Towards a policy improvement algorithm in continuous time

Saul Jacka, Warwick Statistics

Oslo
20 April, 2015

Problem: the optimal control problem: Given a filtered probability space $\Omega, (\mathcal{F}_t), \mathcal{F}, \mathbb{P})$ and a jointly continuous f, find for each x

$$V(x) \stackrel{def}{=} \sup_{\Pi \in \mathcal{A}_x} \mathbb{E}[\int_0^\tau f(X_t^\Pi, \Pi_t)dt + g(X_\tau^\Pi)1_{(\tau<\infty)}]$$

where

(1) X takes value in some topological space S, τ is the first exit from some domain D in S, Π takes values in a (sequentially) compact space A and is adapted.

(2) For each $a \in A$ we assume that the constant process a is in each $\mathcal{A}_x$ and that X^a is a strong Markov process with (martingale) infinitesimal generator L^a and domain $\mathbf{D}^a$. We assume that $\mathbf{C}$ is a nonempty subset of $\cap_{a\in A}\mathbf{D}^a$ with the property that $L^a\phi(x)$ is jointly continuous in $(x, a) \in D \times A$ for each $\phi \in \mathbf{C}$.

(3) $\mathcal{A}_x$ consists of all those adapted processes Π such that there exists a unique adapted, right-continuous process X^Π with

1. $X_0^\Pi = x$;
2. for each $\phi \in \mathbf{C}$,

$$\phi(X_{t\wedge\tau}^\Pi) - \int_0^{t\wedge\tau} L^{\Pi_s}\phi(X_s^\Pi)ds \text{ is a martingale;} \tag{1}$$

   and defining J by

$$J(x,\Pi) = \int_0^\tau f(X_t^\Pi, \Pi_t)dt + g(X_\tau^\Pi)1_{(\tau<\infty)},$$

   we have

3. $$\int_0^{t\wedge\tau} f(X_t^\Pi, \Pi_t)dt + g(X_\tau^\Pi)1_{(\tau<\infty)} \overset{L^1}{\to} J(x,\Pi).$$

We refer to elements of $\mathcal{A}_x$ as controls.

1. *Discounted infinite horizon problem*. Here X^a is a killed Markov process with $S = D \cup \{\partial\}$ with ∂ an isolated cemtery state. Killing to ∂ is at rate α and τ is the death time of the process.
2. *The finite horizon problem*. Here we have Y^a a Markov process on S' with infintesimal generator $\mathcal{G}$. τ is the time to the horizon Then $S = S' \times \mathbb{R}$ and $D = S' \times \mathbb{R}^{++}$, so if $x = (y, T)$ then $X^a_t = (Y^a_t, T - t)$, $\tau = T$ and $L^a = \mathcal{G} - \frac{\partial}{\partial t}$.

We define *Markov policies* as follows: π is a Markov policy if

1. $\pi: \ S \to A$
   and for each $x \in D$ there exists a unique (up to indistinguishability) X such that
2. $X_0 = x$;
3. Π given by $\Pi_t = \pi(X_t)$ is in $\mathcal{A}_x$
4. $X^\Pi = X$.

Hereafter we denote such an X by X^π.

Given x and $\Pi \in \mathcal{A}_x$ we define the *payoff*, $V^\Pi(x) = \mathbb{E}[J(x, \Pi)]$ and in a corresponding fashion for Markov policies π.

We say that a Markov policy is *improvable* if $V^\pi \in \mathbf{C}$ and denote the collection of improvable Markov policies by I.

If π is a Markov policy, we say that π' is an *improvement* of π if,

1. for each $x \in D$

$$\pi'(x) \in \arg\max_{a \in A}[L^a V^\pi(x) + f(x,a)]$$

i.e.

$$L^{\pi'(x)} V^\pi(x) + f(x, \pi'(x)) = \sup_a [L^a V^\pi(x) + f(x,a)],$$

and

2. π' is also a Markov policy.

The PIA works by defining a sequence of improvements and their associated payoffs: so π_{n+1} is the improvement of π_n.

Assumptions A

A1 There exists a non-empty subset I^* of I such that $\pi_0 \in I^*$ implies that $\pi_n \in I^*$ for each n (issue is whether $V^{\pi_n} \in \mathbf{C}$ and whether the sup is attained) and each π_n is continuous.

A2 For $\pi_0 \in I^*$,

$$V^{\pi_{n+1}}(X^{\pi_{n+1}}_{t\wedge\tau}) - V^{\pi_n}(X^{\pi_{n+1}}_{t\wedge\tau}) \xrightarrow{L^1} Z_x \geq 0 \text{ a.s. for each } x \in D.$$

Theorem 1

Under Assumptions A1 and A2,

$$V^{\pi_{n+1}} \geq V^{\pi_n} \text{ for each } n.$$

Assume from now on that Assumptions A1 and A2 hold and that we have fixed a π_0 in I^*.

Assumptions B

A3 V is finite on D.

A4 There is a subsequence $(n_k)_{k\geq 1}$ such that

$$L^{\pi_{n_k+1}} V^{\pi_{n_k}}(x)+f(x,\pi_{n_k+1}(x)) \overset{n\to\infty}{\longrightarrow} 0 \text{ uniformly in } x \in D.$$

A5 For each x, each $\Pi \in \mathcal{A}_x$ and each n

$$V^{\pi_n}(X^{\Pi}_{t\wedge\tau}) \overset{L^1}{\to} g(X^{\Pi}_{t\wedge\tau})1_{(\tau<\infty)}.$$

Theorem 2

Under Assumptions A and B,

$$V^{\pi_n} \uparrow V.$$

Assume from now on that Assumptions A1 to A5 hold and that we have fixed a π_0 in I^*.

Assumptions C

A6 For any $\pi_0 \in I^*$, $(\pi_n)_{n\geq 1}$ is sequentially precompact in the sup norm topology.

A7 For any sequence $\pi_n \in I^*$, if

- $\phi_n \in \mathbf{C}$ for all n and $\phi_n \xrightarrow{n\to\infty} \phi$ pointwise
- $L^{\pi_n}\phi_n \xrightarrow{n\to\infty} Q$
- $\pi_n \xrightarrow{n\to\infty} \pi$ in sup norm

then

$$\phi \in \mathbf{C} \text{ and } L^{\pi}\phi = Q.$$

A8 For each x, each $\Pi \in \mathcal{A}_x$

$$V(X^{\Pi}_{t\wedge\tau}) \xrightarrow{L^1} g(X^{\Pi}_{t\wedge\tau})1_{(\tau<\infty)}.$$

Theorem 3

Under Assumptions A1 to A8, for any π_0 in I^*, there is a subsequence π_{n_k} such that $\pi_{n_k} \stackrel{n\to\infty}{\longrightarrow} \pi^*$ and $V^{\pi^*} = V$

Discounted, infinite horizon controlled diffusion.

Take $D = \mathbb{R}^d$ and $S = \mathbb{R}^d \cup \{\partial\}$ and $\mathbf{C} = C_b^2(\mathbb{R}^d, \mathbb{R})$, the bounded, C^2, real-valued functions on $\mathbb{R}^d$. Suppose that X is a controlled (killed) Ito diffusion in $\mathbb{R}^d$ so that

$$L^a\phi = \frac{1}{2}\sigma(\cdot, a)^T H\phi\sigma(\cdot, a) + \mu(\cdot, a)^T \nabla\phi - \alpha(\cdot, a)\phi,$$

where $H\phi$ is the Hessian $(\frac{\partial^2\phi}{\partial x_i \partial x_j})$. Assume that

Assumption $\aleph 1$ $\sigma(x, a)$, $\mu(x, a)$, $\alpha(x, a)$ and $f(x, a)$ are uniformly (in a) Lipschitz on compacts in $\mathbb{R}^d$ and are continuous in a; α is bounded below by $\lambda > 0$, σ is uniformly elliptic and f is uniformly bounded by M.

Assumption $\aleph 2$ Suppose that the control set A is a compact interval $[a, b]$.

For every $h \in \mathbf{C}$ and $x \in \mathbb{R}^d$, let $I_h(x)$ denote an element of $\arg\max_{a \in A}[L^a h(x,a) + f(x,a)]$.

Assumption $\aleph 3$ If the sequence $(h_n) \in C^2$, if the sequence $(Hh_n)_{n \geq 1}$ is uniformly bounded on compacts, then we may choose the sequence I_{h_n} to be uniformly Lipschitz on compacts.

Remark This assumption is very strong. Neverthless, if σ is independent of a and bounded, $\mu = \mu_1(x) - ma$, $\alpha(x,a) = \alpha_1(x) + ca$ and $f(x,a) = f_1(x) - f_2(a)$ with $f_2 \in C^1$ and with strictly positive derivative on A, and assumptions $\aleph 1$ and $\aleph 2$ hold then $\aleph 3$ holds.

Proposition 4
Under Assumptions $\aleph1$ to $\aleph3$, Assumptions A1 to A8 hold and the PIA converges for π_0 locally Lipschitz.

Proof Note: $L^a\phi$ is jointly continuous if ϕ is in $\mathbf{C}$ and (with the usual trick to deal with killing) (1) holds for any Π such that there is a solution to the killed equation

$$X_t^\Pi = (x + \int_0^t \sigma(X_s^\Pi, \Pi_s)dB_s + \int_0^t \mu(X_s^\Pi, \Pi_s)ds)1_{(t<\tau)} + \partial 1_{(t\geq\tau)}.$$

and any locally Lipschitz π is a Markov policy (by strong uniqueness of the solution to the SDE).

(A1) If π_0 is Lipschitz on compacts then by Assumption $\aleph 3$, A1 holds.

(A3) Boundedness of V (A3) follows from the boundedness of f and the fact that α is bounded away from 0.

(A6) Assumption $\aleph 3$ implies that (π_n) are uniformly Lipschitz and hence sequentially precompact in the sup-norm topology (A6) by the Arzela-Ascoli Theorem.

(A5) $g = 0$ and since α is bounded away from 0, for any Π, $X_t^\Pi \to \partial$. Now $V^n(\partial) = 0$ and so, by bounded convergence, (A5) holds:

$$V^{\pi_n}(X^\Pi_{t\wedge\tau}) \stackrel{L^1}{\to} g(X^\Pi_{t\wedge\tau})1_{(\tau<\infty)}.$$

(A2) Similarly, (A2) holds:

$$V^{\pi_{n+1}}(X^{n+1}_{t\wedge\tau}) - V^{\pi_n}(X^{n+1}_{t\wedge\tau}) \stackrel{L^1}{\to} 0.$$

(A4) (A4) is tricky. Note that we have (A1), (A2) so by Theorem 1, $V^n \uparrow$. Moreover, since (A3) holds, $V^n \uparrow V^{lim}$. Now take a subsequence (n_k) such that $(\pi_{n_k}, \pi_{n_{k+1}}) \to (\pi^*, \tilde{\pi})$ uniformly on compacts. Then the corresponding σ etc. must also converge. Denote the limits by σ^*, $\tilde{\sigma}$ etc. Then (see Friedman [1]), $V^{lim} \in \mathbf{C}_b^2$ and $\nabla V^{n_k}, \nabla V^{n_{k+1}} \to \nabla V^{lim}$, $HV^{n_k}, HV^{n_{k+1}} \to HV^{lim}$ uniformly on compacts and $L^{\tilde{\pi}} V^{lim} + f(\cdot, \tilde{\pi}(\cdot)) = 0$. Now, from the convergence of the derivatives of V^{n_k}, $L^{\pi_{n_{k+1}}} V^{n_k} + f(\cdot, \pi_{n_{k+1}}(\cdot)) \to L^{\tilde{\pi}} V^{lim} + f(\cdot, \tilde{\pi}(\cdot)) = 0$ uniformly on compacts.

(A7) and (A8) From Friedman.

Finite horizon controlled diffusion.
This is very similar to the previous example if we add the requirement that g is Lipschitz and bounded.

Remark In both examples we need to prove that V is continuous before we can apply the usual pde arguments.

Lemma 4
Under Assumptions A1 and A2,

$$L^{\pi_n} V^{\pi_n}(x) + f(x, \pi_n(x)) = 0 \text{ for all } x \in D$$

Proof We know that

$$V^{\pi_n}(X^{\pi_n}_{t\wedge\tau} - \int_0^{t\wedge\tau} L^{\pi_n} V^{\pi_n}(X^{\pi_n}_s)ds$$

is a martingale and the ususal Markovian argument shows that therefore

$$\int_0^{t\wedge\tau} (L^{\pi_n} V^{\pi_n} + f(\cdot, \pi_n(\cdot))(X^{\pi_n}_s)ds = 0.$$

The result then follows from continuity of $L^{\pi_n} V^{\pi_n} + f(\cdot, \pi_n(\cdot))$ and the right continuity of X^{π_n}.

Proof of Theorem 1

Take $\pi_0 \in I^*$ and $x \in D$ and define

$$S_t = (V^{\pi_{n+1}} - V^{\pi_n})(X^{\pi_n}_{t\wedge\tau}).$$

By assumption, $V^{\pi_{n+1}}$ and V^{π_n} are in $\mathbf{C}$ so

$$V^{\pi_k}(X^{\pi_{n+1}}_{t\wedge\tau} - \int_0^{t\wedge\tau} L^{\pi_{n+1}} V^{\pi_k}(X^{\pi_{n+1}}_s)ds$$

is a martingale for $k = n,\ n+1$. So,

$$S_t = (V^{\pi_{n+1}} - V^{\pi_n})(x) + M_{t\wedge\tau} + \int_0^{t\wedge\tau} (L^{\pi_{n+1}} V^{\pi_{n+1}} - L^{\pi_{n+1}} V^{\pi_n})(X^{\pi_{n+1}}_s)ds,$$

where M is a martingale. Thus

$$S_t = (V^{\pi_{n+1}} - V^{\pi_n})(x) + M_{t\wedge\tau} - \int_0^{t\wedge\tau} \sup_a [L^a V^{\pi_n} + f(\cdot, a)](X^{\pi_{n+1}}_s)ds,$$

by Lemma 4 and the definition of π_{n+1} □

Appealing to Lemma 4 again, the integrand is non-negative and hence S is a supermartingale. Taking expectations and letting $t \to \infty$ we obtain the result using A2.

Proof of Theorem 2

From Theorem 1 and A3, $V^{\pi_n} \uparrow V^{lim}$ for a suitable finite limit bounded above by V. Fix x and $\Pi \in \mathcal{A}_x$ and take the subsequence in A4. Set

$$S^k_t = (V^{\pi_{n_k}}(X^{\Pi}_{t\wedge\tau}) + \int_0^{t\wedge\tau} f(X^{\Pi}_s, \Pi_s))ds,$$

It follows that there is a martingale M^k such that

$$\begin{aligned} S^k_t &= S^k_0 + M^k_{t\wedge\tau} + \int_0^{t\wedge\tau}[L^{\Pi_s}V^{\pi_{n_k}} + f(\cdot,\Pi_s)](X^{\Pi}_s)ds \\ &\leq S^k_0 + M^k_{t\wedge\tau} + \int_0^{t\wedge\tau}\max_a[L^a V^{\pi_{n_k}} + f(\cdot,a)](X^{\Pi}_s)ds \\ &= S^k_0 + M^k_{t\wedge\tau} + \int_0^{t\wedge\tau}[L^{\pi_{n_k+1}}V^{\pi_{n_k}} + f(\cdot,\pi_{n_k+1}(\cdot))](X^{\Pi}_s)ds \end{aligned}$$

So

$$\mathbb{E}S^k_t \leq S^k_0 + \mathbb{E}[\int_0^{t\wedge\tau}[L^{\pi_{n_k+1}}V^{\pi_{n_k}} + f(\cdot,\pi_{n_k+1}(\cdot))](X^{\Pi}_s)ds]. \quad (2)$$

Letting $k \to \infty$ in (2) we obtain, by monotone convergence, that

$$\mathbb{E}[(V^{lim}(X^{\Pi}_{t\wedge\tau}) + \int_0^{t\wedge\tau} f(X^{\Pi}_s, \Pi_s))ds] \leq V^{lim}(x).$$

Now by A5, since $V^{lim} \geq V^{\pi_{n_k}}$ for each k we get

$$\liminf_t \mathbb{E}V^{lim}(X^{\Pi}_{t\wedge\tau}) \geq \mathbb{E}g(X^{\Pi}_{t\wedge\tau})1_{(\tau<\infty)},$$

and so $V^{lim} \geq V^{\Pi}$ for each $\Pi \in \mathcal{A}_x$ and so $V^{lim} \geq V$. However $V^{lim} \leq V$ so we have equality □

Proof of Theorem 3

From A5

$$V^{\pi_n}(X^{\Pi}_{t\wedge\tau}) \stackrel{L^1}{\to} g(X^{\Pi}_{t\wedge\tau})1_{(\tau<\infty)}.$$

uniformly in x and now take a subsequence, denoted (m_j) so that

$$\pi_{m_j} \to \pi^*.$$

By A7, $V \in \mathbf{C}$ and $L^{\pi^*}V + f(\cdot, \pi^*(\cdot)) = 0$ so, defining

$$M^j_{t\wedge\tau} \stackrel{def}{=} (V^{\pi^*}(X^{\pi^*}_{t\wedge\tau}) + \int_0^{t\wedge\tau} f((X^{\pi^*}_s, \pi^*(X^{\pi^*}_s))ds,$$

is a martingale. Thus, defining

$$S^j_t \stackrel{def}{=} (V^{\pi_{m_j}}(X^{\pi^*}_{t\wedge\tau}) + \int_0^{t\wedge\tau} f((X^{\pi^*}_s, \pi^*(X^{\pi^*}_s))ds,$$

we have

$$S^j_t = V^{m^j}(X^{\pi^*}_{t\wedge\tau}) + \int_0^{t\wedge\tau} (L^{\pi^*} V^{\pi_{m_j}} + f(\cdot, \pi^*(\cdot))(X^{\pi^*}_s) \qquad (3)$$

Now for each t, $S^j_t \uparrow V(X^{\pi^*}_{t\wedge\tau}) + \int_0^{t\wedge\tau} f((X^{\pi^*}_s, \pi^*(X^{\pi^*}_s))ds$, by A5, while the L^1 limit of the RHS of (3) is $V(x)$ by A4. Letting $t \to \infty$ we obtain the result that $V^{\pi^*} = V$ by A8 $\square$

A Friedman, *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, N.J., 1964.