

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

- 1 Introduction
 - Time consistency notion
 - A first example
- 2 Stochastic optimal control without constraints
 - The classical case
 - The distributed formulation
- 3 Stochastic optimal control with constraints
 - Constraints typology
 - Solving the constrained problem

Time consistency : an informal definition

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

- 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 $\{t_1, \dots, T\}$.
- 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)

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

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

Suppose a solution to this problem exists:

$\rightsquigarrow (u_{t_0}^\#, \dots, u_{T-1}^\#)$: controls indexed by time t ,
 $(x_{t_0}^\#, x_{t_1}^\#, \dots, x_T^\#)$: 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 $t_i \geq t_0$:

$$\begin{aligned} \min_{(u_{t_i}, \dots, u_{T-1}, x_{t_i}, \dots, x_T)} \quad & \sum_{t=t_i}^{T-1} L_t(x_t, u_t) + K(x_T), & (\mathcal{D}_{t_i}) \\ \text{subject to} \quad & x_{t+1} = f_t(x_t, u_t), \quad x_{t_i} \text{ given.} \end{aligned}$$

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 $x_{t_i}^\# = f_{t_i}(x_{t_{i-1}}^\#, u_{t_{i-1}}^\#)$ (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)**:

$\rightsquigarrow (\phi_{t_0}^\#, \dots, \phi_{T-1}^\#)$: controls depending on both t and x .

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

- Right amount of information.** Suppose that one is looking for strategies as feedback functions $\phi_t^\#$ 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.*

- 1 Introduction
- 2 Stochastic optimal control without constraints
 - The classical case
 - The distributed formulation
- 3 Stochastic optimal control with constraints

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

$$\text{dynamics: } \mathbf{x}_{t_0} \text{ given,} \quad (\mathcal{S}_{t_0})$$

$$\mathbf{x}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_{t+1}),$$

$$\text{measurability: } \mathbf{u}_t \preceq (\mathbf{x}_{t_0}, \mathbf{w}_{t_0+1}, \dots, \mathbf{w}_t).$$

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:

$$V_T^\#(x) = K(x),$$

$$V_t^\#(x) = \min_{u \in \mathbb{U}} \mathbb{E} \left(L_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}^\#(f_t(x, u, \mathbf{W}_{t+1})) \right).$$

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

$$\begin{aligned} & \min_{(\mathbf{u}_{t_i}, \dots, \mathbf{u}_{T-1}, \mathbf{x}_{t_i}, \dots, \mathbf{x}_T)} \mathbb{E} \left(\sum_{t=t_i}^{T-1} L_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_{t+1}) + K(\mathbf{x}_T) \right), \\ & \text{subject to: } \mathbf{x}_{t_i} \text{ given,} \\ & \mathbf{x}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_{t+1}), \\ & \mathbf{u}_t \preceq (\mathbf{x}_{t_i}, \mathbf{w}_{t_i+1}, \dots, \mathbf{w}_t). \end{aligned} \quad (\mathcal{S}_{t_i})$$

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:

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

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

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

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

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

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

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

$$\mathcal{V}_{T-1}(\mu) = \min_{\phi} \langle \Lambda_{T-1}^{\phi}, \mu \rangle + \mathcal{V}_T((A_{T-1}^{\phi})^* \mu).$$

Optimal feedback $\Gamma_{T-1}^{\#} : \mu \rightarrow \phi_{\mu}^{\#}(\cdot)$ a priori depends on x and μ .

$$\mathcal{V}_{T-1}(\mu) = \min_{\phi} \langle \Lambda_{T-1}^{\phi} + A_{T-1}^{\phi} K, \mu \rangle,$$

$$= \min_{\phi(\cdot)} \int_{\mathbb{X}} (\Lambda_{T-1}^{\phi} + A_{T-1}^{\phi} K)(x) \mu(dx).$$

Interchanging minimization and expectation operators leads to:

- optimal $\Gamma_{T-1}^{\#}$ does not depend on μ : $\Gamma_{T-1}^{\#} \equiv \phi_{T-1}^{\#}$,
- \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 : $\mathbb{P}(g(\mathbf{X}_T) \leq a) \geq p,$

expectation constraint : $\mathbb{E}(g(\mathbf{X}_T)) \leq a,$

...

A **chance constraint** can be modelled as an **expectation constraint**:

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

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

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

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

subject to constraints

$$\text{dynamics: } \mathbf{X}_{t_0} = x_{t_0}, \quad (\mathcal{C}_{t_0})$$

$$\mathbf{X}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_{t+1}),$$

$$\text{measurability: } \mathbf{u}_t \preceq (\mathbf{x}_{t_0}, \mathbf{w}_{t_0+1}, \dots, \mathbf{w}_t),$$

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

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:

$$\langle \mathbf{g}, \mu_T \rangle \leq a.$$

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

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

$$\begin{aligned} \text{subject to: } \mu_{t_0} &= \delta_{\mathbf{x}_{t_0}}, \\ \mu_{t+1} &= (A_t^{\phi_t})^* \mu_t. \end{aligned}$$

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 $(\phi_{t_0}^\#, \dots, \phi_{T-1}^\#)$ 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**:

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

$$\mathcal{V}_{T-1}(\mu) = \min_{\phi} \left\langle \Lambda_{T-1}^{\phi}, \mu \right\rangle + \mathcal{V}_T \left((A_{T-1}^{\phi})^* \mu \right),$$

leads to feedback functions $\Gamma_{T-1}^\#$ 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**):

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

- 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. Ruszczyński.
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.