Dynamic Programming with Total Variational Distance Uncertainty

Charalambos D. Charalambous, Ioannis Tzortzis and Themistoklis Charalambous

***Abstract*— The aim of this paper is to address optimality of stochastic control strategies via dynamic programming subject to total variational distance uncertainty on the conditional distribution of the controlled process. Utilizing concepts from signed measures, the maximization of a linear functional on the space of probability measures on abstract spaces is investigated, among those probability measures which are within a total variational distance from a nominal probability measure. The maximizing probability measure is found in closed form. These results are then applied to solve minimax stochastic control with deterministic control strategies, under a Markovian assumption on the conditional distributions of the controlled process. The results include: 1) Optimization subject to total variational distance constraints, 2) new dynamic programming recursions, which involve the oscillator seminorm of the value function.**

I. Introduction

The main objective of this paper is to investigate the effect of uncertainties on the conditional distribution of the state given past and present states and controls on dynamic programming. Specifically, by quantifying the uncertainty on the conditional distribution by a ball with respect to the total variational distance metric, centered at a nominal conditional distribution, a new dynamic programming is derived using minimax theory, with two players, the controls and conditional distributions opposing each other actions. The conditional distributions objective is to maximize the cost-to-go, while the controls objective is to minimize it. The maximization over the uncertainty ball of conditional distributions is found by first deriving results related to maximizing linear functionals on a subset of the space of signed measures. Utilizing these results, a new dynamic programming recursion is presented which in addition to the standard terms, includes additional terms that codify the level of uncertainty allowed in the conditional distribution with respect to the total variational distance.

Previous related work on optimization of stochastic uncertain systems subject to total variational distance uncertainty is addressed by the authors in [1], [2]. However, the solution method employed in [2] is fundamentally different; it approaches the maximization problem indirectly, by employing Large Deviations concepts to derive the maximizing measure as a convex combination of a tilted probability measure and the nominal measure, under some restrictions on the uncertain measure.

In this paper, we further elaborate on dynamic programming subject to total variational distance uncertainty from a different point of view, utilizing concepts from signed measures and provide a complete characterization of the maximization of linear functional over the space of conditional distributions subject to a total variational distance constraint. Utilizing these results in dynamic programming and minimax theory, a new dynamic programming recursion is derived which explicitly depends on the uncertainty radius and the closed form expression of the maximizing measure.

The rest of the paper is organized as follows. In Section I-A, a brief formulation of Markov Control Model (MCM), cost-to-go and associated dynamic programming recursion are introduced, followed by its reformulation using total variational distance uncertainty and minimax theory. 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 MCM. Dynamic programming recursions are derived to characterize the optimality of minimax strategies, while the maximizing measure is characterized.

Next, we give a high level discussion on classical dynamic programming for MCM and we present some aspects of the problems and results pursued in this paper.

A. Dynamic Programming

A finite horizon Markov Control Model (MCM) is a six-tuple

$$\text{MCM}: \Big(\{\mathcal{X}\}_{i=0}^{n}, \{\mathcal{U}\}_{i=0}^{n-1}, \{\mathcal{U}_i(x_i): x_i \in \mathcal{X}_i\}_{i=0}^{n-1}, \{Q_i(dx_i|x_{i-1}, u_{i-1}): (x_{i-1}, u_{i-1}) \in \mathcal{X}_{i-1} \times \mathcal{U}_{i-1}\}_{i=0}^{n}, \{f_i\}_{i=0}^{n}, h_n\Big)$$

consisting of

(a) State Space. A sequence of Polish spaces (complete separable metric spaces) $\{\mathcal{X}_i : i = 0, \ldots, n\}$, being the state space of the controlled random process $\{x_j : j = 0, \ldots, n\}$, $x_j \in \mathcal{X}_j$.

(b) Control or Action Space. A sequence of Polish spaces $\{\mathcal{U}_i : i = 0, \ldots, n-1\}$, being the control or action set of the control random process $\{u_j : j = 0, \ldots, n-1\}$, $u_j \in \mathcal{U}_j$.

(c) Feasible Controls or Actions. A family $\{\mathcal{U}_i(x_i) : x_i \in \mathcal{X}_i\}$ of non-empty measurable subsets $\mathcal{U}_i(x_i)$ of $\mathcal{U}_i$, where $\mathcal{U}_i(x_i)$ denotes the set of feasible controls or actions when the controlled process is in state $x_i \in \mathcal{X}_i$, and the feasible state-actions pairs defined by $\mathbb{K}_i \stackrel{\triangle}{=} \Big\{(x_i, u_i) : x_i \in \mathcal{X}_i, u_i \in \mathcal{U}_i(x_i)\Big\}$ are measurable subsets of $\mathcal{X}_i \times \mathcal{U}_i, i = 0, \ldots, n-1$.

This work was not supported by any organization

C. D. Charalambous is with Faculty of Electrical Engineering, University of Cyprus, Nicosia, Cyprus chadcha@ucy.ac.cy

I. Tzortzis is with Faculty of Electrical Engineering, University of Cyprus, Nicosia, Cyprus tzortzis.ioannis@ucy.ac.cy

T. Charalambous, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden themis@ieee.org

(d) Controlled Process Distribution. A collection of conditional distributions or stochastic kernels $Q_i(dx_i|x_{i-1},u_{i-1})$ on $\mathcal{X}_i$ given $(x_{i-1},u_{i-1}) \in \mathbb{K}_{i-1} \subseteq \mathcal{X}_{i-1} \times \mathcal{U}_{i-1}, i = 0,\ldots,n$. The controlled process distribution is described by the sequence of transition probability distributions $\{Q_i(dx_i|x_{i-1},u_{i-1}) : i = 0,\ldots,n-1\}$.
(e) Cost-Per-Stage. A collection of non-negative measurable functions $f_j : \mathbb{K}_j \to [0,\infty]$, such that $f_j(x,\cdot)$ does not take the value $+\infty$ for each $x \in \mathcal{X}_j, j = 0,\ldots,n-1$. f_j is called the cost-per-stage while the running pay-off functional in the minimal pay-off functional is defined in terms of this collection of functions.
(f) Terminal Cost. A bounded measurable non-negative function $h_n : \mathcal{X}_n \to [0,\infty)$ called the terminal cost, in terms of which the pay-off functional at the last stage of the minimal pay-off functional is defined.

The definition of MCM envisions applications of systems described by discrete-time dynamical state space models, which include random external inputs, since such models give rise to a collection of controlled processes distributions $\{Q_i(dx_i|x_{i-1},u_{i-1}) : i = 0,\ldots,n\}$.
Let $\mathcal{X}_{0,j} = \times_{i=0}^{j}\mathcal{X}_i$ and similarly for the rest. The goal in Markov controlled optimization with deterministic strategies is to choose a control strategy or policy $g \stackrel{\triangle}{=} \{g_j : j = 0,1,\ldots,n-1\}$, $g_j : \mathcal{X}_{0,j} \times \mathcal{U}_{0,j-1} \to \mathcal{U}_j(x_j)$, $u_j^g = g_j(x_0^g, x_1^g, \ldots, x_j^g, u_0^g, u_1^g, \ldots, u_{j-1}^g)$, $j = 0,1,\ldots,n-1$ so as to minimize the pay-off functional

$$\mathbb{E}^g\left\{\sum_{j=0}^{n-1} f_j(x_j^g,u_j^g) + h_n(x_n^g)\right\} \tag{I.1}$$

The pay-off (I.1) is a functional of the collection of conditional distributions $\{Q_i(\cdot|\cdot) : i = 0,1,\ldots,n\}$.

For $(i,x) \in \{0,1\ldots,n\} \times \mathcal{X}_i$, let $V_i^0(x) \in \mathbb{R}$ represent the minimal cost-to-go or value function on the time horizon $\{i,i+1,\ldots,n\}$ if the state of the controlled process starts at state $x_i = x$ at time i, defined by

$$V_i^0(x) \stackrel{\triangle}{=} \inf_{\substack{g_k \in \mathcal{U}_k(x_k)\\ k=i,\ldots,n-1}} \mathbb{E}_{i,x}^g\Big\{\sum_{j=i}^{n-1} f_j(x_j^g,u_j^g) + h_n(x_n^g)\Big\} \tag{I.2}$$

The value function satisfies the following dynamic programming recursion [4]

$$V_i^0(x) = \inf_{u\in\mathcal{U}_i(x)}\Big\{f_i(x,u) + \int_{\mathcal{X}_{i+1}} V_{i+1}^0(z)Q_{i+1}(dz|x,u)\Big\}, \quad x \in \mathcal{X}_i \tag{I.3}$$

$$V_n^0(x) = h_n(x), \quad x \in \mathcal{X}_n \tag{I.4}$$

The value function $V_i^0(x)$ defined by (I.2) and the dynamic programming recursion (I.3), (I.4) depend on the complete knowledge of the collection of conditional distributions $\{Q_i(\cdot|\cdot) : i = 0,\ldots,n\}$. Any mismatch of this collection from the true collection of conditional distributions will affect the optimality of the control strategies.

Dynamic Programming with Total Variational Distance Uncertainty
Motivated by the above discussion, the objective of this paper is to investigate dynamic programming under uncertainty of the conditional distributions of the controlled processes $\{Q_i(dx_i|x_{i-1},u_{i-1}) : (x_{i-1},u_{i-1}) \in \mathbb{K}_{i-1}\}, i = 0,\ldots,n$. The uncertainty of the conditional distributions of the controlled process is described via the total variational distance. Specifically, given a collection of nominal controlled process distributions $\{P_i(dx_i|x_{i-1},u_{i-1}) : (x_{i-1},u_{i-1}) \in \mathbb{K}_{i-1}\}, i = 0,\ldots,n$, the corresponding collection of true controlled process distributions $\{Q_i(dx_i|x_{i-1},u_i) : (x_{i-1},u_{i-1}) \in \mathbb{K}_{i-1}\}, i = 0,\ldots,n$ belongs to a set described by the total variational distance centered at the nominal conditional distribution having radius $R_i \in [0,2]$ defined by

$$B_{R_i}(P_i)(x_{i-1},u_{i-1}) \stackrel{\triangle}{=} \Big\{Q_i(\cdot|x_{i-1},u_{i-1}) : ||Q_i(\cdot|x_{i-1},u_{i-1}) - P_i(\cdot|x_{i-1},u_{i-1})||_{TV} \le R_i\Big\}$$

Here $||\cdot||_{TV}$ denotes the total variational distance between two probability measures. This type of uncertainty model is quite general since no assumption is required on the structure of the stochastic control dynamical system model which induces the collection of conditional distributions $\{Q_i(\cdot|\cdot) : i = 0,\ldots,n\}$,$\{P_i(\cdot|\cdot) : i = 0,\ldots,n\}$, hence it includes linear, non-linear, finite and/or countable state space models, etc.
Given the above characterization of uncertainty, the reformulation of value function and dynamic programming recursion is done via minimax theory as follows.

For $(i,x) \in \{0,1\ldots,n\} \times \mathcal{X}_i$, let $V_i(x) \in \mathbb{R}$ represent the minimal cost-to-go on the time horizon $\{i,i+1,\ldots,n\}$ if the state of the controlled process starts at state $x_i = x$ at time i, defined by

$$V_i(x) \stackrel{\triangle}{=} \inf_{\substack{g_k\in\mathcal{U}_k(x_k)\\ k=i,\ldots,n-1}} \sup_{\substack{Q_{k+1}(\cdot|x_k,u_k)\in B_{R_{i+1}}(P_{i+1})(x_i,u_i)\\ k=i,\ldots,n}} \mathbb{E}_{i,x}^g\Big\{\sum_{j=i}^{n-1} f_j(x_j^g,u_j^g) + h_n(x_n^g)\Big\} \tag{I.5}$$

The value function (I.5) satisfies the following dynamic programming recursion

$$V_i(x) = \inf_{u\in\mathcal{U}_i(x)} \sup_{Q_{i+1}(\cdot|x,u)\in B_{R_{i+1}}(P_{i+1})}\Big\{f_i(x,u) + \int_{\mathcal{X}_{i+1}} V_{i+1}(z)Q_{i+1}(dz;x,u)\Big\}, \quad x \in \mathcal{X}_i \tag{I.6}$$

$$V_n(x) = h_n(x), \quad x \in \mathcal{X}_n \tag{I.7}$$

Based on this formulation if $V_{i+1}(\cdot)$ is bounded continuous non-negative, the new dynamic programming equation is

given by

$$V_j(x) \stackrel{\triangle}{=} \inf_{u\in\mathcal{U}(x)} \Big\{ f_j(x,u) + \int_{\mathcal{X}_{j+1}} V_{j+1}(z) P_{j+1}(dz;x,u) + \frac{R_j}{2} \big\{ \sup_{z\in\mathcal{X}_{j+1}} V_{j+1}(z) - \inf_{z\in\mathcal{X}_{j+1}} V_{j+1}(z) \big\} \Big\},\ x \in \mathcal{X}_j \tag{I.8}$$

$$V_n(x) = h_n(x),\ x \in \mathcal{X}_j \tag{I.9}$$

Note that the new term in the right side of (I.8) is the oscillator seminorm of $V_{j+1}(\cdot)$ called the global modulus of continuity of $V_{j+1}(\cdot)$. The issues discussed in this paper are the following.

- Formulation of stochastic optimal control subject to conditional distribution uncertainty described by total variational distance via minimax theory.
- Dynamic programming recursions for nominal MCM under total variational distance uncertainty on the conditional distribution of the controlled process.
- Characterization of the maximizing conditional distribution belonging to the total variational distance set, and a corresponding new dynamic programming recursions.

II. Maximization With Total Variational Distance on Abstract Spaces

The goal on this section is to formulate maximization problems with total variational distance uncertainty on abstract spaces, and then discuss how the maximization over total variation distance constraint is resolved. These material are utilized in subsequent section.

Let (Σ, d_Σ) denote a Polish space, and $(\Sigma, \mathcal{B}(\Sigma))$ the corresponding measurable space. Let $\mathcal{M}_1(\Sigma)$ denote space of countably additive probability measures on $(\Sigma, \mathcal{B}(\Sigma))$.

The total variational distance[1] is a metric $d : \mathcal{M}_1(\Sigma) \times \mathcal{M}_1(\Sigma)$ is defined by

$$d(\alpha,\beta) \equiv ||\alpha-\beta||_{TV} \stackrel{\triangle}{=} \sup_{P\in\mathcal{P}(\Sigma)} \sum_{F_i\in P} |\alpha(F_i) - \beta(F_i)|$$

where $\alpha, \beta \in \mathcal{M}_1(\Sigma)$ and $\mathcal{P}(\Sigma)$ denotes the collection of all finite partitions of Σ.
Given a known or nominal probability measure $\mu \in \mathcal{M}_1(\Sigma)$ the uncertainty set based on total variational distance is defined by

$$B_R(\mu) \stackrel{\triangle}{=} \Big\{ \nu \in \mathcal{M}_1(\Sigma) : ||\nu-\mu||_{TV} \le R \Big\}, \qquad R \in [0,\infty)$$

Since the elements of $\mathcal{M}_1(\Sigma)$ are probability measures $R \in [0,2]$.
Next, we describe the abstract formulation.

Nominal System. The nominal system is a fixed nominal probability measure $\mu \in \mathcal{M}_1(\Sigma)$.

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

Uncertain System. The uncertain system $\nu \in \mathcal{M}_1(\Sigma)$ belongs to the set

$$B_R(\mu) = \{\nu \in \mathcal{M}_1(\Sigma) : d(\nu,\mu) \le R\}, \quad R \ge 0$$

Maximization. Let $BC(\Sigma)$ denote the Banach space of bounded continuous functions $\ell : \Sigma \mapsto \mathbb{R}$ and $BM(\Sigma)$ denote the Banach space of bounded measurable functions on Σ both endowed with the sup norm $||\ell|| \stackrel{\triangle}{=} \sup_{x\in\Sigma} |\ell(x)|$. The subset of $BC(\Sigma)$ consisting of nonnegative functions is denoted by $BC^+(\Sigma)$. For a given $\ell \in BM(\Sigma)$ or $\ell \in BC(\Sigma)$ the uncertain system measure tries to maximize the average pay-off functional over the set $B_R(\mu)$ for a given $\mu \in \mathcal{M}_1(\Sigma)$, defined by

$$\sup_{\nu\in B_R(\mu)} L(\nu) \equiv \sup_{\nu\in B_R(\mu)} \int_\Sigma \ell(x)\nu(dx)$$

In minimax optimization theory, ν, μ, ℓ will depend on another agents strategy as well.

A. Partial Characterization of the Maximizing Measure

The functional $L(\nu)$ will be maximized over the total variational distance constraint for the class of functions $BC^+(\Sigma)$. Note that $BC(\Sigma)$ can be generalized to $L^{\infty,+}(\Sigma, \mathcal{B}(\Sigma), \nu)$, the set of all $\mathcal{B}(\Sigma)$-measurable,non-negative essentially bounded functions defined $\nu - a.e.$ endowed with the essential supremum norm $||\ell||_{\infty,\nu}$. Next, we utilize certain concepts from signed measures to characterize the maximizing measure.
Let $\mathcal{M}_{sm}(\Sigma)$ denote the set of finite signed measures. Then any $\eta \in \mathcal{M}_{sm}(\Sigma)$ has a Jordan decomposition $\{\eta^+, \eta^-\}$ such that $\eta = \eta^+ - \eta^-$, and the total variation of η is defined by $||\eta||_{TV} \stackrel{\triangle}{=} \eta^+(\Sigma) + \eta^-(\Sigma)$. Define the following subset $\mathbb{M}_0(\Sigma) \stackrel{\triangle}{=} \Big\{ \eta \in \mathcal{M}_{sm}(\Sigma) : \eta(\Sigma) = 0 \Big\}$. For a given $\mu \in \mathcal{M}_1(\Sigma)$ the set $B_R(\mu)$ is equivalent to the set

$$\widetilde{B}_R(\mu) \stackrel{\triangle}{=} \Big\{ \xi \in \mathbb{M}_0(\Sigma) : \xi = \nu - \mu, \quad \nu \in \mathcal{M}_1(\Sigma), \quad ||\xi||_{TV} \le R \Big\} \tag{II.10}$$

For $\xi \in \mathbb{M}_0(\Sigma)$, then $\xi(\Sigma) = 0$, which implies that $\xi^+(\Sigma) = \xi^-(\Sigma)$, and hence $\xi^+(\Sigma) = \xi^-(\Sigma) = \frac{||\xi||_{TV}}{2}$. Thus, for any $\xi \in \widetilde{B}_R(\mu)$ then $\xi = (\nu-\mu)^+ - (\nu-\mu)^- \equiv \xi^+ - \xi^-$.
Define $\xi \stackrel{\triangle}{=} \nu - \mu \in \mathbb{M}_0(\Sigma)$. Then since $\ell \in BC^+(\Sigma)$ the following inequalities are obtained.

$$\begin{aligned}
&\int_\Sigma \ell(x)\nu(dx) = \\
&= \int_\Sigma \ell(x)\xi^+(dx) - \int_\Sigma \ell(x)\xi^-(dx) + \int_\Sigma \ell(x)\mu(dx) \\
&\le \sup_{x\in\Sigma} \ell(x)\xi^+(\Sigma) - \inf_{x\in\Sigma} \ell(x)\xi^-(\Sigma) + \int_\Sigma \ell(x)\mu(dx) \\
&= \Big\{ \sup_{x\in\Sigma} \ell(x) - \inf_{x\in\Sigma} \ell(x) \Big\} \frac{||\xi||_{TV}}{2} + \int_\Sigma \ell(x)\mu(dx)
\end{aligned} \tag{II.11}$$

Moreover, the upper bound in the right hand side of (II.11) is achieved by $\xi^* \in \widetilde{B}_R(\mu)$ as follows. Let

$$x^0 \in \Sigma^0 \triangleq \Big\{x \in \Sigma : \ell(x) = \sup\{\ell(x) : x \in \Sigma\} \equiv M\Big\},$$
$$x_0 \in \Sigma_0 \triangleq \Big\{x \in \Sigma : \ell(x) = \inf\{\ell(x) : x \in \Sigma\} \equiv m\Big\}.$$

Take

$$\xi^*(dx) = \nu^*(dx) - \mu(dx) = \frac{R}{2}\Big(\delta_{x^0}(dx) - \delta_{x_0}(dx)\Big) \quad \text{(II.12)}$$

where $\delta_y(dx)$ denotes the Dirac measure concentrated at $y \in \Sigma$. This is indeed a signed measure with total variation $||\xi||_{TV} = ||\nu^* - \mu||_{TV} = R$, and $\int_\Sigma \ell(x)(\nu^* - \mu)(dx) = \frac{R}{2}\Big(M - m\Big)$. Hence, by using (II.12) as a candidate of the maximizing distribution then

$$\int_\Sigma \ell(x)\nu^*(dx) = \frac{R}{2}\Big\{\sup_{x\in\Sigma} \ell(x) - \inf_{x\in\Sigma} \ell(x)\Big\} + E_\mu(\ell) \quad \text{(II.13)}$$

Note that the first right side term in (II.13) is related to the oscillator seminorm of $f \in BM(\Sigma)$ called global modulus of continuity defined by

$$\begin{aligned}\mathrm{osc}(f) &\triangleq \sup_{(x,y)\in\Sigma\times\Sigma} |f(x) - f(y)|, \qquad f \in BM(\Sigma)\\ &= \sup_{x\in\Sigma} |f(x)| - \inf_{x\in\Sigma} |f(x)| = 2 \inf_{\alpha\in\mathbb{R}} ||f - \alpha||\end{aligned}$$

Alternatively, the pay-off $L(\nu^*)$ can be written as

$$\begin{aligned}L(\nu^*) = &\int_{\Sigma^0} M\nu^*(dx) + \int_{\Sigma_0} m\nu^*(dx)\\ &+ \int_{\Sigma\setminus\Sigma^0\cup\Sigma_0} \ell(x)\mu(dx) \quad \text{(II.14)}\end{aligned}$$

Hence, the optimal distribution $\nu^* \in B_R(\mu)$ satisfies

$$\int_{\Sigma^0} \nu^*(dx) = \mu(\Sigma^0) + \frac{R}{2} \in [0,1] \quad \text{(II.15a)}$$
$$\int_{\Sigma_0} \nu^*(dx) = \mu(\Sigma_0) - \frac{R}{2} \in [0,1] \quad \text{(II.15b)}$$
$$\nu^*(A) = \mu(A), \quad \forall A \subseteq \Sigma \setminus \Sigma^0 \cup \Sigma_0 \quad \text{(II.15c)}$$

B. Characterization of the Maximizing Measure for Finite Alphabet Spaces

Here we give a characterization of the maximizing measure for $\Sigma \triangleq \{\ell_1, \ldots, \ell_{|\Sigma|}\}$, $|\Sigma| \triangleq \mathrm{card}(\Sigma)$ finite. Define the maximum and minimum values of the sequence by

$$\ell_{\max} \triangleq \max_{i\in\Sigma} \ell_i, \qquad \ell_{\min} \triangleq \min_{i\in\Sigma} \ell_i$$

and its corresponding support sets by

$$\Sigma^0 \triangleq \{i \in \Sigma : \ell_i = \ell_{\max}\}, \qquad \Sigma_0 \triangleq \{i \in \Sigma : \ell_i = \ell_{\min}\}$$

For all remaining sequence, $\{\ell_i : i \in \Sigma \setminus \Sigma^0 \cup \Sigma_0\}$, and for $1 \le r \le |\Sigma \setminus \Sigma^0 \cup \Sigma_0|$ define recursively

$$\Sigma_k \triangleq \Big\{i \in \Sigma : \ell_i = \min\big\{\ell_\alpha : \alpha \in \Sigma \setminus \Sigma^0 \cup (\bigcup_{j=1}^{k} \Sigma_{j-1})\big\}\Big\}$$

till all the elements of Σ are exhausted, and define the corresponding value of the sequence on these sets by

$$\ell(\Sigma_k) \triangleq \min_{i\in\Sigma\setminus\Sigma^0\cup(\bigcup_{j=1}^{k}\Sigma_{j-1})} \ell_i$$

where $k \in \{1, 2, \ldots, r\}$ and r is the number of Σ_k sets which is at most $|\Sigma \setminus \Sigma^0 \cup \Sigma_0|$. For example when $k = 1$, $\ell(\Sigma_1) = \min_{i\in\Sigma\setminus\Sigma^0\cup\Sigma_0} \ell_i$ and so on.

The maximum pay-off subject to total variation constraint is

$$L(\nu^*) = \ell_{\max}\nu^*(\Sigma^0) + \ell_{\min}\nu^*(\Sigma_0) + \sum_{k=1}^{r} \ell(\Sigma_k)\nu^*(\Sigma_k)$$

The optimal probabilities are

$$\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0} \nu_i^* = \min\left(1, \sum_{i\in\Sigma^0} \mu_i + \frac{R}{2}\right) \quad \text{(II.16a)}$$
$$\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0} \nu_i^* = \left(\sum_{i\in\Sigma_0} \mu_i - \alpha\right)^+ \quad \text{(II.16b)}$$
$$\nu^*(\Sigma_k) \triangleq \sum_{i\in\Sigma_k} \nu_i^* = \left(\sum_{i\in\Sigma_k} \mu_i - \left(\alpha - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i\right)^+\right)^+ \quad \text{(II.16c)}$$
$$\alpha \triangleq \min\left(\frac{R}{2}, 1 - \sum_{i\in\Sigma^0} \mu_i\right), \qquad R \in [0,2] \quad \text{(II.16d)}$$

where $k \in \{1, 2, \ldots, r\}$, and $1 \le r \le |\Sigma \setminus \Sigma^0 \cup \Sigma_0|$ denotes the number of Σ_k sets.

III. Stochastic Control with Total Variational Distance Uncertainty

Define $\mathbb{N}_+ \triangleq \{0, 1, 2, \ldots\}, \mathbb{N}_+^n \triangleq \{0, 1, 2, \ldots, n\}, n \in \mathbb{N}_+$. The state space and the control space are sequences of Polish spaces $\{\mathcal{X}_j : j = 0, 1, \ldots, n\}$ and $\{\mathcal{U}_j : j = 0, 1, \ldots, n-1\}$, respectively. These spaces are associated with their corresponding measurable spaces $(\mathcal{X}_j, \mathcal{B}(\mathcal{X}_j)), j \in \mathbb{N}_+^n$, $(\mathcal{U}_j, \mathcal{B}(\mathcal{U}_j))$, $j \in \mathbb{N}_+^{n-1}$.

Definition 3.1: A finite horizon Feedback Control Model (FCM) is a six-tuple

$$\Big(\mathcal{X}_{0,n}, \mathcal{U}_{0,n-1}, \{\mathcal{U}_i(x_i) : x_i \in \mathcal{X}_i\}_{i=0}^{n-1},$$
$$\{Q_i(dx_i|x^{i-1}, u^{i-1}) : (x^{i-1}, u^{i-1}) \in \mathcal{X}_{0,i-1} \times \mathcal{U}_{0,i-1}\}_{i=0}^{n},$$
$$\{f_i\}_{i=0}^{n}, h_n\Big)$$

consisting of items Section I-A, (a)-(c), (f), while the controlled process distribution in (d) is replaced by the non-Markov collection $\{Q_i(dx_i|x^{i-1}, u^{i-1}) : (x^{i-1}, u^{i-1}) \in \mathcal{X}_{0,i-1} \times \mathcal{U}_{0,i-1}\}_{i=0}^{n}$.

Definition 3.2: A strategy $\pi \triangleq \{\pi : i = 0, \ldots, n-1\} \in \Pi_{0,n-1}$ is called

(a) *deterministic feedback strategy* if there exists a sequence $g \triangleq \{g_j : j = 0, 1, \ldots, n-1\}$ of measurable functions $g_j : \times_{i=0}^{j-1}\mathbb{K}_i \times \mathcal{X}_j \to \mathcal{U}_j$, such

that for all $(x^j, u^{j-1}) \in \times_{i=0}^{j-1}\mathbb{K}_i \times \mathcal{X}_j$, $j \in \mathbb{N}_+^{n-1}$, $g_j(x_0, u_0, x_1, u_1, \ldots, x_{j-1}, u_{j-1}, x_j) \in \mathcal{U}_j(x_j)$.

(b) *deterministic Markov strategy* if there exists a sequence $g \stackrel{\triangle}{=} \{g_j : j = 0, 1, \ldots, n-1\}$ of measurable functions $g_j : \mathcal{X}_j \to \mathcal{U}_j$ satisfying $g_j(x_j) \in \mathcal{U}_j(x_j)$ for all $x_j \in \mathcal{X}_j$, $j \in \mathbb{N}_+^{n-1}$.
The set of deterministic feedback strategies is denoted by $\Pi_{0,n-1}^{DF}$, and the set of deterministic Markov strategies is denoted by $\Pi_{0,n-1}^{DM}$.

A. Variational Distance Uncertainty

The nominal controlled process is described as follows.

Definition 3.3: (Nominal Controlled Process Distribution). A controlled state processes $\{x^g = x_0^g, x_1^g, \ldots, x_n^g : \pi \in \Pi_{0,n-1}^{DM}\}$ corresponds to a sequence of stochastic kernels as follows.

For every $A \in \mathcal{B}(\mathcal{X}_j)$

$$Prob(x_j \in A | x^{j-1}, u^{j-1}) = P_j(A; x_{j-1}, u_{j-1}) - a.s.$$

where $P_j(A; x_{j-1}, u_{j-1}) \in \mathcal{Q}(\mathcal{X}_j; \mathbb{K}_{j-1}), j \in \mathbb{N}_+^n$.

The uncertain controlled process is described as follows.

Definition 3.4: Given a nominal controlled process stochastic kernel of Definition 3.3, and $R_i \in [0, 2], 0 \le i \le n$ the class of uncertain controlled process distributions is defined as follows.

Markov Controlled Nominal Process. Given $P_j(\cdot; x_{j-1}, u_{j-1}) \in \mathcal{Q}(\mathcal{X}_j; \mathcal{X}_{j-1} \times \mathcal{U}_{j-1})$

$$B_{R_i}(P_i)(x^{i-1}, u^{i-1}) \stackrel{\triangle}{=} \Big\{ Q_i(\cdot; x^{i-1}, u^{i-1}) \in \mathcal{M}_1(\mathcal{X}_i) : ||Q_i(\cdot; x^{i-1}, u^{i-1}) - P_i(\cdot; x_{i-1}, u_{i-1})||_{TV} \le R_i \Big\}$$

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

B. Pay-Off Functional

For each $\pi \in \Pi_{0,n-1}^{DF}$ or $\pi \in \Pi_{0,n-1}^{DM}$ the average pay-off is defined by

$$J_{0,n}(g, \mathbb{Q}) \stackrel{\triangle}{=} \mathbb{E}_{\mathbb{Q}}\Big\{ \sum_{j=0}^{n-1} f_j(x_j^g, u_j^g) + h_n(x_n^g) \Big\} \quad \text{(III.17)}$$

where $\mathbb{E}_{\mathbb{Q}}\{\cdot\}$ denotes expectation with respect to the true joint measure $\mathbb{Q}(dx^n; u^{n-1}) \triangleq \otimes_{j=0}^{n-1} Q_j(dx_j; x^{j-1}, u^{j-1}) \in \mathcal{M}_1(\mathcal{X}_{0,n})$ belonging to the total variational distance ball of Definition 3.4.

The following assumption is introduced.

Assumption 3.5: The nominal system family satisfies the following assumption:
The maps $\{f_j : \mathcal{X}_j \times \mathcal{U}_j \longmapsto \mathbb{R} : j = 0, 1, \ldots, n-1\}$, $f_n : \mathcal{X}_n \longmapsto \mathbb{R}$ are bounded continuous and non-negative.

C. Maximization Over Total Variational 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 to deal with uncertainty of the collection of conditional distributions $\{Q_j(dx_j; x^{j-1}, u^{j-1}) \in \mathcal{M}_1(\mathcal{X}_j) : j = 0, \ldots, n\}$ belonging to the total variational distance balls of Definition 3.4.

Dynamic Programming.

Given the above formulation a minimax stochastic controlled problem can be formulated over a total variation distance uncertainty ball. The precise problem statement should thus, be as follows.

Problem 3.6: For a given $\pi \in \Pi_{0,n-1}^{DF}$ assume that the measures $M(\pi)$ induced by the true uncertainty while policy $\pi \in \Pi_{0,n-1}^{DF}$ is applied are $M(\pi) \subset \mathcal{M}_1(\mathcal{X}_{0,n})$. Given a nominal controlled process of Definition 3.3 an admissible policy set $\Pi_{0,n-1}^{DF}$ and an uncertainty class $B_{R_k}(P_k)(x^{k-1}, u^{k-1}), k = 0, 1, ..., n$ find a $\pi^* \in \Pi_{0,n-1}^{DF}$ and a sequence of stochastic kernels $Q_k^*(dx_k; x^{k-1}, u^{k-1}) \in B_{R_k}(P_k)(x^{k-1}, u^{k-1})$, $k = 0, 1, ..., n$ which solve the following minimax optimization problem.

$$J_{0,n}(\pi^*, \{Q_k^*\}_{k=0}^n) = \inf_{\pi \in \Pi_{0,n-1}^{DF}} \sup_{\substack{Q_k(\cdot; x^{k-1}, u^{k-1}) \in B_{R_k}(P_k)(x^{k-1}, u^{k-1}) \\ k=0,1,\ldots,n}} \mathbb{E}_{\mathbb{Q}}\Big\{ \sum_{k=0}^{n-1} f_k(x_k^g, u_k^g) + h_n(x_n^g) \Big\} \quad \text{(III.18)}$$

Define the pay-off associated with the maximization problem

$$J_{0,n}(\pi, \{Q_i^*\}_{i=0}^n) \stackrel{\triangle}{=} \sup_{\substack{Q_k(\cdot; x^{k-1}, u^{k-1}) \in B_{R_k}(P_k)(x^{k-1}, u^{k-1}) \\ k=0,1,\ldots,n}} J_{0,n}(g, \{Q_k\}_{k=0}^n) \quad \text{(III.19)}$$

For a given $\pi \in \Pi_{0,n-1}^{DF}$, which define $\{g_j : j = 0, \ldots, n-1\}$, and $\pi_{[k,m]} \equiv u_{[k,m]}^g$, denoting the restriction of policies in $[k, m]$, $0 \le k \le m \le n-1$ define the conditional expectation taken over the events $\mathcal{G}_{0,j} \stackrel{\triangle}{=} \sigma\{x_0^g, \ldots, x_j^g, u_0^g, \ldots, u_{j-1}^g\}$ maximized over the class $B_{R_k}(P_k)(x^{k-1}, u^{k-1})$, $k = j+1, \ldots, n$, which is the value function of (III.19) as follows:

$$V_j(u_{[j,n-1]}^g, \mathcal{G}_{0,j}) \stackrel{\triangle}{=} \sup_{\substack{Q_k(\cdot; x^{k-1}, u^{k-1}) \in B_{R_k}(P_k)(x^{k-1}, u^{k-1}) \\ k=j+1,\ldots,n}} E_{\mathbb{Q}}\Big\{ \sum_{k=j}^{n-1} f_k(x_k^g, u_k^g) + h_n(x_n^g) | \mathcal{G}_{0,j} \Big\} \quad \text{(III.20)}$$

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

$$V_j(u^g_{[j,n-1]}, \mathcal{G}_{0,j}) = \sup_{Q_{j+1}(\cdot;x^j,u^j)\in B_{R_{j+1}}(P_{j+1})(x^j,u^j)} E_{Q_{j+1}(\cdot;x^j,u^j)}\Big\{ f_j(x^g_j, u^g_j) + V_{j+1}(u^g_{[j+1,n-1]}, \mathcal{G}_{0,j+1})\Big\} \tag{III.21}$$

$$V_n(\mathcal{G}_{0,n}) = h_n(x^g_n) \tag{III.22}$$

where $E_{Q_{j+1}(\cdot;x^j,u^j)}$ denotes expectation with respect to $Q_{j+1}(dx_{j+1}; x^j, u^j)$, $j = 0, \ldots, n-1$.

Note that for finite or countable spaces $\{\mathcal{X}_i : i = 0, \ldots, n-1\}$, $\{\mathcal{U}_i : i = 0, \ldots, n-1\}$ the maximizing distribution is obtained for any $R_j \in [0, 2], j = 0, \ldots, n$ via the analog of the maximizing solution given in (II.16).

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

$$V_j(\mathcal{G}_{0,j}) \stackrel{\triangle}{=} \inf_{\pi\in\Pi^{DF}_{j,n-1}} \sup_{\substack{Q_k(\cdot;x^{k-1},u^{k-1})\in B_{R_k}(P_k)(x^{k-1},u^{k-1}) \\ k=j+1,\ldots,n}} E_{\mathbb{Q}}\Big\{ \sum_{k=j}^{n-1} f_k(x^g_k, u^g_k) + h_n(x^g_n)|\mathcal{G}_{0,j}\Big\} \tag{III.23}$$

$$= \inf_{\pi\in\Pi^{DF}_{j,n-1}} V_j(u^g_{[j,n-1]}, \mathcal{G}_{0,j}) \tag{III.24}$$

Hence, the following dynamic programming recursion

$$V_j(\mathcal{G}_{0,j}) \stackrel{\triangle}{=} \inf_{u_j\in\mathcal{U}_j(x)} \sup_{Q_j(\cdot;x^{j-1},u^{j-1})\in B_{R_j}(P_j)(x^{j-1},u^{j-1})} E_{Q_{j+1}(\cdot;x^j,u^j)}\Big\{ f_j(x^g_j, u^g_j) + V_{j+1}(\mathcal{G}_{0,j+1})|\mathcal{G}_{0,j}\Big\} \tag{III.25}$$

$$V_n(\mathcal{G}_{0,n}) = h_n(x^g_n) \tag{III.26}$$

In view of Section II-A, specifically, the relation between the maximizing distribution and the nominal distribution (II.13), (II.15), (II.16), which also apply for conditional distributions, then one can extract that the maximization conditional distribution $Q^*_i(dx_i; x^{i-1}, u^{i-1})$ is Markovian, hence $Q^*_i(dx_i; x^{i-1}, u^{i-1}) = Q^*_i(dx_i; x_{i-1}, u_{i-1})$ $-$ $a.s.$. Utilizing this observation and the dynamic programming equations for the FCM the following theorem is obtained.

Theorem 3.7: Consider the class of Markov Controlled Process distribution of Definition 3.4. Then $V_j(\mathcal{G}_{0,j}) = V_j(x_j)$ satisfies the following dynamic programming recursion

$$V_j(x) \stackrel{\triangle}{=} \inf_{u\in\mathcal{U}(x)} \sup_{Q_{j+1}(\cdot;x,u)\in B_{R_{j+1}}(P_{j+1})(x,u)} E_{Q_{j+1}(\cdot;x,u)}\Big\{ f_j(x,u) + V_{j+1}(x_{j+1})\Big\}, \quad x \in \mathcal{X}_j \tag{III.27}$$

$$V_n(x) = h_n(x), \quad x \in \mathcal{X}_n \tag{III.28}$$

Also,

$$J_{0,n}(g^*, \{Q^*_i\}^n_{i=0}) = \sup_{Q_0(\cdot)\in B_{R_0}(P_0)} E_{Q_0(\cdot)}\Big\{ V_0(x_0)\Big\}$$

Assume $V_{j+1}(\cdot) : \mathcal{X}_{j+1} \to [0, \infty)$ is bounded continuous in $x \in \mathcal{X}_{j+1}$.
Further, assume

$$P_{j+1}(\Sigma^0_{j+1}; x_j, u_j) + \frac{R_{j+1}}{2} \in [0, 1] \tag{III.29}$$

$$P_{j+1}(\Sigma^0_{j+1}; x_j, u_j) - \frac{R_{j+1}}{2} \in [0, 1] \tag{III.30}$$

where

$$\mathcal{X}^0_{j+1} \stackrel{\triangle}{=} \Big\{ x_{j+1} \in \mathcal{X}_{j+1} : V_{j+1}(x_{j+1}) = \sup\{V_{j+1}(x_{j+1}) : x_{j+1} \in \mathcal{X}_{j+1}\}\Big\}$$

$$\mathcal{X}_{j+1,0} \stackrel{\triangle}{=} \Big\{ x_{j+1} \in \mathcal{X}_{j+1} : V_{j+1}(x_{j+1}) = \inf\{V_{j+1}(x_{j+1}) : x_{j+1} \in \mathcal{X}_{j+1}\}\Big\}$$

Then the dynamic programming recursion is given by

$$V_j(x) \stackrel{\triangle}{=} \inf_{u\in\mathcal{U}(x)} \Big\{ f_j(x,u) + \int_{\mathcal{X}_{j+1}} V_{j+1}(z) P_{j+1}(dz; x, u) + \frac{R_j}{2}\big\{ \sup_{z\in\mathcal{X}_{j+1}} V_{j+1}(z) - \inf_{z\in\mathcal{X}_{j+1}} V_{j+1}(z)\big\}\Big\}, \; x \in \mathcal{X}_j \tag{III.31}$$

$$V_n(x) = h_n(x), \; x \in \mathcal{X}_n \tag{III.32}$$

Proof. Utilize [4].

Remark 3.8: Some observations are discussed.

1) The dynamic programming equation (III.31), (III.32) involves in its right hand side the oscillator seminorm of $V_{j+1}(\cdot)$. To the best of our knowledge, this form of dynamic programming recursion has not appeared in the literature.
2) It is concluded that dynamic programming recursion such as (III.31), (III.32) can be derived for the partially observed case, finite or countable state Markov Decision case, etc. These are subject to future investigations.

REFERENCES

[1] 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, Accepted, pp.42, 2010.

[2] C.D. Charalambous, I. Tzortzis, and F. Rezaei. *Stochastic Optimal Control of Discrete-Time Systems Subject to Conditional Distribution Uncertainty*. In 2011 Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, Florida, December 12-15, 2011.

[3] 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.

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

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

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