Dynamic Programming

Michel De Lara
Cermics, École des Ponts ParisTech
France

École des Ponts ParisTech

January 2, 2020

Outline of the presentation

Outline of the presentation

Basic data

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

History space

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

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

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

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

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

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

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

Noise process and filtration

- For $t = 0, \ldots, T$, let $\mathbf{W}_t$ be a random variable taking values in $\mathbb{W}_t$

$$\mathbf{W}_t : \Omega \to \mathbb{W}_t$$

- We put

$$\mathbf{W}_{[0:t]} = (\mathbf{W}_0, \ldots, \mathbf{W}_t)$$

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

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

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

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

Multistage stochastic optimization problem

- Let

$$\phi : \mathbb{H}_T \to \mathbb{R}$$

be a bounded function, measurable with respect to the field $\mathcal{H}_T$

- We consider the multistage stochastic optimization problem

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

Bellman operators

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

- we define $\mathbb{L}^\infty(\mathbb{H}_t, \mathcal{H}_t)$, the space of bounded measurable real-valued functions over $\mathbb{H}_t$
- We suppose that there exists a regular conditional distribution of the random variable $\mathbf{W}_{t+1}$ knowing the random process $\mathbf{W}_{[0:t]}$, denoted by

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

- we define the Bellman operators

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

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

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

Value functions and Bellman equation

- We define inductively value functions, or Bellman functions,

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

by

$$V_T = \phi \ , \quad V_t = \mathcal{B}_{t+1} V_{t+1} \ , \quad t = 0, \dots, T-1 \ .$$

- that is, solution of the Bellman equation

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

Measurable selection

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

$$\psi_t : (\mathbb{H}_t, \mathcal{H}_t) \to (\mathbb{U}_t, \mathcal{U}_t)$$

such that

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

Proposition

A solution to the multistage stochastic optimization problem

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

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

$$\mathbf{U}_t^\star = \psi_t \circ \mathbf{H}_t^\star \ , \quad t=0,\ldots,T$$

where

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

The minimum is

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

Extension

Constraints of the form

$$\mathbf{H}_0 = \mathbf{W}_0 \ , \ \ \mathbf{H}_{t+1} = (\mathbf{H}_t, \mathbf{U}_t, \mathbf{W}_{t+1})$$

and

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

Outline of the presentation

State reduction and dynamics

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

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

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

- dynamics $f_{:t}$

$$f_{:t} : \mathbb{X}_t \times \mathbb{U}_t \times \mathbb{W}_{t+1} \to \mathbb{X}_{t+1}$$

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

$$\widetilde{\phi} : \mathbb{X}_T \to \mathbb{R}$$

such that the cost ϕ can be factored as

$$\phi = \widetilde{\phi} \circ \theta_T$$

Markovian assumption

- Let $\Delta(\mathbb{W}_{t+1})$ denote the set of probabilities on $(\mathbb{W}_{t+1}, \mathcal{W}_t)$
- Suppose that, for all $t = 0, \ldots, T$, there exists

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

such that

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

Bellman equation

- We define inductively

$$\widetilde{V}_T(x_T) = \widetilde{\phi}(x_T)\ ,\ \ \forall x_T \in \mathbb{X}_T$$

$$\widetilde{V}_t(x_t) = \inf_{u_t \in \mathbb{U}_t} \int_{\mathbb{W}_{t+1}} \widetilde{V}_{t+1}\big(f_{:t}(x_t, u_t, w_{t+1})\big)\mu_t(x_t, dw_{t+1})$$

$$\forall x_t \in \mathbb{X}_t\ ,\ \ t = 0, \ldots, T$$

- We suppose that there exists a measurable selection

$$\widetilde{\psi}_t : (\mathbb{X}_t, \mathcal{X}_t) \to (\mathbb{U}_t, \mathcal{U}_t)\ ,\ \ t = 0, \ldots, T$$

such that

$$\widetilde{\psi}_t(x_t) \in \operatorname*{arg\,min}_{u_t \in \mathbb{U}_t} \int_{\mathbb{W}_{t+1}} \widetilde{V}_{t+1}\big(f_{:t}(x_t, u_t, w_{t+1})\big)\mu_t(x_t, dw_{t+1})$$

Proposition

A solution to the multistage stochastic optimization problem

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

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

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

$$\mathbf{U}_t^\star = \widetilde{\psi}_t(\mathbf{X}_t^\star)\ , \quad t = 0, \ldots, T$$

where

$$\mathbf{X}_0^\star = \mathbf{W}_0\ , \quad \mathbf{X}_{t+1}^\star = f_{:t}(\mathbf{X}_t^\star, \mathbf{U}_t^\star, \mathbf{W}_{t+1})\ , \quad t = 0, \ldots, T$$

The minimum is

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

Extension

Constraints of the form

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

Outline of the presentation

Basic data

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

- let $\mathbb{U}$, $\mathbb{W}$, $\mathbb{X}$ be measurable sets, equipped with σ-fields $\mathcal{U}$, $\mathcal{W}$, $\mathcal{X}$
- dynamics $f_{\cdot t}$

$$f_{\cdot t} : \mathbb{X} \times \mathbb{U} \times \mathbb{W} \to \mathbb{X}_{t+1}$$

- instantaneous costs L_t

$$L_t : \mathbb{X} \times \mathbb{U} \times \mathbb{W} \to \mathbb{R}$$

- final cost K

$$K : \mathbb{X} \to \mathbb{R}$$

- constraints $\mathbb{B}_t$

$$\mathbb{B}_t : \mathbb{X} \rightrightarrows \mathbb{U}$$

Bellman equation

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

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

$$V_t(x) = \inf_{u \in \mathbb{B}_t(x)} \mathbb{E}_{\mathbf{W}_{t+1}} \left[L_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}\left(f_{\cdot t}(x, u, \mathbf{W}_{t+1}) \right) \right]$$

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

- We suppose that there exists a measurable selection

$$\psi_t : (\mathbb{X}, \mathcal{X}) \to (\mathbb{U}, \mathcal{U}) \ , \quad t = 0, \ldots, T-1 \text{ such that}$$

$$\psi_t(x) \in \arg\min_{u \in \mathbb{B}_t(x)} \mathbb{E}_{\mathbf{W}_{t+1}} \left[L_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}\left(f_{\cdot t}(x, u, \mathbf{W}_{t+1}) \right) \right]$$

Proposition

A solution to the multistage stochastic optimization problem

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

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

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

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

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

where

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

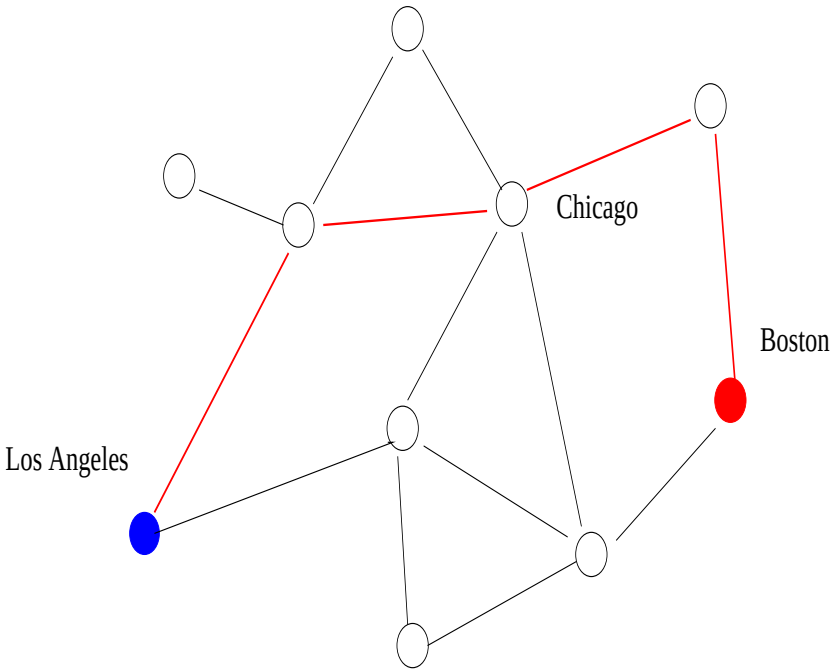
The minimum is

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

Outline of the presentation

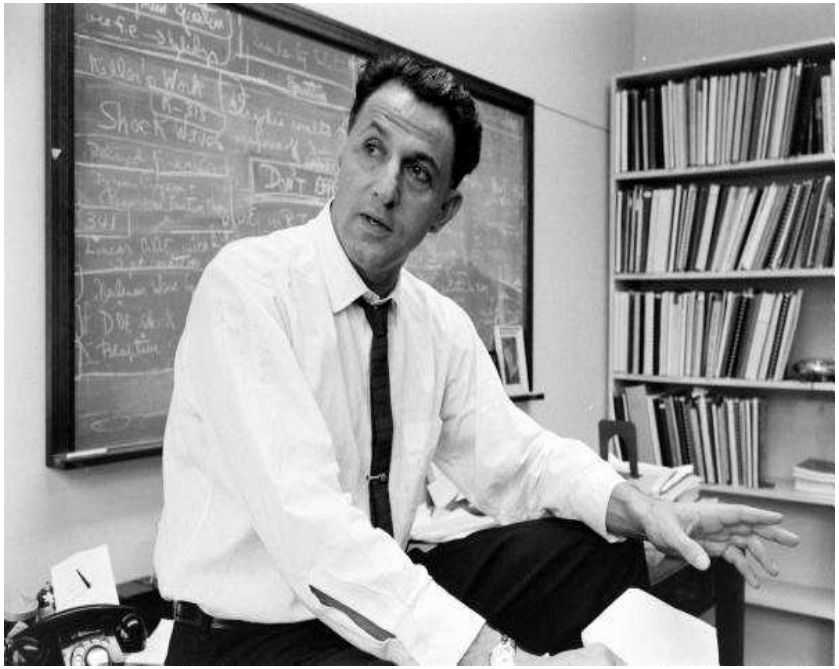
Outline of the presentation

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)

We make the assumption that the primitive random variables are independent

- The set $\mathbb{S} = \mathbb{W}^{T+1-t_0} = \mathbb{R}^q \times \cdots \times \mathbb{R}^q$ of scenarios is equipped with
  - a σ-field $\mathcal{F} = \bigotimes_{t=t_0}^{T} \mathcal{B}(\mathbb{R}^q)$
  - and a probability $\mathbb{P}$, supposed to be of product form

$$\mathbb{P} = \mu_{t_0} \otimes \cdots \otimes \mu_T$$

- Therefore, the primitive random variables

$$\mathbf{W}_{t_0}, \mathbf{W}_{t_0+1}, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$$

  are independent under $\mathbb{P}$, with marginal distributions $\mu_{t_0}, \ldots, \mu_T$
- The notation $\mathbb{E}$ refers to the mathematical expectation over $\mathbb{S}$
  - either under probability $\mathbb{P}$, as in $\mathbb{E}$, $\mathbb{E}_{\mathbf{W}}$, $\mathbb{E}_{\mathbb{P}}$
  - or under the marginal distributions $\mu_{t_0}, \ldots, \mu_T$, as in $\mathbb{E}$, $\mathbb{E}_{\mathbf{W}_{t+1}}$, $\mathbb{E}_{\mu_{t+1}}$

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

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

  where we suppose that $\mathbf{W}_{t_0}, \mathbf{W}_{t_0+1}, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$ 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 $(\omega, u_0, ..., u_{t-1})$, 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

Outline of the presentation

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

Payoff-to-go / value function / Bellman function

Assume that the primitive random variables
$\mathbf{W}_{t_0}, \mathbf{W}_{t_0+1}, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$ are independent under the probability $\mathbb{P}$
The payoff-to-go from state x at time t is

$$V_t(x) = \min_{\mathtt{Pol} \in \mathcal{U}^{ad}} \mathbb{E}\left[\sum_{s=t}^{T-1} \mathtt{L}_s(\mathbf{X}_s, \mathbf{U}_s, \mathbf{W}_{s+1}) + \mathtt{K}(\mathbf{X}_T)\right]$$

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

- The function V_t is called the value function, or the Bellman function
- The original problem is $V_{t_0}(x_0)$

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 $\mathbf{W}_{t_0}, \mathbf{W}_{t_0+1}, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$ are independent under the probability $\mathbb{P}$,
the value function $V_t(x)$ satisfies the following backward induction, where t runs from $T-1$ down to t_0

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

Algorithm for the Bellman functions

initialization $V_T(x) = \sum_{w \in W(T)} \mathbb{P}\{w\}\mathbb{K}(x)$;

for $t = T, T-1, \ldots, t_0$ **do**

  forall $x \in \mathbb{X}$ **do**

    forall $u \in \mathbb{B}_t(x)$ **do**

      forall $w \in \mathbb{W}_t$ **do**

        $l_t(x,u,w) = \mathtt{L}_t(x,u,w) + V_{t+1}\big(f_t(x,u,w)\big)$

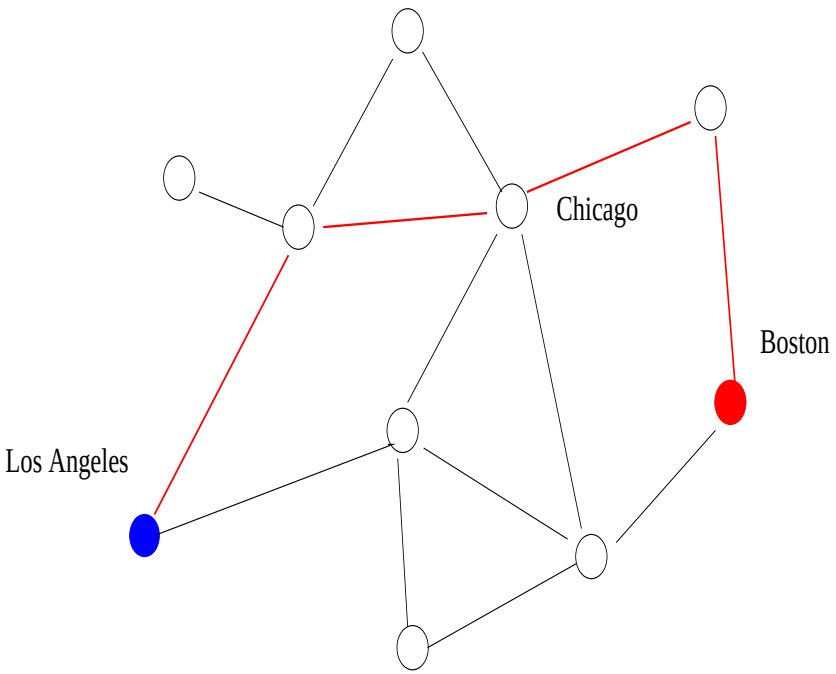
        $\sum_{w \in \mathbb{W}_t} \mathbb{P}\{w\} l_t(x,u,w)$

    $V_t(x) = \min_{u \in \mathbb{B}_t(x)} \sum_{w \in \mathbb{W}_t} \mathbb{P}\{w\} l_t(x,u,w)$;

    $\mathbb{B}_t^\star(x) = \operatorname{argmax}_{u \in \mathbb{B}_t(x)} \sum_{w \in \mathbb{W}_t} \mathbb{P}\{w\} l_t(x,u,w)$

Sketch of the proof in the deterministic case

$$V_t(x) = \min_{u \in \mathbb{B}_t(x)} \Big(\underbrace{\mathrm{L}_t(x,u)}_{\text{instantaneous gain}} + \overbrace{V_{t+1}\big(\underbrace{f_t(x,u)}_{\text{future state}}\big)}^{\text{optimal payoff}} \Big)$$



A decision u at time t in state x provides

- an instantaneous gain $\mathrm{L}_t(x,u)$
- and a future payoff for attaining the new state $f_t(x,u)$

The Bellman equation provides an optimal policy

Proposition

For any time t and state x, assume the existence of the policy $\texttt{Pol}_t^\star(x) \in$

$$\underset{u \in \mathbb{B}_t(x)}{\operatorname{argmax}} \mathbb{E}_{\mathbf{W}_{t+1}} \Big[\mathrm{L}_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathbf{W}_{t+1})\big) \Big]$$

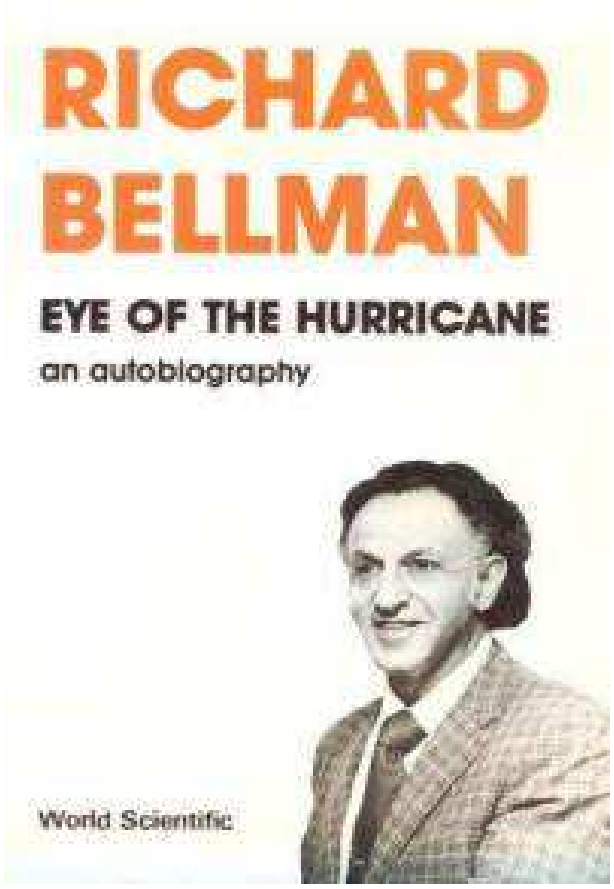
If $\texttt{Pol}_t^\star : x \mapsto \texttt{Pol}^\star t(x)$ *is measurable, then*

- $\texttt{Pol}^\star$ *is an optimal policy*
- *for any initial state x_0, the optimal expected payoff is given by*

$$V_{t_0}(x_0) = \min_{\texttt{Pol} \in \mathcal{U}^{ad}} V^{\texttt{Pol}}_{\texttt{expect}}(t_0, x_0) = V^{\texttt{Pol}^\star}_{\texttt{expect}}(t_0, x_0)$$

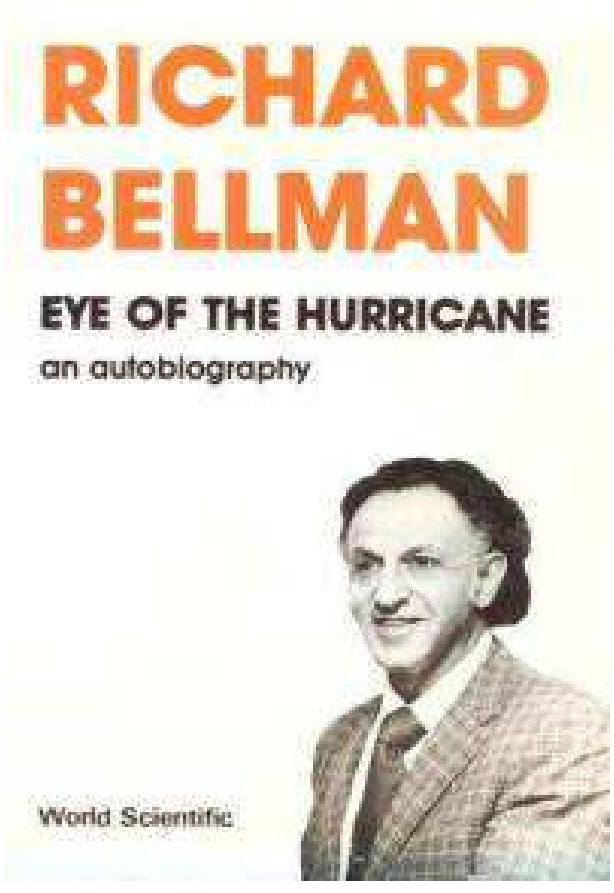
A bit of history (and fun)

"Where did the name, dynamic programming, come from?"



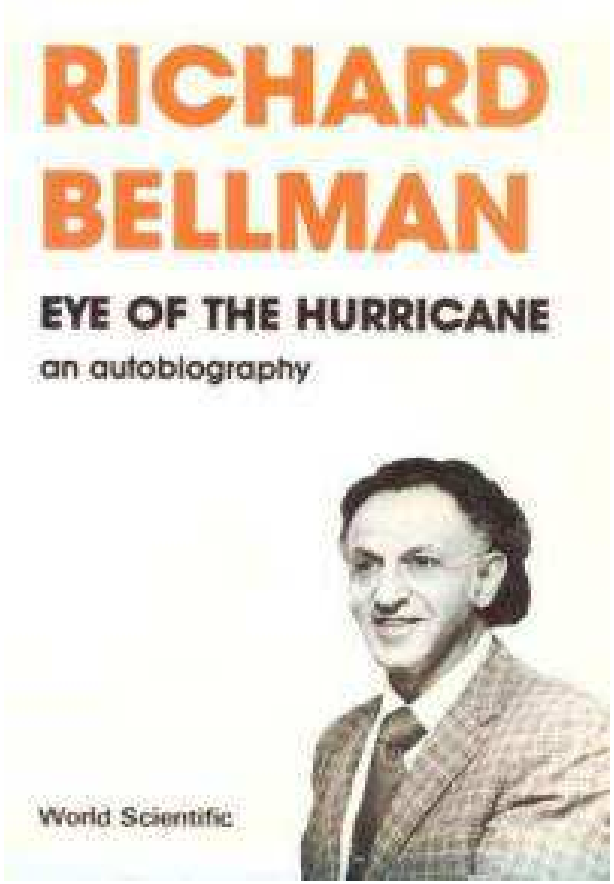
The 1950s were not good years for mathematical research. We had a very interesting gentleman in Washington named Wilson. He was Secretary of Defense, and he actually had a pathological fear and hatred of the word, research. I'm not using the term lightly; I'm using it precisely. His face would suffuse, 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?"



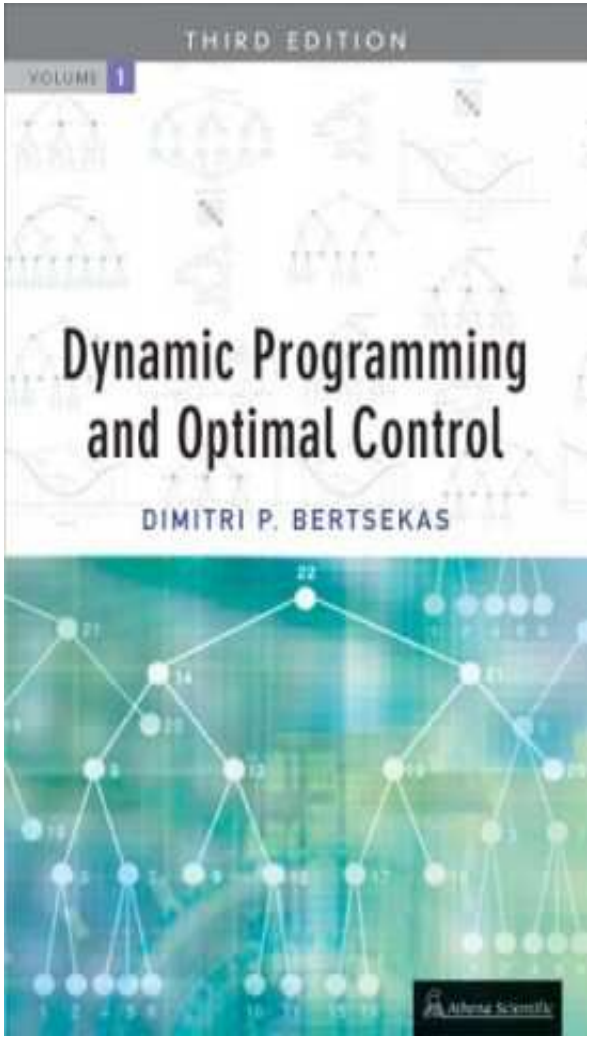
I wanted to get across the idea that this was dynamic, this was multistage, this was time-varying. I thought, let's kill two birds with one stone. Let's take a word that has an absolutely precise meaning, namely dynamic, in the classical physical sense. It also has a very interesting property as an adjective, 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.

Navigating between “backward off-line” and “forward on-line”

Optimal trajectories are calculated forward on-line

1. Initial state $x^\star_{t_0} = x_0$
2. Plug the state $x^\star_{t_0}$ into the "machine" $\texttt{Pol}_{t_0} \rightarrow$ initial decision $u^\star_{t_0} = \texttt{Pol}^\star_{t_0}(x^\star_{t_0})$
3. Run the dynamics $\rightarrow$ second state $x^\star_{t_0+1} = f_{t_0}(x^\star_{t_0}, u^\star_{t_0}, w_{t_0+1})$
4. Second decision $u^\star_{t_0+1} = \texttt{Pol}^\star_{t_0+1}(x^\star_{t_0+1})$
5. And so on $x^\star_{t_0+2} = f_{t_0+1}(x^\star_{t_0+1}, u^\star_{t_0+1}, w_{t_0+2})$
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

The curse of dimensionality :-(

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 3
if additional properties are exploited (linearity, convexity)

Outline of the presentation

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

The uncertainty is observed before making the decision

initialization $V_T(x) = \mathtt{K}(x)$;
for $t = T, T-1, \ldots, t_0$ **do**
  forall $x \in \mathbb{X}$ **do**
    forall $w \in \mathbb{W}_{t+1}$ **do**
      forall $u \in \mathbb{B}_t(x)$ **do**
        $l_t(x,u,w) = \mathtt{L}_t(x,u,w) + V_{t+1}\big(f_t(x,u,w)\big)$
      $\min_{u \in \mathbb{B}_t(x)} l_t(x,u,w)$;
      $\mathbb{B}_t^\star(x,w) = \operatorname*{argmax}_{u \in \mathbb{B}_t(x)} l_t(x,u,w)$
    $V_t(x) = \sum_{w \in \mathbb{W}_{t+1}} \mathbb{P}\{w\} \min_{u \in \mathbb{B}_t(x)} l_t(x,u,w)$

In the linear-quadratic-Gaussian case, optimal policies are linear

- When utilities are quadratic

$$\begin{aligned} \mathtt{K}(x) &= x' S_T x \\ \mathtt{L}t(x,u,w) &= x' S_t x + w' R_t w + u' Q_t u \end{aligned}$$

- and the dynamic is linear

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

- and primitive random variables $\mathbf{W}_{t_0}, \mathbf{W}_{t_0+1}, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$ are Gaussian independent under the probability $\mathbb{P}$
- then, the value functions $x \mapsto V_t(x)$ are quadratic, and optimal policies are linear

$$u_t = K_t x_t$$

How optimal decisions can be computed on-line

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

$$u_t^\star \in \underset{u \in \mathbb{B}_t(x)}{\operatorname{argmax}} \mathbb{E}_{\mathbf{W}_{t+1}} \Big[\mathrm{L}_t(x, u, \mathbf{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \mathbf{W}_{t+1})\big) \Big]$$

- 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
- by estimating basis coefficients,
  when it is known that the value function is quadratic
- by estimating upper affine approximation of the value function,
  when it is known that the value function is concave
- by estimating lower approximation of the value function,
  when restricting the search to a subclass of policies
  (open-loop in OLFC)

In the linear-convex case, value functions are convex

Here, we aim at minimizing expected cumulated costs

$$\mathbb{E}\left[\sum_{t=t_0}^{T-1} \overbrace{\mathtt{L}_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1})}^{\text{instantaneous cost}} + \underbrace{\mathtt{K}(\mathbf{X}_T)}_{\text{final cost}}\right]$$

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

- $(x, u) \mapsto \mathtt{L}_t(x, u, w)$ is jointly convex in state and control
- $x \mapsto \mathtt{K}(x)$ is convex
- the primitive random variables $\mathbf{W}_{t_0}, \mathbf{W}_{t_0+1}, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$ are independent under the probability $\mathbb{P}$
- the dynamics are affine

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

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

A lemma in convex analysis

Let $f : \mathbb{Y} \times \mathbb{Z} \to \mathbb{R}$ be convex, and let $C \subset \mathbb{Y} \times \mathbb{Z}$ be a convex set. Then

$$g(y) = \min_{z \in \mathbb{Z}, (y,z) \in C} f(y, z)$$

is a convex function

The Bellman equation produces convex value functions

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

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

- It can be shown by induction that $x \mapsto V_t(x)$ is convex
- The gradient $\nabla V_t(x_{t+1}^\star)$ at $x_{t+1}^\star$ defines a hyperplane and a lower affine approximation of the value function, calculated by duality

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

$$\underbrace{s_{t+1}}_{\text{future volume}} = \underbrace{s_t}_{\text{volume}} - \underbrace{q_t}_{\text{turbined}} - \underbrace{r_t}_{\text{spilled}} + \underbrace{a_t}_{\text{inflow volume}}$$

- s_t volume (stock) of water at the beginning of period $[t, t+1[$
- a_t, 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 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

$$0 \leq q_t \leq \min\{s_t, q^\sharp\} \text{ and } 0 \leq s_t - q_t + a_t - r_t \leq s^\sharp$$

Stochastic Dual Dynamic Programming (SDDP)

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

- Multiple stocks interconnected by linear dynamics

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

- Water inflows following an auto-regressive model

$$\mathbf{A}_t^i = \sum_{k=1,\ldots,K^i} \alpha_k \mathbf{A}^i(t-k) + \mathbf{W}_t$$

where the random variables $\mathbf{W}_{t_0}, \mathbf{W}_{t_0+1}, \ldots, \mathbf{W}_{T-1}, \mathbf{W}_T$ are independent

Summary

- Bellman's Principle of Optimality breaks
  an intertemporal optimization problem
  into a sequence of interconnected static optimization problems
- The payoff-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