OPTIMAL CONTROL VIA DYNAMIC PROGRAMMING

Stochastic problem

Andrea Scalenghe

Tesi magistrale

Stochastic Optimal Control

What if the system evolves and is controlled stochastically? We enter the reign of stochastic optimal control.

We will study controlled diffusion processes, which can be represented by:

$$dx(r) = f(r, x(r), u(r))dr + \sigma(r, x(r), u(r))dw(r),\ r \in I_0 \tag{1}$$

where I_0 is a time interval and f, σ are drift and volatility coefficients. The control u is itself a random process.

Markov Processes

Definition

A stochastic process x satisfies the Markov property if there exists a function $p : I_0 \times \Sigma \times I_0 \times \mathcal{B}(\Sigma) \to \mathbb{R}$ such that:
- For all t, s, B the function $x \mapsto p(t, x, s, B)$ is Borel measurable on Σ and for all t, x, s the function $A \mapsto p(t, x, s, B)$ is a probability measure on $(\Omega, \mathcal{F})$. The Chapman-Kolmogorov equation holds for all $s, t, r \in I_0$ such that $t < r < s$:

$$p(t, x, s, B) = \int_{\Sigma} p(r, y, s, B)\, p(t, x, r, dy) \tag{2}$$

And such that for all $r, s \in I_0$ where r, s and for all $B \in \mathcal{B}(\Sigma)$ then:

$$P(x(s) \in B \,|\, \mathcal{F}_r^x) = p(r, x(r), s, B) \tag{3}$$

Where $\mathcal{F}_r^x = \sigma\left(x(l) \;:\; l \in [t_0, r]\right)$.

Markov Transition Kernel

Heuristically, the Markov probability kernel gives the probability that the system starting from x at time t will be in B at time s. We define the expected value of a function of the process given the initial data (t, x) as:

$$E_{tx}\phi(x(s)) = \int_{\Sigma} \phi(y)\, p(t, x, s, dy)$$

for a real-valued Borel-measurable function ϕ.

This gives rise to a linear operator over (somehow integrable) functions:

$$T_{t,s}\phi(x) = E_{tx}\phi(x(s))$$

Diffusion Processes

Generally, a diffusion process x is a Markov process with continuous paths. We also impose the existence of the following functions $a_{ij}(t,x), f_{ij}(t,x)$ and limits:

$$\lim_{h\to 0^+} \int_{x-y>\epsilon} p(t,x,t+h,dy) = 0 \tag{4}$$

$$\lim_{h\to 0^+} \int_{x-y\leq\epsilon} (y_i - x_i)\, p(t,x,t+h,dy) = f_i(t,x) \tag{5}$$

$$\lim_{h\to 0^+} \int_{x-y\leq\epsilon} (y_i - x_i)(y_j - x_j)\, p(t,x,t+h,dy) = a_{ij}(t,x). \tag{6}$$

Under stricter conditions on f, a a diffusion process is described by 1.

Backward Evolution Operator

The backward evolution operator A is defined for measurable functions Φ as:

$$A\Phi(t,x) = \lim_{h \to 0+} \frac{E_{tx}\Phi(t+h, x(t+h)) - \Phi(t,x)}{h} \tag{7}$$

We denote $\mathcal{D}(A)$ for the space of function such that the limit exists. In various contexts, Dynkin's formula holds:

$$E_{tx}\Phi(s, x(s)) - \Phi(t,x) = E_{tx} \int_t^s A\Phi(r, x(r))\, dr \tag{8}$$

Dynkin's Formula for diffusion processes

We get Dynkin by taking E_{tx} over:

$$\begin{aligned}
\Phi(s, x(s)) &= \Phi(t, x) + \int_t^s \Phi_s(r, x(r))\, dr && (9) \\
&= \Phi(t, x) + \int_t^s A\Phi(r, x(r))\, dr && (10) \\
&\quad + \int_t^s D_x\Phi \cdot \sigma(r, x(r))\, dw(r), && (11)
\end{aligned}$$

where the last (stochastic) integral can be seen as a martingale. In particular, if we take Φ to have polynomial growth then $D_x\Phi \cdot \sigma \in \mathbb{L}^2(I_0)$[1].

[1]using estimates on SDE solutions.

Markov Control Processes

Moving to controlled dynamics, the controlled SDE becomes:

$$dx(r) = f(r, x(r), u(r))dr + \sigma(r, x(r), u(r))dw(r), \; r \in I_0$$

The objective is to minimize a cost criterion involving a running cost L and terminal cost Ψ.

$$J(t, x; u) = E_{tx} \left\{ \int_t^\tau L(s, x(s), u(s)) \, ds + \Psi(\tau, x(\tau)) \right\}$$

The operator A is defined for $\Phi \in C_p^{1,2}(\overline{Q}_0)$ as:

$$A^v \Phi = \Phi_t + \sum_{i=1}^{n} f_i(t, x, v)\Phi_{x_i} + \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(t, x, v)\Phi_{x_i x_j} \tag{12}$$

Optimality

Definition

A reference probability system is a tuple $(\Omega, \{\mathcal{F}_s\}, P, \omega)$ such that $\nu = (\Omega, \mathcal{F}_{t_1}, P)$ is a probability space, $\{\mathcal{F}_s\}$ is a filtration on Ω, w is an $\mathcal{F}$-adapted Brownian motion on $[t, t_1]$.
We denote with $\mathcal{A}_{t\nu}$ progressively measurable admissible controls u.
We define:

$$V_\nu = \inf_{u \in \mathcal{A}_{t\nu}} J(t, x; u) \tag{13}$$

$$V_{PM} = \inf_{\nu} V_\nu . \tag{14}$$

Equation 13 and respectively define ν-*optimality* and *optimality*

HJB equation derivation

Let us suppose that $O = \mathbb{R}^n$, then by the dynamic programming principle for every $h < t_1 - t$:

$$V(t,x) = \inf_{u \in \mathcal{A}} E_{tx} \left\{ \int_t^{t+h} L(s, x(s), u(s))\, ds + V(t+h, x(t+h)) \right\}.$$

If we take the constant control $u \equiv v$ then by Dynkin's formula we get:

$$0 \leq E_{tx} \int_t^{t+h} A^v V(s, x(s))\, ds + E_{tx} \int_t^{t+h} L(s, x(s), v)\, ds,$$

dividing by h and taking the limit for $h \to 0^+$:

$$0 \leq A^v V(t,x) + L(t,x,v).$$

Hamiltonian and HJB equation

The Hamiltonian $\mathcal{H}$ is defined as:

$$\mathcal{H}(t, x, p, A) = \sup_{v \in U} \left\{ -f \cdot p - \frac{1}{2} \mathrm{tr}\left[a \cdot A \right] - L \right\}$$

The verification theorem uses this Hamiltonian to provide conditions under which a control is optimal. We'll impose the optimality condition:

$$A^v W = 0 \Rightarrow -\frac{\partial W}{\partial t} + \mathcal{H}(t, x, D_x W, D_x^2 W) = 0$$

Verification Theorem

Theorem

Let $W \in C^{1,2}(Q) \cap C_p(\overline{Q})$ *such that:*

$$\begin{cases} -\frac{\partial W}{\partial t} + \mathcal{H}(t, x, D_x W, D_x^2 W) = 0, & \forall \ (t, x) \in Q \\ V(t, x) = \Psi(t, x), & \forall \ (t, x) \in \partial Q. \end{cases} \tag{15}$$

Then, for any system ν*, initial condition* $(t, x) \in Q$ *and any* $u \in \mathcal{A}_{t\nu}$*, we have* $W(t, x) \leq J(t, x; u)$*. If there exists a system* ν^* *and a control* u^* *which realizes the Hamiltonian's minimum, then:*

$$V_{PM}(t, x) = J(t, x; u^*). \tag{16}$$

Proof Idea

For O bounded and $W \in C^{1,2}(\overline{Q})$ we can apply Ito and get the thesis as we did in Dynkin 9.

In O is unbounded we define:

$$O_\rho = O \cap \left\{ x < \rho \,|\, d(x, \partial O) > \frac{1}{\rho} \right\},\ Q_\rho = [t_0, t_1 - \rho^{-1}] \times O_\rho,$$

get thesis for all $\rho^{-1} < t_1 - t_0$ and pass to the limit showing uniform integrability of both addenda, which together with convergence in probability implies L^1 convergence.