

FUNCTIONAL STOCHASTIC GRADIENT ALGORITHM WITH KERNELS

Kengy Barty¹, Jean-Sébastien Roy^{2,3}, Cyrille Strugarek^{1,2,4}

¹École Nationale des Ponts et Chaussées, ²Électricité de France R&D, ³Université Paris VI, ⁴École Nationale Supérieure de Techniques Avancées

INTRODUCTION

We here propose a new family of algorithms based on the mixing of stochastic approximation techniques, with functional approximation and variational algorithms.

These algorithms allow us to solve numerically without any a priori parametrization of the set of solution, stochastic optimization problems with infinite dimensional command variables.

Gradient-type algorithms for problems like

$$\min_{u \in U^f} \mathbb{E}(j(u(\xi), \xi)), \quad \text{or} \quad H(u) = u, \quad (1)$$

in the Hilbert space $L^2(X, Y)$, usually read:

$$\text{Step } k: \quad u^{k+1} = \Pi_{U^f}(u^k + \gamma^k r^k), \quad (2)$$

with $r^k(\cdot) = -\nabla_u j(u(\cdot), \cdot)$ or $r^k = H(u^k) - u^k$. U^f is a subset of $L^2(X, Y)$ and Π_{U^f} denotes the projection over this subset. Since u is infinite dimensional, so is r^k , and (2) is not implementable.

The new algorithms we propose can be written in the following way, for a command variable denoted by $u \in L^2(X, Y)$:

$$\text{Step } k: \quad u^{k+1} = \Pi_{U^f}\left(u^k + \gamma^k r^k(\xi^{k+1}) \frac{1}{\epsilon^k} K^k(\xi^{k+1})\right), \quad (3)$$

where for all $k \in \mathbb{N}$, u^k belongs to $L^2(X, Y)$, $r^k: X \rightarrow Y$ is a descent mapping, ξ^{k+1} is a random variable with values in X , and $K^k: X \rightarrow L^2(X, \mathbb{R})$ is an approximation mapping called *kernel*. (ϵ^k) and (γ^k) are two nonnegative sequence decreasing to 0.

At each iteration, on the basis of a draw ξ^{k+1} , we compute $r^k(\xi^{k+1}) \frac{1}{\epsilon^k} K^k(\xi^{k+1}) \in L^2(X, Y)$, which builds an approximation of the ideal direction r^k . Since we can choose the kernel mappings K^k such that they are known by a finite number of parameters, the current iterate u^k is perfectly known by a finite number of parameters: hence, it represents an infinite dimensional object.

The convergence of these algorithms has been proved using probabilistic quasi-martingale arguments and usual variational and convexity arguments. The proofs can be found with various point of view in [1] for the general setting, or in [2] for the closed-loop stochastic optimization problems, or in [3] for the application to the solution of Bellman-type equations. The mappings K^k , the draws ξ^{k+1} and the sequences (γ^k, ϵ^k) have to fulfill some conditions in order to get the convergence of the algorithm 3 to the solution of the underlying optimization problem:

$$\sum_{k \in \mathbb{N}} \gamma^k \epsilon^k = +\infty, \quad \sum_{k \in \mathbb{N}} \gamma^k (\epsilon^k)^2 < +\infty, \quad \sum_{k \in \mathbb{N}} (\gamma^k)^2 \epsilon^k < +\infty.$$

The kernels typically look like:

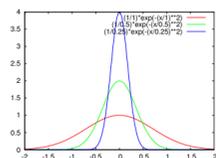


Figure 1: Kernel mappings

LEAST-SQUARE ESTIMATOR

Let us here consider the case of estimating on $[0, 1]$ the real function $x \mapsto \sin\left(\frac{100}{x+1}\right)$. We consider the following cost function:

$$\forall u, x \in \mathbb{R}, \quad j(u, x) = \left(u - \sin\left(\frac{100}{x+1}\right)\right)^2 \quad (4)$$

Let ξ be a real random variable following the uniform law on $[0, 1]$. We define J to be:

$$\forall u \in L^2([0, 1], \mathbb{R}), \quad J(u) = \mathbb{E}(j(u(\xi), \xi)),$$

and we impose the command variable to be a mapping bounded by -0.5 and 0.5 .

We now apply our algorithm to the problem of minimizing J . Figure 2 shows u^k obtained after 50, 200 and 1000 iterations, and the convergence speed of the algorithm.

With the iterations, $[0, 1]$ becomes more and more correctly explored, and hence the feedback converges.

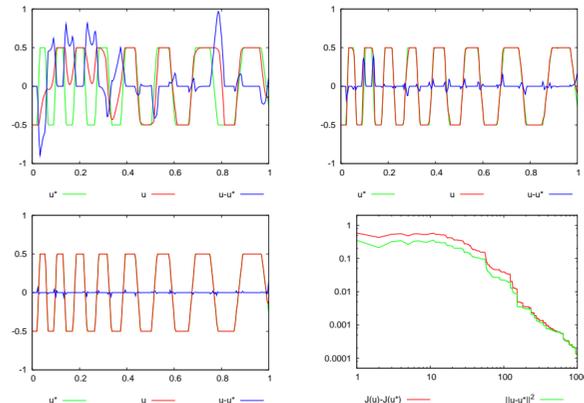


Figure 2: Least Square Problem, feedback after 50, 200 and 1000 iterations, and convergence speed

OPTIMAL CONTROL OF AN HYDRO-POWER PLANT

We consider the problem of managing an hydro-power plant. One has to make two successive production decisions u_1 and u_2 . These decisions have to be taken as feedbacks on successive random selling prices ξ_1, ξ_2 . There is a **measurability constraint** on the first control: the first decision has to be taken prior to any knowledge of the second price, except its conditional law with respect to the first one. Mathematically, we consider the following cost function:

$$j(u_1, u_2, \xi_1, \xi_2) = -u_1 \xi_1 - u_2 \xi_2 - \sqrt{\epsilon + s - u_1 - u_2}, \quad (5)$$

for all $(\xi_1, \xi_2) \in [x_1, \bar{x}_1] \times [x_2, \bar{x}_2]$, and for all $u_1 \in [0, s]$, $u_2 \in [0, s - u_1]$. We take for $i = 1, 2$, ξ_i to be a real random variable with uniform law on $[x_i, \bar{x}_i]$, such that ξ_1 and ξ_2 are independent. Classically, the criterion to be minimized is given by:

$$J(u_1, u_2) = \mathbb{E}(j(u_1(\xi_1), u_2(\xi_1, \xi_2), \xi_1, \xi_2)),$$

with $u_1 \in L^2([x_1, \bar{x}_1], \mathbb{R})$ and $u_2 \in L^2(\Pi_{i=1,2} [x_i, \bar{x}_i], \mathbb{R})$.

We now come to the theoretical solution of this problem. We solve it recursively, using a classical dynamic programming procedure. We first compute the second optimal feedback u_2^* , as a function of the two first prices ξ_1 and ξ_2 and of the first feedback u_1 . It yields:

$$u_2^*(\xi_2, u_1) = \begin{cases} s - u_1 & \text{if } \xi_2 > \frac{1}{2\sqrt{\epsilon}}, \\ \epsilon + s - u_1 - \frac{1}{4(\xi_2)^2} & \text{if } \frac{1}{2\sqrt{\epsilon+s-u_1}} \leq \xi_2 \leq \frac{1}{2\sqrt{\epsilon}}, \\ 0 & \text{if } \xi_2 < \frac{1}{2\sqrt{\epsilon+s-u_1}} \end{cases}$$

We now have to solve the following problem for all ξ_1 , by independence:

$$\min_{u_1 \in [0, s]} -u_1 \xi_1 - \mathbb{E}\left(u_2^*(\xi_2, u_1) \xi_2 + \sqrt{\epsilon + s - u_1 - u_2^*(\xi_2, u_1)}\right)$$

We can express the optimal control $u_1^*(\xi_1)$:

$$u_1^*(\xi_1) = \left[\epsilon + s - \frac{1}{4 \left(\frac{x_2 + \sqrt{2(\bar{x}_2 - x_2)}(\xi_1 - \frac{\bar{x}_2 + x_2}{2}) \right)^2} \right]_0^s.$$

The optimal control u_2^{**} is then given by $u_2^{**}(\xi_1, \xi_2) = u_2^*(\xi_2, u_1^*(\xi_1))$.

We now give few numerical results, with $s = 1, \epsilon = 0.1, x_1 = x_2 = 0.4, \bar{x}_1 = \bar{x}_2 = 2$.

Our algorithm yields the graphs given in Figure 3, giving the evolution of u_1 (top), u_2 (middle) and the error on u_2 (bottom) after respectively 1000, 10000, and 100000 iterations.

Performing the projection on the subset defined by the constraint $u_2 \leq s - u_1$ is quite difficult, requiring the calculation of an expectation which can only be performed numerically. We overcome this difficulty by solving the equivalent penalized problem where u_2 is only constrained to be in $[0, s]$, and $j(u_1, u_2, \xi_1, \xi_2) = a_1 u_1 + a_2 u_2$ for all $u_2 \geq s - u_1$, with a_1 and a_2 being positive penalization constants appropriately chosen.

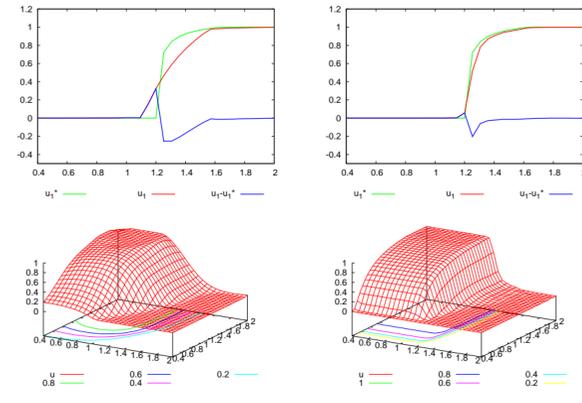


Figure 3: Reservoir Problem with two time periods, feedback at the first time step (top), at the second time step (bottom), after 1000 and 100000 iterations

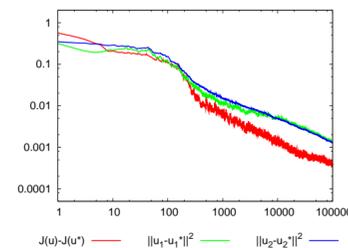


Figure 4: Reservoir Problem with two time periods, convergence speed

PRICING OF A BERMUDAN PUT OPTION

We apply our algorithm to the pricing of a Bermudan put option. A Bermudan put option is an option giving the right to sell the underlying stock at prescribed exercising dates, during a given period, at prescribed prices. It is hence a kind of intermediate between european and american options. In our case, the exercise dates are restricted to equispaced dates t in $0, \dots, T$, and the stock price X_t follows a discretized risk-neutral Black-Scholes dynamics, given by:

$$\forall t \in \mathbb{N}, \quad \ln \frac{X_{t+1}}{X_t} = r - \frac{1}{2} \sigma^2 + \sigma \eta_t$$

where (η_t) is a Gaussian white noise of variance unity, and r is the risk-free interest rate. The strike price is assumed to be s , therefore the intrinsic value of the option when the price is x is $g(x) = \max(0, s - x)$. Let us define the discount factor $\alpha = e^{-r}$. Given the price x_0 at $t = 0$, our objective is to calculate the value of the option:

$$\max_{\tau} \mathbb{E}[\alpha^\tau g(X_\tau) \mid X_0 = x_0],$$

where τ is taken among the stopping times with respect to the filtration generated by the discretized price process (X_t) . In our case, $\tau \in \{0, \dots, T\}$.

Among the multiple methods that have been proposed for option pricing, two share similarities with our approach. [6] describes an approximate dynamic programming approach but neither presents numerical results nor suggests good choices for the basis. Our work directly extends the methodology presented by guaranteeing asymptotic convergence and eliminating the need to choose a basis.

We introduce the Q -functions (Q_t) i.e. the expected payoff at time t if we do not exercise the option. The dynamic programming equation now reads:

$$Q_t(x) = \alpha \mathbb{E}[\max(g(X_{t+1}), Q_{t+1}(X_{t+1})) \mid X_t = x]$$

Such equations can be solved by our algorithm. For the numerical experiment, we take $\mu = 1, \sigma = 1, s = 1, x_0 = 1$ and $r = 0.01$ (and therefore $\alpha = 0.99$).

Figure 5 shows the L^2 error along the iterations, while Figure 6 show the Q -functions $(Q_{t,k})$ along the iterations.

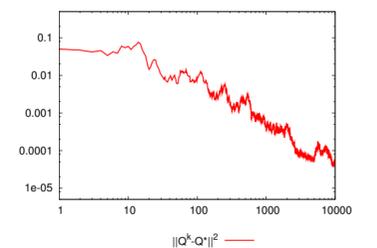


Figure 5: Convergence speed

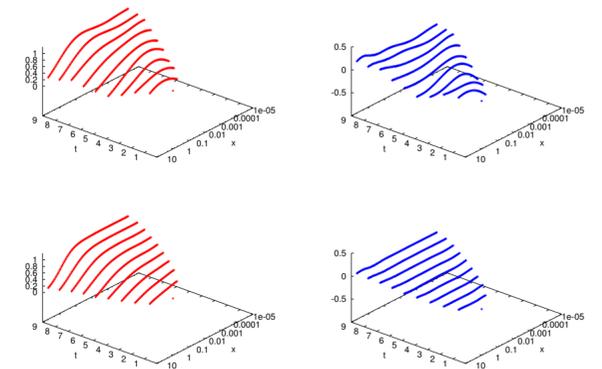


Figure 6: Estimation and error at 100, 10000 iterations.

CONCLUSION

We proposed recently a new approach to solve stochastic optimization problems with closed-loop decisions, and show the convergence of our approach under classical stochastic approximation assumptions.

This new approach improves some theoretical results on Hilbert-valued stochastic gradient schemes (see e.g. [5]), and provides an easier setting than previous ones (see e.g. [4]) to solve infinite dimensional optimization problems.

Moreover, it is easily implementable and yields good numerical results for many interesting applications.

Further work is in progress to use our approach to solve general Q-learning problems, and to improve the convergence speed of our algorithms by tricky choices of the decreasing stepsize sequences.

References

- [1] K. Barty, J.-S. Roy, and C. Strugarek. A perturbed gradient algorithm in Hilbert spaces. *Optimization Online*, 2005. http://www.optimization-online.org/DB_HTML/2005/03/1095.html.
- [2] K. Barty, J.-S. Roy, and C. Strugarek. A stochastic gradient type algorithm for closed loop problems. *submitted to SPEPS*, 2005.
- [3] K. Barty, J.-S. Roy, and C. Strugarek. Temporal difference learning with kernels for pricing american-style options. *submitted to IEEE Trans. Autom. Control*, 2005.
- [4] X. Chen and H. White. Asymptotic properties of some projection-based Robbins-Monro procedures in a Hilbert space. *Stud. Nonlinear Dyn. Econom.*, 6:1-53, 2002.
- [5] J.-B. Hiriart-Urruty. Algorithmes de résolution d'équations et d'inéquations variationnelles. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 33:167-186, 1975.
- [6] J.N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex american-style options. *IEEE Trans. Neural Networks*, 12(4):694-703, July 2001.