Dynamic Programming

Michel De Lara
Cermics, École nationale des ponts et chaussées
IP Paris, France

January 14, 2025

Outline of the presentation

Dam management

- Managed every day, during five years (1,825 stages)
- at stage t (beginning of period $[t, t+1[$), turbining decision Q_t
- during period $[t, t+1[$, random water inflow A_{t+1}

Outline of the presentation

Multistage stochastic optimization

Dynamic programming without state

Dynamic programming with state

Dynamic programming with state and white noise

Dynamic Programming With State and White Noise (Complements)

Stages, decisions, uncertainties

Basic data

Basic data

- Let $(\Omega, \mathcal{A}_\infty, \mathbb{P})$ be a probability space
- Let $T \in \mathbb{N}^*$ be the horizon
- For stages $t = 0, \ldots, T-1$, let $\mathcal{U}_t$ be the control set, a measurable set equipped with σ-field $\mathfrak{U}_t$
- For stages $t = 0, \ldots, T$, let $\mathcal{W}_t$ be the uncertainty set, a measurable set equipped with σ-fields $\mathfrak{W}_t$

In the dam example

- horizon $T = 1,825$
- control set $\mathcal{U}_t = \mathbb{R}$ (equipped with Borel σ-field $\mathfrak{U}_t = \mathcal{B}_{\mathbb{R}}^{\text{o}}$), decision $Q_t = u_t \in \mathcal{U}_t = \mathbb{R}$
- uncertainty set $\mathcal{W}_t = \mathbb{R}$ (equipped with Borel σ-field $\mathfrak{W}_t = \mathcal{B}_{\mathbb{R}}^{\text{o}}$), uncertainty $A_t = w_t \in \mathcal{W}_t = \mathbb{R}$

History space

For $t = 0, \ldots, T$, we define

- the history space $\mathcal{H}_t$

$$\mathcal{H}_t = \mathcal{W}_0 \times \prod_{s=0}^{t-1} (\mathcal{U}_s \times \mathcal{W}_{s+1})$$

  equipped with the history field $\mathfrak{H}_t$

$$\mathfrak{H}_t = \mathfrak{W}_0 \otimes \bigotimes_{s=0}^{t-1} (\mathfrak{U}_s \otimes \mathfrak{W}_{s+1})$$

- A generic element $h_t \in \mathcal{H}_t$ is called a history

$$\begin{aligned}
h_t &= (w_0, u_0, w_1, u_1, w_2, \ldots, u_{t-2}, w_{t-1}, u_{t-1}, w_t) \\
[h_t] &= \big[(w_0, (u_s, w_{s+1})_{s=0,\ldots,t-1})\big] = (w_0, \ldots, w_t) = w_{[0:t]} \\
h_{s:t} &= (u_r, w_{r+1})_{r=s-1,\ldots,t-1} = (u_{s-1}, w_s, \ldots, u_{t-1}, w_t) \\
[h_{s:t}] &= [(u_r, w_{r+1})_{r=s-1,\ldots,t-1}] = (w_s, \ldots, w_t) = w_{[s:t]}
\end{aligned}$$

Noise and noise process

- For $t = 0, \ldots, T$, let
$$\mathbf{W}_t : \Omega \to \mathcal{W}_t$$
be a random variable taking values in $\mathcal{W}_t$ (noise)
- We introduce the past noises, or noise process up to stage t as
$$\mathbf{W}_{[0:t]} = (\mathbf{W}_0, \ldots, \mathbf{W}_t) \in \mathcal{W}_{[0:t]} = \prod_{s=0}^{t} \mathcal{W}_s$$

Nonanticipative solutions (of what?)

Noise filtration and adapted processes

- We introduce the filtration $\mathcal{A} = \big(\mathcal{A}_t\big)_{t=0,\ldots,T-1}$ defined by

$$\mathcal{A}_t = \sigma(\mathbf{W}_0, \ldots, \mathbf{W}_t) \;, \quad t = 0, \ldots, T-1$$

- Let $\mathbb{L}^0(\Omega, \mathcal{A}, \prod_{s=0}^{T-1} \mathcal{U}_s)$ be the space of $\mathcal{A}$-adapted processes $(\mathbf{U}_0, \ldots, \mathbf{U}_{T-1})$ with values in $\prod_{s=0}^{T-1} \mathcal{U}_s$, that is, such that

$$\sigma(\mathbf{U}_0) \subset \mathcal{A}_0, \ldots, \sigma(\mathbf{U}_t) \subset \mathcal{A}_t, \ldots, \sigma(\mathbf{U}_{T-1}) \subset \mathcal{A}_{T-1}$$

Nonanticipativity constraints

$$\boxed{\sigma(\mathbf{U}_t) \subset \sigma(\mathbf{W}_0, \ldots, \mathbf{W}_t) \;, \quad \forall t = 0, \ldots, T-1}$$

Doob Theorem

If every control set $\mathcal{U}_t$ is a separable complete metric space, for $t = 0, \ldots, T-1$, the condition

$$\big(\mathbf{U}_0, \ldots, \mathbf{U}_{T-1}\big) \in \mathbb{L}^0_{\mathcal{A}}(\Omega, \prod_{s=0}^{T-1} \mathcal{U}_s)$$

is equivalent to the existence of measurable mappings

$$\gamma_t : \mathcal{W}_{0:t} \to \mathcal{U}_t \; , \quad t = 0, \ldots, T-1$$

such that

$$\mathbf{U}_t = \gamma_t(\mathbf{W}_{0:t}) = \gamma_t(\mathbf{W}_0, \ldots, \mathbf{W}_t) \; , \quad t = 0, \ldots, T-1$$

Curse of scenario expansion

Assuming that

- the control u_t can take N_u values
- the uncertainty w_t can take N_w values

we obtain

- N_w^T scenarios
- $1 + N_w + \cdots + N_w^T$ nodes in the scenario tree
- $N_u \times \frac{N_w^{T+1}-1}{N_w-1} \approx N_u N_w^T$ elements in the solution space

so that the number of possible solutions
grows exponentially with the number T of stages

Criterion and optimization

In dam management, the traditional economic problem is maximizing the expected payoff

- Suppose that
  - a probability $\mathbb{P}$ is given on the set $\mathbb{R}^T$ of water inflows scenarios $(A_0, \dots, A_{T-1})$
  - turbined water Q_t is sold at price p_t, related to the price at which energy can be sold at time t
  - at the horizon, the final volume S_T has a value $K(S_T)$, the "final value of water"
- The traditional economic problem is to maximize the intertemporal payoff (without discounting if the horizon is short)

$$\max \mathbb{E}\Big[\sum_{t=0}^{T-1} \overbrace{p_t Q_t}^{\text{turbined water payoff}} + \overbrace{K(S_T)}^{\text{final value of water}}\Big]$$

Multistage stochastic optimization problem

- Let be given a (cost) function
$$j : \mathcal{H}_T \to]-\infty, +\infty]$$
bounded below, and measurable with respect to the field $\mathcal{H}_T$
- We consider the multistage stochastic optimization problem
$$\min_{(\mathbf{U}_0,\ldots,\mathbf{U}_{T-1}) \in \mathbb{L}^0_{\mathcal{A}}(\Omega, \prod_{s=0}^{T-1} \mathcal{U}_s)} \mathbb{E}\big[j(\mathbf{W}_0, \mathbf{U}_0, \mathbf{W}_1, \ldots, \mathbf{U}_{T-1}, \mathbf{W}_T)\big]$$
- that is, with criterion $J : \mathbb{L}^0(\Omega, \mathcal{A}, \prod_{s=0}^{T-1} \mathcal{U}_s) \to]-\infty, +\infty]$ given by
$$J(\mathbf{U}_0, \ldots, \mathbf{U}_{T-1}) = \mathbb{E}\big[j(\mathbf{W}_0, \mathbf{U}_0, \mathbf{W}_1, \ldots, \mathbf{U}_{T-1}, \mathbf{W}_T)\big]$$

Complexity of upper and lower bounds

- Upper bound: open loop solution
- Lower bound: anticipative

Outline of the presentation

Bellman operators

For $t = 0, \ldots, T$,

- we define $\mathbb{L}^\infty(\mathcal{H}_t, \mathfrak{H}_t)$, the space of bounded measurable real-valued functions $\mathcal{H}_t \to \mathbb{R}$
- We suppose that there exists a regular conditional distribution

$$\mathbb{P}_{\mathbf{W}_{t+1}}^{\mathbf{W}_{[0:t]}}(w_{[0:t]}, dw_{t+1}) = \mathbb{P}_{\mathbf{W}_{t+1}}^{\mathbf{W}_{[0:t]}}([h_t], dw_{t+1})$$

  of the random variable $\mathbf{W}_{t+1}$ knowing the random process $\mathbf{W}_{[0:t]}$
- we define the Bellman operators

$$\mathcal{B}_{t+1} : \mathbb{L}^\infty(\mathcal{H}_{t+1}, \mathfrak{H}_{t+1}) \to \mathbb{L}^\infty(\mathcal{H}_t, \mathfrak{H}_t) \text{ by}$$

$$(\mathcal{B}_{t+1}\varphi)(h_t) = \inf_{u_t \in \mathcal{U}_t} \int_{\mathcal{W}_{t+1}} \varphi\big((h_t, u_t, w_{t+1})\big) \mathbb{P}_{\mathbf{W}_{t+1}}^{\mathbf{W}_{[0:t]}}([h_t], dw_{t+1})$$

$$\forall \varphi \in \mathbb{L}^\infty(\mathcal{H}_{t+1}, \mathfrak{H}_{t+1}) , \ \forall h_t \in \mathcal{H}_t$$

Value functions and Bellman equation

- We define inductively value functions, or Bellman functions,

$$V_t : \mathcal{H}_t \to \mathbb{R} \;, \quad t = 0, \ldots, T$$

  by

$$V_T = j \;, \quad V_t = \mathcal{B}_{t+1} V_{t+1} \;, \quad t = 0, \ldots, T-1$$

- that is, solution of the Bellman equation

$$V_t(h_t) = \inf_{u_t \in \mathcal{U}_t} \int_{\mathcal{W}_{t+1}} V_{t+1}(h_t, u_t, w_{t+1}) \mathbb{P}_{\mathbf{W}_{t+1}}^{\mathbf{W}_{[0:t]}}([h_t], dw_{t+1})$$

Measurable selection

We suppose that, for all $t = 0, \ldots, T$,
there exists a measurable selection

$$\gamma_t : (\mathcal{H}_t, \mathfrak{H}_t) \to (\mathcal{U}_t, \mathfrak{U}_t)$$

such that

$$\gamma_t(h_t) \in \operatorname*{arg\,min}_{u_t \in \mathcal{U}_t} \int_{\mathcal{W}_{t+1}} V_{t+1}(h_t, u_t, w_{t+1}) \mathbb{P}_{\mathbf{W}_{t+1}}^{\mathbf{W}_{[0:t]}}([h_t], dw_{t+1})$$

$\forall h_t \in \mathcal{H}_t$

Theoretical resolution

Proposition

A solution to the multistage stochastic optimization problem

$$\min_{(\mathbf{U}_0,\ldots,\mathbf{U}_{T-1})\in\mathbb{L}^0_{\mathcal{A}}(\Omega,\prod_{s=0}^{T-1}\mathcal{U}_s)} \mathbb{E}\big[j(\mathbf{W}_0,\mathbf{U}_0,\mathbf{W}_1,\ldots,\mathbf{U}_{T-1},\mathbf{W}_T)\big]$$

is the sequence $\mathbf{U}_0^*,\ldots,\mathbf{U}_{T-1}^*$ of random variables defined inductively by

$$\mathbf{U}_t^* = \gamma_t \circ \mathbf{H}_t^* \ , \quad t = 0,\ldots,T-1$$

$$\textit{where} \quad \mathbf{H}_0^* = \mathbf{W}_0 \ , \ \ \mathbf{H}_{t+1}^* = (\mathbf{H}_t^*,\mathbf{U}_t^*,\mathbf{W}_{t+1}) \ , \quad t = 0,\ldots,T-1$$

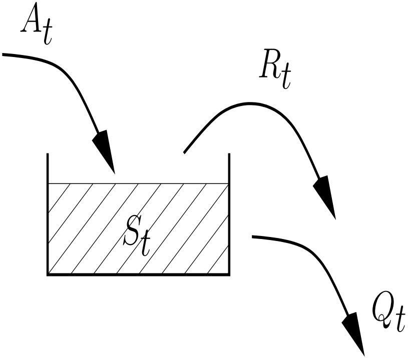
and the minimum is

$$\mathbb{E}\big[V_0(\mathbf{W}_0)\big] = \min_{(\mathbf{U}_0,\ldots,\mathbf{U}_{T-1})\in\mathbb{L}^0_{\mathcal{A}}(\Omega,\prod_{s=0}^{T-1}\mathcal{U}_s)} \mathbb{E}\big[j(\mathbf{W}_0,\mathbf{U}_0,\mathbf{W}_1,\ldots,\mathbf{U}_{T-1},\mathbf{W}_T)\big]$$

Extension with constraints (functions with values in $\mathbb{R}\cup\{+\infty\}$)

Outline of the presentation

Optimal single dam management



A single dam nonlinear dynamical model in decision-hazard

We can model the dynamics of the water volume in a dam by

$$\underbrace{S_{t+1}}_{\text{future volume}} = \min\{S^\sharp, \underbrace{S_t}_{\text{volume}} - \underbrace{Q_t}_{\text{turbined}} + \underbrace{A_{t+1}}_{\text{inflow volume}}\}$$

- S_t volume (stock) of water at the beginning of period $[t, t+1[$
- A_{t+1} inflow water volume (rain, etc.) during $[t, t+1[$
- decision-hazard:
  A_{t+1} is not available at the beginning of period $[t, t+1[$
- Q_t turbined outflow volume during $[t, t+1[$
  - decided at the beginning of period $[t, t+1[$
  - supposed to depend on S_t but not on A_{t+1}
  - chosen such that $0 \leq Q_t \leq S_t$

The traditional economic problem is maximizing the expected payoff

- Suppose that
  - a probability $\mathbb{P}$ is given on the set $\mathbb{R}^T$ of water inflows scenarios $(A_0, \dots, A_{T-1})$
  - turbined water Q_t is sold at price p_t, related to the price at which energy can be sold at time t
  - at the horizon, the final volume S_T has a value $K(S_T)$, the "final value of water"
- The traditional economic problem is to maximize the intertemporal payoff (without discounting if the horizon is short)

$$\max \mathbb{E}\Big[\sum_{t=0}^{T-1} \overbrace{p_t Q_t}^{\text{turbined water payoff}} + \overbrace{K(S_T)}^{\text{final volume utility}}\Big]$$

State reduction and dynamics

For $t = 0, \ldots, T$, suppose that there exists

- state space $\mathcal{X}_t$, a measurable set equipped with σ-field $\mathfrak{X}_t$
- reduction mappings

$$\theta_t : \mathcal{H}_t \to \mathcal{X}_t$$

- dynamics

$$f_t : \mathcal{X}_t \times \mathcal{U}_t \times \mathcal{W}_{t+1} \to \mathcal{X}_{t+1}$$

such that

$$\theta_{t+1}(h_t, u_t, w_{t+1}) = f_t\big(\theta_t(h_t), u_t, w_{t+1}\big) \ , \quad t = 0, \ldots, T-1$$

Cost only depends on final state

Suppose that there exists

$$\tilde{j} : \mathcal{X}_T \to]-\infty, +\infty]$$

such that the cost function $j : \mathcal{H}_T \to]-\infty, +\infty]$ can be factored as

$$j = \tilde{j} \circ \theta_T$$

Markovian assumption

- Let $\Delta(\mathcal{W}_t)$ denote the set of probabilities on $(\mathcal{W}_t, \mathcal{W}_t)$, for $t = 0, \ldots, T$
- Suppose that, for all $t = 0, \ldots, T$, there exists

$$\mu_t : \mathcal{X}_t \times \prod_{s=0}^{t} \mathcal{W}_s \to \Delta(\mathcal{W}_{t+1})$$

such that

$$\mathbb{P}^{\mathbf{W}_{[0:t]}}_{\mathbf{W}_{t+1}}\big([h_t], dw_{t+1}\big) = \mu_t\big(\theta_t(h_t), dw_{t+1}\big)$$

Bellman equation

- We define inductively

$$\tilde{V}_T(x_T) = \tilde{j}(x_T) \ , \ \ \forall x_T \in \mathcal{X}_T$$

$$\tilde{V}_t(x_t) = \inf_{u_t \in \mathcal{U}_t} \int_{\mathcal{W}_{t+1}} \tilde{V}_{t+1}\big(f_t(x_t, u_t, w_{t+1})\big)\mu_t(x_t, dw_{t+1})$$

$$\forall x_t \in \mathcal{X}_t \ , \ \ t = 0, \ldots, T-1$$

- We suppose that there exists a measurable selection

$$\tilde{\gamma}_t^* : (\mathcal{X}_t, \mathfrak{X}_t) \to (\mathcal{U}_t, \mathfrak{U}_t) \ , \ \ t = 0, \ldots, T-1$$

such that

$$\tilde{\gamma}_t^*(x_t) \in \arg\min_{u_t \in \mathcal{U}_t} \int_{\mathcal{W}_{t+1}} \tilde{V}_{t+1}\big(f_t(x_t, u_t, w_{t+1})\big)\mu_t(x_t, dw_{t+1})$$

$\forall x_t \in \mathcal{X}_t$

Proposition

A solution to the multistage stochastic optimization problem

$$\min_{\mathbf{U}_0,\ldots,\mathbf{U}_{T-1}} \mathbb{E}\big[j(\mathbf{W}_0,\mathbf{U}_0,\mathbf{W}_1,\ldots,\mathbf{U}_{T-1},\mathbf{W}_T)\big]$$

$$\sigma(\mathbf{U}_0) \subset \sigma(\mathbf{W}_0),\ldots,\sigma(\mathbf{U}_{T-1}) \subset \sigma(\mathbf{W}_0,\ldots,\mathbf{W}_{T-1})$$

is the sequence $\mathbf{U}_0^*,\ldots,\mathbf{U}_{T-1}^*$ *of random variables defined inductively by*

$$\mathbf{U}_t^* = \tilde{\gamma}_t^*(\mathbf{X}_t^*) \ , \quad t=0,\ldots,T-1$$

$$\textit{where} \quad \mathbf{X}_0^* = \mathbf{W}_0 \ , \quad \mathbf{X}_{t+1}^* = f_t(\mathbf{X}_t^*,\mathbf{U}_t^*,\mathbf{W}_{t+1}) \ , \quad t=0,\ldots,T-1$$

and the minimum is

$$\mathbb{E}\big[\tilde{V}_0(\mathbf{X}_0^*)\big] = \min_{(\mathbf{U}_0,\ldots,\mathbf{U}_{T-1})\in\mathbb{L}^0_{\mathcal{A}}(\Omega,\prod_{s=0}^{T-1}\mathcal{U}_s)} \mathbb{E}\big[j(\mathbf{W}_0,\mathbf{U}_0,\mathbf{W}_1,\ldots,\mathbf{U}_{T-1},\mathbf{W}_T)\big]$$

Extension

$$\mathbf{X}_0 = \mathbf{W}_0 \ , \ \ \mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) \ , \ \ t = 0, \dots, T-1$$

Constraints of the form

$$(\mathbf{X}_t, \mathbf{U}_t) \in \mathcal{C}_t \subset \mathcal{X}_t \times \mathcal{U}_t \ , \ \ \mathbb{P} - \mathrm{a.s} \ , \ \ t = 0, \dots, T-1$$

and

$$\mathbf{X}_T \in \mathcal{C}_T \subset \mathcal{X}_T \ , \ \ \mathbb{P} - \mathrm{a.s}$$

Outline of the presentation

Stochatic optimal control problem formulation

$$\min_{\mathbf{U}_0,\ldots,\mathbf{U}_{T-1}} \mathbb{E}\Big[\sum_{t=0}^{T-1} L_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) + K(\mathbf{X}_T)\Big]$$

$$\sigma(\mathbf{U}_0) \subset \sigma(\mathbf{X}_0), \ldots, \sigma(\mathbf{U}_{T-1}) \subset \sigma(\mathbf{X}_0,\mathbf{W}_1,\ldots,\mathbf{W}_{T-1})$$

$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) \ , \quad t = 0,\ldots,T-1$$

$$\mathbf{U}_t \in \mathcal{B}_t(\mathbf{X}_t) \ , \quad t = 0,\ldots,T-1$$

Basic data

Let $\mathcal{U}$, $\mathcal{W}$, $\mathcal{X}$ be measurable sets, equipped with σ-fields $\mathfrak{U}$, $\mathfrak{W}$, $\mathfrak{X}$ and, for $t = 0, \ldots, T-1$,

- dynamics mapping
$$f_t : \mathcal{X} \times \mathcal{U} \times \mathcal{W} \to \mathcal{X}$$
- instantaneous costs functions
$$L_t : \mathcal{X} \times \mathcal{U} \times \mathcal{W} \to \mathbb{R}$$
- final cost function
$$K : \mathcal{X} \to \mathbb{R}$$
- constraints set-valued mapping
$$\mathcal{B}_t : \mathcal{X} \rightrightarrows \mathcal{U}$$

Bellman equation

- We consider a stochastic process $(\mathbf{W}_1, \ldots, \mathbf{W}_T)$, with values in $\mathcal{W}$
- We define inductively the Bellman functions

$$V_T(x) = K(x) \ , \ \ \forall x \in \mathcal{X}$$

$$V_t(x) = \inf_{u \in \mathcal{B}_t(x)} \mathbb{E}_{\mathbf{W}_{t+1}} \Big[L_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathbf{W}_{t+1})\big) \Big]$$

$$\forall x \in \mathcal{X} \ , \ \ t = 0, \ldots, T-1$$

- We suppose that there exists a measurable selection

$$\gamma_t^* : (\mathcal{X}, \mathfrak{X}) \to (\mathcal{U}, \mathfrak{U}) \ , \ \ t = 0, \ldots, T-1 \ \text{such that}$$

$$\gamma_t^*(x) \in \arg\min_{u \in \mathcal{B}_t(x)} \mathbb{E}_{\mathbf{W}_{t+1}} \Big[L_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathbf{W}_{t+1})\big) \Big]$$

$\forall x \in \mathcal{X}$

White noise assumption

- We suppose that the stochastic process $(\mathbf{W}_1, \ldots, \mathbf{W}_T)$ is a white noise, that is, $\mathbf{W}_1, \ldots, \mathbf{W}_T$ are independent random variables
- We consider a random variable $\mathbf{X}_0$, with values in $\mathcal{X}$, independent of the stochastic process $(\mathbf{W}_1, \ldots, \mathbf{W}_T)$

Bellman result

Proposition

A solution to the multistage stochastic optimization problem

$$
\begin{aligned}
\min_{\mathbf{U}_0,\ldots,\mathbf{U}_{T-1}} \; & \mathbb{E}\Big[\sum_{t=0}^{T-1} L_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) + K(\mathbf{X}_T)\Big] \\
& \sigma(\mathbf{U}_0) \subset \sigma(\mathbf{X}_0), \ldots, \sigma(\mathbf{U}_{T-1}) \subset \sigma(\mathbf{X}_0,\mathbf{W}_1,\ldots,\mathbf{W}_{T-1}) \\
& \mathbf{X}_{t+1} = f_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) \;, \quad t = 0,\ldots,T-1 \\
& \mathbf{U}_t \in \mathcal{B}_t(\mathbf{X}_t) \;, \quad t = 0,\ldots,T-1
\end{aligned}
$$

is the sequence $\mathbf{U}_0^*,\ldots,\mathbf{U}_{T-1}^*$ of random variables defined inductively by

$$
\mathbf{U}_t^* = \gamma_t^*(\mathbf{X}_t^*) \;, \quad t = 0,\ldots,T-1
$$

$$
\textit{where} \quad \mathbf{X}_0^* = \mathbf{X}_0 \;, \quad \mathbf{X}_{t+1}^* = f_t(\mathbf{X}_t^*,\mathbf{U}_t^*,\mathbf{W}_{t+1}) \;, \quad t = 0,\ldots,T-1
$$

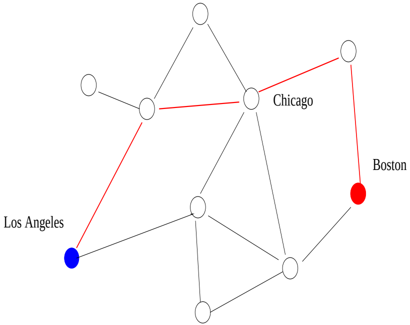
and the minimum is

$$
\mathbb{E}\big[V_0(\mathbf{X}_0)\big] = \min_{\mathbf{U}_0,\ldots,\mathbf{U}_{T-1}} \mathbb{E}\Big[\sum_{t=0}^{T-1} L_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) + K(\mathbf{X}_T)\Big]
$$

Outline of the presentation

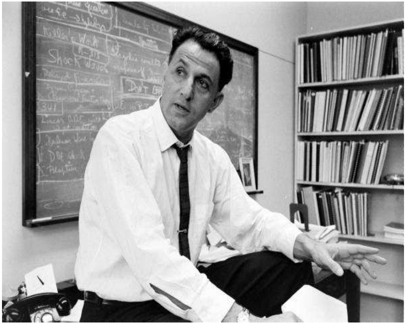
Outline of the presentation

The shortest path on a graph illustrates Bellman's Principle of Optimality



For an auto travel analogy, suppose that the fastest route from Los Angeles to to Boston passes through **Chicago**. *The principle of optimality translates to obvious fact that the Chicago to Boston portion of the route is also the fastest route for a trip that starts from Chicago and ends in Boston. (Dimitri P. Bertsekas)*

Bellman's Principle of Optimality



Richard Ernest Bellman (August 26, 1920 – March 19, 1984)

An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision (Richard Bellman)

What is state and what is noise?

Delineating what is state and what is noise is a modelling issue

When the uncertainties are not independent,
a solution is to enlarge the state

- If the water inflows follow an auto-regressive model, we have

$$\begin{aligned}
\overbrace{\mathbf{S}_{t+1}}^{\text{future stock}} &= \min\{S^\sharp, \overbrace{\mathbf{S}_t}^{\text{stock}} - \overbrace{\mathbf{Q}_t}^{\text{water release}} + \overbrace{\mathbf{A}_{t+1}}^{\text{water inflows}}\} \\
\underbrace{\mathbf{A}_{t+1}}_{\text{future water inflows}} &= \alpha \underbrace{\mathbf{A}_t}_{\text{water inflows}} + \underbrace{\mathbf{W}_{t+1}}_{\text{noise}}
\end{aligned}$$

  where we suppose that $\mathbf{W}_1, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$
  form a sequence of independent random variables

- The couple $x_t = (S_t, A_t)$ is a sufficient summary of past controls
  and uncertainties to do forecasting:
  knowing the state $x_t = (S_t, A_t)$ at time t is sufficient
  to forecast x_{t+1}, given the control Q_t and the uncertainty $\mathbf{W}_{t+1}$

What is a state?

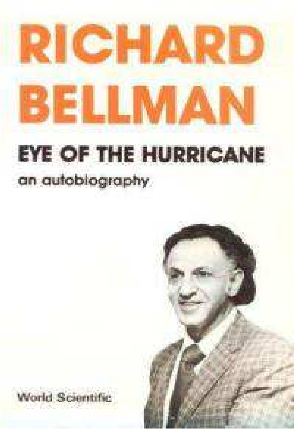
Bellman autobiography, Eye of the Hurricane

> *Conversely, once it was realized that the concept of policy was fundamental in control theory, the mathematicization of the basic engineering concept of 'feedback control,' then the emphasis upon a state variable formulation became natural*

- A state in optimal stochastic control problems is a sufficient statistics for the uncertainties and past controls
(P. Whittle, *Optimization over Time: Dynamic Programming and Stochastic Control*)
- Quoting Whittle, suppose there is a variable x_t
which summarizes past history in that, given t and the value of x_t, one can calculate the optimal u_t and also x_{t+1} without knowledge of the history $(\omega, u_0, \ldots, u_{t-1})$, for all t, where ω represents all uncertainties. Such a variable is termed *sufficient*
- While history takes value in an increasing space as t increases,
a sufficient variable taking values in a space independent of t
is called a state variable

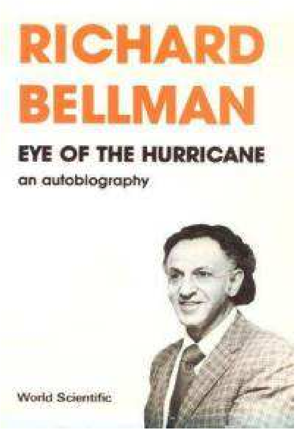
A bit of history (and fun)

“Where did the name, dynamic programming, come from?”



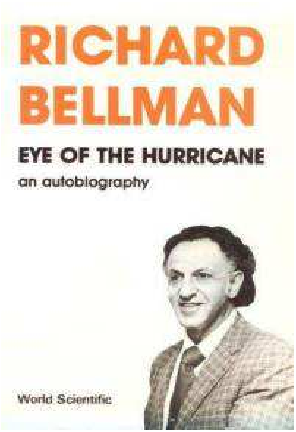
The 1950s were not good years for mathematical research. We had a very interesting gentleman in Washington named Wilson. He was Secretary of Defense, and he actually had a pathological fear and hatred of the word, research. I'm not using the term lightly; I'm using it precisely. His face would suffuse, he would turn red, and he would get violent if people used the term, research, in his presence. You can imagine how he felt, then, about the term, mathematical.

“Where did the name, dynamic programming, come from?”



What title, what name, could I choose? In the first place I was interested in planning, in decision making, in thinking. But planning, is not a good word for various reasons. I decided therefore to use the word, programming.

“Where did the name, dynamic programming, come from?”



I wanted to get across the idea that this was dynamic, this was multistage, this was time-varying. I thought, let’s kill two birds with one stone. Let’s take a word that has an absolutely precise meaning, namely dynamic, in the classical physical sense. It also has a very interesting property as an adjective, and that is it’s impossible to use the word, dynamic, in a pejorative sense. Try thinking of some combination that will possibly give it a pejorative meaning. It’s impossible. Thus, I thought dynamic programming was a good name.

Outline of the presentation

The cost-to-go / value function / Bellman function

Assume that the primitive random variables $\mathbf{W}_1, \dots, \mathbf{W}_{T-1}, \mathbf{W}_T$ are independent under the probability $\mathbb{P}$

Cost-to-go / value function / Bellman function

The cost-to-go from state x at stage t is

$$V_t(x) = \min_{\gamma_t, \dots, \gamma_{T-1}} \mathbb{E}\left[\sum_{s=t}^{T-1} L_s(\mathbf{X}_s, \mathbf{U}_s, \mathbf{W}_{s+1}) + K(\mathbf{X}_T)\right]$$

where $\mathbf{X}_t = x$ and, for $s = t, \dots, T-1$,
$\mathbf{X}_{s+1} = f_s\big(\mathbf{X}_s, \mathbf{U}_s, \mathbf{W}_{s+1}\big)$ and $\mathbf{U}_s = \gamma_s(\mathbf{X}_s)$

- The function $V_t : \mathcal{X} \to \overline{\mathbb{R}}$ is called the value function, or the Bellman function
- The original minimization problem is $V_0(x_0)$

The stochastic dynamic programming equation, or Bellman equation, is a backward equation satisfied by the value function

Stochastic dynamic programming equation

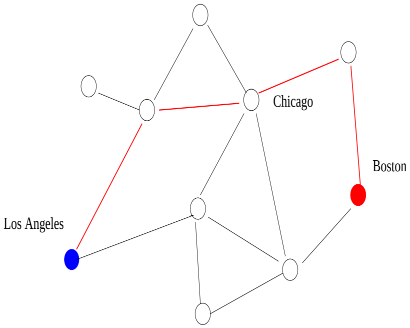
If the primitive random variables $\mathbf{W}_1, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$
are independent under the probability $\mathbb{P}$,
the value function $V_t : \mathcal{X} \to \overline{\mathbb{R}}$ satisfies the following backward induction,
where t runs from $T-1$ down to 0

$$\begin{aligned} V_T(x) &= K(x) \\ V_t(x) &= \min_{u \in \mathcal{B}_t(x)} \mathbb{E}_{\mathbf{W}_{t+1}} \Big[L_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathbf{W}_{t+1})\big) \Big] \end{aligned}$$

$\forall x \in \mathcal{X}$

Sketch of the proof in the deterministic case

$$V_t(x) = \min_{u \in \mathcal{B}_t(x)} \Big(\underbrace{L_t(x,u)}_{\text{instantaneous cost}} + \overbrace{V_{t+1}\big(\underbrace{f_t(x,u)}_{\text{future state}} \big)}^{\text{optimal cost}} \Big)$$



A decision u at time t in state x provides

- an instantaneous cost $L_t(x,u)$
- and a future cost for attaining the new state $f_t(x,u)$

Outline of the presentation

Navigating between “backward offline” and “forward online”

Optimal trajectories are calculated forward online

1. Initial state $x_0^* = x_0$
2. Plug the state x_0^* into the feedback γ_0
   $\rightarrow$ initial decision $u_0^* = \gamma_0^*(x_0^*)$
   or compute the optimal decision u_t^* "on the fly" by

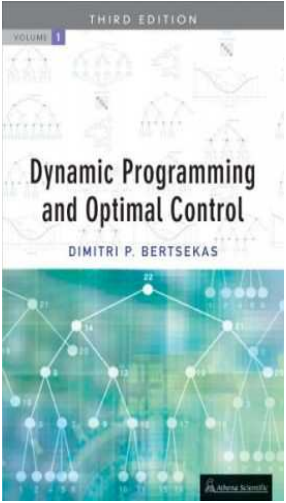
$$u_0^* \in \underset{u \in \mathcal{B}_0(x_0^*)}{\arg\min}\, \mathbb{E}_{\mathbf{W}_1}\Big[L_0(x_0^*, u, \mathbf{W}_1) + V_1\big(f_0(x_0^*, u, \mathbf{W}_1)\big)\Big]$$

3. Run the dynamics $\rightarrow$ second state $x_1^* = f_0(x_0^*, u_0^*, w_1)$
4. Second decision

$$u_1^* \in \underset{u \in \mathcal{B}_1(x_1^*)}{\arg\min}\, \mathbb{E}_{\mathbf{W}_2}\Big[L_1(x_1^*, u, \mathbf{W}_2) + V_2\big(f_1(x_1^*, u, \mathbf{W}_2)\big)\Big]$$

5. And so on $x_2^* = f_1(x_1^*, u_1^*, w_2)$
6. ...

“Life is lived forward but understood backward” (Søren Kierkegaard)



D. P. Bertsekas introduces his book
Dynamic Programming and Optimal Control
with a citation by Søren Kierkegaard

> *"Livet skal forstås baglaens, men leves forlaens"*
>
> *Life is to be understood backwards, but it is lived forwards*

- The value function and the optimal policies are computed backward and offline by means of the Bellman equation
- whereas the optimal trajectories are computed forward and online

How optimal decisions can be computed online

Greedy one-step lookahead algorithm

- If we are able to store the value functions $x \mapsto V_t(x)$
- we do not need to compute and store
  the optimal policy γ_t^* in advance
- Indeed, when we are at state x at time t in real time,
  we can just compute the optimal decision u_t^* "on the fly" by

$$u_t^* \in \arg\min_{u \in \mathcal{B}_t(x)} \mathbb{E}_{\mathbf{W}_{t+1}} \Big[L_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathbf{W}_{t+1})\big) \Big]$$

- In addition to sparing storage, this method makes it possible
  to incorporate in the above program any new information available
  at time t (on the distribution of the noise $\mathbf{W}_{t+1}$, for instance)

So, the question is:
how can we store the value functions?

The effort can be concentrated on computing the value functions

- on a grid, by discretizing the Bellman equation
  (but curse of dimensionality)
- by estimating basis coefficients,
  when it is known that the value functions are quadratic
  (the linear-quadratic case)
- by estimating lower affine approximations of the value functions,
  when it is known that the value function is convex
  (the linear-convex case and SDDP)

Outline of the presentation

The curse of dimensionality :-(

Algorithm for the Bellman functions

initialization $V_T = K$;
for $t = T, T-1, \dots, 0$ **do**
  forall $x \in \mathcal{X}$ **do**
    forall $u \in \mathcal{B}_t(x)$ **do**
      forall $w \in \mathcal{W}_{t+1}$ **do**
        $l_t(x,u,w) = L_t(x,u,w) + V_{t+1}\big(f_t(x,u,w)\big)$
      $$\sum_{w \in \mathcal{W}_{t+1}} \mathbb{P}\{w\} l_t(x,u,w)$$
  $$V_t(x) = \min_{u \in \mathcal{B}_t(x)} \sum_{w \in \mathcal{W}_{t+1}} \mathbb{P}\{w\} l_t(x,u,w) \; ;$$
  $$\mathcal{B}_t^*(x) = \arg\min_{u \in \mathcal{B}_t(x)} \sum_{w \in \mathcal{W}_{t+1}} \mathbb{P}\{w\} l_t(x,u,w)$$

Complexity of the dynamic programming algorithm

Assuming that

- the state x_t can take N_x values
- the control u_t can take N_u values
- the uncertainty w_t can take N_w values

the complexity (number of operations) of the Bellman algorithm is in

$$O(T \times N_x \times N_u \times N_w)$$

which is linear in the number of stages :-)
but exponential in the dimension of the state
(and also control and uncertainty)

The curse of dimensionality is illustrated by the random access memory capacity on a computer: one, two, three, infinity (Gamov)

- On a computer
  - RAM: 8 GBytes $= 8(1\,024)^3 = 2^{33}$ bytes
  - a double-precision real: 8 bytes $= 2^3$ bytes
  - $\Longrightarrow 2^{30} \approx 10^9$ double-precision reals can be handled in RAM
- If a state of dimension 4 is approximated by a grid with 100 levels by components, we need to manipulate $100^4 = 10^8$ reals and
  - do a time loop
  - do a control loop (after discretization)
  - compute an expectation

The wall of dimension can be pushed beyond
if additional properties are exploited (linearity, convexity)

In the linear-quadratic case, value functions are quadratic and optimal policies are linear

- When cost functions are quadratic (convex)

$$
\begin{aligned}
K(x) &= x' S_T x \,(+\text{affine}) \\
L_t(x,u,w) &= x' S_t x + w' R_t w + u' Q_t u \,(+\text{affine})
\end{aligned}
$$

- and the dynamic is affine

$$
f_t(x,u,w) = F_t x + G_t u + H_t w \,(+\text{constant})
$$

- and primitive random variables $\mathbf{W}_1, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$ are square integrable and independent under the probability $\mathbb{P}$
- then, the value functions $x \mapsto V_t(x)$ are quadratic (convex), and optimal policies are affine in the state

$$
u_t = M_t x_t \,(+\text{constant})
$$

Outline of the presentation

Multistage stochastic optimization

Dynamic programming without state

Dynamic programming with state

Dynamic programming with state and white noise

Dynamic Programming With State and White Noise (Complements)

- Bellman's Principle of Optimality
- Cost-to-go and Bellman equation
- Backward offline / forward online
- The curse of dimensionality
- Hazard-decision, linear-convex, SDDP

Stochatic optimal control problem formulation (hazard-decision)

$$
\min_{\mathbf{U}_0,\dots,\mathbf{U}_{T-1}} \mathbb{E}\Big[\sum_{t=0}^{T-1} L_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) + K(\mathbf{X}_T)\Big]
$$

$$
\sigma(\mathbf{U}_0) \subset \sigma(\mathbf{X}_0,\mathbf{W}_1),\dots,\sigma(\mathbf{U}_{T-1}) \subset \sigma(\mathbf{X}_0,\mathbf{W}_1,\dots,\mathbf{W}_T)
$$

$$
\mathbf{X}_{t+1} = f_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1})\ , \quad t = 0,\dots,T-1
$$

$$
\mathbf{U}_t \in \mathcal{B}_t(\mathbf{X}_t,\mathbf{W}_{t+1})\ , \quad t = 0,\dots,T-1
$$

Bellman equation and optimal policies in the hazard-decision information pattern

The uncertainty is observed before making the decision

initialization $V_T = K$;
for $t = T, T-1, \dots, 0$ **do**
  forall $x \in \mathcal{X}$ **do**
    forall $w \in \mathcal{W}_{t+1}$ **do**
      forall $u \in \mathcal{B}_t(x, w)$ **do**
        $l_t(x, u, w) = L_t(x, u, w) + V_{t+1}\big(f_t(x, u, w)\big)$
      $\min_{u \in \mathcal{B}_t(x)} l_t(x, u, w)$;
      $\mathcal{B}_t^*(x, w) = \arg\min_{u \in \mathcal{B}_t(x,w)} l_t(x, u, w)$
    $V_t(x) = \sum_{w \in \mathcal{W}_{t+1}} \mathbb{P}\{w\} \min_{u \in \mathcal{B}_t(x,w)} l_t(x, u, w)$

When spilling decisions are made after knowing the water inflows, we obtain a linear dynamical model

$$\underbrace{S_{t+1}}_{\text{future volume}} = \underbrace{S_t}_{\text{volume}} - \underbrace{Q_t}_{\text{turbined}} + \underbrace{A_{t+1}}_{\text{inflow volume}} - \underbrace{R_{t+1}}_{\text{spilled}}$$

- S_t volume (stock) of water at the beginning of period $[t, t+1[$
- A_{t+1}, inflow water volume (rain, etc.) during $[t, t+1[$;
- Q_t turbined outflow volume
  - decided at the beginning of period $[t, t+1[$ (hazard follows decision)
  - supposed to depend on the stock S_t
- R_{t+1} spilled volume
  - decided at the end of period $[t, t+1[$ (hazard precedes decision)
  - supposed to depend on the stock S_t and on the inflow water A_t

$$0 \leq Q_t \leq S_t \text{ and } 0 \leq S_t - Q_t + A_{t+1} - R_{t+1} \leq S^\sharp$$

In the linear-convex case, value functions are convex

Here, we aim at minimizing expected cumulated costs

$$\mathbb{E}\Big[\sum_{t=0}^{T-1} \overbrace{L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1})}^{\text{instantaneous cost}} + \underbrace{K(\mathbf{X}_T)}_{\text{final cost}}\Big]$$

The value functions $x \mapsto V_t(x)$ are convex whenever

- the instantaneous cost functions $(x, u) \mapsto L_t(x, u, w)$ is jointly convex in state and control $(\forall w)$
- the final cost function $x \mapsto K(x)$ is convex $(\forall w)$
- the dynamics mappings are affine in state and control $(\forall w)$

$$f_t(x, u, w) = F_t(w)x + G_t(w)u + H_t(w)$$

- The constraint sets $\big\{(x, u) \,\big|\, u \in \mathcal{B}_t(x)\big\}$ are convex

The minimum over one variable of a jointly convex function is convex in the other variable

A lemma in convex analysis

Let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be convex, and let $C \subset \mathcal{X} \times \mathcal{Y}$ be a convex set
Then, the so-called marginal function $g : \mathcal{X} \to \mathbb{R}$ defined by

$$g(x) = \min_{y \in \mathcal{Y}, (x,y) \in C} f(x,y) \;, \;\; \forall x \in \mathcal{X}$$

is a convex function

Stochastic Dual Dynamic Programming (SDDP)

- The dynamic programming equation associated with the problem of minimizing the expected costs is

$$
\begin{aligned}
V_T(x) &= \overbrace{K(x)}^{\text{final cost}} \\
V_t(x) &= \min_{u \in \mathcal{B}_t(x)} \mathbb{E}_{\mathbf{W}_{t+1}} \Big[\overbrace{L_t(x, u, \mathbf{W}_{t+1})}^{\text{instantaneous cost}} \\
&\quad + V_{t+1}\big(\underbrace{F_t(\mathbf{W}_{t+1})x + G_t(\mathbf{W}_{t+1})u + H_t(\mathbf{W}_{t+1})}_{\text{future state}} \big) \Big]
\end{aligned}
$$

- It can be shown by induction that $x \mapsto V_t(x)$ is convex
- A subgradient at x^*_{t+1} defines a hyperplane, hence a lower affine approximation of the value function, calculated by duality

SDDP and autoregressive noise models

The property that value functions are convex extends to the following cases

- Multiple stocks interconnected by linear dynamics

$$\mathbf{S}^i_{t+1} = \mathbf{S}^i_t + \mathbf{A}^i_t + \mathbf{Q}^{i-1}_t - \mathbf{Q}^i_t - \mathbf{R}^i_{t+1}$$

- Water inflows following an autoregressive model

$$\mathbf{A}^i_{t+1} = \sum_{k=0,\ldots,K^i} \alpha_k \mathbf{A}^i_{t-k} + \mathbf{W}_{t+1}$$

  where the random variables $\mathbf{W}_1, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$ are independent

Summary

- Bellman's Principle of Optimality breaks
  an intertemporal optimization problem
  into a sequence of interconnected static optimization problems
- The cost-to-go / value function / Bellman function
  is solution of a backward dynamic programming equation,
  or Bellman equation
- The Bellman equation provides an optimal policy,
  a concept of solution adapted to uncertain case
- In numerical practice, the curse of dimensionality
  forbids to use dynamic programming
  for a state with dimension more than four or five
- However, special cases like the linear-quadratic or the linear-convex ones, do not (totally) suffer from the curse of dimensionality