

# DYNAMIC CONSISTENCY FOR STOCHASTIC OPTIMAL CONTROL PROBLEMS

PIERRE CARPENTIER, JEAN-PHILIPPE CHANCELIER, GUY COHEN,  
MICHEL DE LARA, AND PIERRE GIRARDEAU

**ABSTRACT.** For a sequence of dynamic optimization problems, we aim at discussing a notion of consistency over time. This notion can be informally introduced as follows. At the very first time step  $t_0$ , the decision maker formulates an optimization problem that yields optimal decision rules for all the forthcoming time step  $t_0, t_1, \dots, T$ ; at the next time step  $t_1$ , he is able to formulate a new optimization problem starting at time  $t_1$  that yields a new sequence of optimal decision rules. This process can be continued until final time  $T$  is reached. A family of optimization problems formulated in this way is said to be time consistent if the optimal strategies obtained when solving the original problem remain optimal for all subsequent problems. The notion of time consistency, well-known in the field of Economics, has been recently introduced in the context of risk measures, notably by Artzner et al. (2007) and studied in the Stochastic Programming framework by Shapiro (2009) and for Markov Decision Processes (MDP) by Ruszczyński (2009). We here link this notion with the concept of “state variable” in MDP, and show that a significant class of dynamic optimization problems are dynamically consistent, provided that an adequate state variable is chosen.

## 1. INTRODUCTION

Stochastic Optimal Control (SOC) is concerned with sequential decision-making under uncertainty. Consider a dynamical process that can be influenced by exogenous noises as well as decisions one has to make at every time step. The decision maker wants to optimize the behavior of the dynamical system (for instance, minimize a production cost) over a certain time horizon. As the system evolves, observations of the system are made; we here suppose that the decision maker is able to keep in memory all the past observations. Naturally, it is generally more profitable for him to adapt its decisions to the observations he makes of the system. He is hence looking for strategies rather than simple decisions. In other words, he is looking for applications that map every possible history of the observations to corresponding decisions. Because the number of time steps may be large, the representation of such an object is in general numerically intractable.

However, an amount of information lighter than the whole history of the system is often sufficient to make an optimal decision. In the seminal work of Bellman (1957), the minimal information on the system that is necessary to make the optimal decision plays a crucial role; it is called the *state variable* (see Whittle, 1982, for a more formal definition). Moreover, the Dynamic Programming (DP) principle

---

*Date:* May 17, 2010.

*2000 Mathematics Subject Classification.* 93E20, 49L20, 91B70.

*Key words and phrases.* Stochastic optimal control, Dynamic consistency, Dynamic Programming, Risk measures.

This study was made within the Systems and Optimization Working Group (SOWG), which is composed of Laetitia Andrieu, Kengy Barty, Pierre Carpentier, Jean-Philippe Chancelier, Guy Cohen, Anes Dallagi, Michel De Lara and Pierre Girardeau, and based at Université Paris-Est, CERMICS, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France.

provides a way to compute the optimal strategies when the state space dimension is not too large (see Bertsekas, 2000, for a broad overview on DP). The aim of this paper is to establish a link between the concept of state variable and the notion of time consistency<sup>1</sup>.

The notion of dynamic consistency is well-known in the field of economics (see Hammond, 1989) and has been introduced in the context of risk measures (see Artzner et al., 2007; Riedel, 2004; Detlefsen and Scandolo, 2005; Cheridito et al., 2006, for definitions and properties of coherent and consistent dynamic risk measures). Dynamic consistency has then been studied in the stochastic programming framework by Shapiro (2009) and for Markov Decision Processes by Ruszczyński (2009). In this paper, we rather use the (almost equivalent) definition of time consistency given by Ekeland and Lazrak (2006), which is more intuitive and seems better suited in the framework of optimal control problems. In this context, the property of time consistency is loosely stated as follows. The decision maker formulates an optimization problem at time  $t_0$  that yields a sequence of optimal decision rules for  $t_0$  and for the following time steps  $t_1, \dots, t_N = T$ . Then, at the next time step  $t_1$ , he formulates a new problem starting at  $t_1$  that yields a new sequence of optimal decision rules from time steps  $t_1$  to  $T$ . Suppose the process continues until time  $T$  is reached. The sequence of optimization problems is said to be dynamically consistent if the optimal strategies obtained when solving the original problem at time  $t_0$  remain optimal for all subsequent problems. In other words, time consistency means that strategies obtained by solving the problem at the very first stage do not have to be questioned later on.

The notion of information here plays a crucial role. Indeed, we show in this paper that a sequence of problems may be consistent for some information structure while inconsistent for a different one. Consider for example a standard stochastic optimization problem solvable using DP. We will observe that the sequence of problems formulated after the original one at the later time steps are time consistent. Add now a probabilistic constraint involving the state at the final time  $T$ . We will show that such a constraint brings time inconsistency in the sense that optimal strategies based on the usual state variable have to be reconsidered at each time step. This is because, roughly speaking, a probabilistic constraint involves not only the state variable values but their probabilistic distributions. Hence the only knowledge of the usual state variable of the system is insufficient to write consistent problems at subsequent time steps. So, in addition to the usual technical difficulties regarding probabilistic constraints (mainly related to the non-convexity of the feasible set of strategies), an additional problem arises in the dynamic case. We will see that, in fact, this new matter comes from the information on which the optimal decision is based. Therefore, with a well-suited state variable, the sequence of problems regains dynamic consistency.

In §2, we carefully examine the notion of time consistency in the context of a deterministic optimal control problem. The main ideas of the paper are so explained and then extended, in §3, to a sequence of SOC problems. Next, in §4, we show that simply adding a probability constraint (or, equivalently in our context, an expectation constraint) to the problem makes time consistency fall apart, when using the original state variable. We then establish that time consistency can be recovered provided an adequate state variable is chosen. We conclude that, for a broad class of SOC problems, time consistency has to be considered with respect to the notion of a state variable and of DP.

---

<sup>1</sup>We either use the term “dynamically consistent” or “time consistent” to refer to the same notion.

## 2. A FIRST EXAMPLE

We introduce sequential deterministic optimal control problems, indexed by time, and derive the notion of time consistency on this instance. We then illustrate the fact that the decision making process may be time consistent or not, depending on the information on which decisions are based. The discussion is informal, in the sense that we do not enter technical details regarding existence of the solutions for the problems we introduce.

Let us consider a discrete and finite time horizon  $t_0, \dots, t_N = T$ .<sup>2</sup> The decision maker has to optimize (according to a cost function we introduce below) the management of an amount of stock  $x_t$ , which lies in some space  $\mathcal{X}_t$ , at every time step  $t = t_0, \dots, T$ . Let  $\mathcal{U}_t$  be some other space, for every time step  $t = t_0, \dots, T-1$ . At each time step  $t$ , a decision  $u_t \in \mathcal{U}_t$  has to be made. Then a cost  $L_t$  is incurred by the system, depending on the values of the control and on the auxiliary variable  $x_t$  that we call the state of the system. This state variable is driven from time  $t$  to time  $t+1$  by some dynamics  $f_t : \mathcal{X}_t \times \mathcal{U}_t \rightarrow \mathcal{X}_{t+1}$ . The aim of the decision maker is to minimize the sum of the intermediate costs  $L_t$  at all time steps plus a final cost  $K$ .

The problem hence reads:

$$(1a) \quad \min_{x, u} \sum_{t=t_0}^{T-1} L_t(x_t, u_t) + K(x_T),$$

subject to the initial condition:

$$(1b) \quad x_{t_0} \text{ given},$$

and dynamic constraints:

$$(1c) \quad x_{t+1} = f_t(x_t, u_t), \quad \forall t = t_0, \dots, T-1.$$

Note that here the decision at time  $t$  is taken knowing the current time step and the initial condition (the decision is generally termed “open loop”). A priori, there is no need for more information since the model is deterministic.

Suppose a solution to this problem exists. This is a sequence of controls that we denote by  $u_{t_0, t_0}^*, \dots, u_{t_0, T-1}^*$ , where the first index refers to the initial time step and the second index refers to the time step for which the decision applies. Moreover, we suppose a solution exists for each one of the natural subsequent problems, i.e. for every  $t_i = t_1, \dots, T-1$ :

$$(2a) \quad \min_{x, u} \sum_{t=t_i}^{T-1} L_t(x_t, u_t) + K(x_T),$$

$$(2b) \quad \text{s.t. } x_{t_i} \text{ given},$$

$$(2c) \quad x_{t+1} = f_t(x_t, u_t), \quad \forall t = t_i, \dots, T-1.$$

We denote the solutions of these problems by  $u_{t_i, t_i}^*, \dots, u_{t_i, T-1}^*$ , for every time step  $t_i = t_1, \dots, T-1$ . Those notations however make implicit the fact that the solutions do generally depend on the initial condition  $x_{t_i}$ . We now make a first observation.

**Lemma 1** (Independence of the initial condition). *In the very particular case when the solution to Problem (1) and the solutions to Problems (2) for every time step  $t_i = t_1, \dots, T-1$  do not depend on the initial state conditions, problems are dynamically consistent.*

---

<sup>2</sup>where  $t_i + 1 = t_{i+1}$

*Proof.* Let us denote by  $x_{t_0, t_i}^*$  the optimal value of the state variable within Problem (1) at time  $t_i$ . If we suppose that solutions to Problems (2) do not depend on the initial condition, then they are the same as the solutions obtained with the initial condition  $x_{t_0, t_i}^*$ , namely  $u_{t_0, t_i}^*, \dots, u_{t_0, T-1}^*$ . In other words, the sequence of decisions  $u_{t_0, t_0}^*, \dots, u_{t_0, T-1}^*$  remains optimal for the subsequent problems starting at a later date.  $\square$

This property is of course not true in general, but we see in Example 1 hereafter and in §3 that some very practical problems do have this surprising property.

*Example 1.* Let us introduce, for every  $t = t_0, \dots, T-1$ , functions  $l_t : \mathcal{U}_t \rightarrow \mathbb{R}$  and  $f_t : \mathcal{U}_t \rightarrow \mathbb{R}$ , and assume that  $x_t$  is scalar. Let  $K$  be a scalar constant and consider the following deterministic optimal control problem:

$$\begin{aligned} \min_{x, u} \quad & \sum_{t=t_0}^{T-1} l_t(u_t) x_t + K x_T, \\ \text{s.t.} \quad & x_{t_0} \text{ given}, \\ & x_{t+1} = f_t(u_t) x_t, \quad \forall t = t_0, \dots, T-1. \end{aligned}$$

Variables  $x_t$  can be recursively replaced using dynamics  $f_t$ . Therefore, the above optimization problem can be written:

$$\min_u \sum_{t=t_0}^{T-1} l_t(u_t) f_{t-1}(u_{t-1}) \dots f_{t_0}(u_{t_0}) x_{t_0} + K f_{T-1}(u_{T-1}) \dots f_{t_0}(u_{t_0}) x_{t_0}.$$

Hence the optimal cost of the problem is linear with respect to the initial condition  $x_{t_0}$ . Suppose that  $x_{t_0}$  only takes positive values. Then the value of  $x_{t_0}$  has no influence on the minimizer (it only influences the optimal cost). The same argument applies at subsequent time steps  $t_i > t_0$  provided that dynamics are such that  $x_t$  remains positive for every time step  $t = t_1, \dots, T$ . Now, formulate the same problem at a later date  $t_i = t_1, \dots, T-1$ , with initial condition  $x_{t_i}$  given. By the same token as for the first stage problem, the value of the initial condition  $x_{t_i}$  has no influence on the optimal controls. Assumptions made in Lemma 1 are fulfilled, so that the time consistency property holds true for open-loop decisions without reference to initial state conditions.

Although, for the time being, this example may look very special, we will see later on that it is analogous to familiar SOC problems.

As already noticed, Lemma 1 is not true in general. Moreover, the deterministic formulation (1) comes in general from the representation of a real-life process which may indeed be subject to unmodeled disturbances. Think of an industrial context, for example, in which sequential decisions are taken in the following manner.

- At time  $t_0$ , Problem (1) is solved. One obtains a decision  $u_{t_0, t_0}^*$  to apply at time  $t_0$ , as well as decisions  $u_{t_0, t_1}^*, \dots, u_{t_0, T-1}^*$  for future time steps.
- At time  $t_1$ , one formulates and solves the problem starting at time  $t_1$  with initial condition  $x_{t_1} = f_{t_0}(x_{t_0}, u_{t_0, t_0}^*) + \varepsilon_{t_1}$ ,  $\varepsilon_{t_1}$  being some perturbation of the model. There is no reason not to use the observation of the actual value of the variable  $x_{t_1}$  at time  $t_1$  as long as we have it at our disposal.
- Hence a decision  $u_{t_1, t_1}^*$  is obtained, which is different from the initially obtained optimal decision  $u_{t_0, t_1}^*$  (once again, in general).
- The same process continues at times  $t_2, \dots, T-1$ .

Let us now state the two following lemmas.

**Lemma 2** (True deterministic world). *If the deterministic model is actually exact, i.e. if all perturbations  $\varepsilon_{t_i}$  introduced above equal zero, then Problems (2) with initial conditions  $x_{t_i} = x_{t_i}^* \triangleq f_{t_i}(x_{t_{i-1}}^*, u_{t_0, t_{i-1}}^*)$  are dynamically consistent.*

*Proof.* Since decisions  $u_{t_0, t_0}^*, \dots, u_{t_0, T-1}^*$  are optimal for Problem (1), it follows that decisions  $u_{t_0, t_1}^*, \dots, u_{t_0, T-1}^*$  are optimal for the problem:

$$\begin{aligned} \min_{x, u} \quad & L_t(x_{t_0}, u_{t_0, t_0}^*) + \sum_{t=t_1}^{T-1} L_t(x_t, u_t) + K(x_T), \\ \text{s.t.} \quad & x_{t_1} = f_{t_1}(x_{t_0}, u_{t_0, t_0}^*), \\ & x_{t+1} = f_t(x_t, u_t), \quad \forall t = t_1, \dots, T-1, \end{aligned}$$

which has the same arg min as Problem (2) at time  $t_1$ . The same argument applies recursively for subsequent time steps.  $\square$

It is clear that Lemma 2 is not satisfied in real life. Therefore, adding disturbances to the problem seems to bring inconsistency to the sequence of optimization problems. Decisions that are optimal for the first stage problem do not remain optimal for the subsequent problems if we do not let decisions depend on the initial conditions.

In fact, as it is stated next, time consistency is recovered provided we let decisions depend upon the right information.

**Lemma 3** (Right amount of information). *Suppose that one is looking for strategies  $(\Phi_{t_0, t_0}^*, \dots, \Phi_{t_0, T-1}^*)$  as feedback functions depending on the variable  $x$ . Then Problems (2) are time consistent for every time step  $t = t_0, \dots, T-1$ .*

*Proof.* The result is a direct application of the DP principle, which states that there exists such a feedback function  $\Phi_{t_0, t_i}^*$  that is optimal for Problem (1) and is still optimal for Problem (2) at time  $t_i$ , whatever initial condition  $x_{t_i}$  is.  $\square$

We thus retrieve the dynamic consistency property provided that we use the feedback functions  $\Phi_{t_0, t}^*$  rather than the controls  $u_{t_0, t}^*$ . In other words, problems are dynamically consistent as soon as the control strategy is based on a sufficiently rich amount of information (time instant  $t$  **and** state variable  $x$  in the deterministic case).

There is of course an obvious link between these optimal strategies and the controls  $(u_{t_0, t_0}^*, \dots, u_{t_0, T-1}^*)$ , namely:

$$u_{t_0, t}^* = \Phi_{t_0, t}^*(x_{t_0, t}^*), \quad \forall t = t_0, \dots, T-1,$$

where

$$\begin{aligned} x_{t_0, t_0}^* &= x_{t_0}, \\ x_{t_0, t+1}^* &= f_t(x_{t_0, t}^*, \Phi_{t_0, t}^*(x_{t_0, t}^*)), \quad \forall t = t_0, \dots, T-1. \end{aligned}$$

The considerations we made so far seem to be somewhat trivial. However, we shall observe that for SOC problems, which may seem more complicated at first sight, the same considerations remain true. Most of the time, decision making processes are time consistent, provided we choose the correct information on which decisions are based.

## 3. STOCHASTIC OPTIMAL CONTROL WITHOUT CONSTRAINTS

We now consider a more general case in which a controlled dynamical system is influenced by modeled exogenous disturbances. The decision maker has to find strategies to drive the system so as to minimize some objective function over a certain time horizon. This is a sequential decision making process on which we can state the question of dynamic consistency. As in the previous example, the family of optimization problems is derived from the original one by truncating the dynamics and the cost function (the final time step  $T$  remains unchanged in each problem), and strategies are defined relying on the same information structure as in the original problem. In the sequel, random variables will be denoted using bold letters.

**3.1. The classical case.** Consider a dynamical system characterized by state<sup>3</sup> variables  $\mathbf{X} = (\mathbf{X}_t)_{t=t_0, \dots, T}$ , where  $\mathbf{X}_t$  takes values in  $\mathcal{X}_t$ . The system can be influenced by control variables  $\mathbf{U} = (\mathbf{U}_t)_{t=t_0, \dots, T-1}$  and by exogenous noise variables  $\mathbf{W} = (\mathbf{W}_t)_{t=t_0, \dots, T}$  ( $\mathbf{U}_t$  and  $\mathbf{W}_t$  taking values in  $\mathcal{U}_t$  and  $\mathcal{W}_t$  respectively). All random variables are defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . The problem we consider consists in minimizing the expectation of a sum of costs depending on the state, the control and the noise variables over a discrete finite time horizon. The state variable evolves with respect to some dynamics that depend on the current state, noise and control values. The problem starting at  $t_0$  writes:<sup>4</sup>

$$(3a) \quad \min_{\mathbf{X}, \mathbf{U}} \quad \mathbb{E} \left( \sum_{t=t_0}^{T-1} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) + K(\mathbf{X}_T) \right),$$

$$(3b) \quad \text{s.t.} \quad \mathbf{X}_{t_0} \text{ given,}$$

$$(3c) \quad \mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}), \quad \forall t = t_0, \dots, T-1,$$

$$(3d) \quad \mathbf{U}_t \preceq \mathbf{X}_{t_0}, \mathbf{W}_{t_1}, \dots, \mathbf{W}_t, \quad \forall t = t_0, \dots, T-1.$$

Noises that affect the system can be correlated through time. A general approach in optimal control consists in including all necessary information in the variable  $\mathbf{X}$  so that variables  $\mathbf{W}_{t_1}, \dots, \mathbf{W}_T$  are independent through time. At most, one has to include all the past values of the noise variable within the variable  $\mathbf{X}$ . We hence make the following assumption.

*Assumption 1* (Markovian setting). Noises variables  $\mathbf{X}_{t_0}, \mathbf{W}_{t_1}, \dots, \mathbf{W}_T$  are independent.

Using Assumption 1, it is well known (see Bertsekas, 2000) that:

- there is no loss of optimality in looking for the optimal strategy  $\mathbf{U}_t$  at time  $t$  as a feedback function depending on the state variable  $\mathbf{X}_t$ , i.e. as a (measurable) function of the form  $\Phi_{t_0, t} : \mathcal{X}_t \rightarrow \mathcal{U}_t$ ;
- the optimal strategies  $\Phi_{t_0, t_0}^*, \dots, \Phi_{t_0, T-1}^*$  can be obtained by solving the classical DP equation. Let  $V_t(x)$  denote the optimal cost when being at time step  $t$  with state value  $x$ , this equation reads:

$$V_T(x) = K(x),$$

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

<sup>3</sup>The use of the terminology “state” is somewhat abusive until we make Assumption 1.

<sup>4</sup>We here use the notations  $\sim$  for “is distributed according to” and  $\preceq$  for “is measurable with respect to”.

We call this case the *classical case*. It is clear while inspecting the DP equation that optimal strategies  $\Phi_{t_0, t_0}^*, \dots, \Phi_{t_0, T-1}^*$  remain optimal for the subsequent optimization problems:

$$\begin{aligned}
(4a) \quad & \min_{\mathbf{X}, \mathbf{U}} \quad \mathbb{E} \left( \sum_{t=t_i}^{T-1} L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) + K(\mathbf{X}_T) \right), \\
(4b) \quad & \text{s.t.} \quad \mathbf{X}_{t_i} \text{ given}, \\
(4c) \quad & \mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}), \quad \forall t = t_i, \dots, T-1, \\
(4d) \quad & \mathbf{U}_t \preceq \mathbf{X}_{t_i}, \mathbf{W}_{t_{i+1}}, \dots, \mathbf{W}_t, \quad \forall t = t_i, \dots, T-1,
\end{aligned}$$

for every  $t_i = t_1, \dots, T-1$ . In other words, these problems are dynamically consistent provided the information variable at time  $t$  contains at least the state variable  $\mathbf{X}_t$ . While building an analogy with properties described in the deterministic example in §2, the reader should be aware that the case we consider here is closer to Lemma 1 than to Lemma 3, as we explain now in more details.

**3.2. The distributed formulation.** Another consequence of the previous DP equation for Problem (3) is that the optimal feedback functions do not depend on the initial condition  $\mathbf{X}_{t_0}$ . The probability law of  $\mathbf{X}_{t_0}$  only affects the optimal cost value, but not its arg min. In fact, we are within the same framework as in Example 1. Indeed, Problem (3) can be written as a deterministic distributed optimal control problem involving the probability laws of the state variable, the dynamics of which are given by the so-called Fokker-Planck equation. Let us detail this last formulation (see Witsenhausen, 1973).

Let  $\Psi_t$  be the space of  $\mathbb{R}$ -valued functions on  $\mathcal{X}_t$ . Denoting  $\mu_{t_0}$  the probability law of the first stage state  $\mathbf{X}_{t_0}$ , and given feedback laws  $\Phi_t : \mathcal{X}_t \rightarrow \mathcal{U}_t$  for every time step  $t = t_0, \dots, T-1$ , we define the operator  $A_t^{\Phi_t} : \Psi_{t+1} \rightarrow \Psi_t$ , which is meant to integrate cost functions backwards in time, as<sup>5</sup>:

$$\left( A_t^{\Phi_t} \psi_{t+1} \right) (\cdot) \triangleq \mathbb{E} (\psi_{t+1} \circ f_t(\cdot, \Phi_t(\cdot), \mathbf{W}_{t+1})).$$

Given a feedback function  $\Phi_t$  and a cost function  $\psi_{t+1} \in \Psi_{t+1}$ , for every  $x \in \mathcal{X}_t$  the value  $(A_t^{\Phi_t} \psi_{t+1})(x)$  is the expected value of  $\psi_{t+1}(\mathbf{X}_{t+1})$ , knowing that  $\mathbf{X}_t = x$  and that feedback  $\Phi_t$  is used. Thanks to a duality argument, the Fokker-Planck equation, which describes the evolution of the state probability law (as driven by the chosen feedback laws  $\Phi_t$ ), is obtained:

$$\mu_{t+1} = \left( A_t^{\Phi_t} \right)^* \mu_t,$$

with  $(A_t^{\Phi_t})^*$  being the adjoint operator of  $A_t^{\Phi_t}$ . Next we introduce the operator  $\Lambda_t^{\Phi_t} : \mathcal{X}_t \rightarrow \mathbb{R}$ :

$$\Lambda_t^{\Phi_t}(\cdot) \triangleq \mathbb{E} (L_t(\cdot, \Phi_t(\cdot), \mathbf{W}_{t+1})),$$

which is meant to be the expected cost at time  $t$  for each possible state value when feedback function  $\Phi_t$  is applied. Let us define, for every  $\psi_t \in \Psi_t$  and every probability law  $\mu_t$  on  $\mathcal{X}_t$ ,  $\langle \psi_t, \mu_t \rangle$  as  $\mathbb{E} (\psi_t(\mathbf{X}_t))$  when  $\mathbf{X}_t$  is distributed according to  $\mu_t$ . We can now write a deterministic infinite-dimensional optimal control problem that

---

<sup>5</sup>We do not aim at discussing technical details concerning integrability here. We suppose that operators we introduce are well-defined.



is equivalent to Problem (3):

$$\begin{aligned} \min_{\Phi, \mu} \quad & \sum_{t=t_0}^{T-1} \langle \Lambda_t^{\Phi_t}, \mu_t \rangle + \langle K, \mu_T \rangle, \\ \text{s.t.} \quad & \mu_{t_0} \text{ given,} \\ & \mu_{t+1} = \left( A_t^{\Phi_t} \right)^* \mu_t, \quad \forall t = t_0, \dots, T-1. \end{aligned}$$

*Remark 1.* An alternative formulation is:

$$\begin{aligned} \min_{\Phi, \psi} \quad & \langle \psi_{t_0}, \mu_{t_0} \rangle, \\ \text{s.t.} \quad & \psi_T = K, \\ & \psi_t = A_t^{\Phi_t} \psi_{t+1} + \Lambda_t^{\Phi_t}, \quad \forall t = T-1, \dots, t_0. \end{aligned}$$

This may be called “*the backward formulation*” since the “state”  $\psi_t(\cdot)$  follows an affine dynamics which is backward in time, with an initial-only cost function (whereas the previous forward formulation follows a forward linear dynamics with an integral + final cost function). Both formulations are infinite-dimensional linear programming problems which are dual of each other. The functions  $\mu(\cdot)$  and  $\psi(\cdot)$  are the distributed state and/or co-state (according to which one is considered the primal problem) of this distributed deterministic optimal control problem of which  $\Phi$  is the distributed control.

Probability laws  $\mu_t$  are by definition positive and appear only in a multiplicative manner in the problem. Hence we are in a similar case as Example 1. The main difference is rather technical: since we here have probability laws instead of scalars, we need to apply backwards in time intervention theorems between expectation and minimization in order to prove that the solution of the problem actually does not depend on the initial condition  $\mu_{t_0}$ . Indeed, suppose that  $\mu_{T-1}$  is given at time step  $T-1$ . Then the most inner optimization problem reads:

$$\begin{aligned} \min_{\Phi_{T-1}} \quad & \langle \Lambda_t^{\Phi_{T-1}}, \mu_{T-1} \rangle + \langle K, \mu_T \rangle, \\ \text{s.t.} \quad & \mu_T = \left( A_{T-1}^{\Phi_{T-1}} \right)^* \mu_{T-1}, \end{aligned}$$

which is equivalent to:

$$\min_{\Phi_{T-1}} \quad \langle \Lambda_t^{\Phi_{T-1}} + A_{T-1}^{\Phi_{T-1}} K, \mu_{T-1} \rangle.$$

The point is that operators  $\Lambda_t^{\Phi_{T-1}} + A_{T-1}^{\Phi_{T-1}} K$  and  $\mu_{T-1}$  both take values in  $\mathcal{X}_{T-1}$  and that the minimization has to be done “ $x$  by  $x$ ”, so that we are in the case of Example 1 for every  $x$ . Therefore, the minimizer does not depend on  $\mu_{T-1}$ . For a rigorous proof, one needs several technical assumptions concerning measurability, which we do not intend to discuss in this paper (see Rockafellar and Wets, 1998, Theorem 14.60). The same argument applies recursively to every time step before  $T-1$  so that, at time  $t_0$ , the initial condition  $\mu_{t_0}$  only influences the optimal cost of the problem, but not the argument of the minimum itself (here, the feedback laws  $\Phi_{t_0, t}^*$ ).

Hence, following Lemma 1, Problems (4) are naturally time consistent when strategies are searched as feedback functions on  $\mathbf{X}_t$  only. It thus appears that the rather general class of stochastic optimal control problems shaped as Problem (3) is in fact very specific. However, such a property does not remain true when adding new ingredients in the problem, as we show in the next subsection.



## 4. STOCHASTIC OPTIMAL CONTROL WITH CONSTRAINTS

We now give an example in which the state variable, as defined notably by Whittle (1982), cannot be reduced to variable  $\mathbf{X}_t$  as above. Let us make Problem (3) more complex by adding to the model a probability constraint applying to the final time step  $T$ . For instance, we want the system to be in a certain state at the final time step with a given probability:

$$\mathbb{P}(h(\mathbf{X}_T) \geq b) \leq \pi.$$

Such chance constraints can equivalently be modelled as an expectation constraint in the following way:

$$\mathbb{E}(\mathbf{1}_{\{h(\mathbf{X}_T) \geq b\}}) \leq \pi,$$

where  $\mathbf{1}_A$  refers to the indicator function of set  $A$ . Note however that chance constraints bring important theoretical and numerical difficulties, notably regarding connexity and convexity of the feasible set of controls, even in the static case. The interested reader should refer to the work of Prékopa (1995), and to the handbook by Ruszczynski and Shapiro (2003, Ch.5) for mathematical properties and numerical algorithms in Probabilistic Programming (see also Henrion, 2002; Henrion and Strugarek, 2008, for related studies). We do not discuss them here. The difficulty we are interested in is common to both chance and expectation constraints. This is why we concentrate in the sequel on adding an expectation constraint to Problem (3) of the form:

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

The reader familiar with chance constraints might want to see the level  $a$  as a level of probability that one wants to satisfy for a certain event at the final time step.

We now show that when adding such an expectation constraint, the dynamic consistency property falls apart. More precisely, the sequence of SOC problems are not time consistent anymore when using the usual state variable. Nevertheless, we observe that the lack of consistency comes from an inappropriate choice for the state variable. By choosing the appropriate state variable, one regains dynamic consistency.

**4.1. Problem setting.** We now go back to the constrained formulation and introduce a measurable function  $g : \mathcal{X}_T \rightarrow \mathbb{R}$  and  $a \in \mathbb{R}$ . We consider Problem (3) with the additional final expectation constraint:

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

The subsequent optimization problems formulated at an initial time  $t_i > t_0$  are naturally deduced from this problem. The level  $a$  of the expectation constraint remains the same for every problem. One has to be aware that this corresponds to a (naïve) modelling choice for the family of optimization problems under consideration. Such a choice is questionable since the perception of the constraint may evolve over time.

Suppose there exists a solution for the problem at  $t_0$ . As previously, we are looking for the optimal control at time  $t$  as a feedback function  $\Phi_{t_0, t}^*$  depending on the variable  $\mathbf{X}_t$ . The first index  $t_0$  refers to the time step at which the problem is stated, while the second index  $t$  refers to the time step at which the decision is taken.

One has to be aware that these solutions now implicitly depend on the initial condition  $\mathbf{X}_{t_0}$ . Indeed, let  $\mu_T$  be the probability law of  $\mathbf{X}_T$ . Constraint (4) can be written  $\langle g, \mu_T \rangle \leq a$ , so that the equivalent distributed formulation of the initial

time problem is:

$$\min_{\Phi, \mu} \sum_{t=t_0}^{T-1} \langle \Lambda_t^{\Phi_t}, \mu_t \rangle + \langle K, \mu_T \rangle,$$

subject to the Fokker-Planck dynamics:

$$\mu_{t+1} = \left( A_t^{\Phi_t} \right)^* \mu_t, \quad \forall t = t_0, \dots, T-1,$$

$\mu_{t_0}$  being given by the initial condition, and the final expectation constraint:

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

Even though this problem seems linear with respect to variables  $\mu_t$ , the last constraint introduces an additional highly nonlinear term in the cost function, namely:

$$\chi_{\{\langle g, \mu_T \rangle \leq a\}},$$

where  $\chi_A$  stands for the characteristic function<sup>6</sup> of set  $A$ . The dynamics are still linear and variables  $\mu_t$  are still positive, but the objective function is not linear with respect to  $\mu_T$  anymore, and therefore not linear with respect to the initial law  $\mu_{t_0}$  either. Hence there is no reason for feedback laws to be independent of the initial condition as in the case without constraint presented in §3.

Let us now make a remark on this initial condition. Since the information structure is such that the state variable is fully observed, the initial condition is in fact of a deterministic nature:

$$\mathbf{X}_{t_0} = x_{t_0},$$

where  $x_{t_0}$  is a given (observed) value of the system state. The probability law of  $\mathbf{X}_{t_0}$  is accordingly the Dirac function  $\delta_{x_{t_0}}$ .<sup>7</sup> The reasoning made for the problem initiated at time  $t_0$  remains true for the subsequent problems starting at time  $t_i$ : an observation  $x_{t_i}$  of the state variable  $\mathbf{X}_{t_i}$  becomes available before solving Problem (4), so that its natural initial condition is in fact:

$$\mathbf{X}_{t_i} = x_{t_i}.$$

Otherwise stated, the initial state probability law in each optimization problem we consider should correspond to a Dirac function. Note that such a sequence of Dirac functions is not driven by the Fokker-Planck equation, but is in fact associated to some dynamics of the degenerate filter corresponding to this perfect observation scheme. In the sequel, we assume such an initial condition for every problem we consider.

Now, according to Lemma 2, the subsequent optimization problems formulated at time  $t_i$  will be dynamically consistent provided their initial conditions are given by the optimal Fokker-Planck equation:

$$\mu_{t_0, t_i}^* = \left( A_{t_{i-1}}^{\Phi_{t_0, t_{i-1}}} \right)^* \dots \left( A_{t_0}^{\Phi_{t_0, t_0}} \right)^* \mu_{t_0}.$$

However, except for noise free problems, such a probability law  $\mu_{t_0, t_i}^*$  is always different from a Dirac function, which is, as already explained, the natural initial condition for the subsequent problem starting at time  $t_i$ . As a conclusion, the sequence of problems is not time consistent as long as we consider feedback laws  $\Phi_t$  depending on  $\mathbf{X}_t$  only.

---

<sup>6</sup>as defined in convex analysis:  $\chi_A(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{otherwise} \end{cases}$

<sup>7</sup>The initial law  $\mu_{t_0}$  in Problem (3) corresponds to the information available on  $\mathbf{X}_{t_0}$  before  $\mathbf{X}_{t_0}$  is observed, but it seems more reasonable in a practical situation to use all the available information when setting the problem again at each new initial time, and thus to use a Dirac function as the initial condition.

*Remark 2* (Joint probability constraints). Rather than  $\mathbb{P}(g(\mathbf{X}_T) \geq b) \leq a$ , let us consider a more general chance constraint of the form:

$$\mathbb{P}(g_t(\mathbf{X}_t) \geq b_t, \forall t = t_1, \dots, T) \leq a.$$

This last constraint can be modelled, like the previous one, through an expectation constraint by introducing a new binary state variable:

$$\begin{aligned} \mathbf{Y}_{t_0} &= 1, \\ \mathbf{Y}_{t+1} &= \mathbf{Y}_t \times \mathbf{1}_{\{g_{t+1}(\mathbf{X}_{t+1}) \geq b_{t+1}\}}, \quad \forall t = t_0, \dots, T-1, \end{aligned}$$

and considering constraint  $\mathbb{E}(\mathbf{Y}_T) \leq a$ .  $\square$

**4.2. Back to time consistency.** We now show that time consistency can be recovered provided we choose the right state variable on which to base decisions. We hence establish a link between time consistency of a family of optimization problems and the notion of state variable.

We claim that a better-suited state variable for the family of problems with final time expectation constraint introduced above is the probability law of the variable  $\mathbf{X}$ . Let us denote by  $V_t(\mu_t)$  the optimal cost of the problem starting at time  $t$  with initial condition  $\mu_t$ . Using notations of the distributed formulation of a SOC problem, one can write a DP equation depending on the probability laws  $\mu$  on  $\mathcal{X}$ :

$$V_T(\mu) = \begin{cases} \langle K, \mu \rangle & \text{if } \langle g, \mu \rangle \leq a, \\ +\infty & \text{otherwise,} \end{cases}$$

and, for every  $t = t_0, \dots, T-1$  and every probability law  $\mu$  on  $\mathcal{X}$ :

$$V_t(\mu) = \min_{\Phi_t} \left\langle \Lambda_t^{\Phi_t}, \mu \right\rangle + V_{t+1} \left( \left( A_t^{\Phi_t} \right)^* \mu \right).$$

The context is similar to the one of the deterministic example of §2, and Lemma 3 states that solving the deterministic infinite-dimensional problem associated with the constrained problem leads to time consistency provided DP is used. For the problem under consideration, we thus obtain optimal feedback functions  $\Phi_t$  which depend on the probability laws  $\mu_t$ . Otherwise stated, the family of constrained problems introduced in §4.1 is time consistent provided one looks for strategies as feedback functions depending on both the variable  $\mathbf{X}_t$  and the probability law of  $\mathbf{X}_t$ .

Naturally, this DP equation is rather conceptual. The resolution of such an equation is intractable in practice since probability laws  $\mu_t$  are infinite-dimensional objects.

## 5. CONCLUSION

We informally introduced a notion of time consistency of a sequence of decision-making problems, which basically requires that plans that are made from the very first time remain optimal if one rewrites optimization problems at subsequent time steps. We show that, for several classes of optimal control problems, this concept is not new and can be directly linked with the notion of state variable, which is the minimal information one must use to be able to take the optimal decision.

We show that, 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. This is necessary, for example, when the model contains expectation or chance constraints.

Future works will focus on three main directions. The first concern will be to better formalize the state notion in the vein of the works by Witsenhausen (1971,

1973) and Whittle (1982). The second will be to establish the link with the literature concerning risk measures, in particular the work by Ruszczyński (2009). Finally, the last DP equations we introduced are in general intractable. In a forthcoming paper, we will provide a way to get back to a finite-dimensional information variable, which makes a resolution by DP tractable.

## REFERENCES

- Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., and Ku, H. (2007). Coherent multiperiod risk-adjusted values and Bellman's principle. *Annals of Operations Research*, 152(1):5–22.
- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press, New Jersey.
- Bertsekas, D. (2000). *Dynamic Programming and Optimal Control*. Athena Scientific, 2 edition.
- Cheridito, P., Delbaen, F., and Kupper, M. (2006). Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11(3):57–106.
- Detlefsen, K. and Scandolo, G. (2005). Conditional and dynamic convex risk measures. *Finance and Stochastics*, 9(4):539–561.
- Ekeland, I. and Lazrak, A. (2006). Being serious about non-commitment: subgame perfect equilibrium in continuous time. *arXiv*, math.OA 0604264.
- Hammond, P. J. (1989). Consistent plans, consequentialism, and expected utility. *Econometrica*, 57(6):1445–1449.
- Henrion, R. (2002). On the connectedness of probabilistic constraint sets. *Journal of Optimization Theory and Applications*, 112(3):657–663.
- Henrion, R. and Strugarek, C. (2008). Convexity of chance constraints with independent random variables. *Computational Optimization and Applications*, 41(2):263–276.
- Prékopa, A. (1995). *Stochastic Programming*. Kluwer, Dordrecht.
- Riedel, F. (2004). Dynamic coherent risk measures. *Stochastic Processes and their Applications*, 112(2):185 – 200.
- Rockafellar, R. and Wets, R.-B. (1998). *Variational Analysis*. Springer Verlag, Berlin Heidelberg.
- Ruszczyński, A. (2009). Risk-averse dynamic programming for markov decision processes. *Optimization Online*, to appear in Mathematical Programming.
- Ruszczyński, A. and Shapiro, A., editors (2003). *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier.
- Shapiro, A. (2009). On a time consistency concept in risk averse multistage stochastic programming. *Operations Research Letters*, 37(3):143 – 147.
- Whittle, P. (1982). *Optimization over time*. John Wiley & Sons.
- Witsenhausen, H. S. (1971). On information structures, feedback and causality. *SIAM Journal on Control*, 9(2):149–160.
- Witsenhausen, H. S. (1973). A standard form for sequential stochastic control. *Mathematical Systems Theory*, 7(1):5–11.

P. CARPENTIER, ENSTA PARISTECH, 32, BOULEVARD VICTOR, 75739 PARIS CEDEX 15, FRANCE.

*E-mail address:* pierre.carpentier@ensta.fr

J.-PH. CHANCELIER, UNIVERSITÉ PARIS-EST, CERMICS, ÉCOLE DES PONTS PARISTECH, 6 & 8 AVENUE BLAISE PASCAL, 77455 MARNE-LA-VALLÉE CEDEX 2.

*E-mail address:* jpc@cermics.enpc.fr

G. COHEN, UNIVERSITÉ PARIS-EST, CERMICS, ÉCOLE DES PONTS PARISTECH, 6 & 8 AVENUE BLAISE PASCAL, 77455 MARNE-LA-VALLÉE CEDEX 2.

*E-mail address:* guy.cohen@mail.enpc.fr

M. DE LARA, UNIVERSITÉ PARIS-EST, CERMICS, ÉCOLE DES PONTS PARISTECH, 6 & 8 AVENUE BLAISE PASCAL, 77455 MARNE-LA-VALLÉE CEDEX 2.

*E-mail address:* delara@cermics.enpc.fr

P. GIRARDEAU, EDF R&D, 1, AVENUE DU GÉNÉRAL DE GAULLE, F-92141 CLAMART CEDEX, FRANCE, ALSO WITH UNIVERSITÉ PARIS-EST, CERMICS AND ENSTA.

*E-mail address:* pierre.girardeau@cermics.enpc.fr