



Infinite Horizon Average Cost DP Subject to Ambiguity on Conditional Distribution

I. Tzortzis*, **C. D. Charalambous*** and T. Charalambous[‡]



*Department of Electrical and Computer Engineering, University of Cyprus

[‡]Department of Signals and Systems, Chalmers University of Technology



- 1 Introduction - Motivation
- 2 Maximization with TV Distance Ambiguity
- 3 Minimax Stochastic Control
- 4 Example
- 5 Conclusions - Future Work



Outline



- 1 Introduction - Motivation
- 2 Maximization with TV Distance Ambiguity
- 3 Minimax Stochastic Control
- 4 Example
- 5 Conclusions - Future Work



Introduction - Motivation I



General Problem

Address optimality of stochastic control strategies based on infinite horizon minimax Average Cost (AC) criterion via DP subject to Total Variation (TV) distance ambiguity on conditional distribution of controlled process

Methodology:

- Formulate stochastic control problem using minimax theory
 - controller minimizes pay-off
 - conditional distribution maximizes pay-off
- Maximization of linear functionals on the space of probability meas.
- Minimax stochastic control under a Markovian assumption



Introduction - Motivation II



Infinite Horizon Markov Control Model

$$\text{MCM: } (\mathcal{X}, \mathcal{U}, \{\mathcal{U}(x): x \in \mathcal{X}\}, \{Q(dz|x, u): (x, u) \in \mathcal{X} \times \mathcal{U}\}, f)$$

- (a) \mathcal{X} : State Space
- (b) \mathcal{U} : Control Space
- (c) $\mathcal{U}(x)$: Feasible Controls
- (d) $Q(dz|x, u)$: Controlled Process
- (e) f : One-Stage-Cost

Infinite horizon Average Cost Criterion

$$J^0(g, x) \triangleq \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_x^g \left\{ \sum_{k=0}^{n-1} f(x_k, u_k) \right\}$$



Introduction - Motivation III



Markov Control Problem (MCP)

Find a control policy $g^* \in G$ such that

$$J^0(g^*, x) \triangleq \inf_{g \in G} J^0(g, x) = J^{0,*}(x), \quad \forall x \in \mathcal{X}$$

Under appropriate conditions:

- \mathcal{X}, \mathcal{U} are of finite cardinality
- f is non-negative
- For all $g \in G_{SM}$, the controlled process distribution $Q(z|x, u)$ is an irreducible stochastic matrix



Introduction - Motivation IV



Classical DP equation¹

There exists a pair $(J^{0,*}, V^0(x))$, which is the solution of the DP of the infinite horizon MCP

$$J^{0,*} + V^0(x) = \inf_{u \in \mathcal{U}(x)} \left\{ f(x, u) + \sum_{z \in \mathcal{X}} Q(z|x, u) V^0(z) \right\}, \quad x \in \mathcal{X}$$

Remark

- The AC optimality criterion and the DP equation depend on the complete knowledge of $Q(\cdot|x, u)$
- Any mismatch affects the optimality and robustness of the optimal decision strategies

¹[Kumar-Varaiya86][Lerma-Lasserre96][Bertsekas05]



Outline



- 1 Introduction - Motivation
- 2 Maximization with TV Distance Ambiguity
- 3 Minimax Stochastic Control
- 4 Example
- 5 Conclusions - Future Work



Maximization with TV Distance I

Problem Formulation



Notation:

- 1 (Σ, d_Σ) : complete, separable metric space
- 2 $(\Sigma, \mathcal{B}(\Sigma))$: corresponding measurable space
- 3 $\mathcal{M}_1(\Sigma)$: set of probability measures on $\mathcal{B}(\Sigma)$
- 4 $BC^+(\Sigma)$: space of bounded continuous non-negative functions

Nominal and Class of Measures

- Nominal measure: $\mu \in \mathcal{M}_1(\Sigma)$
- Class of measures: $\nu \in \mathcal{M}_1(\Sigma)$ belongs to set

$$\mathbb{B}_R(\mu) = \{\nu \in \mathcal{M}_1(\Sigma) : \|\nu - \mu\|_{TV} \leq R\}, \quad R \geq 0$$



Maximization with TV Distance II

Characterization of Maximizing Measure



Average payoff over $\mathbb{B}_R(\mu)$

$$D^+(R) = \sup_{\nu \in \mathbb{B}_R(\mu)} \mathbb{L}(\nu) \equiv \sup_{\nu \in \mathbb{B}_R(\mu)} \int_{\Sigma} \ell(x) \nu(dx), \quad \ell \in BC^+(\Sigma)$$

- For a given $\mu \in \mathcal{M}_1(\Sigma)$ and $\nu \in \mathbb{B}_R(\mu)$ define the set

$$\tilde{\mathbb{B}}_R(\mu) \triangleq \left\{ \xi \in \mathbb{M}_0(\Sigma) : \xi = \nu - \mu, \nu \in \mathcal{M}_1(\Sigma), \|\xi\|_{TV} \leq R \right\}$$

Solution of Maximization

$$\begin{aligned} \int_{\Sigma} \ell(x) \nu(dx) &= \int_{\Sigma} \ell(x) (\xi^+(dx) - \xi^-(dx)) + \int_{\Sigma} \ell(x) \mu(dx) \\ &\leq \left\{ \sup_{x \in \Sigma} \ell(x) - \inf_{x \in \Sigma} \ell(x) \right\} \frac{\|\xi\|_{TV}}{2} + \int_{\Sigma} \ell(x) \mu(dx) \end{aligned}$$



Maximization with TV Distance III

Characterization of Maximizing Measure



- The upper bound is achieved by $\xi^* \in \widetilde{\mathbb{B}}_R(\mu)$ as follows. Let

$$x^0 \in \Sigma^0 \triangleq \left\{ x \in \bar{\Sigma} : l(x) = \sup\{l(y) : y \in \Sigma\} \right\}$$

$$x_0 \in \Sigma_0 \triangleq \left\{ x \in \bar{\Sigma} : l(x) = \inf\{l(y) : y \in \Sigma\} \right\}$$

- Take $\xi^*(dx) = \nu^*(dx) - \mu(dx) = \frac{R}{2} \left(\delta_{x^0}(dx) - \delta_{x_0}(dx) \right)$

$$\int_{\Sigma} l(x) \nu^*(dx) = \frac{R}{2} \left\{ \sup_{x \in \Sigma} l(x) - \inf_{x \in \Sigma} l(x) \right\} + \int_{\Sigma} l(x) \mu(dx)$$

- The first right side term is related to the oscillator seminorm of a non-negative function



Maximization with TV Distance IV

Characterization of Maximizing Measure



Extremum Pay-Off

$$\int_{\Sigma} \ell(x) \nu^*(dx) = \int_{\Sigma^0} \ell_{\max} \nu^*(dx) + \int_{\Sigma_0} \ell_{\min} \nu^*(dx) + \int_{\Sigma \setminus \Sigma^0 \cup \Sigma_0} \ell(x) \mu(dx)$$

Optimal Distribution² $\nu^* \in \mathbb{B}_R(\mu)$

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

² [C.D. Charalambous et al. IEEETAC-14]



Outline



- 1 Introduction - Motivation
- 2 Maximization with TV Distance Ambiguity
- 3 Minimax Stochastic Control
- 4 Example
- 5 Conclusions - Future Work



Problem Formulation I

The MiniMax Optimization Problem



Nominal Controlled Process Distribution

For each $g \in G_{SM}$,

$$\text{Prob}(x_t \in A | x^{t-1}, u^{t-1}) = Q^o(A | x_{t-1}, u_{t-1}), \quad A \in \mathcal{B}(\mathcal{X})$$

Class of Uncertain Controlled Process Distributions

Given $Q^o(\cdot | x_{t-1}, u_{t-1})$ and $R \in [0, 2]$,

$$B_R(Q^o)(x, u) \triangleq \left\{ Q(\cdot | x, u) : \|Q(\cdot | x, u) - Q^o(\cdot | x, u)\|_{TV} \leq R \right\}$$

Assumption:

- The map $f : \mathcal{X} \times \mathcal{U} \mapsto \mathbb{R}$ is bounded continuous and non-negative



Problem Formulation II

The MiniMax Optimization Problem



Average cost per unit-time for a class

Infinite horizon AC per unit-time when policy $g \in G$ is used,

$$\begin{aligned}
 J(g, x) &\triangleq \limsup_{n \rightarrow \infty} \frac{1}{n} J_n(g, x) \\
 &= \limsup_{n \rightarrow \infty} \frac{1}{n} \left\{ \sup_{Q(\cdot|x, u) \in \mathbb{B}_R(Q^o)(x, u)} \mathbb{E}_x^{g, Q} \left[\sum_{k=0}^{n-1} f(x_k, u_k) \right] \right\}
 \end{aligned}$$

MiniMax MCP

Find a control policy $g^* \in G$ such that

$$J(g^*, x) \triangleq \inf_{g \in G} J(g, x) = J^*(x), \quad \forall x \in \mathcal{X}$$



Problem Solution I

The General Dynamic Programming



n-stage cost

The maximization of the expected n-stage cost is given by

$$\begin{aligned}
 J_n(g, q_0) &= \max_{Q(\cdot|x, u) \in \mathbb{B}_R(Q^o)(x, u)} q_0^T \left\{ \sum_{k=0}^{n-1} Q(g)^k \right\} f(g) \\
 &= q_0^T \left\{ \sum_{k=0}^{n-1} Q^*(g)^k \right\} f(g)
 \end{aligned}$$

- $Q^*(g)$: maximizing conditional distribution

- The infinite horizon AC per unit-time is then

$$J(g, q_0) = \limsup_{n \rightarrow \infty} \frac{1}{n} q_0^T \left\{ \sum_{k=0}^{n-1} Q^*(g)^k \right\} f(g)$$



Problem Solution II

The General Dynamic Programming



Assumption

If for all $g \in G_{SM}$, and $R \in [0, R_{\max}] \in [0, 2]$, the maximizing conditional distribution $Q^*(g)$ is an irreducible stochastic matrix, then

$$J(g, q_0) = q(g)^T f(g)$$

- $q(g)$: unique invariant distribution
- $J(g, q_0) \equiv J(g)$: independent of the initial distribution



Problem Solution III

The General Dynamic Programming



Main DP Theorem

- 1 There exists a solution (V, J^*) to the DP equation

$$J^* + V(x) = \min_{u \in \mathcal{U}} \left\{ f(x, u) + \sum_{z \in \mathcal{X}} Q^*(z|x, u) V(z) \right\}$$

or, equivalently, to the DP equation

$$J^* + V(x) = \min_{u \in \mathcal{U}} \left\{ f(x, u) + \sum_{z \in \mathcal{X}} Q^o(z|x, u) V(z) + \frac{R}{2} \left(\max_{z \in \mathcal{X}} V(z) - \min_{z \in \mathcal{X}} V(z) \right) \right\}$$

- *Oscillator seminorm of the cost-to-go*

- 2 If $g^*(x)$ attains the minimum for every x , then g^* is an AC optimal policy, and the minimum cost is J^*



Problem Solution IV

The Policy Iteration Algorithm



Policy Iteration Algorithm

- Classical policy iteration algorithm performed using the nominal conditional distribution
 - A modified algorithm for average cost DP is proposed based on TV distance ambiguity
-
- Policy improvement and evaluation steps performed using the maximizing conditional distribution as follows:
 - a) Solve DP equation based on nominal conditional distribution
 - b) Identify the support sets and calculate the maximizing distribution
 - c) Solve the general DP equation



Outline



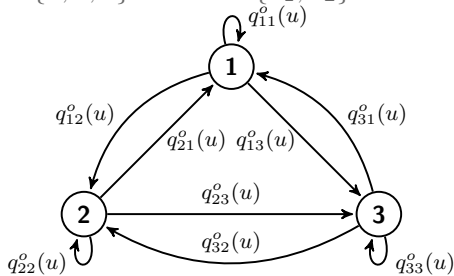
- 1 Introduction - Motivation
- 2 Maximization with TV Distance Ambiguity
- 3 Minimax Stochastic Control
- 4 **Example**
- 5 Conclusions - Future Work



Infinite Horizon MCM Example I



- MCM: $\mathcal{X} = \{1, 2, 3\}$ and $\mathcal{U} = \{u_1, u_2\}$



Nominal transition probabilities

- Under Control u_1 :

$$Q^o(u_1) = \frac{1}{9} \begin{pmatrix} 3 & 1 & 5 \\ 4 & 2 & 3 \\ 1 & 6 & 2 \end{pmatrix}$$

- Under Control u_2 :

$$Q^o(u_2) = \frac{1}{9} \begin{pmatrix} 1 & 2 & 6 \\ 4 & 2 & 3 \\ 4 & 1 & 4 \end{pmatrix}$$



Infinite Horizon MCM Example II



Cost Function for each State

- Under control u_1 : $f(1, u_1) = 2$, $f(2, u_1) = 1$, $f(3, u_1) = 3$
- Under control u_2 : $f(1, u_2) = .5$, $f(2, u_2) = 3$, $f(3, u_2) = 0$

Policy iteration algorithm ($R = 6/9$)

A. Let $m = 0$:

- (1) Initial policies: $g_0(1) = u_1$, $g_0(2) = u_2$, $g_0(3) = u_2$
- (2) **(PE)** Solve $J_{Q^0}(g_0)e + V_{Q^0}(g_0) = f(g_0) + Q^0(g_0)V_{Q^0}(g_0)$,

$$\begin{pmatrix} V_{Q^0}(g_0, 1) \\ V_{Q^0}(g_0, 2) \\ V_{Q^0}(g_0, 3) \end{pmatrix} = \begin{pmatrix} 1.8 \\ 3.375 \\ 0 \end{pmatrix}, \quad J_{Q^0}(g_0) = 1.175.$$

Identify the support sets $\mathcal{X}^+ = \{2\}$, $\mathcal{X}^- = \{3\}$, $\mathcal{X}_1 = \{1\}$
and calculate $Q^*(u_1)$ and $Q^*(u_2)$, i.e.,



Infinite Horizon MCM Example III



$$Q^*(u_1) = \begin{pmatrix} \left(q_{11}^o(u_1) - \left(\frac{R}{2} - q_{13}^o(u_1) \right)^+ \right)^+ & \min \left(1, q_{12}^o(u_1) + \frac{R}{2} \right) & \left(q_{13}^o(u_1) - \frac{R}{2} \right)^+ \\ \left(q_{21}^o(u_1) - \left(\frac{R}{2} - q_{23}^o(u_1) \right)^+ \right)^+ & \min \left(1, q_{22}^o(u_1) + \frac{R}{2} \right) & \left(q_{23}^o(u_1) - \frac{R}{2} \right)^+ \\ \left(q_{31}^o(u_1) - \left(\frac{R}{2} - q_{33}^o(u_1) \right)^+ \right)^+ & \min \left(1, q_{32}^o(u_1) + \frac{R}{2} \right) & \left(q_{33}^o(u_1) - \frac{R}{2} \right)^+ \end{pmatrix}$$

and similarly for $Q^*(u_2)$.

- $Q^*(u)$ remains irreducible!

Solve $J_{Q^*}(g_0)e + V_{Q^*}(g_0) = f(g_0) + Q^*(g_0)V_{Q^*}(g_0)$. The solution is

$$\begin{pmatrix} V_{Q^*}(g_0, 1) \\ V_{Q^*}(g_0, 2) \\ V_{Q^*}(g_0, 3) \end{pmatrix} = \begin{pmatrix} 1.8 \\ 3.375 \\ 0 \end{pmatrix}, \quad J_{Q^*}(g_0) = 2.3.$$



Infinite Horizon MCM Example IV



(3) **(PI)** Calculate $g_1 = \operatorname{argmin}_{g \in \mathbb{R}^3} \{f(g) + Q^*(g)V_{Q^*}(g_0)\}$,

$$g_1(1) = u_2, \quad g_1(2) = u_1, \quad g_1(3) = u_2$$

B. Let $m = 1$: Following the same procedure as in step A we obtain

$$\begin{pmatrix} V_{Q^*}(g_1, 1) \\ V_{Q^*}(g_1, 2) \\ V_{Q^*}(g_1, 3) \end{pmatrix} = \begin{pmatrix} 0.468 \\ 1.125 \\ 0 \end{pmatrix}, \quad J_{Q^*}(g_1) = 0.708.$$

and $g_2(1) = u_2, g_2(2) = u_1, g_2(3) = u_2$.

Since, $g_2 = g_1$, then $g^* = g_1$ is an optimal control policy with minimum cost $J_{Q^*} = 0.708$.



Outline



- 1 Introduction - Motivation
- 2 Maximization with TV Distance Ambiguity
- 3 Minimax Stochastic Control
- 4 Example
- 5 Conclusions - Future Work



Conclusions - Future Work



Conclusions

- New infinite horizon average cost DP equation, and the corresponding policy iteration algorithm

Future Work

- $(\mathcal{X}, \mathcal{U})$: Borel spaces
- $R \in [R_{\max}, 2]$: irreducibility condition is violated
- Develop multichain DP equations

$$J^*(x) = \inf_{u \in \mathcal{U}(x)} \left\{ \int_{\mathcal{X}} Q^\circ(dz|x, u) J^*(z) + \frac{R}{2} \left(\sup_{z \in \mathcal{X}} J^*(z) - \inf_{z \in \mathcal{X}} J^*(z) \right) \right\}$$

$$J^*(x) + V(x) = \inf_{u \in \mathcal{U}(x)} \left\{ f(x, u) + \int_{\mathcal{X}} Q^\circ(dz|x, u) V(z) + \frac{R}{2} \left(\sup_{z \in \mathcal{X}} V(z) - \inf_{z \in \mathcal{X}} V(z) \right) \right\}$$



References



-  **P.R. Kumar and P. Varaiya.**
Stochastic Systems: Estimation, Identification and Adaptive Control.
Prentice Hall, 1986.
-  **O. Hernandez-Lerma and J. B. Lasserre.**
Discrete-Time Markov Control Processes: Basic Optimality Criteria.
Springer Verlag, 1996.
-  **D.P. Bertsekas.**
Dynamic Programming and Optimal Control.
Athena Scientific, 2005.
-  **C.D. Charalambous, I.Tzortzis, S.Loyka and T.Charalambous.**
Extremum Problems with Total Variation Distance and their Applications.
IEEE TAC, Volume 59(9), pp.2353–2368, 2014.
-  **I.Tzortzis, C.D. Charalambous and T.Charalambous.**
Average Cost Dynamic Programming Subject to Ambiguity.
SICON, 2015 (submitted).