2023 American Control Conference (ACC)
San Diego, CA, USA. May 31 - June 2, 2023

On Robust Control of Partially Observed Uncertain Systems with Additive Costs

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

Abstract— **In this paper, we consider the problem of optimizing the worst-case behavior of a partially observed system. All uncontrolled disturbances are modeled as finite-valued uncertain variables. Using the theory of cost distributions, we present a dynamic programming (DP) approach to compute a control strategy that minimizes the maximum possible total cost over a given time horizon. To improve the computational efficiency of the optimal DP, we introduce a general definition for information states and show that many information states constructed in previous research efforts are special cases of ours. Additionally, we define approximate information states and an approximate DP that can further improve computational tractability by conceding a bounded performance loss. We illustrate the utility of these results using a numerical example.**

I. Introduction

In engineering applications, it is common for an agent to operate with limited knowledge of the system state and uncertain system dynamics [1]. This decision-making challenge is typically modeled as a stochastic control problem, where the agent computes a control strategy to minimizes an expected total cost across a time horizon given a prior probability distribution for all uncertainties. This approach has also been utilized in reinforcement learning [2] and decentralized systems [3]. However, the expected total cost may not be an adequate measure of performance in all situations. In fact, many applications require guarantees on a system's worst-case performance, for example: (1) control of systems under attack from an adversary, like cyber-security systems [4], and (2) control of systems where a single event of failure can be damaging, like water reservoirs [5]. Furthermore, the performance of a stochastic control strategy degrades rapidly with a mismatch between the assumed prior distribution and the actual underlying distribution [6]. Consequently, stochastic models are unsuitable for strategy computation when prior distributions are ambigious.

For such applications, we can instead utilize a non-stochastic formulation, where the agent only has access to the feasible sets for all uncertainties, without knowledge of probability distributions. This non-stochastic approach has been utilized in robust control [7]–[9], information theory [10], [11], reinforcement learning [12], [13], and decentralized systems [14], [15]. In this paper, we focus on a centralized non-stochastic control problem where an agent seeks a control strategy to minimize a *maximum* possible cost over a finite-time horizon. It is known that the optimal strategy in such problems can be computed with an offline dynamic program (DP) [8]. However, the growth in the agent's memory with time makes this challenging because the agent's action is a function of the memory and thus, the DP requires solving one optimization problem at each time for each possible realization of the memory. Using an *information state* can address this challenge. Two well known non-stochastic information states are the *conditional range* for terminal cost problems [16], [17] and the *maximum cost-to-come* for additive cost problems [7], [8]. In robust stochastic problems [18] concerns of partial observation have also been addressed using a conditional range [19]. Generalized approximate information states for terminal cost problems were developed in [20]. However, to the best of our knowledge, there is no notion of approximate information states for non-stochastic *additive* cost problems.

The main contributions of this paper are: (1) for additive cost problems, we introduce general information states to compute an optimal strategy (Theorem 2), and (2) we define approximate information states to compute an approximate strategy with a bounded performance loss (Theorems 3 - 4).

The remainder of the paper proceeds as follows. In Section II, we present our model. In Section III, we define information states and the corresponding DP. In Section IV, we define approximate information states, the approximate DP, and derive performance bounds. 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

We use the non-stochastic framework of *uncertain variables* from [10]. For a sample space Ω and a set $\mathscr{X}$, an uncertain variable is a mapping $\mathsf{X} : \Omega \to \mathscr{X}$ written concisely as $\mathsf{X} \in \mathscr{X}$. For any $\omega \in \Omega$, its realization is $\mathsf{X}(\omega) = \mathsf{x} \in \mathscr{X}$. The *marginal range* of an uncertain variable X is the set $[[\mathsf{X}]] := \{\mathsf{X}(\omega) \mid \omega \in \Omega\}$. The *joint range* of two uncertain variables $\mathsf{X} \in \mathscr{X}$ and $\mathsf{Y} \in \mathscr{Y}$ is $[[\mathsf{X}, \mathsf{Y}]] := \{(\mathsf{X}(\omega), \mathsf{Y}(\omega)) \mid \omega \in \Omega\}$. The *conditional range* of X given a realization y of Y is $[[\mathsf{X}|\mathsf{y}]] := \{\mathsf{X}(\omega) \mid \mathsf{Y}(\omega) = \mathsf{y},\ \omega \in \Omega\}$, and $[[\mathsf{X}|\mathsf{Y}]] := \{[[\mathsf{X}|\mathsf{y}]] \mid \mathsf{y} \in [[\mathsf{Y}]]\}$. Next, consider two compact, nonempty subsets $\mathscr{X}, \mathscr{Y}$ of a metric space $(\mathscr{S}, d)$, where $d(\cdot, \cdot)$ is the metric. Then, the Hausdorff distance [21, Chapter 1.12] between the sets is $\mathcal{H}(\mathscr{X}, \mathscr{Y}) := \max\{\max_{\mathsf{x} \in \mathscr{X}} \min_{\mathsf{y} \in \mathscr{Y}} d(\mathsf{x}, \mathsf{y}), \max_{\mathsf{y} \in \mathscr{Y}} \min_{\mathsf{x} \in \mathscr{X}} d(\mathsf{x}, \mathsf{y})\}$.

This research was supported by NSF under Grants CNS-2149520 and CMMI-2219761. 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).

B. Problem Formulation

We consider an agent who controls the evolution of a system over $T \in \mathbb{N}$ discrete time steps. At any time $t = 0, \ldots, T$, the system is denoted by an uncertain variable $X_t \in \mathcal{X}$ and the agent's action is denoted by an uncertain variable $U_t \in \mathcal{U}$. At each t, the system also receives an uncontrolled disturbance $W_t \in \mathcal{W}$. Starting with an initial state $X_0 \in \mathcal{X}$, the state evolves as $X_{t+1} = f_t(X_t, U_t, W_t)$ for all $t = 0, \ldots, T-1$. Before selecting the control action at each t, the agent partially observes the system state as $Y_t = h_t(X_t, N_t) \in \mathcal{Y}$, where $N_t \in \mathcal{N}$ is a noise.

Remark 1. We denote generic uncertain variables by sans-serif upper case alphabets $\mathsf{X} \in \mathscr{X}$ and $\mathsf{Y} \in \mathscr{Y}$, whereas, we denote the state and observation at any t by italicized upper-case alphabets $X_t \in \mathcal{X}$ and $Y_t \in \mathcal{Y}$, respectively.

At each $t = 0, \ldots, T$, the agent stores the history of observations and control actions in their memory, denoted by $M_t := (Y_{0:t}, U_{0:t-1}) \in \mathcal{M}_t$, where $Y_{0:t} := (Y_0, \ldots, Y_t)$. Then, the agent selects an action $U_t = g_t(M_t)$ using a control law $g_t : \mathcal{M}_t \to \mathcal{U}$ and incurs a cost $c_t(X_t, U_t) \in \mathbb{R}_{\geq 0}$. We denote the control strategy by $\boldsymbol{g} := (g_0, \ldots, g_T) \in \mathcal{G}$ and measure its performance using the *worst-case criterion:*

$$\mathcal{J}(\boldsymbol{g}) := \max_{\substack{x_0 \in \mathcal{X}, n_{0:T} \in \mathcal{N}^T, \\ w_{0:T-1} \in \mathcal{W}^{T-1}}} \sum_{t=0}^{T} c_t(X_t, U_t). \tag{1}$$

In (1), we maximize the total cost over all feasible realizations of the *uncontrolled inputs*, i.e., initial state X_0, noises $\{N_t \mid t = 0, \ldots, T\}$, and disturbances $\{W_t \mid t = 0, \ldots, T-1\}$ because they determine all other variables in the system. Next, we state the agent's optimization problem.

Problem 1. We seek to efficiently compute an optimal strategy $\boldsymbol{g}^* = \arg\min_{\boldsymbol{g} \in \mathcal{G}} \mathcal{J}(\boldsymbol{g})$, given the sets $\{\mathcal{X}, \mathcal{U}, \mathcal{Y}, \mathcal{W}, \mathcal{N}\}$ and the functions $\{f_t, h_t, c_t \mid t = 0, \ldots, T\}$.

We impose the following assumptions on our model:

Assumption 1. Each uncontrolled input is independent (see [10, Definintion 2.1]) of all other uncontrolled inputs.

Assumption 1 ensures that the system evolution is Markovian in a non-stochastic sense (see [10, Definintion 2.2]). This assumption will help develop our results.

Assumption 2. Each feasible set $\{\mathcal{X}, \mathcal{U}, \mathcal{Y}, \mathcal{W}, \mathcal{N}\}$ is a finite subset of a metric space $(\mathcal{S}, d)$.

Assumption 2 ensures that all extrema are well defined and that an optimal solution to Problem 1 exists. We will use the metric $d(\cdot, \cdot)$ in Section IV to quantify the distance between two elements in any set.

Assumption 3. All uncertain variables and the cost $c_t(X_t, U_t)$ have a finite maximum value at each t.

Assumption 3, in addition to the finiteness of all feasible sets, ensures that the functions $\{f_t, h_t, c_t \mid t = 0, \ldots, T\}$ are globally Lipschitz . To this end, we will denote the Lipschitz constant of a function f_t by $L_{f_t} \in \mathbb{R}_{\geq 0}$.

III. Dynamic Programs and Information States

In this section, we first present a standard terminal cost DP which can obtain the optimal strategy in Problem 1. Then, in Subsection III-B, we construct a DP which is specialized to the additive cost criterion in (1), and in Subsection III-C, we define information states to simplify it. To begin, we transform Problem 1 into a terminal cost problem by augmenting the state X_t at each t with the *accrued cost*

$$A_t := \sum_{\ell=0}^{t-1} c_\ell(X_\ell, U_\ell), \tag{2}$$

which takes values in a finite set $\mathcal{A}_t \subset \mathbb{R}_{\geq 0}$. Starting with $A_0 := 0$, the accrued cost evolves as $A_{t+1} = A_t + c_t(X_t, U_t)$ for all $t = 0, \ldots, T-1$. Thus, the augmented state (X_t, A_t) evolves as a controlled Markov chain. Furthermore, note that the performance criterion (1) can be written as a function of the terminal augmented state (X_T, A_T), i.e., $\mathcal{J}(\boldsymbol{g}) = \max_{x_0, n_{0:T}, w_{0:T-1}} \big(c_T(X_T, U_T) + A_T\big)$. This construction yields a terminal cost optimization problem in $\boldsymbol{g} \in \mathcal{G}$, where the optimal strategy can be computed using a memory based terminal cost DP [20], as follows. For all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}$, for all $t = 0, \ldots, T-1$, we define the value functions

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

$$V_t^{\text{tm}}(m_t) := \min_{u_t \in \mathcal{U}} Q_t^{\text{tm}}(m_t, u_t), \tag{4}$$

where, at time T, $Q_T^{\text{tm}}(m_T, u_T) := \max_{a_T, x_T \in [[A_T, X_T | m_T, u_T]]} \big(c_T(x_T, u_T) + a_T\big)$ and $V_T^{\text{tm}}(m_T) := \min_{u_T \in \mathcal{U}} Q_T(m_T, u_T)$. The control law at each t is $g_t^{\text{tm}}(m_t) := \arg\min_{u_t \in \mathcal{U}} Q_t(m_t, u_t)$. Using standard arguments, we can conclude that the resulting control strategy $\boldsymbol{g}^{\text{tm}} = (g_0^{\text{tm}}, \ldots, g_T^{\text{tm}})$ is an optimal solution to the terminal cost problem as well as Problem 1 [16]. However, note that the right hand side (RHS) of (4) involves solving a minimization problem for each possible realization $m_t \in \mathcal{M}_t$, at each t. The number of possible realizations $|\mathcal{M}_t|$ increases with time as the agent receives more observations, and consequently, the DP requires a large number of computations for a longer horizon T. To address this, we formulate a DP specialized for additive cost problems in Subsection III-B and simplify it using *information states* in Subsection III-C. We will show (Remark 2) that the specialized DP allows us to define more computationally efficient information states than (3) - (4). To this end, we present a theory of cost distributions in the next subsection which is required to construct the specialized DP.

A. Cost distributions

In this subsection, we develop the mathematical framework of *cost distributions* for finite uncertain variables. Cost distributions were originally defined for $(\max, +)$ algebra [22], and applied to robust control problems [8], [9] independently from the framework of uncertain variables. A cost distribution is a non-stochastic analogue of a probability distribution. Specifically, for a finite sample space Ω with a sigma algebra $\mathcal{B}(\Omega)$, a cost distribution is a function $q : \mathcal{B}(\Omega) \to \{-\infty\} \cup (-\infty, 0]$ satisfying the properties: (1)

$q(\Omega) = 0$, (2) $q(\emptyset) = -\infty$, and (3) $q(B) = \max_{\omega \in B} q(\omega)$ for all $B \in \mathcal{B}(\Omega)$, where, by convention, the maximum over an empty set is $-\infty$. Furthermore, for two sets $B^1, B^2 \in \mathcal{B}(\Omega)$ with $q(B^2) > -\infty$, the conditional cost distribution of B^1 given B^2 is $q(B^1|B^2) := q(B^1, B^2) - q(B^2)$, where $q(B^1, B^2) = \max_{\omega \in B^1 \cap B^2} q(\omega)$. Next, we extend this definition to include finite uncertain variables.

Definition 1. Let $\mathsf{X} : \Omega \to \mathscr{X}$ and $\mathsf{Y} : \Omega \to \mathscr{Y}$ be two finite uncertain variables. The *cost distribution* for any realization $\mathsf{x} \in \mathscr{X}$ is $q(\mathsf{x}) := \max_{\omega \in \{\Omega | \mathsf{X}(\omega) = \mathsf{x}\}} q(\omega)$, and that for any $\mathsf{x} \in \mathscr{X}$ given a realization $\mathsf{y} \in \mathscr{Y}$ with $q(\mathsf{y}) > -\infty$ is $q(\mathsf{x}|\mathsf{y}) = q(\mathsf{x}, \mathsf{y}) - q(\mathsf{y})$, where $q(x, y) = \max_{\omega \in \{\Omega | X(\omega) = x, Y(\omega) = y\}} q(\omega)$.

Any cost distribution given by Definition 1 satisfies the following useful properties.

Lemma 1. *Let* $(\Omega, \mathcal{B}(\Omega))$ *have a cost distribution* $q : \mathcal{B}(\Omega) \to \{-\infty\} \cup (-\infty, 0]$. *Let* $\mathsf{X} : \Omega \to \mathscr{X}$ *and* $\mathsf{Y} : \Omega \to \mathscr{Y}$ *be two finite uncertain variables and let* $f : \mathscr{X} \to \mathscr{Y}$ *such that* $\mathsf{Y} = f(\mathsf{X})$ *and* $f^{-1}(\mathsf{y}) \neq \emptyset$ *for all* $\mathsf{y} \in \mathscr{Y}$. *Then,*

$$q(\mathsf{y}) = \max_{\mathsf{x} \in \{\mathscr{X} | f(\mathsf{x}) = \mathsf{y}\}} q(\mathsf{x}), \quad \forall \mathsf{y} \in \mathscr{Y}, \tag{5}$$

and furthermore, for any function $g : \mathscr{Y} \to \mathbb{R}_{\geq 0}$,

$$\max_{\mathsf{x} \in \mathscr{X}} \big(g(f(\mathsf{x})) + q(\mathsf{x}) \big) = \max_{\mathsf{y} \in \mathscr{Y}} \big(g(\mathsf{y}) + q(\mathsf{y}) \big). \tag{6}$$

Proof. Using Definition 1, $q(\mathsf{y}) = \max_{\omega \in \{\Omega | \mathsf{Y}(\omega) = \mathsf{y}\}} q(\omega)$, where $\{\Omega \mid \mathsf{Y}(\omega) = \mathsf{y}\} = \cup_{\mathsf{x} \in \{\mathscr{X} | f(\mathsf{x}) = \mathsf{y}\}} \{\Omega \mid \mathsf{X}(\omega) = \mathsf{x}\}$. This implies that $q(\mathsf{y}) = \max_{\mathsf{x} \in \{\mathscr{X} | f(\mathsf{x}) = \mathsf{y}\}} \max_{\omega \in \{\Omega | \mathsf{X}(\omega) = \mathsf{x}\}} q(\omega) = \max_{\mathsf{x} \in \{\mathscr{X} | f(\mathsf{x}) = \mathsf{y}\}} q(\mathsf{x})$, where, in the second equality, we used Definition 1. This proves (5). Next, we use (5) in the RHS of (6) as $\max_{\mathsf{y} \in \mathscr{Y}} (g(\mathsf{y}) + q(\mathsf{y})) = \max_{\mathsf{y} \in \mathscr{Y}} (g(\mathsf{y}) + \max_{\mathsf{x} \in \{\mathscr{X} | f(\mathsf{x}) = \mathsf{y}\}} q(\mathsf{x})) = \max_{\mathsf{y} \in \mathscr{Y}} \max_{\mathsf{x} \in \{\mathscr{X} | f(\mathsf{x}) = \mathsf{y}\}} (g(f(\mathsf{x})) + q(\mathsf{x})) = \max_{\mathsf{x} \in \mathscr{X}} (g(f(\mathsf{x})) + q(\mathsf{x}))$, which completes the proof for (6). ☐

B. Specialized Dynamic Program

In this subsection, we construct a specialized DP decomposition for Problem 1 using two specific cost distributions, the first of which is an indicator function.

Definition 2. Let $\mathsf{X} \in \mathscr{X}$ and $\mathsf{Y} \in \mathscr{Y}$ be two finite uncertain variables. The *indicator function* for any $\mathsf{x} \in \mathscr{X}$ is given by

$$\mathbb{I}(\mathsf{x}) := \begin{cases} 0, & \text{if } \mathsf{x} \in [[\mathsf{X}]], \\ -\infty, & \text{if } \mathsf{x} \notin [[\mathsf{X}]], \end{cases} \tag{7}$$

and the conditional indicator function for any $\mathsf{x} \in \mathscr{X}$ given a realization $\mathsf{y} \in \mathscr{Y}$ with $\mathbb{I}(\mathsf{y}) > -\infty$ is

$$\mathbb{I}(\mathsf{x}|\mathsf{y}) := \begin{cases} 0, & \text{if } \mathsf{x} \in [[\mathsf{X}|\mathsf{y}]], \\ -\infty, & \text{if } \mathsf{x} \notin [[\mathsf{X}|\mathsf{y}]]. \end{cases} \tag{8}$$

The indicator function $\mathbb{I}$ can be shown to satisfy the conditions in Definition 1 and thus, it constitutes a valid cost distribution. In addition to Lemma 1, for two uncertain variables $\mathsf{X} \in \mathscr{X}$ and $\mathsf{Y} \in \mathscr{Y}$ and any function $f : \mathscr{X} \to \mathbb{R}$,

$$\max_{\mathsf{x} \in [[\mathsf{X}|\mathsf{y}]]} f(\mathsf{x}) = \max_{\mathsf{x} \in \mathscr{X}} \big(f(\mathsf{x}) + \mathbb{I}(\mathsf{x}|\mathsf{y}) \big), \quad \forall \mathsf{y} \in \mathscr{Y}. \tag{9}$$

We also require the *accrued distribution* for an uncertain variable at each t, defined using the accrued cost $A_t \in \mathcal{A}_t$.

Definition 3. Let $\mathsf{X} \in \mathscr{X}$ and $\mathsf{Y} \in \mathscr{Y}$ be two finite uncertain variables and let $A_t \in \mathcal{A}_t$ be the accrued cost at any $t = 0, \dots, T$. An *accrued distribution* at any t for any $\mathsf{x} \in \mathscr{X}$ is a function $r_t : \mathscr{X} \to \{-\infty\} \cup [-a_t^{\max}, 0]$, given by

$$r_t(\mathsf{x}) := \max_{a_t \in \mathcal{A}_t} \big(a_t + \mathbb{I}(\mathsf{x}, a_t) \big) - \max_{a_t \in \mathcal{A}_t} \big(a_t + \mathbb{I}(a_t) \big), \tag{10}$$

and for $\mathsf{x} \in \mathscr{X}$ given a realization $\mathsf{y} \in \mathscr{Y}$, $\mathbb{I}(\mathsf{y}) > -\infty$, it is a function $r_t : \mathscr{X} \times \mathscr{Y} \to \{-\infty\} \cup [-a_t^{\max}, 0]$, given by

$$r_t(\mathsf{x}|\mathsf{y}) := \max_{a_t \in \mathcal{A}_t} \big(a_t + \mathbb{I}(\mathsf{x}, a_t|\mathsf{y}) \big) - \max_{a_t \in \mathcal{A}_t} \big(a_t + \mathbb{I}(a_t|\mathsf{y}) \big), \tag{11}$$

where $a_t^{\max} := \max \mathcal{A}_t$.

At each $t = 0, \dots, T$, note that the accrued distribution $r_t(\mathsf{x}|\mathsf{y}) = -\infty$ if $\mathsf{x} \notin [[\mathsf{X}|\mathsf{y}]]$ whereas $r_t(\mathsf{x}|\mathsf{y}) \in [-a_t^{\max}, 0]$ if $\mathsf{x} \in [[\mathsf{X}|\mathsf{y}]]$. It satisfies the properties to be a valid cost distribution. Furthermore, we can compute the conditional range $[[X_t, M_{t+1}|m_t, u_t]]$ at any t given the realizations $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}$. Subsequently, we can use Definitions 2 - 3 to derive the accrued distribution $r_t(x_t, m_{t+1}|m_t, u_t)$, for all $x_t \in \mathcal{X}$ and $m_{t+1} \in \mathcal{M}_{t+1}$. Then, we use it in the specialized DP decomposition for Problem 1 as follows. For all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}$, for all $t = 0, \dots, T-1$, we define

$$Q_t(m_t, u_t) := \max_{x_t \in \mathcal{X}, m_{t+1} \in \mathcal{M}_{t+1}} \big(c_t(x_t, u_t) + V_{t+1}(m_{t+1}) + r_t(x_t, m_{t+1}|m_t, u_t) \big), \tag{12}$$

$$V_t(m_t) := \min_{u_t \in \mathcal{U}} Q_t(m_t, u_t), \tag{13}$$

where, at time T, $Q_T(m_T, u_T) := \max_{x_T \in \mathcal{X}} \big(c_T(x_T, u_T) + r_T(x_T|m_T) \big)$ and $V_T(m_T) := \min_{u_T \in \mathcal{U}} Q_T(m_T, u_T)$. We define the corresponding control law at time t as $g_t^*(m_t) := \arg\min_{u_t \in \mathcal{U}} Q_t(m_t, u_t)$ and the control strategy as $\boldsymbol{g}^* = (g_0^*, \dots, g_T^*)$. Next, we show that solving the DP (12) - (13) computes the optimal performance and control strategy.

Due to a lack of space, the proofs for all subsequent results have been archived in our online preprint [23].

Theorem 1. *For all* $m_t \in \mathcal{M}_t$ *and* $u_t \in \mathcal{U}$, *for all* $t = 0, \dots, T$,

$$Q_t^{tm}(m_t, u_t) = Q_t(m_t, u_t) + \max_{a_t \in [[A_t|m_t]]} a_t, \tag{14}$$

$$V_t^{tm}(m_t) = V_t(m_t) + \max_{a_t \in [[A_t|m_t]]} a_t, \tag{15}$$

and furthermore, $\boldsymbol{g}^*$ *is an optimal solution to Problem 1.*

Proof. See Appendix A of our online preprint [23]. ☐

Thoerem 1 establishes that the specialized DP (12) - (13) computes an optimal solution to Problem 1. Note that at each t, the optimization in the RHS of (13) must still be solved for each possible $m_t \in \mathcal{M}_t$, in a manner similar to (3) - (4) Thus, we still require a large number of computations for longer time horizons. In the next subsection, we define *information states* to address this concern.

C. Information States

In this subsection, we introduce information states to construct an optimal DP decomposition for Problem 1.

Definition 4. An *information state* at any $t = 0, \dots, T$ is an uncertain variable $\Pi_t = \sigma_t(M_t)$ taking values in a finite set $\mathcal{P}_t$, where $\sigma_t : \mathcal{M}_t \to \mathcal{P}_t$. Furthermore, for all t, for all $m_t \in \mathcal{M}_t$, $u_t \in \mathcal{U}$, $x_t \in \mathcal{X}$ and $\pi_{t+1} \in \mathcal{P}_{t+1}$, it satisfies:

$$r_t(x_t, \pi_{t+1}|m_t, u_t) = r_t(x_t, \pi_{t+1}|\sigma_t(m_t), u_t), \quad t = 0, \dots, T-1, \quad (16)$$

$$r_T(x_T|m_T) = r_T(x_T|\sigma_t(m_T)). \quad (17)$$

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

$$\bar{Q}_t(\pi_t, u_t) := \max_{x_t \in \mathcal{X}, \pi_{t+1} \in \mathcal{P}_{t+1}} \big(\bar{V}_{t+1}(\pi_{t+1}) + c_t(x_t, u_t) + r_t(x_t, \pi_{t+1}|\pi_t, u_t)\big), \quad (18)$$

$$\bar{V}_t(\pi_t) := \min_{u_t \in \mathcal{U}} \bar{Q}_t(\pi_t, u_t), \quad (19)$$

where, at time T, $\bar{Q}_T(\pi_T, u_T) := \max_{x_T \in \mathcal{X}} \big(c_T(x_T, u_T) + r_T(x_T|\pi_T)\big)$ and $\bar{V}_T(\pi_T) := \min_{u_T \in \mathcal{U}} \bar{Q}_T(\pi_T, u_T)$. The control law at each t is $\bar{g}_t^*(\pi_t) := \arg\min_{u_t \in \mathcal{U}} \bar{Q}_t(\pi_t, u_t)$. Next, we prove that the information state based DP (18) - (19) yields the same value as the specialized DP (12) - (13).

Theorem 2. *Let $\Pi_t = \sigma_t(M_t)$ be an information state at each $t = 0, \dots, T$. Then, for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}$, $Q_t(m_t, u_t) = \bar{Q}_t(\sigma_t(m_t), u_t)$ and $V_t(m_t) = \bar{V}_t(\sigma_t(m_t))$.*

Proof. See Appendix B of our online preprint [23]. ☐

From Theorem 2, the strategy $\bar{\boldsymbol{g}}^* = (\bar{g}_0^*, \dots, \bar{g}_T^*)$ using information states is an optimal solution to Problem 1. In practice, using information states to compute $\bar{\boldsymbol{g}}^*$ is more tractable than using the memory to compute $\boldsymbol{g}^*$ only when the set $\mathcal{P}_t$ has fewer elements than $\mathcal{M}_t$ for most instances of t. This is usually true for systems with long time horizons.

D. Examples of Information States

In this subsection, we present examples of information states which satisfy the conditions in Definition 4.

1) Partially observed systems: Generally, at each $t = 0, \dots, T$ a valid information state which satisfies Definition 4 is the function valued uncertain variable $\Pi_t : \mathcal{X} \to \{-\infty\} \cup [-a_t^{\max}, 0]$. At time t, for a given $m_t \in \mathcal{M}_t$, the realization of Π_t is $p_t(x_t) := r_t(x_t|m_t) = \max_{a_t \in \mathcal{A}_t} \big(a_t + \mathbb{I}(x_t, a_t|m_t)\big) - \max_{a_t \in \mathcal{A}_t} \big(a_t + \mathbb{I}(a_t|m_t)\big)$ for all $x_t \in \mathcal{X}$. Note that this can be interpreted as a normalization [9] of the standard information state from [7], [8].

2) Perfectly observed systems: Consider a system where $Y_t = X_t$ for all t. An information state for such a system is $\Pi_t = X_t$ at each t, i.e, the state itself. This information state is simpler than the one in Case 1.

3) Systems with action dependent costs: Consider a partially observed system where at each t the cost has the form $c_t(U_t) \in \mathbb{R}_{\geq 0}$, and the terminal cost is $c_T(X_T, U_T)$. Then, an information state is the conditional range $\Pi_t = [[X_t|M_t]]$ at each t (see Appendix C of our online preprint [23]).

Remark 2. From [20], we know that the terminal DP (3) - (4) can be used to derive another information state $\Xi_t = [[X_t, A_t|M_t]]$ for each t for Case 1. The conditional range Ξ_t can take $2^{|\mathcal{A}_t| \times |\mathcal{X}|}$ feasible values whereas Π_t from Case 1 can take $|\mathcal{A}_t|^{|\mathcal{X}|}$ values. As $|\mathcal{A}_t|$ grows in size with time t, the number of feasible values of Π_t increases at a slower rate than the number of feasible values of Ξ_t. Thus, Π_t yields a more computationally tractable DP than Ξ_t. This illustrates that constructing information states using the specialized DP (12) - (13) is better than using the terminal DP (3) - (4).

Remark 3. Using Definition 4 we can identify simpler information states for systems with special properties, as shown in Cases 2 - 3. However, in many applications, merely using an information state may not sufficiently improve the tractbility optimal strategies. Thus, we extend Definition 4 to include approximate information states in Section IV.

IV. Approximate Information States

In this section, we define approximate information states and utilize them to develop an approximate DP. We begin by defining a distance between two cost distributions.

Definition 5. Let $\mathscr{X}$ be a finite subset of a metric space $(\mathscr{S}, d)$, with an uncertain variable $X \in \mathscr{X}$ and two distributions $r : \mathscr{X} \to \{-\infty\} \cup [-a^1, 0]$ and $q : \mathscr{X} \to \{-\infty\} \cup [-a^2, 0]$, $a^1, a^2 \in \mathbb{R}_{\geq 0}$. Then:

1) The *finite domains* of r and q are the sets $\mathscr{X}^r := \{\mathsf{x} \in \mathscr{X} \mid r(\mathsf{x}) \neq -\infty\}$ and $\mathscr{X}^q := \{\mathsf{x} \in \mathscr{X} | q(\mathsf{x}) \neq -\infty\}$, respectively.

2) For any $\mathsf{x} \in \mathcal{X}^r \cup \mathcal{X}^q$, the *nearest finite inputs* for r and q are given by $\psi^r(\mathsf{x}) := \arg\min_{\hat{\mathsf{x}} \in \mathscr{X}^r} d(\hat{\mathsf{x}}, \mathsf{x})$, and $\psi^q(\mathsf{x}) := \arg\min_{\hat{\mathsf{x}} \in \mathscr{X}^q} d(\hat{\mathsf{x}}, \mathsf{x})$, respectively.

3) The *distance* between the distributions r and q is

$$\mathcal{R}(r, q) := \max\big(\mathcal{H}(\mathscr{X}^r, \mathscr{X}^q), \max_{\mathsf{x} \in \mathscr{X}^r \cup \mathscr{X}^q} \big|r(\psi^r(\mathsf{x})) - q(\psi^q(\mathsf{x}))\big|\big), \quad (20)$$

where $\mathcal{H}$ is the Hausdorff metric.

Remark 4. Because any cost distribution cannot identically return $-\infty$ for all $\mathsf{x} \in \mathscr{X}$, the sets $\mathscr{X}^r$ and $\mathscr{X}^q$ are non-empty for all distributions r, q on X. Consequently, the distance $\mathcal{R}(r, q)$ always returns a finite value.

Note that $\mathcal{R}$ is the maximum of a metric on a set-space and a metric on a function-space. Thus, it can quantify the distance between two different *accrued distributions* on an uncertain variable $\mathsf{X} \in \mathscr{X}$. Specifically, let $\mathsf{Y} \in \mathscr{Y}$ and $\mathsf{Z} \in \mathscr{Z}$ take realizations $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$, respectively, such that $[[\mathsf{X}, A_t|\mathsf{y}]] \neq \emptyset$ and $[[\mathsf{X}, A_t|\mathsf{z}]] \neq \emptyset$ for some time t. Then, we denote the functional forms of the conditional distributions on X given y and given z as $r_t(\mathsf{X}|\mathsf{y})$ and $r_t(\mathsf{X}|\mathsf{z})$, respectively, and quantify the distance between them as

$$\mathcal{R}\big(r_t(\mathsf{X}|\mathsf{y}), r_t(\mathsf{X}|\mathsf{z})\big) := \max\big(\mathcal{H}\big([[\mathsf{X}|\mathsf{y}]], [[\mathsf{X}|\mathsf{z}]]\big), \max_{\mathsf{x} \in [[\mathsf{X}|\mathsf{y}]] \cup [[\mathsf{X}|\mathsf{z}]]} \big|r_t\big(\psi^{\mathsf{y}}(\mathsf{x})|\mathsf{y}\big) - r_t\big(\psi^{\mathsf{z}}(\mathsf{x})|\mathsf{z}\big)\big|\big), \quad (21)$$

where, the finite domains are $\{\mathsf{x} \in \mathscr{X} \mid r_t(\mathsf{x}|\mathsf{y}) \neq -\infty\} = [[\mathsf{X}|\mathsf{y}]]$ and $\{\mathsf{x} \in \mathscr{X} \mid r_t(\mathsf{x}|\mathsf{z}) \neq -\infty\} = [[\mathsf{X}|\mathsf{z}]]$; and for any

$\mathsf{x} \in [[\mathsf{X}|\mathsf{y}]] \cup [[\mathsf{X}|\mathsf{z}]]$, the nearest finite inputs are $\psi^{\mathsf{y}}(\mathsf{x}) := \arg\min_{\hat{\mathsf{x}} \in [[\mathsf{X}|\mathsf{y}]]} d(\hat{\mathsf{x}}, \mathsf{x})$ and $\psi^{\mathsf{z}}(\mathsf{x}) := \arg\min_{\hat{\mathsf{x}} \in [[\mathsf{X}|\mathsf{z}]]} d(\hat{\mathsf{x}}, \mathsf{x})$. Next, using $\mathcal{R}$ to quantify the approximation gap, we define approximate information states for Problem 1.

Definition 6. An *approximate information state* at any $t = 0, \ldots, T$ is an uncertain variable $\hat{\Pi}_t = \hat{\sigma}_t(M_t)$ taking values in a finite subset $\hat{\mathcal{P}}_t$ of some metric space, where $\hat{\sigma}_t : \mathcal{M}_t \to \hat{\mathcal{P}}_t$. Furthermore, for all t, there exists a parameter $\epsilon_t \in \mathbb{R}_{\geq 0}$ such that for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}$, it satisfies:

$$\mathcal{R}\big(r_t(X_t, \hat{\Pi}_{t+1} \mid m_t, u_t), r_t(X_t, \hat{\Pi}_{t+1} \mid \hat{\sigma}_t(m_t), u_t)\big) \leq \epsilon_t, \quad t = 0, \ldots, T-1, \tag{22}$$

$$\mathcal{R}\big(r_T(X_T \mid m_T), r_T(X_T \mid \hat{\sigma}_T(m_T))\big) \leq \epsilon_T. \tag{23}$$

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

$$\hat{Q}_t(\hat{\pi}_t, u_t) := \max_{x_t \in \mathcal{X}, \hat{\pi}_{t+1} \in \hat{\mathcal{P}}_{t+1}} \big(\hat{V}_{t+1}(\hat{\pi}_{t+1}) + c_t(x_t, u_t) + r_t(x_t, \hat{\pi}_{t+1} | \hat{\pi}_t, u_t)\big), \tag{24}$$

$$\hat{V}_t(\hat{\pi}_t) := \min_{u_t \in \mathcal{U}} \hat{Q}_t(\hat{\pi}_t, u_t), \tag{25}$$

where, at time T, $\hat{Q}_T(\hat{\pi}_T, u_T) := \max_{x_T \in \mathcal{X}} \big(c_T(x_T, u_T) + r_T(x_T | \hat{\pi}_T, u_T)\big)$ and $\hat{V}_T(\hat{\pi}_T) := \min_{u_T \in \mathcal{U}} \hat{Q}_T(\hat{\pi}_T, u_T)$. The control law at each t is $\hat{g}^*_t(\hat{\pi}_t) := \arg\min_{u_t \in \mathcal{U}} \hat{Q}_t(\hat{\pi}_t, u_t)$ and the approximate control strategy is $\hat{\boldsymbol{g}}^* := (\hat{g}^*_0, \ldots, \hat{g}^*_T)$. Next, we bound the performance loss from implementing the approximate control strategy $\hat{\boldsymbol{g}}^*$ in Problem 1. We begin with a preliminary result which will be required subsequently.

Lemma 2. *Let $\mathscr{X}$ be a finite subset of a metric space $(\mathscr{S}, d)$ and consider two cost distributions $r : \mathscr{X} \to \{-\infty\} \cup [-a^1, 0]$ and $q : \mathscr{X} \to \{-\infty\} \cup [-a^2, 0]$, where $a^1, a^2 \in \mathbb{R}_{\geq 0}$. Then, for a Lipschitz function $f : \mathscr{X} \to \mathbb{R}$:*

$$\big| \max_{\mathsf{x} \in \mathscr{X}} \big(f(\mathsf{x}) + r(\mathsf{x})\big) - \max_{\mathsf{x} \in \mathscr{X}} \big(f(\mathsf{x}) + q(\mathsf{x})\big) \big| \leq (L_f + 1) \cdot \mathcal{R}\big(r, q\big). \tag{26}$$

Proof. See Appendix D of our online preprint [23]. □

Next, we bound the maximum error when approximating the value functions in the optimal DP (3) - (4) with the value functions in the approximate DP (24) - (25).

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

$$|Q_t(m_t, u_t) - \hat{Q}_t(\hat{\sigma}_t(m_t), u_t)| \leq \alpha_t, \tag{27}$$

$$|V_t(m_t) - \hat{V}_t(\hat{\sigma}_t(m_t))| \leq \alpha_t, \tag{28}$$

where $\alpha_t = \alpha_{t+1} + (2L_t + 1) \cdot \epsilon_t$, where $L_t := \max\{L_{\hat{V}_{t+1}}, L_{c_t}\}$, for all $t = 0, \ldots, T-1$ and $\alpha_T = (L_{c_T} + 1) \cdot \epsilon_T$.

Proof. See Appendix E of our online preprint [23]. □

Next, we bound the maximum difference in the performance of an approximate control strategy $\hat{\boldsymbol{g}}^* := (\hat{g}^*_0, \ldots, \hat{g}^*_T)$ and optimal strategy $\boldsymbol{g}^*$. Recall that $\hat{g}^*_t(\hat{\pi}_t) = \arg\min_{u_t \in \mathcal{U}} \hat{Q}_t(\hat{\pi}_t, u_t)$ for all $t = 0, \ldots, T$. Then, the equivalent strategy $\boldsymbol{g} = (g_0, \ldots, g_T)$, which utilizes the memory but yield the same actions and performance as $\hat{\boldsymbol{g}}^*$, is constructed as $g_t(m_t) := \hat{g}^*_t(\hat{\sigma}_t(m_t))$ for all t. To compute the performance of $\boldsymbol{g}$ (and consequently, of $\hat{\boldsymbol{g}}^*$), we define for all $t = 0, \ldots, T-1$, for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}$,

$$\Theta_t(m_t, u_t) := \max_{x_t \in \mathcal{X}, m_{t+1} \in \mathcal{M}_{t+1}} \big(\Lambda_{t+1}(m_{t+1}) + c_t(x_t, u_t) + r_t(x_t, m_{t+1} | m_t, u_t)\big), \tag{29}$$

$$\Lambda_t(m_t) := \Theta_t(m_t, g_t(m_t)), \tag{30}$$

where, at time T, $\Theta_T(m_T, u_T) := \max_{x_T \in \mathcal{X}} (c_T(x_T, u_T) + r_T(x_T | m_T, u_T))$ and $V_T(m_T) = \Theta_T(m_T, g_T(m_T))$. Recursively evaluating the value functions (29) - (30) computes the performance of $\boldsymbol{g}$ as $\Lambda_0(m_0)$, where $m_0 = y_0$. Note that the performance of $\boldsymbol{g}^*$ is simply the optimal value. Next, we bound the difference in the performances of $\boldsymbol{g}$ and $\boldsymbol{g}^*$.

Theorem 4. *Let $L_{\hat{V}_{t+1}}$ be the Lipschitz constant of $\hat{V}_{t+1}$ for all $t = 0, \ldots, T-1$. Then, for all $m_t \in \mathcal{M}_t$ and $u_t \in \mathcal{U}$,*

$$|Q_t(m_t, u_t) - \Theta_t(m_t, u_t)| \leq 2\alpha_t, \tag{31}$$

$$|V_t(m_t) - \Lambda_t(m_t)| \leq 2\alpha_t. \tag{32}$$

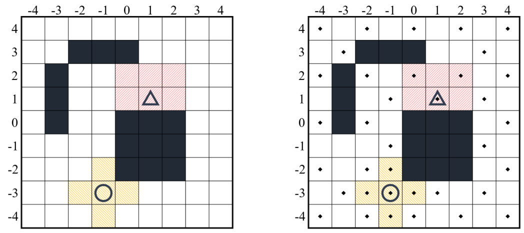
where $\alpha_t = \alpha_{t+1} + (2L_t + 1) \cdot \epsilon_t$ with $L_t := \max\{L_{\hat{V}_{t+1}}, L_{c_t}\}$ for all $t = 0, \ldots, T-1$ and $\alpha_T = (L_{c_T} + 1) \cdot \epsilon_T$.

Proof. See Appendix F of our online preprint [23]. □

V. Numerical Example

For our numerical example, we consider an agent pursuing a target across a 9×9 grid with obstacles. At each $t = 0, \ldots, T$, the agent's position is X^{ag}_t and the target's position is X^{ta}_t, each of which takes values in the set of grid cells $\mathcal{X} = \big\{(-4,-4), (-4,-3), \ldots, (3,4), (4,4)\big\} \setminus \mathcal{O}$, where $\mathcal{O} \subset \mathcal{X}$ is a known set of obstacle cells. Let $\mathcal{W} = \mathcal{N} = \{(-1,0), (1,0), (0,0), (0,1), (0,-1)\}$ and $\mathcal{D} := \{(-1,1), (1,1), (1,-1), (-1,-1)\}$. Starting at $X^{\text{ta}}_0 \in \mathcal{X}$, the target's position evolves as $X^{\text{ta}}_{t+1} = \delta(X^{\text{ta}}_t + W_t \in \mathcal{X}) \cdot (X^{\text{ta}}_t + W_t) + (1 - \delta(X^{\text{ta}}_t + W_t \in \mathcal{X})) \cdot X^{\text{ta}}_t$, where $W_t \in \mathcal{W}$ and $\delta(\cdot)$ returns 1 if the condition in the argument holds and 0 otherwise. At each t, the agent observes their own position perfectly and the target's position as $Y_t = \delta(X^{\text{ta}}_t + N_t \in \mathcal{X}) \cdot (X^{\text{ta}}_t + N_t) + (1 - \delta(X^{\text{ta}}_t + N_t \in \mathcal{X})) \cdot X^{\text{ta}}_t$, where $N_t \in \mathcal{N}$. Then, the agent selects an action $U_t \in \mathcal{U} = \mathcal{W} \cup \mathcal{D}$ and moves as $X^{\text{ag}}_{t+1} = \delta(X^{\text{ag}}_t + U_t \in \mathcal{X}) \cdot (X^{\text{ag}}_t + U_t) + (1 - \delta(X^{\text{ag}}_t + U_t \in \mathcal{X})) \cdot X^{\text{ag}}_t$. The agent incurs an interim cost $c_t(U_t) := 0.5 \cdot \delta(U_t \in \mathcal{D})$ only if it moves diagonally, and a terimal cost $d(X^{\text{ta}}_T, X^{\text{ag}}_T)$ corresponding to the final distance from the target. We illustrate this in Fig. 1(a), where: (1) the black cells are obstacles, (2) the black triangle is the initial position of the agent and the hatched region around it indicates the available actions, and (3) the black circle is the initial observation of the agent and the hatched region around it indicates the possible initial positions of the target.

This formulation is a system with action dependent costs as described in Subsection III-D. For such a system, an information state at time t is $\Pi_t = (X^{\text{ag}}_t, \Lambda_t)$, where $\Lambda_t = [[X^{\text{ta}}_t | M_t]]$. We approximate Λ_t at each t using state quantization. First, we define a static set of quantized states $\hat{\mathcal{X}}$ such



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

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

that $\max_{x_t \in \mathcal{X}} \min_{\hat{x}_t \in \hat{\mathcal{X}}} d(x_t,\ \hat{x}_t) \leq 1$ and a quantization function $\mu(x_t) := \arg\min_{\hat{x}_t \in \hat{\mathcal{X}}} d(x_t, \hat{x}_t)$ using the initial observation of the agent, as illustrated using dots in Fig. 1(b). Note that we use a finer quantization around the point of initial observation and sparser quantization elsewhere. Then, the approximate range at time t is $\hat{\Lambda}_t = \{\mu(x_t) \in \hat{\mathcal{X}} \mid x_t \in \Lambda_t\}$ and the approximate information state is $\hat{\Pi}_t = \big(X_t^{\text{ag}}, \hat{\Lambda}_t, Y_0\big)$. We include Y_0 in $\hat{\Pi}_t$ because it facilitates the update of $\hat{\Lambda}_t$ to $\hat{\Lambda}_{t+1}$. For six initial conditions, we computed the best control strategy using both the optimal DP and approximate DP for $T = 6$. In Fig. 2, we have tabulated the worst-case values (V_0 and $\hat{V}_0$) and run-times in seconds (Run.) for both DPs. We also evaluated the difference between *actual* costs incurred by the approximate strategy and the optimal strategy, respectively, by implementing both of them in 5000 simulations with randomly generated disturbances. We have marked these differences in Fig. 2 and indicated the frequency of each cost difference by the size of the disc marking it. While the approximate strategy is faster to compute than the optimal strategy for all cases, we note that it admits bounded deviations in actual costs.

Initial Conditions		Strategy IS		Strategy AIS	
x_0^{ag}	y_0	V_0	Run. (s)	$\hat{V}_0$	Run. (s)
(2, 3)	(-2, 4)	5	47.8	5	13.5
(1, 1)	(-1,-3)	6	1343.1	6	471.3
(-4, 1)	(-3,-1)	3	524.5	4	235.5
(2, -3)	(0, 4)	8.5	188.2	8.5	34.2
(-4, 4)	(-2, 2)	4	125.9	4.5	23.5
(-3, 3)	(3, -1)	9.5	330.1	8.5	35.8

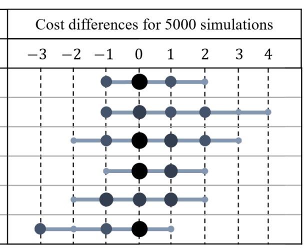


Fig. 2. Results of numerical simulations for $T = 6$.

VI. Conclusion

In this paper, we developed a general theory of information states and approximate information states to tractably compute control strategies in non-stochastic additive cost problems. We used the theoretical framework of cost distributions to present a general definition for information states that compute an optimal control strategy. We showed that specific information states proposed in previous research efforts emerge as special cases of our definition. Then, we extended this definition to approximate information states which can be used to compute approximate control strategies which admit a bounded worst-case performance loss. Finally, using a numerical simulation, we illustrated the trade-off between computational tractability and performance loss inherent in the application of approximate information states. Future work should consider the use of this theory in non-stochastic reinforcement learning problems.

References

[1] A. A. Malikopoulos, "On team decision problems with nonclassical information structures," *IEEE Transactions on Automatic Control*, 2023.

[2] 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.

[3] 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.

[4] 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.

[5] M. Giuliani, J. Lamontagne, P. Reed, and A. Castelletti, "A state-of-the-art review of optimal reservoir control for managing conflicting demands in a changing world," *Water Resources Research*, vol. 57, no. 12, p. e2021WR029927, 2021.

[6] S. Mannor, D. Simester, P. Sun, and J. N. Tsitsiklis, "Bias and variance approximation in value function estimates," *Management Science*, vol. 53, no. 2, pp. 308–322, 2007.

[7] M. R. James, J. S. Baras, and R. J. Elliott, "Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems," *IEEE transactions on automatic control*, vol. 39, no. 4, pp. 780–792, 1994.

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

[9] 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.

[10] 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.

[11] A. Rangi and M. Franceschetti, "Towards a non-stochastic information theory," in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 997–1001, IEEE, 2019.

[12] P. Gradu, J. Hallman, and E. Hazan, "Non-stochastic control with bandit feedback," *Advances in Neural Information Processing Systems*, vol. 33, pp. 10764–10774, 2020.

[13] A. Dave, N. Venkatesh, and A. A. Malikopoulos, "Approximate Information States for Worst-Case Control and Learning in Uncertain Systems," *arXiv:2301.05089 (in review)*, 2023.

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

[15] 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.

[16] 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.

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

[18] W. Wiesemann, D. Kuhn, and B. Rustem, "Robust markov decision processes," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 153–183, 2013.

[19] M. Rasouli and S. Saghafian, "Robust partially observable markov decision processes," 2018.

[20] A. Dave, N. Venkatesh, and A. A. Malikopoulos, "Approximate information states for worst-case control of uncertain systems," in *Proceedings of the 61th IEEE Conference on Decision and Control (CDC)*, pp. 4945–4950, 2022.

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

[22] V. Kolokoltsov and V. P. Maslov, *Idempotent analysis and its applications*, vol. 401. Springer Science & Business Media, 1997.

[23] A. Dave, N. Venkatesh, and A. A. Malikopoulos, "On non-stochastic approximate information states for uncertain systems with additive costs," *arXiv preprint, arXiv:2209.13787*, 2022.