Non-parametric approximation of non-anticipativity

Idea and test case

Jean-Sébastien Roy and Arnaud Lenoir

EDF Research & Development

8 March 2007

Introduction

For multistage problems, stochastic programming is usually performed on scenario trees. These trees are used to represent the non-anticipativity constraint and can be difficult to build if only historical scenarios are available.

We try to provide an alternative to scenario trees that makes use of a non-parametric estimate of non-anticipativity.

We present the idea and apply it to an example. Convergence of the scheme is not currently addressed.

Plan

Overview of stochastic programming

Approximation of non-anticipativity

Hydro management example

Plan

Overview of stochastic programming

Approximation of non-anticipativity

Hydro management example

Plan

Overview of stochastic programming

Approximation of non-anticipativity

Hydro management example

Plan

Overview of stochastic programming

Approximation of non-anticipativity

Hydro management example

Stochastic Optimization

Complexity of multistage programs

Stochastic optimization is difficult problem, even more so in the multistage case [Shapiro, 2006], but appears nevertheless quite frequently in a large number of domains and applications (finance, production, power, logistic, scheduling, ...).

Multistage problems differ from two stages ones by their information structure. In the multistage case, not only an optimal decision must be obtained for the first stage, but also decision rules for the following stages, which are functions of the information acquired at each stage.

This progressive acquisition of information and the associated measurability constraints make this problem much harder.

Therefore, contrarily to the two stages case, it is not possible to use a simple Monte Carlo sampling approach.

I'll put aside dynamic programming based methods which are a separate class with their strengths and weaknesses and I will concentrate on stochastic programming based methods. Among these methods, we can find:

1. Closed loop stochastic gradient methods [Barty et al., 2005] which adapt stochastic gradient techniques to the functional nature of the multistage problems ;
2. Methods of progressive refinement of a discrete approximation of the problem [Casey and Sen, 2006];
3. And the most common methods which first discretize the problem completely before solving it using classical means.

Stochastic Optimization
A priori discretization

Among those methods we can find:

1. Linear decision rules based methods [Thénié and Vial, 2004, Holt et al., 1955] or quantization [Barty, 2004], which discretize the multistage problem into a two stages one ;
2. And the most widely used technique usually associated with "Stochastic Programing" which consists in discretizing the problem into a deterministic finite dimension problem, solve by large scale optimization methods, such as linear programming. This discretization is usually performed with the help of a tree, a structure which intrinsically takes into account measurability constraints on the decisions.

While my approach will be slightly different, I'll first concentrate on that last class of methods.

Two widely different cases can occur:

1. Either we know how to perform conditional sampling on the uncertainties, typically when the processes involved are known, or when the uncertainties are independent. In that case, a large amount of tree building techniques are available, either Monte-Carlo based [Shapiro, 2003, Shapiro, 2006] or Quasi Monte Carlo based [Pennanen and Koivu, 2005].
2. Or we cannot perform such sampling and only scenarios are available, typically when only historical data are available, or the scenario generation procedure is not adapted to conditional sampling (e.g., some physical simulations). In this last case, if the uncertainties are continuous, the scenarios do no lend themselves directly to a tree structure.

When only scenarios are available
Solving from scenarios

In this last case, a few techniques are available (see [Dupačová et al., 2000] for a survey). Mainly:

1. Moment matching techniques [Høyland and Wallace, 2001, Høyland et al., 2003];
2. Probability metric methods [Römisch and Heitsch, 2006, Pflug, 2001, Hochreiter and Pflug, 2006], and among them, the classical tree generation by clustering method.

Plan

Overview of stochastic programming

Approximation of non-anticipativity

Hydro management example

Problem statement

Let $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, two random variables of support $\Xi_1 \subset \mathbb{R}^{p_1}$ and $\Xi_2 \subset \mathbb{R}^{p_2}$, and $\Xi_1 \times \Xi_2$ equipped with the euclidean norm denoted $\|\cdot\|$.

We address the following three stage problem:

$$J = \min_{u_0, u_1(\cdot)} \mathbb{E}\left[f\left(u_0, u_1\left(\boldsymbol{\xi}_1\right), \boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right)\right] \tag{1}$$

Under the constraints

$$\begin{cases} u_0 \in U_0 \subset \mathbb{R}^{n_0} \\ \forall \xi_1 \in \Xi_1,\ u_1\left(\xi_1\right) \in U_1 \subset \mathbb{R}^{n_1} \end{cases}$$

We suppose that for all $u_0 \in U_0$ $u_1\left(\xi_1\right) \in U_1$, ξ_1 and ξ_2, $f\left(u_0, u_1\left(\boldsymbol{\xi}_1\right), \xi_1, \xi_2\right)$ is finite and $f\left(u_0, u_1\left(\boldsymbol{\xi}_1\right), \boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right)$ is measurable and integrable for all u_0 and $u_1(\cdot)$ considered.

Dynamic and Historical scenarios

Hypothesis on the dynamic

We suppose that there exists a function h_1 such that $\boldsymbol{\xi}_2 = h_1(\boldsymbol{\xi}_1, \boldsymbol{\varepsilon}_2)$ with $\boldsymbol{\varepsilon}_2$ a random variables in $\mathbb{R}^{p_2'}$ independent of $\boldsymbol{\xi}_1$.

Scenarii

Let $\left(\xi_1^j, \xi_2^j, \varepsilon_2^j\right)_{j=1,\ldots,N}$ a sample of N independent drawings of $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\varepsilon}_2)$.

We try to solve problem (1), using only the values $\left(\xi_1^j, \xi_2^j\right)_{j=1,\ldots,N}$ that correspond to the historical scenarios observed.

The transfer function h_1 is supposed unknown.

Approximation of non-anticipativity
Equivalent formulation

We can express problem (1) equivalently as

$$J = \min_{u_0, u_1(\cdot)} \mathbb{E}\left[f\left(u_0, \mathbf{u}_1, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right)\right] \tag{2}$$

Under the constraints

$$\begin{cases} u_0 \in U_0 \subset \mathbb{R}^{n_0} \\ \mathbf{u}_1 \in U_1 \subset \mathbb{R}^{n_1}, \; a.s. \end{cases}$$

and the non-anticipativity constraint on $\mathbf{u}_1$:

$$\mathbf{E}\left[\mathbf{u}_1 \,|\boldsymbol{\xi}_1\right] = \mathbf{u}_1$$

Approximation of non-anticipativity
Non-parametric approximation

Let $D > 0$. We relax the non-anticipativity constraint

$$\mathbf{E}\left[\mathbf{u}_1 \left| \xi_1 \right.\right] = \mathbf{u}_1$$

by

$$\left\| \mathbf{E}\left[\mathbf{u}_1 \left| \xi_1 \right.\right] - \mathbf{u}_1 \right\|^2 \leq D^2$$

Then replace the conditional expectation by a non-parametric estimate over the scenarios

$$\sum_{j=1,\ldots,N} \left\| \frac{\sum_{k=1,\ldots,N} K\left(\frac{\left\|\xi_1^j - \xi_1^k\right\|}{\epsilon}\right) u_1^k}{\sum_{k=1,\ldots,N} K\left(\frac{\left\|\xi_1^j - \xi_1^k\right\|}{\epsilon}\right)} - u_1^j \right\|^2 \leq D^2$$

with $K(\cdot)$ a (typically Gaussian) kernel, and $\left(u_1^j\right)_{j=1,\ldots,N}$ the realizations of $\mathbf{u}_1$ on the historical scenario.

Approximation of non-anticipativity
Final form

Finally, we approximate the initial problem (1) by the following deterministic problem:

$$\tilde{J}_N = \min_{u_0, \left(u_1^j\right)_{j=1,\dots,N}} \frac{1}{N} \sum_{j=1,\dots,N} f\left(u_0, u_1^j, \xi_1^j, \xi_2^j\right)$$

subject to the constraints $u_0 \in U_0$ and $\forall j = 1, \dots, N$, $u_1^j \in U_1$ and

$$\sum_{j=1,\dots,N} \left\| \frac{\sum_{k=1,\dots,N} K\left(\frac{\left\|\xi_1^j - \xi_1^k\right\|}{\epsilon}\right) u_1^j}{\sum_{k=1,\dots,N} K\left(\frac{\left\|\xi_1^j - \xi_1^k\right\|}{\epsilon}\right)} - u_1^j \right\|^2 \leq D^2$$

We expect this approximation to converge to the true problem when $D \to 0$, $\epsilon \to 0$ and $N \to \infty$ at the appropriate speeds.

Plan

Context

We consider the problem of managing an hydro-power plant. Two successive production decisions $u_1 \geq 0$ and $u_2 \geq 0$ have to be made.

The reservoir initially contains an amount of energy S, so that we require $u_1 + u_2 \leq S$.

These decisions have to be taken as feedbacks on successive random selling prices $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$.

There is a non-anticipativity constraint on the first decision, i.e., u_1 has to be taken prior to any knowledge of the second price, except its conditional law with respect to the first one.

Problem setting

Mathematically, we consider the following cost function:

$$f(u_1, u_2, \xi_1, \xi_2) = -u_1\xi_1 - u_2\xi_2 - V(S - u_1 + u_2)$$

where $V(x)$ the value of the remaining stock at the end of the two steps, and is in our case a quadratic approximation of $\sqrt{\eta + x}$, i.e., $V(x) = \sqrt{\eta} + ax + bx^2$, with $b = \frac{2}{S^2}\left(\sqrt{\eta} - 2\sqrt{\eta + \frac{S}{2}} + \sqrt{\eta + S}\right)$, $a = \frac{\sqrt{\eta+S} - \sqrt{\eta} - b}{S}$ and $\eta = 0.1$. $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ follow independent uniform laws on $[0.4, 2]$, and $S = 1$.

Optimization problem

Our optimization problem is therefore:

$$J = \min_{u_1(\cdot), u_2(\cdot,\cdot)} \mathbf{E}\left[-u_1\left(\boldsymbol{\xi}_1\right)\boldsymbol{\xi}_1 - u_2\left(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right)\boldsymbol{\xi}_2 - V\left(S - u_1\left(\boldsymbol{\xi}_1\right) + u_2\left(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2\right)\right)\right]$$

subject to, $\forall\left(\xi_1, \xi_2\right) \in [0.4, 2]^2$, $u_1\left(\xi_1\right) \in [0, S]$ and $u_2\left(\xi_1, \xi_2\right) \in \left[0, S - u_1\left(\xi_1\right)\right]$.

The problem can be solved exactly by dynamic programming. We compare this exact solution to our approximation of the non-anticipativity constraint. Let $\left(\xi_1^j, \xi_2^j\right)_{j=1,\ldots,N}$ be N independent realizations of the prices.

Approximate version

We consider the following problem:

$$\tilde{J}_N = \min_{\left(u_1^j, u_2^j\right)_{j=1,\dots,N}} \frac{1}{N} \sum_{j=1,\dots,N} f\left(u_1^j, u_2^j, \xi_1^j, \xi_2^j\right) + \frac{C}{N} \sum_{j=1}^{N} \left\| u_1^j - \frac{\sum\limits_{\substack{k=1 \\ k \neq j}}^{N} K_{\epsilon_1}\left(\xi_1^j, \xi_1^k\right) u_1^k}{\sum\limits_{\substack{k=1 \\ k \neq j}}^{N} K_{\epsilon_1}\left(\xi_1^j, \xi_1^k\right)} \right\|^2$$

where $K_\epsilon(x, y) = e^{-\frac{\|x-y\|^2}{\epsilon^2}}$.

Feedback synthesis

Once this problem is solved, we synthesize the required feedbacks $u_1(\cdot)$ and $u_2(\cdot,\cdot)$ using:

$$u_1(\xi_1) = \frac{\sum_{j=1}^{N} K_{\epsilon_1}\left(\xi_1^j, \xi_1\right) u_1^j}{\sum_{j=1}^{N} K_{\epsilon_1}\left(\xi_1^j, \xi_1\right)}$$

$$u_2(\xi_1, \xi_2) = \frac{\sum_{j=1}^{N} K_{\epsilon_2}\left(\left(\xi_1^j, \xi_2^j\right), (\xi_1, \xi_2)\right) u_2^j}{\sum_{j=1}^{N} K_{\epsilon_2}\left(\left(\xi_1^j, \xi_2^j\right), (\xi_1, \xi_2)\right)}$$

with ϵ_2 chosen to provide the best fit for the point set $\left(\left(\xi_1^j, \xi_2^j\right), u_2^j\right)$. We choose empirically $\epsilon_2 = \sqrt{\frac{\epsilon_1}{\pi}}$. The quality of this approximate is then evaluated by a large Quasi-Monte-Carlo simulation.

Approximation as a function of C and ϵ_1

$\epsilon_1 = 0.02$ $\epsilon_1 = 0.1$ $\epsilon_1 = 0.5$

$C = 1$

$C = 5$

$C = 25$

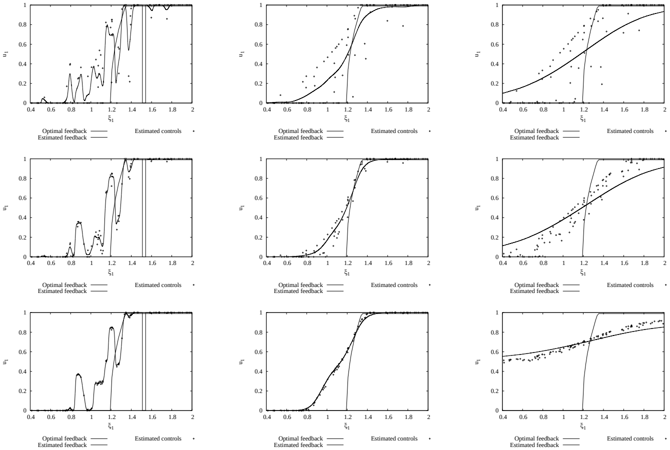


Figure: Approximation of $u_1(\cdot)$ for $N = 100$ scenarios as a function of the bandwidth ϵ_1 and the penalty C.

Approximation as a function of C and ϵ_1

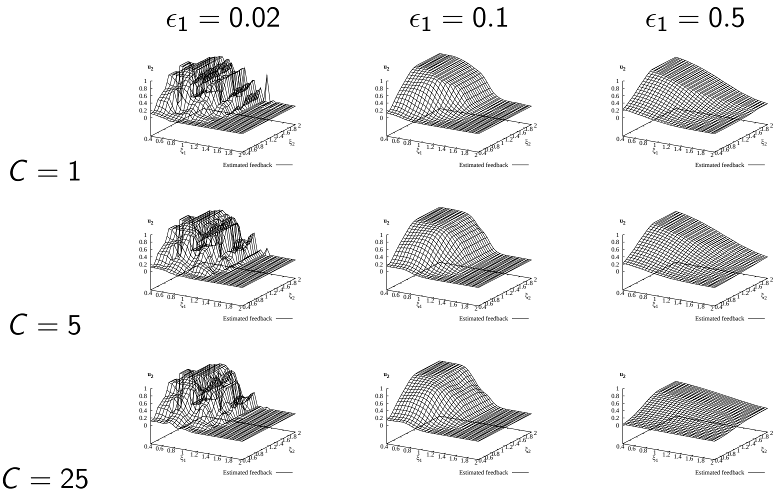


Figure: Approximation of $u_2(\cdot, \cdot)$ for $N = 100$ scenarios as a function of the bandwidth ϵ_1 and the penalty C.

Optimal feedback $u_2(\cdot,\cdot)$

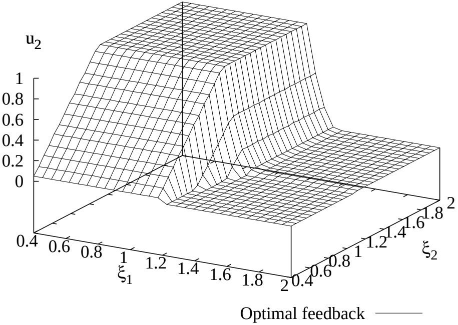


Figure: Optimal feedback $u_2(\cdot,\cdot)$

Comparison

The results must be compared to the optimum value, $J = -1.7414$, the value obtained without the penalty term, i.e., the anticipative solution assuming the future is known, whose value is -1.786, and the value of the solution obtained by synthesizing from the anticipative solution, feedbacks as detailed above, with $\epsilon_1 = 0.1$. The value of this last solution is -1.69563.

Approximation as a function of C and ϵ_1

	$\epsilon_1 = 0.02$	$\epsilon_1 = 0.1$	$\epsilon_1 = 0.5$
$C = 1$	-1.6788	-1.72698	-1.67559
$C = 5$	-1.67757	-1.73394	-1.66757
$C = 25$	-1.67638	-1.72863	-1.58707

Table: Quality of the approximation for $N = 100$ scenarios as a function of the bandwidth ϵ_1 and the penalty C.

Case $C = 0$

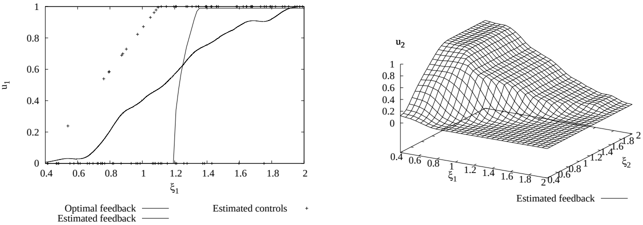


Figure: Approximations of $u_1(\cdot)$ and $u_2(\cdot, \cdot)$, for $N = 100$ scenarios, with bandwidth $\epsilon_1 = 0.1$ and no penalty, i.e., $C = 0$.

Best approximations for various N

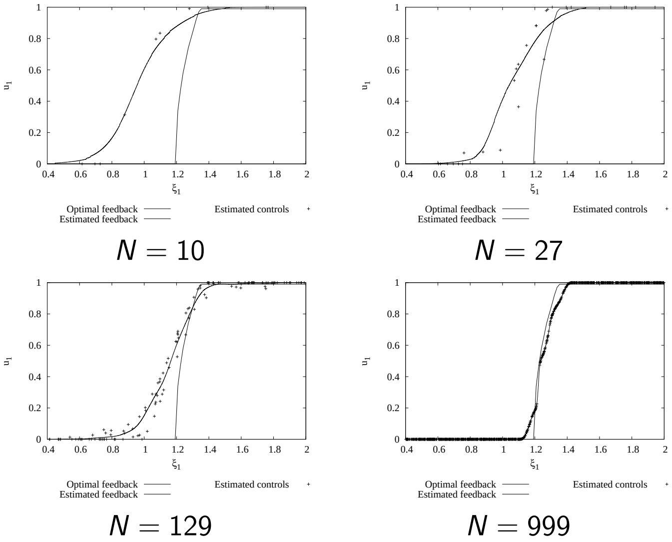
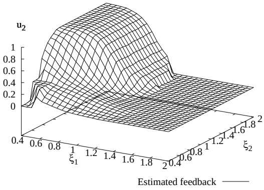
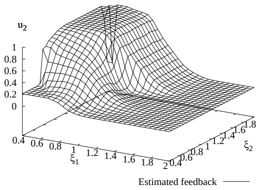


Figure: Best approximation of $u_1(\cdot)$ for $N = 10$, 27, 129 and 999.

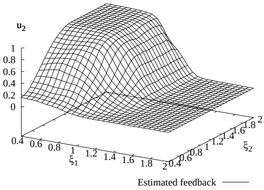
Best approximations for various N



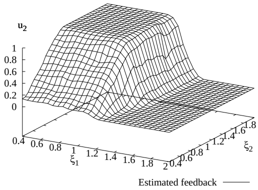
$N = 10$



$N = 27$



$N = 129$



$N = 999$

Figure: Best approximation of $u_2(\cdot, \cdot)$ for $N = 10$, 27, 129 and 999.

Best approximations for various N

N	ϵ_1	C	Value
10	0.215443	5.99484	-1.70561
27	0.129155	2.15443	-1.72187
129	0.0774264	5.99484	-1.73369
999	0.016681	1000	-1.74018

Table: Quality and values of ϵ_1 and C for the best solutions for $N = 10$, 27, 129 and 999.

Conclusion

1. It seems to works ;
2. It is quite simple to implement, and parameters can be defined through cross-validation or simulation ;
3. It should be able to decompose by scenario the problem obtained to speed up the computation ;
4. We are working on prooving convergence.

Bibliography I

Barty, K. (2004).
Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique.
Thèse de doctorat, École Nationale des Ponts et Chaussées.

Barty, K., Roy, J.-S., and Strugarek, C. (2005).
A stochastic gradient type algorithm for closed loop problems.
Stochastic Programming E-Print Series.

Casey, M. and Sen, S. (2006).
The scenario generation algorithm for multistage stochastic linear programming.
Mathematics of Operations Research.
to appear.

Bibliography II

Dupačová, J., Consigli, J., and Wallace, S. (2000).
Generating scenarios for multistage stochastic programs.
Annals of Operations Research, 100:25–53.

Hochreiter, R. and Pflug, G. C. (2006).
Financial scenario generation for stochastic multi-stage decision processes as facility location problem.
Annals of Operations Research.
to appear.

Holt, C., Modigliani, F., and Simon, H. (1955).
A linear decision rule for production and employment scheduling.
Management Science, 2(1):1–30.

Høyland, K., Kaut, M., and Wallace, S. (2003).
A heuristic for moment-matching scenario generation.
Computational Optimization and Applications, 24:169–185.

Bibliography III

Høyland, K. and Wallace, S. W. (2001).
Generating scenario trees for multi-stage decision problems.
Management Science, 47:295–307.

Pennanen, T. and Koivu, M. (2005).
Epi-convergent discretizations of stochastic programs via integration quadratures.
Numerische Mathematik, 100:141–163.

Pflug, G. C. (2001).
Scenario tree generation for multiperiod financial optimization by optimal discretization.
Mathematical Programming, 89:251–271.

Römisch, W. and Heitsch, H. (2006).
Scenrario tree modelling for multistage stochastic programs.
Technical Report Preprint 296, DFG Research Center MATHEON "Mathematics for key technologies".

Bibliography IV

Shapiro, A. (2003).
Inference of statistical bounds for multistage stochastic programming problems.
Math. Meth. Oper. Res., 58:57–68.

Shapiro, A. (2006).
On complexity of multistage stochastic programs.
Operations Research Letters, 34:1–8.

Thénié, J. and Vial, J.-P. (2004).
Stochastic programming with linear decision rules.
In *Computational Management Science Conference*. University of Geneva - Logilab.