Convergence of approximations of stochastic optimization problems subject to measurability constraints.

Pierre Carpentier and SOWG[1]

ENSTA — ENPC — EDF

April 13, 2006

[1]**Systems and Optimization Working Group**:
K. Barty, J.-P. Chancelier, G. Cohen, M. de Lara, A. Dallagi & C. Strugarek

Outline of the presentation

1. Introduction
   - Problem statement
   - Strong convergence topology of σ-fields
   - A two-stage convergence result
2. Counterexample
3. Convergence theorem
4. Conclusions

Prototype Problem

Stochastic optimization problem under consideration:

$$V(\boldsymbol{\xi}, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}\big[j(\mathbf{u}, \boldsymbol{\xi})\big] \ , \quad (1a)$$

$$\text{subject to} \quad \mathbf{u} \ \ \mathcal{F}-\text{measurable} \ . \quad (1b)$$

- $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space,
- $\boldsymbol{\xi}$ a random variable on Ξ (noise),
- $\mathbf{u}$ a random variable on U (control),
- $\mathcal{F}$ a sub σ-field of $\mathcal{A}$ (observation, usually generated by $\mathbf{y}$).

Prototype Problem

Stochastic optimization problem under consideration:

$$V(\xi, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}\big[j(\mathbf{u}, \xi)\big] \ , \tag{1a}$$

$$\text{subject to} \quad \mathbf{u} \quad \mathcal{F}-\text{measurable} \ . \tag{1b}$$

$$\min \mathbb{E}\left[\sum_{t=0}^{T-1} L_{t+1}\big(\mathbf{x}_t, \mathbf{u}_t, \xi_{t+1}\big) + K\big(\mathbf{x}_T\big)\right] \ ,$$

$$\text{subject to} \quad \begin{cases} \mathbf{x}_0 & = f_0\big(\xi_0\big) \\ \mathbf{x}_{t+1} & = f_{t+1}\big(\mathbf{x}_t, \mathbf{u}_t, \xi_{t+1}\big) \end{cases} \ ,$$

$$\mathbf{u}_t \quad \sigma\big(\xi_0, \ldots, \xi_t\big)-\text{measurable} \ .$$

Prototype Problem

Stochastic optimization problem under consideration:

$$V(\boldsymbol{\xi}, \mathcal{F}) = \min_{\mathbf{u} \in L^2(\Omega, \mathcal{A}, \mathbb{P}; U)} \mathbb{E}\big[j(\mathbf{u}, \boldsymbol{\xi})\big] \ , \tag{1a}$$

$$\text{subject to} \quad \mathbf{u} \ \ \mathcal{F}-\text{measurable} \ . \tag{1b}$$

In order to obtain a tractable approximation of problem (1),

1. the expectation in (1a) must be discretized,
2. and the σ-field $\mathcal{F}$ in (1b) must be discretized.

These two discretizations are a priori independent.

The first discretization is somewhat traditional (Monte Carlo), whereas the last one is not so well-known. . .

Strong convergence topology of σ-fields (Neveu)

Coarsest topology such that conditional expectation is continuous w.r.t. the σ-field:

$$\lim_{n \to +\infty} \mathcal{F}_n = \mathcal{F} \iff \lim_{n \to +\infty} \left\| \mathbb{E}\left[f \mid \mathcal{F}_n\right] - \mathbb{E}\left[f \mid \mathcal{F}\right] \right\|_{L^1} = 0 \ \ \forall f \in L^1 .$$

Main properties of the strong topology (Cotter)

- The strong convergence topology is metrizable.
- The set of σ-fields generated by a finite partition is dense.
- If $\mathbf{y}_n \xrightarrow{\mathbb{P}} \mathbf{y}$ and $\sigma(\mathbf{y}_n) \subset \sigma(\mathbf{y})$, then $\sigma(\mathbf{y}_n) \to \sigma(\mathbf{y})$.

Based on these notions, Barty has proposed a discretization scheme in order to approximate problem (1).

Discretization scheme

1. Approximate $\mathcal{F}$ by $\mathcal{F}_k$ generated by a finite partition of Ω:

$$V\big(\xi, \mathcal{F}_k\big) = \min_{\mathbf{u}\ \mathcal{F}_k-\text{measurable}} \mathbb{E}\big[j(\mathbf{u}, \xi)\big] \ .$$

2. Approximate ξ by a finitely valued random variable ξ_n:

$$V\big(\xi_n, \mathcal{F}_k\big) = \min_{\mathbf{u}\ \mathcal{F}_k-\text{measurable}} \mathbb{E}\big[j(\mathbf{u}, \xi_n)\big] \ .$$

Convergence theorem (Barty)

1. **Information structure discretization error:**
$\big|V\big(\xi, \mathcal{F}\big) - V\big(\xi, \mathcal{F}_k\big)\big| \longrightarrow 0$ as $\mathcal{F}_k \longrightarrow \mathcal{F}$ *strongly*.
2. **Mean computation discretization error:**
$\big|V\big(\xi, \mathcal{F}_k\big) - V\big(\xi_n, \mathcal{F}_k\big)\big| \longrightarrow 0$ as $\xi_n \longrightarrow \xi$ *in distribution*.

Discretization scheme independent in ξ and $\mathcal{F}$?

Formulation

- $\mathbf{x}$ and $\mathbf{w}$: independent uniformly distributed random variables on $[-1, 1]$ (initial state and noise),
- $\mathbf{u}$: random variable based on the observation of $\mathbf{x}$ (control),
- $\mathbf{z} = \mathbf{x} + \mathbf{u} + \mathbf{w}$ (final state).
- The problem is formulated on $\big([-1,1]^2, \mathcal{B}_{[-1,1]^2}, \mu\big)$:

$$\min_{\mathbf{u}\ \sigma(\mathbf{x})-\text{measurable}} \mathbb{E}\left[\epsilon \mathbf{u}^2 + \mathbf{z}^2\right] . \qquad (2)$$

Exact resolution using dynamic programming

$$u^\sharp(x) = -\frac{x}{1+\epsilon} \quad \text{and} \quad J^\sharp = V\big((\mathbf{x},\mathbf{w}),\mathcal{F}\big) = \frac{1}{3}\left(1 + \frac{\epsilon}{1+\epsilon}\right) ,$$

with $\mathcal{F} = \mathcal{B}_{[-1,1]} \otimes \{\emptyset, [-1,1]\}$: sub σ-field generated by $\mathbf{x}$.

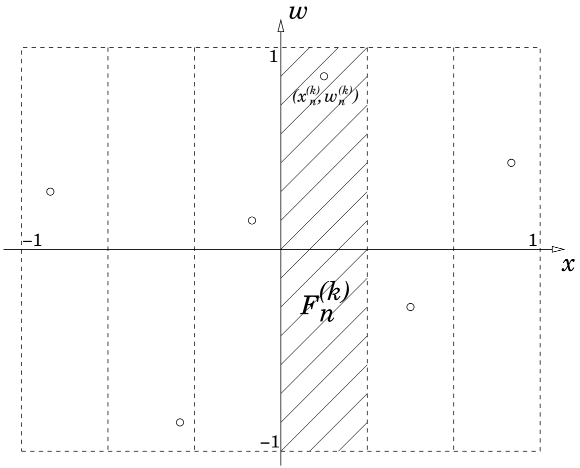


Figure: Partition of $[-1,1]^2$ and associated sample.

Information

Let $n \in \mathbb{N}^\star$. Let $\left(F_n^{(1)}, \ldots, F_n^{(n)}\right)$ be a partition of $[-1,1]^2$, with

$$F_n^{(k)} = \left(\frac{2(k-1)}{n} - 1, \frac{2k}{n} - 1\right] \times [-1,1] \, .$$

Let $\mathcal{F}_n$ be the sub σ-field generated by $\left(F_n^{(1)}, \ldots, F_n^{(n)}\right)$.

- $\left(\mathcal{F}_n\right)_{n\in\mathbb{N}}$ *strongly converges* to $\mathcal{F}$,
- $\mathbf{u}$ is $\mathcal{F}_n-$ measurable $\iff$ $\mathbf{u}$ is constant over each $F_n^{(k)}$

$$\iff \mathbf{u}(x,w) = \sum_{k=1}^{n} u_n^{(k)} \mathbb{I}_{F_n^{(k)}}(x,w) .$$

Random variable

Let $\left(\zeta_n\right)_{n\in\mathbb{N}}$ be a deterministic sequence of elements in $[-1,1]^2$ such that the associated sequence of empirical probability laws *narrowly converges* to μ. For $n \in \mathbb{N}^\star$ and $k \in \{1,\dots,n\}$, let

$$\left(x_n^{(k)}\ ,\ w_n^{(k)}\right) = \left(\frac{2k-1}{n} - 1 + \frac{\zeta_{k,1}}{n}\ ,\ \zeta_{k,2}\right)\ ,$$

and define the approximation $\left(\mathbf{x}_n, \mathbf{w}_n\right)$ of $\left(\mathbf{x}, \mathbf{w}\right)$ by

$$\left(\mathbf{x}_n, \mathbf{w}_n\right) = \sum_{k=1}^{n} \left(x_n^{(k)}, w_n^{(k)}\right) \mathbb{I}_{F_n^{(k)}}\left(\mathbf{x}, \mathbf{w}\right)\ .$$

- $\left(\mathbf{x}_n, \mathbf{w}_n\right)_{n\in\mathbb{N}}$ *converges in distribution* to $\left(\mathbf{x}, \mathbf{w}\right)$,
- $\left(\mathbf{x}_n, \mathbf{w}_n\right)$ is constant over each subset $F_n^{(k)}$.

Approximated problem

$$\min_{\left(u_n^{(1)},\ldots,u_n^{(n)}\right)\in\mathbb{R}^n} \sum_{k=1}^{n} \int_{F_n^{(k)}} \left(\epsilon\left(u_n^{(k)}\right)^2 + \left(x_n^{(k)} + u_n^{(k)} + w_n^{(k)}\right)^2\right) \mu(\mathrm{d}x\mathrm{d}w) \ .$$

Approximated solution

$$\widehat{u}_n^{(k)} = -\frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon} \quad \text{and} \quad \widehat{J}_n = V\big((\mathbf{x}_n, \mathbf{w}_n), \mathcal{F}_n\big) \longrightarrow \frac{2}{3}\left(\frac{\epsilon}{1+\epsilon}\right) \ .$$

Approximated feedback in (2)

$$\widehat{\mathbf{u}}_n(x,w) = -\sum_{k=1}^{n} \frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon} \mathbb{I}_{F_n^{(k)}}(x,w) \ \rightsquigarrow \ \mathbb{E}\left[\epsilon\widehat{\mathbf{u}}_n^2 + \mathbf{z}^2\right] \longrightarrow \frac{2}{3} \ .$$

Discretization fails to asymptotically give the optimal solution.

Same notions of convergence as in the two-stage procedure, but

- $\mathcal{F}$ and ξ are *independently* approximated:

  this "diagonal" discretization makes possible to solve each open-loop subproblem using a *unique* sample of the random variable (a poor way to compute conditional expectations).

- The convergence notion used for ξ is *weak*:

  $\left\{\left(\mathbf{x}_n, \mathbf{w}_n\right)\right\}_{n\in\mathbb{N}}$ does not converge *in probability* to $\left(\mathbf{x}, \mathbf{w}\right)$.

  Note that with an stronger convergence notion (e.g. in L^2), each open-loop subproblem will be correctly approximated.

Question: can we expect a diagonal convergence when using a stronger convergence notion for the random variable?

We go back to problem (1):

- $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and $\mathcal{F}$ a sub σ-field of $\mathcal{A}$,
- $\mathcal{X}$ is the space of integrable Ξ-valued random variables with the topology of convergence in probability $\mathbb{P}$,
- $\mathcal{U}$ is the space $L^p(\Omega, \mathcal{A}, \mathbb{P}; U)$ with $p \in [1, +\infty)$,
- $\Delta(\mathcal{F})$ is the subset of $\mathcal{F}$-measurable random variables of $\mathcal{U}$.

Let j be a normal integrand on $U \times \Xi$, and let J be the associated integral functional on $\mathcal{U} \times \mathcal{X}$: $J(\mathbf{u}, \boldsymbol{\xi}) = \mathbb{E}\big[j(\mathbf{u}, \boldsymbol{\xi})\big]$.

$$V(\boldsymbol{\xi}, \mathcal{F}) = \min_{\mathbf{u} \in \Delta(\mathcal{F})} J(\mathbf{u}, \boldsymbol{\xi}) .$$

Theorem

Under the following assumptions:

H1 J is a continuous function on $\mathcal{U} \times \mathcal{X}$,

H2 $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ strongly converges to $\mathcal{F}$,

H3 $\forall n \in \mathbb{N}$, $\mathcal{F}_n \subset \mathcal{F}$,

H4 $\{\boldsymbol{\xi}_n\}_{n\in\mathbb{N}}$ converges to $\boldsymbol{\xi}$ in probability,

H5 $\{J(\cdot, \boldsymbol{\xi}_n)\}_{n\in\mathbb{N}}$ uniformly converges to $J(\cdot, \boldsymbol{\xi})$ on $\Delta(\mathcal{F})$:

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N},\ \forall n \geq N,\ \forall \mathbf{u} \in \Delta(\mathcal{F}),\ |J(\mathbf{u}, \boldsymbol{\xi}_n) - J(\mathbf{u}, \boldsymbol{\xi})| \leq \epsilon\ ,$$

the convergence of the optimal costs holds true:

$$\lim_{n\to+\infty} V\left(\boldsymbol{\xi}_n, \mathcal{F}_n\right) = V\left(\boldsymbol{\xi}, \mathcal{F}\right)\ . \tag{3}$$

$\limsup_{n \to +\infty} V\big(\xi_n, \mathcal{F}_n\big) \leq V\big(\xi, \mathcal{F}\big)$

- $\forall \mathbf{u} \in \Delta(\mathcal{F})$, define $\mathbf{u}_n = \mathbb{E}[\mathbf{u} \mid \mathcal{F}_n]$. Then, $\mathcal{F}_n \to \mathcal{F} \implies \mathbf{u}_n \to \mathbf{u}$. The set-valued mapping Δ is thus lsc.
- J being usc, we conclude that the marginal function V is also usc.

$\liminf_{n \to +\infty} V\big(\xi_n, \mathcal{F}_n\big) \geq V\big(\xi, \mathcal{F}\big)$

- From $J(\mathbf{u}, \xi_n) = J(\mathbf{u}, \xi) + (J(\mathbf{u}, \xi_n) - J(\mathbf{u}, \xi))$, we obtain:

$$\min_{\mathbf{u} \in \Delta(\mathcal{F}_n)} J(\mathbf{u}, \xi_n) \geq \min_{\mathbf{u} \in \Delta(\mathcal{F}_n)} J(\mathbf{u}, \xi) + \min_{\mathbf{u} \in \Delta(\mathcal{F}_n)} \Big(J(\mathbf{u}, \xi_n) - J(\mathbf{u}, \xi) \Big) .$$

- Using $\mathcal{F}_n \subset \mathcal{F} \implies \Delta(\mathcal{F}_n) \subset \Delta(\mathcal{F})$, we deduce:

$$V(\xi_n, \mathcal{F}_n) \geq V(\xi, \mathcal{F}) + \min_{\mathbf{u} \in \Delta(\mathcal{F})} \Big(J(\mathbf{u}, \xi_n) - J(\mathbf{u}, \xi) \Big) .$$

- The conclusion is then a consequence of **H5**.

1. Same result with:

$$\Delta(\mathcal{F}) = \{\mathbf{u} \in \mathcal{U},\ \mathbf{u}\ \ \mathcal{F} - \text{measurable},\ \mathbf{u}(\omega) \in U^{\mathrm{ad}}\ \ \mathbb{P} - as\} \ ,$$

U^{ad} being a closed convex subset of U.

2. If $\mathcal{F}$ is generated by a Y-valued random variable $\mathbf{y}$, $\mathcal{F}_n$ may be constructed thanks to a *quantification operator* q_n on Y:

$$\mathcal{F}_n = \sigma(q_n \circ \mathbf{y}) \ .$$

3. Assumption **H5** is implied by:

$$\exists \alpha > 0, \forall u \in U,\ \forall (\xi, \xi') \in \Xi \times \Xi,\ |j(u,\xi) - j(u,\xi')| \leq \alpha \|\xi - \xi'\|_{\Xi} \ .$$

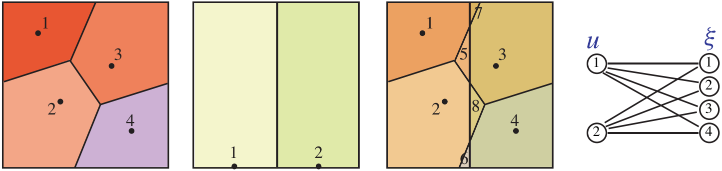
4. Theorem does not make use of the right tool for convergence analysis, namely epi-convergence.

Conclusions

1. Another sight on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.
3. Comparison with Pennanen work.

Conclusions

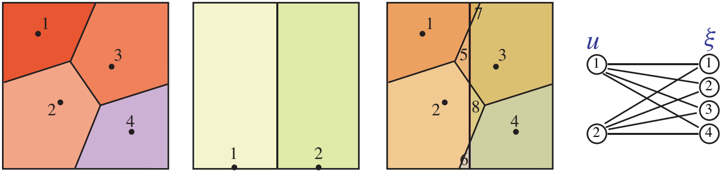
1. Another sight on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.



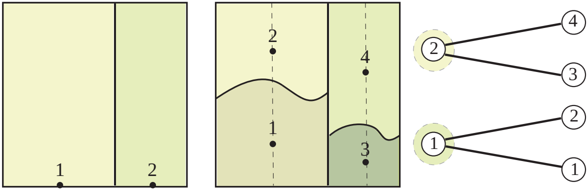
3. Comparison with Pennanen work.

Conclusions

1. Another sight on stochastic approximation.
2. Scenario trees are not built-in in stochastic programming.



3. Comparison with Pennanen work.



K. Barty
Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique.
Thèse de doctorat de l'ENPC, june 2004.

K.D. Cotter
Similarity of information and behavior with a pointwise convergence topology.
Journal of Mathematical Economics, 15, 1986.

T. Pennanen
Epi-Convergent Discretizations of Multistage Stochastic Programs.
Mathematics of Operations Research, Vol. 30, No. 1, 2005.

C. Strugarek and SOWG
On the Fortet-Mourier metric for the stability of Stochastic Optimization Problems, an example.
Stochastic Programming E-Print Series, 25, 2004.

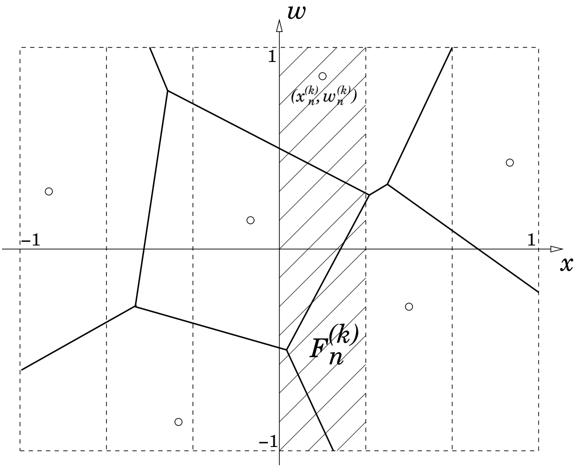


Figure: Partition of $[-1, 1]^2$ and Voronoi cells.