

On the Fortet-Mourier metric for the stability of Stochastic Optimization Problems, an example,

Cyrille Strugarek¹ and SOWG²

May, 19 2004

Abstract

We consider the use of the Fortet-Mourier metric between two probability measures to bound the error term made by an approximated solution of a stochastic program. After a short analysis of usual stability arguments, we propose a simple example of stochastic program which enlightens the importance of the information structure. As a conclusion, we underline the need to take into account both the probability measure and the information structure in the discretization of a stochastic program.

Key words: stochastic optimization, stability, Fortet-Mourier metric, measurability constraints, information structure.

1 Introduction

A stochastic optimization problem typically involves a probability space and random variables on this space. The randomness of such problems appears therefore at two different places :

- through the *calculation of the objective function* of the problem (which involves the random variables and is often the result of an expectation),
- through the probabilistic constraints on the command variables, like measurability constraints (non-anticipativity in a multi-stage model).

From a practical point of view, a relevant question is: how to deal with such a random part ? The law of the random variables is often unknown, and consequently must be approached, for example by a Monte-Carlo simulation method. One usually solves a discretized version of the problem, where an *approached law* is used instead of the real one, and where a *discrete description of the measurable space* replaces the theoretical one.

As a consequence, we may ask now two questions : on the one hand, how to compute the approached objective function (equivalently the approached expectation) ? on the other hand, how to write the measurability constraints on the control variables according to the discrete description of the measurable space ? The answers must take into account the convergence properties of the discretization, because we want of course to find a feasible solution of the initial problem by solving approximated versions of it. These two questions yield a third one : **Is it correct to *discretize* by the same way the calculation of the objective function and the expression of the measurability constraints in the problem ?**

The problem of finding a convergent method to compute the objective function is thoroughly treated in the Stochastic Programming (SP) literature. To fix the ideas, one makes a Monte-Carlo simulation, because the usual input data of SP problems are fans of scenarios. We then can have good stability results using such a way ; see for example Shapiro and Homem-de-Mello [17] who used large deviations to refine usual Monte Carlo convergence results. An other way could be to quantize the space into a finite partition, and to fix a weight to each part of the quantization.

However, those types of methods often hide the difficulties related to measurability, and a common practice is to forget the measurability questions by optimizing on deterministic control variables instead of policies, ie

¹EDF R&D, OSIRIS, 1 avenue du Général de Gaulle, F-92141 Clamart Cedex, France, and, CERMICS-ENPC, 6-8 avenue Blaise Pascal - Cité Descartes Champs-sur-Marne - F-77455 Marne-la-Vallée cedex 2, France
cyrille.strugarek@edf.fr

²Systems Optimization Working Group, composed by Laetitia Andrieu (ENPC), Kengy Barty (ENPC), Pierre Carpentier (ENSTA), Jean-Philippe Chancelier (ENPC), Guy Cohen (ENPC), Michel de Lara (ENPC), Anes Dallagi (ENPC)

random variables. We will use the expression *open-loop problems* for stochastic programs with deterministic control variables, and the expression *closed-loop problems* for stochastic programs with measurability constraints on the control variables.

The usual stability results are often expressed and proved for *open-loop* problems (see for example [11]), and then applied to *closed-loop* problems, as if the measurability questions would not affect the properties of the unconstrained problem.

The main purpose of the present article is to show that the usual stability results based on metrics between probability measures don't easily apply on *closed-loop* problems, for the following reason : the most important part in a stochastic optimization problem is the information constraint, which is independent of all possible analytic metrics between probability measures. The goal here is to show that the two candidates to discretization, namely the objective function and the measurability constraints, must be taken into account separately. Without a cautious point of view concerning the measurability questions, it would be very easy to write false statements.

First of all, we will present in section 2 the framework of a stability result proved by Pflug in [11], and based on a metric between probability measures called Fortet-Mourier metric. We then show how the distinction becomes fundamental between the open-loop and the closed-loop cases. Section 3 is technical and concentrated on a convergence property of the empirical law according to the Fortet-Mourier metric. Section 4 develops a very simple example for which we make various calculations enlightening the importance of a separation in the discretization process between information and computation of the objective function. As a conclusion, we roughly draw the highlights of a method based on the work of Barty ([2]) to treat such problems.

2 Where is the problem ?

2.1 Framework and notations

The general framework we will use here is the following one: we want to solve a stochastic program by solving approached analogous problems. Consequently, we focus on the approximation error term.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(\mathbb{P}_N)_{N \in \mathbb{N}}$ be a sequence of probability measures on this space. We suppose Ω to be a compact metric space, and denote by $c(\cdot, \cdot)$ the associated metric. For all $N \in \mathbb{N}$, let Ω_N be the support of \mathbb{P}_N , and \mathcal{F}_N be the σ -field generated by $\mathcal{F} \cap \Omega_N$.

Let $\mathcal{U} \subset \mathbb{R}^p$ be the control vector space, and $j : \mathcal{U} \times \Omega \rightarrow \mathbb{R}$ be the objective function. We suppose that j is a normal integrand (see [13], 14.D). Consequently, for all measurable function $u : \Omega \rightarrow \mathcal{U}$, $j(u(\cdot), \cdot) : \Omega \rightarrow \mathbb{R}$ is measurable. Let us now define the functional space $\Gamma = \{u : \Omega \rightarrow \mathcal{U} \text{ measurable} : j(u(\cdot), \cdot) \in L_1(\Omega, \mathcal{F}, \mathbb{P})\}$, and the feasible control set $U^{ad} \subset \Gamma$. For notational convenience, let us define $J : \mathcal{U} \rightarrow \mathbb{R}$ to be $J(u) := \mathbb{E}(j(u(\cdot), \cdot))$, $\forall u \in U^{ad}$.

Analogously for all $N \in \mathbb{N}$, let:

$$\Gamma_N = \{u : \Omega_N \rightarrow \mathcal{U} \text{ measurable} : j(u(\cdot), \cdot) \in L_1(\Omega_N, \mathcal{F}_N, \mathbb{P}_N)\} ,$$

and $U_N^{ad} \subset \Gamma_N$ the associated feasible control set. Let us define $J_N(u_N) := \mathbb{E}_N(j(u_N(\omega), \omega))$, $\forall u_N \in U_N^{ad}$.

We are now ready to introduce the two problems:

$$\min_{u \in U^{ad}} \mathbb{E}(j(u(\omega), \omega)) ; \tag{1}$$

$$\forall N \in \mathbb{N}, \min_{u_N \in U_N^{ad}} \mathbb{E}_N(j(u_N(\omega), \omega)) , \tag{2}$$

which figure out respectively the real problem and the approached ones. We do now the following assumptions for all $N \in \mathbb{N}$:

- $\forall u \in U^{ad}, u|_{\Omega_N} \in U_N^{ad}$,

- there exists a onto and measurable operator ³ $\phi_N : \Omega \longrightarrow \Omega_N$ such that $\forall u_N \in U_N^{ad}$, $u_N \circ \phi_N \in U^{ad}$ and $(u_N \circ \phi_N)|_{\Omega_N} = u_N$.

It is therefore possible to compute $J(u_N) = \mathbb{E}(j(u_N \circ \phi_N(\omega), \omega))$ for all $u_N \in U_N^{ad}$ and $J_N(u) = \mathbb{E}_N(j(u|_{\Omega_N}(\omega), \omega))$ for all $u \in U^{ad}$.

These assumptions allow us to build a feasible solution for the problem (1) with the approached ones, and to build a feasible solution for (2) with a feasible solution of the problem (1).

We now refer to the article of Pflug ([11]) who also aims at solving the problem (1) through the approached problems (2). As we wrote, the approximated problems are based on the approximation of the probability law (\mathbb{P}_N is an approximation of \mathbb{P}). The description of the feasible sets U_N^{ad} is then a consequence of the choice of \mathbb{P}_N which fixes the set Ω_N . An other way to express it is to notice that there is a *unique discretization* for the problem, exclusively based on the approximation of the probability.

2.2 A stability result ?

The closed-loop and open-loop problems enter both in the preceding framework: the open-loop problems correspond to U^{ad} be a set of constant functions. In order to minimize the approximation error term, let us denote by $e(u, u_N) = |\mathbb{E}(j(u(\omega), \omega)) - \mathbb{E}(j(u_N \circ \phi_N(\omega), \omega))|$ for all $u \in U^{ad}$, $u_N \in U_N^{ad}$ the error term. This term⁴ compares the calculations of the real objective function on a real feasible control variable, and on an approached control variable made feasible by the use of ϕ .

Let $u^\#$ be the solutions of problem (1), and $u_N^\#$ be the solution for all $N \in \mathbb{N}$ of problems (2). The existence of these solutions has not to be proved here, but could be proved under weak technical assumptions on the objective function and the feasible control sets. The goal is now to give a result on the scalar sequence $(e(u^\#, u_N^\#))_{N \in \mathbb{N}}$:

Proposition 1 (Pflug) *Let $\mathcal{G} = \{g : \Omega \rightarrow \mathbb{R} : \exists u \in U^{ad}, g(\omega) = j(u(\omega), \omega) \forall \omega \in \Omega\}$. Then:*

$$\forall N \in \mathbb{N}, \quad e(u^\#, u_N^\#) \leq 2 \sup_{g \in \mathcal{G}} |\mathbb{E}(g) - \mathbb{E}_N(g)|.$$

Proof : The idea is to bound $e(u^\#, u_N^\#)$. Let us denote $\mathcal{M} = \{u \in U^{ad} : \mathbb{E}(j(u(\omega), \omega)) \leq \mathbb{E}(j(u^\#(\omega), \omega)) + 2\varepsilon\}$, with $\varepsilon = \sup_{g \in \mathcal{G}} |\mathbb{E}(g) - \mathbb{E}_N(g)|$. If $u_N^\# \circ \phi_N \notin \mathcal{M}$, then:

$$\begin{aligned} 2\varepsilon + \mathbb{E}(j(u^\#(\omega), \omega)) &< \mathbb{E}(j(u_N^\# \circ \phi_N(\omega), \omega)) \\ &\leq \varepsilon + \mathbb{E}_N(j(u_N^\#(\omega), \omega)) \text{ by definition of } \varepsilon \\ &\leq \varepsilon + \mathbb{E}_N(j(u^\#(\omega), \omega)) \text{ by minimality} \\ &\leq 2\varepsilon + \mathbb{E}(j(u^\#(\omega), \omega)) \text{ by definition of } \varepsilon, \end{aligned}$$

which is a contradiction, and gives us the result. □

We note then for convenience $\Delta_{max} = \sup_{g \in \mathcal{G}} |\mathbb{E}(g) - \mathbb{E}_N(g)| = \sup_{u \in U^{ad}} |J(u) - J_N(u)|$. The bound on the error term can be seen as a metric between the probability measures involved, having a ζ -form (see [15]). Let us recall the definition of the Fortet-Mourier metric between two probability measures \mathbb{P} and \mathbb{P}_N on Ω :

$$d(\mathbb{P}, \mathbb{P}_N) = \sup_f \left| \int_{\Omega} f d\mathbb{P} - \int_{\Omega} f d\mathbb{P}_N \right|$$

³ We could replace this hypothesis by the following one :

$$\exists \psi_N : U_N^{ad} \longrightarrow U^{ad}, \text{ such that } \forall u \in U^{ad}, u|_{\Omega_N} \in U_N^{ad}, \forall u_N \in U_N^{ad}, (\psi_N \circ u_N)|_{\Omega_N} = u_N.$$

And we could derive the same properties with this hypothesis. The advantage of this hypothesis is that it is very general and allows more ways to build a real feasible solution on the basis of the approximate solution, but for the sake of simplicity, we have chosen here to consider the operator ϕ , and this choice doesn't change at all the questions stated by the present work.

⁴From an industrial point of view, it is the comparison between the real optimal cost (which is unknown) and the real cost due to the use of an approached control. This error term is therefore the quantity we want to be as small as possible!

Let us introduce the following assumption:

$$|j(u(\omega), \omega) - j(u(\omega'), \omega')| \leq c(\omega, \omega') \quad \forall \omega, \omega' \in \Omega, \forall u \in U^{ad}. \quad (3)$$

With assumption (3), we can bound from above the error term by:

$$\forall N \in \mathbb{N}, \quad e(u^\#, u_N^\#) \leq 2 \sup_{u \in U^{ad}} |\mathbb{E}(j(u(\omega), \omega)) - \mathbb{E}_N(j(u(\omega), \omega))| \leq 2d(\mathbb{P}, \mathbb{P}_N). \quad (4)$$

Practically, this result seems to be very useful: it could be a way to build good approximations of the initial problem, by measuring the Fortet-Mourier metric between the initial law and the approximated one which could be found to minimize the metric⁵. The next part offers for example a demonstration of the convergence of the empirical law according to the Fortet-Mourier metric.

But we have now to be very cautious, especially concerning the closed-loop case. Indeed, the regularity assumption (3) takes very various importance depending on the class of problems we consider. When the control variables are taken to be constant (open-loop case), the statement (3) can easily be verified through analytical calculations, and is not so strong for the objective function which only must be Lipschitz continuous, on a compact domain (recall that Ω is supposed to be compact). On the contrary, when the control variables are taken to be functions (closed-loop case), the situation becomes awful⁶: the assumption is not only a requirement on the objective function, but also on the composition of the objective function with all possible control variables!

One could say that such a problem is a consequence of our use of Pflug's stability result. That's true, but it illuminates a more important thing, namely the fact that *by using the Fortet-Mourier metric, we lost all the properties of the problem related with the information, ie the constraints on the control variables.*

Let us now specify the problems for further applications. Let Ω be a compact subset of \mathbb{R}^n , and for all $N \in \mathbb{N}$, Ω_N be a finite subset of Ω of cardinality N (typically a fan of scenarios of size N), and \mathbb{P}_N the associated empirical probability measure (with the same weight $1/N$ for all the elements of Ω_N).

Before we see the discretization problems on a simple example, let us prove a small convergence result for the empirical law of a random variable.

3 Convergence for the Fortet-Mourier metric

Assume that \mathbb{P} is a probability on (Ω, \mathcal{F}) , which is a compact metric space, and let $(\omega_i)_{i \in \mathbb{N}}$ be a sample. Let us define $\mathbb{P}_N := \frac{1}{N} \sum_{i=1}^N \delta_{\omega_i}$ be the associated empirical law of \mathbb{P} . We have :

Theorem 1 *If the support of \mathbb{P} is closed, then:*

$$\lim_{N \rightarrow \infty} d(\mathbb{P}, \mathbb{P}_N) = 0.$$

Proof : By the Glivenko-Cantelli theorem (see [6], theorem 4.4.24), we already know that the empirical laws weakly converge. It means that for all continuous and bounded function $f : \Omega \rightarrow \mathbb{R}$, $\int_{\Omega} f d\mathbb{P}_N \rightarrow \int_{\Omega} f d\mathbb{P}$ when $N \rightarrow \infty$. We now consider $\mathcal{L} = \{f : \Omega \rightarrow \mathbb{R} : |f(\omega) - f(\omega')| \leq c(\omega, \omega') \forall \omega, \omega' \in \Omega\}$ the set of all 1-Lipschitz continuous functions on Ω . It is relatively compact in the set of continuous functions on Ω , because it is a set of equicontinuous compact valued functions (Ascoli theorem). By invoking theorem 5.13.12 of [16], we can say that \mathbb{P}_N uniformly converges to \mathbb{P} on \mathcal{L} . We have therefore the convergence of \mathbb{P}_N to \mathbb{P} with respect to the Fortet-Mourier metric. \square

⁵The methods investigated by Prof. Römisch in this direction to build scenario trees (see [8, 7]) are based on more sophisticated stability results (see [12]), but the purpose here is not to use the most beautiful results but to point out where lies the difficulty!

⁶Assume for example that $\Omega = \mathbb{R}$, and that $\mathcal{U} = \mathbb{R}$. Suppose also that the objective function is given by $j(u, \omega) := u^2 + \omega$. j satisfies of course the assumption (3) when u is constant, but not at all when u is itself a function on Ω , for example when $u(\omega) := \omega$. In this example, Ω isn't compact, but the compactness only plays a role for the convergence result concerning Fortet-Mourier metrics.

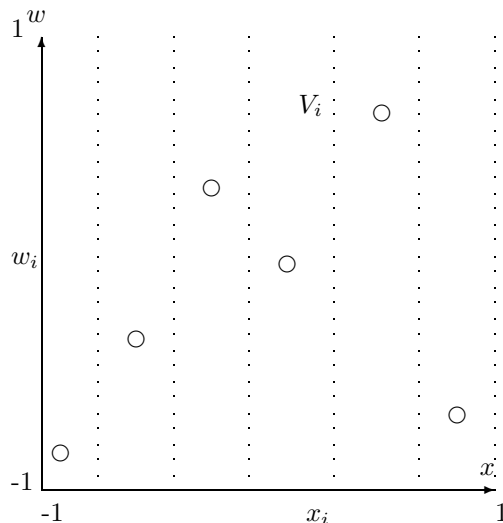


Figure 1: Partition of $[-1, 1]^2$

4 Example

As an application of our preceding warnings on the use of the Fortet-Mourier distance for stochastic stability results, we now consider the following example :

$$\min_u \mathbb{E}(\varepsilon u^2 + \beta y^2), \quad (5)$$

where we set:

$$\begin{aligned} \varepsilon, \beta &> 0, \\ x &\sim \mathcal{U}_{[-1,1]}, \text{ the initial state of the system,} \\ w &\sim \mathcal{U}_{[-1,1]}, \text{ the noise which is independent of } x, \\ y &= x + u + w, \text{ the final state of the system,} \\ u &= \phi(x), \text{ a real closed-loop control on the system.} \end{aligned}$$

To link it with our preceding notations, we have now $\Omega = [-1, 1]^2$, equipped with the product of independent uniform laws on $[-1, 1]$, and with the Borel σ -field.

The simplicity of the problems allows us to solve it explicitly. By using the dynamic programming approach, we introduce the Bellman functions $B_1(y) := \beta y^2$ and $B_0(x) := \min_{u \in \mathbb{R}} \mathbb{E}_w(\varepsilon u^2 + B_1(x + u + w))$. By computing the derivative with respect to u , we find obviously the optimal closed-loop command and the associated cost function:

$$u^\#(x) := -\frac{\beta x}{\varepsilon + \beta}, \quad \forall x \in [-1, 1], \quad \text{and } J^\# := \mathbb{E}_x(B_0(x)) = \frac{1}{3} \left(\frac{\varepsilon \beta}{\varepsilon + \beta} + \beta \right).$$

We now consider a N -sample of (x, w) , denoted by $(x_i, w_i)_{1 \leq i \leq N}$. Let $\mathbb{P}_N := \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, w_i)}$ the corresponding empirical law. We are ready to solve the discretized problem associated to \mathbb{P}_N . The closed-loop constraint on the control allows us to compute one control u_i for each trajectory (x_i, w_i) , for all $1 \leq i \leq N$. The calculation gives us the following optimal controls:

$$u_i^\# = -\frac{\beta}{\varepsilon + \beta}(x_i + w_i) \quad \forall i \in \{1, \dots, N\}.$$

We must then build a feasible solution for the real problem on the base of the $(u_i^\#)_{1 \leq i \leq N}$. It can be done through a partition of $[-1, 1]^2$ which associates to each $x \in [-1, 1]$ a unique trajectory (x_i, w_i) , in order to

preserve the causality of the problem, ie the closed-loop constraint. This operation is called a quantization of the space Ω . We choose the partition $V = \{V_1, \dots, V_N\}$ of $[-1, 1]$ in the following way: for all $i \in \{1, \dots, N\}$, $x_i \in V_i$ and $\forall j \neq i$, $x_j \notin V_i$. We can define with such a partition the operator ϕ by:

$$\forall (x, w) \in [-1, 1]^2, \exists ! i \in \{1, \dots, N\}, x \in V_i, \text{ and } \phi(x, w) = (x_i, w_i) .$$

From the partition V , we build a partition of $[-1, 1]^2$ by making frames. Typically, if we set $V_i = \left[\frac{x_i + x_{i-1}}{2}, \frac{x_i + x_{i-1}}{2} \right]$, for all $i \in \{2, \dots, N-1\}$, we can define a partition of the whole space as is drawn on figure 1. We can now build a feasible solution for the real problem based on the optimal discrete solution. We denote this enlarged solution by $u_N^\#$:

$$\forall x \in [-1, 1], u_N^\#(x) := -\frac{\beta}{\varepsilon + \beta} \sum_{i=1}^N 1_{V_i}(x)(x_i + w_i) . \quad (6)$$

We now investigate the behavior of the error term e introduced before, and its bound as in (4). It is straightforward to see that:

$$|\mathbb{E} \left(\varepsilon u_N^{\#2} + \beta(x + u_N^\# + w)^2 \right) - \mathbb{E}_N \left(\varepsilon u_N^{\#2} + \beta(x + u_N^\# + w)^2 \right)| \leq \Delta_{max} , \quad (7)$$

because by definition:

$$\Delta_{max} = \sup_u |\mathbb{E} (\varepsilon u^2 + \beta(x + u + w)^2) - \mathbb{E}_N (\varepsilon u^2 + \beta(x + u + w)^2)| .$$

We compute separately the two terms on the left hand-side of (7):

$$\begin{aligned} \mathbb{E}_N \left(\varepsilon u_N^{\#2} + \beta(x + u_N^\# + w)^2 \right) &= \frac{\varepsilon \beta^2}{(\varepsilon + \beta)^2} \frac{1}{N} \sum_{i=1}^N (x_i + w_i)^2 + \frac{\beta \varepsilon^2}{(\varepsilon + \beta)^2} \frac{1}{N} \sum_{i=1}^N (x_i + w_i)^2 , \\ &= \frac{\varepsilon \beta}{\varepsilon + \beta} \frac{1}{N} \sum_{i=1}^N (x_i + w_i)^2 \longrightarrow \frac{2}{3} \frac{\varepsilon \beta}{\varepsilon + \beta} \text{ when } N \rightarrow \infty . \end{aligned}$$

And the second term:

$$\begin{aligned} \mathbb{E} \left(\varepsilon u_N^{\#2} + \beta(x + u_N^\# + w)^2 \right) &= \frac{\beta^2 \varepsilon}{(\varepsilon + \beta)^2} \sum_{i=1}^N \mathbb{P}(V_i)(x_i + w_i)^2 + \frac{\beta^3}{(\varepsilon + \beta)^2} \sum_{i=1}^N \mathbb{P}(V_i)(x_i + w_i)^2 \\ &\quad + \frac{2\beta}{3} - \frac{2\beta^2}{\varepsilon + \beta} \sum_{i=1}^N (x_i + w_i) \mathbb{E}(x 1_{V_i}(x)) , \\ &= \frac{\beta^2}{\varepsilon + \beta} \sum_{i=1}^N \mathbb{P}(V_i)(x_i + w_i)^2 + \frac{2\beta}{3} - \frac{2\beta^2}{\varepsilon + \beta} \sum_{i=1}^N (x_i + w_i) \underbrace{\mathbb{E}(x 1_{V_i}(x))} . \\ &\qquad\qquad\qquad \frac{1}{N} \sum_{j=1}^N x_j 1_{V_i}(x_j) = \frac{x_i}{N} \end{aligned}$$

What is important here is the asymptotic behavior of the quantities, when $N \rightarrow \infty$.

The limit of the first term is given by the strong law of large numbers, and equals to $\frac{2\beta^2}{3(\varepsilon + \beta)}$. The strong law of large numbers helps us again to deal with the last term: we first separate by independence x_i and w_i , and easily find the limit: $-\frac{2\beta^2}{3(\varepsilon + \beta)}$.

We can compute the limit of the difference, and conclude with this lower bound for Δ_{max} :

$$\frac{2}{3} \left(\beta - \frac{\varepsilon \beta}{\varepsilon + \beta} \right) \leq \Delta_{max} . \quad (8)$$

As we can set ε as small as we want, the above bound of the error term is bounded from below by $\frac{2}{3}\beta$, and cannot converge to 0 (β is taken as big as we want).

On the contrary, as a conclusion of Theorem 1, \mathbb{P}_N converges to \mathbb{P} with respect to the Fortet-Mourier metric when $N \rightarrow \infty$.

Therefore, the Fortet-Mourier metric cannot asymptotically bound from above the Pflug's bound for the error term. One could nevertheless ask if the Fortet-Mourier metric could help us to govern directly the error term, without using Pflug's bound.

The computation of the error term follows:

$$e(u^\#, u_N^\#) = \underbrace{\mathbb{E} \left(\varepsilon u_N^{\#2} + \beta(x + u_N^\# + w)^2 \right)}_{\rightarrow \frac{2\beta}{3}} - \underbrace{\mathbb{E} \left(\varepsilon u^{\#2} + \beta(x + u^\# + w)^2 \right)}_{\frac{1}{3} \left(\beta + \frac{\beta\varepsilon}{\varepsilon + \beta} \right)}. \quad (9)$$

We see by (9) that $e(u^\#, u_N^\#) \rightarrow \frac{1}{3} \left(\beta - \frac{\beta\varepsilon}{\varepsilon + \beta} \right)$, when $N \rightarrow \infty$. By taking ε as small as we want, it is not possible to find a constant $M \in \mathbb{R}$ such that $e(u^\#, u_N^\#) \leq Md(\mathbb{P}, \tilde{\mathbb{P}})$.

To complete the study, we now show that the Lipschitz property of the closed-loop objective functions fails. It is quite clear when we look at the partition V and at the definition of $u_N^\#$. Let us take (x, w) and $(x', w') \in [-1, 1]^2$. Then define:

$$\Delta(x, x', w, w') := |j(u_N^\#(x), x, w) - j(u_N^\#(x'), x', w')|.$$

When we compute Δ , a part appears to be controlled by $|x - x'|$ and an other part by $|w - w'|$. But if we define i (resp. i') to be the unique index such that $1_{V_i}(x) > 0$ (resp. $1_{V_{i'}}(x') > 0$), we will also encounter terms controlled by $|x_i - x_{i'}|$ and $|w_i - w_{i'}|$. It is but clear that, as V is independent of $(w_n)_{1 \leq n \leq N}$, $|w_i - w_{i'}|$ cannot be controlled neither by $|x - x'|$ nor by $|w - w'|$. It shows that the objective functions are no more Lipschitz continuous in the closed-loop case, ie we cannot bound from above the error term by the Fortet-Mourier metric.

However, the conclusion of this example is not that the difficulty lies in a lack of regularity of the objective function. The conclusion is that we slaughtered the information structure by treating the closed-loop constraint directly on the sampling operation of the probability \mathbb{P} . The main problem is that we cannot control an information gap by a metric between probability measures, as sophisticated as the metric could be. Anyone who wants to approach a stochastic problem must begin to think about the way he will treat the measurability constraints.

5 Conclusion: a new way to proceed

The question of approximating a probability law in a stochastic optimization problem is merely treated in the literature, but we focused here on the difficulties that can happen when one doesn't take care about the information structure of the problem.

The information (namely the measurability constraints) appears to be a very important part in the discretization of a stochastic program, which cannot be solved by the use of metrics between probability measures. Specific tools must be developed to take it into account, and a part of them are introduced by Barty in his PhD thesis ([2]), concerning for example topologies on σ -fields (see [3, 4, 5, 10, 9, 14, 1]).

As a conclusion, we now show a way to treat the two parts where the random side of stochastic programs appears, namely in the calculation of the objective function, and in possible measurability constraints. Let us define with the prior notations:

$$V(\mathbb{P}, U^{ad}) := \min_{u \in U^{ad}} \mathbb{E}(j(u(\cdot), \cdot)).$$

The discrete version of measurability is the piecewise constance on the parts of a partition of the space. Consequently, a natural way is to discretize on the one hand the information structure by quantizing the space and searching control variables as piecewise constant functions on the corresponding partition, and to compute on the other hand the objective function by sampling (Monte Carlo methods!). We refer to Barty's PhD thesis

([2]) or to the report [18] for a more theoretical presentation of these aspects.

The main message of this small article is that stability results for closed-loop optimization problems (for example multi-stage stochastic programs) cannot solely be founded on metrics like Fortet-Mourier metric involving only the probability laws but not the essential part of the problem: information.

References

- [1] Z. Artstein. Sensitivity to σ -fields of information in stochastic allocation. *Stochastic and Stochastics reports*, (36):41–63, 1991.
- [2] K. Barty. *Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d’optimisation stochastique*. Thèse de doctorat, École Nationale des Ponts et Chaussées, 2004.
- [3] E.S. Boylan. Equiconvergence of martingales. *The Annals of Mathematical Statistics*, 42(2):552–559, 1971.
- [4] K.D. Cotter. Similarity of information and behavior with a pointwise convergence topology. *Journal of Mathematical Economics*, (15):25–38, 1986.
- [5] K.D. Cotter. Convergence of information, random variables and noise. *Journal of Mathematical Economics*, (16):39–51, 1987.
- [6] D. Dacunha-Castelle and M. Duflo. *Probabilités et Statistiques, Problèmes à temps fixe*. Masson, 1994.
- [7] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming. An approach using probability metrics. *Mathematical programming*, 95(3):493–511, 2003.
- [8] H. Heitsch and W. Römisch. Scenario reduction algorithms in stochastic programming. *Computational optimization and Applications*, (24):187–206, 2003.
- [9] H. Kudo. A note on the strong convergence of σ -algebras. *Annals of Probability*, 2:76–83, 1974.
- [10] J. Neveu. Note on the tightness of the metric on the set of complete sub σ algebras of a probability space. *The Annals of Mathematical Statistics*, (4):1369–1371, 1972.
- [11] G.Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical programming*, (89):251–271, 2001.
- [12] S.T. Rachev and W. Römisch. Quantitative stability in stochastic programming : the method of probability metrics. *Mathematics of Operations Research*, 27:792–818, 2002.
- [13] R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis*. Springer Verlag, Berlin Heidelberg, 1998.
- [14] L. Rogge. Uniform inequalities for conditional expectations. *Annals of Probability*, 2:486–489, 1974.
- [15] W. Römisch. Stability of stochastic programs. In A. Ruszczynski and A. Shapiro, editors, *Handbooks in Operations Research and Management Science*, volume 10, pages 483–554. Elsevier, Amsterdam, 2003.
- [16] L. Schwartz. *Analyse*. Hermann, 1993.
- [17] A. Shapiro and T. Homem-de Mello. On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *SIAM Journal in Optimization*, 11(1):70–86, 2000.
- [18] C. Strugarek. Quelle approche pour discrétiser un problème d’optimisation stochastique ? *Rapport interne OSIRIS-EDF R&D*, 2004.