

# Simplified Continuous High Dimensional Belief Space Planning with Adaptive Probabilistic Belief-dependent Constraints

Andrey Zhitnikov<sup>1</sup> and Vadim Indelman<sup>2</sup>

<sup>1</sup>Technion Autonomous Systems Program (TASP)

<sup>2</sup>Department of Aerospace Engineering

Technion - Israel Institute of Technology, Haifa 32000, Israel

andreyz@campus.technion.ac.il, vadim.indelman@technion.ac.il

**Abstract**—Online decision making under uncertainty in partially observable domains, also known as Belief Space Planning, is a fundamental problem in robotics and Artificial Intelligence. Due to an abundance of plausible future unravelings, calculating an optimal course of action inflicts an enormous computational burden on the agent. Moreover, in many scenarios, e.g., information gathering, it is required to introduce a belief-dependent constraint. Prompted by this demand, in this paper, we consider a recently introduced probabilistic belief-dependent constrained POMDP. We present a technique to adaptively accept or discard a candidate action sequence with respect to a probabilistic belief-dependent constraint, before expanding a complete set of future observations samples and without any loss in accuracy. Moreover, using our proposed framework, we contribute an adaptive method to find a maximal feasible return (e.g., information gain) in terms of Value at Risk for the candidate action sequence with substantial acceleration. On top of that, we introduce an *adaptive simplification* technique for a probabilistically constrained setting. Such an approach provably returns an identical-quality solution while dramatically accelerating online decision making. Our universal framework applies to any belief-dependent constrained continuous POMDP with parametric beliefs, as well as nonparametric beliefs represented by particles. In the context of an information-theoretic constraint, our presented framework stochastically quantifies if a cumulative information gain along the planning horizon is sufficiently significant (e.g. for, information gathering, active SLAM). We apply our method to active SLAM, a highly challenging problem of high dimensional Belief Space Planning. Extensive realistic simulations corroborate the superiority of our proposed ideas.

## I. INTRODUCTION

A comprehensive approach to craft many online decision-making problems, characterized by the agent situated in an environment and acting under uncertainty, is the Partially Observable Markov Decision Process (POMDP). For most such problems, it is sufficient to assume that the belief-dependent reward is merely the expectation of a state-dependent reward with respect to belief. This assumption is the case in classical POMDP formulations. In contrast, numerous problems in robotics, such as informative planning tasks [1], active Simultaneous Localization and Mapping (SLAM) [2], and sensor placement problem [3] are explicitly concerned with decreasing uncertainty, thereby raising the need for planning with general belief-dependent reward functionals.

General belief-dependent operators were examined in the context of reward but hardly so in the context of the constraint.

In the robotics community, continuous POMDP with belief-dependent information theoretic rewards is known as Belief Space Planning (BSP) [4], [5]. One of the embodiments of BSP, and also the subject of our interest, is active SLAM. In this context, the robot's state comprises the robot's pose trajectory and the map to be estimated. In the Gaussian parametric full SLAM problem, the previous robot poses are not marginalized out but kept to preserve the sparsity of the information matrix [6].

Since belief is to be maintained over an increasingly high-dimensional state, it is not an easy task for an online operating robot. Keeping the whole robot's trajectory causes the state dimension to grow with time. Moreover, the fact that the agent's environment is part of the state, and the agent reveals more area with operation time, also contributes to the state dimensionality. This computational challenge is known as *curse of dimensionality*. With an increasing planning horizon, the number of possible measurements and candidate action sequences grows exponentially, assembling the computationally intractable decision making problem. This phenomenon is usually regarded as the *curse of history*. Many research efforts have targeted both *curses*.

The abundance of possible future observations within the planning phase is often resolved, in robotics, by the Maximum Likelihood (ML) assumption [7]. While widely used, taking into account merely the most likely measurements is highly unrealistic, particularly in the presence of significant uncertainty. It is possible that the largest available reward is not the most likely one, resulting in a substantial error in the objective estimate and consequently suboptimal autonomous behavior. One standing out approach to use a number of sampled observations instead of ML assumption alongside generic belief-dependent rewards builds upon reuse of calculations, alleviating the computational burden [8], [9].

The Artificial Intelligence community also engaged in augmenting the classical POMDP formulation with belief-dependent rewards. The journey started from  $\rho$ -POMDP [10] and significantly advanced through time [11], [12], [13].

Recent methods, merging both worlds, build upon the *simplification* paradigm [14], [15], [16]. These simplification-based methods finally relax limiting assumptions, e.g., Gaussian belief, Piecewise linearity, or Lipschitz continuity of the reward, and permitted universal belief-dependent rewards such as differential entropy of general beliefs. Since the differential entropy operator acts over the belief, which can be

parametrized in various ways, e.g., Gaussian or set of particles, questions of Piecewise linearity or Lipschitz continuity are vague and well defined only when the state is discrete and finite. In a continuous setting they shall be approached individually for each belief parametrization. This fact discards many early approaches [10], [11] to include belief-dependent rewards to POMDP. Another line of *simplification* works alleviate curse of dimensionality in the setting of multivariate Gaussian distributions utilizing sparsification [17] and topological [18] aspects. The simplification paradigm was also applied with Gaussian-mixture distributed beliefs [19], [20]. Another important mechanism for the interplay between computational effort and the quality of the result is adaptivity [21].

All discussed above decision-making methods are concerned with selecting the optimal action, disregarding the actual amount of profit or risk entirely. However, it is essential since preventing the robot from performing unnecessary or self-destructive operations is highly important. This gap can be filled by introducing constraints into decision-making formulation. Some attempts to do so in the context of safe POMDPs include chance constraints [22].

A general belief-dependent constraint, however, has not received proper attention so far except in our previous work [23], where we focused on safety and not information gathering tasks.

In this paper, we continue to investigate the facets of our proposed earlier framework [23] of general belief-dependent constrained continuous POMDP. Motivated by information gathering, also called informative planning tasks, we focus on the cumulative form of the constraint in contrast to the multiplicative form as in our previous paper. One of the specific applications of our framework is stopping exploration. Moreover we provably extend the simplification framework to both forms of the constraints in our novel probabilistically constrained setting. The first form is cumulative and the second is multiplicative.

There are attempts to use differential entropy gain as a constraint to halt exploration in the problem of active SLAM [24] [25]. However, it was never fully explored since, typically, algorithms solving BSP, in particular, active SLAM under partial observability, assume Maximum Likelihood observations [25] to alleviate the computational burden. Stopping exploration is still regarded as an open problem [24].

Our probabilistic belief-dependent constraint of cumulative form, which will become apparent later, generalizes previous approaches. The naive way to threshold a belief-dependent operator under partial observability is to perform expectation with respect to observations. However, even this has gained less attention so far and has not been done to the best of our knowledge since commonly existing approaches take into account only ML observations. In contrast to expectation with respect to future observations, we propose a probabilistic constraint. Our proposed variant is sensitive to the distribution of the belief-dependent constraint, while averaging with respect to future observations is not.

As opposed to a threshold on expectation with respect to observations, we propose two conditions. Interior condition thresholds using  $\delta$  the belief-dependent operator (return) for

given sequence of possible future observations. The exterior condition verifies that the interior one is satisfied with confidence level of at least  $1 - \epsilon$ . To rephrase it, we require that the fraction of the observation sequences fulfilling the interior condition will be at least  $1 - \epsilon$ . In due course, we consider two different problem formulations. In the first problem,  $\delta$  is specified externally by the user. We coin this problem as **optimality under a probabilistic constraint**. In the second problem, that we name **maximal feasible return**,  $\delta$  is a free parameter. In turn, our formulation and approach enable fast adaptive maximization of Value at Risk (VaR) on top of a general belief-dependent return. This problem is highly challenging due to the fact that VaR is not a coherent functional [26].

Our contributions are fourfold. First, we utilize our probabilistically constrained Partially Observable Markov Decision Process (POMDP) in the context of information-theoretic constraint. We analyze Mutual Information constraint in this context, following the expectation approach versus our novel probabilistic constraining. Notably, we did not find any works shifting the Mutual Information from the reward operator to the constraint. Second, we cast unconstrained Risk Aware Belief Space Planning with Value at Risk as a purely constraints-driven problem. As we unveil in this paper, this reformulation enables the decision maker to save time by adaptively expanding the lowest required number of observations without compromising the quality of the solution. Third, we rigorously derive a theory of the *simplification*. Given a converging to the constraint and reward bounds, our approach can be simplified, gaining substantial speedup without any loss in performance. We apply our technique to High Dimensional Belief Space Planning. In particular, our case study is active SLAM.

The remainder of this paper is structured as follows. We start from background and notations in section II. Our next step is the in depth discussion of the problem formulation and our approach (section III). We then present and application of our methods (section IV) and continue to the simulations and results (section V). The conclusions section VI finalizes the paper. We placed the proofs in the appendix to preserve continuity flowing and remove clutter from the central paper ideas.

## II. BACKGROUND AND NOTATIONS

Let us formally introduce the Partially Observable Markov Decision Process with belief-dependent rewards named  $\rho$ -POMDP alias to BSP. By the bold symbols, we denote time vector quantities; by  $\square_{a:b}$ , we mark series annotated by the time discrete indices running from  $a$  to  $b$  inclusive. By  $\mathbb{P}$  we denote probability density function and by  $P$  the probability. By lowercase letter we denote the random quantities or the realizations depending on the context.

The  $\rho$ -POMDP is a tuple  $(\mathcal{X}, \mathcal{A}, \mathcal{Z}, T, O, \rho, \gamma, b_0)$  where  $\mathcal{X}, \mathcal{A}, \mathcal{Z}$  denote state, action, and observation spaces with  $x \in \mathcal{X}$ ,  $a \in \mathcal{A}$ ,  $z \in \mathcal{Z}$  the momentary state, action, and observation, respectively.  $T(x', a, x) = \mathbb{P}_T(x'|x, a)$  is a stochastic transition model from the past state  $x$  to the subsequent  $x'$  through action  $a$ . So far, we have described

the classical components of POMDP. However, in BSP, the observation model  $O(\cdot)$  undergoes a customization that will be apparent later. For now, we left it undefined. Further,  $\gamma \in (0, 1]$  is the discount factor,  $b_0$  is the belief over the initial state (prior), and  $\rho$  is the belief-dependent reward operator.

For conciseness let us denote interchangeably  $\square_{k+}$  and  $\square_{k:k+L-1}$ , as well as  $\square_{(k+1)+}$  and  $\square_{k+1:k+L}$ . In this paper, we deal with static action sequences of variable horizon  $L$ . Namely, our action space is  $\mathcal{A} \triangleq \{a_{k:k+L^i-1}^i\}_{i=1}^{|\mathcal{A}|}$ . Our actions along a particular action sequence are of different lengths. We also can think about such an action sequence as a path  $\mathcal{P}$  comprising motion primitives. However, the action sequence is a much more general notion.

An autonomous robot deployed in an unknown environment repeatedly performs acting, sensing, and planning sessions up until it reaches the required goal or *fails* to do so as we further formulate.

Let  $h_t$  denote history of actions  $a_{0:t-1}$  and observations  $z_{1:t}$  obtained by the agent up to time instant  $t$  and the prior belief  $b_0$ . To clarify, we denote by  $t$  an arbitrary time instant and by  $k$  the time instant of the current planning session. Such that if  $t \geq k$ , the subscript  $t$  regards to future time. Another representation of history is the posterior belief. We define the posterior belief  $b_t$  as a shorthand for the probability density function of the state  $x_t$  given all information up to time instant  $t$ , i.e.  $b_t(x_t) \triangleq \mathbb{P}(x_t|h_t)$ . In this paper the belief converts the history to a more convenient form,  $b_t$  and can be used interchangeably with  $h_t$ , as opposed to our previous work [15].

Often times in BSP problems, the robot's map is unknown and therefore regarded as a random quantity. This allows the robot to operate in unfamiliar environments. We opt for landmarks map representation, so the robot's state is

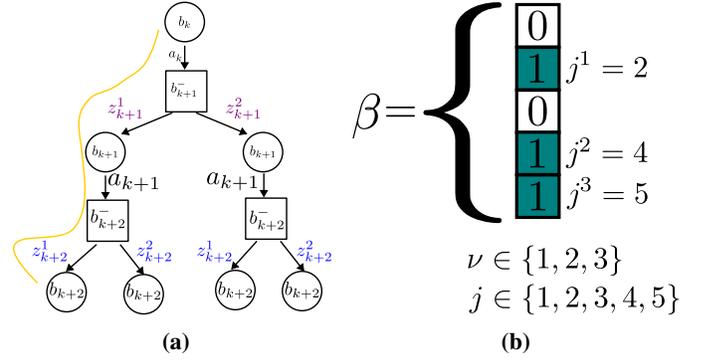
$$\mathbf{x}_t \triangleq \left( x_{0:t}, \{\ell^j\}_{j=1}^{M(k)} \right), \quad (1)$$

where  $M(k)$  is the number of landmarks the robot has observed until time instant  $k$  inclusive. These landmarks represent the unknown robot's environment, specifically the map, to be estimated. To emphasize that  $j$  is not a time index, we denote it by a superscript instead of a subscript.

#### A. Random landmarks configuration vector

Let  $\beta_t \in \{0, 1\}^{M(k)}$  be a random vector of Bernoulli variables, statistically independent given robot's pose  $x_t$ , as will be shortly displayed by (2) and (3). Its dimensionality is the number of landmarks present in the belief. Each realization of  $\beta_t$  defines a subset of visible landmarks. Such a realization has ones at the indexes of visible landmarks and zeros else, such that  $[\beta]^j = 1, \forall j \in \{j^\nu\}_{\nu=1}^{n(\beta)}$ , where  $n(\beta) = \sum_j [\beta]^j$ . (By  $[\cdot]^j$  we indicate the coordinate  $j$  of a vector.) The superscript  $\nu$  defines a subsequence of indices  $j^\nu$  of visible landmarks (Fig. 1b). Let us clarify,  $j^1, j^2, \dots$  represent the strictly increasing with  $\nu$  values of indexes of enumerated landmarks resulting in a random set  $\{j^\nu\}_{\nu=1}^{n(\beta)}$ , such that  $j^\nu = j(\nu)$ .

The mapping from the Boolean vector  $\beta$  to the random finite set of indices  $\{j^1, j^2, \dots\}$  is invertible. Therefore, one can define a probability over the random finite sets [27] instead of boolean vectors.



**Fig. 1:** (a) Visualization of the belief tree given the realization of  $\beta_{k+1:k+2}$  for action sequence  $a_{k+}$ . Here we show two samples of observations per propagated belief. By **purple** and **blue** colors, we denote possibly different dimensionality of the observation stemming from the number of visible landmarks. The yellow lace illustrate the observation sequence  $z_{k+1:k+2}$  (Section III-A); (b) One possible realization of configuration is  $\beta = (01011)^T$ .

One way to define a probabilistic model for visible landmarks configuration given  $x$  is

$$\begin{aligned} \mathbb{P}_\beta([\beta_t]^j = 1|x_t, \ell^j) &= \mathbf{1}_{\{\|x_t - \ell^j\| \leq r\}}, \\ \mathbb{P}_\beta([\beta_t]^j = 0|x_t, \ell^j) &= 1 - \mathbf{1}_{\{\|x_t - \ell^j\| \leq r\}}, \end{aligned} \quad (2)$$

where  $r$  is a visibility radius. Our approach is not limited to this specific model and supports any other model; for instance in more complex scenarios Eqs. (2) would imitate a camera field of view. Eqs. (2) portray that each landmark deterministically has a visibility radius. If the robot is close enough, it receives the signal from the landmark. Overall we arrive at

$$\mathbb{P}_\beta\left(\beta_t \middle| x_t, \{\ell^j\}_{j=1}^{M(k)}\right) = \prod_{j=1}^{M(k)} \mathbb{P}_\beta([\beta_t]^j | x_t, \ell^j). \quad (3)$$

Here, we assumed that  $t \geq k$  and the planner does not reveal new landmarks in planning session, that is,  $M(k)$  depends on the present time but not the future time  $t$ .

Now we are ready to define a customized observation model. This model is used in planning session

$$O(z, \mathbf{x}, \beta) \triangleq \mathbb{P}(z|\mathbf{x}, \beta) = \prod_{\nu=1}^{n(\beta)} \mathbb{P}_Z(z^\nu | x, \ell^{j^\nu}). \quad (4)$$

#### B. The objective

A common BSP objective is given by

$$\mathcal{U}(b_k, a_{k+}) = \mathbb{E}_{\beta_{(k+1)+}} \left[ \mathcal{U}^{\beta_{(k+1)+}}(b_k, a_{k+}) \middle| b_k, a_{k+} \right], \quad (5)$$

where

$$\begin{aligned} \mathcal{U}^{\beta_{(k+1)+}}(b_k, a_{k+}) &= \\ \mathbb{E}_{z_{(k+1)+}} \left[ \sum_{t=k}^{k+L-1} \rho(b_t, a_t, z_{t+1}, b_{t+1}) \middle| b_k, a_{k+}, \beta_{(k+1)+} \right], \end{aligned} \quad (6)$$

and where  $t$  is the running time index and  $k$  is the present time instant. The inner expectation  $\mathcal{U}^{\beta^{(k+1)+}}(b_k, a_{k+})$  (Fig. 1a) corresponds to the utility conditioned on static set of visible landmarks. Therefore, per time index, the dimension of the observation is static (It can be different, however, per time index). Thus, the expectation operator is well defined. The outer expectation performs, weighted in terms of  $\beta^{(k+1)+}$  (Fig. 1b) average of such values. Note that, while it is appealing to fold the conditional expectations in (6) using the law of total expectation, we cannot do that since the dimension of the observation  $z_t$  depends on the number of visible landmarks represented by each specific realization of  $\beta_t$ . By  $\rho(b, a, z', b')$  we denote a general-belief dependent reward depending on two consecutive beliefs and the elements relating them. Further, for clarity, we omit the dependency on the action and the observation.

To summarize this section, BSP accommodates varying dimension of observation conditioned on state and continuous spaces.

### III. PROBLEM FORMULATION AND APPROACH

In this work we tackle two problems. Our **first** problem is the **optimality under a probabilistic constraint**

$$a^* \in \arg \max_{a_{k+} \in \mathcal{A}} \mathcal{U}(b_k, a_{k+}) \text{ subject to} \quad (7)$$

$$P(c(b_{k:k+L}; \phi, \delta) = 1 | b_k, a_{k+}) \geq 1 - \epsilon,$$

where  $c$  is the indicator variable over inner condition as we will shortly see,  $\phi$  is the general belief dependent operator,  $\delta$  and  $0 \leq \epsilon < 1$  are scalars. In this problem the utility conforms to (5). Importantly, as in our previous paper [23]  $\delta$  and  $\epsilon$  are supplied by the user.

The constraint  $c(b_{k:k+L}; \phi, \delta)$  in (7) can be of two forms. The first (cumulative) form is

$$c(b_{k:k+L}; \phi, \delta) \triangleq \mathbf{1}_{\{(\sum_{t=k}^{k+L-1} \phi(b_{t+1}, b_t)) > \delta\}}, \quad (8)$$

and the second (multiplicative) is

$$c(b_{k:k+L}; \phi, \delta) \triangleq \prod_{t=k}^{k+L-1} \mathbf{1}_{\{\phi(b_{t+1}, b_t) \geq \delta\}}. \quad (9)$$

Further, let us refer to the inner inequality as the inner constraint and correspondingly the outer inequality (7) as the probabilistic (outer) constraint. From now on, for clarity, let us denote *constraining* return and the *actual* return operator as  $s(b_{k:k+L}; \phi) \triangleq \sum_{t=k}^{k+L-1} \phi(b_{t+1}, b_t)$  and  $s(b_{k:k+L}; \rho) \triangleq \sum_{t=k}^{k+L-1} \rho(b_{t+1}, b_t)$ , respectively. To include both cases  $\rho$  and  $\phi$  in further discussion, we will denote  $s(b_{k:k+L}; \cdot)$ .

Now, we contemplate what will happen, if  $\delta$  is a free parameter a not pre-determined as before. In this case we would like to select action sequence corresponding to largest maximal feasible return (actual or constraining  $s(b_{k:k+L}; \cdot)$ ) with probability of at least  $1 - \epsilon$ . That is, maximal  $\delta$  yielding that, at most, a single action sequence is feasible. With this insight in mind, we arrive to our **second** problem of **maximal feasible return** defined as follows

$$a^* \in \arg \max_{a_{k+} \in \mathcal{A}} \underbrace{\text{VAR}_\epsilon(s(b_{k:k+L}; \cdot) | b_k, a_{k+})}_{\mathcal{U}(b_k, a_{k+})}, \quad (10)$$

where the Value at Risk (VaR) reads

$$\text{VaR}_\epsilon \left( s(b_{k:k+L}; \cdot) | b_k, a_{k+} \right) \triangleq \delta^* = \sup \left\{ \delta : P \left( s(b_{k:k+L}; \cdot) > \delta | b_k, a_{k+} \right) \geq 1 - \epsilon \right\}. \quad (11)$$

Due to noncompliance to Bellman form of (11) computing (10) is notoriously challenging.

Another way to introduce a belief-dependent constraint to POMDP setting would be by averaging with respect to observations. Namely, the probabilistic constraint in (7) is replaced by  $\mathcal{C}(b_k, a_{k+}) \geq \delta$  given by

$$\mathcal{C}(b_k, a_{k+}) = \mathbb{E}_{\beta^{(k+1)+}} \left[ \mathcal{C}^{\beta^{(k+1)+}}(b_k, a_{k+}) | b_k, a_{k+} \right], \quad (12)$$

where

$$\mathcal{C}^{\beta^{(k+1)+}}(b_k, a_{k+}) = \mathbb{E}_{z^{(k+1)+}} \left[ \sum_{t=k}^{k+L-1} \phi(b_t, a_t, z_{t+1}, b_{t+1}) | b_k, a_{k+}, \beta^{(k+1)+} \right] | b_k, a_{k+}. \quad (13)$$

If  $\phi$  is selected to be Information Gain, (13) is known as Mutual Information (MI). However, if one transfers the utility (5) to the constraint, in other words, when the  $\rho(\cdot) \equiv \phi(\cdot)$  and we use expectation in (12) and in (5) such a constraint appears to be problematic. If  $\mathcal{U}(\cdot) \equiv \mathcal{C}(\cdot)$ , we can always maximize the utility and ask if optimal utility is larger than  $\delta$  ( $\mathcal{U}^* > \delta$ ). In general this is the question of what one verifies first, *optimality* or *feasibility*. As we shall further see, in some cases the order does matter and we can save time by fast feasibility check and cancellation of action sequences. Another option would be to use a maximum likely sequence of observations  $z_{k+1:k+L}^{\text{ML}}$  and check

$$\left( \sum_{\ell=k}^{k+L-1} \phi(b_\ell, a_\ell, z_{\ell+1}^{\text{ML}}, b_{\ell+1}) \right) \geq \delta, \quad (14)$$

where the maximum likelihood observation  $z_{i+1}^{\text{ML}}$  is obtained as follows. We start from a maximum likely state

$$x_{t+1}^{\text{ML}} \in \arg \max_{x_{t+1}} \mathbb{P}(x_{t+1} | b_t, a_t), \quad (15)$$

and then deterministically draw  $\beta_{t+1}$ , using (2). This, in turn, results in

$$z_{t+1}^{\text{ML}} \in \arg \max_{z_{t+1}} \mathbb{P}(z_{t+1} | x_{t+1}^{\text{ML}}, \beta_{t+1}) \quad (16)$$

We can interpret the difference of expected constraint (13) and our probabilistic risk aware constraint (7) as follows. The conventional constraint is not aware of the distribution of the cumulative values of operator  $\phi$ . It decides either the constraint is fulfilled or not solely using the expected value. It is possible that the expected value of the constraint fails to represent adequately the underlying distribution. In contrast, our formulation is **distribution aware**.

In reality to evaluate the probabilistic constraint we shall marginalize over observation sequences to calculate

$$P(c(b_{k:k+L}; \phi, \delta) | b_k, a_{k+}) = \int_{z^{(k+1)+}} P(c(b_{k:k+L}; \phi, \delta) | b_k, a_{k+}, z^{(k+1)+}) \cdot \mathbb{P}(z^{(k+1)+} | b_k, a_{k+}) dz^{(k+1)+}. \quad (17)$$

In the following sections, we develop a universal theory to evaluate our proposed probabilistic inequality  $P(c(b_{k:k+L}; \phi, \delta) = 1; \delta | b_k, a_{k+}) \geq 1 - \epsilon$  *adaptively*. On top of that, we expedite the evaluation process even more by extending *simplification* paradigm to our setting, enjoying the substantially improved celerity versus baseline approaches.

### A. Adaptive Belief Tree

The integral in eq. (17) is not accessible in a general setting. One way to approximately evaluate the (17) is to sample from observation likelihood  $\mathbb{P}(z_{(k+1)+} | b_k, a_{k+})$ . We assume that we have a fixed budget  $m$  of samples of observation laces. Our aim is to leverage the fact that we have a particular structure of the probabilistic condition (17) and to address its evaluation while constructing the belief tree, thereby saving valuable running time or providing a more accurate solution. For clarity, we recite some statements from our previous paper. We remind the reader that this paper focuses on the open-loop setting, namely static action sequences, instead of policies.

Imagine a candidate action sequence  $a_{k:k+L-1}$ . To approximate the utility and the probabilistic constraint (Eq. (7)), an online algorithm at the root (for each candidate action sequence) expands upon termination  $m$  laces appropriate to the drawn observations

$$\{z_{k+1:k+L}^l\}_{l=1}^m. \quad (18)$$

Through the paper we label the laces in the belief tree by the superscript  $l$  (Fig. 1a). Each lace  $l$  corresponds to a particular realization of the sequence of the beliefs, return  $s(b_{k:k+L}; \rho)$  or constraining return  $s(b_{k:k+L}; \phi)$ . The sample approximation of probabilistic constraint (17) is

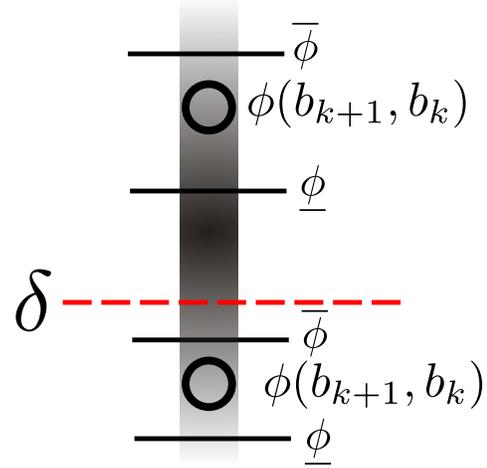
$$\frac{1}{m} \sum_{l=1}^m c(b_{k:k+L}^l; \phi, \delta) \geq 1 - \epsilon. \quad (19)$$

We employ an already expanded part of the belief tree to bound the expression of the probabilistic constraint from each end using the following adaptive upper and lower bounds

$$\underbrace{\frac{1}{m} \sum_{q=1}^{\tilde{m}} c(b_{k:k+L}^{l^q}; \phi, \delta)}_{\text{lb}^{(1)}} \leq \frac{1}{m} \sum_{l=1}^m c(b_{k:k+L}^l; \phi, \delta). \quad (20)$$

$$\frac{1}{m} \sum_{l=1}^m c(b_{k:k+L}^l; \phi, \delta) \leq \underbrace{\frac{m - \tilde{m}}{m} + \frac{1}{m} \sum_{q=1}^{\tilde{m}} c(b_{k:k+L}^{l^q}; \phi, \delta)}_{\text{ub}^{(1)}} \quad (21)$$

where, the algorithm already expanded  $\tilde{m} \leq m$  laces in some order. We denote expanded laces by a sub-sequence  $q \in 1 \dots \tilde{m}$ , such that  $l^q$  is the index of the observation sequence, i.e,  $l^q \in 1 \dots m$ . By the adaptivity we mean the expanding lowest number of laces depending on the situation to accept or discard the candidate action sequence.



**Fig. 2:** Conceptual visualization of our *simplification* approach (Section III-B). For clarity we show a myopic setting. The gradient displays the probability density, i.e., a larger number of samples lands in the area of greater intensity. Using the bounds, we want to assess whether the fraction of the observation laces above  $\delta$  is at least  $1 - \epsilon$ . As we see, we can invalidate bottom sample  $\phi$  using solely the upper bound  $\bar{\phi}$ . In a similar manner, we can validate the upper sample  $\bar{\phi}$  using solely the lower bound  $\underline{\phi}$ . Note that the width of the vertical strip has no role in this visualization.

### B. Adaptive Simplified Constraint Evaluation

As introduced in [15], [14], [17], [19], the simplification paradigm seeks to ease the computational burden in the decision making problem while providing performance guarantees. The latter is achieved by applying bounds over various quantities in the decision making problem (e.g. bounds over a reward function). In this section we extend this concept to our probabilistic belief-dependent constrained POMDP setting of Eq. (7).

Suppose we have adaptive deterministic bounds over  $\phi$ , i.e. these bounds hold for any realization of the beliefs. Further, evaluating these bounds is computationally cheaper than the operator  $\phi$ . Let us present the main theorem of this section, which will shed light on how these bounds can be utilized, propagating their adaptivity further to the adaptive constraint evaluation.

*Theorem 3.1 (Simplification machinery):* Imagine a sample set of the observations laces  $\{z_{k+1:k+L}^q\}_{q=1}^n$ . Assume that  $\forall q$  holds

$$\underline{\phi}(b_{\ell+1}^q, b_{\ell}^q) \leq \phi(b_{\ell+1}^q, b_{\ell}^q) \leq \bar{\phi}(b_{\ell+1}^q, b_{\ell}^q). \quad (22)$$

Let two forms of sample version of inner constraint (8) and (9) variants with bounds be

$$\bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta) \triangleq \mathbf{1}_{\{(\sum_{t=k}^{k+L-1} \bar{\phi}(b_{t+1}^q, b_t^q)) > \delta\}}, \quad (23)$$

$$\underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) \triangleq \mathbf{1}_{\{(\sum_{t=k}^{k+L-1} \underline{\phi}(b_{t+1}^q, b_t^q)) > \delta\}}, \quad (24)$$

and

$$\bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta) \triangleq \prod_{t=k}^{k+L-1} \mathbf{1}_{\{\bar{\phi}(b_{t+1}^q, b_t^q) \geq \delta\}}, \quad (25)$$

$$\underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) \triangleq \prod_{t=k}^{k+L-1} \mathbf{1}_{\{\underline{\phi}(b_{t+1}^q, b_t^q) \geq \delta\}}. \quad (26)$$

Eq. (22), in turn, implies that the following inequalities are satisfied without dependency on the form

$$\sum_{q=1}^{\tilde{m}} \underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) \leq \sum_{q=1}^{\tilde{m}} c(b_{k:k+L}^q; \phi, \delta) \leq \sum_{q=1}^{\tilde{m}} \bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta). \quad (27)$$

We provide a detailed proof of Theorem 3.1 in Appendix C-A.

Let us now show how to speed up the process of evaluation of the probabilistic constraint from (7). The key component of the acceleration is that the adaptivity of the bounds is delegated to adaptivity of the probabilistic constraint bounds (27). Assume the bounds from (22) are adaptive, using insights provided by Theorem 3.1, we first check if

$$\frac{1}{m} \sum_{q=1}^{\tilde{m}} \underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) \stackrel{?}{\geq} 1 - \epsilon. \quad (28)$$

If the above relation holds we declare that the outer constraint is fulfilled. If not, we probe if

$$\frac{1}{m} \sum_{q=1}^{\tilde{m}} \bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta) \stackrel{?}{<} 1 - \epsilon. \quad (29)$$

If yes, we declare that the outer constraint is violated. In case we are not able to say anything (both relations do not hold), we tighten the bounds. In other words, we make the bounds closer to the actual value of  $\phi$  (e.g., by utilizing more particles [15], [16] or mixture belief components [19]). We presented a conceptual visualization of our simplification approach in Fig. 2.

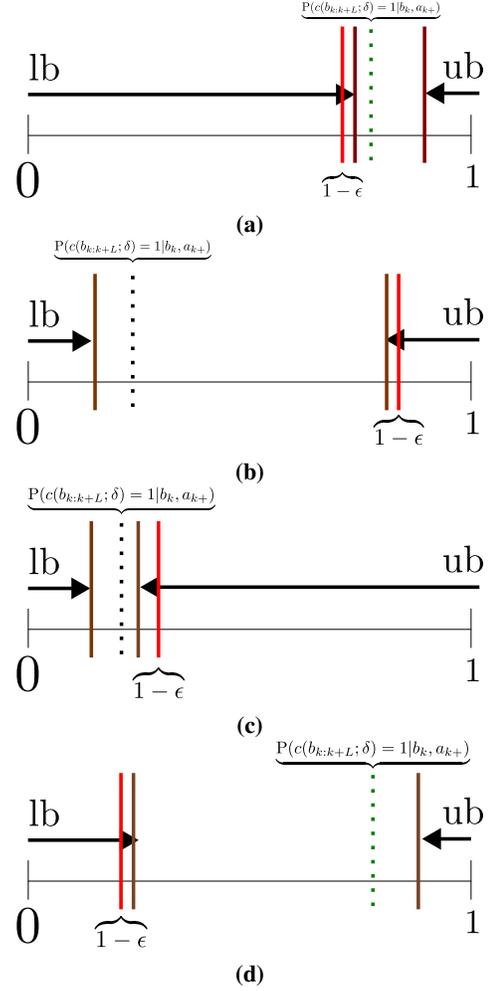
Now our goal is to merge the insights gained in section III-A with the simplification. Clearly from (20) and the left side of (27) we have that

$$1 - \epsilon \stackrel{?}{\leq} \underbrace{\frac{1}{m} \sum_{q=1}^{\tilde{m}} \underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta)}_{\text{lb}^{(2)}} \leq \underbrace{\frac{1}{m} \sum_{q=1}^{\tilde{m}} c(b_{k:k+L}^q; \phi, \delta)}_{\text{lb}^{(1)}}. \quad (30)$$

Similarly from (21) and right side of (27) holds

$$\begin{aligned} & \underbrace{\frac{m - \tilde{m}}{m} + \frac{1}{m} \sum_{q=1}^{\tilde{m}} c(b_{k:k+L}^q; \phi, \delta)}_{\text{ub}^{(1)}} \leq \\ & \leq \underbrace{\frac{m - \tilde{m}}{m} + \frac{1}{m} \sum_{q=1}^{\tilde{m}} \bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta)}_{\text{ub}^{(2)}} \stackrel{?}{<} 1 - \epsilon \end{aligned} \quad (31)$$

By a question mark, we denote the inequalities that shall be fulfilled online to check whether the outer constraint is met (30) or violated (31). If we cannot incur the status of the outer constraint we shall add more laces (adapt the first layer bound) or/and tighten the bounds from (22). Such an approach permits adaptive evaluation of the outer constraint in Eq. (7) before expanding the  $m$  laces of the belief sequences  $b_{k:k+L}$ .



**Fig. 3:** Visualization of adaptation from section III-C. Note in all scenarios the value of dashed line is **unknown**. The red line represents the confidence level  $1 - \epsilon$  to be satisfied with probabilistic constraint. **(a)** Conceptual illustration of challenging scenario. To **accept** such an action the lower bound shall go a long way; **(b)** Conceptual illustration of easy scenario, with a few contractions of the upper bound, the action is **discarded**; **(c)** Another interesting situation, here the upper bound shall go a long way to **discard** an action. **(d)** With a few shrinkage iterations the lower bounds accepts the action sequence.

### C. The adaptation

It occurs that the proposed bounds have riveting properties. To describe a pair of lower ( $\text{lb}^{(1)}, \text{lb}^{(2)}$ ) and a pair of upper bounds ( $\text{ub}^{(1)}, \text{ub}^{(2)}$ ) simultaneously, we omit the superscript. The lower bound is bounded by zero  $0 \leq \text{lb}$  and the upper bound is bounded by one  $\text{ub} \geq 1$ . When we adapt the bounds, we add at most a single lace to the appropriate sum. Therefore, the step of adaptation of the bounds is  $1/m$ .

When we expand a single lace  $\tilde{m} \leftarrow \tilde{m} + 1$ , the lower bound makes a step if  $c(b_{k:k+L}^l; \phi, \delta) = 1$ , otherwise the upper bound makes a step if  $c(b_{k:k+L}^l; \phi, \delta) = 0$ . Alternatively, when we increase the simplification level, some already expanded laces possibly switch from 0 to 1 ( $\underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta)$ ) for some

$l^q$ ), contracting the lower bound, and some from 1 to 0 ( $\bar{c}(b_{k:k+L}^{l^q}; \bar{\phi}, \delta)$  for some  $l^q$ ), tightening the upper bound.

Importantly when we expand a single observation lace and calculate  $\underline{c}(b_{k:k+L}^{l^q}; \underline{\phi}, \delta)$  we will obtain one with probability **at most**  $P(c(b_{k:k+L}; \phi, \delta) = 1 | b_k, a_{k+})$ . Similarly we will obtain  $\bar{c}(b_{k:k+L}^{l^q}; \bar{\phi}, \delta) = 0$  at the new expanded lace with probability at most  $P(c(b_{k:k+L}; \phi, \delta) = 0 | b_k, a_{k+})$ . Both these probabilities are not accessible.

Further, we have four scenarios illustrated in Fig. 3. By analyzing these scenarios, we can speculate about anticipated speedup. In Fig. 3 we show by the red vertical line several positions of the outer threshold  $1 - \epsilon$  from (7). The first scenario, shown in Fig. 3a, is challenging. The unavailable to us probabilistic constraint is fulfilled (shown by green dashed vertical line in Fig. 3a); therefore, no matter how many iterations we perform, invalidation using the calculated ub and Eq. (31) is not possible; only validation using lb and (30) will eventually be possible. As we observe, many contractions of the lb would be required, as we see in Fig. 3a up until lb becomes larger than  $1 - \epsilon$  according to Eq. (30). Conversely, if with a **large margin** the outer constraint is violated as we see in Fig. 3b, we discard the action sequence with a few tightening iterations using ub and (31). We contemplate a similar behavior in reciprocal cases (Figs. 3c and 3d). To conclude the adaptation can be challenging in cases described in Figs. 3a and 3c.

The fact that we have a pair of lower ( $lb^{(1)}, lb^{(2)}$ ) and a pair of upper bounds ( $ub^{(1)}, ub^{(2)}$ ) raises the question which bound from each pair shall we adapt in case that a pair is inconclusive.

When we cannot incur whether the outer constraint from (19) is fulfilled, we shall decide to refine the bounds or add more laces (observation sequences). Luckily for us, these two operations are parallelizable in terms of multithreading. We simultaneously refine the simplification levels, as in [14] of the bounds, and add more laces up until the decision is possible.

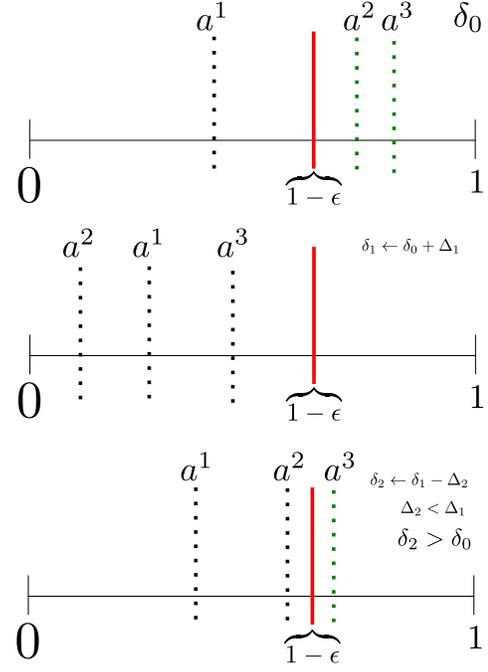
To conclude this section, we proposed a two-layered approach to ease a computational burden. The first layer expresses adaptivity in terms of the number of observation laces. The second layer permits utilization of the adaptive deterministic bounds on realizations of  $\phi | b_k, \pi$ .

One example of using our technique is to save time in open loop planning or spend more time on the action sequences which fulfill the probabilistic constraint. With such an approach, we are able to cut down on the cost of exhaustively validating candidate action sequences. In the setting of Gaussian high dimensional beliefs in SLAM problem, a lower number of observation laces is especially important due to loop closures. Another example is the closed loop setting, where we deal with policies.

Thus far we presented general theory, and now we specifically address the second problem described by Eq. (10).

#### D. Maximal Feasible Return

Picture in your mind that you guess the  $\delta$  and the step size  $\Delta$ . For clarity we drop the dependence of  $s$  on  $b_{k:k+L}$ . However, we shall remember that a single realization of  $s$



**Fig. 4:** Visualization of Alg. 3. We never increase the step size. Therefore, as we see, each candidate action sequence in the bottom visualization is shifted to the left relative to the situation displayed in the top. The action sequence  $a^1$  can be safely discarded in the top illustration (Section III-D).

corresponds to a single lace in the belief tree (Fig. 1a). Observe the following pair of relations

$$P(s > \delta | b_k, a_{k+}) \geq P(s > \delta + \Delta | b_k, a_{k+}), \quad (32)$$

$$P(s > \delta | b_k, a_{k+}) \leq P(s > \delta - \Delta | b_k, a_{k+}). \quad (33)$$

Suppose we fulfill the probabilistic inequality with  $\delta$  for a subset of candidate action sequences, that is,  $P(s > \delta | b_k, a_{k+}) \geq 1 - \epsilon$ . We shall increase  $\delta$  to invalidate more candidate action sequences up until a single candidate action sequence is left. Currently invalidated candidate action sequences can be discarded for eternity, they will never fulfill the outer constraint with larger  $\delta$ , due to the never increasing step size in our approach of alternating increases and decreases of  $\delta$ . Now, suppose we violate the probabilistic inequality with  $\delta$ , that is,  $1 - \epsilon > P(s > \delta | b_k, a_{k+})$  for all the candidate action sequences. We shall decrease the  $\delta$  to render more candidate action sequences feasible. If we will obtain  $\delta$  such that all the candidate action sequences besides the single one are invalidated, we know that this candidate action sequence maximizes (11). This is the underlying principle of Alg. 3. See visualization in Fig. 4. As we see in Fig. 4,  $\delta_2 > \delta_0$  so  $P(s > \delta_0 | b_k, a_{k+}) \geq P(s > \delta_2 | b_k, a_{k+})$ . To the step size, we imply the bisection principle. To rephrase it, we solve

$$a_{k+}^*, \delta^* = \arg \max_{\{a_{k+}\}} \max_{\delta} \quad \delta \quad (34)$$

$$\text{s.t. } \exists a_{k+} \in \mathcal{A} : P(c(b_{k:k+L}; \phi, \delta) | b_k, a_{k+}) \geq 1 - \epsilon$$

This formulation is equivalent to solving the maximal feasible return problem portrayed by equation (10). Before we

---

**Algorithm 1** Optimality under probabilistic constraint

---

```
1: Input:  $\mathcal{A}$  ▷ Set of the action sequences
2:  $a_{k+}^* \leftarrow \text{undef}, \hat{\mathcal{U}}_{(m)}^* \leftarrow -\infty, S \leftarrow \{\}$ 
3: for each  $a_{k+} \in \mathcal{A}$  do
4:   for  $l = 1 : m$  do
5:     Draw observation sequence  $z_{k+1:k+L}^l$ 
6:     Calculate  $c(b_{k:k+L}^l; \phi, \delta)$ ,
7:     if  $1 - \epsilon \leq \frac{1}{m} \sum_{q=1}^l c(b_{k:k+L}^q; \phi, \delta)$  then ▷ Outer constraint is fulfilled
8:        $S \leftarrow S \cup a_{k+}$  ▷ Accept the  $a_{k+}$ 
9:       break ▷ check the next action seq.
10:    else if  $\frac{1}{m} \sum_{q=1}^l c(b_{k:k+L}^q; \phi, \delta) < 1 - \epsilon - \frac{m-l}{m}$  then
▷ Outer constraint is violated
11:      break ▷ check the next action seq.
12:    end if
13:  end for
14: end for
15: for each  $a_{k+}$  in  $S$  do
16:   expand missing laces and get  $\hat{\mathcal{U}}^{(m)}(b_k, a_{k+})$ 
17:   if  $\hat{\mathcal{U}}_{(m)}^* < \hat{\mathcal{U}}^{(m)}(b_k, a_{k+})$  then
18:      $a_{k+}^* \leftarrow a, \hat{\mathcal{U}}_{(m)}^* \leftarrow \hat{\mathcal{U}}^{(m)}(b_k, a_{k+})$ 
19:   end if
20: end for
21: Return  $a_{k+}^*$ 
```

---

---

**Algorithm 2** Optimality under probabilistic constraint (baseline)  $\rho(\cdot) \equiv \phi(\cdot), \mathcal{U}(\cdot) \equiv \mathcal{C}(\cdot)$ 

---

```
1: Input:  $\mathcal{A}$ 
2:  $a_{k+}^* \leftarrow \text{undef}, \hat{\mathcal{U}}_{(m)}^* \leftarrow -\infty,$ 
3: for each  $a_{k+}$  in  $\mathcal{A}$  do
4:   Expand  $m$  laces and get  $\hat{\mathcal{U}}^{(m)}(b_k, a_{k+})$ 
5:   if  $\hat{\mathcal{U}}_{(m)}^* < \hat{\mathcal{U}}^{(m)}(b_k, a_{k+})$  then
6:      $a_{k+}^* \leftarrow a_{k+}, \hat{\mathcal{U}}_{(m)}^* \leftarrow \hat{\mathcal{U}}^{(m)}(b_k, a_{k+})$ 
7:   end if
8: end for
9: if  $\hat{\mathcal{U}}_{(m)}^* \geq \delta$  then
10:  Return  $a_{k+}^*$ 
11: else
12:  Return No feasible action sequence present
13: end if
```

---

continue, note that in Appendix B we discuss sample approximations used in our proposed algorithms. We are ready for the next section, where we formulate algorithms to tackle both of our problems.

### E. Algorithms

In this section, we present four algorithms. All the algorithms receive as input the set of candidate action sequences. How these action sequences are obtained is out of the scope of this paper. For both our problems, we propose our technique and describe the baseline. Our algorithms shall surpass the baseline methods in terms of celerity or/and quality of the solution. Importantly, the overhead from the adaptation shall be neglectable.

---

**Algorithm 3** Maximal feasible return (Bisection method)

---

```
1: Input:  $\mathcal{A}, \delta, \Delta, m$ 
2:  $S \leftarrow \mathcal{A}, T \leftarrow \mathcal{A}, \tilde{\delta} \leftarrow (\delta + \Delta) \cdot 0.5$ 
3:  $\forall a_{k+} \in \mathcal{A}$  expand a single lace and  $\tilde{m} \leftarrow 1$  ▷ warm up
4: while true do ▷ Trials loop
5:   for each  $a_{k+} \in S$  do
6:     if !ADAPTBOUNDS( $a_{k+}, \tilde{m}$ ) then
7:        $S \leftarrow S \setminus a_{k+},$ 
8:     end if
9:   end for
10:  if  $|S| == 1$  then
11:    return  $a_{k+} \in S, \tilde{\delta}$ 
12:  else if  $S \subset \emptyset$  then
13:     $S \leftarrow T, \Delta \leftarrow \tilde{\delta}, \tilde{\delta} \leftarrow (\delta + \tilde{\delta}) \cdot 0.5$ 
14:  next
15:  else
16:     $T \leftarrow S, \delta \leftarrow \tilde{\delta}, \tilde{\delta} \leftarrow (\tilde{\delta} + \Delta) \cdot 0.5$ 
17:  end if
18: end while
19: procedure ADAPTBOUNDS(action seq:  $a_{k+}$ , counter:  $\tilde{m}$ )
20:  while true do
21:    if  $\tilde{m} < m$  then
22:       $\tilde{m} \leftarrow \tilde{m} + 1,$  Draw a lace  $z_{k+1:k+L}^{\tilde{m}}$ ,
23:    end if
24:    if  $\frac{1}{m} \sum_{l=1}^{\tilde{m}} c(b_{k:k+L}^l; \phi, \tilde{\delta}) < 1 - \epsilon - \frac{m-\tilde{m}}{m}$  then
25:      status  $\leftarrow$  false
26:    break
27:    else if  $1 - \epsilon \leq \frac{1}{m} \sum_{l=1}^{\tilde{m}} c(b_{k:k+L}^l; \phi, \tilde{\delta})$  then
28:      status  $\leftarrow$  true
29:    break
30:  end if
31: end while
32: return status
33: end procedure
```

---

---

**Algorithm 4** Baseline maximizing  $\widehat{\text{VaR}}_{\epsilon}^{(m)}$ 

---

```
1: Input:  $\mathcal{A}$ 
2:  $a_{k+}^* \leftarrow \text{undef}, \hat{\mathcal{U}}_{(m)}^* \leftarrow -\infty$ 
3: for each  $a_{k+} \in \mathcal{A}$  do
4:   Expand  $m$  laces and approximate  $\widehat{\text{VaR}}_{\epsilon}^{(m)}$ 
5:   if  $\hat{\mathcal{U}}_{(m)}^* < \widehat{\text{VaR}}_{\epsilon}^{(m)}$  then
6:      $a_{k+}^* \leftarrow a_{k+}, \hat{\mathcal{U}}_{(m)}^* \leftarrow \widehat{\text{VaR}}_{\epsilon}^{(m)}$ 
7:   end if
8: end for
9: Return  $a_{k+}^*$ 
```

---

1) *Optimality under probabilistic constraint:* For the first problem (7), we adaptively check the feasibility of all the action sequences and select the optimal from the set of feasible action sequences Alg. 1. The competing approach is finding the optimal action sequence and verifying feasibility afterwards, see Alg. 2.

2) *Maximal feasible gain:* Here, we propose our adaptive method described in Section III-D and summarized in Alg. 3

and evaluate/compare it versus the brute force maximization of Value at Risk Alg. 4.

Having introduced the algorithms we shall discuss possible drawbacks and overhead.

#### F. Adaptation overhead

When we use presented above adaptation mechanisms, we store laces  $c^l(b_{k:k+L}; \phi, \delta)$  for every expanded  $l$ . Accordingly, the memory consumption is elevated, however not much since these are boolean values. Moreover, we shall evaluate the inner constraint and perform the sum for multiple values of  $\delta$  in Alg. 3. Nevertheless, as we believed and verified by the experiments, this overhead is neglectable compared to the saved time on skipped laces due to loop closures. In addition, these additional operations can be easily parallelized in terms of multithreading.

We can, however, encounter a worst-case scenario. Imagine the  $\epsilon$  is close to 1 from the left. Many action sequences will satisfy the probabilistic constraint. In general, we can say that a more accurate precision of  $\delta$  will be required to differentiate between the action sequences since the working area is closer to zero and the interval  $[0, 1 - \epsilon]$  is shorter. Therefore, more iterations of Alg. 3 will be required. Moreover, a pair of action sequences may be extremely close to each other in terms of Value at Risk, requiring a tremendous amount of iterations of the Alg. 3. To solve this issue, we shall introduce a final precision.

### IV. APPLICATION TO BELIEF SPACE PLANNING

In this section we apply proposed algorithms to informative planning with high dimensional robot's state. We express the exploration problem with our framework (7).

#### A. Belief structure

Let us delve into the mechanics of maintaining and updating high-dimensional belief on top of a stochastic process - sequential decision making. A standard and widely used tool used to maintain a high-dimensional belief is a factor graph [28]. Its building blocks are the probabilistic motion and observation models. These models induce probabilistic dependencies over the state variables. The models are the factors that comprise the factor graph.

In this paper, the stochastic motion and observation models are described by the following dependencies involving the Gaussian distributed sources of stochasticity.

$$x_{i+1} = f(x_t, a_t; w_t), \quad w_t \sim \mathcal{N}(0, W_t), \quad (35)$$

$$z_t^{j^\nu} = g(x_t, \ell^{j^\nu}; v_t), \quad v_t \sim \mathcal{N}(0, V_t), \quad (36)$$

where  $W_t$  and  $V_t$  are covariance matrices. In this paper we assume that the data association is solved. Namely, in general, the belief would be (see, e.g., [29], [30])

$$b_k(\mathbf{x}_k) \triangleq \mathbb{P}(\mathbf{x}_k | b_0, a_{0:k-1}, z_{1:k}) = \sum_{\beta_{1:k}} \mathbb{P}(\mathbf{x}_k | b_0, a_{0:k-1}, z_{1:k}, \beta_{1:k}) \mathbb{P}(\beta_{1:k} | b_0, a_{0:k-1}, z_{1:k}). \quad (37)$$

We, however, assume that given an observation the the realization of corresponding  $\beta$  is known. This fact simplifies the belief structure as such

$$\mathbb{P}(\mathbf{x}_k | b_0, a_{0:k-1}, z_{1:k}) = \mathbb{P}(\mathbf{x}_k | b_0, a_{0:k-1}, z_{1:k}, \beta_{1:k}). \quad (38)$$

Applying the Bayes Rule, we arrive at

$$b_k(\mathbf{x}_k) \propto b_0(x_0) \prod_{i=1}^k \left( \underbrace{\mathbb{P}_T(x_t | x_{t-1}, a_{t-1})}_{\text{motion factor}} \prod_{\nu=1}^{n(\beta_t)} \underbrace{\mathbb{P}_Z(z_t^{\nu_i} | x_t, \ell^{j^{\nu_i}})}_{\text{observation factor}} \right) \quad (39)$$

Eq. (39) can be illustrated as a factor graph [6]. In this paper, we utilize Gaussian probabilistic models (35) and (36) in our simulations.

#### B. Information Gain

Let us address the cumulative form of the inner constraint (8). Similar to [25], we define the operator  $\phi$  as follows

$$\phi(b', b) \triangleq \text{IG}(b, a, z', b'). \quad (40)$$

There are various ways to define the Information Gain (IG) over a pair of the beliefs. One possibility is as follows

$$\text{IG}(b, a, z', b') = -\text{h}(b') + \text{h}(b). \quad (41)$$

The differential entropy  $\text{h}(b)$  is defined by

$$\text{h}(b) \triangleq \int_x b(x) \log b(x) dx. \quad (42)$$

To employ Alg. 3 we require to supply minimal ( $\delta^{\min}$ ) and maximal ( $\delta^{\max}$ ) threshold for inner constraint (eq. (8) and (9)). Let us unveil how we do that.

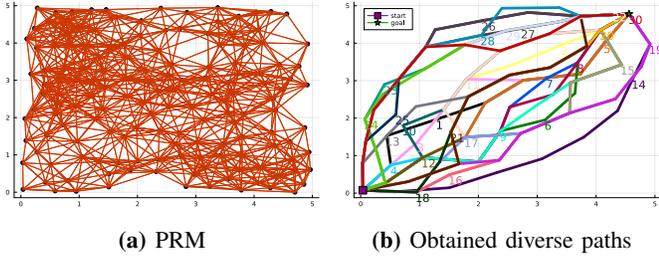
Differential entropy (42) was widely researched in the context of multivariate Gaussian beliefs and led to the formulation of the  $D$ -optimality criterion being the multiplication of eigenvalues of the covariance matrix of the belief (the volume of  $d$ -dimensional parallelepiped proportional to the volume of a hyper-ellipse manifested by the covariance matrix). The information gain becomes

$$0 \underset{\substack{\leq \\ \text{require} \\ I' \geq I}}{\leq} - \underbrace{\sqrt[d]{\prod_i \lambda^{i'}}}_{I' \leq 0} + \underbrace{\sqrt[d]{\prod_i \lambda^i}}_{-I} \leq \sqrt[d]{\prod_i \lambda^i}, \quad (43)$$

where  $d$  is the dimension of the subset of the variables selected from the Gaussian belief. For the reason that our focus is on the uncertainty of the environment surrounding the robot, we select **all the landmarks** as such a subset alongside the **current robot pose**. Since we do not add landmarks in the planning session, the same dimensionality is preserved.

Moreover from (43) we elicit that the maximal feasible  $\delta$  is  $\delta^{\max} \triangleq \sqrt[d]{\prod_i \lambda^d}$ . Meaning, the uncertainty has been reduced to zero in the resulting Gaussian (partial) belief. To continue exploration we select  $\delta^{\min} = 0$ . To summarize in the setting of multivariate Gaussian beliefs  $0 < \delta \leq \sqrt[d]{\prod_i \lambda^i}$ .

Having untangled these aspects, we are keen to demonstrate the superiority of the proposed approach in the following section.



**Fig. 5:** Separate, algorithmically selected paths to the goal **(b)** on top of PRM **(a)**. We show the path number on the vertex, which is removed for finding the subsequent diverse path. The last's path number is shown at its final vertex (the goal).

## V. SIMULATIONS AND RESULTS

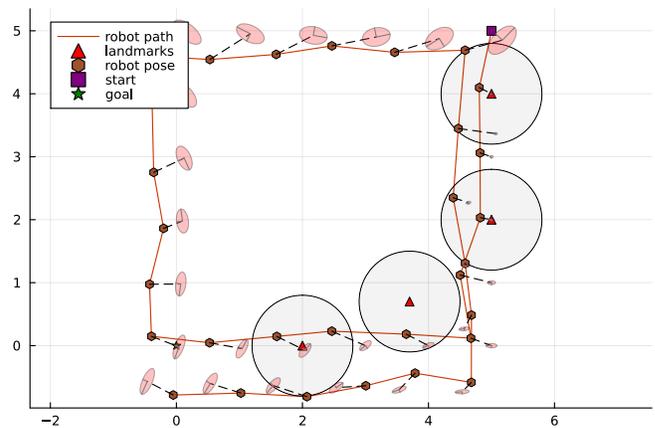
The previous discussion leads us to the actual implementation and simulations of the proposed methods in section III-E. We evaluate our approach by tackling the problem of navigation to the goal in **unknown environments** as an incarnation of BSP. The simulation involves a **highly realistic SLAM** scenario using the gtsam library [31].

As previously mentioned, the generation of candidate paths is not the focus of this paper. Therefore, we generate candidate paths following a similar procedure to [32]. First, we employ a well-studied Probabilistic Road MAP (PRM) method [33]. Then, on top of PRM, to obtain diverse shortest paths, we remove a single vertex from the previous path and utilize Breadth-First Search on the reduced PRM. The path generation requires only the boundaries of an unknown map. In such a way, we obtain  $|\mathcal{A}|$  diverse paths to the goal of various lengths. These paths constitute the space of action sequences  $\mathcal{A}$  (Fig. 5a). To avoid confusion, we recite that any other method for generating candidate paths would be applicable to evaluate our proposed techniques. We illustrated the described above in Fig. 5.

To keep the examination clear, we do not perform re-planning sessions. Instead, we have a preliminary mapping session with manually supplied to the robot action sequence of unit length motion primitives. In the preliminary session, the robot starts from  $b_0$ , detects the landmarks, incorporates them into its state, and obtains the belief  $b_k$ . This belief serves as input to the planning session on top of the candidate paths. After a single planning session, the robot follows the calculated optimal path.

We assume Gaussian sources of stochasticity. For motion (35) and observation (36) models we select  $W_t = \|a\|_2 \cdot \text{diag}(0.015, 0.015, 0.015)$  and  $V_t = \text{diag}(0.001, 0.001)$  respectively. Noticeable, we need to multiply the motion model covariance matrix by the action length since our actions are of variable length. Our prior belief is Gaussian over the robot's pose  $b_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$  with the parameters  $\mu_0 = (0.0, 0.0, 0.0)^T$ ,  $\Sigma_0 = \text{diag}(0.001, 0.001, 0.001)$ . The boundaries of our map are  $[0, 5] \times [0, 5]$ .

We utilize the popular incremental solver ISAM2 [34] to maintain the belief. Noticeably, loop closures impose a computational challenge even with such a sophisticated incremental



**Fig. 6:** Robot first preliminary mapping session, by transparent gray circles, we depict landmarks' visibility radius. The robot starts at the top right corner and moves towards the bottom left corner making two full squares. As we can see, the robot passed inside the visibility radius of the landmarks, detected them and incorporated them to its state. We show covariance ellipses for current robot poses. The landmarks visibility radius is 0.8. By the dashed line we connect estimated robot pose with ground truth.

solver. Especially since we need to perform inference for each posterior node in the constructed belief tree. This fact makes early eliminating or accepting actions highly important for efficient robot's operation.

The robot constructs a belief tree for each candidate path withing planning session. With each promotion of the depth of the belief tree, we reduce the number of observations at each belief node by factor two, up to a possible single observation at the lowest levels. Once the maximal number of observations of the belief node is expanded, we maintain a circular slider that selects the subsequent observation with the following arrival at this belief node.

The advantage of our proposed methods is acceleration without compromising the solution quality. We calculate the speedup using the following equation

$$\frac{t^{\text{baseline}} - t^{\text{our}}}{t^{\text{baseline}}}. \quad (44)$$

Each planning session is initialized by the same seed. In addition, we do the same calculation to the relative fraction of the skipped laces

$$\frac{n^{\text{total}} - n^{\text{expanded}}}{n^{\text{total}}}. \quad (45)$$

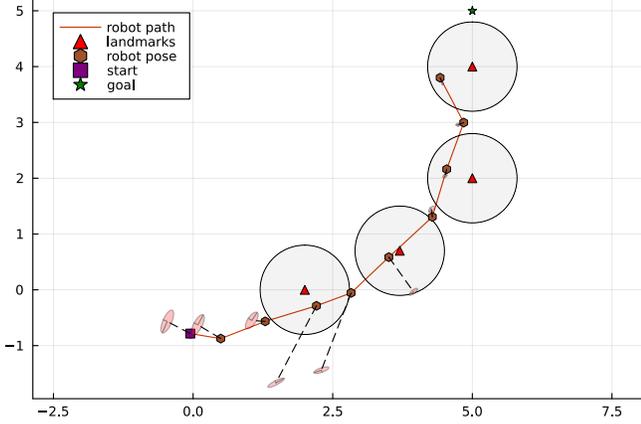
Note that maximal values of (44) and (45) are 1. This means that our approach skipped all the laces ( $n^{\text{expanded}} = 0$  in (45)) and run in zero time ( $t^{\text{our}} = 0$  in (44)).

### A. Optimality under Probabilistic Constraint

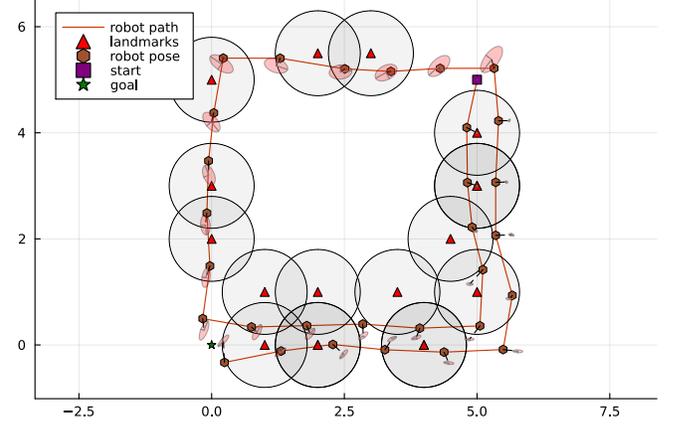
Following the previous discussion, we continue with the experiments. We start from our first problem (7) (optimality under a probabilistic constraint) and study Alg. 1 versus Alg. 2. Noticeable, in Alg. 2 we do not have a mechanism for early

**TABLE I:** Optimality under probabilistic constraint. Here we set 300 observation laces per path. Each quantity was averaged over 5 trials.

	$\epsilon$	$\delta$	$\mathcal{P}^*$	$\hat{U}_{(m)}^*$	N°discarded paths	time [sec] $\pm$ std	speedup (44)	laces frac. (45)	total laces	N°paths	N°land.
Alg. 2	-	0.0	14	$36.98 \cdot 10^{-5}$	-	$1171.21 \pm 74.48$		0	9000/9000	30	4
Alg. 1	0.023	0.0	no feasible	-	30	$77.67 \pm 4.01$	0.934	0.95	459/9000		
Alg. 1	0.3	0.0	14	$36.98 \cdot 10^{-5}$	29	$489.44 \pm 26.46$	0.58	0.60	3559/9000		
Alg. 1	0.5	0.0	14	$36.98 \cdot 10^{-5}$	29	$813.29 \pm 34.27$	0.31	0.37	5685/9000		
Alg. 1	0.7	0.0	14	$36.98 \cdot 10^{-5}$	27	$974.18 \pm 51.14$	0.17	0.134	7794/9000		
Alg. 1	0.8	0.0	14	$36.98 \cdot 10^{-5}$	23	$1099.98 \pm 45.41$	0.06	0.029	8738/9000		
Alg. 1	0.9	0.0	14	$36.98 \cdot 10^{-5}$	0	$1130.77 \pm 56.47$	0.03	0.0	9000/9000		



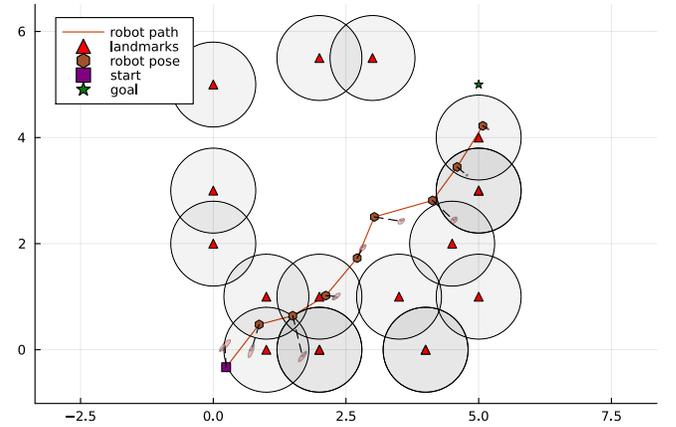
**Fig. 7:** Alg. 2 and Alg. 1 both selected path number 14 from Fig. 5b as optimal. We recognize that a pair of landmarks nearest to starting position (5,5) of preliminary mapping session in Fig. 6 greatly contribute to uncertainty diminishment since the robot twice made a loopclosure there.



**Fig. 8:** Robot second preliminary mapping session, by transparent gray circles we depict landmarks' visibility radius. As we can see that robot detected the landmarks and incorporated to its state. The landmarks visibility radius is 0.8.

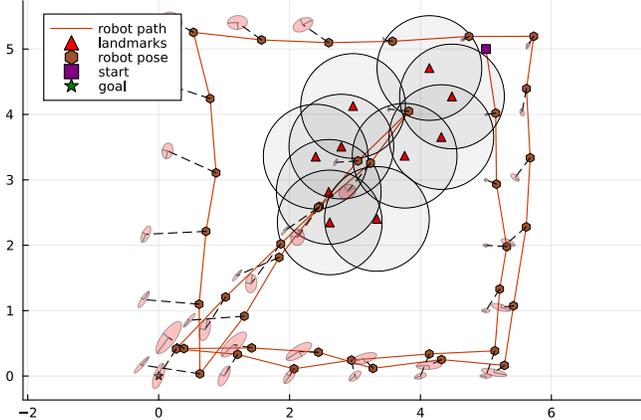
action dismissing until we expand all the observation laces per action sequence. We examine a simple scenario with four landmarks. We show the preliminary mapping session in Fig. 6. We elicit that, as anticipated, the uncertainty over the belief grows until the robot makes a full square and starts to experience loop closures. The path number 14 is highly likely to be optimal from an information perspective since the belief is Gaussian, and this path lies closest to the landmarks. We employ Alg. 1 with  $m = 300$  laces per path from Fig. 5a,  $\delta = 0.0$  and various values of  $\epsilon$ . Our resolution in terms of  $\epsilon$  is  $\Delta\epsilon = 1/m$ . Empirically we found that for  $\epsilon \in [0, 0.023]$ , without dependency on  $m$  as expected, all the paths were discarded as unfeasible (7 from 300 laces given path 14 were violating the inner constraint). Meaning no path is present with the fraction of the laces larger than  $1 - 0.023$  fulfilling inner constraint.

We show a rigorous comparison versus Alg. 2 in Table. I. In Fig. 7 we display the robot following the identified optimal path. Note that with Alg. 1 we do not accelerate decision making when we cannot discard action sequences. We shall note that due to internal gtsam multi-threading, measuring the time speedup is a challenging task. To alleviate that we repeat each run in Table. I five times and report averaged running time and the speedup obtained from averaged value of the



**Fig. 9:** Alg. 4 and Alg. 3 both selected path number 4 from Fig. 5b as optimal.

running time. Remarkably, from the bottom line of Table I we observe that with extremely loose probabilistic constraint ( $\epsilon = 0.9$ ) we do not eliminate any action sequence but the running time is not larger than the baseline. This fact indicate that there is no overhead from adaptation.



**Fig. 10:** Illustration of the third preliminary mapping session with randomly drawn landmarks.

### B. Maximal Feasible Return

We continue to our second problem (maximal feasible return (10)). Let us increase the number of landmarks to obtain more good candidate paths for information gathering. We show our second preliminary mapping session in Fig. 8. Here we need many paths with nonnegative information gain to examine using Alg. 3 early acceptance as well and not only early invalidation as was done in previous section. Our baseline is Alg. 4, which calculates VaR in a straightforward way. We report results in Table. II.

We also have an additional simulation with randomly drawing landmarks. For gtsam stability purposes we add random landmarks uniformly on the square  $[2, 5] \times [2, 5]$ . We also slightly changed the preliminary action sequence (Fig. 10). Results are presented in Table. III. In our simulations we set the final precision to  $1 \cdot 10^{-6} \cdot \delta^{\max}$ . As we witness from Tables II and III we **always** obtain a significant speedup. However, early action elimination appears to be more prominent than early accept. We explain it as follows. It is more likely that will be paths violating the probabilistic constraint and we land at the scenario depicted in Fig. 3b. Conversely, it is less likely that many paths fulfill the probabilistic constraint and we will land at the scenario depicted in Fig. 3d.

### C. Technical details

We used 3 computers with the following characteristics:

- 1) 8 cores Intel(R) Xeon(R) CPU E5-1620 v4 working at 3.50GHz with 80 GB of RAM;
- 2) 8 cores Intel(R) Xeon(R) CPU E5-1620 v4 working at 3.50GHz with 64 GB of RAM;
- 3) 16 cores 11th Gen Intel(R) Core(TM) i9-11900K working 3.50GHz with 64 GB of RAM.

## VI. CONCLUSIONS

We presented a novel adaptive technique to evaluate probabilistic belief-dependent constraints. On top of that, we provably extended the simplification paradigm to our setting. Our rigorous theory is summarized by two novel adaptive

algorithms solving optimality under a probabilistic constraint problem and maximal feasible return problem correspondingly. Our algorithms return identical quality or more profitable solution in a fraction of the baseline running time. By a more profitable solution, we mean elimination of all candidate action sequences using Alg. 1, thereby preventing the robot from redundant actions when the robot is already deployed and operates online, e.g, stopping exploration. Extensive simulations show the superiority of our methods. In the exceptionally challenging problem of active SLAM with a high dimensional state, we obtained a typical speedup of 30%.

## APPENDIX A

### THEORETICAL OBSERVATION LIKELIHOOD

To express the observation in terms of probabilistic models available to our disposal we marginalize over the  $x_{t+1}$

$$\mathbb{P}(z_{t+1}|b_t, a_t, \beta_{t+1})\mathbb{P}(\beta_{t+1}|b_t, a_t) = \int_{\mathbf{x}_{t+1}} \mathbb{P}(z_{t+1}|b_t, a_t, \beta_{t+1}, \mathbf{x}_{t+1}) \cdot \quad (46)$$

$$\mathbb{P}(\mathbf{x}_{t+1}|b_t, a_t, \beta_{t+1})\mathbb{P}(\beta_{t+1}|b_t, a_t) d\mathbf{x}_{t+1} = \int_{\mathbf{x}_{t+1}} \mathbb{P}(z_{t+1}|b_t, a_t, \beta_{t+1}, \mathbf{x}_{t+1}) \cdot \quad (47)$$

$$\mathbb{P}(\mathbf{x}_{t+1}|b_t, a_t)\mathbb{P}(\beta_{t+1}|\mathbf{x}_{t+1}) d\mathbf{x}_{t+1}.$$

All quantities in the (47) are available for us. Such a representation enables us to draw the observations in look-ahead step  $t + 1$ .

## APPENDIX B

### SAMPLE APPROXIMATIONS

The core of our sample approximations is sequential sampling the observations  $z_{t+1}|b_t, a_t, \beta_{t+1}$  using previously sampled  $\beta_{t+1}|b_t, a_t$ . Following the theoretical derivation presented in Appendix A, we leverage the structure verified by (47) in the following way. First, we sample the last pose and the landmarks from the corresponding marginal of the belief. Since our belief is Gaussian, this operation does not introduce a problem. We just pull the appropriate portion of the covariance matrix and the mean value.

$$x_{t+1}^o \sim \underbrace{\mathbb{P}(x_{t+1}, \{\ell^j\}_{j=1}^{M(k)}|b_t, a_t)}_{\text{Gaussian}}. \quad (48)$$

Afterwards, we deterministically decide the configuration of visible landmarks using (2) and draw samples of the observation laces (18) from the observation model (4). Finally, the sample approximation of  $\mathcal{U}$  and  $\mathcal{C}(\cdot)$  are denoted by  $\hat{\mathcal{U}}^{(m)}$  and  $\hat{\mathcal{C}}^{(m)}$  respectively and calculated by sample means of  $\{s(b_{k:k+L}^l)\}_{l=1}^m$ . Similarly  $\widehat{\text{VaR}}_c^{(m)}$  is obtained by sample quantile.

## APPENDIX C

### PROOFS

#### A. Proof of Theorem 3.1 (Simplification machinery)

It is sufficient to show, that for every sample  $z_{k+1:k+L}^q$  holds

$$\underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) \leq c(b_{k:k+L}^q; \phi, \delta) \leq \bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta). \quad (49)$$

**TABLE II:** Solving maximum feasible return problem on top of 30 candidate paths (Fig. 5b) with scenario presented in Fig. 8. In this study the number of observation laces is 64 per path. We observe that the speedup is approximately as the fraction of expanded laces, as expected since it is a little overhead from the adaptation. The values of time are averaged over 10 trials. The speedup is calculated from mean planning time.

	$\epsilon$	$\mathcal{P}^*$	$\delta^*$	time [sec] $\pm$ std	speedup (44)	laces frac. (45)	laces evaluations	N°paths with VaR $_{\epsilon} > 0$	N°paths	N° landmarks
Alg. 4	0.3	4	$9.71 \cdot 10^{-6}$	$2568.97 \pm 6.32$	-	0	1920/1920	18	30	18
Alg. 3		4	$8.93 \cdot 10^{-6}$	$1515.44 \pm 27.35$	0.41	0.57	830/1920			
Alg. 4	0.5	5	$1.06 \cdot 10^{-5}$	$3233.26 \pm 65.59$	-		1920/1920	19		
Alg. 3		5	$1.09 \cdot 10^{-5}$	$2201.36 \pm 38.93$	0.32	0.35	1250/1920			
Alg. 4	0.7	5	$1.73 \cdot 10^{-5}$	$3299.26 \pm 71.26$			1920/1920	20		
Alg. 3		5	$1.59 \cdot 10^{-5}$	$2433.06 \pm 46.19$	0.26	0.27	1411/1920			

**TABLE III:** Analysis of the behavior with randomly drawn landmarks. Each trial we randomly drawn 10 landmarks in the square  $[2, 5] \times [2, 5]$ . Here the visibility radius of the landmarks is 0.8. Note that mean time based speedup and the accumulated time based speedup are identical since the difference in running time in two possibilities is only the division by the number of trials.

N° paths	30	30	30	30	30
accumulated time based speedup	0.35	0.26	0.21	0.14	0.055
min speedup	0.14	0.092	0.14	0.08	0.019
max speedup	0.57	0.43	0.32	0.22	0.081
mean time based speedup	0.35	0.26	0.21	0.14	0.055
mean time [sec] $\pm$ std Alg. 4	$817.79 \pm 132.12$	$771.59 \pm 119.82$	$1670.69 \pm 260.39$	$1607.60 \pm 248.46$	$1671.69 \pm 259.52$
mean time [sec] $\pm$ std Alg. 3	$530.47 \pm 162.30$	$574.55 \pm 130.11$	$1320.71 \pm 216.86$	$1382.89 \pm 245.19$	$1580.00 \pm 230.63$
accumulated time [sec] Alg. 4	8177.92	7715.88	16706.90	16075.98	16716.86
accumulated time [sec] Alg. 3	5304.69	5745.51	13207.11	13828.99	15800.04
accumulated skipped laces frac.	0.36	0.29	0.23	0.15	0.065
accumulated expanded laces Alg. 3	12285	13689	14800	16304	17944
total N° of laces	19200	19200	19200	19200	19200
N°trials	10	10	10	10	10
N° landmarks	10	10	10	10	10
$\epsilon$	0.2	0.3	0.5	0.7	0.9
$\delta^{\min}$	0.0	0.0	0.0	0.0	0.0

We start from the left inequality of (49). Assume that  $\underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) = 0$ . This implies that trivially  $\underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) \leq c(b_{k:k+L}^q; \phi, \delta)$ . Now suppose that  $\underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) = 1$ . If the inner constraint of the form (8) so

$$\left( \sum_{t=k}^{k+L-1} \phi(b_{t+1}^q, b_t^q) \right) \geq \left( \sum_{t=k}^{k+L-1} \underline{\phi}(b_{t+1}^q, b_t^q) \right) > \delta. \quad (50)$$

We have that  $c(b_{k:k+L}^q; \phi, \delta) = 1$ . If the inner constraint of the second form (9) so  $\forall t$

$$\phi(b_{t+1}^q, b_t^q) \geq \underline{\phi}(b_{t+1}^q, b_t^q) > \delta. \quad (51)$$

Again we arrive at  $c(b_{k:k+L}^q; \phi, \delta) = 1$ . To prove the inverse direction, observe that if  $c(b_{k:k+L}^q; \phi, \delta) = 0$ , we behold the following situation with the first form

$$\delta \geq \left( \sum_{t=k}^{k+L-1} \phi(b_{t+1}^q, b_t^q) \right) \geq \left( \sum_{t=k}^{k+L-1} \underline{\phi}(b_{t+1}^q, b_t^q) \right). \quad (52)$$

and with the second form  $\exists t$  such that

$$\delta \geq \phi(b_{t+1}^q, b_t^q) \geq \underline{\phi}(b_{t+1}^q, b_t^q). \quad (53)$$

It follows that for both forms we have that  $\underline{c}(b_{k:k+L}^q; \underline{\phi}, \delta) \leq c(b_{k:k+L}^q; \phi, \delta)$ .

Now we prove the right inequality of (49). If  $\bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta) = 1$  it is trivial. Assume that  $\bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta) = 0$ . For the first form (8) it means that

$$\left( \sum_{t=k}^{k+L-1} \phi(b_{t+1}^q, b_t^q) \right) \leq \left( \sum_{t=k}^{k+L-1} \bar{\phi}(b_{t+1}^q, b_t^q) \right) \leq \delta. \quad (54)$$

Subsequently, for the second form (9) it means that  $\exists t$  such that

$$\phi(b_{t+1}^q, b_t^q) \leq \bar{\phi}(b_{t+1}^q, b_t^q) \leq \delta. \quad (55)$$

Arguing in the similar manner as with the lower bound for inverse direction, suppose that  $c(b_{k:k+L}^q; \phi, \delta) = 1$ . We obtain

$$\delta < \left( \sum_{t=k}^{k+L-1} \phi(b_{t+1}^q, b_t^q) \right) \leq \left( \sum_{t=k}^{k+L-1} \bar{\phi}(b_{t+1}^q, b_t^q) \right), \quad (56)$$

and  $\forall t$

$$\delta < \phi(b_{t+1}^q, b_t^q) \leq \bar{\phi}(b_{t+1}^q, b_t^q). \quad (57)$$

For both forms we have that  $c(b_{k:k+L}^q; \phi, \delta) \leq \bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta)$ .

This concludes the proof. Note that we also land at a identical result for theoretical counterparts of following probabilities and not sample approximations by taking the limit.

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{q=1}^m c(b_{k:k+L}^q; \phi, \delta) \leq \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{q=1}^m \bar{c}(b_{k:k+L}^q; \phi, \delta) \quad (58)$$

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{q=1}^m c(b_{k:k+L}^q; \phi, \delta) \leq \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{q=1}^m \bar{c}(b_{k:k+L}^q; \bar{\phi}, \delta). \quad (59)$$

■

## REFERENCES

- [1] G. A. Hollinger and G. S. Sukhatme, “Sampling-based robotic information gathering algorithms,” *Intl. J. of Robotics Research*, pp. 1271–1287, 2014.
- [2] J. A. Placed, J. Strader, H. Carrillo, N. Atanasov, V. Indelman, L. Carlone, and J. A. Castellanos, “A survey on active simultaneous localization and mapping: State of the art and new frontiers,” *arXiv preprint arXiv:2207.00254*, 2022.
- [3] D. Kopitkov and V. Indelman, “No belief propagation required: Belief space planning in high-dimensional state spaces via factor graphs, matrix determinant lemma and re-use of calculation,” *Intl. J. of Robotics Research*, vol. 36, no. 10, pp. 1088–1130, August 2017.
- [4] 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*, vol. 31, no. 11, pp. 1263–1278, 2012.
- [5] V. Indelman, L. Carlone, and F. Dellaert, “Planning in the continuous domain: a generalized belief space approach for autonomous navigation in unknown environments,” *Intl. J. of Robotics Research*, vol. 34, no. 7, pp. 849–882, 2015.
- [6] F. Dellaert and M. Kaess, “Factor graphs for robot perception,” *Foundations and Trends in Robotics*, vol. 6, no. 1-2, pp. 1–139, 2017.
- [7] R. Platt, R. Tedrake, L. Kaelbling, and T. Lozano-Pérez, “Belief space planning assuming maximum likelihood observations,” in *Robotics: Science and Systems (RSS)*, Zaragoza, Spain, 2010, pp. 587–593.
- [8] E. Farhi and V. Indelman, “ix-bsp: Incremental belief space planning,” <https://arxiv.org/abs/2102.09539>, 2021.
- [9] E. I. Farhi and V. Indelman, “ix-bsp: Belief space planning through incremental expectation,” in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2019.
- [10] M. Araya-López, O. Buffet, V. Thomas, and F. o. Charpillet, “A pomdp extension with belief-dependent rewards,” in *NIPS*, 2010, pp. 64–72.
- [11] M. Fehr, O. Buffet, V. Thomas, and J. Dibangoye, “rho-pomdps have lipschitz-continuous epsilon-optimal value functions,” in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Curran Associates, Inc., 2018, pp. 6933–6943. [Online]. Available: <http://papers.nips.cc/paper/7925-rho-pomdps-have-lipschitz-continuous-epsilon-optimal-value-functions.pdf>
- [12] L. Dressel and M. J. Kochenderfer, “Efficient decision-theoretic target localization,” in *Proceedings of the Twenty-Seventh International Conference on Automated Planning and Scheduling, ICAPS 2017, Pittsburgh, Pennsylvania, USA, June 18-23, 2017*, L. Barbulescu, J. Frank, Mausam, and S. F. Smith, Eds. AAAI Press, 2017, pp. 70–78. [Online]. Available: <https://aaai.org/ocs/index.php/ICAPS/ICAPS17/paper/view/15761>
- [13] Z. Sunberg and M. Kochenderfer, “Online algorithms for pomdps with continuous state, action, and observation spaces,” in *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 28, no. 1, 2018.
- [14] O. Szttyglic, A. Zhitnikov, and V. Indelman, “Simplified belief-dependent reward mcts planning with guaranteed tree consistency,” *arXiv preprint arXiv:2105.14239*, 2021.
- [15] A. Zhitnikov and V. Indelman, “Simplified risk aware decision making with belief dependent rewards in partially observable domains,” *Artificial Intelligence, Special Issue on “Risk-Aware Autonomous Systems: Theory and Practice”*, 2022.
- [16] O. Szttyglic and V. Indelman, “Speeding up online pomdp planning via simplification,” in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022.
- [17] K. Elimelech and V. Indelman, “Simplified decision making in the belief space using belief sparsification,” *Intl. J. of Robotics Research*, vol. 41, no. 5, pp. 470–496, 2022.
- [18] A. Kitanov and V. Indelman, “Topological information-theoretic belief space planning with optimality guarantees,” *arXiv preprint arXiv:1903.00927*, 3 2019.
- [19] M. Shienman and V. Indelman, “D2a-bsp: Distilled data association belief space planning with performance guarantees under budget constraints,” in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2022.
- [20] —, “Nonmyopic distilled data association belief space planning under budget constraints,” in *Proc. of the Intl. Symp. of Robotics Research (ISRR)*, 2022.
- [21] M. Barenboim and V. Indelman, “Adaptive information belief space planning,” in the *31st International Joint Conference on Artificial Intelligence and the 25th European Conference on Artificial Intelligence (IJCAI-ECAI)*, July 2022.
- [22] P. Santana, S. Thiébaux, and B. Williams, “Rao\*: An algorithm for chance-constrained pomdp’s,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.
- [23] A. Zhitnikov and V. Indelman, “Risk aware belief-dependent constrained pomdp planning,” *arXiv preprint arXiv:2209.02679*, 2022.
- [24] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. D. Reid, and J. J. Leonard, “Simultaneous localization and mapping: Present, future, and the robust-perception age,” *arXiv preprint arXiv:1606.05830*, 2016.
- [25] J. A. Placed and J. A. Castellanos, “Enough is enough: Towards autonomous uncertainty-driven stopping criteria,” *arXiv preprint arXiv:2204.10631*, 2022.
- [26] G. C. Pflug and A. Pichler, “Time-consistent decisions and temporal decomposition of coherent risk functionals,” *Mathematics of Operations Research*, vol. 41, no. 2, pp. 682–699, 2016.
- [27] J. Mullane, B.-N. Vo, M. D. Adams, and B.-T. Vo, “A random-finite-set approach to bayesian slam,” *IEEE transactions on robotics*, vol. 27, no. 2, pp. 268–282, 2011.
- [28] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.
- [29] S. Pathak, A. Thomas, and V. Indelman, “A unified framework for data association aware robust belief space planning and perception,” *Intl. J. of Robotics Research*, vol. 32, no. 2-3, pp. 287–315, 2018.
- [30] V. Tchuiev, Y. Feldman, and V. Indelman, “Data association aware semantic mapping and localization via a viewpoint-dependent classifier model,” in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019.
- [31] F. Dellaert, “Factor graphs and GTSAM: A hands-on introduction,” Georgia Institute of Technology, Tech. Rep. GT-RIM-CP&R-2012-002, September 2012.
- [32] V. Indelman, “Cooperative multi-robot belief space planning for autonomous navigation in unknown environments,” *Autonomous Robots*, pp. 1–21, 2017.
- [33] L. Kavraki, P. Svestka, J.-C. Latombe, and M. Overmars, “Probabilistic roadmaps for path planning in high-dimensional configuration spaces,” *IEEE Trans. Robot. Automat.*, vol. 12, no. 4, pp. 566–580, 1996.
- [34] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, “iSAM2: Incremental smoothing and mapping using the Bayes tree,” *Intl. J. of Robotics Research*, vol. 31, no. 2, pp. 217–236, Feb 2012.