO}(d)^n$, since this is the domain on which the estimation problem is defined. Combining these results, we obtain an upper bound on the $\mathrm{SO}(d)$ orbit distance between an initial guess $R^{(0)}$ and a globally optimal solution R^*. Our analysis gives direct control over the mutual deviation between the three quantities of interest: $R^{(0)}$, R^*, and $\underline{R}$ as a function of

the noise magnitude. We conclude with additional remarks about computing these bounds for practical SLAM scenarios and a few straightforward adaptations of the main results.

Recall from Problem 6 that an estimate Φ is a minimizer of Problem 6 if and only if it is composed of a (suitably scaled) orthogonal set of eigenvectors corresponding to the minimum d eigenvalues of $\tilde{Q}$, and that in the noiseless case a minimizer is given by $\underline{R}$. Since a spectral initialization $R^{(0)}$ is obtained as the projection of a solution Φ of Problem 6 onto $\mathrm{SO}(d)^n$, we can bound its estimation error by first bounding the deviation of Φ from $\underline{R}$, then bounding the additional error incurred by projecting onto $\mathrm{SO}(d)^n$.

We will begin our presentation of the main results by giving a bound on the deviation of a solution Φ of Problem 6 from the ground truth $\underline{R}$ via the Davis-Kahan Theorem [147], a classical result relating the perturbation of a matrix's eigenvectors under a symmetric perturbation to the magnitude of that perturbation. Here, we take $\underline{Q}$ to be the matrix under consideration, and define the perturbation $\Delta Q \triangleq \tilde{Q} - \underline{Q}$. The following lemma, which we prove in Appendix A.2.2, gives the desired characterization:

Lemma 3. *Let Φ be a minimizer of Problem 6 and $\underline{R}$ be the corresponding ground truth rotations. Then:*

$$d_{\mathcal{O}}\left(\underline{R}, \Phi\right) \leq \frac{2\sqrt{2dn}\|\Delta Q\|_2}{\lambda_{d+1}(\underline{Q})}. \tag{5.12}$$

Lemma 3 provides control over the deviation of an "unrounded" solution Φ from the ground truth $\underline{R}$. The second technical ingredient we require is the following simple bound controlling the maximum distance between a matrix X and its projection $\Pi_{\mathcal{S}}(X)$ onto $\mathrm{SO}(d)$:

Lemma 4. *Let $X \in \mathbb{R}^{d\times d}$ and $R \in \mathrm{SO}(d)$. Then:*

$$\|\Pi_{\mathcal{S}}(X) - R\|_F \leq 2\|X - R\|_F. \tag{5.13}$$

Proof.

$$\|\Pi_{\mathcal{S}}(X) - R\|_F = \|\Pi_{\mathcal{S}}(X) - X + X - R\|_F \tag{5.14}$$

$$\leq \|\Pi_{\mathcal{S}}(X) - X\|_F + \|X - R\|_F \tag{5.15}$$

$$\leq 2\|X - R\|_F, \tag{5.16}$$

where the last inequality follows from the fact that $\Pi_{\mathcal{S}}(X)$ is a minimizer over $\mathrm{SO}(d)$ of the distance to X with respect to the Frobenius norm, and that, by hypothesis, $R \in \mathrm{SO}(d)$. □

Lemma 4 provides a straightforward approach for converting a bound expressed in the $\mathrm{O}(d)^n$ orbit distance to one expressed in the $\mathrm{SO}(d)^n$ orbit distance. In turn, we obtain the following theorem, which we prove in Appendix A.3.1:

Theorem 5. *Let Φ be a minimizer of Problem 6 and $R^{(0)} = \Pi_{\mathcal{S}}(\Phi) \in \mathrm{SO}(d)^n$ be the corresponding spectral initialization. Finally, let $\underline{R} \in \mathrm{SO}(d)^n$ be the set of ground truth rotations in Problem 5. Then the estimation error of $R^{(0)}$ satisfies:*

$$d_{\mathcal{S}}(\underline{R}, R^{(0)}) \leq \frac{4\sqrt{2dn}\|\Delta Q\|_2}{\lambda_{d+1}(\underline{Q})}. \tag{5.17}$$

The bound (5.17) gives a direct (linear) relationship between the magnitude of the perturbation ΔQ and the worst-case error of a spectral estimate. Moreover, Theorem 5 implies that $d_{\mathcal{S}}(\underline{R}, R^{(0)}) \to 0$ as $\Delta Q \to 0$. That is to say, as the measurements approach their noiseless counterparts, our spectral estimate approaches the ground truth.

Next, we address the issue of furnishing a bound on $d_{\mathcal{S}}(\underline{R}, R^*)$. The following theorem, which we prove in Appendix A.3.2, gives the desired result:

Theorem 6 (Bounding the estimation error for R^*)**.** *Let R^* be a minimizer of Problem 5 and $\underline{R}$ be the set of ground-truth rotations. Then the estimation error of R^* satisfies:*

$$d_{\mathcal{S}}(\underline{R}, R^*) \leq \frac{8\sqrt{dn}\|\Delta Q\|_2}{\lambda_{d+1}(\underline{Q})}. \tag{5.18}$$

To the best of our knowledge, Theorem 6 is the first result to appear in the literature that directly controls the estimation error of the maximum likelihood estimate R^* over $\mathrm{SO}(d)^n$ specifically. Prior work considered the estimation error over $\mathrm{O}(d)^n$ [6, 80, 118]. In our application, however, we are specifically concerned with the estimation error over $\mathrm{SO}(d)^n$; as one can see from inspection, this is the domain on which Problem 5 is defined. Thus, the $\mathrm{SO}(d)^n$ orbit distance corresponds to the actual error one would obtain in practice.

While Theorem 5 establishes error bounds for the spectral estimator, when viewed as an *initialization method*, the distance between the initial guess $R^{(0)}$ and the globally optimal solution is the primary concern. A corollary to Theorems 5 and 6, allows us to control $d_{\mathcal{S}}(R^{(0)}, R^*)$ in terms of the noise matrix ΔQ. We have:

Corollary 7. *The orbit distance between the initialization $R^{(0)}$ and a globally optimal solution R^* satisfies:*

$$d_{\mathcal{S}}(R^{(0)}, R^*) \leq \frac{(8 + 4\sqrt{2})\sqrt{dn}\|\Delta Q\|_2}{\lambda_{d+1}(\underline{Q})}. \tag{5.19}$$

These bounds provide a clear relationship between the spectral properties of $\underline{Q}$ and ΔQ and the deviation between a spectral estimator $R^{(0)}$, maximum likelihood estimator R^*, and the ground-truth $\underline{R}$. An important consequence of these bounds is that as $\Delta Q \to 0$, we have (at least) linear convergence of the estimation error for *both* the spectral estimator and the maximum likelihood estimator to zero. This, in turn, guarantees that $\Delta Q \to 0$ implies $R^*, R^{(0)} \to \underline{R}$ (up to symmetry), which is what we would expect.

In practice, however, we do not have access to $\underline{Q}$. This presents some difficulty in the computation of ΔQ and $\lambda_{d+1}(\underline{Q})$. Fortunately, the noiseless rotation matrices admit a description in terms of quantities that *are* typically assumed to be known. In particular, we have [118, Lemma 8]:

$$\lambda_{d+1}(L(\underline{G}^\rho)) = \lambda_2(L(W^\rho)), \tag{5.20}$$

where $L(W^\rho)$ is the Laplacian of the rotational weight graph. Now, $L(W^\rho)$ depends only on the concentration parameters κ_{ij} attached to each edge, which are generally assumed to be known *a priori* from the noise models (5.1) and (5.3). In the rotation averaging case, we have $\underline{Q} = L(\underline{G}^\rho)$, and therefore the denominator $\lambda_{d+1}(\underline{Q})$ is readily available as $\lambda_2(L(W^\rho))$, the algebraic connectivity of the rotational weight Laplacian.

In the case of pose-graph SLAM, where the matrix $\underline{Q}$ contains the translational terms $\underline{Q}^\tau$, we can use the fact that $\underline{Q} = L(\underline{G}^\rho) + \underline{Q}^\tau$ is the sum of positive-semidefinite matrices (see Rosen et al. [118, Appendix C.3]), so $\lambda_{d+1}(L(\underline{G}^\rho)) \leq \lambda_{d+1}(L(\underline{G}^\rho) + \underline{Q}^\tau) = \lambda_{d+1}(\underline{Q})$. In particular, the (weaker) bounds obtained by substituting $\lambda_{d+1}(\underline{Q})$ with $\lambda_{d+1}(L(\underline{G}^\rho))$ in (5.17) and (5.18) hold.

Moreover, a common SLAM initialization technique is that of *rotation only initialization* – i.e., to compute the initializer $R^{(0)}$ using *only* the relative rotation measurements [30]. This can have computational advantages in practice since $L(\tilde{G}^\rho)$ is generally *sparse*; the same cannot be said for the pose-graph SLAM data matrix $\tilde{Q}$, as it arises via analytic elimination of the translational states, in which case the resulting data matrix $\tilde{Q}$ is formed as a (dense) generalized Schur complement [118, Appendix B]. Interestingly, for pose-graph SLAM, a spectral initialization $R^{(0)}$ computed using the eigenvectors of $L(\tilde{G}^\rho)$ (i.e. ignoring $\tilde{Q}^\tau$) attains the bound:

$$d_{\mathcal{S}}(\underline{R}, R^{(0)}) \leq \frac{4\sqrt{2dn}\|\Delta L(\tilde{G}^\rho)\|_2}{\lambda_{d+1}(L(\underline{G}^\rho))}. \tag{5.21}$$

This bound holds by the same reasoning as Theorem 5, but with the consideration that $\underline{R}^{\mathsf{T}} \in \ker(L(\underline{G}^\rho))$.

As a final consideration, typically we do not have access to ΔQ (if we did, we could recover the true data matrix $\underline{Q}$ as $\tilde{Q} - \Delta Q$). In consequence, we need a method to estimate the likely magnitude of the noise in a given application. One way of achieving this is via simulation from the generative model, given a measurement network and associated measurement precisions.[6] This, in turn, gives a sample set

[6]Simulating measurements in the case of pose-graph SLAM requires knowledge of the ground-truth translation measurement scale, which is typically also unavailable in practice. However, the *rotation-only* initialization bound (5.21) applies in general and depends only upon the rotation

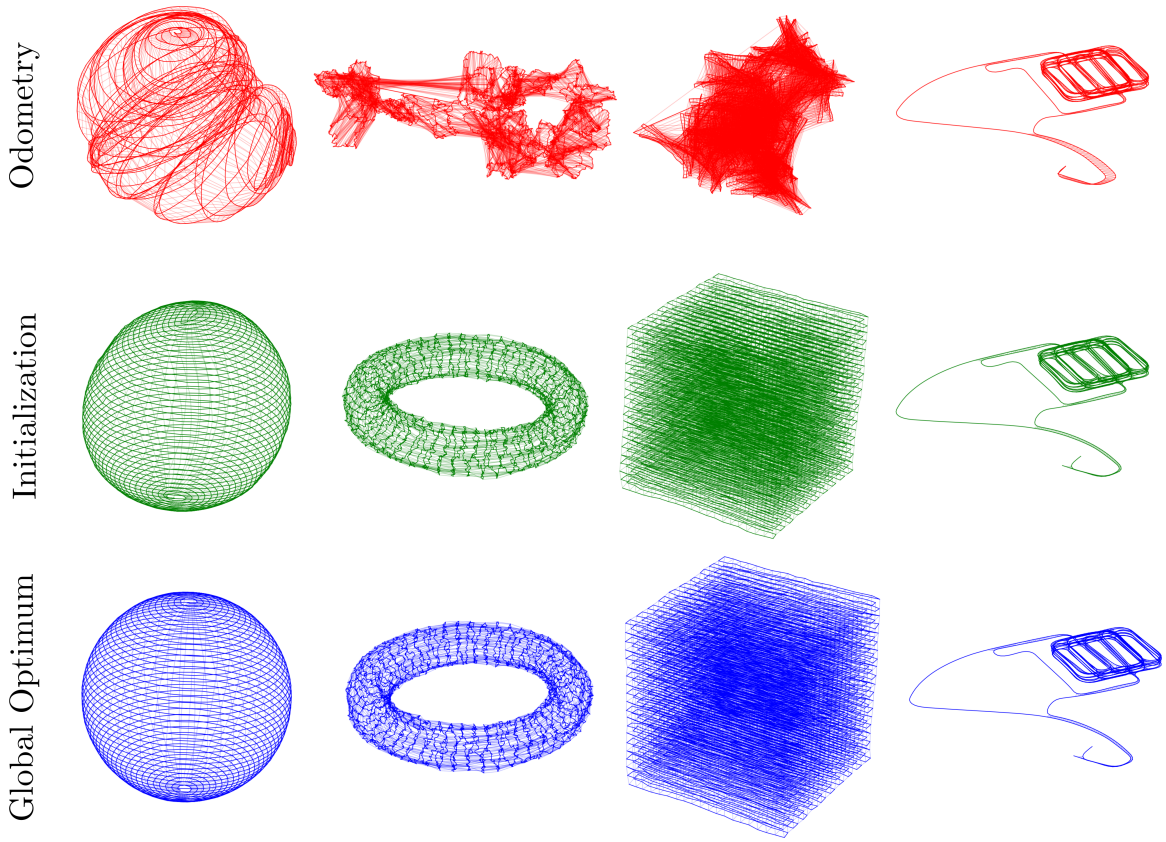


Figure 5-2: **Spectral relaxation produces high-quality initializations.** Qualitative comparison with the globally optimal solution suggests that the spectral relaxation produces estimates that are very close to optimal for a variety of SLAM benchmark datasets. The corresponding *quantitative* comparison is given in Table 5.1.

from a *distribution* over the bounds (5.17), (5.18), and (5.19).

5.4 Experimental results

In this section, we compare the bounds in Theorem 5 to the actual estimation error incurred by the spectral initialization and globally optimal pose-graph SLAM solutions on a variety of simulated problem instances, as well as benchmark SLAM problems. In Section 5.4.1 we construct synthetic pose-graph SLAM scenarios for which the ground-truth poses are known. Since the bounds we have presented depend upon

measurements, which can be simulated to produce an empirical distribution over the spectral norm of the perturbation matrix.

knowledge of the noise magnitude $\|\Delta Q\|_2$ and the spectral gap of the *true* data matrix $\underline{Q}$, which are unknown in practice for pose-graph SLAM, our first set of empirical results shed light on the behavior of these worst-case bounds (as well as the *actual* error realized by different estimators) as we vary the noise parameters controlling the generative model (5.3). In Section 5.4.2, we evaluate the performance of spectral relaxation as a practical initialization method in the context of 3D pose-graph SLAM applications. We show that, consistent with our results on synthetic data, the spectral initialization method offers high-quality initial solutions for pose-graph optimization, and in particular, that the inclusion of translational measurements significantly improves the quality of the spectral estimator versus the common approach of using exclusively rotational measurements.

The spectral initialization method was implemented in C++ using Spectra to efficiently solve large-scale eigenvalue problems [112]. Computation of the bounds in Section 5.4.1 was performed in MATLAB using `eigs`. All experiments were performed on a laptop with a 2.2 GHz Intel i7 CPU. Where (verified) globally optimal solutions were needed, we used the C++ implementation of SE-Sync [118]. We also provide results using the well-known *chordal* initialization method [90], which relaxes the feasible set of Problem 5 to $\mathbb{R}^{d\times dn}$, with the constraint that $R_1^{(0)} = I_d$, for which the solution can be obtained by solving a linear system.

5.4.1 Evaluation on synthetic data

The bounds presented in our analysis depend upon knowledge of the noise magnitude $\|\Delta Q\|_2$, which is unknown in practice. In light of this fact, we examine empirically the behavior of the bounds as a function of the noise parameters using synthetic data. Specifically, we use the Cube dataset [29, 118], which consists of a set of vertices (poses) organized in a three-dimensional cube, with s vertices per dimension. Consecutive poses have an "odometry" edge between them, and loop closures are sampled randomly from the remaining edges with probability p_{LC}. Measurements are generated by randomly sampling from the generative model (5.3) with fixed noise parameters κ and τ for all measurements. Beyond providing access to the ground-

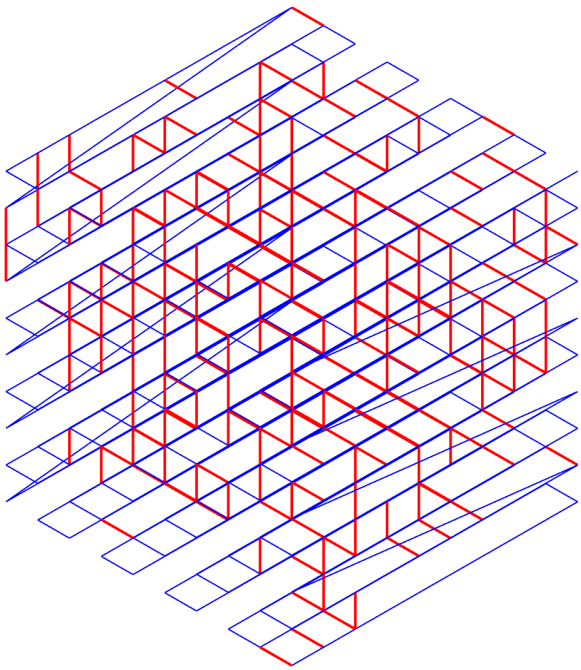


Figure 5-3: **Cube experiments.** Example ground-truth realization of a synthetic Cube dataset [29, 118] with $s = 10$ vertices per side and $p_{LC} = 0.1$. The robot's trajectory is shown in blue with loop closures shown in red.

truth rotations, this setup allows us to compare the worst-case bounds with empirical performance in noise regimes well outside the range typically encountered in real SLAM scenarios. A sample configuration for the Cube dataset is provided in Figure 5-3.

Influence of noise parameters on performance bounds: In Figure 5-4, we study the performance of the spectral initialization approach across a variety of noise configurations. In each case, we provide the worst-case bounds (5.17) and (5.21) along with the empirical error of the different estimators under consideration. In Figure 5-4a, we sample Cube problem instances with logarithmically spaced values of κ while fixing the other parameters: $\tau = 150$ (corresponding to an expected RMS error of 0.14 m), $p_{LC} = 0.2$, and $s = 10$. In Fig. 5-4b, we fix $\kappa = 10^5$ (corresponding to an expected RMS error of approximately 0.1°), $p_{LC} = 0.25$ and $s = 10$ and sample problem instances with logarithmically spaced translation concentration parameter τ. In Fig. 5-4c, we fix $\kappa = 10^5$, $\tau = 150$, $s = 10$ and vary p_{LC} from 0 to 1.

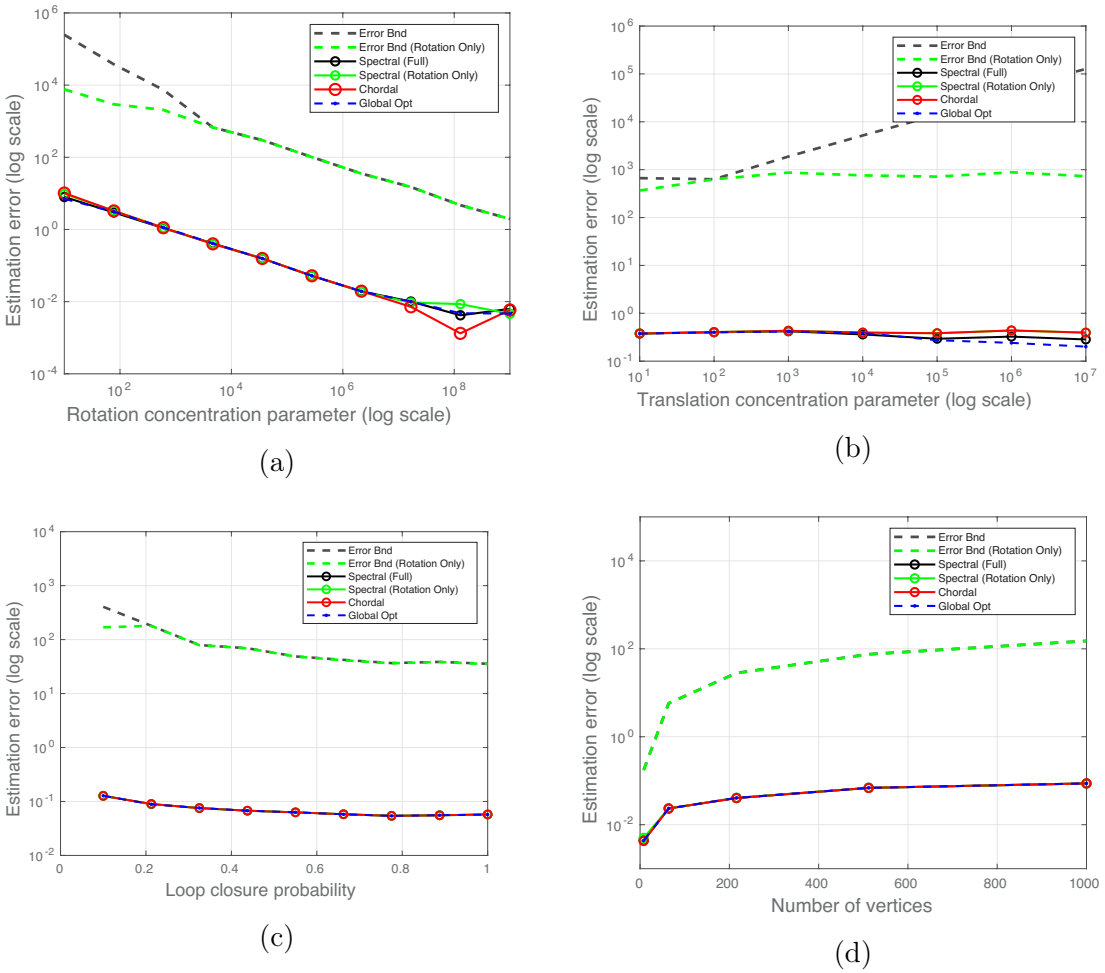


Figure 5-4: **Influence of dataset parameters on the performance bounds for the Cube experiments.** We examine empirically the change in the theoretical bounds (5.17) and (5.21) as well as the estimation error of several pose-graph optimization estimates while varying (a) the rotation concentration parameter κ, (b) the translation concentration parameter τ, (c) the probability of a loop closure p_{LC}, (d) the number of vertices s^3.

Across a wide range of concentration parameters, the spectral initializations attain very similar error to the global optimizer.[7] In particular, their error often improves upon the worst-case bounds (5.17) and (5.21) by orders of magnitude. This is consistent with earlier observations of qualitatively similar bounds for phase synchronization [110]. Moreover, in applications of rotation averaging and pose-graph optimization, previous work has shown that the maximum likelihood estimator often attains expected error close to the Cramér-Rao *lower bound* (see [17] for rotation averaging and [33] for pose-graph optimization). The behavior of the bounds when varying the translation concentration parameter in Figure 5-4b is counterintuitive: while the spectral estimator improves with increasing τ, the bound suggests the opposite worst-case behavior. It seems the form of the bounds we derive (including the translational terms) is not refined enough to capture this behavior, and this certainly warrants further investigation. With this exception, the bounds seem to accurately capture the behavior of the actual estimation error, though they appear to be quite loose with respect to the empirical performance attained by all of the methods. This suggests that, while the bounds we have produced identify key quantities of interest for accurate state estimation and have "reasonable" asymptotic performance (i.e. linear convergence as $\|\Delta Q\|_2 \to 0$), there is significant room for improvement in the bounds themselves.

Dependence on problem dimensionality: Due to the explicit appearance of the problem dimension n in the bounds (5.17), (5.18), and (5.19), it is interesting to consider how the number of rotations to be estimated affects these bounds. In Figure 5-4d, we fix $\kappa = 10^5$, $\tau = 150$, $p_{LC} = 0.2$ and vary the number of vertices in the Cube dataset. Indeed, we find that the behavior of the worst-case bounds suggests an unfavorable scaling in the problem dimension: at $s^3 = 8$ vertices, the worst-case bound overestimates the true error by approximately an order of magnitude; at $s^3 = 1000$, it overestimates the true error by approximately 3 orders of magnitude. It

[7] R^* is the maximum likelihood estimator–the optimal point estimate given the data. Since there is noise in the data, it is conceivable that the maximum likelihood estimate might actually be farther away from the ground truth than a "suboptimal" estimate, which we observe in Fig. 5-4a.

is unclear, at present, whether it is possible to remove this dependence on the problem dimension. A more sophisticated analysis considering the specific structure of these matrices (as defined in Appendix A.1) may yield more refined bounds.

5.4.2 Evaluation on standard SLAM benchmark datasets

In these experiments, we consider evaluation of the spectral initialization method on several standard SLAM benchmark datasets. Figure 5-2 provides a qualitative comparison of three techniques for initialization: odometry only (i.e. composing measurements between consecutive poses), the proposed spectral initialization approach, and the globally optimal solution. We observe that spectral initialization provides solutions that visually resemble the globally optimal solution. Table 5.1 gives our quantitative results. For each method, we provide the computation time, objective value, and number of iterations required for a Riemannian trust-region (RTR) optimization method to converge to a critical point when using that initialization. With the exception of odometry-only initialization, all of the methods considered enabled the recovery of (verifiably) globally optimal solutions; that is, these initialization methods coupled with standard *local* optimization techniques recovered globally optimal solutions *without* the need to explicitly solve a large-scale semidefinite program.

Both of the spectral methods (using the "full" pose-graph optimization data matrix $\tilde{Q}$ and the "rotation only" version using only $L(\tilde{G}^\rho)$) provide estimates competitive with the state-of-the-art chordal initialization method, generally attaining near-optimal objective values.[8] Interestingly, in their work, Moreira et al. [95] found that the rotation-only spectral estimator attains a higher cost on the *Sphere* dataset than alternative methods, as we do here; however, when we include the translation measurements, we find that this discrepancy disappears. Similarly, the chordal estimator also performs well on this dataset, despite the fact that, like the rotation-only spectral initialization, it does not make use of translational measurements.

[8] Our current implementation is aimed at recovering high-precision eigenvector estimates, rather than expedient computation. Despite this, spectral initialization is often faster than the chordal approach, though occasionally this added precision leads to longer computation times than would be necessary to obtain a good estimate, e.g. on the *Garage* dataset.

Dataset		Odom.	Chord.	**Spec. (RO)**	**Spec.**	Glob. Opt.
Sphere	Iter	65	6	8	4	1687
	Cost	1.14×10^9	1971.17	5594.19	1742.75	
	Time (s)	-	0.707	0.602	0.779	
Torus	Iter	32	5	5	4	24227
	Cost	3.87×10^8	24669.2	25833.2	24272.7	
	Time (s)	-	1.316	1.501	1.199	
Grid	Iter	30	6	6	4	84320
	Cost	1.97×10^{10}	87252	86966.1	84486.4	
	Time (s)	-	8.747	18.806	0.25	
Garage	Iter	1028	3	4	4	1.26
	Cost	2.31×10^9	1.42	3.215	2.7	
	Time (s)	-	0.201	0.136	25.7	

Table 5.1: **Standard SLAM benchmarks** Objective value (cost) attained and computation time required for each initialization method on several SLAM benchmarks. We also report the number of iterations (Iter.) required for a Riemannian trust-region optimization method to converge to a critical point. Note that the reported computation time is only the time required to compute the initialization. Proposed approaches are **bold**.

5.5 Summary

In this chapter we presented the first initialization methods equipped with *explicit performance guarantees* adapted to the problems of pose-graph SLAM and rotation averaging. Our approach is based upon a simple *spectral relaxation* of the estimation problem, the form of which permits us to apply eigenvector perturbation bounds to control the distance from our initialization to *both* the (latent) ground-truth *and* the global minimizer of the estimation problem (the *maximum likelihood* estimate) as a function of the measurement noise. Consistent with recent complementary work on information-theoretic aspects [17, 33, 72] and global optimization methods [41, 51, 118] for SLAM and RA, our bounds highlight the central role that spectral properties of the measurement network[9] play in controlling the accuracy of SLAM and RA solutions. Finally, we show experimentally that our spectral estimator is very effective in practice, producing initializations of comparable or superior quality at lower computational cost compared to existing state-of-the-art techniques.

[9]Specifically, the smallest nonzero eigenvalue $\lambda_{d+1}(Q)$, which can be thought of as a generalization of the algebraic connectivity of the classical graph Laplacian.

Chapter 6

Spectral measurement sparsification for pose-graph SLAM

However, as we aim to scale SLAM algorithms to the setting of "lifelong" autonomy, particularly on compute- or memory-limited platforms, a robot must be able to determine *what information should be kept, and what can safely be forgotten* [119]. In particular, in the setting of graph-based SLAM and rotation averaging, the number of edges in a measurement graph determines both the memory required to store a robot's observations as well as the computation time of algorithms employed for state estimation using this measurement graph.

While there has been substantial work on the topic of measurement pruning (or *sparsification*) in lifelong SLAM (e.g. [22, 23, 68, 76, 77]), most existing methods rely on heuristics for sparsification whereby little can be said about the quality of the statistical estimates obtained from the sparsified graph versus the original. Recent work on performance guarantees in the setting of pose-graph SLAM and rotation averaging identified the spectral properties—specifically the *algebraic connectivity* (also known as the *Fiedler value*)—of the measurement graphs encountered in these problems to be central objects of interest, controlling not just the best possible expected performance (per earlier work on Cramér-Rao bounds [17, 33, 72]), but also the *worst-case* error of estimators [48, 118]. These observations suggest the algebraic connectivity as a natural measure of graph quality for assessing SLAM graphs. This motivates our

use of the algebraic connectivity as an *objective* in formulating the graph sparsification problem.

Specifically, we propose a spectral approach to pose graph sparsification which maximizes the algebraic connectivity of the measurement graph subject to a constraint on the number of allowed edges.[1] As we discuss, this corresponds to *E-optimal* design in the setting of pose-graph SLAM [111]. This specific problem turns out to be an instance of the *maximum algebraic connectivity augmentation* problem, which is NP-Hard [96]. To address this, we propose to solve a computationally tractable relaxation and *round* solutions obtained to the relaxed problem to approximate feasible solutions of the original problem. Relaxations of this form have been considered previously; in particular Ghosh and Boyd [58] developed a semidefinite program relaxation to solve problems of the form we consider. However, these techniques do not scale to the size of typical problems encountered in graph-based SLAM. To this end, we propose a first-order optimization approach that we show is practically fast for even quite large SLAM problems. Moreover, we show that the *dual* to our relaxation provides tractable, high-quality bounds on the suboptimality of the solutions we provide *with respect to the original problem.*

In summary, we present an approach for pose graph sparsification by maximizing the *algebraic connectivity* of the measurement graph, a key quantity which has been shown to control the estimation error of pose-graph SLAM solutions. Our method, based on convex relaxation, is *simple* and *computationally inexpensive*, and admits formal *post hoc* performance guarantees on the quality of the solutions it provides. In experiments on several benchmark pose-graph SLAM datasets, we show that our approach quickly produces high-quality sparsification results which retain the connectivity of the graph and better preserve the quality of SLAM solutions compared to a baseline which does not consider graph connectivity.

[1] Our method is related to, but should not be confused with, *spectral sparsification* [130]. Similar to spectral sparsification, we aim to sparsify graphs in a way that preserves their spectral properties. However, our method differs in that we focus only on the algebraic connectivity, whereas traditionally spectral sparsification aims to preserve the entire graph Laplacian spectrum.

6.1 Problem formulation

We consider graph sparsification in the setting of pose-graph SLAM. Recall that pose-graph SLAM is the problem of estimating n unknown values $x_i, \ldots, x_n \in \mathrm{SE}(d)$ given a subset of measurements of their pairwise relative transforms $\tilde{x}_{ij}$. We adopt the same generative model for rotation and translation measurements as in the previous chapter (given in eq. (5.3)). In turn, we obtain the maximum-likelihood estimation procedure given in Problem 4.

In the last chapter, we showed that the smallest (nonzero) eigenvalue of the *rotational weight graph Laplacian* $L(W^\rho)$ (eq. (A.1b)) controls the worst-case error of solutions to Problem 4; this is the algebraic connectivity (or Fiedler value) of the graph having nodes in correspondence with robot poses x_i and edge weights equal to each κ_{ij}. The corresponding eigenvector attaining this value is called the *Fiedler vector*. The method we present in this chapter is applicable to any graph $\mathcal{G}$ with Laplacian L, but because of the specific connections between the rotational weight graph and performance of estimators for SLAM we will take $L = L(W^\rho)$ in all applications we consider here.

The Laplacian of a graph has several well-known properties that we will use here. The Laplacian L of a graph can be written as a sum of the Laplacians of the subgraphs induced by each of its edges. A Laplacian is always positive-semidefinite, and the "all ones" vector $\mathbb{1}$ of length n is always in its kernel. Finally, a graph has positive algebraic connectivity $\lambda_2(L) > 0$ if and only if it is connected.[2]

It will be convenient to partition the edges as $\mathcal{E} = \mathcal{E}^o \cup \mathcal{E}^c,\ \mathcal{E}^o \cap \mathcal{E}^c = \emptyset$ into a *fixed* set of edges $\mathcal{E}^o$ and a set of m *candidate* edges $\mathcal{E}^c$, and where L^o and L^c are the Laplacians of the subgraphs induced by $\mathcal{E}^o$ and $\mathcal{E}^c$. For our purposes, the subgraph induced by $\mathcal{E}^o$ on $\mathcal{V}$ will typically be constructed from sequential odometric measurements (therefore, $|\mathcal{E}^o| = n - 1$), but this is not a requirement of our general approach.[3] It will be helpful in the subsequent presentation to "overload" the defini-

[2]More specifically, the number of zero eigenvalues of a Laplacian is equal to the number of connected components of its corresponding graph.

[3]In particular, to apply our approach we should select L^o and K to guarantee that the feasible set for Problem 7 contains at least one tree. Then, it is clear that the optimization in Problem 7

tion of $L(W^\rho)$. Specifically, let $L : \mathbb{R}^m \to \mathbb{S}^n_+$ be the affine map constructing the total graph Laplacian from a weighted combination of edges in $\mathcal{E}^c$:

$$L(\omega) \triangleq L^o + \sum_{k=1}^{m} \omega_k L^c_k, \tag{6.1}$$

where L^c_k is the Laplacian of the subgraph induced by edge $e_k = \{i_k, j_k\}$ of $\mathcal{E}^c$. Our goal in this work will be to identify a subset of $\mathcal{E}^* \subseteq \mathcal{E}^c$ of fixed size $|\mathcal{E}^*| = K$ (equivalently, the edge selection ω), which maximizes the algebraic connectivity $\lambda_2(L(\omega))$. This corresponds to the following optimization problem:

Problem 7 (Algebraic connectivity maximization)**.**

$$\begin{aligned} p^* = \max_{\omega \in \{0,1\}^m} \ & \lambda_2(L(\omega)) \\ |\omega| = K. & \end{aligned} \tag{6.2}$$

6.2 Approach

Problem 7 is a variant of the *maximum algebraic connectivity augmentation problem*, which is NP-Hard [96]. The difficulty of Problem 7 stems, in particular, from the integrality constraint on the elements of ω. Consequently, our general approach will be to solve a simpler problem obtained by relaxing the integrality constraints of Problem 7, and, if necessary, *rounding* the solution to the relaxed problem to a solution in the feasible set of Problem 7. In particular, we consider the following *Boolean relaxation* of Problem 7:

Problem 8 (Boolean Relaxation of Problem 7)**.**

$$\begin{aligned} \max_{\omega \in [0,1]^m} \ & \lambda_2(L(\omega)) \\ \mathbb{1}^\mathsf{T}\omega = K. & \end{aligned} \tag{6.3}$$

will always return a connected graph, since $\lambda_2(L(\omega)) > 0$ if and only if the corresponding graph is connected. Note that this condition is always easy to arrange: for example, we can start with L^o a tree, as we do here, or (even more simply) take L^o to be the zero matrix and simply take $K \geq n-1$.

Algorithm 2 MAC Algorithm

Input: An initial iterate ω
Output: An approximate solution to Problem 7

function MAC(ω)
    $\omega \leftarrow$FrankWolfeAC(ω) ▷ Solve Problem 8
    return $\Pi(\omega)$ ▷ Round solution; eq. (6.7)
end function

Algorithm 3 Frank-Wolfe Method for Problem 8

Input: An initial feasible iterate ω
Output: An approximate solution to Problem 8

function FrankWolfeAC(ω)
    for $t = 0, \ldots, T$ **do**
        Compute a Fiedler vector y^* of $L(\omega)$
        $\nabla F(\omega)_k \leftarrow y^{*\mathsf{T}} L_k^c y^*, k = 1, \ldots, m$ ▷ Eq. (6.5)
        $s_t \leftarrow \operatorname{argmax}_s s^\mathsf{T} \nabla F(\omega)$ ▷ Prob. 9; eq. (6.6)
        $\alpha \leftarrow 2/(2+t)$ ▷ Compute step size
        $\omega \leftarrow \omega + \alpha\,(s_t - \omega)$
    end for
    return ω
end function

Relaxing the integrality constraints of Problem 7 dramatically alters the difficulty of the problem. In particular, we know (cf. [58]):

Lemma 8. *The function* $F(\omega) = \lambda_2(L(\omega))$ *is* concave *on the set* $\omega \in [0,1]^m, \mathbb{1}^\mathsf{T}\omega = K$.

Consequently, solving Problem 8, then, amounts to maximizing a concave function over a convex set; this is in fact a convex optimization problem (one can see this by simply considering minimization of the objective $-F(\omega)$) and hence *globally solvable* (see, e.g. [11, 19]). Since a solution to Problem 8 need not be feasible for the original problem, we then *round* solutions to the relaxed problem to their nearest correspondents in the feasible set of Problem 7.

6.2.1 Solving the relaxation

There are several methods which could, in principle, be used to solve the relaxation in Problem 8. For example, Ghosh and Boyd [58] consider solving an equivalent semidef-

inite program. This approach has the advantage of fast convergence (in terms of the number of iterations required to compute an optimal solution), but can nonetheless be slow for the large problem instances ($m > 1000$) typically encountered in the SLAM setting. Instead, our algorithm for *maximizing algebraic connectivity* (MAC), summarized in Algorithm 2, employs an inexpensive subgradient (more precisely, *supergradient*) approach to solve Problem 8, then *rounds* its solution to the nearest element of the feasible set for Problem 7.[4]

In particular, MAC uses the *Frank-Wolfe method* (also known as the *conditional gradient method*), a classical approach for solving convex optimization problems of the form in Problem 8 [11]. At each iteration, the Frank-Wolfe method requires (1) linearizing the objective F at a particular ω, (2) maximizing the linearized objective over the (convex) feasible set, and (3) taking a step in the direction of the solution to the linearized problem. The remainder of this section gives a detailed exposition of our adaptation of the Frank-Wolfe method to the problem of algebraic connectivity maximization, which is summarized in Algorithm 3.

The Frank-Wolfe method is particularly advantageous in this setting since the feasible set for Problem 8 is the intersection of the hypercube with the linear subspace determined by $\mathbb{1}^\mathsf{T}\omega = K$ (a linear equality constraint). Consequently this problem amounts to solving a linear program, which can be done easily (and in fact, as we show, admits a simple closed-form solution). In particular, the *direction-finding subproblem* for the Frank-Wolfe method is the following linear program:

Problem 9 (Direction-finding subproblem)**.** Fix an iterate $\omega \in [0,1]^m,\ \mathbb{1}^\mathsf{T}\omega = K$. The direction-finding subproblem is to find the point s solving the following linear program:

$$\begin{aligned} &\max_{s\in[0,1]^m} \ s^\mathsf{T}\nabla F(\omega), \\ &\mathbb{1}^\mathsf{T}s = K. \end{aligned} \tag{6.4}$$

In order to compute the linearized objective in Problem 9 we require a supergradient of the original objective function (in the usual case where F is differentiable at ω,

[4]Supergradients are simply the concave analogue of subgradients; i.e., the tangent hyperplane formed by any supergradient of a concave function F must lie *above* F.

this is simply the gradient of F). It turns out, we can always recover a supergradient of F at a particular ω in terms of a Fiedler vector of $L(\omega)$. Specifically, we have the following theorem (which we prove in Appendix B):

Theorem 9 (Supergradients of $F(\omega)$)**.** *Let $y^*(\omega)$ be any eigenvector of $L(\omega)$ corresponding to $\lambda_2(L(\omega))$. Then:*

$$\begin{aligned} \nabla F(\omega) &= \left[\frac{\partial F}{\partial \omega_1}, \ldots, \frac{\partial F}{\partial \omega_m}\right]^\mathsf{T}, \\ \frac{\partial F}{\partial \omega_k} &= y^*(\omega)^\mathsf{T} L_k^c y^*(\omega), \end{aligned} \tag{6.5}$$

is a supergradient *of F at ω.*

Therefore, supergradient computation can be performed by simply recovering an eigenvector of $L(\omega)$ corresponding to $\lambda_2(L(\omega))$.

Problem 9 is a linear program, for which several solution techniques exist [11]. However, in our case, Problem 9 admits a simple, *closed-form* solution s^* attaining its optimal value (which we prove in Appendix B):

Theorem 10 (A closed-form solution to Problem 9)**.** *Let $\mathcal{S}^*$, $|\mathcal{S}^*| = K$ be the set containing the indices of the K* largest *elements of $\nabla F(\omega)$, breaking ties arbitrarily where necessary. The vector $s^* \in \mathbb{R}^n$ with element k given by:*

$$s_k^* = \begin{cases} 1, & k \in \mathcal{S}^*, \\ 0, & \textit{otherwise}, \end{cases} \tag{6.6}$$

is an optimizer for Problem 9.

In this work, we use a simple decaying step size α to update ω in each iteration. While in principle, we could instead use a line search method [11, Sec. 2.2]), this would potentially require many evaluations of $F(\omega)$ within each iteration. Since every evaluation of $F(\omega)$ requires an eigenvalue computation, this can become a computational burden for large problems.

In the event that the optimal solution to the relaxed problem is integral, we ensure that we have *also* obtained an optimal solution to the original problem. However, this need not be the case in general. In the (typical) case where integrality does not hold, we *project* the solution to the relaxed problem onto the original constraint set. In this case, an integral solution $\Pi(\omega)$ can be obtained by rounding the largest K components of s to 1, and setting all other components to zero:

$$\Pi(\omega)_k \triangleq \begin{cases} 1, & \text{if } \omega_k \text{ is in the largest } K \text{ elements of } \omega, \\ 0, & \text{otherwise.} \end{cases} \tag{6.7}$$

In general, the Frank-Wolfe algorithm offers *sublinear* (i.e. $\mathcal{O}(1/T)$ after T iterations) convergence to the globally optimal solution in the worst case [50]. However, in this context it has several advantages over alternative approaches. First, we can bound the sparsity of a solution after T iterations. In particular, we know that the solution after T iterations has *at most* KT nonzero entries. Second, the gradient computation requires only a single computation of the minimal 2 dimensional eigenspace of an $n \times n$ matrix. This can be performed quickly using a variety of methods (e.g. the preconditioned Lanczos method). Finally, as we showed, the direction-finding subproblem in Problem 9 admits a simple *closed-form solution* (as opposed to a projected gradient method which requires projection onto an ℓ_1-ball). In consequence, despite the fact that gradient-based methods may require many iterations to converge to *globally optimal* solutions, high-quality approximate solutions can be computed fast at the scale necessary for SLAM problems. As we show in the following section, our approach admits *post hoc* suboptimality guarantees even in the event that we terminate optimization prematurely (e.g. when a *fast* but potentially suboptimal solution is required). Critically, these suboptimality guarantees ensure the quality of the solutions of our approach not only with respect to the relaxation, but also with respect to the *original problem*.

6.2.2 Post hoc suboptimality guarantees

Algorithm 3 admits several *post hoc* suboptimality guarantees. Let p^* be the optimal value of the original nonconvex maximization in Problem 7. Since Problem 8 is a relaxation of Problem 7, in the event that optimality attains for a vector ω^*, we know:

$$F(\Pi(\omega^*)) \leq p^* \leq F(\omega^*). \tag{6.8}$$

Therefore, the suboptimality of a rounded solution $\Pi(\omega^*)$ is bounded as follows:

$$p^* - F(\Pi(\omega^*)) \leq F(\omega^*) - F(\Pi(\omega^*)). \tag{6.9}$$

Consequently, in the event that $F(\omega^*) - F(\Pi(\omega^*)) = 0$, we know that $\Pi(\omega^*)$ *must* correspond to an optimal solution to Problem 7.

The above guarantees apply in the event that we obtain a *maximizer* ω^* of Problem 8. This would seem to pose an issue if we aim to terminate optimization before we obtain a verifiable, globally optimal solution to Problem 8 (e.g. in the presence of real-time constraints). Since these solutions are not necessarily globally optimal in the relaxation, we do not know if their objective value is larger or smaller than the optimal solution to Problem 7. However, we can in fact obtain per-instance suboptimality guarantees of the same kind for *any* estimate $\hat{\omega}$ through the *dual* of our relaxation (cf. Lacoste-Julien et al. [78, Appendix D]). Here, we give a derivation of the dual upper bound which uses only the concavity of F.

Since F is concave, for any $x, y \in [0,1]^m,\ \mathbb{1}^\mathsf{T} x = \mathbb{1}^\mathsf{T} y = K$ we have:

$$F(y) \leq F(x) + (y - x)^\mathsf{T} \nabla F(x). \tag{6.10}$$

Consider then the following upper bound:

$$\begin{aligned} F(\omega^*) &\leq F(\hat{\omega}) + (\omega^* - \hat{\omega})^\mathsf{T} \nabla F(\hat{\omega}) \\ &\leq \max_{s \in [0,1]^m, \mathbb{1}^\mathsf{T} s = K} F(\hat{\omega}) + (s - \hat{\omega})^\mathsf{T} \nabla F(\hat{\omega}) \\ &= F(\hat{\omega}) - \hat{\omega}^\mathsf{T} \nabla F(\hat{\omega}) + \max_{s \in [0,1]^m, \mathbb{1}^\mathsf{T} s = K} s^\mathsf{T} \nabla F(\hat{\omega}). \end{aligned} \tag{6.11}$$

We observe that the solution to the optimization in the last line of (6.11) is *exactly* the solution to the direction-finding subproblem (Problem 9). Letting $\hat{s}$ be a vector obtained as a solution to Problem 9 at $\hat{\omega}$, we obtain the following *dual* upper bound:

$$F_D(\hat{\omega}) \triangleq F(\hat{\omega}) + \nabla F(\hat{\omega})^\mathsf{T} (\hat{s} - \hat{\omega}). \tag{6.12}$$

Now, from (6.11), we have $F_D(\omega) \geq F(\omega^*)$ for any ω in the feasible set. In turn, it is straightforward to verify that the following chain of inequalities hold for any estimator $\hat{\omega}$ in the feasible set of the Boolean relaxation:

$$F(\Pi(\hat{\omega})) \leq p^* \leq F_D(\hat{\omega}), \tag{6.13}$$

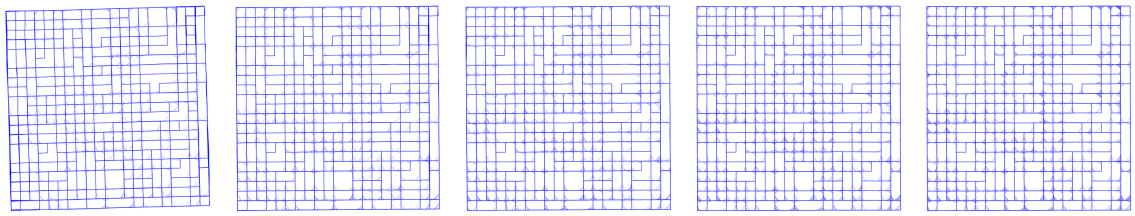
with the corresponding suboptimality guarantee:

$$p^* - F(\Pi(\hat{\omega})) \leq F_D(\hat{\omega}) - F(\Pi(\hat{\omega})). \tag{6.14}$$

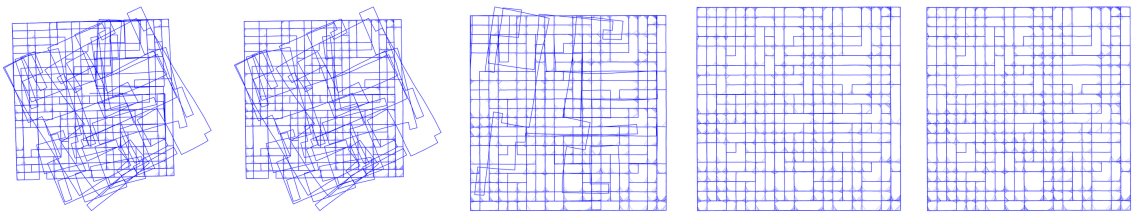
Moreover, we can always recover a suboptimality bound on $\hat{\omega}$ with respect to the optimal value $F(\omega^*)$ to relaxed problem as:

$$F(\omega^*) - F(\hat{\omega}) \leq F_D(\hat{\omega}) - F(\hat{\omega}) \tag{6.15}$$

The expression appearing on the right-hand side of (6.15) is the (Fenchel) *duality gap*. Equation (6.15) also motivates the use of the duality gap as a stopping criterion for Algorithm 3: if the gap is sufficiently close to zero (e.g. to within a certain numerical tolerance), we may conclude that we have reached an optimal solution ω^* to Problem



(a) Our approach.



(b) Naïve baseline.

Figure 6-1: **Qualitative results for pose-graph sparsification**. Pose-graph optimization results for the *City10K* dataset with varying degrees of sparsity using (a) our method and (b) a naïve baseline which selects the most certain measurements. Left to right: 20%, 40%, 60%, 80%, and 100% of the candidate edges.

Dataset	No. of Nodes	No. of Candidate (Loop Closure) Edges
KITTI 02	4661	43
KITTI 05	2761	66
Intel	1728	785
AIS2Klinik	15115	1614
City10K	10000	10688

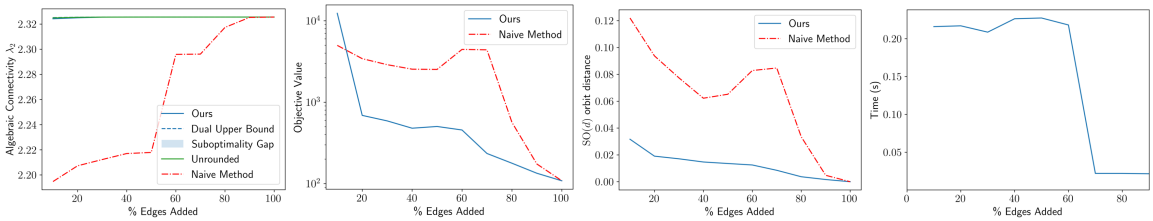
Table 6.1: Summary of the datasets used in our experiments.

8.

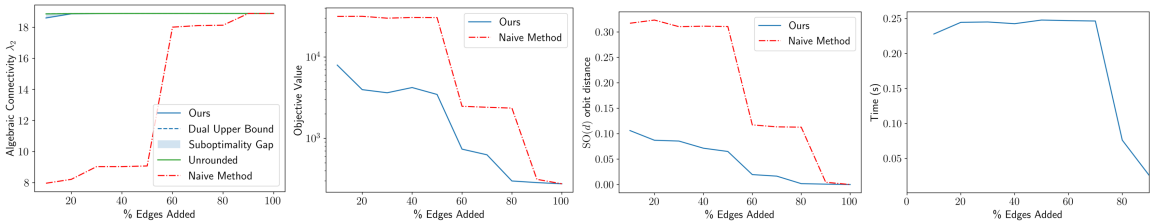
6.3 Experimental results

We implemented the MAC algorithm in Python and all computational experiments were performed on a 2.4 GHz Intel i9-9980HK CPU. For computation of the Fiedler value and the corresponding vector, we use TRACEMIN-Fiedler [89, 125]. In all experiments, we run MAC for a maximum of 20 iterations, or when the duality gap in equation (6.15) reaches a tolerance of 10^{-8}.

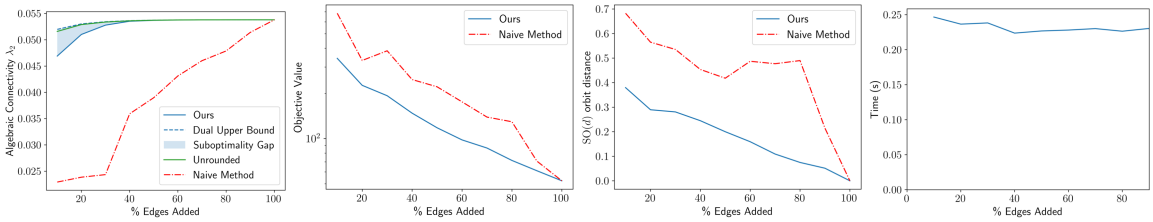
We evaluated our approach using several benchmark pose-graph SLAM datasets.



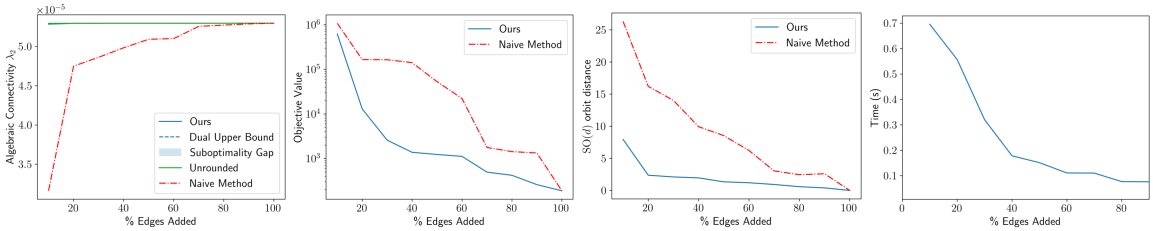
(a) *KITTI 02*



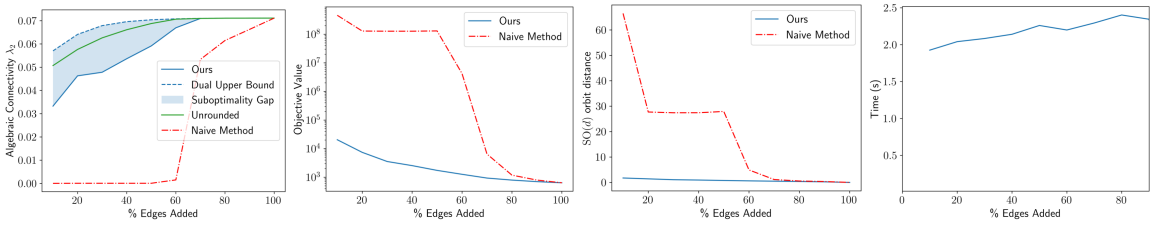
(b) *KITTI 05*



(c) *Intel*



(d) *AIS2Klinik*



(e) *City10K*

Figure 6-2: **Quantitative results for pose-graph sparsification**. Pose-graph optimization results for several benchmark datasets: (a) *KITTI 02* (b) *KITTI 05*, (c) *Intel*, (d) *AIS2Klinik*, and (e) *City10K* with varying degrees of sparsity (as percent of candidate edges added). Left to right: The algebraic connectivity of the graphs obtained by our method versus the naïve baseline (larger is better), the objective value of the maximum-likelihood estimator for each sparsified graph under the *original* objective, i.e. with all edges retained (smaller is better; note the log-scale), the SO(d) orbit distance between a maximum-likelihood estimator computing using the sparsified graph and a one computed for the graph containing *all* of the candidate edges (smaller is better), and the computation time for our approach.

For each dataset, we use odometry edges (between successive poses) to form the base graph and loop closure edges as candidate edges. We consider selection of $10\%, 20\%, \ldots, 100\%$ of the candidate loop closure edges in the sparsification problem. We present results on five datasets in this document (summarized in Table 6.1). In particular, we consider here the *Intel* dataset, the *City10K* dataset, KITTI dataset sequences 02, and 05, and the *AIS2Klinik* dataset. The *Intel* dataset and the *AIS2Klinik* dataset are both obtained from real data, while the *City10K* dataset is synthetic. The *City10K* dataset, however, contains far more candidate edges, and therefore serves as a reasonable "stress test" for the computation time of our approach. We compare our approach to a naïve heuristic method which does not consider graph topology. Specifically, the naïve method selects the edges with the most confident rotation measurements (i.e. the set of K edges $\{i, j\}$ with the largest κ_{ij}). This simple heuristic approach serves two purposes: First, it provides a baseline, topology-agnostic approach to demonstrate the impact of considering graph connectivity in a sparsification procedure; second, we use this method to provide a *sparse* initial estimate to our algorithm. For each method, we compare the graph connectivity (as measured by the Fiedler value) as well as the quality of maximum-likelihood estimators for pose-graph optimization (i.e. solutions to Problem 4) under the edge sets selected by each method. We use SE-Sync [118] to compute the globally optimal estimate of robot poses in each case.[5]

Figure 6-1 gives a qualitative comparison of the results from our approach as compared with the baseline on the *City10K* dataset across a range of candidate loop closures allowed. We observe that even retaining 60% of the candidate edges, the quality of the results provided by the baseline method degrade significantly compared to those of the full set of loop closures. In contrast, our sparsification approach leads to high-quality estimates even with a *significant* reduction in the number of edges.

For a quantitative comparison of each method, we report three performance measures: (1) the algebraic connectivity $\lambda_2(L(\omega))$ of the graphs determined by each edge selection ω, (2) the "full" objective value from Problem 4 (i.e. keeping 100% of the

[5]In all of our experiments, SE-Sync returned *certifiably-optimal* solutions to Problem 4.

edges) attained by globally optimal solutions to the *sparsified* problems, and (3) the SO(d)-orbit distance between the rotational states of a maximum-likelihood estimator for the sparsified problem and those of a maximum-likelihood estimator for the original (full) objective. The SO(d)-orbit distance between two rotational state estimates is defined as:

$$\begin{aligned} d_{\mathcal{S}}(X, Y) &\triangleq \min_{G \in \mathrm{SO}(d)} \|X - GY\|_F, \\ X, Y &\in \mathrm{SO}(d)^n, \end{aligned} \tag{6.16}$$

which can be computed in closed form by means of a singular value decomposition (see Rosen et al. [118, Theorem 5]). The "full" objective value attained by solutions to the sparsified problems serves as one indicator of "how close" solutions to the sparsified problem are to the MLE for the "full" problem. If the "full" objective value attained by the MLE for a sparsified graph is close to that of the MLE computed using 100% of the candidate edges, the MLE for the sparsified graph is likely also a high-quality solution under the full objective. The SO(d)-orbit distance quantifies the actual deviation (up to global symmetry) between the estimated rotational states in each solution. Since the translational states are recovered analytically (per [118]), this serves as a useful measure, independent of the global scale of the translational states, of the degradation in solution quality from the "full" MLE as we sparsify the graph.

Figure 6-2 summarizes our quantitative results on each benchmark dataset. Our approach consistently achieves better connected graphs (as measured by the algebraic connectivity). In most cases, a maximum of 20 iterations was enough to achieve solutions to the relaxation with algebraic connectivity very close to the dual upper bound (and therefore nearly globally optimal). Moreover, maximum-likelihood estimators for Problem 4 computed using the sparsified measurement graphs from our method perform significantly better in terms of their "full" objective value and their deviation from a MLE computed using all of the measurement edges.

Beyond providing high-quality sparse measurement graphs, our approach is also fast. For the *Intel* dataset, all solutions were obtained in less than 250 milliseconds. Sparsifying the (larger) *AIS2Klinik* dataset required up to 700 ms, but only around

100 ms when larger edge selections were allowed, as the duality gap tolerance was reached in fewer than the maximum allowed iterations of Frank-Wolfe method. The largest dataset (in terms of candidate edges) is the *City10K* dataset, with over 10000 loop closure measurements to select from. Despite this, our approach produces near-optimal solutions in just 2 seconds.

With respect to the suboptimality guarantees of our approach, it is interesting to note that on both the *Intel* and *City10K* datasets, the rounding procedure introduces fairly significant degradation in algebraic connectivity - particularly for more aggressive sparsity constraints. In these cases, it seems that the Boolean relaxation we consider leads to fractional optimal solutions, rather than solutions amounting to hard selection of just a few edges. It is not clear in these cases whether the integral solutions obtained by rounding are indeed suboptimal for the Problem 7, or whether this is a consequence of the *integrality gap* between *global* optima of the relaxation and of Problem 7.[6]

At present we do not have access to an implementation of the D-optimal sparsification approaches considered in [73]. However, we evaluate our method on similar (and similarly sized) datasets, and a comparison of the computation times suggests that our approach compares favorably in computation time and (consequently) the scale of problems we can consider. For example, Khosoussi et al. [73, Sec. 9.4] report computation times of "$\gg$ 10 minutes" to solve a convex relaxation of the D-optimal sparsification problem on the *City10K* dataset (versus $\approx$ 2 seconds per Figure 6-2e). In light of this fact, and since both the D-optimality criterion and E-optimality criterion are essentially variance-minimizing criteria, in the event that one requires D-optimal designs specifically, an interesting avenue for future work would be to use our E-optimal designs to supply an initial estimate to, for example, the convex relaxation approach of Khosoussi et al. [73]. A detailed empirical comparison of the impact of different optimal design criteria on the quality of SLAM solutions would be tremendously helpful for practitioners, and would also be an interesting area for

[6] In general, even simply *verifying* the global optimality of solutions to Problem 7 is NP-Hard [96].

future work.

6.4 Summary

In this chapter, we proposed an approach for pose-graph measurement sparsification by maximizing the *algebraic connectivity* of the measurement graphs, a key quantity which has been shown to control the estimation error of pose-graph SLAM solutions. Our algorithm, MAC, is based on a first-order optimization approach for solving a convex relaxation of the maximum algebraic connectivity augmentation problem. The algorithm itself is simple and computationally inexpensive, and, as we showed, admits formal *post hoc* performance guarantees on the quality of the solutions it provides. In experiments on several benchmark pose-graph SLAM datasets, our approach quickly produces high-quality sparsification results which better preserve the connectivity of the graph and, consequently, the quality of SLAM solutions computed using those graphs. An interesting area for future work is the empirical comparison of different optimality criteria for the pose graph sparsification problem. Finally, in this work we consider only the removal of measurement graph *edges*. For lifelong SLAM applications, an important aspect of future work will be to combine these procedures with methods for *node* removal (e.g. [22, 28, 68]).

Chapter 7

Conclusion

This thesis considers the development of lifelong and learning-augmented robot navigation systems. It is desired that any such system be both *robust* and *efficient*. The practical limitations of real systems make this a tremendous challenge: often robustness can be achieved at the expense of efficiency and vice versa, but realizing *both* simultaneously is necessary for true lifelong autonomy. To this end, this thesis makes several contributions:

In Chapter 3, we introduce the DC-SAM library and a new optimization approach for inference in hybrid factor graphs. We demonstrate the utility of the hybrid factor graph framework for solving common inference problems arising in robot perception, emphasizing robust models.

In Chapter 4, we leverage the hybrid factor graph representation and DC-SAM library to combine the (discrete) output of learned object detection and classification models with (continuous) geometric measurements for estimation. In turn, we present an approach to robust object-level semantic SLAM which accounts for uncertainty in semantic predictions and *ambiguity* in measurement-landmark correspondences.

In Chapter 5, we focus on pose-graph SLAM and ask whether it is possible to produce an *initial guess* that achieves provably bounded estimation error and deviation from the globally optimal estimate. We show that it is possible and present an algorithm based on spectral decomposition that admits *formal performance guarantees* on solution quality. In the process, we identify the *algebraic connectivity* or *Fiedler*

value of the measurement graphs arising in SLAM as a key parameter controlling estimation accuracy.

Finally, in Chapter 6 we consider long-term pose-graph SLAM, where controlling the computational burden and memory requirements of SLAM requires *sparsifying* the measurement graph. Motivated by the insights from Chapter 5, we develop the MAC algorithm for pose-graph sparsification, which designs sparse pose-graphs by *maximizing algebraic connectivity*. MAC, based on convex relaxation, admits formal post hoc suboptimality guarantees on the connectivity of the sparse graphs it provides.

7.1 Limitations and Future Work

In retrospect, the central problem of robustness in lifelong robot perception applications lies not in the particulars of how we perform inference within a given model, but rather in the unavoidable fact that a robot building a representation of the world *online* is necessarily "building the plane while flying it." That is to say, every robot perception system to date relies on a set of heuristics that dictate how a model is constructed as new data is collected. Only once that model is constructed can inference be performed, whereby we estimate the parameters of the model. It is in those heuristic procedures that dynamically construct the model that modern perception systems fail. The position of this thesis is not that we should attempt to avoid such rules entirely, but rather to enable the design and engineering of perception systems that make *representing ambiguity* about the model itself an essential consideration. This opinion is summarized concisely by the following quote:

> Furthermore, I believe systems make irreversible decisions (such as feature/no feature) too early in the processing of data. The motivation is usually to reduce the amount of data that need to be processed, but the results can be disastrous. Methods that postpone irreversible commitments should receive more attention.
>
> *Tomás Lozano-Pérez, foreword to Autonomous Robot Vehicles [83]*

The problem of efficiency, on the other hand, lies primarily in the difficulty in

determining "what is important" for the navigation task. Indeed, until relatively recently, it was not deeply understood what aspects of the SLAM problem control estimation error (even in the setting where data association is assumed to be known). In answering this question, the contributions of this thesis point toward spectral graph theory, which seems to hold important keys to understanding what aspects of a measurement graph in SLAM control accuracy and therefore what information can safely be discarded. It is as yet unclear, however, to what extent information should be discarded (as we do in Chapter 6) or instead *summarized* (and if so, how to do it).

It is clear that much work remains to be done on these topics. The following sections provide some key directions for future work and areas for improvement.

7.1.1 Expressive models and robust inference

The work on hybrid factor graphs presented in Chapter 3 and Chapter 4 opens up a number of interesting directions for future work. From a modeling standpoint, the ability to cleanly capture dependencies between discrete or symbolic quantities and continuous (typically, but not necessarily) geometric quantities makes hybrid factor graphs a natural representation for many important robot perception problems.

To this end, one path forward involves using these tools to combine the output of multiple learned models (which may relate discrete quantities, continuous quantities, or a mixture of both). For example, we may want to combine predictions from an object classification model, a category-level object shape model, and real 3D observations from a depth camera. In this way, we could leverage hybrid factor graphs as a generic representation for fusing real sensor data with predictions from learned models.

The methods presented in Chapters 3 and 4 are practically fast in part because they rely on local optimization techniques that may converge to bad solutions if initialized poorly. This is an important limitation of these techniques that warrants further investigation. We discuss this issue in detail in Chapter 3. While we consider the problem of producing "provably good" initial solutions for pose-graph optimization in Chapter 5, addressing this challenge for general hybrid factor graphs is more

challenging (and perhaps particular problems can benefit from application-specific considerations, e.g. additional sources of trusted information, in the same way that one may often rely on odometric measurements for initialization in short-term SLAM problems).

7.1.2 Self-supervised and unsupervised learning

One issue left unexplored by this thesis is *continual* learning, i.e. where a perception model (such as an object detector) is trained (or refined) online as a robot navigates through the environment. The ability to perform robust semantic (object-based) SLAM opens up the possibility of using an optimized map produced by a SLAM method as a supervisory signal for learning-based perception methods like object detectors or object pose estimators. While some initial work to this end has begun [85], this topic remains relatively under-explored.

Beyond self-supervision, we might consider the issue of *discovering* object categories in an unsupervised manner. Outside of the work in this thesis, we have made some of our first steps performing unsupervised multi-robot object category learning [44]. However, this work did not make use of globally consistent three-dimensional world models like those produced by SLAM methods, instead discovering object structure through spatial correlation within images and temporal correlation between subsequent image frames in a video. Combining these ideas with a SLAM method to leverage 3D structure would be a natural area for future development.

From a tooling perspective, closer integration with tools for machine learning would be tremendously helpful. Early examples for *continuous* factor graphs include Theseus [109] and PyPose [141]. However, to date, not of these tools directly support working with hybrid factor graphs.

7.1.3 Performance guarantees for robot perception

From our developments in Chapter 5 it is clear that much remains to be understood about performance guarantees for robot perception methods. While some recent

work has pushed this line of research further (e.g. [24]), the topic of performance guarantees for robot perception still needs to be explored. In the context of the work presented in this thesis, there are a few next steps. First, the bounds we present in Chapter 5 rely on knowledge of the actual error in the data in order to be computed, making them impractical. This could be addressed by considering the statistics of the measurements as determined by a sensor model (rather than the specific measurements obtained by the robot). Matrix concentration inequalities (see, e.g., [137]) are the right mathematical tools for this sort of analysis, and this is an interesting direction for future developments. Second, the bounds themselves appear to be (empirically) quite loose in the scenarios we examined. A more refined analysis (perhaps with less "slack" in the bounds) could be tremendously useful, especially if it could explain the high-quality results of the spectral and chordal initialization techniques. Progress in this direction may point to, for example, a statistical regime in which certain perception problems can be said to be "easier" than others in a formal mathematical sense.

7.1.4 Efficient inference, compression, and hierarchy

A key limitation of the methods presented in Chapter 6 is that we *discard* information during the sparsification process. It stands to reason that perhaps in some cases we might be able to "compensate" for a discarded measurement in a pose-graph by strengthening the influence of other available measurements. Spectral sparsification methods (see, e.g. [130]) allow for the weights of retained edges to be modified (while provably sparsifying graphs) in a manner that aims to preserve not just the algebraic connectivity of the graph but the whole spectrum of the graph Laplacian. Since the spectrum of the graph Laplacian is closely connected to the cost of SLAM solutions for the corresponding pose graph, these techniques may provide better ("compressed") graphs for SLAM than the MAC algorithm. Initial explorations with these spectral sparsification techniques have been promising, but further development of these ideas is a priority for future work. Similar spectral techniques have been explored for *node* sparsification, but have not yet been applied in the context of SLAM problems; this

is another interesting area for future research.

While ideas from spectral graph theory have established mathematical links between graph properties and estimation error for SLAM methods that can be used for graph sparsification (or "compression"), these methods do not exploit *semantics* for the purposes of summarization. A robot equipped with a semantic understanding of the world could use knowledge of *hierarchy* to summarize task-relevant information. For example, a robot may not need to actively update information about the contents of the inside of buildings while navigating down a street, but once it enters a building, that data becomes relevant. Hierarchical representations, such as the object-based representations of Ok et al. [103], the S-graphs of Bavle et al. [9], or 3D dynamic scene graphs (e.g., Hughes et al. [66] and Rosinol et al. [121]) may enable this type of semantically-informed graph compression. Along these lines, it would be interesting to consider whether hybrid factor graph models or the DC-SAM tools developed as part of this thesis may be useful for the development of these technologies.

Appendix A

Proofs for Chapter 5

A.1 Structure of the data matrices

In this appendix, we provide the definitions of the various matrices appearing in the parameterization of the rotation averaging and pose-graph SLAM problems. $L(W^\tau)$ and $L(W^\rho)$ denote the Laplacians of the translational weight graph $W^\tau \triangleq (\mathcal{V}, \mathcal{E}, \{\tau_{ij}\})$ and rotational weight graph $W^\rho \triangleq (\mathcal{V}, \mathcal{E}, \{\kappa_{ij}\})$, respectively, with *undirected edges* $\{i, j\} \in \mathcal{E}$. These are $n \times n$ matrices with i, j-entries:

$$L(W^\tau)_{ij} = \begin{cases} \sum_{e \in \delta(i)} \tau_e, & i = j, \\ -\tau_{ij}, & \{i, j\} \in \mathcal{E}, \\ 0, & \{i, j\} \notin \mathcal{E}, \end{cases} \tag{A.1a}$$

$$L(W^\rho)_{ij} = \begin{cases} \sum_{e \in \delta(i)} \kappa_e, & i = j, \\ -\kappa_{ij}, & \{i, j\} \in \mathcal{E}, \\ 0, & \{i, j\} \notin \mathcal{E}. \end{cases} \tag{A.1b}$$

$L(\tilde{G}^{\rho})$ denotes the *connection Laplacian* for the rotational measurements, which is a $dn \times dn$ symmetric block-diagonal matrix with $d \times d$ blocks determined by:

$$L(\tilde{G}^{\rho})_{ij} \triangleq \begin{cases} d_i^{\rho} I_d, & i = j, \\ -\kappa_{ij} \tilde{R}_{ij}, & \{i,j\} \in \mathcal{E}, \\ 0_{d \times d}, & \{i,j\} \notin \mathcal{E}, \end{cases} \tag{A.2a}$$

$$d_i^{\rho} \triangleq \sum_{e \in \delta(i)} \kappa_e, \tag{A.2b}$$

where $\delta(i)$ denotes the set of edges *incident to* node i. $\tilde{V} \in \mathbb{R}^{n \times dn}$ denotes the $(1 \times d)$-block-structured matrix with (i,j) block given by:

$$\tilde{V}_{ij} \triangleq \begin{cases} \sum_{e \in \delta^-(j)} \tau_e \tilde{t}_e^{\mathsf{T}}, & i = j, \\ -\tau_{ji} \tilde{t}_{ji}^{\mathsf{T}}, & (j,i) \in \vec{\mathcal{E}}, \\ 0_{1 \times d}, & \text{otherwise.} \end{cases} \tag{A.3}$$

Finally, $\tilde{\Sigma} \in \mathrm{SBD}(d,n)$ denotes the symmetric block-structured diagonal matrix given by:

$$\begin{aligned} \tilde{\Sigma} &\triangleq \mathrm{Diag}(\tilde{\Sigma}_1, \ldots, \tilde{\Sigma}_n) \in \mathrm{SBD}(d,n) \\ \tilde{\Sigma}_i &\triangleq \sum_{e \in \delta^-(i)} \tau_e \tilde{t}_e \tilde{t}_e^{\mathsf{T}}, \end{aligned} \tag{A.4}$$

where $\delta^-(i)$ denotes the set of edges *leaving* node i. With these definitions in hand, the translational data matrix $\tilde{Q}^{\tau}$ can be defined as:

$$\tilde{Q}^{\tau} = \tilde{\Sigma} - \tilde{V}^{\mathsf{T}} L(W^{\tau})^{\dagger} \tilde{V}. \tag{A.5}$$

A.2 Analysis of the spectral relaxation

A.2.1 Recovering minimizers of Problem 6 as eigenvectors

In this section we derive a closed-form description of the *global* minimizers Y^* of the spectral relaxation Problem 6. Specifically, we prove the following theorem:

Theorem 11 (Global minimizers of the spectral relaxation)**.** *Let $\lambda_1(\tilde{Q}) \leq \cdots \leq \lambda_d(\tilde{Q})$ be the d smallest eigenvalues of $\tilde{Q}$. Then $Y^* \in \mathbb{R}^{d\times dn}$ is a global minimizer of the spectral relaxation Problem 6 if and only if*

$$Y^* = \sqrt{n} \begin{pmatrix} v_{\sigma(1)} \\ \vdots \\ v_{\sigma(d)} \end{pmatrix} \in \mathbb{R}^{d\times dn} \tag{A.6}$$

where $v_1, \ldots, v_d \in \mathbb{R}^{dn}$ are a set of orthonormal eigenvectors corresponding to the d smallest eigenvalues, and σ is a permutation. The corresponding optimal value of Problem 6 is:

$$p_S^* = n \sum_{i=1}^{d} \lambda_i(\tilde{Q}). \tag{A.7}$$

Proof. Our approach will be to reduce Problem 6 to an equivalent problem whose critical points are already well-understood. To that end, let $Z \triangleq n^{-1/2} Y^{\mathsf{T}} \in \mathbb{R}^{dn\times d}$, so that $Y = \sqrt{n} Z^{\mathsf{T}}$. Substitution into Problem 6 then gives:

$$\begin{aligned} p_S^* = \min_{Z\in\mathbb{R}^{dn\times d}} & \operatorname{tr}\left(n\tilde{Q}ZZ^{\mathsf{T}}\right) \\ \text{s.t. } & Z^{\mathsf{T}}Z = I_d. \end{aligned} \tag{A.8}$$

Observe that $Z^{\mathsf{T}}Z = I_d$ if and only if $Z \in \mathrm{St}(d, dn)$; therefore, we may in turn rewrite (A.8) as the following *unconstrained* minimization over the Stiefel manifold:

$$p_S^* = \min_{Z\in\mathrm{St}(d,dn)} \operatorname{tr}\left(n\tilde{Q}ZZ^{\mathsf{T}}\right). \tag{A.9}$$

Note that we may now recognize (A.9) as the minimization of a generalized Rayleigh

quotient over a Stiefel manifold. This problem has been extensively studied; in particular, Absil et al. [2, Section 4.8.2] provides an elementary proof that

$$Z = (z_1, \ldots, z_d) \in \mathbb{R}^{dn \times d} \tag{A.10}$$

is a critical point of (A.9) *if and only if* its columns $\{z_i\}_{i=1}^d \subset \mathbb{R}^{dn}$ are an orthonormal set of eigenvectors for $n\tilde{Q}$. Moreover, substituting (A.10) into the objective in (A.8) and exploiting the fact that $\{z_i\}_{i=1}^d \subset \mathbb{R}^{dn}$ are pairwise mutually-orthogonal eigenvectors, we find that the corresponding objective value is:

$$\operatorname{tr}\left(n\tilde{Q}ZZ^{\mathsf{T}}\right) = n\sum_{i=1}^{d}\mu_i, \tag{A.11}$$

where μ_i is the eigenvalue corresponding to z_i. Since every critical point of (A.9) is of the form (A.10)–(A.11), it follows that the *global minimizers* Z^* are precisely those critical points whose columns are composed of the eigenvectors $v_1, \ldots, v_d \in \mathbb{R}^{dn}$ corresponding to the d *smallest* eigenvalues of $\tilde{Q}$. Recovering the corresponding optimal Y^* from Z^* then gives (A.6) and (A.7). □

A.2.2 Symmetric perturbations of symmetric matrices

Recall that $\underline{R}$ and Φ are solutions of the noiseless and noisy versions of the spectral relaxation in Problem 6. In Appendix A.2.1 we showed how these can be directly obtained from the Stiefel manifold elements giving the d minimum eigenvectors for their corresponding data matrices. The Davis-Kahan Theorem is a classical result in linear algebra that measures the perturbation of a matrix's eigenvectors under a symmetric perturbation of that matrix [131]. Therefore, we make use of this theorem to derive a bound on the estimation error of a spectral estimator as a function of the noise in the data matrix. In particular, the proof of Lemma 3 (and consequently Theorem 5) relies on a particular variant of the Davis-Kahan $\sin\theta$ Theorem [147, Theorem 2]. Here, we briefly restate the main result of [147] and give a proof of Lemma 3.

Theorem 12 (Yu et al. [147], Theorem 2)**.** *Let Σ, $\hat{\Sigma} \in \mathbb{R}^{p\times p}$ be symmetric, with eigenvalues $\lambda_1 \leq \ldots \leq \lambda_p$ and $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_p$ respectively. Fix $1 \leq r \leq s \leq p$ and assume that $\min(\lambda_r - \lambda_{r-1}, \lambda_{s+1} - \lambda_s) > 0$, where $\lambda_0 \triangleq -\infty$ and $\lambda_{p+1} \triangleq \infty$. Let $d \triangleq s - r + 1$, and let $V = (v_r, v_{r+1}, \ldots, v_s) \in \mathbb{R}^{p\times d}$ and $\hat{V} = (\hat{v}_r, \hat{v}_{r+1}, \ldots, \hat{v}_s) \in \mathbb{R}^{p\times d}$ have orthonormal columns satisfying $\Sigma v_j = \lambda_j v_j$ and $\hat{\Sigma}\hat{v}_j = \hat{\lambda}_j \hat{v}_j$ for $j = r, r+1, \ldots, s$. Then there exists an orthogonal matrix $G \in \mathrm{O}(d)$ such that*

$$\|\hat{V}G - V\|_F \leq \frac{2^{3/2}\min(d^{1/2}\|\hat{\Sigma} - \Sigma\|_{op}, \|\hat{\Sigma} - \Sigma\|_F)}{\min(\lambda_r - \lambda_{r-1}, \lambda_{s+1} - \lambda_s)}. \tag{A.12}$$

With this result in hand, we are ready to prove Lemma 3.

Proof of Lemma 3. The data matrices $\tilde{Q}$ and $\underline{Q}$ are symmetric $dn \times dn$ matrices with eigenvalues $\lambda_1 \leq \ldots \leq \lambda_{dn}$ and $\tilde{\lambda}_1 \leq \ldots \leq \tilde{\lambda}_{dn}$, respectively. From Theorem 11 we have that the d *normalized* eigenvectors corresponding to $\lambda_1, \ldots, \lambda_d$ of $\underline{Q}$ and $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_d$ are exactly $n^{-1/2}\underline{R}^\mathsf{T}$ and $n^{-1/2}\Phi^\mathsf{T}$, respectively. Then, letting $r = 1$ and $s = d$ and applying Theorem 12, there exists an orthogonal matrix $G \in \mathrm{O}(d)$ such that:

$$\frac{1}{\sqrt{n}}\|\Phi^\mathsf{T}G - \underline{R}^\mathsf{T}\|_F \leq \frac{2\sqrt{2d}\|\tilde{Q} - \underline{Q}\|_2}{\lambda_{d+1}(\underline{Q}) - \lambda_d(\underline{Q})}. \tag{A.13}$$

Multiplying both sides of this expression by $\sqrt{n}$, we have:

$$\|\Phi^\mathsf{T}G - \underline{R}^\mathsf{T}\|_F \leq \frac{2\sqrt{2dn}\|\tilde{Q} - \underline{Q}\|_2}{\lambda_{d+1}(\underline{Q}) - \lambda_d(\underline{Q})}. \tag{A.14}$$

Now, by definition $\Delta Q = \tilde{Q} - \underline{Q}$. If we assume $\mathcal{G}$ is connected,[1] from [118, Lemma 8] we have that $\lambda_{d+1}(\underline{Q}) > 0$. Since $\underline{R} \in \ker(\underline{Q})$, we know that $\lambda_d(\underline{Q}) = 0$ and the above expression simplifies to:

$$\|\Phi^\mathsf{T}G - \underline{R}^\mathsf{T}\|_F \leq \frac{2\sqrt{2dn}\|\Delta Q\|_2}{\lambda_{d+1}(\underline{Q})}. \tag{A.15}$$

Taking the transpose of the terms inside the norm gives the desired result. □

[1] It is not particularly restrictive to assume that $\mathcal{G}$ is connected. In the case that $\mathcal{G}$ is not connected, the estimation problem splits over the connected components of $\mathcal{G}$, and all of our results hold separately for each connected component.

A.3 Proof of the main results

In this appendix, we prove the main results, i.e. Theorem 5, Theorem 6, and Corollary 7.

A.3.1 An upper bound for the estimation error in Problem 6

Proof of Theorem 5. To simplify the subsequent derivation, we will assume without loss of generality that $\underline{R}$ and Φ are the representatives of their orbits satisfying $d_{\mathcal{O}}(\underline{R}, \Phi) = \|\underline{R} - \Phi\|_F$. Recall from the definition of $d_{\mathcal{S}}(\underline{R}, R^{(0)})$ that:

$$d_{\mathcal{S}}(\underline{R}, R^{(0)}) = \min_{G \in \mathrm{SO}(d)} \|\underline{R} - GR^{(0)}\|_F. \tag{A.16}$$

Therefore, we have:

$$\begin{aligned} d_{\mathcal{S}}(\underline{R}, R^{(0)})^2 &= \min_{G \in \mathrm{SO}(d)} \|\underline{R} - GR^{(0)}\|_F^2 \\ &\leq \|\underline{R} - R^{(0)}\|_F^2, \\ &= \sum_{i=1}^{n} \|\underline{R}_i - \Pi_{\mathcal{S}}(\Phi_i)\|_F^2, \end{aligned} \tag{A.17}$$

where in the last line we have used the fact that $R^{(0)}$ consists of the projections of individual $(d \times d)$ blocks of Φ onto $\mathrm{SO}(d)$. From Lemma 4, we have that each of the n summands above satisfies:

$$\|\underline{R}_i - \Pi_{\mathcal{S}}(\Phi_i)\|_F^2 \leq 4\|\underline{R}_i - \Phi_i\|_F^2. \tag{A.18}$$

This, in turn, gives a corresponding bound on the summation:

$$\begin{aligned} \sum_{i=1}^{n} \|\underline{R}_i - \Pi_{\mathcal{S}}(\Phi_i)\|_F^2 &\leq 4 \sum_{i=1}^{n} \|\underline{R}_i - \Phi_i\|_F^2 \\ &= 4\|\underline{R} - \Phi\|_F^2. \end{aligned} \tag{A.19}$$

Since, by hypothesis, Φ and $\underline{R}$ are representatives of their orbits satisfying $d_{\mathcal{O}}(\underline{R}, \Phi) = \|\underline{R} - \Phi\|_F$, we have:

$$4\|\underline{R} - \Phi\|_F^2 = 4d_{\mathcal{O}}\left(\underline{R}, \Phi\right)^2. \tag{A.20}$$

Applying Lemma 3, we directly obtain:

$$4d_{\mathcal{O}}\left(\underline{R}, \Phi\right)^2 \leq 4(2\sqrt{2dn})^2 \frac{\|\Delta Q\|_2^2}{\lambda_{d+1}(\underline{Q})^2}. \tag{A.21}$$

In summary, we have:

$$d_{\mathcal{S}}(\underline{R}, R^{(0)})^2 \leq 4(2\sqrt{2dn})^2 \frac{\|\Delta Q\|_2^2}{\lambda_{d+1}(\underline{Q})^2}. \tag{A.22}$$

Taking the square root of both sides of the inequality in the last line gives:

$$d_{\mathcal{S}}(\underline{R}, R^{(0)}) \leq \frac{4\sqrt{2dn}\|\Delta Q\|_2}{\lambda_{d+1}(\underline{Q})}, \tag{A.23}$$

which concludes the proof. □

A.3.2 An upper bound for the estimation error in Problem 5

We begin following the arguments of Preskitt [110, Appendix D.4]. From the optimality of R^* we have:

$$\begin{aligned} \operatorname{tr}(\tilde{Q}\underline{R}^{\mathsf{T}}\underline{R}) &= \operatorname{tr}(\underline{Q}\underline{R}^{\mathsf{T}}\underline{R}) + \operatorname{tr}(\Delta Q\underline{R}^{\mathsf{T}}\underline{R}) \\ &\geq \operatorname{tr}(\underline{Q}R^{*\mathsf{T}}R^*) + \operatorname{tr}(\Delta QR^{*\mathsf{T}}R^*) = \operatorname{tr}(\tilde{Q}R^{*\mathsf{T}}R^*). \end{aligned} \tag{A.24}$$

Since $\operatorname{tr}(\underline{Q}\underline{R}^{\mathsf{T}}\underline{R}) = 0$, we can rearrange the above expression to obtain:

$$\operatorname{tr}(\underline{Q}R^{*\mathsf{T}}R^*) \leq \operatorname{tr}(\Delta Q\underline{R}^{\mathsf{T}}\underline{R}) - \operatorname{tr}(\Delta QR^{*\mathsf{T}}R^*). \tag{A.25}$$

Using the fact that $\operatorname{tr}(\Delta Q \underline{R}^\mathsf{T} \underline{R}) = \operatorname{vec}(\underline{R})^\mathsf{T}(\Delta Q \otimes I_n)\operatorname{vec}(\underline{R})$ (and likewise for $\operatorname{tr}(\Delta Q R^{*\mathsf{T}} R^*)$), we have:

$$\begin{aligned}
\operatorname{tr}(\underline{Q} R^{*\mathsf{T}} R^*) &\leq \operatorname{vec}(\underline{R} - R^*)^\mathsf{T}(\Delta Q \otimes I_n)\operatorname{vec}(\underline{R} + R^*) \\
&\leq \|\operatorname{vec}(\underline{R} - R^*)\|_2 \|\Delta Q \otimes I_n\|_2 \|\operatorname{vec}(\underline{R} + R^*)\|_2 \\
&= \|\underline{R} - R^*\|_F \|\Delta Q\|_2 \|\underline{R} + R^*\|_F \\
&\leq 2\sqrt{dn}\|\underline{R} - R^*\|_F \|\Delta Q\|_2.
\end{aligned} \tag{A.26}$$

In order to lower-bound the right-hand side of (A.26) in terms of the estimation error $d_{\mathcal{S}}(\underline{R}, R^*)$, we will make use of the following technical lemma of Rosen et al. [118]:

Lemma 13 (Lemma 11 of Rosen et al. [118])**.** *Let $R \in \mathrm{O}(d)^n \subset \mathbb{R}^{d \times dn}$ and furthermore let $M = \{WR \mid W \in \mathbb{R}^{d \times d}\} \subset \mathbb{R}^{d \times dn}$ be the subspace of matrices with rows contained in* $\operatorname{image}(R^\mathsf{T})$*. Then*

$$\begin{aligned}
&\operatorname{Proj}_V : \mathbb{R}^{dn} \to \operatorname{image}(R^\mathsf{T}) \\
&\operatorname{Proj}_V(x) = \frac{1}{n} R^\mathsf{T} R x
\end{aligned} \tag{A.27}$$

is the orthogonal projection onto $\operatorname{image}(R^\mathsf{T})$ *with respect to the ℓ_2 inner product, and the map*

$$\begin{aligned}
&\operatorname{Proj}_M : \mathbb{R}^{d \times dn} \to M \\
&\operatorname{Proj}_M(X) = \frac{1}{n} X R^\mathsf{T} R
\end{aligned} \tag{A.28}$$

which applies Proj_V *to the rows of X is the orthogonal projection onto M with respect to the Frobenius inner product.*

Since $\ker(\underline{Q}) = \operatorname{image}(\underline{R}^\mathsf{T})$ and $\dim(\operatorname{image}(\underline{R}^\mathsf{T})) = d$, from Lemma 13, we have:

$$\operatorname{tr}(\underline{Q} R^{*\mathsf{T}} R^*) \geq \lambda_{d+1}(\underline{Q}) \|P\|_F^2, \tag{A.29}$$

where

$$\begin{aligned} R^* &= K + P \\ K &= \mathrm{Proj}_M(R^*) = \frac{1}{n} R^* \underline{R}^\mathsf{T} \underline{R} \\ P &= R^* - \mathrm{Proj}_M(R^*) = R^* - \frac{1}{n} R^* \underline{R}^\mathsf{T} \underline{R} \end{aligned} \tag{A.30}$$

is an orthogonal decomposition of R^* and the rows of P are contained in the orthogonal complement of image($\underline{R}^\mathsf{T}$)

The following lemma provides a bound on $d_{\mathcal{S}}(\underline{R}, R^*)^2$ in terms of $\|P\|_F^2$.

Lemma 14. *Let R^* and $\underline{R}$ be representatives of their orbits such that $d_{\mathcal{S}}(\underline{R}, R^*) = \|\underline{R} - R^*\|_F$, and $P = R^* - \mathrm{Proj}_M(R^*)$ as defined in* (A.30)*. Then:*

$$\frac{1}{4} d_{\mathcal{S}}(\underline{R}, R^*)^2 \leq \|P\|_F^2. \tag{A.31}$$

Proof. Let $X = \frac{1}{n} \underline{R} R^{*\mathsf{T}}$, so that $K = X^\mathsf{T} \underline{R}$. Expanding the left hand side, we have:

$$\begin{aligned} d_{\mathcal{S}}(\underline{R}, R^*)^2 &= \|R^* - \underline{R}\|_F^2 \\ &\leq \|R^* - \Pi_{\mathcal{S}}(X^\mathsf{T}) \underline{R}\|_F^2, \end{aligned} \tag{A.32}$$

from the fact that the orbit distance is obtained as the minimum over $G \in \mathrm{SO}(d)$ of the quantity $\|R^* - G\underline{R}\|_F$, and that by hypothesis this minimum is obtained as $\|R^* - \underline{R}\|_F$. Breaking up the norm into its blockwise summands, and from the orthogonal invariance of the Frobenius norm, we can rearrange this expression as follows:

$$\begin{aligned} \|R^* - \Pi_{\mathcal{S}}(X^\mathsf{T}) \underline{R}\|_F^2 &= \sum_{i=1} \|R_i^* - \Pi_{\mathcal{S}}(X^\mathsf{T}) \underline{R}_i\|_F^2 \\ &= \sum_{i=1}^{n} \|R_i^* \underline{R}_i^\mathsf{T} - \Pi_{\mathcal{S}}(X^\mathsf{T})\|_F^2. \end{aligned} \tag{A.33}$$

From Lemma 4, we know that each summand in the above expression satisfies

$$\|R_i^* \underline{R}_i^\mathsf{T} - \Pi_{\mathcal{S}}(X^\mathsf{T})\|_F^2 \leq 4 \|R_i^* \underline{R}_i^\mathsf{T} - X^\mathsf{T}\|_F^2. \tag{A.34}$$

Since this bound is satisfied for each summand, the total summation satisfies

$$\begin{aligned}\sum_{i=1}^{n} \|R_i^* \underline{R}_i^{\mathsf{T}} - \Pi_{\mathcal{S}}(X^{\mathsf{T}})\|_F^2 &\leq 4 \sum_{i=1}^{n} \|R_i^* \underline{R}_i^{\mathsf{T}} - X^{\mathsf{T}}\|_F^2 \\ &= 4 \sum_{i=1}^{n} \|R_i^* - X^{\mathsf{T}} \underline{R}_i\|_F^2 \\ &= 4\|R^* - X^{\mathsf{T}} \underline{R}\|_F^2.\end{aligned} \tag{A.35}$$

Since $K = X^{\mathsf{T}} \underline{R}$, we have:

$$\begin{aligned}4\|R^* - X^{\mathsf{T}} \underline{R}\|_F^2 &= 4\|R^* - K\|_F^2 \\ &= 4\|P\|_F^2,\end{aligned} \tag{A.36}$$

which gives the desired bound. □

With this result, we are ready to prove Theorem 6.

Proof. From (A.29) and (A.26), we have:

$$\lambda_{d+1}(\underline{Q})\|P\|_F^2 \leq 2\sqrt{dn}\|\underline{R} - R^*\|_F \|\Delta Q\|_2. \tag{A.37}$$

Since, by hypothesis, R^* and $\underline{R}$ are the representatives of their orbits satisfying $d_{\mathcal{S}}(\underline{R}, R^*) = \|\underline{R} - R^*\|_F$, from Lemma 14 we have

$$d_{\mathcal{S}}(\underline{R}, R^*)^2 \leq 4\|P\|_F^2. \tag{A.38}$$

Combining (A.38) with (A.37), we obtain:

$$d_{\mathcal{S}}(\underline{R}, R^*) \leq \frac{8\sqrt{dn}\|\Delta Q\|_2}{\lambda_{d+1}(\underline{Q})}, \tag{A.39}$$

which is what we intended to show. □

A.3.3 An upper bound on $d_{\mathcal{S}}(R^{(0)}, R^*)$

In this section, we give a proof of Corollary 7, bounding the $\mathrm{SO}(d)^n$ orbit distance between the spectral initialization $R^{(0)}$ and the maximum likelihood estimate R^*. First, we establish as the main technical lemma a result that the orbit distances $d_{\mathcal{S}}$ and $d_{\mathcal{O}}$ on $\mathrm{SO}(d)^n$ and $\mathrm{O}(d)^n$ are *pseudometrics*:

Lemma 15 (Orbit distances are pseudometrics)**.** *The orbit distances $d_{\mathcal{S}}$ and $d_{\mathcal{O}}$ are* pseudometrics *on $\mathrm{SO}(d)^n$ and $\mathrm{O}(d)^n$, respectively. In particular, for all $X, Y, Z \in \mathrm{SO}(d)^n$, we have:*

1. $d_{\mathcal{S}}(X, X) = 0$
2. $d_{\mathcal{S}}(X, Y) = d_{\mathcal{S}}(Y, X)$
3. $d_{\mathcal{S}}(X, Z) \leq d_{\mathcal{S}}(X, Y) + d_{\mathcal{S}}(Y, Z)$,

and likewise for $d_{\mathcal{O}}$ on $\mathrm{O}(d)^n$.

Proof. To simplify the subsequent derivation, we prove the result for the orbit distance $d_{\mathcal{S}}$ on $\mathrm{SO}(d)^n$; the same argument applies *mutatis mutandis* to $d_{\mathcal{O}}$ on $\mathrm{O}(d)^n$. A *pseudometric* on $\mathrm{SO}(d)^n$ (resp. $\mathrm{O}(d)^n$) is any nonnegative function $\mathrm{SO}(d)^n \times \mathrm{SO}(d)^n \to \mathbb{R}_{\geq 0}$ satisfying the properties 1–3 [71]. To establish 1, we have:

$$d_{\mathcal{S}}(X, X) = \min_{G \in \mathrm{SO}(d)} \|X - GX\|_F = 0, \tag{A.40}$$

since $\|A\|_F \geq 0$ for all A and taking $G = I$ realizes this minimum value.

For 2, we have:

$$\begin{aligned} d_{\mathcal{S}}(X, Y) &= \min_{G \in \mathrm{SO}(d)} \|X - GY\|_F \\ &= \min_{G \in \mathrm{SO}(d)} \|Y - G^{\mathsf{T}} X\|_F \quad = d_{\mathcal{S}}(Y, X), \end{aligned} \tag{A.41}$$

where the second line follows from the orthogonal invariance of the Frobenius norm, and the last line follows from the fact that since $G^{\mathsf{T}} = G^{-1} \in \mathrm{SO}(d)$, then G^{T} ranges over all of $\mathrm{SO}(d)$ as G does.

Finally, to establish 3, we aim to prove that for any $X, Y, Z \in \mathrm{SO}(d)^n$:

$$d_{\mathcal{S}}(X, Z) \leq d_{\mathcal{S}}(X, Y) + d_{\mathcal{S}}(Y, Z). \tag{A.42}$$

Suppose the orbit distance $d_{\mathcal{S}}(X, Y)$ is attained with minimizer $G^*_{XY} \in \mathrm{SO}(d)$ and likewise the distance $d_{\mathcal{S}}(Y, Z)$ is attained with minimizer $G^*_{YZ} \in \mathrm{SO}(d)$. Define:

$$G' \triangleq G^*_{XY} G^*_{YZ}. \tag{A.43}$$

Now, since G' is itself the product of two elements of $\mathrm{SO}(d)$, we know $G' \in \mathrm{SO}(d)$, and therefore:

$$d_{\mathcal{S}}(X, Z) = \min_{G \in \mathrm{SO}(d)} \|X - GZ\|_F \leq \|X - G'Z\|_F. \tag{A.44}$$

Examining the right-hand side of this expression, we have:

$$\begin{aligned} \|X - G'Z\|_F &= \|X - G^*_{XY}Y + G^*_{XY}Y - G'Z\|_F \\ &\leq \underbrace{\|X - G^*_{XY}Y\|_F}_{d_{\mathcal{S}}(X,Y)} + \|G^*_{XY}Y - G'Z\|_F, \end{aligned} \tag{A.45}$$

where the last line follows from the triangle inequality for the Frobenius norm. Now, substitution of the definition (A.43) into the second term of (A.45) reveals:

$$\begin{aligned} \|G^*_{XY}Y - G'Z\|_F &= \|G^*_{XY}Y - G^*_{XY}G^*_{YZ}Z\|_F \\ &= \|Y - G^*_{YZ}Z\|_F \\ &= d_{\mathcal{S}}(Y, Z), \end{aligned} \tag{A.46}$$

where the second line follows from the orthogonal invariance of the Frobenius norm. Taken together, these results give:

$$d_{\mathcal{S}}(X, Z) \leq \|X - G'Z\|_F \leq d_{\mathcal{S}}(X, Y) + d_{\mathcal{S}}(Y, Z), \tag{A.47}$$

which is what we intended to show. □

Lemma 15 suggests a straightforward proof of Corollary 7.

Proof. From the triangle inequality for $d_{\mathcal{S}}$, we have:

$$d_{\mathcal{S}}(R^{(0)}, R^*) \leq d_{\mathcal{S}}(\underline{R}, R^{(0)}) + d_{\mathcal{S}}(\underline{R}, R^*). \tag{A.48}$$

Substitution of (5.17) and (5.18) into (A.48) gives the desired result. □

A.4 Relationship to the method of Moreira et al. [95]

In their recent work, Moreira et al. [95] also propose an estimator for pose-graph SLAM problems based on eigenvector computations. In this section, we show that their approach is formally equivalent to the *rotation-only* variant of the spectral initialization we discuss in Section 5.3 and therefore has estimation error satisfying the bound (5.21). Moreira et al. [95] specifically consider *unweighted* rotation measurements, which (from an estimation standpoint) is equivalent to considering the generative model (5.1) with *identical* precisions (say $\kappa_{ij} = 1$) for all edges $(i, j) \in \mathcal{E}$.

Their construction begins by considering the matrix $\tilde{M} \in \mathbb{R}^{dn \times dn}$ with $d \times d$ block i, j given by:

$$\tilde{M}_{ij} = \begin{cases} I_d & \text{if } i = j \\ \tilde{R}_{ij}, & \{i, j\} \in \mathcal{E} \\ 0_{d\times d} & \{i, j\} \notin \mathcal{E}. \end{cases} \tag{A.49}$$

They observe that for all stationary points $\hat{R} \in \mathrm{SO}(d)^n \subset \mathbb{R}^{d \times dn}$, there is a corresponding matrix $\Lambda \in \mathbb{R}^{dn \times dn}$ such that:

$$\underbrace{(\Lambda - \tilde{M})}_{\tilde{S}} \hat{R}^{\mathsf{T}} = 0, \tag{A.50}$$

where Λ has the symmetric $d \times d$ block diagonal structure:

$$\Lambda = \begin{bmatrix} \Lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Lambda_n \end{bmatrix}. \tag{A.51}$$

In the *noiseless* case where $\tilde{M} = \underline{M}$,[2] the matrix $\underline{S} = \Lambda - \underline{M}$ is given by [95, Equation 14]:

$$\underline{S} = (\mathcal{L} \otimes J_d) \circ \underline{M}, \tag{A.52}$$

where $\mathcal{L}$ is the scalar (unweighted) rotational graph Laplacian with i, j entry:

$$\mathcal{L}_{ij} = \begin{cases} \delta(i), & i = j, \\ -1, & \{i,j\} \in \mathcal{E}, \\ 0, & \{i,j\} \notin \mathcal{E}, \end{cases} \tag{A.53}$$

$J_d \in \mathbb{R}^{d \times d}$ is an all-ones matrix, and $\circ$ denotes the Hadamard product. Direct comparison of (A.53) with (A.1b) reveals that $\mathcal{L}$ is equivalent to $L(W^\rho)$ when $\kappa_{ij} = 1$ for all $\{i,j\} \in \mathcal{E}$. Expanding (A.52), we have:

$$\underline{S}_{ij} = \begin{cases} \delta(i) I_d, & i = j, \\ -\underline{R}_{ij}, & \{i,j\} \in \mathcal{E}, \\ 0_{d \times d}, & \{i,j\} \notin \mathcal{E} \end{cases}. \tag{A.54}$$

Comparing the definition of $L(\tilde{G}^\rho)$ in (A.2a) and $\underline{S}$ in (A.54), it is straightforward to verify that $\underline{S} = L(\underline{G}^\rho)$ when $\kappa_{ij} = 1$. From the equivalence of $\underline{S}$ and $L(\underline{G}^\rho)$, it follows that $\underline{S} \succeq 0$ and $\underline{R}^\mathsf{T} \in \ker(\underline{S})$, so the ground-truth rotations $\underline{R}$ can be recovered by computing the d eigenvectors of $\underline{S}$ corresponding to the smallest eigenvalues of $\underline{S}$.[3]

[2] In keeping with the notation in the rest of this manuscript, we use the notation $\underline{M}$ to denote the measurement matrix (A.49) constructed from the *ground-truth* relative rotations $\underline{R}_{ij}$.

[3] Recall from Section 5.2 that $\underline{R}$ lie in $\ker(L(\underline{G}^\rho))$ and from Section 5.1 that $L(\underline{G}^\rho) \succeq 0$. The claim then follows from the equivalence of $\underline{S}$ and $L(\underline{G}^\rho)$.

In the case of noisy measurements, Moreira et al. [95] propose to compute, as an approximation, the eigenvectors of $\tilde{S} = (\mathcal{L} \otimes J_3) \circ \tilde{M}$, which has $d \times d$ blocks given by:

$$S_{ij} = \begin{cases} \delta(i) I_d, & i = j, \\ -\tilde{R}_{ij}, & \{i, j\} \in \mathcal{E}, \\ 0_{d \times d}, & \{i, j\} \notin \mathcal{E}. \end{cases} \tag{A.55}$$

The justification given for this approximation is that, in the high signal-to-noise ratio regime, there ought to exist $R \in \mathrm{SO}(d)^n$ such that $\tilde{S}R \approx 0$. Once again, however, directly comparing definitions reveals that the quantity $(\mathcal{L} \otimes J_3) \circ \tilde{M}$ is identical to $L(\tilde{G}^{\rho})$ with $\kappa_{ij} = 1$ (cf. equations (A.55) and (A.2a)). Consequently, Moreira et al. [95]'s method is actually a *particular instance* of the spectral estimator we propose in Section 5.2, corresponding to the special case in which all rotational measurements have *equal weights* and the translational measurements have been discarded (i.e. the rotation-only case discussed in Section 5.3). Moreover, viewing this approach through the lens of the spectral relaxation in Problem 6 provides formal justification for the method and allows us to derive the explicit performance guarantees given in this thesis.

Appendix B

Proofs for Chapter 6

B.1 Subgradients of the Fiedler value

In this appendix we consider the problem of computing a supergradient of $F(\omega) = \lambda_2(L(W^\rho)(\omega))$ with respect to ω. Strictly speaking, F need not be differentiable at a particular ω (which occurs specifically when $\lambda_2(L(W^\rho)(\omega))$ appears with multiplicity greater than 1, i.e. it is not a *simple* eigenvalue). We say that a vector $g \in \mathbb{R}^m$ is a *supergradient* of a concave function F at ω if, for all y, x in the domain of F:

$$F(y) - F(x) \leq g^\mathsf{T}(y - x). \tag{B.1}$$

Equation (B.1) generalizes the notion of differentiability to the scenario where the function F may not be (uniquely) differentiable a particular point. We call the set of all supergradients at a particular value of ω the *superdifferential* of F at ω, denoted $\partial F(\omega)$ [116].

We aim to prove the statement that $\nabla F(\omega)$ as defined in equation (6.5) is a supergradient of F at ω.

Proof of Theorem 9. We aim to prove the claim by way of equation (B.1). Let $u, v \in \mathbb{R}^m$, $\|u\|_2 = \|v\|_2 = 1$ be *any* normalized eigenvectors of $L(W^\rho)(x)$ and $L(W^\rho)(y)$ with corresponding eigenvalues $\lambda_2(L(W^\rho)(x))$ and $\lambda_2(L(W^\rho)(y))$, respectively. By definition, then, u and v are Fiedler vectors of $L(W^\rho)(x)$ and $L(W^\rho)(y)$, respectively.

Then the left-hand side of equation (B.1) can be written as:

$$\begin{aligned} F(y) - F(x) &= \lambda_2(L(W^\rho)(y)) - \lambda_2(L(W^\rho)(x)) \\ &= v^\mathsf{T} L(W^\rho)(y) v - u^\mathsf{T} L(W^\rho)(x) u. \end{aligned} \tag{B.2}$$

Now, substitution of u for the Fiedler vector into the definition in (6.5), reveals that the k-th element of $\nabla F(x)$ is:

$$\nabla F(x)_k = u^\mathsf{T} L_k^c u. \tag{B.3}$$

In turn, the right-hand side of (B.1) can be written as:

$$\nabla F(x)^\mathsf{T}(y - x) = \sum_{k=1}^{m} (y_k - x_k) u^\mathsf{T} L_k^c u. \tag{B.4}$$

Since u and v are minimizers of their respective Rayleigh quotient minimization problems, we know:

$$\begin{aligned} v^\mathsf{T} L(W^\rho)(y) v &\leq u^\mathsf{T} L(W^\rho)(y) u \\ &= u^\mathsf{T} L^o u + \sum_{k=1}^{m} y_k u^\mathsf{T} L_k^c u, \end{aligned} \tag{B.5}$$

where the first line follows from the optimality of v with respect to the Rayleigh quotient for $L(W^\rho)(y)$ and the second line follows from the definition of $L(W^\rho)(y)$. Consider "adding zero" to each y_k in (B.5) as $x_k - x_k$ to obtain an equivalent expression:

$$\begin{aligned} \sum_{k=1}^{m} y_k u^\mathsf{T} L_k^c u &= \sum_{k=1}^{m} (y_k + x_k - x_k) u^\mathsf{T} L_k^c u, \\ &= \sum_{k=1}^{m} x_k u^\mathsf{T} L_k^c u + \sum_{k=1}^{m} (y_k - x_k) u^\mathsf{T} L_k^c u. \end{aligned} \tag{B.6}$$

Comparison to (B.4) reveals that the last term in (B.6) is *exactly* equal to $\nabla F(x)^\mathsf{T}(y-$

x). In turn, substitution of (B.4) into (B.6) gives:

$$\sum_{k=1}^{m} y_k u^\mathsf{T} L_k^c u = \sum_{k=1}^{m} x_k u^\mathsf{T} L_k^c u + \nabla F(x)^\mathsf{T}(y - x). \tag{B.7}$$

Substitution back into (B.5) gives the bound:

$$v^\mathsf{T} L(W^\rho)(y) v \leq u^\mathsf{T} L^o u + \sum_{k=1}^{m} x_k u^\mathsf{T} L_k^c u + \nabla F(x)^\mathsf{T}(y - x). \tag{B.8}$$

Finally, from the definition of $L(W^\rho)(x)$, we obtain

$$v^\mathsf{T} L(W^\rho)(y) v \leq u^\mathsf{T} L(W^\rho)(x) u + \nabla F(x)^\mathsf{T}(y - x). \tag{B.9}$$

Subtracting $u^\mathsf{T} L(W^\rho)(x) u$ from both sides and substituting into (B.2) gives the desired result. □

B.2 Solving the direction-finding subproblem

This appendix aims to prove the claim that (6.6) provides an optimal solution to the linear program in Problem 9.

Proof of Theorem 10. Rewriting the objective from Problem 9 in terms of the elements of s and $\nabla F(\omega)$, we have:

$$\begin{aligned} s^\mathsf{T} \nabla F(\omega) &= \sum_{k=1}^{m} s_k \nabla F(\omega)_k \\ &= \sum_{k=1}^{m} s_k y^*(\omega)^\mathsf{T} L_k^c y^*(\omega), \end{aligned} \tag{B.10}$$

where in the last line we have used the definition of $\nabla F(\omega)_k$ in (6.5). Now, since each $L_k^c \succeq 0$, every component of the gradient must always be nonnegative, i.e. $\nabla F(\omega)_k \geq 0$. Further, since $0 \leq s_k \leq 1$, the objective in (B.10) is itself a sum of nonnegative terms. From this, it follows directly that the objective in (B.10) is maximized (subject to the constraint that $\sum_{k=1}^{m} s_k = K$) specifically by selecting (i.e.

by setting $s_k = 1$) each of the K largest components of $\nabla F(\omega)$, giving the result in (6.6). $\square$

Bibliography

[1] StereoLabs ZED camera. URL `https://www.stereolabs.com/zed/`.

[2] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[3] Pratik Agarwal, Gian Diego Tipaldi, Luciano Spinello, Cyrill Stachniss, and Wolfram Burgard. Robust map optimization using dynamic covariance scaling. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 62–69, 2013.

[4] Sameer Agarwal and Keir Mierle. *Ceres Solver: Tutorial & Reference*. Google Inc.

[5] Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. Spectral synchronization of multiple views in SE(3). *SIAM Journal on Imaging Sciences*, 9(4): 1963–1990, 2016.

[6] Afonso S Bandeira, Nicolas Boumal, and Amit Singer. Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163(1-2):145–167, 2017.

[7] Yaakov Bar-Shalom and Edison Tse. Tracking in a cluttered environment with probabilistic data association. *Automatica*, 11(5):451–460, 1975.

[8] Timothy D Barfoot. *State Estimation for Robotics*. Cambridge University Press, 2017.

[9] Hriday Bavle, Jose Luis Sanchez-Lopez, Muhammad Shaheer, Javier Civera, and Holger Voos. Situational graphs for robot navigation in structured indoor environments. *arXiv preprint arXiv:2202.12197*, 2022.

[10] David P Baxter et al. *Toward robust active semantic SLAM via Max-Mixtures*. PhD thesis, Massachusetts Institute of Technology, 2020.

[11] Dimitri Bertsekas. *Nonlinear Programming*, volume 4. Athena Scientific, 2016.

[12] Paul J Besl and Neil D McKay. Method for registration of 3-D shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, pages 586–606. International Society for Optics and Photonics, 1992.

[13] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[14] Byron Boots and Geoff Gordon. A spectral learning approach to range-only slam. In *International Conference on Machine Learning*, pages 19–26. PMLR, 2013.

[15] Nicolas Boumal. Nonconvex phase synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

[16] Nicolas Boumal, Amit Singer, and P-A Absil. Robust estimation of rotations from relative measurements by maximum likelihood. In *52nd IEEE Conference on Decision and Control*, pages 1156–1161. IEEE, 2013.

[17] Nicolas Boumal, Amit Singer, P-A Absil, and Vincent D Blondel. Cramér–rao bounds for synchronization of rotations. *Information and Inference: A Journal of the IMA*, 3(1):1–39, 2014.

[18] Sean L Bowman, Nikolay Atanasov, Kostas Daniilidis, and George J Pappas. Probabilistic data association for semantic SLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1722–1729, 2017.

[19] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[20] Jesus Briales and Javier Gonzalez-Jimenez. Cartan-Sync: Fast and global SE(d)-synchronization. *IEEE Robotics and Automation Letters*, 2(4):2127–2134, 2017.

[21] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016.

[22] Nicholas Carlevaris-Bianco and Ryan M. Eustice. Long-term simultaneous localization and mapping with generic linear constraint node removal. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Tokyo, Japan, November 2013.

[23] Nicholas Carlevaris-Bianco and Ryan M Eustice. Conservative edge sparsification for graph SLAM node removal. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 854–860. IEEE, 2014.

[24] Luca Carlone. Estimation contracts for outlier-robust geometric perception. *arXiv preprint arXiv:2208.10521*, 2022.

[25] Luca Carlone and Giuseppe C Calafiore. Convex relaxations for pose graph optimization with outliers. *IEEE Robotics and Automation Letters*, 3(2):1160–1167, 2018.

[26] Luca Carlone and Frank Dellaert. Duality-based verification techniques for 2D SLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*.

[27] Luca Carlone, Andrea Censi, and Frank Dellaert. Selecting good measurements via ℓ-1 relaxation: A convex approach for robust estimation over graphs. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2667–2674. IEEE, 2014.

[28] Luca Carlone, Zsolt Kira, Chris Beall, Vadim Indelman, and Frank Dellaert. Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 4290–4297. IEEE, 2014.

[29] Luca Carlone, David M Rosen, Giuseppe Calafiore, John J Leonard, and Frank Dellaert. Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 125–132. IEEE, 2015.

[30] Luca Carlone, Roberto Tron, Kostas Daniilidis, and Frank Dellaert. Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4597–4604. IEEE, 2015.

[31] Luca Carlone, Giuseppe C Calafiore, Carlo Tommolillo, and Frank Dellaert. Planar pose graph optimization: Duality, optimal solutions, and verification. *IEEE Transactions on Robotics*, 32(3):545–565, 2016.

[32] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.

[33] Yongbo Chen, Shoudong Huang, Liang Zhao, and Gamini Dissanayake. Cramér–Rao bounds and optimal design metrics for pose-graph SLAM. *IEEE Transactions on Robotics*, 37(2):627–641, 2021.

[34] Contributors, Ecosystem, and NavAbility. Caesar.jl, v0.11.1, 2021. `https://github.com/JuliaRobotics/Caesar.jl`.

[35] Ingemar J Cox and John J Leonard. Probabilistic data association for dynamic world modeling: A multiple hypothesis approach. In *Advanced Robotics, 1991.'Robots in Unstructured Environments', 91 ICAR., Fifth International Conference on*, pages 1287–1294. IEEE, 1991.

[36] Ingemar J Cox and John J Leonard. Modeling a dynamic environment using a Bayesian multiple hypothesis approach. *Artificial Intelligence*, 66(2):311–344, 1994.

[37] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, pages 303–312, 1996.

[38] Frank Dellaert. Factor graphs and GTSAM: A hands-on introduction. Technical report, Georgia Institute of Technology, 2012.

[39] Frank Dellaert. Factor graphs: Exploiting structure in robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):141–166, 2021. doi: 10.1146/annurev-control-061520-010504. URL `https://doi.org/10.1146/annurev-control-061520-010504`.

[40] Frank Dellaert, Michael Kaess, et al. Factor graphs for robot perception. *Foundations and Trends in Robotics*, 6(1-2):1–139, 2017.

[41] Frank Dellaert, David M Rosen, Jing Wu, Robert Mahony, and Luca Carlone. Shonan rotation averaging: Global optimality by surfing $\mathrm{SO(p)^n}$. In *European Conference on Computer Vision*, pages 292–308. Springer, 2020.

[42] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[43] Kevin Doherty. Robust, Non-Gaussian Semantic Simultaneous Localization and Mapping. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2020.

[44] Kevin Doherty, Genevieve Flaspohler, Nicholas Roy, and Yogesh Girdhar. Approximate Distributed Spatiotemporal Topic Models for Multi-Robot Terrain Characterization. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3730–3737. IEEE, 2018.

[45] Kevin Doherty, Dehann Fourie, and John J Leonard. Multimodal semantic SLAM with probabilistic data association. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2019.

[46] Kevin Doherty, David Baxter, Edward Schneeweiss, and John J. Leonard. Probabilistic data association via mixture models for robust semantic SLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2020.

[47] Kevin J Doherty, Ziqi Lu, Kurran Singh, and John J Leonard. Discrete-Continuous Smoothing and Mapping. *IEEE Robotics and Automation Letters*, October 2022.

[48] Kevin J Doherty, David M Rosen, and John J Leonard. Performance Guarantees for Spectral Initialization in Rotation Averaging and Pose-Graph SLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2022.

[49] Kevin J Doherty, David M Rosen, and John J Leonard. Spectral Measurement Sparsification for Pose-Graph SLAM. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022.

[50] Joseph C Dunn and S Harshbarger. Conditional gradient algorithms with open loop step size rules. *Journal of Mathematical Analysis and Applications*, 62(2): 432–444, 1978.

[51] Taosha Fan, Hanlin Wang, Michael Rubenstein, and Todd Murphey. CPL-SLAM: Efficient and certifiably correct planar graph-based SLAM using the complex number representation. *IEEE Transactions on Robotics*, 36(6):1719–1737, 2020.

[52] Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2):298–305, 1973.

[53] D. Fourie, J.J. Leonard, and M. Kaess. A nonparametric belief solution to the Bayes tree. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Daejeon, Korea, Oct 2016.

[54] Caelan Reed Garrett, Rohan Chitnis, Rachel Holladay, Beomjoon Kim, Tom Silver, Leslie Pack Kaelbling, and Tomás Lozano-Pérez. Integrated Task and Motion Planning. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):265–293, 2021. doi: 10.1146/annurev-control-091420-084139. URL `https://doi.org/10.1146/annurev-control-091420-084139`.

[55] Andreas Geiger, Julius Ziegler, and Christoph Stiller. StereoScan: Dense 3d Reconstruction in Real-time. In *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.

[56] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the KITTI Vision Benchmark Suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[57] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets Robotics: The KITTI Dataset. *Intl. J. of Robotics Research*, 2013.

[58] Arpita Ghosh and Stephen Boyd. Growing well-connected graphs. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 6605–6611. IEEE, 2006.

[59] Giorgio Grisetti, Rainer Kummerle, Cyrill Stachniss, and Wolfram Burgard. A tutorial on graph-based slam. *IEEE Intelligent Transportation Systems Magazine*, 2(4):31–43, 2010.

[60] Giorgio Grisetti, Rainer Kümmerle, Hauke Strasdat, and Kurt Konolige. g2o: A general framework for (hyper) graph optimization. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 9–13, 2011.

[61] Michael Grupp. evo: Python package for the evaluation of odometry and SLAM. `https://github.com/MichaelGrupp/evo`, 2017.

[62] Richard J Hanson and Michael J Norris. Analysis of measurements based on the singular value decomposition. *SIAM Journal on Scientific and Statistical Computing*, 2(3):363–373, 1981.

[63] Ming Hsiao and Michael Kaess. MH-iSAM2: Multi-hypothesis iSAM using Bayes Tree and Hypo-tree. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1274–1280. IEEE, 2019.

[64] G. Huang, M. Kaess, and J.J. Leonard. Consistent sparsification for graph optimization. In *Proc. of European Conference on Mobile Robots (ECMR)*, pages 150–157, Barcelona, Spain, September 25–27, 2013.

[65] Qiangqiang Huang, Can Pu, Dehann Fourie, Kasra Khosoussi, Jonathan P How, and John J Leonard. NF-iSAM: Incremental Smoothing and Mapping via Normalizing Flows. *arXiv preprint arXiv:2105.05045*, 2021.

[66] Nathan Hughes, Yun Chang, and Luca Carlone. Hydra: A real-time spatial perception engine for 3D scene graph construction and optimization. *arXiv preprint arXiv:2201.13360*, 2022.

[67] Fan Jiang, Varun Agrawal, Russell Buchanan, Maurice Fallon, and Frank Dellaert. iMHS: An Incremental Multi-Hypothesis Smoother. *arXiv preprint arXiv:2103.13178*, 2021.

[68] H. Johannsson, M. Kaess, M.F. Fallon, and J.J. Leonard. Temporally scalable visual SLAM using a reduced pose graph. In *RSS Workshop on Long-term Operation of Autonomous Robotic Systems in Changing Environments*, Sydney, Australia, July 2012. Available as MIT CSAIL Technical Report MIT-CSAIL-TR-2012-013.

[69] M. Kaess and F. Dellaert. Covariance recovery from a square root information matrix for data association. *J. of Robotics and Autonomous Systems*, 57(12): 1198–1210, December 2009. doi: 10.1016/j.robot.2009.06.008.

[70] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert. iSAM2: Incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research*, 31:217–236, February 2012.

[71] John L Kelley. *General Topology*. Van Nostrand, New York, 1955.

[72] Kasra Khosoussi, Shoudong Huang, and Gamini Dissanayake. Novel insights into the impact of graph structure on SLAM. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2707–2714. IEEE, 2014.

[73] Kasra Khosoussi, Matthew Giamou, Gaurav S Sukhatme, Shoudong Huang, Gamini Dissanayake, and Jonathan P How. Reliable graphs for SLAM. *The International Journal of Robotics Research*, 38(2-3):260–298, 2019.

[74] A. Kim and R.M. Eustice. Toward AUV survey design for optimal coverage and localization using the Cramer Rao lower bound. In *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, 2009.

[75] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, Cambridge, MA, 2009.

[76] Henrik Kretzschmar and Cyrill Stachniss. Information-theoretic compression of pose graphs for laser-based SLAM. *The International Journal of Robotics Research*, 31(11):1219–1230, 2012.

[77] Gerhard Kurz, Matthias Holoch, and Peter Biber. Geometry-based graph pruning for lifelong SLAM. *arXiv preprint arXiv:2110.01286*, 2021.

[78] Simon Lacoste-Julien, Martin Jaggi, Mark Schmidt, and Patrick Pletscher. Block-coordinate frank-wolfe optimization for structural svms. In *International Conference on Machine Learning*, pages 53–61. PMLR, 2013.

[79] Pierre-Yves Lajoie, Siyi Hu, Giovanni Beltrame, and Luca Carlone. Modeling perceptual aliasing in SLAM via discrete–continuous graphical models. *IEEE Robotics and Automation Letters*, 4(2):1232–1239, 2019.

[80] Shuyang Ling. Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *arXiv preprint arXiv:2008.05341*, 2020.

[81] Huikang Liu, Man-Chung Yue, and Anthony Man-Cho So. A unified approach to synchronization problems over subgroups of the orthogonal group. *arXiv preprint arXiv:2009.07514*, 2020.

[82] S Lowry, N Sunderhauf, P Newman, J. J. Leonard, D Cox, P Corke, and M. Milford. Visual place recognition: A survey. *IEEE Trans. Robotics*, 32(1):1–19, 2016.

[83] Tomás Lozano-Pérez. *Autonomous robot vehicles*. Springer Science & Business Media, 2012.

[84] Ziqi Lu, Qiangqiang Huang, Kevin Doherty, and John J. Leonard. Consensus-Informed Optimization Over Mixtures for Ambiguity-Aware Object SLAM. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2021.

[85] Ziqi Lu, Yihao Zhang, Kevin Doherty, Odin Severinsen, Ethan Yang, and John Leonard. SLAM-supported self-training for 6D object pose estimation. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 2833–2840. IEEE, 2022.

[86] David John Cameron Mackay. Introduction to Monte Carlo Methods. In *Learning in graphical models*, pages 175–204. Springer, 1998.

[87] Danylo Malyuta, Christian Brommer, Daniel Hentzen, Thomas Stastny, Roland Siegwart, and Roland Brockers. Long-duration fully autonomous operation of rotorcraft unmanned aerial systems for remote-sensing data acquisition. *Journal of Field Robotics*, page arXiv:1908.06381, August 2019. doi: 10.1002/rob.21898. URL `https://doi.org/10.1002/rob.21898`.

[88] Joshua G Mangelson, Derrick Dominic, Ryan M Eustice, and Ram Vasudevan. Pairwise consistent measurement set maximization for robust multi-robot map merging. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 2916–2923, 2018.

[89] Murat Manguoglu, Eric Cox, Faisal Saied, and Ahmed Sameh. TRACEMIN-Fiedler: A parallel algorithm for computing the Fiedler vector. In *International Conference on High Performance Computing for Computational Science*, pages 449–455. Springer, 2010.

[90] Daniel Martinec and Tomas Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[91] John McCormac, Ankur Handa, Andrew Davison, and Stefan Leutenegger. SemanticFusion: Dense 3D semantic mapping with convolutional neural networks. In *International Conference on Robotics and Automation (ICRA)*, pages 4628–4635, 2017.

[92] John McCormac, Ronald Clark, Michael Bloesch, Andrew J Davison, and Stefan Leutenegger. Fusion++: Volumetric Object-Level SLAM. *arXiv preprint arXiv:1808.08378*, 2018.

[93] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *AAAI Conf. on Artificial Intelligence*, Edmonton, Canada, 2002. AAAI.

[94] Michael Montemerlo and Sebastian Thrun. Simultaneous localization and mapping with unknown data association using fastSLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, volume 2, pages 1985–1991, 2003.

[95] Gabriel Moreira, Manuel Marques, and Joao Paulo Costeira. Fast Pose Graph Optimization via Krylov-Schur and Cholesky Factorization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1898–1906, 2021.

[96] Damon Mosk-Aoyama. Maximum algebraic connectivity augmentation is NP-hard. *Operations Research Letters*, 36(6):677–679, 2008.

[97] B. Mu, S-Y. Liu, L. Paull, J. Leonard, and J. How. SLAM with objects using a nonparametric pose graph. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2016.

[98] J. Mullane, B.-N. Vo, M.D. Adams, and B.-T. Vo. A random-finite-set approach to Bayesian SLAM. *IEEE Trans. Robotics*, 27(2):268–282, April 2011.

[99] Harsha Nagarajan. On maximizing weighted algebraic connectivity for synthesizing robust networks. *arXiv preprint arXiv:1805.07825*, 2018.

[100] José Neira and Juan D Tardós. Data association in stochastic mapping using the joint compatibility test. *IEEE Transactions on robotics and automation*, 17 (6):890–897, 2001.

[101] Lachlan Nicholson, Michael Milford, and Niko Sünderhauf. QuadricSLAM: Constrained Dual Quadrics from Object Detections as Landmarks in Semantic SLAM. *IEEE Robotics and Automation Letters*, 2018.

[102] Kyel Ok, Katherine Liu, Kris Frey, Jonathan P How, and Nicholas Roy. Robust object-based slam for high-speed autonomous navigation. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 669–675, 2019.

[103] Kyel Ok, Katherine Liu, and Nicholas Roy. Hierarchical object map estimation for efficient and robust navigation. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1132–1139. IEEE, 2021.

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

[105] Edwin Olson and Pratik Agarwal. Inference on networks of mixtures for robust robot mapping. *Intl. J. of Robotics Research*, 32(7):826–840, 2013.

[106] Tim Pfeifer and Peter Protzel. Robust sensor fusion with self-tuning mixture models. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3678–3685. IEEE, 2018.

[107] Tim Pfeifer and Peter Protzel. Expectation-maximization for adaptive mixture models in graph optimization. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3151–3157. IEEE, 2019.

[108] Tim Pfeifer, Sven Lange, and Peter Protzel. Advancing mixture models for least squares optimization. *IEEE Robotics and Automation Letters*, 6(2):3941–3948, 2021.

[109] Luis Pineda, Taosha Fan, Maurizio Monge, Shobha Venkataraman, Paloma Sodhi, Ricky Chen, Joseph Ortiz, Daniel DeTone, Austin Wang, Stuart Anderson, et al. Theseus: A library for differentiable nonlinear optimization. *arXiv preprint arXiv:2207.09442*, 2022.

[110] Brian Patrick Preskitt. *Phase retrieval from locally supported measurements.* PhD thesis, UC San Diego, 2018.

[111] Friedrich Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

[112] Yixuan Qiu. Spectra: C++ library for large scale eigenvalue problems. https://spectralib.org, 2015.

[113] Darío Ramos-López, Andrés R Masegosa, Ana M Martínez, Antonio Salmerón, Thomas D Nielsen, Helge Langseth, and Anders L Madsen. MAP inference in dynamic hybrid Bayesian networks. *Progress in Artificial Intelligence*, 6(2): 133–144, 2017.

[114] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[115] Donald Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24(6):843–854, 1979.

[116] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[117] David M Rosen, Michael Kaess, and John J Leonard. An incremental trust-region method for robust online sparse least-squares estimation. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1262–1269. IEEE, 2012.

[118] David M Rosen, Luca Carlone, Afonso S Bandeira, and John J Leonard. SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. *Intl. J. of Robotics Research*, 38(2-3):95–125, 2019.

[119] David M Rosen, Kevin J Doherty, Antonio Terán Espinoza, and John J Leonard. Advances in Inference and Representation for Simultaneous Localization and Mapping. *Annual Review of Control, Robotics, and Autonomous Systems*, 4 (1):215–242, 2021. doi: 10.1146/annurev-control-072720-082553. URL `https://doi.org/10.1146/annurev-control-072720-082553`.

[120] D.M. Rosen, M. Kaess, and J.J. Leonard. Robust incremental online inference over sparse factor graphs: Beyond the Gaussian case. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Karlsruhe, Germany, May 2013.

[121] Antoni Rosinol, Marcus Abate, Yun Chang, and Luca Carlone. Kimera: an open-source library for real-time metric-semantic localization and mapping. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1689–1696. IEEE, 2020.

[122] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the ICP algorithm. In *Proceedings of the Third International Conference on 3-D Digital Imaging and Modeling*, pages 145–152. IEEE, 2001.

[123] Renato F Salas-Moreno, Richard A Newcombe, Hauke Strasdat, Paul HJ Kelly, and Andrew J Davison. SLAM++: Simultaneous localisation and mapping at the level of objects. In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pages 1352–1359, 2013.

[124] Antonio Salmerón, Rafael Rumí, Helge Langseth, Thomas D Nielsen, and Anders L Madsen. A review of inference algorithms for hybrid bayesian networks. *Journal of Artificial Intelligence Research*, 62:799–828, 2018.

[125] Ahmed H Sameh and John A Wisniewski. A trace minimization algorithm for the generalized eigenvalue problem. *SIAM Journal on Numerical Analysis*, 19 (6):1243–1259, 1982.

[126] Bogdan Savchynskyy et al. Discrete graphical models—an optimization perspective. *Foundations and Trends® in Computer Graphics and Vision*, 11 (3-4):160–429, 2019.

[127] Aleksandr V Segal and Ian D Reid. Hybrid inference optimization for robust pose graph estimation. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 2675–2682, 2014.

[128] Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011.

[129] Randall Smith, Matthew Self, and Peter Cheeseman. Estimating uncertain spatial relationships in robotics. In *Autonomous robot vehicles*, pages 167–193. Springer, 1990.

[130] Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

[131] G.W. Stewart, J.W. Stewart, J. Sun, Academic Press (Londyn)., and Harcourt Brace Jovanovich. *Matrix Perturbation Theory*. Computer Science and Scientific Computing. Elsevier Science, 1990. ISBN 9780126702309.

[132] Erik B Sudderth, Alexander T Ihler, Michael Isard, William T Freeman, and Alan S Willsky. Nonparametric belief propagation. *Communications of the ACM*, 53(10):95–103, 2010.

[133] N. Sünderhauf and P. Protzel. Switchable constraints for robust pose graph SLAM. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, October 2012.

[134] Niko Sünderhauf, Oliver Brock, Walter Scheirer, Raia Hadsell, Dieter Fox, Jürgen Leitner, Ben Upcroft, Pieter Abbeel, Wolfram Burgard, Michael Milford, et al. The limits and potentials of deep learning for robotics. *The International Journal of Robotics Research*, 37(4-5):405–420, 2018.

[135] Harold Szu and Ralph Hartley. Fast simulated annealing. *Physics letters A*, 122(3-4):157–162, 1987.

[136] Yulun Tian, Kasra Khosoussi, David M Rosen, and Jonathan P How. Distributed certifiably correct pose-graph optimization. *arXiv preprint arXiv:1911.03721*, 2019.

[137] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

[138] Vasileios Tzoumas, Pasquale Antonante, and Luca Carlone. Outlier-robust spatial perception: Hardness, general-purpose algorithms, and guarantees. *arXiv preprint arXiv:1903.11683*, 2019.

[139] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(04):376–380, 1991.

[140] Andrew Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13 (2):260–269, 1967.

[141] Chen Wang, Dasong Gao, Kuan Xu, Junyi Geng, Yaoyu Hu, Yuheng Qiu, Bowen Li, Fan Yang, Brady Moon, Abhinav Pandey, et al. Pypose: A library for robot learning with physics-based optimization. *arXiv preprint arXiv:2209.15428*, 2022.

[142] Jinkun Wang and Brendan Englot. Robust exploration with multiple hypothesis data association. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3537–3544. IEEE, 2018.

[143] John Wang and Edwin Olson. AprilTag 2: Efficient and robust fiducial detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4193–4198. IEEE, oct 2016. ISBN 978-1-5090-3762-9. doi: 10.1109/IROS.2016.7759617.

[144] Heng Yang and Luca Carlone. Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization. *IEEE Trans. Pattern Anal. Machine Intell.*

[145] Heng Yang, Pasquale Antonante, Vasileios Tzoumas, and Luca Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. *IEEE Robotics and Automation Letters*, 2019.

[146] Shichao Yang and Sebastian Scherer. Cubeslam: Monocular 3-d object slam. *IEEE Trans. Robotics*, 35(4):925–938, 2019.

[147] Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.