Robust and Stochastic Optimal Sequential Control

Extended from Chapter 8 of
Sustainable Management of Natural Resources. Mathematical Models and Methods
by Luc DOYEN and Michel DE LARA

Michel De Lara
Cermics, École des Ponts ParisTech
Université Paris-Est
France

École des Ponts ParisTech

January 5, 2016

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
2. The stochastic optimality problem and dynamic programming
3. Applications to stochastic resources optimal management
4. The robust optimality problem and dynamic programming
5. Summary

Outline of the presentation

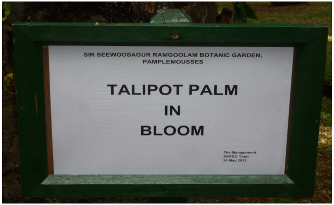
1. Optimization intertemporal criteria under uncertainty
2. The stochastic optimality problem and dynamic programming
3. Applications to stochastic resources optimal management
4. The robust optimality problem and dynamic programming
5. Summary

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

In 2012, the Botanic Garden in Mauritius Island witnessed an exceptional blooming of the talipot palm

- In 2012, at Sir Seewoosagur Ramgoolam Botanic Garden in Mauritius Island, the talipot palm *Corypha umbraculifera* was in bloom
- This remarkable event occurs only once in the life of this species (monocarpic)
- The palm flowers only once, when it is 30 to 80 years old, produces fruits, and dies after fruiting



SIR SEEWOOSAGUR RAMGOOLAM BOTANIC GARDEN,
PAMPLEMOUSSES

TALIPOT PALM
IN
BLOOM

The Management
SSRBG Trust
24 May 2012

Organisms trade growth off for reproduction

Organisms (vegetal, animal) trade growth off for reproduction
to achieve the largest number of offspring

- The bigger a plant today, the bigger tomorrow (leafs and roots capturing more resources)
- Therefore, it might be interesting to postpone reproduction and convert all final biomass into seeds
- But, it the environment is hostile in the sense that the plant faces a sequence of independent death threats, it may be better to start reproducing early
- Fishes and snakes grow and reproduce during all their life time (wild salmon dies after spawning)

What is investing?

Investing is refraining from consuming now
at the benefit of more consumption in one year
at the expense of being dead in one year
(the first reason for discounting the future)

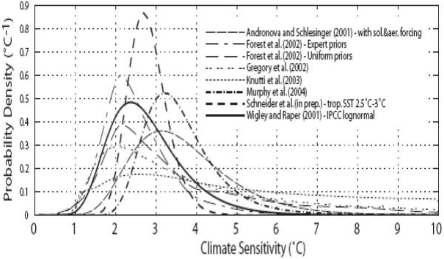
The debate on the timing of decisions for mitigating climate change

Carbon cycle model and uncertain damages

- $\mathrm{CO_2}$ concentration $M(t)$

$$M(t+1) = M(t) \underbrace{-\delta(M(t) - M_{-\infty})}_{\text{natural sinks}} + \alpha \overbrace{\mathtt{Emiss}(t)}^{\text{emissions}} \underbrace{(1 - a(t))}_{\text{abatement}}$$

- decision $a(t) \in [0,1]$ is the abatement rate of $\mathrm{CO_2}$ emissions.


Probability Density (°C⁻¹)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
1
2
3
4
5
6
7
8
9
10
Climate Sensitivity (°C)
Andronova and Schlesinger (2001) - with sol.&aer. forcing
Forest et al. (2002) - Expert priors
Forest et al. (2002) - Uniform priors
Gregory et al. (2002)
Knutti et al. (2003)
Murphy et al. (2004)
Schneider et al. (in prep.) - trop. SST 2.5°C-3°C
Wigley and Raper (2001) - IPCC lognormal

Mitigation for climate change under uncertainty

- Three periods: $t = 0$ (today), $t = 1$ (in twenty-five years), $t = 2$ (in fifty years)
- First-period abatement cost $\texttt{Cost}\big(a(0)\big)$
- Discounted second-period abatement cost $\delta\texttt{Cost}\big(a(1)\big)$, where $\delta = \frac{1}{1+r_e}$
- Discounted final damage cost $\delta^2\texttt{Damage}\big(M(2), \theta(2)\big)$ depends on
  - CO_2 final concentration $M(2)$
  - uncertain damage sensitivity to climate $\theta(2)$ in fifty years

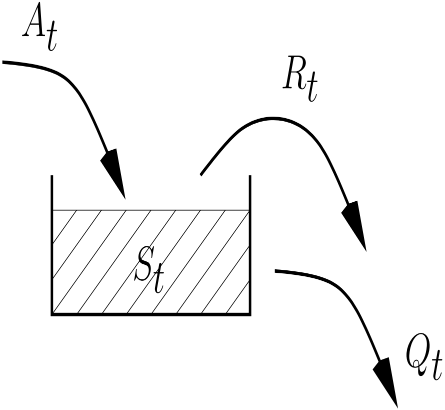
Total costs are $\texttt{Crit}(M(\cdot), a(\cdot), \theta(\cdot)) =$

$$\texttt{Cost}\big(a(0)\big) + \delta\texttt{Cost}\big(a(1)\big) + \delta^2\texttt{Damage}(M(2), \theta(2))$$

and they depend upon the uncertainty $\theta(2)$

Managing a dam requires many decisions over several years

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

$$\underbrace{S(t+1)}_{\text{future volume}} = \min\{S^\sharp, \underbrace{S(t)}_{\text{volume}} - \underbrace{q(t)}_{\text{turbined}} + \underbrace{a(t)}_{\text{inflow volume}} \}$$

- $S(t)$ volume (stock) of water at the beginning of period $[t, t+1[$
- $a(t)$ inflow water volume (rain, etc.) during $[t, t+1[$
- decision-hazard:
  $a(t)$ 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 $a(t)$
  - 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 $\mathbb{S} = \mathbb{R}^{T-t_0}$ of water inflows scenarios $\big(a(t_0), \ldots, a(T-1)\big)$
  - 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 $K\big(S(T)\big)$, the "final value of water"
- The traditional economic problem is to maximize the intertemporal payoff (without discounting if the horizon is short)

$$\max \mathbb{E} \left[\sum_{t=t_0}^{T-1} \overbrace{p(t)q(t)}^{\text{turbined water payoff}} + \overbrace{K\big(S(T)\big)}^{\text{final volume utility}} \right]$$

Let us fix notations and vocabulary

Uncertainty variables are new input variables in a discrete-time nonlinear state-control system

A specific output is distinguished, and is labeled "state" (more on this later), when the system may be written

$$x(t+1) = \mathtt{Dyn}\big(t, x(t), u(t), w(t)\big) , \quad t \in \mathbb{T} = \{t_0, t_0+1, \ldots, T-1\}$$

- time $t \in \overline{\mathbb{T}} = \{t_0, t_0+1, \ldots, T-1, T\} \subset \mathbb{N}$ (the time period $[t, t+1[$ may be a year, a month, etc.)
- state $x(t) \in \mathbb{X} := \mathbb{R}^n$ (biomasses, abundances, etc.)
- control $u(t) \in \mathbb{U} := \mathbb{R}^p$ (catches or harvesting effort)
- uncertainty $w(t) \in \mathbb{W} := \mathbb{R}^q$ (recruitment or mortality uncertainties, climate fluctuations or trends, etc.)
- dynamics $\mathtt{Dyn}$ maps $\mathbb{T} \times \mathbb{X} \times \mathbb{U} \times \mathbb{W}$ into $\mathbb{X}$ (biomass model, age-class model, economic model)

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

Histories and criterion for the single dam optimization problem

Single dam histories

$$\big(S(\cdot), q(\cdot), a(\cdot)\big) = \big(\overbrace{S(t_0), \ldots, S(T)}^{\text{stocks}}, \overbrace{q(t_0), \ldots, q(T-1)}^{\text{turbined}}, \overbrace{a(t_0), \ldots, a(T-1)}^{\text{inflows}}\big)$$

Intertemporal payoff for a single dam

$$\texttt{Crit}\big(S(\cdot), q(\cdot), a(\cdot)\big) = \sum_{t=t_0}^{T-1} \overbrace{\underbrace{p(t)}_{\text{price}} \times \underbrace{q(t)}_{\text{quantity}}}^{\text{turbined water profit}} + \overbrace{K\big(S(T)\big)}^{\text{final stock utility}}$$

An intertemporal criterion attaches a value to a history and performs an aggregation with respect to time, reflecting preferences across time

- The history space is

$$\mathbb{H} := \overbrace{\underbrace{\mathbb{X}^{T+1-t_0}}_{\text{state}} \times \underbrace{\mathbb{U}^{T-t_0}}_{\text{control}} \times \underbrace{\mathbb{W}^{T+1-t_0}}_{\text{uncertainty}}}^{\text{history space}}$$

- A criterion `Crit` is a function

$$\texttt{Crit} : \mathbb{H} \to \mathbb{R}$$

which assigns

- a scalar value $\texttt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) \in \mathbb{R}$
- to a history $\big(x(\cdot), u(\cdot), w(\cdot)\big) \in \mathbb{H}$

Here are the most common intertemporal criteria

The additive criterion is the most common and sums payoffs over time-periods

- The traditional discounted present value is

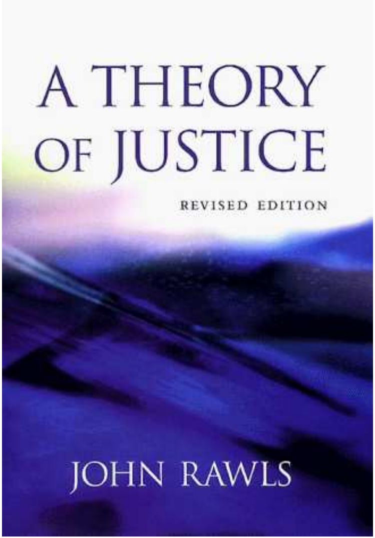
$$\sum_{t=t_0}^{+\infty} \delta^{t-t_0} \mathtt{L}\big(x(t), u(t), w(t)\big)$$

- The time-separable additive criterion includes *discounted present value, Green Golden, Chichilnisky*

$$\begin{aligned} \mathtt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) &= \sum_{t=t_0}^{T-1} \overbrace{\mathtt{L}\big(t, x(t), u(t), w(t)\big)}^{\text{instantaneous gain}} \\ &+ \underbrace{\mathtt{K}\big(x(T), w(T)\big)}_{\text{final gain}} \end{aligned}$$

- The payoffs in one time-period may be compensated by those of other time-periods

The maximin criterion focuses on the worst payoff accross time-periods



- Equity: a focus on the poorest generation
- The maximin form or Rawls criterion is

$$\texttt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) = \underbrace{\min_{t=t_0,\ldots,T-1} \overbrace{\texttt{L}\big(t, x(t), u(t), w(t)\big)}^{\text{generation utility}}}_{\text{worse generation utility}}$$

Summary

- A criterion attaches a value to a history
  and performs an aggregation with respect to time,
  reflecting preferences across time
- How can we attach a value to a policy,
  so that we can rank policies?

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
    - Examples of trade-offs in intertemporal stochastic optimization
    - An intertemporal criterion displays preferences with respect to time
    - How can we rank policies with a criterion under uncertainty?
    - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
    - The payoff-to-go and Bellman's Principle of Optimality
    - Bellman equation and the curse of dimensionality
    - Complements
3. Applications to stochastic resources optimal management
    - A model of plant growth over a finite horizon
    - Biomass linear models
    - The inventory problem
    - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

An example of state and control solution maps

Volume and turbined trajectories under a given policy

Consider a dam modeled as $S(t+1) = S(t) - q(t) + a(t)$, where there is no spilling by supposing that the total volume $S^\sharp = +\infty$, and pick up

- a scenario $a(\cdot) = \big(a(t_0), a(t_0+1), \ldots, a(T)\big)$ of water inflows
- a policy $\mathtt{Pol}(t,S) = \alpha(t)S$ with $0 \leq \alpha(t) \leq 1$
- an initial state (volume) $S(t_0)$

1. Initial decision $q(t_0) = \alpha(t_0)S(t_0)$
2. Second state $S(t_0+1) = \big(1-\alpha(t_0)\big)S(t_0) + a(t_0)$
3. Second decision $q(t_0+1) = \alpha(t_0+1)S(t_0+1) = \alpha(t_0+1)\Big(\big(1-\alpha(t_0)\big)S(t_0) + a(t_0)\Big)$
4. And so on $S(t_0+2) = \big(1-\alpha(t_0+1)\big)S(t_0+1) + a(t_0+1) = \big(1-\alpha(t_0+1)\big)\big(1-\alpha(t_0)\big)S(t_0) + \big(1-\alpha(t_0+1)\big)a(t_0) + a(t_0+1)$
5. ...

A policy and a scenario yield a history that is evaluated by a criterion (time aggregation)

Turbined and final volume payoff under a given policy

Plug the solution maps

1. $q(t_0) = \alpha(t_0) S(t_0)$
2. $S(t_0+1) = \big(1-\alpha(t_0)\big) S(t_0) + a(t_0)$
3. $q(t_0+1) = \alpha(t_0+1)\big(1-\alpha(t_0)\big) S(t_0) + a(t_0)$
4. $S(t_0+2) = \big(1-\alpha(t_0+1)\big)\big(1-\alpha(t_0)\big) S(t_0) + \big(1-\alpha(t_0+1)\big) a(t_0) + a(t_0+1)$
5. ...

into the criterion

$$\texttt{Crit}\big(S(\cdot), q(\cdot), a(\cdot)\big) = \sum_{t=t_0}^{T-1} \overbrace{p(t)q(t)}^{\text{water release profit}} + \overbrace{K\big(S(T)\big)}^{\text{final stock utility}}$$

Let us fix notations and vocabulary

Admissible state feedback policies express control constraints

The control constraints case restricts policies to admissible policies

$$\mathcal{U}^{ad} := \{\texttt{Pol} : \mathbb{T} \times \mathbb{X} \to \mathbb{U} \mid \texttt{Pol}(t,x) \in \mathbb{B}(t,x) \,, \quad \forall (t,x)\}$$

Dam management physical volume constraint

In a water reservoir, the output flow (control) cannot be more than the stock volume (state) and than a capacity constraint

$$0 \leq q(t) \leq S(t) \text{ and } 0 \leq q(t) \leq q^\sharp$$

For instance, a dam management policy of the form

$$\texttt{Pol}(t,S) = \max\{q^\flat, \min\{\alpha(t)S, q^\sharp, S\}\}$$

is admissible, where $0 \leq q^\flat \leq q^\sharp$ captures a requirement of minimal outflow (for biodiversity preservation in downward rivers, for instance)

A policy and a scenario yield a history that is evaluated by a criterion (time aggregation)

- The criterion $\mathtt{Crit}$ maps the history space $\mathbb{H}$ towards $\mathbb{R}$
- For t_0 the initial time, and $x_0 \in \mathbb{X}$ the initial state, the evaluation of the criterion is

$$\mathtt{Crit}^{\mathtt{Pol}}\big(t_0, x_0, w(\cdot)\big) :=$$

$$\mathtt{Crit}\Big(\underbrace{X_{\mathtt{Dyn}}[t_0, x_0, \mathtt{Pol}, w(\cdot)](\cdot)}_{\text{state trajectory}}, \underbrace{U_{\mathtt{Dyn}}[t_0, x_0, \mathtt{Pol}, w(\cdot)](\cdot)}_{\text{control trajectory}}, w(\cdot)\Big) \in \mathbb{R}$$

A policy and a criterion yield a real-valued payoff

Given a policy $\texttt{Pol} \in \mathcal{U}^{ad}$ and a scenario $w(\cdot) \in \mathbb{S}$, we obtain a payoff

$$\texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big) = \texttt{Crit}^{\texttt{Pol}}\big(t_0, x_0, w(\cdot)\big)$$

hence a mapping $\mathcal{U}^{ad} \times \mathbb{S} \to \mathbb{R}$

Policies/Scenarios	$w^A(\cdot) \in \mathbb{S}$	$w^B(\cdot) \in \mathbb{S}$	$\ldots$
$\texttt{Pol}_1 \in \mathcal{U}^{ad}$	$\texttt{Payoff}\big(\texttt{Pol}_1, w^A(\cdot)\big)$	$\texttt{Payoff}\big(\texttt{Pol}_1, w^B(\cdot)\big)$	$\ldots$
$\texttt{Pol}_2 \in \mathcal{U}^{ad}$	$\texttt{Payoff}\big(\texttt{Pol}_2, w^A(\cdot)\big)$	$\texttt{Payoff}\big(\texttt{Pol}_2, w^B(\cdot)\big)$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$

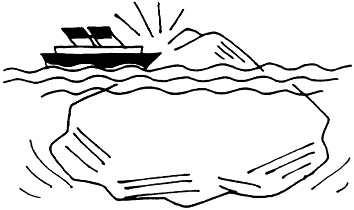
Summary

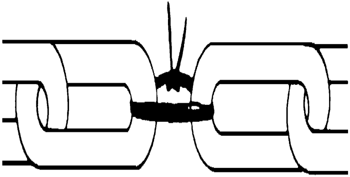
- An intertemporal criterion $\texttt{Crit}$ attaches a value to a history and performs an aggregation with respect to time, reflecting preferences across time
- A policy $\texttt{Pol}$ and a scenario $w(\cdot)$ yield a history, thanks to the state and control solution maps, that is evaluated by a criterion $\texttt{Crit}$ (time aggregation), yielding $\texttt{Crit}^{\texttt{Pol}}\big(t_0, x_0, w(\cdot)\big)$
- A policy $\texttt{Pol}$ and a criterion $\texttt{Crit}$ yield a real-valued mapping $w(\cdot) \in \mathbb{S} \mapsto \texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big) = \texttt{Crit}^{\texttt{Pol}}\big(t_0, x_0, w(\cdot)\big)$ over the scenarios $\mathbb{S}$
- Therefore, comparing policies amounts to comparing mappings over the scenarios $\mathbb{S}$
- For this purpose, we will see how to aggregate the real-valued mapping $w(\cdot) \in \mathbb{S} \mapsto \texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)$ with respect to scenarios

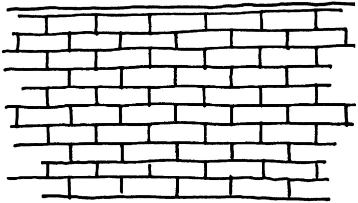
Outline of the presentation

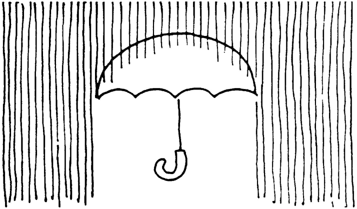
1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

In the robust or pessimistic approach, Nature is supposed to be malevolent, and the DM aims at protection against all odds









In the robust or pessimistic approach, Nature is supposed to be malevolent

- In the robust approach, the DM considers the worst payoff

$$\underbrace{\min_{w(\cdot)\in\Omega} \texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)}_{\text{worst payoff}}$$

- Nature is supposed to be malevolent,
  and specifically selects the worst scenario:
  the DM plays after Nature has played, and maximizes the worst payoff

$$\max_{\texttt{Pol}\in\mathfrak{U}^{ad}} \min_{w(\cdot)\in\Omega} \texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)$$

- Robust, pessimistic, worst-case, maximin, minimax (for costs)

Guaranteed energy production

In a dam, the minimal energy production in a given period,
corresponding to the worst water inflow scenario

The robust approach can be softened with plausibility weighting

- Let $\Theta : \Omega \to \mathbb{R} \cup \{-\infty\}$ be a a plausibility function.
- The higher, the more plausible:
  totally implausible scenarios are those for which $\Theta\big(w(\cdot)\big) = -\infty$
- Nature is malevolent, and specifically selects the worst scenario, but weighs it according to the plausibility function Θ
- The DM plays after Nature has played, and solves

$$\max_{\texttt{Pol} \in \mathfrak{U}^{ad}} \left[\min_{w(\cdot) \in \Omega} \left(\texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big) - \underbrace{\Theta\big(w(\cdot)\big)}_{\text{plausibility}} \right) \right]$$

In the optimistic approach, Nature is supposed to benevolent

Future. That period of time in which our affairs prosper, our friends are true and our happiness is assured.

Ambrose Bierce

- Instead of maximizing the worst payoff as in a robust approach, the optimistic perspective focuses on the most favorable payoff

$$\underbrace{\max_{w(\cdot)\in\Omega} \mathtt{Payoff}\big(\mathtt{Pol}, w(\cdot)\big)}_{\text{best payoff}}$$

- Nature is supposed to benevolent, and specifically selects the best scenario: the DM plays after Nature has played, and solves

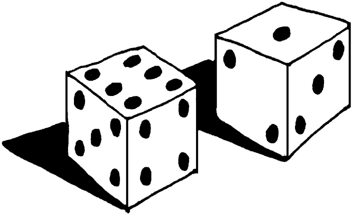
$$\max_{\mathtt{Pol}\in\mathfrak{U}^{ad}} \max_{w(\cdot)\in\Omega} \mathtt{Payoff}\big(\mathtt{Pol}, w(\cdot)\big)$$

The Hurwicz criterion reflects an intermediate attitude between optimistic and pessimistic approaches

A proportion $\alpha \in [0, 1]$ graduates the level of prudence

$$\max_{\texttt{Pol} \in \mathcal{U}^{ad}} \left\{ \alpha \overbrace{\min_{w(\cdot) \in \Omega} \texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)}^{\text{pessimistic}} + (1 - \alpha) \underbrace{\max_{w(\cdot) \in \Omega} \texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)}_{\text{optimistic}} \right\}$$

In the stochastic or expected approach, Nature is supposed to play stochastically



In the stochastic or expected approach, Nature is supposed to play stochastically

- The expected payoff is

$$\overbrace{\mathbb{E}\Big[\texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)\Big]}^{\text{mean payoff}} = \sum_{w(\cdot)\in\mathbb{S}} \mathbb{P}\{w(\cdot)\}\texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)$$

- Nature is supposed to play stochastically, according to distribution $\mathbb{P}$: the DM plays after Nature has played, and solves

$$\max_{\texttt{Pol}\in\mathfrak{U}^{ad}} \mathbb{E}\Big[\texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)\Big]$$

- The discounted expected utility is the special case

$$\mathbb{E}\Big[\sum_{t=t_0}^{+\infty} \delta^{t-t_0}\mathtt{L}\big(x(t), u(t), w(t)\big)\Big]$$

The expected utility approach distorts payoffs before taking the expectation

- We consider a utility function $\mathtt{L}$ to assess the utility of the payoffs (for instance a CARA exponential utility function)
- The expected utility is

$$\underbrace{\mathbb{E}\left[\mathtt{L}\Big(\mathtt{Payoff}\big(\mathtt{Pol}, w(\cdot)\big)\Big)\right]}_{\text{expected utility}} = \sum_{w(\cdot)\in\mathbb{S}} \mathbb{P}\{w(\cdot)\}\mathtt{L}\Big(\mathtt{Payoff}\big(\mathtt{Pol}, w(\cdot)\big)\Big)$$

- The expected utility maximizer solves

$$\max_{\mathtt{Pol}\in\mathfrak{U}^{ad}} \mathbb{E}\left[\mathtt{L}\Big(\mathtt{Payoff}\big(\mathtt{Pol}, w(\cdot)\big)\Big)\right]$$

The ambiguity or multi-prior approach combines robust and expected criterion

- Different probabilities $\mathbb{P}$, termed as beliefs or priors, and belonging to a set $\mathcal{P}$ of admissible probabilities on Ω
- The multi-prior approach combines robust and expected criterion, by taking the worst beliefs in terms of expected payoff

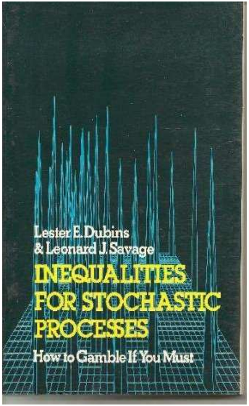
$$\max_{\texttt{Pol}\in\mathfrak{U}^{ad}} \underbrace{\min_{\mathbb{P}\in\mathcal{P}} \overbrace{\mathbb{E}^{\mathbb{P}}\Big[\texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)\Big]}^{\text{mean payoff}}}_{\text{pessimistic over probabilities}}$$

Convex risk measures cover a wide range of risk criteria

- Different probabilities $\mathbb{P}$, termed as beliefs or priors, and belonging to a set $\mathcal{P}$ of admissible probabilities on Ω
- To each probability $\mathbb{P}$ is attached a plausibility $\Theta(\mathbb{P})$

$$\max_{\texttt{Pol}\in\mathfrak{U}^{ad}} \underbrace{\min_{\mathbb{P}\in\mathcal{P}} \overbrace{\mathbb{E}^{\mathbb{P}}\Big[\texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big)\Big]}^{\text{mean payoff}} - \overbrace{\Theta(\mathbb{P})}^{\text{plausibility}}}_{\text{pessimistic over probabilities}}$$

Non convex risk measures can lead to non diversification



How to gamble if you must,
L.E. Dubbins and L.J. Savage,
1965

Imagine yourself at a casino with $1,000. For some reason, you desperately need $10,000 by morning; anything less is worth nothing for your purpose.

The only thing possible is to gamble away your last cent, if need be, in an attempt to reach the target sum of $10,000.

- The question is how to play, not whether. What ought you do? How should you play?
  - Diversify, by playing 1 $ at a time?
  - Play boldly and concentrate,
    by playing 10,000 $ only one time?
- What is your decision criterion?

Savage's minimal regret criterion... "Had I known"

$$\min_{\texttt{Pol} \in \mathfrak{U}^{ad}} \Big\{ \overbrace{\max_{w(\cdot)\in\Omega} \underbrace{\Big[\max_{\text{anticipative policies } \overline{\texttt{Pol}}} \texttt{Payoff}\big(\overline{\texttt{Pol}}, w(\cdot)\big) - \texttt{Payoff}\big(\texttt{Pol}, w(\cdot)\big) \Big]}_{\text{regret}}}^{\text{worst regret}} \Big\}$$

- If the DM knows the future in advance, she solves $\max_{\text{anticipative policies } \overline{\texttt{Pol}}} \texttt{Payoff}\big(\overline{\texttt{Pol}}, w(\cdot)\big)$, for each scenario $w(\cdot) \in \Omega$
- The regret attached to a non-anticipative policy $\texttt{Pol} \in \mathfrak{U}^{ad}$ is the loss due to not being visionary
- The best a non-visionary DM can do with respect to regret is minimizing it

Summary

- A criterion attaches a value to a history
  and performs an aggregation with respect to time,
  reflecting preferences across time
- Off-line information on scenarios makes possible
  different aggregations with respect to uncertainties,
  reflecting risk attitudes
  and preferences across scenarios
- Policies are compared with respect to
  both time and uncertainties payoffs aggregations
- How do we compute optimal policies?

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
2. The stochastic optimality problem and dynamic programming
3. Applications to stochastic resources optimal management
4. The robust optimality problem and dynamic programming
5. Summary

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

Maximizing the expected additive payoff

- The expected additive payoff is $\max_{\mathtt{Pol} \in \mathfrak{U}^{ad}} \mathbb{E}\Big[\mathtt{Payoff}\big(\mathtt{Pol}, w(\cdot)\big)\Big]$ where

$$\mathtt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) = \sum_{t=t_0}^{T-1} \overbrace{\mathtt{L}\big(t, x(t), u(t), w(t)\big)}^{\text{instantaneous gain}} + \underbrace{\mathtt{K}\big(x(T), w(T)\big)}_{\text{final gain}}$$

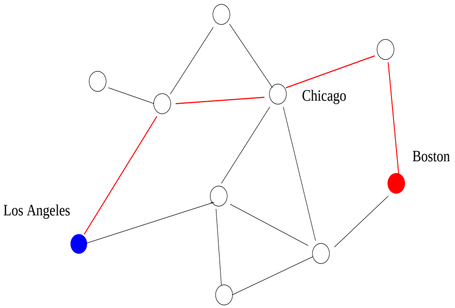
- The optimization problem is traditionally written as

$$\max_{u(\cdot)} \mathbb{E}\Big[\sum_{t=t_0}^{T-1} \mathtt{L}\big(t, x(t), u(t), w(t)\big) + \mathtt{K}\big(x(T), w(T)\big)\Big]$$

  - where the last expression is abusively used, but practical and traditional,
  - in which $x(\cdot)$ and $u(\cdot)$ need to be replaced by

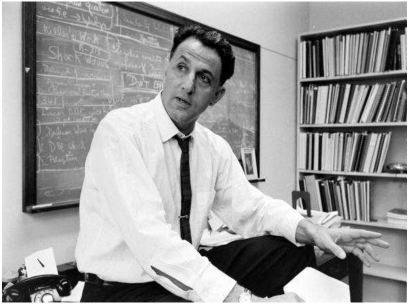
$$x(t) = X_{\mathtt{Dyn}}[t_0, x_0, \mathtt{Pol}, w(\cdot)](t) \text{ and } u(t) = \mathtt{Pol}\big(t, x(t)\big)$$

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 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

$$w(t_0), w(t_0+1), \ldots, w(T-1), w(T)$$

  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 $\mathbb{E}$, $\mathbb{E}_{w(\cdot)}$, $\mathbb{E}_{\mathbb{P}}$
  - or under the marginal distributions $\mu_{t_0}, \ldots, \mu_T$, as in $\mathbb{E}$, $\mathbb{E}_{w(t)}$, $\mathbb{E}_{\mu_t}$

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

$$\begin{aligned}\overbrace{S(t+1)}^{\text{future stock}} &= \min\{S^\sharp, \overbrace{S(t)}^{\text{stock}} - \overbrace{q(t)}^{\text{water release}} + \overbrace{a(t)}^{\text{water inflows}}\} \\ \underbrace{a(t+1)}_{\text{future water inflows}} &= \alpha \underbrace{a(t)}_{\text{water inflows}} + \underbrace{w(t)}_{\text{noise}}\end{aligned}$$

  where we suppose that $\big(w(t_0), \ldots, w(T-1)\big)$ form
  a sequence of independent random variables

- The couple $x(t) = \big(S(t), a(t)\big)$ is a sufficient summary of past controls and uncertainties to do forecasting:
  knowing the state $x(t) = \big(S(t), a(t)\big)$ at time t is sufficient to forecast $x(t+1)$, given the control $q(t)$ and the uncertainty $w(t)$

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 $(\omega, u_0, \dots, u_{t-1})$, 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

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

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

Payoff-to-go / value function / Bellman function

Assume that the primitive random variables $\big(w(t_0), w(t_0+1), \dots, w(T-1), w(T)\big)$ are independent under the probability $\mathbb{P}$. The payoff-to-go from state x at time t is

$$V(t,x) := \max_{\texttt{Pol} \in \mathfrak{U}^{ad}} \mathbb{E}\Big[\sum_{s=t}^{T-1} \texttt{L}\big(s, x(s), u(s), w(s)\big) + \texttt{K}\big(x(T), w(T)\big)\Big]$$

where $x(t) = x$ and, for $s = t, \dots, T-1$, $x(s+1) = \texttt{Dyn}\big(s, x(s), u(s), w(s)\big)$ and $u(s) = \texttt{Pol}\big(s, x(s)\big)$

- The function V 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 $\big(w(t_0), w(t_0+1), \dots, w(T-1), w(T)\big)$ are independent under the probability $\mathbb{P}$, the value function $V(t,x)$ associated with the additive criterion satisfies the following backward induction,
where t runs from $T-1$ down to t_0

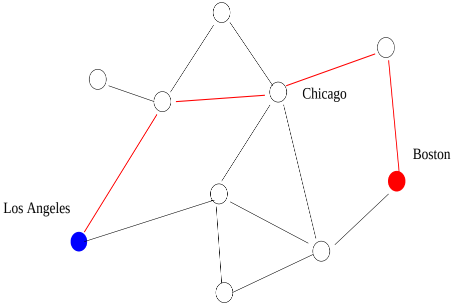
$$
\begin{aligned}
V(T,x) &= \mathbb{E}_{w(T)}\Big[\mathtt{K}\big(x, w(T)\big)\Big] \\
V(t,x) &= \max_{u \in \mathbb{B}(t,x)} \mathbb{E}_{w(t)}\Big[\mathtt{L}\big(t,x,u,w(t)\big) + V\Big(t+1, \mathtt{Dyn}\big(t,x,u,w(t)\big)\Big)\Big]
\end{aligned}
$$

Algorithm for the Bellman functions

```
initialization $V(T,x) = \sum_{w \in \mathbb{S}(T)} \mathbb{P}\{w\} \mathtt{K}(x,w)$;
for $t = T, T-1, \dots, t_0$ do
    forall the $x \in \mathbb{X}$ do
        forall the $u \in \mathbb{B}(t,x)$ do
            forall the $w \in \mathbb{S}(t)$ do
                $l(t,x,u,w) = \mathtt{L}(t,x,u,w) + V\big(t+1, \mathtt{Dyn}(t,x,u,w)\big)$
            $\sum_{w \in \mathbb{S}(t)} \mathbb{P}\{w\} l(t,x,u,w)$
    $V(t,x) = \max_{u \in \mathbb{B}(t,x)} \sum_{w \in \mathbb{S}(t)} \mathbb{P}\{w\} l(t,x,u,w)$;
    $\mathbb{B}^\star(t,x) = \operatorname{argmax}_{u \in \mathbb{B}(t,x)} \sum_{w \in \mathbb{S}(t)} \mathbb{P}\{w\} l(t,x,u,w)$
```

Sketch of the proof in the deterministic case

$$V(t,x) = \max_{u \in \mathbb{B}(t,x)} \Big(\underbrace{\mathtt{L}(t,x,u)}_{\text{instantaneous gain}} + \overbrace{V\big(t+1, \underbrace{\mathtt{Dyn}(t,x,u)}_{\text{future state}}\big)}^{\text{optimal payoff}} \Big)$$



A decision u at time t in state x provides

- an instantaneous gain $\mathtt{L}(t,x,u)$
- and a future payoff for attaining the new state $\mathtt{Dyn}(t,x,u)$

The Bellman equation provides an optimal policy

Proposition

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

$$\operatorname*{argmax}_{u \in \mathbb{B}(t,x)} \mathbb{E}_{w(t)} \Big[\mathtt{L}\big(t,x,u,w(t)\big) + V\Big(t+1, \mathtt{Dyn}\big(t,x,u,w(t)\big)\Big) \Big]$$

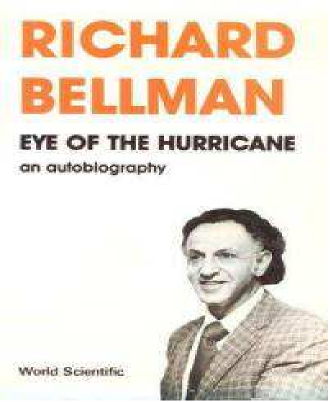
If $\texttt{Pol}^\star : (t,x) \mapsto \texttt{Pol}^\star(t,x)$ *is measurable, then*

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

$$V(t_0,x_0) = \max_{\texttt{Pol} \in \mathfrak{U}^{ad}} \texttt{Crit}^{\texttt{Pol}}_{\texttt{expect}}(t_0,x_0) = \texttt{Crit}^{\texttt{Pol}^\star}_{\texttt{expect}}(t_0,x_0)$$

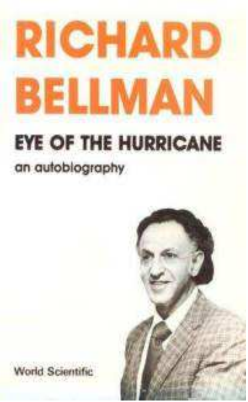
A bit of history (and fun)

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



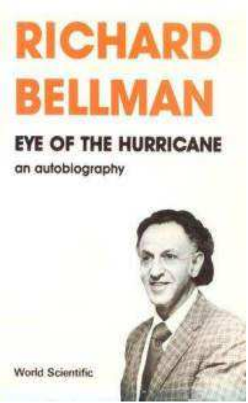
The 1950s were not good years for mathematical research. We had a very interesting gentleman in Washington named Wilson. He was Secretary of Defense, and he actually had a pathological fear and hatred of the word, research. I'm not using the term lightly; I'm using it precisely. His face would suffuse,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



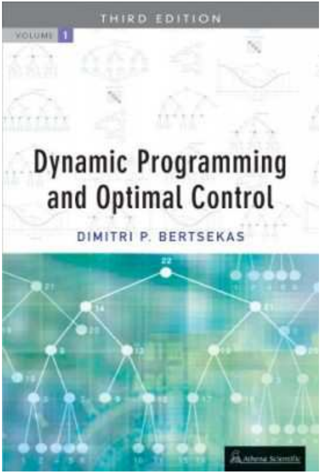
I wanted to get across the idea that this was dynamic, this was multistage, this was time-varying. I thought, let’s kill two birds with one stone. Let’s take a word that has an absolutely precise meaning, namely dynamic, in the classical physical sense. It also has a very interesting property as an adjective, 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" Pol → initial decision $u^{\star}(t_0) = \texttt{Pol}^{\star}(t_0, x^{\star}(t_0))$
3. Run the dynamics → second state $x^{\star}(t_0+1) = \texttt{Dyn}\big(t_0, x^{\star}(t_0), u^{\star}(t_0), w(t_0)\big)$
4. Second decision $u^{\star}(t_0+1) = \texttt{Pol}^{\star}(t_0+1, x^{\star}(t_0+1))$
5. And so on $x^{\star}(t_0+2) = \texttt{Dyn}\big(t_0+1, x^{\star}(t_0+1), u^{\star}(t_0+1)), w(t_0+1)\big)$
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 we discretize a state of dimension 4
  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)

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 practice, the curse of dimensionality
  forbids to use dynamic programming
  for a state with dimension more than three or four

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

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

The uncertainty is observed before making the decision

```
initialization $V(T,x) = \sum_{w \in \mathbb{S}(T)} \mathbb{P}\{w\} \mathtt{K}(x,w)$;
for $t = T, T-1, \dots, t_0$ do
    forall the $x \in \mathbb{X}$ do
        forall the $w \in \mathbb{S}(t)$ do
            forall the $u \in \mathbb{B}(t,x)$ do
                $l(t,x,u,w) = \mathtt{L}(t,x,u,w) + V\big(t+1, \mathtt{Dyn}(t,x,u,w)\big)$
            $\max_{u \in \mathbb{B}(t,x)} l(t,x,u,w)$ ;
            $\mathbb{B}^{\star}(t,x,w) = \operatorname{argmax}_{u \in \mathbb{B}(t,x)} l(t,x,u,w)$
        $V(t,x) = \sum_{w \in \mathbb{S}(t)} \mathbb{P}\{w\} \max_{u \in \mathbb{B}(t,x)} l(t,x,u,w)$
```

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

- When utilities are quadratic

$$\begin{aligned} \mathtt{K}\big(x, w\big) &= x'S(T)x + w'R(T)w \\ \mathtt{L}\big(t, x, u, w\big) &= x'S(t)x + w'R(t)w + u'Q(t)u \end{aligned}$$

- and the dynamic is linear

$$\mathtt{Dyn}(t, x, u, w) = F(t)x + G(t)u + H(t)w$$

- and primitive random variables $\big(w(t_0), w(t_0+1), \ldots, w(T-1), w(T)\big)$ 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 $u^{\star}(t)$ "on the fly" by

$$u^{\star}(t) \in \operatorname*{argmax}_{u \in \mathbb{B}(t,x)} \mathbb{E}_{w(t)} \left[\mathtt{L}\big(t,x,u,w(t)\big) + V\Big(t+1, \mathtt{Dyn}\big(t,x,u,w(t)\big)\Big) \right]$$

- 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 $w(t)$, 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 OLFO)

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
2. The stochastic optimality problem and dynamic programming
3. Applications to stochastic resources optimal management
4. The robust optimality problem and dynamic programming
5. Summary

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

Plants display a large spectrum of life-history patterns

(Mark Kot, *Elements of Mathematical Ecology*)

- Herbs often flower in their first year and then die, roots and all, after setting seed
- Plants that flower once and then die are *monocarpic*
  - Bamboos are grasses but they grow to unusually large size. One Japanese species, *Phyllostachys bambusoides*, waits 120 years to flower (Janzen, 1976)
  - Most trees flower repeatedly. However, Foster (1977) has characterized *Tachigalia versicolor* as a 'suicidal neotropical tree'. After reaching heights of 30-40 m, it flowers once and then dies

A stochastic control model of plant growth

The model is a discrete time one with time variable $t \in \{t_0, \ldots, T\}$

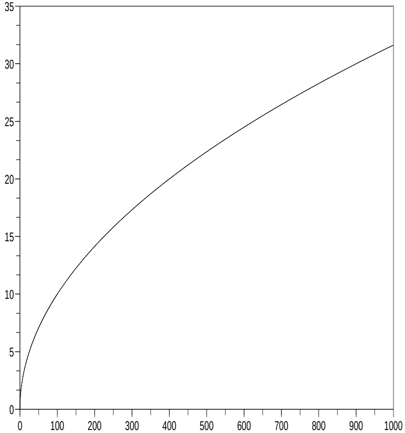
A time unit may typically be either
a day ($t \in \{0, \ldots, 364\}$),
a month ($t \in \{0, \ldots, 11\}$),
or a season ($t \in \{0, 1, 2, 3\}$)

1. At the beginning of each time interval $[t, t+1[$,
   - the plant is characterized by its vegetative biomass $k_t \in [0, +\infty[$
   - and by the cumulated reproductive biomass $S_t \in [0, +\infty[$
2. At the end of each time interval $[t, t+1[$,
   the vegetative biomass k_{t+1} is at most $f(k_t)$,

$$0 \leq k_{t+1} \leq f(k_t)$$

where f is the growth function

The growth of a plant in one period is modeled by a stricly increasing and strictly concave function



Concerning the (gross) growth function $f : \mathbb{R}_+ \to \mathbb{R}$, we make the following assumptions:

- f is continuous
- $f(0) = 0$
- $f(k) > 0$ for $k > 0$
- f is strictly increasing: $f' > 0$
- f is strictly concave: $f'' < 0$

A stochastic control model of plant growth

1. At the beginning of each time interval $[t, t+1[$,
   - the plant is characterized by its vegetative biomass $k_t \in [0, +\infty[$
   - and by the cumulated reproductive biomass $S_t \in [0, +\infty[$
2. During each time interval $[t, t+1[$, where $t+1 < T$,
   - the plant allocates biomass u_t as vegetative biomass, with $0 \leq u_t \leq f(k_t)$
   - and $f(k_t) - u_t$ as reproductive biomass in the interval $[t, t+1[$
3. At the end of each time interval $[t, t+1[$, where $t+1 < T$,
   - the cumulated reproductive biomass is $S_{t+1} = S_0 + \sum_{s=0}^{t} [f(k_s) - u_s] = S_t + [f(k_t) - u_t]$
   - the plant vegetative biomass is k_{t+1}
     - either $k_{t+1} = u_t$ with probability p (survival)
     - or $k_{t+1} = 0$ with probability $1 - p$ (death)
4. At the maximal life span T,
   - the cumulated reproductive biomass S_T is released in the form of independent offspring

Optimization problem and Bellman equation

- Which are the growth strategies $u_t = \mathtt{Pol}(t, k_t)$, or the growth patterns $(u_{t_0},k_{t_0}), \ldots, (u_{T-1},k_{T-1})$ that display the highest expected offspring $\mathbb{E}\left[S_T\right]$?
- The optimization problem is

$$\max \mathbb{E}\left[\sum_{t=t_0}^{T-1} \big(f(k_t) - u_t\big)\right]$$

- The corresponding Bellman equation is

$$\begin{aligned} V(T,k) &= \overbrace{0}^{\text{no final gain}} \\ V(t,k) &= \max_{0 \leq u \leq f(k)} \Big(\underbrace{f(k) - u}_{\text{offspring}} + pV\big(t+1,u\big) + (1-p)V\big(t+1, \underbrace{0}_{\text{death}}\big)\Big) \end{aligned}$$

Optimal growth strategies and patterns

- Since f' is decreasing — and supposing that $f'(0) = +\infty$ and that $\lim_{k\to+\infty} f'(k) = 0$ — define k_p by

$$f'(k_p) = 1/p$$

- In the case where $k_p \leq f(k_p)$, one can show that the optimal stragegy is stationary (except for the ultimate one consisting in dying)

$$\texttt{Pol}^\star(t,k) = \begin{cases} f(k) & \text{if } k \leq f^{-1}(k_p) \\ k_p & \text{if } k \geq f^{-1}(k_p) \end{cases}$$

- Draw optimal trajectories $t \mapsto k_t$ for the vegetative biomass
- What happens to k_p and to the optimal trajectories when the survival probability p decreases from 1 to 0?

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

A biomass linear model over two periods

- We consider the biomass linear model over two periods $T = 2$

$$B(t+1) = R(t)\big(\underbrace{B(t)}_{\text{biomass}} - \underbrace{h(t)}_{\text{catches}} \big) , \quad t = 0, 1$$

  where $R(0)$ and $R(1)$ are two independent random variables representing growth factors

- We aim at maximizing the expectation of the sum of the discounted successive harvesting revenues (with discount factor $\delta = \frac{1}{1+r_e}$)

$$\max \mathbb{E}_{R(0),R(1)} \Big[p(0)h(0) + \delta p(1)h(1) \Big]$$

  - where the harvests satisfy $0 \leq h(0) \leq B(0)$, $0 \leq h(1) \leq B(1)$
  - and the prices $p(0)$ and $p(1)$ are fixed

The Bellman equation (ultimate and penultimate periods)

- Since there is no final term in the criterion, we have

$$V(2,B) = 0$$

- By the Bellman equation, we have

$$\begin{aligned} V(1,B) &= \max_{0 \leq h \leq B} \mathbb{E}_{R(1)}[\delta p(1)h + V\big(2, R(1)(B-h)\big)] \\ &= \max_{0 \leq h \leq B} \mathbb{E}_{R(1)}[\delta p(1)h] \\ &= \delta p(1)B \end{aligned}$$

with a maximum achieved at

$$^{\star}(1,B) = B$$

The Bellman equation (initial period)

- By the Bellman equation, we have

$$\begin{aligned} V(0,B) &= \max_{0 \leq h \leq B} \mathbb{E}_{R(0)}[p(0)h + V\big(1, R(0)(B-h)\big)] \\ &= \max_{0 \leq h \leq B} \mathbb{E}_{R(0)}[p(0)h + \delta p(1) R(0)(B-h)] \\ &= \max_{0 \leq h \leq B} p(0)h + \delta p(1) \mathbb{E}_{R(0)}[R(0)](B-h) \end{aligned}$$

- with a maximum achieved at $h = 0$ or at $h = B$ depending on the sign of $p(0) - \delta \mathbb{E}_{R(0)}[R(0)]p(1)$
  - if $p(0) > \delta \mathbb{E}_{R(0)}[R(0)]p(1)$, then $^\star(0,B) = B$
  - if $p(0) < \delta \mathbb{E}_{R(0)}[R(0)]p(1)$, then $^\star(0,B) = 0$

A biomass linear model over $T - t_0 + 1$ periods

- The dynamic model is

$$B(t+1) = R(t)\big(B(t) - h(t)\big) \, , \quad 0 \leq h(t) \leq B(t)$$

where $R(t_0)$, ..., $R(T-1)$ are independent and identically distributed positive random variables

- We consider expected intertemporal discounted utility maximization

$$\max_{h(t_0),\ldots,h(T-1)} \mathbb{E}\left[\sum_{t=t_0}^{T-1} \delta^{t-t_0} \underbrace{\big(h(t)\big)^{\eta}}_{\text{utility}} + \delta^{T-t_0} \big(B(T)\big)^{\eta}\right]$$

with isoelastic utility

$$0 < \eta < 1$$

Bellman equation

- The dynamic programming equation starts by

$$V(T,B) = \delta^{T-t_0} B^{\eta}$$

- and, for $t = t_0, \ldots, T-1$, gives

$$V(t,B) = \max_{h \in [0,B]} \left(\delta^{t-t_0} h^{\eta} + \mathbb{E}_R \left[V\big(t+1, R(B-h)\big) \right] \right)$$

where R is a random variable
standing for the uncertain growth of the resource
and having the same distribution as any of the random variables
$R(t_0)$, ..., $R(T-1)$

Optimal policy

- The value function $V(t,B)$ is given by

$$V(t,B) = \delta^{t-t_0} b(t)^{\eta-1} B^{\eta}$$

- and the optimal policy is

$$^{\star}(t,B) = b(t)B$$

- where the optimal fraction satisfies

$$\frac{1}{b(t)} = 1 + \frac{1}{a} + \cdots + \frac{1}{a^{T-t}} \text{ with } a = (\delta \widehat{R}^{\eta})^{\frac{1}{\eta-1}}$$

where the certainty equivalent $\widehat{R}$ is defined by

$$\widehat{R} = (\mathbb{E}_R[R^{\eta}])^{1/\eta}$$

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

Inventory control dynamical model

Consider the control dynamical model

$$x(t+1) = x(t) + u(t) - w(t)$$

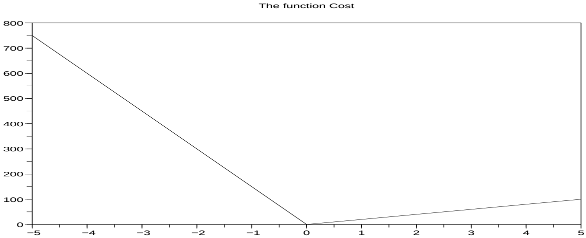
- time $t \in \{t_0, \dots, T\}$ is discrete (days, weeks or months, etc.)
- $x(t)$ is the stock at the beginning of period t, belonging to $\mathbb{X} = \mathbb{R} =]-\infty, +\infty[$
- $u(t)$ is the stock ordered at the beginning of period t, belonging to $\mathbb{U} = \mathbb{R}_+ = [0, +\infty[$
- $w(t)$ is the uncertain demand during the period t, belonging to $\mathbb{W} = \mathbb{R}_+ = [0, +\infty[$

When $x(t) < 0$, this corresponds to a *backlogged demand*, supposed to be filled immediately once inventory is again available

Inventory optimization criterion

- The costs incurred in period t are
  - purchasing costs: $cu(t)$
  - shortage costs: $b\max\{0,-\big(x(t)+u(t)-w(t)\big)\}$
  - holding costs: $h\max\{0,x(t)+u(t)-w(t)\}$
- On the period from t_0 to T, the costs sum up to

$$\sum_{t=t_0}^{T-1}\Big[\underbrace{cu(t)}_{\text{purchasing}}+\overbrace{\underbrace{b\max\{0,-\big(x(t)+u(t)-w(t)\big)\}}_{\text{shortage}}+\underbrace{h\max\{0,x(t)+u(t)-w(t)\}}_{\text{holding}}}^{\texttt{Cost}(x(t)+u(t)-w(t))}\Big]$$



Probabilistic assumptions and the inventory stochastic optimization problem

- We suppose that
  - $w(t)$, the uncertain demand, is a random variable with distribution $p_0, \dots, p_N$ on the set $\{0, \dots, N\}$
  - the sequence of demands $w(t_0)$, ..., $w(T-1)$ is independent
- We consider the inventory sochastic optimization problem

$$\min_{u(\cdot)} \mathbb{E}\Big[\sum_{t=t_0}^{T-1} [cu(t) + \texttt{Cost}\big(x(t) + u(t) - w(t)\big)] \Big]$$

The Bellman equation

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

$$
\begin{aligned}
V(T,x) &= \overbrace{0}^{\text{final cost}} \\
V(t,x) &= \min_{u\geq 0} \mathbb{E}_W \Big[\underbrace{cu + \texttt{Cost}(x+u-W)}_{\text{instantaneous cost}} + V(t+1, \underbrace{x+u-W}_{\text{future stock}}) \Big]
\end{aligned}
$$

where

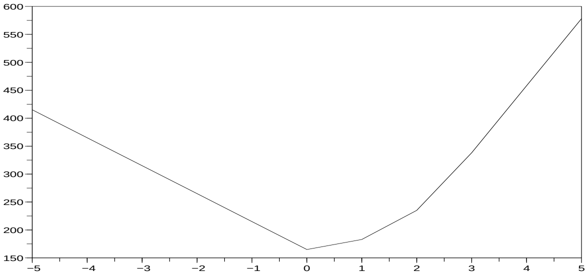
- W is a random variable with the distribution $p_0, \dots, p_N$ on the set $\{0, \dots, N\}$
- the cost function is the piecewise linear function

$$
\texttt{Cost}(x) = b \max\{0, -x\} + h \max\{0, x\}
$$

The value function is convex in the state

$$V(T-1,x) = \min_{u \geq 0} \mathbb{E}[cu + \texttt{Cost}\big(x+u-W\big)] = \min_{u \geq 0}[cu + \sum_{i=0}^{N} p_i \texttt{Cost}\big(x+u-i\big)]$$

Recalling that $b > c > 0$, show that $z \mapsto \mathbb{E}[cz + \texttt{Cost}\big(z-W\big)]$ is a convex function with a minimum $\overline{x}_{T-1}$



The optimal policy is an echelon base-stock policy

- Deduce that an optimal policy is

$$\texttt{Pol}(T-1,x) = \begin{cases} \overline{x}_{T-1} - x & \text{if} \quad x < \overline{x}_{T-1} \\ 0 & \text{if} \quad x \geq \overline{x}_{T-1} \end{cases}$$

- Show by induction that there exist thresholds (echelons) $\overline{x}_{t_0}, \ldots, \overline{x}_{T-1}$ such that the optimal policy at period t is

$$\texttt{Pol}(t,x) = \begin{cases} \overline{x}_t - x & \text{if} \quad x < \overline{x}_t \\ 0 & \text{if} \quad x \geq \overline{x}_t \end{cases}$$

- Interpret this policy

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
   - Examples of trade-offs in intertemporal stochastic optimization
   - An intertemporal criterion displays preferences with respect to time
   - How can we rank policies with a criterion under uncertainty?
   - A risk criterion displays attitudes with respect to uncertainty
2. The stochastic optimality problem and dynamic programming
   - The payoff-to-go and Bellman's Principle of Optimality
   - Bellman equation and the curse of dimensionality
   - Complements
3. Applications to stochastic resources optimal management
   - A model of plant growth over a finite horizon
   - Biomass linear models
   - The inventory problem
   - An introduction to SDDP
4. The robust optimality problem and dynamic programming
5. Summary

In the linear-convex case, value functions are convex

Here, we aim at minimizing expected cumulated costs

$$\mathbb{E}\Big[\sum_{t=t_0}^{T-1} \overbrace{\texttt{Cost}\big(t, x(t), u(t), w(t)\big)}^{\text{instantaneous cost}} + \underbrace{\texttt{CostFin}\big(x(T), w(T)\big)}_{\text{final cost}}\Big]$$

The value functions $x \mapsto V(t,x)$ are convex whenever

- $(x,u) \mapsto \texttt{Cost}\big(t,x,u,w\big)$ is jointly convex in state and control
- $x \mapsto \texttt{CostFin}\big(x,w\big)$ is convex
- $w(t), \ldots, w(T)$ are independent random variables
- the dynamic is linear

$$\texttt{Dyn}(t,x,u,w) = F(t)x + G(t)u + H(t)w$$

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

$$
\begin{aligned}
V(T,x) &= \mathbb{E}_{w(T)}\Big[\overbrace{\texttt{CostFin}\big(x, w(T)\big)}^{\text{final cost}}\Big] \\
V(t,x) &= \min_{u \in \mathbb{B}(t,x)} \mathbb{E}_{w(t)}\Big[\underbrace{\texttt{Cost}\big(t,x,u,w(t)\big)}_{\text{instantaneous cost}} \\
&\qquad + V\big(t+1, \underbrace{F(t)x + G(t)u + H(t)w(t)\big)}_{\text{future state}}\big)\Big]
\end{aligned}
$$

- It can be shown by induction that $x \mapsto V(t,x)$ is convex
- The derivative $\frac{\partial V}{\partial x}$ at $\big(t+1, x^{\star}(t+1)\big)$ 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 $a(t)$

  $0 \leq q(t) \leq \min\{S(t), q^\sharp\}$ and $0 \leq S(t) - q(t) + a(t) - r(t) \leq S^\sharp$

We aim at minimizing cumulated convex costs

On the period from t_0 to T, the costs sum up to

$$\sum_{t=t_0}^{T-1} \overbrace{\texttt{Cost}\big(t, S(t), q(t), a(t)\big)}^{\text{instantaneous cost}} + \underbrace{\texttt{CostFin}\big(T, S(T), a(T)\big)}_{\text{final cost}}$$

where

- $(S, q) \mapsto \texttt{Cost}\big(t, S, q, a\big)$ is jointly convex in state and control
- $S \mapsto \texttt{CostFin}\big(T, S, a\big)$ is convex
- $a(t), \dots, a(T)$ are independent random variables

The Bellman equation produces convex value functions

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

$$
\begin{aligned}
V(T,S) &= \mathbb{E}_{a(T)}\Big[\overbrace{\texttt{CostFin}\big(T,S,a(T)\big)}^{\text{final cost}}\Big] \\
V(t,S) &= \min_{0\leq q\leq \min\{S,q^\sharp\}} \mathbb{E}_{a(t)}\Big[\min_{r\geq 0, 0\leq S-q+a(t)-r\leq S^\sharp} \underbrace{\texttt{Cost}\big(t,S,q,a(t)\big)}_{\text{instantaneous cost}} \\
&\qquad + V\big(t+1, \underbrace{S-q-r+a(t)}_{\text{future stock}}\big)\Big]
\end{aligned}
$$

  and it can be shown by induction that $S \mapsto V(t,S)$ is convex
- The derivative $\frac{\partial V}{\partial S}$ at $\big(t+1, S^\star(t+1)\big)$ defines a hyperplane and a lower approximation of the value function, calculated by duality

Stochastic Dual Dynamic Programming (SDDP)

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

- Multiple stocks interconnected by linear dynamics

$$S_i(t+1) = S_i(t) + a_i(t) + q_{i-1}(t) - q_i(t) - r_i(t)$$

- Water inflows following an auto-regressive model

$$a_i(t) = \sum_{k=1,\ldots,K_i} \alpha_k a_i(t-k) + w(t)$$

where $w(t_0), \ldots, w(T)$ are independent random variables

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
2. The stochastic optimality problem and dynamic programming
3. Applications to stochastic resources optimal management
4. The robust optimality problem and dynamic programming
5. Summary

Maximal worst payoff

- First, we fix an admissible decision rule $\texttt{Pol}$.
- Then, we introduce the worst performance, namely the minimal payoff with respect to the scenarios $w(\cdot) \in \overline{\mathbb{S}} \subset \mathbb{S}$:

$$\texttt{Crit}^{\texttt{Pol}}_{\texttt{worst}}(t_0, x_0) := \min_{w(\cdot)\in\overline{\mathbb{S}}} \texttt{Crit}^{\texttt{Pol}}\big(t_0, x_0, w(\cdot)\big)$$

- Second, we let the decision rule $\texttt{Pol}$ vary, and aim at maximizing this worst payoff by solving the optimization problem

$$\max_{\texttt{Pol}\in\mathfrak{U}^{ad}} \texttt{Crit}^{\texttt{Pol}}_{\texttt{worst}}(t_0, x_0) = \max_{u(\cdot)} \min_{w(\cdot)\in\overline{\mathbb{S}}} \texttt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big)$$

where the last expression is abusively used, but practical and traditional, in which $x(\cdot)$ and $u(\cdot)$ need to be replaced by

$$x(t) = X_{\texttt{Dyn}}[t_0, x_0, \texttt{Pol}, w(\cdot)](t) \text{ and } u(t) = \texttt{Pol}\big(t, x(t)\big)$$

Robust additive dynamic programming equation

$$\mathtt{Crit}\big(x(\cdot), u(\cdot), w(\cdot)\big) = \sum_{t=t_0}^{T-1} \overbrace{\mathtt{L}\big(t, x(t), u(t), w(t)\big)}^{\text{instantaneous gain}} + \underbrace{\mathtt{K}\big(x(T), w(T)\big)}_{\text{final gain}}$$

Proposition

If the scenarios vary within a rectangle $\overline{\mathbb{S}} = \mathbb{S}(t_0) \times \cdots \times \mathbb{S}(T)$ *(corresponding to independence in the stochastic setting), the value functions* $V(t,x)$ *satisfy the following backward induction, where t runs from* $T-1$ *down to* t_0

$$\begin{aligned} V(T,x) &= \min_{w \in \mathbb{S}(T)} \mathtt{K}(x,w) \\ V(t,x) &= \max_{u \in \mathbb{B}(t,x)} \min_{w \in \mathbb{S}(t)} \Big[\mathtt{L}(t,x,u,w) + V\big(t+1, \mathtt{Dyn}(t,x,u,w)\big) \Big] \end{aligned}$$

Optimal robust policies

Proposition

For any time t and state x, assume the existence of the following policy

$$\mathtt{Pol}^{\star}(t,x) \in \operatorname*{argmax}_{u \in \mathbb{B}(t,x)} \min_{w \in \mathbb{S}(t)} \Big[\mathtt{L}(t,x,u,w) + V\big(t+1, \mathtt{Dyn}(t,x,u,w)\big) \Big]$$

Then $\mathtt{Pol}^{\star} \in \mathfrak{U}$ is an optimal policy of the robust problem and, for any initial state x_0, the maximal worst payoff is given by

$$V(t_0,x_0) = \max_{\mathtt{Pol} \in \mathfrak{U}^{ad}} \mathtt{Crit}^{\mathtt{Pol}}_{\mathtt{worst}}(t_0,x_0) = \mathtt{Crit}^{\mathtt{Pol}^{\star}}_{\mathtt{worst}}(t_0,x_0)$$

A biomass linear model over two periods

- The uncertain resource productivity $R(t) \in \mathbb{S}(t) = [R^\flat, R^\sharp] \subset \mathbb{W} = \mathbb{R}$, with $R^\flat < R^\sharp$
- We aim at maximizing the worst benefit namely the minimal sum of the discounted successive harvesting revenues

$$\max_{0 \leq h(0) \leq B(0),\ 0 \leq h(1) \leq B(1)} \min_{R(0), R(1)} \Big[ph(0) + \delta ph(1) \Big]$$

where the resource dynamics is

$$B(1) = R(0)\big(B(0) - h(0)\big) , \quad B(2) = R(1)\big(B(1) - h(1)\big)$$

Outline of the presentation

1. Optimization intertemporal criteria under uncertainty
2. The stochastic optimality problem and dynamic programming
3. Applications to stochastic resources optimal management
4. The robust optimality problem and dynamic programming
5. Summary

Summary

- Time-additive criteria are well adapted to dynamic programming
  in robust and stochastic optimization problems
  (but other criteria also work well)
- Bellman's Principle of Optimality breaks
  an intertemporal optimization problem
  into a sequence of interconnected static optimization problems
- In practice, the curse of dimensionality
  forbids to use dynamic programming for a state
  with dimension more than three or four