Lecture 23 – 2020.07.09 – 12:15 via Zoom

Backward SDEs and non-linear PDEs

Based on the papers:

- N. Perkowski "Backward Stochastic Differential Equations: an Introduction" (lecture notes) (URL)
- N. El Karoui, S. Hamadène, and A. Matoussi. Backward stochastic differential equations and applications, volume 27, pages 267–320. Springer, 2008. (URL)
- N. El Karoui, S. Peng, and M. C. Quenez. "Backward Stochastic Differential Equations in Finance." *Mathematical Finance* 7, no. 1 (January 1997): 1–71. https://doi.org/10.1111/1467-9965.00022.

A new kind of SDEs which have numerous applications:

- *Feynman–Kac like representation formulas for non-linear PDEs*
- Stochastic optimal control (BSDEs give representation formula for the optimal control)
- Pricing of a large class of options in mathematical finance

Let's remind the classical Feynman–Kac formula. Consider the first order differential operator

$$\mathscr{L}f(t,x) = \sum_{i=1}^{d} b^i(t,x)\nabla^i f(t,x) + \sum_{i,j=1}^{d} a^{i,j}(t,x)\cdot\nabla^i\nabla^j f(t,x)$$

where $f\in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^d;\mathbb{R})$ and $b\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^d$, $a\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^{d\times d}$ and b,a are sufficiently regular and $a=\frac{1}{2}\sigma\sigma^T$ for some $\sigma\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^{d\times d}$. We know that the solution of the linear initial value PDE problem

$$\begin{array}{ll}\partial_t u(t,x) = \mathscr{L}u(t,x) + f(x)u(t,x) & \\ u(0,x)=\varphi(x) & \end{array}\qquad x\in\mathbb{R}^d, t\geqslant 0$$

is given (under appropriate condition) by the Feynman–Kac representation formula (we give the formula for b,σ not depending on time)

$$u(t,x) = \mathbb{E}\left[\varphi(X_t^x)\exp\left(\int_0^t f(X_s^x)\mathrm{d}s\right)\right],\qquad t\geqslant 0, x\in\mathbb{R}^d,$$

where $(X_t^x)_{t\geqslant 0}$ is the solution of the SDE

$$\mathrm{d}X_t^x = b(X_t^x)\mathrm{d}t + \sigma(X_t^x)\mathrm{d}W_t$$

with initial condition $X_0^x = x\in\mathbb{R}^d$ and W is a d-dimensional Brownian motion. For this is enough that $u\in C^{1,2}$.

What about non-linear PDEs. There are various ways to represent them using stochastic processes. Mainly it depends on the kind of PDE we are dealing with, in particular on the form of the non-linearity. We consider here a special kind, of the form

$$\partial_t u(t,x) + \mathscr{L}u(t,x) + f(t,x,u(t,x),\nabla u(t,x)) = 0 \tag{1}$$

where $\nabla = \mathrm{D}_x$ is the derivative with respect to the space variable (i.e. the gradient). We would like to have a representation formula like the one above. Assume we write $Y_s = u(s,X_s^{t,x})$ for $s\geqslant t$ where u is a solution of the equation and $X^{t,x}$ is the diffusion process associated to $\mathscr{L}$ which is at $x\in\mathbb{R}^d$ at time t. What is the dynamics of Y? By Ito formula we have (assume again that b,σ do not depends on time)

$$\mathrm{d}Y_s = (\partial_s + \mathscr{L})u(s,X_s^{t,x})\mathrm{d}s + \sigma(X^{t,x})\nabla u(s,X_s^{t,x})\mathrm{d}W_s$$

by using the PDE (1) we have

$$\mathrm{d}Y_s = -f(t, X_s^{t,x}, u(t, X_s^{t,x}), \nabla u(t, X_s^{t,x}))\mathrm{d}s + \sigma(X^{t,x})\nabla u(s, X_s^{t,x})\mathrm{d}W_s$$

Therefore if we consider a slightly less general PDE of the form

$$\partial_t u(t,x) + \mathscr{L}u(t,x) + f(t,x,u(t,x), \sigma(x)\nabla u(t,x)) = 0 \tag{2}$$

It is clear that if σ is invertible then this PDE if equivalent to a PDE of the form (1), indeed we have

$$f(t,x,u(t,x), \nabla u(t,x)) = \tilde{f}(t,x,u(t,x), \sigma(x)\nabla u(t,x))$$

with $\tilde{f}(t,x,y,z) = f(t,x,y,\sigma(x)^{-1}z)$. But in this case we have a nicer dynamics for Y:

$$\mathrm{d}Y_s = -f(t, X_s^{t,x}, Y_s, Z_s)\mathrm{d}s + Z_s\mathrm{d}W_s, \tag{3}$$

with $Z_s = \sigma(X^{t,x})\nabla u(t, X_s^{t,x})$. We are actually going to consider the pair of adapted processes Y, Z as a pair of unknown in this equation. This is the first novelty (not so much, because we arleady seen something similar for reflected equations). The interest of this formulation of the dynamics of (Y,Z) is that it does not depends anymore on the knowledge of u but recall that $Y_s = u(s, X_s^{t,x})$.

Exercise 1. Think about the theory we are going to develop below for the equations of the kind

$$\mathrm{d}Y_s = -f(t, X_s^{t,x}, Y_s, Z_s)\mathrm{d}s + \sigma(X^{t,x})Z_s\mathrm{d}W_s,$$

in this case one would have $Z_s = \nabla u(t, X_s^{t,x})$ with the original formulation (1) of the PDE.

This equation cannot be solved forward in time, indeed even when $f = 0$, in this case we have

$$\mathrm{d}Y_s = Z_s\mathrm{d}W_s,$$

and is clear that this equation has many solutions (just choose Z and then compute Y by giving its initial value). However if we consider it *backwards* in time, things start to be interesting: i.e. assume we give a final condition $Y_T = \xi$ where ξ is some $\mathscr{F}_T$-measurable random variable, then the adapted process $(Y_t)_{t\geqslant 0}$ has to satisfy

$$\xi = Y_T = Y_t + \int_t^T Z_s\mathrm{d}W_s$$

that is

$$\xi = Y_0 + \int_0^T Z_s\mathrm{d}W_s \tag{4}$$

and therefore for all $t \in [0,T]$

$$Y_t = \xi - \int_t^T Z_s\mathrm{d}W_s = Y_0 + \int_0^t Z_s\mathrm{d}W_s$$

with $Y_0 \hat{\in} \mathscr{F}_0$, let's assume that this is the trivial σ-field. Then $Y_0 = \mathbb{E}[\xi]$ and moreover if we are on Brownian filtration (i.e. the probability space is generated by the Brownian motion W) and $\xi \in L^2$, we deduce that there must exists a predictable process $Z \in L^2_{\mathscr{P}}(\mathbb{R}_+ \times \Omega; \mathbb{R})$ such that (4) is statisfied. This by the martingale representation theorem. That the solution is unique is clear since if (Y', Z') is another solution with the same final condition then we have

$$0 = \int_0^T Z_s\mathrm{d}W_s - \int_0^T Z'_s\mathrm{d}W_s$$

but this is only possible if $Z = Z'$ which one shows by computing the expectation of the square of this quantity.

Solution theory for BSDEs

In the following we consider BSDEs of the general form

$$-\mathrm{d}Y_s = f(t,\omega,Y_s,Z_s)\mathrm{d}s - Z_s\mathrm{d}W_s, \qquad Y_T = \xi \tag{5}$$

where $(\Omega,\mathscr{F},\mathbb{P})$ is the canonical d-dimensional Wiener space, $\xi \in L^2(\Omega,\mathscr{F}_T,\mathbb{P};\mathbb{R}^n) = L^2(\mathscr{F}_T;\mathbb{R}^n)$ (i.e. ξ takes values in $\mathbb{R}^n$ and is $\mathscr{F}_T$ measurable) and Y,Z are adapted processes taking values respectively in $\mathbb{R}^n$ and $\mathbb{R}^{n\times d} \approx L(\mathbb{R}^d,\mathbb{R}^n)$. Morever $f\colon \mathbb{R}_+ \times \Omega \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \to \mathbb{R}^n$ (called the *generator* or *driver*) is an adapted process, i.e. $(y,z) \mapsto f(t,\omega,y,z)$ is measurable wrt. $\mathscr{F}_t$. Standard conditions are that

$$f(\cdot,\cdot,0,0) \in L^2_{\mathscr{P}}([0,T]\times\Omega;\mathbb{R}^n)$$

and there exists a constant L such that (Lipshitz condition)

$$|f(t,\omega,y_1,z_1) - f(t,\omega,y_2,z_2)| \leqslant L(|y_1-y_2| + |z_1-z_2|), \qquad y_1,y_2 \in \mathbb{R}^n, z_1,z_2 \in \mathbb{R}^{n\times d}$$

for almost every (t,ω).

Let us note that solutions to BSDEs are by definition only strong (because the given filtration is that of the driving Brownian motion).

Let us introduce the notations

$$L^2_T(V) := L^2_{\mathscr{P}}([0,T]\times\Omega;V).$$

Note that L^2 in the theory of BSDEs plays a particular role because at the core of the solution theory there is the martingale representation theorem.

Note that our driver is quite general and in applications (below) to PDEs one will take

$$f(t,\omega,y,z) = \tilde{f}(t,X^{t_0,x}(\omega),y,z)$$

for example.

Theorem 1. *Under these conditions the BSDE (5) has a unique solution* $(Y,Z) \in L^2_T(\mathbb{R}^n) \times L^2_T(\mathbb{R}^{n\times d})$.

Proof. The idea is to proceed via a fixpoint argument. We consider the map $\Phi\colon (Y,Z) \in L^2_T(\mathbb{R}^n) \times L^2_T(\mathbb{R}^{n\times d}) \mapsto (Y',Z') \in L^2_T(\mathbb{R}^n) \times L^2_T(\mathbb{R}^{n\times d})$ defined as follows. Fixed $(Y,Z) \in L^2_T(\mathbb{R}^n) \times L^2_T(\mathbb{R}^{n\times d})$ we let (Y',Z') be the unique solution to the equation

$$-\mathrm{d}Y'_s = f(t,\omega,Y_s,Z_s)\mathrm{d}s - Z'_s\mathrm{d}W_s, \qquad Y'_T = \xi \tag{6}$$

Note that the solution of this equation is explicitly given by the Brownian martingale representation theorem (MRT). Indeed we need to solve the integral equation

$$Y'_t = \xi - \int_t^T \mathrm{d}Y'_s = \xi + \int_t^T f(t,\omega,Y_s,Z_s)\mathrm{d}s - \int_t^T Z'_s\mathrm{d}W_s,$$

but we have

$$Y'_0 = \xi + \int_0^T f(t,\omega,Y_s,Z_s)\mathrm{d}s - \int_0^T Z'_s\mathrm{d}W_s,$$

so Z' is determined by the MRT applied to the L^2 random variable $\xi + \int_0^T f(t,\omega,Y_s,Z_s)\mathrm{d}s$ and

$$Y'_0 = \mathbb{E}\Big[\xi + \int_0^T f(t,\omega,Y_s,Z_s)\mathrm{d}s\Big].$$

As consequence

$$\mathbb{E}\Big[\xi + \int_0^T f(t,\omega,Y_s,Z_s)\mathrm{d}s\Big|\mathscr{F}_t\Big] = Y'_0 + \int_0^t Z'_s\mathrm{d}W_s = Y'_0 + \int_0^T Z'_s\mathrm{d}W_s - \int_t^T Z'_s\mathrm{d}W_s$$

$$= \underbrace{\xi + \int_t^T f(t,\omega,Y_s,Z_s)\mathrm{d}s - \int_t^T Z'_s\mathrm{d}W_s}_{=Y'_t} + \int_0^t f(t,\omega,Y_s,Z_s)\mathrm{d}s$$

so we concldue that we have

$$Y_t' = \mathbb{E}\left[\xi + \int_0^T f(s,\omega,Y_s,Z_s)\mathrm{d}s \middle| \mathscr{F}_t\right] - \int_0^t f(t,\omega,Y_s,Z_s)\mathrm{d}s$$

which gives an explicit formula for Y_t'. Note that there is no formula for Z' (it is implicitly determined by the MRT). This procedure defines the map Φ.

One has to prove that Φ is a contraction. In order to do this is convenient to use appropriate equivalent norms on $L_T^2(\mathbb{R}^n) \times L_T^2(\mathbb{R}^{n\times d})$ and we replace the L_T^2 norm by the norm

$$\|f\|_{L_{T,\beta}^2}^2 := \mathbb{E}\int_0^t e^{\beta s}|f(s)|^2 \mathrm{d}s$$

for some $\beta \geqslant 0$. And then one can show that Φ is a contraction on $L_{T,\beta}^2(\mathbb{R}^n) \times L_{T,\beta}^2(\mathbb{R}^{n\times d})$ for sufficiently large β. The idea is to take $(Y^1, Z^1), (Y^2, Z^2) \in L_{T,\beta}^2(\mathbb{R}^n) \times L_{T,\beta}^2(\mathbb{R}^{n\times d})$ and let $(\tilde{Y}^1, \tilde{Z}^1) = \Phi(Y^1, Z^1)$, $(\tilde{Y}^2, \tilde{Z}^2) = \Phi(\tilde{Y}^2, \tilde{Z}^2)$ then one uses the Ito formula on the process $t \mapsto e^{\beta t}|\tilde{Y}_t^1 - \tilde{Y}_t^2|^2$ to get

$$\begin{aligned} & e^{\beta t}|\tilde{Y}_t^1 - \tilde{Y}_t^2|^2 + \int_t^T e^{\beta s}|\tilde{Z}_s^1 - \tilde{Z}_s^2|^2 \mathrm{d}s + \beta\int_t^T e^{\beta s}|\tilde{Y}_s^1 - \tilde{Y}_s^2|^2 \mathrm{d}s \\ =& M_T - M_t + 2\int_t^T e^{\beta s}\langle \tilde{Y}_s^1 - \tilde{Y}_s^2, f(s,\omega,Y_s^1,Z_s^1) - f(s,\omega,Y_s^2,Z_s^2)\rangle \mathrm{d}s \end{aligned}$$

where M is uniformly integrable martingale. From this and with some trivial estimates one gets the contrc-tion property, that is for sufficienlty large β one has

$$\|\Phi(Y^1,Z^1) - \Phi(Y^2,Z^2)\|_{L_{T,\beta}^2(\mathbb{R}^n)\times L_{T,\beta}^2(\mathbb{R}^{n\times d})} \leqslant C_\beta \|(Y^1,Z^1) - (Y^2,Z^2)\|_{L_{T,\beta}^2(\mathbb{R}^n)\times L_{T,\beta}^2(\mathbb{R}^{n\times d})}$$

for some $C_\beta \in (0,1)$. Uniqueness is also an easy consequence of the contraction property of the map Φ. □