Some Classes of Imperfect Information Finite State-Space Stochastic Games with Finite-Dimensional Solutions

William M. McEneaney *

October 3, 2003

Abstract

Stochastic games under imperfect information are typically computationally intractable even in the discrete-time/discrete-state case considered here. We consider a problem where one player has perfect information. A function of a conditional probability distribution is proposed as an information state. In the problem form here, the payoff is only a function of the terminal state of the system, and the initial information state is either linear or a sum of max-plus delta functions. When the initial information state belongs to these classes, its propagation is finite-dimensional. The state feedback value function is also finite-dimensional, and obtained via dynamic programming, but has a nonstandard form due to the necessity of an expanded state variable. Under a saddle point assumption, Certainty Equivalence is obtained and the proposed function is indeed an information state.

1 Introduction

A class of discrete stochastic games where one player has imperfect information is considered. Background material on such games can be found in [5], [14], [10]. The focus here will be on minimax-type values [3], [5], [9], [14]. Let us first put the problem difficulty in context by discussing increasingly complex types of problems leading to the problem type under consideration.

For discrete deterministic and stochastic games under perfect information, the value functions and feedback controls are functions of the state (annexed by time for some

*Depts. of Mathematics and Mechanical/Aerospace Engineering, University of California at San Diego, La Jolla, CA 92093-0112, USA, http://math.ucsd.edu/~wmcenean/, wmceneaney@ucsd.edu, Research partially supported by AFRL and DARPA (contract F30602-01-C-0201) and NSF (grant DMS-0307229).

problems). If the state-space is finite, these are functions over finite sets (albeit potentially **large** finite sets).

For stochastic control problems with imperfect information, one must propagate the probability distribution, conditioned on the observations, forward in time. This conditional distribution is an information state, i.e. it contains sufficient information to compute optimal controls. In the finite state-space case with n states, this distribution is a point in $\mathbf{R}^n$ at each time. The value function and optimal control function (if it exists) are functions of the conditional distribution function. In the continuum state-space case, these are functions on an infinite-dimensional space, and in the finite state-space (of size n) case, they are functions on $\mathbf{R}^n$. Under certain conditions, which are typically not verifiable, one has both certainty equivalence and a separation principle [13], [21], which asserts that the optimal control is the state feedback optimal control computed at the (assumed unique) maximum likelihood state (the argmax of the probability distribution).

In recent years, some classes of deterministic games have come under intense study due to the fact that they are equivalent to H_∞ control problems ([3], [18] and the references therein). In H_∞ control, the opposing player is "nature", and the natural imperfect information problem is where only one player (the actual controller) has imperfect, observation–based, information. For these problems, there is an information state which is a function of the state – essentially an integral or summed cost up to the current time, optimized over the opponent's control processes [3], [18], [19], [20]. In general, the value function and optimal controls are functions of this information state function. Under certain conditions, a Certainty Equivalence Principle (weaker than separation) holds [3], [18], [20]. When Certainty Equivalence holds, one need only compute the information state and the state feedback value function. The optimal control is obtained by computing the unique argmax of the sum of the information state and state feedback value functions. The optimal control is the optimal state feedback control evaluated at the argmax point in state-space. Thus, in this case, one avoids computing the control as a function on the space of information state functions.

The problem of interest here is a discrete stochastic game over a finite state-space. Since the interest is also in robust/worst-case control, a minimax upper value is considered (where the opponent is maximizing the cost). Of course, it is implicit in the above that we are interested here only in zero-sum games. Since this problem formulation generalizes both the stochastic control and deterministic game formulations, it is completely unclear what form an information state would take, although heuristically one would expect yet another level of function composition. (Note that in the stochastic control case one had a probability distribution function, and in the deterministic game case, an optimized integral/summed cost function.)

In this paper, an information state for discrete-time/discrete-space stochastic games under imperfect information (for one player) will be proposed. This information state will take the form of an optimal summed cost for any given conditional probability distribution. If the state-space is of finite size n, each conditional probability will be a point on a simplex in $\mathbf{R}^n$, and the information state will thus be a function on this $(n-1)$–dimensional simplex in $\mathbf{R}^n$. (In the continuous state-space case, it would be a function over an infinite-dimensional space.) The payoff will be a function of the true state only, and the initial information will be either piecewise linear or a max-plus sum of max-plus delta functions, and in this case, the information state itself will be finite-dimensional.

Under a saddle point condition, a Certainty Equivalence result will be obtained, and of course, this will implicitly imply that the above function is indeed an information state. Whether it is an information state in general remains an open question. Regardless, the construction of the candidate information state and the state feedback value function are quite technical, and the supporting analysis is substantial. One purpose of this paper is simply the laying of these foundations, and of course the other purpose is the development of the Certainty Equivalence result for this problem class.

The interest here is not only mathematical. There is a motivational application in the realm of military command and control (C^2) for air operations with uninhabited combat air vehicles (UCAVs). For related information, see [2], [6], [7], [15], [16], [17], [23], [24], [26]. In particular, the dynamic models in [23], [24], [26] take the discrete stochastic game form considered here. Notably, the controls for both players affect the observation process for player 1 as well as the game dynamics. Specifically, the choice of task for the player 1 UCAVs can affect the probability that they observe player 2 assets, as well as the game dynamics. At the same time, the control for the player 2 also affects both the probability it will be observed and the game dynamics. A problem model where there is only a terminal cost in terms of the players' remaining assets fits the game formulation studied here as well. Although this motivational application is specific to military C^2 , one could easily imagine other applications which one would hope to formulate similarly, given that the solutions are finite dimensional (although potentially very large).

The paper is organized as follows. In Section 2, the stochastic game problem formulation is laid out. In Section 3, the information state is defined, and its propagation forward in time is discussed. In Section 4, the state feedback value function is defined, and a dynamic programming iteration is demonstrated to compute this value. This state feedback problem is nonstandard. In Section 5, the robustness and Certainty Equivalence results are obtained. Lastly, Section 6 provides a very short discussion of the finiteness of the information state and value function computations.

2 Problem Formulation

Potential states of the system will be represented by $x \in \mathcal{X}$ where $\mathcal{X}$ is some finite set. Time will be discrete, and the state of the system at time t will be denoted by X_t. Each state x will be associated with a unit basis vector in $\mathbf{R}^{(\#\mathcal{X})}$. For instance, one could have $\mathcal{X} = \{1, 2, 3, 4, \ldots, n\}$, and state $x = 3$ would be associated with standard basis vector $(0, 0, 1, 0, \ldots, 0)$. The control for player 1, the minimizing player, will take values $u \in U$ where U is finite. The corresponding controls for player 2 (maximizing) will be $w \in W$ which is also a finite set. Controls for each player at time t will be denoted as u_t and w_t.

We will consider a finite time problem with time taking values in $\{0,1,2,\ldots,T\}$. We will denote the terminal cost as $\mathcal{E}:\ \mathcal{X}\to\mathbf{R}$; the cost of terminal state X_T is $\mathcal{E}(X_T)$. There is no running cost. Player 1 will be minimizing the cost, and player 2 will be maximizing.

We suppose that the state evolves as a controlled Markov chain (where the dynamics are time independent for simplicity of exposition). Let the probability that $X_{t+1}=j$ given $X_t=i$ with controls $u_t=u\in U$ and $w_t=w\in W$ be

$$P_{ij}(u_t,w_t)=\Pr(X_{t+1}=j|X_t=i,u_t=u,w_t=w), \tag{1}$$

and let the $n\times n$ matrix of the elements p_{ij} be denoted as $P(u,w)$ where $n\doteq\#\mathcal{X}$. We will assume that there is an observation process for player 1 (recall that player 2 will know the state perfectly) which can be controlled by both players. Let the observation process be $y_\cdot$ with $y_t\in Y$ where the probability that observation $y_t=\overline{y}$ given $X_t=i$ and controls $u_t=u,w_t=w$ is denoted as

$$R_i\doteq\Pr(y_t=\overline{y}|X_t=i,u_t=u,w_t=w). \tag{2}$$

We take Y to be a finite set for consistency, but that does not appear to be required for the results to follow.

In a deterministic game under imperfect information, the information state for player 1 is a function of the state, and it represents the minimal cost to the opposing player (maximal cost from the point of view of player 1) for the state to be x at current time t given the observations up to the current time. Alternatively, in a stochastic control problem under imperfect information, the information state is simply the probability that $X_t=x$ conditioned on the observations up to the current time t. Here however, player 2 can affect the observation process, so one must consider the cost to player 2 to produce a possibly misleading conditional probability distribution. Thus, it is natural to propose an information state for player 1 as $\mathcal{I}_t:\ Q(\mathcal{X})\to\mathbf{R}$ where $Q(\mathcal{X})$ is the space of probability distributions over state space $\mathcal{X}$; $Q(\mathcal{X})$ is the simplex in the first octant of $\mathbf{R}^n$ defined by the unit basis vectors. For simplicity of presentation, we henceforth refer to $\mathcal{I}_t$ as an information state, although the basis for this designation does not appear until Section 5. We let the initial information state be $\mathcal{I}_0(\cdot)=\phi(\cdot)$. Here, ϕ represents the initial cost to obtain and/or obfuscate initial state information. The case where this information cannot be affected by the players may be represented by a max–plus delta function. That is, ϕ takes the form

$$\phi(q)=\delta_{q_c}(q)=\begin{cases}0 & \text{if } q=q_c\\ -\infty & \text{otherwise.}\end{cases}$$

The problem will be finite-dimensional for initial information states taking the form of finite max-plus sums of max-plus delta functions

$$\phi(q)=\bigoplus_{i=1}^{m}\delta_{q_i}(q)=\max_{i\in\{1,2,\ldots,m\}}\delta_{q_i}(q).$$

A second case that will be tractable is where ϕ is linear or piecewise linear, since the piecewise linear form will be preserved under forward propagation of the information state.

3 Information State Propagation

Let time be denoted by $t \in \{0, 1, 2, \ldots, T\}$ where T is the terminal time. Let a conditional probability of the state at time t be denoted by $q_t \in Q(\mathcal{X})$. It will reduce notation and simplify the presentation if we consider first the case without observations; the observation process will be included further below. In the absence of observations, and for given controls u_t, w_t, the probability distribution propagates according to

$$q_{t+1} = P^T(u_t, w_t)q_t. \tag{3}$$

For simplicity, we will assume the existence of P^{-T} in the standard sense throughout. (This is not broken out as an assumption since it will be superseded by an assumption to appear a little further below.) Note that although this mapping is into $Q(\mathcal{X})$, it is not necessarily onto. Since w_t is not known by player 1, it will be necessary to keep track of a set of feasible conditional probabilities at time t, Q_t. Note that for t prior to the current time, u_t being player 1's control is known by player 1.

Let $w_{[0,t-1]} = \{w_0, w_1, \ldots, w_{t-1}\}$, where each $w_r \in W$, denote a sequence of controls for player 2. Then, if the controls for player 2 were independent of the true state, x, one would have

$$\begin{aligned} Q_t(u_{[0,t-1]}) = \{q \in Q(\mathcal{X}) : \exists w_{[0,t-1]} \in W^t \text{ such that } q_0 \in Q(\mathcal{X}) \text{ where } q_0 \text{ is} \\ \text{given by backward propagation (5) with } q_t = q\,\} \end{aligned} \tag{4}$$

where

$$q_{r-1} = P^{-T}(u_{r-1}, w_{r-1})q_r. \tag{5}$$

However, player 2 has full state knowledge, and consequently, it's control must be allowed to depend on the actual state. The needed notation is most easily handled by the following device. For $u \in U$ and any vector $\vec{w} \in W^n$, define the matrix $\widetilde{P}$ by

$$\widetilde{P}_{ij}(u_t, \vec{w}) \doteq P_{ij}(u_t, \vec{w}_i) \qquad \forall\, i, j \in \{1, 2, \ldots, n\}. \tag{6}$$

Now let $\vec{w}_{[0,t-1]} = \{\vec{w}_0, \vec{w}_1, \ldots, \vec{w}_{t-1}\}$, where each $\vec{w}_r \in W^n$, denote a sequence of state-dependent controls for player 2. That is, the i^{th} component of $\vec{w}_r$ is the player 2 feedback control for state $X_r = i$. One now sees that (in the absence of an observation process) the feasible set at time t should be given by

$$\begin{aligned} Q_t(u_{[0,t-1]}) = \{q \in Q(\mathcal{X}) : \exists \vec{w}_{[0,t-1]} \in [W^n]^t \text{ such that } q_0 \in Q(\mathcal{X}) \text{ where } q_0 \text{ is} \\ \text{given by backward propagation (8) with } q_t = q\,\} \end{aligned} \tag{7}$$

where

$$q_{r-1} = \widetilde{P}^{-T}(u_{r-1}, \vec{w}_{r-1})q_r. \tag{8}$$

Since the following constructions are already quite cumbersome, we make the following assumption throughout.

For all $u \in U$ and $\vec{w} \in W^n$, $\widetilde{P}^{-1}(u, \vec{w})$ exists in the standard sense $\qquad (A3.1)$

(i.e. the Moore-Penrose pseudo-inverse is not needed). The information state at time t, being the worst-case cost, is defined as

$$\mathcal{I}_t(q; u_{[0,t-1]}) \doteq \begin{cases} \sup_{q_0 \in Q_0^{q,u_{[0,t-1]}}} \sup_{\vec{w}_{[0,t-1]} \in [W^n]^t} \mathcal{I}_0(q_0) & \text{if } q \in Q_t(u_{[0,t-1]}) \\ -\infty & \text{otherwise} \end{cases} \tag{9}$$

for $t > 0$ and

$$\mathcal{I}_0(q) = \phi(q) \tag{10}$$

where

$$\begin{aligned} Q_0^{q,u_{[0,t-1]}} \doteq \{ \overline{q} \in Q(\mathcal{X}) : \exists \vec{w}_{[0,t-1]} \in [W^n]^t \text{ such that } q_t = q \text{ given} \\ q_0 = \overline{q} \text{ and propagation (8)} \}. \end{aligned} \tag{11}$$

Note that the maximum here used to compute the player 1 information state allows $\vec{w}$ to be chosen depending on u_t (upper value). For each possible distribution, q, this represents the maximal cost (minimal from player 2's perspective) for the computed conditional probability to be q given the original cost. (Recall that we have only initial and terminal costs in this problem formulation.)

So far we have ignored the possibility of an observation process. Let us now include this in the propagation. We will assume that the observations may occur at each time step, t. We will distinguish between a priori conditional distributions, denoted as q_t, and a posteriori distributions, denoted as $\widehat{q}_t$. That is, $\widehat{q}_t$ incorporates the possible new information in an observation at time t. Suppose the actual observation at time t is $y_t = \overline{y} \in Y$. Recalling the observation discussion of Section 2, and the fact that we are allowing the player 2 control to depend on the true state, we let the vector $\widetilde{R}$ have components

$$\widetilde{R}_i = \widetilde{R}_i(\overline{y}, u, \vec{w}_i) \doteq \Pr(y_t = \overline{y} | X_t = i, u, \vec{w}_i) \tag{12}$$

for each $i \leq n$ where again $\vec{w}$ indicates the possibly state-dependent choice of player 2 control. Let $D(\widetilde{R})$ be the matrix whose i^{th} diagonal element is $\widetilde{R}_i$ for each i, and whose other elements are zero. Then, given any control u and $\vec{w}$, the a posteriori distribution would be given by

$$\widehat{q}_t = \left(\tfrac{1}{\widetilde{R}^T(\overline{y}, u_t, \vec{w}) q_t} \right) D(\widetilde{R}(\overline{y}, u_t, \vec{w})) q_t. \tag{13}$$

The possible set of a posteriori distributions, $\widehat{\mathcal{Q}}_t$ is the set of all $\widehat{q}_t$ given by (13) for some $q_t \in \mathcal{Q}_t$. To reduce the already cumbersome development, we make the following assumption.

Assume that for all $u \in U$, $\vec{w} \in W^n$ and $\overline{y} \in Y$, D is full rank;
i.e. $\widetilde{R}_i(\overline{y}, u_t, \vec{w}) \neq 0$ for all i. $\quad (A3.2)$

Note that each component of $\hat{q}_t$ is given by $\hat{q}_{ti} = \alpha \widetilde{R}_i q_{ti}$ for constant $\alpha = 1/(\widetilde{R}^T(\overline{y}, u_t, \vec{w})q_t)$ independent of i. Inverting this, each component $q_{ti} = (1/\alpha)\widetilde{R}_i^{-1}\hat{q}_{ti}$. Since $\sum_i q_{ti} = 1$, one sees that one must have $1/\alpha = 1/(\sum_i \widetilde{R}_i^{-1}\hat{q}_{t\,i})$. Consequently, each component $q_{ti} = [1/(\sum_i \widetilde{R}_i^{-1}\hat{q}_{ti})]\widetilde{R}_i^{-1}\hat{q}_{ti}$.

With the addition of the observation process, the feasible set now becomes

$$Q_t(u_{[0,t-1]}, y_{[0,t-1]}) = \{q \in Q(\mathcal{X}) : \exists \vec{w}_{[0,t-1]} \in [W^n]^t \text{ such that } q_0 \in Q(\mathcal{X}) \text{ where } q_0 \text{ is given by backward propagation (15) with } q_t = q\,\} \tag{14}$$

where

$$\begin{aligned} q_{r-1} &= \widehat{G}^{-1}(y_{r-1}, u_{r-1}, \vec{w}_{r-1}, q_r) \\ &\doteq \frac{1}{\widehat{R}^T(y_{r-1},u_{r-1},\vec{w}_{r-1})\widetilde{P}^{-T}(u_{r-1},\vec{w}_{r-1})q_r} D^{-1}(\widetilde{R}(y_{r-1}, u_{r-1}, \vec{w}_{r-1}))\widetilde{P}^{-T}(u_{r-1}, \vec{w}_{r-1})q_r \end{aligned} \tag{15}$$

where $\widehat{R}_i(y_{r-1}, u_{r-1}, \vec{w}_{r-1}) \doteq 1/[\widetilde{R}_i(y_{r-1}, u_{r-1}, \vec{w}_{r-1})]$.

Also, with the addition of the observation process, the information state definition (at time t prior to the observation) now becomes

$$\mathcal{I}_t(q; u_{[0,t-1]}, y_{[0,t-1]}) \doteq \begin{cases} \max_{q_0 \in Q_0^{q,u_{[0,t-1]}}} \max_{\vec{w}_{[0,t-1]} \in [W^n]^t} \mathcal{I}_0(q_0) & \text{if } q \in Q_t(u_{[0,t-1]}, y_{[0,t-1]}) \\ -\infty & \text{otherwise} \end{cases} \tag{16}$$

for $t > 0$ and

$$\mathcal{I}_0(q) = \phi(q) \tag{17}$$

where

$$Q_0^{q,u_{[0,t-1]}} \doteq \{q_0 \in Q(\mathcal{X}) : \exists \vec{w}_{[0,t-1]} \in [W^n]^t \text{ such that } q_t = q \text{ given initial } q_0 \text{ and backward propagation (15)}\}. \tag{18}$$

The difference between (16) and (9) is that the propagation is now given by (15) rather than by (8). Note that maxima (as opposed to suprema) are appropriate in (16) since W is finite and since $\mathcal{I}_0$ is either a finite maximum of max-plus delta functions or a piecewise linear function. Also note that we will often suppress the dependence of Q_t on $u_{[0,t-1]}, y_{[0,t-1]}$. The proofs of the next two lemmas are obvious.

Lemma 3.1

$$Q_{t+1} = \{q \in Q(\mathcal{X}) : \exists q_t \in Q_t,\ \vec{w} \in W^n \ \textit{such that}\ q = G(y_t, u_t, \vec{w})[q_t]\} \tag{19}$$

where

$$G(y, u, \vec{w})[q] \doteq \widehat{G}(y, u, \vec{w}, q) = \widetilde{P}^T(u, \vec{w}) \tfrac{1}{\widetilde{R}^T(y,u,\vec{w})q} D(\widetilde{R}(y, u, \vec{w}))q. \tag{20}$$

Note that we are introducing the notation $G(y, u, \vec{w})[q]$ to indicate that $G(y, u, \vec{w})$ is a mapping from $Q(\mathcal{X})$ into $Q(\mathcal{X})$ for each triple $y, u, \vec{w}$; this notation will be useful below.

Lemma 3.2 *$Q_t \neq \emptyset$ for all $t \in \{0, 1, \ldots, T\}$.*

Lemma 3.3

$$\mathcal{I}_{t+1}(q) = \begin{cases} \max_{\vec{w} \in W_t^q} \max_{\widehat{q} \in S_t^{\vec{w},q}} \mathcal{I}_t(\widehat{q}) & \textit{if } W_t^q \neq \emptyset \\ -\infty & \textit{otherwise} \end{cases} \tag{21}$$

where

$$\begin{aligned} S_t^{\vec{w},q} &= S_t^{\vec{w},q}(u_{[0,t]}, y_{[0,t]}) = \{\widehat{q} \in Q_t : q = G(y_t, u_t, \vec{w})[\widehat{q}]\} \\ W_t^q &= \{\vec{w} \in W^n : S_t^{\vec{w},q} \neq \emptyset\}. \end{aligned} \tag{22}$$

Proof. We prove inequalities in both directions. We first show $\mathcal{I}_{t+1}(q)$ is greater than or equal to the right hand side. Suppose $q \in Q_{t+1}$. Let

$$\widetilde{W}_{t-1}^q \doteq \left\{ \vec{w}_{[0,t-1]} : \exists q_0 \in Q(\mathcal{X}) \text{ such that } q = \left[\prod_{r=0}^{t-1} G(y_r, u_r, \vec{w}_r) \right] [q_0] \right\} \tag{23}$$

where the $\prod$ notation indicates operator composition.

Let $\vec{w} \in W_t^q$ and $q_t \in S_t^{\vec{w},q}$ where we assume $W_t^q \neq \emptyset$ since otherwise there is nothing to prove in this direction.

By the definition of $S_t^{\vec{w},q}$ there exists $\tilde{\vec{w}}_{[0,t-1]} \in \widetilde{W}_{t-1}^{q_t}$ and a corresponding $\widetilde{q}_0 \in Q(\mathcal{X})$ such that

$$\mathcal{I}_t(q_t) = \mathcal{I}_0(\widetilde{q}_0) \tag{24}$$

where

$$q_t = \left[\prod_{r=0}^{t-1} G(y_r, u_r, \tilde{\vec{w}}_r) \right] [\widetilde{q}_0]. \tag{25}$$

Define

$$\tilde{\vec{w}}_r = \begin{cases} \tilde{\vec{w}}_r & \text{if } r \leq t-1 \\ \vec{w} & \text{if } r = t. \end{cases}$$

Then $\vec{\widetilde{w}} \in \widetilde{W}_t^q$ (defined similarly to $\widetilde{W}_{t-1}^q$ of course). By the definition of $\mathcal{I}_{t+1}(q)$,

$$\mathcal{I}_{t+1}(q) \geq \mathcal{I}_0(\widetilde{q}_0). \tag{26}$$

Combining (24) and (26) yields $\mathcal{I}_{t+1}(q) \geq \mathcal{I}_t(q_t)$. Since this is true for all $\vec{w} \in W_t^q$ and $q_t \in S_t^{\vec{w},q}$, one has

$$\mathcal{I}_{t+1}(q) \geq \max_{\vec{w} \in W_t^q} \max_{\widetilde{q} \in S_t^{\vec{w},q}} \mathcal{I}_t(\widetilde{q}) \tag{27}$$

which is the inequality in one direction.

Now we turn to the other direction. Suppose $\mathcal{I}_{t+1}(q) \neq -\infty$; otherwise there is nothing to prove. By the finiteness of W, there exists an optimal $\vec{\widetilde{w}}_{[0,t]}$ and corresponding $q_0 \in Q(\mathcal{X})$ given by

$$q = \left[\prod_{r=0}^{t} G(y_r, u_r, \vec{\widetilde{w}}_r) \right] [q_0]$$

such that

$$\mathcal{I}_0(q_0) = \mathcal{I}_{t+1}(q). \tag{28}$$

Then, $\vec{\widetilde{w}}_t \in W_t^q$, $S_t^{\vec{w},q} \neq \emptyset$ and

$$q_t \doteq \left[\prod_{r=0}^{t-1} G(y_r, u_r, \vec{\widetilde{w}}_r) \right] [q_0] \in S_t^{\vec{\widetilde{w}}_t,q}.$$

This implies that $q_t \in Q_t$. By the definition of $\mathcal{I}_t$,

$$\mathcal{I}_t(q_t) \geq \mathcal{I}_0(q_0). \tag{29}$$

Combining (28) and (29) yields

$$\mathcal{I}_{t+1}(q) \leq \mathcal{I}_t(q_t) \leq \max_{\vec{w} \in W_t^q} \max_{\widetilde{q} \in S_t^{\vec{w},q}} \mathcal{I}_t(\widetilde{q}). \tag{30}$$

By (27) and (30), one has the result. □

A potential problem is that the normalization in (13) and (15) induces nonlinearities in the propagation. This is specifically important in the case where I_0 is linear or piecewise linear. Consequently, we will work with the unnormalized distribution. The a priori and a posteriori unnormalized distributions at time t will be denoted as $\widetilde{q}_t$ and $\widehat{\widetilde{q}}_t$, respectively. At any time t, one can renormalize by dividing by $\sum_i [\widetilde{q}_t]_i$ for the a priori distribution, and similarly for the a posteriori. The feasible sets of a priori and a posteriori unnormalized distributions will be denoted by $\widetilde{\mathcal{Q}}_t$ and $\widehat{\widetilde{\mathcal{Q}}}_t$.

If the control processes, $u_{\cdot}$ and $\vec{w}_{\cdot}$, and the observation process, $y_{\cdot}$, are given, then the unnormalized distribution would propagate as

$$\widetilde{q}_{t+1} = \widetilde{P}^T(u_t, \vec{w}_t)\widehat{\widetilde{q}}_t, \qquad \widehat{\widetilde{q}}_t = D(\widetilde{R}(\overline{y}_t, u_t, \vec{w}_t))\widetilde{q}_t \tag{31}$$

for given initial $\widetilde{q}_0 = q_0$. Based opn the above results for the normalized case, the information state as a function of the unnormalized distribution, denoted by $\widetilde{\mathcal{I}}_t$, should propagate by

$$\begin{aligned}\widetilde{\mathcal{I}}_{t+1}(\widetilde{q}) = \max\Big\{&\widetilde{\mathcal{I}}_t[D^{-1}(\widetilde{R}(y_t,u_t,\vec{w}))\widetilde{P}^{-T}(u_t,\vec{w})\widetilde{q}] : \\ &\exists\vec{w}\in W^n \text{ such that } D^{-1}(\widetilde{R}(y_t,u_t,\vec{w}))\widetilde{P}^{-T}(u_t,\vec{w})\widetilde{q}\in\widetilde{\mathcal{Q}}_t\Big\}\end{aligned} \tag{32}$$

where

$$\begin{aligned}\widetilde{\mathcal{Q}}_{t+1} = \Big\{&q\in\mathbf{R}^{\#\mathcal{X}} : \exists\, q_t\in\widetilde{\mathcal{Q}}_t, \vec{w}\in W^n \text{ such that} \\ &q = \widetilde{G}(y_t,u_t,\vec{w}_t)[q_t]\Big\}\end{aligned} \tag{33}$$

where $\widetilde{G}(y,u,\vec{w})[q] \doteq \widetilde{P}^T(u,\vec{w})D(\widetilde{R}(y,u,\vec{w}))[q_t]$ with initial conditions $\widetilde{\mathcal{I}}_0(q) = \phi(q)$ and $\widetilde{\mathcal{Q}}_0 = Q(\mathcal{X})$.

To be fully rigorous, one should first define the unnormalized $\widetilde{\mathcal{I}}_t$ and $\widetilde{\mathcal{Q}}_t$ directly, and then obtain the propagation formulae (32)–(33) in a manner analogous to Lemmas 3.1 through 3.3. Since the variation from the normalized case is trivial, we do not include this.

In the case that the initial cost, ϕ is piecewise linear, we see that, even when including the observation process, the *unnormalized* information state remains finite-dimensional. More specifically,

Theorem 3.4 *If ϕ is linear (where we freely use the term linear to mean affine, i.e. linear plus a constant), then for any time, $t \geq 0$, $\widetilde{\mathcal{I}}_t$ is the maximum of a finite set of linear functions with convex domains with piecewise linear boundaries defined by at most n extremal points. The number of such linear functions required is at most $(\#W^n)^t$. If ϕ is piecewise linear, then $\widetilde{\mathcal{I}}_t$ is the maximum of a finite set of piecewise linear functions with piecewise linear boundaries, and again the number of such functions required is at most $(\#W^n)^t$.*

Proof. Recall that we are assuming, for simplicity, that P^{-T} and D^{-1} exist in the standard sense. Suppose ϕ is linear (more precisely, affine), say $\phi(q) = \gamma^T q + \beta$ for some $\gamma\in\mathbf{R}^n$ and $\beta\in\mathbf{R}$. Note that $\widetilde{Q}_0 = Q(\mathcal{X})$. Suppose $q_0\in\widetilde{Q}_0$, $u_0\in U$, $y_0\in Y$, and let $\vec{w}\in W^n$. Define

$$\begin{aligned}q_1 = q_1(\vec{w}) &\doteq \widetilde{G}(u_0,\vec{w},y_0)[q_0] \\ &= \widetilde{P}^T(u_0,\vec{w})D(\widetilde{R}(y_0,u_o,\vec{w}))q_0.\end{aligned}$$

By (33), $q_1\in\widetilde{Q}_1$. Let $\widetilde{Q}_1^{\vec{w}} = \widetilde{G}(y_0,u_0,\vec{w})[\widetilde{Q}_0]$ which is the image of simplex $\widetilde{Q}_0$ under linear operator $\widetilde{G}(y_0,u_o,\vec{w})$, and so $\widetilde{Q}_1^{\vec{w}}$ is a convex subset of a hyperplane in $\mathbf{R}^n$, and since the boundary of $\widetilde{Q}_0$ has n extremal points, $\widetilde{Q}_1^{\vec{w}}$ has at most n extremal points (the images of the extremal points of $\widetilde{Q}_0$ modulo degeneracy). Then, by (33), $\widetilde{Q}_1 = \cup_{\vec{w}\in W^n}\widetilde{Q}_1^{\vec{w}}$

which is a set of at most $\#W^n$ hyperplane subsets. Proceeding inductively, one obtains $\widetilde{Q}_t$ as the union of at most $(\#W^n)^t$ hyperplane subsets.

Now we turn to the $\widetilde{\mathcal{I}}_t$ themselves. Let $q_1 \in \widetilde{Q}_1$. Also, denote W^n as $W^n = \{\vec{w}^k\}_{k=1}^{\#W^n}$. For each $k \leq \#W^n$, define

$$\begin{aligned}\widetilde{\mathcal{I}}_1^{\vec{w}^k}(q) &= \widetilde{\mathcal{I}}_0(\widetilde{G}^{-1}(u_0, y_0, \vec{w}^k)[q]) \\ &= \gamma^T D^{-1}(\widetilde{R}(y_0, u_0, \vec{w}^k)\widetilde{P}^{-T}(u_0, \vec{w}^k)q + \beta\end{aligned}$$

for all $q \in \widetilde{Q}_1^{\vec{w}^k}$, which is a linear functional over domain $\widetilde{Q}_1^{\vec{w}^k}$. For each $q \in \widetilde{Q}_1$, let $\widehat{W}(q) \doteq \{\vec{w} \in W^n : q = \widetilde{G}(y_0, u_0, \vec{w})[q_0] \text{ for some } q_0 \in \widetilde{Q}_0\}$. Then, by (32), $\widetilde{\mathcal{I}}_1(q) = \max_{\vec{w}^k \in \widehat{W}(q)} \widetilde{\mathcal{I}}_1^{\vec{w}^k}(q)$ which proves the result for $t = 1$. Proceeding inductively, one obtains the result for all t.

The proof for the case where ϕ is piecewise linear is similar, and so we do not include it. □

Although we will not consider the actual computational costs here, the sequential propagation of such information states is tractable in real-time for reasonably small problems. We particularly want to distinguish this propagation from the case where the information state is infinite dimensional.

Alternatively, in the case where ϕ is a max-plus delta function or finite max-plus sum of max-plus delta functions, this is even more tractable. Note that ϕ_k is a max-plus delta function over $Q(\mathcal{X})$ if there exists $q_k \in Q(\mathcal{X})$ such that

$$\phi_k(q) = \begin{cases} 0 & \text{if } q = q_k \\ -\infty & \text{if } q \neq q_k. \end{cases}$$

Also, ϕ is a (finite) max-plus sum of max-plus delta functions if there exist $\{q_k\}_{k=1}^K$ such that

$$\phi(q) = \bigoplus_{k=1}^{K} \phi_k(q) = \max_k \phi_k(q).$$

Theorem 3.5 *If ϕ is a max-plus sum of K max-plus delta functions (where $K \geq 1$), then $I_t(q) : Q(\mathcal{X}) \to \{-\infty, 0\}$ is a max-plus sum of at most $K(\#W^n)^t$ max-plus delta functions.*

The proof is quite trivial by (32), and we do not include it. It is worth noting that in this (max-plus delta functions) case, one does not need to use the unnormalized distributions, which is in contrast to the linear and piecewise linear cases.

4 Value Function

We now turn to the state feedback value function. The full state of the system is now described by the true state taking values $x \in \mathcal{X}$ and the player 1 information state taking values $q \in Q(\mathcal{X})$. We denote the terminal payoff for the game as $\mathcal{E} : \mathcal{X} \to \mathbf{R}$ (where of course this does not depend on the internal information state of player 1). Thus the state feedback value function at the terminal time is

$$V_T(x, q) = \mathcal{E}(x). \tag{34}$$

One issue that arises is the information available to player 2. One option would be to assume that it knows only the actual true state, x. However, with full knowledge of the state and observations, player 2 could also construct the conditional probability, q. This second problem model is more conservative in terms of construction of the player 1 control, and this model will be used here.

The state of the state feedback game at time t is (X_t, q_t) where X_t propagates as a Markov chain with probabilities given by (1) and q_t propagates by (3). Player 1 will have access only to the probability distributions up to the current time, while player 2 will have access to the true state as well.

We define the strategies for player 1 as follows. Throughout, we will continue to use the convention that interval subscripts indicate sequences; for instance, $u_{[\bar{t},t_1]} = \{u_r\}_{r=\bar{t}}^{t_1}$. Since player 1 has access only to probability distributions, the set of strategies for player 1 over time interval $[\bar{t}, T-1]$ is

$$\overline{\Lambda}_{[\bar{t},T-1]} = \Big\{ \overline{\lambda}_{[\bar{t},T-1]} : Q^{T-\bar{t}} \to U^{T-\bar{t}}, \text{ nonanticipative in } q.\, \Big\}. \tag{35}$$

Note that $\overline{\lambda}_{[\bar{t},T-1]}$ is nonanticipative in $q.$ if given any $t \in \{\bar{t}, \bar{t}+1, \ldots, T-1\}$ and any $q_{[\bar{t},T-1]} = \tilde{q}_{[\bar{t},T-1]} \in Q^{T-\bar{t}}$ such that $q_r = \tilde{q}_r$ for all $r \leq t$, then $\overline{\lambda}_t[q_{[\bar{t},T-1]}] = \overline{\lambda}_t[\tilde{q}_{[\bar{t},T-1]}]$. Further, note that $\overline{\lambda}_t$ is independent of x. More specifically, if the true state $X_t \neq \widehat{X}_t$, but $q_r = \tilde{q}_r$ for all $r \leq t$, then one still has $\overline{\lambda}_t[q_{[\bar{t},T-1]}] = \overline{\lambda}_t[\tilde{q}_{[\bar{t},T-1]}]$. For notational simplicity, let $\overline{\lambda}_t \equiv \overline{\lambda}_{[t,t]}$. For reasons of robustness, we will be interested in an upper value (giving advantage to player 2). Consequently, the strategy set for player 2 is naturally

$$\overline{\Theta}_{[\bar{t},T-1]} = \Big\{ \overline{\theta}_{[\bar{t},T-1]} : \mathcal{X}^{T-\bar{t}-1} \times Q^{T-\bar{t}} \to W^{n(T-\bar{t})}, \text{ nonanticipative in } X., q.\, \Big\}.$$

Note that the dependence of $\overline{\theta}_t$ on the current state, X_t is implicit in the fact that $\vec{w}$ is a vector of length n where component i represents the control w to be played if the current state is $X_t = i$. The strategy set $\overline{\Theta}$ corresponds to the closed-loop perfect state (CLPS) information pattern [3], [5], while $\overline{\lambda}$ is similar to CLPS but with the x-portion of the state unobserved.

We note that in this state feedback game definition, player 1 assumes that $q_{\bar{t}}$ is an accurate representation of the true distribution of its lack of information of the true state $X_{\bar{t}}$. Assuming no modeling errors (as always here), q_t then remains an accurate representation for all t for each possible sequence of player 2 moves.

Since player 1 knows only the $q.$ process, the best that could be achieved from player 1's perspective would be

$$V_{\bar{t}}^1(q) = \inf_{\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}} \; \sup_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\} \tag{36}$$

where $\mathbf{E}_q$ represents expectation over X with $P(X = i) = q_i$ for all $i \in \mathcal{X}$, and the dynamics are given by (1), (3), (6) with strategies $\overline{\lambda}$ and $\overline{\theta}$. Since the above formulation is slightly nonstandard, some equivalent formulations follow.

Lemma 4.1 *The optimal player 1 value, $V_{\bar{t}}^1$, satisfies*

$$V_{\bar{t}}^1(q) = \inf_{\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}} \ \max_{\vec{w}_{[\bar{t},T-1]} \in W^{n(T-\bar{t})}} \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\}. \tag{37}$$

Proof. We prove that for any $\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}$,

$$\sup_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\} = \max_{\vec{w}_{[\bar{t},T-1]} \in W^{n(T-\bar{t})}} \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\} \tag{38}$$

Let the left-side of (38) be denoted by $A(q, \overline{\lambda}_{[\bar{t},T-1]})$, and the right-side by $B(q, \overline{\lambda}_{[\bar{t},T-1]})$. Fix any $q \in Q(\mathcal{X})$ and $\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}$. Let $\vec{w}^*_{[\bar{t},T-1]} \in W^{n(T-\bar{t})}$ achieve the maximum on the right in (38). Define $\overline{\theta}^*_t[X_\cdot, q_\cdot] \doteq \overline{w}^*_t$ for all $t \in [\bar{t}, T-1]$. Then the corresponding processes which we denote by $X^*_\cdot$ and $q^*_\cdot$ are identical for both control $\vec{w}^*_{[\bar{t},T-1]}$ and strategy $\overline{\theta}^*_{[\bar{t},T-1]}$, and $B(q, \overline{\lambda}_{[\bar{t},T-1]}) = \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X^*_T) \,|\, X^*_{\bar{t}} = X] \Big\} \leq A(q, \overline{\lambda}_{[\bar{t},T-1]})$.

Now the reverse inequality is proved. Let $\overline{\theta}^\varepsilon_{[\bar{t},T-1]}$ be ε–optimal for the left-side of (38). This yields a $q^\varepsilon_\cdot$ process where $q^\varepsilon_{t+1} = \widetilde{P}^T(\overline{\lambda}_t[q^\varepsilon_{[\bar{t},t]}], \overline{\theta}^\varepsilon_t[X^\varepsilon_{[\bar{t},t-1]}, q^\varepsilon_{[\bar{t},t]}])q^\varepsilon_t$. Let $\vec{w}^\varepsilon_t = \overline{\theta}^\varepsilon_t[X^\varepsilon_{[\bar{t},t-1]}, q^\varepsilon_{[\bar{t},t]}]$ for all $t \in [\bar{t}, T-1]$. Then, the corresponding processes which we denote by $X^\varepsilon_\cdot$ and $q^\varepsilon_\cdot$ are identical for both control $\vec{w}^\varepsilon_{[\bar{t},T-1]}$ and strategy $\overline{\theta}^\varepsilon_{[\bar{t},T-1]}$. Consequently, $A(q, \overline{\lambda}_{[\bar{t},T-1]}) - \varepsilon \leq \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X^\varepsilon_T) \,|\, X^\varepsilon_{\bar{t}} = X] \Big\} \leq B(q, \overline{\lambda}_{[\bar{t},T-1]})$. Since this is true for all ε, the proof is complete. □

Remark 4.2 Note that the $\overline{\theta}^*_\cdot$ constructed in the first half of the proof of Lemma 4.1 is optimal, and consequently, one has

$$\sup_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\} = \max_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\}$$

for any $q \in Q(\mathcal{X})$ and $\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}$, which is to say that one can replace the supremum with a maximum since the least upper bound is achieved by $\overline{\theta}^*$.

Lemma 4.3 *The optimal player 1 value, $V_{\bar{t}}^1$, satisfies*

$$V_{\bar{t}}^1(q) = \inf_{\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}} \mathbf{E}_q \Big\{ \max_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\} \tag{39}$$

$$= \inf_{\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}} \mathbf{E}_q \Big\{ \max_{\vec{w}_{[\bar{t},T-1]} \in W^{n(T-\bar{t})}} \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\}. \tag{40}$$

Proof. We prove only the first assertion. The second follows from the first by a proof similar to that of Lemma 4.1. Fix any $q \in Q(\mathcal{X})$ and $\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}$. We prove that

$$\max_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\} = \mathbf{E}_q \Big\{ \max_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\}. \qquad (41)$$

Let the left-side of (41) be denoted by $A(q, \overline{\lambda}_{[\bar{t},T-1]})$, and the right-side by $B(q, \overline{\lambda}_{[\bar{t},T-1]})$. Let

$$\overline{\theta}^*_{\cdot} \in \operatorname*{argmax}_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}^{n(T-\bar{t})}} \Big[\mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\} \Big],$$

and let $X^*_{\cdot}$ be the corresponding state process. Then

$$A(q, \overline{\lambda}_{[\bar{t},T-1]}) = \mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X^*_T) \,|\, X^*_{\bar{t}} = X] \Big\}$$

where $X^*_{\cdot}$ is generated by strategies $\overline{\theta}^*$ and $\overline{\lambda}$, and since obviously $\overline{\theta}^* \in \overline{\Theta}$,

$$\leq \mathbf{E}_q \Big\{ \max_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\} = B(q, \overline{\lambda}_{[\bar{t},T-1]}).$$

Now we prove the reverse inequality. Recall that $\overline{\theta}_{\cdot}$ is dependent on the state process $X_{\cdot}$ (nonanticipatively). Given any $x \in \mathcal{X}$, let $\overline{\theta}^{x,*}$ be optimal, that is

$$\mathbf{E}[\mathcal{E}(X^{x,*}_T) \,|\, X^{x,*}_{\bar{t}} = x] = \max_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = x] \qquad (42)$$

where $X^{x,*}$ is the state process corresponding to $\overline{\theta}^{x,*}$ with initial condition $X^{x,*}_{\bar{t}} = x$. Now define $\overline{\theta}^* \in \overline{\Theta}$ as follows. For each pair of sequences, $(x_{[\bar{t},T-1]}, q_{[\bar{t},T-1]}) \in \mathcal{X}^{T-\bar{t}} \times Q(\mathcal{X})^{T-\bar{t}}$ such that $x_{\bar{t}} = x$, let $\overline{\theta}^*_{[\bar{t},T-1]} = \overline{\theta}^{x,*}_{[\bar{t},T-1]}$. Note that this defines $\overline{\theta}^*$ uniquely for each process path. Given $\overline{\theta}^*$ and $\overline{\lambda}$, initial $q_{\bar{t}} = q$ and any initial $X_{\bar{t}}$, let $X^*_{\cdot}$ and $q^*_{\cdot}$ be the corresponding processes. By (42) and the definition of $\overline{\theta}^*$,

$$\mathbf{E}_q \Big\{ \mathbf{E}[\mathcal{E}(X^*_T) \,|\, X^*_{\bar{t}} = X] \Big\} = \mathbf{E}_q \Big\{ \max_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}[\mathcal{E}(X^*_T) \,|\, X^*_{\bar{t}} = X] \Big\} = B(q, \overline{\lambda}_{[\bar{t},T-1]}).$$

As above, since $\overline{\theta}^* \in \overline{\Theta}$, this immediately yields

$$A(q, \overline{\lambda}_{[\bar{t},T-1]}) \geq B(q, \overline{\lambda}_{[\bar{t},T-1]})$$

which completes the proof. □

Define

$$M_{\bar{t}}(x, q, \overline{\lambda}_{[\bar{t},T-1]}) \doteq \max_{\overline{\theta}_{[\bar{t},T-1]} \in \overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E}[V_T(X_T, q_T) \,|\, X_{\bar{t}} = x] \qquad (43)$$

so that

$$V^1_{\bar{t}}(q) = \inf_{\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}} \mathbf{E}_q\{M_{\bar{t}}(X, q, \overline{\lambda}_{[\bar{t},T-1]})\}. \tag{44}$$

Noting the fact that U is finite, one sees that there exists an optimal $\overline{\lambda}^0_{[\bar{t},T-1]}$ (see Remark 4.4 just below) given by

$$\overline{\lambda}^0_{[\bar{t},T-1]} = \operatorname*{argmin}_{\overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]}} \mathbf{E}_q\{M_{\bar{t}}(X, q, \overline{\lambda}_{[\bar{t},T-1]})\}. \tag{45}$$

Remark 4.4 Rather than include a full, technical proof of the existence of an optimal $\overline{\lambda}^0_{[\bar{t},T-1]}$, we simply indicate a proof in the one time-step case. Let $F : Q(\mathcal{X}) \times U \to \mathbf{R}$ be any function which is measurable in $Q(\mathcal{X})$ for each $u \in U$. In particular, F will represent $\mathbf{E}_q\{M_{T-1}(X, q, u)\}$ where for expediency, we abuse notation by letting the third argument in M_{T-1} here be an element of U rather than a function with range in U. Let the number of elements of U be N_u. Define

$$A_1 = \{q \in Q(\mathcal{X}) : \, F(q, u_1) \le \min_{i>1} F(q, u_i)\}.$$

Then, for each $j \in \{2, 3, \ldots, N_u - 1\}$, define

$$A_j = \Big\{q \in Q(\mathcal{X}) \setminus \Big[\bigcup_{k<j} A_k\Big] : \, F(q, u_j) \le \min_{i>j} F(q, u_i)\Big\},$$

and finally, let

$$A_{N_u} = Q(\mathcal{X}) \setminus [\bigcup_{k<N_u} A_k].$$

Note that $Q(\mathcal{X}) = \bigcup_{j=1}^{N_u} A_j$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$. Then, $\overline{\lambda}^0_{[T-1,T-1]}[q] \doteq u_j$ for $q \in A_j$ is optimal in (44) with $\bar{t} = T - 1$.

We consider only the upper value of the game given by

$$V_{\bar{t}}(x, q) = M_{\bar{t}}(x, q, \overline{\lambda}^0_{[\bar{t},T-1]}). \tag{46}$$

Note that by (44), (45) and (46)

$$\begin{aligned} V^1_{\bar{t}}(q) &= \mathbf{E}_q[M_{\bar{t}}(X, q, \overline{\lambda}^0_{[\bar{t},T-1]})] \\ &= \mathbf{E}_q[V_{\bar{t}}(X, q)]. \end{aligned} \tag{47}$$

Recall from Remark 4.2, that there also exists an optimal $\overline{\theta}^0_{[\bar{t},T-1]}$ (dependent on $\overline{\lambda}_{[\bar{t},T-1]}$ of course). Let the vector of length n, $\vec{\mathcal{E}}$, be defined by $\vec{\mathcal{E}}_i = \mathcal{E}(i)$. Then one sees that

$$V_{\bar{t}}(i, q) = \left(\Big[\prod_{t=\bar{t}}^{T-1} \widetilde{P}\Big(\overline{\lambda}^0_t[q_{[\bar{t},T-1]}], \overline{\theta}^0_t[X_{[\bar{t},T-1]}, q_{[\bar{t},T-1]}]\Big) \Big] \vec{\mathcal{E}} \right)_i \tag{48}$$

and

$$V_{\bar{t}}^{1}(q) = q^T \left(\Big[\prod_{t=\bar{t}}^{T-1} \widetilde{P}\Big(\overline{\lambda}_t^0[q_{[\bar{t},T-1]}], \overline{\theta}_t^0[X_{[\bar{t},T-1]}, q_{[\bar{t},T-1]}] \Big) \Big] \vec{\mathcal{E}} \right). \tag{49}$$

Now that the state feedback value has been defined, one needs to show how it can be obtained by backward dynamic programming propagation. Let $V_t^i(x,q)$ be the function obtained by the following backward (dynamic programming) iteration. (Note that the i superscript notation does not inidicate an element of a set, but is instead intended to denote the function obtained by this backward *iteration.*) It must be shown that $V_t^i(\cdot,\cdot) = V_t(\cdot,\cdot)$, the value function. Let $V_T^i(x,q) = \mathcal{E}(x)$ for all $x \in \mathcal{X}$ and $q \in Q(\mathcal{X})$. We now suppose that one has $V_{t+1}^i(\cdot,\cdot)$, and demonstrate how one obtains $V_t^i(\cdot,\cdot)$.

1. First, let the vector-valued function $\vec{M}_t$ be given component-wise by

$$[\vec{M}_t]_x(q,u) = \max_{\vec{w}\in W^n} \Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{xj}(u,\vec{w}) V_{t+1}^i(j, q'(q,u,\vec{w})) \Big] \tag{50}$$

where

$$q'(q,u,\vec{w}) = \widetilde{P}^T(u,\vec{w})q \tag{51}$$

and the optimal $\vec{w}$ is

$$\vec{w}_t^0 = \vec{w}_t^0(x,q,u) = \operatorname*{argmax}_{\vec{w}\in W^n} \Big\{ \sum_{j\in\mathcal{X}} \widetilde{P}_{xj}(u,\vec{w}) V_{t+1}^i(j, q'(q,u,\vec{w})) \Big\}. \tag{52}$$

2. Then define L_t as

$$L_t(q,u) = q^T \vec{M}_t(q,u), \tag{53}$$

and note that the optimal u is

$$u_t^0 = u_t^0(q) = \operatorname*{argmin}_{u\in U} L_t(q,u) = \operatorname*{argmin}_{u\in U} q^T \vec{M}_t(q,u). \tag{54}$$

3. With this, one obtains the next iterate from

$$V_t^i(x,q) = \sum_{j\in\mathcal{X}} \widetilde{P}_{xj}(u_t^0, \vec{w}_t^0) V_{t+1}^i(j, q'(q, u_t^0, \vec{w}_t^0)) = [\vec{M}_t]_x(q, u_t^0) \tag{55}$$

and the corresponding best achievable expected result from the player 1 perspective is

$$V_t^{i,1}(q) = q^T \vec{M}_t(q, u_t^0). \tag{56}$$

Consequently, for each $t \in \{0,1,\ldots,T\}$ and each $x \in \mathcal{X}$, $V_t^i(x,\cdot)$ is a piecewise constant function over simplex $Q(\mathcal{X})$. (Once we obtain $V_t \equiv V_t^i$, this will obviously imply the corresponding piecewise constancy of the state feedback value function V_t.) Due to this piecewise constant nature, propagation is relatively straight-forward (more specifically, it is finite-dimensional in contradistinction to the general case). However, this is slightly less critical than the propagation issue for the information state of the unnormalized distribution, $\widetilde{\mathcal{I}}_t$, since the state feedback value may be pre-computed, while the information state must be propagated in real-time.

We now show that in fact, $V_t \equiv V^i_t$ for all $t \in [0, T]$. By definition, $V^i_T(x, q) = \mathcal{E}(x) = V_T(x, q)$ for all $x \in \mathcal{X}$ and $q \in Q(\mathcal{X})$. The next step in proving the equivalence is to prove that V_t satisfies the dynamic programming principle (DPP). For the problem considered here, the DPP takes the form of the following theorem.

Theorem 4.5 *Let $0 \leq t < t_1 \leq T$. Then*

$$V_t(x, q) = M_t(x, q, \overline{\lambda}^0_{[t,T-1]}) = \widetilde{M}_{t,t_1}(x, q, \tilde{\overline{\lambda}}^0_{[t,t_1-1]}) \tag{57}$$

where

$$\widetilde{M}_{t,t_1}(x, q, \overline{\lambda}_{[t,t_1-1]}) = \max_{\overline{\theta}_{[t,t_1-1]} \in \overline{\Theta}_{[t,t_1-1]}} \mathbf{E}[V_{t_1}(X_{t_1}, q_{t_1}) \,|\, X_t = x] \tag{58}$$

$q_t = q$, *and*

$$\tilde{\overline{\lambda}}^0_{[t,t_1-1]} = \operatorname*{argmin}_{\overline{\lambda}_{[t,t_1-1]} \in \overline{\Lambda}_{[t,t_1-1]}} \mathbf{E}_q \left\{ \widetilde{M}_{t,t_1}(X, q, \overline{\lambda}_{[t,t_1-1]}) \right\}. \tag{59}$$

Proof. The first equality in (57) is merely a restatement of the definition (46), and one needs only to prove the second equality in (57). From (58), one has

$$\widetilde{M}_{t,t_1}(x, q, \tilde{\overline{\lambda}}^0_{[t,t_1-1]}) = \max_{\overline{\theta}_{[t,t_1-1]} \in \overline{\Theta}_{[t,t_1-1]}} \mathbf{E}[V_{t_1}(X_{t_1}, q_{t_1}) \,|\, X_t = x] \tag{60}$$

where to be specific, we note that X_{t_1} is generated from $X_t = x$ according to dynamics (1) with controls $\tilde{\overline{\lambda}}^0_{[t,t_1-1]}$ and $\overline{\theta}_{[t,t_1-1]}$, and q_{t_1} is generated similarly from $q_t = q$. By definition (46)

$$V_{t_1}(X_{t_1}, q_{t_1}) = M_{t_1}(X_{t_1}, q_{t_1}, \overline{\lambda}^0_{[t_1,T-1]}) \tag{61}$$

where

$$\overline{\lambda}^0_{[t_1,T-1]} = \operatorname*{argmin}_{\overline{\lambda}_{[t_1,T-1]} \in \overline{\Lambda}_{[t_1,T-1]}} \mathbf{E}_{q_{t_1}} \{ M_{t_1}(X_{t_1}, q_{t_1}, \overline{\lambda}_{[t_1,T-1]}) \}. \tag{62}$$

Note also that by definition (43)

$$M_{t_1}(z, q_{t_1}, \overline{\lambda}^0_{[t_1,T-1]}) = \max_{\overline{\theta}_{[t_1,T-1]} \in \overline{\Theta}_{[t_1,T-1]}} \mathbf{E}[\mathcal{E}(\widehat{X}_T) \,|\, \widehat{X}_{t_1} = z] \tag{63}$$

for any $z \in \mathcal{X}$ where $\widehat{X}_\cdot$ propagates according to dynamics (1) with controls $\overline{\lambda}^0_{[t_1,T-1]}$ and $\overline{\theta}_{[t_1,T-1]}$. Substituting (61) into (60) yields

$$\widetilde{M}_{t,t_1}(x, q, \tilde{\overline{\lambda}}^0_{[t,t_1-1]}) = \max_{\overline{\theta}_{[t,t_1-1]} \in \overline{\Theta}_{[t,t_1-1]}} \mathbf{E}[M_{t_1}(X_{t_1}, q_{t_1}, \overline{\lambda}^0_{[t_1,T-1]}) \,|\, X_t = x]$$

which by (63)

$$= \max_{\tilde{\bar{\theta}}_{[t,t_1-1]}\in\overline{\Theta}_{[t,t_1-1]}} \mathbf{E}\Big\{ \max_{\bar{\theta}_{[t_1,T-1]}\in\overline{\Theta}_{[t_1,T-1]}} \mathbf{E}[\mathcal{E}(\widehat{X}_T)\,|\,\widehat{X}_{t_1}=X_{t_1}]\,|\,X_t=x\Big\}$$

which by Lemma 4.3 and the definition of $\overline{\Theta}$

$$= \max_{\bar{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}\Big\{\mathbf{E}[\mathcal{E}(\widehat{X}_T)\,|\,\widehat{X}_{t_1}=X_{t_1}]\,|\,X_t=x\Big\}$$

$$= \max_{\bar{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}\Big[\mathcal{E}(X_T)\,|\,X_t=x\Big] \tag{64}$$

where $X.$ propagates according to dynamics (1) with controls $\tilde{\bar{\lambda}}^0_{[t,t_1-1]}\cup\bar{\lambda}^0_{[t_1,T-1]}$ and $\bar{\theta}_{[t,T-1]}$.

Note that by (59) and (64),

$$\tilde{\bar{\lambda}}^0_{[t,t_1-1]} = \operatorname*{argmin}_{\bar{\lambda}_{[t,t_1-1]}\in\overline{\Lambda}_{[t,t_1-1]}} \mathbf{E}_q\left\{\max_{\bar{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}\Big[\mathcal{E}(X_T)\,|\,X_t=X\Big]\right\} \tag{65}$$

where $X.$ propagates according to dynamics (1) with controls $\bar{\lambda}_{[t,t_1-1]}\cup\bar{\lambda}^0_{[t_1,T-1]}$ and $\bar{\theta}_{[t,T-1]}$. Now, if $\tilde{\bar{\lambda}}^0_{[t,t_1-1]}\cup\bar{\lambda}^0_{[t_1,T-1]}$ achieved a lower cost in (64) than $\bar{\lambda}^0_{[t,T-1]}$, then this would contradict the optimality of $\bar{\lambda}^0_{[t,T-1]}$. Alternatively, $\tilde{\bar{\lambda}}^0_{[t,t_1-1]}$ achieves the lowest cost when paired with $\bar{\lambda}^0_{[t_1,T-1]}$ by (65), and so $\tilde{\bar{\lambda}}^0_{[t,t_1-1]}\cup\bar{\lambda}^0_{[t_1,T-1]}$ must yield the same cost in (64) as $\bar{\lambda}^0_{[t,T-1]}$. Therefore, by (64) and (43),

$$\widetilde{M}_{t,t_1}(x,q,\tilde{\bar{\lambda}}^0_{[t,t_1-1]}) = M_t(x,q,\bar{\lambda}^0_{[t,T-1]})$$

which completes the proof. □

We now continue with the proof that the $V^i_\cdot$ obtained by the DP iteration is the state feedback value function, $V.$. Recall that by definition

$$V_T(x,q)=\mathcal{E}(x)=V^i_T(x,q) \qquad \forall\, x\in\mathcal{X},\ q\in Q.$$

We will first propagate this equality back a single step.

By (43), for any $\bar{\lambda}_{T-1}:q\to U$,

$$M_{T-1}(x,q,\bar{\lambda}_{T-1}) = \max_{\vec{w}\in W^n}\sum_{j\in\mathcal{X}}\widetilde{P}_{xj}(\lambda_{T-1}[q],\vec{w})\vec{\mathcal{E}}_j \qquad \forall\, x\in\mathcal{X},\ q\in Q.$$

Letting

$$u_q \doteq \bar{\lambda}_{T-1}[q] \qquad \forall\, q\in Q, \tag{66}$$

one then has

$$M_{T-1}(x,q,\bar{\lambda}_{T-1}) = \max_{\vec{w}\in W^n}\sum_{j\in\mathcal{X}}\widetilde{P}_{xj}(u_q,\vec{w})\vec{\mathcal{E}}_j \qquad \forall\, x\in\mathcal{X},\ q\in Q. \tag{67}$$

Let

$$\overline{\lambda}^0_{T-1}[q] \doteq \underset{\overline{\lambda}_{T-1}\in\overline{\Lambda}_{[T-1,T-1]}}{\operatorname{argmin}} \mathbf{E}_q[M_{T-1}(X,q,\overline{\lambda}_{T-1})]. \tag{68}$$

By (50), (66) and (67), for any $q \in Q$, $x \in \mathcal{X}$ and $\overline{\lambda}_{T-1} \in \overline{\Lambda}_{[T-1,T-1]}$,

$$\begin{aligned}[\vec{M}_{T-1}]_x(q,u_q) &= \max_{\vec{w}\in W^n} \sum_{j\in\mathcal{X}} \widetilde{P}_{xj}(u_q,\vec{w})\vec{\mathcal{E}}_j \\ &= M_{T-1}(x,q,\overline{\lambda}_{T-1}). \end{aligned}\tag{69}$$

Note that for any $q \in Q$ and any $\overline{\lambda}_{T-1} \in \overline{\Lambda}_{[T-1,T-1]}$,

$$q^T\vec{M}_{T-1}(q,u_q) = \sum_{i\in\mathcal{X}} q_i[\vec{M}_{T-1}]_i(q,u_q)$$

when u_q is given by (66). By (69)

$$\begin{aligned}&= \sum_{i\in\mathcal{X}} q_i M_{T-1}(i,q,\overline{\lambda}_{T-1}) \\ &= E_q\Big\{M_{T-1}(X,q,\overline{\lambda}_{T-1})\Big\}.\end{aligned}\tag{70}$$

By (70), (45) and (54), for any $q \in Q$,

$$\overline{\lambda}^0_{T-1}[q] = u^0_{T-1}. \tag{71}$$

By (69) and (71),

$$[\vec{M}_{T-1}]_x(q,u^0_{T-1}) = M_{T-1}(x,q,\overline{\lambda}^0_{T-1}) \qquad \forall\, x\in\mathcal{X}, q\in Q. \tag{72}$$

Then, by (72),(46) and (55)

$$V_{T-1}(x,q) = V^i_{T-1}(x,q) \qquad \forall\, x\in\mathcal{X}, q\in Q, \tag{73}$$

and also by (73), (47) and (56)

$$V^1_{T-1}(q) = V^{i,1}_{T-1}(q) \qquad \forall\, q\in Q. \tag{74}$$

This validates the DP iteration for the first (backward) time-step.

Now we validate the DP iteration for all t by induction. Suppose

$$V_{t+1}(x,q) = M_{t+1}(x,q,\overline{\lambda}^0_{t+1}) = [\vec{M}_{t+1}]_x(q,u^0_{t+1}) = V^i_{t+1}(x,q) \qquad \forall\, x\in\mathcal{X}, q\in Q \tag{75}$$

which is true for $t+1 = T-1$ by (73). By Theorem 4.5, for any x, q,

$$\begin{aligned}V_t(x,q) &= M_t(x,q,\overline{\lambda}^0_{[t,T-1]}) \\ &= \widetilde{M}_{t,t+1}(x,q,\widetilde{\overline{\lambda}}^0_t)\end{aligned}$$

which by definition (58)

$$= \max_{\overline{\theta}_t \in \overline{\Theta}_{[t,t]}} \mathbf{E}[V_{t+1}(X_{t+1}, q_{t+1}) \mid X_t = x]$$

where propagation from X_t to X_{t+1} is with controls $\widetilde{\overline{\lambda}}^0_t$, $\overline{\theta}_t$. By (75) and the definition of $\overline{\Theta}$, this is

$$= \max_{\vec{w} \in W^n} \mathbf{E}[V^i_{t+1}(X_{t+1}, q_{t+1}) \mid X_t = x]$$

and since it is easily shown (as in (71)) that $u^0_t = \widetilde{\overline{\lambda}}^0_t[q]$

$$= \max_{\vec{w} \in W^n} \sum_{j \in \mathcal{X}} \widetilde{P}_{xj}(u^0_t, \vec{w}) V^i_{t+1}(j, q_{t+1})$$

which by the notation of (51)

$$= \max_{\vec{w} \in W^n} \sum_{j \in \mathcal{X}} \widetilde{P}_{xj}(u^0_t, \vec{w}) V^i_{t+1}(j, q'(q, u^0_t, \vec{w}))$$

which by (52) and (55)

$$= V^i_t(x, q).$$

Therefore, by induction, one has:

Theorem 4.6

$$V_t = V^i_t \qquad \forall\, t \in [0, T]$$

and of course

$$V^1_t = V^{i,1}_t \qquad \forall\, t \in [0, T].$$

This validates the DP iteration (50)–(56) as a means for computing the state feedback value function, V_t.

5 Robustness

The last step in the computation of the control at each time instant is now discussed. The control computation for such games is typically performed via the use of the Certainty Equivalence Principle (cf. [3], [18]). When the Certainty Equivalence Principle holds, the information state and state feedback value function can be combined to obtain the "optimal" controls which can be shown to be robust in a sense to be discussed below. The chief gain is that this allows one to compute a controller ahead of time, and then only propagate the information state "estimator" forward in time rather than computing the control as a function of the information state in real time. Otherwise, the computational cost would be prohibitive.

To simplify notation, note that by (53), (50) and Theorem 4.6 for any u,

$$L_t(q, u) = \mathbf{E}_q\Big[\max_{\vec{w} \in W^n} \sum_{j \in \mathcal{X}} \widetilde{P}_{xj}(u, \vec{w}) V_{t+1}(j, q'(q, u, \vec{w}))\Big]$$

where the notation $q'(q,u,\vec{w})$ is defined in (51). Let us hypothesize that the optimal control for player 1 is

$$u_t^m \doteq \operatorname*{argmin}_{u\in U}\left[\max_{q\in Q(\mathcal{X})}\{\mathcal{I}_t(q)+L_t(q,u)\}\right]. \tag{76}$$

Note here that this uses $\mathcal{I}_t$ not $\widetilde{\mathcal{I}}_t$ (the function of unnormalized distribution), and one may transform via the transformation from unnormalized $\widetilde{q}$ to normalized q. Alternatively, it may sometimes be computationally more efficient to do the maximization in the unnormalized space since there the unnormalized versions of V and L are piecewise constant while the unnormalized $\widetilde{\mathcal{I}}_t$ remains piecewise linear in the piecewise linear initial information state case. In either case, one has obvious robust game inequalities such as the following. (Note here that u^m will be a *strict minimizer* of a function $f(u)$ if $f(u^m) < f(u)$ for all $u \neq u^m$.)

Theorem 5.1 *Suppose u_t^m is a strict minimizer. Then, given any $\tilde{u}_t \neq u_t^m$, there exist $q^1, \vec{w}^1$ and $\varepsilon > 0$ such that*

$$\begin{aligned}&\Big\{\mathcal{I}_t(q^1)+\mathbf{E}_{q^1}[\sum_{j\in\mathcal{X}}\widetilde{P}_{xj}(\tilde{u}_t,\vec{w}^1)V_{t+1}(j,q'(q^1,\tilde{u}_t,\vec{w}^1))]\Big\} \\ &\quad > \max_{q\in Q(\mathcal{X})}\Big\{\mathcal{I}_t(q)+\mathbf{E}_q\max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}}\widetilde{P}_{xj}(u_t^m,\vec{w})V_{t+1}(j,q'(q,u_t^m,\vec{w}))\Big]\Big\}+\varepsilon\end{aligned} \tag{77}$$

where recall that we set $\mathcal{I}_t(q) = -\infty$ for $q \notin Q_t$.

Proof. By assumption, there exists $\hat{\varepsilon} > 0$ such that

$$\max_{q\in Q(\mathcal{X})}\Big\{\mathcal{I}_t(q)+L_t(q,\tilde{u}_t)\Big\} \geq \max_{q\in Q(\mathcal{X})}\Big\{\mathcal{I}_t(q)+L_t(q,u_t^m)\Big\}+3\hat{\varepsilon}.$$

Letting q^1 be $\hat{\varepsilon}$–optimal for the left-hand side yields

$$\mathcal{I}_t(q^1)+L_t(q^1,\tilde{u}_t) \geq \max_{q\in Q(\mathcal{X})}\Big\{\mathcal{I}_t(q)+L_t(q,u_t^m)\Big\}+2\hat{\varepsilon}.$$

Expanding the left-hand side, this is

$$\begin{aligned}&\mathcal{I}_t(q^1)+\mathbf{E}_{q^1}\Big\{\max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}}\widetilde{P}_{xj}(\tilde{u}_t,\vec{w})V_{t+1}(j,q'(q^1,\tilde{u}_t,\vec{w}))\Big]\Big\} \\ &\quad \geq \max_{q\in Q(\mathcal{X})}\Big\{\mathcal{I}_t(q)+L_t(q,u_t^m)\Big\}+2\hat{\varepsilon}.\end{aligned}$$

Letting $\vec{w}^1$ be $\hat{\varepsilon}$–optimal for the left-hand side yields the result. □

Corollary 5.2 *Let $\mathcal{I}_0$, $u_{[0,t-1]}$ and $y_{[0,t-1]}$ be given. Suppose u_t^m is a strict minimizer. Then, given any $\tilde{u}_t \neq u_t^m$, there exist $q_0^1 \in Q(\mathcal{X})$, $\vec{w}^1 \in [W^n]^t$ and $\varepsilon > 0$ such that*

$$\begin{aligned}&\mathbf{E}_{q_t''}\Big\{\mathcal{I}_0(q_0^1) + \sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(\tilde{u}_t, \vec{w}_t^1)V_{t+1}(j, q'(q_t'', \tilde{u}_t, \vec{w}_t^1))\Big\} \\ &> \max_{q_0 \in Q(\mathcal{X})} \max_{\vec{w}_{[0,t]}\in[W^n]^{t+1}} \mathbf{E}_{q_t'}\Big\{\mathcal{I}_0(q_0) + \sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u_t^m, \vec{w}_t)V_{t+1}(j, q'(q_t', u_t^m, \vec{w}_t))\Big\} + \varepsilon\end{aligned} \tag{78}$$

where conditional distributions q_t'' and q_t' are generated by propagation (15) for given $q_0^1, u_{[0,t-1]}, y_{[0,t-1]}, \vec{w}^1_{[0,t-1]}$ and $q_0, u_{[0,t-1]}, y_{[0,t-1]}, \vec{w}_{[0,t-1]}$, respectively.

Proof. It is sufficient to show that the left and right hand sides of (77) are equivalent to the left and right hand sides of (78). Consider the right hand side of (77). From the definition of $\mathcal{I}_t$ (see (16)), and noting that $Q_t \neq \emptyset$ by Lemma 3.2, one finds that

$$\begin{aligned}&\max_{q\in Q(\mathcal{X})}\Big\{\mathcal{I}_t(q) + \mathbf{E}_q \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u_t^m, \vec{w})V_{t+1}(j, q'(q, u_t^m, \vec{w}))\Big]\Big\} \\ &= \max_{q\in Q_t}\Big\{\mathcal{I}_t(q) + \mathbf{E}_q \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u_t^m, \vec{w})V_{t+1}(j, q'(q, u_t^m, \vec{w}))\Big]\Big\} \\ &= \max_{q\in Q_t} \max_{q_0\in Q_0^{q,u}} \max_{w \in \widetilde{W}_{t-1}^{q_0,q}}\Big\{\mathcal{I}_0(q_0) + \mathbf{E}_q \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u_t^m, \vec{w})V_{t+1}(j, q'(q, u_t^m, \vec{w}))\Big]\Big\}\end{aligned} \tag{79}$$

where $\widetilde{W}_{t-1}^{q_0,q} = \{\vec{w}_{[0,t-1]} \in \widetilde{W}_{t-1}^q : q_t = q$ where q_t is given by propagation (15) with initial $q_0\}$, $\widetilde{W}_{t-1}^q$ is defined in (23), and $Q_0^{q,u} = Q_0^{q,u_{[0,t-1]}}$ is defined in (18). For each $q_0 \in Q(\mathcal{X})$ and $\vec{w}_{[0,t-1]}$, there exists $q = q_t' \in Q_t$ such that q_t' is given by propagation (15), and so $q_0 \in Q_0^{q,u}$ and $\vec{w}_{[0,t-1]} \in \widetilde{W}_{t-1}^{q_0,q}$. Conversely, given any $q \in Q_t$, $q_0 \in Q_0^{q,u}$ and $\vec{w}_{[0,t-1]} \in \widetilde{W}_{t-1}^{q_0,q}$, $q_0 \in Q(\mathcal{X})$ and $\vec{w}_{[0,t-1]} \in [W^n]^t$. Consequently, (79) becomes

$$\begin{aligned}&\max_{q\in Q(\mathcal{X})}\Big\{\mathcal{I}_t(q) + \mathbf{E}_q \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u_t^m, \vec{w})V_{t+1}(j, q'(q, u_t^m, \vec{w}))\Big]\Big\} \\ &= \max_{q_0\in Q(\mathcal{X})} \max_{w_{[0,t-1]}\in[W^n]^t}\Big\{\mathcal{I}_0(q_0) + \mathbf{E}_{q_t'} \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u_t^m, \vec{w})V_{t+1}(j, q'(q_t', u_t^m, \vec{w}))\Big]\Big\}\end{aligned} \tag{80}$$

where q_t' is given by propagation (15) with initial q_0, controls $u_{[0,t-1]}$ and $\vec{w}_{[0,t-1]}$, and observations $y_{[0,t-1]}$, and noting that $\mathcal{I}_t$ is deterministic (given $y_{[0,t-1]}$)

$$= \max_{q_0\in Q(\mathcal{X})} \max_{w_{[0,t-1]}\in[W^n]^t} \mathbf{E}_{q_t'}\Big\{\mathcal{I}_0(q_0) + \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u_t^m, \vec{w})V_{t+1}(j, q'(q_t', u_t^m, \vec{w}))\Big]\Big\}$$

and since W^n consists of state feedback controls

$$= \max_{q_0\in Q(\mathcal{X})} \max_{w_{[0,t]}\in[W^n]^{t+1}} \mathbf{E}_{q_t'}\Big\{\mathcal{I}_0(q_0) + \sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u_t^m, \vec{w}_t)V_{t+1}(j, q'(q_t', u_t^m, \vec{w}_t))\Big\} \tag{81}$$

which is the desired equivalence for the right hand sides. Proceeding similarly with the left hand sides yields the result. □

Theorem 5.1 and Corollary 5.2 provide statements regarding robustness, but this is only with respect to the defined criterion of $I_t(q) + L_t(q, u)$. It still remains to relate this to the original problem definition. That is, one must relate the criterion in these results to an originating imperfect observation value function defined in terms of the worst-case expected cost (from the player 1 point of view). In order to make this section more readable, we will begin by writing down the value function, and then describe the terms within it rather than vice-versa. For technical reasons, it appears best to work with the following value function. This value at any time $\bar{t}$ is

$$Z_{\bar{t}} \doteq \sup_{q_{\bar{t}} \in Q_t} \inf_{\lambda_{[\bar{t},T-1]} \in \Lambda_{[\bar{t},T-1]}} \sup_{\theta_{[\bar{t},T-1]} \in \theta_{[\bar{t},T-1]}} \Big[I_{\bar{t}}(q_{\bar{t}}) + \mathbf{E}_{q_{\bar{t}}} \left\{ \mathbf{E}[\mathcal{E}(X_T) |\, X_{\bar{t}} = X] \right\} \Big]. \tag{82}$$

The expectation uses the (player 1) assumption that the distribution of $X_{\bar{t}}$ is $q_{\bar{t}}$ for each $q_{\bar{t}} \in Q_{\bar{t}}$ and is taken not only over $X_{\bar{t}}$ but also over all observation and dynamic noise from time $\bar{t}$ to terminal time T. Note that this is not a full upper value in that the supremum over $q_{\bar{t}}$ occurs outside the infimum over player 1 controls $\lambda^{[\bar{t},T-1]}$. The strategy set for player 1 is

$$\Lambda^{[\bar{t},T-1]} = \Big\{ \lambda_{[\bar{t},T-1]} : \, Y^{T-\bar{t}} \to U^{T-\bar{t}}, \text{ nonanticipative in } y_{\cdot-1} \Big\}$$

where "nonanticipative in $y_{\cdot-1}$" is defined as follows. A strategy, $\lambda_{[\bar{t},T-1]}$ is nonanticipative in $y_{\cdot-1}$ if given any $t \in [\bar{t}, T-1]$ and any sequences $y_\cdot, \tilde{y}_\cdot$ such that $y_r = \tilde{y}_r$ for all $r \in [\bar{t}, t-1]$, one has $\lambda_t[y] = \lambda_t[\tilde{y}]$. Note that since the infimum over $\lambda_{[\bar{t},T-1]}$ in (82) occurs inside the supremum over $q_{\bar{t}}$, the "optimal" choice of λ may depend on $q_{\bar{t}}$. Also note that the "optimal" choice of $\lambda_{[\bar{t},T-1]}$ may depend on $I_{\bar{t}}(\cdot)$. The strategy set for player 2 (neglecting $q_{\bar{t}}$ as a player 2 control) is naturally

$$\Theta_{[\bar{t},T-1]} = \Big\{ \theta_{[\bar{t},T-1]} : Y^{T-\bar{t}} \to W^{n(T-\bar{t})}, \text{ nonanticipative in } y_{\cdot-1} \Big\}. \tag{83}$$

Also, $Q_t = Q_t(u_{[0,t-1]}, y_{[0,t-1]})$ as given in (14). Since the supremum over $\theta_{[\bar{t},T-1]}$ is inside the infimum, and the $\vec{w}_{[\bar{t},T-1]}$ process is a feedback on the state, then as in Lemma 4.1, one can replace the supremum over $\theta_{[\bar{t},T-1]} \in \Theta_{[\bar{t},T-1]}$ with a maximum over $\vec{w}_{[\bar{t},T-1]} \in W^{n(T-\bar{t})}$, and so

$$Z_{\bar{t}} = \sup_{q_{\bar{t}} \in Q_t} \inf_{\lambda_{[\bar{t},T-1]} \in \Lambda_{[\bar{t},T-1]}} \max_{\vec{w}_{[\bar{t},T-1]} \in W^{n(T-\bar{t})}} \left[I_{\bar{t}}(q_{\bar{t}}) + \mathbf{E}_{q_{\bar{t}}} \left\{ \mathcal{E}(X_T) \right\} \right]. \tag{84}$$

It is helpful to modify the notation of (12), (13) slightly by including the observation dependence in the superscript so that one has observation update (with observation $y_t = y$)

$$\widehat{q}_t = \left(\tfrac{1}{\widetilde{R}^{y\ T}(u_t, \vec{w}) q_t} \right) D(\widetilde{R}^y(u_t, \vec{w})) q_t \tag{85}$$

and dynamics update

$$q_{t+1} = \widetilde{P}^T(u_t, \vec{w}_t) \widehat{q}_t. \tag{86}$$

For $t \geq \bar{t}$, let $\overline{q}_{ti}$ be the probability that $X_t = i$ conditioned only on the observations only up through time $\bar{t}-1$. Then obviously $\overline{q}_{\bar{t}} = q_{\bar{t}}$. For completeness, let $\overline{q}_t = q_t$ for $t < \bar{t}$. Also, given $\overline{q}_t$ and any choice of $\lambda_{[\bar{t},T-1]}, \vec{w}_{[\bar{t},T-1]}$, one obtains $\overline{q}_{t+1}$ from the following (where for compactness, we abuse notation by writing $\widetilde{P}_{j,i}(\lambda_t, \vec{w}_t)$ in place of $\widetilde{P}_{j,i}(\lambda_t[y_{[\bar{t},T-1]}], \vec{w}_t)$ and $\widetilde{R}^y(\lambda_t, \vec{w})$ in place of $\widetilde{R}^y(\lambda_t[y_{[\bar{t},T-1]}], \vec{w})$).

$$\overline{q}_{t+1_i} = \sum_{j\in\mathcal{X}} \sum_{y\in Y} \widetilde{P}_{j,i}(\lambda_t, \vec{w}_t) \left(\frac{1}{\sum_{l\in\mathcal{X}} \widetilde{R}_l^{y\ T}(\lambda_t, \vec{w}_t)\overline{q}_{tl}} \right) \widetilde{R}_j^y(\lambda_t, \vec{w}_t)\overline{q}_{tj}\, P(y_t = y)$$

where $P(y_t = y)$ indicates the probability that observation $y_t = y$

$$\begin{aligned}
&= \sum_{j\in\mathcal{X}} \sum_{l'\in\mathcal{X}} \sum_{y\in Y} \widetilde{P}_{j,i}(\lambda_t, \vec{w}_t) \left(\frac{1}{\sum_{l\in\mathcal{X}} \widetilde{R}_l^{y}(\lambda_t, \vec{w}_t)\overline{q}_{tl}} \right) \widetilde{R}_j^y(\lambda_t, \vec{w}_t)\overline{q}_{tj}\, \widetilde{R}_{l'}^y(\lambda_t, \vec{w}_t)\overline{q}_{tl'} \\
&= \sum_{j\in\mathcal{X}} \sum_{y\in Y} \widetilde{P}_{j,i}(\lambda_t, \vec{w}_t)\widetilde{R}_j^y(\lambda_t, \vec{w}_t)\overline{q}_{tj} \\
&= \sum_{j\in\mathcal{X}} \widetilde{P}_{j,i}(\lambda_t, \vec{w}_t)\overline{q}_{tj}. \qquad (87)
\end{aligned}$$

Using $\overline{q}_T$, (84) may be rewritten as

$$Z_{\bar{t}} = \sup_{q_{\bar{t}}\in Q_{\bar{t}}} \inf_{\lambda_{[\bar{t},T-1]}\in\Lambda_{[\bar{t},T-1]}} \max_{\vec{w}_{[\bar{t},T-1]}\in W^{n(T-\bar{t})}} \left[I_{\bar{t}}(q_{\bar{t}}) + \overline{q}_T^T \vec{\mathcal{E}} \right].$$

which by (87)

$$\begin{aligned}
&= \sup_{q_{\bar{t}}\in Q_{\bar{t}}} \inf_{\lambda_{[\bar{t},T-1]}\in\Lambda_{[\bar{t},T-1]}} \max_{\vec{w}_{[\bar{t},T-1]}\in W^{n(T-\bar{t})}} \left[I_{\bar{t}}(q_{\bar{t}}) + \overline{q}_{T-1}^T \widetilde{P}(\lambda_{T-1}, \vec{w}_{T-1}) \vec{\mathcal{E}} \right] \\
&= \sup_{q_{\bar{t}}\in Q_{\bar{t}}} \inf_{\lambda_{[\bar{t},T-2]}\in\Lambda_{[\bar{t},T-2]}} \inf_{\lambda_{T-1}\in\Lambda_{[T-1,T-1]}} \\
&\qquad \max_{\vec{w}_{[\bar{t},T-1]}\in W^{n(T-\bar{t})}} \left[I_{\bar{t}}(q_{\bar{t}}) + \overline{q}_{T-1}^T \widetilde{P}(\lambda_{T-1}, \vec{w}_{T-1}) \vec{\mathcal{E}} \right]. \qquad (88)
\end{aligned}$$

Note that, not including λ_{T-1}, the dependence of the bracketed term on the $y.$ process is only through $\overline{q}.$. (This is of course simply an instance of the principle that the conditional probability is a sufficient statistic, but in the nonstandard context of a stochastic game.) Thus we may replace the infimum over $\lambda_{T-1} \in \Lambda_{[T-1,T-1]}$ by an infimum over $\overline{\lambda}_{T-1} \in \overline{\Lambda}_{[T-1,T-1]}$ where

$$\overline{\Lambda}_{[t,T-1]} = \{\overline{\lambda}_{[t,T-1]} : Q^{T-t} \to U^{T-t}, \text{ nonanticipative in } \overline{q}.\}$$

In other words,

$$\begin{aligned}
Z_{\bar{t}} &= \sup_{q_{\bar{t}}\in Q_{\bar{t}}} \inf_{\lambda_{[\bar{t},T-2]}\in\Lambda_{[\bar{t},T-2]}} \inf_{\overline{\lambda}_{T-1}\in\overline{\Lambda}_{[T-1,T-1]}} \\
&\qquad \max_{\vec{w}_{[\bar{t},T-1]}\in W^{n(T-\bar{t})}} \left[I_{\bar{t}}(q_{\bar{t}}) + \overline{q}_{T-1}^T \widetilde{P}(\overline{\lambda}_{T-1}, \vec{w}_{T-1}) \vec{\mathcal{E}} \right]. \qquad (89)
\end{aligned}$$

One may step backward another step with this same procedure. It is perhaps worth recalling here that the control at time $T-2$ depends only on the observations up through time $T-3$. Applying (87) to expand $\overline{q}_{T-1}$, (89) becomes

$$\begin{aligned} Z_{\bar t} = \sup_{q_{\bar t}\in Q_{\bar t}} \inf_{\lambda_{[\bar t,T-2]}\in\Lambda_{[\bar t,T-2]}} \inf_{\overline{\lambda}_{T-1}\in\overline{\Lambda}_{[T-1,T-1]}} & \\ \max_{\vec{w}_{[\bar t,T-1]}\in W^{n(T-\bar t)}} & \Big[I_{\bar t}(q_{\bar t}) + \overline{q}^T_{T-2}\widetilde{P}(\lambda_{T-2},\vec{w}_{T-2})\widetilde{P}(\overline{\lambda}_{T-1},\vec{w}_{T-1})\vec{\mathcal{E}}\Big]. \end{aligned} \tag{90}$$

One then notes that, not including λ_{T-2}, the bracketed term depends on $y_{[\bar t,T-3]}$ only through $\overline{q}_{T-2}$. Consequently, (90) becomes

$$\begin{aligned} Z_{\bar t} = \sup_{q_{\bar t}\in Q_{\bar t}} \inf_{\lambda_{[\bar t,T-3]}\in\Lambda_{[\bar t,T-3]}} \inf_{\overline{\lambda}_{[T-2,T-1]}\in\overline{\Lambda}_{[T-2,T-1]}} & \\ \max_{\vec{w}_{[\bar t,T-1]}\in W^{n(T-\bar t)}} & \Big[I_{\bar t}(q_{\bar t}) + \overline{q}^T_{T-2}\widetilde{P}(\overline{\lambda}_{T-2},\vec{w}_{T-2})\widetilde{P}(\overline{\lambda}_{T-1},\vec{w}_{T-1})\vec{\mathcal{E}}\Big]. \end{aligned} \tag{91}$$

Proceeding inductively and recalling that $\overline{q}_{\bar t} = q_{\bar t}$, one obtains

$$Z_{\bar t} = \sup_{q_{\bar t}\in Q_{\bar t}} \inf_{\overline{\lambda}_{[\bar t,T-1]}\in\overline{\Lambda}_{[\bar t,T-1]}} \max_{\vec{w}_{[\bar t,T-1]}\in W^{n(T-\bar t)}} \left[I_{\bar t}(q_{\bar t}) + q^T_{\bar t}\left(\prod_{t=\bar t}^{T-1}\widetilde{P}(\overline{\lambda}_t,\vec{w}_t)\right)\vec{\mathcal{E}}\right] \tag{92}$$

$$= \sup_{q_{\bar t}\in Q_{\bar t}} \inf_{\overline{\lambda}_{[\bar t,T-1]}\in\overline{\Lambda}_{[\bar t,T-1]}} \max_{\vec{w}_{[\bar t,T-1]}\in W^{n(T-\bar t)}} \Big[I_{\bar t}(q_{\bar t}) + \mathbf{E}_{q_{\bar t}}\{\,\mathbf{E}\left[\,\mathcal{E}(X_T)\,|\,X_{\bar t}=X\right]\}\Big]. \tag{93}$$

Comparing (93) to (84), one sees that it has been shown that the value $Z_{\bar t}$ is unchanged if the future player 1 planned controls are assumed to depend only on the conditional probability process rather than the entire observation process. (Again, this is merely a particular instance of the principle that the conditional probability is a sufficient statistic.)

Now, since $I_{\bar t}(q_{\bar t})$ is independent of future control choices, (92), (93) may be rewritten as

$$Z_{\bar t} = \sup_{q_{\bar t}\in Q_{\bar t}} \left[I_{\bar t}(q_{\bar t}) + \inf_{\overline{\lambda}_{[\bar t,T-1]}\in\overline{\Lambda}_{[\bar t,T-1]}} \max_{\vec{w}_{[\bar t,T-1]}\in W^{n(T-\bar t)}} \left\{ q^T_{\bar t}\left(\prod_{t=\bar t}^{T-1}\widetilde{P}(\overline{\lambda}_t,\vec{w}_t)\right)\vec{\mathcal{E}}\right\}\right] \tag{94}$$

$$= \sup_{q_{\bar t}\in Q_{\bar t}} \Big[I_{\bar t}(q_{\bar t}) + \inf_{\overline{\lambda}_{[\bar t,T-1]}\in\overline{\Lambda}_{[\bar t,T-1]}} \max_{\vec{w}_{[\bar t,T-1]}\in W^{n(T-\bar t)}} \mathbf{E}_{q_{\bar t}}\{\,\mathbf{E}\left[\,\mathcal{E}(X_T)\,|\,X_{\bar t}=X\right]\}\Big]. \tag{95}$$

Then, using Lemmas 4.1 and 4.3 and Remark 4.2, this yields

$$Z_{\bar t} = \sup_{q_{\bar t}\in Q_{\bar t}} \left[I_{\bar t}(q_{\bar t}) + \inf_{\overline{\lambda}_{[\bar t,T-1]}\in\overline{\Lambda}_{[\bar t,T-1]}} \mathbf{E}_{q_{\bar t}}\Big\{ \max_{\overline{\theta}_{[\bar t,T-1]}\in\overline{\Theta}_{[\bar t,T-1]}} \mathbf{E}\left[\,\mathcal{E}(X_T)\,|\,X_{\bar t}=X\right]\Big\}\right] \tag{96}$$

$$= \sup_{q_{\bar t}\in Q_{\bar t}} \left[I_{\bar t}(q_{\bar t}) + \inf_{\overline{\lambda}_{[\bar t,T-1]}\in\overline{\Lambda}_{[\bar t,T-1]}} \mathbf{E}_{q_{\bar t}}\Big\{ \max_{\vec{w}_{[\bar t,T-1]}\in W^{n(T-\bar t)}} \mathbf{E}\left[\,\mathcal{E}(X_T)\,|\,X_{\bar t}=X\right]\Big\}\right]. \tag{97}$$

Substituting (39) into (96) (or (40) into (97)), one has

Theorem 5.3

$$Z_{\bar{t}} = \sup_{q_{\bar{t}} \in Q_{\bar{t}}} \Big[I_{\bar{t}}(q_{\bar{t}}) + V^1_{\bar{t}}(q_{\bar{t}}) \Big] \qquad \forall\, \bar{t} \in [0, T]. \tag{98}$$

In other words, the game value $Z_{\bar{t}}$ is the supremum of the sum of the information state, $I_{\bar{t}}$, and the optimal expected state feedback value, $V^1_{\bar{t}}$, from player 1's perspective. It is interesting to note that in the max-plus algebra [8], [4], (98) takes the form

$$Z_{\bar{t}} = \int^{\oplus}_{Q_{\bar{t}}} V^1_{\bar{t}}(q) \otimes I_{\bar{t}}(q)\, dq \tag{99}$$

where $\int^{\oplus}$ indicates max-plus integration. In other words, (99) is the max-plus expectation of $V^1_{\bar{t}}$ with respect to max-plus probability $I_{\bar{t}}$ (see [11], [25], [1] for example).

Also, from (39), by the definition of $\overline{\Lambda}_{[\bar{t}, T-1]}$, one has

$$V^1_{\bar{t}}(q) = \min_{u \in U} \inf_{\overline{\lambda}_{[\bar{t}+1, T-1]} \in \overline{\Lambda}_{[\bar{t}+1, T-1]}} \mathbf{E}_{q_{\bar{t}}} \Big\{ \max_{\vec{w}_{[\bar{t}, T-1]} \in W^{n(T-\bar{t})}} \mathbf{E}[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X] \Big\}$$

which upon using the definition of $M_{\bar{t}}$ (43)

$$= \min_{u \in U} \inf_{\overline{\lambda}_{[\bar{t}+1, T-1]} \in \overline{\Lambda}_{[\bar{t}+1, T-1]}} \mathbf{E}_{q_{\bar{t}}} \Big\{ M_{\bar{t}}(X, q, (u, \overline{\lambda}_{[\bar{t}+1, T-1]})) \Big\} \tag{100}$$

(where it could be noted that we may view $(u, \overline{\lambda}_{[\bar{t}+1, T-1]})$ as an element of $\overline{\Lambda}_{[\bar{t}+1, T-1]}$ which happens to be constant over $Q(\mathcal{X})$ at time $\bar{t}$). Then, defining $\widehat{L}$ to be the term inside the minimum over u,

$$= \min_{u \in U} \widehat{L}_{\bar{t}}(q, u). \tag{101}$$

It is useful to note the following.

Lemma 5.4

$$L_{\bar{t}} = \widehat{L}_{\bar{t}} \qquad \forall\, \bar{t} \in [0, T].$$

For purposes of presentation, the proof of Lemma 5.4 is delayed until after Theorem 5.5 below.

Using Theorem 5.3, Lemma 5.4 and (101), one has

$$\begin{aligned} Z_{\bar{t}} &= \sup_{q_{\bar{t}} \in Q_t} \Big[I_{\bar{t}}(q_{\bar{t}}) + \min_{u \in U} L_{\bar{t}}(q_{\bar{t}}, u) \Big] \\ &= \sup_{q_{\bar{t}} \in Q_t} \min_{u \in U} \Big[I_{\bar{t}}(q_{\bar{t}}) + L_{\bar{t}}(q_{\bar{t}}, u) \Big]. \end{aligned} \tag{102}$$

In order to obtain the Robustness/Certainty Equivalence result to follow, it is sufficient to make the following Saddle Point Assumption. We assume that

$$\sup_{q_{\bar{t}} \in Q_t} \min_{u \in U} \Big[I_{\bar{t}}(q_{\bar{t}}) + L_{\bar{t}}(q_{\bar{t}}, u) \Big] = \min_{u \in U} \sup_{q_{\bar{t}} \in Q_t} \Big[I_{\bar{t}}(q_{\bar{t}}) + L_{\bar{t}}(q_{\bar{t}}, u) \Big] \qquad \forall\, \bar{t} \in [0, T]. \tag{$A5.1$}$$

With Assumption (A5.1), (102) becomes

$$Z_{\bar{t}} = \min_{u\in U} \sup_{q_{\bar{t}}\in Q_t} \Big[I_{\bar{t}}(q_{\bar{t}}) + L_{\bar{t}}(q_{\bar{t}}, u) \Big]. \tag{103}$$

Recall the control choice from (76), $u^m_{\bar{t}}$, the definition of which was as follows (where we recall $\mathcal{I}_{\bar{t}} = -\infty$ off of $Q_{\bar{t}}$)

$$u^m_{\bar{t}} \doteq \operatorname*{argmin}_{u\in U} \left[\sup_{q\in Q_{\bar{t}}} \{ \mathcal{I}_{\bar{t}}(q) + L_{\bar{t}}(q,u) \} \right]. \tag{104}$$

Suppose that $u^m_{\bar{t}}$ is a strict minimizer (where we recall that u^m is a strict minimizier of a function f if $f(u^m) < f(u)$ for all $u \neq u^m$). Suppose $u \neq u^m_{\bar{t}}$. Then there exists $\varepsilon > 0$ (independent of u since U is finite) such that

$$\begin{aligned} Z_{\bar{t}} &= \sup_{q\in Q_{\bar{t}}} \{ \mathcal{I}_{\bar{t}}(q) + L_{\bar{t}}(q, u^m_{\bar{t}}) \} \\ &\leq \sup_{q\in Q_{\bar{t}}} \{ \mathcal{I}_{\bar{t}}(q) + L_{\bar{t}}(q,u) \} - 2\varepsilon \end{aligned} \tag{105}$$

for all $u \neq u^m_{\bar{t}}$.

Fix any $u \neq u^m_{\bar{t}}$. Let

$$Z^u_{\bar{t}} \doteq \sup_{q\in Q_{\bar{t}}} \inf_{\lambda_{[\bar{t},T-1]}\in\Lambda^u_{[\bar{t},T-1]}} \max_{\vec{w}_{[\bar{t},T-1]}\in W^{n(T-\bar{t})}} \Big[I_{\bar{t}}(q_{\bar{t}}) + \mathbf{E}_{q_{\bar{t}}} \Big\{ \mathbf{E} \left[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X \right] \Big\} \Big] \tag{106}$$

where

$$\Lambda^u_{[\bar{t},T-1]} \doteq \{ \lambda_{[\bar{t},T-1]} \in \Lambda_{[\bar{t},T-1]} \mid \lambda_{\bar{t}}[y.] = u \;\; \forall y. \in Y^{T-\bar{t}} \}.$$

Then an analysis essentially identical to that in (88)–(96) holds with Λ^u replacing Λ and letting $\overline{\Lambda}^u_{[\bar{t},T-1]} \doteq \{ \overline{\lambda}_{[\bar{t},T-1]} \in \overline{\Lambda}_{[\bar{t},T-1]} \mid \overline{\lambda}_{\bar{t}}[q.] = u \; \forall q. \in Q(\mathcal{X})^{T-\bar{t}} \}$. Consequently, one obtains

$$Z^u_{\bar{t}} = \sup_{q_{\bar{t}}\in Q_{\bar{t}}} \left[I_{\bar{t}}(q_{\bar{t}}) + \inf_{\overline{\lambda}_{[\bar{t},T-1]}\in\overline{\Lambda}^u_{[\bar{t},T-1]}} \mathbf{E}_{q_{\bar{t}}} \Big\{ \max_{\overline{\theta}_{[\bar{t},T-1]}\in\overline{\Theta}_{[\bar{t},T-1]}} \mathbf{E} \left[\mathcal{E}(X_T) \,|\, X_{\bar{t}} = X \right] \Big\} \right]$$

which by (43) again

$$= \sup_{q_{\bar{t}}\in Q_{\bar{t}}} \left[I_{\bar{t}}(q_{\bar{t}}) + \inf_{\overline{\lambda}_{[\bar{t},T-1]}\in\overline{\Lambda}^u_{[\bar{t},T-1]}} \mathbf{E}_{q_{\bar{t}}} \Big\{ M_{\bar{t}}(X, q, \overline{\lambda}_{[\bar{t},T-1]})) \Big\} \right]$$

and as before

$$\begin{aligned} &= \sup_{q_{\bar{t}}\in Q_{\bar{t}}} \left[I_{\bar{t}}(q_{\bar{t}}) + \inf_{\overline{\lambda}_{[\bar{t}+1,T-1]}\in\overline{\Lambda}_{[\bar{t}+1,T-1]}} \mathbf{E}_{q_{\bar{t}}} \Big\{ M_{\bar{t}}(X, q, (u, \overline{\lambda}_{\bar{t}+1,T-1]})) \Big\} \right] \\ &= \sup_{q_{\bar{t}}\in Q_{\bar{t}}} \Big[I_{\bar{t}}(q_{\bar{t}}) + \widehat{L}_{\bar{t}}(q,u) \Big] \end{aligned}$$

which by Lemma 5.4

$$= \sup_{q_{\bar t}\in Q_{\bar t}} \quad [I_{\bar t}(q_{\bar t}) + L_{\bar t}(q,u)]$$

which by (105)

$$\geq Z_{\bar t} + 2\varepsilon.$$

Combining this with (106), one finds that there exists $q^{\varepsilon}_{\bar t}$ such that

$$\inf_{\lambda_{[\bar t,T-1]}\in\Lambda^u_{[\bar t,T-1]}} \ \max_{\vec w_{[\bar t,T-1]}\in W^{n(T-\bar t)}} \ \Big[I_{\bar t}(q^{\varepsilon}_{\bar t}) + \mathbf{E}_{q^{\varepsilon}_{\bar t}}\Big\{ \mathbf{E}\left[\,\mathcal{E}(X_T)\,|\,X_{\bar t}=X\right]\Big\}\Big] \geq Z_{\bar t} + \varepsilon$$

where $X_\cdot$ is propagated with $\lambda_{[\bar t,T-1]}$ and $\vec w_\cdot$. Consequently, for any $\lambda_{[\bar t,T-1]}\in\Lambda_{[\bar t,T-1]}$ such that $\lambda_{\bar t}[y_\cdot] = u$,

$$\max_{\vec w_{[\bar t,T-1]}\in W^{n(T-\bar t)}} \ \Big[I_{\bar t}(q^{\varepsilon}_{\bar t}) + \mathbf{E}_{q^{\varepsilon}_{\bar t}}\Big\{ \mathbf{E}\left[\,\mathcal{E}(X_T)\,|\,X_{\bar t}=X\right]\Big\}\Big] \geq Z_{\bar t} + \varepsilon$$

where $X_\cdot$ is propagated with $\lambda_{[\bar t,T-1]}$ and $\vec w_\cdot$. This implies that there exists an optimal $\vec w^{\varepsilon}_\cdot$ such that

$$I_{\bar t}(q^{\varepsilon}_{\bar t}) + \mathbf{E}_{q^{\varepsilon}_{\bar t}}\Big\{ \mathbf{E}\left[\,\mathcal{E}(X^{\varepsilon}_T)\,|\,X^{\varepsilon}_{\bar t}=X\right]\Big\} \geq Z_{\bar t} + \varepsilon$$

where $X^{\varepsilon}_\cdot$ is propagated with $\lambda_{[\bar t,T-1]}$ and $\vec w^{\varepsilon}_\cdot$. We summarize this in the following Theorem.

Theorem 5.5 *Let $\bar t\in\{0,T-1\}$. Let $\mathcal{I}_0$, $u_{[0,\bar t-1]}$ and $y_{[0,\bar t-1]}$ be given. Let the player 1 control choice, $u^m_{\bar t}$, given by (104) (also given in (76)) be a strict minimizer. Suppose Saddle Point Assumption ($A5.1$) holds. Then, given any player 1 strategy, $\lambda_{[\bar t,T-1]}$ such that $\lambda_{\bar t}[y_\cdot]\neq u^m_{\bar t}$, there exists $\varepsilon>0$, $q^{\varepsilon}_{\bar t}$ and $\vec w^{\varepsilon}_{[\bar t,T-1]}$ such that*

$$\sup_{q\in Q_{\bar t}}\{\mathcal{I}_{\bar t}(q) + L_{\bar t}(q,u^m_{\bar t})\} = Z_{\bar t} \leq \mathcal{I}_{\bar t}(q^{\varepsilon}_{\bar t}) + \mathbf{E}_{q^{\varepsilon}_{\bar t}}\Big\{\mathbf{E}[\mathcal{E}(X^{\varepsilon}_T)\,|\,X^{\varepsilon}_{\bar t}=X]\Big\} - \varepsilon \qquad (107)$$

where X^{ε} denotes the process propagated with control strategies $\lambda_{[\bar t,T-1]}$ and $\vec w^{\varepsilon}_{[\bar t,T-1]}$.

Remark 5.6 Theorem 5.5 also serves as a basis for referring to $\mathcal{I}_t$ as an information state – at least in the case where Assumption ($A5.1$) holds.

The following proof was delayed for reasons of presentation.

Proof. (proof of Lemma 5.4) Let $t\in\{0,1,\ldots T-1\}$. Note that by definition,

$$\widehat{L}_t(q,u) = \inf_{\overline{\lambda}_{[t+1,T-1]}\in\overline{\Lambda}_{[t+1,T-1]}} \ \mathbf{E}_q\Big\{M_t(X,q,(u,\overline{\lambda}_{[t+1,T-1]}))\Big\}$$

where we note that $(u,\overline{\lambda}_{[t+1,T-1]})\in\overline{\Lambda}_{[t,T-1]}$, and by the definition of M_t (43)

$$= \inf_{\overline{\lambda}_{[t+1,T-1]}\in\overline{\Lambda}_{[t+1,T-1]}} \ \mathbf{E}_q\Big\{\max_{\overline{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \ \mathbf{E}[V_T(X_T,q_T)\,|\,X_t=x]\Big\}$$

where q_T and X_T are obtained by propagation according to (1), (6), (8) with $q_t=q$ and $\overline{\lambda}_t=u$. By Remark 4.2 and Lemma 4.3,

$$= \inf_{\overline{\lambda}_{[t+1,T-1]}\in\overline{\Lambda}_{[t+1,T-1]}} \ \max_{\overline{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}_q\Big\{\mathbf{E}[V_T(X_T,q_T)\,|\,X_t=x]\Big\}. \tag{108}$$

On the other hand, by definitions (53) and (50) and Theorem 4.6,

$$L_t(q,u) = \mathbf{E}_q \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u,\vec{w})V_{t+1}(j,q'(q,u,\vec{w}))\Big] \tag{109}$$

where as before, $q'(q,u,\vec{w}) = \widetilde{P}^T(u,\vec{w})q$. Since the case of $t+1=T$ is rather straightforward, we assume $t+1<T$. Using (46), (109) becomes

$$= \mathbf{E}_q \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u,\vec{w})M_{t+1}(j,q'(q,u,\vec{w}),\overline{\lambda}^0_{[t+1,T-1]})\Big]$$

and by the definition of M_{t+1} (43), this is

$$= \mathbf{E}_q \max_{\vec{w}\in W^n}\Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u,\vec{w}) \max_{\overline{\theta}_{[t+1,T-1]}\in\overline{\Theta}_{[t+1,T-1]}} \mathbf{E}[V_T(X_T,q_T)\,|\,X_{t+1}=j]\Big] \tag{110}$$

where X_T, q_T are obtained by propagation according to (1), (6), (8) with $q_t = q$ and strategies $(u,\overline{\lambda}^0)$ and $\overline{\theta}$. Noting that the supremum over $\vec{w}\in W^n$ allows the control to depend on state X_t, as in Remark 4.2 and Lemma 4.3, (110) becomes

$$L_t(q,u) = \mathbf{E}_q \max_{\overline{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \Big[\sum_{j\in\mathcal{X}} \widetilde{P}_{Xj}(u,\overline{\theta}_t[X_\cdot,q_\cdot])\mathbf{E}[V_T(X_T,q_T)\,|\,X_{t+1}=j]\Big]$$

which by the definition of $\widetilde{P}$ in (6)

$$= \mathbf{E}_q\Big\{\max_{\overline{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}[V_T(X_T,q_T)\,|\,X_t=X]\Big\}$$

and again as in Remark 4.2 and Lemma 4.3,

$$\begin{aligned} &= \max_{\overline{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}_q\Big\{\mathbf{E}[V_T(X_T,q_T)\,|\,X_t=X]\Big\} \\ &= \max_{\overline{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}_{q_{t+1}}\Big\{\mathbf{E}[V_T(X_T,q_T)\,|\,X_{t+1}=X]\Big\} \end{aligned} \tag{111}$$

where to avoid any confusion we note that $\mathbf{E}_{q_{t+1}}$ indicates expectation where the distribution of X is q_{t+1}, where $q_{t+1} = \widetilde{P}^T(u,\overline{\theta}_t)q$, and where X_T, q_T are obtained from X_{t+1}, q_{t+1} by propagation with strategies $\overline{\lambda}^0_{[t+1,T-1]}$ and $\overline{\theta}_{[t+1,T-1]}$. By the definition of $\overline{\lambda}^0$ as optimal, (111) becomes

$$L_t(q,u) = \inf_{\overline{\lambda}_{[t+1,T-1]}\in\overline{\Lambda}_{[t+1,T-1]}} \ \max_{\overline{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}_{q_{t+1}}\Big\{\mathbf{E}[V_T(X_T,q_T)\,|\,X_{t+1}=X]\Big\}$$

where the propagation is with strategies $\overline{\lambda}_{[t+1,T-1]}$ and $\overline{\theta}_{[t+1,T-1]}$, and this is

$$= \inf_{\overline{\lambda}_{[t+1,T-1]}\in\overline{\Lambda}_{[t+1,T-1]}} \; \max_{\overline{\theta}_{[t,T-1]}\in\overline{\Theta}_{[t,T-1]}} \mathbf{E}_q\Big\{\mathbf{E}[V_T(X_T,q_T)\,|\,X_t=X]\Big\} \qquad (112)$$

where the propagation from $X_t = X$, $q_t = q$ is with strategies $(u, \overline{\lambda}_{[t+1,T-1]})$ and $\overline{\theta}_{[t,T-1]}$. Comparing (108) and (112) yields the result. □

6 Computational Tractability

Although one can obtain results such as Corollary 5.2 and Theorem 5.5, another motivation for consideration of games of this form is the claim that they can represent useful applications and, at the same time, lead to reasonably tractable algorithms. The largest problem with tractability for imperfect information problems is the propagation of the information state forward in real-time. A secondary problem is of course the computation of the argmin in (76). We briefly discuss computational tractability for two cases: linear ϕ and max–plus delta function ϕ. A key to the tractability is that the costs are only initial and final, and in particular, the cost to the players to affect the observation process is only indirectly felt through the effects those same controls may have on the state process. (For example, in the military application referred to in the introduction, this effect might be the loss of UCAVs whose controlled trajectories not only lead to observations but also to potential loss of the vehicles.)

Consider the case of linear $\widetilde{\mathcal{I}}_0 = \mathcal{I}_0 = \phi$. The propagation of $\widetilde{\mathcal{I}}_\cdot$ is given by (32) with domain propagation (33) (see the proof of Lemma 3.4 for more details). In the case where there is only one choice of control for the player 2, this would simply be a linear mapping of the underlying distribution, and so $\widetilde{\mathcal{I}}_t(\cdot)$ would remain a linear function. Note that the domain remains a (convex) simplex subset of an affine hyperplane with at most $\#\mathcal{X}$ extremal points, but this may not be the initial simplex $Q(\mathcal{X})$. In the more realistic situation where W is not a single point (but recall that it is still assumed finite), Theorem 3.4 shows that $\widetilde{\mathcal{I}}_t$ is a maximum of linear functions over such convex, simplex subsets. Thus, propagation of the information state forward in time is a finite-dimensional process. On the other hand, for more general problems, $q_t \in \mathbf{R}^n$, and so $\mathcal{I}_t$ is an infinite-dimensional object - a function over $\mathbf{R}^n$. In that case, one might use a finite-element approach (or possibly a max-plus approach, c.f. [12]) to propagate $\mathcal{I}_t$ forward in time. If this requires M_d grid points per space dimension, the problem becomes intractable very quickly as the dimension grows; A four-dimensional problem requiring M_d^4 grid points, and with a very reasonable $M_d = 50$, this is more than 6×10^6 grid points. As noted above, the transformed version of $V_t(x,q)$, $\widetilde{V}_t(x,\widetilde{q}) \doteq V_t(x, q(\widetilde{q}))$, remains piecewise constant. Thus $\widetilde{\mathcal{I}}_t(\widetilde{q}) + \widetilde{V}_t(x,\widetilde{q})$ is a discontinuous piecewise linear function. (That is, it consists of a union of linear pieces, and may be discontinuous along the boundaries of the pieces.) Consequently the argmax computation reduces to a comparison among a finite set of maxima of each of the linear pieces. If the set of not-unreasonable controls for player 2 is small, then it appears that this can be propagated in real-time.

The case where ϕ is a max–plus delta function, i.e. $\phi(q) = \delta_{q_c}(q)$ for some $q_c \in Q(\mathcal{X})$, leads to a particularly tractable problem. Recall that this case corresponds to a model where the initial distribution for player 1 state information is not subject to disruption by some initial control of player 2. (More specifically, such a control is not considered within the game.) In this case, $\mathcal{I}_t$ is 0 only at a finite number of points, and is $-\infty$ elsewhere. Thus, $\mathcal{I}_t$ retains this property. The propagation of these points proceeds by (31) for each possible player 2 control. Thus, the information state is easily propagated. Further, since $\mathcal{I}_t$ is not $-\infty$ at only a finite number of points, the max computation in (76) involves only comparison of a finite (although potentially large) number of values of $\max_x V_t(x, q)$ for these select values of q.

7 Acknowledgments

The author would like to thank Prof. J.W. Helton for helpful discussions.

References

[1] M. Akian, "Densities of Idempotent Measures and Large Deviations", Trans. Amer. Math. Soc. **351** (1999), 4515–4543.

[2] M. Adams, W.M. McEneaney et al., "Mixed Initiative Planning and Control under Uncertainty", Proceedings First AIAA UAV Symposium, Portsmouth, VA, May 22-25, (2002), AIAA-2002-3452.

[3] T. Basar and P. Bernhard, *H_∞ –Optimal Control and Related Minimax Design Problems*, Birkhäuser (1995).

[4] F.L. Baccelli, G. Cohen, G.J. Olsder and J.-P. Quadrat, *Synchronization and Linearity*, John Wiley (1992).

[5] T. Basar and G.J. Olsder, *Dynamic Noncooperative Game Theory*, Classics in Applied Mathematics Series, SIAM (1999), Originally pub. Academic Press (1982).

[6] D.P. Bertsekas, D.A. Castañon, M. Curry and D. Logan, "Adaptive Multi-platform Scheduling in a Risky Environment", Advances in Enterprise Control Symp. Proc., (1999), DARPA–ISO, 121–128.

[7] J.B. Cruz, M.A. Simaan, et al., "Modeling and Control of Military Operations Against Adversarial Control", Proc. 39th IEEE CDC, Sydney (2000), 2581–2586.

[8] R.A. Cuninghame-Green, *Minimax Algebra*, Lecture Notes in Economics and Mathematical Systems 166, Springer, 1979.

[9] R. J. Elliott and N. J. Kalton, “The existence of value in differential games”, Memoirs of the Amer. Math. Society, **126** (1972).

[10] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*, Springer (1997).

[11] W.H. Fleming, “Max-Plus Stochastic Processes” (2003), Sub to App. Math. and Optim.

[12] W.H. Fleming and W.M. McEneaney, “A max–plus based algorithm for an HJB equation of nonlinear filtering”, SIAM J. Control and Optim., 38 (2000), 683–710.

[13] W. H. Fleming and E. Pardoux, “Optimal control of partially–observed diffusions”, SIAM J. Control and Optim., **20** (1982), 261–285.

[14] W.H. Fleming and P.E. Souganidis, “On the existence of value functions of two–player, zero–sum stochastic differential games”, Indiana Univ. Math. Journal, **38** (1989) 293–314.

[15] D. Ghose, M. Krichman, J.L. Speyer and J.S. Shamma, “Game Theoretic Campaign Modeling and Analysis”, Proc. 39th IEEE CDC, Sydney (2000), 2556–2561.

[16] W.D. Hall and M.B. Adams, “Closed-loop, Hierarchical Control of Military Air Operations”, Advances in Enterprise Control Symposium Proc., (1999), DARPA–ISO, 245–250.

[17] S.A. Heise and H.S. Morse, “The DARPA JFACC Program: Modeling and Control of Military Operations”, Proc. 39th IEEE CDC, Sydney (2000), 2551–2555.

[18] J.W. Helton and M.R. James, *Extending H_∞ Control to Nonlinear Systems*, SIAM 1999.

[19] M.R. James and J.S. Baras, “Robust H_∞ output feedback control for nonliner systems”, IEEE Trans. Auto. Control, **40** (1995), 1007–1017.

[20] M.R. James and J.S. Baras, “Partially observed differential games, infinite dimensional HJI equations, and nonlinear H_∞ control”, SIAM J. Control and Optim., **34** (1996), 1342–1364.

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

[22] W.M. McEneaney, “A Class of Reasonably Tractable Partially Observed Discrete Stochastic Games”, Proc. 41st IEEE CDC, Las Vegas (2002).

[23] W.M. McEneaney and B.G. Fitzpatrick, “Control for UAV Operations under Imperfect Information”, Proceedings First AIAA UAV Symposium, Portsmouth, VA, May 22-25, (2002), AIAA-2002-3418.

[24] W.M. McEneaney, B.G. Fitzpatrick and I.G. Lauko, “Stochastic Game Approach to Air Operations”, Submitted to IEEE Trans. on Aerospace and Electronic Sys.

[25] W.M. McEneaney and C.D. Charalambous, “Large Deviations Theory, Induced Log-Plus and Max-Plus Measures and their Applications”, Proc. MTNS (Math. Theory of Networks and Systems) 2000, Perpignan.

[26] W.M. McEneaney and K. Ito, “Stochastic Games and Inverse Lyapunov Methods in Air Operations”, Proc. 39th IEEE CDC, Sydney (2000), 2568–2573.