SLAM as a Stochastic Control with Partial Information Problem: Optimal Solutions and Rigorous Approximations

Ilir Gusija

December 4, 2025

Supervised by:
Dr. Serdar Yuksel
Dr. Fady Alajaji

Department of Mathematics and Statistics
Queen's University

1 Introduction

1.1 Literature Review

The active simultaneous localization and mapping (SLAM) problem involves a robot equipped with sensors, autonomously estimating its pose while constructing a map of its surrounding environment [1]. The necessity for simultaneous operation is driven by applications where pre-mapping is infeasible due to dynamic environments, resource constraints, or inaccessible locations such as underwater, underground, or extraterrestrial exploration [2]. Although classical SLAM approaches focus primarily on estimation [3, 4, 5, 6], it has been shown that active SLAM techniques (which comparatively have received limited attention [7]) significantly improve exploration efficiency and can reduce mapping uncertainty [8]. Deep learning methods for SLAM were also proposed in e.g., [9], although these lack formal convergence and performance guarantees.

Early SLAM approaches relied on extended Kalman filters (EKF), which maintain a Gaussian belief over the current robot pose and landmark positions, updating the belief recursively as new measurements arrive [10]. EKF-based SLAM suffers from inconsistency over long trajectories: linearization errors accumulate, producing overly optimistic covariance estimates that fail to reflect true uncertainty [11, 12]. Additionally, the requirement to maintain a full covariance matrix over all landmarks scales quadratically with map size, limiting applicability to environments with few features.

Rao-Blackwellized particle filters (RBPF) [13], exemplified by FastSLAM [6], addressed computational scaling by representing the robot trajectory with particles while marginalizing landmarks analytically via per-particle Kalman filters. This factorization reduces the effective dimensionality from the full state (trajectory plus all landmarks) to just the trajectory. However, particle filters face consistency issues of their own [14] as well as the issue of particle degeneracy—after repeated resampling, all particles collapse to a few trajectories—and require exponentially many samples to accurately represent high-dimensional distributions [15], limiting their reliability for extended operation.

Graph-based SLAM methods reformulate the problem as maximum *a posteriori* (MAP) estimation over the full robot trajectory and map, representing the posterior as a factor graph where nodes correspond to variables (poses, landmarks) and edges encode probabilistic constraints from motion and observation models [16]. Unlike filtering approaches that maintain only the current belief state, these batch methods solve for the entire history jointly by minimizing a nonlinear least squares objective via iterative optimization [15]. This naturally handles loop closures and avoids filter inconsistency by relinearizing around updated estimates after incorporating all available data, rather than suffering from accumulated linearization errors. However, early batch implementations solved the full optimization problem at each timestep, recomputing matrix factorizations from scratch [17]—while exploiting sparsity via variable ordering reduced complexity substantially, repeatedly computing the full solution remains computationally expensive for real-time operation.

Incremental methods enabled real-time batch optimization by reusing prior computation. iSAM [18] stored the square root information matrix and incrementally updated only affected entries when new measurements arrived, though loop closures necessitated expensive periodic reordering. iSAM2 [19] eliminated batch reordering via the Bayes tree—a data structure encoding conditional independence—and introduced fluid relinearization, updating only variables whose estimates changed significantly. These advances form the foundation of modern optimization libraries like GTSAM [20]. For long-term autonomy, fixed-lag smoothing [21] bounds computational cost by marginalizing old states outside a sliding window, though marginalization destroys sparsity [22]. Hierarchical systems like ORB-SLAM [23] instead schedule computation at multiple granularities: fast motion-only tracking, bounded local bundle adjustment, and infrequent global pose graph optimization. While sacrificing the global optimality of full batch methods, these approaches achieve real-time performance in large-scale environments.

Concurrently with batch method developments, invariant extended Kalman filters (IEKFs) emerged to address the consistency issues plaguing classical EKF-SLAM while preserving the computational efficiency of filtering [24]. Julier and Uhlmann [25] first demonstrated through a counterexample that EKF-SLAM produces inconsistent estimates when the robot is stationary, with the robot orientation uncertainty being identified as the key culprit. Huang and Dissanayake [12] proved that EKF-SLAM incorrectly reduces robot orientation uncertainty when observing new landmarks (i.e., with unknown absolute locations), violating results [12, Corollary 3.3,3.6] showing that such observations cannot reduce robot pose uncertainty. The root cause is that linearization errors accumulate in a way that causes the filter to believe it has more information about the robot's orientation than is actually available from the measurements.

By exploiting the Lie group structure of the state space (e.g., $SE(3)$ for rigid body pose), IEKFs perform linearization in coordinates where unobservable directions (such as global rotation) become independent of the linearization point. This ensures the filter's uncertainty representation remains consistent with the true unobservabilities of the SLAM problem. The resulting filters maintain quadratic computational scaling with map size—feasible for moderate-sized maps on embedded platforms—achieving markedly improved consistency over long trajectories. Barrau and Bonnabel [26, 27] proved that IEKF-SLAM maintains correct observability properties and demonstrated this experimentally on the Victoria Park dataset. Subsequent work extended this framework to visual-inertial SLAM, with [28] showing that unscented Kalman filters on Lie groups (UKF-LG) [29] offer similar performance to IEKFs with simpler implementation and reduced computational cost. The Gaussian assumption remains pervasive across both filtering and batch approaches (save for particle filtering), limiting applicability when measurement distributions are heavy-tailed or when data association is ambiguous—that is, when multiple landmark hypotheses are plausible for a given observation, yielding multimodal posteriors that Gaussian representations cannot capture.

The geometric foundations underlying these invariant filtering approaches were further developed by Mahony and Hamel [30], who formulated the SLAM state space as a quotient manifold $\mathcal{M}_n(3)$ under reference frame transformations and introduced the $\mathrm{SLAM}_n(3)$ symmetry group whose transitive action yields equivariant kinematics. Working in the deterministic observer framework rather than stochastic filtering, they designed nonlinear observers with provable exponential convergence via Lyapunov analysis, demonstrating that the geometric structure enables robust estimation guarantees beyond the Kalman filtering paradigm.

By formulating active SLAM as a partially observable Markov decision process (POMDP), stochastic control techniques have been applied with some success. Much work has focused on the related problem of belief-space planning (BSP)—the general task of computing control policies that optimize over posterior distributions—which encompasses path planning under uncertainty, active localization, and exploration as special cases [31]. However, most BSP approaches to active SLAM rely on parametric belief representations (typically Gaussian) and employ local optimization methods without global convergence guarantees. For instance, [32] applied linear quadratic Gaussian (LQG) control under *maximum likelihood* observation assumptions to compute locally optimal policies. A subsequent approach relaxed the maximum likelihood assumption and instead evaluated the probability of success of a given trajectory, rather than constructing an optimal one [33]. Finally, [31] extended this framework to active SLAM via a dual-layer architecture: an inner estimation layer predicts belief evolution under candidate controls using Gaussian approximations and MAP estimates, while an outer model predictive control (MPC) layer computes locally optimal control sequences. Further work on the approach introduced by [33] introduced constraints on the uncertainty of the belief and relaxed assumptions on the smoothness of the measurement model [34]; this work was then adapted to the active SLAM problem in [35]. A value iteration algorithm for learning to act in POMDPs with continuous state spaces was proposed, for which robotic planning is mentioned as a possible application [36]. Information-theoretic metrics have proven effective as utility functions for active SLAM. Approaches based on information gain [2, 37], Kullback-Leibler divergence [38], and Rényi entropy [39] have been shown to guide robots toward unexplored regions while reducing uncertainty. Conditional mutual information metrics have been proved to provably steer robots toward unexplored areas [40].

Recently, much work was done on learning-based approaches to stochastic control in general state and action space. One approach looks at quantization of the state space, action space, and belief space, showing that under certain conditions the discounted cost of the approximated belief approaches the optimal cost for increasingly finer quantization schemes [41]. An alternate approach is to look at policies of finite memory which incur less computational overhead than the former method [42]. Subsequent work was conducted within both approximation schemes to apply Q-learning, which, under certain conditions, are both shown to converge to near-optimal solutions [43, 44]. These methods were successfully applied to the zero-delay coding problem [45].

In order to leverage these approximation results, we build on recent advances in finite model approximations for POMDPs from [41], [46, Section 4] which provides convergence guarantees for discretized belief spaces. Under appropriate continuity conditions, these finite approximations can achieve provably near-optimal performance — something we show to be the case for the active SLAM problem.

1.2 Contributions

The technical contributions of this papers are as follows:

- In Section 2 we provide the model and cost criteria under a broad formulation and we establish a very general POMDP formulation for the SLAM problem and study several of its properties. We establish, in particular, weak continuity properties, and existence results for the SLAM problem. Theorems 1 and 2 state these properties.
- Building on these properties we provide principled methods to arrive at a rigorously justified approximation algorithm which is near optimal under explicit conditions which we present and validate in the paper. This is given in Algorithm 1 and Theorems 3 and 4.
- Finally, we build a toy example simulating a robot in a 2D plane to illustrate our theoretical findings via simulation results [in progress].

2 Problem Definition

We formulate the active SLAM problem within the framework of POMDPs. Active SLAM treats path planning as an optimal control problem: the robot must select actions that simultaneously reduce uncertainty in its own state estimate and improve map quality. This leads to a coupled estimation-control problem where information gathering and localization are mutually dependent.

The remainder of this section proceeds as follows. We first introduce the state space, dynamics, and observation models in their natural state-space form (Section 2.1). We then reformulate these as stochastic kernels suitable for POMDP analysis (Section 2.2). Finally, we reformulate the POMDP as a fully observable MDP over the space of posterior distributions (Section 2.3).

2.1 Preliminaries

We denote the joint state as $s_t = (x_t, m_t) \in \mathbb{S} = \mathbb{X} \times \mathbb{M}$.

2.1.1 Motion model

Let $\mathbb{X}$ be a Polish space (i.e. a complete and separable metric space) that denotes the set of all possible poses of the robot (which encompasses a wide range of potential state space formulations such as $SE(3)$). Let $x_t \in \mathbb{X}$ denote the pose of the robot at $t \geq 0$. The robot pose evolves in time according to a controlled stochastic process

$$x_{t+1} = f(x_t, u_t, w_t), \quad w_t \overset{\text{iid}}{\sim} \mu_w \tag{1}$$

where $u_t \in \mathbb{U}$ is the control input and $w_t \in \mathbb{W}$ represents an independent and identically distributed (iid) process noise with distribution μ_w (written as $w_t \overset{\text{iid}}{\sim} \mu_w$) on some measurable space $(\mathbb{W}, \mathcal{W})$. Examples for f include unicycle kinematics, differential drive models, or linear dynamics $f(x_t, u_t, w_t) = Ax_t + Bu_t + w_t$ (see Section 5.1.2).

2.1.2 Modelling the map

The representation of the map varies across different approaches in SLAM. This formulation subsumes these various representations including occupancy grids, and landmark-based maps with the key assumption that these maps remain static. Let $\mathbb{M}$ be a Polish space representing the space of maps, possible representations of this map include:

- Occupancy grid maps where $\mathbb{M} \coloneqq \{0,1\}^{H \cdot W}$ where $H, W \geq 1$ denotes the height and width of the map respectively.
- Landmark-based maps where $\mathbb{M} \coloneqq \mathbb{X}^{\bar{\ell}}$ where $\bar{\ell}$ represent the number of landmarks.

- Multi-class maps (i.e. [47]) where $\mathbb{M} = \{0,1\}^{H\cdot W} \times \mathcal{K}$ where $\mathcal{K}$ denotes the set of semantic categories

Let m_t denote the state of the map at time $t \geq 0$. The map m_t is assumed to be static, so its dynamics are simply described as

$$m_{t+1} = m_t, \quad t \geq 0. \tag{2}$$

2.1.3 Sensor model

At each time step $t \geq 0$, the robot receives noisy observations $y_t \in \mathbb{Y}$ of the environment according to a measurement model

$$y_t = g(x_t, m_t, v_t), \quad v_t \overset{\text{iid}}{\sim} \mu_v, \tag{3}$$

where $v_t \in \mathbb{V}$ represents iid measurement noise with distribution μ_v on some measurable space $(\mathbb{V}, \mathcal{V})$.

2.1.4 SLAM as a control problem

The information the decision maker has access to are the past measurements and control actions; we denote the collection of information available at time t as

$$I_t = \{y_{[0,t]}, u_{[0,t-1]}\}, \quad t \geq 0, \quad I_0 = \{y_0\} \tag{4}$$

where we use the notation

$$y_{[0,t]} = \{y_k, 0 \leq k \leq t\}, \quad u_{[0,t-1]} = \{u_k, 0 \leq k \leq t-1\}.$$

We denote the control policy as $\gamma = \{\gamma_t\}_{t\geq 0}$. At time t, we have that $\gamma_t : \mathbb{Y}^t \times \mathbb{U}^{t-1} \to \mathbb{U}$ or

$$u_t = \gamma_t(I_t). \tag{5}$$

We let Γ denote the set of admissible control policies.

In general, for the active SLAM problem, the goal is to minimize the infinite horizon discounted cost given by

$$J(\pi_0, \gamma) \coloneqq \lim_{T\to\infty} \mathbb{E}^{\gamma}_{\pi_0}\left[\sum_{t=0}^{T-1} \beta^t c(s_t, u_t)\right]. \tag{6}$$

where $\beta \in (0,1)$ is the discount factor, $\mathbb{E}^{\gamma}_{\pi_0}$ is the expectation with initial distribution $s_0 \sim \pi_0$ under policy γ, and $c : \mathbb{S} \times \mathbb{U} \to [0, \infty)$ is a one stage cost function which typically encodes information gain (reducing belief uncertainty), map coverage (exploring unmapped regions), collision avoidance, and control effort minimization. Specific instantiations of c will be discussed in Section 5.1.

2.2 SLAM as a POMDP

For the purpose of formulating active SLAM as a POMDP over the joint state space $\mathbb{S} = \mathbb{X} \times \mathbb{M}$, where $s_t = (x_t, m_t) \in \mathbb{S}$, we equivalently describe the dynamics and observations through stochastic kernels. This measure-theoretic formulation, while less common in the SLAM literature, provides the rigor necessary for analyzing continuity properties of the belief dynamics.

We define $\mathcal{P}(\mathbb{X})$ as the set of probability measures on $\mathbb{X}$ endowed with the weak convergence topology. The pose dynamics (1) induce a transition kernel $\mathcal{T} : \mathbb{X} \times \mathbb{U} \to \mathcal{P}(\mathbb{X})$ defined by

$$\mathcal{T}(B \mid x_t, u_t) = \mu_w(\{w \in \mathbb{W} : f(x_t, u_t, w) \in B\}) \tag{7}$$

for any $B \in \mathcal{B}(\mathbb{X})$, the Borel σ-algebra of $\mathbb{X}$.

The map dynamics are trivially given by the Dirac measure

$$\mathbb{P}(m_{t+1} \in \cdot \mid m_t) \coloneqq \delta_{m_t}(\cdot). \tag{8}$$

The explicit specification of the map transition kernel via the Dirac measure, while seemingly trivial, distinguishes our formulation from prior work that typically states informally that "the map is static" without defining the associated probability measure. This precision allows rigorous derivation of the factorized joint kernel (10) rather than assuming it from the outset. Moreover, our formulation naturally extends to time-varying maps by replacing δ_{m_t} with an appropriate stochastic kernel and (2) with an appropriate dynamical model, though we restrict attention to static maps for analytical tractability.

Similarly, the observation model (3) induces a measurement kernel $Q : \mathbb{S} \to \mathcal{P}(\mathbb{Y})$ given by

$$Q(C \mid x_t, m_t) = \mu_v(\{v \in \mathbb{V} : g(x_t, m_t, v) \in C\}) \tag{9}$$

for any $C \in \mathcal{B}(\mathbb{Y})$.

We define active SLAM as a POMDP, fully described by the 7-tuple $(\mathbb{S}, \mathbb{U}, \mathbb{Y}, \mathcal{S}, Q, c, \beta)$ where $\mathbb{S} = \mathbb{X} \times \mathbb{M}$, $\mathbb{U}$, and $\mathbb{Y}$ are Borel state, action, and observation spaces respectively, $\mathcal{S} : \mathbb{S} \times \mathbb{U} \to \mathcal{P}(\mathbb{S})$ is the joint state transition kernel, $Q : \mathbb{S} \to \mathcal{P}(\mathbb{Y})$ is the observation channel from (9), $c : \mathbb{S} \times \mathbb{U} \to \mathbb{R}$ is the stage cost; and $\beta \in (0, 1)$ is the discount factor.

Since the map is static and the pose evolves independently, the joint transition kernel factorizes as

$$\mathcal{S}(B_x \times B_m \mid (x_t, m_t), u_t) = \mathcal{T}(B_x \mid x_t, u_t) \cdot \delta_{m_t}(B_m) \tag{10}$$

for sets $B_x \in \mathcal{B}(\mathbb{X})$ and $B_m \in \mathcal{B}(\mathbb{M})$.

2.3 Belief MDP Reformulation

Since the state s_t is not directly observable, optimal decision-making must condition on the *belief* (information state)

$$\pi_t(\,\cdot\,) := \mathbb{P}(s_t \in \cdot \mid y_{[0,t]}, u_{[0,t-1]}) \tag{11}$$

representing the posterior distribution over states given the observation and action history.

The belief evolves via Bayesian filtering: given the prior belief π_t, action u_t, and new observation y_{t+1}, the posterior belief is

$$\pi_{t+1}(x, m) = \frac{Q(y_{t+1} \mid x, m) \int_{\mathbb{X}} \mathcal{T}(dx \mid x_t, u_t) \pi_t(x_t, m)}{\int_{\mathbb{X}\times\mathbb{M}} Q(y_{t+1} \mid x', m') \int_{\mathbb{X}} \mathcal{T}(dx' \mid x_t, u_t) \pi_t(dx_t, dm')} =: \mathbf{F}(\pi_t, u_t, y_{t+1})(x, m) \tag{12}$$

where we have used the factorization in (10) to integrate out the map component. The belief update (12) computes the posterior $\pi_t(x, m)$ over the joint state space without assuming any factorization structure. By contrast, most active SLAM approaches employ Rao-Blackwellized representations [37, 2] where the robot trajectory posterior $\mathbb{P}(x_{[0,t]} \mid y_{[0,t]}, u_{[0,t-1]})$ is maintained separately from the map posterior $\mathbb{P}(\,m \mid x_{[0,t]}, y_{[0,t]})$. While Rao-Blackwellization offers substantial computational advantages—particularly for particle filter implementations—our joint formulation admits analysis of the belief dynamics on $\mathcal{P}(\mathbb{S})$ without presupposing conditional independence between pose and map estimation errors. This generality is essential for the weak continuity analysis of the filter kernel η developed in our main results.

The distribution over future observations conditioned on the current belief and action, is given by

$$\begin{aligned} \mathbf{H}(C \mid \pi, u) &= \mathbb{P}(y_{t+1} \in C \mid \pi_t = \pi, u_t = u) \\ &= \int\limits_{\mathbb{X}} \int\limits_{\mathbb{X}\times\mathbb{M}} Q(C \mid x_{t+1}, m_t) \mathcal{T}(dx_{t+1} \mid x_t, u_t) \pi_t(dx_t, dm_t) \end{aligned} \tag{13}$$

for any $C \in \mathcal{B}(\mathbb{Y})$.

Crucially, the belief π_t is a sufficient statistic for optimal control [48, 49, 50], enabling reformulation of the POMDP as a fully observable MDP over the belief space $\Pi := \mathcal{P}(\mathbb{S})$. This belief MDP is described by the 4-tuple $(\Pi, \mathbb{U}, \eta, \tilde{c})$ where Π is the space of probability measures on $\mathbb{S}$, $\mathbb{U}$ is the action space (unchanged), $\eta : \Pi \times \mathbb{U} \to \mathcal{P}(\Pi)$ is the belief transition kernel induced by the filter update (12) and predictive observation distribution (13) given by

$$\eta(\,\cdot \mid \pi, u) = \int_{\mathbb{Y}} \delta_{\mathbf{F}(\pi,u,y)}(\,\cdot\,)\, \mathbf{H}(dy \mid \pi, u); \tag{14}$$

and $\tilde{c} : \Pi \times \mathbb{U} \to \mathbb{R}$ is the expected stage cost given by

$$\tilde{c}(\pi_t, u_t) = \int_{\mathbb{S}} c(s_t, u_t)\, \pi_t(ds_t). \tag{15}$$

The optimal policy $\gamma^* : \Pi \to \mathbb{U}$ minimizes the expected discounted cost

$$J^*(\pi_0) = \inf_{\gamma \in \Gamma} \mathbb{E}^{\gamma}_{\pi_0} \left[\sum_{t=0}^{\infty} \beta^t \tilde{c}(\pi_t, u_t) \right], \tag{16}$$

where the expectation is over observation sequences generated under policy γ.

3 Properties of Active SLAM

The belief MDP formulation established in Section 2.3 provides a rigorous mathematical framework for active SLAM, but its practical utility depends on verifying key structural properties of the belief dynamics. This section establishes two fundamental results: first, that the belief process $\{\pi_t, u_t\}_{t\geq 0}$ forms a controlled Markov chain (Theorem 1), confirming that the belief is indeed a sufficient statistic for optimal control; and second, that the belief transition kernel η satisfies a weak continuity property called the weak Feller condition (Theorem 2), which ensures small changes in beliefs and actions produce small changes in posterior distributions.

The weak Feller property is the technical foundation for the finite approximation schemes developed in Section 4. Intuitively, the weak Feller property is a continuity condition on the belief dynamics: if the current belief π and control u change by small amounts, the resulting posterior distribution $\eta(\,\cdot \mid \pi, u)$ after incorporating the next observation also changes by a small amount. This ensures the belief-MDP has smooth dynamics despite operating over an infinite-dimensional state space. Without such regularity, finite approximations of the belief space could fail catastrophically, as arbitrarily close beliefs might evolve into drastically different posteriors. This regularity allows us to approximate the infinite-dimensional belief space with finite grids while controlling the approximation error.

However, verifying the weak Feller property directly from the definition requires establishing probabilistic continuity properties of the transition kernel $\mathcal{T}$ and observation channel Q, which may be unfamiliar to practitioners coming from robotics backgrounds. We therefore provide sufficient conditions (Assumption 2) expressed in terms of standard continuity of the motion and measurement models f and g. These conditions are straightforward to verify for common robotic systems—differential drive robots, Dubins cars, range sensors, etc.—and imply the requisite probabilistic continuity properties.

We begin by formalizing the Markov structure of the belief process.

Theorem 1 (Belief is a controlled Markov chain). *$\{\pi_t, u_t\}_{t\geq 0}$ is a controlled Markov chain.*

To state the weak Feller property precisely, we first recall the notion of weak convergence of probability measures. A sequence of probability measures $\{\mu_n\}_{n\geq 1}$ on a metric space $\mathbb{X}$ converges weakly to μ if

$$\int_{\mathbb{X}} h(x)\, \mu_n(dx) \to \int_{\mathbb{X}} h(x)\, \mu(dx)$$

for every bounded continuous function $h : \mathbb{X} \to \mathbb{R}$. Weak convergence is weaker than convergence in total variation but sufficient for establishing continuity of value functions in dynamic programming.

Definition 1 (Weak Feller Property). *We say that a Markov decision process with transition kernel $\eta(\,\cdot \mid \pi, u)$ has the weak Feller property if η is weakly continuous in (π, u); that is, if $(\pi_n, u_n) \to (\pi, u)$, then $\eta(\,\cdot \mid \pi_n, u_n) \to \eta(\,\cdot \mid \pi, u)$ weakly.*

Establishing the weak Feller property for the belief transition kernel η defined in (14) requires assumptions on the transition kernel $\mathcal{T}$ and measurement channel Q. We present two alternative sets of assumptions that suffice.

We impose the following assumption onto the model

Assumption 1.

1. *The transition probability $\mathcal{T}(\,\cdot \mid x, u)$ is continuous in total variation.*

2. *The observation channel $Q(\,\cdot \mid x, m)$ is independent of the control variable.*

While Assumption 1 is stated in terms of probabilistic continuity of stochastic kernels, it may be difficult to verify directly from system specifications. For practitioners, it is more natural to impose regularity on the motion and measurement functions f and g. The following assumption provides a sufficient condition that is straightforward to check for typical robotic systems.

Assumption 2 (Sufficient conditions for Assumption 1)**.** *A robot is equipped with a motion model* $x_{t+1} = f(x_t, u_t) + w_t$ *that is continuous in* (x_t, u_t) *and a measurement model* $y_t = g(s_t, v_t)$ *that is independent of the control variable.*

We then have the following result, extending [51, Theorem 3.1] to the augmented state space $\mathbb{S} = \mathbb{X} \times \mathbb{M}$:

Theorem 2 (Weak Feller property of the belief transition kernel)**.** *Provided that any one of Assumptions 1–2 holds, the belief transition kernel,* η*, of* (14) *is weak Feller.*

This result establishes that under standard regularity conditions on the system dynamics and observations, the belief dynamics inherit the necessary continuity for approximation methods. Having verified the weak Feller property, we now turn to discretization schemes that reduce the continuous belief MDP to finite models amenable to computational solution.

4 Finite Approximations

The belief MDP formulation developed in Section 2.3 operates over continuous spaces: the belief space $\Pi = \mathcal{P}(\mathbb{S})$ is infinite-dimensional, and the action space $\mathbb{U}$ is typically uncountable. While this generality is necessary for rigorous analysis, it renders the problem of minimizing (16) computationally intractable. Standard dynamic programming methods—such as value iteration or policy iteration—require discrete state and action spaces with finite cardinality.

This section develops a systematic approximation framework that reduces the continuous belief MDP to a finite-state, finite-action MDP while providing theoretical guarantees on the approximation error. Our approach follows the methodology of Saldi et al. [41], which establishes that under the weak Feller property verified in Section 3, optimal policies for finite approximations converge to the optimal policy of the original POMDP as the discretization is refined.

We proceed in three stages. Section 4.1 discretizes the action space $\mathbb{U}$ using $1/n$-nets, establishing that restricting to finitely many actions incurs vanishing suboptimality as the net is refined (Theorem 3). Section 4.2 discretizes the belief space Π via quantization, showing that optimal policies on the quantized belief MDP converge to the optimal policy of the original problem (Theorem 4). Finally, Section 4.3 specializes these results to occupancy grid maps, exploiting the finite structure of $\mathbb{M} = \{0, 1\}^{H \times W}$ to derive explicit quantization schemes using the Wasserstein metric and type lattices from information theory.

Throughout this section, we assume the following.

Assumption 3. *The POMDP has the following properties:*

1. *The one-stage cost function c is bounded and continuous.*
2. *One of Assumptions 1–2 hold.*
3. *$\mathbb{X}$ and $\mathbb{U}$ are compact.*

The above set of assumptions implies the following for the equivalent belief-MDP.

Assumption 4. *The belief-MDP has the following properties:*

1. *The stage cost function $\tilde{c}$ is bounded and continuous.*
2. *The belief transition η is weak Feller.*
3. *$\mathcal{P}(\mathbb{X})$ and $\mathbb{U}$ are compact.*

Note that $\mathbb{X}$ being compact gives us that $\mathcal{P}(\mathbb{X})$ is compact with respect to the weak topology.

4.1 Finite-Action Approximation

Let $d_{\mathbb{U}}$ denote the metric on $\mathbb{U}$. Since we assume $\mathbb{U}$ to be compact then it is necessarily totally bounded. Hence by the definition we get that $\exists\{\Lambda^{(n)}\}_{n\in\mathbb{N}} \subseteq \mathbb{U}$ such that $\Lambda^{(n)} := \{u_1^{(n)}, \ldots, u_{k_n}^{(n)}\}$ such that for each n,

$$\min_{i\in\{1,\ldots,k_n\}} d_{\mathbb{U}}(u, u_i^{(n)}) < \frac{1}{n}, \quad \forall u \in \mathbb{U} \tag{17}$$

Hence, $\Lambda^{(n)}$ is a $1/n$-net in $\mathbb{U}$. The sequence $\{\Lambda^{(n)}\}_{n\geq 1}$ is used by the finite-action model to approximate the belief-MDP and the POMDP.

Theorem 3 (Saldi 2019 [41])**.** *Suppose Assumption 4 holds for the belief MDP. Then we have*

$$\lim_{n\to\infty} \left|\tilde{J}_n^*(\pi) - \tilde{J}^*(\pi)\right| = 0 \quad \forall \pi \in \Pi,$$

where $\tilde{J}_n^*$ is the discounted cost value function of the belief-MDP$_n$ with components $(\Pi, \Lambda_n, \eta, \tilde{c})$ and $\tilde{J}^*$ is the discounted cost value function of the belief-MDP with components $(\Pi, \mathbb{U}, \eta, \tilde{c})$.

4.2 Finite-State Approximation

The finite-state model for the belief-MDP is obtained as in [41]. We let d_Π denote a metric on Π which metrizes the weak topology.

Since Π is compact and thus totally bounded, there exists a sequence $(\{\pi_i^{(M)}\}_{i=1}^{k_M})_{M\geq 1} \subset \Pi$ of "finite grids" in Π such that $\forall M \geq 1$

$$\min_{i\in\{1,\ldots,k_M\}} d_\Pi(\pi, \pi_i^{(M)}) < \frac{1}{M}, \quad \forall \pi \in \Pi.$$

Let $\{B_i^{(M)}\}_{i=1}^{k_M}$ be a *partition* of Π such that $\pi_i^{(M)} \in B_i^{(M)}$ and

$$\max_{\pi\in B_i^{(M)}} d_\Pi(\pi, \pi_i^{(M)}) < \frac{1}{M}, \quad \forall i = 1, \ldots, k_M.$$

Intuitively this can be thought of as a set that contains all the nearest neighbours of $\pi_i^{(M)}$. Now let $\Pi^{(M)} := \{\pi_1^{(M)}, \ldots \pi_{k_M}^{(M)}\}$ and define the quantizer $\mathbf{Q}^{(M)} : \Pi \to \Pi^{(M)}$ by

$$\mathbf{Q}^{(M)}(\pi) = \pi_i^{(M)}, \quad \text{when } \pi \in B_i^{(M)}, \tag{18}$$

i.e., $\mathbf{Q}^{(M)}$ maps π to its nearest neighbour in Π_M. Now, let $\{\nu^{(M)}\}$ be a sequence of probability measures on Π such that $\nu^{(M)}(B_i^{(M)}) > 0$, $\forall i, M$. We let

$$\nu_i^{(M)}(\,\cdot\,) := \frac{\nu^{(M)}(\,\cdot\,)}{\nu^{(M)}(B_i^{(M)})} \tag{19}$$

be the restriction of $\nu^{(M)}$ to $B_i^{(M)}$. We will use $\nu_i^{(M)}$ to define a sequence of finite-state belief-MDPs denoted as MDP$^{(M)}$, which approximate the belief-MDP. Now for each M define the one-state cost $c^{(M)} : \Pi^{(M)} \times \mathbb{U} \to [0, \infty)$ and the transition probability $p^{(M)} : \Pi^{(M)} \times \mathbb{U} \to \mathcal{P}(\Pi^{(M)})$ on $\Pi^{(M)}$ as

$$c^{(M)}(\pi_i^{(M)}, u) := \int\limits_{B_i^{(M)}} \tilde{c}(\pi, u)\nu_i^{(M)}(d\pi) \tag{20}$$

and

$$p^{(M)}(\,\cdot\mid \pi_i^{(M)}, u) := \int\limits_{B_i^{(M)}} [\mathbf{Q}^{(M)} * \eta(\,\cdot\mid \pi, u)]\nu_i^{(M)}(d\pi), \tag{21}$$

where $\mathbf{Q}^{(M)} * \eta(\,\cdot\mid \pi, u) \in \mathcal{P}(\Pi^{(M)})$ is the pushforward of $\eta(\,\cdot\mid \pi, u)$ with respect to $\mathbf{Q}^{(M)}$, i.e.,

$$\mathbf{Q}^{(M)} * \eta(y \mid \pi, u) \in \mathcal{P}(\Pi^{(M)}) = \eta(\{\pi \in \Pi : \mathbf{Q}^{(M)}(\pi) = y\} \mid \pi, u), \quad \forall y \in \Pi^{(M)}. \tag{22}$$

Now, we define MDP$^{(M)} := (\Pi^{(M)}, \mathbb{U}, p^{(M)}, c^{(M)})$. Given this belief-MDP we can then apply [41, Theorem 3.2].

Theorem 4 (Saldi 2019 [41]). *Suppose Assumption 4 holds for the POMDP. Then we have*

$$\lim_{M\to\infty}\left|\tilde{J}(\gamma^{(M)},\mu)-\tilde{J}^*(\mu)\right|=0,.$$

where the policy $\gamma^{(M)}$ is obtained by extending the optimal policy of the MDP$^{(M)}$ from $\Pi^{(M)}$ to Π. Hence, by the equivalence of POMDPs and belief-MDPs we also have

$$\lim_{M\to\infty}\left|J(\pi^{\gamma^{(M)}},\mu)-J^*(\mu)\right|=0.$$

4.3 Belief Space Quantization

We first assume $\mathbb{M}$ to be some finite set which in the feature-based and volumetric paradigms of mapping holds true. In this case, we are primarily concerned with occupancy grid maps, which given some height H and width W are defined as $\mathbb{M} := \{0,1\}^{H\times W}$. We assume Assumption 3 to be true which directly implies Assumption 4 to be true. We can then apply the finite state approximation method from [41, Section V-B]. Since $\mathcal{P}(\mathbb{X})$ is compact, it can be metrized using the Wasserstein metric W_1 where we use the metric on $\mathbb{X}$ induced by the Euclidean norm $\|\cdot\|$ given by

$$W_1(\mu,\nu) := \inf_{\psi\in\mathcal{H}(\mu,\nu)}\int_{\mathbb{X}\times\mathbb{X}}\|x-x'\|\,\psi(dx,dx'),\tag{23}$$

where $\mathcal{H}(\mu,\nu)$ denotes the set of probability measures on $\mathbb{X}\times\mathbb{X}$ with first marginal μ and second marginal ν.

For each $n\geq 1$, let $\mathbf{Q}_n$ be some lattice quantizer on $\mathbb{X}$ such that $\|x-\mathbf{Q}_n(x)\|<\frac{1}{n}$ for all $x\in\mathbb{X}$. Set $\mathbb{X}_n := \mathbf{Q}_n(\mathbb{X})$ with $|\mathbb{X}_n| = m_n$. Since $\mathbb{X}$ is compact, then $\mathbb{X}_n$ is finite. Then, one can approximate any probability measure in $\mathcal{P}(\mathbb{X})$ with probability measures in

$$\mathcal{P}(\mathbb{X}_n) := \{\mu\in\mathcal{P}(\mathbb{X}) : \mu(\mathbb{X}_n)=1\},$$

Indeed, for any $\mu\in\mathcal{P}(\mathbb{X})$, we have

$$\begin{aligned}\inf_{\mu'\in\mathcal{P}(\mathbb{X}_n)}W_1(\mu,\mu') &\leq \inf_{\mathbf{Q}:\mathbb{X}\to\mathbb{X}_n}\int_{\mathbb{X}}\|x-\mathbf{Q}(x)\|\,\mu(dx)\\ &\leq \int_{\mathbb{X}}\|x-\mathbf{Q}_n(x)\|\,\mu(dx)\\ &\leq \frac{1}{n}.\end{aligned}\tag{24}$$

Now, taking $\mathbb{S}_n := \mathbb{X}_n\times\mathbb{M}$ to be our state space, we define the belief space as $\Pi_n := \mathcal{P}(\mathbb{S}_n)$. We should note that since we are now working with Π_n which by definition now works within $\mathbb{X}_n$ we should note that this induces some changes to the dynamics of the state for which the transition kernel is now defined as

$$\mathcal{T}_n(\cdot\mid x_i^n,u) := \int_{B_i^n}\mathbf{Q}_n*\mathcal{T}(\cdot\mid x,u)\nu_i^n(dx)\quad i\in\{1,...,m_n\}\tag{25}$$

where $\mathbb{X}_n=\{x_1^n,\dots,x_{m_n}^n\}$, $B_i^n=\{x\in\mathbb{X}:\mathbf{Q}_n(x)=x_i^n\}$, $\nu_i^n(\cdot)$ is a weighting measure, like the one defined in (33), and $\mathbf{Q}_n*\mathcal{T}(\cdot\mid x,u)\in\mathcal{P}(\mathbb{X}_n)$ is the pushforward of the measure $\mathcal{T}(\cdot\mid x,u)$ with respect to the quantizer $\mathbf{Q}_n$, that is

$$\mathbf{Q}_n*\mathcal{T}(x_i^n\mid x,u)=\mathcal{T}(B_i^n\mid x,u)\quad i=1,\dots,m_n.\tag{26}$$

Similarly, we define the new quantized cost as

$$c_n(x_i^n,m,u)=\int_{B_i^n}c(x,m,u)\nu_i^n(dx)\tag{27}$$

we define a weighting measure $\nu_i^n = \delta_{x_i^n}$ hence for equations (25)—(27) we obtain

$$\mathcal{T}_n(x_j^n \mid x_i^n, u) = \mathcal{T}(B_j^{(n)} \mid x_i^{(n)}, u) \quad i, j \in \{1, \dots, m_n\} \tag{28}$$

$$c_n(x_i^n, m, u) = c(x_i^n, m, u) \tag{29}$$

this defines a POMDP $(\mathbb{S}_n, \mathbb{U}, \mathbb{Y}, \mathcal{S}_n, Q, c_n, \beta)$ where $\mathcal{S}_n$ is defined similarly to (10) with sole change being in its use of $\mathcal{T}_n$ over $\mathcal{T}$. The equivalent belief-MDP$_n$ is then defined as $(\Pi_n, \mathbb{U}, \eta_n, \tilde{c}_n)$ where η_n follows the same construction as in (12)—(14) but with $\mathcal{T}_n$ and $\mathbb{X}_n$ substituted in for their non-quantized counterparts. The same connection holds for $\tilde{c}_n$ and (15).

We note that Π_n is a simplex in $\mathbb{R}^{N_n}$ since $\mathbb{S}_n$ is finite with $|\mathbb{S}_n| = m_n 2^{HW} =: N_n$. This fact allows us to use quantization methods as seen in [52]. In fact we can use this method to quantize Π_n in a nearest neighbour manner. In order to achieve this, for each $M \geq 1$ we define

$$\Pi_n^{(M)} := \left\{ (p_1, \dots, p_{N_n}) \in \mathbb{Q}^{N_n} : p_i = \frac{k_i}{M}, \sum_{i=1}^{N_n} k_i = M \right\} \tag{30}$$

where $\mathbb{Q}$ is the set of rational numbers and $k_1, \dots, k_m \in \mathbb{Z}_+$. Parameter M serves as a common denominator to all fractions, and can be used to control the density and number of points in $\Pi_n^{(M)}$. The quantized belief space, $\Pi_n^{(M)}$ is called the *type lattice* by analogy with the concept of *types* in information theory. Then, the algorithm that computes the nearest neighbour levels can be described as follows.

Algorithm 1 Reznik [52, Algorithm 1]

Require: $z = (z_1, \dots, z_{N_n})$, $\sum z_i = 1$, integer M $\quad\triangleright$ Given $z_n \in \Pi_n$, find nearest $y \in \Pi_n^{(M)}$

Ensure: $y = \left(\frac{k_1}{M}, \dots, \frac{k_{N_n}}{M}\right)$ with $\sum k_i = M$

$k_i' \leftarrow \lfloor M z_i + \frac{1}{2} \rfloor$ for $i = 1, \dots, N_n$

$M' \leftarrow \sum k_i'$, $\Delta \leftarrow M' - M$

if $\Delta = 0$ **then return** $\left(\frac{k_1'}{M}, \dots, \frac{k_{N_n}'}{M}\right)$

end if

$\delta_i \leftarrow k_i' - M z_i$ for $i = 1, \dots, N_n$

Sort δ so that $\delta_{i_1} \leq \dots \leq \delta_{i_{N_n}}$

if $\Delta > 0$ **then**

$$k_{i_j} \leftarrow \begin{cases} k_{i_j}' & j = 1, \dots, N_n - \Delta \\ k_{i_j}' - 1 & j = N_n - \Delta + 1, \dots, N_n \end{cases}$$

else

$$k_{i_j} \leftarrow \begin{cases} k_{i_j}' + 1 & j = 1, \dots, |\Delta| \\ k_{i_j}' & j = |\Delta| + 1, \dots, N_n \end{cases}$$

end if

return $\left(\frac{k_1}{M}, \dots, \frac{k_{N_n}}{M}\right)$

The number of types in lattice $\Pi_n^{(M)}$ depends on the size of the quantization term, M. It is essentially the number of partitions of M into N_n terms $k_1 + \dots + k_{N_n} = M$:

$$|\Pi_n^{(M)}| = \binom{M + N_n - 1}{N_n - 1}. \tag{31}$$

Also, according to [41], we can compute the maximum radius of the quantization regions for this algorithm. We have

$$b_\infty := \max_{\pi \in \Pi} \min_{\hat{\pi} \in \Pi_n^{(M)}} d_\infty(\pi, \hat{\pi}) = \frac{1}{M}\left(1 - \frac{1}{N_n}\right)$$

$$b_2 := \max_{\pi \in \Pi} \min_{\hat{\pi} \in \Pi_n^{(M)}} d_2(\pi, \hat{\pi}) := \frac{1}{M}\sqrt{\frac{a(N_n - a)}{N_n}}$$

$$b_1 \coloneqq \max_{\pi\in\Pi} \min_{\hat\pi\in\Pi_n^{(M)}} d_1(\pi,\hat\pi) = \frac{1}{M}\frac{2a(N_n-a)}{N_n}, \tag{32}$$

where $a = \lfloor N_n/2 \rfloor$. Hence, for each $M \geq 1$, the set $\Pi_n^{(M)}$ is a b_j-net in Π with respect to the d_j metric, where $j \in \{1, 2, \infty\}$. First we note that $\Pi_n^{(M)} \subset \Pi_n \subset \Pi$. Then, for any $\mu \in \Pi$ and $\mu_n \in \Pi_n$ such that $\mu_n = \arg\min_{\mu_n^*\in\Pi_n} W_1(\mu, \mu_n^*)$,

$$\begin{aligned}
\inf_{\mu_n^{(M)}\in\Pi_n^{(M)}} W_1(\mu,\mu_n^{(M)}) &\leq W_1(\mu,\mu_n) + \inf_{\mu_n^{(M)}\in\Pi_n^{(M)}} W_1(\mu_n,\mu_n^{(M)}) && \text{(triangle inequality)}\\
&\leq \frac{1}{n} + \mathrm{diam}(\mathbb{S}_n)\cdot \inf_{\mu_n^{(M)}\in\Pi_n^{(M)}} \|\mu_n-\mu_n^{(M)}\|_{TV} && \text{by (24) and [53, } Theorem\ 6.15]\\
&= \frac{1}{n} + \frac{\mathrm{diam}(\mathbb{S}_n)}{2}\cdot \inf_{\mu_n^{(M)}\in\Pi_n^{(M)}} \|\mu_n-\mu_n^{(M)}\|_1 && \text{by [54, } Proposition\ 4.2]\\
&\leq \frac{1}{n} + \frac{D}{2M}\max_{\mu_n\in\Pi_n}\min_{\mu_n^{(M)}\in\Pi_n^{(M)}} \|\mu_n-\mu_n^{(M)}\|_1\\
&\leq \frac{1}{n} + \frac{D}{M}\frac{a(N_n-a)}{N_n}, && \text{by (32)}
\end{aligned}$$

where $D \coloneqq \mathrm{diam}(\mathbb{S}_n)$. The second inequality is due to (24) and [53, Theorem 6.15] where for Polish $\mathbb{S}_n$ we have

$$W_1(\mu,\mu') \leq \mathrm{diam}(\mathbb{S}_n)\|\mu-\mu'\|_{TV}.$$

The third line uses the fact that $\mathbb{S}_n$ is countable to apply [54, Proposition 4.2]:

$$\|\mu-\mu'\|_{TV} = \frac{1}{2}\|\mu-\mu'\|_1.$$

To wrap up this section we need to define the weighting measure $\nu^{(M)}$, which for any $\pi \in \Pi_n$ is

$$\nu_i^{(M)}(\pi) \coloneqq \delta_{\pi_i^{(M)}}(\pi) \tag{33}$$

satisfying (19). We can then define the quantized cost $\tilde c_n^{(M)}$:

$$\tilde c_n^{(M)}(\pi_i^{(M)}, u) \coloneqq \tilde c_n(\pi_i^{(M)}, u) \quad i = 1, ..., N_n,\ \forall \pi \in B_i^{(M)} \tag{34}$$

as in (20) and the transition probability $p^{(M)}$

$$p_n^{(M)}(\pi_j^{(M)} \mid \pi_i^{(M)}, u) \coloneqq \eta_n(B_j^{(M)} \mid \pi_i^{(M)}, u) \quad i, j \in \{1, ..., N_n\} \tag{35}$$

as in (21).

Theorem 5. *Suppose Assumption 4 holds for the POMDP. Let $(M_1^n, M_2^n)_{n\geq 1}$ be a sequence of quantization parameters with $M_1^n, M_2^n \to \infty$ as $n \to \infty$. Let γ_n denote the policy obtained by extending the optimal policy of MDP$_{M_1^n}^{(M_2^n)}$ from $\Pi_{M_1^n}^{(M_2^n)}$ to $\Pi_{M_1^n}$ and again to Π. Then*

$$\lim_{n\to\infty} \left|\tilde J(\gamma_n,\mu) - \tilde J^*(\mu)\right| = 0.$$

Hence, by the equivalence of POMDPs and belief-MDPs we also have

$$\lim_{n\to\infty} |J(\pi^{\gamma_n},\mu) - J^*(\mu)| = 0.$$

5 Simulation results

5.1 Model description

We instantiate the active SLAM framework of Sections 2–4 for a mobile robot equipped with a range-finding sensor navigating an unknown environment represented as an occupancy grid map (OGM). This section specifies the motion model, sensor model, map representation, and cost function used in our simulations.

5.1.1 Map representation

As stated in Section 4.3, we represent the environment using an occupancy grid map $\mathbb{M} = \{0,1\}^{H\times W}$, where the workspace is discretized into an $H \times W$ grid of cells. Each cell $m^{ij} \in \{0,1\}$ indicates whether the cell at position (i,j) is occupied ($m^{ij} = 1$) or free ($m^{ij} = 0$). The map is assumed static throughout the mission, consistent with (2).

5.1.2 Motion model

We model the robot as a point mass with holonomic dynamics. This abstraction is appropriate for platforms where orientation is either mechanically decoupled from translational motion (e.g., omnidirectional wheeled robots with mecanum or omni-wheels) or where orientation control operates on a faster timescale than position control (e.g., quadrotors maintaining level flight). Such models are standard in multi-agent coordination [55, 56] and cooperative control [57] where the focus is on information-driven motion planning rather than individual vehicle kinematics.

The robot's state is described by its position $p(t) = [p^x(t), p^y(t)]^\top$ and velocity $v(t) = [v^x(t), v^y(t)]^\top$, yielding state $x(t) = [p(t)^\top, v(t)^\top]^\top \in \mathbb{R}^4$. The continuous-time dynamics follow a linear time-invariant double integrator model [58]:

$$dx(t) = (Ax(t) + Bu(t))dt + GdB_t \tag{36}$$

where $u(t) \in \mathbb{R}^2$ is the acceleration control input, B_t denotes Brownian motion with covariance $\sigma_w^2 \mathbf{I}$ and with $\mathbf{I}$ and $\mathbf{0}$ denoting the 2×2 identity and zero matrices the matrices in (36) are defined as

$$A = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad B = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}, \quad G = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}. \tag{37}$$

This model captures the essential second-order dynamics common to many robotic platforms while avoiding nonholonomic constraints.

We define the state space as

$$\mathbb{X} := \mathbb{P} \times \mathbb{V} \subset \mathbb{R}^4$$

where $\mathbb{P} = [0, L_x] \times [0, L_y]$ is the spatial domain covered by the occupancy grid map and $\mathbb{V} = [-v_{\max}, v_{\max}]^2$ bounds the robot's velocity, with $v_{\max} > 0$ representing the maximum achievable speed. The compactness of $\mathbb{X}$ satisfies Assumption 3-(3). Control inputs are constrained to $u_t \in \mathbb{U} = [-a_{\max}, a_{\max}]^2$ where $a_{\max} > 0$ is the maximum achievable acceleration, ensuring Assumption 3-(3) holds.

For computational tractability, we discretize (36) using zero-order hold on the control input with sampling period $\Delta t > 0$, yielding [59, Section 2.4]:

$$x_{t+1} = P_{\mathbb{P}}(Ax_t + Bu_t + w_t) \tag{38}$$

where

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} I & \Delta t\, I \\ \mathbf{0} & I \end{bmatrix}, \quad B = \begin{bmatrix} \frac{\Delta t^2}{2} & \mathbf{0} \\ \mathbf{0} & \frac{\Delta t^2}{2} \\ \Delta t & \mathbf{0} \\ \mathbf{0} & \Delta t \end{bmatrix} = \begin{bmatrix} \frac{\Delta t^2}{2} I \\ \Delta t\, I \end{bmatrix},$$

and $w_t \overset{iid}{\sim} \mathcal{N}(0, P^s)$ modeling discretized process noise with covariance [60, Section 6.2.2]

$$P^s = \begin{bmatrix} \frac{(\Delta t)^3}{3} & 0 & \frac{(\Delta t)^2}{2} & 0 \\ 0 & \frac{(\Delta t)^3}{3} & 0 & \frac{(\Delta t)^2}{2} \\ \frac{(\Delta t)^2}{2} & 0 & \Delta t & 0 \\ 0 & \frac{(\Delta t)^2}{2} & 0 & \Delta t \end{bmatrix} \sigma_w^2 = \begin{bmatrix} \frac{(\Delta t)^3}{3} I & \frac{(\Delta t)^2}{2} I \\ \frac{(\Delta t)^2}{2} I & \Delta t\, I \end{bmatrix} \sigma_w^2. \tag{39}$$

This discretization preserves the controllability and observability properties of the continuous-time system. The induced transition kernel

$$\mathcal{T}(D \mid x, u) = \int_D \frac{1}{(2\pi)^2 |Q|^{1/2}} \exp\left\{ -\frac{1}{2}(x' - Ax - Bu)^\top Q^{-1}(x' - Ax - Bu) \right\} \lambda(dx') \tag{40}$$

for any $D \in \mathcal{B}(\mathbb{X})$ is Gaussian with mean $Ax + Bu$ continuous in (x,u), so $\mathcal{T}$ is continuous in total variation, satisfying Assumption 1-(1).

5.1.3 Sensor model

Range finders measure the range to nearby objects and are often measured along a beam or within a cone which are good models of laser range finders (e.g., LIDAR) and ultrasonic sensors (e.g. sonar) respectively [5, pp 153]. Given the prevalence of LIDAR in modern autonomous robotic systems we chose to use a beam model of a range finder inspired by [61, Section II-A].

Let, $B \in \mathbb{N}$ be the number of beams emitted by the range finder, spaced evenly radially around the robot; $r_{\max} \in \mathbb{R}_{\geq 0}$ be the maximum range of any given beam; $\mathcal{R} := [0, r_{\max}]$, a Borel set, be the range of measurements that a beam can return; $\mathbb{Y} := \mathcal{R}^B$ be the measurement space. Any observation at time t is denoted as $y_t = [r_t^1, \ldots, r_t^B]^\top$ with each r_t^i being an $\mathcal{R}$-valued measurement returned from the i^{th} beam on a $i \cdot \frac{360}{B}$ angle from the x-axis at time t.

For each beam $k \in \{1, \ldots, B\}$, the measurement is governed by the following dynamics

$$r_t^k = \|p_t - c_k(p_t, m_t, \theta_i)\| + v_t, \tag{41}$$

where $v_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_{\text{hit}}^2)$ is Gaussian, $c_k : \mathbb{X} \times \mathbb{M} \times [0, 2\pi) \to \mathbb{R}^2$ is the *ray-casting function* that returns the position of the first occupied cell intersected by beam k emanating from position p_t at angle θ_k, or returns a point at distance $r_{\max}$ if no obstacle is encountered. The independence of measurements across beams and the continuity of the ray-casting function (away from map boundary transitions) ensure that the observation channel

$$Q(A \mid x, m) := \prod_{k=1}^{B} \int_{R^k} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(y - \|p - c_k(p, m, \theta_k)\|)^2}{2\sigma_v^2}\right) \lambda(dy) \tag{42}$$

for $A = R^1 \times \cdots \times R^B \in \mathcal{B}(\mathbb{Y})$ satisfies the continuity requirements of Assumption 2.

5.1.4 Cost function

The stage cost $c : \mathbb{S} \times \mathbb{U} \to \mathbb{R}_{\geq 0}$ balances coverage, localization, and control effort. It comprises three terms:

Coverage: Encourages map coverage, want to maximize

$$c_{\text{cov}}(m_t) = \sum_{i,j} \left| m_t^{ij} - \frac{1}{2} \right|. \tag{43}$$

Control effort: Penalizes large control inputs to encourage energy-efficient trajectories, want to minimize:

$$c_{\text{effort}}(u_t) := \|u_t\|^2. \tag{44}$$

Obstacle avoidance: Discourages motion toward occupied regions using a virtual force field (VFF) approach [62], want to minimize:

$$c_{\text{vff}}(x_t, m_t, u_t) := \max\left\{0, -\frac{\mathbf{F}_r(p_t, m_t) \cdot u_t}{\|\mathbf{F}_r(p_t, m_t)\|\|u_t\| + \epsilon}\right\}, \tag{45}$$

where $\mathbf{F}_r(p_t, m_t)$ is the repulsive force vector computed from nearby occupied cells and $\epsilon > 0$ prevents division by zero. The total stage cost is

$$c(s_t, u_t) = -\alpha_1 c_{\text{cov}}(m_t) + \alpha_2 c_{\text{effort}}(u_t) + \alpha_3 c_{\text{vff}}(x_t, m_t, u_t), \tag{46}$$

where $\alpha_1, \alpha_2, \alpha_3 > 0$ are weighting coefficients that balance the objectives. The expected cost (15) then becomes

$$\tilde{c}(\pi_t, u_t) = \int_{\mathbb{S}} c(s_t, u_t)\, \pi_t(ds_t). \tag{47}$$

The active SLAM problem is to find a policy $\gamma^* : \Pi \to \mathbb{U}$ that minimizes the infinite-horizon discounted cost (16) under the dynamics (38), observations (42), and cost (46)

5.2 Simulation details

For the simulation a quantization level of $n = 2$ was chosen leaving us with $m_n = 16$ and $|\mathbb{M}| = 16$ and $N_n = 256$. For the belief space the quantization level was set to be $M = 3$. As a result the size of the quantized belief space is $|\Pi_n^{(M)}| = \binom{N_n+M-1}{N_n-1} = 2,829,056$.

In order to run value iteration on the belief $\mathrm{MDP}_n^{(M)}$ we need to first pre-compute the quantized belief transition kernel from (35), which is for quantization level $n = 2$ a $2,829,056 \times 2,829,056 \times 4$ matrix.

Now in order to precompute this the transition $\eta_n(B_j^{(M)} \mid \pi_i^{(M)}, u)$ needs to be computed $\forall u \in \mathbb{U}_n$ and $\forall i, j \in \{1, \dots, N_n\}$. This computation (after extensive computational optimization and parallelization) takes approximately one second per computation but given that this needs to be computed $2,829,056 \cdot 2,829,056 \cdot 4 = 3.2 \times 10^{13}$ times which would approximately take one million years.

6 Localization or Mapping

This section will investigate localization and mapping separately through the lens of these finite approximations. We will begin with localization as this is the more straight forward of the two. We will then investigate the mapping problem which requires an augmented state space of its own, combining a quantized belief and a quantized state.

6.1 Active Localization

In contrast to the active SLAM problem, active localization assumes full knowledge of the map and the goal in this case is to navigate the robot in the environment in an effort to localize itself within the map.

The setup remains the same as in Section 2.1 with the sole difference being that access to m_t is now assumed for all $t \geq 0$. Again, similarly to active SLAM we formulate active localization as POMDP, fully described by the 7-tuple $(\mathbb{X}, \mathbb{U}, \mathbb{Y}, \mathcal{T}, Q, c, \beta)$, note that the sole difference from our definition in Section 2.2 is the definition of the transition kernel which here is instead equation (7). We then define belief as

$$\pi_t := \mathbb{P}(x_t \in \cdot \mid y_{[0,t]}, u_{[0,t-1]}) \tag{48}$$

which evolves via the Bayesian filtering equation

$$\pi_{t+1}(x) = \frac{Q(y_{t+1} \mid x) \int_{\mathbb{X}} \mathcal{T}(x \mid x_t, u_t)\pi_t(dx_t)}{\int_{\mathbb{X}} Q(y_{t+1} \mid x_{t+1}) \int_{\mathbb{X}} \mathcal{T}(dx_{t+1} \mid x_t, u_t)\pi_t(dx_t)} =: \mathbf{F}(\pi_t, u_t, y_{t+1})(x) \tag{49}$$

Again like in Section 2.3 we see the belief π_t is a sufficient statistic for optimal control allowing us to reformulate this as a belief MDP defined as the 4-tuple $(\mathcal{P}(\mathbb{X}), \mathbb{U}, \eta, \tilde{c})$. The distribution over future observations conditioned on the current belief and action, is given by

$$\begin{aligned} \mathbf{H}(C \mid \pi, u) &= \mathbb{P}(y_{t+1} \in C \mid \pi_t = \pi, u_t = u) \\ &= \int\limits_{\mathbb{X}} \int\limits_{\mathbb{X}} Q(C \mid x_{t+1})\mathcal{T}(dx_{t+1} \mid x_t, u_t)\pi_t(dx_t) \end{aligned} \tag{50}$$

for any $C \in \mathcal{B}(\mathbb{Y})$. Finally, the belief transition kernel induced by the filter update (49) and predictive observation distribution (50) is given by

$$\eta(\cdot \mid \pi, u) = \int\limits_{\mathbb{Y}} \delta_{\mathbf{F}(\pi,u,y)}(\cdot)\mathbf{H}(dy \mid \pi, u) \tag{51}$$

and $\tilde{c} : \mathcal{P}(\mathbb{X}) \times \mathbb{U} \to \mathbb{R}$ is the expected stage cost given by

$$\tilde{c}(\pi_t, u_t) = \int_{\mathbb{X}} c(x_t, u_t)\, \pi_t(dx_t). \tag{52}$$

The optimal policy $\gamma^* : \mathcal{P}(\mathbb{X}) \to \mathbb{U}$ minimizes the expected discounted cost

$$J^*(\pi_0) = \inf_{\gamma \in \Gamma} \mathbb{E}_{\pi_0}^{\gamma} \left[\sum_{t=0}^{\infty} \beta^t \tilde{c}(\pi_t, u_t) \right], \tag{53}$$

6.2 Active Mapping

In the active mapping problem, we reverse the assumptions of active localization, this time assuming full knowledge of our position x_t for any time $t \geq 0$ but now treating our map m_t as the partially observed quantity. We will attempt to treat this problem more carefully as it has some finer differences from SLAM and active localization that force us to our POMDP and belief MDP in a different fashion.

First, we would like to consider our information structure, which in this case is described by the following information variables:

$$I_t = \{x_{[0,t]}, y_{[0,t]}, u_{[0,t-1]}\}, \quad I_0 = \{x_0, y_0\}. \tag{54}$$

Now, since our map is not observable we would like to condition our decision making instead on our belief which is defined as the posterior distribution of the map given the state, observation, and action history

$$\pi_t(\cdot) \coloneqq \mathbb{P}(m_t \in \cdot \mid x_{[0,t]}, y_{[0,t]}, u_{[0,t-1]}). \tag{55}$$

We first make the claim that given the entire history of the states and observations that the map is conditionally independent of the history of actions (this only holds when the observation channel is independent of the control variable which is the case in standard robotics measurement models). This allows us to equivalently define the belief as

$$\pi_t(\cdot) \coloneqq \mathbb{P}(m_t \in \cdot \mid x_{[0,t]}, y_{[0,t]}) \tag{56}$$

which evolves via the following Bayesian recursion

$$\pi_{t+1}(m) = \frac{Q(y_{t+1} \mid x_{t+1}, m)\pi_t(m)}{\int_{\mathbb{M}} Q(y_{t+1} \mid x_{t+1}, m_t)\pi_t(dm_t)} \eqqcolon \mathbf{F}(\pi_t, x_{t+1}, y_{t+1})(m) \tag{57}$$

which uses the fact that $m_t \perp\!\!\!\perp x_t \mid x_{[0,t-1]}, y_{[0,t-1]}$. Now, due to the stochasticity of our state we must define an augmented state for both our state and belief on our map which we define as $b_t \coloneqq (x_t, \pi_t) \in \mathbb{X} \times \mathcal{P}(\mathbb{M})$ and we define our new space that this state lives on as $\mathfrak{B} \coloneqq \mathbb{X} \times \mathcal{P}(\mathbb{M})$. We then claim $\{b_t, u_t\}$ is a controlled Markov chain that allows us to define an equivalent MDP as the 4-tuple $(\mathfrak{B}, \mathbb{U}, \eta, \tilde{c})$ where $\eta : \mathfrak{B} \times \mathbb{U} \to \mathcal{P}(\mathfrak{B})$ is the transition kernel induced by the augmented state space's filter update (57) given by

$$\begin{aligned}
\eta(b_{t+1} \mid b_t, u_t) &= \eta(\pi_{t+1}, x_{t+1} \mid x_t, \pi_t, u_t) && (58)\\
&= \mathcal{T}(x_{t+1} \mid x_t, u_t) \int_{\mathbb{Y}} \delta_{\mathbf{F}(\pi_t, x_{t+1}, y_{t+1})}(\pi_{t+1}) \int_{\mathbb{M}} Q(y_{t+1} \mid x_{t+1}, m)\pi_t(dm) && (59)\\
&= \mathcal{T}(x_{t+1} \mid x_t, u_t) \cdot \mathbb{P}(\pi_{t+1} \mid x_{t+1}, \pi_t) && (60)
\end{aligned}$$

where

$$\mathbb{P}(\pi_{t+1} \mid x_{t+1}, \pi_t) \coloneqq \int_{\mathbb{Y}} \delta_{\mathbf{F}(\pi_t, x_{t+1}, y_{t+1})}(\pi_{t+1}) \int_{\mathbb{M}} Q(y_{t+1} \mid x_{t+1}, m)\pi_t(dm) \tag{61}$$

is the posterior distribution of the map.

References

[1] C. Stachniss, J. J. Leonard, and S. Thrun, "Simultaneous localization and mapping," *Springer Handbook of Robotics*, pp. 1153–1176, 2016.

[2] C. Stachniss, B. Siciliano, O. Khatib, and F. C. A. Groen, *Robotic mapping and exploration*. Springer Tracts in Advanced Robotics, 55, Berlin ;: Springer Berlin Heidelberg, 1st ed. 2009. ed., 2009.

[3] C.-C. Wang, C. Thorpe, S. Thrun, M. Hebert, and H. Durrant-Whyte, "Simultaneous localization, mapping and moving object tracking," *The International Journal of Robotics Research*, vol. 26, no. 9, pp. 889–916, 2007.

[4] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[5] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA: MIT Press, 2005.

[6] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit, *et al.*, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *Proc. International Joint Conference on Artificial Intelligence*, vol. 3, pp. 1151–1156, 2003.

[7] J. A. Placed, J. Strader, H. Carrillo, N. Atanasov, V. Indelman, L. Carlone, and J. A. Castellanos, "A survey on active simultaneous localization and mapping: State of the art and new frontiers," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1686–1705, 2023.

[8] R. Sim, "Stable exploration for bearings-only SLAM," in *Proc. IEEE International Conference on Robotics and Automation*, pp. 2411–2416, 2005.

[9] K. Tateno, F. Tombari, I. Laina, and N. Navab, "CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[10] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous Robot Vehicles* (I. J. Cox and G. T. Wilfong, eds.), pp. 167–193, New York, NY: Springer New York, 1990.

[11] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, "Consistency of the EKF-SLAM algorithm," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3562–3568, 2006.

[12] S. Huang and G. Dissanayake, "Convergence and consistency analysis for extended Kalman filter based slam," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 1036–1049, 2007.

[13] K. Murphy and S. Russell, "Rao-blackwellised particle filtering for dynamic bayesian networks," in *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. de Freitas, and N. Gordon, eds.), pp. 499–515, New York, NY: Springer New York, 2001.

[14] T. Bailey, J. Nieto, and E. Nebot, "Consistency of the fastSLAM algorithm," in *Proc. IEEE International Conference on Robotics and Automation. (ICRA)*, pp. 424–429, 2006.

[15] T. D. Barfoot, *State estimation for robotics*. Cambridge University Press, 2024.

[16] S. Thrun and M. Montemerlo, "The graph SLAM algorithm with applications to large-scale mapping of urban structures," *The International Journal of Robotics Research*, vol. 25, no. 5-6, pp. 403–429, 2006.

[17] F. Dellaert and M. Kaess, "Square root SAM: Simultaneous localization and mapping via square root information smoothing," *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1181–1203, 2006.

[18] M. Kaess, A. Ranganathan, and F. Dellaert, “iSAM: Incremental smoothing and mapping,” *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1365–1378, 2008.

[19] 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*, vol. 31, no. 2, pp. 216–235, 2012.

[20] F. Dellaert and G. Contributors, “borglab/gtsam,” May 2022.

[21] T.-C. Dong-Si and A. I. Mourikis, “Consistency analysis for sliding-window visual odometry,” in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5202–5209, 2012.

[22] R. M. Eustice, H. Singh, and J. J. Leonard, “Exactly sparse delayed-state filters for view-based slam,” *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1100–1114, 2006.

[23] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, “ORB-SLAM: A versatile and accurate monocular slam system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[24] A. Barrau and S. Bonnabel, “The invariant extended kalman filter as a stable observer,” *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1797–1812, 2017.

[25] S. J. Julier and J. K. Uhlmann, “A counter example to the theory of simultaneous localization and map building,” in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, vol. 4, pp. 4238–4243, IEEE, 2001.

[26] A. Barrau and S. Bonnabel, “An EKF-SLAM algorithm with consistency properties,” *CoRR*, vol. abs/1510.06263, 2015.

[27] A. Barrau and S. Bonnabel, “Invariant Kalman filtering,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. Volume 1, 2018, pp. 237–257, 2018.

[28] M. Brossard, S. Bonnabel, and A. Barrau, “Invariant Kalman filtering for visual inertial slam,” in *Proc. 21st International Conference on Information Fusion (FUSION)*, pp. 2021–2028, 2018.

[29] M. Brossard, S. Bonnabel, and J.-P. Condomines, “Unscented Kalman filtering on lie groups,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2485–2491, 2017.

[30] R. Mahony and T. Hamel, “A geometric nonlinear observer for simultaneous localisation and mapping,” in *Proc. IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 2408–2415, 2017.

[31] V. Indelman, L. Carlone, and F. Dellaert, “Planning in the continuous domain: A generalized belief space approach for autonomous navigation in unknown environments,” *The International Journal of Robotics Research*, vol. 34, no. 7, pp. 849–882, 2015.

[32] R. Platt Jr, R. Tedrake, L. Kaelbling, and T. Lozano-Perez, “Belief space planning assuming maximum likelihood observations,” in *Proc. Robotics: Science and Systems Conference*, 2010.

[33] J. Van Den Berg, S. Patil, and R. Alterovitz, “Motion planning under uncertainty using iterative local optimization in belief space,” *The International Journal of Robotics Research*, vol. 31, no. 11, pp. 1263–1278, 2012.

[34] S. Rahman and S. L. Waslander, “Uncertainty-constrained differential dynamic programming in belief space for vision based robots,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3112–3119, 2021.

[35] S. Koga, A. Asgharivaskasi, and N. Atanasov, “Active SLAM over continuous trajectory and control: A covariance-feedback approach,” in *Proc. American Control Conference (ACC)*, pp. 5062–5068, 2022.

[36] J. M. Porta, M. T. J. Spaan, and N. Vlassis, “Robot planning in partially observable continuous domains,” in *Robotics: Science and Systems*, (Cambridge, USA), June 2005.

[37] C. Stachniss, G. Grisetti, and W. Burgard, "Information gain-based exploration using rao-blackwellized particle filters.," in *Robotics: Science and Systems*, vol. 2, pp. 65–72, 2005.

[38] L. Carlone, J. Du, M. Kaouk Ng, B. Bona, and M. Indri, "Active SLAM and Exploration with particle filters using Kullback-Leibler divergence," *Journal of Intelligent & Robotic Systems*, vol. 75, pp. 291–311, Aug. 2014.

[39] H. Carrillo, P. Dames, V. Kumar, and J. A. Castellanos, "Autonomous robotic exploration using a utility function based on Rényi's general theory of entropy," *Autonomous Robots*, vol. 42, pp. 235–256, Feb. 2018.

[40] B. J. Julian, S. Karaman, and D. Rus, "On mutual information-based control of range sensing robots for mapping applications," *The International Journal of Robotics Research*, vol. 33, no. 10, pp. 1375–1392, 2014.

[41] N. Saldi, S. Yüksel, and T. Linder, "Asymptotic optimality of finite model approximations for partially observed Markov decision processes with discounted cost," *IEEE Transactions on Automatic Control*, vol. 65, no. 1, pp. 130–142, 2019.

[42] A. Kara and S. Yüksel, "Near optimality of finite memory feedback policies in partially observed Markov decision processes," *Journal of Machine Learning Research*, vol. 23, no. 11, pp. 1–46, 2022.

[43] A. Kara, N. Saldi, and S. Yüksel, "Q-learning for MDPs with general spaces: Convergence and near optimality via quantization under weak continuity," *Journal of Machine Learning Research*, vol. 24, no. 199, pp. 1–34, 2023.

[44] A. D. Kara and S. Yüksel, "Convergence of finite memory Q learning for POMDPs and near optimality of learned policies under filter stability," *Mathematics of Operations Research*, vol. 48, no. 4, pp. 2066–2093, 2023.

[45] L. Cregg, T. Linder, and S. Yüksel, "Reinforcement learning for near-optimal design of zero-delay codes for Markov sources," *IEEE Transactions on Information Theory*, vol. 70, no. 11, pp. 8399–8413, 2024.

[46] A. D. Kara and S. Yüksel, "Partially observed optimal stochastic control: Regularity, optimality, approximations, and learning," in *Proc. IEEE 63rd Conference on Decision and Control (CDC)*, pp. 6709–6721, 2024.

[47] A. Asgharivaskasi and N. Atanasov, "Semantic octree mapping and shannon mutual information computation for robot exploration," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1910–1928, 2023.

[48] A. A. Yushkevich, "Reduction of a controlled Markov model with incomplete data to a problem with complete information in the case of borel state and control space," *Theory of Probability & Its Applications*, vol. 21, no. 1, pp. 153–158, 1976.

[49] D. Rhenius, "Incomplete information in Markovian decision models," *The Annals of Statistics*, vol. 2, no. 6, pp. 1327–1334, 1974.

[50] D. Blackwell, "Memoryless strategies in finite-stage dynamic programming," *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 863–865, 1964.

[51] A. D. Kara, N. Saldi, and S. Yüksel, "On weak Feller continuity properties of non-linear filters," in *Proc. IEEE 58th Conference on Decision and Control (CDC)*, pp. 654–659, 2019.

[52] Y. A. Reznik, "An algorithm for quantization of discrete probability distributions," in *Proc. Data Compression Conference*, pp. 333–342, IEEE, 2011.

[53] C. Villani, *Optimal transport: Old and new*. Grundlehren der mathematischen Wissenschaften ; 338, Berlin: Springer, 2009.

[54] D. A. Levin and Y. Peres, *Markov chains and mixing times*, vol. 107. American Mathematical Soc., 2017.

[55] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[56] Y. Zhang, Y. Liu, J. Qiu, and J. Cao, "Coordination on double-integrator antagonistic dynamics robust to velocity constraint," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 599–609, 2023.

[57] D. Meng and Y. Jia, "Robust consensus algorithms for multiscale coordination control of multi-vehicle systems with disturbances," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 2, pp. 1107–1119, 2016.

[58] P. S. Maybeck, *Stochastic models, estimation and control*, vol. 1 of *Mathematics in science and engineering*. New York: Academic Press, 1979.

[59] F. L. Lewis, *Optimal estimation : with an introduction to stochastic control theory*. New York: Wiley, 1986.

[60] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*. New York: Wiley, 2001.

[61] G. Caroleo, A. Albini, D. De Martini, T. D. Barfoot, and P. Maiolino, "Tiny lidars for manipulator self-awareness: Sensor characterization and initial localization experiments," *arXiv preprint arXiv:2503.03449*, 2025.

[62] J. Borenstein and Y. Koren, "Real-time obstacle avoidance for fast mobile robots," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 19, no. 5, pp. 1179–1187, 1989.