Dynamic Consistency for Stochastic Optimal Control Problems

Cadarache Summer School CEA/EDF/INRIA 2012

Pierre Carpentier Jean-Philippe Chancelier Michel De Lara
SOWG

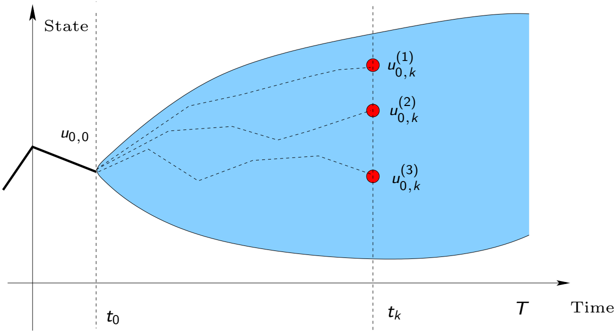
June 2012

Lecture outline

Informal definition

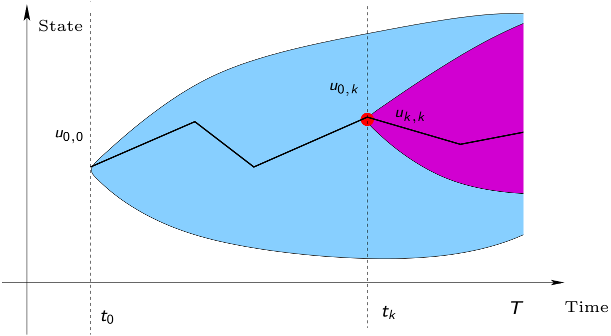
- For a sequence of dynamic optimization problems, we aim at discussing a notion of consistency over time.
- At the very first time step t_0, formulate an optimization problem that yields optimal decision rules for all the forthcoming time steps $t_0, t_1, \dots, T$;
- At the next time step t_1, formulate a new optimization problem starting at time t_1 that yields a new sequence of optimal decision rules. This process can be continued until the final time T is reached.
- A family of optimization problems is said to be dynamically consistent if the optimal strategies obtained when solving the original problem remain optimal for all subsequent problems.

Decision at time t_0



depends on decisions computed at later times for possible future states.

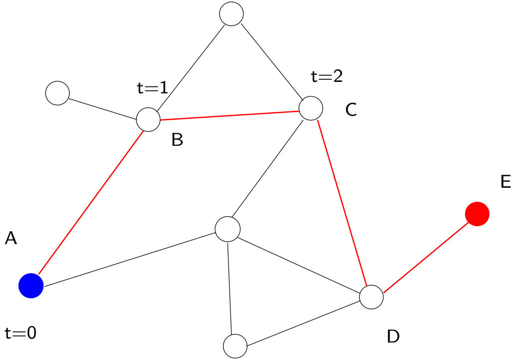
Time consistent decision at time t_k



Decision $u_{k,k}$ computed using a problem formulated at time t_k knowing "where I am" at time t_k or decision $u_{0,k}$ computed at time t_0 and to be applied at time t_k for the "same position" should be the same.

An example: deterministic shortest path

If $A \to B \to C \to D \to E$ is optimal then $C \to D \to E$ is optimal.
Arriving at C, If I behave rationally I should not deviate from the originally chosen route starting at A.



Dynamic Programming and dynamically consistent problems

- Dynamic Programing gives dynamically consistent controls
- It gives a family of problems indexed by time: the solution of problem at time t uses the solution of problem at time $t+1$:

$$V_k(x) = \min_u \quad \mathbb{E}\left[L_k\left(x, u, \mathbf{W}_{k+1}\right) + V_{k+1}\left(f_{t_k}(x, u, \mathbf{W}_{k+1})\right)\right]$$

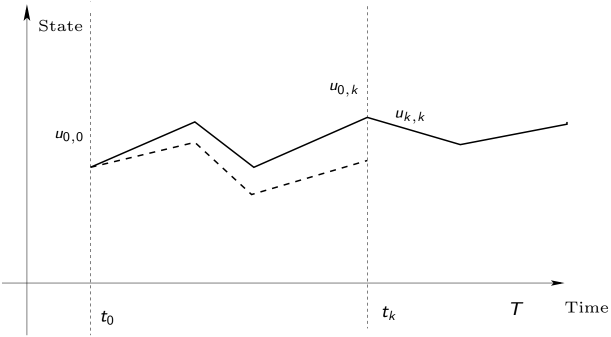
Choose control at time t_k optimizing the sum of a current cost and of a running cost. The running cost has an interpretation as the optimal cost function of a problem at time t_{k+1}.

Dynamic Programming and dynamically consistent problems (2)

- The problem at time t_{k+1} is the following:

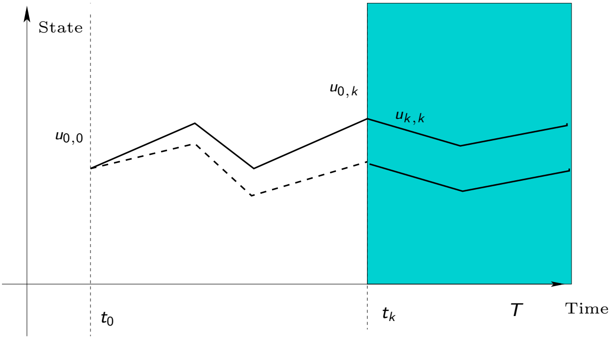
$$
\begin{aligned}
V_{k+1}(x) = \min_{\mathbf{X},\mathbf{U}} \quad & \mathbb{E}\left[\sum_{t=t_{k+1}}^{T-1} L_t\left(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\right) + K\left(\mathbf{X}_T\right) \,|\mathbf{X}_{t_{k+1}} = x \right], \\
\text{s.t.} \quad & \mathbf{X}_{t+1} = f_t\left(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\right), \quad \forall t = t_{k+1}, \dots, T-1, \\
& \mathbf{U}_t \preceq \mathbf{X}_t \qquad \forall t = t_{k+1}, \dots, T-1.
\end{aligned}
$$

Deterministic case



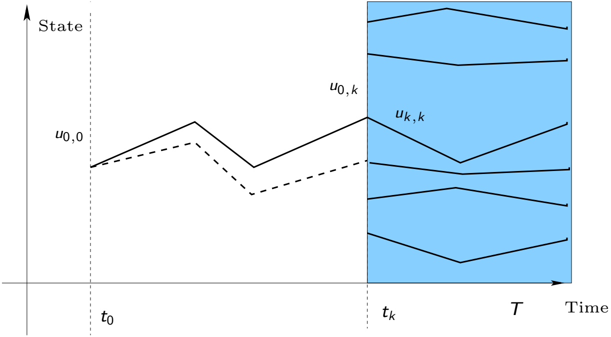
Exact state match at time t_k

Deterministic case



Independence of the initial condition.

Deterministic case



Taking care of all possible initial states at time t_k through state feedback.

A first example in the deterministic case (1)

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

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

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.

- 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.

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.

Independence of the initial condition

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

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

Independence of the initial condition (2)

Variables x_t can be recursively replaced using dynamics g_t leading to:

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

- The optimal cost is linear with respect to x_{t_0}.
- Suppose that x_{t_0} only takes positive values. Then x_{t_0} has no influence on the minimizer.
- Remains true at subsequent time steps provided that dynamics are such that x_t remains positive for every time step.
- The dynamic consistency property holds true.
- This example which may look very special but we will see later on that it is analogous to familiar SOC problems.

The guiding principle of the lecture

To obtain Dynamical Consistency, let the decision rules depend upon a sufficiently rich information set.

A classical SOC Problem

Control variables $\mathbf{U} = (\mathbf{U}_t)_{t=t_0,\dots,T-1}$. Noise variables $\mathbf{W} = (\mathbf{W}_t)_{t=t_1,\dots,T}$. Markovian setting: noise variables $\mathbf{X}_{t_0}, \mathbf{W}_{t_1}, \dots, \mathbf{W}_T$ are independent.
The problem $(\mathcal{S}_{t_0})$ starting at t_0 writes:

$$
\begin{aligned}
\min_{\mathbf{X},\mathbf{U}} \quad & \mathbb{E}\left[\sum_{t=t_0}^{T-1} L_t\left(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\right) + K\left(\mathbf{X}_T\right)\right], \\
\text{s.t.} \quad & \mathbf{X}_{t_0} \text{ given}, \\
& \mathbf{X}_{t+1} = f_t\left(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\right), \\
& \mathbf{U}_t \preceq \mathbf{X}_{t_0}, \mathbf{W}_{t_1}, \dots, \mathbf{W}_t, \qquad \forall t = t_0, \dots, T-1,
\end{aligned}
\tag{1}
$$

We know that there is no loss of optimality in looking for the optimal strategy as a state feedback control $\mathbf{U}_t^\sharp = \phi_t(\mathbf{X}_t)$.

Stochastic optimal control: the classical case (2)

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:

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

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

An equivalent distributed formulation

- Ξ_t a linear space of $\mathbb{R}$-valued measurable functions on $\mathcal{X}_t$.
- Υ_t space of signed measures on $\mathcal{X}_t$.
- We consider a dual pairing between Ξ_t and Υ_t by considering the bilinear form:

$$\langle \xi, \mu \rangle = \int_{\mathcal{X}_t} \xi(x) \mathrm{d}\mu(x) \quad \text{for } \xi \in \Xi_t \text{ and } \mu \in \Upsilon_t \, .$$

- When $\mathbf{X}_t$ is a random variable taking values in $\mathcal{X}_t$ and distributed according to μ_t. Then

$$\langle \xi_t, \mu_t \rangle = \mathbb{E}\big(\xi_t(\mathbf{X}_t)\big) \, .$$

Fokker-Planck equation: state probability law dynamics

For a given feedback laws $\phi_t : \mathcal{X}_t \to \mathcal{U}_t$ we define $\mathbb{T}_t^{\phi_t} : \Xi_{t+1} \to \Xi_t$:

$$\left(\mathbb{T}_t^{\phi_t} \xi_{t+1}\right)(x) \overset{\text{def}}{=} \mathbb{E}\left(\xi_{t+1}(\mathbf{X}_{t+1}) \mid \mathbf{X}_t = x\right) \quad \forall x \in \mathcal{X}_t \; .$$

Using state dynamics and Markovian setting we obtain:

$$\left(\mathbb{T}_t^{\phi_t} \xi_{t+1}\right)(x) = \mathbb{E}\Big(\xi_{t+1}\big(f_t(x, \phi_t(x), \mathbf{W}_{t+1})\big)\Big) \quad \forall x \in \mathcal{X}_t \; .$$

Denoting by $\left(\mathbb{T}_t^{\phi_t}\right)^\star$ the adjoint operator of $\mathbb{T}_t^{\phi_t}$

$$\left\langle \mathbb{T}_t^{\phi} \xi \, , \mu \right\rangle = \left\langle \xi \, , \left(\mathbb{T}_t^{\phi}\right)^\star \mu \right\rangle$$

The state probability law driven by the chosen feedback follows:

$$\mu_{t+1} = \left(\mathbb{T}_t^{\phi_t}\right)^\star \mu_t \quad \forall t = t_0, \ldots, T-1, \quad \mu_{t_0} \text{ given} \; .$$

Revisiting the expected cost

- Next we introduce the operator $\Lambda_t^{\phi_t} : \mathcal{X}_t \to \mathbb{R}$:

$$\Lambda_t^{\phi_t}(x) \stackrel{\text{def}}{=} \mathbb{E}\Big(L_t\big(x, \phi_t(x), \mathbf{W}_{t+1}\big) \Big)$$

  meant to be the expected cost at time t for each possible state value when feedback function ϕ_t is applied.
- Using $\Lambda_t^{\phi_t}$ we have:

$$\mathbb{E}\Big(L_t\big(\mathbf{X}_t, \phi_t(\mathbf{X}_t), \mathbf{W}_{t+1}\big) \Big) = \mathbb{E}\big(\Lambda_t^{\phi_t}(\mathbf{X}_t)\big) = \big\langle \Lambda_t^{\phi_t}, \mu_t \big\rangle .$$

Equivalent deterministic infinite-dimensional problem

We obtain an equivalent deterministic infinite-dimensional optimal control problem $(\mathcal{D}_{t_0})$:

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

Very similar to the Independence of the initial condition example

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

$$\begin{aligned}\mathcal{V}_{T-1}(\mu) &= \min_{\phi} \left\langle \Lambda^{\phi}_{T-1} + A^{\phi}_{T-1} K \, , \mu \right\rangle , \\ &= \min_{\phi(\cdot)} \int_{\mathbb{X}} \big(\Lambda^{\phi}_{T-1} + A^{\phi}_{T-1} K \big)(x) \mu(\mathrm{d}x).\end{aligned}$$

Solving $(\mathcal{D}_{t_0})$ using Dynamic Programming (2)

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.

Constraints in stochastic optimal control

Different kinds of constraints in stochastic optimization:

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

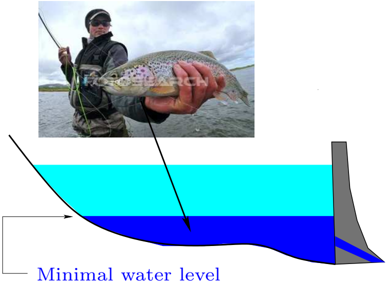
A chance constraint can be modelled as an expectation constraint:

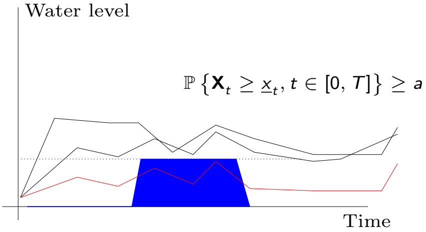
$$\mathbb{P}\big(g(\mathbf{X}_T) \leq a\big) = \mathbb{E}\big(\mathbf{1}_{\mathbb{X}^{\mathrm{ad}}}(\mathbf{X}_T)\big),$$

(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.

Probability constraint on state trajectories





Transformed into a Probability constraints on final state : $\mathbb{E}\left[g(\tilde{\mathbf{X}}_T)\right] \leq a$

- Motivated by a joint probability constraint:

$$\mathbb{P}\left\{\gamma_t\left(\mathbf{X}_t\right) \geq b_t, \forall t = t_0, \ldots, T\right\} \geq a.$$

- Introducing a new binary state variable $\mathbf{Y}_t \in \{0,1\}$:

$$\mathbf{Y}_{t_0} = 1, \quad \mathbf{Y}_{t+1} = \mathbf{Y}_t \times \mathbf{1}_{\left\{\gamma_{t+1}\left(\mathbf{X}_{t+1}\right) \geq b_{t+1}\right\}},$$

$$\mathbb{E}\left[\mathbf{Y}_T\right] = 1 \times \mathbb{P}\left\{\gamma_t\left(\mathbf{X}_t\right) \geq b_t, \forall t = t_0, \ldots, T\right\} \geq a$$

Problem setting with expectation constraint

Control variables $\mathbf{U} = (\mathbf{U}_t)_{t=t_0,\dots,T-1}$. Noise variables $\mathbf{W} = (\mathbf{W}_t)_{t=t_1,\dots,T}$. Markovian setting: noises variables $\mathbf{X}_{t_0}, \mathbf{W}_{t_1}, \dots, \mathbf{W}_T$ are independent.
The problem starting at t_0 writes:

$$\begin{aligned}
\min_{\mathbf{X},\mathbf{U}} \quad & \mathbb{E}\left[\sum_{t=t_0}^{T-1} L_t\left(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\right) + K\left(\mathbf{X}_T\right)\right], \\
\text{s.t.} \quad & \mathbf{X}_{t_0} \text{ given}, \\
& \mathbf{X}_{t+1} = f_t\left(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\right), \\
& \mathbf{U}_t \preceq \mathbf{X}_{t_0}, \mathbf{W}_{t_1}, \dots, \mathbf{W}_t, \qquad \forall t = t_0, \dots, T-1, \\
& \mathbb{E}\left[g\left(\mathbf{X}_T\right)\right] \leq a, \quad a \in \mathbb{R}
\end{aligned}$$

Problem at time t_i

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

- Not dynamically consistent with the usual state variable. With new appropriate state variable, one regains dynamical consistency.
- Optimal strategy at time t is a function of $\Phi_{t_0,t} : \mathcal{X} \to \mathcal{U}$ since information structure is not modified by the last constraint.

Equivalent deterministic infinite-dimensional problem

We obtain an equivalent deterministic infinite-dimensional optimal control problem.

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

$$\text{s.t.} \quad \mu_{t_0} \text{ given,}$$

$$\mu_{t+1} = \left(\mathbb{T}_t^{\Phi_t}\right)^{\star} \mu_t, \qquad \forall t = t_0, \ldots, T-1,$$

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

The last constraint implies that the optimal feedback laws depend on the initial condition μ_{t_0} and we do not have dynamical consistency.

Simple case-study: a discrete controlled Markov chain

- The state space $\mathbb{X} = \{1, 2, 3\}$: Discrete possible values of a reservoir level. $\underline{x} = 1$ (resp. $\overline{x} = 3$) the lower (resp. upper) level of the reservoir.
- The control action u takes values in $\mathbb{U} = \{0, 1\}$, the value 1 corresponding to using some given water release to produce some electricity.
- The noise variable: stochastic inflow takes its values in $\mathbb{W} = \{-1, 0, 1\}$.
- The discrete probability law of each noise variable $\mathbf{W}_t$ being characterized by the weights $\{\varpi_-, \varpi_0, \varpi_+\}$.
- Finally, the reservoir dynamics is:

$$\mathbf{X}_{t+1} = \min\left(\overline{x}, \max(\underline{x}, \mathbf{X}_t - \mathbf{U}_t + \mathbf{W}_{t+1})\right).$$

Simple case-study: a discrete controlled Markov chain

Let μ_t be the discrete probability law associated with the state random variable $\mathbf{X}_t$, the initial probability law μ_{t_0} being given. In such a discrete case, it is easy to compute the transition matrix P^u giving the Markov chain transitions for each possible value u of the control, with the following interpretation:

$$
\begin{aligned}
P^u_{ij} &= \mathbb{P}\big(\mathbf{X}_{t+1} = j \,\big|\, \mathbf{X}_t = i, \mathbf{U}_t = u\big) \\
&= \mathbb{P}\Big(\big\{ \min\big(\overline{x}, \max(\underline{x}, i - u + \mathbf{W}_{t+1})\big) = j \big\} \Big) .
\end{aligned}
$$

Simple case-study: a discrete controlled Markov chain

We obtain for the reservoir problem:

$$P^0 = \begin{pmatrix} \varpi_- + \varpi_0 & \varpi_+ & 0 \\ \varpi_- & \varpi_0 & \varpi_+ \\ 0 & \varpi_- & \varpi_0 + \varpi_+ \end{pmatrix} \quad , \quad P^1 = \begin{pmatrix} 1 & 0 & 0 \\ \varpi_- + \varpi_0 & \varpi_+ & 0 \\ \varpi_- & \varpi_0 & \varpi_+ \end{pmatrix}$$

and we denote by P_i^u the i^{th} row of matrix P^u.

Simple case-study: a discrete controlled Markov chain

Let $\phi_t : \mathbb{X} \to \mathbb{U}$ be a feedback law at time t. We denote by $\mathbf{\Phi}_t$ the set of admissible feedbacks at time t. In our discrete case, $\mathrm{card}(\mathbf{\Phi}_t) = \mathrm{card}(\mathbb{U})^{\mathrm{card}(\mathbb{X})} = 8$ for all t. The transition matrix P^{ϕ_t} associated with such a feedback is obtained by properly selecting rows of the transition matrices P^u, namely:

$$P^{\phi_t} = \begin{pmatrix} P_1^{\phi_t(1)} \\ P_2^{\phi_t(2)} \\ P_3^{\phi_t(3)} \end{pmatrix} .$$

Then the dynamics given by the Fokker-Planck equation writes:

$$\mu_{t+1} = \left(P^{\phi_t}\right)^\top \mu_t ,$$

and the state μ_t involved in this dynamic equation is a three-dimensional column vector $\mu_t = (\mu_{1,t}, \mu_{2,t}, \mu_{3,t})^\top$

- The cost at time t is supposed to be linear, equal to $p_t \mathbf{U}_t$, p_t being a (negative) deterministic price.
- No final cost ($K \equiv 0$).
- The reservoir level at final time T must be equal to $\overline{x}$ with a probability level at least equal to π.

The distributed formulation associated with the reservoir control problem is:

$$\min_{\{\phi_t \in \boldsymbol{\Phi}_t\}_{t=t_0}^{T-1}} \quad \sum_{t=t_0}^{T-1} p_t \big\langle \phi_t \, , \mu_t \big\rangle \, , \tag{6a}$$

$$\text{subject to:} \quad \mu_{t+1} = \big(P^{\phi_t}\big)^\top \mu_t \, , \quad \mu_{t_0} \text{ given} \, , \tag{6b}$$

$$\big\langle \mathbf{1}_{\{\overline{x}\}} \, , \mu_T \big\rangle \geq \pi \, , \tag{6c}$$

with $\mathbf{1}_{\{\overline{x}\}}(x) = 1$ if $x = \overline{x}$, and 0 otherwise.

- The previous problem can be solved using DP: the state equation is a two-dimensional ($\mu_{1,t} + \mu_{2,t} + \mu_{3,t} = 1$)
- Suppose now that the number of reservoir discrete levels is n (rather than 3), with n big enough: the resolution of the distributed formulation suffers from the curse of dimensionality ($(n-1)$-dimensional state),
- The ultimate curse being to consider that the level takes values in the interval $[\underline{x}, \overline{x}]$, so that the μ_t's are continuous probability laws.

Back to Dynamical Consistency

We can use Dynamic Programing:
Let $V_t(\mu_t)$ be the optimal cost of the problem starting at time t with initial condition μ_t.

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

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

$$V_t(\mu) = \min_{\Phi_t} \left\langle \Lambda_t^{\Phi_t}, \mu \right\rangle + V_{t+1}\left(\left(\mathbb{T}_t^{\Phi_t} \right)^{\star} \mu \right).$$

Optimal feedback functions Φ_t depend on μ_t. Consistency recovered for strategies depending on $\mathbf{X}_t$ and μ_t.

Back to Dynamical Consistency (2)

- Optimal feedback functions Φ_t depend on μ_t

$$\min_{\Phi_t} \left\langle \Lambda_t^{\Phi_t}, \mu \right\rangle + V_{t+1}\left(\left(\mathbb{T}_t^{\Phi_t} \right)^{\star} \mu \right) .$$

- But it also depends on $\mathbf{X}_t$ because of the distributed formulation $\mathbf{U}_t = \phi_t(\mathbf{X}_t)$!
- Consistency recovered for strategies depending on $\mathbf{X}_t$ and μ_t.
- The new state variable to consider (X_t, μ_t) is an infinite dimensional object !

Equivalent problem: added state process $\mathbf{Z}$ and control $\mathbf{V}$

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

$$\begin{aligned} \mathbf{X}_{t_0} &= x_{t_0}, & \mathbf{X}_{t+1} &= f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}), \\ \mathbf{Z}_{t_0} &= 0, & \mathbf{Z}_{t+1} &= \mathbf{Z}_t + \mathbf{V}_t, \end{aligned}$$

- Almost sure final constraint: $g(\mathbf{X}_T) - \mathbf{Z}_T \leq a$..
- Measurability constraints: $\mathbf{U}_t \preceq \mathcal{F}_t$, $\mathbf{V}_t \preceq \mathcal{F}_{t+1}$.
- Additional time constraints: $\mathbb{E}\big(\mathbf{V}_t \mid \mathcal{F}_t\big) = 0$.

Equivalent problem (2)

- Let $(\mathbf{U}_{t_0},\dots,\mathbf{U}_{T-1})$ and $(\mathbf{X}_{t_0},\dots,\mathbf{X}_T)$ satisfying the constraints of Initial problem.
- We define $(\mathbf{V}_{t_0},\dots,\mathbf{V}_{T-1})$ and $(\mathbf{Z}_{t_0},\dots,\mathbf{Z}_T)$ as follow:

$$\mathbf{V}_t = \mathbb{E}\big(g(\mathbf{X}_T) \mid \mathcal{F}_{t+1}\big) - \mathbb{E}\big(g(\mathbf{X}_T) \mid \mathcal{F}_t\big), \quad \mathbf{Z}_{t+1} = \mathbf{Z}_t + \mathbf{V}_t.$$

  Noting that the σ-field $\mathcal{F}_{t_0}$ is the minimal σ-field and that the random variable $\mathbf{X}_T$ is $\mathcal{F}_T$-measurable, we have:

$$\mathbf{Z}_T = g(\mathbf{X}_T) - \mathbb{E}\big(g(\mathbf{X}_T)\big).$$

- Hence, **U**, **X**, **V** and **Z** satisfy the set of constraints of the new problem.

Equivalent problem (3)

- Let $(\mathbf{U}_{t_0}, \dots, \mathbf{U}_{T-1})$, $(\mathbf{X}_{t_0}, \dots, \mathbf{X}_T)$, $(\mathbf{V}_{t_0}, \dots, \mathbf{V}_{T-1})$ and $(\mathbf{Z}_{t_0}, \dots, \mathbf{Z}_T)$ satisfying the constraints of $(\mathbf{V}, \mathbf{Z})$-problem be given.
- Then, $\mathbf{Z}_T = \sum_{t=t_0}^{T-1} \mathbf{V}_t$.
- Thus we have:

$$\mathbb{E}\big(\mathbf{Z}_T\big) = \sum_{t=t_0}^{T-1} \mathbb{E}\Big(\mathbb{E}\big(\mathbf{V}_t \mid \mathcal{F}_t\big)\Big) = 0.$$

- If follows that $g(\mathbf{X}_T) - \mathbf{Z}_T \leq a \Rightarrow \mathbb{E}\big(g(\mathbf{X}_T)\big) \leq a$, and therefore the processes $(\mathbf{U}_{t_0}, \dots, \mathbf{U}_{T-1})$ and $(\mathbf{X}_{t_0}, \dots, \mathbf{X}_T)$ satisfy the constraints of Initial problem

To conclude, since the two problems share the same sets as admissible constraints and share the same criteria to optimize, they are equivalent.

Equivalent problem (4)

1. Dynamic Programming will give feedbacks as functions of $(\mathbf{X}_t, \mathbf{Z}_t)$.
2. The new problem is more intricate:
   - added state and added control,
   - $\mathbf{V}_t$ depends on $\mathbf{W}_{t+1}$ (*Hazard–Decision*)
   - some new constraints on the controls are to be taken into account.
3. $\mathbb{E}\big(g(\mathbf{X}_T) \bigm| \mathcal{F}_t\big)$: perception of the risk constraint at time t.
4. $\mathbf{V}_t^\sharp = \mathbb{E}\big(g(\mathbf{X}_T) \bigm| \mathcal{F}_{t+1}\big) - \mathbb{E}\big(g(\mathbf{X}_T) \bigm| \mathcal{F}_t\big)$: variation between time t and time $t+1$ of this perception.
5. $\mathbf{Z}_t^\sharp$: cumulative variation over time of the risk constraint.

Dynamic programing

In order to use Dynamic programming on problem (7), we will first recall on a simplified model how Dynamic Programming works on a classical SOP.

$$
\begin{aligned}
\min_{\mathbf{X},\mathbf{U}} \quad & \mathbb{E}\left[\sum_{t=t_0}^{T-1} L_t\left(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\right) + K\left(\mathbf{X}_T\right)\right], \\
\text{s.t.} \quad & \mathbf{X}_{t_0} \text{ given}, \\
& \mathbf{X}_{t+1} = f_t\left(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\right), \\
& \mathbf{U}_t \preceq \mathcal{F}_t \quad \forall t = t_0, \dots, T-1,
\end{aligned}
$$

Then we will show the modifications induced by hazard-decision control or $\mathbb{E}\left(\mathbf{V}_t \mid \mathcal{F}_t\right) = 0$.

Dynamic programing steps in classical framework

1. Rewrites criterion using conditional expectations:

$$\mathbb{E}\left(\mathbb{E}\left(L_0(\mathbf{X}_0, \mathbf{U}_0) + \mathbb{E}\left(L_1(\mathbf{X}_1, \mathbf{U}_1) + K(\mathbf{X}_2) \;\middle|\; \mathcal{F}_1\right) \;\middle|\; \mathcal{F}_0\right)\right)$$

2. minimization with respect to $\mathbf{U}_t$ and $\mathbf{X}_t$ enters up to the conditional expectation with respect to $\mathcal{F}_t$.
3. Recursive computation from the most internal problem to the outer one ($t = t_0$). The most internal ($t = T - 1$):

$$\min_{\mathbf{U}_{T-1} \preceq \mathcal{F}_{T-1}} \mathbb{E}\left(h(\mathbf{X}_{T-1}, \mathbf{U}_{T-1}, \mathbf{W}_T) \;\middle|\; \mathcal{F}_{T-1}\right).$$

$$h(x, u, w) = L_{T-1}(x, u) + K\left(f_{T-1}(x, u, w)\right)$$

Hazard–Decision framework

- Hazard–Decision: $\mathbf{U}_t \preceq \mathcal{F}_{t+1}$
- Interchange of expectation and minimization can be done one step further.

$$\mathbb{E}\left(\min_{\mathbf{U}_{T-1} \preceq \mathcal{F}_T} h(\mathbf{X}_{T-1}, \mathbf{U}_{T-1}, \mathbf{W}_T) \;\middle|\; \mathcal{F}_{T-1} \right).$$

with

$$h(x, u, w) = L_{T-1}(x, u, w) + K\big(f_{T-1}(x, u, w)\big)$$

- Optimal feedback $\mathbf{U}^{\sharp}_{T-1}$ can be chosen as a $(\mathbf{X}_{T-1}, \mathbf{W}_T)$-measurable function.

Hazard–Decision framework with constraint

- Control constraint: $\mathbb{E}\big(\mathbf{U}_t \bigm| \mathcal{F}_t\big) = 0\,.$
- We are led to the following problem at time $t = T-1$:

$$\min_{\mathbf{U}_{T-1} \preceq \mathcal{F}_T} \mathbb{E}\Big(h(\mathbf{X}_{T-1}, \mathbf{U}_{T-1}, \mathbf{W}_T) \Bigm| \mathcal{F}_{T-1}\Big)\,,$$
$$\text{subject to:} \quad \mathbb{E}\big(\mathbf{U}_{T-1} \bigm| \mathcal{F}_{T-1}\big) = 0\,.$$

- Optimal feedback $\mathbf{U}^\sharp_{T-1}$ can still be chosen as a $(\mathbf{X}_{T-1}, \mathbf{W}_T)$-measurable function?
- Yes and the proof is actually available when $\mathcal{F}_{T-1}$ is generated by finite valued random variables.

Back to the equivalent problem

The Hazard-Decision ($\mathbf{V}$) and Decision-Hazard ($\mathbf{U}$) controls.

- Initialization of the Bellman function at time T:

$$V_T(x,z) = K(x) + \chi_{G^{\mathrm{ad}}}(x,z) \, ,$$

$\chi_{G^{\mathrm{ad}}}$ characteristic function of $G^{\mathrm{ad}} = \big\{(x,z) \mid g(x) - z \leq a\big\}$;

- Bellman function at time t:

$$V_t(x,z) = \min_{u \in \mathbb{U}} \min_{\mathbf{V} \preceq \mathbf{W}_{t+1}} \mathbb{E}\Big(h(x,u,z,\mathbf{V},\mathbf{W}_{t+1})\Big) \, ,$$

$$\text{subject to:} \quad \mathbb{E}\big(\mathbf{V}\big) = 0 \, .$$

$h(x,u,z,v,w) \stackrel{\mathrm{def}}{=} L_t(x,u,w) + V_{t+1}\big(f_t(x,u,w), z+v\big)$
$\mathbf{U}_t^\sharp \preceq (\mathbf{X}_t, \mathbf{Z}_t)$ and $\mathbf{V}_t^\sharp \preceq (\mathbf{X}_t, \mathbf{Z}_t, \mathbf{W}_{t+1})$.