PDE in Finance, Spring 2008,
`http://www.math.nyu.edu/faculty/goodman/teaching/PDEfin/index.html`
Written by Jonathan Goodman. Please report errors, sources of confusion, and typos on the class Blackboard site.
Last updated April 5, 2008.

Section 7: Deterministic control and viscosity solutions

These notes outline the theory of viscosity solutions to first order Hamilton Jacobi equations that arise from deterministic dynamic optimal control problems. There are merely an outline. The complete theory involves subtleties and technicalities that we have no time for here. The book *Partial Differential Equations*, by L. C. Evans, has a chapter on this subject that is a great place to look if you want the complete mathematical picture. The book *Applied Optimal Control: Optimization, Estimation, and Control*, by Arthur Bryson and Yu-Chi Ho, has a beautiful discussion that shows how hard this subject was before the viscosity solution theory. It also has a great discussion of the calculus of variations approach to deterministic optimal control, and of Kalman filtering.

1 The issue

The issue is clear in a simple example: minimize t with $|X(t)| = 1$ starting from $X(0) = x \in [-1, 1]$ and dynamics $\left|\dot{X}\right| \leq 1$, or (in notation closer to the general situation), $dX = \alpha(t)dt$, where $|\alpha(t)| \leq 1$ for all t. The answer to this optimization problem is obvious: if $X(0) = x > 0$ then go to $x = 1$ as fast as possible. The time will be $u(x) = 1 - x$. If $x < 0$, go to -1. The minimum time for that will be $u(x) = 1 + x$. The overall value function for this optimization problem is is

$$u(x) = 1 - |x| \ . \tag{1}$$

The Hamilton Jacobi equation is derived as in Section 6 from

$$u(x,t) \ = \ u(x) \ + \ dt \ + \ \min_{|\alpha| \leq 1} \partial_x u \cdot \alpha \ ,$$

or

$$|\partial_x u| \ = \ 1 \ . \tag{2}$$

Our solution (1) satisfies this equation for every $x \in (-1, 1)$ except at $x = 0$, where the derivative $\partial_x u$ does not exist. It also satisfies the obvious boundary conditions $u(-1) = 0$ and $u(1) = 0$. The problem is that the function $v(x) = -1 + |x|$ has all these properties too. It also satisfies the PDE (2) at every x except $x = 0$ and the boundary conditions. The *viscosity solution* theory is a criterion that allows us to distinguish correct solutions such as (1) from incorrect or non-slutions such as v. The criterion is not that the PDE should be satisfied at almost every point. Both our candidates do that.

Moreover, the problem

$$v(x) \;=\; \max\; -t \;\text{ with }\; |X(t)| \leq 1 \;,$$

with $X(0) = x$ and $\left|\dot{X}\right| \leq 1$ has (2) as its Hamilton Jacobi equation, but this time $v(x) = -1 + |x|$ is the correct solution and (1) the incorrect one. Thus, choosing right from wrong among solutions involves the original optimization problem as well as the formal PDE.

I have a pet peeve with Evans' book and Kohn's notes. They both call the Hamilton Jacobi equation $\partial_t u = |\nabla u|$, or $|\nabla u| = 1$, *the* eikonal equation (time dependent or steady state). There are two things wrong with this. First, *eikonal equation* is a term from geometric optics. Many other eikonal equations that arise in that way. It would be better to say *an* eikonal equation. Second, the theory of viscosity solutions is not relevant in geometric optics. The viscosity solution is not relevant in geometric optics. If the solution is not smooth, you are not interested in it.

2 Verification theorem

A *verification theorem* states that if u is a differentiable solution of the Hamilton Jacobi equation and satisfies the proper initial, final, or boundary conditions, then u is actually the value function. It says that our heuristic derivation (with Δt, etc.) gives a correct conclusion. This would be important in any mathematical theory of optimal deterministic control. It also leads to a comparison principle related to the maximum principle from Section 3. This comparison principle is the main ingredient of the theory of viscosity solutions.

There are different verification theorems for different situations. We give (a sketch of) the proof here for a simple time dependent problem

$$u(x,t) \;=\; \max \int_t^T h(X(s),\alpha(s))\,ds \;, \tag{3}$$

with (see Section 6)

$$dX(s) \;=\; a(X(s),\alpha(s))ds \quad \text{for } s \geq t, \text{ and} \quad X(t) = x \;.$$

The Hamilton Jacobi problem is (again, Section 6)

$$0 = \partial_t u \;+\; H(x,\nabla u) \;, \tag{4}$$

$$H(x,p) \;=\; \max_\alpha \left\{\; h(x,\alpha) \;+\; a(x,\alpha)\cdot p \;\right\} \tag{5}$$

We say that u is a *classical solution* of (4) if:

1. It is a continuously differentiable in x and t for all x and $t \in [0,T]$.
2. These derivatives satisfy (4) for every x and $t \in [0,T]$.

3. $u(x,T) = 0$ for all x (the final condition).

The *verification theorem* states that if u is a classical solution of (4), then the minimum in (3) is attained and that minimum is u.

We give a proof of the verification theorem that uses the method of upper and lower bounds. If you want to show that $A = B$ you prove the upper bound, $A \leq B$, and the lower bound, $A \geq B$. Let $v(x,t)$ be a classical solution of the Hamilton Jacobi equation (4) (with final condition $v(x,T) = 0$) and u the value function (3). We start with the upper bound $v(x,t) \leq u(x,t)$ for all x and $t \in [0,T]$. To do this, we show that $v(x,t)$ is the value you get if you use the control $\alpha_*(x,t)$ that solves the maximization problem (5). (This is where we use the hypothesis that the maximum in (5) is achieved. This would not work if we replaced max with sup in (5), and the verification theorem might not be true.) If v corresponds to some control and u corresponds to the optimal control, then $v \leq u$.

More precisely, suppose we start with $X_*(t) = x$ and use the closed loop dynamics $dX_*(s) = a(\, X_*(s),\, \alpha_*(X_*(s), s)\,)\, ds$. The $v(x,t)$ is the value realized with this strategy:

$$v(x,t) \;=\; \int_t^T h(X_*(s), \alpha_*(s))\, ds\ . \tag{6}$$

This we do by differentiation in time over the whole trajectory. If $X_*(t)$ is a closed loop trajectory, we show

$$\frac{d}{dt}\, v(X(t),t) \;=\; \frac{d}{dt} \int_t^T h(X_*(s), \alpha_*(s))\, ds\ . \tag{7}$$

Clearly (7) implies (6) if we know the final condition $v(x,T) = 0$, because the right side of (6) is also equal to zero when $t = T$.

With all that, the actual differentiation is simple. On the left we have

$$\begin{aligned}
\frac{d}{dt}\, v(X_*(t),t) &= \nabla v \cdot \dot{X}_*(t) \;+\; \partial_t v \\
&= a(X_*(t), \alpha_*(X_*(t),t)) \cdot \nabla v(X_*(t),t) \;+\; \partial_t v(X_*(t),t) \\
&= H(X_*(t), \nabla v(X_*(t),t)) \;-\; h(X_*(t), \alpha_*(X(t),t) \;+\; \partial_t v(X_*(t),t) \\
&= -h((X_*(t), \alpha_*(X(t),t))\ .
\end{aligned}$$

In these steps we used all the hypotheses. The chain rule calculation is correct if v is continuously differentiable, and $a \cdot \nabla v = H - h$ if α_* solves the maximization problem (5), and (for the last line) v satisfies (4). On the right of (7) we have

$$\frac{d}{dt} \int_t^T h(X_*(s), \alpha_*(s)\, ds \;=\; -h(X_*(t))\ .$$

This shows that the right and left sides of (7) are equal, which completes the proof of the upper bound.

A subtlety of this proof is that there may be more than one α_* for a given x and p that optimizes (5). Therefore, "the" trajectory $X_*(s)$ might actually be a family of trajectories depending on which choice of α_* we made. The calculation above holds for any choice of α_* (if it optimizes (5)) and X_*.

The lower bound is a longer version of this argument. We know from (6) that using the control α_* gives value $v(x,t)$. The only way to produce a better value is to use a different control. Therefore, suppose $\alpha(s)$, defined for $s \in [t,T]$, is a control that may be different from α_*. Let $X(s)$ be the corresponding open loop controlled trajectory starting from $X(t) = x$ using the control $\alpha(s)$:

$$d\widetilde{X}(s) \;=\; a(\widetilde{X}(s), \widetilde{\alpha}(s))\, ds \;.$$

A calculation like the one above will show that

$$v(x,t) \;\geq\; \int_t^T h(X(s), \alpha(s))\, ds \;. \tag{8}$$

But $u(x,t)$ is defined as the supremum of the right side of (8) for all choices of α. Therefore, (6) implies that $u(x,t) \leq v(x,t)$ and completes the proof of the verification theorem.

There is another mathematicians' subtlety in this wording. We did not assume that there is a control that achieves the value $u(x,t)$. If u is the supremum but not the maximum, then for any $\epsilon > 0$ there is a control that achieves value $u(x,t) - \epsilon$. This control may depend on ϵ in a way that the limit as $\epsilon \to 0$ does not exist. That is what happened in problem 2 of Assignment 6. Part of the statement of the verification theorem is that the maximum is achieved whenever the solution to the Hamilton Jacobi equation is continuously differentiable. This implies that the value function for problem 2 of Assignment 6 is not continuously differentiable.

The possible inequality in (8), as opposed to the strict equality in (6), comes from the possibility that $\alpha(s)$ is not an optimizer in (5) corresponding to the state $X(s)$. For any y, s, α, define in terms of v as

$$R(y,s,\alpha) \;=\; H(x, \nabla v(y,s)) \;-\; \Big\{\; h(y,\alpha) \;+\; a(y,\alpha)\nabla v(y,s) \;\Big\} \;\geq\; 0 \;.$$

Clearly, $R = 0$ only if $\alpha = \alpha_*$. Now to repeat the calculation (7) for (8), (using $v(x,T) = 0$ for all x)

$$
\begin{aligned}
v(x,t) &= -\int_t^T \frac{d}{ds} v(X(s),s)\, ds \\
&= -\int_t^T \nabla v(X(s),s) \frac{dX}{ds} \;+\; \partial_s v(X(s),s)\, ds \\
&= -\int_t^T \Big(\nabla v(X(s),s) \cdot a(X(s),\alpha(s)) \;+\; h(X(s),\alpha_*(s)) \;+\; \partial_s v(X(s),s) \\
&\qquad - h(X(s),\alpha_*(s)) \;+\; h(X(s),\alpha(X(s),s)) \\
&\qquad - h(X(s),\alpha(X(s),s)) \Big)\, ds \\
&= -\int_t^T \Big(H(X(s),\nabla v(X(s),s)) \;+\; \partial_s v(X(s),s) \Big)\, ds \\
&\quad + \int_t^T R(X(s),s,\alpha(s))\, ds \\
&\quad + \int_t^T h(X(s),s,\alpha(s))\, ds \; .
\end{aligned}
$$

The top line on the right of the last equality is zero because v satisfies the Hamilton Jacobi equation. The middle line is negative (more precisely, not positive) by the definition of R. The last line is the right side of (8), which completes the proof of (8) and, with that, the verification theorem.

3 Comparison principle

Local comparison principles underlie the theory of viscosity solutions. As in Section 3, we start with the case of a time independent value function. The extension to time dependent value functions is more or less as we did there. We will talk about a function $u(x)$ that maximizes (3) with $T = \infty$. We call a function *smooth* if it is continuously differentiable (Warning: Context determines how many derivatives it takes to be "smooth". It could be two or more for second order PDEs.) We do not assume u is smooth.

Suppose D is a small region of space, such as a ball $|x - x_0| \leq \epsilon$. Let Γ be the boundary of D. A smooth function $w(x)$ is a *supersolution* if it satisfies the Hamilton Jacobi inequality

$$H(x, \nabla w) \leq 0 \; , \tag{9}$$

and a *subsolution* of

$$H(x, \nabla w) \geq 0 \; , \tag{10}$$

The comparison principle is that if $u(x)$ is the value function for the optimal control problem and w is a smooth supersolution defined for $x \in D \cup \Gamma$, and $u(x) \leq w(x)$ for $x \in \Gamma$, then $u(x) \leq w(x)$ for $x \in D$. This statement is similar to the comparison principle of Section 3. The proof is different here because we do not assume that u is smooth.

There is a mathematical subtlety in the definitions (9) and (10). For smooth solutions, the Hamilton Jacobi equation $H(x, \nabla u) = 0$ is equivalent to the

equation $-H(x, \nabla u) = 0$. For example, the equation $|\nabla u| - 1 = 0$ would seem to be equivalent to $1 - |\nabla u| = 0$. But a *super*solution of $|\nabla u| - 1 = 0$ is a *sub*solution of $1 - |\nabla u| = 0$. The viscosity solutions also are different. It is important to take H with the sign it is given in (5).

We can prove this by using the verification theorem (that's why we proved the verification theorem) to show that w is the value function for an optimization related to (3). The inequality $u \leq w$ on D is obvious from the optimization problems. Indeed, this is almost easier done than said. Define

$$R(x,t) \;=\; \partial_t w + H(x, \nabla w) \; .$$

Then w is a supersolution if and only if $R \leq 0$. For all $x \in \Gamma$, define $g(x,t) = w(x,t)$. For all $x \in D \cup \Gamma$, define $f(x) = w(x,T)$. For a trajectory $X(t)$, the *exit time* is $\tau = \min \{t \text{ with } X(t) \in \Gamma \text{ or } t \geq T\}$. In other words, if $X(t)$ touches the boundary before time T, τ is the first such time. Otherwise, $\tau = T$. The optimization problem is

$$\begin{aligned} v(x,t) \;\; &= \;\; \max \int_t^\tau \big(h(X(s), \alpha(s)) \;-\; R(X(s), s) \big) \, ds \\ &\quad + \begin{cases} g(x,\tau) & \text{if } \tau < T \\ f(x) & \text{if } \tau = T. \end{cases} \end{aligned}$$

The Hamilton Jacobi equation for this is

$$0 \;=\; \partial_t v \;+\; H(x, \nabla v) - R(x,t) \; .$$

The verification theorem tells us that if w is a smooth solution that is equal to its boundary values everywhere, then w is the corresponding value function. On the other hand, the principle of dynamic programming tells us that $u(x,t)$ is the optimal value function for the same local problem but with $R = 0$ and boundary values $u \leq w$ on $x \in \Gamma$ or $t = T$. (This is "clear" but might take some thought.) Lowering the boundary payouts and the running payout (by setting $R = 0$) can only lower the value functions, which implies that $u \leq w$ inside D as well.

4 Viscosity solution, definition

We saw in the example of Section 1 above that the value function of a dynamic optimal control problem may not be differentiable. Just using the Hamilton Jacobi equation at places where the value function is differentiable, we are unable to distinguish the correct value function from something that is not the value function.

The theory of viscosity solutions addresses this problem. The theory starts with a definition of viscosity solution that applies to any continuous function. This definition is given below. The theory then shows that the Hamilton Jacobi equation has a viscosity solution, that the value function is a viscosity solution, and that the viscosity solution is unique. (Uniqueness is the hard part. The proof is very clever.) This shows that the value function is the unique viscosity solution. Any approximation scheme for Hamilton Jacobi equations (such as finite difference approximations) that converges to the viscosity solution converges to the value function. This makes the Hamilton Jacobi equation a good way to calculate the value function.

We have a function, u, a hamiltonian $H(x, u, p)$, and a point, x_0. We want to know whether u satisfies the Hamilton Jacobi equation $H(x, u(x), \nabla u(x)) = 0$ at the specific point x_0. If ∇u exists at x_0, this is a simple question of calculus. The special definition allows us to answer this question for any x_0 whether or not u is differentiable there. This viscosity solution definition agrees with the usual one at points where ∇u is defined.

The viscosity solution definition calls for finding comparison functions $\phi(x)$ that are differentiable at x_0 and checking the sign of $H(x_0, u(X_0), \nabla\phi(x_0))$. One set of comparison functions is those that *touch* u at x_0 *from above*. We say ϕ touches u from above at x_0 if

1. $\phi(x)$ is defined for all x close enough to x_0 (there is an $\epsilon > 0$ so that $\phi(x)$ is defined for all x with $|x - x_0| < \epsilon$).

2. $\phi(x_0) = u(x_0)$. This is the "touching" part.

3. $\phi(x) \geq u(x)$ for all x for which ϕ is defined (or for all x with $|x - x_0| < \epsilon$). This is the "from above" part.

What matters for the viscosity solution definition is $p = \nabla\phi(x_0)$. The order one *superjet* of u at x_0 is written $\mathcal{J}_1^+(x_0)$ ($\mathcal{J}$ is how my LaTeX makes a fancy J.). It is the set of possible gradients of comparison functions. The definition is that $p \in \mathcal{J}_1^+(x_0)$ if there is a continuously differentiable function, ϕ that touches u from above at x_0 with $p = \nabla\phi(x_0)$. This is similar to the definition of *subdifferential* in the theory of convex functions (for those who know this theory).

Similarly, we can say what it means for ϕ to touch u from below at x_0. Then $p \in \mathcal{J}_1^-(x_0)$ if there is a continuously differentiable ϕ that touches u from below at x_0 so that $p = \nabla\phi(x_0)$.

It is also possible to define higher order subjets and superjets. The order k superjet depends derivatives up to order k. For example, let p be an n component vector and A an $n \times n$ symmetric matrix. We say $(p, A) \in \mathcal{J}_2^+(x_0)$ if there is a ϕ that is twice continuously differentiable and that touches u from above at x_0 so that $p = \nabla\phi(x_0)$ and $Q = D^2u(x_0)$ (the Hessian matrix of second partial derivatives). Order 2 subjets and superjets enter into the definition of viscosity solution of the Hamilton Jacobi equations that describe value functions for stochastic optimal control problems.

Back to first order superjets and subjets. If $p = \nabla u(x_0)$ is well defined, then the superjet and subjet are nothing but p. More precisely, $\mathcal{J}_1^+(x_0) = \mathcal{J}_1^-(x_0) = \{p\}$. Here, $\{p\}$ is the set consisting of the single element, p. For example, consider the one dimensional case and suppose $p = \partial_x u(x_0)$ and that $q = \partial_x\phi(x_0)$ has $q > p$. Then if $u(x_0) = \phi(x_0)$, it is clear that $u(x) > \phi(x)$ for x slightly larger than x_0 and $u(x) < \phi(x)$ for x slightly smaller than x_0. This implies that ϕ does not touch u from above or below at x_0 and that $q \notin \mathcal{J}_1^+(x_0)$ and $q \notin \mathcal{J}_1^-(x_0)$.

The situation is different if $\nabla\phi$ is not defined at x_0. Consider the one dimensional example $u(x) = 1 - |x|$. This is differentiable except at $x = 0$. Consider

lines that touch the graph of u from above at $x_0 = 0$. These must have slope in the set $[-1, 1]$. The lines with slope strictly between -1 and 1 touch the graph only at $x = 0$ and otherwise are strictly above u. The lines with slope $+1$ or -1 touch from above and remain above u, but not strictly so. This is allowed by the definition of touching from above. Therefore, every number $p \in [-1, 1]$ is in $\mathcal{J}_1^+(0)$. If $p > 1$, $\phi(0) = u(0) = 1$, and $\partial_x \phi(0) = p$, then $\phi(x) < u(x)$ for x slightly less than zero. This implies that $p \notin \mathcal{J}_1^+(0)$. Altogether, we see that $\mathcal{J}_1^+(0) = [-1, 1]$. To summarize, u is not differentiable at $x_0 = 0$, the superjet of u at x_0 is not a single number (as it would be if u were differentiable at 0), but a whole interval.

The subjet of u at $x_0 = 0$ is the empty set, which we write $\{\}$. This is because no differentiable function can touch u from below at x_0. If $\phi(0) = u(0) = 1$, and $p = \partial_x \phi(0)$ has $|p| < 1$, then $\phi(x) > u(x)$ for sufficiently small but non-zero x. If $p = \partial_x \phi(0) > 0$, and $\phi(0) = u(0)$, then $\phi(x) > u(x)$ for $x > 0$ but sufficiently small. You get the picture.

The general definition of viscosity solutions is that $H(x_0, u(x_0), p)$ should have a definite sign whenever $p \in \mathcal{J}_1^+(x_0)$ or $p \in \mathcal{J}_1^-(x_0)$. Which sign depends on whether we are solving a maximization or a minimization problem. I suggest figuring out the sign in whatever situation you're in by checking an example like u. For the minimization problem that defined u, the definition is that $H(x_0, p) \geq 0$ for all $p \in \mathcal{J}_1^+(x_0)$, and $H(x_0, p) \leq 0$ for all $p \in \mathcal{J}_1^-(x_0)$. Let us check these for our specific u, starting with the superjet. The Hamiltonian was $H(p) = 1 - |p|$. We saw that $p \in \mathcal{J}_1^+(0)$ if and only if $|p| \leq 1$. For these p, we have $H \geq 0$, as needed. The subjet at $x_0 = 0$ is empty. Therefore it is true in the mathematical sense that every $p \in \mathcal{J}_1^-(0)$ has $H(p) \leq 0$. There can be no counter-examples because there are no examples, period.

On the other hand, our false solution $v(x) = -1 + |x|$ has $\mathcal{J}_1^-(0) = [-1, 1]$ and $\mathcal{J}_1^+(0) = \{\}$. The superjet part of the viscosity solution definition is OK because this time the superjet is empty. The subjet part of the definition is the requirement that $H(p) = 1 - |p| \leq 0$ for all $p \in \mathcal{J}_1^-(0) = [-1, 1]$. But we just saw that $|p| < 1$ implies that $H(p) > 0$, which violates the requirement that $H(p) \leq 0$ for $p \in \mathcal{J}_1^-(0)$. Thus, v is not a viscosity solution.

You can find some discussion of the proofs of the existence and uniqueness theorems for viscosity solutions in the PDE book of L. C. Evans. Here is a little motivation. Suppose a smooth function ϕ satisfies $H(\nabla\phi(x)) = R(x) \geq 0$. Then ϕ is the value function for the optimization problem involving $\widetilde{H} = H - R(x)$. The comparison theorem above then makes it impossible for ϕ to touch u from below at a point. On the contrary, if $\phi(x) < u(x)$ when $|x - x_0| = \epsilon$ then $\phi(x) < u(x)$ when $|x - x_0| \leq \epsilon$. (As above, the signs and senses of inequalities are uncertain and depend on whether we are minimizing or maximizing.) This shows that the value function is a viscosity solution. The uniqueness of viscosity solutions is more technical.

5 Converging to a viscosity solution

The term *viscosity solution* comes from an analogy with the theory of shock waves in a compressible gas. The dynamics of a compressible gas may be described by a system of first order nonlinear partial differential equations for the density, pressure, and velocity as a function of x and t. Solutions to these nonlinear equations develop discontinuities that are characterized by a discontinuous jump in pressure – a shock wave. The physical mechanism for this is that sound travels faster in higher pressure air, so if there is a sound wave, the high pressure parts catch up with the lower pressure parts, forming a jump in pressure and density.

A more detailed description of the gas includes small second derivative terms that model friction forces in the gas, *viscosity*. These terms have the effect of spreading out the sharp discontinuity to a thin layer where the density changes rapidly from one value to the other. The width of this *viscous layer* is proportional to this viscosity so that in the limit as the viscosity goes to zero, the solution converges to the perfect discontinuity.

There is a strong mathematical analogy (if not physical analogy) between the pressure as a function of x and the slope, $\partial_x u(x)$, of the value function. We have seen that the slope of the value function can be discontinuous at a point, just as the pressure can be discontinuous. Furthermore, adding a small viscosity term $\epsilon \partial_x^2 u$ to the Hamilton Jacobi equation will change the discontinuous change in slope to a rapid but smooth transition. Technically, suppose $u^\epsilon(x)$ satisfies

$$H(x, u^\epsilon, \nabla u^\epsilon) \; + \; \epsilon \triangle u^\epsilon \; = \; 0 \; . \tag{11}$$

Part of the viscosity solution theory is that as $\epsilon \to 0$, we have $u^\epsilon(x) \to u(x)$, where u is the viscosity solution. We have seen in Section 3 that solutions of second order PDEs like this satisfy comparison principles. These comparison principles are used to show that the limit, u, is a viscosity solution by the subjet/superjet definition.

The equation (11) has the interpretation as the second order Hamilton Jacobi equation for a stochastic optimal control problem where the level of random forcing is small. This is covered in the next class. It makes sense that the value function for the noisy problem should converge to the value function of the noise free problem as the level of noise converges to zero.

The theory of viscosity solutions also leads to finite difference methods that produce approximations to even nondifferentiable viscosity solutions.