Backward Stochastic Differential Equations: an Introduction

Nicolas Perkowski

Abstract

This is a short introduction to the theory of Backward Stochastic Differential Equations (BSDEs). The main focus is on stochastic representations of Partial Differential Equations (PDEs) or Stochastic Partial Differential Equations (SPDEs). Proofs are mostly only sketched, references to the literature are given. I do not strive for the greatest generality, but rather attempt to give heuristic explanations.

Contents

1 Classical stochastic representations of PDEs

Say we want to find a stochastic representation for a certain class of PDEs. The easiest representation of this kind, and in fact the starting point from which we will motivate the generalizations, is the Kolmogorov backward equation: Let X be the solution of the Stochastic Differential Equation (SDE)

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{1}$$

where $X_t \in \mathbb{R}^d$, and W is an n-dimensional standard Brownian motion (and therefore σ is a $d \times n$-matrix). Define $\mathcal{L}$ as the action of the infinitesimal generator of X on C^2-functions:

$$\mathcal{L}\varphi(x) = \sum_{i=1}^{d} b_i(x)\partial_i\varphi(x) + \sum_{i,j=1}^{d} a_{ij}(t,x)\partial_{i,j}\varphi(x)$$

where $a_{ij} = 1/2(\sigma\sigma^*)_{ij}$, σ^* is the transpose of σ. Denote by $\mathbb{P}_x$ the probability measure under which X satisfies equation (1) with $X_0 = x$. If μ is a distribution on $\mathbb{R}^d$, then we can define a measure $\mathbb{P}$ under which X_0 has distribution μ and X satisfies the SDE (1) by setting

$$\mathbb{P}(A) = \int_{\mathbb{R}^d} \mathbb{P}_x(A)\mu(dx)$$

An application of Itô's formula implies the **Kolmogorov forward equation**:

$$\frac{d}{dt}\mathbb{E}(\varphi(X_t)) = \mathbb{E}(\mathcal{L}\varphi(X_t))$$

for $\varphi \in C_b^2$, the space of twice continuously differentiable bounded functions with their first and second derivative bounded. This equation is also known among physicists and engineers as the **Fokker-Planck equation**. Assume that X_t has a density $p(t,\cdot)$ for any $t \geq 0$. Then a formal calculation gives

$$\frac{d}{dt}\int_{\mathbb{R}^d} \varphi(x)p(t,x)dx = \int_{\mathbb{R}^d} \mathcal{L}\varphi(x)p(t,x)dx = \int_{\mathbb{R}^d} \varphi(x)\mathcal{L}^*p(t,x)dx$$

where $\mathcal{L}^*$ is the L^2-adjoint of $\mathcal{L}$. So at least formally we obtain the Fokker-Planck equation for the density of X_t:

$$\partial_t p(t,x) = \mathcal{L}^*p(t,x), \quad p(0,x) = p_0(x)$$

Note that this is a way to **characterize the solution of an SDE by solving a PDE**.

To get the opposite direction, i.e. to characterize a PDE by solving an SDE, we need to consider the backward equation. Let $u(t,x) := \mathbb{E}_x(\varphi(X_t))$. Then under certain regularity assumptions u solves a PDE:

$$\begin{aligned} u(t+h,x) &= \mathbb{E}_x(\mathbb{E}_x(\varphi(X_{t+h})|\mathcal{F}_h)) = \mathbb{E}_x(u(t,X_h)) \\ &= \mathbb{E}_x\left(u(t,X_0) + \int_0^h \mathcal{L}u(t,X_s)ds\right) \end{aligned}$$

where in the second step we used the Markov property of X, and in the third step we applied Itô's formula to $s \mapsto u(t,X_s)$. Hence

$$\begin{aligned} \partial_t u(t,x) &= \mathcal{L}u(t,x) \\ u(0,x) &= \varphi(x) \end{aligned} \tag{2}$$

This equation is called the **Kolmogorov backward equation**, and it does exactly what we wanted: instead of solving the PDE (2), we can now (at least theoretically) solve the SDE (1) and calculate $E_x(\varphi(X_t))$.

At this point a comment on the names of those two equations is appropriate: In the forward equation, $d/dt\mathbb{E}_x(\varphi(X_t)) = \mathbb{E}_x(\mathcal{L}\varphi(X_t))$, we perturb the final position X_t:

$$\mathcal{L}\varphi(X_t) = \lim_{h\to 0} \frac{\mathbb{E}_x(\varphi(X_{t+h}) - \varphi(X_t)|\mathcal{F}_t)}{h}$$

whereas in the backward equation, $d/dt\mathbb{E}_x(\varphi(X_t)) = \mathcal{L}\mathbb{E}_x(\varphi(X_t))$, we perturb the starting position: in $\mathcal{L}\mathbb{E}_x(\varphi(X_t))$, $\mathcal{L}$ acts on the initial condition x.

The Kolmogorov backward equation allows us to represent equations of the form $\partial_t u = \mathcal{L}u$. But we can even find stochastic representations for more general equations via diffusion processes: The classical **Feynman-Kac representation** works for linear equations of the type

$$\begin{aligned} \partial_t u(t,x) &= \mathcal{L}u(t,x) + f(x)u(t,x) + g(x) \\ u(0,x) &= \varphi(x) \end{aligned} \tag{3}$$

Let us consider only the case $g = 0$. Then under certain regularity assumptions, a solution of the above equation is given by

$$u(t,x) = \mathbb{E}_x\left(\varphi(X_t)\exp\left(\int_0^t f(X_s)ds\right)\right)$$

Formally we show this exactly as for the backward equation:

$$\begin{aligned} u(t+h,x) &= \mathbb{E}_x\left(\mathbb{E}_x\left(\varphi(X_{t+h})e^{\int_0^{t+h} f(X_s)ds}|\mathcal{F}_h\right)\right) = \mathbb{E}_x\left(u(t,X_h)e^{\int_0^h f(X_s)ds}\right) \\ &= \mathbb{E}_x\left(u(t,X_0) + \int_0^h \mathcal{L}u(t,X_s)e^{\int_0^s f(X_r)dr}ds + \int_0^h u(t,X_s)f(X_s)e^{\int_0^s f(X_r)dr}ds\right) \end{aligned}$$

and therefore

$$\partial_t u(t,x) = \mathcal{L}u(t,x) + f(x)u(t,x)$$

However the equations of the type (3) are all linear. In the following chapter we will see how to represent **semilinear parabolic PDEs**.

Bibliographic Notes The material treated here is classic. Good references for the Kolmogorov forward and backward equations are e.g. Revuz and Yor (1999), Chapter VII, or Karatzas and Shreve (1988), Chapter 5. The Feynman-Kac representation in a very general form can be found in Karatzas and Shreve (1988), Theorem 5.7.6.

2 Backward Stochastic Differential Equations

2.1 Motivation

We want to find a generalization of the Feynman-Kac formula, more precisely we want to be able to represent semilinear parabolic PDEs of the type

$$\partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),D_x u(t,x)) = 0$$

Let us assume for now that that this equation has a solution u (and we do not worry about initial conditions yet). If we can describe the dynamics of $Y_t = u(t,X_t)$, then we could for every (t,x) consider a version $X^{t,x}$ of X that starts at time t in x. This would imply $u(t,x) = u(t,X_t^{t,x}) = Y_t^{t,x}$. What would the dynamics of Y have to be? By Itô's formula

$$\begin{aligned}
dY_t &= (\partial_t u(t,X_t) + \mathcal{L}u(t,X_t))dt + D_x u(t,X_t)\sigma(X_t)dW_t \\
&= -f(t,X_t,Y_t,D_x u(t,X_t))dt + D_x u(t,X_t)\sigma(X_t)dW_t
\end{aligned}$$

This suggests to consider equations of a slightly less general type than above:

$$\partial_t u(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),D_x u(t,X_t)\sigma(X_t)) = 0 \tag{4}$$

If u solves this equation, we obtain for $Y_t = u(t,X_t)$ and $Z_t = D_x u(t,X_t)\sigma(X_t)$:

$$dY_t = -f(t,X_t,Y_t,Z_t)dt + Z_t dW_t \tag{5}$$

A solution to this equation consists of a pair of processes (Y,Z). Note that this equation does not make any sense if we consider it as a forward equation: For $f=0$ we obtain $dY_t = Z_t dW_t$. Of course then we can choose Z independently of the initial condition, and therefore there would be infinitely many solutions. However, if we consider it as a backward equation, then there is hope: We consider again the case $f=0$. If Y is adapted we get for any adapted and square integrable Z

$$Y_t = \mathbb{E}(Y_t|\mathcal{F}_t) = \mathbb{E}\left(\xi - \int_t^T Z_s dW_s \middle| \mathcal{F}_t\right) = \mathbb{E}(\xi|\mathcal{F}_t)$$

Therefore Y is a martingale. If the filtration F is now generated by the Brownian motion W, then by the martingale representation property (cf. Revuz and Yor (1999), Theorem (3.9) of Chapter V) there exists a unique predictable process Z such that $Y_t = Y_0 + \int_0^t Z_s dW_s$ which yields

$$Y_t = Y_T - \int_t^T Z_s dW_s = \xi - \int_t^T Z_s dW_s$$

We will show now that the equation (5) is well defined. For this we will consider equations of a more general type. After showing existence and uniqueness for the more general equations, we will return to (5) and show that under certain assumptions it gives indeed a stochastic representation for PDEs of type (4).

2.2 Existence and Uniqueness

Let $(\Omega, (\mathcal{F}_t), \mathcal{F}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions. Let W be a n-dimensional standard Brownian motion on Ω, and assume $(\mathcal{F}_t)$ is the filtration generated by W (as we have seen above, this will be important in what follows; usually BSDEs cannot be solved with respect to general filtrations). We borrow the following notation from El Karoui et al. (1997):

- $L^2_T(\mathbb{R}^d)$ is the space of $\mathcal{F}_T$-measurable random variables ξ satisfying $\mathbb{E}(|\xi|^2) < \infty$
- $H^2_T(\mathbb{R}^d)$ is the space of predictable processes Y s.t. $||Y||^2 = \mathbb{E}\left(\int_0^T |Y_t|^2 dt\right) < \infty$
- $H^2_{T,\beta}(\mathbb{R}^d)$ denotes $H^2_T(\mathbb{R}^d)$ equipped with the equivalent norm $||Y||^2_\beta = \mathbb{E}\left(\int_0^T e^{\beta t}|Y_t|^2 dt\right)$

We are interested in equations of the type

$$-dY_t = f(t, \omega, Y_t, Z_t)dt - Z_t dW_t, \qquad Y_T = \xi \tag{6}$$

or equivalently

$$Y_t = \xi + \int_t^T f(s, \omega, Y_s, Z_s)ds - \int_t^T Z_s dW_s$$

where $Y_t \in \mathbb{R}^d$, ξ is $\mathcal{F}_T$-measurable, and f is $\mathcal{P} \otimes \mathcal{B}^d \otimes \mathcal{B}^{d\times n}$-measurable. $\mathcal{P}$ is the predictable σ-algebra, and $\mathcal{B}^d$ is the Borel σ-algebra on $\mathbb{R}^d$. f will be called the **generator** of the BSDE. Note that this has nothing to do with the infinitesimal generator of a Markov process. A solution is a process (Y, Z) such that Y is continuous and **adapted**, and Z is **predictable** and satisfies $\int_0^T |Z_s|^2 ds < \infty$ a.s. That means that unlike in the deterministic case, BSDEs can not be considered as time-reversed SDEs. At time t, (Y_t, Z_t) is $\mathcal{F}_t$-measurable, so the process does not "know" the terminal condition yet (which is in $\mathcal{F}_T$)! Here we are only concerned with the "standard" case: The parameters f and ξ are called **standard parameters** if

- $\xi \in L^2_T(\mathbb{R}^d)$
- $f(\cdot, \cdot, 0, 0) \in H^2_T(\mathbb{R}^d)$
- f is uniformly Lipschitz: There exists L s.t. Lebesgue $\otimes\, \mathbb{P}$-a.s.

$$|f(t, \omega, y_1, z_1) - f(t, \omega, y_2, z_2)| \leq L(|y_1 - y_2| + |z_1 - z_2|)$$

  for any y_1, y_2, z_1, z_2.

Theorem 1. *Given standard parameters, the BSDE (6) has a unique solution (Y, Z) in $H^2_T(\mathbb{R}^d) \times H^2_T(\mathbb{R}^{n\times d})$*

We will only sketch the proof. It is based on the martingale representation property used already above, and on a Picard iteration scheme.

Proof. 1. For $f = 0$ we have already seen that for any terminal condition $\xi \in L^2_T$, there is a unique solution (Y, Z). This solution actually is in $H^2_T(\mathbb{R}^d) \times H^2_T(\mathbb{R}^{d\times n})$.

2. For general f and given processes $(y, z) \in H^2_T(\mathbb{R}^d) \times H^2_T(\mathbb{R}^{d\times n})$ we consider the following equation:

$$-dY_t = f(t, y_t, z_t)dt - Z_t dW_t, \quad Y_T = \xi \tag{7}$$

The ω-dependence of f will no longer be explicitly noted. Let (Y^0, Z^0) solve the equation with generator 0 and terminal condition $\xi + \int_0^T f(s, y_s, z_s)ds$. Then $(Y, Z) = (Y^0 - \int_0^\cdot f(s, y_s, z_s)ds, Z^0)$ is a solution to equation (7):

$$\begin{aligned} Y_t &= Y^0_t - \int_0^t f(s, y_s, z_s)ds \\ &= Y^0_T - \int_t^T (Z^0_s)dW_s - \int_0^t f(s, y_s, z_s)ds \\ &= \xi + \int_t^T f(s, y_s, z_s)ds - \int_t^T (Z^0_s)dW_s \end{aligned}$$

It is also in $H^2_T(\mathbb{R}^d) \times H^2_T(\mathbb{R}^{d\times n})$.

3. One can check that the map $\Phi(y, z) = (Y, Z)$, where (Y, Z) is the solution of (7), is a contraction on $H^2_{T,\beta}(\mathbb{R}^d) \times H^2_{T,\beta}(\mathbb{R}^{d\times n})$ for a suitable β. Hence there exists a unique fixed point (Y, Z) such that

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s dW_s$$

Further if we define (Y^0, Z^0) as the solution of

$$-dY^0_t = f(t, 0, 0)dt - (Z^0_t)dW_t, \quad Y^0_T = \xi$$

and then $(Y^{k+1}, Z^{k+1}) = \Phi(Y^k, Z^k)$, then (Y^k, Z^k) will converge to (Y, Z) in $H^2_{T,\beta}(\mathbb{R}^d) \times H^2_{T,\beta}(\mathbb{R}^{d\times n})$. But $H^2_{T,\beta}(\mathbb{R}^d) \times H^2_{T,\beta}(\mathbb{R}^{d\times n})$ is just $H^2_T(\mathbb{R}^d) \times H^2_T(\mathbb{R}^{d\times n})$ with an equivalent norm. Thus (Y, Z) is the unique solution of (6) in $H^2_T(\mathbb{R}^d) \times H^2_T(\mathbb{R}^{d\times n})$.

□

The technical difficulty of this proof consists in actually showing that Φ is a contraction. A proof of this fact can be found in El Karoui et al. (1997) ("a-priori-estimates").

2.3 BSDEs and PDEs: Forward-Backward SDEs

We now return to equation (5): Let for $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ $X^{t,x}$ be the solution of

$$\begin{aligned} dX^{t,x}_s &= b(s, X^{t,x}_s)ds + \sigma(s, X^{t,x}_s)dW_s,\ s \geq t \\ X^{t,x}_s &= x,\ s \leq t \end{aligned} \tag{8}$$

with $X_s^{t,x} \in \mathbb{R}^d$, Borel measurable $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \mapsto \mathbb{R}^{d\times n}$, and an n-dimensional standard Brownian motion W. Associate a BSDE to this SDE:

$$\begin{aligned} -dY_s^{t,x} &= f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})ds - Z_s^{t,x} dW_s,\ 0 \le s \le T \\ Y_T &= \Psi(X_T^{t,x}) \end{aligned} \tag{9}$$

With $Y_s^{t,x} \in \mathbb{R}^m$, $Z_s^{t,x} \in \mathbb{R}^{m\times n}$ and Borel measurable $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times n} \to \mathbb{R}^m$ and $\Psi : \mathbb{R}^d \to \mathbb{R}^m$. This system is called a **Forward-Backward Stochastic Differential Equation (FBSDE).** It is called **uncoupled** because the solution $(Y^{t,x}, Z^{t,x})$ of (9) does not interfere with the dynamics of the forward SDE (8).

In the following we will show that $(Y^{t,x}, Z^{t,x})$ can be expressed in terms of a deterministic function of time and state process $(X^{t,x})$. Under some regularity assumptions this function will solve a PDE. Conversely, the solution of the FBSDE will be a solution of the associated PDE.

We need the following Lipschitz conditions and growth constraints on the coefficient functions: Assume there is $C > 0$ and $p \ge 1/2$ s.t.

$$\begin{aligned} &|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \le C|x-y| \\ &|\sigma(t,x)| + |b(t,x)| \le C(1+|x|) \\ &|f(t,x,y_1,z_1) - f(t,x,y_2,z_2)| \le C(|y_1-y_2| + |z_1-z_2|) \\ &|f(t,x,0,0)| + |\Psi(x)| \le C(1+|x|^p) \end{aligned}$$

In this case the SDE (8) has a unique strong solution $X^{t,x}$ satisfying for all $p \ge 1/2$:

$$\sup_{s\le T} \mathbb{E}(|X_s^{t,x}|^{2p}) \le K(1+|x|^{2p})$$

for some $K > 0$ (cf. Karatzas and Shreve (1988), Thm. 5.2.9, for the statement with $p = 1$, for general p it follows easily with the Burkholder-Davis-Gundy inequality and Gronwall). This means that we obtain standard parameters for the BSDE (9) which therefore has a unique solution as well.

Let us show that $(Y_s^{t,x})$ can be expressed as $u(s, X_s^{t,x})$ for some deterministic function u. In a first step we will show that $(Y^{t,x}, Z^{t,x})$ inherits measurability properties of $X^{t,x}$.

Proposition 2. *Let $(\mathcal{F}_{t,s})_{s\in[t,T]}$ denote the completed right-continuous future σ-algebra of W after t, i.e.*

$$\mathcal{F}_{t,s}^0 = \cap_{r>s} \sigma(W_u - W_t : t \le u \le r)$$

and $\mathcal{F}_{t,s}$ is the completion of $\mathcal{F}_{t,s}^0$. Then the solution $(Y^{t,x}, Z^{t,x})_{s\in[t,T]}$ of (9) is $(\mathcal{F}_{t,\cdot})$-adapted. In particular, $Y_t^{t,x}$ is deterministic as well as $Y_s^{t,x}$ for $0 \le s \le t$.

Sketch of Proof.

- Consider the translated Brownian Motion $W'_s := W_{t+s} - W_t$ and its completed right-continuous filtration $\mathcal{F}'_s$.

- Consider the FBSDE

$$
\begin{aligned}
dX'_s &= b(s+t, X'_s)ds + \sigma(s+t, X'_s)dW'_s, \quad X'_0 = x \\
-dY'_s &= f(s+t, X'_s, Y'_s, Z'_s)ds - Z'_s dW'_s, \quad Y'_{T-t} = \Psi(X'_{T-t})
\end{aligned}
$$

- (X', Y', Z') is $(\mathcal{F}')$-adapted and $(X'_{s+t}, Y'_{s+t}, Z'_{s+t})$ solves the original FBSDE.
- Uniqueness of solution: $(Y^{t,x}_s, Z^{t,x}_s)$ is $(\mathcal{F}'_{s-t}) = (\mathcal{F}_{t,s})$-adapted.
- To see that $Y^{t,x}_s$ is deterministic for $s \le t$, consider $W'_u := W_{u+s} - W_s$ and repeat the above proof.

□

Proposition 3. *There exist two deterministic measurable functions u and v s.t. the solution $(Y^{t,x}, Z^{t,x})$ of (9) satisfies*

$$
\begin{aligned}
Y^{t,x}_s &= u(s, X^{t,x}_s) \\
Z^{t,x}_s &= v(s, X^{t,x}_s)\sigma(s, X^{t,x}_s)
\end{aligned}
$$

Sketch of Proof. 1. If f does not depend on (y, z):

$$
Y^{t,x}_s = \mathbb{E}\left(\int_s^T f(r, X^{t,x}_r)dr + \Psi(X^{t,x}_T) \middle| \mathcal{F}_s \right) = u(s, X^{t,x}_s)
$$

where $u(s,y) := \mathbb{E}\left(\int_s^T f(r, X^{s,y}_r)dr + \Psi(X^{s,y}_T)\right)$. This is true by Theorem 6.27 in Cinlar et al. (1980) which also implies the existence of $v : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{m\times d}$ s.t. $Z^{t,x}_s = v(s, X^{t,x}_s)\sigma(s, X^{t,x}_s)$

2. For general f use the iterative procedure from the proof of Theorem 1

   - define $Y^0 := Z^0 := 0$
   - define Y^{k+1}, Z^{k+1} as solution of

$$
-dY^{k+1}_s = f(s, X^{t,x}_s, Y^k_s, Z^k_s)ds - Z^{k+1}_s dW_s, \; Y^{k+1}_T = \Psi(X^{t,x}_T)
$$

   - recursively $Y^k_s = u^k(s, X^{t,x}_s)$
   - set $u_i(s,x) := \limsup_{k\to\infty}(u^k(s,x))_i$
   - $u_i(s, X^{t,x}_s) = \limsup_{k\to\infty}(u^k(s, X^{t,x}_s))_i = \lim_{k\to\infty}(Y^k_s)_i = (Y^{t,x}_s)_i$
   - same for v and Z

□

Consider the semilinear parabolic PDE associated to our problem:

$$
\begin{aligned}
&\partial_t u(t,x) + \mathcal{L}_t u(t,x) + f\left(t, x, u(t,x), D_x u(t,x)\sigma(t,x)\right) = 0 \\
&u(T,x) = \Psi(x)
\end{aligned}
\tag{10}
$$

where $\mathcal{L}_t$ is again the differential operator associated to X:

$$\mathcal{L}_t\varphi(x) = \sum_{i=1}^{d} b_i(t,x)\partial_i\varphi(x) + \sum_{i,j=1}^{d} a_{ij}(t,x)\partial_{i,j}\varphi(x)$$

for $\varphi \in C^2(\mathbb{R}^d, \mathbb{R})$. For multidimensional φ, $(\mathcal{L}_t\varphi)_i = \mathcal{L}_t\varphi_i$. Again $a_{ij} = 1/2$ $(\sigma(t,x)\sigma(t,x)^*)_{ij}$.

Theorem 4 (Generalization of the Feynman-Kac formula)**.** *Assume $u \in C^{1,2}([0,T]\times \mathbb{R}^d, \mathbb{R}^m)$ is a solution of (10) s.t. for some $C > 0$ and some $k \in \mathbb{N}$*

$$|u(s,x)| + |D_x u(s,x)\sigma(s,x)| \leq C(1+|x|^k) \quad \forall (s,x) \in [0,T]\times\mathbb{R}^d$$

Then

$$Y_s^{t,x} = u(s, X_s^{t,x}),\ Z_s^{t,x} = D_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x})\ \forall s \in [0,T]$$

where $(X^{t,x}, Y^{t,x}, Z^{t,x})$ is the unique solution of the associated FBSDE (8), (9). **In particular, u has the representation** *$u(t,x) = Y_t^{t,x}$, and therefore such a solution u is unique.*

Proof. We simply apply Itô's formula and use the fact that u solves (10):

$$\begin{aligned} du(s, X_s^{t,x}) &= \left(\partial_s u(s, X_s^{t,x}) + \mathcal{L}_s u(s, X_s^{t,x})\right) ds + D_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x}) dW_s \\ &= -f\left(s, X_s^{t,x}, u(s, X_s^{t,x}), D_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x})\right) ds \\ &\quad + D_x u(s, X_s^{t,x})\sigma(s, X_s^{t,x}) dW_s \end{aligned}$$

Furthermore we have $u(T, X_T^{t,x}) = \Psi(X_T^{t,x})$ and therefore $(u(s, X_s^{t,x}),\ D_x u(s,\ X_s^{t,x})$ $\sigma(s, X_s^{t,x}))$ is the unique solution of BSDE (9). □

This result is exactly what we set out to find: a stochastic representation of solutions of semilinear PDEs. In fact we can do even better than that: under strong regularity assumptions on the coefficients, one can show that $u(t,x) = Y_t^{t,x}$ is a solution of the associated PDE (not assuming a priori that a solution exists). In dimension $d = 1$, one only needs Lipschitz assumptions and growth constraints on the coefficients to show that $Y_t^{t,x}$ is a "viscosity solution" of the PDE. These results can be for example found in El Karoui et al. (1997), Theorem 4.2, or Pardoux and Peng (1992), Theorem 3.2 and Theorem 4.3. We will not go into detail here.

2.4 BSDEs and Stochastic Control

To see how BSDEs arise naturally in the stochastic version of Pontryagin's Maximum Principle, let us briefly recall the deterministic version: Let the dynamics of x be governed by a controlled ordinary differential equation:

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0$$

and consider a payoff function

$$J(u) = \int_0^T r(t, x(t), u(t))dt + g(x(T))$$

which we want to maximize over the set of admissible controls. Let $u_0(t), t \in [0, T]$ be an optimal control, i.e. $J(u_0) \geq J(u')$ for all admissible u'. Introduce a perturbed version of u, $u_\varepsilon = u_0 + \varepsilon u$ for some u (it is not clear that there exists an admissible u_ε defined like this, but we are only arguing formally.) Denote x_0 the solution corresponding to u_0, and x_ε the one corresponding to u_ε. Write

$$y(t) = \frac{d}{d\varepsilon} x_\varepsilon(t)\bigg|_{\varepsilon=0}$$

Then we obtain

$$\begin{aligned}
y(t) &= \lim_{\varepsilon\to 0} \frac{x_\varepsilon(t) - x_0(t)}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^t \frac{d}{ds}(x_\varepsilon(s) - x_0(s))ds \\
&= \int_0^t \lim_{\varepsilon\to 0} \frac{f(x_\varepsilon(s), u_\varepsilon(s)) - f(x_0(s), u_0(s))}{\varepsilon} ds \\
&= \int_0^t \left(D_x f(x_0(s), u_0(s))y(s) + D_u f(x_0(s), u_0(s))u(s)\right) ds
\end{aligned}$$

so that

$$\dot{y}(t) = D_x f(x_0(t), u_0(t))y(t) + D_u f(x_0(t), u_0(t))u(t), \quad y(0) = 0$$

Let us vary the control and see how the payoff changes:

$$\begin{aligned}
\frac{d}{d\varepsilon} J(u_\varepsilon)\bigg|_{\varepsilon=0} &= \int_0^T (D_x r(t, x_0(t), u_0(t))y(t) + D_u r(t, x_0(t), u_0(t))u(t))dt \\
&\quad + D_x g(x_0(T))y(T) = 0
\end{aligned} \tag{11}$$

because $\varepsilon \mapsto J(u_\varepsilon)$ has a local maximum in 0. We would like to have only terms of the form $\int(\dots)u(t)dt$, because if this is zero for all admissible controls u, and there are "enough" admissible controls, then $(\dots)$ must already be zero. Let us introduce an adjoint variable p_0 that will take care of the "bad" terms:

$$\begin{aligned}
\dot{p}_0(t) &= -D_x f(t, x_0(t), u_0(t))^* p_0(t) - D_x r(t, x_0(t), u_0(t))^* \\
p_0(T) &= D_x g(x_0(T))^*
\end{aligned}$$

We obtain with this choice of p_0 and because $y(0) = 0$:

$$\begin{aligned}
D_x g(x_0(T))y(T) &= p_0(T)^* y(T) - p_0(0)^* y(0) \\
&= \int_0^T (\dot{p}_0(t)^* y(t) + p_0(t)^* \dot{y}(t))dt \\
&= \int_0^T ((-p_0^* D_x f - D_x r)y + p_0^*(D_x f y + D_u f u))dt \\
&= \int_0^T (p_0^* D_u f u - D_x r y)dt
\end{aligned} \tag{12}$$

Combining (11) and (12) gives

$$\frac{d}{d\varepsilon} J(u_\varepsilon)\bigg|_{\varepsilon=0} = \int_0^T (p_0^* D_u f + D_u r) u dt$$

so that heuristically $p_0^* D_u f - D_u r = 0$. But if we define the **Hamiltonian**

$$H(t, u, x, p) = p^* f(t, x, u) + r(t, x, u)$$

this is just $D_u H$. This implies heuristically that for any t, $u \mapsto H(t, u, x_0(t), p_0(t))$ has a local extremum in $u_0(t)$. To see that this should actually be a maximum, note that for $\varepsilon > 0$:

$$0 \geq J(u_\varepsilon) - J(u_0) \simeq \varepsilon \int_0^T (p_0 D_u f - D_u r) u dt = \int_0^T D_u H(u_\varepsilon - u_0) dt$$

We obtain the Pontryagin Maximum Principle:

Theorem. *Let u_0 be an optimal control, and let x_0 be the associated trajectory. Then under suitable assumptions there exists p_0 such that*

$$\begin{aligned} \frac{d}{dt} x_0(t) &= D_p H(t, u_0(t), x_0(t), p_0(t)), \quad x_0(0) = x_0 \\ \frac{d}{dt} p_0(t) &= -D_x H(t, u_0(t), x_0(t), p_0(t)), \quad p_0(T) = D_x g(x_0(T))^* \end{aligned}$$

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

$$H(t, u_0(t), x_0(t), p_0(t)) = \max_u H(t, u, x_0(t), p_0(t))$$

The map $t \mapsto H(t, u_0(t), x_0(t), p_0(t))$ is constant.

How do we find a candidate for an optimal control? Well, solve for all t, x, p

$$u(t, x, p) = \max_u H(t, u, x, p)$$

and then solve

$$\begin{aligned} &\frac{d}{dt} \begin{pmatrix} x(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} D_p H(t, u(t, x(t), p(t)), x(t), p(t)) \\ -D_x H(t, u(t, x(t), p(t)), x(t), p(t)) \end{pmatrix} \\ &x(0) = x_0 \quad p(T) = D_x g(x(T)) \end{aligned}$$

Finally set $u_0(t) = u(t, x(t), p(t))$.

In the stochastic case we consider the following control problem:

$$\begin{aligned} dX_t &= b(X_t, u_t) dt + \sigma(X_t, u_t) dW_t \\ X_0 &= x_0 \end{aligned}$$

where $X_t \in \mathbb{R}^d$, W is an n-dimensional standard Brownian motion, and u is a predictable control process that takes its values in some set $\mathbb{U}$. We want to maximize

$$J(u) = \mathbb{E}\left(\int_0^T r(s, X_s, u_s)ds + g(X_T)\right)$$

for some measurable $r : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{U} \to \mathbb{R}$ and concave $g : \mathbb{R}^d \to \mathbb{R}$. It is possible to derive the maximum principle along the same lines as in the deterministic case. This is done in Yong and Zhou (1999), but it is very technical, its core argument being a stochastic Taylor expansion. Here we start with the right Hamiltonian and show the maximum principle under some concavity condition, using only integration by parts. This proof is less intuitive, that is why I included the deterministic case, to give the right intuition as to where the Hamiltonian and the adjoint equation come from. Assume there exists $C > 0$ s.t.

- $|b(x,u) - b(y,u)| + |\sigma(x,u) - \sigma(y,u)| \leq L|x - y|$
- $r(\cdot,\cdot,u)$ is a continuous function of (t,x) for all u
- g is concave and C^1
- $|r(t,x,u)| + |g(x)| \leq C(1 + |x|^2)$ for all t, x, u
- b, σ and r are all differentiable in x

Define the generalized Hamiltonian

$$\begin{aligned}
&H : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{U} \times \mathbb{R}^d \times \mathbb{R}^{d\times n} \to \mathbb{R} \\
&H(t,x,u,y,z) = b(x,u)^* y + \mathrm{tr}(\sigma(x,u)^* z) + r(t,x,u)
\end{aligned}$$

and let the adjoint (Y, Z) be the solution of

$$\begin{aligned}
-dY_t &= D_x H(t, X_t, u_t, Y_t, Z_t)^* dt - Z_t dW_t \\
Y_T &= D_x g(X_T)^* \qquad (13)
\end{aligned}$$

for a given control process u.

Theorem 5. *Let $\hat{u}$ be a control process. Suppose $(\hat{Y}, \hat{Z})$ solves the adjoint BSDE (13). Suppose for all $t \in [0,T]$*

$$H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t) = \max_{u\in\mathbb{U}} H(t, \hat{X}_t, u, \hat{Y}_t, \hat{Z}_t)$$

and that a.s.

$$(x,u) \mapsto H(t,x,u,\hat{Y}_t,\hat{Z}_t) \qquad (14)$$

is a concave function for all $t \in [0,T]$. Then $\hat{u}$ is optimal, i.e. $J(\hat{u}) = \max_u J(u)$.

Sketch of Proof. Let u be another control process, and denote X, Y, Z the associated processes. It suffices to show $J(\hat{u}) \geq J(u)$.

$$J(\hat{u}) - J(u) = \mathbb{E}\left(\int_0^T r(t, \hat{u}_t, \hat{X}_t) - r(t, u_t, X_t)dt + g(\hat{X}_T) - g(X_T)\right) \tag{15}$$

Let us rewrite this expression:

$$\begin{aligned} r(t, \hat{u}_t, \hat{X}_t) - r(t, u_t, X_t) = H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t) - H(t, X_t, u_t, \hat{Y}_t, \hat{Z}_t) \\ - [b(\hat{X}_t, \hat{u}_t) - b(X_t, u_t)]^* \hat{Y}_t \\ - \operatorname{tr}[(\sigma(\hat{X}_t, \hat{u}_t) - \sigma(X_t, u_t))^* \hat{Z}_t] \end{aligned} \tag{16}$$

on the other hand because g is concave and differentiable:

$$\begin{aligned} &\mathbb{E}(g(\hat{X}_T) - g(X_T)) \geq \mathbb{E}(D_x g(\hat{X}_T)(\hat{X}_T - X_T)) = \mathbb{E}(\hat{Y}_T^*(\hat{X}_T - X_T)) \\ &\quad = \mathbb{E}\left(\int_0^T -D_x H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t)(\hat{X}_t - X_t)dt + \int_0^T \hat{Y}_t^*[b(\hat{X}_t, \hat{u}_t) - b(X_t, u_t)]dt \right. \\ &\qquad \left. + \int_0^T \operatorname{tr}[\hat{Z}_t^*(\sigma(\hat{X}_t, \hat{u}_t) - \sigma(X_t, u_t))]dt\right) \end{aligned} \tag{17}$$

where we used integration by parts, the fact that $X_0 = \hat{X}_0$, and that the expectation of the stochastic integrals should be zero. Combining (15), (16) and (17) we obtain

$$\begin{aligned} J(\hat{u}) - J(u) \geq \mathbb{E}\left(\int_0^T [H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t) - H(t, X_t, u_t, \hat{Y}_t, \hat{Z}_t) \right. \\ \left. - D_x H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t)(\hat{X}_t - X_t)]dt\right) \end{aligned}$$

because of the concavity condition (14),

$$\begin{aligned} H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t) - H(t, X_t, u_t, \hat{Y}_t, \hat{Z}_t) \geq D_x H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t)(\hat{X}_t - X_t) \\ + D_u H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t)(\hat{u}_t - u_t) \end{aligned}$$

which implies

$$J(\hat{u}) - J(u) \geq \mathbb{E}\left(\int_0^T -D_u H(t, \hat{X}_t, \hat{u}_t, \hat{Y}_t, \hat{Z}_t)(u_t - \hat{u}_t)dt\right)$$

but because $u \mapsto H(t, \hat{X}_t, u, \hat{Y}_t, \hat{Z}_t)$ takes its maximum in $\hat{u}_t$, this is always nonnegative. □

Note however that to obtain a candidate for an optimal control, in general we need to solve a **coupled** FBSDE, where the backward components (Y, Z) influence the dynamics of the forward component X:

1. solve $u(t, x, y, z) = \max_u H(t, x, u, y, z)$ for all t, x, y, z

2. solve the coupled FBSDE

$$
\begin{aligned}
dX_t &= b(X_t, u(t, X_t, Y_t, Z_t))dt + \sigma(X_t, u(t, X_t, Y_t, Z_t))dW_t, \quad X_0 = x_0 \\
-dY_t &= H(t, X_t, u(t, X_t, Y_t, Z_t), Y_t, Z_t)dt - Z_t dW_t, \quad Y_T = D_x g(X_T)^*
\end{aligned}
$$

3. set $\hat{u}_t = u(t, X_t, Y_t, Z_t)$

Morally, it is not surprising that we do not end up with an uncoupled FBSDE: Those represent semilinear PDEs, whereas the HJB equation from the dynamic programming principle for optimal diffusion control is fully nonlinear. Therefore also the FBSDE from the maximum principle should be more difficult than an uncoupled FBSDE.

Bibliographic Notes The material on Feynman-Kac representations and on existence and uniqueness of solutions is essentially taken from El Karoui et al. (1997), which is a good introduction to BSDEs. An updated version of that article is El Karoui et al. (2008). BSDEs also arise naturally in mathematical finance. They were introduced by Bismut in 1973, but their systematic study began with Pardoux and Peng (1990). It is also possible to give stochastic representations for fully nonlinear parabolic PDEs in terms of so-called second order BSDEs. This was done in Cheridito et al. (2007), and it allows e.g. to find stochastic representations for the HJB equations arising in stochastic control. The material on deterministic optimal control is from Evans, while the material on stochastic control is taken pretty much one-to-one from Pham (2009).

3 BDSDEs

3.1 Motivation

In the previous sections we were concerned with PDEs and their stochastic representations. In this section we will find stochastic representations for SPDEs. This may sound odd, since SPDEs are themselves stochastic equations. But what we will do is in fact find a finite-dimensional stochastic (ordinary) differential equation that represents the SPDE. Heuristically, we will need one "stochasticity" to represent the partial differential operator, and another "stochasticity" to represent the random noise in the SPDE. Hence this will be a "doubly" stochastic differential equation.

Just as in the deterministic case, we will find representations for backward equations. They will be of the type

$$
\begin{aligned}
u(t, x) = \Psi(x) &+ \int_t^T (\mathcal{L}_s u(s, x) + f(s, x, u(s, x), D_x u(s, x)\sigma(s, x)))ds \\
&+ \int_t^T g(s, x, u(s, x), D_x u(s, x)\sigma(s, x))d\overleftarrow{B}_s
\end{aligned} \tag{18}
$$

We have to specify what we mean by backward equation: This has not much to do with the BSDEs from the last section. There, the solution (Y, Z) was adapted to the past, i.e. $\mathcal{F}_t$, and the stochastic integral was just Itô's integral. Here we are looking for solutions adapted to the future of B: Define

$$\mathcal{F}^{0,B}_{t,s} = \cap_{r<t}\sigma(B_u - B_r : r \le u \le s)$$

and $\mathcal{F}^B_{t,s}$ is the completion of $\mathcal{F}^{0,B}_{t,s}$. A solution u will have to satisfy $u(t,x) \in \mathcal{F}^B_{t,T}$ for all (t,x). Under this assumption we can interpret $d\overleftarrow{B}$ as backward Itô integral: The backward integral of a simple function $H_s = \sum_{i=1}^n H_i \mathbb{1}_{[T_{i-1},T_i)}(s)$ with $H_i \in \mathcal{F}_{T_i,T}$ is defined as

$$\int H_s d\overleftarrow{B}_s = \sum_{i=1}^n H_i(B_{T_i} - B_{T_{i-1}})$$

and for general locally square integrable H, predictable w.r.t. $(\mathcal{F}_{\cdot,T})$, it is extended via the Itô isometry, just like the forward integral. So in fact for $H'_t = H_{T-t}$ and for the Brownian motion $B'_t = B_T - B_{T-t}$

$$\int_0^t H_s d\overleftarrow{B}_s = \int_{T-t}^T H'_s dB'_s$$

where on the right hand side we just have Itô's forward integral. This allows us to reverse time in (18) to obtain a forward equation: Let u be a solution and define $u'(t,x) = u(T-t,x)$ and B' as above. Then u' solves

$$\begin{aligned} u'(t,x) = \Psi(x) &+ \int_0^t (\mathcal{L}_{T-s}u'(s,x) + f(T-s,x,u'(s,x),D_x u'(s,x)\sigma(T-s,x)))ds \\ &+ \int_0^t g(T-s,x,u'(s,x),D_x u'(s,x)\sigma(T-s,x))dB'_s \end{aligned} \tag{19}$$

in the classical sense. This means that we do not need any new theory for backward SPDEs of the above type, and we can just solve (19) and then reverse time. It also means that we can always reverse time for forward equations like (19) and end up with a backward equation of type (18). Note however that we do not allow the noise dB' to depend on the space variable. That is, we only allow for finite dimensional noise. Luckily the SPDEs that arise in nonlinear filtering are of this type.

Just like in the previous chapter we will find an $\mathbb{R}^m$-valued process Y s.t. $Y_s = u(s,X_s)$ where X is the diffusion associated to $\mathcal{L}_s$, driven by a Brownian motion W that is independent of B. Let us again derive heuristically what the dynamics of Y should be:

$$\begin{aligned} Y_{t+h} - Y_t &= u(t+h,X_{t+h}) - u(t,X_{t+h}) + u(t,X_{t+h}) - u(t,X_t) \\ &= -\int_t^{t+h} (\mathcal{L}_s u(s,X_{t+h}) + f(s,X_{t+h},u(s,X_{t+h}),D_x u(s,X_{t+h})\sigma(s,X_{t+h})))ds \\ &\quad - \int_t^{t+h} g(s,X_{t+h},u(s,X_{t+h}),D_x u(s,X_{t+h})\sigma(s,X_{t+h}))d\overleftarrow{B}_s \\ &\quad + \int_t^{t+h} \mathcal{L}_s u(t,X_s)ds + \int_t^{t+h} D_x u(t,X_s)\sigma(s,X_s)dW_s \end{aligned}$$

We see that in the limit the differential operator terms will vanish, and if we set again $Z_t = D_x u(t, X_t)\sigma(t, X_t)$, we obtain

$$dY_t = -f(s, X_s, Y_s, Z_s)ds - g(s, X_s, Y_s, Z_s)d\overleftarrow{B}_s + Z_s dW_s$$

3.2 Existence and Uniqueness

The basic ideas to prove existence and uniqueness are the same as for normal BSDEs: in the proofs, the martingale representation property of Brownian motion and a Picard iteration scheme are used. The actual proofs however are more involved. For this reason we only cite the results.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space carrying two independent standard Brownian motions $(W_t : t \geq 0)$ with values in $\mathbb{R}^n$ and $(B_t : t \geq 0)$ with values in $\mathbb{R}^k$. Let $\mathcal{F}^B_{t,s}$ be as above, and define $\mathcal{F}^{0,W}_t = \cap_{s>t}\sigma(W_r : 0 \leq r \leq s)$ and $\mathcal{F}^W_t$ as its completion. Fix $T > 0$ and define for $t \leq T$:

$$\mathcal{F}_t = \mathcal{F}^B_{t,T} \vee \mathcal{F}^W_t$$

Note that this is not a filtration, it is neither decreasing nor increasing in t. Analogously to the BSDE section, we introduce the following notation

- $L^2_T(\mathbb{R}^m)$ is the space of $\mathbb{R}^m$-valued $\mathcal{F}_T$-measurable random variables ξ satisfying $\mathbb{E}(|\xi|^2) < \infty$
- $H^2_T(\mathbb{R}^m)$ is the space of $\mathbb{R}^m$-valued processes Y s.t. Y_t is $\mathcal{F}_t$-measurable and $\mathbb{E}\left(\int_0^T |Y_t|^2 dt\right) < \infty$
- $S^2_T(\mathbb{R}^m)$ is the space of **continuous** $\mathbb{R}^m$-valued processes Y s.t. Y_t is $\mathcal{F}_t$-measurable and $\mathbb{E}\left(\sup_{0\leq t\leq T} |Y_t|^2\right) < \infty$

The type of equation we want to solve is

$$\begin{aligned} -dY_t &= f(t, Y_t, Z_t)dt + g(t, Y_t, Z_t)d\overleftarrow{B}_t - Z_t dW_t \\ Y_T &= \xi \end{aligned} \tag{20}$$

where Y_t is $\mathbb{R}^m$-valued, Z_t is $\mathbb{R}^{m\times n}$-valued, and $f : \mathbb{R}_+ \times \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times n} \to \mathbb{R}^m$ and $g : \mathbb{R}_+ \times \Omega \times \mathbb{R}^m \times \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times k}$ are measurable. (Y, Z) will be called solution of equation (20) if $(Y, Z) \in S^2_T(\mathbb{R}^m) \times H^2_T(\mathbb{R}^{m\times n})$ and it solves the integral equation. We will equip all matrix spaces with the Frobenius norm: $||A||^2 = \mathrm{tr}(AA^*)$.

f, g and ξ will be called **standard parameters** if

- $\xi \in L^2_T(\mathbb{R}^m)$
- for any $(y, z) \in \mathbb{R}^m \times \mathbb{R}^{m\times n}$: $f(\cdot, \cdot, y, z) \in H^2_T(\mathbb{R}^m)$ and $g(\cdot, \cdot, y, z) \in H^2_T(\mathbb{R}^{m\times k})$

- f and g satisfy Lipschitz conditions and g is a contraction in z: there exist $L > 0$ and $0 < \alpha < 1$ s.t. for any (ω, t) and y_1, y_2, z_1, z_2:

$$|f(t,\omega,y_1,z_1) - f(t,\omega,y_2,z_2)|^2 \leq L(|y_1 - y_2|^2 + ||z_1 - z_2||^2)$$
$$||g(t,\omega,y_1,z_1) - g(t,\omega,y_2,z_2)||^2 \leq L|y_1 - y_2|^2 + \alpha||z_1 - z_2||^2$$

Theorem 6. *Given standard parameters, the BDSDE (20) has a unique solution (Y, Z) in $S_T^2(\mathbb{R}^m) \times H_T^2(\mathbb{R}^{m\times n})$.*

The proof can be found in Pardoux and Peng (1994), Theorem 1.1.

3.3 BDSDEs and SPDEs

Now we show that solutions of backward SPDEs can be represented as solutions of BDSDEs. Recall equation (18):

$$\begin{aligned} u(t,x) = \Psi(x) &+ \int_t^T (\mathcal{L}_s u(s,x) + f(s,x,u(s,x),D_x u(s,x)\sigma(s,x)))ds \\ &+ \int_t^T g(s,x,u(s,x),D_x u(s,x)\sigma(s,x))d\overleftarrow{B}_s \end{aligned}$$

with $u : \mathbb{R}_+ \times \Omega \times \mathbb{R}^d \to \mathbb{R}^m$, and measurable $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times n} \to \mathbb{R}^m$, $g : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times k}$, and $\Psi : \mathbb{R}^d \to \mathbb{R}^m$. $\mathcal{L}_s$ is again a differential operator corresponding to a diffusion process (cf. (8)):

$$\begin{aligned} dX_s^{t,x} &= b(s, X_s^{t,x})ds + \sigma(s, X_s^{t,x})dW_s,\ s > t \\ X_s^{t,x} &= x,\ s \leq t \end{aligned}$$

with $X_t \in \mathbb{R}^d$. Associate a BDSDE to (18) and (8):

$$\begin{aligned} -dY_s^{t,x} &= f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})ds + g(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})d\overleftarrow{B}_s - Z_s^{t,x}dWs \\ Y_T^{t,x} &= \Psi(X_T^{t,x}) \end{aligned} \tag{21}$$

Assume there exists $p \geq 1$, $C > 0$, $0 < \alpha < 1$ s.t.

$$\begin{aligned} &|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq C|x - y| \\ &|\sigma(t,x)| + |b(t,x)| \leq C(1 + |x|) \\ &|f(t,x,y_1,z_1) - f(t,x,y_2,z_2)|^2 \leq L(|y_1 - y_2|^2 + ||z_1 - z_2||^2) \\ &||g(t,x,y_1,z_1) - g(t,x,y_2,z_2)||^2 \leq L|y_1 - y_2|^2 + \alpha||z_1 - z_2||^2 \\ &|f(t,x,y,z)|^2 + |\Psi(x)|^2 \leq C(1 + |x|^p) \\ &gg^*(t,x,y,z) \leq zz^* + C(||g(t,x,0,0)||^2 + |y|^2)I \\ &\Psi \in C^2(\mathbb{R}^d, \mathbb{R}^m) \end{aligned}$$

where $A \leq B$ for symmetric positive semi-definite matrices A and B means that $B - A$ is symmetric and positive semi-definite. Define $f'(s,\omega,y,z) = f(s, X_s^{t,x}(\omega), y, z)$, analogously for g. Under the above assumptions, f', g' and $\Psi(X_T^{t,x})$ are standard parameters, and therefore (21) has a unique solution.

Theorem 7. *Under the assumptions stated above, let u be a solution of (18): $u(t,x) \in \mathcal{F}^B_{t,T}$ for any (t,x), $u \in C^{0,2}([0,T] \times \mathbb{R}^d, \mathbb{R}^m)$, and u solves the integral equation (18). Denote $X^{t,x}$ and $(Y^{t,x}, Z^{t,x})$ the solutions of (8) respectively (21). Then for any $s \in [0,T]$:*

$$Y^{t,x}_s = u(s, X^{t,x}_s)$$

and in particular $u(t,x) = Y^{t,x}_t$.

The proof is given in Pardoux and Peng (1994), Theorem 3.1. This theorem is great except for one fact: It requires a priori knowledge of the existence of a strong solution of the SPDE, i.e. we need to know that the solution u is twice **strongly differentiable**. This is in general not given, since in SPDE theory one usually works on Sobolev spaces with weak derivatives. We could work on Sobolev spaces of higher order, and then use Sobolev embedding to obtain strong derivatives, but this requires much extra work. Therefore the following converse result will be more useful in practice - even though it requires much stronger assumptions!

Let b, σ, f, g satisfy all the above assumptions, and in addition let

- $b \in C^{0,3}_b(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R}^d)$, $\sigma \in C^{0,3}_b(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R}^{d\times n})$
- $f \in C^{0,3}_b(\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times n}, \mathbb{R}^m)$, $g \in C^{0,3}_b(\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times n}, \mathbb{R}^{m\times k})$
- $D_z g(t,x,y,z)\theta\theta^*(D_z g(t,x,y,z))^* \leq \theta\theta^*$ for any (t,x,y,z) and θ - again in the sense of positive symmetric matrices
- $\Psi \in C^3_p(\mathbb{R}^d, \mathbb{R}^m)$ where C^3_p are the functions in C^3 which, together with all their partial derivatives, grow at most polynomially at infinity.

Theorem 8. *Under the above assumptions, $u(t,x) = Y^{t,x}_t$ is the unique classical solution of (18).*

This is Theorem 3.2 in Pardoux and Peng (1994). Note that here we also work with strong derivatives - but we do not a priori require them, the theorem actually tells us that the solution will be strongly differentiable!

3.4 Applications to Filtering

Consider the following nonlinear filtering problem: There is an unobservable random dynamical system which is given as the solution of the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$$

with $X_t \in \mathbb{R}^d$, $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times n}$ Borel measurable, and an n-dimensional standard Brownian motion W. Assume we have the following observation processes

$$Y_t = \int_0^t h(s, X_s)ds + B_t$$

with $Y_t \in \mathbb{R}^n$, and B is an n-dimensional Brownian motion independent of W. Then under very general conditions (essentially only an ellipticity condition on X and some boundedness assumptions), the law of X conditioned on Y can be described by the solution of an SPDE, the so-called Zakai equation. More precisely, assume

- σ is Lipschitz continuous and bounded with bounded derivative
- b and h are Lipschitz continuous and bounded
- there exists $\alpha > 0$ s.t. $|\sigma(t,x)^*\theta|^2 \geq \alpha|\theta|^2$ for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, $\theta \in \mathbb{R}^d$

Then there exist unique strong solutions X and Y of the above equations. Denote the filtration generated by B and W by $\mathcal{F}_t$, and denote the filtration generated by Y as $\mathcal{G}_t$. We are interested in describing $\mathbb{E}(f(X_t)|\mathcal{G}_t)$. For this define the equivalent measure $\tilde{\mathbb{P}}$ on $\mathcal{F}_T$ under which Y is a Brownian motion:

$$\left.\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t^{-1},\ 0 \leq t \leq T$$

with

$$Z_t = \exp\left(\int_0^t h(s,X_s)^* dY_s - \frac{1}{2}\int_0^t |h(s,X_s)|^2 ds\right)$$

and thus

$$Z_t^{-1} = \exp\left(-\int_0^t h(s,X_s)^* dW_s - \frac{1}{2}\int_0^t |h(s,X_s)|^2 ds\right)$$

By Bayes' formula we have

$$\mathbb{E}(f(X_t)|\mathcal{G}_t) = \frac{\tilde{\mathbb{E}}(f(X_t)Z_t|\mathcal{G}_t)}{\tilde{\mathbb{E}}(Z_t|\mathcal{G}_t)}$$

and we can describe $\tilde{\mathbb{E}}(f(X_t)Z_t|\mathcal{G}_t)$ through the solution of the Zakai equation: Let $\mathcal{L}_t$ be the differential operator associated to X and let $\mathcal{L}_t^*$ be its L^2-adjoint. Consider the SPDE

$$\begin{aligned} du(t,x) &= \mathcal{L}^* u(t,x)dt + h(t,x)^* u(t,x) dY_t \\ &= \mathcal{L}^* u(t,x)dt + h(t,x)^* u(t,x) h(t,X_t)dt + h(t,x)^* u(t,x) dB_t \\ u(0,x) &= p_0(x) \end{aligned} \tag{22}$$

where p_0 is the density of the initial distribution X_0. This equation describes the evolution of the unnormalized conditional density u: For any $f \in L^\infty$,

$$\tilde{\mathbb{E}}(f(X_t)Z_t|\mathcal{G}_t) = \langle f, u(t,\cdot)\rangle$$

where $\langle \cdot, \cdot \rangle$ is the L^2 inner product. This result is shown in Pardoux (1979), Corollary 3.2 of section II. In particular

$$\mathbb{E}(f(X_t)|\mathcal{G}_t) = \frac{\langle f, u(t, \cdot)\rangle}{\langle 1, u(t, \cdot)\rangle}$$

Therefore (22) is somewhat like a conditional Kolmogorov forward equation (or conditional Fokker-Planck equation) - although the actual conditional Fokker-Planck equation would be the Kushner equation, which we do not treat here. We are able to reverse time and obtain a backward equation for which we can find a stochastic representation in terms of a BDSDE. The forward process for this BDSDE will then not be X, but the diffusion associated to the differential operator part of $\mathcal{L}_t^*$. However it turns out that there is also something like a conditional Kolmogorov backward equation, and that it is maybe more natural to find a BDSDE representation for this backward equation: For $f \in L^\infty(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ consider the backward SPDE

$$\begin{aligned} dv(t,x) &= -\mathcal{L}_t v(t,x)dt - h(t,x)^* v(t,x) d\overleftarrow{Y}_t \\ &= -\mathcal{L}_t v(t,x)dt - h(t,x)^* v(t,x) h(t, X_t)dt - h(t,x)^* v(t,x) d\overleftarrow{B}_t \\ v(T,x) &= f(x) \end{aligned}$$

The solution v of this equation satisfies for any (t, x) (cf. Pardoux (1979), Theorem 2.1 of section II)

$$v(t,x) = \tilde{\mathbb{E}}(f(X_T^{t,x}) Z_{t,T} | \mathcal{G}_{t,T})$$

where as previously $X^{t,x}$ is the version of X that starts in x at time t, $\mathcal{G}_{t,T}$ is the future σ-algebra of Y after t, and

$$Z_{t,T} = \exp\left(\int_t^T h(s, X_s)^* dY_s - \frac{1}{2} \int_t^T |h(s, X_s)|^2 ds \right)$$

Also, it is adjoint to u in the following sense:

Theorem 9. *Almost all trajectories of $(\langle u(t, \cdot), v(t, \cdot)\rangle : 0 \le t \le T)$ are constant.*

This is Theorem 3.1 of section II in Pardoux (1979).

Again, of course the "real" conditional Kolmogorov backward equation will be the adjoint of the Kushner equation.

In a sense, the expression of v and u as conditional expectations is already a stochastic representation of the solutions of the SPDEs, and it is strongly reminiscent of the Kolmogorov equations. But let us also represent v as solution of a BDSDE. This is natural because it is precisely the type of equation that we treated in the last section. We have two options to do so: under the measure $\mathbb{P}$, the drift term $h(t,x)^* v(t,x) h(t, X_t)dt$ is stochastic. This was not permitted in the representation in Pardoux and Peng (1994), but we could slightly generalize their results to include our case. For this it is necessary that X_t is $\mathcal{F}_{t,T}^B \vee \mathcal{F}_t^W$-measurable. But this is of course the case. Assume all the conditions from the previous section are satisfied. Then we obtain the BDSDE

$$\begin{aligned}
-dR_s^{t,x} &= h(s, X_s^{t,x})^* R_s^{t,x} h(s, X_s) ds + h(s, X_s^{t,x})^* R_s^{t,x} d\overleftarrow{B}_s - Z_s^{t,x} dW_s \\
R_T^{t,x} &= f(X_T^{t,x})
\end{aligned}$$

where X is as previously, starting at time 0 with density p_0, and $X^{t,x}$ starts at time t in x. The "filtration" for which we search an adapted solution is $\mathcal{F}_{s,T}^B \vee \mathcal{F}_s^W$.

The simpler and maybe more natural way to find a representation is to work under the equivalent probability $\tilde{\mathbb{P}}$ for which Y is a Brownian motion. Since then we have $\tilde{\mathbb{P}}$-a.s. $v(s, X_s^{t,x}) = R_s^{t,x}$ for the solution R of the BDSDE, this equality will also hold $\mathbb{P}$-a.s. Under $\tilde{\mathbb{P}}$ we have two choices for the second Brownian motion: we can choose W, which is a standard Brownian motion that is independent of Y under $\tilde{\mathbb{P}}$. But we can also choose a completely new Brownian motion $\tilde{W}$, independent of W and Y, and define $\tilde{X}$ as solution of (8) driven by $\tilde{W}$. In actual calculations, the second option might be more convenient. Consider the BDSDE

$$\begin{aligned}
-dR_s^{t,x} &= h(s, \tilde{X}_s^{t,x})^* R_s^{t,x} d\overleftarrow{Y}_s - Z_s^{t,x} d\tilde{W}_s \\
R_T^{t,x} &= f(\tilde{X}_T^{t,x})
\end{aligned}$$

under the measure $\tilde{\mathbb{P}}$ and with the "filtration" $\mathcal{G}_{s,T} \vee \mathcal{F}_{t,s}^{\tilde{W}}$, $s \geq t$. Under the right assumptions on h and f, there is a unique solution $R^{t,x}$, that additionally satisfies $\tilde{\mathbb{P}}$-a.s. and thus $\mathbb{P}$-a.s.

$$R_s^{t,x} = v(s, \tilde{X}_s^{t,x})$$

Here it does not matter whether $\tilde{W} = W$ or whether $\tilde{W}$ is independent of W.

Bibliographic Notes Most of the material is from Pardoux and Peng (1994), where BDSDEs were first introduced. The material on filtering is from Pardoux (1979).

References

P. Cheridito, H. M. Soner, N. Touzi, and N. Victoir. Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs. *Communications on Pure and Applied Mathematics*, 60(7):1081–1110, July 2007. ISSN 00103640. doi: 10.1002/cpa.20168. URL http://doi.wiley.com/10.1002/cpa.20168.

E. Cinlar, J. Jacod, P. Protter, and M. J. Sharpe. Semimartingales and Markov Processes. *Probability Theory and Related Fields*, 54(2):161219, 1980. URL http://www.springerlink.com/index/T5447V78377606U4.pdf.

N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Mathematical Finance*, 7(1):1–71, Jan. 1997. ISSN 0960-1627. doi: 10.1111/1467-9965.00022. URL http://www.blackwell-synergy.com/links/doi/10.1111/1467-9965.00022.

N. El Karoui, S. Hamadène, and A. Matoussi. *Backward stochastic differential equations and applications*, volume 27, pages 267–320. Springer, 2008. URL http://www.springerlink.com/index/MH434U18150M2530.pdf.

L. C. Evans. An Introduction to Mathematical Optimal Control Theory. URL http://math.berkeley.edu/~{}evans/control.course.pdf.

I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*. Springer Verlag, 1988. URL http://books.google.com/books?hl=en&lr=&id=eYQeKZ7aC2AC&oi=fnd&pg=PR7&dq=Brownian+motion+and+stochastic+calculus&ots=V8MsgyVj3J&sig=tlYX6ZMWPR1c1hOe_6bKG_HF-3I.

E. Pardoux. Stochastic partial differential equations and filtering of diffusion processes. *Stochastics An International Journal of Probability and Stochastic Processes*, 3:127–167, 1979. ISSN 1744-2508. doi: 10.1080/17442507908833142. URL http://www.informaworld.com/openurl?genre=article&doi=10.1080/17442507908833142&magic=crossref||D404A21C5BB053405B1A640AFFD44AE3.

E. Pardoux and S. Peng. *Backward Stochastic Differential Equations and Quasilinear Parabolic Stochastic Differential Equations*, volume 176, pages 200–217. Springer Verlag, Aug. 1992. doi: 10.1016/S0304-4149(98)00030-1.

E. Pardoux and S. Peng. Backward Doubly Stochastic Differential Equations and Systems of Quasi-linear SPDEs. *Probab. Theory Relat. Fields*, 98:209–227, Sept. 1994. ISSN 1558-2639. doi: 10.1109/TNB.2010.2070805. URL http://www.ncbi.nlm.nih.gov/pubmed/20840904.

E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Systems & Control Letters*, 14:55–61, 1990.

H. Pham. *Continuous-time Stochastic Control and Optimization with Financial Applications*. Springer Verlag, 2009. ISBN 3540894993. URL http://books.google.com/books?hl=en&lr=&id=xBsagiBp1SYC&oi=fnd&pg=PR8&dq=Continuous-time+Stochastic+Control+and+Optimization+with+Financial+Applications&ots=lCcOmCdUE1&sig=kf6XIDQZJMgPT-gCvrPzfDRxYEE.

D. Revuz and M. Yor. *Continuous martingales and Brownian motion*. Springer Verlag, 3rd edition, 1999. URL http://books.google.com/books?hl=en&lr=&id=1ml95FLM5koC&oi=fnd&pg=PA1&dq=Continuous+martingales+and+Brownian+motion&ots=-6imnfBOGw&sig=FB9xdfGdsxZEMgXRsEXlIiOnrGc.

J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*. Springer Verlag, 1999.