PDE for Finance Notes, Spring 2003 – Section 5

Notes by Robert V. Kohn, Courant Institute of Mathematical Sciences. For use only in connection with the NYU course PDE for Finance, G63.2706, Spring 2000.

Stochastic optimal control. Stochastic optimal control is like deterministic optimal control except that (i) the equation of state is a *stochastic* differential equation, and (ii) the goal is to maximize or minimize the *expected* utility or cost. To see the structure of the theory in a simple, uncluttered way, we begin by examining what becomes of a standard deterministic utility maximization problem when the state equation is perturbed by a little noise. Then we present a finance classic: Merton's analysis of optimal consumption and investment, in the simplest meaningful case (a single risk-free asset and a risk-free account). Our treatment follows more or less Chapter VI of Fleming and Rishel. For those who wish to learn more, Merton's original paper is highly recommended; it is, in my view, easier to read than Fleming and Rishel. The reference is: "Optimal consumption and portfolio rules in a continuous-time model", *J. Economic Theory* 3, 1971, 373-413 (reprinted as Chapter 5 of his book *Continuous Time Finance*).

****

Perturbation of a deterministic problem by small noise. We've discussed at length the deterministic dynamic programming problem with state equation

$$dy/ds = f(y(s), \alpha(s)) \text{ for } t < s < T, \quad y(t) = x,$$

controls $\alpha(s) \in A$, and objective

$$\max_{\alpha} \left\{ \int_t^T h(y(s), \alpha(s))\, ds + g(y(T)) \right\}.$$

Its value function satisfies the HJB equation

$$u_t + H(\nabla u, x) = 0 \text{ for } t < T, \quad u(x,T) = g(x),$$

with Hamiltonian

$$H(p,x) = \max_{a \in A} \{ f(x,a) \cdot p + h(x,a) \}. \tag{1}$$

Let us show (heuristically) that when the state is perturbed by a little noise, the value function of resulting stochastic control problem solves the perturbed HJB equation

$$u_t + H(\nabla u, x) + \frac{1}{2}\epsilon^2 \Delta u = 0 \tag{2}$$

where H is still given by (1), and $\Delta u = \sum_i \frac{\partial^2 u}{\partial x_i^2}$.

Our phrase "perturbing the state by a little noise" means this: we replace the ODE governing the state by the stochastic differential equation (SDE)

$$dy = f(y, \alpha)ds + \epsilon dw,$$

keeping the initial condition $y(t) = x$. Here dw is a standard, vector-valued Brownian motion (each component w_i is a scalar-valued Brownian motion, and different components are independent).

The evolution of the state is now stochastic, hence so is the value of the utility. Our goal in the stochastic setting is to maximize the *expected* utility. The value function is thus

$$u(x,t) = \max_{\alpha} E_{y(t)=x}\left\{\int_t^T h(y(s),\alpha(s))\,ds + g(y(T))\right\}.$$

There is some subtlety to the question: what is the class of admissible controls? Of course we still restrict $\alpha(s) \in A$. But since the state is random, it's natural for the control to be random as well – however its value at time s should depend only on the past and present, not on the future (which is after all unknown to the controller). Such controls are called "non-anticipating." A simpler notion, sufficient for most purposes, is to restrict attention to *feedback* controls, i.e. to assume that $\alpha(s)$ is a deterministic function of s and $y(s)$. One can show (under suitable hypotheses, when the state equation is a stochastic differential equation) that these two different notions of "admissible control" lead to the same optimal value.

Courage. Let's look for the HJB by applying the usual heuristic argument, based on the principle of dynamic programming applied to a short time interval:

$$u(x,t) \approx \max_{a\in A}\left\{h(x,a)\Delta t + E_{y(t)=x}u(y(t+\Delta t),t+\Delta t)\right\}.$$

The term $h(x,a)\Delta t$ approximates $\int_t^{t+\Delta t} h(y(s),a)\,ds$, because we assume h is smooth and $y(s) = x +$ terms tending to 0 with Δt. Notice that $h(x,a)\Delta t$ is deterministic. The expression $u(y(t+\Delta t), t+\Delta t)$ is the optimal expected utility starting from time $t+\Delta t$ and spatial point $y(t+\Delta t)$. We must take its expected value, because $y(t+\Delta t)$ is random. (If you think carefully you'll see that the Markov property of the process $y(s)$ is being used here.)

We're almost in familiar territory. In the deterministic case the next step was to express $u(y(t+\Delta t), t+\Delta t)$ using the state equation and the Taylor expansion of u. Here we do something analogous: use Ito's lemma and the stochastic differential equation. Ito's lemma says the process $\phi(s) = u(y(s),s)$ satisfies

$$\begin{aligned} d\phi &= \frac{\partial u}{\partial s}ds + \sum_i \frac{\partial u}{\partial y_i}dy_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 u}{\partial y_i \partial y_j}dy_i dy_j \\ &= u_t(y(s),s)ds + \nabla u\cdot(f(y(s),\alpha(s))ds + \epsilon dw) + \frac{1}{2}\epsilon^2\Delta u\,ds. \end{aligned}$$

The real meaning of this statement is that

$$\begin{aligned} u(y(t'),t') - u(y(t),t) &= \int_t^{t'} [u_t(y(s),s) + \nabla u(y(s),s)\cdot(f(y(s),\alpha(s)) + \frac{1}{2}\epsilon^2\Delta u(y(s),s)]ds \\ &\quad + \int_t^{t'} \epsilon\nabla u(y(s),s)\cdot dw. \end{aligned}$$

The expected value of the second integral is 0, so

$$E_{y(t)=x}[u(y(t+\Delta t),t+\Delta t)]-u(x,t)\approx[u_t(x,t)+\nabla u(x,t)\cdot f(x,a)+\frac{1}{2}\epsilon^2\Delta u(x,t)]\Delta t.$$

Assembling these ingredients, we have

$$u(x,t)\approx\max_{a\in A}\left\{h(x,a)\Delta t+u(x,t)+[u_t(x,t)+\nabla u(x,t)\cdot f(x,a)+\frac{1}{2}\epsilon^2\Delta u(x,t)]\Delta t\right\}.$$

This is almost identical to the relation we got in the deterministic case. The only difference is the new term $\frac{1}{2}\epsilon^2\Delta u(x,t)\Delta t$ on the right. It doesn't depend on a, so the optimal a is unchanged – it still maximizes $h(x,a)+f(x,a)\cdot\nabla u$ – and we conclude, as asserted, that u solves (2).

Before going to another topic, let's link this discussion to the notion of "viscosity solution." We noted in Section 4 that the solution of the deterministic HJB equation can be nonunique. (For example, our geometric Example 2 has the HJB equation $|\nabla u|=1$ with boundary condition $u=0$ at the target; it clearly has many almost-everywhere solutions, none of them smooth). We also mentioned in Section 4 that this difficulty can be resolved by working with the "viscosity solution." One characterization of the viscosity solution is this: it is the solution obtained by including a little noise in the problem formulation (with variance ϵ, as above), then taking the limit $\epsilon\to 0$.

****

Optimal portfolio selection and consumption. This is the simplest of a class of problems solved by Robert Merton in his paper "Optimal consumption and portfolio rules in a continuous-time model", *J. Economic Theory* 3, 1971, 373-413 (reprinted in his book *Continuous Time Finance*.) As you'll see, the math is almost the same as our Example 1 – though the finance is more interesting.

We consider a world with one risky asset and one risk-free asset. The risk-free asset grows at a constant risk-free rate r, i.e. its price per share satisfies $dp_1/dt=p_1r$. The risky asset executes a geometric Brownian motion with constant drift $\mu>r$ and volatility σ, i.e. its price per share solves the stochastic differential equation $dp_2=\mu p_2dt+\sigma p_2dw$.

The control problem is this: an investor starts with initial wealth x at time t. His control variables are

$$\begin{aligned}\alpha_1(s)&=\text{fraction of total wealth invested in the risky asset at time } s\\ \alpha_2(s)&=\text{rate of consumption at time } s.\end{aligned}$$

It is natural to restrict these controls by $0\le\alpha_1(s)\le 1$ and $\alpha_2(s)\ge 0$. We ignore transaction costs. The state is the investor's total wealth y as a function of time; it solves

$$dy=(1-\alpha_1)yrdt+\alpha_1y(\mu dt+\sigma dw)-\alpha_2dt$$

so long as $y(s) > 0$. We denote by τ the first time $y(s) = 0$ if this occurs before time T, or $\tau = T$ (a fixed horizon time) otherwise. The investor seeks to maximize the discounted total utility of his consumption. We therefore consider the value function

$$u(x,t) = \max_{\alpha_1,\alpha_2} E_{y(t)=x} \int_t^\tau e^{-\rho s} h[\alpha_2(s)]\, ds$$

where $h[\cdot]$ is a specified utility function (monotone increasing and concave, with $h(0) = 0$). We shall specialize eventually to the power-law utility $h(\alpha_2) = \alpha_2^\gamma$ with $0 < \gamma < 1$. (We have chosen, as in Example 1, to work with the utility discounted to time 0. It is also possible, as in HW1, to work with the utility discounted to time t. The latter choice would give an autonomous HJB equation, i.e. time would not appear explicitly in the equation.)

We find the HJB equation by essentially the same method used above. The principle of dynamic programming applied on a short time interval gives:

$$u(x,t) \approx \max_{a_1,a_2} \left\{ e^{-\rho t} h(a_2)\Delta t + E_{y(t)=x} u(y(t+\Delta t), t+\Delta t) \right\}.$$

To evaluate the expectation term, we use Ito's lemma again. Using the state equation

$$dy = [(1-\alpha_1)yr + \alpha_1 y\mu - \alpha_2]dt + \alpha_1 y\sigma dw$$

and skipping straight to the conclusion, we have

$$u(y(t'),t') - u(y(t),t) = \int_t^{t'} [u_t + u_y[(1-\alpha_1)yr + \alpha_1 y\mu - \alpha_2] + \frac{1}{2}u_{yy}y^2\alpha_1^2\sigma^2]dt + \int_t^{t'} \alpha_1\sigma y u_y dw.$$

The expected value of the second integral is 0, so

$$E_{y(t)=x}[u(y(t+\Delta t),t+\Delta t)] - u(x,t) \approx [u_t + u_y[(1-\alpha_1)yr + \alpha_1 y\mu - \alpha_2 + \frac{1}{2}u_{yy}y^2\alpha_1^2\sigma^2]\Delta t.$$

Assembling these ingredients,

$$u(x,t) \approx \max_{a_1,a_2} \left\{ e^{-\rho t}h(a_2)\Delta t + u(x,t) + [u_t + u_x[(1-a_1)xr + a_1 x\mu - a_2] + \frac{1}{2}u_{xx}x^2a_1^2\sigma^2]\Delta t \right\}.$$

Cleaning up, and taking the limit $\Delta t \to 0$, we get

$$u_t + \max_{a_1,a_2} \left\{ e^{-\rho t}h(a_2) + [(1-a_1)xr + a_1 x\mu - a_2]u_x + \frac{1}{2}x^2a_1^2\sigma^2 u_{xx} \right\} = 0.$$

This is the relevant HJB equation. It is to be solved for $t < T$, with $u(x,T) = 0$ since we have associated no utility associated to final-time wealth.

That looks pretty horrible, but it isn't really so bad. First of all, if we constrain $a_1 = 0$ it reduces to the HJB equation from Example 1. (Well, it has to: if $a_1 = 0$ then all investment is in the risk-free asset, and the problem *is* Example 1.) So we charge ahead.

Let us assume $u_x > 0$ (practically obvious: larger initial wealth should produce larger total utility; what comparison argument would you use to prove it?). Let's also assume $u_{xx} < 0$ (not quite so obvious: this reflects the concavity of the utility function; it will be easy to check it on our explicit solution at the end). Then the optimal a_1 (ignoring the constraint $0 \le a_1 \le 1$) is

$$a_1^* = -\frac{(\mu - r)u_x}{\sigma^2 x u_{xx}}$$

which is positive. We proceed, postponing till later the verification that $a_1^* \le 1$. The optimal a_2 satisfies

$$h'(a_2^*) = e^{\rho t} u_x;$$

we can be sure this a_2^* is positive by assuming that $h'(0) = \infty$.

To go further we now specialize to the power-law utility $h(a_2) = a_2^\gamma$ with $0 < \gamma < 1$. The same argument we used in the deterministic case shows that the solution must have the form

$$u(x,t) = g(t)x^\gamma.$$

The associated a_1^* and a_2^* are evidently

$$a_1^* = \frac{(\mu - r)}{\sigma^2(1-\gamma)}, \quad a_2^* = \left[e^{\rho t} g(t)\right]^{1/(\gamma-1)} x.$$

We assume henceforth that $\mu - r < \sigma^2(1-\gamma)$ so that $a_1^* < 1$. Substituting these values into the HJB equation gives, after some arithmetic,

$$\frac{dg}{dt} + \nu\gamma g + (1-\gamma) g (e^{\rho t} g)^{1/(\gamma-1)} = 0$$

with

$$\nu = r + \frac{(\mu - r)^2}{2\sigma^2(1-\gamma)}.$$

We must solve this with $g(T) = 0$. This is the same nonlinear equation we dealt with in Example 1 – with ν in place of r. So we can go straight to the answer: $u = g(t)x^\gamma$ with

$$g(t) = e^{-\rho t}\left[\frac{1-\gamma}{\rho - \nu\gamma}\left(1 - e^{-\frac{(\rho-\nu\gamma)(T-t)}{1-\gamma}}\right)\right]^{1-\gamma}.$$

It should not be surprising that we had to place some restrictions on the parameters to get this solution. When these restrictions fail, inequalities that previously didn't bother us become important (namely the restrictions $0 \le a_1 \le 1$, which prohibit borrowing and short-selling).

We have solved the HJB equation; but have we found the value function? The answer is yes, as we now show using a verification argument.

****

The verification argument. In the deterministic case we used a heuristic argument to derive the HJB equation, but then showed completely honestly that a (sufficiently smooth) solution of the HJB equation (satisfying appropriate boundary or final-time conditions) provides a bound on the value attainable by any control. A similar result holds in the stochastic setting.

Rather than give a general result at this time, let's focus on the example just completed (Merton's optimal selection and consumption problem). All the ideas required for the general case are already present here. Brief review of our task: the state equation is

$$dy = [(1-\alpha_1)yr + \alpha_1 y\mu - \alpha_2]dt + \alpha_1 y\sigma dw$$

which we shall write for simplicity as

$$dy = f(y, \alpha_1, \alpha_2)dt + \alpha_1 y\sigma dw.$$

The value function is

$$u(x,t) = \max_{\alpha} E_{y(t)=x} \int_t^{\tau} e^{-\rho s} h[\alpha_2(s)]\, ds$$

where τ is either the first time $y = 0$ (if this happens before time T) or $\tau = T$ (if y doesn't reach 0 before time T). The HJB equation is

$$v_t + \max_{a_1,a_2} \left\{ e^{-\rho t} h(a_2) + f(x, a_1, a_2)v_x + \frac{1}{2}x^2 a_1^2 \sigma^2 v_{xx} \right\} = 0$$

for $t < T$, with $v = 0$ at $t = T$. We didn't fuss over it before, but clearly v should also satisfy $v(0, s) = 0$ for all s. We write v instead of u, to reserve notation u for the optimal value. The goal of the verification argument is to show that $v \geq u$, i.e. to show that no control strategy can achieve an expected discounted utility better than v. Our argument will also show that the feedback strategy associated with the HJB calculation – namely

$$\alpha_1(s) = -\frac{(\mu - r)v_x}{\sigma^2 x v_{xx}}(y(s), s), \quad h'(\alpha_2)(s) = e^{\rho s} v_x(y(s), s) \tag{3}$$

does indeed achieve expected discounted value v; in other words $v \leq u$. This suffices of course to show $v = u$.

Consider any control $\tilde{\alpha}(s)$, and the associated state $\tilde{y}(s)$ starting from $\tilde{y}(t) = x$. Of course we assume $\tilde{\alpha}$ is non-anticipating, i.e. it depends only on knowledge of $\tilde{y}(s)$ in the present and past, not the future. (If this condition confuses you, just assume $\tilde{\alpha}$ is given by a feedback law, i.e. $\tilde{\alpha}(s) = F(y(s), s)$ for some deterministic function $F(y, s)$. Such controls are automatically non-anticipating.) We wish to show that

$$v(x,t) \geq E_{y(t)=x} \int_t^{\tilde{\tau}} e^{-\rho s} h[\tilde{\alpha}_2(s)]\, ds.$$

Consider $\phi(s) = v(\tilde{y}(s), s))$: by the Ito calculus it satisfies

$$\begin{aligned} d\phi &= v_s ds + v_y d\tilde{y} + \frac{1}{2} v_{yy} d\tilde{y} d\tilde{y} \\ &= v_s ds + v_y [f(\tilde{\alpha}, \tilde{y})ds + \tilde{\alpha}_1(s)\tilde{y}(s)\sigma dw] + \frac{1}{2} v_{yy} \tilde{\alpha}_1^2(s)\tilde{y}^2(s)\sigma^2 ds. \end{aligned}$$

Therefore

$$v(\tilde{y}(t'), t') - v(\tilde{y}(t), t) = \int_t^{t'} [v_s + v_y f + \frac{1}{2} v_{yy} \tilde{y}^2 \tilde{\alpha}_1^2 \sigma^2] dt + \int_t^{t'} \sigma \tilde{\alpha}_1 \tilde{y} v_y dw$$

where each integrand is evaluated at $y = \tilde{y}(s)$, $\alpha = \tilde{\alpha}(s)$ at time s. The expected value of the second integral is 0 (here is where we use that α is nonanticipating; we will return to this when we discuss stochastic integrals). Thus taking the expectation, and using the initial condition:

$$E\left[v(\tilde{y}(t'), t')\right] - v(x, t) = E\left[\int_t^{t'} (v_s + v_y f + \frac{1}{2} v_{yy} \tilde{y}^2 \tilde{\alpha}_1^2 \sigma^2) dt\right].$$

Now from the definition of the Hamiltonian we have

$$v_t(\tilde{y}(s), s) + \left\{ e^{-\rho s} h(\tilde{\alpha}_2(s)) + f(\tilde{y}(s), \tilde{\alpha}(s)) v_y(\tilde{y}(s), s) + \frac{1}{2} \tilde{y}^2(s) \tilde{\alpha}_1^2(s) \sigma^2 v_{yy}(\tilde{y}(s), s) \right\} \le 0. \tag{4}$$

Combining this with the preceding relation gives

$$E\left[v(\tilde{y}(t'), t')\right] - v(x, t) \le -E\left[\int_t^{t'} e^{-\rho s} h(\tilde{\alpha}_2(s) ds\right].$$

Taking $t' = \tilde{\tau}$ and using the fact that $v(\tilde{y}(t'), t') = 0$, we conclude that

$$v(x, t) \ge E\left[\int_t^{\tilde{\tau}} e^{-\rho s} h(\tilde{\alpha}(s) ds\right].$$

Maximizing the right hand side over all $\tilde{\alpha}$ we conclude that

$$v \ge u$$

For the special feedback law associated with the HJB equation, which fixes the control α by (3), relation (4) becomes equality. This shows that $v \le u$, since v is the value achieved by a specific control strategy and u is the maximum value over all possible strategies. Thus $v = u$. In summary: the function v, defined by solving the HJB equation with appropriate boundary and initial conditions, is in fact the value function of this stochastic control problem, and the control strategy (3) is indeed optimal.

Notice that this calculation rests on pretty much the same tools we used to derive the HJB: (a) the Ito calculus, to get a representation of $u(\tilde{y}(s), s)$, and (b) the fact that any integral "dw" has expected value 0.