Optimal Path Planning and Resource Allocation for Active Target Localization

Charles Freundlich, Philippos Mordohai, and Michael M. Zavlanos

Abstract— **This paper addresses optimal path planning and resource allocation for active multi-target localization. For each target, we solve a local Dynamic Program (DP) that plans optimal trajectories in the joint state-space of robot positions and target location uncertainties, captured by a cumulative error covariance matrix. The transitions in the space of robot positions are governed by the robot dynamics, while the transitions in the space of target uncertainties are regulated by a Kalman filter (KF) that fuses new information about the target locations with the current beliefs. The fused target uncertainties enter the objective function of the local DP using the trace of the associated covariance matrix. Using the optimal sensing policies local to each target, we construct a global DP to determine how far along the single target optimal trajectories the sensor should travel before transitioning to the next target. The integrated system jointly optimizes the collective target localization uncertainty and the total distance traveled by the sensing agent. The proposed control scheme is more computationally efficient than methods that use only the sensor configuration to compute future uncertainty and more exact than methods that abstract away the filtered sensing uncertainty.**

I. Introduction

Localization focuses on determining the exact location of a source, robot, person, or otherwise interesting object in a possibly noisy or cluttered environment. For a growing number of commercial and defense applications, this needs to be done with very high accuracy over long periods of time and for large numbers of objects. Compared to static arrays or remotely operated mobile sensors, autonomous robots have the potential to accomplish this task more efficiently and with less human intervention.

In this paper, we address the problem of optimal path planning and resource allocation for active multi-target localization. We compute an optimal control policy that guides a mobile sensing agent, i.e. a robot, through a sequence of configurations that maximize the expected target localization accuracy while minimizing the total distance travelled. General approaches to problems such as this formulate a Partially Observable Markov Decision Process (POMDP), where the targets take on the characteristics of a Hidden Markov Model (HMM) [1]–[3]. Finding the optimal control policy for a POMDP is notoriously difficult because these models reason over all possible future control outcomes and observations. Consequently, specialized approximation algorithms that are intended to mitigate the path dependence that arises are required [4]–[6]. In this work, we assume that the robot is self localized and its exteroceptive sensor is corrupted by approximately Gaussian noise, which is propagated to the target locations also as a Gaussian uncertainty. Under these assumptions, the active sensing problem has received significant attention in the controls community as far back as 1967 [7], recently [8], [9], and in the active vision community [10]–[15]. The key idea that enables our work is to formulate the active sensing problem as a Dynamic Program (DP) that incorporates the target uncertainty (along with the sensor positions) in the state-space and employs a Kalman filter (KF) to regulate the dynamics in this uncertainty space. By defining the state-space in this way, the instantaneous reward only relies on the current state-action pair. In typical POMDP [1]–[3], optimal control [7]–[9], and receding horizon [16] approaches, the objective function at any given control instance depends on the full history of observations.

To obtain a method that is tractable for problems of practical size, defined in terms of the number of targets, in this paper we develop a hierarchical control scheme for multi-target active localization that consists of the following two components: a *local DP* that finds the optimal sequence of sensor configurations in the vicinity of a single target and a *global DP* that balances between the time spent at every target following the locally optimal sensing policy and the total distance travelled in order to best localize all targets. Both DPs have reward functions that capture the positive impact of uncertainty reduction and the negative impact of the distance the sensor must travel between measurements. Like the Traveling Salesman Problem (TSP) approaches to active sensing [17]–[19], our method uses a binary flag to represent whether or not a target has been sensed, except that for us, the robot only deems a given target “sensed” when the potential for reward elsewhere outweighs the potential for reward at the current site. Compared to POMDP approaches, the complexity reduction allows us to solve problems of practical size, using exact value iteration on the global DP, that would otherwise be intractable.

The paper is organized as follows. In Section II, we formulate the active localization problem using Dynamic Programming. In Sections III and IV, we detail our hierarchical control scheme that integrates the local and global DPs in a multi-target active localization mission. In Section V, we show simulations of an active stereo vision sensor that localizes several targets.

This work is supported in part by the National Science Foundation under awards No. DGE-1068871 and IIS-1217797.

Charles Freundlich and Michael M. Zavlanos are with the Dept. of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA {charles.freundlich, michael.zavlanos}@duke.edu. Philippos Mordohai is with the Dept. of Computer Science, Stevens Institute of Technology, Hoboken, NJ 07030, USA mordohai@cs.stevens.edu.

II. Problem Formulation

Consider a mobile sensing robot with configuration $\mathbf{p} \in \mathcal{W} \subset \mathbb{R}^2 \times [0, 2\pi)$, which corresponds to a position in $\mathbb{R}^2$ and orientation in the half open interval $[0, 2\pi) \subset \mathbb{R}$. In this paper, the workspace $\mathcal{W}$ is a finite set. The robot has deterministic dynamics $\boldsymbol{\phi} \colon \mathcal{W} \times \mathcal{U} \to \mathcal{W}$, where $\mathcal{U}$ is the set of admissible control inputs, which, in this paper, are combinations of physical motion and measurement action, e.g., 'move north and observe target i.' Assume there are M static point targets at locations $\left\{\mathbf{x}_i \in \mathbb{R}^2\right\}_{i=1}^{M}$, of which the agent can only observe one at a time. Every observation of target i, which we denote by $\mathbf{y}_i$, is corrupted by zero mean Gaussian noise with error covariance matrix $\mathbf{Q}\left(\mathbf{x}_i, \mathbf{p}\right) \in \mathbb{S}_+^2$ that depends on both range and bearing, or simply $\mathbf{Q}_{i,\mathbf{p}}$ when the meaning is clear. We use the symbol $\mathbb{S}_+^n$ to denote the set of $n \times n$ symmetric positive semidefinite matrices. The sensing agent acquires a sequence of observations $\{\mathbf{y}_{i,k}\}$ of target i with measurement error covariances $\{\mathbf{Q}_{i,\mathbf{p}_k}\}$ from various vantage points along a controlled trajectory $\{\mathbf{p}_k\} \subset \mathcal{W}$, where k denotes a time index. A KF fuses observations as they are acquired, producing a sequence of estimates $\{\hat{\mathbf{x}}_{i,k}\}$ and their associated error covariances $\{\boldsymbol{\Sigma}_{i,k}\}$. In particular, if the robot observes target i at the k-th time step while being in configuration $\mathbf{p}_k$, then the KF fuses the prior belief $\mathcal{N}\left(\hat{\mathbf{x}}_{i,k-1}, \boldsymbol{\Sigma}_{i,k-1}\right)$ with the new observation $\mathcal{N}\left(\mathbf{y}_{i,k}, \mathbf{Q}_{i,\mathbf{p}_k}\right)$ to obtain the new target location distribution $\mathcal{N}\left(\hat{\mathbf{x}}_{i,k}, \boldsymbol{\Sigma}_{i,k}\right)$, where $\mathcal{N}(\hat{\mathbf{x}}, \boldsymbol{\Sigma})$ denotes the normal distribution with mean $\hat{\mathbf{x}}$ and covariance matrix $\boldsymbol{\Sigma}$. Initialization of the KF assumes that the robot has access to an *a priori* set of target location estimates $\{\hat{\mathbf{x}}_{i,0}\}_{i=1}^{M}$, which are considered to be normally distributed random variables with zero mean and error covariance matrices $\{\boldsymbol{\Sigma}_{i,0}\}_{i=1}^{M}$.

As discussed in Section I, our goal is to minimize the variance in the target locations as well as the distance the robot needs to travel while observing the targets in order to achieve this goal. We denote by $\boldsymbol{\psi} \colon \mathcal{U} \to \mathbb{R}_+$ a metric that measures the distance the agent needs to travel as a result of actions in $\mathcal{U}$. Then, given a factor $\rho \in [0, 1]$ that controls the tradeoff between target uncertainty reduction and distance travelled, and a sufficiently large horizon length K, we seek a control sequence $\{\mathbf{u}_k\}_{k=0}^{K-1}$ that minimizes the objective

$$H\left(\{\mathbf{u}_k\}_{k=0}^{K-1}\right) = (1-\rho)\sum_{i=1}^{M}\left(\mathbf{tr}(\boldsymbol{\Sigma}_{i,K})\right)^{1/2} + \rho\sum_{k=0}^{K-1}\boldsymbol{\psi}(\mathbf{u}_k).$$

Letting the horizon length $K \to \infty$, it can be shown that the infinite horizon DP,

$$\max_{\{\mathbf{u}_k\}\subset\mathcal{U}}\sum_{k=0}^{\infty}\gamma^k\left[(1-\rho)\sum_{i=1}^{M}\Big(\mathbf{tr}(\boldsymbol{\Sigma}_{i,k}-\boldsymbol{\Sigma}_{i,k+1})\Big)^{\frac{1}{2}} - \rho\boldsymbol{\psi}(\mathbf{u}_k)\right]$$

$$\text{s.t.}\quad \mathbf{p}_{k+1} = \boldsymbol{\phi}\left(\mathbf{p}_k, \mathbf{u}_k\right) \tag{1a}$$

$$\boldsymbol{\Sigma}_{i,k+1} = \begin{cases}\left(\boldsymbol{\Sigma}_{i,k}^{-1} + \mathbf{Q}_{i,\mathbf{p}_{k+1}}^{-1}\right)^{-1} & \text{if } i = \mathbf{i}_{k+1}\\ \boldsymbol{\Sigma}_{i,k} & \text{else}\end{cases}, \tag{1b}$$

with initial condition $\mathbf{p}_0$ and *a priori* error covariance $\boldsymbol{\Sigma}_{i,0}$ for each i is a valid approximation of the minimization problem involving the objective $H(\cdot)$ above. The parameter ρ controls the trade-off between the effort spent sensing targets and traveling to obtain new measurements. As localization uncertainty and distance travelled can differ by orders of magnitude, ρ provides a way to linearly scale those quantities so that they can be fairly compared. The robot dynamics (1a) are problem-specific; we define a simple possibility in Section III. In (1b), $\mathbf{i}_{k+1} \in \{1, \ldots, M\}$ is a scalar indicator function that determines which target is observed at time step $k+1$. Therefore, the condition $i = \mathbf{i}_{k+1}$ means that target i is the only target that is observed at time $k+1$. The particular dynamics of $\mathbf{i}_{k+1}$ are discussed in details in Section IV. The constant $\gamma \in (0, 1]$ in (1) is a user-specified discount factor for future rewards, which is usually set to 0.9 or higher.

In the next two sections, we propose a hierarchical control scheme to solve problem (1). First, we solve a set of M local Dynamic Programs (DPs), one for every target. Each local DP is defined as in (1), assuming $M = 1$. The state-space of each local DP consists of two parts: a finite local configuration space confined to the vicinity of every target and a finite representation of the set of target uncertainties, or covariance matrices, that are reachable by any observation sequence generated by the robot. Transitions in this joint state-space are governed by the robot dynamics (1a) and the KF update of the target error covariance (1b), respectively, and the reward function in (1) balances between uncertainty reduction and distance traveled by the sensor. Then, we construct a second, global DP that determines for how long the agent should follow the local optimal sensing policy at each target before transitioning to the next target. The integrated system jointly optimizes the collective target localization uncertainty and the total distance traveled by the mobile sensor, as per the requirements of problem (1).

III. The Single Target Case

In this section, we propose a method to solve problem (1) for one target. We call this the *local DP*. We solve M instances of the local DP, one for every target, and the obtained local solutions constitute the inputs to the global DP discussed later in Section IV. To construct the local state-space for target i, we approximate the space of reachable covariances by a discrete subset of the positive semidefinite matrices, which we denote by $\mathcal{C}_i \subset \mathbb{S}_+^2$. We discuss the specifics of designing $\mathcal{C}_i$ in Section III-A. We use the finite set $\mathcal{C}_i$ as a component of the local state-space, given by

$$\mathcal{S}_i \triangleq \mathcal{W}_i \times \mathcal{C}_i = \left\{\mathbf{s}_i \triangleq (\mathbf{p}_i, \boldsymbol{\Sigma}_i) \mid \mathbf{p}_i \in \mathcal{W}_i \text{ and } \boldsymbol{\Sigma}_i \in \mathcal{C}_i\right\},$$

where $\mathcal{W}_i$ denotes the discrete set of robot configurations in the vicinity of target i. A state $\mathbf{s}_i' \in \mathcal{S}_i$ is accessible from $\mathbf{s}_i \in \mathcal{S}_i$ if there exists a control input $\mathbf{u} \in \mathcal{U}_i$ that satisfies the joint dynamical equation

$$\mathbf{p}_i' = \boldsymbol{\phi}\left(\mathbf{p}_i, \mathbf{u}\right) \tag{2a}$$

$$\boldsymbol{\Sigma}_i' = \Pi_{\mathcal{C}_i}\left[\boldsymbol{\Sigma}_i^{-1} + \mathbf{Q}_{i,\boldsymbol{\phi}(\mathbf{p}_i,\mathbf{u})}^{-1}\right]^{-1}, \tag{2b}$$

where $\mathcal{U}_i$ is the set of control options for the robot when it is in the local state-space $\mathcal{S}_i$, $\boldsymbol{\phi} \colon \mathcal{W}_i \times \mathcal{U}_i \to \mathcal{W}_i$ is the deterministic robot dynamics, and $\Pi_{\mathcal{C}_i}$ in (2b) is the projection of the updated KF covariance to the finite set $\mathcal{C}_i$. The projection in (2b) is necessary because we require that $\mathbf{\Sigma}_i' \in \mathcal{C}_i$, and $\mathcal{C}_i$, as a finite set, is inherently much smaller than $\mathbb{S}_+^2$. Let $\mathbf{\Phi}_i \colon \mathcal{S}_i \times \mathcal{U}_i \to \mathcal{S}_i$ denote the joint dynamics (2). It constitutes the transition function for the local DP. Then, denote by $R_i \colon \mathcal{S}_i \times \mathcal{U}_i \to \mathbb{R}$ the instantaneous reward from problem (1), given by

$$R_i(\mathbf{s}_i, \mathbf{u}) \triangleq -\rho\psi(\mathbf{u}) + (1-\rho)\left(\mathbf{tr}\left[\mathbf{\Sigma}_i - \mathbf{\Sigma}_i'\right]\right)^{1/2}, \quad (3)$$

where $\mathbf{s}_i' = \mathbf{\Phi}_i(\mathbf{s}_i, \mathbf{u})$. We seek the optimal policy $\boldsymbol{\mu}_i^* : \mathcal{S}_i \to \mathcal{U}_i$ that solves the following problem, which, because of the covariance projection, is an approximation to (1) for $M = 1$:

$$\max_{\boldsymbol{\mu}_i} \sum_{k=0}^{\infty} \gamma^k R_i(\mathbf{s}_i, \boldsymbol{\mu}_i(\mathbf{s}_i)) \quad (4)$$

We use value iteration to determine the local *optimal value function* $V_i^* \colon \mathcal{S}_i \to \mathbb{R}$, which maps local states in $\mathcal{S}_i$ to a scalar representing the potential for discounted future reward. Computing V_i^* yields the local *optimal policy* $\boldsymbol{\mu}_i^* \colon \mathcal{S}_i \to \mathcal{U}_i$ as a byproduct.

A. The uncertainty state-space and transition function

If $\mathcal{C}_i$ is manageable in size, the local DP in (4) admits a solution using value iteration. If we used all of $\mathbb{S}_+^2$ as the covariance component of the state-space, an approximate form of the value function would be necessary; see, e.g., [20]. In the context of the present problem, function approximation is not straightforward, and, instead, we propose a specialized method to discretize $\mathbb{S}_+^2$ based on the parameters of the problem. We emphasize that, in general, finding the optimal discrete representation of $\mathbb{S}_+^n$ is not trivial [21]. The following method works well within the active sensing context for two dimensional ($n = 2$) environments.

Let $\boldsymbol{\lambda}_1 \colon \mathbb{S}_+^2 \to \mathbb{R}_+$ denote the maximum eigenvalue function of a 2×2 positive semidefinite matrix, $\boldsymbol{\lambda}_2 \colon \mathbb{S}_+^2 \to \mathbb{R}_+$ denote the minimum eigenvalue function, and $\boldsymbol{\theta} \colon \mathbb{S}_+^2 \to [0, \pi)$ denote the counter-clockwise angle between a line parallel to the eigenvector corresponding to the maximum eigenvalue and the positive x-axis. Note that this angle is always contained in $[0, \pi)$. In what follows, we use $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$, and $\boldsymbol{\theta}$ as features of matrices in $\mathbb{S}_+^2$. We discretize the space of feature vectors using the properties of the KF. We begin by the following lemma, which will be useful for determining a bound on the trace of the target location error covariance matrices that are reachable in our state-space by any sequence of observations that the sensor can make. The proof is similar to that of Lemma 2.7 in [22].

Lemma 3.1: Let $n \in \mathbb{N}$. Denote by $\mathbb{S}_{++}^n$ the set of $n \times n$ symmetric *positive definite* matrices. If $\mathbf{A}, \mathbf{B} \in \mathbb{S}_{++}^n$, then

$$\mathbf{tr}\left(\mathbf{A}^{-1} + \mathbf{B}^{-1}\right)^{-1} < \mathbf{tr}\mathbf{A}. \quad (5)$$

Note that Lemma 3.1 does not apply to matrices $\mathbf{A} \in \mathbb{S}_+^n \setminus \mathbb{S}_{++}^n$. The first useful implication of Lemma 3.1 is that, since the trace is the sum of the eigenvalues, the trace of the largest instantaneous covariance bounds the maximum eigenvalue of the reachable covariance matrices. Define this bound as

$$\beta_i \triangleq \max\nolimits_{\mathbf{p}_i \in \mathcal{W}_i} \ \mathbf{tr}\mathbf{Q}_{i,\mathbf{p}}. \quad (6)$$

Lemma 3.1 also implies that the trace is decreasing with additional independent measurements. Therefore, the set of reachable covariances is contained in $\left\{\mathbf{\Sigma}_i \in \mathbb{S}_{++}^2 \mid \boldsymbol{\lambda}_1(\mathbf{\Sigma}_i) \le \beta_i\right\}$.

Lemma 3.1 and the resulting bound in (6) can be used to obtain a discretization of $\mathbb{S}_{++}^2$. In particular, we define the logspace set

$$\mathcal{L}_i \triangleq \{\beta_i e^{\kappa_{\mathcal{L}_i}(j - N_{\mathcal{L}_i})/N_{\mathcal{L}_i}}\}_{j=1}^{N_{\mathcal{L}_i}} \subset (0, \beta_i], \quad (7)$$

where $N_{\mathcal{L}_i}$ is the number of maximum eigenvalue samples and $\kappa_{\mathcal{L}_i}$ is a sampling gain that controls how clustered the samples are toward zero. The set $\mathcal{L}_i$ represents the set of maximal eigenvalues that we allow to occur during the mission.

The finite sampling of the smaller eigenvalue is not as straightforward as (7). In particular, we need to satisfy the obvious constraint that $\boldsymbol{\lambda}_1(\mathbf{\Sigma}_i) \ge \boldsymbol{\lambda}_2(\mathbf{\Sigma}_i)$ for every triplet $(\boldsymbol{\lambda}_1(\mathbf{\Sigma}_i), \boldsymbol{\lambda}_2(\mathbf{\Sigma}_i), \boldsymbol{\theta}(\mathbf{\Sigma}_i))$. We address this by sampling the eigenvalue ratio $\boldsymbol{\alpha}(\mathbf{\Sigma}_i) \triangleq \frac{\boldsymbol{\lambda}_2(\mathbf{\Sigma}_i)}{\boldsymbol{\lambda}_1(\mathbf{\Sigma}_i)}$. This ratio is related to the ratio between lengths of the minor and major axes of the confidence ellipse defined by the error covariance matrix $\mathbf{\Sigma}_i$. Depending on the sensor model, we can tune the sampling to account for "skinnier" ellipses by using the logspace distribution with a variable gain parameter. Define

$$\mathcal{A}_i \triangleq \{e^{\kappa_{\mathcal{A}_i}(j - N_{\mathcal{A}_i})/N_{\mathcal{A}_i}}\}_{j=1}^{N_{\mathcal{A}_i}} \subset (0, 1]. \quad (8)$$

In (8), $N_{\mathcal{A}_i}$ is the number of eigenvalue ratio samples and $\kappa_{\mathcal{A}_i}$ is a sampling gain that controls the "skinniness" of the confidence ellipse. Note that, given any pair $(\boldsymbol{\lambda}_1(\mathbf{\Sigma}_i), \boldsymbol{\alpha}(\mathbf{\Sigma}_i)) \in \mathcal{L}_i \times \mathcal{A}_i$, we immediately have $\boldsymbol{\lambda}_2(\mathbf{\Sigma}_i) = \boldsymbol{\alpha}(\mathbf{\Sigma}_i)\boldsymbol{\lambda}_1(\mathbf{\Sigma}_i)$.

We discretize the angular dimension of $\mathcal{C}_i$ linearly,

$$\mathcal{T}_i \triangleq \{\pi(j-1)/N_{\mathcal{T}_i}\}_{j=1}^{N_{\mathcal{T}_i}} \subset [0, \pi), \quad (9)$$

where $N_{\mathcal{T}_i}$ is the number of angles in the discretization. Let

$$\mathcal{C}_i \triangleq \left\{\mathbf{\Sigma}_i \in \mathbb{S}_{++}^2 \,\middle|\, \boldsymbol{\lambda}_1(\mathbf{\Sigma}_i) \in \mathcal{L}_i, \boldsymbol{\alpha}(\mathbf{\Sigma}_i) \in \mathcal{A}_i, \boldsymbol{\theta}(\mathbf{\Sigma}_i) \in \mathcal{T}_i\right\} \cup \{\mathbf{0}\}$$

be the covariance part of the state-space $\mathcal{S}_i$, where $\mathbf{0}$ is the 2×2 matrix of all zeros. The $\mathbf{0}$ covariance is an artificial state that we include in the state-space $\mathcal{C}_i$ to denote that no more uncertainty remains in the variable being estimated, i.e., localization is complete to the user-specified tolerance, defined by the minimal element of the set $\mathcal{L}_i$.

As new information is acquired by the sensor, the target error covariance $\mathbf{\Sigma}_i$ is updated according to (2b). The projection operator $\Pi_{\mathcal{C}_i}$ guarantees that the fusion of the current covariance state $\mathbf{\Sigma}_i$ and the new measurement covariance $\mathbf{Q}_{i,\boldsymbol{\phi}(\mathbf{p}_i,\mathbf{u})}$ is a member of $\mathcal{C}_i$, namely that $\mathcal{S}_i$ is closed under actions in $\mathcal{U}_i$. In particular, we define the projection

$\Pi_{\mathcal{C}_i} : \mathbb{S}_+ \to \mathcal{C}_i$ as

$$\Pi_{\mathcal{C}_i}\left[\mathbf{\Sigma}_i^{-1} + \mathbf{Q}_{i,\boldsymbol{\phi}(\mathbf{p}_i,\mathbf{u})}^{-1}\right]^{-1} \triangleq \mathbf{R}_\theta \mathrm{diag}[\lambda_1, \alpha\lambda_1]\mathbf{R}_\theta^\top, \quad (10)$$

where $\lambda_1, \alpha,$ and θ are given by

$$\lambda_1 = \operatorname*{argmin}_{\lambda' \in \mathcal{L}_i \cup \{0\}} \left| \lambda' - \boldsymbol{\lambda}_1\left(\left[\mathbf{\Sigma}_i^{-1} + \mathbf{Q}_{i,\boldsymbol{\phi}(\mathbf{p}_i,\mathbf{u})}^{-1}\right]^{-1}\right)\right|,$$

$$\alpha = \operatorname*{argmin}_{\alpha' \in \mathcal{A}_i} \left| \alpha' - \boldsymbol{\alpha}\left(\left[\mathbf{\Sigma}_i^{-1} + \mathbf{Q}_{i,\boldsymbol{\phi}(\mathbf{p}_i,\mathbf{u})}^{-1}\right]^{-1}\right)\right|,$$

$$\theta = \operatorname*{argmin}_{\theta' \in \mathcal{T}_i \cup \{\pi\}} \left| \theta' - \boldsymbol{\theta}\left(\left[\mathbf{\Sigma}_i^{-1} + \mathbf{Q}_{i,\boldsymbol{\phi}(\mathbf{p}_i,\mathbf{u})}^{-1}\right]^{-1}\right)\right|.$$

In (10), $\mathbf{R}_\theta$ is the rotation matrix with parameter θ. If the value for θ found by solving (10) is π, we set $\theta = 0$. Additionally, if $\lambda_1 = 0$, then we do not compute α or θ, and instead the projected covariance is assigned $\mathbf{0}$.

IV. The Multi-target Case

In this section, we use the solutions of the local DPs discussed in Section III to develop a global Dynamic Program that determines for how long the robot should follow the local optimal policies for every target before transitioning to a new target. Essentially, the goal of the global DP is to balance between the time spent at every target and the total distance traveled in order to best localize all targets. In order to formulate the proposed global DP we first need to define some relevant notation. Let $\mathbf{v} \in \{0,1\}^M$ denote a vector containing the visitation history, so that $\mathbf{v}_i = 1$ if target i has been visited by the robot and $\mathbf{v}_i = 0$ otherwise. Moreover, define an indicator function $\mathbf{i} \in \{1, \ldots, M\}$ that returns the index of the target that is currently being sensed by the robot, an indicator function $\mathbf{j} \in \mathcal{E}_i$ that returns the entry state from where the robot begins sensing target i, where $\mathcal{E}_i = \partial\mathcal{W}_i \times \{\mathbf{\Sigma}_{i,0}\} \subset \mathcal{S}_i$ is the set of boundary states of the local configuration space $\mathcal{W}_i \times \mathcal{C}_i$ of target i, and the number $\mathbf{k}$ of steps that the robot follows the local optimal policy for target $\mathbf{i}$. Together, this information comprises the global state $\mathbf{s} \triangleq (\mathbf{v}, \mathbf{i}, \mathbf{j}, \mathbf{k}) \in \mathcal{S}$, where $\mathcal{S}$ is the state-space of the global DP. Then our goal is to find the optimal policy $\boldsymbol{\mu}$ that solves the following optimization problem:

$$\max\nolimits_{\boldsymbol{\mu}} \sum\nolimits_{k=0}^{\infty} \gamma^k R(\mathbf{s}, \boldsymbol{\mu}(\mathbf{s})), \quad (11)$$

where R denotes the global reward function that, along with the state-space $\mathcal{S}$ and the global transition function $\mathbf{\Phi}$, is defined in Section IV-A. As in Section III, value iteration can be used to determine the global optimal value function $V^* \colon \mathcal{S} \to \mathbb{R}$, which provides the global optimal policy $\boldsymbol{\mu}^* \colon \mathcal{S} \to \mathcal{U}$ for (11) as a byproduct.

A. The Global State-Space, Transition Function, and Reward Function

In this section, we first associate transitions in the global DP with transitions in the local state-spaces of every target. Then, we define the necessary global state-space, transition function, and reward function. Consider a global state $\mathbf{s} = (\mathbf{v}, \mathbf{i}, \mathbf{j}, \mathbf{k})$, as defined previously, and define by $\{\mathbf{s}_i^j\}_{j=1}^{|\mathcal{E}_i|}$ the set of all entry states of target i that are contained in $\mathcal{E}_i$. Here, the superscript j is an index used to uniquely identify the entry states using an integer between 1 and $|\mathcal{E}_i|$. Define also the k-times *recursive local optimal transition function* for target i as $\mathbf{\Phi}_i^{*k} \colon \mathcal{E}_i \to \mathcal{S}_i$, given by

$$\begin{aligned}
\mathbf{\Phi}_i^{*0}(\mathbf{s}_i^j) &= \mathbf{s}_i^j \\
\mathbf{\Phi}_i^{*1}(\mathbf{s}_i^j) &= \mathbf{\Phi}_i\left(\mathbf{s}_i^j, \boldsymbol{\mu}_i^*(\mathbf{s}_i^j)\right) \\
\mathbf{\Phi}_i^{*2}(\mathbf{s}_i^j) &= \mathbf{\Phi}_i\left(\mathbf{\Phi}_i^{*1}(\mathbf{s}_i^j), \boldsymbol{\mu}_i^*(\mathbf{\Phi}_i^{*1}(\mathbf{s}_i^j))\right) \\
&\vdots \\
\mathbf{\Phi}_i^{*k}(\mathbf{s}_i^j) &= \mathbf{\Phi}_i\left(\mathbf{\Phi}_i^{*k-1}(\mathbf{s}_i^j), \boldsymbol{\mu}_i^*(\mathbf{\Phi}_i^{*k-1}(\mathbf{s}_i^j))\right). \quad (12)
\end{aligned}$$

The function $\mathbf{\Phi}_i^{*k}$ denotes the local state at which the robot will land when starting observing target i at state $\mathbf{s}_i^j$ and after following the local optimal policy $\boldsymbol{\mu}_i^*$ for k time steps. Therefore, given a global state $\mathbf{s}$ containing the current target $\mathbf{i}$, the index $\mathbf{j}$ of the entry state to this target, and the number of steps $\mathbf{k}$ that the robot has followed the local optimal policy at this target, the function $\mathbf{\Phi}_{\mathbf{i}}^{*\mathbf{k}}(\mathbf{s}_{\mathbf{i}}^{\mathbf{j}}) \in \mathcal{S}_{\mathbf{i}}$ returns the local state $\mathbf{s}_{\mathbf{i}} = (\mathbf{p}_{\mathbf{i}}, \mathbf{\Sigma}_{\mathbf{i}})$ containing the current robot position and current uncertainty of target $\mathbf{i}$. Using this notation we can associate transitions in the global DP with transitions in the local state-space of target $\mathbf{i}$. Specifically, let $\mathbf{u}_{\mathbf{i}}$ denote the action that directs the robot to continue observing the current target $\mathbf{i}$ according to the local optimal policy $\boldsymbol{\mu}_{\mathbf{i}}^*$. Then, under action $\mathbf{u}_{\mathbf{i}}$ the global state $\mathbf{k}$ transitions to $\mathbf{k}+1$, and as result the local state $\mathbf{\Phi}_{\mathbf{i}}^{*\mathbf{k}}(\mathbf{s}_{\mathbf{i}}^{\mathbf{j}})$ transitions to $\mathbf{\Phi}_{\mathbf{i}}^{*\mathbf{k}+1}(\mathbf{s}_{\mathbf{i}}^{\mathbf{j}})$. Note that, since the optimal policy is stationary, there is always an optimal action to take, implying that local optimal trajectories are infinitely long, i.e., it is possible to extend $\mathbf{k} \to \infty$. The following proposition truncates any such sequence by showing that any optimal trajectory reaches an absorbing state. The proof is omitted due to space limitations.

Proposition 4.1: $\forall i \in \{1, \ldots, M\}$ and $\mathbf{s}_i^j \in \mathcal{E}_i$, there exists $K_i \in \mathbb{N}$ such that $\mathbf{\Phi}_i^{*K_i}(\mathbf{s}_i^j) = \mathbf{\Phi}_i^{*k}(\mathbf{s}_i^j)\ \forall k \geq K_i$.

As a result of Proposition 4.1, we do not store local optimal trajectories longer than $\max_{i \in \{1,\ldots,M\}} K_i$. We can now define the global state-space as the product of finite sets

$$\mathcal{S} \triangleq \{0,1\}^M \times \{1, \ldots, M\} \times \left\{1, \ldots, \max_{i\in\{1,\ldots,M\}} |\mathcal{E}_i|\right\} \times \left\{0, \ldots, \max_{i\in\{1,\ldots,M\}} K_i\right\}. \quad (13)$$

We define the set of admissible controls in the global DP as $\mathcal{U} \triangleq \{\mathbf{u}_0, \ldots, \mathbf{u}_M\}$. We denote the global transition function that captures the set of all admissible global transitions by $\mathbf{\Phi} \colon \mathcal{S} \times \mathcal{U} \to \mathcal{S}$. When in state $\mathbf{s} = (\mathbf{v}, \mathbf{i}, \mathbf{j}, \mathbf{k})$, we have already discussed the result of taking action $\mathbf{u}_{\mathbf{i}}$, which we now define formally as

$$\mathbf{\Phi}(\mathbf{s}, \mathbf{u}_{\mathbf{i}}) = \begin{cases} (\mathbf{v}, \mathbf{i}, \mathbf{j}, \mathbf{k}+1) & \text{if } \mathbf{k} < K_{\mathbf{i}} \\ (\mathbf{v}, \mathbf{i}, \mathbf{j}, \mathbf{k}) & \text{if } \mathbf{k} = K_{\mathbf{i}} \end{cases}, \quad (14)$$

The action $\mathbf{u}_0$ is the *null action*, i.e. $\mathbf{\Phi}(\mathbf{s}, \mathbf{u}_0) = \mathbf{s}$. For $i \neq \mathbf{i}$ and $i \neq 0$, the action $\mathbf{u}_i$ transitions the robot to the local state-space of the i-th target via the closest entry state in $\mathcal{E}_i$ in the i-th configuration space $\mathcal{W}_i \times \mathcal{C}_i$. To determine this new local state, note that $\mathbf{s_i} = \mathbf{\Phi}_\mathbf{i}^{*\mathbf{k}}(\mathbf{s}_\mathbf{i}^\mathbf{j})$ denotes the current local state corresponding to the global state $\mathbf{s} = (\mathbf{v}, \mathbf{i}, \mathbf{j}, \mathbf{k})$ and, therefore, $\mathbf{p}_{\mathbf{s_i}} \triangleq \mathbf{p}_{\mathbf{\Phi}_\mathbf{i}^{*\mathbf{k}}(\mathbf{s}_\mathbf{i}^\mathbf{j})} \in \mathcal{W}_\mathbf{i}$ denotes the current robot position associated with the global state $\mathbf{s}$. Then, we can define the index of the new local state resulting from taking action $\mathbf{u}_i$ by

$$j_i^*(\mathbf{p}_{\mathbf{s_i}}) \triangleq \underset{j \in \{1,\ldots,|\mathcal{E}_i|\}}{\operatorname{argmin}} \left\| \mathbf{p}_{\mathbf{s}_i^j} - \mathbf{p}_\mathbf{s} \right\| \tag{15}$$

The global transition function in this case is

$$\mathbf{\Phi}(\mathbf{s}, \mathbf{u}_i) = (\mathbf{J_i v},\ i,\ j_i^*(\mathbf{p_s}),\ 0)\,, \tag{16}$$

where $\mathbf{J_i}$ is the $M \times M$ identity matrix with the $\mathbf{i}$-th diagonal entry equal to zero. Recalling that $\mathbf{v}$ is a binary vector, the matrix $\mathbf{J_i}$ operates on $\mathbf{v}$ by turning the $\mathbf{i}$-th entry in $\mathbf{v}$ from 1 to 0, i.e., flagging target $\mathbf{i}$ as 'observed,' which prevents it from offering any further positive reward in the way of uncertainty reduction. The combination of (14) for the transition when $\mathbf{u} = \mathbf{u_i}$, the null transition for $\mathbf{u} = \mathbf{u}_0$, and (16) for $\mathbf{u} = \mathbf{u}_{i \neq \mathbf{i}}$, comprise the global transition function $\mathbf{\Phi}: \mathcal{S} \times \mathcal{U} \to \mathcal{S}$.

For a global state $\mathbf{s} = (\mathbf{v}, \mathbf{i}, \mathbf{j}, \mathbf{k})$, let the reward for following the local optimal policy be

$$R(\mathbf{s}, \mathbf{u_i}) = \begin{cases} (1-\rho)\left(\mathbf{tr}\left[\mathbf{\Sigma}_{\mathbf{s}_i} - \mathbf{\Sigma}_{\mathbf{s}_i{'}}\right]\right)^{1/2} & \text{if } \mathbf{e}_\mathbf{i}^\top \mathbf{v} = 1 \\ \qquad -\rho \psi(\mathbf{u}^*) & \\ -\rho \left\|\mathbf{p}_{\mathbf{s}_i} - \mathbf{p}_{\mathbf{s}_i'}\right\| & \text{if } \mathbf{e}_\mathbf{i}^\top \mathbf{v} = 0 \end{cases}, \tag{17}$$

where

$$\mathbf{s}_i = \mathbf{\Phi}_\mathbf{i}^{*\mathbf{k}}(\mathbf{s}_\mathbf{i}^\mathbf{j}), \qquad \mathbf{u}^* = \boldsymbol{\mu}_\mathbf{i}^*\left(\mathbf{\Phi}_\mathbf{i}^{*\mathbf{k}}(\mathbf{s}_\mathbf{i}^\mathbf{j})\right), \text{ and}$$

$$\mathbf{s}_i' = \begin{cases} \mathbf{\Phi}_\mathbf{i}^{*\mathbf{k}+1}(\mathbf{s}_\mathbf{i}^\mathbf{j}) & \text{if } k < K_\mathbf{i} \\ \mathbf{\Phi}_\mathbf{i}^{*\mathbf{k}}(\mathbf{s}_\mathbf{i}^\mathbf{j}) & \text{if } k = K_\mathbf{i} \end{cases},$$

where $\mathbf{e_i}$ is the vector or all zeros except for the i-th entry equal to 1. The purpose of introducing the unit vector $\mathbf{e}_i$ is so that if $\mathbf{e}_i^\top \mathbf{v} = 1$, then the robot has not deemed localization for the $\mathbf{i}$-th target complete. If $\mathbf{e}_\mathbf{i}^\top \mathbf{v} = 0$, then no positive reward can be gained by visiting target $\mathbf{i}$.

The remaining actions $\mathbf{u}_i$ for $i \neq \mathbf{i}$ have the nonpositive rewards given by

$$R(\mathbf{s}, \mathbf{u}_{i \neq \mathbf{i}}) = \begin{cases} 0 & \text{if } i = 0 \\ -\rho \left\|\mathbf{p_s} - \mathbf{p}_{\mathbf{s}'}\right\| & \text{else} \end{cases}, \tag{18}$$

where $\mathbf{s}' = \mathbf{\Phi}(\mathbf{s}, \mathbf{u}_i)$. The combination of (17) and (18) comprise the reward function for the global DP.

V. Numerical Simulations

In this section, we present simulations of the proposed hierarchical control scheme. In our simulations, we use two cameras in a rigid fronto-parallel configuration, i.e., stereo vision, as the exteroceptive sensor model. Stereo vision, due to its light weight, low cost, low power consumption, and analytical covariance model, has shown great promise in active sensing. We omit a direct derivation of the covariance function $\mathbf{Q}(\hat{\mathbf{x}}_i, \mathbf{p})$ for stereo vision, and instead we refer the reader to our prior work on this subject [13], [14]. The

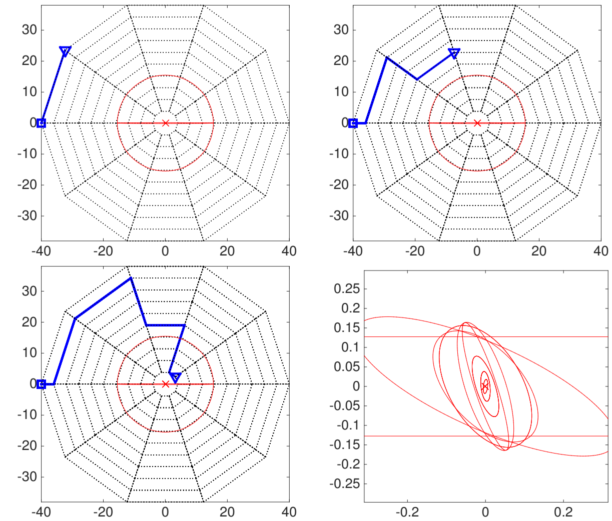

Fig. 1: Various values for ρ and the resulting optimal trajectories in the single target simulations. Top-left: $\rho = 5 \times 10^{-2}$. Top-right: $\rho = 5 \times 10^{-3}$. Bottom: $\rho = 5 \times 10^{-4}$ and the 95% confidence ellipses. The robot starts at the square and ends at the triangle. All units in meters.

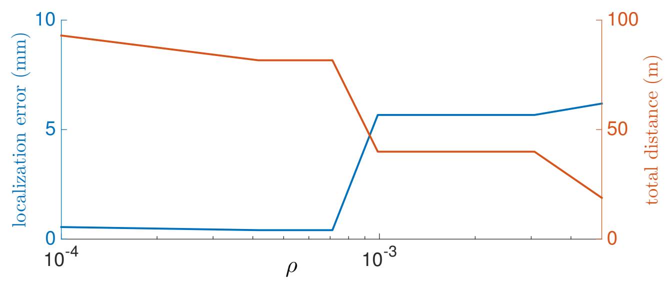

Fig. 2: Plotting terminal localization error and total distance traveled by the sensor versus ρ in the single target simulations. Note the different length units on the vertical axes. Lines are drawn to guide the eye.

stereo rig model we use employs two 1024×1024 resolution cameras with identical $70°$ fields of view. The baseline of the rig is 5 cm. Our simulations also focus on the case where $\mathcal{W}_i = \mathcal{W}_j$ and $\mathbf{\Sigma}_{i,0} = \mathbf{\Sigma}_{j,0}$ for all $i, j \in \{1, \ldots, M\}$.

We determined the optimal policy for the single-target case using a polar grid as the pose space and stereo vision as the sensing model. The polar grid $\mathcal{W}$ was centered at the initial guess of the estimated target location $\hat{\mathbf{x}}_0$ and was 80 m in diameter, so that views in $\partial\mathcal{W}$ had disparity 1, i.e., they were at the limit of the sensing range. There were 100 total views in $\mathcal{W}$ at ten equally spaced radii and angles. We note that in practice, other discretizations, e.g. [23], may be preferable, and our method extends to them as well. To discretize the covariance space, we set $N_\mathcal{L} = 10, N_\mathcal{A} = 10, N_\mathcal{T} = 15, \kappa_\mathcal{L} = 15$, and $\kappa_\mathcal{A} = 12$. The discount factor γ was 0.9.

The value of the uncertainty reduction gain ρ controls the tradeoff between traveling cost and uncertainty reduction due to more images. Fig. 1 shows example optimal trajectories using three different values for ρ. The general trend is as expected: the smaller ρ is, the further the robot will travel to reduce localization uncertainty. Fig. 2 shows the sensitivity of the terminal localization performance as it depends on ρ.

We also present several simulations of the proposed hier-

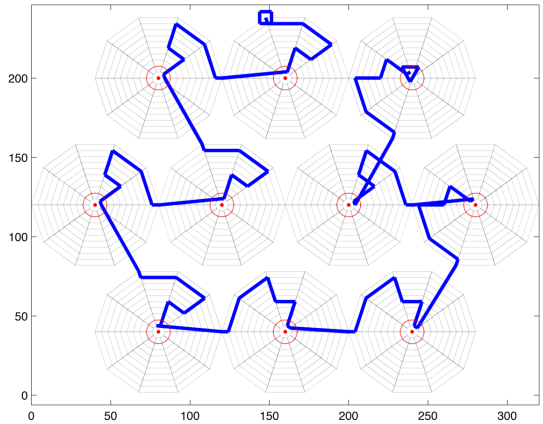

Fig. 3: Path planning and resource allocation for 10 targets on a hexagonal pattern. All units are in meters.

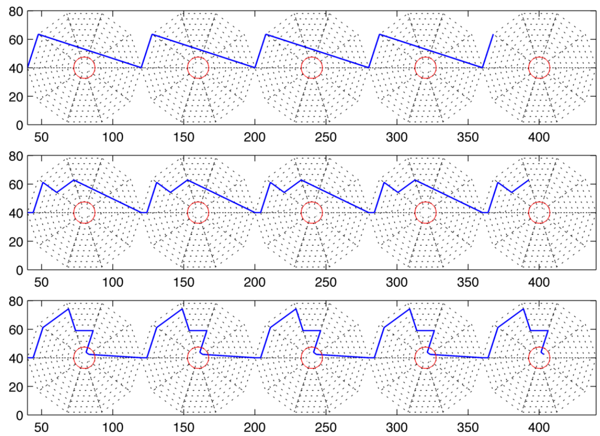

Fig. 4: The effect of varying ρ in the global DP. From top to bottom, the values of ρ are 5×10^{-2}, 5×10^{-3}, and 5×10^{-4}. All units are in meters.

archical scheme for multi-target active localization. In Fig. 3, we run our hierarchical control scheme for ten targets, which results in about 1.3 million states. We use the empirically determined value of $\rho = 5 \times 10^{-4}$ in these simulations. Since the time horizon is longer in these simulations, we used a higher discount factor of $\gamma = 0.9999$ to generate these trajectories. Exact value iteration converges for this case in about 5 minutes using Matlab™ on a Macbook Air™ with a 1.7 GHz Intel Core i7 processor. In Fig 4, we show the effect of varying ρ in the global DP for multiple targets.

VI. Conclusion

In this paper, we addressed the optimal path planning and resource allocation problem for active multi-target localization. We framed the problem using tools from optimal control, ultimately proposing a hierarchical Dynamic Programming solution. The hierarchical approach solves a local DP for each target that simultaneously minimizes target localization uncertainty and distance traveled. We determined the reduction in variance by using a sensor model and a Kalman filter to fuse observations. We achieved lower computational complexity compared to more general POMDP methods by including the error covariance matrix of a target's location in the local state-spaces. Then, we combined the local optimal trajectories in a global DP that balances between reducing the uncertainty in the targets and traveling among configuration spaces. Simulations show automatically generated trajectories that address uncertainty reduction in terms of an information-theoretic objective function while remaining computationally tractable, a novel pair of characteristics that compare favorably with POMDP and TSP approaches.

References

[1] A. Logothetis *et al.*, "An information theoretic approach to observer path design for bearings-only tracking," in *IEEE Conf. on Decision and Control (CDC)*, vol. 4, pp. 3132–3137, 1997.

[2] S. Singh *et al.*, "Simulation-based optimal sensor scheduling with application to observer trajectory planning," *Automatica*, vol. 43, no. 5, pp. 817–830, 2007.

[3] A. Ryan and J. K. Hedrick, "Particle filter based information-theoretic active sensing," *Robotics and Autonomous Systems*, vol. 58, no. 5, pp. 574–584, 2010.

[4] S. Singh *et al.*, "Stochastic approximation for optimal observer trajectory planning," in *IEEE Conf. on Decision and Control (CDC)*, vol. 6, pp. 6313–6318, 2003.

[5] M. T. Spaan, "A point-based POMDP algorithm for robot planning," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, vol. 3, pp. 2399–2404, IEEE, 2004.

[6] P. Skoglar *et al.*, "On information measures based on particle mixture for optimal bearings-only tracking," in *Aerospace conference*, pp. 1–14, IEEE, 2009.

[7] L. Meier *et al.*, "Optimal control of measurement subsystems," *IEEE Transactions on Automatic Control*, vol. 12, no. 5, pp. 528–536, 1967.

[8] J. Le Ny and G. J. Pappas, "On trajectory optimization for active sensing in gaussian process models," in *IEEE Conf. on Decision and Control (CDC)*, pp. 6286–6292, IEEE, 2009.

[9] I. Hussein, "Kalman filtering with optimal sensor motion planning," in *IEEE American Control Conference (ACC)*, pp. 3548–3553, 2008.

[10] S. Wenhardt *et al.*, "An information theoretic approach for next best view planning in 3-d reconstruction," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 103 –106, 2007.

[11] S. Ponda and E. Frazzoli, "Trajectory optimization for target localization using small unmanned aerial vehicles," in *AIAA Conf. on Guidance, Navigation, and Control*, 2009.

[12] E. Dunn *et al.*, "Developing visual sensing strategies through next best view planning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4001 –4008, oct. 2009.

[13] C. Freundlich *et al.*, "A hybrid control approach to the next-best-view problem using stereo vision," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 4478 –4483, 2013.

[14] C. Freundlich *et al.*, "Hybrid control for mobile target localization with stereo vision," in *IEEE Conf. on Decision and Control (CDC)*, pp. 2635– 2640, 2013.

[15] N. Atanasov *et al.*, "Hypothesis testing framework for active object detection," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 4216–4222, 2013.

[16] N. Adurthi *et al.*, "Optimal information collection for nonlinear systems- an application to multiple target tracking and localization," in *IEEE American Control Conference (ACC)*, pp. 3864–3869, 2013.

[17] J. Le Ny *et al.*, "Resource allocation for signal detection with active sensors," in *IEEE Conf. on Decision and Control (CDC)*, pp. 8561–8566, IEEE, 2009.

[18] G. A. Hollinger *et al.*, "Active planning for underwater inspection and the benefit of adaptivity," *The International Journal of Robotics Research*, vol. 32, no. 1, pp. 3–18, 2013.

[19] E. Galceran and M. Carreras, "A survey on coverage path planning for robotics," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1258–1276, 2013.

[20] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Athena Scientific Belmont, MA, 1st ed., 1996.

[21] D. Hardin and E. Saff, "Discretizing manifolds via minimum energy points," *Notices of the AMS*, vol. 51, no. 10, pp. 1186–1194, 2004.

[22] E. Carlen, "Trace inequalities and quantum entropy: an introductory course," *Entropy and the quantum*, vol. 529, pp. 73–140, 2010.

[23] B. Englot and F. Hover, "Planning complex inspection tasks using redundant roadmaps," in *Proc. Int. Symp. Robotics Research*, 2011.