Taylor-based pseudo-metrics for random process fitting in dynamic programming : expected loss minimization and risk management

Sylvain Gelly*, Jérémie Mary*, Olivier Teytaud* **
*TAO-Inria, LRI, UMR 8623(CNRS - Universite Paris-Sud),
bat 490 Universite Paris-Sud 91405 Orsay Cedex France
email: jeremie.mary@lri.fr
**Artelys, 12 rue du 4 Septembre, 75002 Paris, www.artelys.com

July 5, 2005

Abstract

Stochastic optimization is the research of x optimizing $E\ C(x, A)$, the expectation of $C(x, A)$, where A is a random variable. Typically $C(x, a)$ is the cost related to a strategy x which faces the realization a of the random process.

Many stochastic optimization problems deal with multiple time steps, leading to computationally difficult problems ; efficient solutions exist, for example through Bellman's optimality principle, but only provided that the random process is represented by a well structured process, typically an inhomogeneous Markovian process (hopefully with a finite number of states) or a scenario tree. The problem is that in the general case, A is far from being Markovian. So, we look for A', "looking like A", but belonging to a given family $\mathcal{A}'$ which does not at all contain A. The problem is the numerical evaluation of "A' looks like A".

A classical method is the use of the Kantorovitch-Rubinstein distance or other transportation metrics [Pflug, 2001], justified by straightforward bounds on the deviation $|EC(x, A) - EC(x, A')|$ through the use of the Kantorovitch-Rubinstein distance and uniform lipschitz conditions. These approaches might be better than the use of high-level statistics [Keefer, 1994]. We propose other (pseudo-)distances, based upon refined inequalities, guaranteeing a good choice of A'. Moreover, as in many cases, we indeed prefer the optimization with risk management, e.g. optimization of $EC(x, noise(A))$ where $noise(.)$ is a random noise modelizing the lack of knowledge on the precise random variables, we propose distances which can deal with a user-defined noise. Tests on artificial data sets with realistic loss functions show the relevance of the method.

1 Introduction, notations

1.1 Informal overview

Consider a deterministic system in discrete time that depends on an exogenous random process and on your decision. The question is : what is the decision you should take ?

If the dynamics of the system are known, if the random process has a structured representation, if the dimension of the state space is not too large, a very classical solution is dynamic programming. Unfortunately, in real life, the random process can often only be represented in a Markovian manner by a huge random process, the random process made of the few relevant variables being far from Markovian. We therefore have to design a Markovian random process, without explosion of the state space, that is close to the real one, which is only known through a use Markov random process or through a non-Markovian random process.

A classical solution is the replacement of the non-Markovian random process by a Markovian one, minimizing the distance between the original random process and the created one. In this paper :

- we recall that the Kantorovitch-Rubinstein distance corresponds to the minimization of a 0-degree Taylor's expansion of the objective function ;
- we show how to use a degree 1 or degree 2 Taylor's expansion ;
- we show how to extend this to optimization robust to distribution shifts of the random process;
- we show experimental results illustrating the improved behavior, in particular with the degree 2.

1.2 The problem

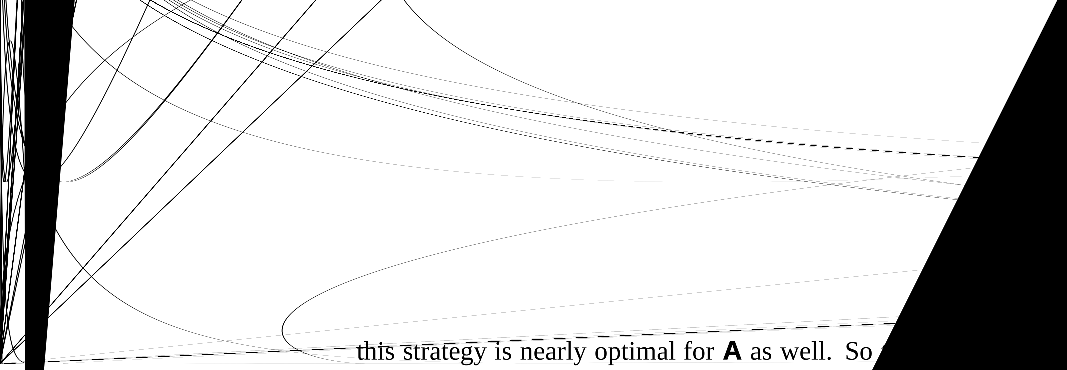
Let $\mathbf{S}$ be a set of strategies. Let $\mathbf{A}$ be a random variable belonging to a given set $\mathcal{A}$ of random variables, with domain $\mathbf{X}$. Let $\mathbf{s};\mathbf{a} \mapsto \mathbf{C}(\mathbf{s};\mathbf{a})$ be a cost function (where $\mathbf{s} \in \mathbf{S}$ is a strategy and $\mathbf{a}$ is a realization of $\mathbf{A}$). Our goal is to find a strategy $\mathbf{s}*$ which ensure a low cost (i.e. the smallest $\mathbf{E}(\mathbf{C}(\mathbf{s}*;\mathbf{A}))$ possible).

We suppose we already have an optimization method $\mathbf{Opt}$ which can find this optimum for a subset $\mathcal{A}' \subset \mathcal{A}$. So if our random variable $\mathbf{A}$ is in $\mathcal{A}'$, there is no problem. The trouble comes when we have a random variable $\mathbf{A} \notin \mathcal{A}'$.

We look for $\mathbf{s} \in \mathbf{S}$ such that $\mathbf{E}\mathbf{C}(\mathbf{s};\mathbf{A})$, cost expectation, is minimal, and the trouble is that $\mathbf{A} \notin \mathcal{A}'$.

In all the paper, $\mathbf{s}^*$ is an optimal strategy for $\mathbf{A}$. $\mathbf{s}^*$ is unknown. We assume that we can use $\mathbf{s}_0$, a strategy close to $\mathbf{s}^*$ (which will be used for the evaluation of derivatives in Taylor's expansions). In usual cases, such a $\mathbf{s}_0$ exist ; indeed in many practical cases, different versions coming from experts or from simplified models are used for comparison.

So we'll look for $\mathbf{A}' \in \mathcal{A}'$, such that $\mathbf{E}_{\mathcal{A}}\mathbf{C}(\mathbf{opt}(\mathbf{A}');\mathbf{A})$ is as small as possible, i.e. we'll use as strategy for $\mathbf{A}$ the strategy which is only optimal for $\mathbf{A}'$, expecting that



this strategy is nearly optimal for $\mathbf{A}$ as well. So t
choose $\mathbf{A}'$ close to $\mathbf{A}$ is very important. The K
proposed ([Gröwe-Kuska et al, 2003]) but it do
cost function of the problem. We propose a
about the cost function to make a better cho

In particular, we'll show that in orde
be chosen through $\mathbf{A}' \in \mathbf{argmin}_{A' \in \mathcal{A}'}$
This distance can be computed from us
cost function $\mathbf{C}(\mathbf{:})$ at a reasonably goo
based upon $\mathbf{A}' \in \mathcal{A}'$ and $\mathbf{A} \in \mathcal{A}$. Fig

Random pr

([Aïd et al, 2003][Coraluppi et al, 1999][Geibel, 2001][Heger, 1994]), multi-actors environments ([Tesauro, 1999][Nagayuki et al, 2000]). It has been used in the domain of energy, in particular for its ability to deal with particular form of dynamic optimization such as multi-actor optimization or optimization with particular treatment of random processes ([Riedmiller et al, 2001][Aïd et al, 2003]).

These approaches, however, are notably slow in face of problems where the natural convexity (as in stock management) allows the use of efficient approaches with stable convergences : stochastic dynamic programing ([Bertsekas, 1995]). Whenever it is someone said that dynamic programming is a particular case of reinforcement learning, it is in fact a very particular case (it converges by Lagrangian decomposition and not by fix-point iterations), and the use of decomposition methods makes possible to treat real problems with very large size ; the convergence is proved and the precision is very good. However, theses methods are very constrained from the point of view of the modelization of random processes ; only very structured random processes can be taken into account. The trouble is that in many cases, the random processes can only be known empirically through empirical time series and tendencies.

So the sum of the two constraints "reinforcement learning is too unstable and slow for the problem to be treated" and "stochastic dynamic programming can not deal with our random processes" leads to the following problem : how to choose a given family of random processes so as to fit a given random process ? Moreover, how to take risk into account ?

Assume that a criterion for choosing in a family of random process is provided. Then, the algorithmic resolution has been successfully performed in different papers ; in particular, [Hoyland et al, 2001] show that a complex criterion (therein based upon the respect of some high-level statistics) can be performed efficiently through non-linear programming, whenever the problem is non-convex ; it is easily differentiable and non-convexities can be handled thanks to multiple random initializations. Thus, a wide family of criterions can be concretely optimized.

Thus, the optimization of the criterion is tractable and the dynamic optimization thereafter is optimal within a very good precision after a long but controlled dynamic programming procedure. The choice of the criterion for choosing a model is then the main problem. In some cases the criterion was the preservation of some statistics, and in some cases the minimum Kantorovitch-Rubinstein distance. The latter is derived using bounds on the loss of precision depending upon the Kantorovitch-Rubinstein distance between the real random-process and its model. However, the bound is far from being optimal as it uses uniform bounds on the derivative. The purpose of this paper is the use of refined inequalities, using information about the derivatives, and to derive from these inequalities some improved distances for the choice of a model of the random process. Following [Gröwe-Kuska et al, 2003], we try to define the sense to be given to the term "look like", so that when a random process "looks like" the real random process, the optimum for this approximate random process is nearly optimal for the real random process. In particular, we search a distance more relevant than the Kantorovitch-Rubinstein distance. Moreover, we adapt our results to the case with risk.

The paper is organized as follows. Section 2 provides theoretical foundations (link with distances in 2.1, case of expectation 2.2, case of risk management 2.3). Section 3 provides experiments on artificial (reality-inspired) data sets. Section 4 concludes.

2 Theoretical results

Section 2.1 presents the general ideas. Section 2.2 presents and justifies a few distances for the optimization in expectation. Section 2.3 adapts these distances to the case of robust-to-noise optimization.

2.1 Stochastic optimization and distances

We want to optimize $\mathbf{s} \mapsto \mathbf{EC}(\mathbf{s};\mathbf{A})$. We look for $\mathbf{A}' \in \mathcal{A}'$ such that $\mathbf{E}_A\mathbf{C}(\mathbf{Opt}(\mathbf{A}');\mathbf{A})$ is as small as possible. We note that :

<u>**Lemma 1 :**</u>

$$\mathbf{EC}(\mathbf{Opt}(\mathbf{A}');\mathbf{A}) \leq$$

$$\mathbf{inf}_s\mathbf{EC}(\mathbf{s};\mathbf{A}) + 2\sup_s |\mathbf{EC}(\mathbf{s};\mathbf{A}) - \mathbf{EC}(\mathbf{s};\mathbf{A}')|$$

<u>**Interpretation :**</u> This simply shows that if $\mathbf{EC}(\mathbf{s};\mathbf{A})$ is uniformly close to $\mathbf{EC}(\mathbf{s};\mathbf{A}')$, then the cost in generalization, for the strategy optimized against $\mathbf{A}'$ instead of $\mathbf{A}$, is surely close to the optimal strategy for $\mathbf{A}$.

<u>**Proof :**</u>

Let's note $\mathbf{s}' \in \mathbf{Opt}(\mathbf{A}')$, $\mathbf{s}^* \in \mathbf{Opt}(\mathbf{A})$, and $= \sup_s |\mathbf{EC}(\mathbf{s};\mathbf{A}) - \mathbf{EC}(\mathbf{s};\mathbf{A}')|$.

$$\mathbf{EC}(\mathbf{s}';\mathbf{A}) \leq \mathbf{EC}(\mathbf{s}';\mathbf{A}') +$$

$$\leq \mathbf{EC}(\mathbf{s}^*;\mathbf{A}') + \quad \leq \mathbf{EC}(\mathbf{s}^*;\mathbf{A}) + 2$$

■

2.2 Optimization of the expectation

We will note $\mathbf{A}' = \ (\mathbf{A})$. This (slight abuse of notation) notes that $\mathbf{A}'$ and $(\mathbf{A})$ have the same law. As noticed in lemma 1, we are interested in $\sup_s |\mathbf{EC}(\mathbf{s};\mathbf{A}') - \mathbf{EC}(\mathbf{s};\mathbf{A})|$; i.e. we look for a distance which is very related to this quantity (so that if we replace $\mathbf{A}$ by $\mathbf{A}'$ close to $\mathbf{A}$ for this distance, then by lemma 1 the optimal strategy for $\mathbf{A}'$ will be nearly optimal for $\mathbf{A}$). We consider

$$_s(\mathbf{A};\mathbf{A}') = |\mathbf{EC}(\mathbf{s};\mathbf{A}') - \mathbf{EC}(\mathbf{s};\mathbf{A})|$$

$$_s(\mathbf{A};\mathbf{A}') = \Big|\mathbf{E}_A\Big(\ \nabla_A\mathbf{C}(\mathbf{s};\mathbf{A})(\ \ (\mathbf{A}) - \mathbf{A})$$

$$+\ \ \| \frac{@^2\mathbf{C}}{@\mathbf{A}^2} \|^{\infty}_{S,[A,\pi(A)]} (\mathbf{A} - \ \ (\mathbf{A}))^2\mathbf{=}2\Big)\Big|$$

where $\in [-1;1]$. We will now assume that $\mathbf{d}(\mathbf{s};\mathbf{s}^*) \leq$ and $\mathbf{d}(\mathbf{s}_0;\mathbf{s}^*) \leq$. Then

$$_s(\mathbf{A};\mathbf{A}') = \Big|\mathbf{E}_A\Big(\ \nabla_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})(\ \ (\mathbf{A}) - \mathbf{A})$$

$$+\ \ \| \frac{@^2\mathbf{C}}{@\mathbf{A}^2} \|^{\infty}_{S,[A,\pi(A)]} (\mathbf{A} - \ \ (\mathbf{A}))^2\mathbf{=}2$$

$$+2\ ' \ \ \| \frac{@\mathbf{C}}{@\mathbf{S}} \|^{\infty}_{S,[A,A']} \Big)\Big|$$

where $\ \in [-1;1]$ and $\ ' \in [0;1]$. So

$$_s(\mathbf{A};\mathbf{A}') \leq \textbf{distance}(\ ;\mathbf{A};\mathbf{s}_0)+$$

$$\Big|\mathbf{E}_A\Big(\ \ \| \frac{@^2\mathbf{C}}{@\mathbf{A}^2} \|^{\infty}_{S,[A,\pi(A)]} (\mathbf{A} - \ \ (\mathbf{A}))^2 \mathbf{=} 2$$

$$+2\ ' \ \ \| \frac{@\mathbf{C}}{@\mathbf{S}} \|^{\infty}_{S,[A,A']} \Big)\Big|$$

with $\textbf{distance}(\ ;\mathbf{A};\mathbf{s}_0) = |\textbf{deviation}(\ ;\mathbf{A};\mathbf{s}_0)|$ and $\textbf{deviation}(\ ;\mathbf{A};\mathbf{s}_0) = \mathbf{E}_A\nabla_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})(\ \ (\mathbf{A}) - \mathbf{A})$. This "distance" (which is not formally a distance), contains negative terms, and can be close to 0 whenever some terms are very large. We thus have another version which looks more safe : $\textbf{distance}'(\ ;\mathbf{A};\mathbf{s}_0) = \mathbf{E}_A|\nabla_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})(\ \ (\mathbf{A}) - \mathbf{A})|$. However, we see below that this is not necessary.

We can go further with a Taylor expansion of order 2, the second derivative being indeed classically the criterion showing the transition to the risky domain :

$$_s(\mathbf{A};\mathbf{A}') \leq \textbf{distance}^{(2)}(\ ;\mathbf{A};\mathbf{s}_0)+$$

$$\Big|\mathbf{E}_A\Big(\ \ \| \frac{@^3\mathbf{C}}{@\mathbf{A}^3} \|^{\infty}_{S,[A,\pi(A)]} (\mathbf{A} - \ \ (\mathbf{A}))^3 \mathbf{=} 6$$

$$+2\ ' \ \ \| \frac{@\mathbf{C}}{@\mathbf{S}} \|^{\infty}_{S,[A,A']} \Big)\Big|$$

with $\textbf{distance}^{(2)}(\ ;\mathbf{A};\mathbf{s}_0) = |\textbf{deviation}^{(2)}(\ ;\mathbf{A};\mathbf{s}_0)|$ with $\textbf{deviation}^{(2)}(\ ;\mathbf{A};\mathbf{s}_0) = \mathbf{E}_A\nabla_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})(\ \ (\mathbf{A})-\mathbf{A})+\frac{1}{2}(\ \ (\mathbf{A})-\mathbf{A})^t\mathbf{H}_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})(\ \ (\mathbf{A})-\mathbf{A})$ with $\mathbf{H}_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})$ the Hessian of $\mathbf{C}(\mathbf{s}_0;\mathbf{:})$. As previously, one can prefer $\textbf{distance}'^{(2)}(\ ;\mathbf{A};\mathbf{s}_0) = \mathbf{E}_A(|\nabla_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})(\ \ (\mathbf{A}) - \mathbf{A})| + |\frac{1}{2}(\ \ (\mathbf{A}) - \mathbf{A})^t\mathbf{H}_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})(\ \ (\mathbf{A}) - \mathbf{A})|)$.

We will consider as well a distance $\textbf{distance}^{(3)}$ in our experiments, Taylor expansion to order 3. The third derivative is not always very stable ; we prefer a term modelling the size of the error committed by the degree 2 approximation than a refined expansion. Thus, we replace the term of degree 3 by its absolute value.

Conclusion :

$$\mathbf{EC}(\textbf{Opt}(\mathbf{A}');\mathbf{A}) \leq \mathbf{EC}(\mathbf{s}^*;\mathbf{A}) + 2\textbf{distance}(\ ;\mathbf{A};\mathbf{s}_0)$$

$$+\textbf{error term}$$

(error term of degree 1)

$$\mathbf{EC}(\textbf{Opt}(\mathbf{A}');\mathbf{A}) \leq \mathbf{EC}(\mathbf{s}^*;\mathbf{A}) + 2\textbf{distance}^{(2)}(\ ;\mathbf{A};\mathbf{s}_0)$$

$$+\textbf{error term}_2$$

(error term of degree 2) (where the error terms are upper bounded as explained above), i.e. we can optimize the choice of $\mathbf{A}'$ with the help of derivatives at $\mathbf{s}_0$.

2.3 Optimization with risk management

We will now consider an optimization in the sense of the expectation for risk ([Aïd et al, 2003][Coraluppi et al, 1999][Geibel, 2001][Heger, 1994]), here for a modified random process. We consider that the real random process is not $\mathbf{A} \in \mathcal{A}$, but an unknown $'(\mathbf{A})$, where $'$ is random. Instead of optimizing $\mathbf{EC}(\mathbf{s};\mathbf{A})$, we thus want to optimize $\mathbf{EC}(\mathbf{s};\ '(\mathbf{A}))$, where the expectation is with respect to both $\mathbf{A}$ and $'$. Unfortunately, the straightforward application of the method above uses $\nabla_{\pi'(A)}\mathbf{C}(\mathbf{s}_0;\ '(\mathbf{A}))$ and we can only evaluate efficiently $\nabla_A\ \mathbf{C}(\mathbf{s}_0;\ '(\mathbf{A}))$. We need a distance between elements in $\mathcal{A}$ and elements in $\mathcal{A}'$, taking into account high-level informations about the noise.

Precisely, we define $'(\mathbf{a}) = \mathbf{a} + \mathbf{b}$, where $\mathbf{b}$ is an additive random noise ($'$ is thus a random function). We will then optimize $\mathbf{EC}(\mathbf{s};\ '(\mathbf{A}))$ instead of $\mathbf{EC}(\mathbf{s};\mathbf{A})$.

We will so look for a distance between $\mathbf{A}'$ and $\mathbf{A}$, that can be evaluated from available elements, such that $\mathbf{EC}(\mathbf{Opt}(\mathbf{A}');\ '(\mathbf{A}))$ is as small as possible. As previously, we will use lemma 1 and so try to optimize $\mathbf{sup}_s|\mathbf{EC}(\mathbf{s};\mathbf{A}') - \mathbf{EC}(\mathbf{s};\ '(\mathbf{A})|$. We recall that we are interested in distances using only derivatives of $\mathbf{C}$ and distances between $\mathbf{A}$ and $(\mathbf{A})$ for $\mathbf{A} \in \mathcal{A}$ and such that the law of $(\mathcal{A})$ is in $\mathcal{A}'$.

So we have, with $(\mathbf{A})$ of the same law as $\mathbf{A}'$ if $\mathbf{A}'$ represents $\mathbf{A}$ in $\mathcal{A}'$:

$$\mathbf{EC}(\mathbf{s};\ '(\mathbf{A})) - \mathbf{EC}(\mathbf{s};\ (\mathbf{A}))$$

$$= \mathbf{EC}(\mathbf{s};\ '(\mathbf{A})) - \mathbf{EC}(\mathbf{s};\mathbf{A}) + \mathbf{EC}(\mathbf{s};\mathbf{A}) - \mathbf{EC}(\mathbf{s};\ (\mathbf{A}))$$

$$= \mathbf{deviation}^{(2)}(\ ';\mathbf{A};\mathbf{s}_0) - \mathbf{deviation}(\ ;\mathbf{A};\mathbf{s}_0)$$

$$+\mathbf{error\ term}$$

If $\mathbf{b}$ is invariant by rotation, the term of order 1 in $\mathbf{deviation}^{(2)}$ disappears by symmetry and the remainder is

$$= \frac{1}{2}\mathbf{E\ Tr}(\mathbf{H};\mathbf{A})\mathbf{K} - \mathbf{deviation}^{(2)}(\ ;\mathbf{A};\mathbf{s}_0)$$

where $\mathbf{K}$ (chosen by the user) is $\mathbf{L}_2(\mathbf{b})^2$ and $\mathbf{Tr}(\mathbf{H};\mathbf{A})$ is the trace of the Hessian matrix $\mathbf{H}_A\mathbf{C}(\mathbf{s}_0;\mathbf{A})$. So, the distance proposed for the case of risk is :

$$\mathbf{distance}_K^{(r)}(\ ;\mathbf{A};\mathbf{s}_0) =$$

$$|\mathbf{deviation}^{(2)}(\ ;\mathbf{A};\mathbf{s}_0) - \frac{1}{2}\mathbf{E\ Tr}(\mathbf{H};\mathbf{A})\mathbf{K}|$$

It is not surprising that as $\mathbf{K}$ increases $(\mathbf{A})$ gets further from $\mathbf{A}$. Note that other forms of noise could be considered, for example with $\mathbf{K}$ depending upon $\mathbf{A}$.

3 Experiments

We explain in 3.1 the overall paradigm which is tested here. The chosen problem is specified in 3.2. The pseudo-code of the experiments is provided in 3.3. The results are presented in section 3.4.

3.1 Proposed approach

Our goal is the comparison of different distances between random processes $\mathbf{A}$ and $\mathbf{A}'$ for the method illustrated by the figure 1. However, the distances we propose are based on a Taylor polynomial approximation and so we restrict $\mathbf{A}'$ to be "not too far" from $\mathbf{A}$, so that the approximation is valid. The protocol is then the one illustrated in figure 1, but $\mathbf{A}'$ is chosen through optimization of $\mathbf{d}(\mathbf{A};\mathbf{A}')$ under the constraint that $\mathbf{dK}(\mathbf{A};\mathbf{A}') \leq \mathbf{dK}_0$, where $\mathbf{dK}$ is the Kantorovitch-Rubinstein distance.

3.2 Problem specifications

Let $\mathbf{S} = \mathbf{X} = [0;1]^d$. Consider a cost function $\mathbf{C}(:;:)$ and a random process $\mathbf{A}$. Here, $\mathbf{A}$ is a sum of $\mathbf{m}$ Dirac masses , drawn uniformly on $[0;1]^d$, with $\mathbf{d}$ the dimension of the considered problem (which is both the dimension of the random process and the dimension of the strategy). We have used typically $\mathcal{A}$ the family of distributions equal to sum of $\mathbf{m}$ Dirac masses ; $\mathbf{A}$ is randomly selected in $\mathcal{A}$ by placing $\mathbf{m}$ Dirac masses uniformly in $[0;1]^d$. $\mathcal{A}'$ is the family of sums of $\mathbf{p}$ Dirac masses, where $\mathbf{p} \in [[10;50]]$ depending upon the number of Dirac masses used for $\mathbf{A}$. The function mapping $\mathbf{A}$ to $\mathbf{A}'$ is then defined in the following way :

- $\mathbf{m{=}p}$ Dirac masses of $\mathbf{A}$ are mapped to each Dirac mass of $\mathbf{A}'$. We choose $\mathbf{m{=}p}$ to be integer to make the experiments easier.
- The way used to map the masses of $\mathbf{A}$ to the masses of $\mathbf{A}'$ is such as $\mathbf{A} - \ (\mathbf{A})$ is not large to ensure that the Taylor-approximations are relevant. In dimension 1, we simply sort the masses $\mathbf{A}$ and $\mathbf{A}'$, and we map the $\mathbf{m{=}p}$ smallest values of $\mathbf{A}$ to the smallest value of $\mathbf{A}'$ and so on. In dimension > 1, the method to map $\mathbf{A}$ to $\mathbf{A}'$ is more tricky, but the principle is the same.

The figure 2 gives, for a given distribution $\mathbf{A}$ the expectation of the cost in function of the strategies for a problem in dimension 1.

3.3 Algorithm

The experimental setup is as follow, with $\mathbf{A}$ a randomly drawn random process :

- Draw randomly a large number of $\mathbf{A}'$ ($\mathbf{p}$ Dirac masses, with $\mathbf{p} < \mathbf{m}$). We reject the $\mathbf{A}'$ which are not enough close to $\mathbf{A}$, because the approximations we have made are only valid close to $\mathbf{A}$. We construct $\ (\mathbf{A}) = \mathbf{A}'$ as stated above.
- For each random process $\mathbf{A}'$ calculate the optimal strategy $\mathbf{s}$ ($\mathbf{s} = \mathbf{Opt}(\mathbf{A}')$), and the expected associated cost (for the real random process $\mathbf{A}$) $\mathbf{c}(\mathbf{s}) = \mathbf{EC}(\mathbf{s};\mathbf{A})$.
- For each distance considered, sort the Dirac masses of the random process $\mathbf{A}'$ by their distance to $\mathbf{A}$.
- Plot the expected cost of the optimal strategy ($\mathbf{c}(\mathbf{s})$) for each distance : at abscissa $\mathbf{n}$, plot the expected cost of the optimal strategy for the $\mathbf{n}^{th}$ closest $\mathbf{A}'$ according to this distance. It the distance is efficient, then the cost should get lower and lower as the distance decreases.

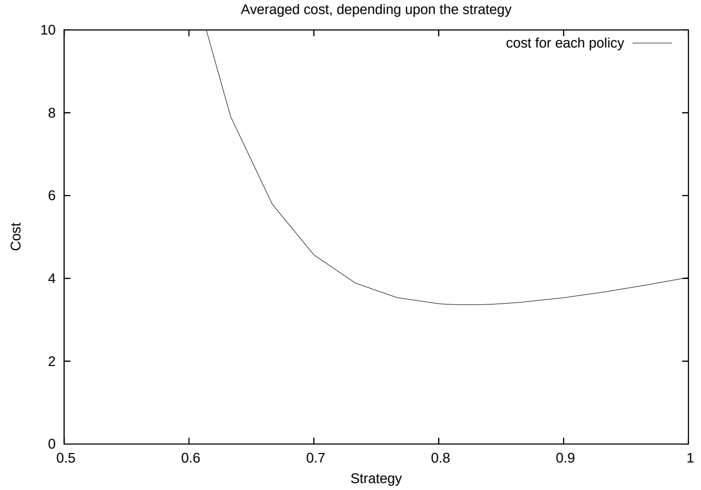


Figure 2: **Dimension 1 :** expectation of the cost for a given distribution A, in function of the different strategies. We see that as in many supply chain problems, if the strategy is too short, then the cost is very large (strong derivative for small values of the strategy), whereas if the strategy is too long, the marginal cost is much smaller (small derivative). This means that being too short is much more detrimental than being too long (for the same deviation).

3.4 Results without noise

The figures 3 and 4 show the expected cost for random process $\mathbf{A}$ for the optimal strategy optimized for random process $\mathbf{A}'$, in function of the rank of the distance between $\mathbf{A}$ and $\mathbf{A}'$.

For a "good" distance we expect that small distances are associated to a small cost. We can see that the Kantorovitch-Rubinstein distance gives very bad results because the cost is not correlated to the distance, whereas the $\mathbf{distance}^{(3)}$ is the best, and $\mathbf{distance}^{(2)}$ gives also good results.

Hence, if one chooses a random process $\mathbf{A}'$ because of its small Kantorovitch-Rubinstein distance to the true random process $\mathbf{A}$, and uses this $\mathbf{A}'$ as a model for the optimization he might get a very bad strategy, whenever the cost for $\mathbf{A}'$ is optimal.

We see on the zoomed part (below) of the figure 3 that $\mathbf{distance}^{(2)}$ and $\mathbf{distance}^{(3)}$ are very efficient. Indeed, the expected cost is very low in a wide range of chosen $\mathbf{A}'$. The non zoomed part (above) of the figure 3 shows that the distances we have proposed measures efficiently the distance between $\mathbf{A}$ and $\mathbf{A}'$ in the sense that when the distance is large, the optimization of the cost for $\mathbf{A}'$ is not efficient as a strategy for $\mathbf{A}$.

Figure 4 shows that the results remain in dimension 2.

3.5 Results with an additive noise

We keep the same protocol except we add a Gaussian noise of standard deviation $0{:}07$ (quite important noise in regard to the data) as defined in section 2.3. As previously, we focus on the very first points of the curves (the smallest distances). As we have shown in section 3.4 that $\mathbf{distance}^{(2)\prime}$ has no interest, we do not represent it any more. We always have $\mathbf{distance}^{(3)}$ better than $\mathbf{distance}^{(2)}$, itself better than the Kantorovitch-Rubinstein distance. The new distance "$\mathbf{distance}^{(2)}$ with risk", tries to take in account the additive noise. Notice that we only assume we are able to evaluate the variance of this noise (a realistic hypothesis) but we verify it on a "nice" distribution shift, with the same law and the same parameters as in the derivation of the distance. This nevertheless shows the mathematical consistency of the derivation. The curve of the "$\mathbf{distance}^{(2)}$ with risk" (see 5) is far better than the three other ones. This figure shows that our way to integrate noise into the distance can lead to significant improvements on the evaluations of the distance between random variables. Note that the level of noise is of course usually not available ; in our experiment, we used the *real* $\mathbf{K}$ value.

4 Conclusions

We have shown how to define a distance between $\mathbf{A}$ and $\mathbf{A}'$ taking into account derivatives of $\mathbf{C}(\mathbf{s}_0; \mathbf{A})$ in the research of a good represent ant $\mathbf{A}'$ of $\mathbf{A}$. Roughly speaking, the interpretation of the criterion leads to two ways of selecting $\mathbf{A}'$:

- put more points where the second derivative is strong ;
- counterbalance over-estimates $(\mathbf{A})$ of $\mathbf{A}$ for some values of the random process $\mathbf{A}$ by under-estimates of similar amplitude.

Our approach quantifies numerically these elements. Figure 3 shows a very clear correlation between our distance and the loss function in $\mathbf{A}$, which fully confirms our theoretical evaluations.

Another important feature of our method is that the risk can be taken into account in a very simple and efficient manner. Results of figure 5 show a very convincing improvement in the case of risk : adding the term of risk in our distance leads to a much better expectation on the noisy version of $\mathbf{A}$. Our framework is natural and avoids the use of ad-hoc adaptations of the notion of risk such as conditional Value-At-Risk ; our notion of risk can take into account additive noise (possibly *dependent* on the realization of $\mathbf{A}$ through $\mathbf{K} = \mathbf{K}(\mathbf{A})$), thus modelizing a variable uncertainty on $\mathbf{A}$; increasing $\mathbf{K}$ increasing the robustness to noise, in a simple, natural and convincing manner.

Our theoretical results are mathematically simple and sound (simply the use of Taylor's expansions). We point out the fact that they are just the natural extension of the derivation of Kantorovitch-Rubinstein bounds with more terms in the expansion. Our experiments confirm the theoretical results both for optimization in expectation and for the difficult problem of optimization with risk management.

We consider that these results are significant in the sense that :

- providing a framework that naturally extends to uncertainty management is important itself. In particular, the main argument for risk management is uncertainty ; dealing with risk management by the introduction of a noise on the random process is thus natural. This does not introduce any complexity in the model.

- stochastic dynamic optimization by dynamic programming methods is a very important real-world problem, which has not been much investigated by the AI community yet, whereas many future advances, according to us, will deal with the representation of random process and can come from statistics and paradigms of artificial intelligence ; dynamic programming remains the only reliable solution for many industrial problems ;

- only little attention has been devoted to the representation of random processes in stochastic dynamic optimization, whereas this part of the process appears in many cases as the most important part ; working on complex random processes in dynamic programming leads to very huge state spaces (the Bellman value has to be indexed on the full memory as soon as the random processes involved are not markovian) ; the nice representation of random processes is thus decisive with regard to the curse of dimensionality ;

- derivatives required here are often easily available (e.g., through Bellman values) ; the first derivative is a direct computation from the Bellman values and the transition function, the second derivative is more problem specific but follows naturally in many frameworks (e.g. parametric Bellman values parameterized by both the random process and the state space, or stock management with)

- the (relative) simplicity of our arguments is one more argument in favor of the method ;

- the optimization under constraint of our distance is easily performed (it is not convex, but it is polynomial and far from the state of the art of non-linear programming ; see e.g. [Hoyland et al, 2001]).

We now insist on the fact that our approach is directly operational and fully tractable. Implementing our approach requires 1) restricting the attention to processes such that $(\mathbf{A}) - \mathbf{A}$ is small enough, for any realization, to ensure that the Taylor expansion is valid 2) compute the required distances 3) optimize one of our Taylor-based distances. 1) is performed by introducing constraints in the optimization. 2) is immediate for the first derivative and the diagonal of the Hessian if approximate Bellman values are available ; the general case of the second derivatives is more tricky and problem-specific but is tractable in many cases and in particular in stock management, which is an important area of dynamic programming, or whenever the bellman values are parametric and parameterized by both the state space and a random process value. 3) can be handled with non-linear programming with automatic differentiation for an optimization of the Taylor-based distance under constraint on $(\mathbf{A}) - \mathbf{A}$; non-convexity can be handled thanks to multiple initial points as in [Hoyland et al, 2001].

In particular, direct applications are resource management for energy production, where modelization of short, middle and long-term meteorological random processes are very important and which is a main provider of real-world problems for stochastic dynamic programming, or portfolio management, where many different forms of unproperly defined random processes are concerned.

References

[Bertsekas, 1995] D. Bertsekas. Dynamic programming and optimal control. Athena Scientific, 1995.

[Aïd et al, 2003] R. Aïd, V. Grellier, A. Renaud, O. Teytaud. Application de l'apprentissage par renforcement à la gestion du risque. Proceedings of CAP 2003, in press.

[Bertsekas et al, 1996] D.P. Bertsekas, J.N. Tsitsiklis. Neuro-dynamic programming, Athena Scientific.

[Coraluppi et al, 1999] S. Corallupi, S. Marcus, Risk-sensitive and minimax control of discrete-time, finite state Markov Decision Processes. Automatica, 35, 301-309.

[Geibel, 2001] P. Geibel, Reinforcement Learning With Bounded Risk, In: C. E. Brodley, and A. P. Danyluk, editors, "Machine Learning - Proceedings of the Eighteenth International Conference (ICML01)", pages 162-169. Morgan Kaufmann Publishers, San Francisco,CA.

[Gröwe-Kuska et al, 2003] N. Gröwe-Kuska, H. Heitsch, W. Römisch, Scenario Reduction and Scenario Tree Construction for Power Management Problems, IEEE Bologna Power Tech Proceedings (A. Borghetti, C.A. Nucci, M. Paolone eds.), 2003.

[Heger, 1994] M. Heger, Consideration of risk in reinforcement learning. Proceedings of ECML pp105-111, Morgan Kaufman, 1994.

[Hoyland et al, 2001] K. Hoyland, S.W. Wallace, Generating Scenario Trees for Multistage Decision Problems, Informs 47, 2, 2001.

[Keefer, 1994] D.L. Keefer, Certainty equivalents for three-point discrete-distribution approximations. Management Sciences 40, 760-773.

[Nagayuki et al, 2000] Y. Nagayuki, S. Ishii, and K. Doya. Multi-agent reinforcement learning: An approach based on the other agent's internal model. In Proc. of the 4th Intl. Conf. on MultiAgent Systems.

[Neuneier, 1996] R. Neuneir, Optimal asset allocation using adaptive dynamic programming, in Advances in Neural Information Processings Systems, D.S. Touretzky, M.C. Mozer, M.E. Hasjselmo, eds, vol. 8, MIT Press.

[Neuneier, 1998] R. Neuneier, Enhancing Q-learning for optimal asset allocation, in Advances in Neural Information Processing Systems, M.I. Jordan, M.J. Kearns, S.A. Solla, eds, vol.10, MIT Press.

[Neuneier et al, 1999] R. Neuneier, O. Mihatsch, Risk-sensitive reinforcement learning. MIT Press, NIPS'99.

[Pflug, 2001] G.C. Pflug, Scenario Tree Generation for Multi-Period Financial Optimization by Optimal Discretization, Mathematical Programming, Series B 89:251-271, 2001.

[Riedmiller et al, 2001] M. Riedmiller, A. Moore, J. Schneider, Reinforcement Learning for Cooperating and Communicating Reactive Agents in Electrical Power Grids, Balancing Reactivity and Social Deliberation in Multi-Agent Systems.

[Sutton et al, 1998] R.S. Sutton, A.G. Barto, Reinforcement learning.

[Tesauro, 1989] G. Tesauro. Neurogammon wins Computer Olympiad. Neural Computation 1, 321-323.

[Tesauro, 1999] G. Tesauro. Pricing in agent economies using neural networks and multi-agent Q-learning. Proceedings of Workshop ABS-3: Learning About, From and With other Agents (held in conjunction with IJCAI '99).

[Watkins, 1989] C. Watkins, Learning from Delayed Reward, Ph.D thesis, Kings College, University of Cambridge.

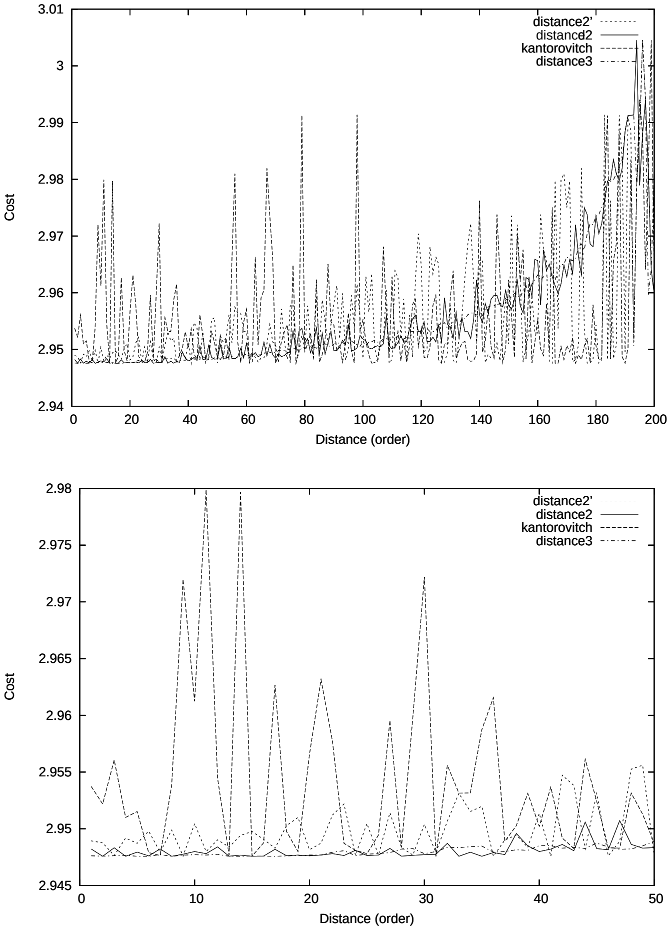


Figure 3: Results in **dimension 1** without noise. X-coordinate is the rank of A' according to the considered distance. For a "good" distance we expect that small distances are associated to a small cost. The second graphic is a zoom on the smallest distances between A and $\pi(A) = A'$. We can see that the Kantorovitch-Rubinstein distance gives very bad results because the cost is not correlated to the distance, whereas the $distance^{(3)}$ is the best, and $distance^{(2)}$ gives also good results.Interpretation : a small rank for distance $distance^{(2)}(A, A')$ or $distance^{(3)}(A, A')$ leads to a good behavior when $Opt(A')$ is used on the random process A and these distances are highly correlated to the cost in generalization.

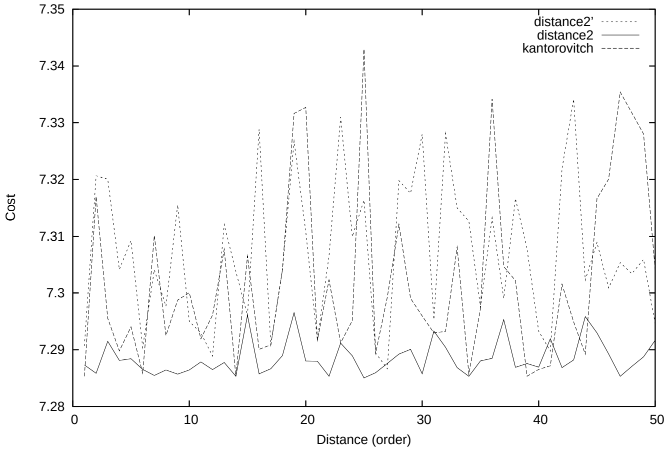


Figure 4: Results in **dimension 2**. This is the same graphic (zoom on the 50 smallest distances) using a two dimensions function instead of dimension one. distance3 is not plotted because we haven't implemented it in dimension greater than 1. We see that we get the same results in dimension 2, that is to say that the Kantorovitch-Rubinstein distance is not correlated to the cost, whereas the $distance^{(2)}$ leads to good results. Hence, choosing a random process A' "close" to A in the Kantorovitch-Rubinstein distance leads to have a very bad estimation of the best cost, and leads to get an apparently-optimal solution (when looking at the cost for A') very far from the real optimum.

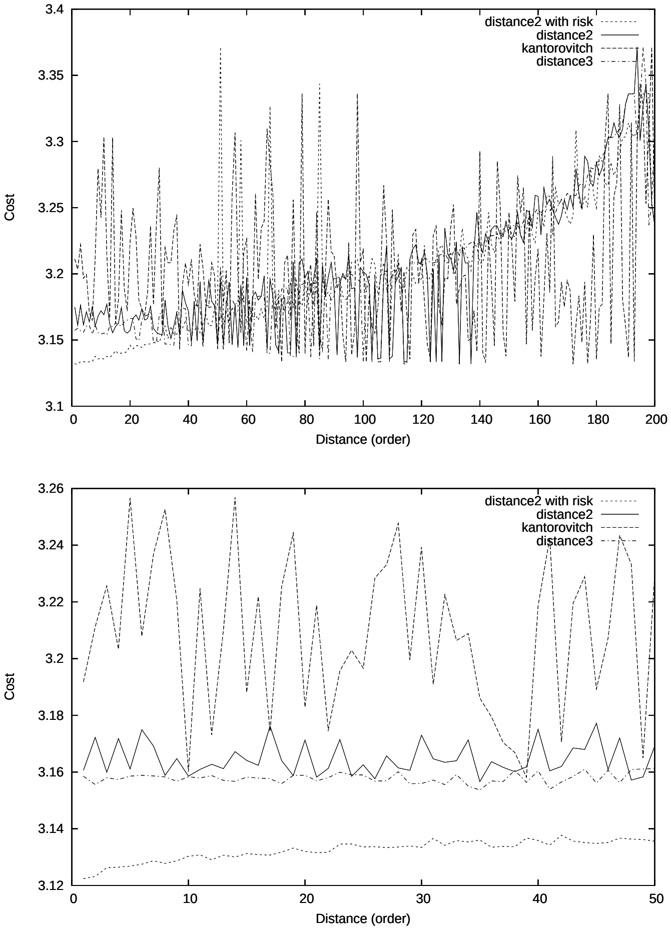


Figure 5: Results in **dimension 1 with noise**. X-coordinate is the rank of A' according to the considered distance. For a "good" distance we expect that small distances are associated to a small cost. The second graphic is a zoom on the smallest distances between A and $\pi(A) = A'$ and with more experiments (the graphic becomes unreadable with too much points on it). "$distance^{(2)}$ with risk" corresponds to the distance introduced in the section 2.3. We can see that the Kantorovitch-Rubinstein distance gives again bad results since the cost is not correlated to the distance, whereas the $distance^{(2)}$, $distance^{(3)}$ and "$distance^{(2)}$ with risk" are better. The really interesting point here is that the "$distance^{(2)}$ with risk" gives much better results than even $distance^{(3)}$ which worked very well without noise. Hence, the added term in $distance^{(2)}$ with risk" ($\frac{1}{2}E\,Tr(H, A)K$) is here validated.