Contribution to Decomposition Methods in Stochastic Optimization

V. Leclère

PhD Defense
2014, June 25

École des Ponts ParisTech

UNIVERSITÉ PARIS-EST

Mulstistage Stochastic Optimization: an Example

How to manage a chain of dam producing electricity from the turbine water to optimize the gain?

$$\mathbb{E}\left[\sum_{i=1}^{N}\sum_{t=0}^{T-1} L_t^i\big(\underbrace{\mathbf{X}_t^i}_{\text{state}}, \underbrace{\mathbf{U}_t^i}_{\text{control}}, \underbrace{\mathbf{W}_{t+1}}_{\text{noise}}\big)\right]$$

Constraints:

- dynamics:
  $\mathbf{X}_{t+1} = f_t\big(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\big)$,
- nonanticipativity:
  $\mathbf{U}_t \preceq \mathcal{F}_t$,
- spatial coupling:
  $\mathbf{Z}_t^{i+1} = g_t^i\big(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}^i\big)$.

Figures/three_dams_VL.png

Mulstistage Stochastic Optimization: an Example

How to manage a chain of dam producing electricity from the turbine water to optimize the gain?

$$\mathbb{E}\left[\sum_{i=1}^{N}\sum_{t=0}^{T-1} L_t^i(\underbrace{\mathbf{X}_t^i}_{\text{state}}, \underbrace{\mathbf{U}_t^i}_{\text{control}}, \underbrace{\mathbf{W}_{t+1}}_{\text{noise}})\right]$$

Constraints:

- dynamics:
  $\mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1})$,
- nonanticipativity:
  $\mathbf{U}_t \preceq \mathcal{F}_t$,
- spatial coupling:
  $\mathbf{Z}_t^{i+1} = g_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}^i)$.

Figures/three_dams_VL.png

Mulststage Stochastic Optimization: an Example

How to manage a chain of dam producing electricity from the turbine water to optimize the gain?

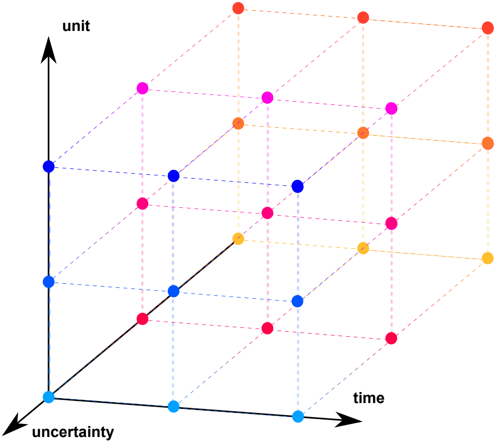
$$\mathbb{E}\left[\sum_{i=1}^{N}\sum_{t=0}^{T-1} L_t^i(\underbrace{\mathbf{X}_t^i}_{\text{state}}, \underbrace{\mathbf{U}_t^i}_{\text{control}}, \underbrace{\mathbf{W}_{t+1}}_{\text{noise}})\right]$$

Constraints:

- dynamics:
  $\mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1})$,
- nonanticipativity:
  $\mathbf{U}_t \preceq \mathcal{F}_t$,
- spatial coupling:
  $\mathbf{Z}_t^{i+1} = g_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}^i)$.

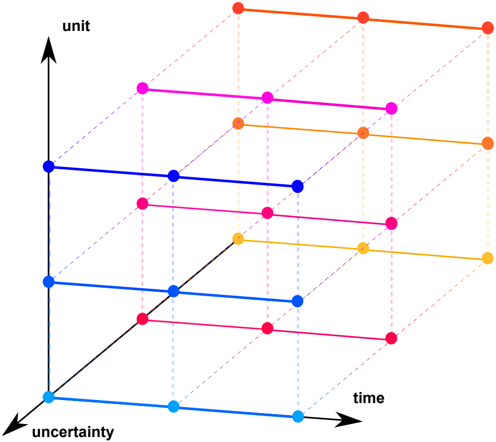
Figures/three_dams_VL.png

Couplings for Stochastic Problems



$$\min \sum_{\omega} \sum_{i} \sum_{t} \pi_\omega L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

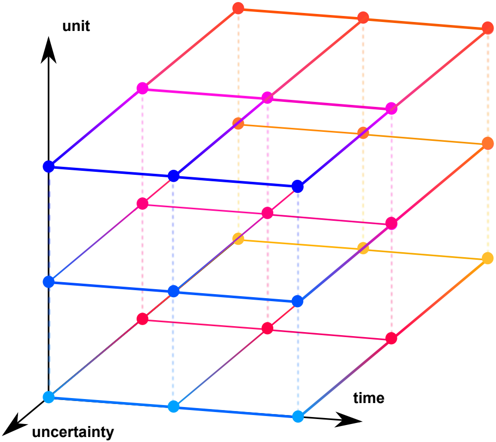
Couplings for Stochastic Problems: in Time



$$\min \sum_{\omega} \sum_{i} \sum_{t} \pi_\omega L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\text{s.t.} \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

Couplings for Stochastic Problems: in Uncertainty

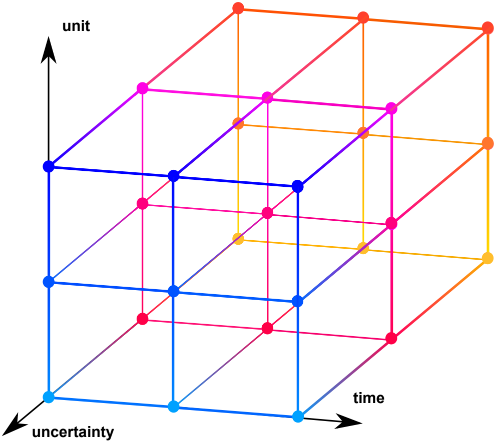


$$\min \sum_{\omega} \sum_{i} \sum_{t} \pi_\omega L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\text{s.t.} \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\mathbf{U}_t^i \preceq \mathcal{F}_t = \sigma\big(\mathbf{W}_1, \dots, \mathbf{W}_t\big)$$

Couplings for Stochastic Problems: in Space



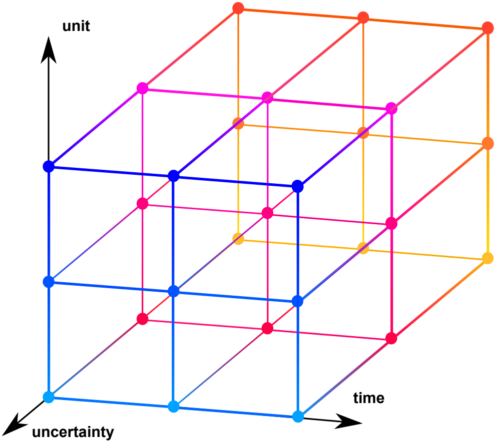
$$\min \sum_{\omega} \sum_{i} \sum_{t} \pi_\omega L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\text{s.t.} \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\mathbf{U}_t^i \preceq \mathcal{F}_t = \sigma\big(\mathbf{W}_1, \dots, \mathbf{W}_t\big)$$

$$\sum_i \Theta_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i) = 0$$

Couplings for Stochastic Problems: a Complex Problem



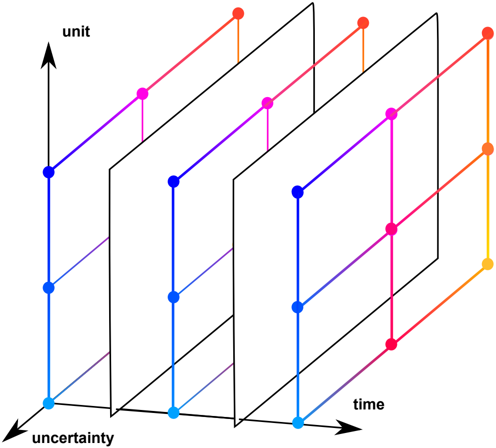
$$\min \sum_{\omega} \sum_{i} \sum_{t} \pi_\omega L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\text{s.t.} \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\mathbf{U}_t^i \preceq \mathcal{F}_t = \sigma\big(\mathbf{W}_1, \dots, \mathbf{W}_t\big)$$

$$\sum_i \Theta_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i) = 0$$

Decompositions for Stochastic Problems: in Time



$$\min \sum_{\omega} \sum_{i} \sum_{t} \pi_\omega L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\text{s.t. } \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

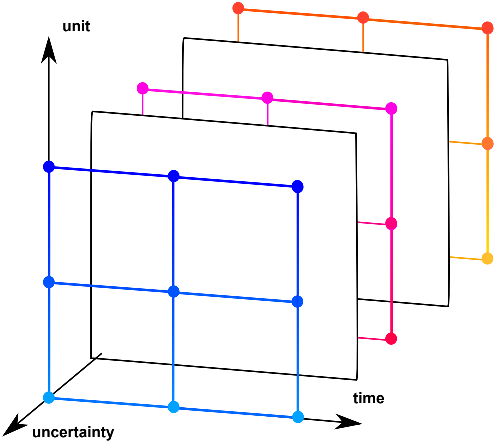
$$\mathbf{U}_t^i \preceq \mathcal{F}_t = \sigma\big(\mathbf{W}_1, \dots, \mathbf{W}_t\big)$$

$$\sum_i \Theta_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i) = 0$$

Dynamic Programming

Bellman (56)

Decompositions for Stochastic Problems: in Uncertainty



$$\min \sum_{\omega} \sum_{i} \sum_{t} \pi_\omega L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\text{s.t. } \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

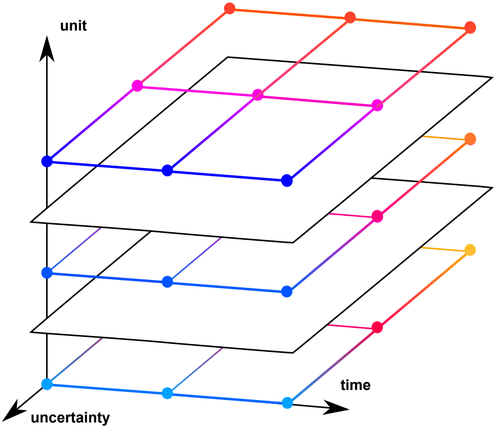
$$\mathbf{U}_t^i \preceq \mathcal{F}_t = \sigma\big(\mathbf{W}_1, \dots, \mathbf{W}_t\big)$$

$$\sum_i \Theta_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i) = 0$$

Progressive Hedging

Rockafellar - Wets (91)

Decompositions for Stochastic Problems: in Space



$$\min \sum_{\omega} \sum_{i} \sum_{t} \pi_\omega L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

s.t. $$\mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1})$$

$$\mathbf{U}_t^i \preceq \mathcal{F}_t = \sigma\big(\mathbf{W}_1, \dots, \mathbf{W}_t\big)$$

$$\sum_i \Theta_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i) = 0$$

Dual Approximate
Dynamic Programming

Thesis Outline

1. Prelimaries

2. Time-Consistency: from Optimization to Risk Measures
3. Stochastic Dual Dynamic Programming Algorithm

4. Constraint Qualification in Stochastic Optimization
5. Constraint Qualification in $\left(\mathrm{L}^\infty, \mathrm{L}^1\right)$
6. Uzawa Algorithm in L^∞

7. Epiconvergence of Relaxed Stochastic Problems
8. Dual Approximate Dynamic Programming Algorithm

Thesis Outline

Thesis Outline

1. Prelimaries
2. Time-Consistency: from Optimization to Risk Measures
3. Stochastic Dual Dynamic Programming Algorithm
4. Constraint Qualification in Stochastic Optimization
5. Constraint Qualification in $\left(\mathrm{L}^\infty, \mathrm{L}^1\right)$
6. Uzawa Algorithm in L^∞
7. Epiconvergence of Relaxed Stochastic Problems
8. Dual Approximate Dynamic Programming Algorithm

Presentation Outline

Presentation Outline

Classical Discrete Time Stochastic Optimization Problem

$$\min_{\mathbf{U}} \quad \mathbb{E}\Big[\overbrace{L_0(\mathbf{X}_0,\mathbf{U}_0,\mathbf{W}_1)}^{\text{instantaneous cost}} + \cdots + L_{T-1}(\mathbf{X}_{T-1},\mathbf{U}_{T-1},\mathbf{W}_T) + \overbrace{K(\mathbf{X}_T)}^{\text{final cost}}\Big]$$

$$\text{s.t.} \quad \mathbf{X}_0 = x_0$$

$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t,\mathbf{U}_t,\mathbf{W}_{t+1}) \qquad \text{(dynamic)}$$

$$\mathbf{U}_t \preceq \sigma\big(\mathbf{W}_1,\dots,\mathbf{W}_t\big) \qquad \text{(non-anticipativity)}$$

- $\mathbf{X}_t$: state (r.v. with value in $\mathbb{X}_t$),
- $\mathbf{U}_t$: control (r.v. with value in $\mathbb{U}_t$),
- $\mathbf{W}_t$: uncertainty (r.v. with value in $\mathbb{W}_t$)

$\rightsquigarrow$ time independence assumption!

Classical Discrete Time Stochastic Optimization Problem

$$
\begin{aligned}
\min_{\pi} \quad & \mathbb{E}\Big[\overbrace{L_0(\mathbf{X}_0, \mathbf{U}_0, \mathbf{W}_1)}^{\text{instantaneous cost}} + \cdots + L_{T-1}(\mathbf{X}_{T-1}, \mathbf{U}_{T-1}, \mathbf{W}_T) + \overbrace{K(\mathbf{X}_T)}^{\text{final cost}} \Big] \\
\text{s.t.} \quad & \mathbf{X}_0 = x_0 \\
& \mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) && \text{(dynamic)} \\
& \mathbf{U}_t = \pi_t(\mathbf{X}_t) && \text{(non-anticipativity)}
\end{aligned}
$$

- $\mathbf{X}_t$: state (r.v. with value in $\mathbb{X}_t$),
- $\mathbf{U}_t$: control (r.v. with value in $\mathbb{U}_t$),
- $\mathbf{W}_t$: uncertainty (r.v. with value in $\mathbb{W}_t$)

$\rightsquigarrow$ time independence assumption!

Risk Measure Formulation

$$\min_{\pi} \quad \varrho_{0,T}\Big\{ \overbrace{L_0(\mathbf{X}_0, \mathbf{U}_0, \mathbf{W}_1)}^{\text{instantaneous cost}}, \cdots, L_{T-1}(\mathbf{X}_{T-1}, \mathbf{U}_{T-1}, \mathbf{W}_T), \overbrace{K(\mathbf{X}_T)}^{\text{final cost}} \Big\}$$

$$\text{s.t.} \quad \mathbf{X}_0 = x_0$$

$$\mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) \qquad \text{(dynamic)}$$

$$\mathbf{U}_t = \pi_t(\mathbf{X}_t) \qquad \text{(non-anticipativity)}$$

- $\mathbf{X}_t$: state (r.v. with value in $\mathbb{X}_t$),
- $\mathbf{U}_t$: control (r.v. with value in $\mathbb{U}_t$),
- $\mathbf{W}_t$: uncertainty (r.v. with value in $\mathbb{W}_t$)

Example of Conditional Risk Measures

- $\varrho_{0,T}\left\{\mathbf{C}_0,\dots,\mathbf{C}_T\right\} = \mathbb{E}\Big[\sum_{t=0}^{T}\mathbf{C}_t\Big]$ (Classical framework)
- $\varrho_{0,T}\left\{\mathbf{C}_0,\dots,\mathbf{C}_T\right\} = \mathbb{E}\Big[\sum_{t=0}^{T} r^t\mathbf{C}_t\Big]$
- $\varrho_{0,T}\left\{\mathbf{C}_0,\dots,\mathbf{C}_T\right\} = \sup_{\mathbb{P}\in\mathcal{P}} \left\{\mathbb{E}_{\mathbb{P}}\Big[\sum_{t=0}^{T}\mathbf{C}_t\Big]\right\}$
- $\varrho_{0,T}\left\{\mathbf{C}_0,\dots,\mathbf{C}_T\right\} = \sup_{\mathbb{P}\in\mathcal{P}} \left\{\mathbb{E}_{\mathbb{P}}\Big[\prod_{t=0}^{T}\mathbf{C}_t\Big]\right\}$

Example of Conditional Risk Measures

- $\varrho_{0,T}\big\{\mathbf{C}_0, \dots, \mathbf{C}_T\big\} = \mathbb{E}\Big[\sum_{t=0}^{T} \mathbf{C}_t\Big]$ (Classical framework)
- $\varrho_{0,T}\big\{\mathbf{C}_0, \dots, \mathbf{C}_T\big\} = \mathbb{E}\Big[\sum_{t=0}^{T} r^t \mathbf{C}_t\Big]$
- $\varrho_{0,T}\big\{\mathbf{C}_0, \dots, \mathbf{C}_T\big\} = \sup_{\mathbb{P} \in \mathcal{P}} \Big\{\mathbb{E}_{\mathbb{P}}\Big[\sum_{t=0}^{T} \mathbf{C}_t\Big]\Big\}$
- $\varrho_{0,T}\big\{\mathbf{C}_0, \dots, \mathbf{C}_T\big\} = \sup_{\mathbb{P} \in \mathcal{P}} \Big\{\mathbb{E}_{\mathbb{P}}\Big[\prod_{t=0}^{T} \mathbf{C}_t\Big]\Big\}$

Dynamic Programming: Classical Framework

The sequence of Bellman functions $(V_t)_{t \in [\![0,T]\!]}$ defined by

$$V_t(x) = \min_{\pi \in \Pi} \quad \mathbb{E}\Big[\sum_{\tau=t}^{T-1} L_\tau(\mathbf{X}_\tau, \pi_\tau(\mathbf{X}_\tau), \mathbf{W}_{\tau+1}) + K(\mathbf{X}_T)\Big]$$

$$\text{s.t.} \quad \mathbf{X}_t = x$$

$$\mathbf{X}_{\tau+1} = f_\tau(\mathbf{X}_\tau, \pi_\tau(\mathbf{X}_\tau), \mathbf{W}_{\tau+1})$$

satisfies the Bellman equation $\rightsquigarrow$ Time Decomposition!

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

Question: what about other risk measures?

Dynamic Programming: Classical Framework

The sequence of Bellman functions $(V_t)_{t\in[\![0,T]\!]}$ defined by

$$
\begin{aligned}
V_t(x) = \min_{\pi\in\Pi} \quad & \mathbb{E}\Big[\sum_{\tau=t}^{T-1} L_\tau(\mathbf{X}_\tau, \pi_\tau(\mathbf{X}_\tau), \mathbf{W}_{\tau+1}) + K(\mathbf{X}_T)\Big] \\
\text{s.t.} \quad & \mathbf{X}_t = x \\
& \mathbf{X}_{\tau+1} = f_\tau(\mathbf{X}_\tau, \pi_\tau(\mathbf{X}_\tau), \mathbf{W}_{\tau+1})
\end{aligned}
$$

satisfies the Bellman equation $\rightsquigarrow$ Time Decomposition!

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

Question: what about other risk measures?

Uncertainty Aggregators

- Global uncertainty aggregator $\mathbb{G} : \mathcal{L}\big(\mathbb{W}_1 \times ... \times \mathbb{W}_T; \mathbb{R}\big) \to \mathbb{R}$
  - $\mathbb{G}\big[f\big] = \mathbb{E}\Big[f\big(\mathbf{W}_1, \dots, \mathbf{W}_T\big)\Big]$
  - $\mathbb{G}\big[f\big] = \max_{w \in \mathbb{W}_1 \times \dots \times \mathbb{W}_T} f\big(w_1, \dots, w_T\big)$
- Time-step uncertainty aggregator $\mathbb{G}_t : \mathcal{L}\big(\mathbb{W}_t; \mathbb{R}\big) \to \mathbb{R}$
  - $\mathbb{G}_t\big[f_t\big] = \mathbb{E}\Big[f_t\big(\mathbf{W}_t\big)\Big]$
  - $\mathbb{G}_t\big[f_t\big] = \max_{w_t \in \mathbb{W}_t} f_t\big(w_t\big)$
- Composition of aggregators: $\mathbb{G}_t\Big[w_t \mapsto \mathbb{G}_{t+1}\big[f(w_t, w_{t+1})\big]\Big]$

$$\max_{w \in \mathbb{W}_1 \times \dots \times \mathbb{W}_T} f\big(w_1, \dots, w_T\big)$$

$$= \max_{w_1}\Big[\max_{w_2}\Big[\cdots \max_{w_T}\big[f\big(w_1, \dots, w_T\big)\big]\Big]\Big]$$

Time Aggregators

- Global time aggregator $\Phi : \mathbb{R}^{T+1} \to \mathbb{R}$
  - $\Phi\big\{c_0, \dots, c_T\big\} = \sum_{t=0}^{T} c_t$
  - $\Phi\big\{c_0, \dots, c_T\big\} = \prod_{t=0}^{T} c_t$
- Time-step time aggregator $\Phi_t : \mathbb{R}^2 \to \mathbb{R}$
  - $\Phi_t\big\{c_1, c_2\big\} = c_1 + c_2$
  - $\Phi_t\big\{c_1, c_2\big\} = c_1 \times c_2$
- Composition of aggregators $\Phi_t\Big\{c_t, \Phi_{t+1}\big\{c_{t+1}, c_{t+2}\big\}\Big\}$

$$\sum_{t=0}^{T} c_t = c_0 + \Big\{c_1 + \Big\{\dots + \big\{c_{T-1} + c_T\big\}\Big\}\Big\}$$

Constructing Optimization Problems

Time - then - Uncertainty (TU)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\cdots,\mathbf{C}_T\Big) = \mathbb{G}\Big[\Phi\Big\{\mathbf{C}_0,\cdots,\mathbf{C}_T\Big\}\Big]$$

Uncertainty - then - Time (UT)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\cdots,\mathbf{C}_T\Big) = \Phi\Big\{\mathbb{G}_0\big[\mathbf{C}_0\big],\cdots,\mathbb{G}_T\big[\mathbf{C}_T\big]\Big\}$$

(TU) examples:

- $\mathbb{E}\Big[\sum_{t=0}^{T}\mathbf{C}_t\Big]$
- $\mathbb{E}\Big[\sum_{t=0}^{T} r^t\mathbf{C}_t\Big]$
- $\max_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\Big[\sum_{t=0}^{T} r^t\mathbf{C}_t\Big]$

(UT) examples:

- $\sum_{t=0}^{T}\mathbb{E}\big[\mathbf{C}_t\big]$
- $\sum_{t=0}^{T} r^t\mathbb{E}\Big[\mathbf{C}_t\Big]$
- $\sum_{t=0}^{T} r^t\max_{\mathbb{P}_t\in\mathcal{P}_t}\mathbb{E}_{\mathbb{P}_t}\Big[\mathbf{C}_t\Big]$

Constructing Optimization Problems

Nested - Time - then - Uncertainty (NTU)

$$\varrho_{T,T}\Big(\mathbf{C}_T\Big) = \mathbb{G}_0\Big[\Phi_0\Big\{\mathbf{C}_0, \mathbb{G}_1\Big[\Phi_1\Big\{\cdots \mathbb{G}_{T-1}\Big[\Phi_{T-1}\big\{\mathbf{C}_{T-1}, \mathbb{G}_T\big[\mathbf{C}_T\big]\big\}\Big]\cdots\Big\}\Big]\Big\}\Big]$$

Nested - Uncertainty - then - Time (NUT)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\dots,\mathbf{C}_T\Big) = \Phi_0\Big\{\mathbb{G}_0\big[\mathbf{C}_0\big], \Phi_1\Big\{\mathbb{G}_1\big[\mathbf{C}_1\big], \mathbb{G}_1\Big[\cdots \Phi_{T-1}\Big\{\mathbb{G}_{T-1}\big[\mathbf{C}_{T-1}\big], \mathbb{G}_{T-1}\Big[\mathbb{G}_T\big[\mathbf{C}_T\big]\Big]\Big\}\cdots\Big]\Big\}\Big\}$$

Constructing Optimization Problems

Nested - Time - then - Uncertainty (NTU)

$$\varrho_{T-1,T}\Big(\mathbf{C}_{T-1},\mathbf{C}_T\Big) = \mathbb{G}_0\Big[\Phi_0\Big\{\mathbf{C}_0,\mathbb{G}_1\Big[\Phi_1\Big\{\cdots \mathbb{G}_{T-1}\Big[\Phi_{T-1}\big\{\mathbf{C}_{T-1},\mathbb{G}_T\big[\mathbf{C}_T\big]\big\}\Big]\cdots\Big\}\Big]\Big\}\Big]$$

Nested - Uncertainty - then - Time (NUT)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\dots,\mathbf{C}_T\Big) = \Phi_0\Big\{\mathbb{G}_0\big[\mathbf{C}_0\big],\Phi_1\Big\{\mathbb{G}_1\big[\mathbf{C}_1\big],\mathbb{G}_1\Big[\cdots \Phi_{T-1}\Big\{\mathbb{G}_{T-1}\big[\mathbf{C}_{T-1}\big],\mathbb{G}_{T-1}\Big[\mathbb{G}_T\big[\mathbf{C}_T\big]\Big]\Big\}\cdots\Big]\Big\}\Big\}$$

Constructing Optimization Problems

Nested - Time - then - Uncertainty (NTU)

$$\varrho_{T-1,T}\Big(\mathbf{C}_{T-1},\mathbf{C}_T\Big) = \mathbb{G}_0\Big[\Phi_0\Big\{\mathbf{C}_0, \mathbb{G}_1\Big[\Phi_1\Big\{\cdots \mathbb{G}_{T-1}\Big[\Phi_{T-1}\big\{\mathbf{C}_{T-1}, \mathbb{G}_T\big[\mathbf{C}_T\big]\big\}\Big]\cdots\Big\}\Big]\Big\}\Big]$$

Nested - Uncertainty - then - Time (NUT)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\dots,\mathbf{C}_T\Big) = \Phi_0\Big\{\mathbb{G}_0\big[\mathbf{C}_0\big], \Phi_1\Big\{\mathbb{G}_1\big[\mathbf{C}_1\big], \mathbb{G}_1\Big[\cdots \Phi_{T-1}\Big\{\mathbb{G}_{T-1}\big[\mathbf{C}_{T-1}\big], \mathbb{G}_{T-1}\Big[\mathbb{G}_T\big[\mathbf{C}_T\big]\Big]\Big\}\cdots\Big]\Big\}\Big\}$$

Constructing Optimization Problems

Nested - Time - then - Uncertainty (NTU)

$$\varrho_{1,T}\Big(\mathbf{C}_1,\cdots,\mathbf{C}_T\Big) = \mathbb{G}_0\Big[\Phi_0\Big\{\mathbf{C}_0, \mathbb{G}_1\Big[\Phi_1\Big\{\cdots \mathbb{G}_{T-1}\Big[\Phi_{T-1}\big\{\mathbf{C}_{T-1}, \mathbb{G}_T\big[\mathbf{C}_T\big]\big\}\Big]\cdots\Big\}\Big]\Big\}\Big]$$

Nested - Uncertainty - then - Time (NUT)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\dots,\mathbf{C}_T\Big) = \Phi_0\Big\{\mathbb{G}_0\big[\mathbf{C}_0\big], \Phi_1\Big\{\mathbb{G}_1\big[\mathbf{C}_1\big], \mathbb{G}_1\Big[\cdots \Phi_{T-1}\Big\{\mathbb{G}_{T-1}\big[\mathbf{C}_{T-1}\big], \mathbb{G}_{T-1}\Big[\mathbb{G}_T\big[\mathbf{C}_T\big]\Big]\Big\}\cdots\Big]\Big\}\Big\}$$

Constructing Optimization Problems

Nested - Time - then - Uncertainty (NTU)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\cdots,\mathbf{C}_T\Big) = \mathbb{G}_0\Big[\Phi_0\Big\{\mathbf{C}_0,\mathbb{G}_1\Big[\Phi_1\Big\{\cdots \mathbb{G}_{T-1}\Big[\Phi_{T-1}\big\{\mathbf{C}_{T-1},\mathbb{G}_T\big[\mathbf{C}_T\big]\big\}\Big]\cdots\Big\}\Big]\Big\}\Big]$$

Nested - Uncertainty - then - Time (NUT)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\dots,\mathbf{C}_T\Big) = \Phi_0\Big\{\mathbb{G}_0\big[\mathbf{C}_0\big],\Phi_1\Big\{\mathbb{G}_1\big[\mathbf{C}_1\big],\mathbb{G}_1\Big[\cdots \Phi_{T-1}\Big\{\mathbb{G}_{T-1}\big[\mathbf{C}_{T-1}\big],\mathbb{G}_{T-1}\Big[\mathbb{G}_T\big[\mathbf{C}_T\big]\Big]\Big\}\cdots\Big]\Big\}\Big\}$$

Constructing Optimization Problems

Nested - Time - then - Uncertainty (NTU)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\cdots,\mathbf{C}_T\Big) = \mathbb{G}_0\Big[\Phi_0\Big\{\mathbf{C}_0, \mathbb{G}_1\Big[\Phi_1\Big\{\cdots \mathbb{G}_{T-1}\Big[\Phi_{T-1}\big\{\mathbf{C}_{T-1}, \mathbb{G}_T\big[\mathbf{C}_T\big]\big\}\Big]\cdots\Big\}\Big]\Big\}\Big]$$

Nested - Uncertainty - then - Time (NUT)

$$\varrho_{0,T}\Big(\mathbf{C}_0,\dots,\mathbf{C}_T\Big) = \Phi_0\Big\{\mathbb{G}_0\big[\mathbf{C}_0\big], \Phi_1\Big\{\mathbb{G}_1\big[\mathbf{C}_1\big], \mathbb{G}_1\Big[\cdots \Phi_{T-1}\Big\{\mathbb{G}_{T-1}\big[\mathbf{C}_{T-1}\big], \mathbb{G}_{T-1}\Big[\mathbb{G}_T\big[\mathbf{C}_T\big]\Big]\Big\}\cdots\Big]\Big\}\Big\}$$

Conditions for a Dynamic Programming Principle (NTU)

De Lara - L.

Assume that the time-step aggregators $\mathbb{G}_t$ and Φ_t are monotonous. Define the value functions

$$\begin{cases} V_T^{\mathrm{NTU}}(x) & = K(x) \\ V_t^{\mathrm{NTU}}(x) & = \inf_{u \in \mathbb{U}_t} \mathbb{G}_t \Big[\Phi_t \Big\{ L_t(x,u,\cdot), V_{t+1}^{\mathrm{NTU}} \circ f_t(x,u,\cdot) \Big\} \Big] \end{cases}$$

Assume that there exists an admissible strategy $\pi^\sharp$ such that

$$\pi_t^\sharp(x) \in \arg\min_{u \in \mathbb{U}_t} \mathbb{G}_t \Big[\Phi_t \Big\{ L_t(x,u,\cdot), V_{t+1}^{\mathrm{NTU}} \circ f_t(x,u,\cdot) \Big\} \Big]$$

Then, $\pi^\sharp$ is an optimal policy.

Conditions for a Dynamic Programming Principle (NUT)

De Lara - L.

Assume that the time-step aggregators $\mathbb{G}_t$ and Φ_t are monotonous. Define the value functions

$$\begin{cases} V_T^{\mathrm{NUT}}(x) & = K(x) \\ V_t^{\mathrm{NUT}}(x) & = \inf_{u \in \mathbb{U}_t} \Phi_t \Big\{ \mathbb{G}_t\Big[L_t(x,u,\cdot) \Big], \mathbb{G}_t\Big[V_{t+1}^{\mathrm{NUT}} \circ f_t(x,u,\cdot) \Big] \Big\} \end{cases}$$

Assume that there exists an admissible strategy $\pi^\sharp$ such that

$$\pi_t^\sharp(x) \in \arg\min_{u \in \mathbb{U}_t} \Phi_t \Big\{ \mathbb{G}_t\Big[L_t(x,u,\cdot) \Big], \mathbb{G}_t\Big[V_{t+1}^{\mathrm{NUT}} \circ f_t(x,u,\cdot) \Big] \Big\}$$

Then, $\pi^\sharp$ is an optimal policy.

Commutation

Commutation

Uncertainty aggregator $\mathbb{G}_{t+1}$ and time-aggregator Φ_t are said to be commuting when, for all functions f and g

$$\mathbb{G}_{t+1}\Big[\Phi_t\Big\{f(\mathbf{W}_t), g(\mathbf{W}_{t+1})\Big\}\Big] = \Phi_t\Big\{f(\mathbf{W}_t), \mathbb{G}_{t+1}\big[g(\mathbf{W}_{t+1})\big]\Big\}$$

Examples:

- $\mathbb{E}_{\mathbb{P}_{t+1}}\Big[f(\mathbf{W}_t) + g(\mathbf{W}_{t+1})\Big] = f(\mathbf{W}_t) + \mathbb{E}_{\mathbb{P}_{t+1}}\big[g(\mathbf{W}_{t+1})\big]$
- commutation with sum $\iff$ translation equivariance property
- $\mathbb{E}_{\mathbb{P}_{t+1}}\Big[f(\mathbf{W}_t) \times g(\mathbf{W}_{t+1})\Big] = f(\mathbf{W}_t) \times \mathbb{E}_{\mathbb{P}_{t+1}}\big[g(\mathbf{W}_{t+1})\big]$

Conditions for a Dynamic Programming Principle (TU)

De Lara - L.

Assume that

- the global aggregators are a composition of time-step aggregators,
- the time-step aggregators $\mathbb{G}_t$ and Φ_t are monotonous,
- the time-step aggregators $\mathbb{G}_t$ and Φ_s $(s < t)$ commute.

Then, the nested and not nested formulations are equivalent, and we have a DP equation.

Presentation Outline

Time-Consistency of a Sequence of Optimization Problems

$$
(\mathcal{P}_t) \qquad \min_{\pi} \quad \varrho_{t,T}\Big(L_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}), \cdots , L_{T-1}(\mathbf{X}_{T-1}, \mathbf{U}_{T-1}, \mathbf{W}_T), K(\mathbf{X}_T) \Big)
$$

$$
\text{s.t.} \quad \mathbf{X}_t = x, \quad \mathbf{X}_{\tau+1} = f_\tau(\mathbf{X}_\tau, \mathbf{U}_\tau, \mathbf{W}_\tau), \quad \mathbf{U}_\tau = \pi_\tau(\mathbf{X}_\tau)
$$

The sequence of problems $(\mathcal{P}_t)_{t\in[\![0,T-1]\!]}$ is said to be time consistent if there exists an optimal strategy of Problem $(\mathcal{P}_{t_0})$ such that its restriction is optimal for $(\mathcal{P}_{t_1})$, $(t_1 > t_0)$.

Time-Consistency of a Dynamic Risk Measure

A sequence of conditional risk measures $\big(\rho_{0,T}, \rho_{1,T}, \dots, \rho_T\big)$ is time-consistent if for any two sequences of costs $\big(\mathbf{C}_0, \dots, \mathbf{C}_T\big)$ $\big(\mathbf{C}'_0, \dots, \mathbf{C}'_T\big)$ we have

$$\left.\begin{array}{rcl} \big(\mathbf{C}_{t_1}, \cdots, \mathbf{C}_{t_2-1}\big) & = & \big(\mathbf{C}'_{t_1}, \cdots, \mathbf{C}'_{t_2-1}\big) \\ \rho_{t_2,T}\big(\mathbf{C}_{t_2}, \cdots, \mathbf{C}_T\big) & \leq & \rho_{t_2,T}\big(\mathbf{C}'_{t_2}, \cdots, \mathbf{C}'_T\big) \end{array}\right\}$$

$$\Longrightarrow \qquad \rho_{t_1,T}\big(\mathbf{C}_{t_1}, \cdots, \mathbf{C}_{t_2}, \cdots \mathbf{C}_T\big) \leq \rho_{t_1,T}\big(\mathbf{C}'_{t_1}, \cdots, \mathbf{C}'_{t_2}, \cdots \mathbf{C}'_T\big)$$

Time-Consistency Result

Nested formulation - De Lara, L.

If the time-step aggregators are monotonous, the induced:

- sequence of optimization problems
- sequence of conditional risk measures

are time consistent.

Non-Nested Formulation - De Lara, L.

If the global aggregators are composition of monotonous and commuting time-step aggregators, the induced

- sequence of optimization problems
- sequence of conditional risk measures

are time consistent.

Markovian Case

- We have extended the framework to allow for Markovian aggregators:

$$\mathbb{G}_t \rightsquigarrow \mathbb{G}_t^x \qquad \Phi_t \rightsquigarrow \Phi_t^x$$

- Examples:
  - Conditional expectation: $\mathbb{G}_t^x = \mathbb{E}\big[\, \cdot \;\big|\; \mathbf{X}_t = x \big]$,
  - Markov risk measure (Ruszczynski 2010).

Presentation Outline

Classical Extension: Multiplicative Case

A stochastic viability problem can be written

$$\begin{aligned} \max_{\pi \in \Pi} \quad & \mathbb{P}\Big(\big\{ \mathbf{X}_t \in \mathcal{X}_t, \quad \forall t \in [\![0, T]\!] \big\} \Big) \\ \text{s.t} \quad & \mathbf{X}_{t+1} = f_t(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}) \\ & \mathbf{U}_t = \pi_t(\mathbf{X}_t) \end{aligned}$$

With the following DP equation

$$\begin{cases} V_T(x) & = \mathbb{E}\Big[\mathbb{1}_{\{x \in \mathcal{X}_T\}} \Big] \\ V_t(x) & = \max_{u \in \mathbb{U}_t} \quad \mathbb{E}\Big[\mathbb{1}_{\{x \in \mathcal{X}_t\}} \cdot V_{t+1} \circ f_t(x, u, \mathbf{W}_{t+1}) \Big] \end{cases}$$

Classical Extension: Multiplicative Case

A stochastic viability problem can be written

$$\begin{aligned} \max_{\pi \in \Pi} \quad & \mathbb{E}\Big[\prod_{t=0}^{T} \mathbb{1}_{\{\mathbf{X}_t \in \mathcal{X}_t\}}\Big] \\ \text{s.t} \quad & \mathbf{X}_{t+1} = f_t\big(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\big) \\ & \mathbf{U}_t = \pi_t(\mathbf{X}_t) \end{aligned}$$

With the following DP equation

$$\begin{cases} V_T(x) & = \mathbb{E}\Big[\mathbb{1}_{\{x \in \mathcal{X}_T\}}\Big] \\ V_t(x) & = \max_{u \in \mathbb{U}_t} \quad \mathbb{E}\Big[\mathbb{1}_{\{x \in \mathcal{X}_t\}} \cdot V_{t+1} \circ f_t(x, u, \mathbf{W}_{t+1})\Big] \end{cases}$$

Classical Extension: Multiplicative Case

A stochastic viability problem can be written

$$\begin{aligned} \max_{\pi \in \Pi} \quad & \mathbb{E}\Big[\prod_{t=0}^{T} \mathbb{1}_{\{\mathbf{X}_t \in \mathcal{X}_t\}}\Big] \\ \text{s.t} \quad & \mathbf{X}_{t+1} = f_t\big(\mathbf{X}_t, \mathbf{U}_t, \mathbf{W}_{t+1}\big) \\ & \mathbf{U}_t = \pi_t(\mathbf{X}_t) \end{aligned}$$

With the following DP equation

$$\begin{cases} V_T(x) & = \mathbb{E}\Big[\mathbb{1}_{\{x \in \mathcal{X}_T\}}\Big] \\ V_t(x) & = \max_{u \in \mathbb{U}_t} \quad \mathbb{E}\Big[\mathbb{1}_{\{x \in \mathcal{X}_t\}} \cdot V_{t+1} \circ f_t(x, u, \mathbf{W}_{t+1})\Big] \end{cases}$$

Coherent Risk Measure

Consider the following sequence of conditional risk measures.

$$\varrho_{t,T}(\mathbf{C}) = \sup_{\mathbb{P}_t \in \mathcal{P}_t} \mathbb{E}_{\mathbb{P}_t}\Big[\cdots \sup_{\mathbb{P}_T \in \mathcal{P}_T} \mathbb{E}_{\mathbb{P}_T}\Big[\sum_{s=t}^{T}\Big(\alpha_s(\mathbf{C}_s)\prod_{r=t}^{s-1}\beta_r(\mathbf{C}_r)\Big)\Big]\cdots\Big]$$

The associated optimization problem is solved by the following DP equation (if $\beta_t \geq 0$)

$$\begin{cases} V_T(x) & = K(x) \\ V_t(x) & = \inf_{u} \sup_{\mathbb{P}_t \in \mathcal{P}_t} \Big\{\mathbb{E}_{\mathbb{P}_t}\Big[\alpha_t\big(L_t(x,u,\cdot)\big) + \beta_t\big(L_t(x,u,\cdot)\big) V_{t+1} \circ f_t(x,u,\cdot)\Big]\Big\} \end{cases}$$

Elements of proof

- The problem is of (TU) form where the global aggregators are composition of the following time-step aggregators:

$$\begin{cases} \mathbb{G}_t\big[\,\cdot\,\big] & = \sup_{\mathbb{P}_t \in \mathcal{P}_t} \mathbb{E}_{\mathbb{P}_t}\big[\,\cdot\,\big] \\ \Phi_t\big\{c, c'\big\} & = \alpha_t(c) + \beta_t(c)c' \end{cases}$$

- The time-step aggregators are monotonous.
- The time-step aggregators commute:

$$\begin{aligned} \mathbb{G}_t\Big[\Phi_s\big\{\mathbf{C}_s, \mathbf{C}_t\big\}\Big] &= \sup_{\mathbb{P}_t \in \mathcal{P}_t} \Big(\mathbb{E}_{\mathbb{P}_t}\Big[\alpha_s(\mathbf{C}_s) + \beta_s(\mathbf{C}_s)\mathbf{C}_t\Big]\Big) \\ &= \sup_{\mathbb{P}_t \in \mathcal{P}_t} \Big(\alpha_s(\mathbf{C}_s) + \beta_s(\mathbf{C}_s)\mathbb{E}_{\mathbb{P}_t}\big[\mathbf{C}_t\big]\Big) && \text{Translation-equiv.} \\ &= \alpha_s(\mathbf{C}_s) + \beta_s(\mathbf{C}_s) \sup_{\mathbb{P}_t \in \mathcal{P}_t} \Big(\mathbb{E}_{\mathbb{P}_t}\big[\mathbf{C}_t\big]\Big) && \text{Pos. Homogeneity} \\ &= \Phi_s\Big\{\mathbf{C}_s, \mathbb{G}_t\big[\mathbf{C}_t\big]\Big\} \end{aligned}$$

Elements of proof

- The problem is of (TU) form where the global aggregators are composition of the following time-step aggregators:

$$\begin{cases} \mathbb{G}_t\big[\,\cdot\,\big] & = \sup_{\mathbb{P}_t \in \mathcal{P}_t} \mathbb{E}_{\mathbb{P}_t}\big[\,\cdot\,\big] \\ \Phi_t\big\{c, c'\big\} & = \alpha_t(c) + \beta_t(c)c' \end{cases}$$

- The time-step aggregators are monotonous.
- The time-step aggregators commute:

$$\begin{aligned} \mathbb{G}_t\Big[\Phi_s\big\{\mathbf{C}_s, \mathbf{C}_t\big\}\Big] &= \sup_{\mathbb{P}_t \in \mathcal{P}_t} \Big(\mathbb{E}_{\mathbb{P}_t}\Big[\alpha_s(\mathbf{C}_s) + \beta_s(\mathbf{C}_s)\mathbf{C}_t\Big]\Big) \\ &= \sup_{\mathbb{P}_t \in \mathcal{P}_t} \Big(\alpha_s(\mathbf{C}_s) + \beta_s(\mathbf{C}_s)\mathbb{E}_{\mathbb{P}_t}\big[\mathbf{C}_t\big]\Big) && \text{Translation-equiv.} \\ &= \alpha_s(\mathbf{C}_s) + \beta_s(\mathbf{C}_s) \sup_{\mathbb{P}_t \in \mathcal{P}_t} \Big(\mathbb{E}_{\mathbb{P}_t}\big[\mathbf{C}_t\big]\Big) && \text{Pos. Homogeneity} \\ &= \Phi_s\Big\{\mathbf{C}_s, \mathbb{G}_t\big[\mathbf{C}_t\big]\Big\} \end{aligned}$$

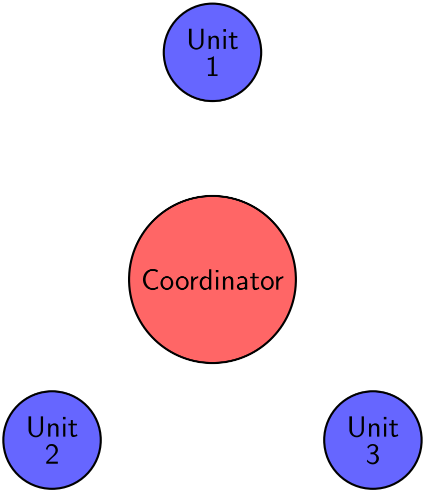
Conclusion of Part I

- We have presented a generic framework for stochastic optimization problem and conditions to write a chained time decomposition through a DP equation. We extended it to a Markovian framework.
- We show that our conditions lead to time-consistency of
  - the sequence of induced optimization problems,
  - and the induced dynamic risk measure.
- This part was concerned with formulation of problem in a time-consistent way, and time decomposition. However, it is still affected by the so-called "curse of dimensionality".

Presentation Outline

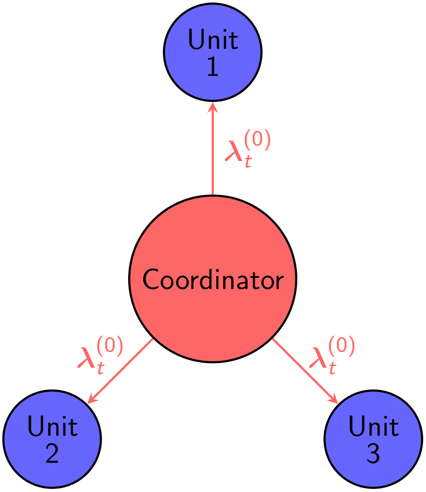
Intuition of Spatial Decomposition

- Satisfy a demand (over T time step) with N units of production at minimal cost.
- Price decomposition:
  - the coordinator sets a sequence of price λ_t,
  - the units send their production planning $\mathbf{U}_t^{(i)}$,
  - the coordinator compares total production and demand and updates the price,
  - and so on...



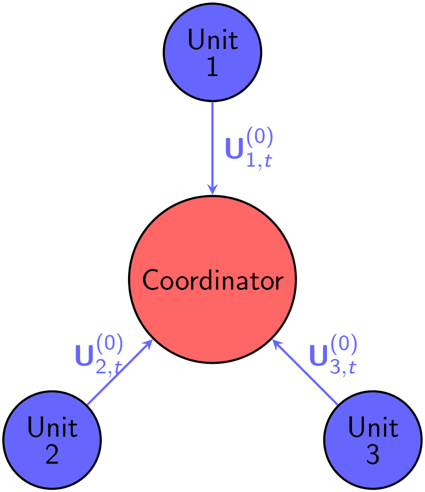
Intuition of Spatial Decomposition

- Satisfy a demand (over T time step) with N units of production at minimal cost.
- Price decomposition:
  - the coordinator sets a sequence of price λ_t,
  - the units send their production planning $\mathbf{U}_t^{(i)}$,
  - the coordinator compares total production and demand and updates the price,
  - and so on...



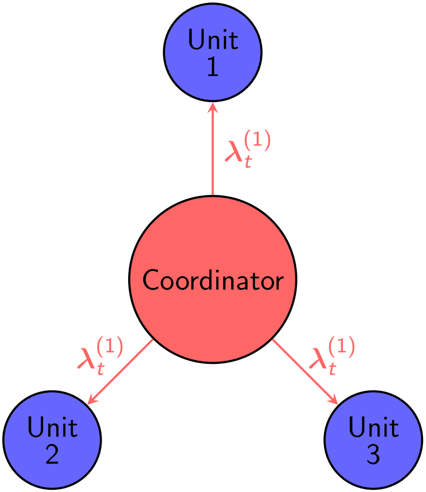
Intuition of Spatial Decomposition

- Satisfy a demand (over T time step) with N units of production at minimal cost.
- Price decomposition:
  - the coordinator sets a sequence of price λ_t,
  - the units send their production planning $\mathbf{U}_t^{(i)}$,
  - the coordinator compares total production and demand and updates the price,
  - and so on...



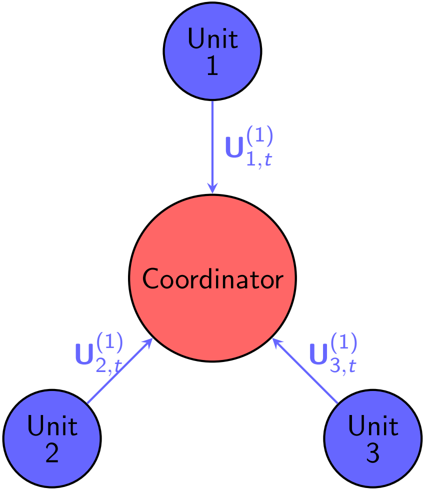
Intuition of Spatial Decomposition

- Satisfy a demand (over T time step) with N units of production at minimal cost.
- Price decomposition:
  - the coordinator sets a sequence of price λ_t,
  - the units send their production planning $\mathbf{U}_t^{(i)}$,
  - the coordinator compares total production and demand and updates the price,
  - and so on...



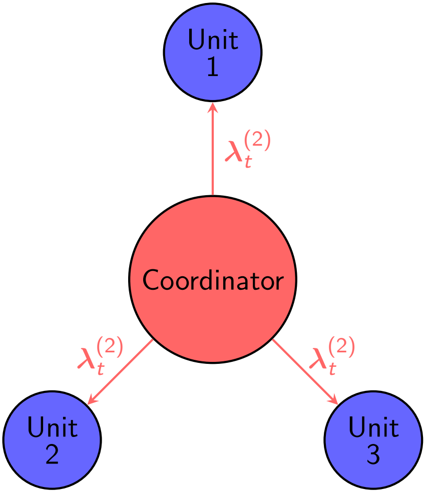
Intuition of Spatial Decomposition

- Satisfy a demand (over T time step) with N units of production at minimal cost.
- Price decomposition:
  - the coordinator sets a sequence of price λ_t,
  - the units send their production planning $\mathbf{U}_t^{(i)}$,
  - the coordinator compares total production and demand and updates the price,
  - and so on...



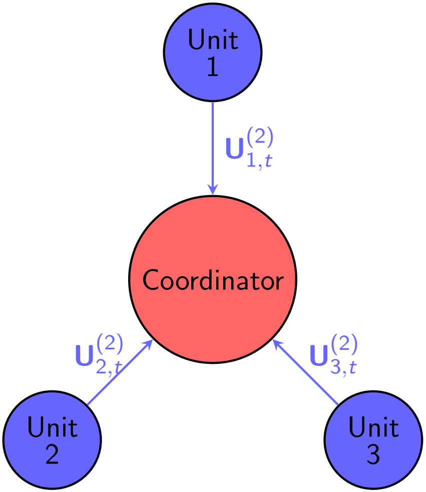
Intuition of Spatial Decomposition

- Satisfy a demand (over T time step) with N units of production at minimal cost.
- Price decomposition:
  - the coordinator sets a sequence of price λ_t,
  - the units send their production planning $\mathbf{U}_t^{(i)}$,
  - the coordinator compares total production and demand and updates the price,
  - and so on...



Intuition of Spatial Decomposition

- Satisfy a demand (over T time step) with N units of production at minimal cost.
- Price decomposition:
  - the coordinator sets a sequence of price λ_t,
  - the units send their production planning $\mathbf{U}_t^{(i)}$,
  - the coordinator compares total production and demand and updates the price,
  - and so on...



Primal Problem

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

$$\forall\, i, \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}), \quad \mathbf{X}_0^i = x_0^i,$$

$$\forall\, i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t,$$

$$\sum_{i=1}^{N} \theta_t^i(\mathbf{U}_t^i) = 0$$

Solvable by DP with state $(\mathbf{X}_1, \dots, \mathbf{X}_N)$ (under noise independence assumption)

Primal Problem

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

$$\forall\, i, \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}), \quad \mathbf{X}_0^i = x_0^i,$$

$$\forall\, i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t,$$

$$\sum_{i=1}^{N} \theta_t^i(\mathbf{U}_t^i) = 0 \qquad \rightsquigarrow \boldsymbol{\lambda}_t \quad \text{multiplier}$$

Solvable by DP with state $(\mathbf{X}_1, \dots, \mathbf{X}_N)$ (under noise independence assumption)

Primal Problem with Dualized Constraint

$$\min_{\mathbf{X},\mathbf{U}} \max_{\boldsymbol{\lambda}} \sum_{i=1}^{N} \mathbb{E}\Big[\sum_{t=0}^{T} L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}) + \langle \boldsymbol{\lambda}_t, \theta_t^i(\mathbf{U}_t^i) \rangle + K^i(\mathbf{X}_T^i) \Big]$$

$$\forall\, i, \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}), \quad \mathbf{X}_0^i = x_0^i,$$

$$\forall\, i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t,$$

Coupling constraint dualized $\Longrightarrow$ remaining constraints are i by i

Dual Problem

$$\max_{\lambda} \min_{\mathbf{X},\mathbf{U}} \sum_{i=1}^{N} \mathbb{E}\Big[\sum_{t=0}^{T} L_t^i\big(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}\big) + \big\langle \boldsymbol{\lambda}_t, \theta_t^i(\mathbf{U}_t^i)\big\rangle + K^i(\mathbf{X}_T^i)\Big]$$

$$\forall\, i, \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_{t+1}), \quad \mathbf{X}_0^i = x_0^i,$$

$$\forall\, i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t,$$

Exchange operator min and max to obtain a new problem

Decomposed Dual Problem

$$\max_{\boldsymbol{\lambda}} \sum_{i=1}^{N} \min_{\mathbf{X}^i,\mathbf{U}^i} \mathbb{E}\Big[\sum_{t=0}^{T} L_t^i(\mathbf{X}_t^i,\mathbf{U}_t^i,\mathbf{W}_{t+1}) + \langle \boldsymbol{\lambda}_t, \theta_t^i(\mathbf{U}_t^i)\rangle + K^i(\mathbf{X}_T^i)\Big]$$

$$\mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i,\mathbf{U}_t^i,\mathbf{W}_{t+1}), \quad \mathbf{X}_0^i = x_0^i,$$

$$\mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t,$$

For a given λ, minimum of sum is sum of minima

Inner Minimization Problem

$$\min_{\mathbf{X}^i,\mathbf{U}^i} \mathbb{E}\Big[\sum_{t=0}^{T} L_t^i(\mathbf{X}_t^i,\mathbf{U}_t^i,\mathbf{W}_{t+1}) + \langle \boldsymbol{\lambda}_t, \theta_t^i(\mathbf{U}_t^i)\rangle + K^i(\mathbf{X}_T^i)\Big]$$

$$\mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i,\mathbf{U}_t^i,\mathbf{W}_{t+1}), \quad \mathbf{X}_0^i = x_0^i,$$

$$\mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t,$$

We have N smaller subproblems. Can they be solved by DP?

Inner Minimization Problem

$$\min_{\mathbf{X}^i,\mathbf{U}^i} \mathbb{E}\Big[\sum_{t=0}^{T} L_t^i(\mathbf{X}_t^i,\mathbf{U}_t^i,\mathbf{W}_{t+1}) + \langle \boldsymbol{\lambda}_t, \theta_t^i(\mathbf{U}_t^i)\rangle + K^i(\mathbf{X}_T^i)\Big]$$

$$\mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i,\mathbf{U}_t^i,\mathbf{W}_{t+1}), \quad \mathbf{X}_0^i = x_0^i,$$

$$\mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t,$$

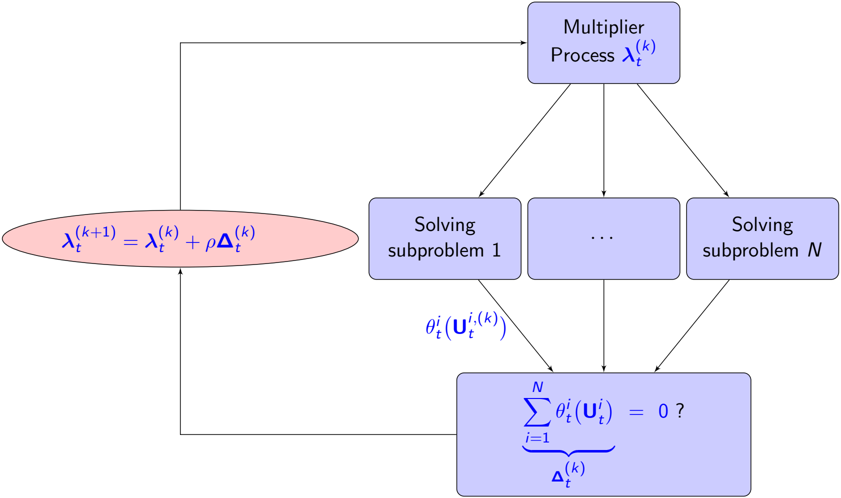
No: $\boldsymbol{\lambda}$ is a time-dependent noise $\rightsquigarrow$ $\mathbf{X}_t^i$ is not a proper state, but rather $(\mathbf{W}_1,\dots,\mathbf{W}_t)$

A Few Questions

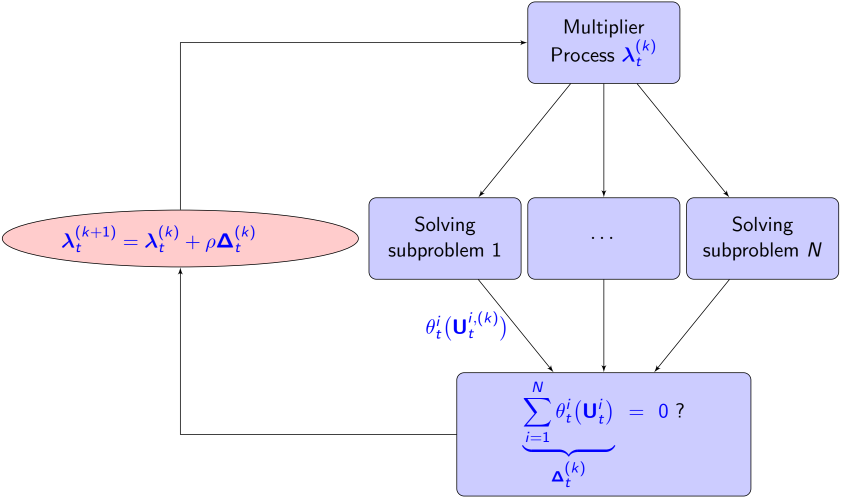
- What is the duality scheme ? In which space lives the multiplier process $\boldsymbol{\lambda}$?
  - L^2
  - L^1
  - $\left(\mathrm{L}^\infty\right)^\star$
- What are the relations between the primal and dual problem?
- Can we solve the subproblems by Dynamic Programming?
  $\rightsquigarrow$ No! (with small enough state)
- How to update the multiplier process?
  $\rightsquigarrow$ "gradient step":

$$\boldsymbol{\lambda}_t^{(k+1)} = \boldsymbol{\lambda}_t^{(k)} + \rho \sum_{i=1}^{N} \theta_t^i\big(\mathbf{U}_t^{i,k}\big)$$

Stochastic spatial decomposition scheme

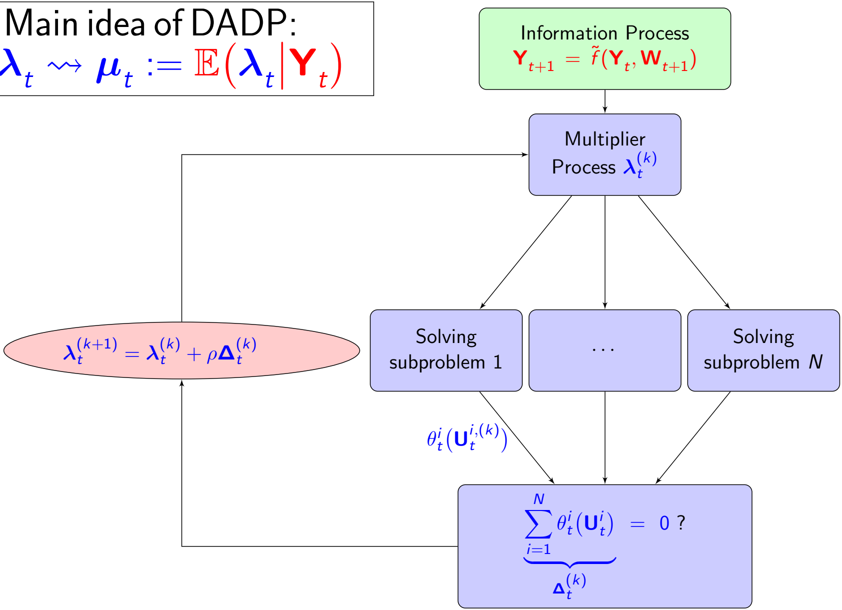


Main idea of DADP:
$\boldsymbol{\lambda}_t \rightsquigarrow \boldsymbol{\mu}_t := \mathbb{E}\big(\boldsymbol{\lambda}_t \big| \mathbf{Y}_t\big)$

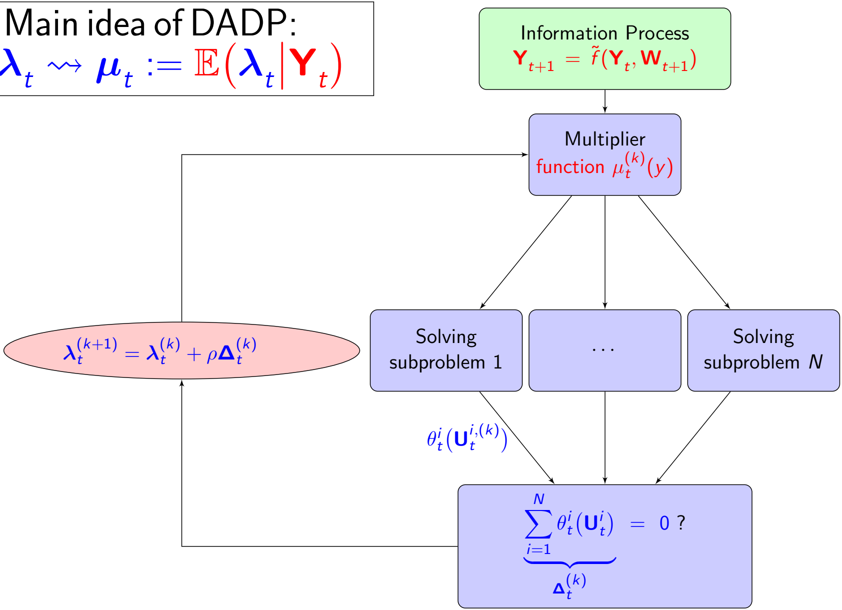


Main idea of DADP:

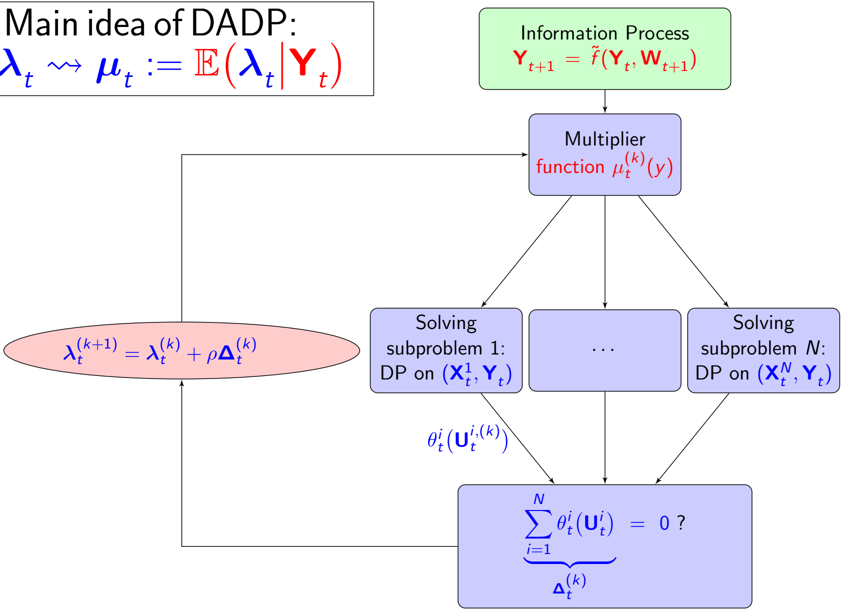
$$\boldsymbol{\lambda}_t \rightsquigarrow \boldsymbol{\mu}_t := \mathbb{E}\big(\boldsymbol{\lambda}_t \big| \mathbf{Y}_t\big)$$



Main idea of DADP:
$\boldsymbol{\lambda}_t \rightsquigarrow \boldsymbol{\mu}_t := \mathbb{E}\big(\boldsymbol{\lambda}_t \big| \mathbf{Y}_t\big)$

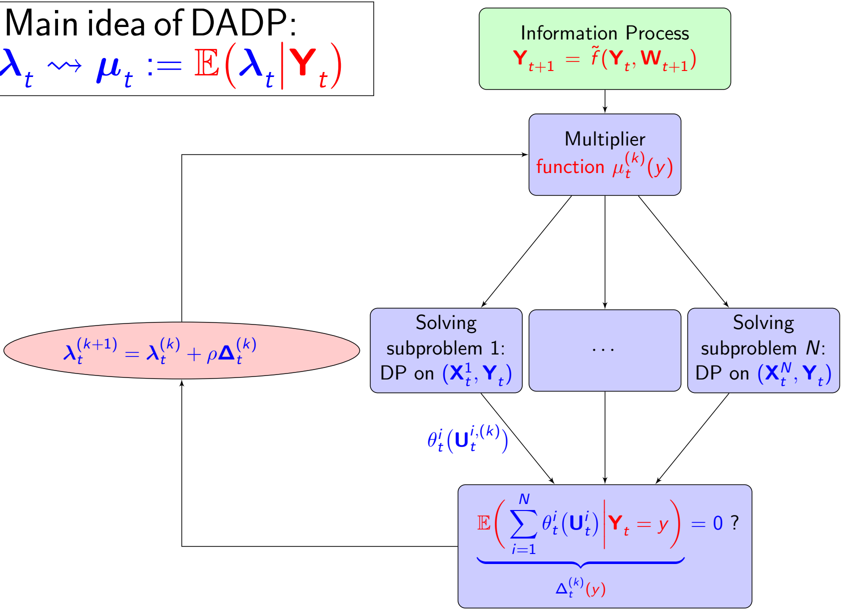


Main idea of DADP:
$\boldsymbol{\lambda}_t \rightsquigarrow \boldsymbol{\mu}_t := \mathbb{E}\big(\boldsymbol{\lambda}_t \big| \mathbf{Y}_t\big)$



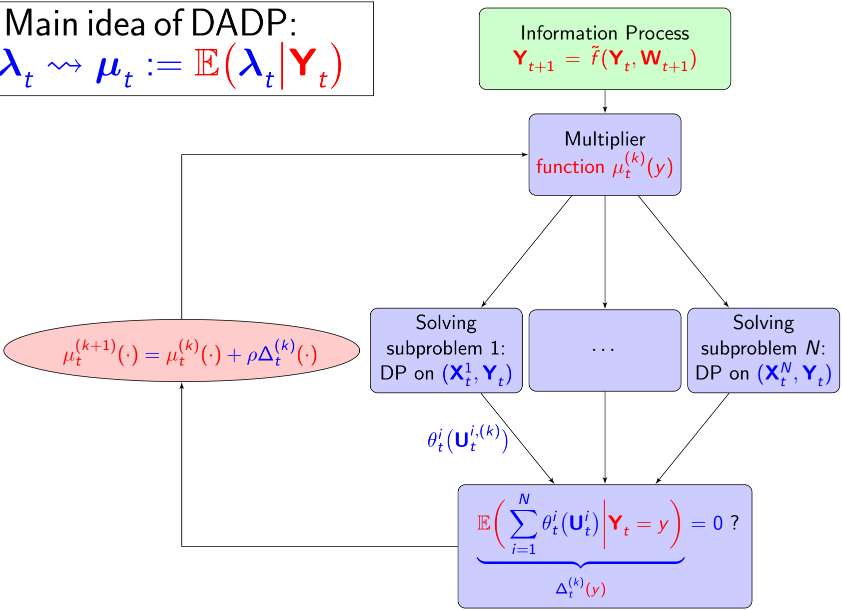
Main idea of DADP:

$$\boldsymbol{\lambda}_t \rightsquigarrow \boldsymbol{\mu}_t := \mathbb{E}\big(\boldsymbol{\lambda}_t \big| \mathbf{Y}_t\big)$$

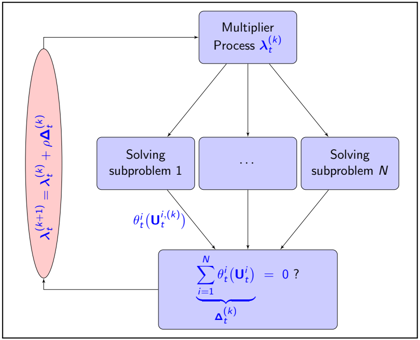


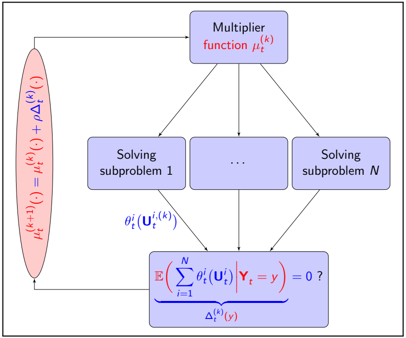
Main idea of DADP:

$$\boldsymbol{\lambda}_t \rightsquigarrow \boldsymbol{\mu}_t := \mathbb{E}\big(\boldsymbol{\lambda}_t \big| \mathbf{Y}_t\big)$$



Main idea of DADP: $\boldsymbol{\lambda}_t \rightsquigarrow \boldsymbol{\mu}_t := \mathbb{E}\big(\boldsymbol{\lambda}_t \big| \mathbf{Y}_t\big)$





Main problems:

- Subproblems not easily solvable by DP
- $\boldsymbol{\lambda}^{(k)}$ live in a huge space

Advantages:

- Subproblems solvable by DP with state $\big(\mathbf{X}_t^i, \mathbf{Y}_t\big)$
- $\mu^{(k)}$ live in a smaller space

Presentation Outline

1. Time-Consistency: from Optimization to Risk Measures
   - A Framework for Dynamic Programming
   - Conditions for Time-Consistency
   - Examples
2. Spatial Stochastic Decomposition Method
   - Spatial Decomposition
   - Theoretical Results

Three Interpretations of DADP

- DADP as an approximation of the optimal multiplier

$$\boldsymbol{\lambda}_t \quad \rightsquigarrow \quad \mathbb{E}\left(\boldsymbol{\lambda}_t \middle| \mathbf{Y}_t\right) .$$

- DADP as a decision-rule approach in the dual

$$\max_{\boldsymbol{\lambda}} \min_{\mathbf{U}} L\left(\boldsymbol{\lambda}, \mathbf{U}\right) \quad \rightsquigarrow \quad \max_{\boldsymbol{\lambda}_t \preceq \mathbf{Y}_t} \min_{\mathbf{U}} L\left(\boldsymbol{\lambda}, \mathbf{U}\right) .$$

- DADP as a constraint relaxation in the primal

$$\sum_{i=1}^{n} \theta_t^i\left(\mathbf{U}_t^i\right) = 0 \quad \rightsquigarrow \quad \mathbb{E}\left(\sum_{i=1}^{n} \theta_t^i\left(\mathbf{U}_t^i\right) \middle| \mathbf{Y}_t\right) = 0 .$$

Consistence of the Approximation Scheme

- The DADP algorithm solves a relaxation $(\mathcal{P}_{\mathbf{Y}})$ of the original problem $(\mathcal{P})$ where

$$\sum_{i=1}^{n} \theta_t^i(\mathbf{U}_t^i) = 0 \qquad \rightsquigarrow \qquad \mathbb{E}\left(\sum_{i=1}^{n} \theta_t^i(\mathbf{U}_t^i) \,\middle|\, \mathbf{Y}_t \right) = 0$$

- Question: if we consider a sequence of information processes $\left\{\mathbf{Y}^{(n)}\right\}_{n\in\mathbb{N}}$, such that the information converges

$$\sigma\left(\mathbf{Y}_t^{(n)}\right) \to \sigma\left(\mathbf{W}_0, \cdots, \mathbf{W}_t\right)$$

does the associated sequence $\left(\mathbf{U}^{\mathbf{Y}^{(n)}}\right)$ of optimal control converges toward an optimal control of $(\mathcal{P})$?

Epiconvergence of Approximation

Epiconvergence result -L.

Assume that

- the cost functions L_t^i, dynamic functions f_t^i and constraint functions θ_t^i are continuous;
- the noise variables $\mathbf{W}_t$ are essentially bounded;
- the constraint sets $\mathcal{U}_{i,t}^{\mathrm{ad}}$ are bounded.

Consider a sequence of information process $\left\{\mathbf{Y}^{(n)}\right\}_{n\in\mathbb{N}}$ such that $\sigma\left(\mathbf{Y}^{(n)}\right) \to \mathcal{F}_\infty$. Let $\mathbf{U}^{(n)}$ be an ε_n-optimal solution to the relaxed problem $\left(\mathcal{P}^{\mathbf{Y}^{(n)}}\right)$.

Then, every cluster point[a] of $\left\{\mathbf{U}^{(n)}\right\}_{n\in\mathbb{N}}$ is an optimal solution of the relaxation corresponding to $\mathcal{F}_\infty$.

[a] for the topology of the convergence in probability

Convergence of Coordination Method

- We consider a given information process $\mathbf{Y}$.
- Question: does the algorithm
  1. solve the N subproblems
  2. update the multiplier by a gradient-step

  yield a converging sequence of controls $\mathbf{U}^{(k)}$?
- It is an application of the so-called Uzawa algorithm. This algorithm take naturally place in an Hilbert space, here L^2 is the natural choice. However, existence of saddle-point in L^2 is difficult to prove. Hence we adapt the algorithm to a non-reflexive Banach space: L^∞.

Coordination-Convergence Result

Convergence result - Carpentier,L.

Assume that,

- the set of uncertainties is finite;
- the local cost L_t^i are Gâteaux-differentiable functions, strongly convex (in (x, u)) and continuous (in w);
- the evolution functions f_t are affine (in (x, u, w));
- the coupling functions θ_t^i are affine;
- the admissible set $\mathcal{U}_{i,t}^{\mathrm{ad}} \neq \emptyset$ is a weak* closed, convex set;
- there exists an admissible control;
- the coupling constraint admits an optimal multiplier in L^2 .

For a step $\rho > 0$ small enough, the sequence of control generated by DADP converges in L^∞ toward the optimal control of the relaxed problem.

Coordination-Convergence Result

Convergence result - Carpentier,L.

Assume that,

- the set of uncertainties is finite;
- the local cost L_t^i are Gâteaux-differentiable functions, strongly convex (in (x, u)) and continuous (in w);
- the evolution functions f_t are affine (in (x, u, w));
- the coupling functions θ_t^i are affine;
- the admissible set $\mathcal{U}_{i,t}^{\mathrm{ad}} \neq \emptyset$ is a weak* closed, convex set;
- there exists an admissible control;
- the coupling constraint admits an optimal multiplier in L^1 .

For a step $\rho > 0$ small enough, there exists a subsequence of the sequence of control generated by DADP converging in L^∞ toward the optimal control of the relaxed problem.

Existence of Multiplier

Existence of multiplier -L.

Assume that

- the random noises $\mathbf{W}_t$ are essentially bounded;
- the local cost functions L_t^i are finite and convex in (x_i, u_i), continuous in w;
- the dynamic functions f_t^i are affine in (x_i, u_i), continuous in w;
- the constraint functions θ_t^i are affine;
- there is no bound constraints on $\mathbf{U}_t^i$ and $\mathbf{X}_t^i$.

Then, the coupling constraint admits a multiplier in L^1, hence the relaxed coupling constraint admits a multiplier in L^1.

Existence of Multiplier

Existence of multiplier -L.

Assume that

- the random noises $\mathbf{W}_t$ are essentially bounded;
- the local cost functions L_t^i are finite and convex in (x_i, u_i), continuous in w;
- the dynamic functions f_t^i are affine in (x_i, u_i), continuous in w;
- the constraint functions θ_t^i are affine;
- there is no bound constraints on $\mathbf{U}_t^i$ and $\mathbf{X}_t^i$.

Then, the coupling constraint admits a multiplier in L^1, hence the relaxed coupling constraint admits a multiplier in L^1.

Bounds over the Original Problem

Upper and lower bounds

Lower Bound : For a given $\mu^{(k)}$ we have a lower bound of the dual of the relaxed problem $(\mathcal{P}^{\mathbf{Y}})$, hence a lower bound of the original problem $(\mathcal{P})$.

Upper bound : Through an heuristic (using the DP equation) we can construct an admissible (for the original problem $(\mathcal{P})$) solution and hence obtain an upper bound (by Monte Carlo).

In practice, on a simple problem:

- around 3% gap with minimal information ($\mathbf{Y}_t \equiv 0$),
- around 2% gap with dynamic information.

Validity a posteriori

Validity

If we obtain a multiplier $\boldsymbol{\mu}^\sharp$ leading to a solution $\mathbf{U}(\mu^\sharp)$ satisfying the (relaxed) constraint:

$$\mathbb{E}\Big[\sum_{i=1}^{N} \theta_t\big(\mathbf{U}_t^i(\mu^\sharp)\big) \ \Big|\ \mathbf{Y}_t\Big] = 0$$

then the solution $\mathbf{U}(\mu^\sharp)$ is optimal (for the relaxed problem $(\mathcal{P}_{\mathbf{Y}})$).

Consequences:

- A Posteriori conclusion even if abstract conditions not verified,
- use of improved multiplier update step.

Conclusion of Part II

- Summing up DADP:
  - Choose an information process $\mathbf{Y}$ following $\mathbf{Y}_{t+1} = \tilde{f}_t(\mathbf{Y}_t, \mathbf{W}_{t+1})$.
  - We relax the almost sure coupling constraint into a conditional expectation one and apply a price decomposition scheme to the relaxed problem.
  - The subproblems can be solved by dynamic programming with the state $(\mathbf{X}_t^i, \mathbf{Y}_t)$.
- We give:
  - a consistency result (family of information process),
  - a convergence result (fixed information process),
  - an existence of multiplier condition.

Thesis Outline

1. Prelimaries

2. Time-Consistency: from Optimization to Risk Measures
3. Stochastic Dual Dynamic Programming Algorithm

4. Constraint Qualification in Stochastic Optimization
5. Constraint Qualification in $(\mathrm{L}^\infty, \mathrm{L}^1)$
6. Uzawa Algorithm in L^∞

7. Epiconvergence of Relaxed Stochastic Problems
8. Dual Approximate Dynamic Programming Algorithm

Thesis Outline

1. Prelimaries
2. Time-Consistency: from Optimization to Risk Measures
3. Stochastic Dual Dynamic Programming Algorithm
4. Constraint Qualification in Stochastic Optimization
5. Constraint Qualification in $(\mathrm{L}^\infty, \mathrm{L}^1)$
6. Uzawa Algorithm in L^∞
7. Epiconvergence of Relaxed Stochastic Problems
8. Dual Approximate Dynamic Programming Algorithm

Thesis Outline

Thesis Outline

Thesis Outline

Thesis Outline

Thesis Outline

1. Prelimaries
2. Time-Consistency: from Optimization to Risk Measures
3. Stochastic Dual Dynamic Programming Algorithm
4. Constraint Qualification in Stochastic Optimization
5. Constraint Qualification in $(\mathrm{L}^\infty, \mathrm{L}^1)$
6. Uzawa Algorithm in L^∞
7. Epiconvergence of Relaxed Stochastic Problems
8. Dual Approximate Dynamic Programming Algorithm

Papers

- P. Girardeau, V. Leclère and A. Philpott
  On the convergence of decomposition methods for multi-stage stochastic convex programs.
  Accepted in Mathematics of Operations Research, 2014
- M. De Lara and V. Leclère
  Time-Consistency: from Optimization to Risk Measures
  Submitted, 2014
- V. Leclère
  Epiconvergence of relaxed stochastic optimization problem.
  Submitted, 2013
- M. Grasselli, M. Ludkovski and V. Leclère
  Priority option: the value of being a leader.
  IJTAF, 16, 2013

Conclusion: the next steps

2. Dynamic Programming
   - extension of state
   - more generic links
3. SDDP
   - noise with compact support
   - convergence estimation
4. L^1 multiplier
   - bounds on control via Relatively Complete Recourse
   - conditions for L^2 multiplier
6. Uzawa in L^∞
   - reflexions around the strong-convexity
   - use ε-convergence theory
7. Epiconvergence
   - obtain a non-asymptotical bound
8. DADP
   - Numerical test on big scale
   - Method to construct $\mathbf{Y}$
   - Interactions with SDDP

The end

Thank you for your attention!