Epiconvergence of relaxed stochastic optimization problem

V. Leclère (CERMICS, ENPC).

October 4, 2013

Contents

Model presentation

Parameters:

- storage level $\mathbf{X}_t^i$,
- hydroturbine outflows $\mathbf{U}_t^i$,
- external inflows $\mathbf{W}_t^i$.

Objective function:

$$\mathbb{E}\left(\sum_{i=1}^{N}\sum_{t=0}^{T-1} L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t) + K_i(\mathbf{X}_T^i)\right)$$

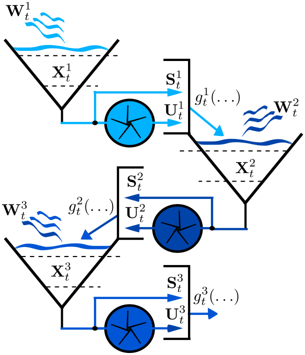
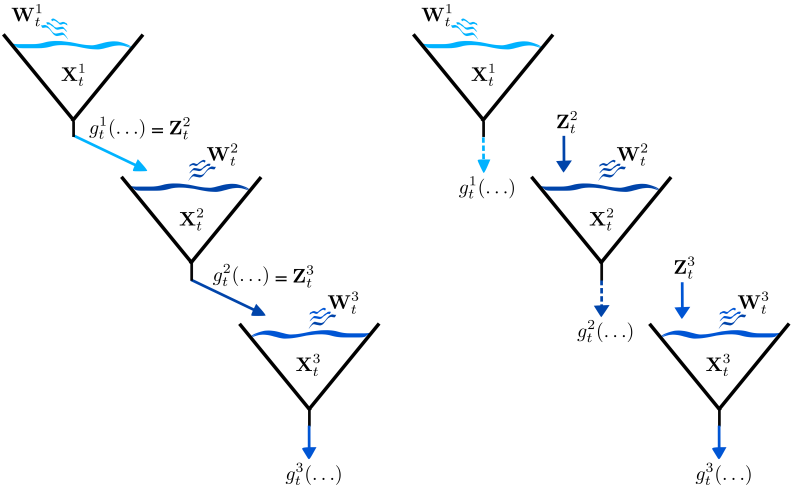


Figure: The river chain model

Decomposition Principle



(a) (b)

Contents

Abstract formulation of the problem

$$
\begin{array}{rll}
\min_{\mathbf{X},\mathbf{U}} & \mathbb{E}\Big(\sum_{t=0}^{T-1} \sum_{i=1}^{N} L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t) + K^i(\mathbf{X}_T^i) \Big) & \\
\text{(dynamic)} & \forall\, i, \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t) & \\
 & \forall\, i, \quad \mathbf{X}_0^i = x_0^i & \\
\text{(bounds constraint)} & \forall\, i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad} & \\
\text{(information constraint)} & \forall\, i, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t & \text{i.e. } \mathbf{U}_t^i \text{ is } \mathcal{F}_t \text{ meas.} \\
\text{(coupling constraint)} & \sum_{i=1}^{N} \theta_t^i(\mathbf{U}_t^i) = 0 & a.s.
\end{array}
$$

Primal problem

$$
\begin{aligned}
\min_{\mathbf{X},\mathbf{U}} \sum_{i=1}^{N} \quad & \mathbb{E}\left(\sum_{t=0}^{T} L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t) + K^i(\mathbf{X}_T^i) \right) \\
& \forall\, i, \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t) \\
& \forall\, i, \quad \mathbf{X}_0^i = x_0^i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t, \\
& \sum_{i=1}^{N} \theta_t^i(\mathbf{U}_t^i) = 0
\end{aligned}
$$

Primal problem

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

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

$$\forall\, i, \quad \mathbf{X}_0^i = x_0^i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t .$$

Dual problem

$$
\begin{aligned}
\max_{\boldsymbol{\lambda}} \min_{\mathbf{X},\mathbf{U}} \sum_{i=1}^{N} \quad & \mathbb{E}\Big(\sum_{t=0}^{T} L_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t) + \langle \boldsymbol{\lambda}_t, \theta_t^i(\mathbf{U}_t^i) \rangle + K^i(\mathbf{X}_T^i) \Big) \\
& \forall\, i, \quad \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i, \mathbf{U}_t^i, \mathbf{W}_t) \\
& \forall\, i, \quad \mathbf{X}_0^i = x_0^i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t.
\end{aligned}
$$

Decomposed problem

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

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

$$\mathbf{X}_0^i = x_0^i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t.$$

General scheme of the decomposition algorithm

The price decomposition algorithm is done as follows, given a multiplier process $(\boldsymbol{\lambda}_t)_{t\in[\![0,T-1]\!]}$:

1. solve N problems with only one dam,
2. update the multiplier by a gradient step.

We need to specify:

- How to solve the one-dam problem?
- How to update the multiplier?

Contents

Recalling the one-dam problem

Recall that the inner problem reads:

$$\begin{aligned} \min_{\mathbf{X}^i,\mathbf{U}^i} \quad & \mathbb{E}\bigg(\sum_{t=0}^{T} L_t^i(\mathbf{X}_t^i,\mathbf{U}_t^i,\mathbf{W}_t) + \langle \boldsymbol{\lambda}_t^{(k)}, \theta_t^i(\mathbf{U}_t^i)\rangle + K^i(\mathbf{X}_T^i)\bigg) \\ & \mathbf{X}_{t+1}^i = f_t^i(\mathbf{X}_t^i,\mathbf{U}_t^i,\mathbf{W}_t) \\ & \mathbf{X}_0^i = x_0^i, \quad \mathbf{U}_t^i \in \mathcal{U}_{t,i}^{ad}, \quad \mathbf{U}_t^i \preceq \mathcal{F}_t \end{aligned}$$

This problem can be solved by Dynamic Programming with the extended state $(\mathbf{X}_t^i, \boldsymbol{\lambda}_{[t]}^{(k)})$, where $\boldsymbol{\lambda}_{[t]}^{(k)} = \big(\boldsymbol{\lambda}_0^{(k)}, \cdots, \boldsymbol{\lambda}_t^{(k)}\big)$.

Using a conditional expectation: $\boldsymbol{\lambda} \rightsquigarrow \mathbb{E}\big(\boldsymbol{\lambda} \bigm| \mathbf{Y}_t\big)$

Idea behind the Dual Approximate Dynamic Programming algorithm :

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

We will see that it is equivalent to

$$\sum_{i=1}^{N} \theta_t^i(\mathbf{U}_t^i) = 0 \quad \mathbb{P} - a.s. \rightsquigarrow \mathbb{E}\Big(\sum_{i=1}^{N} \theta_t^i(\mathbf{U}_t^i) \Big| \mathbf{Y}_t\Big) = 0 \quad \mathbb{P} - a.s.$$

If $(\mathbf{Y}_t)_{t\in[\![0,T-1]\!]}$ is a Markovian process, then the problem can be solved by Dynamic Programming with the state $(\mathbf{X}_t^i, \mathbf{Y}_t)$ which is numerically tractable if $\mathbf{Y}$ lives in a reasonably "small" state space.

Using a conditional expectation: $\boldsymbol{\lambda} \rightsquigarrow \mathbb{E}\big(\boldsymbol{\lambda} \bigm| \mathbf{Y}_t\big)$

Idea behind the Dual Approximate Dynamic Programming algorithm :

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

We will see that it is equivalent to

$$\sum_{i=1}^{N} \theta_t^i(\mathbf{U}_t^i) = 0 \quad \mathbb{P}-a.s. \rightsquigarrow \mathbb{E}\Big(\sum_{i=1}^{N} \theta_t^i(\mathbf{U}_t^i) \Big| \mathbf{Y}_t\Big) = 0 \quad \mathbb{P}-a.s.$$

If $(\mathbf{Y}_t)_{t \in [\![0, T-1]\!]}$ is a Markovian process, then the problem can be solved by Dynamic Programming with the state $(\mathbf{X}_t^i, \mathbf{Y}_t)$ which is numerically tractable if $\mathbf{Y}$ lives in a reasonably "small" state space.

Dual approximation as constraint relaxation 1/2

In an abstract point of view where J and Θ incorporate the dynamics of the system the original problem is

$$\begin{aligned} \min_{\mathbf{U}\in\mathcal{U}} \quad & J(\mathbf{U}) \\ s.t. \quad & \Theta(\mathbf{U}) = 0 \end{aligned}$$

where $J(\mathbf{U}) = \mathbb{E}\big(j(\mathbf{U})\big)$, which can be written as

$$\min_{\mathbf{U}\in\mathcal{U}} \max_{\boldsymbol{\lambda}} \quad J(\mathbf{U}) + \mathbb{E}\big(\langle \boldsymbol{\lambda}, \Theta(\mathbf{U})\rangle\big)$$

Dual approximation as constraint relaxation 2/2

Subsituting $\boldsymbol{\lambda}$ by $\mathbb{E}(\boldsymbol{\lambda} | \mathbf{Y})$ gives

$$\min_{\mathbf{U} \in \mathcal{U}} \max_{\boldsymbol{\lambda}} \quad J(\mathbf{U}) + \mathbb{E}\Big(\langle \mathbb{E}(\boldsymbol{\lambda} | \mathbf{Y}), \Theta(\mathbf{U}) \rangle\Big)$$

which is equal to

$$\min_{\mathbf{U} \in \mathcal{U}} \max_{\boldsymbol{\lambda}} \quad J(\mathbf{U}) + \mathbb{E}\Big(\langle \boldsymbol{\lambda}, \mathbb{E}(\Theta(\mathbf{U}) | \mathbf{Y}) \rangle\Big)$$

and thus equivalent to

$$\begin{aligned} \min_{\mathbf{U} \in \mathcal{U}} \quad & J(\mathbf{U}) \\ s.t. \quad & \mathbb{E}(\Theta(\mathbf{U}) | \mathbf{Y}) = 0 \end{aligned}$$

Contents

The approximation studied

We consider the stochastic optimization problem:

$$(\mathcal{P}) \qquad \min_{\mathbf{U}\in\mathcal{U}} \quad J(\mathbf{U}) \; ,$$
$$s.t. \quad \Theta(\mathbf{U}) \in -C \; .$$

And its approximation

$$(\mathcal{P}_n) \qquad \min_{\mathbf{U}\in\mathcal{U}} \quad J(\mathbf{U}) \; ,$$
$$s.t. \quad \mathbb{E}\big(\Theta(\mathbf{U}) \big| \mathcal{F}_n\big) \in -C \; .$$

We give convergence results of the approximation $\big(\mathcal{P}_n\big)$ toward the original problem $(\mathcal{P})$.

Contents

Kudo convergence of $\sigma-$algebras in a nutshell

A sequence $(\mathcal{F}_n)_{n\in\mathbb{N}}$ of $\sigma-$algebras Kudo-converges toward $\mathcal{F}_\infty$, iff

$$\forall X \in L^1(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}), \qquad \mathbb{E}\big(\mathbf{X} \bigm| \mathcal{F}_n\big) \longrightarrow_{L^1} \mathbb{E}\big(\mathbf{X} \bigm| \mathcal{F}_\infty\big).$$

The following result is shown by Piccinini (1998):

Proposition

Assume that the sequence of σ-algebras Kudo-converge and the sequence of random variable converges in L^p:

- $\mathcal{F}_n \to \mathcal{F}_\infty$,
- $\mathbf{X}_n \to_{L^p} \mathbf{X}$,

then

$$\mathbb{E}\big(\mathbf{X}_n \bigm| \mathcal{F}_n\big) \to_{L^p} \mathbb{E}\big(\mathbf{X} \bigm| \mathcal{F}_\infty\big).$$

Some properties of Kudo-convergence

A few properties on the Kudo-convergence of σ-algebras:

1. Kudo-convergence's topology is metrizable;
2. the set of σ-fields generated by partition of Ω is dense in the set of all σ-algebras;
3. if a sequence of random variables $(\mathbf{X}_n)_{n\in\mathbb{N}}$ converges in probability toward $\mathbf{X}$ and for all $n \in \mathbb{N}$ we have $\sigma(\mathbf{X}_n) \subset \sigma(\mathbf{X})$, then we have the Kudo-convergence of $\big(\sigma(\mathbf{X}_n)\big)_{n\in\mathbb{N}}$ toward $\sigma(\mathbf{X})$.

Painlevé-Kuratovski convergence of set

- E is a topologic space,
- $A_n \subset E$,
- $J_n : E \to \mathbb{R} \cup \{+\infty\}$.

We define outer and inner limits of sequence of subset of a topological set E.

$$\underline{\lim}_n A_n = \{x \in E \quad | \quad \forall n \in \mathbb{N}, \quad x_n \in A_n, \quad \lim_{k\to\infty} x_n = x\} ,$$

$$\overline{\lim}_n A_n = \{x \in E \quad | \quad \exists (n_k)_{k\in\mathbb{N}}, \quad \forall k \in \mathbb{N}, \quad x_{n_k} \in A_{n_k}, \quad \lim_{k\to\infty} x_{n_k} = x\} .$$

And $(A_n)_{n\in\mathbb{N}}$ converges toward A iff

$$\overline{\lim}_n A_n = \underline{\lim}_n A_n = A .$$

Epi-convergence in a nutshell

The epiconvergence of the sequence of functions $J_n : E \to \mathbb{R} \cup \{+\infty\}$ toward J is given as the convergence of their epigraphs :

$$J_n \to_e J \qquad \text{iff} \qquad \overline{\lim}_n \, epi(J_n) = \underline{\lim}_n \, epi(J_n) = epi(J) \; .$$

Epiconvergence is the right notion of convergence in optimization as epiconvergence of J_n toward J almost implies:

- the convergence of $\min J_n$ toward $\min J$,
- the convergence of $\arg\min J_n$ toward $\arg\min J$.

More information can be found in the book by Rockafellar and Wets (1995).

Contents

Convergence result

Theorem

Assume that

- the set of controls $\mathcal{U}$ is endowed with a topology τ,
- the image space of the constraint operator Θ, is $\mathcal{V} = L^p\big(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{V}\big)$, with $p \in [1, \infty)$, (strong or weak topo.),
- the cone of constraint C is such that $\mathbb{E}\big(C \bigm| \mathcal{F}_n\big) \subset C$,
- the sequence of σ-algebras $(\mathcal{F}_n)_{n\in\mathbb{N}}$ converges towards $\mathcal{F}$,
- the constraint operator $\Theta : (\mathcal{U}, \tau) \to (\mathcal{V}, \tau_{L^p})$ and the objective operator $J : (\mathcal{U}, \tau) \to \mathbb{R}$ are continuous.

Then $\tilde{J}_n$ epi-converges toward $\tilde{J}$, where

$$\tilde{J}_n(\mathbf{U}) = \begin{cases} J(\mathbf{U}) & \text{if } \mathbf{U} \text{ satisfies the constraint of } (\mathcal{P}_n), \\ +\infty & \text{otherwise} \end{cases}$$

Practical consequences

Consider a sequence of ϵ_n-solution of $(\mathcal{P}_n)$ denoted $\mathbf{U}_n$, i.e.

$$\tilde{J}_n(\mathbf{U}_n) < \inf_{\mathbf{U}\in\mathcal{U}} \tilde{J}_n(\mathbf{U}) + \varepsilon_n \ .$$

Under the conditions of the preceding theorem for every converging sub-sequence $(\mathbf{U}_{n_k})_{k\in\mathbb{N}}$, we have

$$\tilde{J}\big(\lim_k \mathbf{U}_{n_k}\big) = \min_{\mathbf{U}\in\mathcal{U}} \tilde{J}(\mathbf{U}) = \lim_k \tilde{J}_{n_k}\big(\mathbf{U}_{n_k}\big) \ .$$

Which means that

- the optimal value of the relaxed problem converges toward the optimal value of the original problem,
- and if the optimal strategies $\mathbf{U}_n$ of the relaxed problems converges, then their limit is a solution of the original problem.

Contents

Proving continuity under $\tau_{\mathbb{P}}$

Lemma

Let $\Theta : E \to F$, where $(E, \tau_{\mathbb{P}})$ is a space of random variables endowed with the topology of convergence in probability, and (F, τ) is a topological space. Assume that Θ is such that

$$\mathbf{U}_n \longrightarrow_{a.s.} \mathbf{U} \qquad \Longrightarrow \qquad \Theta(\mathbf{U}_n) \to_\tau \Theta(\mathbf{U}).$$

Then Θ is a continuous operator from $(E, \tau_{\mathbb{P}})$ into (F, τ).

Remarks :

- There is no topology $\tau_{a.s.}$.
- However $\tau_{\mathbb{P}}$ is very close to what would be $\tau_{a.s.}$.

Collection of continuous operators

j and θ are assumed to be continuous and bounded. Recall that $\mathcal{V} = L^p(\Omega, \mathcal{F}, \mathbb{P})$.

Operator	Hypothesis	$(\mathcal{U}, \tau)$
$J(\mathbf{U}) = \mathbb{E}\big(j(\mathbf{U})\big)$		$\mathcal{U} = L^0(\Omega, \mathcal{F}, \mathbb{P}; \tau_{\mathcal{L}})$
$J(\mathbf{U}) = \rho\big(j(\mathbf{U})\big)$	ρ c.r.m	$\mathcal{U} = L^0(\Omega, \mathcal{F}, \mathbb{P}; \tau_{\mathbb{P}})$
$\Theta(\mathbf{U}) = \theta(\mathbf{U})$		$\mathcal{U} = L^0(\Omega, \mathcal{F}, \mathbb{P}; \tau_{\mathbb{P}})$
$\Theta(\mathbf{U}) = \mathbb{E}\big(\mathbf{U} \bigm\vert \mathcal{B}\big) - \mathbf{U}$	$p' \geq p$	$\mathcal{U} = L^{p'}(\Omega, \mathcal{F}, \mathbb{P}; \tau_{L^{p'}})$
$\Theta(\mathbf{U}) = \rho(\mathbf{U})$	ρ cond. c.r.m	$\mathcal{U} = L^p(\Omega, \mathcal{F}, \mathbb{P}; \tau_{\mathbb{P}})$
$\Theta(\mathbf{U}) = VaR_\alpha\big(\theta(\mathbf{U})\big)$		cont. distribution, $\tau_{\mathcal{L}}$

c.r.m = convex risk measures

remark: there is no $\tau_{a.s.}$.

Conclusion

- We can apply price-decomposition methods in stochastic setting. However the subproblem have the same complexity that the original one.
- One idea is to approximate the stochastic multiplier process by its conditional expectation. This is equivalent to solve an approximate problem, where the almost sure constraint is relaxed in a conditional expectation constraint.
- We give epi-convergence results of the relaxation toward the original problem which relies on:
  - stability of cone of constraints w.r.t conditionning,
  - Kudo-convergence of the conditionning σ-algebras,
  - continuity of cost and constraint operators.
- The results are the same for a finite number of constraint operator (with different conditionning).
- Finally most cost and constraint operators found in stochastic optimization are continuous for the right topology (often $\tau_{\mathbb{P}}$).

Conclusion

- We can apply price-decomposition methods in stochastic setting. However the subproblem have the same complexity that the original one.
- One idea is to approximate the stochastic multiplier process by its conditional expectation. This is equivalent to solve an approximate problem, where the almost sure constraint is relaxed in a conditional expectation constraint.
- We give epi-convergence results of the relaxation toward the original problem which relies on:
  - stability of cone of constraints w.r.t conditionning,
  - Kudo-convergence of the conditionning σ-algebras,
  - continuity of cost and constraint operators.
- The results are the same for a finite number of constraint operator (with different conditionning).
- Finally most cost and constraint operators found in stochastic optimization are continuous for the right topology (often $\tau_{\mathbb{P}}$).