2021 60th IEEE Conference on Decision and Control (CDC)
December 13-15, 2021. Austin, Texas

A Distributionally Robust LQR for Systems with Multiple Uncertain Players

Ioannis Tzortzis, Charalambos D. Charalambous, and Christoforos N. Hadjicostis

2021 60th IEEE Conference on Decision and Control (CDC) | 978-1-6654-3659-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/CDC45484.2021.9682976

***Abstract*— In this paper, we study the robust linear quadratic regulator (LQR) problem for a class of discrete-time dynamical systems composed of several uncertain players with unknown or ambiguous distribution information. A distinctive feature of the assumed model is that each player is prescribed by a nominal probability distribution and categorized according to an uncertainty level of confidence. Our approach is based on minimax optimization. By following a dynamic programming approach a closed-form expression of the robust control policy is derived. The effect of ambiguity on the performance of the LQR is studied via a sequential hierarchical game with one leader and several followers. The equilibrium solution is obtained through a maximizing, time-varying probability distribution characterizing each player's optimal policy. The behavior of the proposed method is demonstrated through an application to a drop-shipping retail fulfillment model.**

I. Introduction

The linear quadratic regulator (LQR) is an important class of optimal control problems that provides state feedback control policies to enable dynamical systems stability and optimal performance. Due to its importance, the LQR finds applications in many areas of engineering such as communication, robotics, network control systems, power systems, and others. In the presence of uncertainty, however, the optimality of the LQR is not guaranteed [1]–[3]. Moreover, in situations of multiple sources of uncertainty with ambiguous distribution information the design of robust LQR control policies is highly desirable, but it requires a proper understanding of (i) the effect of these sources of uncertainty on the behavior of the system's response, and (ii) their impact on the performance of the linear quadratic regulator. This is one of the most fundamental and challenging issues in the practical implementation of the robust LQR.

During the last few years, several robustness approaches have been developed in the area of LQR to deal with uncertainty [4]–[7], and ambiguous distribution information [8]–[11]. The purpose of this paper, in contrast to existing literature, is to extend the robust LQR problem studied in our early work [8], to systems with multiple sources of uncertainty of ambiguous distribution information. Considering the LQR problem for such systems is of great importance given that in many practical applications the most commonly used approach to deal with uncertainty is the Monte Carlo method [12], applied under the assumption of known distribution information. That is, to characterize and determine the probability distribution of the system's total uncertainty from a set of individual uncertain quantities, each characterized by a known probability distribution. In specific applications, however, estimating the system's response probability distribution might be insufficient, especially, if one does not have the same amount of confidence in all the individual uncertain quantities. In other words, it might be difficult, or even impossible, to define a sensible probability distribution for the system's total uncertainty if we are ambiguous about the probability distribution of the individual contributors of uncertainty. As an illustration, for the class of systems we consider in this work, let us consider the following motivating example.

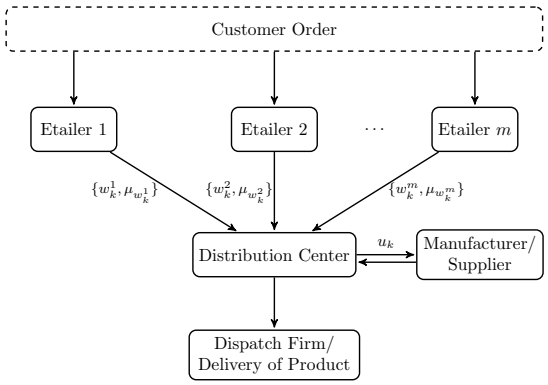


Fig. 1: Drop–shipping model.

An illustration: Consider a drop–shipping retail fulfillment model as shown in Fig. 1, where m etailers (online retailers) sell a homogeneous product to consumers. Etailers do not hold any inventory of the product; instead, they pass the customer's order to a distribution center responsible for the fulfillment and delivery of the product. The distribution center serves as a central location (which may also hold inventory but with limited capacity), and places orders u_k at the beginning of the kth period to a manufacturer/supplier based on a stochastic demand w_k^i of which the probability distribution $\mu_{w_k^i}$ is assumed to be given. The demand provided by each of the $i = 1, 2, \ldots, m$ etailers is not known in advance, and its value cannot be controlled by the distribution center. In particular, the value of the demand w_k^i will be known on the next time period $k+1$. Hence, the objective of the distribution center is to maximize its profit (or equivalently, minimize its cost) with minimum inventory investment, without impacting customer satisfaction levels due to shortage of inventory.

The above illustration is an example of an optimal decision problem with uncertainty, and an equal amount of confidence

This work was partially funded by the European Regional Development Fund and the Republic of Cyprus through the Cyprus Research and Innovation Foundation (Project: POST-DOC/0916/0139).

The authors are with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus. E-mails: {tzortzis.ioannis,chadcha,chadjic}@ucy.ac.cy

placed between all the etailers. Now, let us suppose that the distribution center has less confidence in some of the etailer's stochastic demand than others. Such a case can occur due to several reasons, for example due to distribution center experience, and/or due to well-established etailers with strong distribution networks versus some other etailers with not so strong distribution networks, and/or due to new-entrant etailers, etc. The question is how to choose the optimal decision policy so that the distribution center balances the desire for high profits (or equivalently, low costs) with the undesirability of scenarios with high uncertainty.

In the developed approach, a key issue in studying the effects of ambiguous sources of uncertainty is to consider a sequential hierarchical game with multiple uncertain players trying to choose their optimal policy so that their common expected pay-off is maximized. The distinctive feature of the assumed model is that each player is prescribed by a "nominal" probability distribution, and categorized (either as a leader or as follower) according to an uncertainty level of confidence defined by the Total Variation (TV) distance metric. Moreover, this sequence of leader-followers is sequentially decided during the different time intervals, and hence, it is not considered fixed at the initial time or kept constant during the whole interval of optimization. For this sequential game, a maximizing, time-varying, probability distribution for each player is provided which when applied, leads to an optimal equilibrium solution. On the other hand, by following a Dynamic Programming (DP) approach, a closed-form expression of the robust LQR control policy is derived. The main feature of the derived control policy is that it preserves its linearity, similar to the classical case, and its evaluation is performed based on the multiple players maximizing probability distribution.

This paper is organized as follows. In Section II the distributionally robust LQR problem is formulated, and the TV distance ambiguity class is introduced. In Section III the solution of the optimization problem is derived, along with a maximizing, time-varying probability distribution, and an LQR optimal control policy. In Section IV the drop-shipping retail fulfillment model is considered as an illustration of the proposed methodology. Finally, concluding remarks are given in Section V.

II. Problem Formulation

A. Model description

Consider a discrete-time control system of the form

$$x_{k+1} = A_k x_k + B_k u_k + \sum_{i=1}^{m} C_k^i w_k^i, \quad x_0 = x \qquad (1)$$

where $x_k \in \mathcal{X}_k \triangleq \mathbb{R}^{n_x}$, $u_k \in \mathcal{U}_k \triangleq \mathbb{R}^{n_u}$, are the state and control processes, respectively. The matrices A_k, B_k, and C_k^i are matrices with compatible dimensions. Let us denote by $\mathbb{N}_n \triangleq \{0, 1, 2, \ldots, n\}$ the set of the first n natural numbers. For each $i = 1, 2, \ldots, m$, the sequence $\{w_k^i, k \in \mathbb{N}_{N-1}\}$ is an independent sequence of random vectors such that for each k, $w_k^i \in \mathcal{W}_k^i \triangleq \mathbb{R}^{n_{w_i}}$, with unknown probability distribution $\{\nu_{w_k^i}(dw), k \in \mathbb{N}_{N-1}\}$. Here, it is assumed that the basic random vectors $\{x_0, w_0^1, w_1^1, \ldots, w_{N-1}^1, \ldots, w_0^m, w_1^m, \ldots, w_{N-1}^m\}$, are all mutually independent. Next, the information structure available to the controller is specified.

B. Control policies

For the construction of u_k at any time $k \in \mathbb{N}_{N-1}$, it is assumed that the controller has complete state information about x_k. The set of control policies, G, is the set of measurable Markov feedback control policies $g_k : \mathcal{X}_k \longmapsto \mathcal{U}_k$. Then, associated with the open-loop system (1), the closed-loop system is defined by

$$x_{k+1}^g = A_k x_k^g + B_k g_k(x_k^g) + \sum_{i=1}^{m} C_k^i w_k^i, \quad x_0^g = x \qquad (2)$$

with the control policy $g_k \in G$ and the associated control process related by $u_k^g = g_k(x_k^g)$. The notation x^g, u^g is used to emphasize the dependence of both the state and the control process on the control policy $g \in G$. Note that, the super index g will be omitted for the rest of the paper, for notational simplicity. The ambiguity class is defined next.

C. Ambiguity class

The ambiguity class will be formulated in terms of the TV distance metric.

Definition 2.1: Let $(\mathcal{W}, \mathcal{B}(\mathcal{W}))$ denote an arbitrary measurable space, and $\mathcal{M}_1(\mathcal{W})$ the set of probability distributions on $\mathcal{W}$. The TV distance between two probability distributions is a function $||\cdot||_{TV} : \mathcal{M}_1(\mathcal{W}) \times \mathcal{M}_1(\mathcal{W}) \longmapsto [0, \infty)$, defined by

$$||\alpha - \beta||_{TV} \triangleq \sup_{P \in \mathcal{P}(\mathcal{W})} \sum_{F_i \in P} |\alpha(F_i) - \beta(F_i)| \qquad (3)$$

where $\alpha, \beta \in \mathcal{M}_1(\mathcal{W})$ and $\mathcal{P}(\mathcal{W})$ denotes the collection of all finite partitions $P = \{F_1, F_2, \ldots, F_{n_p}\}$ on $\mathcal{W}$.

Given a collection of nominal probability distributions $\{\mu_{w_k^i}(dw), k \in \mathbb{N}_{N-1}\}$ the corresponding collection of true probability distributions $\{\nu_{w_k^i}(dw), k \in \mathbb{N}_{N-1}\}$, $i = 1, 2, \ldots, m$, is modeled by a ball with respect to the TV distance metric, centered at the nominal probability distribution with radius R_k^i, defined by

$$\mathbb{B}_{R_k^i}(\mu_{w_k^i}) \triangleq \big\{\nu_{w_k^i}(\cdot) \in \mathcal{M}_1(\mathcal{W}_k^i) : ||\nu_{w_k^i}(\cdot) - \mu_{w_k^i}(\cdot)||_{TV} \leq R_k^i\big\}, \; R_k^i \in [0, 2]. \qquad (4)$$

Here, the value of the TV distance parameter R_k^i, for each $i = 1, 2, \ldots, m$, is pre-specified by the decision maker, and can be interpreted as the level of confidence on player's i nominal probability distribution. Next, the optimal stochastic control problem is introduced.

D. Optimal stochastic control problem

Define the N-stage expected cost by

$$J_N(g, \nu) \triangleq \mathbb{E}_\nu^{g,x}\big[\sum_{k=0}^{N-1} (x_k^T Q_k x_k + u_k^T R_k u_k) + x_N^T Q_N x_N\big]$$

where $\mathbb{E}_{\nu}^{g,x}[\cdot]$ indicates the dependence of the expectation operation on the policy $g \triangleq \{g_k, k \in \mathbb{N}_{N-1}\}$ for a given initial state $x_0 = x$, and induced by the probability distribution $\nu \triangleq \{\nu_{w_k^i}(\cdot), k \in \mathbb{N}_{N-1}, i = 1, 2, \ldots, m\}$. We assume that the stage cost matrices $Q_k \succeq 0$, $k \in \mathbb{N}_N$, and the input cost matrices $R_k \succ 0$, $k \in \mathbb{N}_{N-1}$, are known.

Minimax stochastic control problem: Find an optimal control policy $g^* \in G$, and a maximizing probability distribution ν^* within the TV distance ambiguity set, that causes the closed-loop system (2) to maintain the state vector close to the origin, by solving the optimization problem

$$J^* = J_N(g^*, \nu^*) \triangleq \min_{g \in G} \max_{\substack{\nu_{w_k^i} \in \mathbb{B}_{R_k^i} \\ k=0,1,\ldots,N-1, i=1,\ldots,m}} J_N(g, \nu). \quad (5)$$

A distinctive feature of (5) is that, by appropriately adjusting the TV distance parameter $R_k^i \in [0, 2]$ we can control (i) the size of the ambiguity set (4), and (ii) the degree of conservatism of the optimization problem. In particular, for $R_k^i = 0$, the distributionally robust LQR problem reduces to the classical robust LQR problem without ambiguity. On the other hand, by letting R_k^i to increase, then highly ambiguous scenarios are considered. In the next section the solution of the minimax stochastic control problem is derived.

III. Solution of the Minimax Stochastic Control Problem

A. Dynamic programming

For $(k, x) \in \{0, 1, \ldots, N\} \times \mathcal{X}$ let $V_k(x)$ denote the minimal cost-to-go or value function on the time horizon $\{k, k+1, \ldots, N\}$, given an optimal policy $g_t^*(\cdot)$, $t = 0, 1, \ldots, k-1$, and optimal probability distribution $\nu_{w_t^i}^* \in \mathbb{B}_{R_t^i}(\mu_{w_t^i})$, $i = 1, 2, \ldots, m$, $t = 0, 1, \ldots, k-1$, defined by

$$V_k(x) = \min_{u_k \in \mathcal{U}_k(x)} \max_{\substack{\nu_{w_t^i} \in \mathbb{B}_{R_t^i} \\ t=k,k+1,\ldots,N-1, i=1,\ldots,m}} \mathbb{E}_{\nu}^{g,x}\Big[\sum_{t=k}^{N-1} \big(x_t^T Q_t x_t + u_t^T R_t u_t\big) + x_N^T Q_N x_N\Big] \quad (6)$$

where $\mathbb{E}_{\nu}^{g,x}[\cdot]$ denotes conditional expectation given that $x_k^g = x$ for fixed x. The DP algorithm gives [1]

$$V_N(x_N) = x_N^T Q_N x_N \quad (7a)$$

$$V_k(x) = \min_{u_k \in \mathcal{U}_k(x)} \max_{\substack{\nu_{w_k^i} \in \mathbb{B}_{R_k^i} \\ i=1,\ldots,m}} \mathbb{E}_{\nu}^{g,x}\big[x^T Q_k x + u_k^T R_k u_k + V_{k+1}(x_{k+1})\big], \quad x \in \mathcal{X}. \quad (7b)$$

Notice that (7) relates the value function $V_k(\cdot)$ and $V_{k+1}(\cdot)$ for all $k = N-1, N-2, \ldots, 0$, and generates $V_N(\cdot), V_{N-1}(\cdot), \ldots, V_0(\cdot)$ by backward recursion. Next, we will show by backward induction that the solution is of the following form

$$V_k(x) = x^T P_k x + x^T F_k + r_k, \quad k = 0, 1, \ldots, N \quad (8)$$

for $F_k \in \mathbb{R}^{n_x}$, $r_k \in \mathbb{R}$, and some matrices $P_k \succeq 0$.

The induction hypothesis is true for $k = N$, with $P_N = Q_N$, $F_N = 0$ and $r_N = 0$. Then $P_N = P_N^T \succeq 0$ and $V_N(x) = x^T P_N x + x^T F_N + r_N$. Let us assume that for $t = k+1, k+2, \ldots, N$, $P_t = P_t^T \succeq 0$, $F_t = F_t^T \succeq 0$ and $V_t(x) = x^T P_t x + x^T F_t + r_t$. It will be shown that then $P_k = P_k^T \succeq 0$, $F_k = F_k^T \succeq 0$ and $V_k(x) = x^T P_k x + x^T F_k + r_k$. Toward this end, we write (7b) as follows

$$V_k(x) = \min_{u_k \in \mathcal{U}_k(x)} \Big\{x^T Q_k x + u_k^T R_k u_k + \max_{\substack{\nu_{w_k^i} \in \mathbb{B}_{R_k^i} \\ i=1,\ldots,m}} \mathbb{E}_{\nu}^{g,x}\big[\ell_k(x_k, u_k, w_k^1, \ldots, w_k^m)\big]\Big\}, \quad x \in \mathcal{X} \quad (9)$$

where the functional $\ell_k(\cdot)$ is defined by

$$\ell_k(x_k, u_k, w_k^1, \ldots, w_k^m) \triangleq V_{k+1}(A_k x_k + B_k u_k + \sum_{i=1}^m C_k^i w_k^i).$$

Next, we address the maximization in (9).

B. Maximization of a linear functional

To solve the inner optimization in (9), that is

$$\max_{\substack{\nu_{w_k^i} \in \mathbb{B}_{R_k^i} \\ i=1,\ldots,m}} \mathbb{E}_{\nu}^{g,x}\big[\ell_k(x_k, u_k, w_k^1, \ldots, w_k^m)\big] \quad (10)$$

we consider a sequential game with m players where each player $i = 1, 2, \ldots, m$ is identified by (i) its nominal probability distribution $\mu_{w_k^i}$, and (ii) its TV distance parameter R_k^i. In such a game, one player acts as the leader (L) and the remaining $m-1$ players act as the followers (F). This classification of the m players is described by a classification function ϕ which is sequentially decided during the different time intervals of optimization.

Definition 3.1: Let $\mathcal{M} = \{1, 2, \ldots, m\}$ and $\mathcal{N} = \{L, F_1, F_2, \ldots, F_{m-1}\}$ be two finite sets with cardinality $|\mathcal{M}| = |\mathcal{N}|$. A classification function $\phi : \mathcal{M} \longmapsto \mathcal{N}$ is a bijective function from $\mathcal{M}$ to $\mathcal{N}$.

Problem Statement: The m-player problem is to find an optimal classification function $\phi_k : \mathcal{M} \longmapsto \mathcal{N}$, for each k, and a maximizing transition probability distribution $\nu_{w_k^t}$ for all players $t \in \mathcal{N}$, to solve

$$\max_{\substack{\phi_k : \mathcal{M} \longmapsto \mathcal{N} \\ \nu_{w_k^t} \in \mathbb{B}_{R_k^t}, \forall t \in \mathcal{N}}} \mathbb{E}_{\nu}^{g,x}\big[\ell_k(x_k, u_k, w_k^L, w_k^{F_1}, \ldots, w_k^{F_{m-1}})\big]. \quad (11)$$

The main difficulty in solving (11) lies in finding the optimal classification of the m players (in total, there are $m!$ different classifications to choose from) for each time k. For any given (fixed) classification $\phi_k : \mathcal{M} \longmapsto \mathcal{N}$, the hierarchical model described above can be defined as an m-stage game model. In particular, each player $t \in \mathcal{N}$ solves the optimization problem to find its maximizing probability distribution for fixed probability distribution (within the TV distance ambiguity set) of its predecessors, and given the optimal probability distribution of its successors, that is

$$\nu_{w_k^t}^{*,\phi} = \arg\max_{\nu_{w_k^t}^{\phi} \in \mathbb{B}_{R_k^t}} \mathbb{E}_{\nu^{\phi}}^{g,x}\big[\ell_k(x_k, u_k, w_k^L, \ldots, w_k^t, \ldots, w_k^{F_{m-1}})\big]$$

$$\text{where } t \in \mathcal{N},\ \nu_{w_k^L}^{\phi} = \text{fixed}, \ldots, \nu_{w_k^{F_{t-1}}}^{\phi} = \text{fixed},$$

$$\text{and } \nu_{w_k^{F_{t+1}}}^{*,\phi}, \ldots, \nu_{w_k^{F_{m-1}}}^{*,\phi}. \quad (12)$$

Then, the solution of this sequential game for all players $t \in \mathcal{N}$ forms an equilibrium distribution $(\nu^{*,\phi}_{w_k^L}, \ldots, \nu^{*,\phi}_{w_k^{F_{m-1}}})$. The super index ϕ is used to emphasize that the solution is obtained for a fixed classification function. In what follows, the super index ϕ will be dropped from our notation. Next, an example is considered to clarify the procedure.

Example 3.2: To exemplify the basic steps of the above procedure let us consider a 2-player game where player 1 is identified by $(\mu^1_{w_k}, R^1_k)$ and acts as the leader (i.e., $\phi_k(1) = L$), and player 2 is identified by $(\mu^2_{w_k}, R^2_k)$ and acts as the follower (i.e., $\phi_k(2) = F$). For a fixed probability distribution of the leader $\nu_{w_k^L} \in \mathbb{B}_{R_k^L}(\mu_{w_k^L})$, the follower tries to find its maximizing probability distribution which solves

$$\max_{\nu_{w_k^F} \in \mathbb{B}_{R_k^F}} \mathbb{E}^{g,x}_{\nu_{w_k^L}=\text{fixed}, \nu_{w_k^F}} \big[\ell_k(x_k, u_k, w_k^L, w_k^F)\big]$$
$$\triangleq \max_{\nu_{w_k^F} \in \mathbb{B}_{R_k^F}} \sum_{w_k^F \in \mathcal{W}_k^F} \Big(\sum_{w_k^L \in \mathcal{W}_k^L} \ell_k(x_k, u_k, w_k^L, w_k^F) \nu_{w_k^L} \Big) \nu_{w_k^F}.$$

Note that, the maximizing probability distribution $\nu^*_{w_k^F} \in \mathbb{B}_{R_k^F}(\mu_{w_k^F})$ of the follower is obtained subject to the leader's choice, and is such that

$$\mathbb{E}^{g,x}_{\nu_{w_k^L}, \nu^*_{w_k^F}} \big[\ell_k(x_k, u_k, w_k^L, w_k^F)\big]$$
$$\geq \mathbb{E}^{g,x}_{\nu_{w_k^L}, \nu_{w_k^F}} \big[\ell_k(x_k, u_k, w_k^L, w_k^F)\big], \ \forall \nu_{w_k^F} \in \mathbb{B}_{R_k^F}(\mu_{w_k^F}).$$

Next, the leader tries to find its maximizing probability distribution by knowing that in equilibrium the follower will choose $\nu^*_{w_k^F}$ as above, that is

$$\max_{\nu_{w_k^L} \in \mathbb{B}_{R_k^L}} \mathbb{E}^{g,x}_{\nu_{w_k^L}, \nu^*_{w_k^F}} \big[\ell_k(x_k, u_k, w_k^L, w_k^F)\big]$$
$$\triangleq \max_{\nu_{w_k^L} \in \mathbb{B}_{R_k^L}} \sum_{w_k^L \in \mathcal{W}_k^L} \Big(\sum_{w_k^F \in \mathcal{W}_k^F} \ell_k(x_k, u_k, w_k^L, w_k^F) \nu^*_{w_k^F} \Big) \nu_{w_k^L}.$$

For all equilibria $(\nu^*_{w_k^L}, \nu^*_{w_k^F})$, it holds that

$$\mathbb{E}^{g,x}_{\nu^*_{w_k^L}, \nu^*_{w_k^F}} \big[\ell_k(x_k, u_k, w_k^L, w_k^F)\big]$$
$$\geq \mathbb{E}^{g,x}_{\nu_{w_k^L}, \nu^*_{w_k^F}} \big[\ell_k(x_k, u_k, w_k^L, w_k^F)\big], \ \forall \nu_{w_k^L} \in \mathbb{B}_{R_k^L}(\mu_{w_k^L}).$$

Before we proceed with the solution of the m-player problem, we first introduce some definitions. We define the linear functional for each player $t \in \mathcal{N}$ and fixed ϕ_k, by

$$L_k(x_k, u_k, w_k^t) \triangleq \sum_{w_k^{F_{m-1}}} \cdots \sum_{w_k^{F_{t+1}}} \sum_{w_k^{F_{t-1}}} \cdots \sum_{w_k^L} \tag{13}$$
$$\ell_k(x_k, u_k, w_k^L, \ldots, w_k^{F_{m-1}}) \nu_{w_k^L} \ldots \nu_{w_k^{F_{t-1}}} \nu^*_{w_k^{F_{t+1}}} \ldots \nu^*_{w_k^{F_{m-1}}}$$

and the sets of elements which achieve the maximum and minimum values of (13) with respect to $w_k^t \in \mathcal{W}_k^t$, $t \in \mathcal{N}$, by

$$\Sigma^0(k,t) \triangleq \big\{ w_k^t \in \mathcal{W}_k^t : L_k(x_k, u_k, w_k^t) = \max_{w_k^t \in \mathcal{W}_k^t} L_k(x_k, u_k, w_k^t) \big\}$$
$$\Sigma_0(k,t) \triangleq \{ w_k^t \in \mathcal{W}_k^t : L_k(x_k, u_k, w_k^t) = \min_{w_k^t \in \mathcal{W}_k^t} L_k(x_k, u_k, w_k^t) \}.$$

For all remaining elements for which $\Sigma^0(k,t) \cup \Sigma_0(k,t) \subset \mathcal{W}_k^t$, and for $1 \leq r \leq |\mathcal{W}_k^t \setminus \{\Sigma^0(k,t) \cup \Sigma_0(k,t)\}|$, we define recursively the set of elements $\Sigma_j(k,t)$, $j \in \{1,2,\ldots,r\}$, for which (13) achieves its $(j+1)$st smallest value (till all the elements of $\mathcal{W}_k^t$ are exhausted), that is

$$\Sigma_j(k,t) \triangleq \big\{ w_k^t \in \mathcal{W}_k^t : L_k(x_k, u_k, w_k^t) = \min_{a_k^t \in \mathcal{W}_k^t} \big\{ L_k(x_k, u_k, a_k^t) : a_k^t \in \mathcal{W}_k^t \setminus \Sigma^0(k,t) \cup (\cup_{i-1}^j \Sigma_{i-1}(k,t)) \big\} \big\}, \ j = 1, 2, \ldots, r.$$

The following theorem provides the solution of the m-player problem for any given fixed classification function ϕ. For convenience, we denote by $(x)^+ \triangleq \max\{0, x\}$.

Theorem 3.3: Let ϕ_k be a classification function defined on $\mathcal{M}$ with range $\mathcal{N}$. Consider (12) with the constraint that ϕ_k is fixed. Then, the maximizing, time-varying, probability distribution for each player $t \in \mathcal{N}$ is given by

$$\nu^*_{w_k^t}(\Sigma^0(k,t)) = \mu_{w_k^t}(\Sigma^0(k,t)) + \frac{\alpha_k^t}{2}, \tag{14a}$$
$$\nu^*_{w_k^t}(\Sigma_0(k,t)) = \Big(\mu_{w_k^t}(\Sigma_0(k,t)) - \frac{\alpha_k^t}{2} \Big)^+, \tag{14b}$$
$$\nu^*_{w_k^t}(\Sigma_j(k,t)) = \Big(\mu_{w_k^t}(\Sigma_j(k,t)) - \Big(\frac{\alpha_k^t}{2} - \sum_{z=1}^{j} \sum_{i \in \Sigma_{z-1}(k,t)} \mu_{w_k^t}(\Sigma_i(k,t)) \Big)^+ \Big)^+, \ j = 1, \ldots, r, \tag{14c}$$
$$\alpha_k^t = \min \Big(R_k^t, 2\big(1 - \mu_{w_k^t}(\Sigma^0(k,t))\big) \Big). \tag{14d}$$

Proof: The proof is similar to that of the problem addressed in [13], [14], and hence, it is omitted. ■

By Theorem 3.3, (9) becomes

$$V_k(x) = \min_{u_k \in \mathcal{U}_k(x)} \big\{ x^T Q_k x + u_k^T R_k u_k + \mathbb{E}^{g,x}_{\nu^*} \big[\ell_k(x_k, u_k, w_k^1, \ldots, w_k^m) \big] \big\}, \quad x \in \mathcal{X}. \tag{15}$$

The expectation in (15) is performed with respect to the maximizing probability distribution $\nu^*_{w_k^{\phi_k^{-1}(t)}}$, for all $t \in \mathcal{N}$, where for $t \in \mathcal{N}$, $\phi_k^{-1}(t)$ denotes the i-th player in $\mathcal{M}$.

C. Optimal control policy and optimal cost

By definition of $\ell_k(\cdot)$, and by induction hypothesis (8), then (15) becomes

$$V_k(x) = \min_{u_k \in \mathcal{U}_k(x)} \Big\{ \begin{bmatrix} x^T & u_k^T \end{bmatrix} \begin{bmatrix} H_{11}(k) & H_{12}(k) \\ H_{12}^T(k) & H_{22}(k) \end{bmatrix} \begin{bmatrix} x \\ u_k \end{bmatrix} + \begin{bmatrix} x^T & u_k^T \end{bmatrix} \begin{bmatrix} A_k^T F_{k+1} \\ B_k^T F_{k+1} \end{bmatrix} + \Big(F_{k+1}^T + 2(A_k x + B_k u_k)^T P_{k+1} \Big) \sum_{i=1}^{m} C_k^i \mathbb{E}^{g,x}_{\nu^*}[w_k^i] + \mathbb{E}^{g,x}_{\nu^*} \Big[\sum_{i=1}^{m} (C_k^i w_k^i)^T P_{k+1} \sum_{i=1}^{m} (C_k^i w_k^i) \Big] + r_{k+1} \Big\} \tag{16}$$

where $H_{11}(k) \triangleq A_k^T P_{k+1} A_k + Q_k$, $H_{12}(k) \triangleq A_k^T P_{k+1} B_k$, and $H_{22}(k) \triangleq R_k + B_k^T P_{k+1} B_k$. Differentiating (16) with

respect to u_k, and setting the derivative equal to zero, we obtain

$$u_k^* = -L_k x - S_k, \quad \text{for } x_k = x \tag{17}$$

where

$$L_k \triangleq H_{22}^{-1}(k) H_{12}^T(k)$$

$$S_k \triangleq H_{22}^{-1}(k)\Big(B_k^T P_{k+1} \sum_{i=1}^{m} C_k^i \mathbb{E}_{\nu^*}^{g,x}[w_k^i] + \frac{1}{2} B_k^T F_{k+1}\Big).$$

By our assumption on $P_k \succeq 0$ and $R_k \succ 0$, it follows that $H_{22} = H_{22}^T \succ 0$, and the inverse exists. Substituting (17) back into (16), it follows that (8) holds, that is, $V_k(x) = x^T P_k x + x^T F_k + r_k$, with

$$P_k = H_{11}(k) - H_{12}(k) H_{22}^{-1}(k) H_{12}^T(k) \tag{18a}$$

$$F_k = \big(A_k^T - H_{12}(k) H_{22}^{-1}(k) B_k^T\big) \tag{18b}$$
$$\times \big(F_{k+1} + 2P_{k+1} \sum_{i=1}^{m} C_k^i \mathbb{E}_{\nu^*}^{g,x}[w_k^i]\big)$$

$$r_k = r_{k+1} - \frac{1}{4} F_{k+1}^T B_k H_{22}^{-1}(k) B_k^T F_{k+1} \tag{18c}$$
$$+ F_{k+1}^T \big(I - B_k H_{22}^{-1}(k) B_k^T P_{k+1}\big) \sum_{i=1}^{m} C_k^i \mathbb{E}_{\nu^*}^{g,x}[w_k^i]$$
$$- \sum_{i=1}^{m} \mathbb{E}_{\nu^*}^{g,x}[(C_k^i w_k^i)^T] P_{k+1} B_k H_{22}^{-1}(k) B_k^T P_{k+1}$$
$$\times \sum_{i=1}^{m} \mathbb{E}_{\nu^*}^{g,x}[C_k^i w_k^i] + \mathbb{E}_{\nu^*}^{g,x}[\sum_{i=1}^{m} (C_k^i w_k^i)^T P_{k+1} \sum_{i=1}^{m} (C_k^i w_k^i)].$$

The optimal cost for the minimax stochastic control problem (5) is given by $J^* = V_0(x_0) = x_0^T P_0 x_0 + x_0^T F_0 + r_0$. Next, we illustrate an application of the drop-shipping model for inventory control which was mentioned in the introduction.

IV. The Drop-Shipping Problem

Consider the drop-shipping retail fulfillment model discussed in Section I (as shown in Fig. 1). Let us denote the various parameters of interest as follows:

- N: planning horizon;
- m: number of etailers;
- x_k: stock available at the beginning of the kth period;
- u_k: stock ordered at the beginning of the kth period;
- w_k^i: demand of etailer $i = 1, 2, \ldots, m$ during the kth period with given nominal probability distribution $\mu_{w_k^i}$;
- h_k, c_k, p_k: holding, ordering, and shortage cost per unit item, respectively.

The state dynamics are

$$x_{k+1} = \max\big(0, x_k + u_k - \sum_{i=1}^{m} w_k^i\big)$$

with the total sample pay-off over N periods given by

$$\sum_{k=0}^{N-1} \Big(c_k u_k + p_k\big(\min\big(0, x_k + u_k - \sum_{i=1}^{m} w_k^i\big)\big)^2 + h_k\big(\min\big(0, \sum_{i=1}^{m} w_k^i - x_k - u_k\big)\big)^2\Big).$$

The above problem is formulated as a minimax stochastic control problem defined by

$$\min_{u_k \in \mathcal{U}_k(x_k)} \max_{\substack{\nu_{w_k^i} \in \mathbb{B}_{R_k^i} \\ k=0,1,\ldots,N-1, i=1,\ldots,m}} \mathbb{E}_{\nu}^{g,x}\Big[\sum_{k=0}^{N-1} \Big(c_k u_k + p_k\big(\min\big(0, x_k + u_k - \sum_{i=1}^{m} w_k^i\big)\big)^2 + h_k\big(\min\big(0, \sum_{i=1}^{m} w_k^i - x_k - u_k\big)\big)^2\Big)\Big].$$

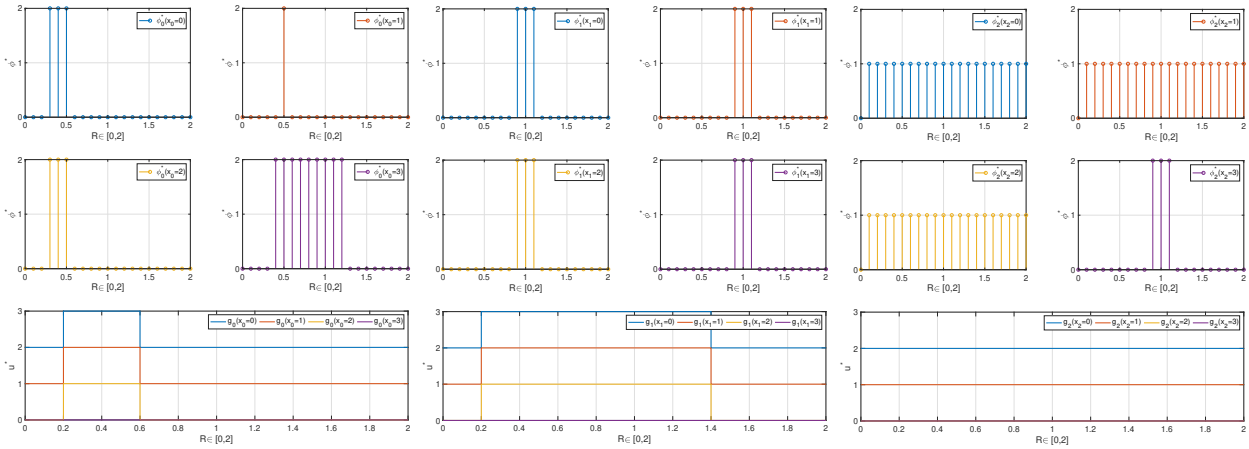
Let $N = 3$, $m = 2$, $\mathcal{U} = \{0, 1, 2, 3\}$, with inventory and demand being non-negative integer variables. Suppose that $\{x_0, w_0^1, w_1^1, w_2^1, w_0^2, w_1^2, w_2^2\}$ are independent random variables. Moreover, let w_k^i be an independent identically distributed sequence with $\mu_{w_k^1} = [0.4\ 0.2\ 0.4]^T$ and $\mu_{w_k^2} = [0.1\ 0.1\ 0.8]^T$. Choose the TV distance parameters $R_k^1 = R \in [0, 2]$, for all $k = 0, 1, 2$, and $R_0^2 = R_0^1$, $R_1^2 = R_1^1/2$, and $R_2^2 = R_2^1/4$. Based on this specific selection of the TV distance parameters: (i) at stage $k = 0$, player 1 is considered to be equally reliable with player 2, and (ii) at stages $k = 1$ and $k = 2$, player 1 is considered to be less reliable compared to player 2. For ease of computation, we further assume that the maximum capacity is $x_k + u_k \leq 3$, and that excess demand is lost. We choose $h_k = c_k = p_k = 1$ for all $k = 0, 1, 2$. The DP is given by: $V_3(x_3) = 0$ and

$$V_k(x_k) = \min_{0 \leq u_k \leq 3 - x_k} \max_{\substack{\nu_{w_k^i} \in \mathbb{B}_{R_k^i} \\ i=1,2}} \mathbb{E}_{\nu}^{g,x}\big[u_k + \big(\min\big(0, x_k + u_k - w_k^1 - w_k^2\big)\big)^2 + \big(\min\big(0, w_k^1 + w_k^2 - x_k - u_k\big)\big)^2 + V_{k+1}(\max(0, x_k + u_k - w_k^1 - w_k^2))\big]$$
$$= \min_{0 \leq u_k \leq 3 - x_k} \max_{\substack{\nu_{w_k^i} \in \mathbb{B}_{R_k^i} \\ i=1,2}} \mathbb{E}_{\nu}^{g,x}\big[\ell_k(x_k, u_k, w_k^1, w_k^2)\big], k = 0, 1, 2.$$

Fig. 2 depicts the optimal solution of the 2-player drop-shipping problem for each possible state $x_k = 0, 1, 2, 3$ and for each stage $k = 0, 1, 2$. Top and middle row graphs depict the optimal classification as a function of the TV distance parameter $R \in [0, 2]$. For the solution of the inner optimization problem (as described in Section III-B), there are two possible classification functions $\phi^{[i]} : \mathcal{M} \longmapsto \mathcal{N}$, $i = 1, 2$, $\mathcal{M} = \{1, 2\}$, $\mathcal{N} = \{L, F\}$, to choose from. Specifically, we consider : (i) $\phi_k^{[1]}(1) = F$ and $\phi_k^{[1]}(2) = L$, where player 2 acts as the leader and player 1 acts as the follower, and (ii) $\phi_k^{[2]}(1) = L$ and $\phi_k^{[2]}(2) = F$, where player 1 acts as the leader and player 2 acts as the follower. Then, the optimal classification is as shown in Fig. 2, where

$$\phi_k^*(x_k) = \begin{cases} 2, & \phi^{[2]} \text{ is the optimal classification} \\ 1, & \phi^{[1]} \text{ is the optimal classification} \\ 0, & \text{both } \phi^{[1]}, \phi^{[2]} \text{ are optimal classifications.} \end{cases}$$

In our simulations the classification function was evaluated at each stage k, over all possible states x_k and controls u_k (that is, we considered a time-varying classification function). In Fig. 2, the optimal classification over the optimal



(a) Stage 0 (b) Stage 1 (c) Stage 2

Fig. 2: Solution of the drop-shipping problem as a function of the TV distance metric. Top and middle row graphs depict the optimal classification of the 2-player game. Bottom row graphs depict the optimal control policy.

ordering policy is illustrated. On the other hand, the graphs at the bottom row of Fig. 2 depict the optimal ordering policy as a function of the TV distance parameter. In particular, for any given $R \in [0, 2]$, the ordering policy can be extracted from Fig. 2. We distinguish the following two cases:

C1) $R_k^1 = R_k^2 > 0$. The optimal ordering policy is obtained by assuming an equal amount of confidence between the two players. This case is illustrated in Fig. 2(a).

C2) $R_k^1 \neq R_k^2 > 0$. The optimal ordering policy is obtained by assuming a different amount of confidence between the two players. This case is illustrated in Fig. 2(b)–(c).

As an illustration of the optimal ordering policy and the optimal cost for cases C1 and C2 above, let us choose $R = 1$, that is, $R_k^1 = 1$, $\forall k = 0, 1, 2$, and $R_0^2 = 1$, $R_1^2 = 0.5$, and $R_2^2 = 0.25$. Then, the optimal ordering policy is given by

$$g_k^*(x_k) = \begin{cases} 2, & \text{if } x_k = 0 \\ 1, & \text{if } x_k = 1 \\ 0, & \text{if } x_k = 2 \\ 0, & \text{if } x_k = 3 \end{cases}, \quad g_1^*(x_1) = \begin{cases} 3, & \text{if } x_1 = 0 \\ 2, & \text{if } x_1 = 1 \\ 1, & \text{if } x_1 = 2 \\ 0, & \text{if } x_1 = 3 \end{cases}$$

for $k = 0, 2$. The optimal cost is equal to $V_0^*(x_0) = [16.45\ 15.45\ 14.45\ 14.29]^T$.

The solution of the minimax optimization problem guarantees the robustness of the optimal ordering policy with respect to ambiguous probability distributions. At the same time, however, higher costs are obtained. For this reason, the designer always needs to balance the desire for low costs with the undesirability of scenarios with high ambiguity.

V. Conclusion

A robust LQR method for discrete-time dynamical systems with multiple sources of uncertainty of ambiguous distribution information is proposed. To study the effects of ambiguity on the performance of the LQR, a sequential hierarchical game with multiple uncertain players is considered, and a maximizing equilibrium policy characterizing each player's optimal policy is provided. In addition, by following a DP approach, a robust feedback control policy is derived. Finally, through an application to a drop-shipping retail fulfillment model, the proposed solution is illustrated.

References

[1] I. Tzortzis, C. D. Charalambous, and T. Charalambous, "Dynamic programming subject to total variation distance ambiguity," *SIAM J. Control Optim.*, vol. 53, no. 4, pp. 2040–2075, Jul. 2015.

[2] I. Petersen, M. James, and P. Dupuis, "Minimax optimal control of stochastic uncertain systems with relative entropy constraints," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 398–412, Mar. 2000.

[3] A. D. Kara and S. Yüksel, "Robustness to incorrect system models in stochastic contol," *SIAM J. Control Optim.*, vol. 58, no. 2, pp. 1144–1182, Apr. 2020.

[4] M. H. Terra, J. P. Cerri, and J. Y. Ishihara, "Optimal robust linear quadratic regulator for systems subject to uncertainties," *IEEE Transactions on Automatic Control*, vol. 59, no. 9, pp. 2586–2591, 2014.

[5] A. Scampicchio and G. Pillonetto, "A convex approach to robust LQR," in *59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 3705–3710.

[6] A. Tsiamis, D. S. Kalogerias, L. F. O. Chamon, A. Ribeiro, and G. J. Pappas, "Risk-constrained linear-quadratic regulators," in *59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 3040–3047.

[7] W. Jongeneel, T. Summers, and P. M. Esfahani, "Robust linear quadratic regulator: Exact tractable reformulation," in *58th IEEE Conference on Decision and Control (CDC)*, 2019, pp. 6742–6747.

[8] I. Tzortzis, C. D. Charalambous, T. Charalambous, C. K. Kourtellaris, and C. N. Hadjicostis, "Robust linear quadratic regulator for uncertain systems," in *55th IEEE Conference on Decision and Control (CDC)*, 2016, pp. 1515–1520.

[9] I. Yang, "Wasserstein distributionally robust stochastic control: A data-driven approach," *IEEE Transactions on Automatic Control*, pp. 1–8, 2020. [Online]. Available: 10.1109/TAC.2020.3030884

[10] K. Kim and I. Yang, "Minimax control of ambiguous linear stochastic systems using the Wasserstein metric," in *59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 1777–1784.

[11] K. Kim and I. Yang, "Distributional robustness in minimax linear quadratic control with Wasserstein distance," *arXiv:2102.12715*, 2021.

[12] J. Hammersley, *Monte Carlo Methods*, ser. Monographs on Statistics and Applied Probability. Springer Netherlands, 2013.

[13] C. D. Charalambous, I. Tzortzis, S. Loyka, and T. Charalambous, "Extremum problems with total variation distance and their applications," *IEEE Transactions on Automatic Control*, vol. 59, no. 9, pp. 2353–2368, Sep. 2014.

[14] I. Tzortzis, C. D. Charalambous, T. Charalambous, C. N. Hadjicostis, and M. Johansson, "Approximation of Markov processes by lower dimensional processes via total variation metrics," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1030–1045, Mar. 2017.