Dynamic Consistency for Stochastic Optimal Control Problems

Pierre Carpentier and SOWG [1]

ENSTA ParisTech

August 16, 2010

[1] **Systems and Optimization Working Group**:
L. Andrieu, K. Barty, P. Carpentier, J.-P. Chancelier, G. Cohen, A. Dallagi, M. De Lara and P. Girardeau, with the support of Électricité de France R&D.

Presentation outline

Time consistency : an informal definition

Consider a discrete time optimal control problem on $\{t_0, t_1, \ldots, T\}$.

- The decision maker solves the problem at time t_0, that yields a sequence of optimal decision rules for time step t_0 and for the following time steps $\{t_1, \ldots, T\}$.
- At the next time step t_1, suppose that he is able to formulate a new problem starting at t_1, that yields a new sequence of optimal decision rules for time steps t_1 to T.
- This process can be continued until final time T is reached.

Such a family of optimization problems is said to be time consistent if the optimal strategies obtained when solving the original problem at time t_0 remain optimal for all subsequent problems.

A first example in the deterministic case (1)

$$\min_{(u_{t_0},\dots,u_{T-1}, x_{t_0},\dots,x_T)} \sum_{t=t_0}^{T-1} L_t\big(x_t, u_t\big) + K\big(x_T\big), \qquad (\mathcal{D}_{t_0})$$

$$\text{subject to} \quad x_{t+1} = f_t\big(x_t, u_t\big)\ , \quad x_{t_0} \ \text{given.}$$

Suppose a solution to this problem exists:

$\rightsquigarrow$ $(u^\sharp_{t_0}, \dots, u^\sharp_{T-1})$: controls indexed by time t,

$(x_{t_0}, x^\sharp_{t_1}, \dots, x^\sharp_{T})$: optimal path for the state variable.

No need for more information since the model is deterministic.

One has to note that

- these controls depend on the hidden parameter x_{t_0},
- these controls are usually not optimal for $x'_{t_0} \neq x_{t_0}$.

A first example in the deterministic case (2)

Consider the natural subsequent problems for every $t_i \geq t_0$:

$$\min_{(u_{t_i},\dots,u_{T-1},x_{t_i},\dots,x_T)} \sum_{t=t_i}^{T-1} L_t\big(x_t,u_t\big) + K\big(x_T\big), \qquad (\mathcal{D}_{t_i})$$

$$\text{subject to} \quad x_{t+1} = f_t\big(x_t,u_t\big) \ , \quad x_{t_i} \ \text{given}.$$

One makes the two following observations.

1. Independence of the initial condition. In the very particular case where the solution to Problem $(\mathcal{D}_{t_i})$ does not depend on x_{t_i}, Problems $\{(\mathcal{D}_{t_i})\}_{t_i}$ are dynamically consistent.
2. True deterministic world. Suppose that the initial condition for Problem $(\mathcal{D}_{t_i})$ is given by $x^\sharp_{t_i} = f_{t_i}\big(x^\sharp_{t_{i-1}}, u^\sharp_{t_{i-1}}\big)$ (exact model), then Problems $\{(\mathcal{D}_{t_i})\}_{t_i}$ are dynamically consistent.

Otherwise, adding disturbances to the problem brings inconsistency.

A first example in the deterministic case (3)

Solve now Problem $(\mathcal{D}_{t_0})$ using Dynamic Programming (DP):
$\rightsquigarrow$ $(\phi^\sharp_{t_0}, \ldots, \phi^\sharp_{T-1})$: controls depending on both t and x.

The following result is a direct application of the DP principle.

3. Right amount of information. Suppose that one is looking for strategies as feedback functions $\phi^\sharp_t$ depending on state x. Then Problems $\{(\mathcal{D}_{t_i})\}_{t_i}$ are dynamically consistent.

As a first conclusion, time consistency is recovered provided we let the decision rules depend upon a sufficiently rich information.

Aim of the talk: *enlighten the link between the notions of time consistency and state variable in Markov Decision Processes.*

Stochastic optimal control: the classical case (1)

$$\min_{(\mathbf{U}_{t_0},\dots,\mathbf{U}_{T-1},\mathbf{X}_{t_0},\dots,\mathbf{X}_T)} \mathbb{E}\left(\sum_{t=t_0}^{T-1} L_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) + K(\mathbf{X}_T)\right),$$

subject to constraints

$$\begin{aligned} \text{dynamics:}\quad & \mathbf{X}_{t_0} \text{ given,} & (\mathcal{S}_{t_0}) \\ & \mathbf{X}_{t+1} = f_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}), & \\ \text{measurability:}\quad & \mathbf{U}_t \preceq (\mathbf{X}_{t_0},\mathbf{W}_{t_0+1},\dots,\mathbf{W}_t). & \end{aligned}$$

In the Markovian setting (noises $\mathbf{X}_{t_0},\mathbf{W}_{t_0+1},\dots,\mathbf{W}_T$ independent), there is no loss of optimality in looking for the optimal strategy $\mathbf{U}_t$ at t as a feedback function ϕ_t depending on the state variable $\mathbf{X}_t$.

Stochastic optimal control: the classical case (2)

Problem $(\mathcal{S}_{t_0})$ can be solved using Dynamic Programming:

$$V_T^\sharp(x) = K(x),$$
$$V_t^\sharp(x) = \min_{u \in \mathbb{U}} \mathbb{E}\Big(L_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}^\sharp\big(f_t(x, u, \mathbf{W}_{t+1})\big)\Big).$$

It is clear while inspecting the DP equation that optimal strategies $\{\phi_t^\sharp\}_{t \geq t_0}$ remain optimal for the subsequent optimization problems:

$$\min_{(\mathbf{U}_{t_i},\dots,\mathbf{U}_{T-1},\mathbf{X}_{t_i},\dots,\mathbf{X}_T)} \mathbb{E}\left(\sum_{t=t_i}^{T-1} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) + K(\mathbf{X}_T) \right),$$

$$\text{subject to: } \quad \mathbf{X}_{t_i} \text{ given}, \qquad (\mathcal{S}_{t_i})$$
$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}),$$
$$\mathbf{U}_t \preceq (\mathbf{X}_{t_i}, \mathbf{W}_{t_i+1}, \dots, \mathbf{W}_t).$$

In consequence, problems $\{(\mathcal{S}_{t_i})\}_{t_i}$ are dynamically consistent.

A distributed formulation for Problem $(\mathcal{S}_{t_0})$

Dynamics of the probability laws (see [Witsenhausen, 1973]).

- Markovian setting: $\mathbf{U}_t = \phi_t(\mathbf{X}_t)$.
- Let μ_{t_0} be the probability law of the initial state $\mathbf{X}_{t_0}$, and μ_t be the probability law of the state at time t.

Introducing the two following operators:

$$A_t^{\phi} V(\cdot) := \mathbb{E}\Big(V\big(f_t(\cdot, \phi(\cdot), \mathbf{W}_{t+1})\big)\Big) \;,\; \Lambda_t^{\phi}(\cdot) := \mathbb{E}\big(L_t(\cdot, \phi(\cdot), \mathbf{W}_{t+1})\big),$$

$(\mathcal{S}_{t_0})$ is equivalent to the infinite-dimensional deterministic problem:

$$\min_{(\phi_{t_0},\dots,\phi_{T-1}\,,\,\mu_{t_0},\dots,\mu_T)} \sum_{t=t_0}^{T-1} \Big\langle \Lambda_t^{\phi_t} \,, \mu_t \Big\rangle + \langle K \,, \mu_T \rangle \,,$$

$$\text{subject to:} \quad \mu_{t_0} \text{ given,} \qquad (\mathcal{D}_{t_0})$$

$$\mu_{t+1} = \left(A_t^{\phi_t}\right)^{\star} \mu_t \quad \text{(Fokker-Planck).}$$

Solving $(\mathcal{D}_{t_0})$ using Dynamic Programming

$$\mathcal{V}_T(\mu) = \langle K \,, \mu \rangle \,.$$

$$\mathcal{V}_{T-1}(\mu) = \min_{\phi} \left\langle \Lambda^{\phi}_{T-1} \,, \mu \right\rangle + \mathcal{V}_T\Big(\big(A^{\phi}_{T-1}\big)^{\star} \mu \Big).$$

Optimal feedback $\Gamma^{\sharp}_{T-1} : \mu \rightarrow \phi^{\sharp}_{\mu}(\cdot)$ a priori depends on x and μ.

$$\mathcal{V}_{T-1}(\mu) = \min_{\phi} \left\langle \Lambda^{\phi}_{T-1} + A^{\phi}_{T-1} K \,, \mu \right\rangle ,$$

$$= \min_{\phi(\cdot)} \int_{\mathbb{X}} \big(\Lambda^{\phi}_{T-1} + A^{\phi}_{T-1} K\big)(x)\mu(\mathrm{d}x).$$

Interchanging minimization and expectation operators leads to:

- optimal $\Gamma^{\sharp}_{T-1}$ does not depend on μ : $\Gamma^{\sharp}_{T-1} \equiv \phi^{\sharp}_{T-1}$,
- $\mathcal{V}_{T-1}$ again depends on μ in a multiplicative manner.

Close to the very particular case mentioned in the first example.

1. Introduction
2. Stochastic optimal control without constraints
3. Stochastic optimal control with constraints
   - Constraints typology
   - Solving the constrained problem

Constraints in stochastic optimal control

Different kinds of constraints in stochastic optimization:

almost-sure constraint	:	$g(\mathbf{X}_T) \leq a \quad \mathbb{P}\text{-a.s.},$
chance constraint	:	$\mathbb{P}\big(g(\mathbf{X}_T) \leq a\big) \geq p,$
expectation constraint	:	$\mathbb{E}\big(g(\mathbf{X}_T)\big) \leq a,$
...		

A chance constraint can be modelled as an expectation constraint:

$$\mathbb{P}\big(g(\mathbf{X}_T) \leq a\big) = \mathbb{E}\big(\mathbf{1}_{\mathbb{X}^{\mathrm{ad}}}(\mathbf{X}_T)\big),$$

(with $\mathbb{X}^{\mathrm{ad}} = \{x \in \mathbb{X},\ g(x) \leq a\}$).

Chance constraints bring both theoretical and numerical difficulties, especially convexity [Prékopa, 1995]. However the difficulty we are interested in is common to chance and expectation constraints.

In the sequel, we concentrate on adding an expectation constraint.

Stochastic optimal control with expectation constraints

$$\min_{(\mathbf{U}_{t_0},\ldots,\mathbf{U}_{T-1},\mathbf{X}_{t_0},\ldots,\mathbf{X}_T)} \mathbb{E}\left(\sum_{t=t_0}^{T-1} L_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) + K(\mathbf{X}_T)\right),$$

subject to constraints

$$\text{dynamics:} \quad \mathbf{X}_{t_0} = x_{t_0}, \qquad (\mathcal{C}_{t_0})$$
$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}),$$

$$\text{measurability:} \quad \mathbf{U}_t \preceq (\mathbf{X}_{t_0},\mathbf{W}_{t_0+1},\ldots,\mathbf{W}_t),$$

$$\text{expectation:} \quad \mathbb{E}\big(g(\mathbf{X}_T)\big) \leq a.$$

Note that the initial state condition is equivalent to: $\mu_{t_0} = \delta_{x_{t_0}}$.
It corresponds to the full observation of the state variable at t_0.

Distributed formulation of Problem $(\mathcal{C}_{t_0})$

Again, there is no loss of optimality in looking for $\mathbf{U}_t$ as $\phi_t(\mathbf{X}_t)$.
The expectation constraint writes:

$$\langle g \, , \mu_T \rangle \leq a.$$

The distributed formulation of Problem $(\mathcal{C}_{t_0})$ writes:

$$\min_{(\phi_{t_0},\dots,\phi_{T-1}\,,\,\mu_{t_0},\dots,\mu_T)} \sum_{t=t_0}^{T-1} \left\langle \Lambda_t^{\phi_t} \, , \mu_t \right\rangle + \langle K \, , \mu_T \rangle + \chi_{\{\langle g,\mu\rangle \leq a\}}(\mu_T),$$

$$\text{subject to:} \quad \mu_{t_0} = \delta_{x_{t_0}},$$

$$\mu_{t+1} = \left(A_t^{\phi_t}\right)^{\star} \mu_t.$$

The expectation constraint introduces an additional nonlinear term in the cost function: there is no reason for the feedback laws to be independent of the initial condition μ_{t_0} as in the previous case.

Back to time consistency

Solving the previous deterministic problem leads to feedback laws $(\phi^{\sharp}_{t_0}, \ldots, \phi^{\sharp}_{T-1})$ depending on the initial condition μ_{t_0}.

According to the second observation made on our first example, time consistency holds if one follows the optimal path given by the Fokker-Planck equation. But the best observation available at time t_i for the probability law of $\mathbf{X}_{t_i}$ is a Dirac function...

Solving the problem using Dynamic Programming:

$$\mathcal{V}_T(\mu) = \langle K \,, \mu \rangle + \chi_{\{\langle g, \mu \rangle \leq a\}}(\mu),$$

$$\mathcal{V}_{T-1}(\mu) = \min_{\phi} \left\langle \Lambda^{\phi}_{T-1} \,, \mu \right\rangle + \mathcal{V}_T\Big(\big(A^{\phi}_{T-1}\big)^{\star} \mu \Big),$$

leads to feedback functions $\Gamma^{\sharp}_{T-1}$ depending on both x and μ. The context is similar to the one of the first example: time consistency holds, but the state variable is an infinite dimensional object...

Conclusions

- For several classes of optimal control problems, the concept of time consistency can be directly linked with the notion of state variable.
- In general, feedback laws have to depend on the probability law of the usual state variable for stochastic optimal control problems to be time consistent.
- The family of optimization problems we introduced in the constrained case makes use of the same level a of constraint, whatever the initial time step t_i (modelling choice):

$$\mathbb{E}\big(g(\mathbf{X}_T)\big) \leq a, \qquad \forall t_i \geq t_0.$$

- The Dynamic Programming equations we introduced in the constrained case are in general intractable since probability laws are infinite dimensional objects.

P. Carpentier, J. P. Chancelier, G. Cohen, M. De Lara, P. Girardeau.
Dynamic consistency for stochastic optimal control problems.
arXiv, 1005.3605[math.OC], 2010.

P. J. Hammond.
Consistent plans, consequentialism, and expected utility.
Econometrica, 57(6):1445–1449.

A. Ruszczynski.
Risk-averse dynamic programming for Markov decision processes.
Mathematical Programming, 2009.

A. Shapiro.
On a time consistency concept in risk averse multistage stochastic programming.
Operations Research Letters, 37(3):143–147, 2009.

H. S. Witsenhausen.
A standard form for sequential stochastic control.
Mathematical Systems Theory, 7(1):5–11, 1973.