Policy iteration algorithm for zero-sum stochastic differential games with ergodic payoff

Departamento de Matemáticas Cinvestav-IPN

50 ANIVERSARIO 1961-2011

DeLaSalle | Universidad La Salle

José Daniel López-Barrientos

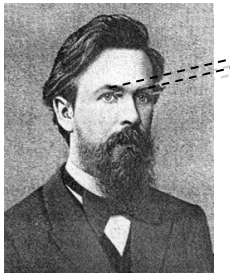
Coloquio de Sistemas Estocásticos

Hotel Holiday Inn – Zócalo

October 28, 2011

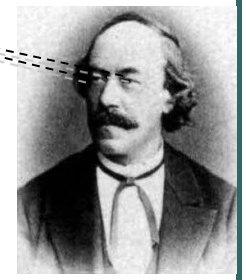
Main problem

For each $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ and $T \geq 0$,



$$J_T(x, \pi^1, \pi^2) := \mathbb{E}_x^{\pi^1, \pi^2}\left[\int_0^T r(x(t), \pi^1, \pi^2)\, dt\right]$$

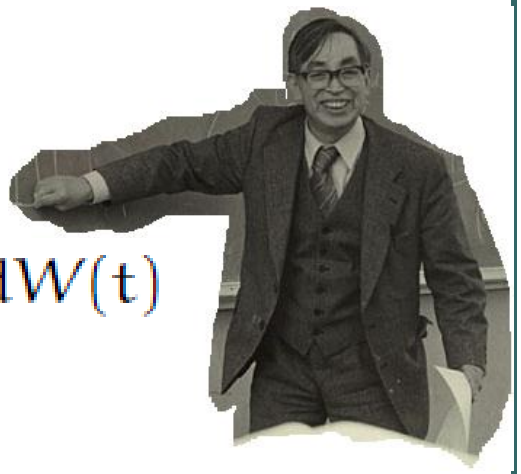
$$J(x, \pi^1, \pi^2) := \limsup_{T \to \infty} \frac{1}{T} J_T(x, \pi^1, \pi^2).$$



Namely, we are interested in pairs of policies $(\pi^1_*, \pi^2_*) \in \Pi^1 \times \Pi^2$ for which

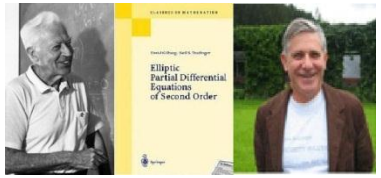
$$J(\pi^1, \pi^2_*) \leq J(\pi^1_*, \pi^2_*) \leq J(\pi^1_*, \pi^2) \quad \text{for every} \quad (\pi^1, \pi^2) \in \Pi^1 \times \Pi^2.$$

Main assumptions



$$dx(t) = b\left(x(t), u_1(t), u_2(t)\right) dt + \sigma\left(x(t)\right) dW(t)$$

For (u_1, u_2) in $U^1 \times U^2$ and h in $\mathcal{C}^2(\mathbb{R}^n)$, let



$$\begin{aligned}
\mathbb{L}^{u_1,u_2} h(x) &:= \langle \nabla h(x), b\left(x, u_1, u_2\right)\rangle + \frac{1}{2}\mathrm{Tr}\left[\mathbb{H}h(x) \cdot a(x)\right] \\
&= \sum_{i=1}^{n} \partial_{x_i} h(x) b_i\left(x, u_1, u_2\right) + \frac{1}{2}\sum_{i,j=1}^{n} \partial^2_{x_i x_j} h(x) a_{ij}(x)
\end{aligned}$$

Main assumptions

When using randomized Markov policies (π^1, π^2) in $\Pi^1 \times \Pi^2$, we will write, for $x \in \mathbb{R}^n$,

$$b(x, \pi^1, \pi^2) := \int_{U^2} \int_{U^1} b(x, u_1, u_2)\, \pi^1(du_1|x)\pi^2(du_2|x).$$

for $h \in \mathcal{C}^2(\mathbb{R}^n)$, let

$$\mathbb{L}^{\pi^1,\pi^2} h(x) := \int_{U^2} \int_{U^1} \mathbb{L}^{u_1,u_2} h(x)\pi^1(du_1|x)\pi^2(du_2|x).$$

Assumption 2.7. *There exists a function* $w \geq 1$ *in* $\mathcal{C}^2(\mathbb{R}^n)$ *and constants* $d \geq c > 0$ *such that*

(a) $\lim_{|x|\to\infty} w(x) = \infty$.

(b) $\mathbb{L}^{u_1,u_2} w(x) \leq -cw(x) + d$ *for all* (u_1, u_2) *in* $U^1 \times U^2$ *and* x *in* $\mathbb{R}^n$.

The PIA

Step 1. Select a strategy $\pi_0^2 \in \Pi^2$, set $m = 0$ and define $J\left(\pi_{-1}^1, \pi_{-1}^2\right) := -\infty$.

Step 2. Find a policy $\pi_m^1 \in \Pi^1$, a constant $J\left(\pi_m^1, \pi_m^2\right)$, and a function $h_m : \mathbb{R}^n \to \mathbb{R}$ such that $\left(J\left(\pi_m^1, \pi_m^2\right), h_m\right)$ is a solution of the so–called *Poisson equation* (6.2); that is,

$$\begin{aligned} J\left(\pi_m^1, \pi_m^2\right) &= \sup_{\varphi \in V^1} \left[r\left(x, \varphi, \pi_m^2\right) + \mathbb{L}^{\varphi, \pi_m^2} h_m(x) \right] && (6.1) \\ &= r\left(x, \pi_m^1, \pi_m^2\right) + \mathbb{L}^{\pi_m^1, \pi_m^2} h_m(x) \quad \text{for all } x \in \mathbb{R}^n. && (6.2) \end{aligned}$$

Observe that

$$J\left(\pi_m^1, \pi_m^2\right) \geq \inf_{\psi \in V^2} \left[r\left(x, \pi_m^1, \psi\right) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right] \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{6.3}$$

If $J\left(\pi_m^1, \pi_m^2\right) = J\left(\pi_{m-1}^1, \pi_{m-1}^2\right)$, then $J\left(\pi_m^1, \pi_m^2\right) = J\left(\pi_*^1, \pi_*^2\right)$. Terminate PIA.

Step 3. Determine a strategy $\pi_{m+1}^2 \in \Pi^2$ that attains the minimum on the right hand side of (6.3), i.e., for all $x \in \mathbb{R}^n$

$$r\left(x, \pi_m^1, \pi_{m+1}^2\right) + \mathbb{L}^{\pi_m^1, \pi_{m+1}^2} h_m(x) = \inf_{\psi \in V^2} \left[r\left(x, \pi_m^1, \psi\right) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right]. \tag{6.4}$$

Step 4. Increase m in 1 and go to step 2.

The PIA

Step 1. Select a strategy $\pi_0^2 \in \Pi^2$, set $m = 0$ and define $J(\pi_{-1}^1, \pi_{-1}^2) := -\infty$.

Step 2. Find a policy $\pi_m^1 \in \Pi^1$, a constant $J(\pi_m^1, \pi_m^2)$, and a function $h_m : \mathbb{R}^n \to \mathbb{R}$ such that $(J(\pi_m^1, \pi_m^2), h_m)$ is a solution of the so–called *Poisson equation* (6.2); that is,

$$J(\pi_m^1, \pi_m^2) = \sup_{\varphi \in V^1} \left[r(x, \varphi, \pi_m^2) + \mathbb{L}^{\varphi, \pi_m^2} h_m(x) \right] \tag{6.1}$$

$$= r(x, \pi_m^1, \pi_m^2) + \mathbb{L}^{\pi_m^1, \pi_m^2} h_m(x) \quad \text{for all } x \in \mathbb{R}^n. \tag{6.2}$$

Observe that

$$J(\pi_m^1, \pi_m^2) \geq \inf_{\psi \in V^2} \left[r(x, \pi_m^1, \psi) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right] \quad \text{for all } x \in \mathbb{R}^n. \tag{6.3}$$

If $J(\pi_m^1, \pi_m^2) = J(\pi_{m-1}^1, \pi_{m-1}^2)$, then $J(\pi_m^1, \pi_m^2) = J(\pi_*^1, \pi_*^2)$. Terminate PIA.

Step 3. Determine a strategy $\pi_{m+1}^2 \in \Pi^2$ that attains the minimum on the right hand side of (6.3), i.e., for all $x \in \mathbb{R}^n$

$$r(x, \pi_m^1, \pi_{m+1}^2) + \mathbb{L}^{\pi_m^1, \pi_{m+1}^2} h_m(x) = \inf_{\psi \in V^2} \left[r(x, \pi_m^1, \psi) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right]. \tag{6.4}$$

Step 4. Increase m in 1 and go to step 2.

Super reference

Nowak, A.S. (1985) *Measurable selection theorems for minimax stochastic optimization problems.* SIAM J. Control Optimization **23**, 466–476.

The PIA

Step 1. Select a strategy $\pi_0^2 \in \Pi^2$, set $m = 0$ and define $J(\pi_{-1}^1, \pi_{-1}^2) := -\infty$.

Step 2. Find a policy $\pi_m^1 \in \Pi^1$, a constant $J(\pi_m^1, \pi_m^2)$, and a function $h_m : \mathbb{R}^n \to \mathbb{R}$ such that $(J(\pi_m^1, \pi_m^2), h_m)$ is a solution of the so–called *Poisson equation* (6.2); that is,

$$J(\pi_m^1, \pi_m^2) = \sup_{\varphi \in V^1} \left[r(x, \varphi, \pi_m^2) + \mathbb{L}^{\varphi, \pi_m^2} h_m(x) \right] \tag{6.1}$$

$$= r(x, \pi_m^1, \pi_m^2) + \mathbb{L}^{\pi_m^1, \pi_m^2} h_m(x) \quad \text{for all } x \in \mathbb{R}^n. \tag{6.2}$$

Observe that

$$J(\pi_m^1, \pi_m^2) \geq \inf_{\psi \in V^2} \left[r(x, \pi_m^1, \psi) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right] \quad \text{for all } x \in \mathbb{R}^n. \tag{6.3}$$

If $J(\pi_m^1, \pi_m^2) = J(\pi_{m-1}^1, \pi_{m-1}^2)$, then $J(\pi_m^1, \pi_m^2) = J(\pi_*^1, \pi_*^2)$. Terminate PIA.

Step 3. Determine a strategy $\pi_{m+1}^2 \in \Pi^2$ that attains the minimum on the right hand side of (6.3), i.e., for all $x \in \mathbb{R}^n$

$$r(x, \pi_m^1, \pi_{m+1}^2) + \mathbb{L}^{\pi_m^1, \pi_{m+1}^2} h_m(x) = \inf_{\psi \in V^2} \left[r(x, \pi_m^1, \psi) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right]. \tag{6.4}$$

Step 4. Increase m in 1 and go to step 2.

T. G. I. Lyapunov!

$$\mu_{\pi^1,\pi^2}(w) := \int_{\mathbb{R}^n} w(x)\mu_{\pi^1,\pi^2}(dx)$$

Proposition 5.1. *The payoff rate* r *is* μ_{π^1,π^2}*–integrable.*

$$J(\pi^1,\pi^2) := \mu_{\pi^1,\pi^2}\left(r\left(\cdot,\pi^1,\pi^2\right)\right) = \int_{\mathbb{R}^n} r\left(x,\pi^1,\pi^2\right)\mu_{\pi^1,\pi^2}(dx).$$

The PIA

Step 1. Select a strategy $\pi_0^2 \in \Pi^2$, set $m = 0$ and define $J(\pi_{-1}^1, \pi_{-1}^2) := -\infty$.

Step 2. Find a policy $\pi_m^1 \in \Pi^1$, a constant $J(\pi_m^1, \pi_m^2)$, and a function $h_m : \mathbb{R}^n \to \mathbb{R}$ such that $(J(\pi_m^1, \pi_m^2), h_m)$ is a solution of the so–called *Poisson equation* (6.2); that is,

$$J(\pi_m^1, \pi_m^2) = \sup_{\varphi \in V^1} \left[r(x, \varphi, \pi_m^2) + \mathbb{L}^{\varphi, \pi_m^2} h_m(x) \right] \tag{6.1}$$

$$= r(x, \pi_m^1, \pi_m^2) + \mathbb{L}^{\pi_m^1, \pi_m^2} h_m(x) \text{ for all } x \in \mathbb{R}^n. \tag{6.2}$$

Observe that

$$J(\pi_m^1, \pi_m^2) \geq \inf_{\psi \in V^2} \left[r(x, \pi_m^1, \psi) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right] \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{6.3}$$

If $J(\pi_m^1, \pi_m^2) = J(\pi_{m-1}^1, \pi_{m-1}^2)$, then $J(\pi_m^1, \pi_m^2) = J(\pi_*^1, \pi_*^2)$. Terminate PIA.

Step 3. Determine a strategy $\pi_{m+1}^2 \in \Pi^2$ that attains the minimum on the right hand side of (6.3), i.e., for all $x \in \mathbb{R}^n$

$$r(x, \pi_m^1, \pi_{m+1}^2) + \mathbb{L}^{\pi_m^1, \pi_{m+1}^2} h_m(x) = \inf_{\psi \in V^2} \left[r(x, \pi_m^1, \psi) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right]. \tag{6.4}$$

Step 4. Increase m in 1 and go to step 2.

An Assumption, a Definition and a Proposition

Assumption . *for each pair* $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, *the process* $x^{\pi^1,\pi^2}(\cdot)$ *is* w*–exponentially–ergodic, that is, there exist constants* C, $\delta > 0$ *such that*

$$\sup_{(\pi^1,\pi^2)\in\Pi^1\times\Pi^2} \left|\mathbb{E}_x^{\pi^1,\pi^2} \nu(x(t)) - \mu_{\pi^1,\pi^2}(\nu)\right| \leq Ce^{-\delta t}\|\nu\|_w w(x)$$

for all $x \in \mathbb{R}^n$, $t \geq 0$, *and* $\nu \in \mathcal{B}_w(\mathbb{R}^n)$.

Definition 6.3. *Let* $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$. *The* bias *of* (π^1, π^2) *is the function* $h_{\pi^1,\pi^2}(x) \in \mathcal{B}_w(\mathbb{R}^n)$ *given by*

$$h_{\pi^1,\pi^2}(x) := \int_0^\infty \left[\mathbb{E}_x^{\pi^1,\pi^2} r(x(t), \pi^1, \pi^2) - J(\pi^1, \pi^2)\right] dt.$$

Proposition 6.5. *For each* $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$, *the pair* $(J(\pi^1, \pi^2), h_{\pi^1,\pi^2})$ *is the unique solution of the Poisson equation* (6.2) *for which the* μ_{π^1,π^2}*–expectation is zero:*

$$\mu_{\pi^1,\pi^2}(h_{\pi^1,\pi^2}) = \int_{\mathbb{R}^n} h_{\pi^1,\pi^2}(x)\mu_{\pi^1,\pi^2}(dx) = 0. \tag{6.9}$$

Moreover, h_{π^1,π^2} *is in* $\mathcal{C}^2(\mathbb{R}^n) \cap \mathcal{B}_w(\mathbb{R}^n)$.

The PIA

Step 1. Select a strategy $\pi_0^2 \in \Pi^2$, set $m = 0$ and define $J(\pi_{-1}^1, \pi_{-1}^2) := -\infty$.

Step 2. Find a policy $\pi_m^1 \in \Pi^1$, a constant $J(\pi_m^1, \pi_m^2)$, and a functi… $(J(\pi_m^1, \pi_m^2), h_m)$ is a solution of the so–called *Poisson equation* (6.2); that is.

$$J(\pi_m^1, \pi_m^2) = \sup \Big[\qquad \qquad \Big] \tag{6.1}$$

$$\qquad \qquad h_m(x) \quad \text{for all } x \in \mathbb{R}^n. \tag{6.2}$$

Observe that

$$\inf_{\psi \in V^2} \left[r(x, \pi_m^1, \psi) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right] \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{6.3}$$

If …$_{m-1}$), then $J(\pi_m^1, \pi_m^2) = J(\pi_*^1, \pi_*^2)$. Terminate PIA.

…mine a strategy $\pi_{m+1}^2 \in \Pi^2$ that attains the minimum on the right hand side of (6.3), i.e., for all

$$r(x, \pi_m^1, \pi_{m+1}^2) + \mathbb{L}^{\pi_m^1, \pi_{m+1}^2} h_m(x) = \inf_{\psi \in V^2} \left[r(x, \pi_m^1, \psi) + \mathbb{L}^{\pi_m^1, \psi} h_m(x) \right]. \tag{6.4}$$

Step 4. Increase m in 1 and go to step 2.

Remark 6.2. The PIA is said to converge if $J(\pi_m^1, \pi_m^2) \to J(\pi_*^1, \pi_*^2) = v_*$.

Convergence of PIA 1/2

Lemma 6.6. *Let* $(\pi^1, \pi^2) \in \Pi^1 \times \Pi^2$ *be an arbitrary pair of randomized stationary strategies. Let* $\pi^1_* \in \Pi^1$ *be such that*

$$J(\pi^1_*, \pi^2) = \sup_{\varphi \in V^1} \left[r(x, \varphi, \pi^2) + \mathbb{L}^{\varphi, \pi^2} h(x) \right] \tag{6.10}$$

$$= r(x, \pi^1_*, \pi^2) + \mathbb{L}^{\pi^1_*, \pi^2} h_{\pi^1_*, \pi^2}(x) \quad \textit{for all} \quad x \in \mathbb{R}^n. \tag{6.11}$$

Let $\pi^2_* \in \Pi^2$ *be such that*

$$\inf_{\psi \in V^2} \left[r(x, \pi^1_*, \psi) + \mathbb{L}^{\pi^1_*, \psi} h_{\pi^1_*, \pi^2}(x) \right] = r(x, \pi^1_*, \pi^2_*) + \mathbb{L}^{\pi^1_*, \pi^2_*} h_{\pi^1_*, \pi^2}(x) \tag{6.12}$$

for all $x \in \mathbb{R}^n$. *Then (a)* $J(\pi^1_*, \pi^2_*) \leq J(\pi^1_*, \pi^2)$, *and (b) if* $J(\pi^1, \pi^2_*) \leq J(\pi^1_*, \pi^2_*)$, *then* (π^1_*, π^2_*) *is a saddle point of the SDG with average payoff.*


yield: the existence of a pair of policies (π^1_*, π^2_*) *in* $\Pi^1 \times \Pi^2$ *that is the limit in the sense of Schäl of the sequence* $\{(\pi^1_m, \pi^2_m) : m = 1, 2, \ldots\}$.

Convergence of PIA 2/2

Theorem 6.8. *let* (π_m^1, π_m^2) *be a pair of randomized stationary policies generated by the PIA. Then* $\{(\pi_m^1, \pi_m^2) : m = 1, 2, ...\}$ *converges in the sense of Schäl to a saddle point* (π_*^1, π_*^2) *of the average SDG. Therefore the PIA converges.*

Proof. By Proposition 6.5 we can ensure the existence of a function $h_m \in \mathcal{C}^2(\mathbb{R}^n) \cap \mathcal{B}_w(\mathbb{R}^n)$ such that (6.1)–(6.2) hold for each $m = 1, 2, ...$ Then we apply Theorem 3.4 with $\xi_m := J(\pi_m^1, \pi_m^2)$ and $\alpha_m := 0$ for each $m = 1, 2, ...$ (the verification of its assumptions is straightforward) to ensure the existence of the function $h_{\pi_*^1, \pi_*^2}$.

The Remark 6.7 asserts the existence of the limit (in the sense of Schäl) of the sequence of policies $\{(\pi_m^1, \pi_m^2)\}$ generated by the PIA.

Observe that (6.1) in step 2 of the PIA ensures that (6.10) holds. Moreover, (6.4) in step 3 yields (6.12). The condition in step 2 of the PIA is accepted when the accummulation point referred to in the last paragraph is attained; and, by Lemma 6.6(b), (π_*^1, π_*^2) is a saddle point of the ergodic game. The result follows from Theorem 5.5. □

A useful result

$$\hat{b}(x, u_1, u_2, h, \alpha) := \langle \nabla h(x), b(x, u_1, u_2) \rangle - \alpha h(x) + r(x, u_1, u_2)$$

$$\hat{\mathbb{L}}_\alpha h(x) := \sup_{u_1 \in U^1} \inf_{u_2 \in U^2} \hat{b}(x, u_1, u_2, h, \alpha) + \frac{1}{2} \mathrm{Tr}\left[\mathbb{H} h(x) a(x)\right]$$

Theorem 3.4. *assume that there exist sequences* $\{h_m\} \subset \mathcal{W}^{2,p}(\Omega)$ *and* $\{\xi_m\} \subset \mathcal{L}^p(\Omega)$, *with* $p > 1$, *and a sequence* $\{\alpha_m\}$ *of positive numbers satisfying that:*

(a) $\hat{\mathbb{L}}_{\alpha_m} h_m = \xi_m$ *in* Ω *for* $m = 1, 2, ...$

(b) There exists a constant M_1 *such that* $\|h_m\|_{\mathcal{W}^{2,p}(\Omega)} \leq M_1$ *for* $m = 1, 2, ...$

(c) ξ_m *converges in* $\mathcal{L}^p(\Omega)$ *to some function* ξ.

(d) α_m *converges to some* α.

Then:

(i) There exist a function $h \in \mathcal{W}^{2,p}(\Omega)$ *and a subsequence* $\{m_k\} \subset \{1, 2, ...\}$ *such that* $h_{m_k} \to h$ *in the norm of* $\mathcal{W}^{1,p}(\Omega)$ *as* $k \to \infty$. *Moreover,*

$$\hat{\mathbb{L}}_\alpha h = \xi \quad in \quad \Omega. \tag{3.4}$$

(ii) If $p > n$, *then* $h_{m_k} \to h$ *in the norm of* $\mathcal{C}^{0,\eta}(\bar{\Omega})$ *for* $\eta < 1 - \frac{n}{p}$. *If, in addition,* ξ *is in* $\mathcal{C}^{0,\beta}(\Omega)$, *with* $\beta \leq \eta$, *then* h *belongs to* $\mathcal{C}^{2,\beta}(\Omega)$.

References

Gilbarg, D., Trudinger, N.S. (1998) *Elliptic Partial Differential Equations of Second Order*. Reprinted version, Springer. Heidelberg.

Hashemi, S.N., Heunis, A.J. (2005) On the Poisson equation for singular diffusions. Stochastics **77**, 155–189.

Hernández–Lerma, O., Lasserre, J.B. (1991) *Policy iteration for Markov control processes on Borel spaces*. Acta Appl. Math. **47**, 125–154.

Hoffman, A.K., Karp, R.M. (1966) *On nonterminating stochastic games*. Management Science, **12**, 359–370.

Howard, R.A. (1960) *Dynamic programming and Markov processes*. MIT Press, Cambridge, Massachusetts.

López–Barrientos J.D. (2011) *Policy Iteration for zero–sum stochastic differential games with ergodic payoff*. Submitted.

Nowak, A.S. (1985) *Measurable selection theorems for minimax stochastic optimization problems*. SIAM J. Control Optimization **23**, 466–476.