Derivation of the optimality equation in the context of relaxed control for PDMP's. Proof of the existence of a measurable selector.

ANR-09-SEGI-004

Final report of team CQFD on task 3.3.1

This work deals with the long run average continuous control problem of piecewise deterministic Markov processes (PDMP's) taking values in a general Borel space and with compact action space depending on the state variable. The control variable acts on the jump rate and transition measure of the PDMP, and the running and boundary costs are assumed to be positive but not necessarily bounded. Our first main result is to obtain an optimality equation for the long run average cost in terms of a discrete-time optimality equation related to the embedded Markov chain given by the post-jump location of the PDMP. Our second main result guarantees the existence of a feedback measurable selector for the discrete-time optimality equation by establishing a connection between this equation and an integro-differential equation. Our final main result is to obtain some sufficient conditions for the existence of a solution for a discrete-time optimality inequality and an ordinary optimal feedback control for the long run average cost using the so-called vanishing discount approach.

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

A general family of non-diffusion stochastic models suitable for formulating many optimization problems in several areas of operations research, namely piecewise-deterministic Markov processes (PDMP's), was introduced in [8], and [10]. These processes are determined by three local characteristics; the flow ϕ, the jump rate λ and the transition measure Q. Starting from x the motion of the process follows the flow $\phi(x,t)$ until the first jump time T_1 which occurs either spontaneously in a Poisson-like fashion with rate λ or when the flow $\phi(x,t)$ hits the boundary of the state-space. In either case the location of the process at the jump time T_1 is selected by the transition measure $Q(\phi(x,T_1),.)$ and the motion restarts from this new point as before. A suitable choice of the state space and the local characteristics ϕ, λ, and Q provide stochastic models covering a great number of problems of operations research [10].

As introduced by M.H.A. Davis in [10, page 134], there exist two types of control for PDMP's: *continuous control*, in which the control variable acts at all times on the process through the characteristics (ϕ, λ, Q), and *impulse control*, used to describe control actions that intervene on the process by moving it to a new point of the state space at some specific times. This work deals with 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 is to minimize the long run average cost, which is composed of a running cost and a boundary cost (which is added each time the PDMP touches the boundary). Both costs are assumed to be positive but not necessarily bounded. As far as the authors are aware of, this is the first time that this kind of problem is considered in the literature. Indeed, results are available for the long run average cost problem but for impulse control see Costa [5], Gatarek [16] and the book by M.H.A. Davis [10] (see the references therein). On the other hand, the continuous control problem has been studied only for discounted costs by A. Almudevar [1], M.H.A. Davis [9, 10], M.A.H. Dempster and J.J. Ye [11, 12], Forwick, Schäl, and Schmitz [15], M. Schäl [29], A.A. Yushkevich [33].

2 Approach

Our approach to study the long run average control problem of PDMP's is related to the analysis of Markov Decision Processes (MDP's in short). MDP's have received considerable attention in the literature both in the discrete and continuous-time context. Without attempting to present an exhaustive panorama of MDP's, the interested reader may consult the surveys [2, 19] and the books [4, 21, 22, 28, 30] and the references therein to get a rather complete view of this research field. A possible framework to study continuous-time Markov Decision Processes (MDP's) consists of reducing the original continuous-time control problem into a semi-Markov or discrete-time MDP [3, 14, 28, 31, 32]. For a detailed discussion about these reduction techniques the reader is referred to the recent reference [14]. The reduction method proposed in [14] consists of two steps. First the original continuous-time MDP is converted into a Semi-Markov Decision Process (SMDP) in which the decisions are selected only at the jumps epoch. Second, within the discounted cost context, the SMDP is reduced into a discrete-time MDP. Regarding PDMP's, the idea developed by M.H.A. Davis is somehow related to the reduction technique previously described in the context of MDP's. It consists of reformulating the optimal control problem of a PDMP for a discounted cost as an equivalent discrete-time Markov decision model in which the stages are the jump times T_n of the PDMP. A somewhat different approach to the problem of controlling a PDMP through an embedded discrete time MDP is also considered in [1], in which the decision function space is made compact by permitting piecewise construction of an open-loop control function. It must be stressed the fact that one of the key points in the development of these methods is that the control problem under consideration is concerned with the discounted cost criteria. Obviously, as pointed out in [14], it is well known that a SMDP with discounted cost can be reduced to a MDP with discounted cost. Similarly, the approach adopted in [9] for PDMP's with discounted cost is very natural since the key idea is to re-write the integral cost as a sum of integrals between two consecutive jump times of the PDMP and, by doing this, naturally obtaining the one step cost function for the discrete-time Markov decision model. However, this decomposition for the long run average cost is no longer possible to be done and, therefore, a more specific approach has to be developed. This is one of the goals of the present work. It must be pointed out that there exists another framework for studying continuous-time MDP's in which the controller can choose *continuously* in time the actions to be applied to the process. There exists an extensive literature within this context, see for example [18, 19, 20, 27] and the references therein. This could be another way of studying the control problem for PDMP's with average cost. However, as far as the authors are aware of, it is an open problem to convert a control problem for a PDMP into a continuous-time MDP. In particular, the main problem is how to write explicitly the transition rate of a PDMP in terms of its parameters: the state space E, its boundary ∂E and (ϕ, λ, Q).

3 Results

We consider in this work that the control acts only on (λ, Q). The main difficulty in considering the control acting also on the flow comes from the fact that in such a situation the time $t_*(x)$ which the flow takes to hit the boundary starting from x and the first order differential operator $\mathcal{X}$ associated to the flow would depend on the control. Under these conditions, it is far from obvious to write an optimality equation for the long run average cost in terms of a discrete-time optimality equation related to the embedded Markov chain given by the post-jump location of the PDMP. This step is easier to derive in the situation studied in [9] which considers the control acting on all the local characteristics (ϕ, λ, Q) of the PDMP since, as noted previously, for a discounted cost, it is very natural to re-write the integral cost as a sum of integrals between two consecutive jump

times of the PDMP obtaining naturally the one step cost function for the discrete-time Markov decision model. However, this decomposition for the long run average cost is no longer possible to be done and consequently, due this technical difficulty, the present approach may only be applied to PDMP's in which the control acts on the jump rate and transition measure. Nevertheless this work seems to be the first attempt to study the average continuous control of PDMP's. Furthermore it should be noticed that, as illustrated in the example, in some cases the set up developed in this work can cover some problems in which it is desired to control the flow in a "bang-bang" fashion.

Our first main result is to propose another approach for obtaining an optimality equation for the long run average cost. It is shown that if there exist a measurable function h, a parameter ρ and a measurable selector satisfying a discrete-time optimality equation related to the embedded Markov chain given by the post-jump location of the PDMP, and also that an extra condition involving the function h is verified then an optimal control can be obtained from the measurable selector and ρ is the optimal cost.

Our second main result is to remove the hypothesis of the existence of a measurable selector mentioned in the previous theorem and in fact, to guarantee the existence of a feedback measurable selector (that is, a selector that depends on the present value of the state variable, provided that the function h and parameter ρ satisfy the optimality equation. This is done by establishing a link between the discrete-time optimality equation and an integro-differential equation (using the weaker concept of absolute continuity along the flow of the value function). The common approach for the existence of a measurable selector is to impose semicontinuity properties of the cost function and to introduce the class of relaxed controls to get a compactness property for the action space. By doing this one obtains an existence result but within the class of relaxed controls. However, what is desired is to show the existence of an optimal control in the class of ordinary controls. Combining the existence result within the class of relaxed controls with the connection between the integro-differential equation and the discrete-time equation we can show that the optimal control is non-relaxed and in fact it is an ordinary feedback control.

In general it is a hard task to get the equality in the solution of the discrete-time optimality equation and verify the extra condition. A common approach to avoid this is to consider an inequality instead of equality for the optimality equation, and to use an Abelian result to get the reverse inequality (see for instance [21]). Our last main result is to obtain some sufficient conditions, based on the value function of the discounted control problems, that guarantee the existence of a solution for the discrete-time optimality inequality using the so-called vanishing discount approach (see [21], page 83). The idea of using the vanishing discount approach to get an optimality condition (i.e. a condition for the existence of an average policy) has been widely developed in the literature. Different methods have been proposed based on conditions for ensuring the existence of a solution to the average cost optimality equality, see for example [2, 25], and to the average cost optimality inequality, see for example [23, 24, 26]. More recently, a new approach was proposed in [17, 20]. Combining our result with the link between the integro-differential equation and the discrete-time equation we obtain the existence of an ordinary optimal feedback control for the long run average cost. In order to do that we need first to establish an optimality equation for the discounted control problem. It is worth mentioning that the sufficient condition to be derived in this work is mainly based on the relative difference of the α-discount value functions while in [7] the main goal was to derive conditions directly related to the primitive data of the PDMP to ensure that the vanishing discount approach yields sufficient conditions for the existence of an optimal control.

A closely related paper to our work, but considering the discounted control case, is the paper

by Forwick, Schäl, and Schmitz [15], which also considers unbounded costs and relaxed controls, and obtain sufficient conditions for the existence of ordinary feedback controls. However, in [15] the authors do not consider the long run average cost case neither the related limit problem associated to the vanishing discount approach. Besides, unlike in [15], we consider here boundary jumps and the control action space depending on the state variable. Note however that control on the flow is not considered here, while it was studied in [15]. Finally it is worth mentioning that the authors are studying in a companion work the important question of deriving sufficient stability conditions (like those presented in [6], [13]) under which the conditions on the discounted value function used in the vanishing discount approach are satisfied, tracing a parallel with the discrete-time case (see, for instance, [17, 21]).

4 Dissemination of results

The theoretical part of this work the presentation of academic examples have been published in an international peer-reviewed journal *SIAM Journal of Control and Optimization* Vol. 48, No. 7, pp. 4262-4291, 2010 and is co-authored with O.L.V. Costa (University of Sao Paulo, Brasil).

References

[1] A. Almudevar. A dynamic programming algorithm for the optimal control of piecdewise deterministic Markov processes. *SIAM J. of Control and Optim.*, 40(2):525–539, 2001.

[2] Aristotle Arapostathis, Vivek S. Borkar, Emmanuel Fernández-Gaucherand, Mrinal K. Ghosh, and Steven I. Marcus. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM J. Control Optim.*, 31(2):282–344, 1993.

[3] Dimitri P. Bertsekas. *Dynamic programming and optimal control. Vol. II.* Athena Scientific, Belmont, MA, second edition, 2001.

[4] D.P. Bertsekas and S.E. Shreve. *Stochastic optimal control: The discrete time case*, volume 139 of *Mathematics in Science and Engineering*. Academic Press Inc., New York, 1978.

[5] O.L.V. Costa. Average impulse control of piecewise deterministic processes. *IMA J. Math. Control Inform.*, 6(4):375–397, 1989.

[6] O.L.V. Costa and F. Dufour. Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, 47(2):1053–1077, 2008.

[7] O.L.V. Costa and F. Dufour. The vanishing approach for the average continuous control of piecewise deterministic Markov processes. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 3817 – 3822, Cancun, Mexico, December, 2008.

[8] 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.

[9] M.H.A. Davis. Control of piecewise-deterministic processes via discrete-time dynamic programming. In *Stochastic differential systems (Bad Honnef, 1985)*, volume 78 of *Lecture Notes in Control and Inform. Sci.*, pages 140–150. Springer, Berlin, 1986.

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

[11] M.A.H. Dempster and J.J. Ye. Necessary and sufficient optimality conditions for control of piecewise deterministic processes. *Stochastic and Stochastics Reports*, 40:125–145, 1992.

[12] M.A.H. Dempster and J.J. Ye. Generalized Bellman-Hamilton-Jacob optimality conditions for a control problem with boundary conditions. *Appl. Math. Optimization*, 33:211–225, 1996.

[13] F. Dufour and O.L.V. Costa. Stability of piecewise-deterministic Markov processes. *SIAM J. Control Optim.*, 37(5):1483–1502, 1999.

[14] Eugene A. Feinberg. Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Oper. Res.*, 29(3):492–524, 2004.

[15] L. Forwick, M. Schäl, and M. Schmitz. Piecewise deterministic Markov control processes with feedback controls and unbounded costs. *Acta Appl. Math.*, 82(3):239–267, 2004.

[16] D. Gatarek. Impulsive control of piecewise-deterministic processes with long run average cost. *Stochastics Stochastics Rep.*, 45(3-4):127–143, 1993.

[17] 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.

[18] Xianping Guo and Onésimo Hernández-Lerma. Continuous-time controlled Markov chains with discounted rewards. *Acta Appl. Math.*, 79(3):195–216, 2003.

[19] Xianping Guo, Onésimo Hernández-Lerma, and Tomás Prieto-Rumeau. A survey of recent results on continuous-time Markov decision processes. *Top*, 14(2):177–261, 2006.

[20] Xianping Guo and Ulrich Rieder. Average optimality for continuous-time Markov decision processes in Polish spaces. *Ann. Appl. Probab.*, 16(2):730–756, 2006.

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

[22] 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.

[23] Onésimo Hernández-Lerma. Existence of average optimal policies in Markov control processes with strictly unbounded costs. *Kybernetika (Prague)*, 29(1):1–17, 1993.

[24] Onésimo Hernández-Lerma and Jean B. Lasserre. Average cost optimal policies for Markov control processes with Borel state space and unbounded costs. *Systems Control Lett.*, 15(4):349–356, 1990.

[25] Onésimo Hernández-Lerma, Raúl Montes-de Oca, and Rolando Cavazos-Cadena. Recurrence conditions for Markov decision processes with Borel state space: a survey. *Ann. Oper. Res.*, 28(1-4):29–46, 1991.

[26] Raúl Montes-de Oca and Onésimo Hernández-Lerma. Conditions for average optimality in Markov control processes with unbounded costs and controls. *J. Math. Systems Estim. Control*, 4(1):19 pp. (electronic), 1994.

[27] Tomás Prieto-Rumeau and Onésimo Hernández-Lerma. Bias optimality for continuous-time controlled Markov chains. *SIAM J. Control Optim.*, 45(1):51–73 (electronic), 2006.

[28] Martin L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1994. A Wiley-Interscience Publication.

[29] M. Schäl. On piecewise deterministic Markov control processes: control of jumps and of risk processes in insurance. *Insurance Math. Econom.*, 22(1):75–91, 1998.

[30] Linn I. Sennott. *Stochastic dynamic programming and the control of queueing systems*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1999. A Wiley-Interscience Publication.

[31] Richard F. Serfozo. An equivalence between continuous and discrete time Markov decision processes. *Oper. Res.*, 27(3):616–620, 1979.

[32] A.A. Yushkevich. On reducing a jump controllable Markov model to a model with discrete time. *Theory Probab. Appl.*, 25:58–69, 1980.

[33] A.A. Yushkevich. Verification theorems for Markov decision processes with controlled deterministic drift and gradual and impulsive controls. *Theory Probab. Appl.*, 34(3):474–496, 1989.