Multi-Robot Active SLAM with Relative Entropy Optimization

Michail Kontitsis[1], Evangelos A. Theodorou[2] and Emanuel Todorov[3]

Abstract— **This paper presents a new approach for Active Simultaneous Localization and Mapping that uses the Relative Entropy(RE) optimization method to select trajectories which minimize both the localization error and the corresponding uncertainty bounds. To that end we construct a planning cost function which includes, besides the state and control cost, a term that encapsulates the uncertainty of the state. This term is the trace of the state covariance matrix produced by the estimator, in this case an Extended Kalman Filter. The role of the RE method is to iteratively guide the selection of the trajectories towards the ones minimizing the aforementioned cost. Once the method has converged, the result is a near-optimal path in terms of achieving the pre-defined goal in the state space while also improving the localization error and the total uncertainty. In essence the method integrates motion planning with robot localization. To evaluate the approach we consider scenarios with single and multiple robots navigating in presence of obstacles and various conditions of landmark densities. The results show a behavior consistent with our expectations.**

I. INTRODUCTION

Over the last 10 years there has been a significant amount of work in the area of **S**imultaneous **L**ocalization **A**nd **M**apping or otherwise SLAM with a plethora of applications which include single robot, multi-robot exploration scenarios. Research in this area relates to testing, comparison and evaluation of different nonlinear estimation methods such as Particle, Unscented and Extended Kalman Filters for the purpose of fusion information gathered by the sensors onboard. The map precision, the computational complexity and consistency of the underlying state estimators and the existence of provable upper bounds on the map uncertainty and robot localization error are only just few of the desirable specifications in robot localization and exploration scenarios.

In a typical localization scenario, the goal for a robot is to localize its position while at the same time reducing the uncertainty of the map of the environment under exploration. The central idea is that the detection of a landmark that has been previously seen, causes map uncertainty reduction. Stochastic estimation is separated from control and planning and therefore estimation is treated without investigation of how feedforward policies for planning or feedback policies for control affect the robot localization error. The aforementioned epistemological approach has its origin in the separation principle of stochastic optimal control [2], [3], [8] according to which, the control and estimation problems are dual and they can be solved completely separately. In fact the estimation Ricatti and control Ricatti equations are separated in the sense that control policy is not a function of the Kalman gain and vice versa. This is however true only for linear systems with additive process and observation noise.

In this work we integrate planning and stochastic optimization with robot localization in order to perform planning under uncertainly. The approach is known as Active SLAM and it has been studied elsewhere [5]–[7]. In particular, in [6] and [5] model predictive control was used for planning trajectories which improve SLAM performance. Simulations and experiments were performed for a single robot exploration scenario. In [7] active SLAM was performed for the case of single robot while the metric used for planning was the information gain.

Here we study the case where the robot reaches a desired target while simultaneously minimizing its localization error and map uncertainty. However, in contrast to previous work we consider the problem of multi-robot active SLAM which takes into account optimal trajectory planning and uncertainty reduction for a team of robots. We assume that localization processing occurs in a central station or robot and thus the only task that the robots have to perform is to gather information and send it to the central station for processing. Robots besides measuring their relative position with respect to Landmarks, can also detect each other. Consequently each robot can measure its relative position with respect to other robots in the team. We demonstrate multi robot optimal coordination and planning which results in trajectories that do not only reach pre-specified targets but also reduce the robots' localization error and map uncertainty while avoiding obstacles. To do so we make use of stochastic optimization for continuous state actions spaces based on Relative Entropy (RE) minimization a.k.a. Cross Entropy (CE) [1]. The RE optimization method has been used for Unmanned Aerial Vehicle (UAV) path planning and obstacle avoidance in [4]. Here the RE method is applied to Active SLAM for the cases of single-robot as well as homogeneous multi-robot exploration and planning scenarios.

The paper is organized as follows: in Section II we provide a brief description of Optimal Control followed by Section III where we present the Cross Entropy method. In Section IV we provide the EKF equations. In Section V we present the state space as well as observation models for the single robot and multirobot cases. Section VI contains all of our simulation experiments. We conclude this paper in Section VII with a discussion and future extensions of this work.

[1] Michail Kontitsis is a Lecturer with the Department of Electrical Engineering, University of Denver, Colorado. mkontits@du.edu

[2] Evangelos A. Theodorou is a Postdoctoral Research Associate with the Department of Computer Science and Engineering, University of Washington, Seattle. etheodor@cs.washington.edu

[3] Emanuel Todorov is an Associate Professor with the Department of Computer Science and Engineering and the Applied Math Department, University of Washington, Seattle. todorov@cs.washington.edu

II. Optimal Control

The optimal control is defined as a constrained optimization problem with the cost function:

$$\min_{\mathbf{u}} \mathbb{E}_p[\boldsymbol{\mathcal{L}}(\mathbf{x})] = \min_{\mathbf{u}} \mathbb{E}_p \left(\sum_{t=t_0}^{t_N} l\left(\mathbf{x}(t), \mathbf{u}(t)\right) dt \right) \tag{1}$$

subject to dynamics:

$$\mathbf{x}(t_{k+1}) = \mathbf{f}\Bigg(\mathbf{x}(t_k), \mathbf{u}(t_k)\Bigg) + \boldsymbol{\mathcal{C}}(\mathbf{x}(t_k))\mathbf{w}(t_k) \tag{2}$$

with $\mathbf{x} \in \Re^n$ is the state, $\mathbf{u} \in \Re^p$ the control parameters and $\mathbf{w}(t) \in \Re^l$ is the process zero mean Gaussian noise with covariance $\boldsymbol{\Sigma}_{\mathbf{w}}$. The functions $\mathbf{f}$, $\boldsymbol{\mathcal{C}}$ are defined as $\mathbf{f}(\mathbf{x}, \mathbf{u}) : \Re^p \times \Re^n \rightarrow \Re^n$, $\boldsymbol{\mathcal{C}}(\mathbf{x}) : \Re^n \rightarrow \Re^n \times \Re^l$. In this work we parameterize the time varying controls $\mathbf{u}(t) \in \Re^p \times \Re^{t_N - t_0}$ as a function a parameter $\boldsymbol{\theta} \in \Re^m$ and therefore we will have that $\mathbf{u}(t) = \mathbf{u}(t; \boldsymbol{\theta})$. Instead therefore of minimizing the cost function in (1) with respect to $\mathbf{u}(t)$ we will minimize it with respect to the lower dimensional vector $\boldsymbol{\theta}$.

III. Relative Entropy Optimization

In this section we present stochastic optimization based on relative Entropy minimization and explain how this method is used for stochastic optimal control. In particular, we consider the estimation [1] of the cost function in (1) written as:

$$\mathcal{J}(\mathbf{x}) = \mathbb{E}_{p(\mathbf{x})}[\boldsymbol{\mathcal{L}}(\mathbf{x})] = \int p(\mathbf{x})\boldsymbol{\mathcal{L}}(\mathbf{x})d\mathbf{x} \tag{3}$$

with $\boldsymbol{\mathcal{L}}(\mathbf{x})$ a non-negative cost function and $p(\mathbf{x})$ the probability density. This is the probability density corresponding to sampling trajectories based on (2). Under the parameterization of the baseline probability density we will have that $p(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\mu})$. We perform importance sampling from a proposal probability density $q(\mathbf{x})$ and evaluate the expectation as follows:

$$\mathcal{J}(\mathbf{x}) = \int \frac{p(\mathbf{x}; \boldsymbol{\mu})}{q(\mathbf{x})}\boldsymbol{\mathcal{L}}(\mathbf{x})q(\mathbf{x})d\mathbf{x} = \mathbb{E}_q\left[\frac{p(\mathbf{x}; \boldsymbol{\mu})}{q(\mathbf{x})}\boldsymbol{\mathcal{L}}(\mathbf{x})\right] \tag{4}$$

The expression above is numerically evaluated as:

$$\hat{\mathcal{J}}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{p(\mathbf{x}_i; \boldsymbol{\mu})}{q(\mathbf{x}_i)}\boldsymbol{\mathcal{L}}(\mathbf{x}_i)\right] \tag{5}$$

with $\hat{\mathcal{J}}(\mathbf{x})$ being an unbiased estimator. The probability density that minimizes the variance of the estimator $\hat{\mathcal{J}}$ is:

$$q^*(\mathbf{x}) = \underset{q}{\operatorname{argmin}}\, Var\left[\boldsymbol{\mathcal{L}}(\mathbf{x})\frac{p(\mathbf{x}; \boldsymbol{\mu})}{q(\mathbf{x})}\right] \tag{6}$$

The solution to which is $q^*(\mathbf{x}) = \frac{p(\mathbf{x}; \boldsymbol{\mu})\boldsymbol{\mathcal{L}}(\mathbf{x})}{\hat{\mathcal{J}}(\mathbf{x})}$ and it is the optimal importance sampling density. Inside the relative entropy optimization framework [1] the main idea is to find the parameters $\boldsymbol{\theta}$ within the parametric class of pdfs $p(\mathbf{x}, \boldsymbol{\theta})$ such that the probability distribution $p(\mathbf{x}; \boldsymbol{\theta})$ is close to the optimal distribution $q^*(\mathbf{x})$. We use the Kullback-Leibler divergence as a distance metric between $q^*(\mathbf{x})$ and $p(\mathbf{x}; \boldsymbol{\theta})$ and thus we will have .

$$\begin{aligned}\mathbb{D}\left(q^*(\mathbf{x}||p(\mathbf{x}; \boldsymbol{\theta})\right) = & \int q^*(\mathbf{x}) \ln q^*(\mathbf{x})d\mathbf{x} \\ & - \int q^*(\mathbf{x}) \ln p(\mathbf{x}; \boldsymbol{\theta})d\mathbf{x}\end{aligned}$$

The minimization problem is now specified as:

$$\begin{aligned}\boldsymbol{\theta}^* &= \operatorname{argmin} \mathbb{D}\left(q^*(\mathbf{x})||p(\mathbf{x}; \boldsymbol{\theta})\right) \\ &= \operatorname{argmax} \int q^*(\mathbf{x}) \ln p(\mathbf{x}; \boldsymbol{\theta})d\mathbf{x} \\ &= \operatorname{argmax} \int \frac{p(\mathbf{x}; \boldsymbol{\mu})\boldsymbol{\mathcal{L}}(\mathbf{x})}{\hat{\mathcal{J}}(\mathbf{x})} \ln p(\mathbf{x}; \boldsymbol{\theta})d\mathbf{x} \\ &= \operatorname{argmax} \int p(\mathbf{x}; \boldsymbol{\mu})\boldsymbol{\mathcal{L}}(\mathbf{x}) \ln p(\mathbf{x}; \boldsymbol{\theta})d\mathbf{x} \\ &= \operatorname{argmax} \mathbb{E}_{p(\mathbf{x}; \boldsymbol{\mu})}\left[\boldsymbol{\mathcal{L}}(\mathbf{x}) \ln p(\mathbf{x}; \boldsymbol{\theta})\right]\end{aligned}$$

The optimal parameters can be approximated numerically as:

$$\boldsymbol{\theta}^* = \operatorname{argmax} \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\mathcal{L}}(\mathbf{x}_i) \ln p(\mathbf{x}_i; \boldsymbol{\theta}) \tag{7}$$

with sample paths evaluated under the density $p(\mathbf{x}; \boldsymbol{\mu})$.

A. Optimization as Estimation of Rare-Event Probabilities

We consider the problem of estimating the probability that a trajectory $\boldsymbol{\tau}$ sampled from the distribution $p(\mathbf{x}; \boldsymbol{\mu})$ has a cost that is smaller than a constant γ. More precisely we will have:

$$\mathbb{P}\left(\boldsymbol{\mathcal{L}} \leq \gamma\right) = \mathbb{E}_{p(\mathbf{x}; \boldsymbol{\mu})}[I_{\{\boldsymbol{\mathcal{L}} \leq \gamma)\}}]$$

The probability above can be numerically approximated by the equation:

$$\hat{\mathbb{P}}\left(\boldsymbol{\mathcal{L}} \leq \gamma\right) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{p(\mathbf{x}_i; \boldsymbol{\mu})}{p(\mathbf{x}_i; \boldsymbol{\theta})}I_{\{\boldsymbol{\mathcal{L}}(\mathbf{x}_i) \leq \gamma)\}}\right]$$

where $\mathbf{x}_i$ are i.i.d samples from $p(\mathbf{x}; \boldsymbol{\theta})$. The goal here is to find the optimal $\boldsymbol{\theta}^*$ which is defined as:

$$\boldsymbol{\theta}^* = \operatorname{argmax} \frac{1}{N}\sum_{i=1}^{N} I_{\{\boldsymbol{\mathcal{L}}(\mathbf{x}_i) \leq \gamma)\}} \ln p(\mathbf{x}; \boldsymbol{\theta}) \tag{8}$$

where now the samples $\mathbf{x}_i$ are generated according to probability density $p(\mathbf{x}; \boldsymbol{\mu})$. Since the event $\{\boldsymbol{\mathcal{L}} \leq \gamma\}$ is rare its probability is difficult to be estimated. Instead of keeping the γ fixed, an alternative approach is to start with a γ_1 for which the probability of the event $\{\boldsymbol{\mathcal{L}} \leq \gamma_1\}$ is equal to $\rho > 0$. Thus the value γ_1 is set to the ρ -th quantile of $\boldsymbol{\mathcal{L}}(\mathbf{x})$ which means that γ_1 is the largest number for which

$$\mathbb{P}\left(\boldsymbol{\mathcal{L}}(\mathbf{x}) \leq \gamma_1\right) = \rho. \tag{9}$$

The parameter γ_1 can be found by sorting the samples according to their cost in a increasing order and setting $\gamma_1 = \boldsymbol{\mathcal{L}}_{\lceil \rho N \rceil}$. The optimal parameter $\boldsymbol{\theta}_1$ for the level γ_1 is calculated according to (7). The iterative procedure terminates when $\gamma_k \leq \gamma$ in this case the corresponding parameter $\boldsymbol{\theta}_k$ is the optimal one and thus $\boldsymbol{\theta}^* = \boldsymbol{\theta}_k$.

For the case of continuous optimal control, optimization is treated as an *estimation problem of rare event probability*. In particular the cost function optimum is defined as $\gamma^* = \min \boldsymbol{\mathcal{L}}(\mathbf{x})$. To find the optimal trajectory $\mathbf{x}^*$ and optimal parameters $\boldsymbol{\theta}^*$ we iterate the process of estimating rare event probabilities until $\gamma \rightarrow \gamma^*$. Since γ^* is not know *a-priori* we choose as γ^* the value γ for which no further improvement in the iterative process is observed.

IV. EKF based Active SLAM

We consider the stochastic dynamics as expressed as in (2) together with the observation model:

$$\mathbf{y}(t_{k+1}) = \mathbf{h}(\mathbf{x}(t_{k+1})) + \mathbf{v}(t_{k+1}) \tag{10}$$

The parameters $\mathbf{y} \in \Re^q$ correspond to measurements while $\mathbf{v}(t) \in \Re^q$ is the observation noise that is also zero mean with covariance $\boldsymbol{\Sigma}_{\mathbf{v}}$ while the function $\mathbf{h}(\mathbf{x}) : \Re^n \rightarrow \Re^q$.

In this work we make use of the Extended Kalman Filter for state estimation. There are two phases in the estimation process, namely the propagation and the update phase. The equations of the propagation phase of EKF are given as follows:

$$\begin{aligned}
\hat{\mathbf{x}}(t_{k+1|k}) &= \mathbf{f}\Bigg(\hat{\mathbf{x}}(t_{k,k}), \mathbf{u}(\hat{\mathbf{x}}(t_{k,k}), t; \boldsymbol{\theta})\Bigg) \\
\boldsymbol{\Phi} &= \nabla_{\mathbf{x}} \mathbf{f}\Bigg(\hat{\mathbf{x}}(t_{k,k}), \mathbf{u}(\hat{\mathbf{x}}(t_{k,k}), t_k; \boldsymbol{\theta}), 0\Bigg) \\
\boldsymbol{\Sigma}(t_{k+1|k}) &= \boldsymbol{\Phi}\boldsymbol{\Sigma}(t_{k+1|k})\boldsymbol{\Phi}^T + \boldsymbol{\mathcal{C}}\boldsymbol{\Sigma}_{\mathbf{w}}\boldsymbol{\mathcal{C}}^T
\end{aligned} \tag{11}$$

For the update phase we consider the Joseph form of EKF. The update equation are expressed as follows:

$$\begin{aligned}
\boldsymbol{\mathcal{H}}(\mathbf{x}(t)) &= \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}(t)) \\
\hat{\mathbf{y}}(t_{k+1}) &= \boldsymbol{\mathcal{H}}(\hat{\mathbf{x}}(t_{k+1|k})) \\
\mathbf{y}(t_{k+1}) &= \boldsymbol{\mathcal{H}}(\mathbf{x}(t_{k+1})) + \mathbf{v}(t_{k+1}) \\
\mathbf{r}(t_{k+1}) &= \mathbf{y}(t_{k+1}) - \hat{\mathbf{y}}(t_{k+1}) \\
\boldsymbol{\mathcal{S}}(t_{k+1}) &= \boldsymbol{\mathcal{H}}\boldsymbol{\Sigma}(t_{k+1|k})\boldsymbol{\mathcal{H}}^T + \boldsymbol{\Sigma}_{\mathbf{v}} \\
\boldsymbol{\mathcal{L}}(t_{k+1}) &= \boldsymbol{\Sigma}(t_{k+1|k})\boldsymbol{\mathcal{H}}^T \boldsymbol{\mathcal{S}}(t_{k+1})^{-1} \\
\hat{\mathbf{x}}_{t_{k+1|k+1}} &= \hat{\mathbf{x}}_{t_{k+1|k}} + \boldsymbol{\mathcal{L}}(t_{k+1})\mathbf{r}(t_{k+1}) \\
\boldsymbol{\Sigma}(t_{k+1|k+1}) &= (I - \boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{H}})\,\boldsymbol{\Sigma}(t_{k+1|k})\,(I - \boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{H}})^T \\
&\quad + \boldsymbol{\mathcal{L}}\boldsymbol{\Sigma}_{\mathbf{v}}\boldsymbol{\mathcal{L}}^T
\end{aligned} \tag{12}$$

Note that in the equation of the update of the covariance matrix both matrices are symmetric, the first being positive definite and the second positive semidefinite. Due to the aforementioned form of the covariance-update equation, the Joseph form of Kalman Filtering ensures symmetry and positive definiteness of $\boldsymbol{\Sigma}(t_{k+1|k+1})$.

V. State Space Models

In this section we present the state space and observation models for the single-robot and multi-robot case and provide the corresponding changes in the update and propagation equation of Kalman Filter as presented in the previous section.

A. Single Robot Case

In this subsection we describe the state space model used for our simulations. In particular the kinematics of the robot are expressed as:

$$x(t_{k+1}) = x(t_k) + V_m \cos(\phi(t_k))\delta t \tag{13}$$

$$y(t_{k+1}) = y(t_k) + V_m \sin(\phi(t_k))\delta t \tag{14}$$

$$\phi(t_{k+1}) = \phi(t_k) + \omega_m(t_k)\delta t \tag{15}$$

where x, y correspond to the position and ϕ to the orientation of the robot. V_m and ω_m are the linear and rotational velocities as measured. The aforementioned velocities are further expressed as functions of the wheel velocities V_{1m} and V_{2m} measured by the odometers. More precisely we will have: $V_m = \frac{V_{1m}+V_{2m}}{2}$ and $\omega_m = \frac{V_{1m}-V_{2m}}{\alpha}$ with $V_{1m} = V_1 + w_1(t)$ and $V_{2m} = V_2 + w_2(t)$. The quantities V_1 and V_2 are the true velocities. Clearly the measured linear and rotational velocity can be further written as:

$$\begin{aligned}
V_m &= \frac{V_{1m}(t) + V_{2m}(t)}{2} = V(t) + \frac{w_1(t) + w_2(t)}{2} \\
\omega_m &= \frac{V_{1m}(t) - V_{2m}(t)}{\alpha} = \boldsymbol{\omega}(t) + \frac{w_1(t) - w_2(t)}{\alpha}
\end{aligned}$$

We define $\epsilon_1(t_k) = \frac{w_1(t_k)+w_2(t_k)}{2}$ and $\epsilon_2(t_k) = \frac{w_1 - w_2}{\alpha}$ and substitute back to the kinematics of the mobile robot which yields for the state $\mathbf{x} = [x, y, \phi]$:

$$\mathbf{x}(t_{k+1}) = \mathbf{f}(\mathbf{x}(t_k) + \mathbf{g}(\mathbf{x}(t_k))\mathbf{w}(t_k) \tag{16}$$

where $\mathbf{f}(\mathbf{x}_{t_k})$, $\mathbf{g}(\mathbf{x}_{t_k})$ are defined as:

$$\mathbf{f}(\mathbf{x}(t_k)) = \begin{pmatrix} x(t_k) + V\cos(\phi(t_k))\delta t \\ y(t_k) + V\sin(\phi(t_k))\delta t \\ \phi(t_k) + \omega(t_k)\delta t \end{pmatrix} \tag{17}$$

$$\mathbf{g}(\mathbf{x}(t_k)) = \begin{pmatrix} 0.5\cos(\phi(t_k)) & 0.5\cos(\phi(t_k)) \\ 0.5\sin(\phi(t_k)) & 0.5\sin(\phi(t_k)) \\ \frac{1}{\alpha} & -\frac{1}{\alpha} \end{pmatrix} \tag{18}$$

and $\mathbf{w}(t_k) = (w_1(t_k), w_2(t_k))^T$. We assume that landmarks are fixed and therefore: $\frac{d\mathbf{p}_i}{dt} = 0, \quad \forall i = 1, 2, 3..., N$ where N is the number of Landmarks. The complete state space model is expressed as:

$$\begin{pmatrix} \mathbf{x}(t_{k+1}) \\ \mathbf{p}_1(t_{k+1}) \\ ... \\ \mathbf{p}_N(t_{k+1}) \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{x}(t_k)) \\ \mathbf{p}_1(t_k) \\ ... \\ \mathbf{p}_N(t_k) \end{pmatrix} + \begin{pmatrix} \mathbf{g}(\mathbf{x}(t_k)) \\ 0_{2\times 2} \\ ... \\ 0_{2\times 2} \end{pmatrix} \mathbf{w}(t_k) \tag{19}$$

The nonlinear equation above matches the form of the state space model in (2) with the state $\mathbf{x} \in \Re^{N+3}$ defined as $\mathbf{x} = (x, y, \phi, \mathbf{p}_1, ..., \mathbf{p}_N)$ and $\mathbf{w} \in \Re^2$. The observation model is nonlinear and it has the form:

$$\mathbf{y}(t_k) = {}^{R}_{G}\mathbf{R}(\phi(t_k))\Big({}^{G}\mathbf{p}_i(t_k) - {}^{G}\mathbf{p}_R(t_k)\Big) + \mathbf{v}(t_k) \tag{20}$$

with ${}^{G}\mathbf{p}_i(t_k)^T = \big({}^{G}p_{x_i}(t_k),\ \ {}^{G}p_{y_i}(t_k)\big)$ and ${}^{G}\mathbf{p}_R(t_k) = \big({}^{G}x(t_k),\ \ {}^{G}y(t_k)\big)$ being the position of the landmarks and the robot with respect to the global frame of reference $\{G\}$ and $\mathbf{v}$ Gaussian distributed zero mean noise with covariance $\mathbf{\Sigma}_{\mathbf{v}} = \mathbf{diag}(\sigma_{v1}^2,\ \sigma_{v2}^2)$. The matrix ${}^{R}_{G}\mathbf{R}(\phi(t_k))$ is the rotational matrix which expresses rotation from the global $\{G\}$ to local(=robot) frame of reference $\{R\}$.

B. Multirobot Case

For the multi-robot case we consider two robots with linear and rotational velocities V_{R1}, V_{R2} and ω_{R1}, ω_{R2} respectively. The state space model for the two robot case is expressed as follows:

$$\mathbf{X}(t_{k+1}) = \mathbf{F}(\mathbf{X}(t_k)) + \mathbf{G}(\mathbf{X}(t_k))\mathbf{w}(t_k) \tag{21}$$

where the state $\mathbf{X} \in \Re^{6+2\times N}$ in the equation above is defined as $\mathbf{X} = \big(\mathbf{x}_1^T, \mathbf{x}_2^T, \mathbf{p}_1^T, ..., \mathbf{p}_N^T\big)^T = (x_1, y_1, \phi_1, x_2, y_2, \phi_2, p_{x_1}, p_{y_1}, ..., p_{x_N}, p_{y_N})^T$ and the noise term $\mathbf{w} \in \Re^4$ is defined as $\mathbf{w}^T = \big(\mathbf{w}_{R1}^T,\ \ \mathbf{w}_{R2}^T\big)$ with $\mathbf{w}_{R1}^T(t_k) = \Big(w_{R1}^{(1)}(t_k),\ \ w_{R1}^{(2)}(t_k)\Big)$ and $\mathbf{w}_{R2}^T(t_k) = \Big(w_{R2}^{(1)}(t_k),\ \ w_{R2}^{(2)}(t_k)\Big)$ corresponding to process noise for robot 1 and 2 with covariance matrices $\mathbf{\Sigma}_{\mathbf{w}_1} = \mathbf{diag}(\sigma_1^2,\ \sigma_2^2)$ and $\mathbf{\Sigma}_{\mathbf{w}_2} = \mathbf{diag}(\sigma_3^2,\ \sigma_4^2)$. The drift $\mathbf{F}(\mathbf{X}(t_k))$ and diffusion $\mathbf{G}(\mathbf{X}(t_k))$ are expressed as follows:

$$\mathbf{F}(\mathbf{X}(t_k)) = \begin{pmatrix} \mathbf{f}(\mathbf{x}_1(t_k)) \\ \mathbf{f}(\mathbf{x}_2(t_k)) \\ \mathbf{p}_1(t_k) \\ ... \\ \mathbf{p}_N(t_k) \end{pmatrix}, \mathbf{G}(\mathbf{X}(t_k)) = \begin{pmatrix} \mathbf{G}(\mathbf{x}_1(t_k)) \\ \mathbf{G}(\mathbf{x}_2(t_k)) \\ 0 \\ ... \\ 0 \end{pmatrix} \tag{22}$$

Observation for the multi-robot case consists of 4 models. The first model corresponds to the detection of a landmark p_i by robot 1. The second model corresponds to the detection of a landmark p_i by robot 2. The 3rd and 4th models correspond to detection of one robot by the other. In mathematical terms we will have:

$${}^{R1}\mathbf{y}_1(t_k) = {}^{R1}_{G}\mathbf{R}(\phi_1(t_k))\Big({}^{G}\mathbf{p}_i(t_k) - {}^{G}\mathbf{p}_{R1}(t_k)\Big) + \mathbf{v}_1(t_k) \tag{23}$$

$${}^{R2}\mathbf{y}_2(t_k) = {}^{R2}_{G}\mathbf{R}(\phi_2(t_k))\Big({}^{G}\mathbf{p}_i(t_k) - {}^{G}\mathbf{p}_{R2}(t_k)\Big) + \mathbf{v}_2(t_k) \tag{24}$$

$${}^{R1}\mathbf{y}_3(t_k) = {}^{R1}_{G}\mathbf{R}(\phi_1(t_k))\Big({}^{G}\mathbf{p}_{R2}(t_k) - {}^{G}\mathbf{p}_{R1}(t_k)\Big) + \mathbf{v}_1(t_k) \tag{25}$$

$${}^{R2}\mathbf{y}_4(t_k) = {}^{R2}_{G}\mathbf{R}(\phi_2(t_k))\Big({}^{G}\mathbf{p}_{R1}(t_k) - {}^{G}\mathbf{p}_{R2}(t_k)\Big) + \mathbf{v}_2(t_k) \tag{26}$$

with ${}^{G}\mathbf{p}_i(t_k)^T = \big({}^{G}p_{x_i}(t_k),\ \ {}^{G}p_{y_i}(t_k)\big)$ and ${}^{G}\mathbf{p}_{R1}(t_k) = \big({}^{G}x_1(t_k),\ \ {}^{G}y_1(t_k)\big)$, ${}^{G}\mathbf{p}_{R2}(t_k) = \big({}^{G}x_2(t_k),\ \ {}^{G}y_2(t_k)\big)$ being the position of landmarks, the robot 1 and robot 2 with respect to a global frame of reference $\{G\}$. The parameters $\mathbf{v}_1$ and $\mathbf{v}_2$ correspond to the observation noise of robot 1 and 2 with covariance matrices $\mathbf{\Sigma}_{\mathbf{v}_1} = \mathbf{diag}(\sigma_5^2,\ \sigma_6^2)$ and $\mathbf{\Sigma}_{\mathbf{v}_2} = \mathbf{diag}(\sigma_7^2,\ \sigma_8^2)$ respectively. The matrices ${}^{R1}_{G}\mathbf{R}(\phi(t_k))$ and ${}^{R2}_{G}\mathbf{R}(\phi(t_k))$ are the rotational matrices which express rotational transformations from the global $\{G\}$ to local(=robot1) and local(=robot2) frame of reference $\{R1\}$, $\{R2\}$. In a compact form the observation model is formulated as:

$$\mathbf{y}_j(t_k) = \mathbf{H}_j\left(\mathbf{x}(t_k), \mathbf{v}(t_k)\right), \quad \forall j = 1, 2, 3, 4. \tag{27}$$

The function $\mathbf{H}_j\left(\mathbf{x}(t_k), \mathbf{v}(t_k)\right) : \Re^n \times \Re^q \rightarrow \Re^q$ is defined such that it matches the observation scenarios as expressed in (23),(24),(25) and (26). Given the observation model for the multi-robot case the Kalman Filter update equations are:

$$\begin{aligned} \boldsymbol{\mathcal{H}}_j(\mathbf{x}(t)) &= \nabla_{\mathbf{x}}\mathbf{H}_j(\mathbf{x}(t)), \quad \forall j = 1, 2, 3, 4. \\ \hat{\mathbf{y}}_j(t_{k+1}) &= \boldsymbol{\mathcal{H}}_j(\hat{\mathbf{x}}(t_{k+1|k})) \\ \mathbf{y}_j(t_{k+1}) &= \boldsymbol{\mathcal{H}}_j(\mathbf{x}(t_{k+1})) + \mathbf{v}_i(t_{k+1}) \\ \mathbf{r}_j(t_{k+1}) &= \mathbf{y}_j(t_{k+1}) - \hat{\mathbf{y}}_j(t_{k+1}) \\ \boldsymbol{\mathcal{S}}(t_{k+1}) &= \boldsymbol{\mathcal{H}}_j\mathbf{\Sigma}(t_{k+1|k})\boldsymbol{\mathcal{H}}_j^T + \mathbf{\Sigma}_{\mathbf{v}_i} \\ \boldsymbol{\mathcal{L}}(t_{k+1}) &= \mathbf{\Sigma}(t_{k+1|k})\boldsymbol{\mathcal{H}}_j^T\boldsymbol{\mathcal{S}}(t_{k+1})^{-1} \\ \hat{\mathbf{x}}_{t_{k+1|k+1}} &= \hat{\mathbf{x}}_{t_{k+1|k}} + \boldsymbol{\mathcal{L}}(t_{k+1})\mathbf{r}_j(t_{k+1}) \\ \mathbf{\Sigma}(t_{k+1|k+1}) &= \left(I - \boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{H}}_j\right)\mathbf{\Sigma}(t_{k+1|k})\left(I - \boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{H}}_j\right)^T \\ &\quad + \boldsymbol{\mathcal{L}}\mathbf{\Sigma}_{\mathbf{v}}\boldsymbol{\mathcal{L}}^T \end{aligned} \tag{28}$$

There are many observation scenarios depending on the location of the robot and the landmark as well as the sensing range and the sensing capability of each robot. In particular, these scenarios include cases in which 1) only one robot detects landmarks 2) both robots detect landmarks and they detect each other 3) robots detect only landmarks but they can not detect each other 4) robots do not detect landmarks but they can detect each other. All these scenarios result in a number of updates based on the observation models in (28) which cause a reduction of the total covariance and an improvement in the localization of landmarks and robot position state.

VI. Simulation- Results

Our simulation results consist of a number of a single robot and multi-robot experiments. The assumptions used in this work are summarized as follows:

- We assume that all landmarks have been previously detected. Thus the map of the area where the robot navigates is partially known. With the term partially we mean the map is known with an initial uncertainty that is potentially large. The task for the robot is to navigate through the environment and reach pre-specified goals while at the same time reducing its localization error as well as the map uncertainty.

- We consider that computation and processing takes place in a centralized fashion either in one of the robots or in a central base station. In that sense all that the robots do is to gather information with their proprioceptive and exteroceptive sensors and transmit this information to the central station.
- We assume that there are no communication latencies and therefore the information gathered locally is sent with zero delay to the central station-Robot.
- In our experiments none of the robot has access to GPS sensor measurements.
- We assume that the robot have the same sensors onboard which include an odometer to measure the linear and rotational velocity and a laser scanner which measures relative distance between robot-landmarks and robot-robot. In that sense we are dealing with a homogeneous team of robots.
- The robots can perfectly associate the detected landmarks with the corresponding ones stored in the state vector. Therefore, we are not dealing here with the problem of data association and its impact in the performance of localization.
- The policies constructed for planning are feedforward and they do not include feedback terms.
- We parameterize the rotational velocity with a differential equation of the form:

$$\begin{aligned}\omega(t_k) &= \mathbf{u}\Big(\omega(t_{k-1}), \alpha_\omega(t_{k-1};\boldsymbol{\theta})\Big)\\ &= \omega(t_{k-1}) + \alpha_\omega(t_{k-1};\boldsymbol{\theta})\delta t\end{aligned} \qquad (29)$$

The parameter $\alpha_\omega(t_{k-1};\boldsymbol{\theta})$ can be thought as rotational acceleration and it is parameterized as a trajectory split in 6 parts. For each part i there are two parameters, the duration ΔT_i and the acceleration α_i which remains constant in time interval ΔT_i. The sum of all time intervals is fixed and equal to the pre-specified time horizon, thus $\sum_i^6 \Delta T_i = T_N$. The parameter vector $\boldsymbol{\theta}$ is defined as $\boldsymbol{\theta}^T = (\Delta T_1, \alpha_1, ..., \Delta T_6, \alpha_6)$.

The choice of parameterizing the rotational velocity as in (29) is made so that to ensure smooth sampled trajectories during learning and planning. The smoothness of the trajectories helps to keep the EKF consistent. From the stochastic estimation point of view, keeping EKF consistent is critical such that the Gaussian approximation of the underlying distribution of the transition dynamics remains accurate.

In all of the experiments that involve two robots we provide the optimal trajectories as learned by the cross entropy method as well as the 3 sigma bounds of the error state of the robots to demonstrate the EKF preserved consistency as well as to show how the aforementioned error bounds are affected by the cost function design.

A. *Single-Robot*

The task for the robot is to reach the target at $p^* = [5, \quad 14]^T$. The linear velocity is constant $V = 0.31$ while the discretization is $dt = 0.05$ and the horizon $t_N = 1000$ timesteps. The cost function under minimization is: $\boldsymbol{\mathcal{L}}(\mathbf{x}) = \phi(\mathbf{x}_{t_N}) + \int_0^{t_N}\Big(q(\mathbf{x}) + \frac{1}{2}\mathbf{u}^T R\mathbf{u}\delta t\Big)$ where the terminal cost is $\phi(\mathbf{x}_{t_N}) = ||\mathbf{p}^* - \mathbf{p}_R||^2 + w_{\boldsymbol{\Sigma}} \times trace\left(\boldsymbol{\Sigma}(t_N)\right)$ with $\mathbf{p}_R = \left({}^G x, \quad {}^G y\right)$ and $w_{\boldsymbol{\Sigma}}$ the weight for the trace of the covariance given by the EKF at terminal time t_N. For this experiment we assume that the state dependent cost accumulated over the time horizon $q(\mathbf{x}) = 0$. The results are illustrated in Figure 1. In that Figure, the blue line is the optimal trajectory when $w_{\boldsymbol{\Sigma}} = 0$. In this case the robot selects a path to go to the goal without considering the localization as well as the map uncertainty. Marked with the red, is the optimal trajectory when $w_{\boldsymbol{\Sigma}} = 10^5$. Clearly, in the latter case the task is to reach the desired target but also minimize the localization error. For this reason the optimal path is different than the first case as it visits the part of the state space with maximum landmark visibility. Thus the robot can see more landmarks than in the first case, update more times its state and further reduce the total uncertainty.

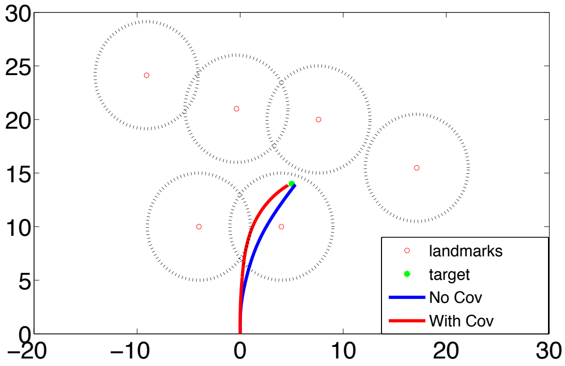


Fig. 1. The trajectories with (red) and without (blue) the trace of the covariance in the cost. The dotted lines encircle the areas within which each landmark is visible to the robot's sensor.

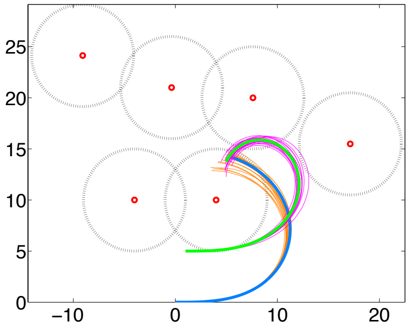


Fig. 2. Trajectories of two robots reaching a common goal. The sampled trajectories are marked orange and magenta for robots 1 and 2 respectively. The minimum cost trajectory is marked with blue for robot 1 and green for robot 2.

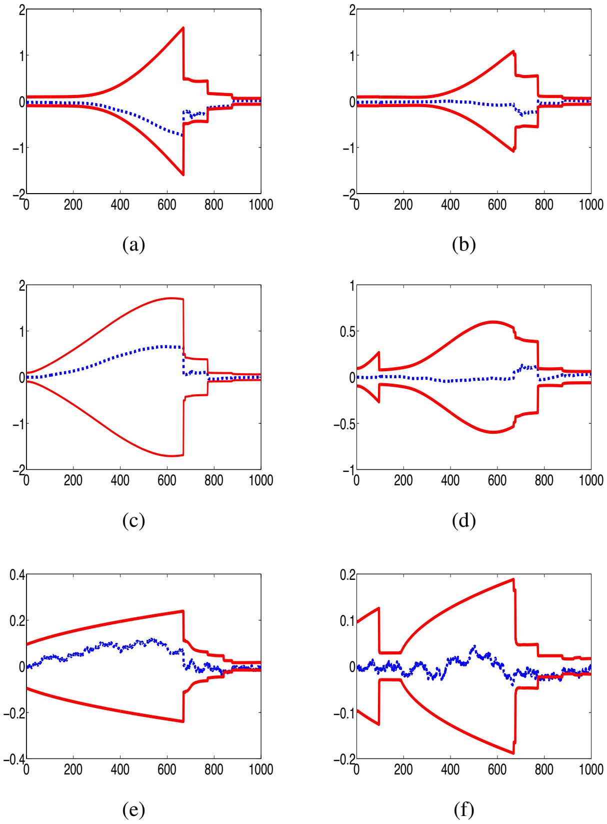


Fig. 3. The 3σ bounds (red) for the localization error (blue) of the robots for the trajectories shown in Figure 2. Subfigures (a)(c),(e) correspond to $\mathbf{x}, \mathbf{y}, \phi$ of Robot1 and (b)(d)(f) to the $\mathbf{x}, \mathbf{y}, \phi$ of Robot2

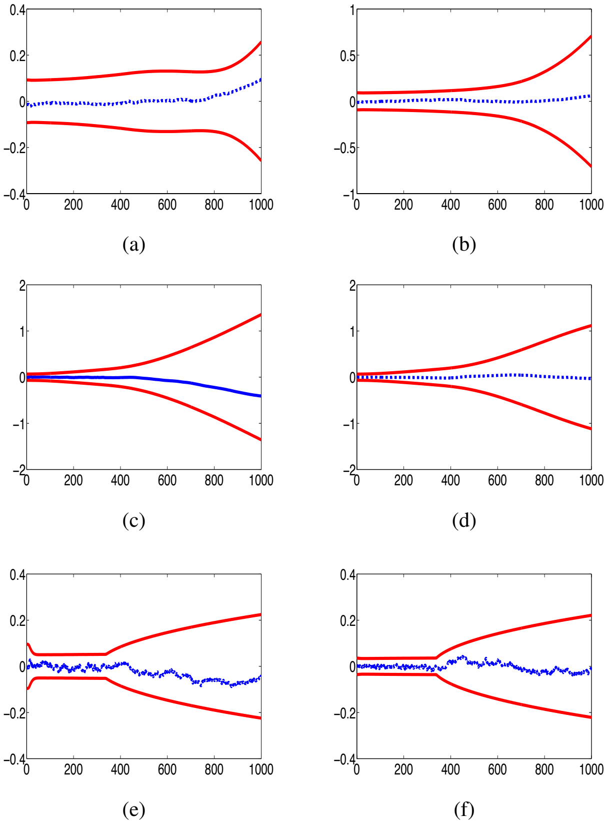


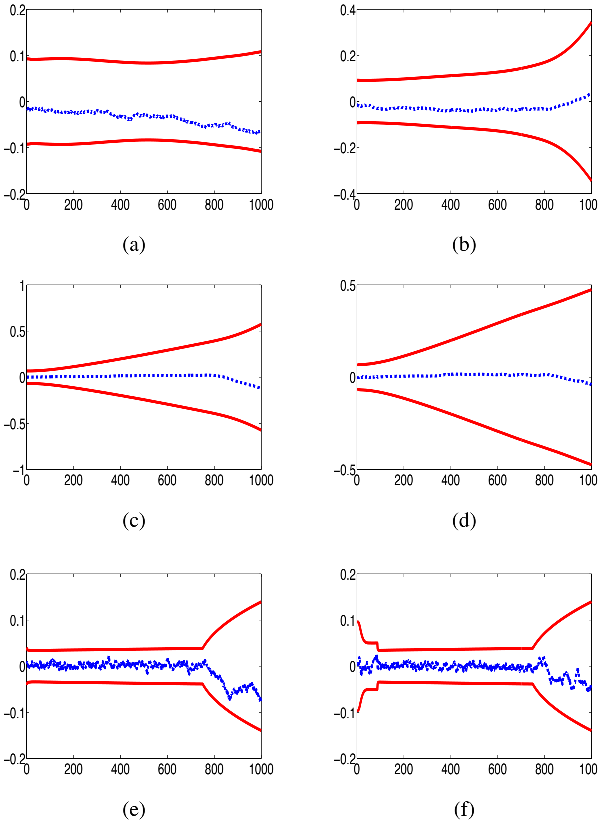
Fig. 4. The sub-figures (a)(c),(e) correspond to $\mathbf{x}, \mathbf{y}, \phi$ state of Robot 1 and (b)(d)(f) $\mathbf{x}, \mathbf{y}, \phi$ state of Robot 2. In this experiment no localization error is considered in the cost for planning.

B. Multi - Robot

To further investigate the behavior of the algorithm, we conducted four experiments that involved more than one robot. In this multi robot scenario, we assume that we have two identical robots. The task for the first experiment is for the robots to meet at $p_1{}^* = p_2{}^* = [5\ 14]^T$. The linear velocity of each robot, the discretization and the time horizon remain the same as in the case of the single robot described in the previous paragraph. Figure 2 shows the resulting trajectories where it can be seen that the robots reach the aforementioned point. Furthermore, the EKF remains consistent through the trajectory as it can be deduced by the 3σ bound plots shown in Figure 3.

The second experiment illustrates the influence of the inclusion of the trace of the covariance matrix into the cost. In this case the task is to reach $p_1{}^* = [15\ 0]^T$ and $p_2{}^* = [13\ 8]^T$ in the absence of landmarks. When the cost includes the trace of the covariance the planner choses trajectories that will bring the robots within sensing distance for a longer period of time as can be seen in Figure 7. Conversely, when the trace of the covariance is not considered, the selected trajectories make the robots visible to each other for a shorter part of the path as shown in Figure 6. Note, that the points where the two robots can detect each other are marked with a wider trajectory line. Figures 4 and 5 show the estimation error for the x, y, ϕ state variables of each robot along with the 3σ bounds. Again, the estimator remains consistent. It is also evident, that the uncertainty bounds for the estimates are tighter when the localization error is considered in the planning cost (Figure 5) than when it is not (Figure 4).

In the third experiment the task is to reach another set of points $p_1{}^* = [-9.5\ 17]^T$ and $p_2{}^* = [-4.5\ 18]^T$ in the presence of various landmarks along the possible paths. Since this minimization problem is significantly more difficult than those of the previous experiments, we allowed for a higher number of path segments as well as for a longer horizon of $t_N = 1500$ timesteps. The linear velocity remains constant at $V = 0.3$. The resulting trajectories are shown in Figures 9 and 8. By comparing these two Figures it is again evident that, when the localization error is included in the cost for planning the resulting trajectories bring the robots within sensing range of each other for a longer part of the path (see Figure 9) than the trajectories which resulted from planning with no consideration of the localization error (see Figure 8). In Figure 10 the covariance bounds together with the position error in x for the two robots are illustrated. By combining the covariance profiles in Figure 10 and the robot trajectories in Figure 9 we see that robot 1 (on the right) first detects a landmark. This detection reduces only the localization error of robot 1 since the robots have not been in a close enough range to detect each other. When robot to robot detection occurs this event causes the drastic



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 5. The sub-figures (a)(c),(e) correspond to $\mathbf{x}, \mathbf{y}, \phi$ state of Robot 1 and (b)(d)(f) $\mathbf{x}, \mathbf{y}, \phi$ state of Robot 2. In this experiment the trace of the covariance is considered for planning.

reduction of the localization error in robot 2 (on the left) even though robot 2 has not detected any landmark. The localization error of robot 1 is not drastically reduced from this event due to continuous landmark detection that keeps its localization error small and bounded. Finally, detection of a new landmark in later phase of the trajectory results in a simultaneous reduction of the localization error for both robots since they are now in the proximity to detect each other.

The fourth and final experiment incorporates an obstacle in the map placed at $p^*_{obs} = [-1.5\ 10]^T$. It is assumed that the robots cannot at any point enter a circle or radius $R_{safety} = 2$ centered at the object. Any trajectory that enters the circle is discarded. The target points are set at ${p_1}^* = [-7\ 17]^T$ and ${p_2}^* = [-4.5\ 18]^T$ while the time horizon and linear velocity remain the same as in the third experiment. The trajectories presented in Figure 11 show the robots reaching their target while avoiding the obstacle that is in their way.

VII. Discussion

We present a new approach for active SLAM and consider the single-robot and multi-robot cases. We integrate ideas on stochastic optimization [1] together with nonlinear stochastic estimation and SLAM for addressing the problem of planning under uncertainty. The main idea in this work is how to perform planning in a partially known environment and consider the localization error as well as the uncertainty of the map of the environment.

Designing a cost function for active SLAM is not straight forward. Previous work in this area suggests the maximiza-

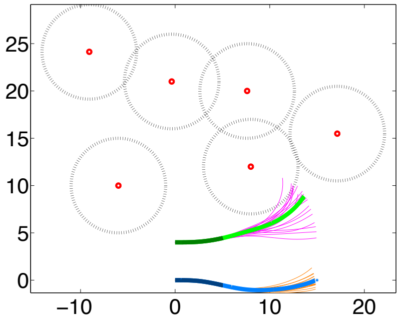


Fig. 6. Trajectories of two robots without considering the trace of the covariance in the cost. The sampled trajectories are orange and magenta for robots 1 and 2 respectively. The minimum cost trajectory is marked with blue for robot 1 and green for 2.

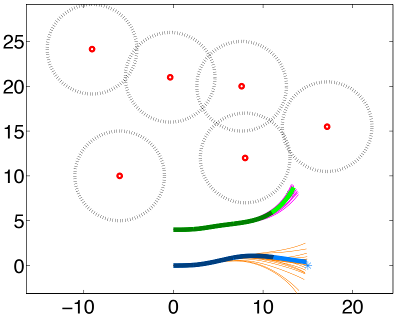


Fig. 7. Trajectories of two robots when the trace of the covariance is included in the cost. The sampled trajectories are orange and magenta for robots 1 and 2 respectively. The minimum cost trajectory is marked with blue for robot 1 and green for 2.

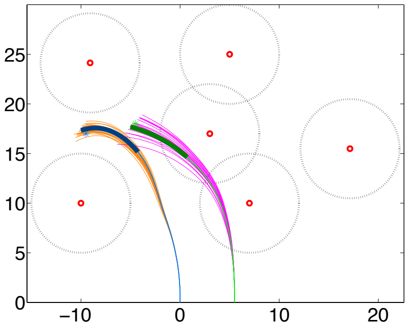


Fig. 8. Trajectories of two robots reaching individual targets through a landmark rich environment when the localization error is not included is the planning cost. The wider and darker trajectory lines corresponds to points where the robots are visible to each other.

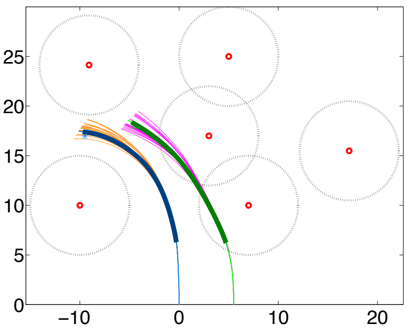


Fig. 9. Trajectories of two robots reaching individual targets through a landmark rich environment when the localization error is included is the planning cost. The wider and darker trajectory lines corresponds to points where the robots are visible to each other.

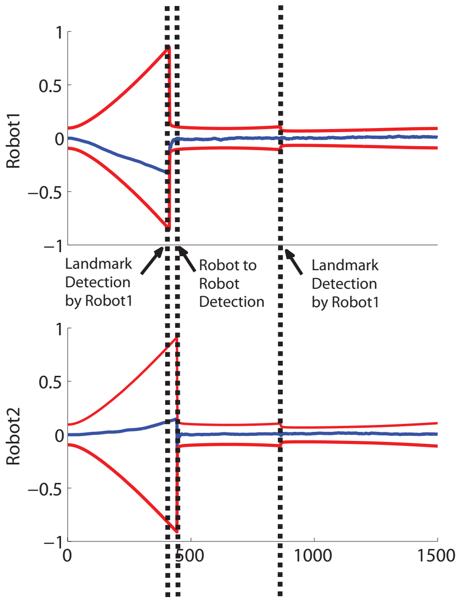


Fig. 10. The covariance bounds and the position error in x for the two robots for experiment 3.

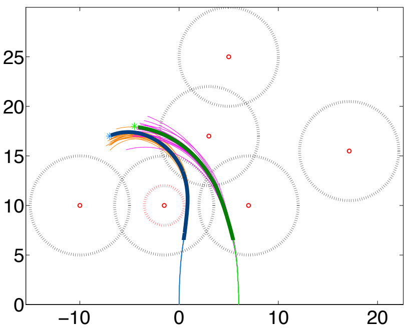


Fig. 11. Trajectories of two robots reaching individual targets while avoiding an obstacle indicated with the red dashed circle. The wider and darker trajectory lines correspond to points where the robots are visible to each other.

tion of information-gain as part of the objective function. The information gain at time t_{k+1} is defined as $\mathbf{I}_{Gain}(t_{k+1}) = trace\left(\mathbf{\Sigma}(t_{k+1|k+1})\right) - trace\left(\mathbf{\Sigma}(t_{k+1|k})\right)$ where the terms $\mathbf{\Sigma}(t_{k+1|k+1})$ and $\mathbf{\Sigma}(t_{k+1|k})$ are the state covariances after the update and propagation phases of EKF [7]. The total information gain gathered over a trajectory is defined as: $\mathbf{I}_{Gain}(t_0 \rightarrow t_N) = \sum_{i=0}^{N} \mathbf{I}_{Gain}(t_i)$.

Maximization of the information gain over state trajectories may result either from the summation of many uncertainty reductions which are small(=high landmark density regions), or from the summation of few but rather large reductions in uncertainty(=low landmark density regions). To avoid this ambiguity in this work we use the trace of covariance matrix at the terminal state as a measure of the uncertainty of the robot localization error and map uncertainty. As we have demonstrated with our simulations, our cost function formulation results in expected behaviors that validate basic intuitions regarding how robots should navigate to reach desired targets while minimizing their localization error.

We are currently working towards incorporating different communication protocols between the robots and the central station. On the reinforcement learning side, we will compare RE with other methods such as Policy Improvement with Path Integrals [9] and free energy based policy gradients [10].

References

[1] Kroese D.P. Rubinstein R.Y Botev, Z.I. and P. L'Ecuyer. *The Cross-Entropy Method for Optimization*, volume 31:Machine Learning. In Handbook of Statistics, 2011.

[2] Peter Dorato, Vito Cerone, and Chaouki Abdallah. *Linear Quadratic Control: An Introduction*. Krieger Publishing Co., Inc., Melbourne, FL, USA, 2000.

[3] W. H. Fleming and H. Mete Soner. *Controlled Markov processes and viscosity solutions*. Applications of mathematics. Springer, New York, 2nd edition, 2006.

[4] Marin Kobilarov. *Cross-entropy Randomized Motion Planning*. In Robotics: Science and Systems, 2011.

[5] C Leung, S. Huang, and G. Dissanayake. Active slam using model predictive control and attractor based exploration. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2006, October 9-15, 2006, Beijing, China*, pages 5026–5031. IEEE, 2006.

[6] Cindy Leung, Shoudong Huang, and Gamini Dissanayake. Active slam in structured environments. In *ICRA*, pages 1898–1903. IEEE, 2008.

[7] R. Sim and N. Roy. Active exploration planning for slam using extended information filters. In *In Proc. 20th Conf. Uncertainty in AI*, 2004.

[8] Robert F. Stengel. *Optimal control and estimation*. Dover books on advanced mathematics. Dover Publications, New York, 1994.

[9] E. Theodorou, J. Buchli, and S. Schaal. A generalized path integral approach to reinforcement learning. *Journal of Machine Learning Research*, (11):3137–3181, 2010.

[10] E. Theodorou, J Najemnik, and E. Todorov. Free energy based policy gradient methods. In *Proceedings of the IEEE on Adaptive Dynamic Programming and Reinforcement Learning*, 2013.