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 28, 2008.

Section 9: Approximations and perturbations.

All models are wrong. Some models are useful.
George Box, statistician.

Some of the most powerful techniques in PDE are not exact solutions methods but methods of approximate solution. Analytic approximation is as useful as numerical solution for getting numbers form PDE models. We will see that approximation often work together, the approximation being far less challenging for numerical solution than the original.

Here is a piece of philosophy related to the famous and wise words quoted above. Suppose we have a model with a parameter,[1] ϵ, that we expect to be small. Setting ϵ to zero may make the model so much easier to solve that we consider doing it, even though $\epsilon = 0$ is a less faithful model than $\epsilon \neq 0$. An intermediate possibility is to use perturbation theory to estimate the difference between the desired small ϵ model and the simpler $\epsilon = 0$ one. This *correction*is not exact. But it may be preferable to the alternative of leaving ϵ out altogether. If ϵ is small enough that we contemplate setting it to zero, it may well be small enough that perturbation theory estimates the correction accurately. The perturbation correction is not exact, but it is more useful than no correction at all.

As an example, consider the effect of fluctuating interest rates on the price of equity options. Typically this effect is small. Short dated options prices do not depend much on interest rates anyway, and interest rates may not fluctuate much during the lifetime of the option. This is why is makes sense to use the Black Scholes model with fixed risk free rate, r. On the other hand, options do carry some interest rate risk. We could model this using a short rate model for interest rates driven by a Brownian motion $W_2(t)$. Together with the standard equity price model, this gives a system of two SDEs such as

$$\begin{aligned} dS(t) &= r(t)S(t)dt \ + \ \sigma S(t)dW_1(t) \ , & (1) \\ dr(t) &= a\,(\,\overline{r} - r(t)\,)\,dt \ + \ \gamma dW_2(t) \ , & (2) \\ \rho &= \mathrm{corr}(\,W_1(t),\,W_2(t)\,) \ . \end{aligned}$$

Replacing $r(t)$ by $\overline{r}$ in (1) is the usual geometric Brownian motion Black Scholes model. This may be a good approximation if $r(0) - \overline{r}$ is small, if γ is small (the volatility of interest rates), and if the final time, T is not too large. In these cases, we can estimate the effect of stochastic non-constant interest rates from

[1] Calling a parameter ϵ, rather than, say, M, indicates that it will soon be small. This is like putting a movie cowboy on a white horse to indicate that he's the good guy.

Black Scholes and perturbation corrections to it. We do not need to solve the full two dimensional PDE.

When you're done with Computational Methods in Finance, you will not be afraid of two dimensional PDEs. But suppose we use a multi-factor interest rate model in place of (2) and throw in stochastic volatility on top. Then we would be relieved to be able to reduce the dimensionality by one.

You might have noticed that none of the smallness assumptions above was described by an explicit parameter called ϵ. Part of the formal perturbation theory process is identifying the effects that are small and finding ways to express their smallness as the smallness of a parameter.

Notice also the disclaimer that T may not be too large. The effect of stochastic interest rates is larger for longer dated options. More abstractly, saying that the effect of an ϵ perturbation is $O(\epsilon)$ means that there is a C so that the perturbation is not more than $C|\epsilon|$. It may happen that C is finite for any particular T, but $C \to \infty$ as $T \to \infty$. This indicates that the perturbation correction may lose accuracy as T grows. In this situation we say the perturbation accuracy is not *uniform* as $T \to \infty$. Non-uniform approximations are the nightmares of every asymptotician.

1 Kinds of approximations

Let $A(\epsilon)$ be the desired quantity, the *answer*. The simplest way A can depend on ϵ is as a normal power series

$$A(\epsilon) \;=\; A_0 \;+\; \epsilon A_1 \;+\; \epsilon^2 A_2 + \cdots \;. \tag{3}$$

The *unperturbed* answer is A_0. The *first correction*, or *leading order* correction is ϵA_1. The *next* correction is $\epsilon^2 A_2$. For small epsilon, the leading order correction normally is much larger than the next correction. It is not that common that we are able to calculate beyond the leading order correction[2].

Sometimes, we can calculate the coefficients A_1, A_2, etc. using Taylor series (see Section 3 below). Other times (more commonly) we calculate the coefficients by assuming an expansion of the form (3) – a *scaling ansatz* (see Section 2) – and computing the terms by substitution into the equation defining $A(\epsilon)$.

A rule of thumb is that the error is roughly the size of the first neglected term For example the error from using no correction is $A(\epsilon) - A_0$. This is roughly equal to ϵA_1. Further, the error after the leading correction is $A(\epsilon) - (A_0 + \epsilon A_1)$. This is roughly equal to $\epsilon^2 A_2$. The rule of thumb is not a theorem both because it is not precisely stated and because there are exceptions to it. It remains a useful if somewhat suspect heuristic.

The perturbation expansion may take another form. Maybe the correction is on the order of $\sqrt{\epsilon}$. Maybe $A(\epsilon) \to \infty$ as $\epsilon \to 0$.

[2]See, however, the PhD Thesis of Edward Hayes.

2 Scalings

The creative problem solving side of perturbation theory comes out when we are uncertain what order of magnitude a small ϵ addition to the equation has on the solution. A *scaling* is a partial ansatz that puts in powers of ϵ in various places in a proposed solution. The trick is to do this in a way that the resulting perturbation calculations lead to an actual *asymptotic approximation* (defined shortly). There may be more than one way to do this.

A famous example that illustrates some possibilities is finding the roots of the quadratic equation

$$\epsilon x^2 \ - \ x \ + \ 1 \ = \ 0 \, . \tag{4}$$

If we simply set $\epsilon = 0$, there is a solution $x = 1$. We can find corrections to this solution by assuming a power series ansatz, $x(\epsilon) = 1 + \epsilon x_1 + \epsilon^2 x_2 + \cdots$ and substituting this into (4). We get

$$\epsilon \left(1 + \epsilon 2x_1 + \epsilon^2 (2x_2 + x_1^2) + \cdots\right) \ - \ \left(1 + \epsilon x_1 + \epsilon^2 x_2 + \cdots\right) \ + \ 1 \ = \ 0 \, .$$

If this is to hold for every ϵ, the terms of various powers of ϵ must balance separately. This can be expressed by collecting the coefficients of ϵ to various powers:

$$\begin{aligned} O\left(\epsilon^0\right): &\qquad -1 \ + \ 1 \ = \ 0 \\ O\left(\epsilon^1\right): &\qquad 1 \ - \ x_1 \ = \ 0 \\ O\left(\epsilon^2\right): &\qquad 2x_1 \ - \ x_2 \ = \ 0 \end{aligned}$$

The first equation works out because we chose $x_0 = 1$ (the coefficient of ϵ^0). The second equation determines $x_1 = 1$. The third, together with the value of x_1, determines that $x_2 = 2$. This gives

$$x(t) \ \approx \ 1 \ + \ \epsilon \ + \ 2\epsilon^2 \, . \tag{5}$$

There are several different ways to find this result that give us some confidence. One is to suppose there is a solution $x(\epsilon)$ that is a smooth function of ϵ near $\epsilon = 0$. We then compute $\dot{x} = \partial_\epsilon x$ and $\ddot{x} = \partial_\epsilon^2 x$, both evaluated at $\epsilon = 0$. Differentiating (4) with respect to ϵ gives

$$x^2 + 2\epsilon x\dot{x} \ + \ \dot{x} \ = \ 0 \, .$$

Setting $\epsilon = 0$ then gives $\dot{x} = 1$. Differentiating again gives

$$2x\dot{x} \ + \ 2x\dot{x} \ + \ 2\epsilon x\ddot{x} \ + \ \dot{x}^2 \ - \ \ddot{x} \ = \ 0 \, .$$

We set $\epsilon = 0$ and use the known value of $\dot{x}$ to get $\ddot{x} = 4$. The Taylor series with this information is

$$x(\epsilon) \ \approx \ x(0) \ + \epsilon\dot{x} \ + \ \frac{1}{2}\epsilon^2\ddot{x} \ = \ 1 \ + \ \epsilon \ + \ 2\epsilon^2 \, ,$$

which is the same as (5). A final check would be to use the quadratic formula for $x(\epsilon)$ and expand the square root. This also gives (5).

We're not quite done. A quadratic equation has two solutions, and we only found one. If you think about it, you will see that if we assume the solution has the form $x(\epsilon)$ that is a continuous function of ϵ for $\epsilon \geq 0$ so that $\lim_{\epsilon \to 0} X(\epsilon)$ exists, then it must be the solution we just found. The other solution therefore is not a continuous function of ϵ at $\epsilon = 0$. It must blow up. To find it, let's try the scaling ansatz

$$x(\epsilon) \;=\; \epsilon^{-1} x_{-1} \;+\; x_0 \;+\; \epsilon x_1 \;+ \cdots \,. \tag{6}$$

Substitution into (4) gives

$$\epsilon \left(\epsilon^{-2} x_{-1}^2 \;+\; 2\epsilon^{-1} x_{-1} x_0 \;+ 2 x_{-1} x_1 + x_0^2 + \cdots \right) - \epsilon^{-1} x_{-1} - x_0 - \epsilon x_1 + \cdots + 1 \;=\; 0 \,.$$

Equating powers of ϵ as above gives:

$$\begin{aligned} O\left(\epsilon^{-1}\right): &\qquad x_{-1}^2 \;-\; x_{-1} \;=\; 0 \\ O\left(\epsilon^{0}\right): &\qquad 2 x_{-1} x_0 - x_0 \;+\; 1 \;=\; 0 \\ O\left(\epsilon^{1}\right): &\qquad 2 x_{-1} x_1 + x_0^2 - x_1 \;=\; 0 \end{aligned}$$

The first equation has solutions $x_{-1} = 0$ and $x_{-1} = 1$. We choose the second because the first gives a trivial solution. The next equation gives $x_0 = -1$. The last gives $x_1 = -1$. Altogether, this solution takes the form

$$x(\epsilon) \;\approx\; \epsilon^{-1} \;-\; 1 \;-\; \epsilon \;\pm\; \cdots \,. \tag{7}$$

There are several ways to check this answer. One is to use the quadratic formula. One of the roots looks like (7). Another way is to write $x(\epsilon) = \epsilon^{-1} y(\epsilon)$ to get (after multiplying by ϵ)

$$y^2 \;-\; y \;+\; \epsilon \;=\; 0 \,. \tag{8}$$

The Taylor series solution for y with $y(0) = 1$ is equivalent to the solution (7).

There are several takeaways from this example. One is that scaling can change which terms look big and which look small. In the scaling (4) the quadratic term looks small. In the scaling (8) it's the constant term that looks small. Another is that scaling is an art. The scaling ansatz (6) is arrived at by trial and error.

3 Short time

Suppose $f(x,t) = E_{x,t}\left[V(X(T))\right]$, and $X(t)$ is a multivariate diffusion,

$$dX \;=\; a(X)dt \;+\; b(X)dW \,. \tag{9}$$

The value function satisfies the backward equation (with the summation convention)

$$0 \;=\; \partial_t f \;+\; \mathcal{L} f \;=\; \partial_t f \;+\; \frac{1}{2} b_{jl}(x) b_{kl}(x) \partial_{x_j} \partial_{x_k} f \;+\; a_j(x) \partial_{x_j} f \,. \tag{10}$$

There are useful approximations that apply when $T - t$ is small.

The simplest such case is when the payout, $V(x)$, is a smooth function of x. This is not so common. But one case is the log contract, $V(x) = \log(x)$, that is useful in hedging options on realized variance or realized volatility[3]. Anyway, if $V(x)$ is a smooth function of x, then $f(x,t)$ is a smooth function of x and T all the way to the expiration time, T. This implies that one can approximate $f(x,t)$ for t near T using ordinary Taylor series in time. The first such approximation is

$$\begin{aligned} f(x,t) &\approx V(x) - (T-t)\,\mathcal{L}V \\ &= V(x) - (T-t)\left\{ \frac{1}{2} b_{jl}(x) b_{kl}(x) \partial_{x_j} \partial_{x_k} V - a_j(x) \partial_{x_j} V \right\} . \end{aligned}$$

One could compute the second term

$$f(x,t) \approx V(x) - (T-t)\,\mathcal{L}V - \frac{1}{2}(T-t)^2\,\mathcal{L}^2 V .$$

This is complicated, but it might be worth it if the alternative is numerical solution. Moreover, even if we don't use the second order term, it gives an idea how accurate the first term is likely to be. If the fourth derivatives are so large that the second term is larger than the first, then the first term does not give an accurate approximation.

None of these approximations makes sense for a vanilla put or call because even the second derivative is infinite (here $H(x)$ is the Heaviside function $H(x) = 0$ if $x \leq 0$ and $H(x) = 1$ for $x > 0$):

$$\partial_x^2 (x-K)_+ = \partial_x H(x-K) = \delta(x-K) .$$

It makes no sense (and certainly is false) to say $f(x,t) \approx (x-K)_+ + C(T-t)\delta(x-K) + (\text{other stuff})$. The solution is not infinite at the strike price.

As an example, consider the log contract in the Black Scholes model. There we have $\mathcal{L}V = \frac{\sigma^2 s^2}{2}\partial_s^2 V + rs\partial_s V - rV$. With $V = \log(s)$, this becomes $\mathcal{L}V = \frac{\sigma^2}{2} - r - r\log(s)$, so the approximation is

$$f(s,t) \approx \Big(1 - r(T-t)\Big)\log(s) - (T-t)\left(\frac{\sigma^2}{2} - r\right) .$$

This is not an exact solution, but it is reasonably accurate close to expiration. If you examine this approximate solution, particularly after calculating the second order correction, you may discover the simple exact closed form solution for this problem.

More commonly, these simple expansions do not work. One looks for the

[3] Ask Peter Carr for details.

4 Approximate equilibrium

Here we present a simplified version of work of Fouque, Papanicolaou, and Sircar. The method is much older but they seem to be the first to apply it to financial models. Consider the stochastic volatility model

$$dS(t) \;=\; rS(t)dt \;+\; \sigma(t)S(t)dW_1(t)\;, \tag{11}$$

and

$$d\sigma(t) \;=\; a\,(\overline{\sigma} - \sigma(t))\,dt \;+\; \gamma dW_2(t)\;. \tag{12}$$

We will assume that the two driving processes are independent, though it is not hard to redo these calculations without that assumption. The hypothesis will be that volatility has a large uncertainty, but that this has a rapidly decaying memory. The authors mentioned present statistical evidence of rapid memory loss for the stochastic volatility, but I am skeptical of this. Still, their model provides useful approximations and is a great example of the general method.

That means that the constants γ and a must be large. The parameter a determines the rate of memory loss, while γ determines the level of uncertainty in the statistical steady state. The proper scaling to make this happen is to multiply a by ϵ^{-1} and γ by $\epsilon^{-1/2}$. It will become clearer where these scalings come from. In (12) we replace a with $\epsilon^{-1}a$ and γ with $\epsilon^{-1/2}\gamma$, which gives

$$d\sigma(t) \;=\; \epsilon^{-1}a\,(\overline{\sigma} - \sigma(t))\,dt \;+\; \epsilon^{-1/2}\gamma dW_2(t)\;. \tag{13}$$