Dynamic Programming with Total Variational Distance Uncertainty

Charalambos D. Charalambous[@] and Ioannis Tzortzis[@]
and Themistoklis Charalambous[†]

University of Cyprus, Department of Electrical and Computer Engineering[@]
Royal Institute of Technology, School of Electrical Engineering[†]

51st IEEE Conference on Decision and Control
December 10-13, 2012, Maui, Hawaii

Outline

Outline

Introduction I

General Problem

- Model Uncertainty via Total Variational Distance between Measures
- Optimization of Stochastic Uncertain Systems via Dynamic Programming subject to Total Variational Distance Uncertainty
- Minimax Dynamic Programming Recursions for Markov Chain Models

Motivation I

Markov Control Model (MCM)

A finite horizon MCM is a six-tuple

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

Consisting of

1. State Space
2. Control or Action Space
3. Feasible Controls or Actions
4. Controlled Process Distribution
5. Cost-Per-Stage
6. Terminal Cost

Motivation II

Pay-Off Functional

Choose a control policy $g \stackrel{\triangle}{=} \{g_j : j = 0, 1, \ldots, n-1\}$ so as to minimize the pay-off functional

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

Remarks

1. Pay-off is a functional of conditional distributions $\{Q_i(\cdot|\cdot)\}_{i=0}^n$.
2. Classical DP recursion depend on complete knowledge of conditional distributions.
3. Any mismatch from true conditional distributions affect the optimality of control strategies.

Motivation III

DP Under Uncertainty of Conditional Distributions

Uncertainty of true controlled process via Total Variational (TV) distance:

$$B_{R_i}(P_i)(x_{i-1}, u_{i-1}) \triangleq \Big\{ Q_i(\cdot|x_{i-1}, u_{i-1}) : \\ ||Q_i(\cdot|x_{i-1}, u_{i-1}) - P_i(\cdot|x_{i-1}, u_{i-1})||_{TV} \leq R_i \Big\}$$

- $\{P_i(dx_i|x_{i-1}, u_{i-1})\}$: Nominal controlled process distributions
- $\{Q_i(dx_i|x_{i-1}, u_i)\}$: True controlled process distributions
- $R_i \in [0, 2]$

Remark

The uncertainty model is general, and includes linear, non-linear, finite and/or countable state space models, etc.

Outline

Problem Formulation I

Notation

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

Total Variational Distance

Total variational distance is a metric $||\cdot||_{TV} : \mathcal{M}_1(\Sigma) \times \mathcal{M}_1(\Sigma) \mapsto [0, \infty)$ defined by

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

where $\alpha, \beta \in \mathcal{M}_1(\Sigma)$ and $\mathcal{P}(\Sigma)$ denotes the collection of all finite partitions of Σ.

Problem Formulation II

Nominal and Uncertain System

- The nominal system is a fixed nominal probability measure $\mu \in \mathcal{M}_1(\Sigma)$.
- The uncertain system $\nu \in \mathcal{M}_1(\Sigma)$ belongs to the set

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

Maximization

The uncertain system measure tries to maximize the average pay-off functional over the set $\mathbb{B}_R(\mu)$ for a given $\mu \in \mathcal{M}_1(\Sigma)$, defined by

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

Characterization of Maximizing Measure I

1. Let $\mathcal{M}_{sm}(\Sigma)$ denote the set of finite signed measures
2. Define $\mathbb{M}_0(\Sigma) \triangleq \Big\{\eta \in \mathcal{M}_{sm}(\Sigma) : \eta(\Sigma) = 0\Big\}$
3. The total variation of $\xi \in \mathbb{M}_0(\Sigma)$ is $||\xi||_{TV} = \xi^+(\Sigma) + \xi^-(\Sigma)$, where, $\{\xi^+, \xi^-\}$ is the Hanh-Jordan decomposition of ξ into $\xi = \xi^+ - \xi^-$
4. $\xi(\Sigma) = 0$ implies $\xi^+(\Sigma) = \xi^-(\Sigma) = \frac{||\xi||_{TV}}{2}$
5. Define $\xi \triangleq \nu - \mu \in \mathbb{M}_0(\Sigma)$ and let $\ell \in BC^+(\Sigma)$

Characterization of Maximizing Measure II

Solution of Maximization

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

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

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

Characterization of Maximizing Measure III

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 $\xi^*(dx) = \nu^*(dx) - \mu(dx) = \frac{R}{2}\Big(\delta_{x^0}(dx) - \delta_{x_0}(dx)\Big)$
- Using above equation as a candidate of the maximizing distribution

$$\int_\Sigma \ell(x)\nu^*(dx) = \frac{R}{2}\Big\{\sup_{x\in\Sigma}\ell(x) - \inf_{x\in\Sigma}\ell(x)\Big\} + E_\mu(\ell) \qquad (1)$$

1. ν^* satisfies the constraint $||\xi^*||_{TV} = ||\nu^* - \mu||_{TV} = R$.
2. $\nu^*(\Sigma) = 1$ and $0 \le \nu^*(A) \le 1$ on any $A \in \mathcal{B}(\Sigma)$.

Characterization of Maximizing Measure IV

Average Pay-Off, $\mathbb{L}(\ell, \nu^*)$

Can be written as

$$L(\nu^*) \quad = \int_{\Sigma^0} M\nu^*(dx) + \int_{\Sigma_0} m\nu^*(dx) + \int_{\Sigma \setminus \Sigma^0 \cup \Sigma_0} \ell(x)\mu(dx)$$

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

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

for any $R \in [0, 2]$.

Finite Alphabet Case I

- $|\Sigma| \triangleq \text{card}(\Sigma)$ finite
- Let $\ell \triangleq \{\ell_1, \ldots, \ell_{|\Sigma|}\}$ denote a non-negative sequence of real numbers.
- Define $\xi_i \triangleq \nu_i - \mu_i, i = 1, \ldots, |\Sigma|$

Maximization Problem

Then

$$\max_{\nu \in \mathbb{B}_R(\mu)} \sum_{i \in \Sigma} \ell_i \nu_i \longrightarrow \sum_{i \in \Sigma} \ell_i \mu_i + \max_{\xi \in \widetilde{\mathbb{B}}_R(\mu)} \sum_{i \in \Sigma} \ell_i \xi_i$$

and $\sum_{i \in \Sigma} |\xi_i| \leq R, \quad \sum_{i \in \Sigma} \xi_i = 0, \quad 0 \leq \xi_i + \mu_i \leq 1, \quad \forall i \in \Sigma$

- **Total Variation** $\alpha = \sum_{i \in \Sigma} |\xi_i| = \sum_{i \in \Sigma} \xi_i^+ + \sum_{i \in \Sigma} \xi_i^- \leq R$
- **Average Pay-Off** $\sum_{i \in \Sigma} \ell_i \xi_i = \sum_{i \in \Sigma} \ell_i \xi_i^+ - \sum_{i \in \Sigma} \ell_i \xi_i^-$

Finite Alphabet Case II

Computation of the Partition of Σ

- Define the maximum and minimum values of the sequence and its corresponding support sets by

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

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

- For all $k = 1, 2, \ldots, |\Sigma \setminus \Sigma^0 \cup \Sigma_0|$ define recursively

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

and the corresponding value of the sequence on these sets

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

Finite Alphabet Case III

Maximum Pay-Off

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

where the optimal probabilities are given by

- $\nu^*(\Sigma^0) \triangleq \sum_{i\in\Sigma^0} \nu_i^* = \min\left(1, \sum_{i\in\Sigma^0} \mu_i + \frac{R}{2}\right)$
- $\nu^*(\Sigma_0) \triangleq \sum_{i\in\Sigma_0} \nu_i^* = \left(\sum_{i\in\Sigma_0} \mu_i - \alpha\right)^+$
- $\nu^*(\Sigma_k) \triangleq \sum_{i\in\Sigma_k} \nu_i^* = \left(\sum_{i\in\Sigma_k} \mu_i - \left(\alpha - \sum_{j=1}^{k} \sum_{i\in\Sigma_{j-1}} \mu_i\right)^+\right)^+$
- $\alpha = \min(\frac{R}{2}, 1 - \sum_{i\in\Sigma^0} \mu_i), \quad R \in [0,2]$
- $k = 1, 2, \ldots, r$ and r denotes the number of Σ_k sets.

Outline

Problem Formulation I

Nominal Controlled Process

A controlled state processes corresponds to a sequence of stochastic kernels as follows. For every $A \in \mathcal{B}(\mathcal{X}_j)$

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

Uncertain Controlled Process

Given $P_j(\cdot; x_{j-1}, u_{j-1})$

$$B_{R_i}(P_i)(x_{i-1}, u_{i-1}) \triangleq \Big\{ Q_i(\cdot; x_{i-1}, u_{i-1}) \in \mathcal{M}_1(\mathcal{X}_i) : \\ ||Q_i(\cdot; x_{i-1}, u_{i-1}) - P_i(\cdot; x_{i-1}, u_{i-1})||_{TV} \leq R_i \Big\}$$

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

Problem Formulation II

Assumption

The nominal system family satisfies: 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.

Problem Formulation III

MiniMax Stochastic Controlled Problem

Given a nominal controlled process an admissible feedback policy set $\Pi_{0,n-1}^{DF}$ and an uncertainty class $B_{R_k}(P_k)(x_{k-1}, u_{k-1})$, find a $\pi^* \in \Pi_{0,n-1}^{DF}$ and a sequence of stochastic kernels $Q_k^*(dx_k; x_{k-1}, u_{k-1}) \in B_{R_k}(P_k)(x_{k-1}, u_{k-1})$, which solve the following minimax optimization problem.

$$J_{0,n}(\pi^*, \{Q_k^*\}_{k=0}^n) = \inf_{\pi \in \Pi_{0,n-1}^{DF}} \sup_{\substack{Q_k(\cdot; x_{k-1}, u_{k-1}) \in B_{R_k}(P_k)(x_{k-1}, u_{k-1}) \\ k=0,1,\ldots,n}} \mathbb{E}_{\mathbb{Q}} \left\{ \sum_{k=0}^{n-1} f_k(x_k^g, u_k^g) + h_n(x_n^g) \right\}$$

Principle of Optimality I

MiniMax Cost-to-Go $V_j(\mathcal{G}_{0,j})$ during $\{j, j+1, ..., n\}$

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

where $\mathcal{G}_{0,j} \stackrel{\triangle}{=} \sigma\{x_0^g, \ldots, x_j^g, u_0^g, \ldots, u_{j-1}^g\}$

Assumptions

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

Principle of Optimality II

General Dynamic Programming Recursion

By (1),

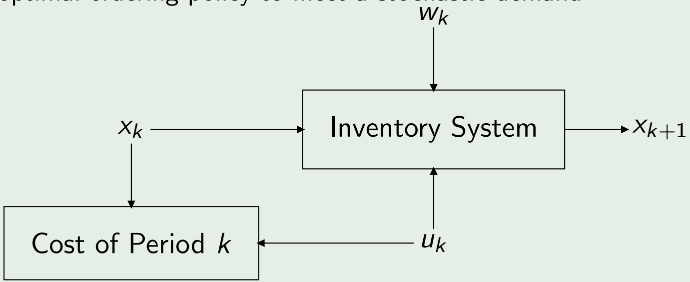
$$
\begin{aligned}
V_j(x) \triangleq& \inf_{u \in \mathcal{U}(x)} \Big\{ f_j(x,u) + \int_{\mathcal{X}_{j+1}} V_{j+1}(z) P_{j+1}(dz; x, u) \\
&+ \frac{R_j}{2} \Big\{ \sup_{z \in \mathcal{X}_{j+1}} V_{j+1}(z) - \inf_{z \in \mathcal{X}_{j+1}} V_{j+1}(z) \Big\} \Big\}, \ x \in \mathcal{X}_j \\
V_n(x) =& h_n(x), \ x \in \mathcal{X}_n
\end{aligned}
$$

- Next we give an example.

Example I

Inventory Control

Find an optimal ordering policy to meet a stochastic demand



- k, indexes discrete time
- x_k, stock available at *kth* period
- u_k, stock ordered at *kth* period
- w_k, demand *kth* period with a given probability distribution

Example II

Nominal System and Pay-Off

- Nominal Model

$$x_{k+1} = \max(0, x_k + u_k - w_k), \qquad k = 0, 1, 2$$

- Pay-off over $N = 2$ periods

$$E_Q\left\{ \sum_{k=0}^{2} [u_k + \max(0, w_k - x_k - u_k) + \max(0, x_k + u_k - w_k)] \right\}$$

Specific example:

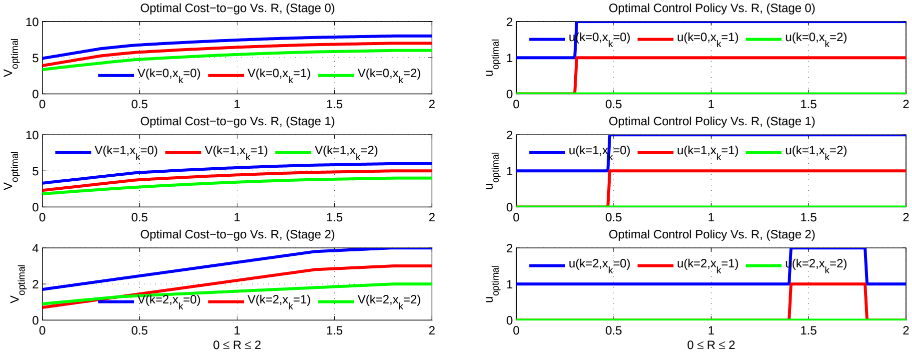
1. $x_k, u_k, w_k \in \{0, 1, 2\}$.
2. Maximum Capacity for stock $(x_k + u_k)$ is 2 units.
3. Nominal Probability Distribution: $P_{w_k}(w_k = 0) = 0.1$, $P_{w_k}(w_k = 1) = 0.7$, $P_{w_k}(w_k = 2) = 0.2$.

Example III

DP Algorithm Results

	Stock	Stage.0 Cost-to-go	Stage.0 Optimal Stock to Purchase	Stage.1 Cost-to-go	Stage.1 Optimal Stock to Purchase	Stage.2 Cost-to-go	Stage.2 Optimal Stock to Purchase
R=0	0	4.9	1	3.3	1	1.7	1
	1	3.9	0	2.3	0	0.7	0
	2	3.352	0	1.82	0	0.9	0
R=0.2	0	5.8	1	3.9	1	2	1
	1	4.8	0	2.9	0	1	0
	2	3.96	0	2.22	0	1.1	0
R=1	0	7.44	2	5.44	2	3.2	1
	1	6.44	1	4.44	1	2.2	0
	2	5.44	0	3.44	0	1.6	0

Example IV



- $R = 0$: the uncertain class reduces to the nominal demand probability distribution.
- $0 < R < 2$: balance the desire for low costs with the undesirability of scenarios with high uncertainty.
- $R = 2$: optimal ordering policies which are more robust with respect to uncertainty, but with the sacrifice of low present and future costs.

Outline

Conclusions-Future Work I

Conclusions

- Optimization of Stochastic Uncertain Systems Subject to Total Variation Distance Uncertainty
- Characterization of the Maximizing Measure belonging to the Total Variational Distance Set
- Generalized Dynamic Programming Algorithm

Future Work

- Derive Dynamic Programming Recursion for the Partially Observed Markov Decision Case
- Application to Decision and Information Theory

Outline

References I

F. Rezaei, C.D. Charalambous, N.U. Ahmed.
Optimal Control of Uncertain Stochastic Systems Subject to Total Variation Distance Uncertainty.
SIAM Journal of Control and Optimization, Vol. 50, No. 5, September 2012, pp.2683-2725.

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

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, No. 2, April 1981, pp.312-332.

References II

- P.R. Kumar and P. Varaiya.
  Stochastic Systems: Estimation, Identification and Adaptive Control.
  Prentice Hall, 1986.
- N. Dunford, and J. T. Schwartz.
  Linear Operators: Part 1: General Theory.
  Interscience Publishers, Inc., New York, 1957.
- D.P. Bertsekas.
  Dynamic Programming and Optimal Control.
  Athena Scientific, 2005.