In the context of the so-called vanishing discount approach, derivation of sufficient conditions for the existence of optimal feedback control.

ANR-09-SEGI-004

Final report of team CQFD on task 3.3.2

AGENCE NATIONALE DE LA RECHERCHE ANR

INRIA centre de recherche BORDEAUX - SUD-OUEST

This work is concerned with the existence of an optimal control strategy for the long run average continuous control problem of piecewise deterministic Markov processes (PDMP's). In a previous work (see report 3.3.1), sufficient conditions were derived to ensure the existence of an optimal control by using the vanishing discount approach. These conditions were mainly expressed in terms of the relative difference of the α-discount value functions. The main goal of this work is to derive tractable conditions directly related to the primitive data of the PDMP to ensure the existence of an optimal control. The present work can be seen as a continuation of the results derived in our previous work (report 3.3.1). Our main assumptions are written in terms of some integro-differential inequalities related to the so-called expected growth condition, and geometric convergence of the post-jump location kernel associated to the PDMP. An example based on the capacity expansion problem is presented illustrating the possible applications of the results developed in this work.

The partner involved is INRIA CQFD. Professor O. Costa (Escola Politécnica da Universidade de Sao Paulo, Brazil) international expert on optimal stochastic control has worked with the team on this task as an external provider of services.

1 Context

Piecewise-deterministic Markov processes (PDMP's) have been introduced in the literature by M.H.A. Davis [1, 2] as a general class of stochastic models suitable for formulating optimization problems in queueing and inventory systems, maintenance/replacement models, investment scheduling, and many other areas of operations research. PDMP's are a family of Markov processes involving deterministic trajectories punctuated by random jumps. On a state space E the motion of a PDMP is determined by three parameters: the flow ϕ, the jump rate λ, and the transition measure Q. Between two jumps, the trajectory of a PDMP follows the flow ϕ. The jumps occur either spontaneously in a Poisson-like fashion with rate λ or when the flow hits the boundary of the state-space. In either case the location of the process at the jump time is selected by the transition measure Q.

Control of PDMP's has received considerable attention in the literature and the interested reader may consult the book by M.H.A. Davis [2] and the references therein to get a rather complete view of the general theory related to this class of processes. In the previous work (see report 3.3.1) it was studied the long run average continuous control problem of PDMP's taking values in a general Borel space. At each point x of the state space a control variable is chosen from a compact action set $\mathbb{U}(x)$ and is applied on the jump parameter λ and transition measure Q. The goal was to minimize the long run average cost, which was composed of a running cost and a boundary cost (which is added each time the PDMP touches the boundary). Both costs were assumed to be positive but not necessarily bounded. As far as the authors are aware of, this was the first time that this kind of problem was considered in the literature. The approach developed in our previous work (report 3.3.1) to study the long run average control problem of PDMP's was somehow related to the classical analysis of Markov Decision Processes. In particular sufficient conditions were derived in our previous work (see report 3.3.1) to ensure the existence of an optimal control by using the so-called vanishing discount approach. These conditions were mainly expressed in terms of the relative difference $h_\alpha(x) = \mathcal{J}_{\mathcal{D}}^{\alpha}(x) - \mathcal{J}_{\mathcal{D}}^{\alpha}(x_0)$ of the α-discount value functions $\mathcal{J}_{\mathcal{D}}^{\alpha}$. Roughly speaking, it was shown in our previous work (see report 3.3.1) that if there exists a fixed state x_0 such that $\alpha\mathcal{J}_{\mathcal{D}}^{\alpha}(x_0)$ is bounded in a neighborhood of $\alpha = 0$ and if the relative difference h_α satisfies $-K_h \leq h_\alpha(x) \leq b(x)$ for a non-negative constant K_h and a measurable function b then there exists an optimal control. From a practical point of view this result is not completely satisfactory due to the fact that these conditions depend on the α-discount value function $\mathcal{J}_{\mathcal{D}}^{\alpha}$ which may be difficult to be obtained explicitly even for simple examples.

2 Results

The aim of the present work is to overcome this difficulty by providing tractable conditions that are directly related to the primitive data (ϕ, λ, Q) of the PDMP to ensure the existence of an optimal control. The problem of finding sufficient conditions based on the primitive data of the process is a traditional subject in the literature of the long run average control problem for Markov Decision Processes. Without attempting to present an exhaustive panorama on this topic, the interested reader may consult [3, 4, 5, 6, 7] and the references therein for detailed discussions on this problem. More precisely, we present in this paper some assumptions based on integro-differential inequalities related to a positive test function g and $\overline{r}$, and on the geometric convergence of the post-jump location kernel associated to the PDMP, so that, under these hypotheses, we can show that $\alpha\mathcal{J}_{\mathcal{D}}^{\alpha}(x_0)$ is bounded in a neighborhood of $\alpha = 0$ and that the relative difference of the α-discount value function h_α belongs to a weighted-norm space of functions, labeled $\mathbb{B}_g(E)$. Another important difference with respect to the previous work (see report 3.3.1) is that in the present work h_α is not necessarily bounded below by a constant. Indeed, this last property was crucial in the previous work (see report 3.3.1) to show that there exists a solution to an average cost optimality inequality (ACOI) having this same boundedness property, leading to the existence of an optimal control for the PDMP. Here we can only show that a solution to the ACOI exists but belongs to $\mathbb{B}_g(E)$. Consequently, the approach presented in the previous work (see report 3.3.1) cannot be used in the present context and has to be refined.

3 Dissemination of results

The theoretical part of this work has been published in an international peer-reviewed journal *Journal of Applied Probability* Vol. 46, No. 4, pp. 1157-1183, 2009 and is co-authored with O.L.V. Costa (University of Sao Paulo, Brasil).

References

[1] M.H.A. Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *J.Royal Statistical Soc. (B)*, 46:353–388, 1984.

[2] M.H.A. Davis. *Markov Models and Optimization.* Chapman and Hall, London, 1993.

[3] X. Guo and U. Rieder. Average optimality for continuous-time Markov decision processes in polish spaces. *The Annals of Applied Probability*, 16:730–756, 2006.

[4] X. Guo and Q. Zhu. Average optimality for Markov decision processes in Borel spaces: A new condition and approach. *Journal of Applied Probability*, 43:318–334, 2006.

[5] O. Hernández-Lerma and J.B. Lasserre. *Discrete-time Markov control processes*, volume 30 of *Applications of Mathematics.* Springer-Verlag, New York, 1996. Basic optimality criteria.

[6] O. Hernández-Lerma and J.B. Lasserre. *Further topics on discrete-time Markov control processes*, volume 42 of *Applications of Mathematics.* Springer-Verlag, New York, 1999.

[7] Q. Zhu. Average optimality for continuous-time Markov decision processes with a policy iteration approach. *Journal of Mathematical Analysis and Applications*, 339:691–704, 2008.