MF921 Topics in Dynamic Asset Pricing

Week 9

Yuanhui Zhao

Boston University

Chapter 12 Backward Stochastic Differential Equations

Motivation and Definition

Consider the following question: Find a random variable Y_0 and a progressively measurable process $Z_t \in \mathbb{R}^d$, such that

$$-dY_t = f(t, Y_t, Z_t)dt - Z_t^T dW_t, \quad Y_T = \xi,$$

where T means transpose, ξ is a constant ($\mathcal{F}_T$ measurable r.v?), and W_t is a standard d-dimensional Brownian motion. This is an example of backward stochastic differential equations (BSDE), which can be viewed as a hedging problem to match the final payoff $Y_T = \xi$ by finding the initial price Y_0 and the hedging strategy Z_t.

In addition to the link with option pricing, there are at least four applications.
(1) BSDE is linked to recursive utilities.
(2) BSDE has been used to study continuous-time principle-agent problems, starting from Sannikov (2008, Review of Economic Studies).
(3) There is a link between BSDE and certain classes of semi-linear parabolic PDE's, as an extension of the Feynman-Kac formula.

Motivation and Definition

Consider the following question: Find a random variable Y_0 and a progressively measurable process $Z_t \in \mathbb{R}^d$, such that

$$-dY_t = f(t, Y_t, Z_t)dt - Z_t^T dW_t, \quad Y_T = \xi,$$

where T means transpose, ξ is a constant ($\mathcal{F}_T$ measurable r.v?), and W_t is a standard d-dimensional Brownian motion. This is an example of backward stochastic differential equations (BSDE), which can be viewed as a hedging problem to match the final payoff $Y_T = \xi$ by finding the initial price Y_0 and the hedging strategy Z_t.

In addition to the link with option pricing, there are at least four applications.
(1) BSDE is linked to recursive utilities.
(2) BSDE has been used to study continuous-time principle-agent problems, starting from Sannikov (2008, Review of Economic Studies).
(3) There is a link between BSDE and certain classes of semi-linear parabolic PDE's, as an extension of the Feynman-Kac formula.

(4) There is a natural link between BSDE's and stochastic control problems. For example, the above BSDE problem can be formulated as a special stochastic control problem such that

$$\min_{y,Z} E\left[\left\{Y_T^{y,Z} - \xi\right\}^2\right] = 0,$$

where $Y_T^{y,Z}$ is the solution of

$$-dY_t = f(t, Y_t, Z_t)dt - Z_t^T dW_t, \quad Y_0 = y.$$

This minimization problem can be solved by using a neural network, by learning $Y_T^{y,Z}$ to match ξ as close as possible. This connection leads to a fast way to solve some semi-linear PDE's, especially in the high-dimensional case, by using neural networks.

As a comparison, the HJB equation is a continuous analogy of dynamic programming and is in general a nonlinear parabolic PDE, which is challenging to solve numerically, especially in high dimensions. However, for certain special stochastic control problem, the HJB may become a semi-linear parabolic PDE, which can be solved by using neural networks via backward stochastic differential equations (BSDEs). In general, instead of using BSDEs and PDEs, one can use iterated procedure to build a neural network to solve stochastic control problems.

Motivation and Definition

(4) There is a natural link between BSDE's and stochastic control problems. For example, the above BSDE problem can be formulated as a special stochastic control problem such that

$$\min_{y,Z} E\left[\left\{Y_T^{y,Z} - \xi\right\}^2\right] = 0,$$

where $Y_T^{y,Z}$ is the solution of

$$-dY_t = f(t, Y_t, Z_t)dt - Z_t^T dW_t, \quad Y_0 = y.$$

This minimization problem can be solved by using a neural network, by learning $Y_T^{y,Z}$ to match ξ as close as possible. This connection leads to a fast way to solve some semi-linear PDE's, especially in the high-dimensional case, by using neural networks.

As a comparison, the HJB equation is a continuous analogy of dynamic programming and is in general a nonlinear parabolic PDE, which is challenging to solve numerically, especially in high dimensions. However, for certain special stochastic control problem, the HJB may become a semi-linear parabolic PDE, which can be solved by using neural networks via backward stochastic differential equations (BSDEs). In general, instead of using BSDEs and PDEs, one can use iterated procedure to build a neural network to solve stochastic control problems.

Motivation and Definition

(4) There is a natural link between BSDE's and stochastic control problems. For example, the above BSDE problem can be formulated as a special stochastic control problem such that

$$\min_{y,Z} E\left[\left\{Y_T^{y,Z} - \xi\right\}^2\right] = 0,$$

where $Y_T^{y,Z}$ is the solution of

$$-dY_t = f(t, Y_t, Z_t)dt - Z_t^T dW_t, \quad Y_0 = y.$$

This minimization problem can be solved by using a neural network, by learning $Y_T^{y,Z}$ to match ξ as close as possible. This connection leads to a fast way to solve some semi-linear PDE's, especially in the high-dimensional case, by using neural networks.

As a comparison, the HJB equation is a continuous analogy of dynamic programming and is in general a nonlinear parabolic PDE, which is challenging to solve numerically, especially in high dimensions. However, for certain special stochastic control problem, the HJB may become a semi-linear parabolic PDE, which can be solved by using neural networks via backward stochastic differential equations (BSDEs). In general, instead of using BSDEs and PDEs, one can use iterated procedure to build a neural network to solve stochastic control problems.

Motivation and Definition

A formal definition of one-dimensional BSDE for a given pair (ξ, f) satisfying the regularity conditions.
For the pair (ξ, f) we require:

- (i) $\xi \in \mathcal{F}_T$ is a L^2 random variable.
- (ii) $f(\cdot, t, y, z) : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$, denoted for simplicity as $f(t, y, z)$, is progressively measurable for all y and z, such that $E[\int_0^T f^2(t, 0, 0)dt] < \infty$.

A solution to the BSDE at the beginning is a pair (Y, Z), both progressively measurable, such that $E[\sup_{0 \le t \le T} |Y_t|^2] < \infty$, $E\left[\int_0^T |Z_t|^2 \, dt\right] < \infty$, and

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T.$$

Here ξ and f are called the terminal condition and the driver of the BSDE, respectively. Given $\xi \in L^2$ and f satisfying a uniform Lipschitz condition, i.e. there exists a constant C_f such that

$$|f(t, x_1, y_1) - f(t, x_2, y_2)| \le C_f \left(|x_1 - x_2| + |y_1 - y_2|\right),$$

there exists a unique solution (Y, Z) to the BSDE. Unfortunately, the uniform Lipschitz condition does not hold in many cases.

Motivation and Definition

A formal definition of one-dimensional BSDE for a given pair (ξ, f) satisfying the regularity conditions.
For the pair (ξ, f) we require:

- (i) $\xi \in \mathcal{F}_T$ is a L^2 random variable.
- (ii) $f(\cdot, t, y, z) : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$, denoted for simplicity as $f(t, y, z)$, is progressively measurable for all y and z, such that $E[\int_0^T f^2(t, 0, 0)dt] < \infty$.

A solution to the BSDE at the beginning is a pair (Y, Z), both progressively measurable, such that $E[\sup_{0 \le t \le T} |Y_t|^2] < \infty$, $E\left[\int_0^T |Z_t|^2 \, dt\right] < \infty$, and

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T.$$

Here ξ and f are called the terminal condition and the driver of the BSDE, respectively.
Given $\xi \in L^2$ and f satisfying a uniform Lipschitz condition, i.e. there exists a constant C_f such that

$$|f(t, x_1, y_1) - f(t, x_2, y_2)| \le C_f \left(|x_1 - x_2| + |y_1 - y_2|\right),$$

there exists a unique solution (Y, Z) to the BSDE. Unfortunately, the uniform Lipschitz condition does not hold in many cases.

Motivation and Definition

A formal definition of one-dimensional BSDE for a given pair (ξ, f) satisfying the regularity conditions.
For the pair (ξ, f) we require:

- (i) $\xi \in \mathcal{F}_T$ is a L^2 random variable.
- (ii) $f(\cdot, t, y, z) : \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$, denoted for simplicity as $f(t, y, z)$, is progressively measurable for all y and z, such that $E[\int_0^T f^2(t, 0, 0)dt] < \infty$.

A solution to the BSDE at the beginning is a pair (Y, Z), both progressively measurable, such that $E[\sup_{0 \le t \le T} |Y_t|^2] < \infty$, $E\left[\int_0^T |Z_t|^2 \, dt\right] < \infty$, and

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T.$$

Here ξ and f are called the terminal condition and the driver of the BSDE, respectively. Given $\xi \in L^2$ and f satisfying a uniform Lipschitz condition, i.e. there exists a constant C_f such that

$$|f(t, x_1, y_1) - f(t, x_2, y_2)| \le C_f \left(|x_1 - x_2| + |y_1 - y_2|\right),$$

there exists a unique solution (Y, Z) to the BSDE. Unfortunately, the uniform Lipschitz condition does not hold in many cases.

Linear BSDE

First consider a special case $f = 0$, in which the BSDE becomes

$$Y_t = \xi - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T. \tag{*}$$

Recall that the martingale representation theorem yields for every $\mathcal{F}_T$ measurable and square integrable random variable ξ, there exists a unique progressively measurable process β_t, $E\left[\int_0^T |\beta_t|^2 dt\right] < \infty$, such that

$$\xi = E[\xi] + \int_0^T \beta_s^T dW_s.$$

We shall prove that the unique solution of the BSDE $(*)$ is given by

$$Y_t = E[\xi|\mathcal{F}_t], \quad Z_t = \beta_t.$$

Linear BSDE

First consider a special case $f = 0$, in which the BSDE becomes

$$Y_t = \xi - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T. \tag{*}$$

Recall that the martingale representation theorem yields for every $\mathcal{F}_T$ measurable and square integrable random variable ξ, there exists a unique progressively measurable process β_t, $E\left[\int_0^T |\beta_t|^2 dt\right] < \infty$, such that

$$\xi = E[\xi] + \int_0^T \beta_s^T dW_s.$$

We shall prove that the unique solution of the BSDE $(*)$ is given by

$$Y_t = E[\xi|\mathcal{F}_t], \quad Z_t = \beta_t.$$

Linear BSDE

First consider a special case $f = 0$, in which the BSDE becomes

$$Y_t = \xi - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T. \tag{*}$$

Recall that the martingale representation theorem yields for every $\mathcal{F}_T$ measurable and square integrable random variable ξ, there exists a unique progressively measurable process β_t, $E\left[\int_0^T |\beta_t|^2 dt\right] < \infty$, such that

$$\xi = E[\xi] + \int_0^T \beta_s^T dW_s.$$

We shall prove that the unique solution of the BSDE $(*)$ is given by

$$Y_t = E[\xi|\mathcal{F}_t], \quad Z_t = \beta_t.$$

Linear BSDE

First note that by

$$\begin{aligned}
Y_t &= E[\xi|\mathcal{F}_t] \\
&= E[\xi] + E\left[\int_0^T \beta_s^T dW_s|\mathcal{F}_t\right] \\
&= \xi - \int_0^T Z_s^T dW_s + \int_0^t Z_s^T dW_s + E\left[\int_t^T Z_s^T dW_s|\mathcal{F}_t\right] \\
&= \xi - \int_t^T Z_s^T dW_s + E\left[\int_t^T Z_s^T dW_s|\mathcal{F}_t\right] \\
&= \xi - \int_t^T Z_s^T dW_s,
\end{aligned}$$

because $E\left[\int_t^T Z_s^T dW_s|\mathcal{F}_t\right] = 0$ as the local martingale of the stochastic integral becomes a martingale, as will be shown in the following problem.

Linear BSDE

First note that by

$$\begin{aligned} Y_t &= E[\xi|\mathcal{F}_t] \\ &= E[\xi] + E\left[\int_0^T \beta_s^T dW_s|\mathcal{F}_t\right] \\ &= \xi - \int_0^T Z_s^T dW_s + \int_0^t Z_s^T dW_s + E\left[\int_t^T Z_s^T dW_s|\mathcal{F}_t\right] \\ &= \xi - \int_t^T Z_s^T dW_s + E\left[\int_t^T Z_s^T dW_s|\mathcal{F}_t\right] \\ &= \xi - \int_t^T Z_s^T dW_s, \end{aligned}$$

because $E\left[\int_t^T Z_s^T dW_s|\mathcal{F}_t\right] = 0$ as the local martingale of the stochastic integral becomes a martingale, as will be shown in the following problem.

Linear BSDE

Prove that the local martingale $M_t = \int_0^t Z_s^T dW_s$ is actually a martingale.

We know that $E\left[\int_0^T |Z_t|^2 dt\right] < \infty$ and the quadratic variation of $\int_0^t Z_s^T dW_s$ is $\int_0^T |Z_t|^2 dt$. Thus, by the Burkholder-Davis-Gundy inequality for the quadratic variation of a martingale, with $p = 2$, there exists a positive constant C_2 such that

$$E\left[\left(\sup_{0\le t\le T}\left|\int_0^t Z_s^T dW_s\right|\right)^2\right] \le C_p E\left[\left(\int_0^T |Z_t|^2 dt\right)^{p/2}\right] = C_2 E\left[\int_0^T |Z_t|^2 dt\right] < \infty.$$

Define stopping times $\tau_n := \inf\{t : \langle M\rangle_t \ge n\} \wedge T$. Each stopped process $M_t^{\tau_n} := M_{t\wedge\tau_n}$ is a bounded L^2-martingale. Because of BDG, the family $\{M_T^{\tau_n}\}_n$ is uniformly integrable. Hence we can pass to the limit $n \to \infty$ in the martingale property:

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = \lim_{n\to\infty} \mathbb{E}[M_t^{\tau_n} \mid \mathcal{F}_s] = \lim_{n\to\infty} M_s^{\tau_n} = M_s, \quad 0 \le s \le t \le T,$$

where we used DCT, justified by the UI bound above. Therefore M is a martingale.

Linear BSDE

Prove that the local martingale $M_t = \int_0^t Z_s^T dW_s$ is actually a martingale.

We know that $E\left[\int_0^T |Z_t|^2 dt\right] < \infty$ and the quadratic variation of $\int_0^t Z_s^T dW_s$ is $\int_0^T |Z_t|^2 dt$. Thus, by the Burkholder-Davis-Gundy inequality for the quadratic variation of a martingale, with $p = 2$, there exists a positive constant C_2 such that

$$E\left[\left(\sup_{0\le t\le T}\left|\int_0^t Z_s^T dW_s\right|\right)^2\right] \le C_p E\left[\left(\int_0^T |Z_t|^2 dt\right)^{p/2}\right] = C_2 E\left[\int_0^T |Z_t|^2 dt\right] < \infty.$$

Define stopping times $\tau_n := \inf\{t : \langle M\rangle_t \ge n\} \wedge T$. Each stopped process $M_t^{\tau_n} := M_{t\wedge\tau_n}$ is a bounded L^2-martingale. Because of BDG, the family $\{M_T^{\tau_n}\}_n$ is uniformly integrable. Hence we can pass to the limit $n \to \infty$ in the martingale property:

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = \lim_{n\to\infty} \mathbb{E}[M_t^{\tau_n} \mid \mathcal{F}_s] = \lim_{n\to\infty} M_s^{\tau_n} = M_s, \quad 0 \le s \le t \le T,$$

where we used DCT, justified by the UI bound above. Therefore M is a martingale.

Linear BSDE

Prove that the local martingale $M_t = \int_0^t Z_s^T dW_s$ is actually a martingale.

We know that $E\left[\int_0^T |Z_t|^2 dt\right] < \infty$ and the quadratic variation of $\int_0^t Z_s^T dW_s$ is $\int_0^T |Z_t|^2 dt$. Thus, by the Burkholder-Davis-Gundy inequality for the quadratic variation of a martingale, with $p = 2$, there exists a positive constant C_2 such that

$$E\left[\left(\sup_{0\le t\le T}\left|\int_0^t Z_s^T dW_s\right|\right)^2\right] \le C_p E\left[\left(\int_0^T |Z_t|^2 dt\right)^{p/2}\right] = C_2 E\left[\int_0^T |Z_t|^2 dt\right] < \infty.$$

Define stopping times $\tau_n := \inf\{t : \langle M\rangle_t \ge n\} \wedge T$. Each stopped process $M_t^{\tau_n} := M_{t\wedge\tau_n}$ is a bounded L^2-martingale. Because of BDG, the family $\{M_T^{\tau_n}\}_n$ is uniformly integrable. Hence we can pass to the limit $n \to \infty$ in the martingale property:

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = \lim_{n\to\infty} \mathbb{E}[M_t^{\tau_n} \mid \mathcal{F}_s] = \lim_{n\to\infty} M_s^{\tau_n} = M_s, \quad 0 \le s \le t \le T,$$

where we used DCT, justified by the UI bound above. Therefore M is a martingale.

Linear BSDE

To show the uniqueness of the BSDE $(*)$, without using the general theorem of the BSDE based on the uniform Lipschitz condition, consider a solution pair (Y_t, Z_t). Note that when $t = 0$

$$Y_0 = \xi - \int_0^T Z_s^T dW_s,$$

yielding

$$Y_0 = E[\xi] - E\left[\int_0^T Z_s^T dW_s\right] = E[\xi],$$

because $\int_0^T Z_s^T dW_s$ has a martingale property. Furthermore, by $(*)$,

$$dY_t = Z_t^T dW_t, \quad Y_t = Y_0 + \int_0^t Z_s^T dW_s = E[\xi] + \int_0^t Z_s^T dW_s. \qquad (**)$$

Thus,

$$\xi = Y_t + \int_t^T Z_s^T dW_s = E[\xi] + \int_0^t Z_s^T dW_s + \int_0^T Z_s^T dW_s = E[\xi] + \int_0^T Z_s^T dW_s.$$

Such Z must be unique according to the martingale representation theorem. Hence (Y_t, Z_t) must be unique due to $(**)$.

Linear BSDE

To show the uniqueness of the BSDE $(*)$, without using the general theorem of the BSDE based on the uniform Lipschitz condition, consider a solution pair (Y_t, Z_t). Note that when $t = 0$

$$Y_0 = \xi - \int_0^T Z_s^T dW_s,$$

yielding

$$Y_0 = E[\xi] - E\left[\int_0^T Z_s^T dW_s\right] = E[\xi],$$

because $\int_0^T Z_s^T dW_s$ has a martingale property. Furthermore, by $(*)$,

$$dY_t = Z_t^T dW_t, \quad Y_t = Y_0 + \int_0^t Z_s^T dW_s = E[\xi] + \int_0^t Z_s^T dW_s. \qquad \text{(**)}$$

Thus,

$$\xi = Y_t + \int_t^T Z_s^T dW_s = E[\xi] + \int_0^t Z_s^T dW_s + \int_0^T Z_s^T dW_s = E[\xi] + \int_0^T Z_s^T dW_s.$$

Such Z must be unique according to the martingale representation theorem. Hence (Y_t, Z_t) must be unique due to $(**)$.

Linear BSDE

To show the uniqueness of the BSDE $(*)$, without using the general theorem of the BSDE based on the uniform Lipschitz condition, consider a solution pair (Y_t, Z_t). Note that when $t = 0$

$$Y_0 = \xi - \int_0^T Z_s^T dW_s,$$

yielding

$$Y_0 = E[\xi] - E\left[\int_0^T Z_s^T dW_s\right] = E[\xi],$$

because $\int_0^T Z_s^T dW_s$ has a martingale property. Furthermore, by $(*)$,

$$dY_t = Z_t^T dW_t, \quad Y_t = Y_0 + \int_0^t Z_s^T dW_s = E[\xi] + \int_0^t Z_s^T dW_s. \qquad \text{(**)}$$

Thus,

$$\xi = Y_t + \int_t^T Z_s^T dW_s = E[\xi] + \int_0^t Z_s^T dW_s + \int_0^T Z_s^T dW_s = E[\xi] + \int_0^T Z_s^T dW_s.$$

Such Z must be unique according to the martingale representation theorem. Hence (Y_t, Z_t) must be unique due to $(**)$.

The solution for the linear BSDE via the Girsanov theorem

Now we consider BSDE with a linear driver, i.e.,

$$f(t, y, z) = A_t + B_t y + C_t^\top z,$$

where A_t, B_t are one-dimensional progressive measurable processes and C_t is a d-dimensional progressive measurable process, such that

$$E\left[\int_0^T |A_t|^2 \, dt\right] < \infty,$$

and B_t and C_t are bounded processes. We attempt to reduce this case to the case of zero driver, via the Girsanov theorem and other transforms.

To eliminate the term $C_t^\top z$ in the driver, we consider a new probability measure Q, defined as

$$\frac{dQ}{dP} = \exp\left\{\int_0^T C_t^\top dW_t - \frac{1}{2}\int_0^T |C_t|^2 \, dt\right\},$$

and a new process

$$\bar{W}_t = W_t - \int_0^t C_s ds.$$

The solution for the linear BSDE via the Girsanov theorem

Now we consider BSDE with a linear driver, i.e.,

$$f(t, y, z) = A_t + B_t y + C_t^\top z,$$

where A_t, B_t are one-dimensional progressive measurable processes and C_t is a d-dimensional progressive measurable process, such that

$$E\left[\int_0^T |A_t|^2 \, dt\right] < \infty,$$

and B_t and C_t are bounded processes. We attempt to reduce this case to the case of zero driver, via the Girsanov theorem and other transforms.

To eliminate the term $C_t^\top z$ in the driver, we consider a new probability measure Q, defined as

$$\frac{dQ}{dP} = \exp\left\{\int_0^T C_t^\top dW_t - \frac{1}{2}\int_0^T |C_t|^2 \, dt\right\},$$

and a new process

$$\bar{W}_t = W_t - \int_0^t C_s ds.$$

The solution for the linear BSDE via the Girsanov theorem

By Girsanov theorem, $\bar{W}_t$ is a standard Brownian motion under the new measure Q. Note that the Novikov condition is satisfied because

$$E\left[\exp\left(\frac{1}{2}\int_0^T |C_s|^2\, ds\right)\right] < \infty.$$

Using $\bar{W}_t$, the dynamic of Y_t can be re-written as

$$\begin{aligned} -dY_t &= (A_t + B_t Y_t + C_t^\top Z_t)\, dt - Z_t^\top dW_t \\ &= (A_t + B_t Y_t + C_t^\top Z_t)\, dt - Z_t^\top (d\bar{W}_t + C_t dt) \\ &= (A_t + B_t Y_t)\, dt - Z_t^\top d\bar{W}_t. \end{aligned}$$

Next, to eliminate the term $B_t y$ in the driver, we consider the discounted version of Y. More precisely, introduce

$$Y_{D,t} = Y_t \exp\left\{\int_0^t B_s ds\right\}, \quad Z_{D,t} = Z_t \exp\left\{\int_0^t B_s ds\right\}.$$

The solution for the linear BSDE via the Girsanov theorem

By Girsanov theorem, $\bar{W}_t$ is a standard Brownian motion under the new measure Q. Note that the Novikov condition is satisfied because

$$E\left[\exp\left(\frac{1}{2}\int_0^T |C_s|^2\, ds\right)\right] < \infty.$$

Using $\bar{W}_t$, the dynamic of Y_t can be re-written as

$$\begin{aligned}
-dY_t &= (A_t + B_tY_t + C_t^\top Z_t)\, dt - Z_t^\top dW_t \\
&= (A_t + B_tY_t + C_t^\top Z_t)\, dt - Z_t^\top (d\bar{W}_t + C_t dt) \\
&= (A_t + B_tY_t)\, dt - Z_t^\top d\bar{W}_t.
\end{aligned}$$

Next, to eliminate the term $B_t y$ in the driver, we consider the discounted version of Y. More precisely, introduce

$$Y_{D,t} = Y_t \exp\left\{\int_0^t B_s ds\right\}, \quad Z_{D,t} = Z_t \exp\left\{\int_0^t B_s ds\right\}.$$

The solution for the linear BSDE via the Girsanov theorem

By Girsanov theorem, $\bar{W}_t$ is a standard Brownian motion under the new measure Q. Note that the Novikov condition is satisfied because

$$E\left[\exp\left(\frac{1}{2}\int_0^T |C_s|^2\, ds\right)\right] < \infty.$$

Using $\bar{W}_t$, the dynamic of Y_t can be re-written as

$$\begin{aligned}
-dY_t &= (A_t + B_t Y_t + C_t^\top Z_t)\, dt - Z_t^\top dW_t \\
&= (A_t + B_t Y_t + C_t^\top Z_t)\, dt - Z_t^\top (d\bar{W}_t + C_t dt) \\
&= (A_t + B_t Y_t)\, dt - Z_t^\top d\bar{W}_t.
\end{aligned}$$

Next, to eliminate the term $B_t y$ in the driver, we consider the discounted version of Y. More precisely, introduce

$$Y_{D,t} = Y_t \exp\left\{\int_0^t B_s ds\right\}, \quad Z_{D,t} = Z_t \exp\left\{\int_0^t B_s ds\right\}.$$

The solution for the linear BSDE via the Girsanov theorem

Then by the Ito formula

$$
\begin{aligned}
dY_{D,t} &= \exp\left\{\int_0^t B_s ds\right\} dY_t + Y_t \exp\left\{\int_0^t B_s ds\right\} B_t dt \\
&= \exp\left\{\int_0^t B_s ds\right\} \left\{-A_t dt - B_t Y_t dt + Z_t^\top d\bar{W}_t + Y_t B_t dt\right\} \\
&= -A_t \exp\left\{\int_0^t B_s ds\right\} dt + Z_{D,t}^\top d\bar{W}_t,
\end{aligned}
$$

with the terminal condition $Y_{D,T} = \xi \exp\left\{\int_0^T B_s ds\right\}$.

Finally to eliminate A_t, we define

$$
\bar{Y}_t = Y_{D,t} + \int_0^t A_u \exp\left\{\int_0^u B_s ds\right\} du.
$$

Then

$$
d\bar{Y}_t = dY_{D,t} + A_t \exp\left\{\int_0^t B_s ds\right\} dt = Z_{D,t}^\top d\bar{W}_t,
$$

The solution for the linear BSDE via the Girsanov theorem

Then by the Ito formula

$$
\begin{aligned}
dY_{D,t} &= \exp\left\{\int_0^t B_s ds\right\} dY_t + Y_t \exp\left\{\int_0^t B_s ds\right\} B_t dt \\
&= \exp\left\{\int_0^t B_s ds\right\} \left\{-A_t dt - B_t Y_t dt + Z_t^\top d\bar{W}_t + Y_t B_t dt\right\} \\
&= -A_t \exp\left\{\int_0^t B_s ds\right\} dt + Z_{D,t}^\top d\bar{W}_t,
\end{aligned}
$$

with the terminal condition $Y_{D,T} = \xi \exp\left\{\int_0^T B_s ds\right\}$.
Finally to eliminate A_t, we define

$$
\bar{Y}_t = Y_{D,t} + \int_0^t A_u \exp\left\{\int_0^u B_s ds\right\} du.
$$

Then

$$
d\bar{Y}_t = dY_{D,t} + A_t \exp\left\{\int_0^t B_s ds\right\} dt = Z_{D,t}^\top d\bar{W}_t,
$$

The solution for the linear BSDE via the Girsanov theorem

with the terminal condition

$$\begin{aligned}
\bar{Y}_T &= Y_{D,T} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du \\
&= \xi \exp\left\{\int_0^T B_s ds\right\} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du.
\end{aligned}$$

Thus, we can apply the previous result $Y_t = E[\xi|\mathcal{F}_t]$, $Z_t = \beta_t$ about BSDE with zero driver, we have the unique solution is given by

$$\bar{Y}_t = E^Q\left[\xi \exp\left\{\int_0^T B_s ds\right\} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du|\mathcal{F}_t\right],$$

whence

$$\begin{aligned}
Y_{D,t} &= \bar{Y}_t - \int_0^t A_u \exp\left\{\int_0^u B_s ds\right\} du \\
&= E^Q\left[\xi \exp\left\{\int_0^T B_s ds\right\} + \int_t^T A_u \exp\left\{\int_0^u B_s ds\right\} du|\mathcal{F}_t\right].
\end{aligned}$$

with the terminal condition

$$\begin{aligned}\bar{Y}_T &= Y_{D,T} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du \\ &= \xi \exp\left\{\int_0^T B_s ds\right\} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du.\end{aligned}$$

Thus, we can apply the previous result $Y_t = E[\xi|\mathcal{F}_t]$, $Z_t = \beta_t$ about BSDE with zero driver, we have the unique solution is given by

$$\bar{Y}_t = E^Q\left[\xi \exp\left\{\int_0^T B_s ds\right\} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du|\mathcal{F}_t\right],$$

whence

$$\begin{aligned}Y_{D,t} &= \bar{Y}_t - \int_0^t A_u \exp\left\{\int_0^u B_s ds\right\} du \\ &= E^Q\left[\xi \exp\left\{\int_0^T B_s ds\right\} + \int_t^T A_u \exp\left\{\int_0^u B_s ds\right\} du|\mathcal{F}_t\right].\end{aligned}$$

The solution for the linear BSDE via the Girsanov theorem

with the terminal condition

$$\begin{aligned}\bar{Y}_T &= Y_{D,T} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du \\ &= \xi \exp\left\{\int_0^T B_s ds\right\} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du.\end{aligned}$$

Thus, we can apply the previous result $Y_t = E[\xi|\mathcal{F}_t],\ Z_t = \beta_t$ about BSDE with zero driver, we have the unique solution is given by

$$\bar{Y}_t = E^Q\left[\xi \exp\left\{\int_0^T B_s ds\right\} + \int_0^T A_u \exp\left\{\int_0^u B_s ds\right\} du|\mathcal{F}_t\right],$$

whence

$$\begin{aligned}Y_{D,t} &= \bar{Y}_t - \int_0^t A_u \exp\left\{\int_0^u B_s ds\right\} du \\ &= E^Q\left[\xi \exp\left\{\int_0^T B_s ds\right\} + \int_t^T A_u \exp\left\{\int_0^u B_s ds\right\} du|\mathcal{F}_t\right].\end{aligned}$$

The solution for the linear BSDE via the Girsanov theorem

Thus, the unique solution of the linear BSDE is given by

$$
\begin{aligned}
Y_t &= Y_{D,t} \exp\left\{ -\int_0^t B_s ds \right\} \\
&= E^Q \left[\xi \exp\left\{ \int_t^T B_s ds \right\} + \int_t^T A_u \exp\left\{ \int_t^u B_s ds \right\} du | \mathcal{F}_t \right].
\end{aligned}
$$

Let

$$\Psi(t) = \exp\left\{ \int_0^t B_s ds \right\} \frac{dQ}{dP}\bigg|_{\mathcal{F}_t}.$$

Then we can show

$$\frac{d\Psi(t)}{\Psi(t)} = B_t dt + C_t^\top dW_t, \quad \Psi(0) = 1,$$

and the unique solution of the BSDE is given by

$$Y_t = \frac{1}{\Psi(t)} E^P \left[\xi \Psi(T) + \int_t^T A_u \Psi(u) du | \mathcal{F}_t \right].$$

The solution for the linear BSDE via the Girsanov theorem

Thus, the unique solution of the linear BSDE is given by

$$\begin{aligned} Y_t &= Y_{D,t} \exp\left\{-\int_0^t B_s ds\right\} \\ &= E^Q\left[\xi \exp\left\{\int_t^T B_s ds\right\} + \int_t^T A_u \exp\left\{\int_t^u B_s ds\right\} du \middle| \mathcal{F}_t\right]. \end{aligned}$$

Let

$$\Psi(t) = \exp\left\{\int_0^t B_s ds\right\} \frac{dQ}{dP}\bigg|_{\mathcal{F}_t}.$$

Then we can show

$$\frac{d\Psi(t)}{\Psi(t)} = B_t dt + C_t^\top dW_t, \quad \Psi(0) = 1,$$

and the unique solution of the BSDE is given by

$$Y_t = \frac{1}{\Psi(t)} E^P\left[\xi \Psi(T) + \int_t^T A_u \Psi(u) du \middle| \mathcal{F}_t\right].$$

Connection with semi-linear parabolic PDE's

It can be shown that the above BSDE is linked to a semi-linear parabolic PDE

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}(\Delta_x u)(t,x) + f(t, u(t,x), (\nabla_x u)(t,x)) = 0,\ u(T,x) = g(x),$$

where $\Delta_x u$ is the Laplace operator and $\nabla_x u$ is the gradient, i.e.

$$\Delta_x u = \sum_{i=1}^{d} \frac{\partial^2 u}{\partial x_i^2},\ \nabla_x u = \left(\frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_d}\right).$$

This equation is semi-linear because the second derivative is linear despite that f on the first derivative is nonlinear. The link can be stated in two parts.

(1) If the semi-linear PDE has a C^2 solution, then subject to some regularity conditions on u, such as a linear growth condition, the solution of BSDE

$$Y_t = g(W_T) + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s^T dW_s,\ 0 \le t \le T.$$

can be given in terms of u as

$$Y_t = u(t, W_t),\ Z_t = (\nabla_x u)(t, W_t).$$

Connection with semi-linear parabolic PDE's

It can be shown that the above BSDE is linked to a semi-linear parabolic PDE

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}(\Delta_x u)(t,x) + f(t, u(t,x), (\nabla_x u)(t,x)) = 0,\ u(T,x) = g(x),$$

where $\Delta_x u$ is the Laplace operator and $\nabla_x u$ is the gradient, i.e.

$$\Delta_x u = \sum_{i=1}^{d} \frac{\partial^2 u}{\partial x_i^2},\ \nabla_x u = \left(\frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_d}\right).$$

This equation is semi-linear because the second derivative is linear despite that f on the first derivative is nonlinear. The link can be stated in two parts.

(1) If the semi-linear PDE has a C^2 solution, then subject to some regularity conditions on u, such as a linear growth condition, the solution of BSDE

$$Y_t = g(W_T) + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s^T dW_s,\ 0 \le t \le T.$$

can be given in terms of u as

$$Y_t = u(t, W_t),\ Z_t = (\nabla_x u)(t, W_t).$$

Connection with semi-linear parabolic PDE's

This is essentially from the Itô formula, because

$$u(T, W_T) = g(W_T),$$

and

$$\begin{aligned} du(t, W_t) &= \frac{\partial u}{\partial t} dt + \frac{1}{2}(\Delta_x u)(t, W_t) dt + (\nabla_x u)(t, W_t) dW_t \\ &= -f(t, u(t, W_t), (\nabla_x u)(t, W_t)) + (\nabla_x u)(t, W_t) dW_t \\ &= -f(t, u(t, W_t), Z_t) + Z_t dW_t, \end{aligned}$$

which means that with $Y_t = u(t, W_t)$ we have

$$Y_t = g(W_T) + \int_t^T f(s, Y_s, Z_s) ds - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T.$$

(2) Conversely, suppose the BSDE has a solution (Y_t, Z_t), where $Y_t = u(t, W_t)$ for a continuous function u. Then u is a viscosity solution to the semi-linear PDE.

Given this link with BSDE, one can use the neural network to solve the semi-linear parabolic PDE to get $u(0, \xi)$, because $u(0, \xi) = Y_0$.

Connection with semi-linear parabolic PDE's

This is essentially from the Itô formula, because

$$u(T, W_T) = g(W_T),$$

and

$$\begin{aligned} du(t, W_t) &= \frac{\partial u}{\partial t} dt + \frac{1}{2} (\Delta_x u)(t, W_t) dt + (\nabla_x u)(t, W_t) dW_t \\ &= -f(t, u(t, W_t), (\nabla_x u)(t, W_t)) + (\nabla_x u)(t, W_t) dW_t \\ &= -f(t, u(t, W_t), Z_t) + Z_t dW_t, \end{aligned}$$

which means that with $Y_t = u(t, W_t)$ we have

$$Y_t = g(W_T) + \int_t^T f(s, Y_s, Z_s) ds - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T.$$

(2) Conversely, suppose the BSDE has a solution (Y_t, Z_t), where $Y_t = u(t, W_t)$ for a continuous function u. Then u is a viscosity solution to the semi-linear PDE.

Given this link with BSDE, one can use the neural network to solve the semi-linear parabolic PDE to get $u(0, \xi)$, because $u(0, \xi) = Y_0$.

Connection with semi-linear parabolic PDE's

This is essentially from the Itô formula, because

$$u(T, W_T) = g(W_T),$$

and

$$\begin{aligned}
du(t, W_t) &= \frac{\partial u}{\partial t}dt + \frac{1}{2}(\Delta_x u)(t, W_t)dt + (\nabla_x u)(t, W_t)dW_t \\
&= -f(t, u(t, W_t), (\nabla_x u)(t, W_t)) + (\nabla_x u)(t, W_t)dW_t \\
&= -f(t, u(t, W_t), Z_t) + Z_t dW_t,
\end{aligned}$$

which means that with $Y_t = u(t, W_t)$ we have

$$Y_t = g(W_T) + \int_t^T f(s, Y_s, Z_s)ds - \int_t^T Z_s^\top dW_s, \quad 0 \le t \le T.$$

(2) Conversely, suppose the BSDE has a solution (Y_t, Z_t), where $Y_t = u(t, W_t)$ for a continuous function u. Then u is a viscosity solution to the semi-linear PDE.

Given this link with BSDE, one can use the neural network to solve the semi-linear parabolic PDE to get $u(0, \xi)$, because $u(0, \xi) = Y_0$.

An example

Let

$$f(t,u,x_1,...,x_d)=\beta\sum_{i=1}^d x_i^2.$$

Then the semi-linear parabolic PDE becomes

$$\frac{\partial u}{\partial t}(t,x)+\frac{1}{2}(\Delta_x u)(t,x)+\beta\sum_{i=1}^d\left(\frac{\partial u}{\partial x_i}\right)^2=0,\quad u(T,x)=g(x).$$

The BSDE becomes

$$dY(t)=-\beta\sum_{i=1}^d Z_i^2(t)dt+\sum_{i=1}^d Z_i(t)dW_i(t),\quad Y(T)=g(x+W(T)).$$

Rewriting the BSDE, we get a related stochastic control problem

$$\min_{y,Z} E\left[\left\{Y_T^{y,Z}-g(x+W_T)\right\}^2\right]=0,$$

where $Y_T^{y,Z}$ is the solution of

$$dY(t)=-\beta\sum_{i=1}^d Z_i^2(t)dt+\sum_{i=1}^d Z_i(t)dW_i(t),\quad Y(0)=y.$$

An example

Let

$$f(t, u, x_1, ..., x_d) = \beta \sum_{i=1}^{d} x_i^2.$$

Then the semi-linear parabolic PDE becomes

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}(\Delta_x u)(t,x) + \beta \sum_{i=1}^{d} \left(\frac{\partial u}{\partial x_i} \right)^2 = 0, \quad u(T,x) = g(x).$$

The BSDE becomes

$$dY(t) = -\beta \sum_{i=1}^{d} Z_i^2(t)dt + \sum_{i=1}^{d} Z_i(t)dW_i(t), \quad Y(T) = g(x + W(T)).$$

Rewriting the BSDE, we get a related stochastic control problem

$$\min_{y,Z} E\left[\left\{ Y_T^{y,Z} - g(x + W_T) \right\}^2 \right] = 0,$$

where $Y_T^{y,Z}$ is the solution of

$$dY(t) = -\beta \sum_{i=1}^{d} Z_i^2(t)dt + \sum_{i=1}^{d} Z_i(t)dW_i(t), \quad Y(0) = y.$$

An example

Let

$$f(t, u, x_1, ..., x_d) = \beta \sum_{i=1}^{d} x_i^2.$$

Then the semi-linear parabolic PDE becomes

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}(\Delta_x u)(t,x) + \beta \sum_{i=1}^{d} \left(\frac{\partial u}{\partial x_i}\right)^2 = 0, \quad u(T,x) = g(x).$$

The BSDE becomes

$$dY(t) = -\beta \sum_{i=1}^{d} Z_i^2(t)dt + \sum_{i=1}^{d} Z_i(t)dW_i(t), \quad Y(T) = g(x + W(T)).$$

Rewriting the BSDE, we get a related stochastic control problem

$$\min_{y,Z} E\left[\left\{Y_T^{y,Z} - g(x + W_T)\right\}^2\right] = 0,$$

where $Y_T^{y,Z}$ is the solution of

$$dY(t) = -\beta \sum_{i=1}^{d} Z_i^2(t)dt + \sum_{i=1}^{d} Z_i(t)dW_i(t), \quad Y(0) = y.$$

An example

We're solving semi-linear parabolic PDE

$$\partial_t u + \frac{1}{2}\Delta u + \beta|\nabla u|^2 = 0, \quad u(T,x) = g(x).$$

Define

$$v(t,x) := \exp(2\beta u(t,x)).$$

Compute v_t and Δv

$$v_t = \frac{\partial}{\partial t} e^{2\beta u} = e^{2\beta u} \cdot 2\beta u_t = 2\beta v u_t.$$

For each coordinate x_i:

$$\partial_{x_i} v = \partial_{x_i}(e^{2\beta u}) = e^{2\beta u} \cdot 2\beta u_{x_i} = 2\beta v u_{x_i}.$$

Differentiate once more:

$$\partial_{x_i x_i} v = \partial_{x_i}(2\beta v u_{x_i}) = 2\beta((2\beta v u_{x_i})u_{x_i} + v u_{x_i x_i}) = 2\beta v u_{x_i x_i} + (2\beta)^2 v u_{x_i}^2.$$

Sum over $i = 1, \dots, d$

$$\Delta v = \sum_{i=1}^{d} \partial_{x_i x_i} v = 2\beta v \Delta u + (2\beta)^2 v \sum_{i=1}^{d} u_{x_i}^2 = 2\beta v \Delta u + 4\beta^2 v |\nabla u|^2.$$

An example

We're solving semi-linear parabolic PDE

$$\partial_t u + \frac{1}{2}\Delta u + \beta|\nabla u|^2 = 0, \quad u(T,x) = g(x).$$

Define

$$v(t,x) := \exp(2\beta u(t,x)).$$

Compute v_t and Δv

$$v_t = \frac{\partial}{\partial t} e^{2\beta u} = e^{2\beta u} \cdot 2\beta u_t = 2\beta v u_t.$$

For each coordinate x_i:

$$\partial_{x_i} v = \partial_{x_i}(e^{2\beta u}) = e^{2\beta u} \cdot 2\beta u_{x_i} = 2\beta v u_{x_i}.$$

Differentiate once more:

$$\partial_{x_i x_i} v = \partial_{x_i}(2\beta v u_{x_i}) = 2\beta((2\beta v u_{x_i}) u_{x_i} + v u_{x_i x_i}) = 2\beta v u_{x_i x_i} + (2\beta)^2 v u_{x_i}^2.$$

Sum over $i = 1, \ldots, d$

$$\Delta v = \sum_{i=1}^{d} \partial_{x_i x_i} v = 2\beta v \Delta u + (2\beta)^2 v \sum_{i=1}^{d} u_{x_i}^2 = 2\beta v \Delta u + 4\beta^2 v |\nabla u|^2.$$

An example

We're solving semi-linear parabolic PDE

$$\partial_t u + \frac{1}{2}\Delta u + \beta|\nabla u|^2 = 0, \quad u(T,x) = g(x).$$

Define

$$v(t,x) := \exp(2\beta u(t,x)).$$

Compute v_t and Δv

$$v_t = \frac{\partial}{\partial t} e^{2\beta u} = e^{2\beta u} \cdot 2\beta u_t = 2\beta v u_t.$$

For each coordinate x_i:

$$\partial_{x_i} v = \partial_{x_i}(e^{2\beta u}) = e^{2\beta u} \cdot 2\beta u_{x_i} = 2\beta v u_{x_i}.$$

Differentiate once more:

$$\partial_{x_i x_i} v = \partial_{x_i}(2\beta v u_{x_i}) = 2\beta((2\beta v u_{x_i})u_{x_i} + v u_{x_i x_i}) = 2\beta v u_{x_i x_i} + (2\beta)^2 v u_{x_i}^2.$$

Sum over $i = 1, \ldots, d$

$$\Delta v = \sum_{i=1}^{d} \partial_{x_i x_i} v = 2\beta v \Delta u + (2\beta)^2 v \sum_{i=1}^{d} u_{x_i}^2 = 2\beta v \Delta u + 4\beta^2 v |\nabla u|^2.$$

An example

Combine v_t and Δv

$$v_t + \frac{1}{2}\Delta v = (2\beta v u_t) + \frac{1}{2}(2\beta v \Delta u + 4\beta^2 v|\nabla u|^2) = v\left(2\beta u_t + \beta\Delta u + 2\beta^2|\nabla u|^2\right).$$

Now use the PDE:

$$u_t + \frac{1}{2}\Delta u + \beta|\nabla u|^2 = 0 \implies 2\beta u_t + \beta\Delta u + 2\beta^2|\nabla u|^2 = 0.$$

Therefore

$$v_t + \frac{1}{2}\Delta v = 0.$$

Also, from $u(T,x) = g(x)$,

$$v(T,x) = e^{2\beta g(x)}.$$

v satisfies the backward heat equation

$$v_t + \frac{1}{2}\Delta v = 0, \quad v(T,x) = e^{2\beta g(x)}.$$

An example

Combine v_t and Δv

$$v_t + \frac{1}{2}\Delta v = (2\beta v u_t) + \frac{1}{2}(2\beta v \Delta u + 4\beta^2 v|\nabla u|^2) = v\left(2\beta u_t + \beta\Delta u + 2\beta^2|\nabla u|^2\right).$$

Now use the PDE:

$$u_t + \frac{1}{2}\Delta u + \beta|\nabla u|^2 = 0 \implies 2\beta u_t + \beta\Delta u + 2\beta^2|\nabla u|^2 = 0.$$

Therefore

$$v_t + \frac{1}{2}\Delta v = 0.$$

Also, from $u(T,x) = g(x)$,

$$v(T,x) = e^{2\beta g(x)}.$$

v satisfies the backward heat equation

$$v_t + \frac{1}{2}\Delta v = 0, \quad v(T,x) = e^{2\beta g(x)}.$$

An example

The standard Feynman–Kac representation is for the forward heat equation with initial data at time 0:

$$w_s = \frac{1}{2}\Delta w, \quad w(0,x) = h(x) \implies w(s,x) = \mathbb{E}[h(x+W_s)].$$

Define

$$s := T - t, \quad w(s,x) := v(T-s,x).$$

Then $w(0,x) = v(T,x) = e^{2\beta g(x)}$, differentiate

$$w_s(s,x) = -v_t(T-s,x), \quad \Delta w(s,x) = \Delta v(T-s,x).$$

Since $v_t + \frac{1}{2}\Delta v = 0$, we have $-v_t = \frac{1}{2}\Delta v$. Hence

$$w_s = \frac{1}{2}\Delta w, \quad w(0,x) = e^{2\beta g(x)}.$$

So w solves the forward heat equation with initial condition $e^{2\beta g}$. Apply Feynman–Kac:

$$w(s,x) = \mathbb{E}\left[e^{2\beta g(x+W_s)}\right].$$

An example

The standard Feynman–Kac representation is for the forward heat equation with initial data at time 0:

$$w_s = \frac{1}{2}\Delta w, \quad w(0,x) = h(x) \implies w(s,x) = \mathbb{E}[h(x+W_s)].$$

Define

$$s := T - t, \quad w(s,x) := v(T-s,x).$$

Then $w(0,x) = v(T,x) = e^{2\beta g(x)}$, differentiate

$$w_s(s,x) = -v_t(T-s,x), \quad \Delta w(s,x) = \Delta v(T-s,x).$$

Since $v_t + \frac{1}{2}\Delta v = 0$, we have $-v_t = \frac{1}{2}\Delta v$. Hence

$$w_s = \frac{1}{2}\Delta w, \quad w(0,x) = e^{2\beta g(x)}.$$

So w solves the forward heat equation with initial condition $e^{2\beta g}$. Apply Feynman–Kac:

$$w(s,x) = \mathbb{E}\left[e^{2\beta g(x+W_s)}\right].$$

An example

The standard Feynman–Kac representation is for the forward heat equation with initial data at time 0:

$$w_s = \frac{1}{2}\Delta w, \quad w(0,x) = h(x) \implies w(s,x) = \mathbb{E}[h(x+W_s)].$$

Define

$$s := T - t, \quad w(s,x) := v(T-s,x).$$

Then $w(0,x) = v(T,x) = e^{2\beta g(x)}$, differentiate

$$w_s(s,x) = -v_t(T-s,x), \quad \Delta w(s,x) = \Delta v(T-s,x).$$

Since $v_t + \frac{1}{2}\Delta v = 0$, we have $-v_t = \frac{1}{2}\Delta v$. Hence

$$w_s = \frac{1}{2}\Delta w, \quad w(0,x) = e^{2\beta g(x)}.$$

So w solves the forward heat equation with initial condition $e^{2\beta g}$. Apply Feynman–Kac:

$$w(s,x) = \mathbb{E}\left[e^{2\beta g(x+W_s)}\right].$$

An example

Finally the change of variables $s = T - t$:

$$v(t,x) = w(T-t,x) = \mathbb{E}\left[e^{2\beta g(x+W_{T-t})}\right].$$

Since $v = e^{2\beta u}$,

$$u(t,x) = \frac{1}{2\beta}\ln\left(E\left[\exp\{2\beta g(x+W_{T-t})\}\right]\right).$$

In particular, $u(T,x) = g(x)$ and for the BSDE

$$Y(0) = u(0,x) = \frac{1}{2\beta}\ln\left(E\left[\exp\{2\beta g(x+W_T)\}\right]\right).$$

Note that the solution of the BSDE is given by the pair (Y_t, Z_t), where

$$Y_t = u(t,x+W_t), \quad Z_t = (\nabla_x u)(t,x+W_t), \quad 0 \le t \le T,$$

$$Y_T = u(T,x+W_T) = g(x+W_T).$$

This semi-linear parabolic PDE can also be seen as an HJB equation for a simple quadratic-control problem.

An example

Finally the change of variables $s = T - t$:

$$v(t,x) = w(T-t,x) = \mathbb{E}\left[e^{2\beta g(x+W_{T-t})}\right].$$

Since $v = e^{2\beta u}$,

$$u(t,x) = \frac{1}{2\beta}\ln\left(E\left[\exp\{2\beta g(x+W_{T-t})\}\right]\right).$$

In particular, $u(T,x) = g(x)$ and for the BSDE

$$Y(0) = u(0,x) = \frac{1}{2\beta}\ln\left(E\left[\exp\{2\beta g(x+W_T)\}\right]\right).$$

Note that the solution of the BSDE is given by the pair (Y_t, Z_t), where

$$Y_t = u(t, x+W_t), \quad Z_t = (\nabla_x u)(t, x+W_t), \quad 0 \le t \le T,$$

$$Y_T = u(T, x+W_T) = g(x+W_T).$$

This semi-linear parabolic PDE can also be seen as an HJB equation for a simple quadratic-control problem.

An example

Finally the change of variables $s = T - t$:

$$v(t,x) = w(T-t,x) = \mathbb{E}\left[e^{2\beta g(x+W_{T-t})}\right].$$

Since $v = e^{2\beta u}$,

$$u(t,x) = \frac{1}{2\beta}\ln\left(E\left[\exp\{2\beta g(x+W_{T-t})\}\right]\right).$$

In particular, $u(T,x) = g(x)$ and for the BSDE

$$Y(0) = u(0,x) = \frac{1}{2\beta}\ln\left(E\left[\exp\{2\beta g(x+W_T)\}\right]\right).$$

Note that the solution of the BSDE is given by the pair (Y_t, Z_t), where

$$Y_t = u(t, x+W_t), \quad Z_t = (\nabla_x u)(t, x+W_t), \quad 0 \le t \le T,$$

$$Y_T = u(T, x+W_T) = g(x+W_T).$$

This semi-linear parabolic PDE can also be seen as an HJB equation for a simple quadratic-control problem.

An example

Take a d-dimensional control $\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{d,t})$ and the controlled state

$$dX_t = -2\sqrt{\beta}\lambda_t dt + dW_t, \quad X_t = x.$$

Running cost is $\|\lambda_t\|^2 = \sum_i \lambda_{i,t}^2$. The value function (at time t) is

$$u(t,x) = \min_{\lambda} \mathbb{E}\left[\int_t^T \|\lambda_s\|^2 ds + g(X_T) \,\middle|\, \mathcal{F}_t\right].$$

Indeed, the HJB equation is given by

$$0 = \partial_t u(t,x) + \inf_{\lambda}\left\{(-2\sqrt{\beta}\lambda)\cdot\nabla u(t,x) + \frac{1}{2}\Delta u(t,x) + \|\lambda\|^2\right\}, \quad u(T,x) = g(x).$$

Pull out the term not depending on λ:

$$0 = \partial_t u + \frac{1}{2}\Delta u + \inf_{\lambda}\left\{\|\lambda\|^2 - 2\sqrt{\beta}\lambda\cdot\nabla u\right\}.$$

An example

Take a d-dimensional control $\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{d,t})$ and the controlled state

$$dX_t = -2\sqrt{\beta}\lambda_t dt + dW_t, \quad X_t = x.$$

Running cost is $\|\lambda_t\|^2 = \sum_i \lambda_{i,t}^2$. The value function (at time t) is

$$u(t,x) = \min_{\lambda} \mathbb{E}\left[\int_t^T \|\lambda_s\|^2 ds + g(X_T) \,\middle|\, \mathcal{F}_t\right].$$

Indeed, the HJB equation is given by

$$0 = \partial_t u(t,x) + \inf_{\lambda}\left\{(-2\sqrt{\beta}\lambda)\cdot\nabla u(t,x) + \frac{1}{2}\Delta u(t,x) + \|\lambda\|^2\right\}, \quad u(T,x) = g(x).$$

Pull out the term not depending on λ:

$$0 = \partial_t u + \frac{1}{2}\Delta u + \inf_{\lambda}\left\{\|\lambda\|^2 - 2\sqrt{\beta}\lambda\cdot\nabla u\right\}.$$

An example

For fixed $p := \nabla u(t,x)$,

$$\|\lambda\|^2 - 2\sqrt{\beta}\lambda \cdot p = \|\lambda - \sqrt{\beta}p\|^2 - \beta\|p\|^2.$$

Hence

$$\inf_{\lambda}\{\cdots\} = -\beta\|\nabla u\|^2, \quad \lambda^*(t,x) = \sqrt{\beta}\nabla u(t,x).$$

Plugging back:

$$\partial_t u + \frac{1}{2}\Delta u - \beta\|\nabla u\|^2 = 0, \quad u(T,x) = g(x).$$

The above theory can be extended to include forward state dynamics, which can a¤ect both the drift and volatility, resulting in forward-backward stochastic di¤erential equations (FBSDE)

An example

For fixed $p := \nabla u(t,x)$,

$$\|\lambda\|^2 - 2\sqrt{\beta}\lambda \cdot p = \|\lambda - \sqrt{\beta}p\|^2 - \beta\|p\|^2.$$

Hence

$$\inf_{\lambda}\{\cdots\} = -\beta\|\nabla u\|^2, \quad \lambda^*(t,x) = \sqrt{\beta}\nabla u(t,x).$$

Plugging back:

$$\partial_t u + \frac{1}{2}\Delta u - \beta\|\nabla u\|^2 = 0, \quad u(T,x) = g(x).$$

The above theory can be extended to include forward state dynamics, which can a¤ect both the drift and volatility, resulting in forward-backward stochastic di¤erential equations (FBSDE)

Forward backward stochastic differential equations

Consider the following forward backward stochastic differential equation (FBSDE): Find a random variable Y_0 and an adapted process Z_t such that

$$\begin{aligned} dX_t &= \mu(t, X_t)dt + \sigma(t, X_t)^T dW_t, \quad X_0 = x \\ dY_t &= -f(t, X_t, Y_t, Z_t)dt + Z_t^T dW_t, \\ Y_T &= g(W_T). \end{aligned}$$

where T means transpose and W_t is a standard d-dimensional Brownian motion. This can be viewed as a hedging problem to match the final payoff Y_T by finding the initial price Y_0 and the hedging strategy Z_t, but with the stochastic state dynamic X_t for the drift and volatility of Y_t.

The above FBSDE problem can be formulated as a special stochastic control problem such that

$$\min_{y,Z} E\left[\left\{Y_T^{y,Z} - g(W_T)\right\}^2\right] = 0,$$

where $Y_T^{y,Z}$ is the solution of

$$\begin{aligned} dX_t &= \mu(t, X_t)dt + \sigma(t, X_t)^T dW_t, \quad X_0 = x \\ dY_t &= -f(t, X_t, Y_t, Z_t)dt + Z_t^T dW_t. \end{aligned}$$

Forward backward stochastic differential equations

Consider the following forward backward stochastic differential equation (FBSDE): Find a random variable Y_0 and an adapted process Z_t such that

$$\begin{aligned} dX_t &= \mu(t, X_t)dt + \sigma(t, X_t)^T dW_t, \quad X_0 = x \\ dY_t &= -f(t, X_t, Y_t, Z_t)dt + Z_t^T dW_t, \\ Y_T &= g(W_T). \end{aligned}$$

where T means transpose and W_t is a standard d-dimensional Brownian motion. This can be viewed as a hedging problem to match the final payoff Y_T by finding the initial price Y_0 and the hedging strategy Z_t, but with the stochastic state dynamic X_t for the drift and volatility of Y_t.

The above FBSDE problem can be formulated as a special stochastic control problem such that

$$\min_{y,Z} E\left[\left\{Y_T^{y,Z} - g(W_T)\right\}^2\right] = 0,$$

where $Y_T^{y,Z}$ is the solution of

$$\begin{aligned} dX_t &= \mu(t, X_t)dt + \sigma(t, X_t)^T dW_t, \quad X_0 = x \\ dY_t &= -f(t, X_t, Y_t, Z_t)dt + Z_t^T dW_t. \end{aligned}$$

Connection with a semi-linear parabolic PDE

This minimization problem can again be solved by using a neural network, by learning $Y_T^{y,Z}$ to match $g(\xi + W_T)$ as close as possible.
Time grid $0 = t_0 < \cdots < t_N = T, \Delta t_k$.
Forward Euler for X:

$$X_{k+1} = X_k + \mu(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)^T \Delta W_k.$$

Parameterize $Z : Z_{t_k} \approx Z_\theta(t_k, X_k) \in \mathbb{R}^d$(NN). Treat Y_0 as a trainable scalar y_θ.
Backward Euler-Maruyama for Y:

$$Y_{k+1} = Y_k - f(t_k, X_k, Y_k, Z_\theta(t_k, X_k))\Delta t_k + Z_\theta(t_k, X_k)^T \Delta W_k, \quad Y_0 = y_\theta.$$

The loss function, empirical mean of terminal mismatch

$$\mathcal{L}(\theta) = \frac{1}{M}\sum_{m=1}^{M}\left(Y_N^{(m)} - g(W_T^{(m)})\right)^2.$$

Minimize by SGD/backprop through the simulation; at optimum $\mathcal{L} \approx 0$ and $y_\theta \approx Y_0$.

Connection with a semi-linear parabolic PDE

This minimization problem can again be solved by using a neural network, by learning $Y_T^{y,Z}$ to match $g(\xi + W_T)$ as close as possible.
Time grid $0 = t_0 < \cdots < t_N = T, \Delta t_k$.
Forward Euler for X:

$$X_{k+1} = X_k + \mu(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)^T \Delta W_k.$$

Parameterize $Z : Z_{t_k} \approx Z_\theta(t_k, X_k) \in \mathbb{R}^d$(NN). Treat Y_0 as a trainable scalar y_θ.
Backward Euler-Maruyama for Y:

$$Y_{k+1} = Y_k - f(t_k, X_k, Y_k, Z_\theta(t_k, X_k))\Delta t_k + Z_\theta(t_k, X_k)^T \Delta W_k, \quad Y_0 = y_\theta.$$

The loss function, empirical mean of terminal mismatch

$$\mathcal{L}(\theta) = \frac{1}{M} \sum_{m=1}^{M} \left(Y_N^{(m)} - g(W_T^{(m)}) \right)^2.$$

Minimize by SGD/backprop through the simulation; at optimum $\mathcal{L} \approx 0$ and $y_\theta \approx Y_0$.

Connection with a semi-linear parabolic PDE

This minimization problem can again be solved by using a neural network, by learning $Y_T^{y,Z}$ to match $g(\xi + W_T)$ as close as possible.
Time grid $0 = t_0 < \cdots < t_N = T, \Delta t_k$.
Forward Euler for X:

$$X_{k+1} = X_k + \mu(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)^T \Delta W_k.$$

Parameterize $Z : Z_{t_k} \approx Z_\theta(t_k, X_k) \in \mathbb{R}^d$(NN). Treat Y_0 as a trainable scalar y_θ.
Backward Euler-Maruyama for Y:

$$Y_{k+1} = Y_k - f(t_k, X_k, Y_k, Z_\theta(t_k, X_k))\Delta t_k + Z_\theta(t_k, X_k)^T \Delta W_k, \quad Y_0 = y_\theta.$$

The loss function, empirical mean of terminal mismatch

$$\mathcal{L}(\theta) = \frac{1}{M} \sum_{m=1}^{M} \left(Y_N^{(m)} - g(W_T^{(m)}) \right)^2.$$

Minimize by SGD/backprop through the simulation; at optimum $\mathcal{L} \approx 0$ and $y_\theta \approx Y_0$.

Connection with a semi-linear parabolic PDE

It can be shown that the above FBSDE is linked to a semi-linear parabolic PDE

$$0 = \frac{\partial u}{\partial t}(t,x) + Au(t,x) + f(t,x,u(t,x),\sigma(t,X_t)^T(\nabla_x u)(t,x)),$$

$$u(T,x) = g(x),$$

where the generator

$$Au(t,x) = \frac{1}{2}\mathrm{Tr}\left[\sigma\sigma^T(t,x)(\mathrm{Hess}_x u)(t,x)\right] + (\nabla_x u)(t,x)\cdot\mu(t,x)$$

where $\mathrm{Hess}_x u$ is the Hessian matrix, and Tr denotes the trace of the matrix. This equation is again semi-linear because the second derivative is linear despite that f on the first derivative is nonlinear.
The solution of FBSDE can be given in terms of u as

$$Y_t = u(t,X_t), \quad Z_t = \sigma(t,X_t)^T(\nabla_x u)(t,X_t).$$

In particular, one can also use the neural network to solve the semi-linear parabolic PDE to get $u(0,\xi)$, because

$$u(0,x) = Y_0.$$

Connection with a semi-linear parabolic PDE

It can be shown that the above FBSDE is linked to a semi-linear parabolic PDE

$$0 = \frac{\partial u}{\partial t}(t,x) + Au(t,x) + f(t,x,u(t,x),\sigma(t,X_t)^T(\nabla_x u)(t,x)),$$

$$u(T,x) = g(x),$$

where the generator

$$Au(t,x) = \frac{1}{2}\mathrm{Tr}\left[\sigma\sigma^T(t,x)(\mathrm{Hess}_x u)(t,x)\right] + (\nabla_x u)(t,x)\cdot\mu(t,x)$$

where $\mathrm{Hess}_x u$ is the Hessian matrix, and Tr denotes the trace of the matrix. This equation is again semi-linear because the second derivative is linear despite that f on the first derivative is nonlinear.

The solution of FBSDE can be given in terms of u as

$$Y_t = u(t,X_t), \quad Z_t = \sigma(t,X_t)^T(\nabla_x u)(t,X_t).$$

In particular, one can also use the neural network to solve the semi-linear parabolic PDE to get $u(0,\xi)$, because

$$u(0,x) = Y_0.$$

Connection with a semi-linear parabolic PDE

It can be shown that the above FBSDE is linked to a semi-linear parabolic PDE

$$0 = \frac{\partial u}{\partial t}(t,x) + Au(t,x) + f(t,x,u(t,x),\sigma(t,X_t)^T(\nabla_x u)(t,x)),$$

$$u(T,x) = g(x),$$

where the generator

$$Au(t,x) = \frac{1}{2}\mathsf{Tr}\left[\sigma\sigma^T(t,x)(\mathsf{Hess}_x u)(t,x)\right] + (\nabla_x u)(t,x)\cdot\mu(t,x)$$

where $\mathsf{Hess}_x u$ is the Hessian matrix, and Tr denotes the trace of the matrix. This equation is again semi-linear because the second derivative is linear despite that f on the first derivative is nonlinear.
The solution of FBSDE can be given in terms of u as

$$Y_t = u(t,X_t), \quad Z_t = \sigma(t,X_t)^T(\nabla_x u)(t,X_t).$$

In particular, one can also use the neural network to solve the semi-linear parabolic PDE to get $u(0,\xi)$, because

$$u(0,x) = Y_0.$$