Stochastic Optimal Control of Discrete-Time Systems Subject to Conditional Distribution Uncertainty

Charalambos D. Charalambous, Ioannis Tzortzis and Farzad Rezaei

Abstract— The aim of this paper is to address optimality of control strategies for stochastic discrete time control systems subject to conditional distribution uncertainty. This type of uncertainty is motivated from the fact that the value function involves expectation with respect to the conditional distribution. The issues which will be discussed are the following. 1) Optimal stochastic control systems subject to conditional distribution uncertainty, 2) optimality criteria for stochastic control systems with conditional distribution uncertainty, including principle of optimality and dynamic programming.

I. INTRODUCTION

The objective of this paper is to investigate stochastic optimal control problems in discrete time when the conditional distribution associated with the classical value function is uncertain. By invoking a total variation distance model to describe the uncertainty of conditional distribution, a new dynamic programming equation is derived. The issues which will be discussed are the following.

- Formulation of optimal stochastic control systems subject to conditional distribution uncertainty.
- Optimality criteria for stochastic control systems with conditional distribution uncertainty, when the conditional distribution belongs to a ball centered at a nominal conditional distribution with respect to total variational distance.

The mathematical model used to describe uncertain systems is the total variational distance developed in earlier work [1], [2]. Here the goal is to extend the previous work to uncertain discrete-time systems using conditional distribution uncertainty and provide an illustrative example. The formulation is based on minimax theory, in which nature attempts to maximize the pay-off while the designers objective is to minimize it. The main objective is to characterize the solution of the minimax game via dynamic programming. It turns out that once the maximizing measure is found and substituted into the pay-off the equivalent optimization problem to be solved is a stochastic optimal control problem.

The rest of the paper is organized as follows. In Section II the abstract formulation is introduced while in Section II-A the maximizing measure is characterized. In section III, the abstract setup is applied to a stochastic discrete-time uncertain controlled systems. A dynamic programming equation is derived to characterize the optimality of minimax strategies.

II. ABSTRACT FORMULATION

Let (Σ, d_Σ) denote a complete, separable metric space (a Polish space), and $(\Sigma, \mathscr{B}(\Sigma))$ the corresponding measurable space, in which $\mathscr{B}(\Sigma)$ is the σ-algebra generated by open sets in Σ. Let $\mathscr{M}_1(\Sigma)$ denote space of countably additive probability measures on $(\Sigma, \mathscr{B}(\Sigma))$. Given a known or nominal probability measure $\mu \in \mathscr{M}_1(\Sigma)$ the uncertainty set based on total variation distance is defined by

$$B_R(\mu) \triangleq \left\{ \nu \in \mathscr{M}_1(\Sigma) : ||\nu - \mu|| \leq R \right\}$$

where $R \in [0, \infty)$. The total variation distance[1] on $\mathscr{M}_1(\Sigma) \times \mathscr{M}_1(\Sigma)$ is defined by

$$d(\alpha, \beta) \equiv ||\alpha - \beta|| \triangleq \sup_{P \in \mathscr{P}(\Sigma)} \sum_{F_i \in P} |\alpha(F_i) - \beta(F_i)|$$

where $\alpha, \beta \in \mathscr{M}_1(\Sigma)$ and $\mathscr{P}(\Sigma)$ denotes the collection of all finite partitions of Σ. Note that the distance metric induced by the total variation norm does not require absolute continuity of measures when defining the uncertainty ball, i.e., singular measures are admissible. It covers the case when $\mu_0 \in \mathscr{M}_1(\tilde{\Sigma})$ and $\nu \in \mathscr{M}_1(\Sigma)$, where $\tilde{\Sigma} \subset \Sigma$ and $\mu \in \mathscr{M}_1(\Sigma)$ is the extension of $\mu_0 \in \mathscr{M}_1(\tilde{\Sigma})$ on Σ. Since the elements of $\mathscr{M}_1(\Sigma)$ are probability measures the radius of uncertainty belongs to the restricted set $R \in [0, 2]$.

Let Σ be a locally compact separable metric space and $\mathscr{B}(\Sigma)$ its Borel σ-algebra. Let $\mathscr{X}_0 \triangleq C_0(\Sigma)$ denote the Banach space of continuous functions on Σ that vanish at infinity, $\mathscr{X}_1 \triangleq BC(\Sigma)$ the Banach space of bounded continuous functions on Σ, and $\mathscr{X}_2 \triangleq BM(\Sigma)$ the Banach space of bounded measurable functions on Σ, all equipped with the sup-norm. Clearly, $\mathscr{X}_0 \subset \mathscr{X}_1 \subset \mathscr{X}_2$.

It is known that the topological dual of $\mathscr{X}_0$ [9] denoted by $\mathscr{X}_0^*$ is isometrically isomorphic to $\mathscr{M}_{rca}(\Sigma)$, the Banach space of finite signed regular Borel measures on $(\Sigma, \mathscr{B}(\Sigma))$ (also known as Radon measures), the dual space $\mathscr{X}_1^*$ is isometrically isomorphic to $\mathscr{M}_{rba}(\Sigma)$, the Banach space of finitely additive finite regular signed measures on $(\Sigma, \mathscr{B}(\Sigma))$, and the dual space $\mathscr{X}_2^*$ is isometrically isomorphic to $\mathscr{M}_{ba}(\Sigma)$, the Banach space of finitely additive finite signed measures on $(\Sigma, \mathscr{B}(\Sigma))$. Note that when Σ is compact, the elements of $\mathscr{X}_1^*$ have countably additive extensions, thus $\mathscr{X}_1^*$ is isometrically isomorphic to $\mathscr{M}_{rca}(\Sigma)$.

At the abstract level, systems are represented by measures $\theta \in \mathscr{M}_1(\Sigma)$ induced by the underlying random processes,

---

[1]The definition of total variation distance is defined for signed measures as well.

which are defined on an appropriate Polish space. The set of admissible controls denoted by $\mathcal{U}_{ad}$ is assumed to be a compact subset of an appropriate Polish space ($\mathcal{U}$ which may be endowed with a metric compatible with the topology). The pay-off is represented by a linear functional on the space of probability measures $\mathcal{M}_1(\Sigma)$.

Nominal System. The nominal system is defined as follows. By choosing a control policy $u \in \mathcal{U}_{ad}$ for the nominal system (which is perfectly known), then the nominal system induces a nominal probability measure $\mu^u \in \mathcal{M}_1(\Sigma)$.

Uncertain System. For a given $u \in \mathcal{U}_{ad}$, let $M(u) \subset \mathcal{M}_1(\Sigma)$ denote the set of probability measures induced by the perturbed system while control $u \in \mathcal{U}_{ad}$ is applied. The perturbed system or uncertain system $v^u \in M(u)$ is further restricted to the following constraint described by the variational norm.

$$B_R(\mu^u) = \{v^u \in M(u) : d(v^u, \mu^u) \leq R\}, \quad R \geq 0$$

Mini-Max Optimization. Let $\ell^u : \Sigma \to \Re$ be a real-valued bounded non-negative measurable function. The uncertain system measure tries to maximize the average pay-off functional denoted by

$$L(v^u) \equiv \int_\Sigma \ell^u(x) v^u(dx) \tag{II.1}$$

over the set $B_R(\mu^u)$ for a given $u \in \mathcal{U}_{ad}$. The effect of uncertainty leads to the following maximization problem:

$$\sup_{v \in B_R(\mu^u)} L(v) \equiv \sup_{v \in B_R(\mu^u)} \int_\Sigma \ell^u(x) v^u(dx), \quad \text{for every} \quad u \in \mathcal{U}_{ad}$$

The designer on the other hand, tries to choose a control policy to minimize the worst case average cost. This gives rise to the min-max problem

$$\inf_{u \in \mathcal{U}_{ad}} \sup_{v \in B_R(\mu^u)} L(v^u) \tag{II.2}$$

Note that many stochastic optimization problems involve operation as in (II.2); specifically, value functions in optimal stochastic control are defined via (II.2) for a fixed v^u (e.g., without the supremum over the measure on which expectation is taken).

A. Characterization of the Maximizing Measure

In this section we drop the dependence on the control u of the various measures and functions. Suppose ℓ is a non negative element in $BC(\Sigma)$. Let $\mu \in \mathcal{M}_1(\Sigma) \subset \mathcal{M}_{rba}(\Sigma)$ be a given probability measure referred to as the nominal measure. Define the uncertainty set by

$$B_R(\mu) \stackrel{\triangle}{=} \left\{ v \in \mathcal{M}_1(\Sigma) : d(v, \mu) \leq R \right\}$$

The objective is to find the worst case (supremum) of average pay-off over the uncertainty set $B_R(\mu)$. The average pay-off is defined as a linear functional on the Banach space $\mathcal{M}_{rba}(\Sigma)$

given by $L(v) \equiv \int_\Sigma \ell(x) v(dx)$, subject to the constraint that $v \in B_R(\mu)$. Hence, the problem is the following

$$\tilde{L}(\mu) \equiv \sup_{v \in B_R(\mu)} L(v) = \int_\Sigma \ell(x) v(dx) \tag{II.3}$$

where $\mu \in \mathcal{M}_1(\Sigma)$ is fixed. The optimization problem in $(II.3)$ is solved as follows. Introduce the set $\mathcal{M}_0(\Sigma) \stackrel{\triangle}{=} \{\eta \in \mathcal{M}_{rca}(\Sigma) : \eta(\Sigma) = 0\}$. The constraint set $B_R(\mu)$ is equivalent to the set

$$B_R(\mathcal{M}_0(\Sigma)) \stackrel{\triangle}{=} \Big\{ \xi \in \mathcal{M}_0(\Sigma) : \xi = v - \mu, \quad v \in \mathcal{M}_1(\Sigma),$$
$$\mu \in \mathcal{M}_1(\Sigma) \text{ is fixed}, \quad ||\xi|| \leq R \Big\} \tag{II.4}$$

For $\xi \in \mathcal{M}_0(\Sigma)$, the total variation of ξ is defined by $||\xi|| \stackrel{\triangle}{=} \xi^+(\Sigma) + \xi^-(\Sigma)$, where $\{\xi^+, \xi^-\}$ is the Hanh-Jordan decomposition of ξ into $\xi = \xi^+ - \xi^-$. Moreover, $\xi(\Sigma) = 0$ implies that $\xi^+(\Sigma) = \xi^-(\Sigma)$, and hence $\xi^+(\Sigma) = \xi^-(\Sigma) = \frac{||\xi||}{2}$. For $\xi \in B_R(\mathcal{M}_0(\Sigma))$ then $\xi = (v - \mu)^+ - (v - \mu)^- \equiv \xi^+ - \xi^-$.

Define $\xi \stackrel{\triangle}{=} v - \mu \in \mathcal{M}_0(\Sigma)$. Then from (II.3), since $\ell \in BC(\Sigma)$ and nonnegative, we have have the following.

$$L(v) \leq \frac{R}{2} \left\{ \sup_{x \in \Sigma} \ell(x) - \inf_{x \in \Sigma} \ell(x) \right\} + E_\mu(\ell) \tag{II.5}$$

where $E_\mu(\ell) \stackrel{\triangle}{=} \int \ell d\mu$. Moreover, the upper bound in the right hand side of (II.5) is achieved by $\xi^* \in B_R(\mathcal{M}_0(\Sigma))$ as follows. Let

$$x^0 \in \Sigma^0 \quad \stackrel{\triangle}{=} \quad \left\{ x \in \Sigma : \ell(x) = \sup\{\ell(x) : x \in \Sigma\} \equiv M \right\},$$
$$x_0 \in \Sigma_0 \quad \stackrel{\triangle}{=} \quad \left\{ x \in \Sigma : \ell(x) = \inf\{\ell(x) : x \in \Sigma\} \equiv m \right\}$$

Take

$$\xi^*(dx) = v^*(dx) - \mu(dx) = \frac{R}{2} \Big(\delta_{x^0}(dx) - \delta_{x_0}(dx) \Big), \tag{II.6}$$

where $\delta_y(dx)$ denotes the Dirac measure concentrated at $y \in \Sigma$. This is indeed a signed measure with total variation $||v^* - \mu|| = R$, and $\int_\Sigma \ell(x)(v^* - \mu)(dx) = \frac{R}{2} \Big(M - m \Big)$. Hence, by using (II.6) as a candidate of the maximizing measure then

$$\int_\Sigma \ell(x) v^*(dx) = \frac{R}{2} \left\{ \sup_{x \in \Sigma} \ell(x) - \inf_{x \in \Sigma} \ell(x) \right\} + E_\mu(\ell), \tag{II.7}$$

where ξ^* satisfies the constraint $||\xi^*|| = ||v^* - \mu|| = R$, it is normalized $\xi^*(\Sigma) = 1$, while the additional condition $0 \leq v^*(A) \leq 1$ on any $A \in \mathcal{B}(\Sigma)$ should hold. Alternatively, (II.7) is expressed as

$$\int_\Sigma \ell(x) v^*(dx) = \int_{\Sigma^0} M v^*(dx) + \int_{\Sigma_0} m v^*(dx)$$
$$+ \int_{\Sigma \setminus \Sigma^0 \cup \Sigma_0} \ell(x) \mu(dx) \tag{II.8}$$

Hence,

$$\int_{\Sigma^0} v^*(dx) = \mu(\Sigma^0) + \frac{R}{2}, \quad \int_{\Sigma_0} v^*(dx) = \mu(\Sigma_0) - \frac{R}{2},$$
$$v^*(A) = \mu(A), \quad \forall A \subseteq \Sigma \setminus \Sigma^0 \cup \Sigma_0 \tag{II.9}$$

Lemma 2.1: [2]. Suppose $\ell : \Sigma \to \Re$ is a bounded non-negative measurable function, and η is a countably additive non-negative finite measure defined on $(\Sigma, \mathscr{B}(\Sigma))$. Then

$$\sup_{s>0} \frac{\int_{\Sigma} \ell(x) e^{s\ell(x)} \eta(dx)}{\int_{\Sigma} e^{s\ell(x)} \eta(dx)} = ||\ell||_{\infty, \eta} \qquad (\text{II.10})$$

where $||\ell||_{\infty, \eta} = \eta\text{-ess sup}_{x \in \Sigma} \ell(x) \stackrel{\triangle}{=} \inf_{\Delta \in \mathscr{N}_{\eta}} \sup_{x \in \Delta^c} ||\ell(x)||$ and $\mathscr{N}_{\eta} = \{A \in \mathscr{B}(\Sigma) : \eta(A) = 0\}$

Next, we state the main theorem which characterizes the maximizing measure as a convex combination of two probability measures.

Theorem 2.2: [2] Suppose $\ell \in BC(\Sigma)$ is non-negative, $\inf_{x \in \Sigma} \ell(x) = 0$, and $R \in [0, 2)$. Then there exists a family of probability measures which attain the supremum in (II.3) given by

$$v^*(E) = \frac{\beta}{\beta+1} \frac{\int_E e^{s_0 \ell(x)} \eta(dx)}{\int_{\Sigma} e^{s_0 \ell(x)} \eta(dx)} + \frac{1}{1+\beta} \mu(E) \qquad (\text{II.11})$$

where $E \in \mathscr{B}(\Sigma)$, $2\beta \in (R, \infty)$, $s_0 \in (0, \infty)$ and $\eta \in \mathscr{M}^+_{rca}(\Sigma)$ is arbitrary. Moreover, β and η satisfy $||v^* - \mu|| = R$.

Clearly, $v^*(dx)$ is a convex combination of the tilted measure $\frac{e^{s_0 \ell(x)} \eta(dx)}{\int_{\Sigma} e^{s_0 \ell(x)} \eta(dx)}$ and $\mu(dx)$. Hence, the initial optimization problem is equivalent to a linear combination of L_1 and L_{∞} optimization problem as follows.

$$\sup_{v \in B_R(\mu)} \int_{\Sigma} \ell(x) v(dx) = \frac{R}{2} . ||\ell||_{\infty} + E_{\mu}(\ell), \quad \text{for} \quad \inf_{x \in \Sigma} \ell(x) = 0,$$

$$= (\beta+1) \int_{\Sigma} \ell(x) v^*(dx), \quad f_{\eta, \ell}(s_0) = \frac{R}{2\beta} ||\ell||_{\infty}. \qquad (\text{II.12})$$

The rest of the paper deals with the application of the above results to discrete-time controlled stochastic systems (e.g., $\inf_{x \in \Sigma} \ell(x) = 0$).

III. FULLY OBSERVED UNCERTAIN CONTROL SYSTEMS

Define $\mathbb{N}_+ \stackrel{\triangle}{=} \{0, 1, 2, 3, \ldots\}$, $\mathbb{N}_+^n \stackrel{\triangle}{=} \{0, 1, 2, \ldots, n\}, n \in \mathbb{N}_+$. All processes are defined on the probability space $(\Omega, \mathbb{F}, \mathbb{Q})$ with filtration $\{\mathbb{F}_{0,j}\}_{j=0}^n, n \in \mathbb{N}_+$. Let $\mathscr{F}_{0,j} \subset \mathbb{F}_{0,j}, j = 0, 1, \ldots, n$ be a sub-sigma field. The state space, the control space, and the noise spaces are sequences of Polish spaces $\{\mathscr{X}_j : j = 0, 1, \ldots, n\}$, $\{\mathscr{U}_j : j = 0, 1, \ldots, n-1\}$, $\{\mathscr{W}_j : j = 1, 2, \ldots, n-1\}$, respectively. These spaces are associated with their corresponding measurable spaces $(\mathscr{X}_j, \mathscr{B}(\mathscr{X}_j)), j \in \mathbb{N}_+^n$, $(\mathscr{U}_j, \mathscr{B}(\mathscr{U}_j))$, $(\mathscr{W}_j, \mathscr{B}(\mathscr{W}_j))$, $j \in \mathbb{N}_+^{n-1}$. Thus, sequences of state spaces, control spaces, noise spaces are identified with their product spaces, $(\mathscr{X}_{0,n}, \mathscr{B}(\mathscr{X}_{0,n})), (\mathscr{U}_{0,n-1}, \mathscr{B}(\mathscr{U}_{0,n-1}))$, $(\mathscr{W}_{0,n-1}, \mathscr{B}(\mathscr{W}_{0,n-1}))$, respectively, $n \in \mathbb{N}^n$. The state process is denoted by $x \stackrel{\triangle}{=} \{x_j : j = 0, 1, \ldots, n\}, x_j : \Omega \longmapsto \mathscr{X}_j, j \in \mathbb{N}_+^n$, the control process is denoted by $u \stackrel{\triangle}{=} \{u_j : j = 0, 1, \ldots, n-1\}$, $u_j : \Omega \longmapsto \mathscr{U}_j, j \in \mathbb{N}_+^{n-1}$, and noise process is denoted by $w \stackrel{\triangle}{=} \{w_j : j = 1, 2, \ldots, n-1\}, w_j : \Omega \longmapsto \mathscr{W}_j, j \in \mathbb{N}_+^{n-1}$.

Denote by $\tilde{\mathscr{U}}_{ad}[0, n-1]$ the set of the $\mathscr{U}_{0,n-1}$−valued control processes u such that u_j is $\mathscr{F}_{0,j}$−measurable, $j \in \mathbb{N}_+^{n-1}$. Note that state constrained controls may be included in the formulation by further assuming that u_j take values in a nonempty subset $\mathscr{U}_j(x_j) \subset \mathscr{U}_j, \forall x_j \in \mathscr{X}_j, j = 0, 1, \ldots, n-1$.

Define two additional classes of admissible control laws as follows. $\mathscr{U}_{ad}[0, n-1] \subseteq \tilde{\mathscr{U}}_{ad}[0, n-1]$ denoting those controls u_j which are $\mathscr{G}_{0,j} \stackrel{\triangle}{=} \sigma\{x_0, \ldots, x_j, u_0, \ldots, u_{j-1}\}$−measurable, and $\mathscr{U}_{ad}^M[0, n-1] \subseteq \mathscr{U}_{ad}[0, n-1]$ denoting those controls which are $\sigma\{x_j\}$-measurable, called feedback control strategies, and Markovian control strategies, respectively. Thus, $u \in \mathscr{U}_{ad}[0, n-1]$ implies that there exists a sequence of measurable function called control laws or strategies $g \stackrel{\triangle}{=} \{g_j : j = 0, 1, \ldots, n-1\}$, $g_j : \mathscr{X}_{0,j} \times \mathscr{U}_{0,j-1} \to \mathscr{U}_j$, $u_j^g = g_j(x_0^g, x_1^g, \ldots, x_j^g, u_0^g, u_1^g, \ldots, u_{j-1}^g)$ and similarly for the rest. Conditional distributions are represented by stochastic kernels defined below.

Definition 3.1: Given the measurable spaces $(\mathscr{X}, \mathscr{B}(\mathscr{X})), (\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$, a stochastic Kernels on $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ conditioned on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ is a mapping $P : \mathscr{B}(\mathscr{Y}) \times \mathscr{X} \to [0, 1]$ satisfying the following two properties:
1) For every $x \in \mathscr{X}$, the set function $P(\cdot; x)$ is a probability measure (possibly finitely additive) on $\mathscr{B}(\mathscr{Y})$;
2) for every $A \in \mathscr{B}(\mathscr{Y})$, the function $P(A; \cdot)$ is $\mathscr{B}(\mathscr{X})$-measurable.
The set of all stochastic Kernels $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ conditioned on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ are denoted by $\mathscr{M}(\mathscr{Y}; \mathscr{X})$.

A. Problem Formulation

Below, the stochastic dynamics, pay-off, assumptions and uncertain system definitions are introduced.

Nominal Stochastic Dynamical Model. For each $u \in \tilde{\mathscr{U}}_{ad}[0, n-1]$ the nominal state process giving rise to a nominal measure is described by the following discrete-time difference equation.

Definition 3.2: (Nominal System). A nominal system family of state processes $\{x^g = x_0^g, x_1^g, \ldots, x_n^g : u \in \tilde{\mathscr{U}}_{ad}[0, n-1]\}$ corresponds to a sequence of stochastic kernels $\{P_{w_j}(dw; x, u) : j = 0, 1, \ldots, n-1\}$, and functions $\{b_j : \mathscr{X}_j \times \mathscr{U}_j \times \mathscr{W}_j \longmapsto \mathscr{X}_{j+1} : j = 0, 1, \ldots, n-1\}$ if for all $u \in \tilde{\mathscr{U}}_{ad}[0, n-1]$, there exists noise processes $\{w_j : j = 0, 1, \ldots, n-1\}$ such that the following hold.

1) For each $j \in \mathbb{N}_+^{n-1}$, w_j is $\mathscr{F}_{0,j}$− measurable and $\{x_0^g, x_1^g, \ldots, x_n^g\}$ are generated by the recursion

$$x_{j+1}^g = b_j(x_j^g, u_j^g, w_j), x_0^g = x_0 \qquad (\text{III.13})$$

which implies that if x_0 is $\mathscr{F}_{0,0}$−measurable then x_j^g is $\mathscr{F}_{0,j-1}$−measurable.

2) For every $A \in \mathcal{B}(\mathcal{W}_j), j \in \mathbb{N}_+^{n-1}$

$$Prob(w_j \in A|\mathcal{G}_{0,j}) = P_{w_j}(A; x_j^g, u_j^g), a.s.$$

3) $Prob(x_0^g = x_0) = 1, \forall u \in \tilde{\mathcal{U}}_{ad}[0, n-1].$

Uncertainty Stochastic Model. The uncertainty model is described by the conditional distribution, $\mathbb{Q}_{w_j|\mathcal{G}_{0,j}}(dw_j|\mathcal{G}_{0,j}) \in \mathcal{M}_1(\mathcal{W}_j), 0 \leq j \leq n-1$, as follows.

Definition 3.3: Given a nominal system of Definition 3.2, a fixed nominal stochastic kernel $P_{w_i}(dw_i; x_i^g, u_i^g) \in \mathcal{M}(\mathcal{W}_i; \mathcal{X}_i \times \mathcal{U}_i)$, and $R_i \in [0, 2]$, the class of measures is defined by

$$B_{R_i}(P_{w_i})(\mathcal{G}_{0,i}) \triangleq \left\{ \mathbb{Q}_{w_i}(\cdot|\mathcal{G}_{0,i}) \in \mathcal{M}_1(\mathcal{W}_{0,i-1}) : \right.$$
$$\left. ||\mathbb{Q}_{w_i}(\cdot; \mathcal{G}_{0,i}) - P_{w_i}(\cdot; x_i^g, u_i^g)||_{var} \leq R_i \right\}$$

for $i = 0, 1, \ldots, n-1$.

The above model is motivated by the fact that dynamic programming involves conditional expectation with respect to $\mathbb{Q}_{w_i}(dw_i; \mathcal{G}_{0,i})$.

Pay-Off Functional. The sample pay-off is functional of x^g, u^g, w, and for each $u \in \tilde{\mathcal{U}}_{ad}[0, n-1]$ the average pay-off is defined by

$$J_{0,n}(g, \mathbb{Q}) \triangleq E_{\mathbb{Q}} \left\{ \sum_{j=0}^{n-1} f_j(x_j^g, u_j^g, w_j) + h_n(x_n^g) \right\} \quad \text{(III.14)}$$

where $E_{\mathbb{Q}}(\cdot)$ denotes expectation with respect to the true joint measure $\mathbb{Q} \in \mathcal{M}_1(\mathcal{X}_{0,n} \times \mathcal{U}_{0,n-1} \times \mathcal{W}_{0,n-1})$

The following assumptions are introduced.

Assumption 3.4: The nominal system family satisfies the following assumptions:

1) $(\mathcal{U}_{0,n-1}, d)$ is Polish space. The control $\{u_j^g : j \in \mathbb{N}_+^{n-1}\}$ is non anticipative.

2) The maps $\{b_j : \mathcal{X}_j \times \mathcal{U}_j \times \mathcal{W}_j \longmapsto \mathcal{X}_{j+1} : j = 0, 1, \ldots, n-1\}$ are bounded continuous, and the maps $\{f_j : \mathcal{X}_j \times \mathcal{U}_j \times \mathcal{W}_j \longmapsto \mathbb{R} : j = 0, 1, \ldots, n-1\}, f_n : \mathcal{X}_n \longmapsto \mathbb{R}$ are bounded continuous and non-negative.

Notice that for $u \in \tilde{\mathcal{U}}_{ad}[0, n-1]$ the nominal system state x_j^g is a measurable function of $\{w_k : k = 0, 1, \ldots, j-1\}$ and $\{u_k^g : k = 0, 1, \ldots, j-1\}$, and hence x_j^g is $\mathcal{F}_{0,j-1}$−measurable for $j \in \mathbb{N}_+^n$.

For $u \in \mathcal{U}_{ad}[0, n-1]$ the nominal system state x_j^g is a measurable function of $\{w_k : k = 0, 1, \ldots, j-1\}$ and $\{u_k^g : k = 0, 1, \ldots, j-1\}$.

B. Maximization over a Class of Measures and Dynamic Programming

Section II describes at the abstract level, how to construct the maximizing measure of a linear functional over a total variational distance constraint. Similar arguments can be carried out for the case of discrete time stochastic controlled

systems, to deal with uncertainty of the conditional distribution $\mathbb{Q}_{w_j|\mathcal{G}_{0,j}}(dw_j|\mathcal{G}_{0,j}) \in \mathcal{M}_1(\mathcal{W}_j), 0 \leq j \leq n-1$.

Given the above formulation a minimax stochastic controlled problem can be formulated over a total variation distance uncertainty ball, centered at the nominal conditional distribution $P_{w_i}(dw_i; x_i^g, u_i^g) \in \mathcal{M}(\mathcal{W}_i; \mathcal{X}_i \times \mathcal{U}_i)$ having radius $R_i \in [0, 2]$, for $i = 0, 1, \ldots, n-1$ with respect to the total variation distance metric. The precise problem statement should thus, be as follows.

Problem 3.5: For a given $u \in \mathcal{U}_{ad}[0, n-1]$ assume that the measures $M(u)$ induced by the true uncertainty while control $u \in \mathcal{U}_{ad}[0, n-1]$ is applied are $M(u) \subset \mathcal{M}_1(\mathcal{W}_{0,n-1})$. Given a nominal system of Definition 3.2 an admissible control set $\mathcal{U}_{ad}[0, n-1]$ and an uncertainty class $B_{R_k}(\mathbb{P}_{w_k})(\mathcal{G}_{0,k}), k = 0, 1, \ldots, n-1$ find a $u^* \in \mathcal{U}_{ad}[0, n-1]$ and a sequence of stochastic kernels $\mathbb{Q}_{w_k}^*(dw_k; \mathcal{G}_{0,k}) \in B_{R_k}(P_{w_k})(\mathcal{G}_{0,k}), k = 0, 1, \ldots, n-1$ which solve the following minimax optimization problem.

$$J_{0,n}(g^*, \{\mathbb{Q}_{w_k}^*\}_{k=0}^{n-1}) = \inf_{\substack{u \in \mathcal{U}_{ad}[0,n-1] \\ }} \sup_{\substack{\mathbb{Q}_{w_k}(dw_k; \mathcal{G}_{0,k}) \in B_{R_k}(P_{w_k})(\mathcal{G}_{0,k}) \\ k=0,1,\ldots,n-1}}$$

$$E_{\mathbb{Q}} \left\{ \sum_{k=0}^{n-1} f_k(x_k^g, u_k^g, w_k) + h_n(x_n^g) \right\} \quad \text{(III.15)}$$

Dynamic Programming for Maximization over Conditional Distributions.

Define the pay-off associated with the maximization problem

$$J_{0,n}(g, \{\mathbb{Q}_{w_i}^*\}_{i=0}^{n-1}) \triangleq$$
$$\sup_{\substack{\mathbb{Q}_{w_k}(dw_k; \mathcal{G}_{0,k}) \in B_{R_k}(P_{w_k})(\mathcal{G}_{0,k}) \\ k=0,1,\ldots,n-1}} J_{0,n}(g, \{\mathbb{Q}_{w_k}\}_{k=0}^{n-1}) \quad \text{(III.16)}$$

Define the conditional expectation taken over the events $\mathcal{G}_{0,j}$ maximized over the class $B_{R_k}(P_{w_k})(\mathcal{G}_{0,k}), k = j, j+1, \ldots, n-1$, which is the value function of (III.16) as follows:

$$V_j(u_{[j,n-1]}^g, \mathcal{G}_{0,j}) \triangleq \sup_{\substack{\mathbb{Q}_{w_k}(dw_k; \mathcal{G}_{0,k}) \in B_{R_k}(P_{w_k})(\mathcal{G}_{0,k}) \\ k=j,j+1,\ldots,n-1}}$$

$$E_{\mathbb{Q}} \left\{ \sum_{k=j}^{n-1} f_k(x_k^g, u_k^g, w_k) + h_n(x_n^g)|\mathcal{G}_{0,j} \right\} \quad \text{(III.17)}$$

Then $V_j(u_{[j,n-1]}^g, \mathcal{G}_{0,j})$ satisfies the following dynamic programming equation.

$$V_j(u_{[j,n-1]}^g, \mathcal{G}_{0,j}) = \sup_{\mathbb{Q}_{w_j}(dw_j; \mathcal{G}_{0,j}) \in B_{R_j}(P_{w_j})(\mathcal{G}_{0,j})} E_{\mathbb{Q}_{w_j}(dw_j; \mathcal{G}_{0,j})}$$

$$\left\{ f_j(x_j^g, u_j^g, w_j) + V_{j+1}(u_{[j+1,n-1]}^g, \mathcal{G}_{0,j+1}) \right\} \quad \text{(III.18)}$$

$$V_n(\mathcal{G}_{0,n}) = h_n(x_n^g) \quad \text{(III.19)}$$

where $E_{\mathbb{Q}_{w_j}(dw_j; \mathcal{G}_{0,j})}$ denotes expectation with respect to $\mathbb{Q}_{w_j}(dw_j; \mathcal{G}_{0,j})$.

Theorem 3.6: Assume $f_j(x_j^g, u_j, \cdot) + V_{j+1}(u_{[j+1,n-1]}^g, \cdot) : \mathscr{W}_j \to \mathbb{R}_+$ in (III.18) is bounded continuous with zero infimum value.
Then

$$V_j(u_{[j,n-1]}^g, \mathscr{G}_{0,j}) = \frac{R_j}{2} \sup_{w_j \in \mathscr{W}_j} \left\{ f_j(x_j^g, u_j^g, w_j) \right.$$

$$+ V_{j+1}(u_{[j+1,n-1]}^g, \mathscr{G}_{0,j+1}) \right\} + E_{P_{w_j}} \left\{ f_j(x_j^g, u_j^g, w_j) \right.$$

$$+ V_{j+1}(u_{[j+1,n-1]}^g, \mathscr{G}_{0,j+1}) | \mathscr{G}_{0,j} \right\} \tag{III.20}$$

$$V_n(\mathscr{G}_{0,n}) = h_n(x_n^g) \tag{III.21}$$

and the supremum in (III.18) is attained at

$$\mathbb{Q}_{w_j}^*(dw_j; \mathscr{G}_{0,j})$$

$$= \frac{\gamma_j e^{s_j\left(f_j(x_j^g, u_j^g, w_j) + V_{j+1}(u_{[j+1,n-1]}^g, \mathscr{G}_{0,j+1}) \right)}}{E_{\mathbb{V}_{w_j}} \left\{ e^{s_j\left(f_j(x_j^g, u_j^g, w_j) + V_{j+1}(u_{[j+1,n-1]}^g, \mathscr{G}_{0,j+1}) \right)} | \mathscr{G}_{0,j} \right\}}$$

$$\times \mathbb{V}_{w_j}(dw_j; \mathscr{G}_{0,j}) + (1 - \gamma_j) P_{w_j}(dw_j; x_j^g, u_j^g) \tag{III.22}$$

where $\gamma_j = \frac{\beta_j}{1+\beta_j}$. Also,

$$J_{0,n}(g, \{\mathbb{Q}_{w_i}^*\}_{i=0}^{n-1}) = E\left\{ V_0(u_{[0,n-1]}, \mathscr{G}_{0,0}) \right\} \tag{III.23}$$

Moreover, if $\mathbb{V}_{w_j}(dw_j; \mathscr{G}_{0,j}) = P_{w_j}(dw_j; x_j^g, u_j^g)$, a.s., then $V_j(u_{[j,n-1]}, \mathscr{G}_{0,j}) = V_j(u_{[j,n-1]}, x_j^g)$, a.s.

Proof. (III.20), (III.21) and (III.23), follow from dynamic programming arguments. (III.22) is an application of Theorem 2.2. The last statement follows from the assumption.

Dynamic Programming for the Minimax Problem.

Let $V_j(\mathscr{G}_{0,j})$ represent the minimax pay-off on the future time horizon $\{j, j+1, ..., n\}$ at time $j \in \mathbb{N}_+^n$ defined by

$$V_j(\mathscr{G}_{0,j}) \stackrel{\triangle}{=} \inf_{\substack{u \in \mathscr{U}_{ad}[j,n-1]}} \sup_{\substack{Q_{w_k}(dw_k; \mathscr{G}_{0,k}) \in B_{R_k}(P_{w_k})(\mathscr{G}_{0,k}) \\ k=j,j+1,...,n-1}}$$

$$E_{\mathbb{Q}} \left\{ \sum_{k=j}^{n-1} f_k(x_k^g, u_k^g, w_k) + h_n(x_n^g) | \mathscr{G}_{0,j} \right\} \tag{III.24}$$

$$= \inf_{u \in \mathscr{U}_{ad}[j,n-1]} V_j(u_{[j,n-1]}^g, \mathscr{G}_{0,j}) \tag{III.25}$$

Then by reconditioning one obtains

$$V_j(\mathscr{G}_{0,j}) \stackrel{\triangle}{=} \inf_{\substack{u \in \mathscr{U}_{ad}[j,n-1]}} \sup_{\substack{Q_{w_k}(dw_k; \mathscr{G}_{0,k}) \in B_{R_k}(P_{w_k})(\mathscr{G}_{0,k}) \\ k=j,j+1,...,n-1}}$$

$$E_{\mathbb{Q}} \left\{ f_k(x_k^g, u_k^g, w_k) + E_{\mathbb{Q}} \left\{ \sum_{k=j+1}^{n-1} f_k(x_k^g, u_k^g, w_k) \right.\right.$$

$$\left.\left. + h_n(x_n^g) | \mathscr{G}_{0,j+1} \right\} | \mathscr{G}_{0,k} \right\} \tag{III.26}$$

Hence, the following dynamic programming recursion

$$V_j(\mathscr{G}_{0,j}) \stackrel{\triangle}{=} \inf_{\substack{u \in \mathscr{U}_{ad}[j,j]}} \sup_{\substack{Q_{w_j}(dw_j; \mathscr{G}_{0,j}) \in B_{R_j}(P_{w_j})(\mathscr{G}_{0,j})}}$$

$$E_{\mathbb{Q}_{w_j}} \left\{ f_j(x_j^g, u_j^g, w_j) + V_{j+1}(\mathscr{G}_{0,j+1}) | \mathscr{G}_{0,j} \right\} \tag{III.27}$$

$$V_n(\mathscr{G}_{0,n}) = h_n(x_n^g) \tag{III.28}$$

Next, invoke the following additional assumption.

Assumption 3.7: The maximizing measure (III.22) in Theorem 3.6 is chosen so that $\mathbb{Q}_{w_j}^*(dw_j; \mathscr{G}_{0,j}) = Q_{w_j}^*(dw_j; x_j^g, u_j^g)$, a.s.

The above assumption is satisfied provided the tilted measure in (III.22) is constructed from the nominal measure $P_{w_j}(dw_j; x_j^g, u_j^g)$, via $\mathbb{V}_{w_j}(dw_j; \mathscr{G}_{0,j}) = P_{w_j}(dw_j; x_j^g, u_j^g)$, a.s. Next, we state the main theorem.

Theorem 3.8: Suppose Assumption 3.7 holds. Then $V_j(\mathscr{G}_{0,j}) = V_j(x)$ satisfies the following dynamic programming recursion

$$V_j(x) \stackrel{\triangle}{=} \inf_{\substack{u \in \mathscr{U}_{ad}[j,j]}} \sup_{\substack{Q_{w_j}(dw_j; x, u) \in B_{R_j}(P_{w_j})(x,u)}}$$

$$E_{Q_{w_j}} \left\{ f_j(x, u, w) + V_{j+1}(b_j(x, u, w)) \right\} \tag{III.29}$$

$$V_n(x) = h_n(x) \tag{III.30}$$

Also,

$$J_{0,n}(g^*, \{\mathbb{Q}_{w_i}^*\}_{i=0}^{n-1}) = E\left\{ V_0(x) \right\} \tag{III.31}$$

Moreover, if $f_j(x_j, u_j, \cdot) + V_{j+1}(b_j(x_j, u_j, \cdot)) : \mathscr{W}_j \to \mathbb{R}_+$ is bounded continuous with zero infimum value then

$$V_j(x) \stackrel{\triangle}{=} \inf_{\substack{u \in \mathscr{U}_{ad}[j,j]}} \left\{ \frac{R_j}{2} \sup_{w \in \mathscr{W}_j} \left\{ f_j(x, u, w) + V_{j+1}(b_j(x, u, w)) \right\} \right.$$

$$+ E_{P_{w_j}(dw; x, u)} \left\{ f_j(x, u, w) + V_{j+1}(b_j(x, u, w)) \right\} \right\} \tag{III.32}$$

$$V_n(x) = h_n(x) \tag{III.33}$$

Remark. 1) The point to be made here is that the dynamic programming equation (III.32) involves in its right hand side the supremum of the cost-to-go in addition to the standard term, which to the best of our knowledge, has not appeared in the literature. When the zero infimum value is removed the right side of (III.32) will include the term $-\frac{R_j}{2} \inf_{w \in \mathscr{W}_j} \left\{ f_j(x, u, w) + V_{j+1}(b_j(x, u, w)) \right\}$ as in (II.7). 2) One may also conclude that dynamic programming recursion such as (III.32), (III.33) can be derived for finite state Markov Decision problems as well.

C. Working Example

To illustrate an application of the minimax problem, consider an inventory control example inspired by [12]. Specifically, an optimal inventory ordering policy of a quantity of a certain

item at each of N periods must be found so as to meet a stochastic demand. Let us denote

- x_k, stock available at the beginning of the kth period
- u_k, stock ordered at the beginning of the kth period
- w_k, demand during kth period with given probability distribution
- h, holding cost per unit item remaining unsold at the end of the kth period
- c, cost per unit stock ordered
- p, shortage cost per unit demand unfilled

The random disturbance w_k may depend on values of x_k and u_k but not on values of prior disturbances $w_0, ..., w_{k-1}$. Excess demand is backlogged and filled as soon as additional inventory becomes available. The noise of previous section belong to a class described by $B_{R_i}(P_{w_i})(\mathcal{G}_{0,i})$ (see Definition 3.3), and inventory and demand are non-negative integers variables. Thus, the nominal system equation is

$$x_{k+1} = \max(0, x_k + u_k - w_k) \qquad \text{(III.34)}$$

The total cost over N periods is

$$E\left\{ \sum_{k=0}^{N-1} [cu_k + p\max(0, w_k - x_k - u_k) \right.$$
$$\left. + h\max(0, x_k + u_k - w_k)] \right\} \qquad \text{(III.35)}$$

Assume that the maximum capacity $(x_k + u_k)$ for stock is 2 units, that the planning horizon $N = 3$ periods, and that the holding cost h and the ordering cost c are both 1 unit. The shortage cost p is assumed to be 3 units. The demand w_k has a nominal probability distribution given by, $P_{w_k}(w_k = 0) = 0.1$, $P_{w_k}(w_k = 1) = 0.7$, and $P_{w_k}(w_k = 2) = 0.2$. There is no fixed cost and the initial stock is zero. The dynamic programming algorithm for the minimax problem takes the form (ignoring the $-\frac{R}{2}\min_{w_k \in \{0,1,2\}}\{\cdot\}$ term)

$$V_k(x_k) = \min_{0 \le u_k \le 2-x_k} \left\{ \frac{R}{2} \max_{w_k \in \{0,1,2\}} \left[u_k + \max(0, x_k + u_k - w_k) \right. \right.$$
$$\left. + 3\max(0, w_k - x_k - u_k) + V_{k+1}(\max(0, x_k + u_k - w_k)) \right]$$
$$+ E_{w_k}\{u_k + \max(0, x_k + u_k - w_k) + 3\max(0, w_k - x_k - u_k)$$
$$\left. + V_{k+1}(\max(0, x_k + u_k - w_k))\} \right\} \qquad \text{(III.36)}$$

$$V_3(x_3) = 0 \qquad \text{(III.37)}$$

where $R_k = R \in [0, 2]$. By setting $R = 0$, the uncertain class reduces to the nominal demand probability distribution, which implies that the maximizing term in (III.36) is removed (as in [12]). The problem is solved for two possible values of R for each period resulting in optimal ordering policies as shown in Table IV.1. By setting $R = 1$, we choose to calculate the optimal ordering policy, when $P_{w_k}(w_k \ne i)$, $k = 0, 1, 2$, $i = 0, 1, 2$ is uncertain. Taking fully into consideration the maximizing term in (III.36) the dynamic programming algorithm results in optimal ordering policies which are more robust with respect to uncertainty, but with the sacrifice of low present and future costs. In cases where the planner needs to balance the desire for low costs with the undesirability of scenarios

with high uncertainty, he must choose values of R between 0 and 1.

IV. CONCLUSION AND FUTURE WORK

The paper derives a new dynamic programming equation when the conditional distribution which appear in dynamic programming equation is described by a ball with respect to the total variation distance uncertainty. Future work will treat the general case described by II.8 and II.9 and will determine how $v^*(dx)$ behaves as a function of R.

REFERENCES

[1] F. Rezaei, C.D. Charalambous, and N.U. Ahmed. *Optimization of Stochastic Uncertain Systems with Variational Norm Constraints*. In 2007 Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, U.S.A., December 12-14, 2007.

[2] F. Rezaei, C.D. Charalambous, and N.U. Ahmed. *Optimal Control of Uncertain Stochastic Systems Subject to Total Variation Distance Uncertainty*. SIAM Journal on Control and Optimization, Submitted, pp.35, 2010.

[3] M.S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, 1964.

[4] P.R. Kumar, and J.H., Van Schuppen. *On the Optimal Control of Stochastic Systems with an Exponential-of-Integral Performance Index*. Journal of Mathematical Analysis and Applications, Vol. 80, pp.312-332, 1981.

[5] W.M. McEneany. *Connections between Risk-Sensitive Stochastic Control, Differential Games, and H^∞ Control: The Nonlinear Case*. PhD Dissertation, Brown University, 1993.

[6] C.D. Charalambous. *Partially Observable Nonlinear Risk-Sensitive Control Problems: Dynamic Programming and Verification Theorems*. IEEE Transactions on Automatic Control, Vol. 42, No. 8, pp.1130-1138, August 1997.

[7] W. Rudin. *Functional Analysis*. McGraw-Hill, New York, 1991.

[8] J. Yong and X. Y. Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. New York: Springer-Verlag, 1999.

[9] N. Dunford, and J. T. Schwartz. *Linear Operators: Part 1: General Theory*. Interscience Publishers, Inc., New York, 1957.

[10] H.V. Poor. *On Robust Wiener Filtering*. IEEE Transactions on Automatic Control, Vol. 25, No. 3, pp.531-536, June 1980.

[11] K.S. Vastola and H.V. Poor. *On Robust Wiener-Kolmogorov Theory*. IEEE Transactions on Information Theory, Vol. 30, No. 2, pp.315-327, March 1984.

[12] D.P. Bertsekas *Dynamic Programming and Optimal Control*. Athena Scientific, 2005.

	R = 1			R = 0.2	
Stock	Stage.0 Cost-to-go	Stage.0 Optimal Stock to Purchase	Stock	Stage.0 Cost-to-go	Stage.0 Optimal Stock to Purchase
0	15.03	2	0	6.72	1
1	13.53	1	1	5.62	0
2	12.03	0	2	4.68	0
Stock	Stage.1 Cost-to-go	Stage.1 Optimal Stock to Purchase	Stock	Stage.1 Cost-to-go	Stage.1 Optimal Stock to Purchase
0	8.32	2	0	4.30	1
1	6.82	1	1	3.20	0
2	5.32	0	2	2.44	0
Stock	Stage.2 Cost-to-go	Stage.2 Optimal Stock to Purchase	Stock	Stage.2 Cost-to-go	Stage.2 Optimal Stock to Purchase
0	3.7	1	0	2.10	1
1	2.2	0	1	1	0
2	1.9	0	2	1.10	0

TABLE IV.1

RESULTS OF THE DP ALGORITHM