Discrete-Continuous Smoothing and Mapping

Kevin J. Doherty, Ziqi Lu, Kurran Singh, and John J. Leonard

Abstract—We describe a general approach for maximum *a posteriori* (MAP) inference in a class of *discrete-continuous factor graphs* commonly encountered in robotics applications. While there are openly available tools providing flexible and easy-to-use interfaces for specifying and solving inference problems formulated in terms of *either* discrete *or* continuous graphical models, at present, no similarly general tools exist enabling the same functionality for *hybrid* discrete-continuous problems. We aim to address this problem. In particular, we provide a library, DC-SAM, extending existing tools for inference problems defined in terms of factor graphs to the setting of discrete-continuous models. A key contribution of our work is a novel solver for efficiently recovering approximate solutions to discrete-continuous inference problems. The key insight to our approach is that while joint inference over continuous and discrete state spaces is often hard, many commonly encountered discrete-continuous problems can naturally be split into a "discrete part" and a "continuous part" that can individually be solved easily. Leveraging this structure, we optimize discrete and continuous variables in an alternating fashion. In consequence, our proposed work enables straightforward representation of and approximate inference in discrete-continuous graphical models. We also provide a method to approximate the uncertainty in estimates of both discrete and continuous variables. We demonstrate the versatility of our approach through its application to distinct robot perception applications, including robust pose graph optimization, and object-based mapping and localization.

Index Terms—SLAM, Localization, Mapping, Hybrid Inference, Factor Graphs

SUPPLEMENTAL MATERIAL

The DC-SAM library is currently available at https://www.github.com/MarineRoboticsGroup/dcsam. Several components of DC-SAM are presently undergoing migration into GTSAM at https://www.github.com/borglab/gtsam. Finally, an extended technical report can be found at [1].

I. INTRODUCTION

PROBABILISTIC graphical models have 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 [2], which provide flexible and modular languages for specifying and solving inference problems in these models (typically in terms of factor graphs). Among the models relevant to robotics
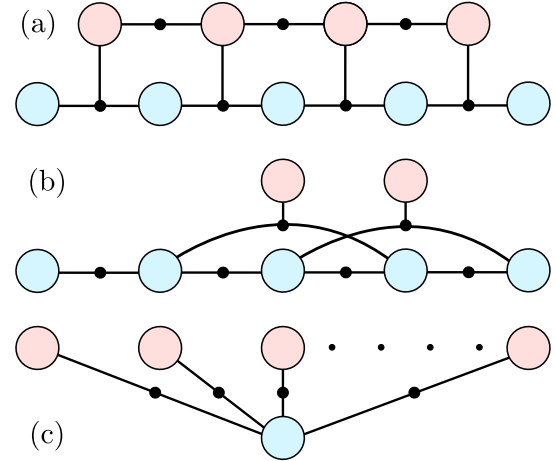
Fig. 1. **Discrete-continuous factor graphs in robotics.** Factor graphs modeling several relevant discrete-continuous robot perception problems. Discrete variable nodes are colored red, continuous variable nodes are blue, and factor nodes are black. (a) Switching systems: discrete states control the 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.

applications, *discrete-continuous graphical models* capture a great breadth of key problems arising in robot perception, task and motion planning [3, Sec 3.2], and navigation, including data association, outlier rejection, and semantic simultaneous localization and mapping (SLAM) [4] (see Figure 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. This is the problem that we consider in this paper.

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. [5, 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 [5, 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 [6] and uncertainty estimation (cf. [7]) 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 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 robust pose graph optimization, and semantic SLAM. In the former case, our general approach naturally recovers well-known solutions. In semantic SLAM, our method produces high-quality trajectory and semantic map estimates incrementally during navigation.

II. RELATED WORK

The problem of inference in discrete-continuous graphical models arises in many domains and intersects a number of communities, even within the field of robotics. Since our focus in this paper will be on applications in robot perception, we primarily discuss related works in these settings. The interested reader may refer to Dellaert [8] for a discussion of these models in broader robotics applications or Koller and Friedman [5, 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 our alternating minimization approach in relation to existing methods, it is important to note: the mere availability of a consistent framework in which these solutions could be implemented 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 [9] and early SLAM applications [10]. MH-iSAM2 [11] extends the capabilities of iSAM2 [6] 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 1; see also [12]), outlier rejection,[1] and as we explore in Section V-B, 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 [12] takes a qualitatively similar approach to MH-iSAM2, but focus 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 [14] for a review). Prior solution methods focus on either *specific* models, such as conditional linear Gaussian models (e.g. [15]) or attempt to approximate more general models in a manner amenable to 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 [4], and solutions are often required quickly. This precludes *direct* application of these techniques to the problems we explore in Section V.

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 [16] 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) [17] and NF-iSAM [18] perform incremental non-Gaussian inference for continuous-valued variables using nonparametric belief propagation [19] 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. [20, 21]) do so by marginalizing out discrete variables and using smooth local optimization techniques on the resulting *continuous-only* estimation problem.

[1]Though we do not explore the issue of outlier rejection problems with correlations, the interested reader may see Lajoie et al. [13] for a formulation in the setting of SLAM.

Consequently, they do not permit the explicit estimation of discrete states, as we consider here.

Existing Tools. Several existing solvers perform optimization with models that can be represented in terms of factor graphs. Ceres [22] and g2o [23] 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 1. GTSAM [2] provides incremental nonlinear least-squares solvers, like iSAM2 [6], 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 [24] implements mm-iSAM [17], 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.

III. BACKGROUND AND PRELIMINARIES

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. In our setting, we are interested in determining the most probable assignment to a set of discrete variables D and continuous variables C given a set of *realized* measurements Z. Under the assumption that each measurement 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 $p(C, D \mid Z)$ into a product of measurement factors f_k, each of which depends only on a subset of variables $\mathcal{V}_k$:

$$
\begin{aligned}
p(C, D \mid Z) &\propto \prod_k f_k(\mathcal{V}_k), \\
\mathcal{V}_k &\triangleq \{v_i \in \mathcal{V} \mid (f_k, v_i) \in \mathcal{E}\},
\end{aligned}
\tag{1}
$$

where each factor f_k is in correspondence with either a measurement likelihood of the form $p(z_k \mid \mathcal{V}_k)$ or a prior $p(\mathcal{V}_k)$. From (1), the posterior $p(C, D \mid Z)$ 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 as follows:

$$
\begin{aligned}
C^*, D^* &= \underset{C,D}{\arg\max}\, p(C, D \mid Z) \\
&= \underset{C,D}{\arg\max} \prod_k f_k(\mathcal{V}_k) \\
&= \underset{C,D}{\arg\min} \sum_k -\log f_k(\mathcal{V}_k).
\end{aligned}
\tag{2}
$$

That is to say, we can maximize the posterior probability $p(C, D \mid Z)$ 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 non-linear least-squares problem, which permits the application of

incremental nonlinear least-squares solvers like iSAM2 [6].[2] In particular, we consider discrete-continuous factors $f_k(C_k, D_k)$ admitting a description as:

$$
\begin{aligned}
-\log f_k(C_k, D_k) &= \|r_k(C_k, D_k)\|_2^2, \\
C_k \subseteq C, \ &D_k \subseteq D,
\end{aligned}
\tag{3}
$$

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.

IV. OVERVIEW OF THE APPROACH

The following subsections describe our approach to solving optimization problems of the form in (2). In Section IV-A, we outline our core alternating minimization procedure and prove that our approach guarantees monotonic descent. In Section IV-B, we describe how our approach can easily benefit from existing incremental optimization techniques to efficiently solve large-scale estimation problems. Finally, in Section IV-C, we consider the issue of estimating *uncertainties* for the solutions provided by our method.

A. Alternating Minimization

In general, the MAP inference problem in (2) is computationally intractable [5, 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 [4]. 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 [6], [25]. 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

---

[2]This turns out not to be particularly restrictive; as demonstrated in Rosen et al. [21, Theorem 1], 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.

optimality using max-product variable elimination [5, 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:

$$p(D \mid C, Z) \propto \prod_j p(D_j \mid C, Z). \tag{4}$$

It is straightforward to verify from the mutual exclusivity of each set D_j that the problem of optimizing the conditional in (4) then breaks up into subproblems involving each D_j:

$$\max_D p(D \mid C, Z) \propto \prod_j \left[\max_{D_j} p(D_j \mid C, Z) \right]. \tag{5}$$

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 1b and 1c, depicting robust pose graph optimization and point-cloud registration, respectively, admit a decomposition of the form in equation (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 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 [26].

In turn, we will use these ideas to construct an algorithm for efficiently producing solutions to problems of the form in (2).[3] Consider the negative log posterior, defined as:

$$\mathcal{L}(C, D) \triangleq -\log p(C, D \mid Z). \tag{6}$$

From (2), the joint optimization problem of interest can be formulated as:

$$C^*, D^* = \underset{C,D}{\operatorname{argmin}} \mathcal{L}(C, D). \tag{7}$$

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

$$D^{(i+1)} = \underset{D}{\operatorname{argmin}} \mathcal{L}(C^{(i)}, D) \tag{8a}$$

$$C^{(i+1)} = \underset{C}{\operatorname{argmin}} \mathcal{L}(C, D^{(i+1)}). \tag{8b}$$

We may then repeat (8a) and (8b) until the relative decrease in $\mathcal{L}(C, D)$ is sufficiently small or we have reached

---

[3]The approach we present does not *require* that a model admit a conditional factorization like the one in equation (4), though it improves computational efficiency considerably (see Section VI-A for a discussion).

a maximum desired number of iterations. Finding minimizers for the subproblems (8a) and (8b) may still be challenging. Fortunately, one need not find a minimizer for the subproblems (8a) and (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:

$$\mathcal{L}(C^{(i)}, D^{(i+1)}) \leq \mathcal{L}(C^{(i)}, D^{(i)}) \tag{9a}$$

$$\mathcal{L}(C^{(i+1)}, D^{(i+1)}) \leq \mathcal{L}(C^{(i)}, D^{(i+1)}). \tag{9b}$$

There are many methods for updating the discrete and continuous states that satisfy (9a) and (9b), respectively. For the discrete states, the descent criterion in (9a) can be ensured by using the max-product algorithm to compute the *optimal solution* to the subproblem in (8a). For the continuous states, the descent criterion in (9b) can be guaranteed by, for instance, using a trust region method (e.g. [25]) to refine the continuous states with respect to the objective in (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* (9a) *and likewise for the continuous update in* (9b). *Then, the estimates* $C^{(i)}, D^{(i)}$ *obtained by alternating optimization satisfy:*

$$\mathcal{L}(C^{(0)}, D^{(0)}) \geq \mathcal{L}(C^{(1)}, D^{(1)}) \geq \ldots \geq \mathcal{L}(C^{(T)}, D^{(T)}), \tag{10}$$

i.e., this procedure monotonically improves the objective.

B. 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 [6] 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 (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 demonstrate in Section V-B, in which we produce solutions to online SLAM problems.

C. 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* [27, 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 clique tree propagation [5, Ch. 10]. The marginals we recover, then are:

$$p(D_j \mid \hat{C}, Z) = \sum_{D \setminus D_j} p(D \mid \hat{C}, Z), \quad D_j \subseteq D, \qquad (11a)$$

$$p(C_j \mid \hat{D}, Z) = \int_{C \setminus C_j} p(C \mid \hat{D}, Z), \quad C_j \subseteq C. \qquad (11b)$$

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 can easily become intractable. In contrast, by making use of the conditional factorization in (4), solving for the discrete marginals in (11a) is often tractable.[4] Notably, our approach to marginal recovery does not require that one use the alternating minimization strategy outlined in Section IV-A; any method of providing an estimate $(\hat{C}, \hat{D})$ will suffice.

The continuous marginals in (11b) are estimated using the Laplace approximation [7]. In our derivation, it will be convenient to consider the continuous states as a vector $C \in \mathbb{R}^d$. Assume the point $(\hat{C}, \hat{D})$ is a critical point of the continuous subproblem (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 $\hat{C}$:

$$\mathcal{L}(C, \hat{D}) \approx \mathcal{L}(\hat{C}, \hat{D}) - \frac{1}{2} A \left(C - \hat{C} \right), \qquad (12)$$

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

$$A \triangleq -\nabla^2 \mathcal{L}(C, \hat{D})|_{\hat{C}}. \qquad (13)$$

Exponentiating both sides of (12) and appropriately normalizing the result gives the linear Gaussian approximation:

$$p(C \mid \hat{D}, Z) \approx \frac{|A|^{1/2}}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} \|C - \hat{C}\|_{A^{-1}}^2 \right\}, \qquad (14)$$

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), 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 $A \succeq 0$ [29]. 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. [7]).

V. EXAMPLE APPLICATIONS

In the following sections we provide example applications motivated by typical robot perception problems. In Section V-A, we consider the problem of *robust pose graph optimization*, where we aim to estimate a set of poses given only

---

[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 [28].

noisy measurements between a subset of them, and some fraction of those measurements may be outliers. We implement a straightforward approach to solving this problem using DC-SAM and show that it produces competitive results. In Section V-B, we demonstrate the application of our solver to a tightly-coupled semantic SLAM problem, where the variables of interest are robot poses, landmark locations, and semantic classes of each landmark.

A. Robust Pose Graph Optimization

In this section we consider *robust* pose graph optimization. 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:

$$\min_{x_i \in \mathrm{SE}(3)} \sum_{\{i,j\} \in \vec{\mathcal{E}}} \big\| \underbrace{\log \left(\tilde{x}_{ij}^{-1} x_i^{-1} x_j \right)^\vee}_{r_{ij}(x_i, x_j)} \big\|_\Sigma^2, \qquad (15)$$

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. [30, 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. [31–34]), 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:

$$\min_{x_i \in \mathrm{SE}(3)} \sum_{\{i,j\} \in \vec{\mathcal{E}}_\mathcal{O}} \|r_{ij}(x_i, x_j)\|_\Sigma^2 + \sum_{\{i,j\} \in \vec{\mathcal{E}}_\mathcal{L}} e_{ij}(x_i, x_j, d_{ij}),$$
$$(16)$$

where

$$e_{ij}(x_i, x_j, d_{ij}) \triangleq \begin{cases} -\log \omega_0 + \|r_{ij}(x_i, x_j)\|_\Sigma^2, & d_{ij} = 0, \\ -\log \omega_1 + \|r_{ij}(x_i, x_j)\|_{\tilde{\Sigma}}^2, & d_{ij} = 1, \end{cases}$$
$$(17)$$

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 ω_0, ω_1 and discrete-continuous factors of the form in (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. 1b shows the corresponding graph).
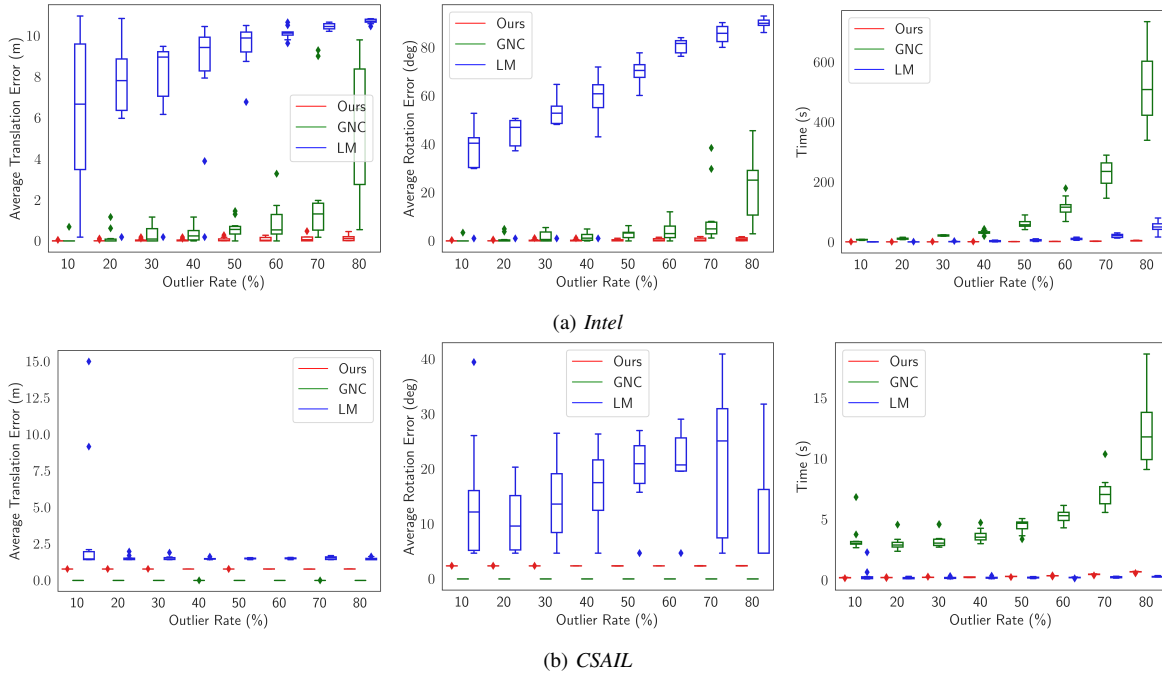
(a) *Intel*

(b) *CSAIL*

Fig. 2. **Robust pose graph optimization.** Average trajectory errors on (a) the *Intel* dataset, (b) the *CSAIL* 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.

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 [35, Section VI.C]). Based on the prior work of Olson and Agarwal [31], we made the outlier covariance model isotropic with variance 10^7 times larger than the inlier variance and set the weights ω_0, ω_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 [36].[5] Our results are summarized in Figure 2. 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 suboptimal performance on the CSAIL dataset). We will revisit this issue in Section VI-B.

B. Tightly-coupled Semantic SLAM

Several recent works have considered the problem of jointly inferring a robot's trajectory and a set of landmark po-

[5]We use the GNC approach implemented in GTSAM with the truncated least-squares cost. We use the default parameters.

[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.
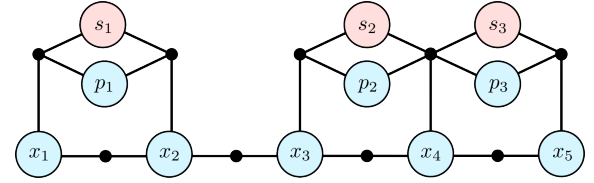

Fig. 3. **Factor graph representing semantic SLAM.** Here robot poses x_i are connected by odometry measurements, and joint geometric-semantic measurements are made between poses and landmarks (p_j, s_j). Measurements with multiple possible associations, represented as mixture factors in (18), are connected to multiple landmarks. To avoid clutter, the measurements depicted have only two hypotheses at most, but we allow for larger hypothesis sets.

Seq	Method	Translation Error (m)		Rotation Error (deg)	
		Mean	RMSE	Mean	RMSE
00	VISO2 [37]	11.457	13.136	2.410	2.562
	Ours	2.883	3.260	3.682	3.805
05	VISO2 [37]	6.227	7.2772	2.489	2.735
	Ours	2.175	2.442	1.486	1.665
08	VISO2 [37]	8.586	9.797	3.003	3.283
	Ours	8.468	9.429	6.264	6.732

TABLE I. **KITTI datasets.** Absolute translation and rotation errors (in meters and degrees, respectively) for our approach and VISO2.

sitions and classes with unknown measurement correspondences, i.e. semantic SLAM [38, 39]. In particular, here we apply our approach to optimize jointly for robot poses $x_i \in \mathrm{SE}(3), i = 1, \ldots, n$, and landmark locations and classes $\ell_j = (p_j, s_j),\ p_j \in \mathbb{R}^3,\ s_j \in \mathcal{C},\ j = 1, \ldots, m$, where $\mathcal{C}$ is an *a priori* known set of discrete semantic classes.

We adopt the general methodology of Doherty et al. [38] for data association: given a range-bearing measurement and associated semantic class obtained from an object detector,

we apply a threshold on the measurement likelihood to determine whether the measurement corresponds to a *new* or *old* landmark (for this, we employ the approximate marginal computations in (11a) and (14)). If it is a new landmark, we simply add the new measurement and landmark to our graph. If it is an old landmark, we add it to the graph as a mixture with a single component for each landmark that passes the likelihood threshold:

$$f_k(x_i, \mathcal{H}_k) \triangleq \max_{\ell_j \in \mathcal{H}_k} f_k(x_i, \ell_j), \tag{18}$$

where $\mathcal{H}_k \subseteq L$ is a subset of landmarks and the landmark measurement factor $f_k(x_i, \ell_j)$ decomposes as:

$$f_k(x_i, \ell_j) \triangleq \phi_k(s_j)\psi_k(x_i, p_j). \tag{19}$$

Here $\phi_k(s_j)$ is a categorical distribution and $\psi_k(x_i, p_j)$ is Gaussian with respect to the range and bearing between x_i and p_j. We also incorporate odometry factors of the form used in (15).[7] The overall graphical model specifying this problem is depicted in Figure 3. Since the measurement factors in equation (19) involve both continuous and discrete variables, it is nontrivial to implement this within any existing framework. In the hybrid factor graph representation, however, problems of this form admit a concise description and solution using discrete-continuous factors.

In this demonstration, we consider semantic SLAM using stereo camera data from the KITTI dataset [40]. We sample keyframes every two seconds, using VISO2 [37] to obtain stereo odometry measurements and YOLO [41] for noisy detections of two types of objects: *cars* and *trucks*. 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.[8] Table I 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.[9]

VI. DISCUSSION

A. When is alternating minimization efficient?

The conditional factorization in equation (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 (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 [5]. 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.

---

[7]For a more detailed exposition of this approach, see [38].

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

[9]Qualitative results visualizing the semantic map output from our method are available in the supplement [1] .

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 [42].

B. 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 [4].[10] 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) [36] as discussed in Section V-A, 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 [43, 44]. Similarly, stochastic gradient descent is a classical approach for nonconvex optimization (and has appeared in the setting of robust pose-graph SLAM [45]), which could reasonably be adapted to our approach. Finally, heuristics have been considered which use consistency of measurements to filter out unlikely hypotheses [46] or to *re-initialize* estimates for factor graphs [47].

VII. CONCLUSION

In this work 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

---

[10]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 [4, Sec 2] for a review) and is otherwise itself an open problem.

for addressing the issue of recovering uncertainties associated with estimates in the 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, and semantic SLAM.

REFERENCES

[1] K. J. Doherty, Z. Lu, K. Singh, and J. J. Leonard, "Discrete-Continuous Smoothing and Mapping," *arXiv preprint arXiv:2204.11936*, 2022.

[2] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Institute of Technology, Tech. Rep., 2012.

[3] C. R. Garrett, R. Chitnis, R. Holladay, B. Kim, T. Silver, L. P. Kaelbling, and T. Lozano-Pérez, "Integrated Task and Motion Planning," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 265–293, 2021. [Online]. Available: https://doi.org/10.1146/annurev-control-091420-084139

[4] D. M. Rosen, K. J. Doherty, A. T. Espinoza, and J. J. Leonard, "Advances in Inference and Representation for Simultaneous Localization and Mapping," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 215–242, 2021. [Online]. Available: https://doi.org/10.1146/annurev-control-072720-082553

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

[6] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Intl. J. of Robotics Research*, vol. 31, pp. 217–236.

[7] M. Kaess and F. Dellaert, "Covariance recovery from a square root information matrix for data association," *J. of Robotics and Autonomous Systems*, vol. 57, no. 12.

[8] F. Dellaert, "Factor graphs: Exploiting structure in robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 141–166, 2021. [Online]. Available: https://doi.org/10.1146/annurev-control-061520-010504

[9] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.

[10] I. J. Cox and J. J. Leonard, "Modeling a dynamic environment using a Bayesian multiple hypothesis approach," *Artificial Intelligence*, vol. 66, no. 2, pp. 311–344, 1994.

[11] M. Hsiao and M. Kaess, "MH-iSAM2: Multi-hypothesis iSAM using Bayes tree and hypo-tree," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2019, pp. 1274–1280.

[12] F. Jiang, V. Agrawal, R. Buchanan, M. Fallon, and F. Dellaert, "iMHS: An Incremental Multi-Hypothesis Smoother," *arXiv preprint arXiv:2103.13178*, 2021.

[13] P.-Y. Lajoie, S. Hu, G. Beltrame, and L. Carlone, "Modeling perceptual aliasing in SLAM via discrete–continuous graphical models," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1232–1239, 2019.

[14] A. Salmerón, R. Rumí, H. Langseth, T. D. Nielsen, and A. L. Madsen, "A review of inference algorithms for hybrid Bayesian networks," *Journal of Artificial Intelligence Research*, vol. 62, pp. 799–828, 2018.

[15] D. Ramos-López, A. R. Masegosa, A. M. Martínez, A. Salmerón, T. D. Nielsen, H. Langseth, and A. L. Madsen, "MAP inference in dynamic hybrid Bayesian networks," *Progress in Artificial Intelligence*, vol. 6, no. 2, pp. 133–144, 2017.

[16] M. Montemerlo and S. Thrun, "Simultaneous localization and mapping with unknown data association using fastSLAM," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, vol. 2, 2003, pp. 1985–1991.

[17] D. Fourie, J. Leonard, and M. Kaess, "A nonparametric belief solution to the Bayes tree," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2016, pp. 2189–2196.

[18] Q. Huang, C. Pu, D. Fourie, K. Khosoussi, J. P. How, and J. J. Leonard, "NF-iSAM: Incremental Smoothing and Mapping via Normalizing Flows," *arXiv preprint arXiv:2105.05045*, 2021.

[19] E. B. Sudderth, A. T. Ihler, M. Isard, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," *Communications of the ACM*, vol. 53, no. 10, pp. 95–103, 2010.

[20] T. Pfeifer, S. Lange, and P. Protzel, "Advancing Mixture Models for Least Squares Optimization," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3941–3948, 2021.

[21] 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)*, 2013, pp. 1025–1032.

[22] S. Agarwal and K. Mierle, *Ceres Solver: Tutorial & Reference*, Google Inc.

[23] G. Grisetti, R. Kümmerle, H. Strasdat, and K. Konolige, "g2o: A general framework for (hyper) graph optimization," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2011, pp. 9–13.

[24] Contributors, Ecosystem, and NavAbility. (2021) Caesar.jl, v0.11.1. https://github.com/JuliaRobotics/Caesar.jl.

[25] D. M. Rosen, M. Kaess, and J. J. Leonard, "An incremental trust-region method for robust online sparse least-squares estimation," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2012, pp. 1262–1269.

[26] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[27] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

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

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

[31] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," *Intl. J. of Robotics Research*, vol. 32, no. 7, pp. 826–840, 2013.

[32] N. Sünderhauf and P. Protzel, "Switchable constraints for robust pose graph SLAM," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2012, pp. 1879–1884.

[33] P. Agarwal, G. D. Tipaldi, L. Spinello, C. Stachniss, and W. Burgard, "Robust map optimization using dynamic covariance scaling," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 62–69.

[34] A. V. Segal and I. D. Reid, "Hybrid inference optimization for robust pose graph estimation," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2014, pp. 2675–2682.

[35] V. Tzoumas, P. Antonante, and L. Carlone, "Outlier-robust spatial perception: Hardness, general-purpose algorithms, and guarantees," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019, pp. 5383–5390.

[36] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, "Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection," *arXiv preprint arXiv:1909.08605*, 2019.

[37] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: Dense 3d Reconstruction in Real-time," in *IEEE Intelligent Vehicles Symposium*, 2011.

[38] K. Doherty, D. Baxter, E. Schneeweiss, and J. J. Leonard, "Probabilistic data association via mixture models for robust semantic SLAM," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2020.

[39] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas, "Probabilistic data association for semantic SLAM," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2017, pp. 1722–1729.

[40] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset," *Intl. J. of Robotics Research*, 2013.

[41] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[42] B. Savchynskyy *et al.*, "Discrete graphical models—an optimization perspective," *Foundations and Trends® in Computer Graphics and Vision*, vol. 11, no. 3-4, pp. 160–429, 2019.

[43] D. J. C. Mackay, "Introduction to Monte Carlo Methods," in *Learning in graphical models*. Springer, 1998, pp. 175–204.

[44] H. Szu and R. Hartley, "Fast simulated annealing," *Physics letters A*, vol. 122, no. 3-4, pp. 157–162, 1987.

[45] 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)*, pp. 2262–2269.

[46] J. G. Mangelson, D. Dominic, R. M. Eustice, and R. Vasudevan, "Pairwise consistent measurement set maximization for robust multi-robot map merging," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2018, pp. 2916–2923.

[47] Z. Lu, Q. Huang, K. Doherty, and J. J. Leonard, "Consensus-Informed Optimization Over Mixtures for Ambiguity-Aware Object SLAM," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2021.