2021 61st IEEE Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Approximate Information States for Worst-case Control of Uncertain Systems

Aditya Dave, *Student Member, IEEE,* Nishanth Venkatesh, *Student Member, IEEE,*
Andreas A. Malikopoulos, *Senior Member, IEEE*

Abstract— **In this paper, we investigate a worst-case-scenario control problem with a partially observed state. We consider a non-stochastic formulation, where noises and disturbances in our dynamics are uncertain variables which take values in finite sets. In such problems, the optimal control strategy can be derived using a dynamic program (DP) with respect to the memory. The computational complexity of this DP can be improved using a conditional range of the state instead of the memory. We present a more general definition of an information state which is sufficient to construct a DP without loss of optimality, and show that the conditional range is an example of an information state. Next, we extend this notion to define an approximate information state and an approximate DP. We also bound the maximum loss of optimality when using an approximate DP to derive the control strategy. Finally, we illustrate our results in a numerical example.**

I. Introduction

A decision maker in a cyber-physical system must generate control actions using noisy observations and uncertain predictions of the state's evolution [1], [2]. Such a decision making problem is usually modelled using one of the following approaches: (1) Stochastic control: We consider that all uncertainties have known distributions and seek a control strategy which minimizes the expected total cost over a time horizon [3]. The resulting control strategy performs optimally on average across multiple runs of the system. (2) Non-stochastic control: We consider that all uncertainties take values in known sets with unknown distributions and seek a control strategy which minimizes the maximum cost that can be incurred across a time horizon [4]. The resulting control strategy is more conservative, but has concrete performance guarantees for any evolution of the system. Thus, this approach is more useful in applications where safety guarantees are critical, or where the probability distributions of external disturbances are unknown a priori. Under either approach, the optimal control strategy can be derived offline using a dynamic program (DP) [4]. In general, the optimal action at any time is a function of the historical data in the decision maker's memory. This memory grows with time as more data is added. Subsequently, the domain of the corresponding control strategy grows with time which makes it computationally challenging to solve the DP for long time horizons. In stochastic control, this problem is alleviated by writing the DP using *information states* instead of the memory [5]. In fact, the notion of an information state is fundamental in the study of stochastic systems has also been extended to reinforcement learning [6], [7] and decentralized problems [8]–[10]. The most frequently used information state for stochastic control is the *belief state*, i.e., a distribution on the feasible states conditioned on the current memory [3]. However, in some systems with large state spaces, using an information state imposes severe computational challenges in DP. Thus, recent efforts focus on identifying *approximate information states* [7], [11]. In contrast, for non-stochastic systems, DP has been simplified using a *conditional range*, i.e., the set of state realizations consistent with the current memory [12]–[15]. This concept has also been applied to decentralized systems [16]–[18]. However, to the best of our knowledge, there is no general definition of an information state or approximate information state for non-stochastic systems.

The contributions of this paper are (1) we introduce a general definition of an information state (Definition 1) along with a proof of the optimality of the resulting DP (Theorem 1) and (2) we introduce an approximate information state for non-stochastic systems (Definition 2), use it to formulate an approximate DP and present an upper bound on the resulting loss in performance (Theorems 2 - 3). We illustrate our results using numerical simulations.

The remainder of the paper proceeds as follows. In Section II, we present our model. In Section III, we define the notion of the information state and the corresponding DP. In Section IV, we present the approximate information state, approximate DP, and bounds on the approximation loss. In Section V, we present a numerical example to illustrate our results. Finally, in Section VI, we draw concluding remarks and discuss ongoing work.

II. Model

A. Notation and Preliminaries

In this paper, we utilize the mathematical framework for *uncertain variables*, which was presented in the context of non-stochastic information theory in [19]. An uncertain variable is a non-stochastic analogue of a random variable which take values in a known set and has an unknown distribution. Thus, for a sample space Ω and a given set $\mathcal{X}$, an uncertain variable is a mapping $X : \Omega \to \mathcal{X}$. For any $\omega \in \Omega$, the uncertain variable has the realization $X(\omega) = x \in \mathcal{X}$. The *range* of an uncertain variable is analogous to the distribution of a random variable. The *marginal range* of X is the set $[[X]] := \{X(\omega) \mid \omega \in \Omega\}$. For two uncertain

This research was supported by the Sociotechnical Systems Center (SSC) at the University of Delaware.

The authors are with the Department of Mechanical Engineering, University of Delaware, Newark, DE 19716 USA (email: adidave@udel.edu; nish@udel.edu; andreas@udel.edu).

variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, their *joint range* is the set $[[X, Y]] := \big\{\big(X(\omega), Y(\omega)\big) \mid \omega \in \Omega\big\}$. For a given realization y of Y, the *conditional range* of X is the set $[[X|y]] := \big\{X(\omega) \mid Y(\omega) = y,\ \omega \in \Omega\big\}$ and in general, the conditional range of X given Y is $[[X|Y]] := \big\{[[X|y]] \mid y \in [[Y]]\big\}$.

Next, consider that the feasible sets $\mathcal{X}, \mathcal{Y}$ are compact, nonempty subsets of a metric space $(\mathcal{S}, d)$, where $d(x, y)$ is the distance between any feasible realization $x \in \mathcal{X}$ of the uncertain variable X and $y \in \mathcal{Y}$ of the uncertain variable Y. Furthermore, the distance between the sets $\mathcal{X}, \mathcal{Y}$ is measured using the Hausdorff metric [20, Chapter 1.12]

$$\mathcal{H}(\mathcal{X}, \mathcal{Y}) := \max\{\max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} d(x, y), \max_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} d(x, y)\}. \quad (1)$$

B. Problem Formulation

We consider a partially observed system where an agent selects control actions over $T \in \mathbb{N}$ discrete time steps. At any time $t = 0, \ldots, T$, the state of the system is denoted by an uncertain variable X_t which takes values in a finite set $\mathcal{X}_t$. The action at time t is denoted by an uncertain variable U_t which takes values in a finite set $\mathcal{U}_t$. Starting with the initial state $X_0 \in \mathcal{X}_0$, the state evolves as $X_{t+1} = f_t\left(X_t, U_t, W_t\right)$ for all $t = 0, \ldots, T-1$. The uncertain variable W_t denotes an uncontrolled disturbance acting on the state at time t, which takes values in a finite set $\mathcal{W}_t$. At each $t = 0, \ldots, T$, the agent partially observes the state as an uncertain variable $Y_t = h_t(X_t, N_t)$ which takes values in a finite set $\mathcal{Y}_t$. The uncertain variable N_t denotes a measurement noise which takes values in a finite set $\mathcal{N}_t$. The external disturbances $\{W_t : t = 0, \ldots, T\}$, noises in measurement $\{N_t : t = 0, \ldots, T\}$, and initial state X_0 are collectively called *primitive variables*. We consider that each primitive variable is independent of each other. This ensures that the system evolution is Markovian in a non-stochastic sense [12], [19]. We also consider that the agent has perfect memory and thus, at each $t = 0, \ldots, T$, the agent's memory is the set of uncertain variables $M_t := \{Y_{0:t}, U_{0:t-1}\}$ which takes values in a collection of sets $\mathcal{M}_t$. Note that $Y_{0:t} = \{Y_0, \ldots, Y_t\}$. After updating their memory, at each t the agent selects the action $U_t = g_t(M_t)$ using the control law $g_t : \mathcal{M}_t \to \mathcal{U}_t$. We denote the control strategy by $\boldsymbol{g} := (g_0, \ldots, g_T)$ and the set of all feasible control strategies by $\mathcal{G}$. After T time steps, the agent incurs a terminal cost $c_T(X_T, U_T) \in \mathbb{R}_{\geq 0}$. We measure the system's performance by the *maximum terminal cost*

$$\mathcal{J}(\boldsymbol{g}) := \max_{X_0, W_{0:T}, N_{0:T}} c_T(X_T, U_T). \quad (2)$$

Problem 1. The optimization problem is $\min_{\boldsymbol{g}\in\mathcal{G}} \mathcal{J}(\boldsymbol{g})$, given the feasible sets $\{\mathcal{X}_0, \mathcal{W}_t, \mathcal{N}_t : t = 0, \ldots, T\}$, the system dynamics $\{f_t, h_t : t = 0, \ldots, T\}$ and the terminal cost function c_T.

Our aim is to tractably compute an optimal control strategy $\boldsymbol{g}^* \in \mathcal{G}$ for Problem 1. This strategy is guaranteed to exist because all variables take values in finite sets. In our modeling framework, we impose the following assumptions:

Assumption 1. We consider that the feasible sets $\{\mathcal{X}_t, \mathcal{Y}_t : t = 0, \ldots, T\}$ are both subsets of metric spaces.

Assumption 1 allows us to measure the distance between any two realizations of X_t and Y_t, respectively, for all t. To this end, we denote a generic metric by $d(\cdot, \cdot)$.

Assumption 2. We consider that all uncertain variables at each t and the cost $c_T(X_T, U_T)$ have a finite maximum value.

Since all feasible sets are finite, Assumption 2 ensures that the functions $\{f_t, h_t : t = 0, \ldots, T\}$ and c_T are globally Lipschitz. To this end, we will denote the Lipschitz constant of a function f_t by $L_{f_t} \in \mathbb{R}_{\geq 0}$.

Remark 1. While we derive our results for terminal cost problems, we present an extension of our results to additive cost problems in Subsection III-B.

III. Dynamic Programs and Information States

In this section, we present a DP to derive the optimal control strategy for Problem 1, and then define an information state which can simplify it. For all t, for each possible realization $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$ of the memory M_t and action U_t, respectively, we recursively define the *value functions* of the DP at $t = 0, \ldots, T-1$ as

$$Q_t(m_t, u_t) := \max_{m_{t+1}\in[[M_{t+1}|m_t, u_t]]} V_{t+1}(m_{t+1}), \quad (3)$$

$$V_t(m_t) := \min_{u_t\in\mathcal{U}_t} Q_t(m_t, u_t), \quad (4)$$

and $Q_T(m_T, u_T) := \max_{x_T\in[[X_T|m_T]]} c_T(x_T, u_T)$ and $V_T(m_T) := \min_{u_T\in\mathcal{U}_T} Q_T(m_T, u_T)$ at time T. The control law at each t is $g_t^*(m_t) := \arg\min_{u_t\in\mathcal{U}_t} Q_t(m_t, u_t)$. The control strategy $\boldsymbol{g}^* = (g_0^*, \ldots, g_T^*)$ can be shown to be the optimal solution to Problem 1 using standard arguments [4]. Note that at each t, the optimization in the RHS of (4) must be solved for each possible $m_t \in \mathcal{M}_t$. As t increases, the size of the $\mathcal{M}_t$ increases with the addition of new information. Subsequently, a large number of computations are required to solve the DP for a long time horizon T. This issue motivates us to seek an uncertain variable, called an information state, which can be used in the DP instead of the memory, without loss of optimality.

A. Information States

In this subsection, we define an information state which is sufficient to construct a DP and prove that it yields an optimal control strategy.

Definition 1. An *information state* for Problem 1 at each $t = 0, \ldots, T$ is an uncertain variable $\Pi_t = \sigma_t(M_t)$ taking values in a finite set $\mathcal{P}_t$ and generated by $\sigma_t : \mathcal{M}_t \to \mathcal{P}_t$. Furthermore, for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$, and for all $t = 0, \ldots, T$, it satisfies the following properties:

1) *Sufficient to evaluate terminal cost:*

$$\max_{x_T\in[[X_T|m_T]]} c_T(x_T, u_T) = \max_{x_T\in[[X_T|\sigma_T(m_T)]]} c_T(x_T, u_T). \quad (5)$$

2) *Sufficient to predict itself:*

$$[[\Pi_{t+1}|m_t,u_t]] = [[\Pi_{t+1}|\sigma_t(m_t),u_t]]. \tag{6}$$

In the corresponding DP, for all t, and for all $\pi_t \in \mathcal{P}_t$ and $u_t \in \mathcal{U}_t$, we recursively define the value functions at $t = 0,\dots,T-1$ as

$$\bar{Q}_t(\pi_t,u_t) := \max_{\pi_{t+1}\in[[\Pi_{t+1}|\pi_t,u_t]]} \bar{V}_{t+1}(\pi_{t+1}), \tag{7}$$

$$\bar{V}_t(\pi_t) := \min_{u_t\in\mathcal{U}_t} \bar{Q}_t(\pi_t,u_t), \tag{8}$$

and $\bar{Q}_T(\pi_T,u_T) := \max_{x_T\in[[X_T|\pi_T]]} c_T(x_T,u_T)$ and $V_T(\pi_T) := \min_{u_T\in\mathcal{U}_T} Q_T(\pi_T,u_T)$ at time T. The control law at each t is $g^*_t(\pi_t) := \arg\min_{u_t\in\mathcal{U}_t} \bar{Q}_t(\pi_t,u_t)$. Next, we prove that the DP decomposition with information states (7) - (8) yields the same optimal value as the DP decomposition in (3) - (4) that uses the system's memory at each $t = 0,\dots,T$.

Theorem 1. *Let $\Pi_t = \sigma_t(M_t)$ be an information state for Problem 1 for all $t = 0,\dots,T$. Then, for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$, $Q_t(m_t,u_t) = \bar{Q}_t(\sigma_t(m_t),u_t)$ and $V_t(m_t) = \bar{V}_t(\sigma_t(m_t))$.*

Proof. Let $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$ be given realizations of M_t and U_t, respectively, for all t. We prove the result using mathematical induction, starting with T where $Q_T(m_T,u_T) = \max_{x_T\in[[X_T|m_T]]} c_T(x_T,u_T) = \max_{x_T\in[[X_T|\sigma_T(m_T)]]} c_T(x_T,u_T) = \bar{Q}_T(\sigma_T(m_T),u_T)$ holds as a direct result of (5) in Definition 1. Subsequently, by taking the minimum on both sides with respect to $u_t \in \mathcal{U}_t$, it holds that $V_T(m_T) = \bar{V}_T(\sigma_T(m_T))$. With this as the basis, for any $t = 0,\dots,T-1$, we consider the induction hypothesis $V_{t+1}(m_{t+1}) = \bar{V}_{t+1}(\sigma_{t+1}(m_{t+1}))$. Next, we first prove that $Q_t(m_t,u_t) = \bar{Q}_t(\sigma_t(m_t),u_t)$ at t by showing that the RHS of (3) is equal to the RHS of (7). Using the induction hypothesis in the RHS of (3), $\max_{m_{t+1}\in[[M_{t+1}|m_t,u_t]]} V_{t+1}(m_{t+1}) = \max_{m_{t+1}\in[[M_{t+1}|m_t,u_t]]} \bar{V}_{t+1}(\sigma_{t+1}(m_{t+1})) = \max_{\sigma_{t+1}(m_{t+1})\in[[\Pi_{t+1}|\sigma_t(m_t),u_t]]} \bar{V}_{t+1}(\sigma_{t+1}(m_{t+1}))$, where, in the second equality, we use the fact that $[[\Pi_{t+1}|m_t,u_t]] = \{\sigma_{t+1}(m_{t+1}) \in \mathcal{P}_{t+1} | m_{t+1} \in [[M_{t+1}| m_t,u_t]]\}$ and (6). Thus, at time t, it holds that $Q_t(m_t,u_t) = \bar{Q}_t(\sigma_t(m_t),u_t)$. Subsequently, we can prove $V_t(m_t) = \bar{V}_t(\sigma_t(m_t))$ by minimizing both sides with $u_t \in \mathcal{U}_t$. This proves the induction hypothesis at time t, and the result follows by mathematical induction. □

Theorem 1 implies that the control strategy computed using (7) - (8) is an optimal solution to Problem 1. In practice, using the information state in the DP decomposition only improves computational tractability when the set $\mathcal{P}_t$ has fewer elements than $\mathcal{M}_t$ for most time steps. This is usually true for systems with long time horizons.

B. Examples of Information States

In this subsection, we present examples of information states which satisfy the conditions in Definition 1.

1) Perfectly Observed Systems: Consider a system where $Y_t = X_t$ for all $t = 0,\dots,T$. Then, the state is a valid information state [4], i.e., $\Pi_t = X_t$, which takes values in the set $\mathcal{X}_t$ and satisfies (5) - (6) for all t.

2) Partially Observed Systems: The conditional range $\Pi_t = [[X_t|M_t]]$ is an information state for each $t = 0,\dots,T$ in any partially observed system [4]. This is a set-valued uncertain variable which takes values in the power set $2^{\mathcal{X}_t}$. Explicitly, for a given realization of the memory $m_t \in \mathcal{M}_t$ at time t, the conditional range takes the realization $P_t := \{x_t \in \mathcal{X}_t | \exists x_0 \in \mathcal{X}_0, w_{0:t-1} \in \prod_{\ell=0}^{t-1} \mathcal{W}_\ell, n_{0:t} \in \prod_{\ell=0}^{t} \mathcal{N}_\ell$ such that $y_t = h_t(x_t,n_t), x_{\ell+1} = f_\ell(x_\ell,u_\ell,w_\ell), y_\ell = h_\ell(x_\ell,n_\ell)$ for all $\ell = 0,\dots,t-1\}$. We denote the realization by P_t instead of π_t to highlight that it is a set.

3) Additive Cost Problems: Consider a partially observed system where the agent incurs a cost $c_t(X_t,U_t) \in \mathbb{R}_{\geq 0}$ at each $t = 0,\dots,T$ and the performance is measured by an additive performance criterion $\mathcal{J}^{\text{ad}}(\boldsymbol{g}) := \max_{X_0,W_{0:T},N_{0:T}} \sum_{t=0}^{T} c_t(X_t,U_t)$. We can construct a DP and an information state for an additive cost problem by recasting it as a terminal cost problem [12]. At $t = 0$, we define $A_0 := 0$ and for all $t = 1,\dots,T$, we recursively define an uncertain variable $A_t \in \mathcal{A}_t$ as $A_t := A_{t-1} + c_{t-1}(X_{t-1},U_{t-1})$. Then, at each t, we consider an augmented state $S_t = (X_t,A_t)$ and note that it yields a terminal cost $A_T + c_T(X_T,U_T)$. Thus, we can derive the optimal control strategy using the terminal cost DP. The information state at time t is the conditional range $\Pi_t = [[X_t,A_t|M_t]]$ which takes values in the power set $2^{\mathcal{X}_t\times\mathcal{A}_t}$.

Remark 2. While the conditional range is an information state for all partially observed systems, the more general conditions in Definition 1 can enable us to identify simpler information states for cases like systems with perfect observation. However, in many applications we may seek to construct an information state using limited data or seek to improve computational tractability of the DP by approximation. To account for these applications, in Section IV, we extend Definition 1 to define approximate information states.

IV. Approximate Information State

In this section, we define an approximate information state by relaxing the conditions in Definition 1. Then, we use it to develop an approximate DP and derive an upper bound on the resulting loss of optimality.

Definition 2. An *approximate information state* for Problem 1 is an uncertain variable $\hat{\Pi}_t = \hat{\sigma}_t(M_t)$, at each $t = 0,\dots,T$, taking values in a finite set $\hat{\mathcal{P}}_t$ and generated by $\hat{\sigma}_t : \mathcal{M}_t \to \hat{\mathcal{P}}_t$. Furthermore, for all t, there exist parameters $\epsilon_T, \delta_t \in \mathbb{R}_{\geq 0}$ such that for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$ it is:

1) Sufficient to approximate terminal cost:

$$|\max_{x_T\in[[X_T|m_T]]} c_T(x_T,u_T) - \max_{x_T\in[[X_T|\hat{\sigma}_T(m_T)]]} c_T(x_T,u_T)| \leq \epsilon_T. \tag{9}$$

2) Sufficient to approximate evolution: We define $\mathcal{K}_{t+1} := [[\hat{\Pi}_{t+1} \mid m_t,u_t]]$ and $\hat{\mathcal{K}}_{t+1} := [[\hat{\Pi}_{t+1} \mid \hat{\sigma}_t(m_t),u_t]]$. Then,

$$\mathcal{H}(\mathcal{K}_{t+1},\hat{\mathcal{K}}_{t+1}) \leq \delta_t, \tag{10}$$

where recall that $\mathcal{H}(\cdot)$ is the Hausdorff metric from (1).

In the approximate DP, for all $t = 0, \dots, T-1$, for all $\hat{\pi}_t \in \hat{\mathcal{P}}_t$ and $u_t \in \mathcal{U}_t$, we recursively define the value functions

$$\hat{Q}_t(\hat{\pi}_t, u_t) := \max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1}|\hat{\pi}_t, u_t]]} \hat{V}_{t+1}(\hat{\pi}_{t+1}), \tag{11}$$

$$\hat{V}_t(\hat{\pi}_t) := \min_{u_t \in \mathcal{U}_t} \hat{Q}_t(\hat{\pi}_t, u_t), \tag{12}$$

and $\hat{Q}_T(\hat{\pi}_T, u_T) := \max_{x_T \in [[X_T|\hat{\pi}_T]]} c_T(x_T, u_T)$ and $\hat{V}_T(\hat{\pi}_T) := \min_{u_T \in \mathcal{U}_T} \hat{Q}_T(\hat{\pi}_T, u_T)$ at time T. The control law at each t is $\hat{g}_t^*(\hat{\pi}_t) := \arg\min_{u_t \in \mathcal{U}_t} \hat{Q}_t(\hat{\pi}_t, u_t)$ and the approximately optimal strategy is $\hat{\boldsymbol{g}}^* = (\hat{g}_0^*, \dots, \hat{g}_T^*)$. Next, in Theorem 2, we establish an error bound when the value functions for the optimal DP (3) - (4) are approximated by (11) - (12). We begin with two preliminary lemmas.

Lemma 1. *Consider a metric space $(\mathcal{A}, d)$ and two finite subsets $\mathcal{A}, \mathcal{B} \subset \mathcal{X}$. Let $f : \mathcal{X} \to \mathbb{R}$ be a function with a global Lipschitz constant $L_f \in \mathbb{R}_{\geq 0}$. Then,*

$$\big|\max_{a \in \mathcal{A}} f(a) - \max_{b \in \mathcal{B}} f(b)\big| \leq L_f \cdot \mathcal{H}(\mathcal{A}, \mathcal{B}). \tag{13}$$

Proof. We prove this result by considering two cases which are mutually exclusive but cover all the possibilities. Case 1: $\max_{a \in \mathcal{A}} f(a) \geq \max_{b \in \mathcal{B}} f(b)$, which implies $|\max_{a \in \mathcal{A}} f(a) - \max_{b \in \mathcal{B}} f(b)| = \max_{a \in \mathcal{A}} f(a) - \max_{b \in \mathcal{B}} f(b)$. We define the non-empty set $\mathcal{A}^1 := \{a \in \mathcal{A} | f(a) \geq \max_{b \in \mathcal{B}} f(b)\}$ and note that $\max_{a \in \mathcal{A}} f(a) - \max_{b \in \mathcal{B}} f(b) = \max_{a \in \mathcal{A}^1} f(a) - \max_{b \in \mathcal{B}} f(b) = \max_{a \in \mathcal{A}^1} \min_{b \in \mathcal{B}} (f(a) - f(b)) \leq \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} |f(a) - f(b)| \leq L_f \cdot \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} |a - b|$. We can complete the proof for case 1 by invoking the definition of the Hausdorrf metric in (1) to conclude that $|\max_{a \in \mathcal{A}} f(a) - \max_{b \in \mathcal{B}} f(b)| \leq L_f \cdot \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} |a - b| \leq L_f \cdot \mathcal{H}(\mathcal{A}, \mathcal{B})$. Case 2: $\max_{a \in \mathcal{A}} f(a) < \max_{b \in \mathcal{B}} f(b)$ and can prove the result using similar arguments as case 1. □

Lemma 2. *Consider a finite set $\mathcal{X}$ and two functions $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{X} \to \mathbb{R}$ with bounded outputs. Then,*

$$|\max_{x \in \mathcal{X}} f(x) - \max_{x \in \mathcal{X}} g(x)| \leq \max_{x \in \mathcal{X}} |f(x) - g(x)|, \tag{14}$$

$$|\min_{x \in \mathcal{X}} f(x) - \min_{x \in \mathcal{X}} g(x)| \leq \max_{x \in \mathcal{X}} |f(x) - g(x)|. \tag{15}$$

Proof. First, we prove (14) by considering two mutually exclusive cases which cover all possibilities. Case 1: We consider $\max_{x \in \mathcal{X}} f(x) \geq \max_{x \in \mathcal{X}} g(x)$, which implies $|\max_{x \in \mathcal{X}} f(x) - \max_{x \in \mathcal{X}} g(x)| = \max_{x \in \mathcal{X}} f(x) - \max_{x \in \mathcal{X}} g(x)$. Then, we define $x^* := \arg\max_{x \in \mathcal{X}} f(x)$ and note $\max_{x \in \mathcal{X}} f(x) - \max_{x \in \mathcal{X}} g(x) = f(x^*) - \max_{x \in \mathcal{X}} g(x) \leq f(x^*) - g(x^*) \leq \max_{x \in \mathcal{X}} |f(x) - g(x)|$. Case 2: $\max_{x \in \mathcal{X}} f(x) < \max_{x \in \mathcal{X}} g(x)$. The proof can be completed using similar arguments as in Case 1. The proof for (15) follows from similar arguments as (14). Due to space limitations, it is omitted. □

Theorem 2. *Let $L_{\hat{V}_{t+1}}$ be the Lipschitz constant of $\hat{V}_{t+1}(\cdot)$ for all $t = 0, \dots, T$. Then, for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$:*

$$|Q_t(m_t, u_t) - \hat{Q}_t(\hat{\sigma}_t(m_t), u_t)| \leq \alpha_t, \tag{16}$$

$$|V_t(m_t) - \hat{V}_t(\hat{\sigma}_t(m_t))| \leq \alpha_t, \tag{17}$$

where $\alpha_T = \epsilon_T$ and $\alpha_t = \alpha_{t+1} + L_{\hat{V}_{t+1}} \cdot \delta_t$ for all $t = 0, \dots, T-1$.

Proof. For all t, let $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$ be the realizations of M_t and U_t, respectively. We prove both results by mathematical induction, starting at time T, where (16) follows from (9) in Definition 2. For (17), we can expand the LHS as $|V_T(m_T) - \hat{V}_T(\hat{\sigma}_T(m_T))| = |\min_{u_T \in \mathcal{U}_T} Q_T(m_T, u_T) - \min_{u_t \in \mathcal{U}_T} \hat{Q}_T(\hat{\sigma}_T(m_T), u_T)| \leq \max_{u_T \in \mathcal{U}_T} |Q_T(m_T, u_T) - \hat{Q}_T(\hat{\sigma}_T(m_T), u_T)| \leq \epsilon_T$, where in the first inequality, we use (15) from Lemma 2 and in the second inequality we use (16). Next, for all t, we consider the hypothesis $|V_{t+1}(m_{t+1}) - \hat{V}_{t+1} \ (\hat{\sigma}_{t+1}(m_{t+1}))| \leq \alpha_{t+1}$. Then, $|Q_t(m_t, u_t) - \hat{Q}_t(\hat{\sigma}_t(m_t), u_t)| = \big|\max_{m_{t+1} \in [[M_{t+1}|m_t, u_t]]} V_{t+1}(m_{t+1}) - \max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1}|\hat{\sigma}_t(m_t), u_t]]} \hat{V}_{t+1}(\hat{\pi}_{t+1})\big|$. Then,

$$\begin{aligned} |Q_t(m_t, u_t) - \hat{Q}_t(\hat{\sigma}_t(m_t), u_t)| \leq \Big| \max_{m_{t+1} \in [[M_{t+1}|m_t, u_t]]} \\ V_{t+1}(m_{t+1}) - \max_{\hat{\sigma}_{t+1}(m_{t+1}) \in [[\hat{\Pi}_{t+1}|m_t, u_t]]} \hat{V}_{t+1}(\hat{\sigma}_{t+1}(m_{t+1}))\Big| + \\ \Big| \max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1}|m_t, u_t]]} \hat{V}_{t+1}(\hat{\pi}_{t+1}) - \max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1}|\hat{\sigma}_t(m_t), u_t]]} \hat{V}_{t+1}(\hat{\pi}_{t+1})\Big|, \end{aligned} \tag{18}$$

where, we use the triangle inequality. In the first term, $\big|\max_{m_{t+1} \in [[M_{t+1}|m_t, u_t]]} V_{t+1}(m_{t+1}) - \max_{\hat{\sigma}_{t+1}(m_{t+1}) \in [[\hat{\Pi}_{t+1}|m_t, u_t]]} \hat{V}_{t+1}(\hat{\sigma}_{t+1}(m_{t+1}))\big| \leq \max_{m_{t+1} \in [[M_{t+1}|m_t, u_t]]} \big|V_{t+1}(m_{t+1}) - \hat{V}_{t+1}(\hat{\sigma}_{t+1}(m_{t+1}))\big| \leq \alpha_{t+1}$, where, in the first inequality, we note that $[[\hat{\Pi}_{t+1}|m_t, u_t]] = \{\hat{\sigma}_{t+1}(m_{t+1}) \in \hat{\mathcal{P}}_t | m_{t+1} \in [[M_{t+1}|m_t, u_t]]\}$ and use (14) from Lemma 2; and, in the second inequality, we use the induction hypothesis. The second term in the RHS of (18) satisfies $\big|\max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1}|m_t, u_t]]} \hat{V}_{t+1}(\hat{\pi}_{t+1}) - \max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1}|\hat{\sigma}_t(m_t), u_t]]} \hat{V}_{t+1}(\hat{\pi}_{t+1})\big| \leq L_{\hat{V}_{t+1}} \cdot \delta_t$ using (13) from Lemma 1 and (10) from Definition 2. Substituting the inequality for each term in the RHS of (18) yields $|Q_t(m_t, u_t) - \hat{Q}_t(\hat{\sigma}_t(m_t), u_t)| \leq \alpha_{t+1} + L_{\hat{V}_{t+1}} \cdot \delta_t$. Subsequently, we can prove (17) as $V_t(m_t) - \hat{V}_t(\hat{\sigma}_t(m_t)) = |\min_{u_t \in \mathcal{U}_t} Q_t(m_t, u_t) - \min_{u_t \in \mathcal{U}_t} \hat{Q}_t(\hat{\sigma}_t(m_t), u_t)| \leq \max_{u_t \in \mathcal{U}_t} |Q_t(m_t, u_t) - \hat{Q}_t(\hat{\sigma}_t(m_t), u_t)| \leq \alpha_t$, where, in the first inequality, we use (15) from Lemma 2. This completes our proof by induction for all $t = 0, \dots, T$. □

After bounding the approximation error for value functions, we also seek to bound the optimality gap from using the approximate strategy. Consider the approximately optimal strategy $\hat{\boldsymbol{g}} := (\hat{g}_0, \dots, \hat{g}_T)$ with $\hat{g}_t(\hat{\pi}_t) = \arg\min_{u_t \in \mathcal{U}_t} \hat{Q}_t(\hat{\pi}_t, u_t)$ for all t. Then, the equivalent strategy $\boldsymbol{g} = (g_0, \dots, g_T)$ using memory has $g_t(m_t) := \hat{g}_t(\hat{\sigma}_t(m_t))$ for all t. To compute the performance of $\boldsymbol{g}$, we define for all $t = 0, \dots, T-1$, for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$:

$$\Theta_t(m_t, u_t) := \max_{m_{t+1} \in [[M_{t+1}|m_t, u_t]]} \Lambda_{t+1}(m_{t+1}), \tag{19}$$

$$\Lambda_t(m_t) := \Theta_t(m_t, g_t(m_t)), \tag{20}$$

and $\Theta_T(m_T, u_T) := \max_{x_T \in [[X_T|m_T]]} c_T(x_T, u_T)$ and $\Lambda_T(m_T) := \Theta_T(m_T, g_T(m_T))$ for time T. Then, because

$m_0 = \{y_0\}$, the performance of $\boldsymbol{g}$ is $\Lambda_0(y_0)$ for any $y_0 \in \mathcal{Y}_0$. Next, we bound the difference in the performance of the approximate strategy $\boldsymbol{g}$ and the optimal strategy.

Theorem 3. *Let $L_{\hat{V}_{t+1}}$ be the Lipschitz constant of $\hat{V}_{t+1}(\cdot)$ for all $t = 0, \dots, T$. Then, for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}_t$,*

$$|Q_t(m_t, u_t) - \Theta_t(m_t, u_t)| \leq 2\alpha_t, \tag{21}$$

$$|V_t(m_t) - \Lambda_t(m_t)| \leq 2\alpha_t, \tag{22}$$

where $\alpha_T = \epsilon_T$ and $\alpha_t = \alpha_{t+1} + L_{\hat{V}_{t+1}} \cdot \delta_t$ for all $t = 0, \dots, T-1$.

Proof. For all $t = 0, \dots, T-1$ and for each $\hat{\pi}_t \in \hat{\mathcal{P}}_t$ and $u_t \in \mathcal{U}_t$, let $\hat{\Theta}_t\ (\hat{\pi}_t, u_t) := \max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1} | \hat{\pi}_t, u_t]]}$ $\hat{\Lambda}_{t+1}(\hat{\pi}_{t+1})$ and $\hat{\Lambda}_t(\hat{\pi}_t) := \hat{\Theta}_t(\hat{\pi}_t, \hat{g}_t(\hat{\pi}_t))$. At $t = T$, $\hat{\Theta}_T(\hat{\pi}_T, u_T) := \max_{x_T \in [[X_T | \hat{\pi}_T]]} c_T(x_T, u_T)$ and $\hat{\Lambda}_T$ $(\hat{\pi}_T) := \hat{\Theta}_T(\hat{\pi}_T, \hat{g}_T(\hat{\pi}_T))$. Note that $\hat{\Theta}_t(\hat{\pi}_t, u_t) = \hat{Q}_t(\hat{\pi}_t, u_t)$ and $\hat{\Lambda}_t(\hat{\pi}_t) = \hat{V}_t(\hat{\pi}_t)$ for all t since $\hat{g}_t(\hat{\pi}_t) = \arg\min_{u_t \in \mathcal{U}_t} \hat{Q}_t(\hat{\pi}_t, u_t)$. Next, at each $t = 0, \dots, T$, we use the triangle inequality in the LHS of (21):

$$\begin{aligned} |Q_t(m_t, u_t) - \Theta_t(m_t, u_t)| &\leq |Q_t(m_t, u_t) - \\ \hat{Q}_t(\hat{\sigma}_t(m_t), u_t)| &+ |\hat{\Theta}_t(\hat{\sigma}_t(m_t), u_t) - \Theta_t(m_t, u_t)| \\ &\leq \alpha_t + |\hat{\Theta}_t(\hat{\sigma}_t(m_t), u_t) - \Theta_t(m_t, u_t)|, \end{aligned} \tag{23}$$

where, in the second inequality, we use (16) from Theorem 2. Next, we prove that $|\hat{\Theta}_t(\hat{\sigma}_t(m_t), u_t) - \Theta_t(m_t, u_t)| \leq \alpha_t$ and $|\hat{\Lambda}_t(\hat{\sigma}_t(m_t)) - \Lambda_t(m_t)| \leq \alpha_t$ for all $t = 0, \dots, T$ using backward mathematical induction starting at time T. At time T, from (10) in Definition 2, $|\hat{\Theta}_T(\hat{\sigma}_T(m_T), u_T) - \Theta_T(m_T, u_T)| = |\max_{x_T \in [[X_T | \hat{\sigma}_T(m_T)]]}$ $c_T(x_T, u_T) - \max_{x_T \in [[X_T | m_T]]} c_T(x_T, u_T)| \leq \epsilon_T$. Furthermore, $|\hat{\Lambda}_T(\hat{\sigma}_T(m_T)) - \Lambda_T(m_T)| = |\hat{\Theta}_T(\hat{\sigma}_T(m_T), \hat{g}_T(\hat{\sigma}_T(m_T))) - \Theta_T(m_T, g_T(m_T))| \leq \epsilon_T$, where the inequality holds because $g_T(m_T) = \hat{g}_T(\hat{\sigma}_T(m_T))$. With this as the basis, for any $t = 0, \dots, T-1$, we consider the induction hypothesis $|\hat{\Lambda}_{t+1}(\hat{\sigma}_{t+1}(m_{t+1})) - \Lambda_{t+1}(m_{t+1})| \leq \alpha_{t+1}$. Then, using the definitions of the value functions, $|\hat{\Theta}_t(\hat{\sigma}_t(m_t), u_t) - \Theta_t(m_t, u_t)| = |\max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1} | \hat{\sigma}_t(m_t), u_t]]} \hat{\Lambda}_{t+1}(\hat{\pi}_{t+1}) - \max_{m_{t+1} \in [[M_{t+1} | m_t, u_t]]} \Lambda_{t+1}(m_{t+1})|$. Next, we expand the RHS using the triangle inequality to write that

$$\begin{aligned} |\hat{\Theta}_t(\hat{\sigma}_t(m_t), u_t) - \Theta_t(m_t, u_t)| &\leq \Big| \max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1} | \hat{\sigma}_t(m_t), u_t]]} \\ \hat{\Lambda}_{t+1}(\hat{\pi}_{t+1}) - \max_{\hat{\pi}_{t+1} \in [[\hat{\Pi}_{t+1} | m_t, u_t]]} \hat{\Lambda}_{t+1}(\hat{\pi}_{t+1})\Big| &+ \Big| \max_{m_{t+1} \in [[M_{t+1} | m_t, u_t]]} \\ \Lambda_{t+1}(m_{t+1}) - \max_{\hat{\sigma}_{t+1}(m_{t+1}) \in [[\hat{\Pi}_{t+1} | m_t, u_t]]} &\hat{\Lambda}_{t+1}(\hat{\sigma}_{t+1}(m_{t+1}))\Big| \\ &\leq L_{\hat{V}_{t+1}} \cdot \delta_t + \alpha_{t+1} = \alpha_t, \end{aligned} \tag{24}$$

where, in the second inequality, the first term is upper bounded by noting that $\hat{\Lambda}_{t+1}(\hat{\pi}_{t+1}) = \hat{V}_{t+1}(\hat{\pi}_{t+1})$, using (13) from Lemma 1 and then using (10) from Definition 2, whereas the second term is bounded using the induction hypothesis. Following the same sequence of arguments as time T, $|\hat{\Lambda}_t(\hat{\sigma}_t(m_t)) - \Lambda_t(m_t)| \leq \alpha_t$ as a direct consequence of (24). This completes the induction. The proof is complete by substituting this result into the RHS of (23) to show (21) and consequently, (22). □

A. Examples

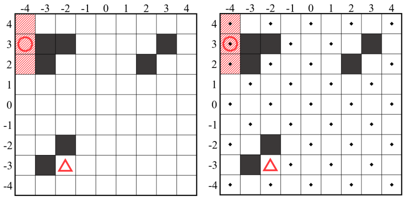
In this subsection, we present two approximate information states which satisfy Definition 2. They are both inspired by state quantization [21]. Specifically, at any $t = 0, \dots, T$, let $\mathcal{X}_t$ be the set of feasible states. Then, a subset $\hat{\mathcal{X}}_t \subset \mathcal{X}_t$ is a *set of quantized states* with parameter $\gamma_t \in \mathbb{R}_{\geq 0}$ if $\max_{x_t \in \mathcal{X}_t} \min_{\hat{x}_t \in \hat{\mathcal{X}}_t} d(x_t, \hat{x}_t) \leq \gamma_t$. The corresponding *quantization function* $\mu_t : \mathcal{X}_t \to \hat{\mathcal{X}}_t$ is defined as $\mu_t(x_t) := \arg\min_{\hat{x}_t \in \hat{\mathcal{X}}_t} d(x_t, \hat{x}_t)$. Note that by construction, $d(x_t, \mu_t(x_t)) \leq \gamma_t$ for all $x_t \in \mathcal{X}_t$, for all t.

1) Perfectly Observed Systems: Consider a system where $Y_t = X_t$ for all $t = 0, \dots, T$. Recall from Subsection III-B that the information state is simply $\Pi_t = X_t$ and it takes values in $\mathcal{X}_t$ for all t. Then, an approximate information state for such a system is the quantized state $\hat{\Pi}_t := \mu_t(X_t)$ with $\epsilon_T = 2L_{c_T} \cdot \gamma_T$ and $\delta_t = 2\gamma_{t+1} + 2L_{f_t} \cdot \gamma_t$, where $\gamma_{T+1} = 0$, and L_{c_T} and L_{f_t} are the Lipschitz constants for $c_T(\cdot)$ and $f_t(\cdot)$, respectively. The derivation for the values of ϵ_T and δ_t can be found in Appendix A of our online preprint [22].

2) Partially Observed Systems: For a partially observed system, recall from Section III-B that an information state is given by the conditional range $\Pi_t = [[X_t | m_t]]$. We approximate the conditional range by quantizing each element in Π_t. This is generated by the mapping $\nu_t : 2^{\mathcal{X}_t} \to 2^{\hat{\mathcal{X}}_t}$ which yields the approximate range $\nu_t(\Pi_t) := \{\mu_t(x_t) \in \hat{\mathcal{X}}_t | x_t \in \Pi_t\}$. Then, $\hat{\Pi}_t = \nu_t(\Pi_t)$ is an approximate information state for partially observed systems for all $t = 0, \dots, T$ with $\epsilon_T = 2L_{c_T} \cdot \gamma_T$ and $\delta_t = 2\gamma_{t+1} + 2L_{\bar{f}_t} \cdot L_{h_{t+1}} \cdot L_{f_t} \cdot \gamma_t$, where $\gamma_{T+1} = 0$, and L_{c_T}, $L_{\bar{f}_t}$, $L_{h_{t+1}}$ and L_{f_t} are Lipschitz constants of $c_T(\cdot)$, $\bar{f}_t(\cdot)$, $h_{t+1}(\cdot)$, and $f_t(\cdot)$, respectively. The derivation of the values of ϵ_T and δ_t can be found in Appendix B of our online preprint [22].

V. Numerical Example

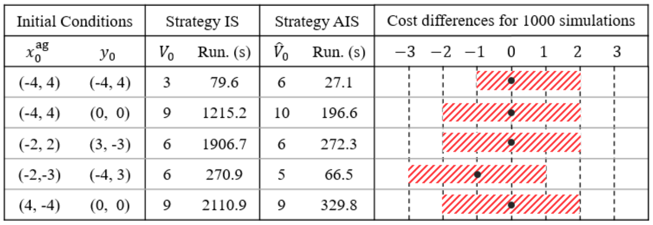
In this section, we present a numerical example illustrating the performance of the approximate conditional range for a gridworld pursuit problem. We consider an agent who seeks to catch a moving target at the end of a time horizon T on 9×9 grid with static obstacles. For all $t = 0, \dots, T$, we denote the position of the agent by X_t^{ag} and the position of the target by X_t^{ta}, each of which takes values in the set of grid cells $\mathcal{X} = \big\{(-4,-4), (-4,-3), \dots, (3,4), (4,4)\big\} \setminus \mathcal{O}$, where $\mathcal{O} \subset \mathcal{X}$ is the set of obstacle cells. Let $\mathcal{U}_t = \mathcal{W}_t = \mathcal{N}_t = \{(-1,0), (1,0), (0,0), (0,1),\ (0,-1)\}$ for all t. Then, starting at $X_0^{\text{ta}} \in \mathcal{X}$, the target's position is updated as $X_{t+1}^{\text{ta}} = \mathbb{I}(X_t^{\text{ta}} + W_t \in \mathcal{X}) \cdot (X_t^{\text{ta}} + W_t) + (1 - \mathbb{I}(X_t^{\text{ta}} + W_t \in \mathcal{X})) \cdot X_t^{\text{ta}}$, where $W_t \in \mathcal{W}_t$ and $\mathbb{I}(\cdot)$ is the indicator function. At each t, the agent receives an observation of the target $Y_t = \mathbb{I}(X_t^{\text{ta}} + N_t \in \mathcal{X}) \cdot (X_t^{\text{ta}} + N_t) + (1 - \mathbb{I}(X_t^{\text{ta}} + N_t \in \mathcal{X})) \cdot X_t^{\text{ta}}$, where $N_t \in \mathcal{N}_t$. The agent observes their own position perfectly. Then, the agent selects an action $U_t \in \mathcal{U}_t$ and updates their position to $X_{t+1}^{\text{ag}} = \mathbb{I}(X_t^{\text{ag}} + U_t \in \mathcal{X}) \cdot (X_t^{\text{ag}} + U_t) + (1 - \mathbb{I}(X_t^{\text{ag}} + U_t \in \mathcal{X})) \cdot X_t^{\text{ag}}$. The terminal cost after T time steps is $d(X_T^{\text{ta}}, X_T^{\text{ag}})$, where $d(\cdot,\cdot)$ is the shortest path between two cells while avoiding all obstacles. The distance between two adjacent cells is 1 unit. The grid and an initial state are illustrated in Fig. 1(a). Here, the black colored cells mark the obstacles. The red triangle and the red circle indicate the position and observation of the agent, respectively, at $t = 0$. The red hatched region indicates the possible locations of the target at $t = 0$.



(a) The original grid (b) The quantized grid

Fig. 1. The gridworld pursuit problem with the initial conditions $x_0^{\text{ag}} = (-2, -3)$ and $y_0 = (-4, 3)$.

Recall from Subsection III-B that an information state at time t is $\Pi_t = \big(X_t^{\text{ag}}, [[X_t^{\text{ta}}|M_t]]\big) \in \mathcal{X} \times 2^{\mathcal{X}}$. We construct an approximation of the conditional range $[[X_t^{\text{ta}}|M_t]]$ at time t using the quantization approach from Subsection IV-A. The set of quantized states $\hat{\mathcal{X}}$, with $\gamma_t = 1$ for all t, is illustrated in Fig. 1(b) by marking the relevant cells with dots. Recall that $\mu_t(x_t) = \arg\min_{\hat{x}_t \in \hat{\mathcal{X}}} d(x_t, \hat{x}_t)$ and the approximate range at time t is $\hat{A}_t = \big\{\mu_t(x_t) \in \hat{\mathcal{X}} | x_t \in [[X_t^{\text{ta}}|M_t]]\big\}$. We consider the approximate information state $\hat{\Pi}_t = \big(X_t^{\text{ag}}, \hat{A}_t, Y_0\big) \in \mathcal{X} \times 2^{\hat{\mathcal{X}}} \times \mathcal{X}$ for all t. The initial observation Y_0 in $\hat{\Pi}_t$ makes the prediction of $\hat{A}_{t+1}$ more accurate. For five initial conditions, we compute the best control strategy for $T = 6$ using both the information state (IS) and the approximate information state (AIS). In Fig. 2, we present the worst-case costs for both the DPs (V_0 and $\hat{V}_0$) and the computational times (Run.) in seconds. Note that the approximate DP has a significantly faster runtime. We also implement both strategies on the system with random disturbances. In Fig. 2, we illustrate differences in *actual* costs across 1000 such simulations when implementing the approximate strategy and the optimal strategy. The dots indicate the most frequently observed difference for each case. Note that the difference in costs is bounded.



Initial Conditions		Strategy IS		Strategy AIS		Cost differences for 1000 simulations
x_0^{ag}	y_0	V_0	Run. (s)	$\hat{V}_0$	Run. (s)	−3 −2 −1 0 1 2 3
(-4, 4)	(-4, 4)	3	79.6	6	27.1	
(-4, 4)	(0, 0)	9	1215.2	10	196.6	
(-2, 2)	(3, -3)	6	1906.7	6	272.3	
(-2,-3)	(-4, 3)	6	270.9	5	66.5	
(4, -4)	(0, 0)	9	2110.9	9	329.8	

Fig. 2. Results of numerical simulations for $T = 6$.

VI. Conclusion

In this paper, we presented a theory of information states for non-stochastic control problems. We characterized the information states by their properties and proved that they can be used to derive the optimal control strategy. Then, we proposed a definition for an approximate information state which yielded an approximate DP. We provide upper bounds on the approximation error if an agent uses the approximately optimal strategy to generate their control actions. Future work should consider constructing approximate information states using only partial knowledge of system dynamics.

References

[1] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.

[2] A. Dave and A. A. Malikopoulos, "The Prescription Approach for Decentralized Stochastic Control with Word-of-Mouth Communication," *preprint, arXiv:1907.12125*, 2021.

[3] P. R. Kumar and P. P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[4] P. Bernhard, "Minimax - or feared value - L_1 / L_∞ control," *Theoretical computer science*, vol. 293, no. 1, pp. 25–44, 2003.

[5] J. Subramanian and A. Mahajan, "Approximate information state for partially observed systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 1629–1636, IEEE, 2019.

[6] M. Littman and R. S. Sutton, "Predictive representations of state," *Advances in neural information processing systems*, vol. 14, 2001.

[7] J. Subramanian, A. Sinha, R. Seraj, and A. Mahajan, "Approximate information state for approximate planning and reinforcement learning in partially observed systems," *Journal of Machine Learning Research*, vol. 23, no. 12, pp. 1–83, 2022.

[8] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1644–1658, 2013.

[9] A. A. Malikopoulos, "On team decision problems with nonclassical information structures," *IEEE Transactions on Automatic Control*, 2022 (in press) arXiv:2101.10992.

[10] A. Dave, N. Venkatesh, and A. A. Malikopoulos, "On decentralized control of two agents with nested accessible information," in *2022 American Control Conference (ACC)*, pp. 3423–3430, IEEE, 2022.

[11] A. D. Kara and S. Yuksel, "Near optimality of finite memory feedback policies in partially observed markov decision processes," *preprint arXiv:2010.07452*, 2020.

[12] D. Bertsekas and I. Rhodes, "Sufficiently informative functions and the minimax feedback control of uncertain dynamic systems," *IEEE Transactions on Automatic Control*, vol. 18, no. 2, pp. 117–124, 1973.

[13] P. Bernhard, "A separation theorem for expected value and feared value discrete time control," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 1, pp. 191–206, 1996.

[14] T. Başar and P. Bernhard, *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.

[15] M. Rasouli, E. Miehling, and D. Teneketzis, "A scalable decomposition method for the dynamic defense of cyber networks," in *Game Theory for Security and Risk Management*, pp. 75–98, Springer, 2018.

[16] M. Gagrani and A. Nayyar, "Decentralized minimax control problems with partial history sharing," in *2017 American Control Conference (ACC)*, pp. 3373–3379, IEEE, 2017.

[17] M. Gagrani, Y. Ouyang, M. Rasouli, and A. Nayyar, "Worst-case guarantees for remote estimation of an uncertain source," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1794–1801, 2020.

[18] A. Dave, N. Venkatesh, and A. A. Malikopoulos, "On decentralized minimax control with nested subsystems," in *2022 American Control Conference (ACC)*, pp. 3437–3444, IEEE, 2022.

[19] G. N. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Transactions on automatic control*, vol. 58, no. 6, pp. 1497–1510, 2013.

[20] M. F. Barnsley, *Superfractals*. Cambridge University Press, 2006.

[21] D. Bertsekas, "Convergence of discretization procedures in dynamic programming," *IEEE Transactions on Automatic Control*, vol. 20, no. 3, pp. 415–419, 1975.

[22] A. Dave, N. Venkatesh, and A. A. Malikopoulos, "Approximate information states for worst-case control of uncertain systems," *arXiv preprint arXiv:2203.15271*, 2022.