Rapport de recherche du CERMICS 2007-345 Avril 2007

Approximations of stochastic optimization problems subject to measurability constraints

Pierre Carpentier[1], Jean-Philippe Chancelier[2] & Michel de Lara[2]

[1] *Ecole Nationale Supérieure de Techniques Avancées, 32 Boulevard Victor, 75739 Paris Cedex 15, France*

[2] *ENPC, ParisTech, Champs sur Marne, 77455 Marne la Vallée Cedex 2, France*

CERMICS — ENPC
6 et 8 avenue Blaise Pascal
Cité Descartes - Champs sur Marne
77455 Marne la Vallée Cedex 2

http://cermics.enpc.fr

Approximations of stochastic optimization problems subject to measurability constraints

Pierre Carpentier
École Nationale Supérieure de Techniques Avancées, 32 Boulevard Victor, 75739 Paris Cedex 15, France
email: Pierre.Carpentier@ensta.fr

Jean-Philippe Chancelier
École Nationale des Ponts et Chaussées, 6 et 8 avenue Blaise Pascal, 77455, Marne la Vallée, Cedex, France
email: chancelier@cermics.enpc.fr

Michel De Lara
École Nationale des Ponts et Chaussées, 6 et 8 avenue Blaise Pascal, 77455, Marne la Vallée, Cedex, France
email: delara@cermics.enpc.fr

Motivated by the numerical resolution of stochastic optimization problems subject to measurability constraints, we focus upon the issue of how to discretize the components arising in the problem formulation. By means of a counterexample based on Monte Carlo approximation, we emphasize the importance of independent discretization of, on the one side, the random variable modelling uncertainties (noise) and, on the other side, the σ-field modelling the knowledge (information). Then, we present conditions under which the discretized problems converge to the original one. The focus is put on the probabilistic convergence notions ensuring the convergence.

Key words: stochastic programming; measurability constraints; discretization

MSC2000 Subject Classification: Primary: 90C15, 49M25, 62L20

OR/MS subject classification: Primary: Programming/Stochastic

History: Received: ; Revised: .

---

1. Introduction Let $\big(\Omega, \mathcal{A}, \mathbb{P}\big)$ be a probability space, and let $\big(\Xi, \mathcal{B}_\Xi\big)$ and $\big(U, \mathcal{B}_U\big)$ be $\mathbb{R}^n$ and $\mathbb{R}^p$ with their associated Borel σ-fields. Given a random variable $\boldsymbol{\xi}$ with values in Ξ and a subfield $\mathcal{F}$ of $\mathcal{A}$, which respectively represent the noise and the observation, we are concerned with the following stochastic optimization problem:

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

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

Here $j : U \times \Xi \to \mathbb{R}$, and $\mathbb{E}$ is the mathematical expectation under probability $\mathbb{P}$.

REMARK 1.1 *Problem (1) can be easily extended to the sequential control case with direct observation of the noises. Then,* $\boldsymbol{u} = (\boldsymbol{u_0}, \ldots, \boldsymbol{u_{T-1}})$*, each* $\boldsymbol{u_t}$ *being measurable with respect to the* σ*-field generated by the noises prior to* t*. Practical instances are multi-stage stochastic programming problems:*

$$\min_{(\boldsymbol{u}, \boldsymbol{x})} \mathbb{E}\left[\sum_{t=0}^{T-1} L_{t+1}\big(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\xi}_{t+1}\big) + K\big(\boldsymbol{x}_T\big)\right] \,, \tag{2a}$$

$$\textit{subject to} \quad \left\{ \begin{array}{ll} \boldsymbol{x}_0 & = f_0\big(\boldsymbol{\xi}_0\big) \\ \boldsymbol{x}_{t+1} & = f_{t+1}\big(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\xi}_{t+1}\big) \end{array} \right. \,, \tag{2b}$$

$$\boldsymbol{u}_t \textit{ is } \sigma\big(\boldsymbol{\xi}_0, \ldots, \boldsymbol{\xi}_t\big) - \textit{measurable} \,. \tag{2c}$$

Two polar cases are worth mentioning.

- The *full information* case corresponds to $\mathcal{F} = \mathcal{A}$. In this case, under technical assumptions ensuring the interchange of minimization and expectation (see [10, Theorem 14.60]), problem (1) becomes
$$\mathbb{E}\big[\min_{u \in U} j(u, \boldsymbol{\xi})\big] \,.$$
Once the noise $\boldsymbol{\xi}$ is discretized by a random variable $\boldsymbol{\xi}_n$, the approximate solution is given by $\min_{u \in U} j(u, \boldsymbol{\xi}_n)$, which returns a $\boldsymbol{\xi}_n$-measurable solution, thus an $\mathcal{A}$-measurable one.

- The *open loop* case arises when $\mathcal{F} = \{\emptyset, \Omega\}$. In this case, the problem is of deterministic nature provided that $\mathbb{E}\big[j(u, \boldsymbol{\xi})\big]$ and its gradient are readily available for each $u \in U$. Otherwise, the stochastic gradient method (see [9]) gives the optimal solution using samples of $\boldsymbol{\xi}$.

In these two cases, one has to deal with only *one* stochastic approximation. However, in the general case, *two* different components of the problem have to be taken into account in order to discretize (1):

(i) the σ-field $\mathcal{F}$ in (1b) must be approximated in order to deal with tractable constraints,

(ii) the expectation in (1a) must be approximated in order to be computable.

Observe that these two points are rather independent, in the sense that there is no reason for one of these approximations to be deduced from the other. The last point, related to the convergence of measures and random variables, is somewhat "traditional" in probability theory, whereas the first point is not so well-known.

Let us recall some results about the space $\mathcal{A}^\star$ of the sub σ-fields of $\mathcal{A}$ (see [6] and [5] for further details). The *strong convergence* topology on $\mathcal{A}^\star$ is the coarsest topology such that the conditional expectation is continuous with respect to the σ-field:

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

Note that this definition depends on the probability $\mathbb{P}$. The main properties of $\mathcal{A}^\star$ equipped with the strong convergence topology are the following.

P_1 The strong convergence topology on $\mathcal{A}^\star$ is metrizable.

P_2 The set of σ-fields generated by a finite partition of Ω is dense in $\mathcal{A}^\star$.

P_3 If $\boldsymbol{y}_n \longrightarrow \boldsymbol{y}$ in probability and $\sigma\big(\boldsymbol{y}_n\big) \subset \sigma\big(\boldsymbol{y}\big)$, then $\sigma\big(\boldsymbol{y}_n\big)$ strongly converges to $\sigma\big(\boldsymbol{y}\big)$.

According to [8, Theorem 2.3.1], the notion of strong convergence of σ-fields, given using $L^1(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R})$, can be equivalently defined using $L^r(\Omega, \mathcal{F}, \mathbb{P}; U)$, for $r \in [1, +\infty)$.

PROPOSITION 1.1 *Let $r \in [1, +\infty)$. The two following statements are equivalent.*

- $\lim_{n\to+\infty} \mathcal{F}_n = \mathcal{F}$.
- $\forall f \in L^r(\Omega, \mathcal{A}, \mathbb{P}; U),\ \lim_{n\to+\infty} \|\mathbb{E}\left[f \mid \mathcal{F}_n\right] - \mathbb{E}\left[f \mid \mathcal{F}\right]\|_{L^r} = 0$.

In this paper, we aim at proving a convergence result for approximations of problem (1). For this, we first illustrate by an example in what a naive approach ignoring the measurability constraint specificity may lead to sub-optimality (§ 2). In § 3, we present a convergence result. Contrarily to the two polar cases of full or null information, the functional $J\big(\boldsymbol{u}, \boldsymbol{\xi}\big) := \mathbb{E}\big[j(\boldsymbol{u}, \boldsymbol{\xi})\big]$ now plays a central role. The continuity of J turns out to be crucial for convergence, and is related to the convergence notions used in the approximation. Ultimately, we review in § 4 the convergence results obtained in [2] and [7] about the same problem.

2. Counterexample We present here an example illustrating how convergence notions matter in order to accurately discretize problem (1). This example has already been used to show that the Fortet-Mourier metric is not a suitable tool when discretizing a stochastic optimal control problem (see [12] or [11]).

2.1 Formulation and exact solution We consider a two-stage dynamical system whose initial state $\boldsymbol{x}$ is a random variable on $[-1, 1]$ with uniform distribution. The final state of the system is defined as

$$\boldsymbol{z} := \boldsymbol{x} + \boldsymbol{u} + \boldsymbol{w}\ , \tag{3a}$$

$\boldsymbol{w}$ being another uniformly distributed random variable on $[-1, 1]$ independent of $\boldsymbol{x}$ and the control $\boldsymbol{u}$ being a random variable measurable with respect to the initial state $\boldsymbol{x}$. Let $\epsilon > 0$ and consider the following problem:

$$\min_{\boldsymbol{u} \text{ is } \sigma(\boldsymbol{x})-\text{measurable}} \mathbb{E}\left[\epsilon \boldsymbol{u}^2 + \boldsymbol{z}^2\right]\ . \tag{3b}$$

The probability space associated with problem (3) is $\big([-1,1]^2, \mathcal{B}_{[-1,1]^2}, \mu\big)$, where $\mathcal{B}_{[-1,1]^2}$ is the Borel σ-field on $[-1,1]^2$ and μ is the product of two independent uniform probability laws on $[-1,1]$. The random variables $\boldsymbol{x}$ and $\boldsymbol{w}$ are the two components of the identity application $\mathrm{Id}_{[-1,1]^2}$ on $[-1,1]^2$, the real valued control variable $\boldsymbol{u}$ being defined on $[-1,1]^2$. Problem (3) is thus equivalent to:

$$\min_{\boldsymbol{u} \text{ is } \sigma(\boldsymbol{x})-\text{measurable}} \int_{[-1,1]^2} \Big(\epsilon\big(\boldsymbol{u}(x,w)\big)^2 + \big(x + \boldsymbol{u}(x,w) + w\big)^2\Big)\mu\big(\mathrm{d}x\mathrm{d}w\big) \ . \tag{4}$$

This problem is a Markovian stochastic optimal control problem which can be solved using dynamic programming. Introducing the Bellman functions

$$V_1(z) := z^2 \ \ , \ \ V_0(x) := \min_{u\in\mathbb{R}} \mathbb{E}\left[\epsilon u^2 + V_1(x + u + \boldsymbol{w})\right] \ ,$$

we obtain the optimal feedback law $u^\sharp$ and the associated optimal cost $J^\sharp$ $(:= \mathbb{E}\left[V_0(\boldsymbol{x})\right])$:

$$u^\sharp(x) = -\frac{x}{1+\epsilon} \ \ , \ \ J^\sharp = \frac{1}{3}\left(1 + \frac{\epsilon}{1+\epsilon}\right) \ . \tag{5}$$

2.2 Discretization Let $\big(\zeta_n\big)_{n\in\mathbb{N}^*}$ be a deterministic sequence of elements in $[-1,1]^2$, with $\zeta_n = \big(\zeta_{n,1}, \zeta_{n,2}\big)$. We make the assumption that the sequence $\big(\mu_n\big)_{n\in\mathbb{N}^*}$ of empirical probability laws associated with $\big(\zeta_n\big)_{n\in\mathbb{N}^*}$, that is

$$\mu_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{\zeta_k} \ ,$$

weakly converges to the probability measure μ (see [3]).

REMARK 2.1 *Such a sequence $\big(\zeta_n\big)_{n\in\mathbb{N}^*}$ is usually obtained as the realization of a sequence $\big(\boldsymbol{\zeta}_n\big)_{n\in\mathbb{N}^*}$ of i.i.d. random variables on $[-1,1]^2$ with law μ. The weak convergence assumption is then, almost surely, a consequence of the Glivenko-Cantelli theorem.*

Let $n \in \mathbb{N}^\star$; for any $k \in \{1,\dots,n\}$, we define

$$\big(x_n^{(k)} \ , \ w_n^{(k)}\big) := \left(\frac{2k-1}{n} - 1 + \frac{\zeta_{k,1}}{n} \ , \ \zeta_{k,2}\right) \ , \tag{6}$$

and

$$I_n^{(k)} := \left(\frac{2k-2}{n} - 1, \frac{2k}{n} - 1\right] \quad , \quad F_n^{(k)} := I_n^{(k)} \times [-1,1] \, . \tag{7}$$

By construction, $\big(F_n^{(1)}, \dots, F_n^{(n)}\big)$ is a partition of $[-1,1]^2$, made of vertical stripes as in Figure 1, and $\big(x_n^{(k)}, w_n^{(k)}\big) \in F_n^{(k)}$ $\forall k \in \{1,\dots,n\}$.

We are now ready to discretize problem (4).

Random variable. Let $q_n : [-1,1]^2 \to [-1,1]^2$ be the function defined by

$$q_n\big(x,w\big) := \sum_{k=1}^{n} \big(x_n^{(k)}, w_n^{(k)}\big)\mathbb{I}_{F_n^{(k)}}\big(x,w\big) \ ,$$

that is $q_n\big(x,w\big) = \big(x_n^{(k)}, w_n^{(k)}\big)$ if $\big(x,w\big) \in F_n^{(k)}$. We then define the sequence $\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big)_{n\in\mathbb{N}^*}$ of random variables by

$$\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big) := q_n\big(\boldsymbol{x}, \boldsymbol{w}\big) \ . \tag{8}$$

According to this definition, the discretized random variable $\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big)$ is constant over each subset $F_n^{(k)}$.

LEMMA 2.1 *The sequence $\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big)_{n\in\mathbb{N}^*}$* converges in distribution *to $\big(\boldsymbol{x}, \boldsymbol{w}\big)$ as $n \to +\infty$.*

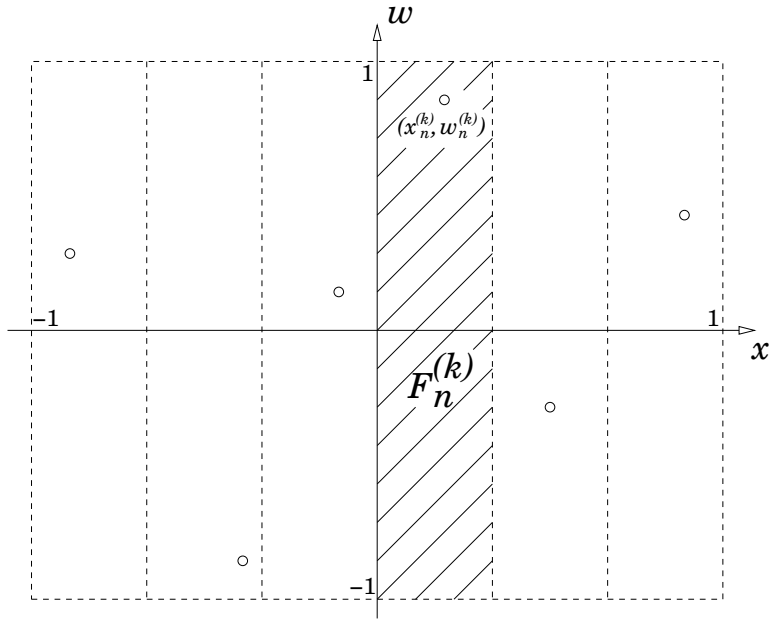

Figure 1: Partition of $[-1,1]^2$ and associated sample.

PROOF. Consider the empirical distribution function F_n of $\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big)$:

$$F_n\big(x,w\big) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_{[-1,x]\times[-1,w]}\big(x_n^{(k)}, w_n^{(k)}\big) \ .$$

For given $x \in [-1,1]$ and $n \in \mathbb{N}^*$, let k_0 be the index such that $x \in I_n^{(k_0)}$ (see (7)) and let ν_0 be equal to 0 if $x \leq x_n^{(k_0)}$ and equal to 1 otherwise. Then,

$$\begin{aligned} F_n\big(x,w\big) &= \frac{1}{n}\sum_{k=1}^{k_0-1} \mathbb{I}_{[-1,w]}\big(w_n^{(k)}\big) + \frac{\nu_0}{n}\mathbb{I}_{[-1,w]}\big(w_n^{(k_0)}\big) \\ &= \frac{k_0-1}{n}\left(\frac{1}{k_0-1}\sum_{k=1}^{k_0-1} \mathbb{I}_{[-1,w]}\big(w_n^{(k)}\big)\right) + \frac{\nu_0}{n}\mathbb{I}_{[-1,w]}\big(w_n^{(k_0)}\big) \ . \end{aligned}$$

The index k_0 goes to infinity as n goes to infinity (for any $x > -1$). We thus conclude that $F_n\big(x,w\big)$ converges to $F\big(x,w\big) = \frac{(1+x)(1+w)}{4}$, the distribution function of μ, the uniform probability on the square $[-1,1]^2$. □

Information. Since $\boldsymbol{x}$ is the first component of $\mathrm{Id}_{[-1,1]^2}$, the sub σ-field $\sigma(\boldsymbol{x})$ of $\mathcal{B}_{[-1,1]^2}$ generated by the random variable $\boldsymbol{x}$ is

$$\mathcal{F} = \mathcal{B}_{[-1,1]} \otimes \{\emptyset, [-1,1]\} \ .$$

For a given $n \in \mathbb{N}^\star$, we approximate $\mathcal{F}$ by the σ-field $\mathcal{F}_n$ generated by the partition $\big(F_n^{(1)}, \ldots, F_n^{(n)}\big)$:

$$\mathcal{F}_n = \sigma\big(F_n^{(1)}, \ldots, F_n^{(n)}\big) \ . \tag{9}$$

Note that the approximated information constraint "$\boldsymbol{u}$ is $\mathcal{F}_n-$measurable" is equivalent to "$\boldsymbol{u}$ is constant over each subset $F_n^{(k)}$", that is constant on each vertical stripe of Figure 1. Such a control variable $\boldsymbol{u}$ is thus parameterized by the values $u_n^{(k)}$ taken on each subset $F_n^{(k)}$:

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

Notice that $\mathbb{I}_{F_n^{(k)}}\big(x,w\big)$ does not depend upon w, and therefore $\boldsymbol{u}\big(x,w\big)$ depends only upon x.

LEMMA 2.2 *The sequence* $\big(\mathcal{F}_n\big)_{n\in\mathbb{N}^*}$ strongly converges *to* $\mathcal{F}$ *as* $n \to +\infty$.

PROOF. Since the values $x_n^{(k)}$, $k = 1,\dots, n$, defined by (6) and taken by the random variable $\boldsymbol{x}_n$ are two by two distinct, then $\mathcal{F}_n = \sigma\big(\boldsymbol{x}_n\big)$. Following Property $\boldsymbol{P_3}$ (page 2), it is sufficient to show that $\boldsymbol{x}_n \longrightarrow \boldsymbol{x}$ in probability. This last convergence is obvious from the definition of $x_n^{(k)}$. □

2.3 Approximated solution Approximating problem (4) consists in replacing $\mathcal{F}$ and $\big(\boldsymbol{x}, \boldsymbol{w}\big)$ by their discretized versions $\mathcal{F}_n$ and $\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big)$. The resulting function to be minimized is constant over each F_n^k, and the approximated problem is written accordingly

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

which is equivalent to:

$$\min_{\left(u_n^{(1)},\dots,u_n^{(n)}\right)\in\mathbb{R}^n} \ \sum_{k=1}^{n} \Big(\epsilon\big(u_n^{(k)}\big)^2 + \big(x_n^{(k)} + u_n^{(k)} + w_n^{(k)}\big)^2\Big)\ . \tag{10}$$

Given $n \in \mathbb{N}^*$, each $u_n^{(k)}$, $k = 1,\dots,n$, in problem (10) can be optimized separately. The optimal values are the solutions of a quadratic minimization problem, yielding

$$\widehat{u}_n^{(k)} = -\frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon}\ ,$$

and the associated optimal control variable $\widehat{\boldsymbol{u}}_{\boldsymbol{n}}$ is:

$$\widehat{\boldsymbol{u}}_{\boldsymbol{n}}\big(x, w\big) = -\sum_{k=1}^{n} \frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon}\mathbb{I}_{F_n^{(k)}}\big(x, w\big)\ .$$

To evaluate the quality of the approximation, let us compute the cost $\widehat{J}_n$ obtained by plugging the approximated optimal control variable $\widehat{\boldsymbol{u}}_{\boldsymbol{n}}$ into the original cost:

$$\widehat{J}_n := \mathbb{E}\left[\epsilon\widehat{\boldsymbol{u}}_{\boldsymbol{n}}{}^2 + \big(\boldsymbol{x} + \widehat{\boldsymbol{u}}_{\boldsymbol{n}} + \boldsymbol{w}\big)^2\right]. \tag{11}$$

LEMMA 2.3 *The sequence* $\big(\widehat{J}_n\big)_{n\in\mathbb{N}}$ *is such that:* $\lim\limits_{n\to+\infty} \widehat{J}_n = \dfrac{2}{3}$.

PROOF. We have:

$$\begin{aligned}
\widehat{J}_n &= \int_{[-1,1]^2} \Big(\epsilon\big(\widehat{\boldsymbol{u}}_{\boldsymbol{n}}(x,w)\big)^2 + \big(x + \widehat{\boldsymbol{u}}_{\boldsymbol{n}}(x,w) + w\big)^2\Big)\mu\big(\mathrm{d}x\mathrm{d}w\big) \\
&= \sum_{k=1}^{n} \int_{F_n^{(k)}} \left(\epsilon\left(\frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon}\right)^2 + \left(x + w - \frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon}\right)^2\right)\mu\big(\mathrm{d}x\mathrm{d}w\big)\ .
\end{aligned}$$

Developing the last quadratic term in the previous expression leads to:

$$\begin{aligned}
\widehat{J}_n &= \frac{2}{3} + \frac{1}{n}\sum_{k=1}^{n} \frac{\big(x_n^{(k)} + w_n^{(k)}\big)^2}{1+\epsilon} - 2\sum_{k=1}^{n}\left(\frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon}\right)\int_{F_n^{(k)}} \big(x + w\big)\mu\big(\mathrm{d}x\mathrm{d}w\big) \\
&= \frac{2}{3} + \frac{1}{n}\sum_{k=1}^{n} \frac{\big(x_n^{(k)} + w_n^{(k)}\big)^2}{1+\epsilon} - \frac{2}{n}\sum_{k=1}^{n}\left(\frac{2k-1}{n} - 1\right)\left(\frac{x_n^{(k)} + w_n^{(k)}}{1+\epsilon}\right)\ .
\end{aligned}$$

Using the convergence in distribution of $\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big)_{n\in\mathbb{N}^*}$ and the fact that $\left|x_n^{(k)} - \big(\frac{2k-1}{n} - 1\big)\right| \le \frac{1}{n}$ holds for every k, we obtain by (6):

$$\lim_{n\to+\infty} \widehat{J}_n = \frac{2}{3}\ .$$

□

By definition, $\widehat{J}_n$ is the "true" cost associated with the control variable $\widehat{\boldsymbol{u}}_{\boldsymbol{n}}$. It is thus comparable with the optimal cost $J^\sharp$ given by (5). Taking $\epsilon \approx 0$, we obtain $J^\sharp \approx \frac{1}{3}$, whereas the limit given by Lemma 2.3 is equal to $\frac{2}{3}$ (this limit does not depend on ϵ). **We conclude that the proposed discretization fails to asymptotically give the optimal solution of the problem**.

REMARK 2.2 *It is easy to verify that the optimal cost $\widetilde{J}_n$ of problem (10) is such that:*

$$\lim_{n\to+\infty} \widetilde{J}_n = \frac{2}{3}\left(\frac{\epsilon}{1+\epsilon}\right) .$$

Once again, this cost is different from $J^\sharp$.[1] *Note however that $\widetilde{J}_n$ and $J^\sharp$ are not easily comparable, because the former does not correspond to the evaluation of an admissible control variable for problem (3).*

2.4 What has gone wrong? We have used the notion of strong convergence for σ-fields and the notion of convergence in distribution for random variables, and we have *simultaneously* discretized the random variable and the σ-field. As illustrated by the counterexample, this "diagonal" discretization procedure makes possible to solve each open-loop subproblem using a *unique* sample of the random variable: this poor way to compute the underlying conditional expectations explains the convergence failure.

REMARK 2.3 *Note however that the convergence notions used here may lead to a positive convergence result if the approximations of $\mathcal{F}$ and $\boldsymbol{\xi}$ are implemented in a nested manner (see § 4).*

A feature of this example is the weakness of the convergence notion used for the random variable. If the sequence $\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big)_{n\in\mathbb{N}^*}$ defined by (8) converges in distribution to $\big(\boldsymbol{x}, \boldsymbol{w}\big)$, it is clear that *the convergence does not hold in probability*. Indeed, let $\tau > 0$ be given. Consider the norm $\|(x,w)\| = \sup\{|x|, |w|\}$ on $[-1,1]^2$, and let A_n be the subset of $[-1,1]^2$ defined by:

$$A_n := \big\{(x,w) \in [-1,1]^2 \ , \ \ \big\|\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big)(x,w) - (x,w)\big\| \le \tau\big\} \ .$$

The subset A_n is expressed as the disjoint union of n subsets $A_n^{(k)}$, with

$$A_n^{(k)} := A_n \cap F_n^{(k)} = \Big\{(x,w) \in F_n^{(k)} \ , \ \ \sup\Big\{\Big|x_n^{(k)} - x\Big|, \Big|w_n^{(k)} - w\Big|\Big\} \le \tau\Big\} \ .$$

From the definition of $F_n^{(k)}$ and A_n, the subset $A_n^{(k)}$ is included in a $\frac{2}{n} \times 2\tau$ rectangle. We thus obtain $\mu\big(A_n^{(k)}\big) \le \frac{\tau}{n}$, and then $\mu\big(A_n\big) \le \tau$ by summation. This demonstrates that

$$\mu\left(\big\|\big(\boldsymbol{x}_n, \boldsymbol{w}_n\big) - \big(\boldsymbol{x}, \boldsymbol{w}\big)\big\| > \tau\right) \ge 1 - \tau \ .$$

A quite natural question arising at this point is: can we expect a diagonal convergence if we use an stronger convergence notion for the random variables? We shall give a positive answer to this question in the next section.

3. Convergence theorem We go back to the initial problem (1). The framework of the study is the following.

- The underlying probability space is $\big(\Omega, \mathcal{A}, \mathbb{P}\big)$, and we consider $\mathcal{F}$ a sub σ-field of $\mathcal{A}$,
- The control variable $\boldsymbol{u}$ belongs to the subset $\Delta\big(\mathcal{F}\big)$ of the $\mathcal{F}-$measurable random variables:

  $$\Delta(\mathcal{F}) := \{\boldsymbol{u} \in L^r(\Omega, \mathcal{A}, \mathbb{P}; U), \ \boldsymbol{u} \text{ is } \mathcal{F}-\text{measurable}\} = L^r(\Omega, \mathcal{F}, \mathbb{P}; U) \ .$$

  Here, $1 \le r < +\infty$ and $L^r(\Omega, \mathcal{A}, \mathbb{P}; U)$ is equipped with the strong topology.
- The random variable $\boldsymbol{\xi}$ belongs to $L^q(\Omega, \mathcal{A}, \mathbb{P}; \Xi)$, where $1 \le q < +\infty$, equipped with the strong topology.
- The cost function J, defined on $L^r(\Omega, \mathcal{A}, \mathbb{P}; U) \times L^q(\Omega, \mathcal{A}, \mathbb{P}; \Xi)$, is given by:

  $$J\big(\boldsymbol{u}, \boldsymbol{\xi}\big) := \mathbb{E}\big[j(\boldsymbol{u}, \boldsymbol{\xi})\big] \ .$$

  Here j is a normal integrand on $U \times \Xi$, J being the associated integral functional (see[10, Chapter 14]).

[1] Consider the case $\epsilon \approx 0$.

Using these notations, we want to compute the optimal value $V(\boldsymbol{\xi}, \mathcal{F})$ of problem (1):

$$V(\boldsymbol{\xi}, \mathcal{F}) := \min_{\boldsymbol{u} \in \Delta(\mathcal{F})} J(\boldsymbol{u}, \boldsymbol{\xi}) \ . \tag{12}$$

REMARK 3.1 *There is no additional difficulty in incorporating in $\Delta(\mathcal{F})$ pointwise constraints such as $\boldsymbol{u}(\omega) \in U^{\mathrm{ad}}$ $\mathbb{P}$-a.s., U^{ad} being a closed convex set of U.*

To approximate problem (12), we choose a sequence $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ of sub σ-fields of $\mathcal{A}$ and a sequence $\{\boldsymbol{\xi}_n\}_{n\in\mathbb{N}}$ of random variables in $L^q(\Omega, \mathcal{A}, \mathbb{P}; \Xi)$, and we consider the approximated problem:

$$V(\boldsymbol{\xi}_n, \mathcal{F}_n) := \min_{\boldsymbol{u} \in \Delta(\mathcal{F}_n)} J(\boldsymbol{u}, \boldsymbol{\xi}_n) \ . \tag{13}$$

Here is a result which emphasizes the role of adequate probabilistic convergences, with rather strong assumptions on the criterion. Weaker assumptions may be found in a companion paper [4].

THEOREM 3.1

Under the following assumptions:

- $\boldsymbol{H_1}$ *the sequence $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ strongly converges to $\mathcal{F}$, and $\mathcal{F}_n \subset \mathcal{F}$,*
- $\boldsymbol{H_2}$ *the sequence $\{\boldsymbol{\xi}_n\}_{n\in\mathbb{N}}$ converges to $\boldsymbol{\xi}$ in $L^q(\Omega, \mathcal{A}, \mathbb{P}; \Xi)$,*
- $\boldsymbol{H_3}$ *the normal integrand j is such that:*

$$\forall (u, u') \in U^2, \ \forall (\xi, \xi') \in \Xi^2, \ \left| j(u, \xi) - j(u', \xi') \right| \leq \alpha \left\| u - u' \right\|_U^r + \beta \left\| \xi - \xi' \right\|_\Xi^q \ ,$$

the convergence of the approximated optimal costs holds true:

$$\lim_{n \to +\infty} V(\boldsymbol{\xi}_n, \mathcal{F}_n) = V(\boldsymbol{\xi}, \mathcal{F}) \ . \tag{14}$$

PROOF.

Step 1: $\limsup_{n\to+\infty} V(\boldsymbol{\xi}_n, \mathcal{F}_n) \leq V(\boldsymbol{\xi}, \mathcal{F})$.

For any $\boldsymbol{u} \in \Delta(\mathcal{F})$, we define $\boldsymbol{u}_n = \mathbb{E}[\boldsymbol{u} \mid \mathcal{F}_n]$. Note that $\boldsymbol{u} = \mathbb{E}[\boldsymbol{u} \mid \mathcal{F}]$ $\mathbb{P}$- almost surely. Using Assumption $\boldsymbol{H_1}$ and Proposition 1.1, we obtain the convergence of the sequence $\{\boldsymbol{u}_n\}_{n\in\mathbb{N}}$ to $\boldsymbol{u}$ in $L^r(\Omega, \mathcal{A}, \mathbb{P}; U)$. This implies that the set-valued mapping Δ is lower semicontinuous ([1, Definition 1.4.2]).

From $\boldsymbol{H_3}$, we then deduce that the integral functional J is continuous and therefore upper semicontinuous.

Using [1, Theorem 1.4.16], we conclude that the marginal function V is also upper semicontinuous:

$$\limsup_{n\to+\infty} V(\boldsymbol{\xi}_n, \mathcal{F}_n) \leq V(\boldsymbol{\xi}, \mathcal{F}) \ . \tag{15}$$

Step 2: $\liminf_{n\to+\infty} V(\boldsymbol{\xi}_n, \mathcal{F}_n) \geq V(\boldsymbol{\xi}, \mathcal{F})$.

Starting from:

$$J(\boldsymbol{u}, \boldsymbol{\xi}_n) = J(\boldsymbol{u}, \boldsymbol{\xi}) + \left(J(\boldsymbol{u}, \boldsymbol{\xi}_n) - J(\boldsymbol{u}, \boldsymbol{\xi}) \right) \ ,$$

we obtain by minimization over $\Delta(\mathcal{F}_n)$:

$$\begin{aligned} \min_{\boldsymbol{u} \in \Delta(\mathcal{F}_n)} J(\boldsymbol{u}, \boldsymbol{\xi}_n) &\geq \min_{\boldsymbol{u} \in \Delta(\mathcal{F}_n)} J(\boldsymbol{u}, \boldsymbol{\xi}) + \min_{\boldsymbol{u} \in \Delta(\mathcal{F}_n)} \Big(J(\boldsymbol{u}, \boldsymbol{\xi}_n) - J(\boldsymbol{u}, \boldsymbol{\xi}) \Big) \\ &\geq \min_{\boldsymbol{u} \in \Delta(\mathcal{F})} J(\boldsymbol{u}, \boldsymbol{\xi}) + \min_{\boldsymbol{u} \in \Delta(\mathcal{F})} \Big(J(\boldsymbol{u}, \boldsymbol{\xi}_n) - J(\boldsymbol{u}, \boldsymbol{\xi}) \Big) \ , \end{aligned}$$

the last inequality being true because $\mathcal{F}_n \subset \mathcal{F}$ implies $\Delta(\mathcal{F}_n) \subset \Delta(\mathcal{F})$. We thus obtain:

$$V(\boldsymbol{\xi}_n, \mathcal{F}_n) \geq V(\boldsymbol{\xi}, \mathcal{F}) + \min_{\boldsymbol{u} \in \Delta(\mathcal{F})} \Big(J(\boldsymbol{u}, \boldsymbol{\xi}_n) - J(\boldsymbol{u}, \boldsymbol{\xi}) \Big) \, . \tag{16}$$

From assumptions $\boldsymbol{H_2}$ and $\boldsymbol{H_3}$, the last term in (16) converges to 0 as n goes to infinity, which proves that V is lower semicontinuous:

$$\liminf_{n \to +\infty} V(\boldsymbol{\xi}_n, \mathcal{F}_n) \geq V(\boldsymbol{\xi}, \mathcal{F}) \, . \tag{17}$$

Gathering (15) and (17) leads to the result. □

The assumptions made in Theorem 3.1 are far from being minimal. More specifically, the continuity assumption $\boldsymbol{H_3}$ on j is rather restrictive and can be alleviated using the tools of *epi-convergence* (see [4]). Nevertheless, the important point is that, to the difference of evaluating an expectation by the Monte Carlo method, a continuity property on J cannot be obtained by the *convergence in distribution*. The discretization process requires a stronger convergence notion, as shown by the following example:

- $\big(\Omega, \mathcal{A}, \mathbb{P}\big) = \big([-1,1], \mathcal{B}_{[-1,1]}, \mu\big)$, μ being the uniform law on $[-1,1]$, and $U = \Xi = \Omega$,
- $j(u, \xi) = u\xi$,
- $\boldsymbol{\xi}_n(\omega) = \left\{ \begin{array}{ll} (-1)^n & \text{if } \omega \geq 0 \\ (-1)^{n+1} & \text{otherwise} \end{array} \right.$, $\boldsymbol{u}(\omega) = \left\{ \begin{array}{ll} +1 & \text{if } \omega \geq 0 \\ -1 & \text{otherwise.} \end{array} \right.$

Being stationary in law, the sequence $\{\boldsymbol{\xi}_n\}_{n \in \mathbb{N}}$ is converging in distribution, whereas $J\big(\boldsymbol{u}, \boldsymbol{\xi}_n\big) = (-1)^n$.

From the numerical point of view, problem (13) is a tractable approximation of problem (12) provided that the range of $\boldsymbol{\xi}_n$ is finite and that $\mathcal{F}_n$ is generated by a finite partition of Ω. Indeed, let

- $\big(\Omega_n^{(1)}, \ldots, \Omega_n^{(n)}\big)$ be a partition of Ω generating the σ-field $\mathcal{F}_n$, $u_n^{(i)}$ denoting the (constant) value of a $\mathcal{F}_n$-measurable control $\boldsymbol{u}$ on the subset $\Omega_n^{(i)}$,
- $\big(\mho_n^{(1)}, \ldots, \mho_n^{(n)}\big)$ be a partition of Ω generated by $\boldsymbol{\xi}_n$, $\xi_n^{(l)}$ denoting the (constant) value of the random variable $\boldsymbol{\xi}_n$ on the subset $\mho_n^{(l)}$.

Problem (13) is then equivalent to:

$$\min_{(u_n^{(1)}, \ldots, u_n^{(n)}) \in U^n} \sum_{i=1}^{n} \sum_{l=1}^{n} \mathbb{P}\big(\Omega_n^{(i)} \cap \mho_n^{(l)}\big) \, j\big(u_n^{(i)}, \xi_n^{(l)}\big) \, .$$

The numerical solution of this last problem is obtained using classical optimization techniques. Note however that different implementations can be considered, which do not necessarily lead to the "scenario tree" approach widely used in the field of stochastic programming.

4. Conclusion We have proposed an approximation scheme in which the discretization of the noise and the discretization of the information are designed *independently*. It is interesting to compare this approach with others also taking into account the whole discretization process (noise and information) for stochastic optimal control problems.

Barty's approach. In his PhD thesis (in French), K. Barty proves the convergence of a discretization scheme for problem (1). The result he gives ([2, Theorem IV.28]) makes use of the same notions of convergence as those used in § 2 for the counterexample, both for the σ-field and the random variable. Its approach involves two consecutive steps.

(i) The σ-field $\mathcal{F}$ is approximated by a σ-field $\mathcal{F}_k \subset \mathcal{F}$ which is generated by a finite partition $\mathcal{P}_k = \{\Omega_1, \ldots, \Omega_k\}$ of Ω, and problem (1) is replaced by:

$$V\big(\boldsymbol{\xi}, \mathcal{F}_k\big) = \min_{\boldsymbol{u} \text{ is } \mathcal{F}_k - \text{measurable}} \mathbb{E}\big[j(\boldsymbol{u}, \boldsymbol{\xi})\big] \, . \tag{18}$$

The optimal value $V\big(\boldsymbol{\xi}, \mathcal{F}_k\big)$ of (18) converges towards the optimal value $V\big(\boldsymbol{\xi}, \mathcal{F}\big)$ of (1) as $\mathcal{F}_k$ strongly converges to $\mathcal{F}$ ([2, Theorem IV.21]). Note that an $\mathcal{F}_k-$measurable random variable $\boldsymbol{u}$ is constant over each subset Ω_l constituting $\mathcal{P}_k$: such a random variable $\boldsymbol{u}$ is characterized by a k-uple $\big(u_1, \ldots, u_k\big) \in U^k$, and the minimization in (18) is thus performed over a finite dimensional space.

(ii) For a given index k, the random variable $\boldsymbol{\xi}$ is approximated by a finitely valued random variable $\boldsymbol{\xi}_n$ and problem (18) is replaced by:

$$V\big(\boldsymbol{\xi}_n, \mathcal{F}_k\big) = \min_{\boldsymbol{u} \text{ is } \mathcal{F}_k-\text{measurable}} \mathbb{E}\big[j(\boldsymbol{u}, \boldsymbol{\xi}_n)\big] \ . \tag{19}$$

The optimal value $V\big(\boldsymbol{\xi}_n, \mathcal{F}_k\big)$ of (19) converges with n toward the optimal value $V\big(\boldsymbol{\xi}, \mathcal{F}_k\big)$ of (18) as $\boldsymbol{\xi}_n$ converges in distribution toward $\boldsymbol{\xi}$ ([2, Theorem IV.26]). Note that this step only involves open-loop problems, which are approximated using the traditional Monte Carlo approach.

In this approach, the global discretization error $\big|V\big(\boldsymbol{\xi}, \mathcal{F}\big) - V\big(\boldsymbol{\xi}_n, \mathcal{F}_k\big)\big|$ is bounded from above by the sum of two terms:

- $\big|V\big(\boldsymbol{\xi}, \mathcal{F}\big) - V\big(\boldsymbol{\xi}, \mathcal{F}_k\big)\big|$: *information structure discretization error*,
- $\big|V\big(\boldsymbol{\xi}, \mathcal{F}_k\big) - V\big(\boldsymbol{\xi}_n, \mathcal{F}_k\big)\big|$: *mean computation discretization error.*

Apart form the convergence result itself, this approach enlightens the fact that it is not sufficient to properly deal with the last term (Monte Carlo) in order to obtain a "good" approximation of problem (1).

The main difference is thus that problem (1) is *nestedly* approximated in Barty's approach, whereas we *simultaneously* approximate the σ-field and the random variable.

Pennanen's approach. In [7], T. Pennanen adresses a stochastic optimization problem very similar to problem (1). He assumes that the observation $\boldsymbol{y}$ is a function[2] of the noise $\boldsymbol{\xi}$:

$$\boldsymbol{y} = h\big(\boldsymbol{\xi}\big) \ .$$

Then the problem can be formulated on the probability space $\big(\Xi, \mathcal{B}_\Xi, \mu\big)$, μ being the probability distribution of ξ, rather than on the probability space $\big(\Omega, \mathcal{A}, \mathbb{P}\big)$. Pennanen chooses a quantification operator q_n on Ξ, leading to an approximated random variable $\boldsymbol{\xi}_n$

$$\boldsymbol{\xi}_n = q_n\big(\boldsymbol{\xi}\big) \ ,$$

and then deduces the information quantization from the noise quantization by setting:

$$\boldsymbol{y}_n = h\big(\boldsymbol{\xi}_n\big) = h \circ q_n\big(\boldsymbol{\xi}\big) \ .$$

Now, there is no reason for the quantified observation $\boldsymbol{y}_n$ to be measurable with respect to the initial observation $\boldsymbol{y}$. In order to overcome the difficulty, Pennanen assumes that, in terms of sub σ-fields of $\mathcal{B}$, the following inclusion holds:

$$\sigma\big(h \circ q_n\big) \subset \sigma\big(h\big) \ . \tag{20}$$

Note that condition (20) means that, if two samples ξ and ξ' of $\boldsymbol{\xi}$ lead to identical observations, so do the quantified noises $q_n\big(\xi\big)$ and $q_n\big(\xi'\big)$. In the dynamic framework of problem (2), this assumption implies that the sampled trajectories of the noise are organized in a scenario tree.

The main difference is thus that the approximation of the σ-field is intimately related to the approximation of the random variable in Pennanen's approach, requiring the additional assumption (20), whereas these two approximations are designed *independently* in our approach.[3]

Ultimately, our approach seems to be more generic as the ones we reviewed. Moreover, it leads to more general numerical schemes as the scenarios trees widely used in stochastic programming.

Acknowledgments. This paper is based on earlier work by the Systems & Optimization Working Group (SOWG, École Nationale des Ponts et Chaussées), currently composed with Kengy Barty, Pierre Carpentier, Jean-Philippe Chancelier, Guy Cohen, Anes Dallagi, Michel De Lara, Babacar Seck and Cyrille Strugarek.

[2] an assumption which can be made without loss of generality

[3] Note however that Pennanen's approach is also designed to handle extended-real-valued functions.

References

[1] J-P. Aubin and H. Frankowska, *Set-valued analysis*, Birkhäuser, 1988.

[2] K. Barty, *Contributions à la discrétisation des contraintes de mesurabilité, pour les problèmes d'optimisation stochastique*, Ph.D. thesis, Thèse de doctorat de l'ENPC, june 2004, http://cermics.enpc.fr/theses.

[3] P. Billingsley, *Convergence of probability measures*, Wiley, 1999.

[4] J.-P. Chancelier and SOWG, *Epi-convergence of stochastic optimization problems involving both random variables and measurability constraints approximations*, Prépublication du CERMICS (2006), http://cermics.enpc.fr/reports.

[5] K.D. Cotter, *Similarity of information and behavior with a pointwise convergence topology*, Journal of Mathematical Economics **15** (1986), no. 1, 25–38.

[6] H. Kudo, *A note on the strong convergence of σ-algebras*, Annals of Probability **2** (1974), no. 1, 76–83.

[7] T. Pennanen, *Epi-convergent discretizations of multistage stochastic programs*, Mathematics of Operations Research **30** (2005), no. 1, 245–256.

[8] L. Piccinini, *Convergence of nonmonotone sub-σ-fields and convergence of associated subspaces $L^p(\mathcal{B}_n)$ ($p \in [1, +\infty]$)*, Journal of Mathematical Analysis and Applications **225** (1998), no. 1, 73–90.

[9] B.T. Polyak, *Convergence and convergence rate of iterative stochastic algorithms*, Automation and Remote Control **37** (1976), no. 12, 1858–1868.

[10] R.T. Rockafellar and R.J.-B. Wets, *Variational analysis*, A series of Comprehensive Studies in Mathematics, vol. 317, Springer Verlag, 1998.

[11] C. Strugarek, *Approches variationnelles et autres contributions en optimisation stochastique*, Ph.D. thesis, Thèse de doctorat de l'ENPC, may 2006, http://cermics.enpc.fr/theses.

[12] C. Strugarek and SOWG, *On the fortet-mourier metric for the stability of stochastic optimization problems, an example*, `http://speps.org`, Stochastic Programming E-Print Series **25** (2004).