Towards Planning in Generalized Belief Space

Vadim Indelman, Luca Carlone, and Frank Dellaert

Abstract We investigate the problem of planning under uncertainty, which is of interest in several robotic applications, ranging from autonomous navigation to manipulation. Recent effort from the research community has been devoted to design planning approaches working in a continuous domain, relaxing the assumption that the controls belong to a finite set. In this case robot policy is computed from the current robot belief (*planning in belief space*), while the environment in which the robot moves is usually assumed to be known or partially known. We contribute to this branch of the literature by relaxing the assumption of known environment; for this purpose we introduce the concept of *generalized belief space* (GBS), in which the robot maintains a joint belief over its state and the state of the environment. We use GBS within a Model Predictive Control (MPC) scheme; our formulation is valid for general cost functions and incorporates a dual-layer optimization: the outer layer computes the best control action, while the inner layer computes the generalized belief given the action. The resulting approach does not require prior knowledge of the environment and does not assume maximum likelihood observations. We also present an application to a specific family of cost functions and we elucidate on the theoretical derivation with numerical examples.

1 Introduction

Planning is an important component of robot navigation and manipulation, and it is crucial in application endeavours in which the robot operates in full or partial autonomy, e.g., multi-robot exploration, autonomous surveillance, and robotic surgery. The planning problem consists in establishing a map between the state space and the control space, such that the robot can autonomously determine a suitable action (e.g., a motion command), depending on its current state (e.g., current robot pose).

V. Indelman, L. Carlone, F. Dellaert
College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA,
e-mail: indelman@cc.gatech.edu, luca.carlone@gatech.edu, frank@cc.gatech.edu

The complexity of the problem stems from the fact that robot dynamics are stochastic; in most practical applications, the state of the robot is not directly observable, and can only be inferred from observations. Therefore, the robot maintains a probability distribution, the *belief*, over possible states, and computes a control policy using the current belief. The corresponding problem falls in the general framework of *Partially Observable Markov Decision Process* (POMDP).

The literature on planning under uncertainty can be sectioned in three main categories: *simulation-based approaches, infinite-horizon strategies,* and *receding horizon strategies*. Simulation-based approaches generate a few potential plans and select the best policy according to a given metric (e.g., *information gain*). They are referred to as *simulation-based* approaches, since they simulate the evolution of the belief for each potential plan, in order to quantify its quality. Examples of this approach are the work of Stachniss et al. [23, 22], Blanco et al. [4], and Du et al. [7], in which particle filters are used as inference engine. Martinez-Cantin et al. [17, 18] and Bryson and Sukkarieh [5] investigate simulation-based approaches in conjunction with the use of EKF as inference engine. Carrillo et al. [6] provide an analysis of the uncertainty metrics used in EKF-based planning. Other examples of simulation-based approaches are [14, 24, 25] in which the belief is assumed to be a Gaussian over current and past poses of the robot. These works assume *maximum likelihood observations*: since future observations are not given at planning time, the robot assumes that it will acquire the measurements that are most likely given the simulated belief. We notice that, while all the previous examples are applied to mobile robot navigation problems (the corresponding problem is usually referred to as *active Simultaneous Localization and Mapping*), similar strategies can be found with application to manipulation and computer vision (e.g., *next best view* problem [19]).

In the second category, infinite-horizon strategies, the search space is usually discretized (e.g., robot can only move between nodes of a uniformly spaced grid) and the plan may be subject to given budget constraints. The corresponding problem is also referred to as *informative path planning*. These problems are characterized by a combinatorial complexity [8], which increases with the available budget. A greedy strategy for informative path planning is proposed by Singh et al. [21] while a branch and bound approach is proposed by Binney et al. in [3]. More recently, Hollinger et. al [8] propose more efficient algorithms, based on rapidly-exploring random tree and probabilistic roadmap. The approaches falling in these second category usually assumes that the robot moves in a known environment; a remarkable property of these techniques is that they approach optimality when increasing the runtime (which is exponential in the size of the problem). A recent example of infinite-horizon planning is the work [1], in which Bai et al. apply a Monte Carlo sampling technique to update an initial policy, assuming maximum likelihood observations.

Finally, receding horizon strategies compute a policy over the next L control actions, where L is a given horizon. Huang et al. [9] propose a model predictive control (MPC) strategy, associated with EKF-SLAM. Leung et al. [16] propose an approach in which the MPC strategy is associated with a heuristic based on global attractors. Sim and Roy [20] propose A-optimal strategies for solving the active SLAM problem. While these approaches are based on a discretization of the state space [20], or of the space of possible controls [16], recent efforts of the research community are pushing towards the use of continuous-domain models in which controls belong to a continuous set. Continuous models appear as more natural representations for real problems, in which robot states (e.g., poses) and controls (e.g., steering angles) are not constrained to few discrete values. These approaches are usually referred to as *planning in the belief space* (BS). Platt et al. [10] assume maximum likelihood observations and apply linear quadratic regulation (LQR) to compute locally optimal policies. Kontitsis et al. [15] recently propose a sampling based approach for solving a constrained optimization problem in which the constraints correspond to state dynamics, while the objective function to optimize includes uncertainty and robot goal. A hierarchical goal regression for mobile manipulation is proposed by Kaelbling et al. in [11, 12, 13]. While this branch of the literature has already produced excellent results in real problem instances, it still relies on two basic assumptions: (i) future observations are assumed to reflect current robot belief (*maximum likelihood observations*), and (ii) the environment in which the robot moves is partially or completely known. Van den Berg et al. [2] deal with the former issue and propose a general planning strategy in which maximum likelihood assumption is relaxed: future observations are treated as random variables and the future (predicted) belief preserves the dependence on these random variables. In the present work, instead, we deal with the second issue, as we assume no prior knowledge of the environment.

Our contribution belongs to the last category, as we use a receding horizon strategy. We introduce the concept of *generalized belief space* (GBS): the robot keeps a joint belief over *both* the state of the robot and the state of the surrounding environment. This allows relaxing the assumption that the environment is known or partially known, and enables applications in completely unknown and unstructured scenarios. Planning in GBS, similarly to planning in BS, is done in a continuous domain and avoids the maximum likelihood assumption that characterize earlier works. Our planning strategy, described in Section 2, comprises two layers: an inner layer that performs inference in the GBS, and an outer layer that computes a locally-optimal control action. In Section 3, we also present an application to a specific family of cost functions and we elucidate on the theoretical derivation with a numerical example in which a robot has to reach a goal while satisfying a soft bound on the admissible position estimation uncertainty. Conclusions are drawn in Section 4.

2 Planning in Generalized Belief Space (GBS)

2.1 Notation and Probabilistic Formulation

Let x_i and $\mathscr{W}_i$ denote the *robot state* and the *world state* at time t_i. For instance, in mobile robots navigation, x_i may describe robot pose at time t_i and $\mathscr{W}_i$ may describe the positions of landmarks in the environment observed by the robot by time t_i. In a manipulation problem, instead, x_i may represent the pose of the end effector of the manipulator, and $\mathscr{W}_i$ may describe the pose of an object to be grasped. The world state $\mathscr{W}_i$ is time-dependent in general (e.g., to account for possible variations in the environment, or to model the fact that the robot may only have observed a subset of the environment by time t_i) and for this reason we keep the index i in $\mathscr{W}_i$. Let

z_i denote the available observations at time t_i and u_i the control action applied at time t_i. We define the joint state at time t_k as $X_k \doteq \{x_0, \ldots, x_k, \mathscr{W}_k\}$, and we write the probability distribution function (pdf) over the joint state as:

$$p\left(X_k | \mathscr{Z}_k, \mathscr{U}_{k-1}\right), \tag{1}$$

where $\mathscr{Z}_k \doteq \{z_0, \ldots, z_k\}$ represent all the available observations until time t_k, and $\mathscr{U}_{k-1} \doteq \{u_0, \cdots, u_{k-1}\}$ denotes all past controls. The probabilistic motion model given the control u_i and the state robot x_i is

$$p\left(x_{i+1} | x_i, u_i\right). \tag{2}$$

We consider a general observation model that involves at time t_i a subset of joint states $X_i^o \subseteq X_i$:

$$p\left(z_i | X_i^o\right). \tag{3}$$

The basic observation model, commonly used in motion planning, e.g., [2], involves only the current robot state x_i at each time t_i and is a particular case of the above general model (3).

The joint pdf (1) at the current time t_k can be written according to the motion and observation models (2) and (3) as

$$p\left(X_k | \mathscr{Z}_k, \mathscr{U}_{k-1}\right) = priors \cdot \prod_{i=1}^{k} p\left(x_i | x_{i-1}, u_{i-1}\right) p\left(z_i | X_i^o\right). \tag{4}$$

The *priors* term includes $p\left(x_0\right)$ and any other available prior information.

2.2 Approach Overview

The aim of this paper is to present a general strategy that allows a robot (autonomous vehicle, UAV, etc.) to plan a suitable control strategy to accomplish a given *task*. Task accomplishment is modelled through an objective function to be optimized; for instance the objective can penalize the distance to a goal position (*path planning*), the uncertainty in the state estimate (*active sensing*), or can model the necessity to visit new areas (*exploration*). We adopt a standard *model predictive control* (MPC) strategy in which the robot has to plan an optimal sequence of controls $u_{k:k+L-1} = \{u_k, \ldots, u_{k+L-1}\}$ for L look-ahead steps, so that a given objective function is minimized over the time horizon (see Figure 1). The presentation of this section is general and does not assume a specific cost function, while in Section 3 we discuss practical choices of the cost function.

At planning time t_k, the optimal control minimizes an objective function $J_k(u_{k:k+L-1})$ for L look-ahead steps. The objective function involves L immediate costs, one for each look-ahead step. We consider a *general* immediate cost c_{k+l} that may involve any subset of states $X_{k+l}^c \subseteq X_{k+l}$, where $l \in \{0, \ldots, L-1\}$ is the lth look-ahead step. The immediate cost c_{k+l} can be therefore written as

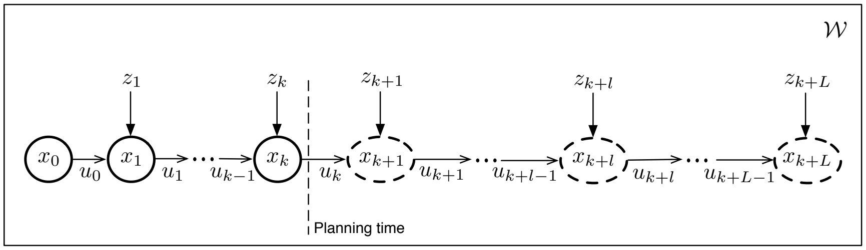


Fig. 1: Illustration of the L look-ahead steps planning problem. Past observations $\mathscr{Z}_k = \{z_0, \ldots, z_k\}$ and past controls $\mathscr{U}_{k-1} = \{u_0, \cdots, u_{k-1}\}$ are known at the planning time t_k. Future observations $z_{k+1:k+L}$ are instead unknown and treated as random variables. The objective is to compute a suitable control strategy $u_{k:k+L-1} = \{u_k, \ldots, u_{k+L-1}\}$ for L look-ahead steps. The figure only illustrates the temporal evolution of the system and we note that each observation may involve generic subset of the states (robot states and world state $\mathscr{W}$) according to the observation model (3).

$$c_{k+l}\left(p\left(X^c_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}\right), u_{k+l}\right), \tag{5}$$

where the notation $a_{i:j} \doteq \left\{a_i, \ldots, a_j\right\}$ is used. As seen in Eq. (5), the immediate cost function c_{k+l} directly involves a distribution over the subset of states X^c_{k+l}. This distribution is conditioned on past measurements and controls $\mathscr{Z}_k, \mathscr{U}_{k-1}$ (that are known at planning time), as well as on future controls and observations $z_{k+1:k+l}, u_{k:k+l-1}$ While the actual observations $z_{k+1:k+l}$ are not given at planning time, the corresponding observation model is known (Eq. (3)) and involves additional subsets of states X^o_{k+j} with $j = [0, \ldots, l]$. Therefore, calculating the pdf $p\left(X^c_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}\right)$ involves an extended subset of the joint state X_{k+l}. For clarity of presentation, however, we will proceed with the *entire* joint state X_{k+l} which contains X^c_{k+l}.[1]

We thus define the *generalized belief space* (GBS) at the lth planning step as

$$gb\left(X_{k+l}\right) \doteq p\left(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}\right). \tag{6}$$

The objective function $J_k\left(u_{k:k+l-1}\right)$ can now be defined as

$$J_k\left(u_{k:k+L-1}\right) \doteq \underset{z_{k+1:k+L}}{\mathbb{E}}\left\{\sum_{l=0}^{L-1} c_l\left(gb\left(X_{k+l}\right), u_{k+l}\right) + c_L\left(gb\left(X_{k+L}\right)\right)\right\} \tag{7}$$

where the expectation is taken to account for all the possible observations during the planning lag, since these are not given at planning time and are stochastic in nature.

[1] In principle, for planning it is only necessary to maintain a distribution over the states X^c_{k+l} while marginalizing out the remaining states. This would avoid performing computation over a large state space, hence resulting in a computational advantage. We leave the investigation of this aspect as an avenue for future research.

Since the expectation is a linear operator we rewrite the objective function (7) as:

$$J_k\left(u_{k:k+L-1}\right) \doteq \sum_{l=0}^{L-1} \underset{z_{k+1:k+l}}{\mathbb{E}}\left[c_l\left(gb\left(X_{k+l}\right), u_{k+l}\right)\right] + \underset{z_{k+1:k+L}}{\mathbb{E}}\left[c_L\left(gb\left(X_{k+L}\right)\right)\right]. \tag{8}$$

The optimal control $u^*_{k:k+L-1} \doteq \left\{u^*_k, \ldots, u^*_{k+L-1}\right\}$ is the control policy π:

$$u^*_{k:k+L-1} = \pi\left(gb\left(X_k\right)\right) = \underset{u_{k:k+L-1}}{\arg\min}\ J_k\left(u_{k:k+L-1}\right). \tag{9}$$

Calculating the optimal control policy (9) involves the optimization of the objective function $J_k\left(u_{k:k+L-1}\right)$. According to (8), the objective depends on the (known) GBS at planning time t_k, on the predicted GBS at time $t_{k+1}, \ldots, t_{k+L}$, and on the future controls $u_{k:k+L-1}$. Since in general the immediate costs $c_l\left(gb\left(X_{k+l}\right), u_{k+l}\right)$ are non-linear functions, $\underset{z_{k+1:k+l}}{\mathbb{E}}\left[c_l\left(gb\left(X_{k+l}\right), u_{k+l}\right)\right] \neq c_l\left(\underset{z_{k+1:k+l}}{\mathbb{E}}\left[gb\left(X_{k+l}\right)\right], u_{k+l}\right)$, and we have to preserve the dependence of the belief $gb\left(X_{k+l}\right)$ on the observations $z_{k+1:k+l}$. The latter are treated as random variables. Therefore, the belief at the lth look-ahead step depends on $u_{k:k+l-1}$ (which is our optimization variable) and $z_{k+1:k+l}$ (which is a random variable).

In order to optimize the objective function (8) we resort to an iterative optimization approach, starting from a known initial guess on the controls. The overall approach can be described as a *dual-layer inference*: the inner layer performs inference to calculate the GBS at each of the look-ahead steps, for a given $u_{k:k+L-1}$. The outer layer performs inference over the control $u_{k:k+L-1}$, minimizing the objective function (8). A schematic representation of the approach is provided in Figure 2, while in the next sections we describe in detail each of these two inference processes, starting from the outer layer: inference over the control.

2.3 Outer Layer: Inference over the Control

Finding a locally-optimal control policy $u^*_{k:k+L-1}$ corresponds to minimizing the general objective function (8). The *outer* layer is an iterative optimization over the non-linear function $J_k\left(u_{k:k+L-1}\right)$. In each iteration of this layer we are looking for the delta vector $\Delta u_{k:k+L-1}$ that is used to update the current values of the controls:

$$u^{(i+1)}_{k:k+L-1} \leftarrow u^{(i)}_{k:k+L-1} + \Delta u_{k:k+L-1}, \tag{10}$$

where i denotes the iteration number. Calculating $\Delta u_{k:k+L-1}$ involves computing the GBS of all the L look-ahead steps based on the current value of the controls $u^{(i)}_{k:k+L-1}$. This process of calculating the GBS is by itself a non-linear optimization and represents the *inner* layer inference in our approach. We describe this inference in Section 2.4. The GBS $gb\left(X_{k+l}\right)$, given the current values of the controls $u_{k:k+L-1}$, is represented by the mean $\hat{X}^*_{k+l}\left(z_{k+1:k+l}\right)$ and the information matrix I_{k+l}. The mean is a linear function in the unknown observations $z_{k+1:k+l}$. The immediate

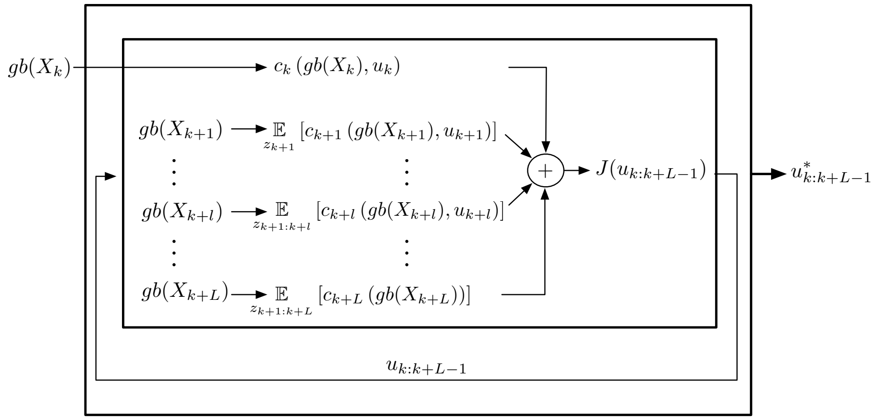


Fig. 2: Illustration of the dual-layer inference planning approach. The algorithm takes as an input the GBS at the current time t_k, $gb\left(X_k\right)$, and produces as output a locally-optimal control $u^*_{k:k+L-1}$. The outer layer performs inference over the control $u_{k:k+L-1}$, while the inner layer evaluates the GBS for a given value of $u_{k:k+L-1}$. Note that the GBS at the lth look-ahead step is a function of controls $u_{k:k+l-1}$ as well as of the random observations $z_{k+1:k+l}$: $gb\left(X_{k+l}\right) = p\left(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}\right)$.

cost function, in the general case, may involve both the mean and the information matrix, and is therefore also a function of $z_{k+1:k+l}$. Taking the expectation over these random variables produces the expected cost that is only a function of $u_{k:k+L-1}$ and captures the effect of the current controls on the lth look-ahead step.

We conclude this section by noting that the control update (10) is performed in a continuous domain and can be realized using different optimization techniques (e.g., dynamic programming, gradient descent, Gauss-Newton).

2.4 Inner Layer: Inference in GBS

In this section we focus on calculating the GBS at the lth look-ahead step $gb\left(X_{k+l}\right) \equiv p\left(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}\right)$, with $l \in [1, L]$. As this inference is performed as part of the higher-level optimization over the control (see Section 2.2), the current values for $u_{k:k+l-1}$ are given in the inner inference layer.

The GBS $gb\left(X_{k+l}\right)$ can be expressed in terms of the joint pdf at the planning time t_k and the individual motion and observation models applied since then:

$$gb\left(X_{k+l}\right) = p\left(X_k|\mathscr{Z}_k, \mathscr{U}_{k-1}\right) \prod_{j=1}^{l} p\left(x_{k+j}|x_{k+j-1}, u_{k+j-1}\right) p\left(z_{k+j}|X^o_{k+j}\right). \tag{11}$$

We consider the case of motion and observation models with additive Gaussian noise:

$$x_{i+1} = f(x_i, u_i) + w_i \quad , \quad w_i \sim N(0, \Omega_w) \tag{12}$$

$$z_i = h(X_i^o) + v_i \quad , \quad v_i \sim N(0, \Omega_v), \tag{13}$$

where for notational convenience, we use $\varepsilon \sim N(\mu, \Omega)$ to denote a Gaussian random variable ε with mean μ and information matrix Ω (inverse of the covariance matrix). Then the distribution $p(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1})$ can also be modeled as a Gaussian and can be expressed as

$$p(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}) \sim N\left(\hat{X}^*_{k+l}, I_{k+l}\right). \tag{14}$$

Our goal is then to calculate the mean $\hat{X}^*_{k+l}$ and the information matrix I_{k+l} describing the GBS at the lth look ahead step, bearing in mind that the observations $z_{k+1:k+l}$ are unknown at planning time t_k.

At this point, it is convenient to assume that the joint pdf at planning time t_k can be parametrized by a Gaussian distribution

$$p(X_k|\mathscr{Z}_k, \mathscr{U}_{k-1}) \sim N\left(\hat{X}_k, I_k\right), \tag{15}$$

with known $\hat{X}_k, I_k$. Taking the negative log of $p(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1})$ from Eq. (11) results in

$$-\log p(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}) = \left\|X_k - \hat{X}_k\right\|^2_{I_k} + \sum_{i=1}^{l} \left[\left\|x_{k+i} - f(x_{k+i-1}, u_{k+i-1})\right\|^2_{\Omega_w} + \left\|z_{k+i} - h\left(X^o_{k+i}\right)\right\|^2_{\Omega_v}\right], \tag{16}$$

where we use the standard notation $\|y-\mu\|^2_\Omega = (y-\mu)^T \Omega (y-\mu)$ for the Mahalanobis norm. The maximum a posteriori (MAP) estimate of X_{k+l} is then given by

$$\hat{X}^*_{k+l} = \underset{X_{k+l}}{\arg\min} \; -\log p(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}). \tag{17}$$

This optimization problem lies at the core of the inner inference layer of our planning approach. In principle, solving (17) involves iterative nonlinear optimization. A standard way to solve the minimization problem is the Gauss-Newton method, where a single iteration involves linearizing the above equation about the current estimate $\bar{X}_{l+l}$, calculating the delta vector ΔX_{k+l} and updating the estimate $\bar{X}_{l+l} \leftarrow \bar{X}_{l+l} + \Delta X_{k+l}$. This process should be repeated until convergence. While this is standard practice in information fusion, what makes it interesting in the context of planning is that the observations $z_{k+1:k+l}$ are unknown and considered instead as random variables.

In order to perform a Gauss-Newton iteration on (17), we linearize the motion and observation models in Eq. (16) about the linearization point $\bar{X}_{k+l}(u_{k:k+l-1})$. The linearization point for the existing states at planning time is set to $\hat{X}_k$, while the future states are predicted via the motion model (12) using the current values of the controls $u_{k:k+l-1}$:

$$\bar{X}_{k+l}\left(u_{k:k+l-1}\right) \equiv \begin{pmatrix} \bar{X}_k \\ \bar{x}_{k+1} \\ \bar{x}_{k+2} \\ \vdots \\ \bar{x}_{k+l} \end{pmatrix} \doteq \begin{pmatrix} \hat{X}_k \\ f\left(\hat{x}_{k|k}, u_k\right) \\ f\left(\bar{x}_{k+1}, u_{k+1}\right) \\ \vdots \\ f\left(\bar{x}_{k+l-1}, u_{k+l-1}\right) \end{pmatrix}. \tag{18}$$

Using this linearization point, Eq. (16) turns into:

$$-\log p\left(X_{k+l}|\mathscr{Z}_k, \mathscr{U}_{k-1}, z_{k+1:k+l}, u_{k:k+l-1}\right) = \|\Delta X_k\|^2_{I_k} + \sum_{i=1}^{l}\left[\left\|\Delta x_{k+i} - F_i \Delta x_{k+i-1} - b_i^f\right\|^2_{\Omega_w} + \left\|H_i \Delta X^o_{k+i} - b_i^h\right\|^2_{\Omega_v}\right], \tag{19}$$

where the Jacobian matrices $F_i \doteq \nabla_x f$ and $H_i \doteq \nabla_x h$ are evaluated about $\bar{X}_{k+l}\left(u_{k:k+l-1}\right)$. The right hand side vectors b_i^f and b_i^h are defined as

$$b_i^f \doteq f\left(\bar{x}_{k+i-1}, u_{k+i-1}\right) - \bar{x}_{k+i} \quad , \qquad b_i^h\left(z_{k+i}\right) \doteq z_{k+i} - h\left(\bar{X}^o_{k+i}\right) \tag{20}$$

Note that b_i^h is a function of the random variable z_{k+i}. Also note that under the maximum-likelihood assumption this terms would be nullified: assuming maximum likelihood measurements essentially means assuming zero *innovation*, and b_i^h is exactly the innovation for measurement z_{k+i}. We instead keep, for now, the observation z_{k+i} as a variable and we will compute the expectation over this random variable only when evaluating the objective function (8). In order to calculate the update vectors ΔX_k and $\Delta x_{k+1}, \ldots, \Delta x_{k+l}$, it is convenient to write Eq. (19) in a matrix formulation, which can be compactly represented as:

$$\left\|\mathscr{A}_{k+l}\left(u_{k:k+l-1}\right)\Delta X_{k+l} - \breve{b}_{k+l}\left(u_{k:k+l-1}, z_{k+1:k+l}\right)\right\|^2, \tag{21}$$

where we used the relation $\|a\|^2_\Omega \equiv \left\|\Omega^{1/2}a\right\|^2$, and $\mathscr{A}_{k+l}$ and $\breve{b}_{k+l}$ are of the following form:

$$\mathscr{A}_{k+l} \doteq \begin{bmatrix} \begin{bmatrix} I_k^{1/2} & 0 \end{bmatrix} \\ \mathscr{F}_{k+l} \\ \mathscr{H}_{k+l} \end{bmatrix} \quad , \quad \breve{b}_{k+l} = \begin{pmatrix} 0 \\ \Omega_w^{1/2}\breve{b}^f_{k+l} \\ \Omega_v^{1/2}\breve{b}^h_{k+l} \end{pmatrix}. \tag{22}$$

Here, $\mathscr{F}_{k+l}$ and $\mathscr{H}_{k+l}$ include all the Jacobian-related entries $\Omega_w^{1/2}F_i$ and $\Omega_v^{1/2}H_i$ (for all $i \in [1,l]$), respectively, and zeros in appropriate locations. Likewise, the vectors $\breve{b}^f_{k+l}$ and $\breve{b}^h_{k+l}$ respectively collect the terms b_i^f and $b_i^h\left(z_{k+i}\right)$. The term $\begin{bmatrix} I_k^{1/2} & 0 \end{bmatrix}$ includes a matrix of zeros of appropriate size for padding.

The information matrix I_{k+l} can now be calculated from Eq. (21) as

$$I_{k+l}\left(u_{k:k+l-1}\right) \doteq \mathscr{A}^T_{k+l}\mathscr{A}_{k+l}, \tag{23}$$

and the update vector ΔX_{k+l}, that minimizes (21), is given by

$$\Delta X_{k+l}\left(u_{k:k+l-1}, z_{k+1:k+l}\right) \doteq \left(\mathscr{A}_{k+l}^{T} \mathscr{A}_{k+l}\right)^{-1} \mathscr{A}_{k+l}^{T} \breve{b}_{k+l} = I_{k+l}^{-1} \mathscr{A}_{k+l}^{T} \breve{b}_{k+l}. \quad (24)$$

Noting that the right hand side vectors b_i^f are zero for the linearization point (18), the only non-zero entries in the vector $\breve{b}_{k+l}$ are the right hand side vectors b_i^h, which depend *linearly* on $z_{k+1}, \ldots, z_{k+L}$. Using the definitions (22), Eq. (24) can hence be written as $\Delta X_{k+l}\left(u_{k:k+l-1}, z_{k+1:k+l}\right) = I_{k+l}^{-1} \mathscr{H}_{k+l}^{T} \breve{\Omega}_v \breve{b}_{k+l}^h$, where $\breve{\Omega}_v$ is an appropriate block diagonal matrix with Ω_v elements. The updated estimate is then calculated as

$$\hat{X}_{k+l}\left(u_{k:k+l-1}, z_{k+1:k+l}\right) = \bar{X}_{k+l} + \Delta X_{k+l} = \bar{X}_{k+l} + I_{k+l}^{-1} \mathscr{H}_{k+l}^{T} \breve{\Omega}_v \breve{b}_{k+l}^h. \quad (25)$$

The estimate $\hat{X}_{k+l}\left(u_{k:k+l-1}, z_{k+1:k+l}\right)$ is the outcome of a single iteration of the nonlinear optimization (17). We remark that for a single iteration, the information matrix $I_{k+l}\left(u_{k:k+l-1}\right)$ does not depend on $z_{k+1:k+l}$ and the mean depends on $z_{k+1:k+l}$ *linearly*. This fact greatly helps when taking the expectation over $z_{k+1:k+l}$ of the immediate cost function (8). Considering more iterations would better capture the dependence of the estimate on the measurements; however, more iterations would make $\hat{X}_{k+l}\left(u_{k:k+l-1}, z_{k+1:k+l}\right)$ a nonlinear function of $z_{k+1}, \ldots, z_{k+L}$. We currently assume a single iteration sufficiently captures the effect of the measurements for a certain control action on the GBS. Therefore,

$$\hat{X}_{k+l}^{*}\left(u_{k:k+l-1}, z_{k+1:k+l}\right) = \hat{X}_{k+l}\left(u_{k:k+l-1}, z_{k+1:k+l}\right).$$

The difference with the maximum-likelihood observations assumption is evident from Eq. (25): in that case only the first term would appear in the above equation. Lastly, we notice that the same linearization point (18) will be used also for the next look-ahead step $(l+1)$, in which only a few additional terms will be added to Eq. (19); this allows large re-use of calculations. Moreover we notice the matrices appearing in Eqs. (23) and (24) are sparse. We leave the investigation of these computational aspects to future research.

3 Application to a Specific Family of Cost Functions

3.1 Choice of the Cost Functions

The exposition of the approach thus far has been given for general immediate cost functions. To demonstrate the effectiveness of our approach we will now focus on a specific family of cost functions. We define:

$$c_l\left(gb\left(X_{k+l}\right), u_{k+l}\right) \doteq \left\| E_{k+l}^{G} \hat{X}_{k+l}^{*} - X^{G} \right\|_{M_x} + tr\left(M_{\Sigma} I_{k+l}^{-1} M_{\Sigma}^{T}\right) + \left\| \zeta\left(u_{k+l}\right) \right\|_{M_u} \quad (26)$$

$$c_L\left(gb\left(X_{k+L}\right)\right) \doteq \left\| E_{k+L}^{G} \hat{X}_{k+L}^{*} - X^{G} \right\|_{M_x} + tr\left(M_{\Sigma} I_{k+L}^{-1} M_{\Sigma}^{T}\right). \quad (27)$$

Here M_Σ, M_u and M_x are given weight matrices, and $\zeta(u)$ is some known function that, depending on the application, quantifies the usage of control u. X^G is the goal for some subset of states (e.g. the last pose), and E^G_{k+l} is a selection matrix, such that the matrix $E^G_{k+l}\hat{X}^*_{k+l}$ contains a subset of states for which we want to impose a goal. Similarly, the matrix M_Σ may also be used to choose the covariance of some subset of states from the joint covariance I^{-1}_{k+l} (e.g. consider only uncertainty of the landmarks in the environment).

Plugging Eqs. (26) and (27) into Eq. (8), taking the expectation, and rearranging the terms, we get

$$J_k\left(u_{k:k+L-1}\right) \doteq \sum_{l=0}^{L-1} \left\|\zeta\left(u_{k+l}\right)\right\|_{M_u} + \sum_{l=0}^{L} tr\left(M_\Sigma I^{-1}_{k+l} M^T_\Sigma\right) + \sum_{l=0}^{L} \underset{z_{k+1:k+l}}{\mathbb{E}}\left[\left\|E^G_{k+l}\hat{X}^*_{k+l} - X^G\right\|_{M_x}\right].$$

We recall that the posterior $\hat{X}^*_{k+l}$ at a generic step l is a function of the observations $z_{k+1:k+l}$, which are random variables. In order to obtain the final expression of the objective function we have to compute the expectation in the last summand in the above equation. We omit the complete derivation for space reasons; in this article we report the final result after taking the expectation:

$$J_k\left(u_{k:k+L-1}\right) \doteq \sum_{l=0}^{L-1} \left\|\zeta\left(u_{k+l}\right)\right\|_{M_u} + \sum_{l=0}^{L} tr\left(M_\Sigma I^{-1}_{k+l} M^T_\Sigma\right) + \underbrace{\sum_{l=0}^{L} \left[\left\|E^G_{k+l}\bar{X}_{k+l} - X^G\right\|_{M_x} + tr\left(Q_{k+l}\left(\mathscr{H}^T_{k+l}\bar{I}^{-1}_{k+l}\mathscr{H}^T_{k+l} + \breve{\Omega}^{-1}_v\right)\right)\right]}_{(a)}, \qquad (28)$$

where $\bar{X}_{k+l}$ is the nominal belief (18), $\bar{I}_{k+l}$ is the information matrix of the nominal belief $\bar{X}_{k+l}$ and $Q_{k+l} = \left(E^G_{k+l}I^{-1}_{k+l}\mathscr{H}^T_{k+l}\breve{\Omega}_v\right)^T M_x \left(E^G_{k+l}I^{-1}_{k+l}\mathscr{H}^T_{k+l}\breve{\Omega}_v\right)$. The first sum contains the terms penalizing the control actions; the second sum contains terms penalizing uncertainty (captured by the information matrix I_{k+l} of the belief); the last term (a) was derived from $\underset{z_{k+1:k+l}}{\mathbb{E}}\left[\left\|E^G_{k+l}\hat{X}^*_{k+l} - X^G\right\|_{M_x}\right]$ and represents the expected incentive in reaching the goal. We notice that the term (a), thus being connected to goal achievement, also contains a term, $tr\left(Q_{k+l}\left(\mathscr{H}^T_{k+l}\bar{I}^{-1}_{k+l}\mathscr{H}^T_{k+l} + \breve{\Omega}^{-1}_v\right)\right)$, that depends on the uncertainty. This term appears because we did not assume maximum likelihood observations, therefore the random nature of the estimate $\hat{X}^*_{k+l}$ (as a function of the random variables $z_{k+1:k+l}$) is preserved.

3.2 Choice of the Weight Matrices

In this section we discuss how to properly choose the weight matrices M_u, M_Σ, and M_x. Most related work assume these matrices are given, while in practice their choice can be scenario dependent and can largely influence the control policy. The matrix M_u, appearing in the summand (a) of (28) has a very intuitive function: a larger M_u induces conservative policies that will penalize large controls (or large variations in the controls, depending on the definition of the function $\zeta(u)$). Consequently, M_u can be tuned to have smoother trajectories or when it is important to keep the controls small (e.g., in presence of strict fuel/power constraints).

The choice of the matrices M_x and M_Σ is instead less intuitive. A balance between these two matrices is crucial for letting the robot satisfy the concurrent tasks of reaching a goal and minimizing the estimation uncertainty. In this section, we propose a grounded way to select these matrices. For simplicity we first assume that the two matrices can be written as $M_x = \alpha_x \bar{M}_x$ and $M_\Sigma = \sqrt{\alpha_\Sigma}\bar{M}_\Sigma$ for some constant and known matrices $\bar{M}_x$ and $\bar{M}_\Sigma$. For instance, $\bar{M}_x$ can simply be a selection matrix that "extract" the subset of states for which we want to set a goal; similarly $\bar{M}_\Sigma$ can be a selection matrix extracting from I_{k+l}^{-1} the marginal covariance that we want to minimize at planning time. Under these assumption the objective function becomes:

$$J_k(u_{k:k+L-1}) = \sum_{l=0}^{L-1} \|\zeta(u_{k+l})\|_{M_u} + \alpha_\Sigma \sum_{l=0}^{L} tr\left(\bar{M}_\Sigma I_{k+l}^{-1} \bar{M}_\Sigma^T\right) + \alpha_x \left[\sum_{l=0}^{L}\left[\left\|E_{k+l}^G \bar{X}_{k+l} - X^G\right\|_{\bar{M}_x} + tr\left(\bar{Q}_{k+l}\left(\mathscr{H}_{k+l}^T \bar{I}_{k+l}^{-1} \mathscr{H}_{k+l}^T + \check{\Omega}_v^{-1}\right)\right)\right]\right].$$

with $\bar{Q}_{k+l} = \left(E_{k+l}^G I_{k+l}^{-1} \mathscr{H}_{k+l}^T \check{\Omega}_v\right)^T \bar{M}_x \left(E_{k+l}^G I_{k+l}^{-1} \mathscr{H}_{k+l}^T \check{\Omega}_v\right)$. The scalar α_x controls the "attraction" towards the goal; similarly α_Σ represents the "importance" of minimizing the uncertainty of the selected states. In order to determine suitable α_x and α_Σ we notice that we can divide the cost function by a constant term, without altering the solution of the optimization problem. Therefore, we divide the cost by $(\alpha_x + \alpha_\Sigma)$, obtaining:

$$J_k(u_{k:k+L-1}) = \sum_{l=0}^{L-1} \|\zeta(u_{k+l})\|_{\bar{M}_u} + (\alpha) \sum_{l=0}^{L} tr\left(\bar{M}_\Sigma I_{k+l}^{-1} \bar{M}_\Sigma^T\right) + (1-\alpha)\left[\sum_{l=0}^{L}\left[\left\|E_{k+l}^G \bar{X}_{k+l} - X^G\right\|_{\bar{M}_x} + tr\left(\bar{Q}_{k+l}\left(\mathscr{H}_{k+l}^T \bar{I}_{k+l}^{-1} \mathscr{H}_{k+l}^T + \check{\Omega}_v^{-1}\right)\right)\right]\right]. \tag{29}$$

where $\alpha = \frac{\alpha_\Sigma}{\alpha_x + \alpha_\Sigma}$ and $\bar{M}_u = \frac{1}{\alpha_x + \alpha_\Sigma} M_u$. The previous expression highlights the trade-off between the last two terms in the cost function (uncertainty reduction VS goal achievement). In order to set α we assume the robot is given an upper bound β on $tr\left(\bar{M}_x I_{k+L}^{-1} \bar{M}_x^T\right)$ (which represents the uncertainty of a selected set of states at the end of the horizon) and we want to compute α so that this upper bound is satisfied. Therefore, we set $\alpha = \frac{tr\left(\bar{M}_x I_{k+L}^{-1} \bar{M}_x^T\right)}{\beta}$ such that for values of $tr\left(\bar{M}_x I_{k+L}^{-1} \bar{M}_x^T\right)$ close to the bound β, the ratio α is closer to 1 and the robot will give more importance to the second summand in (29) (i.e., it will prefer minimizing the uncertainty). Conversely, when the uncertainty is far from the upper bound, the term $(1-\alpha)$ will be large and the robot will prefer reaching the goal. We notice that the quantity $tr\left(\bar{M}_x I_{k+L}^{-1} \bar{M}_x^T\right)$ can eventually become larger than β, as we are not imposing a hard constraint on this term, and for this reason we set $\alpha = \min\left(\frac{tr\left(\bar{M}_x I_{k+L}^{-1} \bar{M}_x^T\right)}{\beta}, 1\right)$.

3.3 Simulation Results

We demonstrate our approach in simulated scenarios in which the robot has to navigate to different goals while operating in an unknown environment, comprising randomly scattered landmarks. The control is found by minimizing the objective function (29), according to our dual-layer inference approach. We use a gradient descent method for optimizing the outer layer and Gauss-Newton for calculating inference in the inner layer. The number of look-ahead steps (L) is set to 5.

For simplicity we assume the robot can only control its heading angle while keeping the velocity constant. The control effort $\zeta(u)$ in Eq. (29) is therefore defined as the change in the heading angle. We assume on-board camera and range sensors with measurements corrupted by Gaussian noise with standard deviation of 0.5 pixels and 1 meters, respectively. Using these sensors the robot can detect and measure the relative positions of nearby landmarks. The corresponding measurements can be described by the observation model (3), where the subset of states X^o comprises the robot's pose and the observed landmark. The motion model is represented by a zero-mean Gaussian with standard deviation of 0.05 meters in position and 0.5 degrees in orientation. The matrices $\bar{M}_x$ and $\bar{M}_\Sigma$ are set to extract the current robot state and covariance at each of the look-ahead steps; M_u is chosen to be the identity matrix.

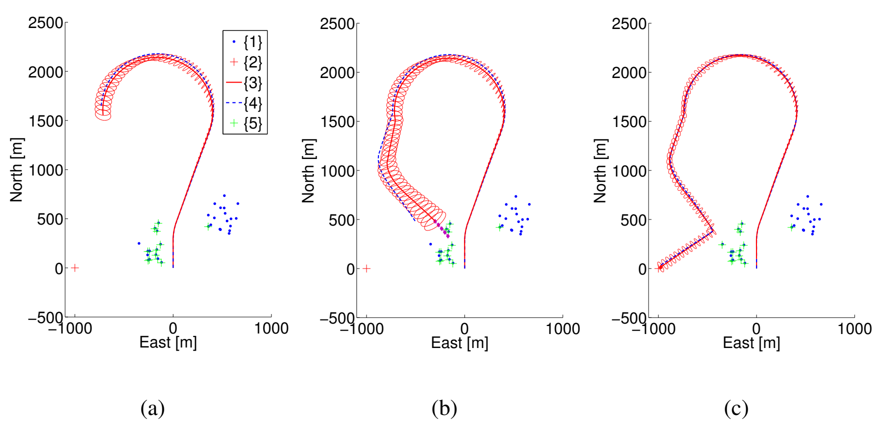
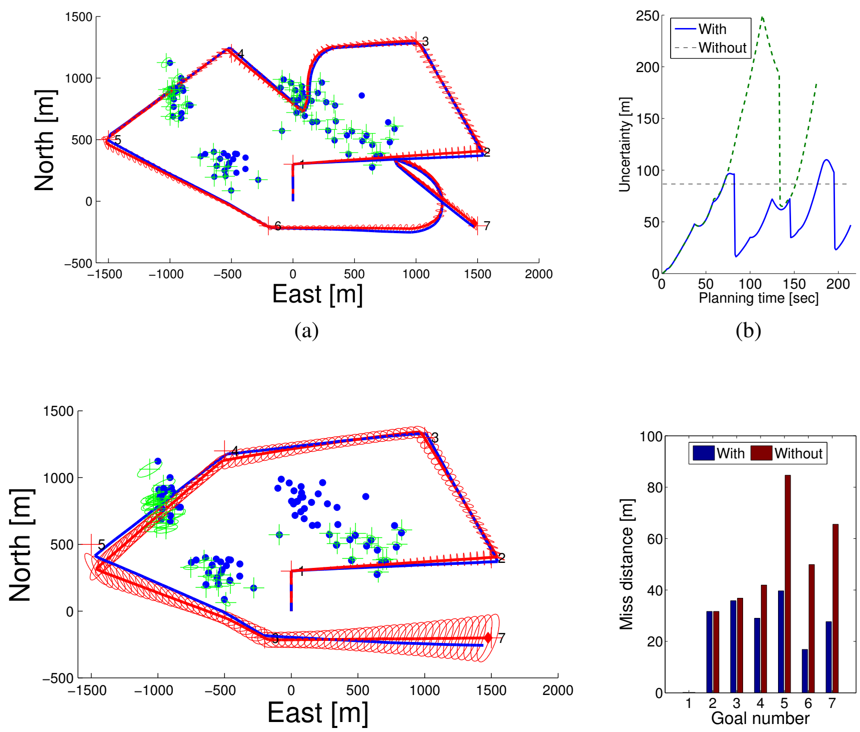


Fig. 3: A mobile robot starts from the configuration (a) and has to plan a motion strategy to reach a goal position, while reducing estimation uncertainty. **Legend**: $\{1\}$ unknown environment $\mathscr{W}$; $\{2\}$ Goal; $\{3\text{-}4\}$ Actual and estimated trajectories of the robot with 1σ uncertainty bounds; $\{5\}$ mapped environment $\mathscr{W}_k$. (a) Robot trajectory before planning begins. (b) Trajectory before loop closure. Planned motion (for $L = 5$ look ahead steps) is shown by 'diamond' marks. (c) Final trajectory after reaching the goal.

We first present a basic scenario where the robot needs to navigate to a single goal (Figure 3). For explanation purposes, in this first example we consider the planning phase starts from the configuration shown in Figure 3a. During the first planning steps the distance to the goal is the dominant component in the objective function (28) and the calculated control guides the robot towards the goal. However, as the robot uncertainty increases, the parameter α increases causing more weight to be placed on uncertainty reduction. The planner then guides the robot towards previously observed landmarks in the environment (Figure 3b). After the robot makes observations of these landmarks (*loop closure*), the uncertainty is reduced, hence the parameter α drops to lower values and more weight is put on guiding the robot to the goal (Figure 3c).



(a) (b)

(c) (d)

Fig. 4: Multi-goal planning example. The same notations as in Figure 3 are used. Resulting trajectories when planning with and without uncertainty terms in the objective function are shown in (a) and (c), respectively. (b) Covariance evolution in the two cases, with the covariance threshold β indicated by a dashed-dotted line. The drops in the covariance values correspond to loop closure events. (d) Miss distance at each of the goals.

We now consider a more complicated scenario, where the planning is carried out from the beginning and the robot has to navigate to a series of goals. The resulting robot trajectory using our approach is shown in Figure 4a. As seen, robot uncertainty exceeds the threshold β twice (on the way to goals 4 and 7), see also Figure 4b, and our planning strategy leads it to re-visiting previously observed landmarks. For comparison, Figure 4c shows the trajectory when the objective function does not account for the uncertainty (i.e. without second and last terms in Eq. (29)). Neglecting the uncertainty during planning leads to much larger covariances, as shown in Figure 4c. Moreover, this results in higher estimation errors, which leads to larger *miss distances*[2] with respect to the goals (Figure 4d).

4 Conclusion

This work presents an approach for planning in the belief space assuming no prior knowledge of the environment in which the robot operates. In order to deal with the uncertainty about the surrounding environment and its state, the robot maintains a joint belief over its own state and the state of the world; this joint belief is used in the computation of a suitable control policy, leading to the concept of *generalized belief space* (GBS) planning. Our approach for planning in the GBS includes two layers of inference: the inner layer performs inference to calculate the belief at each of the look-ahead steps, for a given control action; the outer layer performs inference over the control, minimizing a suitable objective function. The approach does not assume maximum likelihood observations and allows planning in a continuous domain (i.e., without assuming a finite set of possible control actions). We elucidate on the theoretical derivation by presenting an application to a specific family of cost functions and discussing the policies computed in simulated examples.

References

1. H. Bai, H. David, and W.S. Lee. Integrated perception and planning in the continuous space: A POMDP approach. In *Robotics: Science and Systems (RSS)*, 2013.
2. J. Van Den Berg, S. Patil, and R. Alterovitz. Motion planning under uncertainty using iterative local optimization in belief space. *Intl. J. of Robotics Research*, 31(11):1263–1278, 2012.
3. J. Binney and G. S. Sukhatme. Branch and bound for informative path planning. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 2147–2154, 2012.
4. J. Blanco, J. Fernandez-Madrigal, and J. Gonzalez. A novel measure of uncertainty for mobile robot SLAM with Rao-Blackwellized particle filters. *Intl. J. of Robotics Research*, 27(1):73–89, 2008.
5. M. Bryson and S. Sukkarieh. Observability analysis and active control for airborne SLAM. *IEEE Trans. Aerosp. Electron. Syst.*, 44:261–280, 2008.
6. H. Carrillo, I. Reid, and J.A. Castellanos. On the comparison of uncertainty criteria for active SLAM. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 2080–2087, 2012.

[2] The robot considers a goal as achieved when its estimated position coincides with the goal; therefore, the miss distance is defined as the mismatch between the goal and the actual position at that time.

7. J. Du, L. Carlone, M. Kaouk Ng, B. Bona, and M. Indri. A comparative study on active SLAM and autonomous exploration with particle filters. In *IEEE/ASME Intl. Conf. on Advanced Intelligent Mechatronics (AIM)*, pages 916–923, 2011.
8. G. Hollinger and G. Sukhatme. Stochastic motion planning for robotic information gathering. In *Robotics: Science and Systems (RSS)*, 2013.
9. S. Huang, N. Kwok, G. Dissanayake, Q. Ha, and G. Fang. Multi-step look-ahead trajectory planning in SLAM: Possibility and necessity. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 1091–1096, 2005.
10. R. Platt Jr, R. Tedrake, L.P. Kaelbling, and T. Lozano-Pérez. Belief space planning assuming maximum likelihood observations. In *Robotics: Science and Systems (RSS)*, pages 587–593, 2010.
11. L.P. Kaelbling and T. Lozano-Pérez. Pre-image backchaining in belief space for mobile manipulation. In *Proc. of the Intl. Symp. of Robotics Research (ISRR)*, 2011.
12. L.P. Kaelbling and T. Lozano-Pérez. Unifying perception, estimation and action for mobile manipulation via belief space planning. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 2952–2959, 2012.
13. L.P. Kaelbling and T. Lozano-Pérez. Integrated task and motion planning in belief space. *Intl. J. of Robotics Research*, 2013.
14. A. Kim and R.M. Eustice. Perception-driven navigation: Active visual SLAM for robotic area coverage. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2012.
15. M. Kontitsis, E.A. Theodorou, and E. Todorov. Multi-robot active slam with relative entropy optimization. In *American Control Conference*, 2013.
16. C. Leung, S. Huang, and G. Dissanayake. Active SLAM using model predictive control and attractor based exploration. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 5026–5031, 2006.
17. R. Martinez-Cantin, N. de Freitas, E. Brochu, J.A. Castellanos, and A. Doucet. A bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot. *Autonomous Robots*, 27(2):93–103, 2009.
18. R. Martinez-Cantin, N. De Freitas, A. Doucet, and J.A. Castellanos. Active policy learning for robot planning and exploration under uncertainty. In *Robotics: Science and Systems (RSS)*, 2007.
19. C. Potthast and G. Sukhatme. Next best view estimation with eye in hand camera. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2011.
20. R. Sim and N. Roy. Global A-optimal robot exploration in SLAM. pages 661–666, 2005.
21. A. Singh, A. Krause, C. Guestrin, and W.J. Kaiser. Efficient informative sensing using multiple robots. *J. of Artificial Intelligence Research*, 34:707–755, 2009.
22. C. Stachniss, G. Grisetti, and W. Burgard. Information gain-based exploration using rao-blackwellized particle filters. In *Robotics: Science and Systems (RSS)*, pages 65–72, 2005.
23. C. Stachniss, D. Haehnel, and W. Burgard. Exploration with active loop-closing for Fast-SLAM. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2004.
24. R. Valencia, J. Valls Miró, G. Dissanayake, and J. Andrade-Cetto. Active pose SLAM. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, pages 1885–1891, 2011.
25. R. Valencia, M. Morta, J. Andrade-Cetto, and J.M. Porta. Planning reliable paths with pose SLAM. *IEEE Trans. Robotics*, 2013.