Cost-Effective Sensing for Goal Inference: A Model Predictive Approach

Ran Tian, Nan Li, Anouck Girard, Ilya Kolmanovsky, and Masayoshi Tomizuka

***Abstract*— Goal inference is of great importance for a variety of applications that involve interaction, coordination, and/or competition with goal-oriented agents. Typical goal inference approaches use as many pointwise measurements of the agent's trajectory as possible to pursue a most accurate a-posteriori estimate of the goal. However, taking frequent measurements may not be preferred in situations where sensing is associated with high cost (e.g., sensing + perception may involve high computational/bandwidth cost and sensing may raise security concerns in privacy-critical/data-sensitive applications). In such situations, a sensible tradeoff between the information gained from measurements and the cost associated with sensing actions is highly desirable. This paper introduces a cost-effective sensing strategy for goal inference tasks based on hybrid Kalman filtering and model predictive control. Our key insights include: 1) a model predictive approach can be used to predict the amount of information gained from new measurements over a horizon and thus to optimize the tradeoff between information gain and sensing action cost, and 2) the high computational efficiency of hybrid Kalman filtering can ensure real-time feasibility of such a model predictive approach. We evaluate the proposed cost-effective sensing approach in a goal-oriented task, where we show that compared to standard goal inference approaches, our approach takes a considerably reduced number of measurements while not impairing the speed, accuracy, and reliability of goal inference by taking measurements smartly.**

I. Introduction

Goal inference is an important problem in a variety of applications that involve interaction, coordination, and/or competition with goal-oriented agents (e.g., humans). For instance, in human-robot collaboration or shared-autonomy tasks where the a-priori uncertain intention of human can be represented as one of multiple goals, fast and accurate estimation of human's goal can inform robot's behavior and trajectory planning to improve the overall performance of human-robot interaction [1]–[3].

Bayesian inference-based approaches [4]–[7] and data-driven approaches [1], [8] have been widely considered for goal inference problems. These approaches typically use as many pointwise measurements of the human's/robot's trajectory as possible, e.g., a measurement at every sample time instant, to pursue a most accurate a-posteriori estimate of the goal. However, for certain circumstances, taking measurements at a high frequency can be costly or infeasible: *1) Computational cost.* When the primary sensor is a camera or radar/Lidar, taking a measurement involves running a computer vision algorithm or sampling and analyzing a point cloud online, which generally entails intensive computations without advanced hardware [9]. *2) Assets or material usage cost.* In space situational awareness, a finite number of assets are used to track multiple orbital objects, so there is a cost for pointing them in certain directions. In agricultural applications, taking measurements (e.g., tracking herd locations or collecting soil samples) involves manpower/unmanned system (e.g., quadcopter) employment and/or material consumption, which can be costly. *3) Time or bandwidth resource cost.* For systems such as those involving both magnetic sensors and magnetic actuators for which sensing and control interfere with each other and thus must operate in a disjunctive manner, sensing takes the total time resource for both sensing and control [10]. In such situations, a sensible trade-off between the information gained from new measurements and the cost associated with sensing actions is highly desirable.

Ran Tian and Masayoshi Tomizuka are with the UC Berkeley. {rantian,tomizuka}@berkeley.edu.

Nan Li, Anouck Girard, and Ilya Kolmanovsky are with the University of Michigan, Ann Arbor. {nanli,anouck,ilya}@umich.edu.

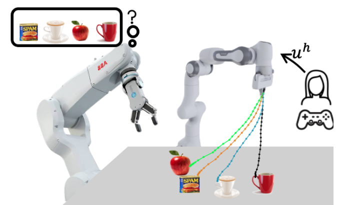

Fig. 1: Human-robot collaboration: an autonomous robot (left) collaborates with a dexterous robotic manipulator teleoperated by a human (right). The robot infers the goal of the human operator to inform motion planning and avoid potential conflicts.

In this paper, we consider cost-effective sensing in a goal inference setting. Here, "cost-effective" means taking as few measurements as possible while not significantly impairing the speed and accuracy of goal inference. Our key insights include: 1) a model predictive approach can be used to predict the amount of information gained from new measurements over a horizon and thus to optimize the trade-off between information gain and sensing action cost, and 2) the high computational efficiency of hybrid Kalman filtering can ensure real-time feasibility of such a model predictive approach. Specifically, with these insights, we first cast the goal inference problem as a hybrid estimation problem and extend a hybrid Kalman filtering algorithm for this estimation problem. We then exploit the model predictive control framework to determine a cost-effective sensing strategy for the extended hybrid Kalman filter.

We evaluate our cost-effective sensing approach for goal inference in a goal-oriented task (see Fig. 1), where we show that compared to baseline goal inference methods, our approach takes a considerably reduced number of measurements while not impairing the speed, accuracy, and reliability of goal inference by taking measurements in a smart way.

II. Preliminaries

A. Dynamical model

We consider a general system (e.g., a human body, a robot, etc.) the dynamics of which can be described by a model as follows,

$$x_{t+1} = f_o(x_t, u_t), \tag{1}$$

where the subscript $t \in \mathbb{Z}_{\geq 0}$ represents the discrete time instant, $x_t \in \mathbb{R}^{n_x}$ represents the state of the system (e.g., generalized positions, velocities, etc.), $u_t \in \mathbb{R}^{n_u}$ represents an action input (e.g., generalized forces, etc.), and f_o is a nonlinear function which is continuously differentiable (i.e., $\mathcal{C}^1$) with respect to x_t and u_t. We consider a sensing system (sensor + perception algorithm) that, when triggered, can (partially or indirectly) measure the state of (1) as follows,

$$y_t = g(x_t), \tag{2}$$

where $y_t \in \mathbb{R}^{n_y}$ represents a (partial or indirect) measurement of x_t, and g is a nonlinear $\mathcal{C}^1$ function.

We assume that this system is controlled by a human or a higher-level planning algorithm with the control task represented by a goal, $\hat{\theta}$, which belongs to a specified finite set of goals,

$$\Theta = \{\theta^1, \theta^2, \dots, \theta^{n_\theta}\}, \tag{3}$$

however, which element of Θ corresponds to the current actual goal $\hat{\theta}$ is a-priori uncertain. Our objective is to infer $\hat{\theta}$ through measurements y_t assuming that the goal affects the action input as further detailed below.

B. Goal-corresponding policies

We assume that for each goal $\theta \in \Theta$, the corresponding average/nominal human's or higher-level planning algorithm's behavior can be approximated by a (possibly time-dependent) control policy $\pi_t^\theta : \mathbb{R}^{n_x} \to \mathbb{R}^{n_u}$, which maps the current state of the system x_t to an action u_t. Such a goal-corresponding policy π_t^θ can be obtained via various strategies. For instance, when pre-collected demonstrations corresponding to each goal θ are available, a policy π_t^θ that most likely explains the behavior of these trajectories may be learned using goal-conditioned imitation learning [11]. Alternatively, π_t^θ may be constructed as an optimal policy maximizing a goal-corresponding return [12], e.g.,

$$\pi_t^\theta \in \arg\max_{\pi_t} \sum_{t=0}^{N} \gamma^t l^\theta(x_t, \pi_t(x_t)), \tag{4a}$$

$$\text{s.t. } x_{t+1} = f_o(x_t, \pi_t(x_t)), \tag{4b}$$

where $N \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$ is a horizon length, $\gamma \in (0, 1]$ is a discount factor, and l^θ is a goal-corresponding reward function which may be learned from demonstrations using, e.g., inverse reinforcement learning [13]. With the policies π_t^θ, the evolution of system state x_t under a specific goal $\theta \in \Theta$ can be represented by the following closed-loop model:

$$x_{t+1} = f_t^\theta(x_t) = f_o(x_t, \pi_t^\theta(x_t)). \tag{5}$$

Here, "closed-loop" means that the above model represents the dynamics of the overall system consisting of (1) and the human/higher-level planning algorithm with the goal θ.

Remark 1: In some cases, for a goal θ its corresponding policy π_t^θ may not be unique. For instance, if the policies are obtained from a data-driven approach, there could exist multiple policies that explain a given set of demonstrations equally well. In addition, if the policies are obtained from model-based optimizations, non-uniqueness of π_t^θ may be due to non-uniqueness of optimal policies for a given reward function or due to multiple reward functions for a given goal to represent different behavioral preferences. In our goal inference approach based on hybrid Kalman filtering that will be described in what follows, such non-uniqueness of π_t^θ can be handled by treating them as separate goals.

III. Cost-Effective Sensing for Goal Inference

In this section, we cast the goal inference problem as a hybrid estimation problem, and present our model predictive approach for cost-effective sensing.

A. Goal inference using a hybrid Kalman filter

In this paper, we adopt a hybrid Kalman filtering algorithm for the task of goal inference. The original algorithm was developed in [14] for discrete-time linear hybrid systems. We extend the algorithm of [14] to nonlinear models following the general idea of the extended Kalman filter (EKF) [15]. We choose to adopt such a Kalman filter-based approach because it achieves sufficiently high inference accuracy, as will be shown in our experiments in Section V, while has low computational complexity, making it suitable for online tasks. We now present the algorithm as follows.

At each discrete time instant t, we consider the following set of linear models with disturbances:

$$\delta x_{t+1}^\theta = A_t^\theta \delta x_t^\theta + w_t^\theta, \tag{6a}$$

$$\delta y_t^\theta = C_t^\theta \delta x_t^\theta + v_t^\theta, \tag{6b}$$

for each possible goal $\theta \in \Theta$, where δx_t^θ represents the deviation of the state x_t from a nominal state value corresponding to the goal θ and the time t, $x_t^{\theta,\text{nom}}$ (i.e., $\delta x_t^\theta = x_t - x_t^{\theta,\text{nom}}$), and δy_t^θ represents the deviation of the output y_t from the nominal output value $y_t^{\theta,\text{nom}} = g(x_t^{\theta,\text{nom}})$. The matrices A_t^θ and C_t^θ represent local linear approximations of the functions f_t^θ in (5) and g in (2), respectively. The disturbances w_t^θ and v_t^θ account for 1) linearization errors, 2) deviations of the nominal policy π_t^θ from actual human/higher-level planning algorithm behavior, which may be due to human sub-optimality and variability or errors of the assumed reward function in (4) from the actual one of the planning algorithm, and 3) other disturbance inputs and measurement noises.

There are generally two strategies to determine the linearization matrices A_t^θ and C_t^θ. The first strategy is to set them as the Jacobian matrices of f_t^θ and g at the current state estimate $\hat{x}_t$. This strategy is frequently adopted in standard EKF. The second strategy is to consider a nominal/reference trajectory corresponding to each goal $\theta \in \Theta$, $\xi^{\theta,\text{nom}} = \{x_0^{\theta,\text{nom}}, \dots, x_N^{\theta,\text{nom}}\}$, and set A_t^θ and C_t^θ as the Jacobian matrices of f_t^θ and g along this nominal trajectory, i.e.,

$$A_t^\theta = \left.\frac{\partial f_t^\theta}{\partial x}\right|_{x_t^{\theta,\text{nom}}}, \qquad C_t^\theta = \left.\frac{\partial g}{\partial x}\right|_{x_t^{\theta,\text{nom}}}. \tag{7}$$

In this paper, we adopt this second strategy. The nominal trajectory $\xi^{\theta,\text{nom}}$ may be learned as the average trajectory from a pre-collected set of trajectory data corresponding to each goal $\theta \in \Theta$. Alternatively, given initial condition x_0, one may simulate the nominal system (5) and treat the obtained trajectory as the nominal trajectory $\xi^{\theta,\text{nom}}$. The reason for us to adopt this second strategy is as follows: Using the first strategy, the computation of the linearization matrices $A_t^\theta = \frac{\partial f_t^\theta}{\partial x}\big|_{\hat{x}_t}$ and $C_t^\theta = \frac{\partial g}{\partial x}\big|_{\hat{x}_t}$ and the filtering process must be performed in an alternating serial manner [16]. In contrast, using the second strategy, once the nominal trajectory $\xi^{\theta,\text{nom}}$ is given, the matrices A_t^θ and C_t^θ along this trajectory can either be pre-computed offline and stored for online use or be computed in a parallel manner with the filtering process, because the A_t^θ and C_t^θ values do not depend on the filtering results. Note that the nominal policies π_t^θ are frequently represented using functional approximators, such as neural networks, and the output function g in (2) may be determined by a perception algorithm. In such cases, the Jacobian matrices of f_t^θ and g do not admit explicit functional forms and must be evaluated numerically, e.g., using a finite-difference method, which entails non-negligible computational effort. Therefore, the second strategy, which enables pre- or parallel computation of A_t^θ and C_t^θ, improves online implementation feasibility.

For the disturbances w_t^θ and v_t^θ, we model them as Gaussian random variables with zero mean and known covariances W^θ and V^θ, i.e., $w_t^\theta \sim \mathcal{N}(0, W^\theta)$ and $v_t^\theta \sim \mathcal{N}(0, V^\theta)$. The covariance matrices W^θ and V^θ may be estimated from trajectory data or manually tuned to achieve satisfactory inference performance. In principle, the covariances can be time-varying. We have chosen to consider time-invariant covariances to relax reliance on big data and tuning effort.

The hybrid Kalman filtering algorithm uses a set of Kalman filters matched to the different modes of the hybrid system to calculate a set of estimates of the continuous state x_t, each corresponding to the premise that the system is in a specific mode. Meanwhile, the algorithm estimates the probabilities of the modes according to the measurement residuals of these Kalman filters (i.e., the errors between the actual measurement and the output values corresponding to their state estimates). In our setting, modes of the system correspond to the goals $\theta \in \Theta$, and thereby mode estimation corresponds to goal inference. Specifically, the hybrid Kalman filtering algorithm applied to our goal inference problem is represented by the following equations:

State prediction: $$\hat{x}_{t|t-1}^\theta = f_{t-1}^\theta(\hat{x}_{t-1|t-1}^\theta), \tag{8a}$$

Covariance prediction:
$$\Sigma_{t|t-1}^\theta = A_{t-1}^\theta \Sigma_{t-1|t-1}^\theta (A_{t-1}^\theta)^\top + W^\theta, \tag{8b}$$

Measurement residual: $$z_t^\theta = y_t - g(\hat{x}_{t|t-1}^\theta), \tag{8c}$$

Residual covariance: $$\Xi_t^\theta = C_t^\theta \Sigma_{t|t-1}^\theta (C_t^\theta)^\top + V^\theta, \tag{8d}$$

Kalman gain: $$L_t^\theta = \Sigma_{t|t-1}^\theta (C_t^\theta)^\top (\Xi_t^\theta)^{-1}, \tag{8e}$$

State estimate: $$\hat{x}_{t|t}^\theta = \hat{x}_{t|t-1}^\theta + L_t^\theta z_t^\theta, \tag{8f}$$

Covariance estimate: $$\Sigma_{t|t}^\theta = (I - L_t^\theta C_t^\theta)\Sigma_{t|t-1}^\theta. \tag{8g}$$

Finally, the algorithm calculates the a-posteriori belief in each goal $\theta \in \Theta$ conditioned on all measurements up to time t, $\mathbf{y}_t = \{y_0, \ldots, y_t\}$, according to:

$$b_{t|t}(\theta) = \mathbb{P}(\hat{\theta}_t = \theta \,|\, \mathbf{y}_t) = \frac{1}{c_t}\, \Lambda_t^\theta(z_t^\theta)\, b_{t|t-1}(\theta), \tag{9}$$

where $\Lambda_t^\theta(z_t^\theta)$ is the likelihood function of goal θ at time t, $\Lambda_t^\theta(\cdot) = \mathsf{pdf}(\cdot \,|\, 0, \Xi_t^\theta)$, evaluated at the residual z_t^θ, with $\mathsf{pdf}(\cdot \,|\, 0, \Xi_t^\theta)$ denoting the probability density function of a Gaussian distribution with mean 0 and covariance Ξ_t^θ, $b_{t|t-1}(\theta)$ is a probabilistic prediction of the goal at time t before taking into account the newest measurement y_t, and $c_t = \sum_{\theta' \in \Theta} \Lambda_t^{\theta'}(z_t^{\theta'})\, b_{t|t-1}(\theta')$ is a normalization factor.

Remark 2: When taking into the consideration that the agent may switch goals during the task, the transitions between goals can be modeled as a Markov process with transition probabilities $\pi_{ij} = \mathbb{P}(\hat{\theta}_t = \theta_j \,|\, \hat{\theta}_{t-1} = \theta_i)$ and the goal prediction $b_{t|t-1}(\theta)$ is calculated according to

$$b_{t|t-1}(\theta_j) = \sum_{i=1}^{n_\theta} \left(b_{t-1|t-1}(\theta_i)\, \pi_{ij}\right). \tag{10}$$

In this paper, we assume the agent has a constant goal, i.e., $\hat{\theta}_t \equiv \hat{\theta}$. In this case, we have $b_{t|t-1}(\theta) = b_{t-1|t-1}(\theta)$ for all $\theta \in \Theta$ and (9) reduces to

$$b_{t|t}(\theta) = c_t^{-1}\, \Lambda_t^\theta(z_t^\theta)\, b_{t-1|t-1}(\theta). \tag{11}$$

B. Cost-effective sensing

We use $u_t = 1$ to denote a sensing action (i.e., to take a measurement) at time t, and use $u_t = 0$ to denote not to take a measurement at t. Note that the equations (8) and (11) correspond to the hybrid Kalman filtering algorithm where a measurement y_t is taken at every time instant t. To account for time instants where measurements are not taken (i.e., with $u_t = 0$), we modify the algorithm as follows:

$$\hat{x}_{t|t}^\theta = \begin{cases} \hat{x}_{t|t-1}^\theta + L_t^\theta(y_t - g(\hat{x}_{t|t-1}^\theta)), & \text{if } u_t = 1, \\ \hat{x}_{t|t-1}^\theta, & \text{if } u_t = 0, \end{cases} \tag{12a}$$

$$\Sigma_{t|t}^\theta = \begin{cases} (I - L_t^\theta C_t^\theta)\Sigma_{t|t-1}^\theta, & \text{if } u_t = 1, \\ \Sigma_{t|t-1}^\theta, & \text{if } u_t = 0, \end{cases} \tag{12b}$$

$$b_{t|t}(\theta) = \begin{cases} \frac{1}{c_t}\Lambda_t^\theta(y_t - g(\hat{x}_{t|t-1}^\theta))\, b_{t-1|t-1}(\theta), & \text{if } u_t = 1, \\ b_{t-1|t-1}(\theta), & \text{if } u_t = 0, \end{cases} \tag{12c}$$

where $\hat{x}_{t|t-1}^\theta = f_{t-1}^\theta(\hat{x}_{t-1|t-1}^\theta)$ and $\Sigma_{t|t-1}^\theta = A_{t-1}^\theta \Sigma_{t-1|t-1}^\theta (A_{t-1}^\theta)^\top + W^\theta$ are the a-priori state and covariance estimates, respectively, and L_t^θ is the Kalman gain computed as

$$L_t^\theta = \Sigma_{t|t-1}^\theta (C_t^\theta)^\top (C_t^\theta \Sigma_{t|t-1}^\theta (C_t^\theta)^\top + V^\theta)^{-1}. \tag{13}$$

The above modified algorithm resembles equations (8) and (11) for the case of taking a measurement ($u_t = 1$) and sets the "a-posteriori" estimates $\hat{x}_{t|t}^\theta$, $\Sigma_{t|t}^\theta$, and $b_{t|t}(\theta)$ to their a-priori values $\hat{x}_{t|t-1}^\theta$, $\Sigma_{t|t-1}^\theta$, and $b_{t-1|t-1}(\theta)$ if a measurement is not taken ($u_t = 0$).

We assume a constant cost associated with each sensing action, $w_a > 0$. Our objective is to minimize the sensing cost while ensuring a fast, accurate, and reliable goal inference. In particular, we use the information entropy of $b_{t|t}$ to quantify uncertainty in our goal inference,

$$H(b_{t|t}) = -\sum_{\theta \in \Theta} b_{t|t}(\theta) \log\left(b_{t|t}(\theta)\right). \tag{14}$$

Recall that $b_{t|t}$ is a probability distribution on the space of possible goals, Θ, describing our a-posteriori beliefs in each goal $\theta \in \Theta$. In general, a higher entropy $H(b_{t|t})$ represents a higher degree of uncertainty in the goal inference [17]. For instance, when we are 100% certain about the goal (i.e., $b_{t|t}(\theta') = 1$ for some $\theta' \in \Theta$), the entropy $H(b_{t|t})$ attains its minimum value 0. In contrast, when we have equal beliefs in every goal (i.e., $b_{t|t}$ is a uniform distribution on Θ), $H(b_{t|t})$ is maximized. Therefore, minimizing uncertainty in the goal inference can be expressed as minimizing the entropy $H(b_{t|t})$. In particular, to optimize a trade-off between inference uncertainty and sensing cost, we consider the following objective function

$$\varphi_t(u_t) = w_h H_{u_t}(b_{t|t}) + w_a u_t, \tag{15}$$

where $w_h > 0$ is a weighting factor and the subscript u_t of $H(b_{t|t})$ represents its dependence on the value of u_t, that is, we aim to minimize a weighted sum of information entropy and sensing action cost.

Remark 3: Considering the objective function (15) is equivalent to comparing the "information gain" (the change in information entropy from a-priori estimate to a-posteriori estimate after taking a measurement),

$$IG(b_{t|t}, u_t) = H_{u_t=0}(b_{t|t}) - H_{u_t=1}(b_{t|t}), \tag{16}$$

and the sensing cost w_a to determine whether it is worth taking a sensing action $u_t = 1$. When the weighted information gain $w_h IG(b_{t|t}, u_t)$ is higher than the sensing cost w_a, minimizing (15) leads to $u_t = 1$, and vice versa. Specifically, if $w_h IG(b_{t|t}, u_t) = w_h H_{u_t=0}(b_{t|t}) - w_h H_{u_t=1}(b_{t|t}) \geq w_a$, then $\varphi_t(u_t = 0) = w_h H_{u_t=0}(b_{t|t}) \geq w_h H_{u_t=1}(b_{t|t}) + w_a = \varphi_t(u_t = 1)$, i.e., $u_t = 1$ corresponds to a smaller value of the objective function (15).

Note that the belief distribution $b_{t|t}$, computed according to (12c), depends not only on the sensing decision u_t, but also on the received value of measurement, y_t, for $u_t = 1$. Consequently, the objective function (15) also depends on y_t. Recall that we model y_t as a random variable in (6b) to formulate our hybrid Kalman filtering-based algorithm above. In this case, before a measurement y_t is actually received, (15) can be considered as a random variable (related to y_t) the distribution of which depends on u_t.

C. Model predictive approach

Our proposed cost-effective sensing strategy for goal inference is based on the model predictive control framework. Specifically, at each time instant t, we solve the following optimization problem:

$$\min_{\mathbf{u}_t = \{u_{t|t}, \ldots, u_{t+N|t}\}} J(\mathbf{u}_t, b_{t-1|t-1}), \tag{17}$$

where the cost index J is

$$\begin{aligned} J(\mathbf{u}_t, b_{t-1|t-1}) &= \mathbb{E}\left[\sum_{\tau=0}^{N} \varphi_{t+\tau}(u_{t+\tau}) \middle| \mathbf{u}_t, b_{t-1|t-1}\right] \qquad (18) \\ &= w_h \mathbb{E}\left[\sum_{\tau=0}^{N} H(b_{t+\tau|t+\tau}) \middle| \mathbf{u}_t, b_{t-1|t-1}\right] + w_a \sum_{\tau=0}^{N} u_{t+\tau|t}. \end{aligned}$$

In (17) and (18), each $u_{t+\tau|t} \in \{0, 1\}$, $\tau = 0, \ldots, N$, is a predicted sensing decision for the future time $t+\tau$ (predicted at the current time t) and is an optimization variable, $b_{t-1|t-1}$ is the previous belief distribution and is a known constant at t, and $\mathbb{E}\big[\,\cdot\,\big|\mathbf{u}_t, b_{t-1|t-1}\big]$ denotes a conditional expectation conditioned on given values of $\mathbf{u}_t$ and $b_{t-1|t-1}$. We consider the expected value because, as discussed above, the objective function depends on the realized values of measurements $y_{t+\tau}$ over the prediction horizon, which have randomness. The solution to (17) is an optimal sequence of predicted sensing decisions, $\mathbf{u}_t^* = \{u_{t|t}^*, \ldots, u_{t+N|t}^*\}$. We then decide whether to take a measurement at the current time t according to the value of $u_{t|t}^*$, i.e., we determine the current sensing decision as $u_t = u_{t|t}^*$.

Exactly and analytically evaluating $\mathbb{E}\big[\sum_{\tau=0}^{N} H(b_{t+\tau|t+\tau}) \big|\mathbf{u}_t, b_{t-1|t-1}\big]$ involves integration of multivariate distribution and is computationally difficult. Therefore, we adopt a scenario method (also called Monte-Carlo method) to approximate the expected value. Specifically, for each $\theta \in \Theta$, we randomly generate n_s scenarios, $\big\{(x_t^{\theta,i}, \mathbf{w}_t^{\theta,i}, \mathbf{v}_t^{\theta,i})\big\}_{i=1}^{n_s}$. Each scenario $(x_t^{\theta,i}, \mathbf{w}_t^{\theta,i}, \mathbf{v}_t^{\theta,i})$ is a state $x_t^{\theta,i} \sim \mathcal{N}(\hat{x}_{t|t-1}^{\theta}, \Sigma_{t|t-1}^{\theta})$, a sequence of process disturbances $\mathbf{w}_t^{\theta,i} = \{w_t^{\theta,i}, \ldots, w_{t+N-1}^{\theta,i}\} \sim \mathcal{N}(0, W^\theta)$, and a sequence of measurement noises $\mathbf{v}_t^{\theta,i} = \{v_t^{\theta,i}, \ldots, v_{t+N}^{\theta,i}\} \sim \mathcal{N}(0, V^\theta)$. Each scenario determines the sequence of measurement values $\{y_t^{\theta,i}, \ldots, y_{t+N}^{\theta,i}\}$ through iteration of (6). Then, for any given sequence of sensing decisions $\mathbf{u}_t$, we compute a corresponding sequence of belief distributions $b_{t+\tau|t+\tau}$ using (12) and entropies $H(b_{t+\tau|t+\tau})$ using (14). Finally, we approximate the expected cumulative entropy according to

$$\begin{aligned} &\mathbb{E}\left[\sum_{\tau=0}^{N} H(b_{t+\tau|t+\tau}) \middle| \mathbf{u}_t, b_{t-1|t-1}\right] \approx \sum_{\theta \in \Theta} b_{t-1|t-1}(\theta) \cdot \\ &\cdot \left(\frac{1}{n_s} \sum_{i=1}^{n_s} \left[\sum_{\tau=0}^{N} H(b_{t+\tau|t+\tau}) \,|\, \mathbf{u}_t, (x_t^{\theta,i}, \mathbf{w}_t^{\theta,i}, \mathbf{v}_t^{\theta,i})\right]\right), \qquad (19) \end{aligned}$$

where $H(b_{t+\tau|t+\tau}) \,|\, \mathbf{u}_t, (x_t^{\theta,i}, \mathbf{w}_t^{\theta,i}, \mathbf{v}_t^{\theta,i})$ denotes the predicted entropy at time $t + \tau$ corresponding to decision sequence $\mathbf{u}_t$ and scenario $(x_t^{\theta,i}, \mathbf{w}_t^{\theta,i}, \mathbf{v}_t^{\theta,i})$, computed using the procedure described above.

Computational feasibility. For any given $\mathbf{u}_t$, evaluation of (18), with the first term approximated using the scenario method as (19), involves calculations that are fully vectorizable/parallelizable. Then, the optimization problem (17) reduces to evaluations of (18) for all decision candidates $\mathbf{u}_t \in \{0, 1\}^{N+1}$ and selection of the one corresponding to the smallest (18) value, which is real-time feasible.

IV. Experiment Design

A. Goal-oriented task

We evaluate the proposed cost-effective sensing approach for goal inference in a human-robot collaboration (HRC) task. We consider a setting where an autonomous robot collaborates with a robotic manipulator that is teleoperated by a human (see Fig. 1). We assume that both the autonomous robot and the human are tasked with grasping, inspecting, and moving objects on a table. In order to ensure effective collaboration, the autonomous robot needs to infer the human operator's goal (i.e., the object to grasp) to inform its motion planning and avoid potential conflicts. The considered robotic manipulator has 4 degrees of freedom, $x = (\theta_1, \ldots, \theta_4)$, where θ_i denotes the angle of the ith joint. The manipulator end-effector's position in the world frame, $y = (x_e, y_e, z_e)$, is assumed to be measurable to infer the human's goal. In implementing our approach, we use Chomp (a model-based motion optimizer [18]) to model the closed-loop dynamics (4) of the robotic manipulator, and use the optimal trajectories corresponding to each goal obtained via Chomp as the nominal trajectories $\xi^{\theta,\mathrm{nom}}$.

Remark 4: We have chosen to consider such a HRC task for evaluating our approach because: 1) goal inference is an essential component in HRC, 2) well-tuned goal inference baseline methods for such a task are available, and 3) a sensing action in such a task may entail a considerable amount of computation due to 3D image processing (assuming the primary sensor is a 3D camera) and high-DOF state estimation. Note, however, that computational cost is only one type of sensing costs. Tasks that involve other types of sensing costs have been exemplified in Section I, to which our approach can also be applied.

B. Independent variable

The independent variable is the implementation of different goal inference algorithms. In addition to our Goal Inference approach based on Hybrid Kalman Filtering with Cost-Effective sensing (HKF-CE-GI), we consider two other baselines:

Bayesian-GI. The first baseline algorithm is the Bayesian Goal Inference algorithm [5], [7] that exploits the Boltzmann noisily-rational behavioral model [19]. In Bayesian-GI, the agent's goal is inferred according to:

$$\theta^* = \arg\max_{\theta \in \Theta} \mathbb{P}(\xi_t|\theta)\mathbb{P}_0(\theta), \tag{20a}$$

$$\mathbb{P}(\xi_t|\theta) = \frac{\exp(-\mathcal{C}_\theta(\xi_t)) \int_{\tilde{\xi}_{x_t \to x_\theta}} \exp(-\mathcal{C}_\theta(\tilde{\xi}_{x_t \to x_\theta}))\, \mathrm{d}\tilde{\xi}_{x_t \to x_\theta}}{\int_{\tilde{\xi}_{x_0 \to x_\theta}} \exp(-\mathcal{C}_\theta(\tilde{\xi}_{x_0 \to x_\theta}))\, \mathrm{d}\tilde{\xi}_{x_0 \to x_\theta}}, \tag{20b}$$

where $\mathbb{P}_0(\theta)$ is a prior distribution over Θ, $\mathcal{C}_\theta$ is a goal-conditioned cost function that drives the agent to reach the goal, $\tilde{\xi}_t$ denotes the observed agent's trajectory at t, $\tilde{\xi}_{x_i \to x_f}$ denotes a sampled trajectory from state x_i to state x_f, and x_θ denotes the state of goal θ. The integrals over trajectories can be approximated using Laplace's method, with which (20b) reduces to

$$\mathbb{P}(\xi_t|\theta) = \frac{\exp(-\mathcal{C}_\theta(\xi_{x_0 \to x_t}) - \mathcal{C}_\theta(\xi^*_{x_t \to x_\theta}))}{\exp(-\mathcal{C}_\theta(\xi^*_{x_0 \to x_\theta}))}, \tag{21}$$

where $\xi^*_{x_0 \to x_\theta}$ denotes the optimal trajectory from x_0 to x_θ with resp. to $\mathcal{C}_\theta$ and can be computed using a motion planner.

Remark 5: The Boltzmann model assumes that the agent chooses her trajectory ξ with a probability proportional to its exponential cost. Consequently, the more ξ is deviated from the optimal trajectory (in terms of the cost), the less likely the agent is choosing ξ. This assumption agrees with the assumption behind the stochastic model (6) based on which our goal inference approach based on hybrid Kalman filtering is developed: the more a trajectory is deviated from the nominal trajectory $\xi^{\theta,\mathrm{nom}}$ (which can be the optimal trajectory), the less likely the agent is choosing that trajectory. Furthermore, our belief update rule (11) implements a principle that is analogous to Bayesian-GI: if the agent's behavior is more aligned with the nominal/optimal trajectory of a goal, then that goal is more likely to be the intended one. However, Bayesian-GI requires solving for the optimal trajectories from the current state to all possible goals at every sample time instant in order to update the belief distribution over the goal set, which entails intensive computations and is thus not as suitable as our hybrid Kalman filtering-based approach for implementing our model predictive-based cost-effective sensing strategy.

HKF-GI. The second baseline algorithm is an implementation of the Goal Inference approach based on Hybrid Kalman Filtering with persistent measurement (i.e., taking a measurement at every step), described in Section III-A.

C. Dependent measures

In order to validate our proposed approach and compare it against baselines, we choose to evaluate the *cost-effectiveness* and the *computational feasibility* of each goal inference algorithm. Specifically, we count the time taken and the number of measurements taken for each algorithm to build up 90% confidence in the agent's actual goal ($b_{t|t}(\theta^*) = \mathbb{P}(\hat{\theta} = \theta^* \,|\, \mathbf{y}_t) \geq 90\%$) for evaluating the cost-effectiveness, and count the average computation time to perform one belief update for evaluating the computational feasibility.

V. Results and Analysis

A. User studies

We conduct our user study on a simulated robotic manipulator developed in [20]. In this study, the robotic manipulator is controlled by a human operator via a joystick to grasp one of the four objects on the table (see Fig. 2(a)). The solid lines show the end-effector's nominal trajectories corresponding to each goal obtained from Chomp. We let the human control the manipulator to reach her intended goal and record the end-effector's entire trajectory during this process (by taking a position measurement at every sample time instant with a sampling period of $0.2[s]$). We then replay this trajectory when evaluating each goal inference algorithm, to eliminate the effect of trajectory variations.

Fig. 2(b) and (c) show the time histories of the beliefs in each goal and of the sensing decision (i.e., whether or not to take a measurement at each sample time instant) using our proposed HKF-CE-GI algorithm. Fig. 2(d) and (e) show the time histories of the beliefs in each goal using HKF-GI and using Bayesian-GI, respectively. It can be seen that all three algorithms accurately predict the human operator's intended goal with similar inference speeds. Note that our HKF-CE-GI algorithm did not take any measurements for $t \leq 1.2[s]$ and $t \geq 4.2[s]$. This is because our algorithm predicted that 1) little information about the human operator's intended goal could be gained from measurements at the initial stage of the manipulator's motion, and 2) little more information could be gained from measurements at the terminal stage (because we had already built up a high confidence in the intended goal), which are also verified in Fig. 2(d) and (e): the beliefs changed only slightly from their initial values for $t \leq 1.2[s]$

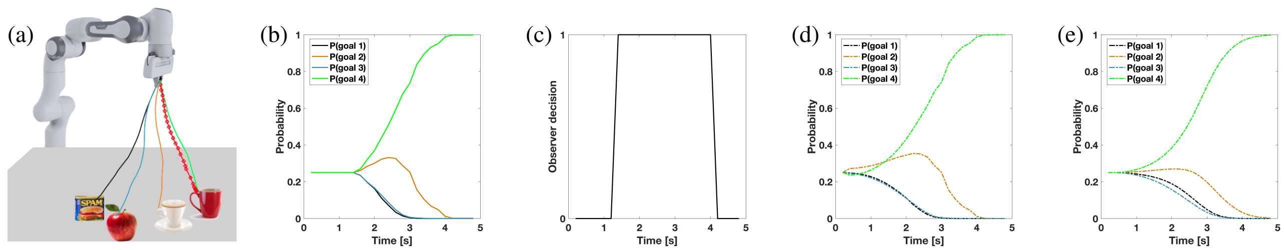

Fig. 2: (a): Manipulator end-effector's nominal trajectories corresponding to each goal (solid lines) and a fully-measured sample trajectory of a human operator for reaching her intended goal (red dotted line). (b): Time-history of the beliefs in each goal using HKF-CE-GI. (c): Time-history of the sensing decision using HKF-CE-GI. (d): Time-history of the beliefs in each goal using HFK-GI. (e): Time-history of the beliefs in each goal using Bayesian-GI.

even though HKF-GI and Bayesian-GI took measurements at every sample time instant, and had almost converged at $t = 4.2[s]$, which caused any measurements for $t \geq 4.2[s]$ less valuable.

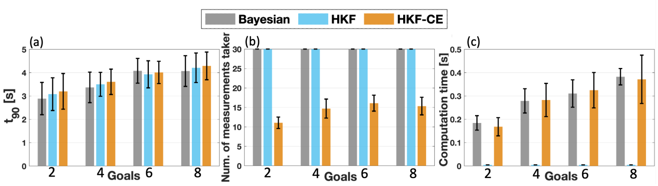

Fig. 3: Evaluation results for different goal set sizes. (a): Average t_{90}. (b): Average number of measurements taken during the entire process. (c): Average computation time for one belief distribution update. The black lines denote the standard deviations.

B. Evaluations

We quantitatively evaluate our HKF-CE-GI algorithm against baselines. Before evaluation, we randomly generate 20 goals. For each goal, we let a human operator control the manipulator to reach the goal and record the end-effector's trajectory with a sampling period of $0.2[s]$. During evaluation, we sample a subset of the generated goals as the goal set, randomly select a goal as the "actual goal," and replay the corresponding recorded trajectory. Fig. 3 shows the evaluation results for different goal set sizes. Specifically, we show the average t_{90} (the time taken to build up 90% confidence) in Fig. 3(a), the average number of measurements taken during the entire inference process in Fig. 3(b), and the average computation time for one belief update in Fig. 3(c). The numbers are calculated using statistics from 40 trials (the goal set and actual goal are re-sampled for each trial). It can be seen that in all cases, HKF-CE-GI achieves similar t_{90} performance compared with the baselines but takes significantly less measurements. With regard to computational feasibility, HKF-GI takes negligible computation times to perform belief updates thanks to the high computational efficiency of Kalman filtering, while HKF-CE-GI takes slightly more computation times than Bayesian-GI. Note that the computations of HKF-CE-GI are for predicting the information gain and optimizing the sensing decisions to achieve cost-effective sensing, which are not attainable by HKF-GI and Bayesian-GI. We also note that the computation times of HKF-CE-GI can be reduced by exploiting parallel computation.

C. Robustness to sub-optimal behaviors

We aim to evaluate the robustness of HKF-CE-GI to sub-optimal behaviors. Because it is difficult to control real human's sub-optimal behaviors, we train a simulated human to perform the task using the reinforcement learning approach in [21] and use controlled perturbations to model sub-optimal behaviors. The evaluation procedure is similar as before except that the actions from the simulated human are subject to perturbations sampled from $\mathcal{N}(0, \lambda\mathbb{I}_{4\times 4})$, where λ characterizes the perturbation level. Fig. 4 shows the average t_{90} and average number of measurements taken of 40 trials under two different perturbation levels. It can be seen that HKF-CE-GI outperforms the baselines in terms of cost-effectiveness for both perturbation levels.

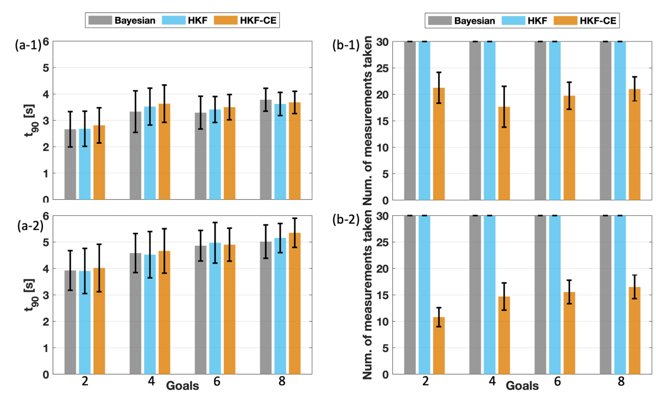

Fig. 4: Evaluation results with simulated human subject to sub-optimal behaviors. (a-1) and (a-2): Average t_{90} for $\lambda = 0.1$ and $\lambda = 1$. (b-1) and (b-2): Average number of measurements taken during the inference process for $\lambda = 0.1$ and $\lambda = 1$. The black lines denote the standard deviations.

VI. Conclusion

Goal inference is an important problem for a variety of applications that involve goal-oriented agents. In this work, we advocated that a sensible trade-off between the information gained from measurements and the cost associated with sensing actions should be considered in situations where sensing is associated with high cost. We cast the goal inference problem in a goal-oriented task as a hybrid estimation problem and developed a hybrid Kalman filtering-based algorithm for goal inference. We then exploited the model predictive control framework to define a cost-effective sensing strategy. Our evaluation showed that, compared to baselines, our approach takes a considerably reduced number of measurements while not impairing the inference speed, accuracy, and reliability.

REFERENCES

[1] C. Pérez-D'Arpino and J. A. Shah, "Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 6175–6182.

[2] C. Liu, J. B. Hamrick, J. F. Fisac, A. D. Dragan, J. K. Hedrick, S. S. Sastry, and T. L. Griffiths, "Goal inference improves objective and perceived performance in human-robot collaboration," in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, ser. AAMAS '16, 2016, p. 940–948.

[3] L. Bruckschen, K. Bungert, N. Dengler, and M. Bennewitz, "Predicting human navigation goals based on bayesian inference and activity regions," *Robotics and Autonomous Systems*, vol. 134, p. 103664, 2020.

[4] C. L. Baker, J. B. Tenenbaum, and R. R. Saxe, "Goal inference as inverse planning," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 29, no. 29, 2007.

[5] A. D. Dragan and S. S. Srinivasa, *Formalizing assistive teleoperation*. MIT Press, July, 2012.

[6] T. Zhi-Xuan, J. Mann, T. Silver, J. Tenenbaum, and V. Mansinghka, "Online bayesian goal inference for boundedly rational planning agents," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[7] M. Zurek, A. Bobu, D. S. Brown, and A. D. Dragan, "Situational confidence assistance for lifelong shared autonomy," *arXiv preprint arXiv:2104.06556*, 2021.

[8] Y. Hu, W. Zhan, and M. Tomizuka, "Scenario-transferable semantic graph reasoning for interaction-aware probabilistic prediction," *arXiv preprint arXiv:2004.03053*, 2020.

[9] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, "Computer vision algorithms and hardware implementations: A survey," *Integration*, vol. 69, pp. 309–320, 2019.

[10] R. L. Sutherland, I. V. Kolmanovsky, A. R. Girard, F. A. Leve, and C. D. Petersen, "Optimal strategies for disjunctive sensing and control," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 4712–4717.

[11] Y. Ding, C. Florensa, P. Abbeel, and M. Phielipp, "Goal-conditioned imitation learning," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.

[12] J. Mainprice, R. Hayne, and D. Berenson, "Goal set inverse optimal control and iterative replanning for predicting human reaching motions in shared workspaces," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 897–908, 2016.

[13] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey *et al.*, "Maximum entropy inverse reinforcement learning." in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.

[14] I. Hwang, H. Balakrishnan, and C. Tomlin, "State estimation for hybrid systems: applications to aircraft tracking," *IEE Proceedings-Control Theory and Applications*, vol. 153, no. 5, pp. 556–566, 2006.

[15] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[16] T. Lefebvre, H. Bruyninckx, and J. De Schutter, "Kalman filters for nonlinear systems: A comparison of performance," *International Journal of Control*, vol. 77, no. 7, pp. 639–653, 2004.

[17] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[18] M. Zucker, N. Ratliff, A. D. Dragan, M. Pivtoraiko, M. Klingensmith, C. M. Dellin, J. A. Bagnell, and S. S. Srinivasa, "Chomp: Covariant hamiltonian optimization for motion planning," *The International Journal of Robotics Research*, vol. 32, no. 9-10, pp. 1164–1193, 2013.

[19] J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior*. Princeton university press, 2007.

[20] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," http://pybullet.org, 2016–2021.

[21] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel, and W. Zaremba, "Hindsight experience replay," *arXiv preprint arXiv:1707.01495*, 2017.