Dynamic Programming

Jean-Philippe Chancelier and SOWG
CERMICS, École des Ponts ParisTech

CIRM, Luminy
November 2019

Outline of the presentation

Outline of the presentation

The All-Pairs Shortest Path Problem

- (V, E) a given oriented graph, nodes V, edges $E \subset V \times V$.
- P : sequences $(v_1, \ldots, v_n)$, s.t $v_i \in V$, $(v_i, v_{i+1}) \in E$. given $\ell : E \to \mathbb{R} \cup \{+\infty\}$ we extend $\ell : P \to \mathbb{R} \cup \{+\infty\}$ by

$$\ell\big((v_1, \ldots, v_n)\big) := \sum_{i=1}^{n-1} \ell\big((v_i, v_{i+1})\big) \text{ (path length)}$$

- Shortest path from v_i to v_j :

$$sp(v_i, v_j) \in \arg\min \left\{ \ell(p) \,\middle|\, p = (w_1, \ldots, w_n) \in P, w_1 = v_i, w_n = v_j \right\}$$

Find $sp(v, w)$ for all $(v, w) \in V^2$ is called the all-pairs shortest path problem.

- Assumption : there does not exists cycles with negative length

$$p = (w_1, \ldots, w_n) \in P, w_1 = w_n \text{ then } \ell(p) \geq 0$$

Matrix Notations

- To simplify the notation we consider that $V = \{1, \ldots, n\}$
- $\ell : E \to \mathbb{R} \cup \{+\infty\}$ is represented by a matrix L

$$L_{i,j} = \begin{cases} 0 & \text{when } i = j, \\ +\infty & \text{when } (i,j) \notin E \\ \ell(i,j) & \text{when } (i,j) \in E \end{cases}$$

- We define $(D^{(k)})_{k \in \mathbb{N}}$, a sequences of matrices given by

$$D^{(k)}_{i,j} = \begin{cases} 0 & \text{when } i = j, \\ lsp_k(i,j) & \end{cases}$$

$$lsp_k(i,j) = \min \left\{ \ell(p) \,\middle|\, p = (i, \ldots, j) \in P, |p| \leq k+1 \right\}$$

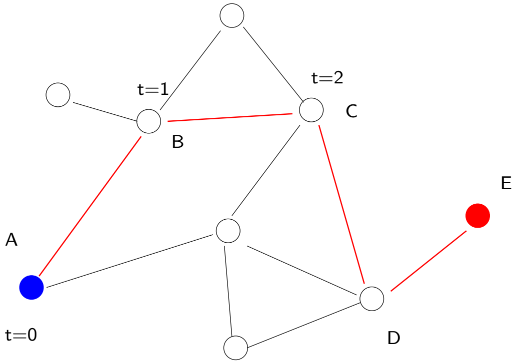
Dynamic Programming Approach

- We have that $D^{(1)}_{i,j} = L_{i,j}$
- The no-negative-cycle assumption implies that $D^{(k)}_{i,j} = \ell\big(sp(i,j)\big)$ when $k \geq n-1$ ($n = |V|$).
- We thus have to compute $D^{(1)}, \dots, D^{(n-1)}$
- We have introduced a family of problems, the original problem being one of them
- We compute value functions (value of optimal path lengths in set of constrained paths) instead of optimal paths.

A recursion formula giving the value function of problem k given the value of problem $k-1$ will be obtained through an optimality principle.

Optimality Principle

A sub-path of an optimal path is an optimal path
If the path $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$ is a shortest path form A to E
Then, the path $C \rightarrow D \rightarrow E$ is a shortest path from C to E.



Optimality Principle at Work

- Consider $D_{i,j}^{(m)}$ the value of the shortest path going from i to j with at most $m+1$ nodes.
- If the shortest path have at most m nodes then $D_{i,j}^{(m)} = D_{i,j}^{(m-1)}$
- If the shortest path have exactly $m+1$ nodes : then it is composed of a path from i to k with m nodes and an edge (k,j) whose length is $L_{k,j}$.
- Optimality principle : The path from i to k must be optimal in the set of path with at most m nodes. Thus,

$$D_{i,j}^{(m)} = D_{i,k}^{(m-1)} + L_{k,j}$$

Gathering all the cases

$$D_{i,j}^{(m)} = \min\left(D_{i,j}^{(m-1)}, \min_{k\neq j}\left(D_{i,k}^{(m-1)} + L_{k,j}\right)\right) = \min_{k}\left(D_{i,k}^{(m-1)} + L_{k,j}\right)$$

Find the optimal path using $D^{(1)}, \ldots, D^{(n-1)}$

- Compute the shortest path from i to j in $p^\sharp$:
  1. if $D^{(n-1)}_{i,j} = \infty$ then stop (there's no path from i to j)
  2. Fix $last = j$, set $q = 1$, and $p^\sharp = ()$
  3. If $n - q = 1$ then $p^\sharp = (i, p^\sharp)$ and go to end 7
  4. let $k^\sharp \in \arg\min_k \left(D^{(n-q)}_{s,k} + L_{k,last} \right)$.
  5. If $k^\sharp \neq last$ then $p^\sharp = (k^\sharp, p^\sharp)$.
  6. $q = q + 1$, $last = k^\sharp$ and iterate at 3.
  7. $p^\sharp$ gives the optimal path

If $k^\sharp \in \arg\min_k \left(D^{(p-1)}_{i,k} + L_{k,j} \right)$

$$\underbrace{i \rightsquigarrow j}_{p\text{-shortest path}} \quad = \quad \underbrace{i \rightsquigarrow k^\sharp}_{p-1 \text{ shortest path}} \quad \mapsto j$$

Recursive functions

- $Fib(n) = (n <= 1)?1 : Fib(n-1) + Fib(n-2)$.

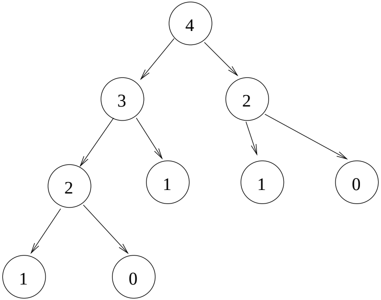


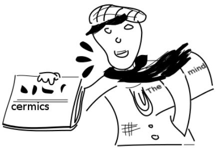
Figure – Fibonacci

Recursion

- Exponential complexity of the naive recursive algorithm.
- Solution 1 : iterative computation from $(Fib(0), Fib(1))$
- Solution 2 : use recursion but keep track of already computed values (memoization).

Outline of the presentation

The One Day Newsvendor Problem



- Each morning, The newsvendor orders a number of newspaper $u \in \mathbb{U} = \{0, 1, \ldots\}$ at unit price $c > 0$.
- The demand is incertain $w \in \mathbb{W} = \{0, 1, \ldots\}$
  - If at the end of the day, if it remains unsold newspapers, the incurred cost is

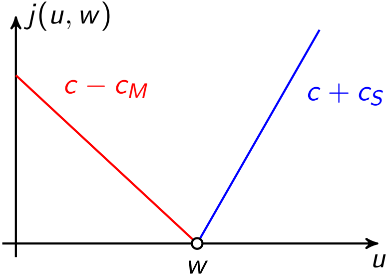
$$c_S(u - w)_+ = c_S \underbrace{\max(u - w, 0)}_{\text{unsold}} \quad \text{with } c > -c_S$$

  - If during the day facing the demand was not possible, the incurred cost for unsatisfied demand is

$$c_M(w - u)_+ = c_M \underbrace{\max(w - u, 0)}_{\text{missing}}$$

Uncertain Demand Induces Uncertain Cost

$$j(u,w) = \underbrace{cu}_{\text{purchases}} + \underbrace{c_S(u-w)_+}_{\text{unsold}} + \underbrace{c_M(w-u)_+}_{\text{missing}}$$



$$\text{Argmin}_{u \in \mathbb{U}} j(u,w) = \{w\}$$: unknown quantity w !

Attitude Toward Risk

The choice of the cost to minimize depends on the risk attitude of the newsvendor. We will use to aggregate random values through expectation

$$\min_{u \in \mathbb{U}} J(u) \quad \text{with } J(u) := \underbrace{\mathbb{E}_w\big[j(u,w)\big]}_{\text{Expectation}}$$

A pessimistic newsvendor could decide to use the worse case among all the possible values of the demand

$$\min_{u \in \mathbb{U}} J(u) \quad \text{with } J(u) := \underbrace{\max_{w \in \overline{\mathbb{W}}} j(u,w)}_{\text{Worse case}}$$

Expected cost

- The demand, $\boldsymbol{W}$, is a random variable. Te newsvendor knows the distribution $\mathbb{P}_{\boldsymbol{W}}$, that is the law of $\boldsymbol{W}$.
- The cost to be minimized is

$$J(u) = \mathbb{E}_{\boldsymbol{W}}\left[cu + c_S(u - \boldsymbol{W})_+ + c_M(\boldsymbol{W} - u)_+\right]$$

- The newsvendor problem is to find

$$u^\star \overset{?}{\in} \operatorname*{Argmin}_{u \in \mathbb{U}} J(u) \,.$$

The newsvendor takes a deterministic decision facing a future random variable. He knows the law of the random variable but not the realization (Decision-Hazard framework).

Optimal Control

The optimal control $u^\star$ when $u \in \mathbb{R}_+$:

- If $c_M \leq c$ then J is a non-decreasing function and the optimal value is not to order $u^\star = 0$
- If $c_M > c$ then the minimum of J is reached at $u^\star$:

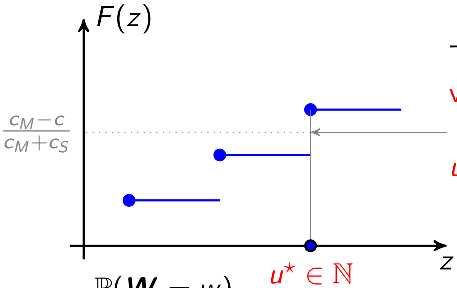
$$u^\star = \inf\{z \in \mathbb{R} \,|\, F(z) \geq (c_M - c)/(c_M + c_S)\}$$

  where F is the cumulative distribution function of $\boldsymbol{W}$, $F(z) = \mathbb{P}(\boldsymbol{W} \leq z)$.
- If $\boldsymbol{W}$ is a discrete random variable taking values in $\mathbb{N}$ then the optimal value of the relaxed problem $u^\star$ is in $\mathbb{N}$

$$\begin{aligned} J(u) &= c_M \mathbb{E}_{\boldsymbol{W}}[\boldsymbol{W}] + (c - c_M)u + (c_M + c_S)\mathbb{E}_{\boldsymbol{W}}[(u - \boldsymbol{W})_+] \\ &= c_M \mathbb{E}_{\boldsymbol{W}}[\boldsymbol{W}] + (c - c_M)u + (c_M + c_S)\int_0^u F(z)dz \end{aligned}$$

J is continuous and coercive. When $c_M > c$ it is non-increasing then non-decreasing with a minimum at $u^\star$.

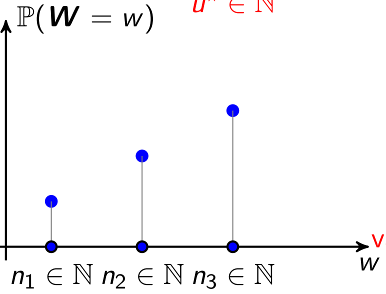
Optimal Control (II)



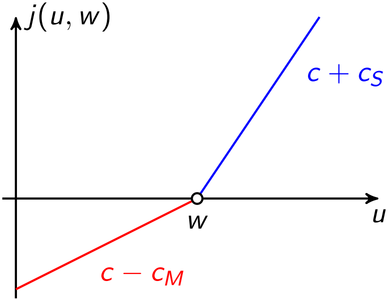
The optimal control takes integer values

$$u^\star = \inf\left\{ z \in \mathbb{R} \mid F(z) \geq \frac{c_M - c}{c_M + c_S} \right\} ,$$

If the demand takes integer values



The case $c_M \leq c$

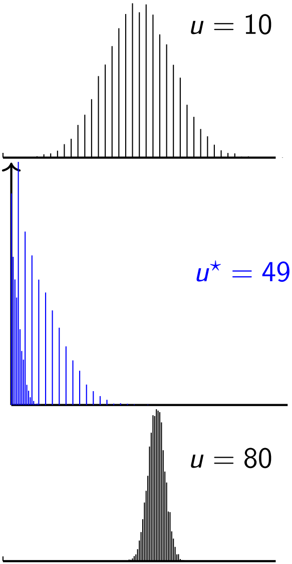


We note that

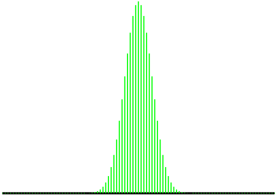
$$\min_{u \in \mathbb{U}} J(u) \geq \min_{u=\phi(\boldsymbol{W})} J(u) = \mathbb{E}_{\boldsymbol{W}}[\min_{u \in \mathbb{U}} j(u, \boldsymbol{W})]$$

$\min_{u \in \mathbb{U}} j(u, w)$ is attained by $u = 0$ for all w and gives an admissible command for the problem $\min_{u \in \mathbb{U}} j(u, w)$.

Costs Distribution $j(u, \boldsymbol{W})$



Distribution of the demand $\mathrm{bin}(100, 1/2)$.



Initial Stock and Fixed Cost for Buying Newspapers

We change the model :

- We start with an initial stock $x \in \mathbb{Z}$
- When increasing stock a fixed cost c_F occurs

$$
\begin{aligned}
\widetilde{J}(u) &= \mathbb{E}_{\boldsymbol{W}}\left[c_F \mathbb{I}_{\{u>0\}} + cu + c_s(x+u-\boldsymbol{W})_+ + c_M(\boldsymbol{W}-x-u)_+\right] \\
&= c_F \mathbb{I}_{\{u>0\}} - cx + \mathbb{E}_{\boldsymbol{W}}\left[j(u+x,\boldsymbol{W})\right] \\
&= c_F \mathbb{I}_{\{u>0\}} + J(u+x) - cx
\end{aligned}
$$

- The optimal control $u^\star(x)$ (for $c_M > c$) is a function of x
- The optimal control depends on two bounds (s, S) :

$$u^\star(x) = (S - x)\mathbb{I}_{\{x \le s\}}$$

  - If the stock is smaller than s, buy newspapers to bring the stock to S.
  - If the stock is bigger than s, do not buy newspapers
  - The value of S is given by $\mathrm{Argmin}_{u \in \mathbb{U}} J(u)$.

Initial Stock and Fixed Cost for Buying Newspapers

Let S given by $\{S\} = \mathrm{Argmin}_{u \in \mathbb{U}} J(u)$.

- If $x \geq S$, not buying is optimal since $J(\cdot) \nearrow$ and $c_F \geq 0)$:

$$\widetilde{J}(0) = J(x) - cx \leq J(x+u) - cx + c_F = \widetilde{J}(u)$$

- If $x \leq S$, since $J(\cdot)$ is minimal for S :
  - If buying, we need to order $u = S - x$ whose cost is

$$\widetilde{J}(u) = J(S) - cx + c_F$$

  - If not buying the cost is $J(x) - cx$

The solution for minimizing the costs is to fill the stock up to S if $J(x) \geq c_F + J(S)$ and do nothing otherwise. Noting that

$$\{x \,|\, J(x) \geq c_F + J(S)\} = \{x \,|\, x \leq s\}$$

where s is given by

$$s := \sup \{z \in (-\infty, S) \,|\, J(z) \geq c_F + J(S)\}$$

Initial Stock and Fixed Cost for Buying Newspapers

Let $\boldsymbol{X}_0 = x$ and $\boldsymbol{X}_1 = f(\boldsymbol{X}_0, u)$ with $f(x, u) := x + u - w$ et
$\widetilde{j}(u, x_1) := c_F \mathbb{I}_{\{u>0\}} + cu + c_S(x_1)_+ + c_M(-x_1)_+$
The newsvendor problem is

$$\min_{\boldsymbol{U} \in \mathcal{U}, \boldsymbol{X}_1, \boldsymbol{X}_0} \mathbb{E}_{\boldsymbol{W}}[\widetilde{j}(\boldsymbol{u}, \boldsymbol{X}_1)]$$

$$\boldsymbol{X}_0 = x \quad \boldsymbol{X}_1 = f(\boldsymbol{X}_0, \boldsymbol{u}, \boldsymbol{W})$$

With a non-anticipative constraint
$\mathcal{U} = \{\boldsymbol{U} : \Omega \to \mathbb{N} \mid \boldsymbol{U}(\omega) = \phi(\boldsymbol{X}_0(\omega))\}$
Note that $u^\sharp(x)$ is obtained by

$$u^\sharp(x) \in \operatorname*{Argmin}_{u \in \mathbb{N}} h(u, x) \text{ with } h(u, x) = \mathbb{E}_{\boldsymbol{W}}[\bar{j}(u, x, \boldsymbol{W})]$$

with $\bar{j}(u, x, \boldsymbol{W}) = \widetilde{j}(u, f(u, x, w))$.

Initial Stock and Fixed Cost for Buying Newspapers

- Suppose now that $\boldsymbol{X}_0$ is a given random variable
- The newsvendor problem becomes

$$\min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\left[\bar{j}\big(\boldsymbol{U}, \boldsymbol{X}_0, \boldsymbol{W}\big)\right]$$

  Where $\mathcal{U} = \{\boldsymbol{U} : \Omega \to \mathbb{N} \mid \boldsymbol{U}(\omega) = \phi\big(\boldsymbol{X}_0(\omega)\big)\}$
- Suppose that $\boldsymbol{X}_0$ and $\boldsymbol{W}$ are independent r.v
  - $h\big(\gamma(\boldsymbol{X}_0), \boldsymbol{X}_0\big) = \mathbb{E}\left[\bar{j}\big(\gamma(\boldsymbol{X}_0), \boldsymbol{X}_0, \boldsymbol{W}\big) \middle| \boldsymbol{X}_0\right]$
  - If $h\big(u^\sharp(\boldsymbol{X}_0), \boldsymbol{X}_0\big) \leq h\big(\gamma(\boldsymbol{X}_0), \boldsymbol{X}_0\big)$ for all γ then $\mathbb{E}\left[\bar{j}\big(u^\sharp(\boldsymbol{X}_0), \boldsymbol{X}_0, \boldsymbol{W}\big) \middle| \boldsymbol{X}_0\right] \leq \mathbb{E}\left[\bar{j}\big(\gamma(\boldsymbol{X}_0), \boldsymbol{X}_0, \boldsymbol{W}\big) \middle| \boldsymbol{X}_0\right]$
  - If $h\big(u^\sharp(\boldsymbol{X}_0), \boldsymbol{X}_0\big) \leq h\big(\gamma(\boldsymbol{X}_0), \boldsymbol{X}_0\big)$ for all γ then $\mathbb{E}\left[\bar{j}\big(u^\sharp(\boldsymbol{X}_0), \boldsymbol{X}_0, \boldsymbol{W}\big)\right] \leq \mathbb{E}\left[\bar{j}\big(\gamma(\boldsymbol{X}_0), \boldsymbol{X}_0, \boldsymbol{W}\big)\right]$
- The optimal control is $\boldsymbol{U}^\sharp = u^\sharp(\boldsymbol{X}_0)$
- We have

$$\min_{\boldsymbol{U} \in \mathcal{U}} \mathbb{E}\left[\bar{j}\big(\boldsymbol{U}, \boldsymbol{X}_0, \boldsymbol{W}\big)\right] = \mathbb{E}\left[\min_{u \in \mathbb{N}} \mathbb{E}\left[\bar{j}\big(u, \boldsymbol{X}_0, \boldsymbol{W}\big) \middle| \boldsymbol{X}_0\right]\right]$$

Dynamics as a Markov Chain

The two stocks $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ can be seen as two consecutive states of a controlled Markov chain.

- Assume that $u \in \mathbb{N}$ is fixed, the transition matrix is

$$P^u_{x_0,x_1} = \mathbb{P}(\boldsymbol{X}_1 = x_1 | \boldsymbol{X}_0 = x_0)$$

$$P^u_{x_0,x_1} = \begin{cases} \mathbb{P}(\boldsymbol{W} = w_0) & \text{if } x_1 = x_0 + u - w_0 \\ 0 & \text{if not} \end{cases}$$

- Assume that $\boldsymbol{U}$ is chosen as a function of X_0, $U = \phi(\boldsymbol{X}_0)$ then

$$P^\phi_{x_0,x_1} = \begin{cases} \mathbb{P}(\boldsymbol{W} = w_0) & \text{if } x_1 = x_0 + \phi(x_0) - w_0 \\ 0 & \text{if not} \end{cases}$$

Outline of the presentation

From One Stage To Finite Horizon Problem

The one stage problem with initial stock

$$\min_{u \in \mathcal{U}, \boldsymbol{X}_1, \boldsymbol{X}_0} \mathbb{E}_{\boldsymbol{W}} \big[\widetilde{J}(u, \boldsymbol{X}_1) \big]$$

$$\text{s.t. } \boldsymbol{X}_0 = x \quad \boldsymbol{X}_1 = f(\boldsymbol{X}_0, u, \boldsymbol{W}_1)$$

with

$$f(x, u, w) := x + u - w$$

$$\widetilde{J}(u, x) := c_F \mathbb{I}_{\{u>0\}} + cu + \alpha \left(c_S (x)_+ + c_M (-x)_+ \right)$$

The stock $\boldsymbol{X}_t$ can be positive (physical stock) or negative (the opposite of missing newspapers)
The law of the demand, $\boldsymbol{W}$, is known (finite expectation)

Finite Horizon problem

The newsvendor minimizes the costs over a period T

$$\min_{\boldsymbol{U} \in \mathcal{U}, \boldsymbol{X}} \mathbb{E}_{\boldsymbol{W}} \Big[\sum_{t=0}^{T-1} \alpha^t \mathfrak{J}(\boldsymbol{U}_t, \boldsymbol{X}_{t+1}) \Big]$$

$$\boldsymbol{X}_0 = x \quad \boldsymbol{X}_{t+1} = f(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1})$$

- $(\boldsymbol{X}_0, \boldsymbol{W}_1, \boldsymbol{W}_2, \dots, \boldsymbol{W}_T)$ are independent
- $\alpha \in (0, 1]$ is an actualization rate.

Canonical Form

The newsvendor minimizes the costs over a T period of time

$$\min_{\boldsymbol{U} \in \mathcal{U}, \boldsymbol{X}} \mathbb{E}_{\boldsymbol{W}} \Big[\sum_{t=0}^{T-1} \alpha^t c_t(\boldsymbol{U}_t, \boldsymbol{X}_t) + \alpha^T K(\boldsymbol{X}_T) \Big]$$
$$\boldsymbol{X}_0 = x \quad \boldsymbol{X}_{t+1} = f(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1})$$

with

$$\begin{aligned} c_t(u,x) &:= c_F \mathbb{I}_{\{u>0\}} + cu + c_S(x)_+ + c_M(-x)_+ \\ c_0(u,x) &:= c_F \mathbb{I}_{\{u>0\}} + cu \\ K(x) &:= c_S(x)_+ + c_M(-x)_+ \end{aligned}$$

Canonical Form (II)

$$
\begin{aligned}
\tilde{J}(u_0, x_1) + \alpha \tilde{J}(u_1, x_2) &= \overbrace{c_F \mathbb{I}_{\{u_0>0\}} + c u_0 + \alpha \left(c_S (x_1)_+ + c_M (-x_1)_+ \right)}^{\tilde{J}(u_0, x_1)} \\
&\quad + \alpha \overbrace{\left(c_F \mathbb{I}_{\{u_1>0\}} + c u_1 + \alpha \left(c_S (x_2)_+ + c_M (-x_2)_+ \right) \right)}^{\tilde{J}(u_1, x_2)} \\
&= \overbrace{c_F \mathbb{I}_{\{u_0>0\}} + c u_0}^{c_0(u_0, x_0)} \\
&\quad + \alpha \overbrace{\left(c_F \mathbb{I}_{\{u_1>0\}} + c u_1 + c_S (x_1)_+ + c_M (-x_1)_+ \right)}^{c_1(u_1, x_1)} \\
&\quad + \alpha^2 \overbrace{\left(c_S (x_2)_+ + c_M (-x_2)_+ \right)}^{K(x_2)} \\
&= c_0(u_0, x_0) + \alpha c_1(u_1, x_1) + \alpha^2 K(x_2)
\end{aligned}
$$

Non Anticipativity

- The newsvendor collects over time the demand of each day.
  - At time t, he knows $(\boldsymbol{W}_1, \cdots, \boldsymbol{W}_t)$ and $\boldsymbol{X}_0$ and can use this information to compute $\boldsymbol{U}_t$. He could also collect the past controls.
- Under the independence assumption of the r.v $(\boldsymbol{X}_0, \boldsymbol{W}_1, \cdots, \boldsymbol{W}_t)$ the optimal control at time t only depends of the stock $\boldsymbol{X}_t$.

Policies

$$\min_{\boldsymbol{U}\in\mathcal{U}} J(\boldsymbol{U}) \text{ with } J(\boldsymbol{U}) = \mathbb{E}_{\boldsymbol{W}}[\sum_{t=0}^{T-1} \alpha^t c_t(\boldsymbol{U}_t, \boldsymbol{X}_t) + \alpha^T K(\boldsymbol{X}_T)]$$

$$\boldsymbol{X}_0 = x \quad \boldsymbol{X}_{t+1} = f(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1})$$

1. $\min_{\boldsymbol{U}\in\mathcal{U}_H} J(\boldsymbol{U})$: $\mathcal{U}_H$ space of history dependent controls depending on past states, past controls, past noises
2. $\leq \min_{\boldsymbol{U}\in\mathcal{U}_M} J(\boldsymbol{U})$: $\mathcal{U}_M$ space of markovian controls depending on current state
3. $\leq \min_{\boldsymbol{U}\in\mathcal{U}_{OL}} J(\boldsymbol{U})$: $\mathcal{U}_{OL}$ space of open-loop controls deterministic (constant random variables)

We have that $\min_{\boldsymbol{U}\in\mathcal{U}_H} J(\boldsymbol{U}) = \min_{\boldsymbol{U}\in\mathcal{U}_M} J(\boldsymbol{U})$

State Feedback Versus Open-Loop Feedback

$$\min_{\boldsymbol{U}_1,\boldsymbol{U}_2} \mathbb{E}\Big[\boldsymbol{X}_1^2 + \boldsymbol{X}_2^2\Big]$$

s.t $\quad \boldsymbol{X}_2 = \boldsymbol{X}_1 - \boldsymbol{U}_1 + \boldsymbol{W}_2 \;,\; \boldsymbol{X}_1 = \boldsymbol{X}_0 - \boldsymbol{U}_0 + \boldsymbol{W}_1 \;,\; \boldsymbol{X}_0 = 0$

$\quad \boldsymbol{W}_0, \boldsymbol{W}_1$ i.i.d (Bernoulli with $p = 1/2$)

- State Feedback $\boldsymbol{U}_0 = \gamma_0(\boldsymbol{X}_0)$ $\boldsymbol{U}_1 = \gamma_1(\boldsymbol{X}_1)$
  - $V_2(x) = x^2$
  - $V_1(x) = \min_u \mathbb{E}\big[x^2 + V_2(x - u + W_2)\big]$
    $= \min_u x^2 + (x-u+1)^2/2 + (x-u)^2/2 = x^2 + 1/4$
  - $V_0(x) = \min_u \mathbb{E}\big[V_1(x - u + W_1)\big]$
    $= \min_u 1/4 + (x-u+1)^2/2 + (x-u)^2/2 = 1/2$
- Open Loop Controls $\boldsymbol{U}_0 = u_0$, $\boldsymbol{U}_1 = u_1$
  - $\min_{u_0,u_1} \mathbb{E}\Big[(-u_0 + \boldsymbol{W}_1)^2 + (-u_0 + \boldsymbol{W}_1 - u_1 + \boldsymbol{W}_2)^2\Big]$
    $= \min_{u_0,u_1} 2u_0^2 + 2u_0u_1 + u_1^2 - 3u_0 - 2u_1 + 2 = 3/4$

$$1/2 = \min_{\boldsymbol{U}_0,\boldsymbol{U}_1} \mathbb{E}\Big[J(\boldsymbol{U}_0,\boldsymbol{U}_1)\Big] < \min_{u_0,u_1} \mathbb{E}\Big[J(u_0,u_1)\Big] = 3/4$$

Outline of the presentation

Start with a Simplified Problem $\to$ just a Final Cost

Problem $(\mathcal{P}_0)$ starting at position x at initial time $t = 0$:

$$
\begin{aligned}
V_0(x) = \min_{\boldsymbol{X}, \boldsymbol{U} \in \mathcal{U}} \quad & \mathbb{E}\left[K\left(\boldsymbol{X}_T \right) \right], \\
\text{s.c.} \quad & \boldsymbol{X}_0 = x \ , \\
& \boldsymbol{X}_{t+1} = f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1} \right), \qquad \forall t = 0, \dots, T-1,
\end{aligned}
$$

- Noises $\boldsymbol{W} = (\boldsymbol{W}_t)_{t=1,\dots,T}$ (The demand)
- Controls $\boldsymbol{U} = (\boldsymbol{U}_t)_{t=0,\dots,T-1}$ (Newspaper to order)
- States $(\boldsymbol{X}_t)_{t=0,\dots,T-1}$ (Stock of Newspaper)

Markovian Dynamics

The noises and initial state $\boldsymbol{X}_0, \boldsymbol{W}_1, \dots, \boldsymbol{W}_T$ are independent r.v

- Transition Matrix : uncontrolled case $f : \mathbb{X} \times \mathbb{W} \to \mathbb{X}$

$$\boldsymbol{X}_{t+1} = f(\boldsymbol{X}_t, \boldsymbol{W}_{t+1}), \quad P(x, y) = \mathbb{P}\big(f(x, \boldsymbol{W}_1) = y\big)$$

- Transition Matrix : controlled case $f : \mathbb{X} \times \mathbb{U} \times \mathbb{W} \to \mathbb{X}$ with markovian policy $(\phi_s)_{s \in [0, T-1]}$, $\boldsymbol{U}_t = \phi_t(\boldsymbol{X}_t)$

$$\boldsymbol{X}_{t+1} = f(\boldsymbol{X}_t, \phi_t(\boldsymbol{X}_t), \boldsymbol{W}_{t+1}), \quad P_t(x, y) = \mathbb{P}\big(f(x, \phi_t(x), \boldsymbol{W}_1) = y\big)$$

For $u \in \mathbb{U}$, let $P^u := \mathbb{P}\big(f(x, u, \boldsymbol{W}_1) = y\big)$. For a Markovian policy $(\phi_s)_{s \in [0, T-1]}$, P_t^{ϕ} is defined by $P_t^{\phi}(x, y) := P^{\phi_t(x)}(x, y)$

$$\boldsymbol{X}_{t+1} = f(\boldsymbol{X}_t, \phi_t(\boldsymbol{X}_t), \boldsymbol{W}_{t+1}), \quad P_t^{\phi}(x, y) = P^{\phi_t(x)}(x, y)$$

A Family of problems

- Problem $(\mathcal{P}_{t_0})$ starting with stock x at time t_0 :

$$
\begin{aligned}
V_{t_0}(x) = \min_{\boldsymbol{X},\phi(\cdot)} \quad & V^{\phi}_{t_0}(x) \\
& V^{\phi}_{t_0}(x) = \mathbb{E}\left[K\left(\boldsymbol{X}_T\right)\right], \\
& \text{s.t} \quad \boldsymbol{X}_{t_0} = x, \quad \boldsymbol{X}_{t+1} = f_t\left(\boldsymbol{X}_t, \phi_t(\boldsymbol{X}_t), \boldsymbol{W}_{t+1}\right)
\end{aligned}
$$

- Problem $(\mathcal{P}'_{t_0})$ démarrant en μ à t_0 :

$$
\begin{aligned}
\mathcal{V}_{t_0}(\mu) = \min_{\mu,\phi(\cdot)} \quad & \mathcal{V}^{\phi}_{t_0}(\mu) \\
& \mathcal{V}^{\phi}_{t_0}(\mu) = \sum_x \mu_T(x) K(x) \\
& \text{with} \quad \mu_{t_0} = \mu, \quad \mu_{t+1} = \mu_t P^{\phi}_t
\end{aligned}
$$

  - Dynamic $\mu_{t+1}(y) = \sum_x \mu_t(x) P^{\phi(x)}_{x,y}$

Links between $\mathcal{V}^{\phi}_{t_0}(\cdot)$ and $V^{\phi}_{t_0}(\cdot)$

We have that

$$\mathcal{V}^{\phi}_{t_0}(\mu) = \left\langle \mu, V^{\phi}_{t_0} \right\rangle := \sum_{x} \mu(x) V^{\phi}_{t_0}(x)\,, \text{and} \quad V^{\phi}_{t_0}(x) = \mathcal{V}^{\phi}_{t_0}(\delta_x(\cdot))$$

Indeed :

- Problem $(\mathcal{P}_{t_0})$ starting with stock x at t_0 :

$$V^{\phi}_{t_0}(x) = (P^{\phi}_{t_0} \cdots P^{\phi}_{T-1} K)(x)$$

$$\text{(Ex. for } t_0 = T-1 \quad V^{\phi}_{T-1}(x) = \sum_{y \in \mathbb{X}} P^{\phi}_{T-1}(x,y) K(y))$$

- Problem $(\mathcal{P}'_{t_0})$ starting with μ at t :

$$\mathcal{V}^{\phi}_{t_0}(\mu) = \mu P^{\phi}_{t_0} \cdots P^{\phi}_{T-1} K$$

$$\text{(Ex. for } t_0 = T-1 \quad \mathcal{V}^{\phi}_{T-1}(\mu) = \sum_{x,y \in \mathbb{X}} \mu(x) P^{\phi}_{T-1}(x,y) K(y))$$

The case $T = 1$

$X_0 = x$

$$V(x) = \mathbb{E}\left[K(\boldsymbol{X}_1)\right] = \sum_y P_{x,y} K(y)$$

Assume that the law of $\boldsymbol{X}_0$ is μ

$$\mathcal{V}(\mu) = \mathbb{E}\left[K(\boldsymbol{X}_1)\right] = \sum_x \mu(x) \sum_y P_{x,y} K(y)$$

We obtain

$$\mathcal{V}(\mu) = \sum_y \mu(x) V(x) = \langle \mu, V \rangle \ ,$$

and

$$\mathcal{V}(\delta_{x'}) = \sum_x \delta_{x'}(x) \sum_y P_{x,y} K(y) = \sum_y \mu(x) P_{x',y} K(y) = V(x') \, .$$

Recursive Computation of $\mathcal{V}_t$

We have that :

$$\mathcal{V}_t(\mu) = \min_{\phi_t} \mathcal{V}_{t+1}(\mu P_t^{\phi})$$

Proof : The problem $(\mathcal{P}'_t)$ starting with μ at time t :

$$\mathcal{V}_t(\mu) = \min_{\phi(\cdot)} \quad \mu P_t^{\phi} \cdots P_{T-1}^{\phi} K$$

- At time t, P_t^{ϕ} only depends on ϕ_t.
- The raw vector $\mu P_{t_0}^{\phi}$ is non negative (its a probability law)

$$\mathcal{V}_t(\mu) = \min_{\phi_t} \left\langle \mu P_t^{\phi}, \min_{(\phi_s)_{s>t}} P_{t+1}^{\phi} \cdots P_{T-1}^{\phi} K \right\rangle$$

$$\mathcal{V}_t(\mu) = \min_{\phi_t} \left\langle \mu P_t^{\phi}, V_{t+1}(\cdot) \right\rangle = \min_{\phi_t} \mathcal{V}_{t+1}(\mu P_t^{\phi})$$

Recursive Equation for V_t with $t \in \{0, \cdots, T\}$

Bellman Equation :

$$V_t(x) = \min_{u \in \mathbb{U}} \mathbb{E}\big[V_{t+1}(f_t(x, u, \boldsymbol{W}_{t+1}))\big]$$

$$u^\sharp(x) \in \operatorname*{Argmin}_{u \in \mathbb{U}} \mathbb{E}\big[V_{t+1}(f_t(x, u, \boldsymbol{W}_{t+1}))\big]$$

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

Proof : We already have for $\mathcal{V}_t$ that

$$\mathcal{V}_t(\mu) = \min_{\phi_t} \mathcal{V}_{t+1}(\mu P_t^\phi)$$

- $V_t(x) = \mathcal{V}_t(\delta_x(\cdot))$
- $\mathcal{V}_{t+1}(\delta_x(\cdot)P_t^\phi) = \mathbb{E}\big[V_{t+1}(f_t(x, \phi_t, \boldsymbol{W}_{t+1}))\big]$

Recursive Equation for V_t with $t \in \{0, \cdots, T\}$

Bellman Equation

$$V_t(x) = \min_{u \in \mathbb{U}} \sum_y P^u_{x,y} V_{t+1}(y)$$
$$V_T(x) = K(x)$$

Optimal control

$$u^\sharp(x) \in \operatorname*{Argmin}_{u \in \mathbb{U}} \sum_y P^u_{x,y} V_{t+1}(y)$$

Proof :

$$\sum_y P^u_{x,y} V_{t+1}(y) = \sum_y \mathbb{P}\big(f(x, u, \boldsymbol{W}_{t+1}) = y\big) V_{t+1}(y)$$
$$= \mathbb{E}\big[V_{t+1}(f_t(x, u, \boldsymbol{W}_{t+1}))\big]$$

Finite Horizon with instantaneous costs

The problem $(\mathcal{P}_0)$ starting with x at initial time $t = 0$:

$$
\begin{aligned}
V_0(x) = \min_{\boldsymbol{X},\boldsymbol{U} \in \mathcal{U}} \quad & \mathbb{E}\left[\sum_{t=0}^{T-1} L_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right) + K\left(\boldsymbol{X}_T\right)\right] , \\
\text{s.t.} \quad & \boldsymbol{X}_0 = x \ , \\
& \boldsymbol{X}_{t+1} = f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right) , \qquad \forall t = 0, \dots, T-1,
\end{aligned}
$$

Dynamic Programming Equation (Bellman Equation)

$$
\begin{aligned}
V_t(x) &= \min_{u \in \mathbb{U}} \mathbb{E}\left[L_t(x, u, \boldsymbol{W}_{t+1}) + V_{t+1}(f_t(x, u, \boldsymbol{W}_{t+1}))\right] \\
V_T(x) &= K(x)
\end{aligned}
$$

Finite Horizon with instantaneous costs

The cost of problem $(\mathcal{P}_0)$ starting at x at $t = 0$ equals $\widetilde{V}(0, x)$

$$\begin{aligned}\widetilde{V}_0(z, x) = \min_{\boldsymbol{X}, \boldsymbol{U} \in \mathcal{U}} \quad & \mathbb{E}\left[Z_T + K\left(\boldsymbol{X}_T\right)\right], \\ \text{s.t.} \quad & \boldsymbol{X}_0 = x, \boldsymbol{X}_{t+1} = f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right), \\ & \boldsymbol{Z}_0 = z, \boldsymbol{Z}_{t+1} = \boldsymbol{Z}_t + L_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right),\end{aligned}$$

- $(\boldsymbol{Z}, \boldsymbol{X})$ is Markovian for feedback controls $\phi_t(z, x)$.
- The Bellman Equation is $\widetilde{V}_T(z, x) = z + K(x)$ and

$$\widetilde{V}_t(z, x) = \min_{u \in \mathbb{U}} \mathbb{E}\left[\widetilde{V}_{t+1}(z + L_t(x, u, \boldsymbol{W}_{t+1}), f_t(x, u, \boldsymbol{W}_{t+1}))\right]$$

Finite Horizon with instantaneous costs

- We recursively show that $\widetilde{V}_t(z, x) = z + V_t(x)$
- true for $t = T$ since $\widetilde{V}_T(z, x) = z + K(x)$
- at time t

$$\widetilde{V}_t(z, x) = \min_{u \in \mathbb{U}} \mathbb{E}\big[z + L_t(x, u, \boldsymbol{W}_{t+1}) + V_{t+1}(f_t(x, u, \boldsymbol{W}_{t+1}))\big]$$

$$\widetilde{V}_t(z, x) = z + \underbrace{\min_{u \in \mathbb{U}} \mathbb{E}\big[L_t(x, u, \boldsymbol{W}_{t+1}) + V_{t+1}(f_t(x, u, \boldsymbol{W}_{t+1}))\big]}_{V_t(x)}$$

- the minimization for $u \in \mathbb{U}$ only depends on x. Thus, the optimal control is a feedback on the state x.
- we note that $\widetilde{V}_t(0, x) = V_t(x)$, giving the Bellman Equation for the problem with instantaneous cost.

Final Result : The Bellman Equation

$$
\begin{aligned}
V_0(x) = \min_{\boldsymbol{X},\boldsymbol{U}\in\mathcal{U}} \quad & \mathbb{E}\left[\sum_{t=0}^{T-1} L_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right) + K\left(\boldsymbol{X}_T\right)\right], \\
\text{s.t.} \quad & \boldsymbol{X}_0 = x\;, \\
& \boldsymbol{X}_{t+1} = f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right), \qquad \forall t = 0, \dots, T-1,
\end{aligned}
$$

Dynamic Programming Equation (Bellman Equation)

$$
\begin{aligned}
V_t(x) &= \min_{u\in\mathbb{U}} \mathbb{E}\Big[L_t(x, u, \boldsymbol{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \boldsymbol{W}_{t+1})\big)\Big] \\
V_T(x) &= K(x) \\
u^\star(x) &\in \operatorname*{Argmin}_{u\in\mathbb{U}} \mathbb{E}\Big[L_t(x, u, \boldsymbol{W}_{t+1}) + V_{t+1}\big(f_t(x, u, \boldsymbol{W}_{t+1})\big)\Big]
\end{aligned}
$$

Computing value functions

1. Time loop backward to compute V_t for all t
2. State loop for all x
3. Find the optimal control (control loop, L.P., Q.P)
4. Loop on random values to compute the expected cost

State Feedback Versus Open-Loop Feedback (II)

$$\min_{\boldsymbol{U}_1,\boldsymbol{U}_2} \mathbb{E}\Big[\boldsymbol{X}_1^2 + \boldsymbol{X}_2^2\Big]$$

$$\text{s.t} \qquad \boldsymbol{X}_2 = \boldsymbol{X}_1 - \boldsymbol{U}_1 + \boldsymbol{W}_2 \ , \ \boldsymbol{X}_1 = \boldsymbol{X}_0 - \boldsymbol{U}_0 + \boldsymbol{W}_1 \ , \ \boldsymbol{X}_0 = 0$$

$$\boldsymbol{W}_0, \boldsymbol{W}_1 \text{ i.i.d (Bernoulli with } p = 1/2)$$

- State Feedback $\boldsymbol{U}_0 = \gamma_0(\boldsymbol{X}_0)$ $\boldsymbol{U}_1 = \gamma_1(\boldsymbol{X}_1)$
  - $V_2(x) = x^2$
  - $V_1(x) = \min_u \mathbb{E}\big[x^2 + V_2(x - u + W_2)\big]$
    $= \min_u x^2 + (x - u + 1)^2/2 + (x - u)^2/2 = x^2 + 1/4$
  - $V_0(x) = \min_u \mathbb{E}\big[V_1(x - u + W_1)\big]$
    $= \min_u 1/4 + (x - u + 1)^2/2 + (x - u)^2/2 = 1/2$
- Open Loop Controls $\boldsymbol{U}_0 = u_0$, $\boldsymbol{U}_1 = u_1$
  - $\min_{u_0,u_1} \mathbb{E}\Big[(-u_0 + \boldsymbol{W}_1)^2 + (-u_0 + \boldsymbol{W}_1 - u_1 + \boldsymbol{W}_2)^2\Big]$
    $= \min_{u_0,u_1} 2u_0^2 + 2u_0u_1 + u_1^2 - 3u_0 - 2u_1 + 2 = 3/4$

$$1/2 = \min_{\boldsymbol{U}_0,\boldsymbol{U}_1} \mathbb{E}\Big[J(\boldsymbol{U}_0, \boldsymbol{U}_1)\Big] < \min_{u_0,u_1} \mathbb{E}\Big[J(u_0, u_1)\Big] = 3/4$$

A Farmer problem

- when annual production is x units of a certain crop
  - he stores $(1-u)x$ units,
  - he uses the remaining ux units for next year production,

  where $u \in (0,1)$
- then, the level of next year production will be $\boldsymbol{W}ux$, where $\boldsymbol{W}$ is a positive random variable not depending on x or u with known expectation $\mathbb{E}[\boldsymbol{W}] = \overline{w}$.
- Optimization problem : find the optimal investment policy that maximizes the total expected product stored over N years

$$\mathbb{E}\Big[\sum_{k=0}^{N-1}(1-\boldsymbol{U}_k)\boldsymbol{X}_k + \boldsymbol{X}_N\Big] ,$$

assuming that $\boldsymbol{X}_{k+1} = \boldsymbol{W}_{k+1}\boldsymbol{U}_k\boldsymbol{X}_k$.

Bellman Equation

$$V_n(x) = \max_{\boldsymbol{U}_n, \boldsymbol{U}_{n+1}, \dots, \boldsymbol{U}_{N-1}} \mathbb{E}\Big[\sum_{k=n}^{N-1}(1-\boldsymbol{U}_k)\boldsymbol{X}_k + \boldsymbol{X}_N\Big]$$

$$\boldsymbol{X}_{k+1} = \boldsymbol{W}_{k+1}\boldsymbol{U}_k\boldsymbol{X}_k \text{ and } \boldsymbol{X}_n = x \ .$$

We obtain that $V_N(x) = x$ and

$$V_n(x) = \max_{u \in [0,1]} (1-u)x + \mathbb{E}\Big[V_{n+1}\big(\boldsymbol{W} ux\big)\Big]$$

Assume that $V_{n+1}(x) = a_{n+1}x$ then we have that

$$V_n(x) = \begin{cases} x & \text{when } a_{n+1}\overline{w} \le 1 \\ a_{n+1}\overline{w}x & \text{when } a_{n+1}\overline{w} \ge 1 \end{cases}$$

That is $V_n(x) = a_n x$ with $a_n = \max\big(1, a_{n+1}\overline{w}\big)$ (with $a_N = 1$).

Increasing the State Space

How to solve

$$
\begin{aligned}
V_0(x) = \min_{\boldsymbol{X},\boldsymbol{U}\in\mathcal{U}} \quad & \mathbb{E}\Big[K\big(\max_{s\in\{0,\dots,T\}} \boldsymbol{X}_s\big)\Big], \\
\text{s.t.} \quad & \boldsymbol{X}_0 = x \ , \\
& \boldsymbol{X}_{t+1} = f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right), \qquad \forall t = 0, \dots, T-1,
\end{aligned}
$$

The noises and initial state $\boldsymbol{X}_0, \boldsymbol{W}_1, \dots, \boldsymbol{W}_T$ are independent r.v

- $\boldsymbol{Y}_t = \max_{s\in\{0,\dots,t\}} \boldsymbol{X}_s$, is not a Markov chain.
- $\left(\boldsymbol{X}_t, \boldsymbol{Y}_t\right)$ is a Markov chain. $\left(\boldsymbol{X}_0, \boldsymbol{Y}_0\right) = (x, x)$ and

$$
\begin{aligned}
\boldsymbol{X}_{t+1} &= f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right) \\
\boldsymbol{Y}_{t+1} &= \max_{s\in\{0,\dots,t+1\}} \boldsymbol{X}_s = \max\left(\boldsymbol{Y}_t, \boldsymbol{X}_{t+1}\right) \\
&= \max\left(\boldsymbol{Y}_t, f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1})\right)
\end{aligned}
$$

Increasing the State Space

How to solve

$$
\begin{aligned}
V_0(x) = \min_{\boldsymbol{X},\boldsymbol{U} \in \mathcal{U}} \quad & \mathbb{E}\Big[K\big(\boldsymbol{X}_T\big)\Big], \\
\text{s.t.} \quad & \boldsymbol{X}_0 = x \text{ fixed}, \\
& \boldsymbol{X}_{t+1} = f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right), \qquad \forall t = 0, \dots, T-1,
\end{aligned}
$$

The noises and initial state $\boldsymbol{X}_0, \boldsymbol{W}_1, \dots, \boldsymbol{W}_T$ are not independent

- $\boldsymbol{W}_{t+1} = g_t\big(\boldsymbol{W}_t, \overline{\boldsymbol{W}}_{t+1}\big)$
- $\boldsymbol{X}_0, \overline{\boldsymbol{W}}_1, \dots, \overline{\boldsymbol{W}}_T$ are independent
- $\big(\boldsymbol{X}_t, \boldsymbol{W}_t\big)$ is a Markov chain

$$
\begin{aligned}
\boldsymbol{X}_{t+1} &= f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_{t+1}\right) = f_t\Big(\boldsymbol{X}_t, \boldsymbol{U}_t, g_t\big(\boldsymbol{W}_t, \overline{\boldsymbol{W}}_{t+1}\big)\Big) \\
\boldsymbol{W}_{t+1} &= g_t\big(\boldsymbol{W}_t, \overline{\boldsymbol{W}}_{t+1}\big)
\end{aligned}
$$