Optimal Stochastic Control: Numerical Methods

Diffusion processes are a very useful tool for modelling differential equations perturbed by a noise. In this context the control of such processes is quite natural. The optimality conditions are obtained by the dynamic programming method (see *Dynamic Programming: Introduction*), which leads to the solution of a nonlinear partial differential equation called the Hamilton–Jacobi equation (see *Hamilton–Jacobi–Bellman Equation*). The simplest is

$$\frac{\partial}{\partial t} y + \min_u \left[\sum_{i=1}^{n} b_i(x,u) \frac{\partial y}{\partial x_i} + c(x,u) \right] + \Delta y = 0$$

with $y(T, x)$ given. In general this equation cannot be solved analytically and numerical finite-element methods are used to approximate its solution.

Another method of approximation is based on discretization of the stochastic control problem. It leads to solving control of Markov chains.

One of the main difficulties is that very often the dimension of the state is large ($x \in \mathbb{R}^n$, n large). In this case we cannot apply the above methods and we search for suboptimal control. For this purpose we discuss three kinds of approach:

(a) optimization in the class of local feedbacks;

(b) Monte Carlo techniques;

(c) the small-noise case.

These three methods lead to the computation of feedbacks of practical interest.

1. *Stochastic Control Problem*

Defined on some probability space $(\Omega, F_t, F, \mathbb{P})$ we consider the controlled diffusion process

$$\mathrm{d}X_t = b[X_t, u(X_t)]\,\mathrm{d}t + \sigma\,\mathrm{d}W_t \tag{1}$$

where X_t denotes the state $\in \mathbb{R}^n$, W_t is a Brownian perturbation, $u: \mathbb{R}^n \to \mathbb{R}^m$ is a feedback control and σ is an (n, n) matrix.

Given $\mathbb{O}$ an open set of $\mathbb{R}^n$ of boundary Γ, U a closed set of $\mathbb{R}^m$, τ a stopping time defined by

$$\tau(\omega) = \arg\min_t \{X_t(\omega) \notin \mathbb{O}\} \tag{2}$$

$c: \mathbb{R}^n_x \times \mathbb{R}^m_u \to \mathbb{R}^+_{c(x,u)}$ an instantaneous cost, $f: \mathbb{R}^n \to \mathbb{R}^+$ a final cost, T a time horizon and λ an actualization rate, we wish to solve one of the following problems

$$y(x,s) = \min_{u \in U} \mathrm{E}\left[\int_s^{T\wedge\tau} c[X_t, u(X_t)]\,\mathrm{d}t + f(X_{T\wedge\tau}) | X(s) = x \right] \tag{3}$$

$$y(x) = \min_{u \in U} \mathrm{E}\left[\int_0^{\tau} \mathrm{e}^{-\lambda t} c(X_t, u(X_t))\,\mathrm{d}t + \mathrm{e}^{-\lambda\tau} f(X_\tau) | X(0) = x \right] \tag{4}$$

$$y(x) = \min_{\mathbb{O}} \mathrm{E}\left[\int_0^{\tau} \mathrm{e}^{-\lambda t} c(X_t)\,\mathrm{d}t + \mathrm{e}^{-\lambda\tau} f(X_\tau) | X(0) = x \right] \tag{5}$$

We suppose in the last case that c is constant in u.

More general situations can be studied, for example modulation of the noise intensity of jump processes, impulsive control and the ergodic control problem. The examples discussed here present the most numerical difficulties of the general situation.

Denoting

$$L(x,u,p,q) = \sum_{i=1}^{n} b_i(x,u) p_i + \sum_{\substack{i=1\\j=1}}^{n} a_{ij} q_{ij} + c(x,u) \tag{6}$$

$$a = \tfrac{1}{2}\sigma\sigma^* \tag{7}$$

the dynamic programming equations of Eqns. (3–5) are, respectively,

$$D_0 y + \min_{u \in U} L(x,u,Dy,D^2y) = 0, \quad x \in \mathbb{O},$$
$$y(x,t) = f(x,t), \quad x \in \Gamma \quad \text{or} \quad t = T \tag{8}$$

$$-\lambda y + \min_{u \in U} L(x,u,Dy,D^2y) = 0, \quad x \in \mathbb{O},$$
$$y(x) = f(x), \quad x \in \Gamma \tag{9}$$

$$\min[-\lambda y + L(x,Dy,D^2y), g - y] = 0, \quad x \in \mathbb{R}^n \tag{10}$$

where

$$D_0 = \partial/\partial t, D_i = \partial/\partial x_i, D_{ij} = \partial^2/\partial x_i \partial x_j,$$
$$D = [D_1, \ldots, D_n] \quad \text{and} \quad D^2 = [D_{ij}]$$

In general these equations cannot be solved analytically, so we discretize them and then solve numerically the corresponding discretized problem. For discretization, two kinds of approach are possible:

(a) discretization of the dynamic programming equation;

(b) discretization of the stochastic control problem.

These two points of view are the subject of the following two sections.

2. *Finite-Element Approximation of the Dynamic Programming Equation*

We discretize Eqn. (9) by the finite-element method, give some convergence results and discuss the resolution of the discretized problem.

2.1 *Variational Formulation of the Dynamic Programming Equation*

We suppose that

$$\underset{(x,u)}{\mathbb{O} \times U} \to \underset{[b(x,u),c(x,u)]}{\mathbb{R}^n \times \mathbb{R}^+}$$

is bounded, Borelian, Lipschitz uniformly in u,

$$\sum_{\substack{i=1\\j=1}}^{n} a_{ij}h_ih_j \geqslant \alpha \sum_{i=1}^{n} h_i^2$$

$$\forall h \in \mathbb{R}^n, \quad \text{for } \alpha \in \mathbb{R} > 0 \quad \text{given} \qquad (11)$$

and $f = 0$ to simplify the discussion. We denote

$$V = H_0^1(\mathbb{O}) = \{z : z \in L^2(\mathbb{O}) \quad \text{and} \quad Dz \in L^2(\mathbb{O}), z(x) = 0 \quad \forall x \in \Gamma\}$$

$$a_\lambda(y, v) = \int_{\mathbb{O}} \left(\sum_{i,j} a_{ij} D_i y D_j v + \lambda y v\right) \mathrm{d}\mathbb{O} \qquad (12)$$

$$H(x, p, u) = \sum_i b_i(x, u)p_i + c(x, u) \qquad (13)$$

$$\hat{H}(x, p) = \min_{u \in U} H(x, p, u) \qquad (14)$$

Then a precise definition of the solution of the dynamic programming equation is the unique solution (Bensoussan and Lions 1978), belonging to V of

$$a_\lambda(y, v) - (\hat{H}(Dy), v) = 0, \quad \forall v \in V \qquad (15)$$

where (,) denotes the scalar product in $L^2(\mathbb{O})$. $\hat{H}$ denotes the operator

$$\underset{z(x)}{L^2(\mathbb{O})} \rightarrow \underset{\hat{H}(x, z(x))}{L^2(\mathbb{O})}$$

The interest of this formulation is that no second derivative appears in Eqn. (15); thus this approach is a method of generalizing the sense of Eqn. (9). Moreover, in this way it is very easy to prove the existence and the unicity of Eqn. (15) (Bensoussan and Lions 1978).

2.2 *Finite-Dimensional Approximation*

Let (V_ξ, B_ξ, C_ξ) be an internal approximation of V (Aubin 1972). That is, $\xi \in \mathbb{R}^+$, $V_\xi = \mathbb{R}^N$, $B_\xi : V_\xi \rightarrow V$ is a linear injection and $C_\xi : V \rightarrow V_\xi$ is linear, such that

$$B_\xi C_\xi v \underset{\xi \to 0}{\rightarrow} v, \quad \forall v \in V$$

A typical example is obtained taking parallelepipedic finite elements Q_1 (linear on each component) (Ciarlet 1978).

To obtain a control interpretation of the discretized problem we approximate:

(a) the set of feedbacks using an internal approximation $(W_\eta, \bar{B}_\eta, \bar{C}_\eta)$ of $L^p(\mathbb{O}; \mathbb{R}^m)$, and

(b) the Hamiltonian $\hat{H}(x, p)$, using an internal approximation $(Z_\eta, B_\eta C_\eta)$ of $L^2(\mathbb{O})$ which satisfies

$$f_1 \geqslant f_2 \Rightarrow B_\eta C_\eta f_1 \geqslant B_\eta C_\eta f_2 \qquad (16)$$

These two internal approximations must be such that

$$\hat{H}_\eta(z) = \min_{\substack{u(x) \in U\\ \forall x \in \mathbb{O}}} B_\eta C_\eta H(x, \bar{B}_\eta \bar{C}_\eta u, z)$$

$$\text{exists } \forall z \in L^2(\mathbb{O}, \mathbb{R}^n) \qquad (17)$$

This condition is realized, for example, when $B_\eta Z_\eta$ and $\bar{B}_\eta W_\eta$ are piecewise-constant functions, on the same partition.

Then the discrete problem is defined by

$$a_\lambda(B_\xi y_\xi, B_\xi v_\xi) - (\hat{H}_\eta(DB_\xi v_\xi), B_\xi v_\xi) = 0, \quad \forall v_\xi \in V_\xi \qquad (18)$$

Let us take the example where $\mathbb{O} =]0, 1[$, $U = [0, 1]$, $L(x, u, p, q) = b(x, u)p + c(x, u) + q$, $\xi = \eta = 1/N$, $B_\xi V_\xi$ piecewise-linear functions, $\bar{B}_\eta W_\eta$ and $B_\eta Z_\eta$ piecewise-constant functions on the partition of $[0, 1]$ defined by $\{]i/N, (i+1)/N], i = 0, \ldots, N-1\}$. Then, for an interior point, ($x \neq 0$, and $x \neq 1$), Eqn. (17) becomes

$$-\frac{\lambda\xi}{6}[y^{i-1} + 4y^i + y^{i+1}]$$
$$+ \tfrac{1}{2} \min_{u^{i-1}} H_{i-1}(u^{i-1}, y^i - y^{i-1})$$
$$+ \tfrac{1}{2} \min_{u^i} H_i(u^i, y^{i+1} - y^i)$$
$$+ \frac{1}{\xi}[y^{i-1} - 2y^i + y^{i+1}] = 0,$$
$$i = 1, \ldots, N-1 \qquad (19)$$

in which we have dispensed with the indices ξ and η and used the notation

$$H_i(u, p) = b_i(u)p + c_i(u), \quad u \in U \qquad (20)$$

$$b_i(u) = \frac{1}{\xi}\int_{i\xi}^{(i+1)\xi} b(u, x)\,\mathrm{d}x, \quad u \in U \qquad (21)$$

$$c_i(u) = \int_{i\xi}^{(i+1)\xi} c(u, x)\,\mathrm{d}x, \quad u \in U \qquad (22)$$

2.3 *Convergence Results*

THEOREM 1. *If the solution of Eqn. (15) belongs to $H^2(\mathbb{O}) = \{f : f, Df, D^2f \in L^2(\mathbb{O})\}$ and y_ξ is the solution of the discretized problem, Eqn. (18), we have*

$$\|y - B_\xi y_\xi\|_V \leqslant k\xi \qquad (23)$$

where k is a constant.

In general we have

$$B_\xi y_\xi \underset{\xi \to 0}{\rightarrow} y \qquad (24)$$

Proofs of such results are given in Quadrat (1975), Goursat and Quadrat (1976) and Cortay-Dumont (1979).

2.4 *Resolution of the Discretized Problem*

The nonlinear operator defined by Eqn. (17) can be written

$$\min_{u_n} [A_{\xi,\eta}(u_\eta)y_\xi + C_{\xi,\eta}(u_\eta)] = 0 \qquad (25)$$

and we can choose an internal approximation (V_ξ, B_ξ, C_ξ) such that all the out-diagonal terms are positive. But any internal approximation does not satisfy this condition; for example, Q_1 finite element for $n \geq 4$. Where the out-diagonal terms are positive, a generalized Howard algorithm gives the solution. The difference from the classical controlled Markov theory is Eqn. (25), which has only a vectorial meaning. The same u_i appears in more than one row (see the example of Eqn. (19)). Nevertheless, the minimizations defining u_i are the same, thanks to the condition of Eqn. (17).

HOWARD ALGORITHM (not using indices ξ and η).
Step 1: $u \to y$ solving $A(u)y + C(u) = 0$.
Step 2: $y \to u$ solving $\min_u A(u)y + C(u)$.

In this way we define a decreasing sequence $(y^r, r \in \mathbb{N})$, $y^r \geq 0$, which converges to the solution of Eqn. (25). An efficient algorithm is obtained by solving step 1 with an iterative method. Then we have to optimize the distribution of the computation effort between steps 1 and 2. For other algorithms see Lions and Mercier (1980).

3. *Discretization of the Stochastic Control Problem*

The approach in Sect. 2 leads to good numerical results, but the implementation of the method is cumbersome and the hypothesis of Eqn. (11) is not very often satisfied. A completely different approach gives convergence in the general situation. We approximate the initial control problem by a Markov chain control problem and we prove the convergence of the probability law of the optimally controlled Markov chain to the law of optimal diffusion. To obtain a precise convergence result we need a generalized meaning of the diffusion process.

3.1 Martingale Problem for Set-Valued Local Characteristics of Diffusion Processes

This approach is a generalization of the results of Stroock and Varadhan (1979) for the problem of Eqn. (3). Given

(a) a probability space (Ω, F_t, F), where $\Omega = \mathscr{C}([0, T]; \mathbb{R}^n)$ is the set of continuous trajectories, F_t is the smallest σ algebra such that $(X_s, s \leq t)$ becomes measurable and $F = F_T$; and

(b) an upper semicontinuous convex set-valued function in a fixed compact set

$$C: \underset{t}{[0, T]} \times \underset{x}{\mathbb{R}^n} \to \mathbb{R}^n \times \underset{C(t,x)=\{(b,a)\}}{\mathscr{S}_n^+} \tag{26}$$

where $\mathscr{S}_n^+$ denotes the set of matrices of order n, symmetric, nonnegative, and C defines for each time t and state x the set of admissible drift terms b and diffusion terms a,

we can give a sense to the set of admissible diffusion processes as the set of probability laws $\mathscr{P}$ defined on (Ω, F_t, F), such that $P \in \mathscr{P}$ satisfies

(a) $$P(X(0) = x) \tag{27}$$

(b) $\exists c(s, \omega) = (b(s, \omega), a(s, \omega))$, F_s measurable, $\forall s \in [0, T]$ such that for all $\phi \in \mathscr{C}_b^{1,2}([0, T], \mathbb{R}^n)$ (the set of functions with first derivative in time and second derivative in x continuous and bounded),

$$c(s, \omega) \in C(s, X_s(\omega)) \tag{28}$$

$$\phi(t, X_t(\omega)) - \phi(t, x) - \int_0^t L_c\phi(s, \omega)\, ds \text{ is a } (P, F_t) \text{ martingale} \tag{29}$$

with

$$L_c = D_0 + \sum_{i=1}^{n} b_i D_i + \sum_{\substack{i=1\\j=1}}^{n} a_{ij} D_{ij} \tag{30}$$

We have generalized to the stochastic situation the well-known notion of set-valued differential equations used in deterministic control theory (Ekeland and Temam 1974).

3.2 Existence of the Stochastic Control Problem

THEOREM 2 (Quadrat 1980). *$\mathscr{P}$ is a compact set.*

Given a final cost function $f: \mathbb{R}^n \to \mathbb{R}$ lower semicontinuous and bounded, we have the following.

COROLLARY. *The control problem*

$$\min_{p \in \mathscr{P}} \mathrm{E} f(X_T) \tag{31}$$

has a solution.

3.3 Time Discretization of the Stochastic Control Problem

As we accept the degeneracy of the diffusion term, the integral term of the cost can be added in the dynamics. Thus there is no loss of generality in taking into account only the final cost in Eqn. (31). With this notion of the diffusion control problem, it is possible to define a finite-dimensional approximation of Eqn. (31). For this purpose we discretize the problem in time.

Let us denote by h the time discretization

$$C_h(s, x) = \mathscr{C}\text{onvx}\left[\left(\bigcup_{\substack{s<t\leq s+h\\|z-x|\leq h^\gamma}} C(t, y) + V(0, \rho h^{1-2\gamma})\right) \cap (\mathbb{R}^n \times \mathscr{S}_n^+)\right] \tag{32}$$

where $0 < \gamma < \frac{1}{2}$, $\rho > 0$, $\mathscr{C}$onvx denotes the convex closure of a set and $V(0, r)$ denotes the sphere of center 0 and radius r. We remark that $C_h \searrow C$ when $h \searrow 0$.

$$\Pi^h(s,x) = \Big\{\pi \in \mathcal{M}_+^1 : \Big[\int (z-x)\pi(\mathrm{d}y), \int (z-x)^{\otimes 2}\pi(\mathrm{d}y)\Big] \in C_h(s,x)h$$

$$\beta > 2, \alpha > 1, \rho > 0;$$

$$\int |z-x|^\beta \pi(\mathrm{d}y) \leq \rho h^\alpha \Big\} \quad (33)$$

Π^h describes precisely the set of admissible transition probabilities of the discrete time process approximating the admissible diffusions. The first condition tells us that the first and second moments of its increase "belong" to C; the second condition is necessary to obtain some compactness properties.

Now we can define the discrete-time dynamic programming equation:

$$\left.\begin{aligned} y(ih,x) &= \min_{\pi\in\Pi_h(ih,x)} \int y((i+1)h,z)\,\pi(\mathrm{d}z) \\ y(T,x) &= f(x) \end{aligned}\right\} \quad (34)$$

We denote $\hat{P}_h$ the probability law of the stochastic process with continuous trajectories obtained by linear interpolation of the discrete-time Markov chain defined by

$$\hat{P}_h(\mathrm{d}x_0, \mathrm{d}x_1, \ldots, \mathrm{d}x_N) = \delta_x(\mathrm{d}x_0)\hat{\pi}_{0,x_0}(\mathrm{d}x_1)\,\hat{\pi}_{h,x_1}(\mathrm{d}x_2)\ldots\hat{\pi}_{(N-1)h,x_{N-1}}(\mathrm{d}x_N) \quad (35)$$

where $\hat{\pi}_{ih,x}$ is a solution of Eqn. (34). Then we have the following convergence result.

THEOREM 3. *$\hat{P}_h$ converges weakly to $\hat{P}$, the optimal solution of Eqn.* (31).

The weak convergence here means that for all $\phi:\Omega\to\mathbb{R}$ bounded, continuous,

$$\int \phi(\omega)\hat{P}_h(\mathrm{d}\omega) \underset{h\to 0}{\to} \int \phi(\omega)\hat{P}(\mathrm{d}\omega)$$

This convergence is sufficient to prove the convergence of the optimal discrete cost to the optimal continuous one.

The discretization in space is obtained by approximating $x\to\Pi^h(t,x)$ by a piecewise-constant set-valued function. Then it is possible to bring back the minimization problem to finite-dimensional linear programs (Quadrat 1980).

The dynamic programming method is useful only when the dimension of the state is small. If this is not so, we search for suboptimal controls. Three kinds of approach are described in the following sections.

4. Optimization in the Class of Local Feedbacks

In general this problem is more difficult to solve than computing the optimal global feedback (Quadrat 1982). In the following particular cases this approach is of practical interest.

4.1 Uncoupled Dynamic Systems

We denote by $I = \{1, 2, \ldots, n\}$ the set of subsystems; for simplicity their states are here only of dimension one. We study the particular case where b_i is a function of x_i and u_i, $\forall i \in I$:

$$b_i : \underset{t}{\mathbb{R}} \times \underset{x_i}{\mathbb{R}} \times \underset{u_i}{V_i} \to \underset{b_i(t,x_i,u_i)}{\mathbb{R}}$$

The noises are not coupled between the subsystems, that is, σ is a diagonal matrix, and set $U = \Pi_{i\in I} U_i$, $U_i \subset \mathbb{R}$. We denote by R a local strategy, that is,

$$R = (R_1, R_2, \ldots, R_n) \text{ with } R_i : \mathbb{R}^+ \times \underset{(t,x_i)}{\mathbb{R}} \to \underset{u_i = R_i(t,x_i)}{U_i}$$

In this situation we have

$$p^R = \prod_{i\in I} p_i^{R_i}$$

with $p_i^{R_i}$ the solution of

$$L_{i,R_i}^{\mathrm{T}} p_i^{R_i} = 0, \qquad p_i^{R_i}(0,\cdot) = \mu_i \quad (36)$$

with $\mu = \Pi_i \mu_i$, superior T denoting transposition, and

$$L_{i,R_i} = D_0 + b_i \circ R_i D_i + a_{ii} D_{ii}$$

where $b \circ R$ denotes $b(t, x, R(t,x))$. Let us denote by

$$c_i^R \circ R_i : \underset{t}{\mathbb{R}^+} \times \underset{x_i}{\mathbb{R}} \to \underset{\int c\circ R(t,x)\Pi_{j\neq i} p_j^{R_j}(t,x_j)\mathrm{d}x_j}{\mathbb{R}^+} \quad (37)$$

the conditional expectation of the instantaneous cost, knowing the information only for the local subsystem i. We have the following sufficient conditions for a strategy to be optimal player by player.

THEOREM 4. *A sufficient condition for a strategy R to be optimal player by player is that*

$$\min_{R_i} [L_{i,R_i} y_i + c_i^R \circ R_i] = 0, \quad i \in I \quad (38)$$

with $c_i^R \circ R_i$ defined by Eqns. (36, 37).

The optimal cost is $\mu_1(y_1) \ldots = \mu_n(y_n)$ with

$$\mu_i(y_i) = \int_{\mathbb{R}} \mu_i(\mathrm{d}x_i) y_i(0, x_i)$$

Let us consider the following algorithm.

ALGORITHM. Given: ϵ, $\nu \in \mathbb{R}^+$

Step 1:

(a) Choose $i \in I$

(b) Solve Eqn. (38)

(c) If $\mu_i(y_i) \leqslant \nu - \epsilon$ then $\nu := \mu_i(y_i)$
$R_i := \arg\min_{R_i} \{L_{i,R_i} y_i + c_i^R \circ R_i\}$

(d) Otherwise choose another $i \in I$ until $\mu_i(y_i) \geqslant \nu - \epsilon$, $\forall i \in I$.

Step 2: When $\mu_i(y_i) \geqslant \nu - \epsilon$, $\forall i \in I$, then $\epsilon := \epsilon/2$, go to step 1.

By this algorithm we obtain a decreasing sequence $\nu^{(n)}$ which converges to a cost optimal player by player. A proof of a discrete version of this algorithm is given in Quadrat and Viot (1980). We have to solve a coupled system of PDEs, but each of them is defined on a space of dimension one. In this way we can optimize, in the class of local feedbacks, systems which are not attainable by the direct approach.

4.2 Systems Having the Product-Form Property

The property that a system has its dynamic uncoupled is very restrictive. In this section we present a class of systems having an uncoupled invariant measure. They are limits of networks of queues of Jackson type. This property can be used to apply to them the results of Sect. 4.1. for the corresponding ergodic control problem, that is,

$$\min_S \lim_{T\to\infty} \frac{1}{T}\int_0^T c \circ S(\omega_t)\, dt$$

Given B, a generator of a Markov chain defined on I, a function $I \times \mathbb{R}_{(i,x)} \to \mathbb{R}_{u_i(x)}$, a matrix $\sigma \in S_n$, $A = \frac{1}{2}\sigma\sigma^*$ and Λ a diagonal matrix satisfying

$$\Lambda B^* + B\Lambda + 2A = 0 \tag{39}$$

we have the following theorem.

THEOREM 5. *The invariant measure of probability p of the diffusion ($b = Bu$, $a = A$) such that Eqn. (39) is true has the product-form property, that is,*

$$p(x) = k \prod_{i=1}^n p_i(x_i), \quad i \in I \tag{40}$$

$$p_i(x_i) = \exp\left(-\frac{1}{\lambda_{ii}}\right)\int_0^{x_i} u_i(s)\, ds \tag{41}$$

where k is a constant of normalization.

Demonstration. The invariant measure p satisfies

$$-\mathrm{div}[bp] + \mathrm{div}[A\,\mathrm{grad}\,p] = 0 \tag{42}$$

Making the change of variables $p = \exp V$ in Eqn. (42), we obtain

$$(\mathrm{grad}\,V, b - A\,\mathrm{grad}\,V) + \mathrm{div}(b - A\,\mathrm{grad}\,V) = 0$$

using Eqn. (41), we have

$$-(\Lambda^{-1}u, (B + A\Lambda^{-1})u) + \mathrm{tr}[(B + A\Lambda^{-1})\,\mathrm{grad}\,u] = 0 \tag{43}$$

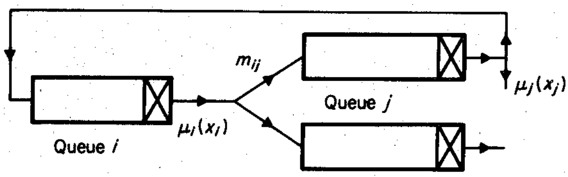

Figure 1
Jackson network of queues for diffusion processes

The quadratic part in (u) of Eqn. (43) is equal to 0 if and only if

$$\Lambda^{-1}B + B^*\Lambda^{-1} + 2\Lambda^{-1}A\Lambda^{-1} = 0$$

which is equivalent to Eqn. (39).

We have also $\mathrm{tr}[B + A\Lambda^{-1}]\,\mathrm{grad}\,u = 0$. Indeed, grad u is diagonal because u_i is a function of x_i only and the coefficient of $D_i u_i$ is $b_{ii} + a_{ii}/\lambda_{ii}$, which is equal to zero, thanks to Eqn. (39).

These diffusion processes are quite natural if we see them as the limit process when $N \to \infty$, obtained from a Jackson network of queues by the scaling $x \to x/N$, $t \to t/N^2$ (Fig. 1), where $\mu_i(x_i)$ is the output rate of the queue i and m_{ij} is the probability of a customer leaving the queue i to go to the queue j.

The correlation of the noise given by Eqn. (39) corresponds to a system for which the noise satisfies a conservation law (for example, the total number of customers in a closed network of queues).

We can now apply the result of Sect. 4.1 to compute the optimal local feedback for systems having the product-form property and an ergodic criterion. Indeed,

$$\min \frac{1}{T}\int_0^T c \circ S(\omega_t)\, dt = \int c \circ S(x) p(x)\, dx$$

$$p(x) = \prod_{i=1}^n p_i(x_i)$$

and p_i satisfies

$$-D_i[u_i p_i] + D_{ii}[\lambda_{ii} p_i] = 0, \quad i \in I$$

$$\int p_i(x_i)\, dx_i = 1$$

4.3 Remarks on Decoupling Feedbacks

Another way of using the results of Sect. 4.1 when the dynamic is coupled is to perform a change of feedback. Let us consider the simpler case:

$$b: \mathbb{R}^n_x \times U_u \to \mathbb{R}^n_{b(x,u)}, \quad U \subset \mathbb{R}^n$$

We use the feedback transformation $v = b(x, u)$ to decouple the drift terms. Now v is the control and we can apply the results of Sect. 4.1 to compute the best local feedback $v_i = R_i(x_i)$. Then the solution in u of

$$b(x, u) = R(x) \tag{44}$$

gives the best feedback among the class that we can call local decoupling feedbacks. The new control must belong to a hypercube, that is, $v \in V = \Pi_{i=1}^{n} V_i \subset b(U)$, $v_i \subset \mathbb{R}$, which in general leads to a loss of optimality.

5. *Optimization in a Parametrized Class of Feedbacks by Monte Carlo Techniques*

In Sect. 4 we computed the optimal local feedback in particular cases. Sometimes the local information is not good; moreover, we may have *a priori* an idea of a better feedback, and would like to use this *a priori* information to solve a simpler problem. A method of doing this is to parametrize the feedback and optimize the open-loop parameter by a Monte Carlo technique. More precisely, given the stochastic control problem

$$\left.\begin{aligned} &\mathrm{d}X_t = b(t, X_t, U_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_t \in \mathbb{R}^n, \quad U_t \in \mathbb{R}^m \\ &\min \mathrm{E} \int_0^T c(t, X_t, U_t)\,\mathrm{d}t \end{aligned}\right\} \tag{45}$$

we make the feedback transformation

$$U(t) = S(t, X_t, v_t), \quad v_t \in \mathbb{R}^p \tag{46}$$

where $S: \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$ is given.

For the approximation of the probability law of the noise M, we use the distribution

$$\mu = \frac{1}{r} \sum_{j=1}^{r} \delta_{\omega_j}(\omega)$$

where ω_j are trajectories of the noise obtained by random generation. We have to solve

$$\left.\begin{aligned} &\mathrm{d}x_t^j = b(t, x_t^j, S(t, x_t^j, v_t))\,\mathrm{d}t + \mathrm{d}w_t^j \\ &\min_v \frac{1}{r} \sum_{j=1}^{r} \int_0^T c(t, x_t^j, S(t, x_t^j, v_t))\,\mathrm{d}t \end{aligned}\right\} \tag{47}$$

where w_t^j denotes a particular trajectory of the noise. Thus in the end we have to solve a deterministic dynamic control problem, for which we can use a gradient technique or the Pontryagin principle. In practice we discretize the problem in time to avoid the difficulty of the generation of diffusion trajectories.

The idea of the stochastic gradient method is the same as the method of Sect. 4 but we use a recursive method of optimization, the recursivity being on the index of the trajectory of the noise generated. The problem of Eqns. (45, 46) can be reduced to the problem (Polyak 1978, Kushner and Clark 1976):

$$\min_{v \in V} \mathrm{E}\, J(v) \tag{48}$$

where

$$J(v) = \int_0^T c(t, X_t, S(t, X_t, v_t))\,\mathrm{d}t \tag{49}$$

We suppose that we are able to compute DJ by an adjoint-state technique. At least after discretization v is finite-dimensional. The stochastic gradient algorithm is the following recursive method of improving the parameter v:

$$v_{r+1} = P_V\{v_r - \rho_r DJ(v_r, \omega_r)\}, \quad \rho_r \in \mathbb{R}^+, \quad \forall r \in N \tag{50}$$

$$\sum_{r \in N} \rho_r = \infty, \qquad \sum_{r \in N} \rho_r^2 < \infty \tag{51}$$

where D denotes $\partial/\partial v$ and ω_r denotes a generated random realization of the stochastic parameter in the definition of $J(v)$; P_V denotes the projection on the set V of admissible parameters.

In a convex situation we have global convergence results. Unfortunately this hypothesis is not true, in general, for the problem of Eqns. (36, 37).

THEOREM 6. *On the hypothesis*

$$v \to J(\omega, v) \text{ convex } \forall \omega$$

$$\int J(\omega, v) M(\mathrm{d}\omega) < \infty\, \forall v$$

$$\sup_{\substack{v \in V \\ \omega \in \Omega}} |DJ(v, \omega)| \leq q \tag{52}$$

$$\mathrm{E}\, J(v) - \hat{J} \geq k l^2(v) \tag{53}$$

(where $\hat{J}$ denotes the optimal cost and $l(v)$ denotes the distance of v to the set of optimal solutions of Eqn. (48))

V a bounded set

we have $\lim \tau \mathrm{E}\, l^2(v_n) = 0$, *and moreover if*

$$\rho_r = 1 \bigg/ \left(kr + \frac{q^2}{y_0 k}\right)$$

with $y_0 = \mathrm{E}\, l^2(v_0)$ *we have*

$$\mathrm{E}\, l^2(v_r) \leq 1 \bigg/ \left(\frac{k^2}{q^2} r + \frac{1}{y_0}\right)$$

The proof of this theorem can be found in Dodu *et al.* (1981).

We suppose now that

(a) the noise is finite-valued and we denote by v_μ arg min $\mathrm{E}_\mu J(v)$, and

(b) $v \to J(\omega, v)$ is twice differentiable and uniformly convex $\forall \omega \in \Omega$.

Then the following result gives a bound on the optimal speed of convergence.

THEOREM 7 (Dodu *et al.* 1981).

$$\mathrm{E}(\hat{v} - v_\mu)^{\otimes 2} \geq \frac{1}{r} H_\mu^{-1} Q_\mu H_\mu^{-1} \tag{54}$$

with

$$H_\mu = D^2 \mathrm{E}_\mu J(v) \quad (55)$$

$$Q_\mu = \mathrm{E}_\mu (DJ(v_\mu))^{\otimes 2} \quad (56)$$

for all unbiased statistics $\hat{v}$ *of* v_μ *defined on* $(\Omega, \mu)^{\otimes r}$.

If we note that $1/k$ is an estimate of H_μ^{-1} and q^2 an estimate of Q_μ, we see that in a certain sense the speed of convergence of the stochastic gradient technique is optimal.

6. Perturbation Methods

By perturbation methods (see *Optimal Control: Perturbation Methods*) we can approximate a difficult problem by a simpler one. In this section we discuss the small noise intensity case. It is possible to construct an affine control which leads to ϵ^4 error with respect to the optimal control, where ϵ denotes the diffusion term. This situation is the common case.

We consider the following stochastic control problem:

$$\left.\begin{array}{l} \mathrm{d}X_t = f(X_t, U_t)\,\mathrm{d}t + \epsilon\,\mathrm{d}W_t, \quad x_t \in \mathbb{R}^n, \quad u_t \in \mathbb{R}^m \\ y^\epsilon(0,x) = \min\limits_u \mathrm{E}\left[\int_0^T c(x_t, u_t)\,\mathrm{d}t \,|\, X(0) = x\right] \end{array}\right\} \quad (57)$$

where ϵ belongs to $\mathbb{R}^+$ and is small. We denote

$$H(x, u, p) = pf(x, u) + c(x, u) \quad (58)$$

and we suppose that

$$u \rightarrow f(x, u) \text{ is linear} \quad (59)$$

$$(v, c_{uu} v) \geqslant k|v|^2 \quad (60)$$

where k is a positive constant, $\forall x$, and we use the notation c_{uu} for $D_{uu}c$. Consider the deterministic control problem

$$\left.\begin{array}{l} \mathrm{d}X_t = f(X_t, U_t)\,\mathrm{d}t \\ y(0,x) = \min\limits_u \left\{\int_0^T c(x_t, u_t)\,\mathrm{d}t \,|\, X(0) = x\right\} \end{array}\right\} \quad (61)$$

and denote by $u_0(t)$ the optimal open-loop deterministic control.

The second-variation calculus around the optimal trajectory of Eqn. (61) gives the osculatory quadratic problem. This quadratic form is defined by the (n, n) time-dependent matrix P, the solution of the Riccati equation

$$\dot{P} + PA + A^*P - PSP + Q = 0, \quad P(T) = 0 \quad (62)$$

where the matrices

$$A = f_x - f_u H_{uu}^{-1} H_{ux} \quad (63)$$

$$S = f_u H_{uu}^{-1} f_u^* \quad (64)$$

$$Q = H_{xx} - H_{ux}^* H_{uu}^{-1} H_{ux} \quad (65)$$

are evaluated along the optimal trajectory of Eqn. (61), with the notation of Eqn. (60).

To give a meaning to Eqns. (63–65), we suppose that

$$H_{uu} > 0, \qquad H_{xx} - H_{ux}^* H_{uu}^{-1} H_{ux} \geqslant 0 \quad (66)$$

Then consider the following affine control:

$$u_f(t, X(t)) = u_0(t) + K(t)[X(t) - X_0(t)] \quad (67)$$

where $X_0(t)$ denotes the optimal trajectory of the deterministic control problem, Eqn. (61), $X(t)$ denotes the actual trajectory of the diffusion process (Eqn. 57) when the control is Eqn. (67) and $K(t)$ is defined by

$$K(t) = H_{uu}^{-1}(H_{ux} + f_u^* P)(t) \quad (68)$$

evaluated along the optimal trajectory $X_0(t)$.

The quality of this affine control is given by the following theorem.

THEOREM 8. *On the hypotheses of Eqns.* (59, 60, 66) *and* (f, c) *twice differentiable, the affine control constructed on the deterministic control problem, used in the stochastic control problem, leads to a loss of optimality of* $\mathcal{O}(\epsilon^4)$.

The proof follows from the results of Cruz (1972) and Fleming (1970). New results in this direction are given in Bensoussan (1987).

See also: Optimal Control: Perturbation Methods; Optimal Feedback: Linear Quadratic Problem; Optimal Feedback: Linear Time-Optimal Control Problem; Stochastic Control: Introduction; Stochastic Maximum Principle; Optimal Stochastic Control: General Aspects; Optimal and Suboptimal Stochastic Control: Discrete-Time Systems; Stochastic Adaptive Systems Stability: Martingale Theory

Bibliography

Aubin J P 1972 *Approximation of Elliptic Boundary-Value Problems.* Wiley–Interscience, New York

Bensoussan A 1987 *Perturbations regulières et singulières en contrôle optimal.* Dunod, Paris

Bensoussan A, Lions J L 1978 *Applications des inéquations variationnelles en contrôle stochastique.* Dunod, Paris

Ciarlet P F 1978 *The Finite Element Method for Elliptic Problems.* North-Holland, Amsterdam

Cortey-Dumont P 1979 Contribution à l'étude de l'approximation par la méthode des éléments finis d'inéquations quasi-variationnelles. *C.R. Hebd. Seances. Acad. Sci., Ser. A.* **288**, 141

Cruz J B 1972 *Feedback Systems.* McGraw-Hill, New York

Dodu J C, Goursat M, Hertz A, Quadrat J P, Viot M 1981 Méthodes de gradient stochastique pour l'optimisation des investissements dans un réseau électrique. *Bull. Dir. Etud. Rech. Bull. Ser. C (Electr. Fr.)* **2**, 134–164

Ekeland I, Temam R 1974 *Analyse convexe et problèmes variationnels.* Dunod Paris

Fleming W H 1971 Control for small noise intensities. *SIAM j. control* **9**, 473–517

Goursat M, Quadrat J P 1975 *Analyse numérique d'inéquations quasi-variationnelles elliptiques associées à des problèmes de contrôle impulsionnelle,* INRIA Report, No. 154 and 186. INRIA, Le Chesnay, France

Kushner H J 1977 *Probability Methods in Stochastic Control and for Elliptic Equations*. Academic Press, New York
Kushner H J, Clark D S 1976 *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer, Berlin
Lions R L, Mercier B 1980 Approximation numériques des équations de Hamilton–Jacobi–Bellman. *Rev. Fr. Autom., Inf. Rech. Oper.* **14**, 369–93
Polyak B T 1978 Subgradient methods: A survey of Soviet research in nonsmooth optimization. In: Lemarechal C, Mifflin R (eds.) *Nonsmooth Optimization*. Pergamon, Oxford
Quadrat J P 1975 *Analyse numérique de l'équation de Bellman stochastique*, INRIA Report No. 140. INRIA, Le Chesnay, France
Quadrat J P 1980 Existence de solution et algorithme de résolution numériques de problèmes stochastiques dégénérées ou non. *SIAM j. control.* **18**, 199–226
Quadrat J P 1982 On optimal stochastic control problem of large systems. *Advances in Filtering and Optimal Stochastic Control*, Lecture Notes in Control and Computer Science Vol. 42. Springer, Berlin
Quadrat J P, Viot M 1980 Product form and optimal local feedback for multi-index Markov chains. *Proc. 18th Allerton Conf. Com. Control and Computing*, pp. 870–81
Stroock F, Varadhan S R S 1979 *Multidimensional Diffusion Process*. Springer, Berlin

M. Goursat and J. P. Quadrat