ZERO-SUM DISCOUNTED MARKOV GAMES WITH IMPULSE CONTROLS

ARNAB BASU* AND ŁUKASZ STETTNER †

1. Introduction. Given a locally compact separable metric space E, let $\Omega \stackrel{def}{=} \mathcal{D}([0,\infty);E)$ be the space of *cadlag* functions from $\mathbb{R}_+$ to E. We consider the state process $\{X_s\}_{s\geq 0}$ to be a *standard* Markov process ($X_s(\omega) = \omega(s), \omega \in \Omega$) defined on a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ taking values in E. For any space S, we denote by $C(S)$ the space of bounded, continuous and real-valued functions on S. Let $f \in C(E)$ and $h_1, h_2 \in C(E \times E)$ be given. Let $\mathcal{F}_t^X$ and $\mathcal{F}^X$ (resp.) denote the completions of $\sigma\{X_s : s \leq t\}$ and $\sigma\{X_s : s \leq \infty\}$. Let U_1, U_2 be compact subsets of E. We consider a zero-sum game between two players I and II where player I chooses strategy $V_1 \stackrel{def}{=} \{\tau_1, \xi_1; \tau_2, \xi_2; \ldots\}$ to maximize his payoff (described below) and player II chooses strategy $V_2 \stackrel{def}{=} \{\sigma_1, \zeta_1; \delta_2, \zeta_2; \ldots\}$ to minimize the same where $\{\tau_i\}_{i=1,2,\ldots}, \{\sigma_i\}_{i=1,2,\ldots}$ are $\{\mathcal{F}_t^X\}$-measurable stopping times and $\{\xi_i\}_{i=1,2,\ldots}, \{\zeta_i\}_{i=1,2,\ldots}$ are (resp.) $\{\mathcal{F}_{\tau_i}^X\}, \{\mathcal{F}_{\sigma_i}^X\}$-measurable random variables taking values in (resp.) U_1, U_2. To describe the evolution of the controlled Markov process under impulse controls V_1 and V_2 we have to consider an extended probability space $\tilde{\Omega}$ together with probability measure P^{V_1,V_2} (see [3] or [1] for the construction). We denote by $V_i^{>(n)}$ the suffix of V_i starting after the n-th impulse i.e. $V_i^{>(n)} = \Theta_{\rho_n} \circ V_i', V_i' \in \mathcal{V}_i$ and Θ_t is the time-translation operator with $\rho = \tau$ or σ (resp.) depending upon $i = I, II$. The infinite-horizon discounted payoff under strategy-tuple (V_1, V_2) starting at time t at the (random) point $X_t \in E$ is defined as

$$\mathcal{J}^{V_1,V_2}(X_t,t) \stackrel{def}{=} e^{\alpha t} E_{X_t,t}^{V_1,V_2}\left[\int_t^\infty e^{-\alpha s} f(X_s)ds + \sum_{i=1}^{\infty} e^{-\alpha(\tau_i\wedge\sigma_i)}\left(\mathbf{1}_{\{\tau_i\leq\sigma_i\}}h_1(X_{\tau_i}^-,\xi_i) + \mathbf{1}_{\{\sigma_i<\tau_i\}}h_2(X_{\sigma_i}^-,\zeta_i)\right)|\mathcal{F}_t^X\right] \tag{1.1}$$

where, to avoid infinitely many shifts for gain by any player, $h_1(\cdot,\cdot) \leq c < 0$ and $h_2(\cdot,\cdot) \geq d > 0$, $\alpha > 0$ is the discount factor and X^- denotes the value just before the corresponding impulsive shift is made. The interpretation is that player I (resp. player II) chooses a random time τ_i (resp. σ_i) and shifts the process from $X_{\tau_i} \in E$ (resp. $X_{\sigma_i} \in E$) to a point $\xi_i \in U_1$ (resp. $\zeta_i \in U_2$) thereby incurring a negative payoff $h_1(X_{\tau_i}, \xi_i)$ (resp. positive payoff $h_2(X_{\sigma_i}, \zeta_i)$) and this goes on ad infinitum. There is a running payoff denoted by a bounded function $f(\cdot)$ which accumulates over the entire time horizon. We are interested to study the game which starts at $t = 0$ from a given

*Decision Sciences and Information Systems Area, Indian Institute of Management Bangalore, India 560076. (arnab.basu@iimb.ernet.in).

†Department of Probability Theory and Mathematics of Finance, Institute of Mathematics, Polish Academy of Sciences, Warszawa, Poland 00-656, also Vistula University. (stettner@impan.pl).

arbitrarily fixed $x \in E$ with payoff

$$\mathcal{J}^{V_1,V_2}(x) \stackrel{def}{=} E_x^{V_1,V_2}\Bigg[\int_0^\infty e^{-\alpha s} f(X_s)ds + \sum_{i=1}^{\infty} e^{-\alpha(\tau_i \wedge \sigma_i)} \left(\mathbf{1}_{\{\tau_i \le \sigma_i\}} h_1(X_{\tau_i}^-, \xi_i) + \mathbf{1}_{\{\sigma_i < \tau_i\}} h_2(X_{\sigma_i}^-, \zeta_i)\right)\Bigg] \tag{1.2}$$

Note that we omit the notational dependence on t when $t = 0$. The upper and lower values of such a game, starting at $x \in E$, are defined (resp.) as follows:

$$\begin{aligned} \overline{v}(x) &\stackrel{def}{=} \inf_{V_2 \in \mathcal{V}_2} \sup_{V_1 \in \mathcal{V}_1} \mathcal{J}^{V_1,V_2}(x), \\ \underline{v}(x) &\stackrel{def}{=} \sup_{V_1 \in \mathcal{V}_1} \inf_{V_2 \in \mathcal{V}_2} \mathcal{J}^{V_1,V_2}(x) \end{aligned} \tag{1.3}$$

where $\mathcal{V}_1$ and $\mathcal{V}_2$ (resp.) denote the space of strategies of player I and II. It is to be noted here that the game described above is an *online* (and not *offline*) game as might be incorrectly interpreted from the value functions defined in (1.3) above. What this actually means is that the upper value game is as follows:

$$\overline{v}(x) \equiv \inf_{(\sigma_1,\zeta_1)} \sup_{(\tau_1,\xi_1)} E_x^{V_1,V_2}\Bigg[\int_0^{\tau_1 \wedge \sigma_1} e^{-\alpha s} f(X_s)ds + e^{-\alpha(\tau_1 \wedge \sigma_1)} \big(\mathbf{1}_{\{\tau_1 \le \sigma_1\}} h_1(X_{\tau_1}^-, \xi_1) + \mathbf{1}_{\{\sigma_1 < \tau_1\}} h_2(X_{\sigma_1}^-, \zeta_1)\big) + e^{-\alpha(\tau_1 \wedge \sigma_1)} ess\inf_{V_2^{>(1)} \in \mathcal{V}_2} ess\sup_{V_1^{>(1)} \in \mathcal{V}_1} J^{V_1^{>(1)}, V_2^{>(1)}}(\xi_1 \mathbf{1}_{\{\tau_1 \le \sigma_1\}} + \zeta_1 \mathbf{1}_{\{\sigma_1 < \tau_1\}}, \tau_1 \wedge \sigma_1)\Bigg] \tag{1.4}$$

where the essential optima are to be understood again in a recursive sense and the lower value $\underline{v}(x)$ can be correspondingly interpreted. The main results of this paper are to show that such a game has a *value* $v(\cdot)$ i.e. $\overline{v}(x) = \underline{v}(x) \equiv v(x)$ for all $x \in E$ and that there exists optimal saddle-point strategies attaining this value obtained via the unique solution to a corresponding Isaacs' equation for this game. The paper generalizes [1] where similar game was considered under the assumption that the players make their decisions with a deterministic constant delay $h > 0$. In the proofs we also use some results from [2].

REFERENCES

[1] L. Stettner, *Zero-sum Markov Games with Stopping and Impulsive Strategies*, J. Appl. Math. Optimiz. 9 (1982), 1–24,

[2] L. Stettner, *Penalty method for finite horizon stopping problems*, SIAM J. Control Optim., 49 (2011), 1078–1999

[3] L. Stettner, *On Impulsive Control with Long Run Average Cost Criterion*, Studia Math. 76 (1983), 279-298,