Note 3 – version 3 – 2022.7.7 – Please send remarks and corrections to gubinelli@iam.uni-bonn.de

The martingale representation theorem and applications

In a Brownian filtrered probability space all martingales are stochastic integrals. In this part we will prove this theorem which is at the basis of the analysis on Wiener space and consider some applications: the Boué–Dupuis formula and backward SDEs.

Table of contents

1 Brownian martingale representation theorem 1

2 Boué–Dupuis formula 6

Applications to functional analysis 12

3 Applications to probabilitistic problems 13

4 Large deviations of diffusion 14

5 Backward SDEs and non-linear PDEs 22

5.1 Solution theory for BSDEs 25

5.2 Representation formula for non-linear PDEs 27

1 Brownian martingale representation theorem

We concentrate now in the study of the probability space generated by a Brownian motion (maybe multidimensional, taking values in $\mathbb{R}^d$). We assume in this part that $(\Omega, \mathscr{F}, (\mathscr{F}_t)_t, \mathbb{P})$ is the canonical d-dimensional Wiener space, i.e. $\Omega = \mathscr{C}^d = C(\mathbb{R}_+, \mathbb{R}^d)$, $X_t(\omega) = \omega(t)$, $\mathbb{P}$ is the law of the Brownian motion and $(\mathscr{F}_t)_{t \geqslant 0}$ is the right continuous $\mathbb{P}$-completed filtration generated by the canonical process $(X_t)_{t \geqslant 0}$ in particular we have $\mathscr{F}_\infty = \mathscr{F} = \overline{\mathscr{B}(\Omega)}^{\mathbb{P}}$. This is called a Brownian probability space.

Theorem 1. *Let $\Phi \in L^2(\Omega, \mathscr{F}, \mathbb{P})$, then there exists a unique predictable process $F \in L^2_{\mathscr{P}}(\mathbb{R}_+ \times \Omega; \mathbb{R}^d)$ such that*

$$\Phi = \mathbb{E}[\Phi] + \sum_{k=1}^{d} \int_0^\infty F_s^{(k)} \mathrm{d}X_s^{(k)}, \qquad \mathbb{P}\text{-}a.s.$$

This theorem says that any mean zero L^2 random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ can be written as a stochastic integral wrt. the Brownian motion. It will have as a consequence that any martingale on $(\Omega, \mathscr{F}, \mathbb{P})$ is a stochastic integral wrt. to (the given) Brownian motion and therefore it has a continuous modification. This rules out the possibility that martingales on a Brownian probability space has jumps, "informations comes in in a continuous way".

We will give a "Markovian" proof. In the next exercise sheet you will be asked to give a "Gaussian" proof.

We need this technical lemma.

Lemma 2. *Let $p \geqslant 1$ and $\mathscr{C} \subseteq L^\infty(\Omega, \mathscr{F}, \mathbb{P})$ be the algebra generated by the random variables*

$$\Phi^\alpha(f) := \int_0^\infty e^{-\alpha t} f(X_t)\mathrm{d}t$$

where $\alpha > 0$ and $f \in C_c^\infty(\mathbb{R}^d)$ (smooth and compact support). Then $\mathscr{C}$ is dense in $L^p(\Omega, \mathscr{F}, \mathbb{P})$.

The interest of this algebra of functions is that it behaves *nicely* wrt. Markov processes. (The proof really uses only the continuity of the trajectories of X and the fact that $\mathscr{F}$ is the filtration generated by X.

Proof. (of Theorem 1) If $F \in \mathscr{C}$ we can give an explicit martingale representation because conditional expectations of elements in $\mathscr{C}$ can be computed explicitly. Take for example $\Phi^\alpha(f)$, then we have by the Markov property

$$\begin{aligned}
\mathbb{E}[\Phi^\alpha(f)|\mathscr{F}_t] = \mathbb{E}\left[\int_0^\infty e^{-\alpha s} f(X_s)\mathrm{d}s\middle|\mathscr{F}_t\right] &= \int_0^t e^{-\alpha s} f(X_s)\mathrm{d}s + \int_t^\infty e^{-\alpha s} \underbrace{\mathbb{E}[f(X_s)|\mathscr{F}_t]}_{(P_{s-t}f)(X_t)}\mathrm{d}s \\
&= \int_0^t e^{-\alpha s} f(X_s)\mathrm{d}s + \int_t^\infty e^{-\alpha s}(P_{s-t}f)(X_t)\mathrm{d}s \\
&= \int_0^t e^{-\alpha s} f(X_s)\mathrm{d}s + e^{-\alpha t}\underbrace{\int_0^\infty e^{-\alpha s} P_s f(X_t)\mathrm{d}s}_{=:U^\alpha(f)(X_t)}
\end{aligned}$$

where we let $U^\alpha f(x) := \int_0^\infty e^{-\alpha t} P_t f(x)\mathrm{d}x$ for any $\alpha > 0$ (the resolvent operator) and $f \in C(\mathbb{R}^d)$ and with P_t the transition operator for the Brownian motion:

$$P_t f(x) = \frac{1}{(2\pi t)^{d/2}} \int_{\mathbb{R}^d} f(y) e^{-|x-y|^2/t^2}\mathrm{d}y.$$

Recall that a generic element of $\mathscr{C}$ is a finite linear combination of monomials of the form

$$\prod_{i=1}^n \Phi^{\alpha_i}(f_i)$$

for some $\alpha_1, \ldots, \alpha_n > 0$ and $f_1, \ldots, f_n \in C_0^\infty(\mathbb{R}^d)$. This can be written as (where S_n is the set of permutations of n elements, and $t \geqslant 0$ is arbitrary)

$$\prod_{i=1}^{n} \Phi^{\alpha_i}(f_i) = \sum_{\sigma \in S_n} \int_{0<s_1<\cdots<s_n} \left[\prod_{i=1}^{n} e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] \mathrm{d}s_1 \cdots \mathrm{d}s_n$$

$$= \sum_{\sigma \in S_n} \int_{0<s_1<\cdots<s_n} \left[\prod_{i=1}^{n} (\mathbb{1}_{s_i \leqslant t} + \mathbb{1}_{s_i > t}) e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] \mathrm{d}s_1 \cdots \mathrm{d}s_n$$

$$= \sum_{k=0}^{n} \sum_{\sigma \in S_n} \int_{0<s_1<\cdots<s_n} \mathbb{1}_{s_k \leqslant t} \mathbb{1}_{s_{k+1} > t} \left[\prod_{i=1}^{n} e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] \mathrm{d}s_1 \cdots \mathrm{d}s_n$$

$$= \sum_{k=0}^{n} \sum_{\sigma \in S_n} \int_{0<s_1<\cdots<s_k<t} \left[\prod_{i=1}^{k} e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] \mathrm{d}s_1 \cdots \mathrm{d}s_k \int_{t \leqslant s_{k+1} < s_n} \left[\prod_{i=k+1}^{n} e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] \mathrm{d}s_{k+1} \cdots \mathrm{d}s_n$$

$$= \sum_{k=0}^{n} \sum_{\sigma \in S_n} V_t^{\sigma,k}(X) \int_{t \leqslant s_{k+1} < s_n} \left[\prod_{i=k+1}^{n} e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] \mathrm{d}s_{k+1} \cdots \mathrm{d}s_n$$

where we use the convention that $s_0 = 0$ and $s_{n+1} = +\infty$ and where we let

$$V_t^{\sigma,k}(X) = \int_{0<s_1<\cdots<s_k<t} \left[\prod_{i=1}^{k} e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] \mathrm{d}s_1 \cdots \mathrm{d}s_k.$$

A computation using the Markov property inductively gives

$$\mathbb{E}\left[\int_{t \leqslant s_{k+1} < s_n} \left[\prod_{i=k+1}^{n} e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] \mathrm{d}s_{k+1} \cdots \mathrm{d}s_n \middle| \mathscr{F}_t\right]$$

$$= \mathbb{E}\left[\int_{t \leqslant s_{k+1} < s_{n-1}} \left[\prod_{i=k+1}^{n-1} e^{-\alpha_{\sigma(i)} s_i} f_{\sigma(i)}(X_{s_i})\right] e^{-\alpha_{\sigma(n)} s_{n-1}} U^{\alpha_{\sigma(n)}} f_{\sigma(n)}(X_{s_{n-1}}) \mathrm{d}s_{k+1} \cdots \mathrm{d}s_{n-1} \middle| \mathscr{F}_t\right]$$

$$= e^{-\alpha(\sigma,k)t} U^{\alpha(\sigma,k)}(f_{\sigma(k+1)} U^{\alpha(\sigma,k+1)}(f_{\sigma(k+2)} \cdots (f_{\sigma(n-1)} U^{\alpha(\sigma,n-1)}(f_{\sigma(n)})))) (X_t)$$

$$= e^{-\alpha(\sigma,k)t} U^{\alpha(\sigma,k)}(H^{\sigma,k})(X_t)$$

where $\alpha(\sigma,k) = \alpha_{\sigma(k+1)} + \alpha_{\sigma(k+2)} + \cdots + \alpha_{\sigma(n)}$ and

$$H^{\sigma,k}(x) := f_{\sigma(k+1)}(x) U^{\alpha(\sigma,k+1)}(f_{\sigma(k+2)} \cdots (f_{\sigma(n-1)} U^{\alpha(\sigma,n)}(f_{\sigma(n)})))(x) = f_{\sigma(k+1)}(x) U^{\alpha(\sigma,k+1)}(H^{\sigma,k+1})(x)$$

$$H^{\sigma,n}(x) := f_{\sigma(n)}(x).$$

We conclude that

$$M_t = \mathbb{E}\left[\prod_{i=1}^{n} \Phi^{\alpha_i}(f_i) \middle| \mathscr{F}_t\right] = \sum_{k=0}^{n} \sum_{\sigma \in S_n} V_t^{\sigma,k}(X) e^{-\alpha(\sigma,k)t} U^{\alpha(\sigma,k)}(H^{\sigma,k})(X_t). \tag{1}$$

This formula shows that the martingale $(M_t)_{t\geqslant 0}$ is continuous in $t\in\mathbb{R}$ since this is so for the the r.h.s. since $V_t^{\sigma,k}(X)$ is an integral and therefore continuous in t and $U^{\alpha(\sigma,k)}(H^{\sigma,k})(x)$ a smooth function of x. Note that $t\mapsto V_t^{\sigma,k}(X)e^{-\alpha(\sigma,k)t}$ is a bounded variation process. So the only contributions to the martingale M_t must come from the processes $t\mapsto U^{\alpha(\sigma,k)}(H^{\sigma,k})(X_t)$. By Ito formula we have

$$\mathrm{d}U^{\alpha(\sigma,k)}(H^{\sigma,k})(X_t)=\nabla(U^{\alpha(\sigma,k)}(H^{\sigma,k}))(X_t)\mathrm{d}X_t+\text{bounded variation part}$$

we do not care about the bounded variation part since it has to cancel with the bounded variation part coming from $t\mapsto V_t^{\sigma,k}(X)e^{-\alpha(\sigma,k)t}$ (maybe, as an exercise, you can check it). By equating the two continuous local martingales on the l.h.s. and r.h.s. of eq. (1) we deduce that

$$M_t-M_0=\int_0^t F_s\cdot\mathrm{d}X_s$$

where

$$F_s:=\sum_{k=0}^{n}\sum_{\sigma\in S_n}V_s^{\sigma,k}(X)e^{-\alpha(\sigma,k)s}\nabla(U^{\alpha(\sigma,k)}(H^{\sigma,k}))(X_s).$$

By taking $t\to\infty$ this shows that (by martingale convegence theorem in L^2)

$$\prod_{i=1}^{n}\Phi^{\alpha_i}(f_i)=\mathbb{E}\left[\prod_{i=1}^{n}\Phi^{\alpha_i}(f_i)\right]+\int_0^\infty F_s\cdot\mathrm{d}X_s$$

indeed note that by Ito isometry

$$\left(\mathbb{E}\left[\prod_{i=1}^{n}\Phi^{\alpha_i}(f_i)\right]\right)^2+\mathbb{E}\left[\left(\int_0^\infty F_s\cdot\mathrm{d}X_s\right)^2\right]=\mathbb{E}\left[\left(\prod_{i=1}^{n}\Phi^{\alpha_i}(f_i)\right)^2\right]<\infty.$$

Any $\Phi\in\mathscr{C}$ can be written as a stochastic integral wrt. Brownian motion plus a constant.

For general $\Phi\in L^2(\Omega,\mathscr{F},\mathbb{P})$ we can choose a sequence $(\Phi_n)_{n\geqslant 1}\subset\mathscr{C}$ such that $\Phi_n\to\Phi$ in L^2. Now let $M_t^n:=\mathbb{E}[\Phi^n|\mathscr{F}_t]$ and $M_t=\mathbb{E}[\Phi|\mathscr{F}_t]$.

By the previous step we know there exists adapted functions $F^n\in L^2_{\mathscr{P}}(\mathbb{R}_+\times\Omega)$ such that

$$M_t^n=\mathbb{E}[\Phi^n]+\int_0^t F_s^n\mathrm{d}X_s,$$

therefore by Ito isometry and $n,m\geqslant 1$

$$\mathbb{E}[(M_t^n-M_t^m)^2]=\mathbb{E}[[M^n-M^m]_t]=\mathbb{E}\int_0^t|F_s^n-F_s^m|_{\mathbb{R}^d}^2\mathrm{d}s,\qquad t\geqslant 0.$$

therefore

$$\mathbb{E}\int_0^\infty|F_s^n-F_s^m|_{\mathbb{R}^d}^2\mathrm{d}s\leqslant\sup_t\mathbb{E}[(M_t^n-M_t^m)^2]\leqslant\mathbb{E}\left[\sup_{t\geqslant 0}(M_t^n-M_t^m)^2\right]=o_{n,m}(1)$$

By martingale convergence theorem we have that $M_t^n \to M_t$ a.s. and in L^2 and by Doob's maximal inequality this convegence is uniform in t (here we need that the filtration is right-continuous). This implies also that $(F^n)_{n\geqslant 1}$ is a Cauchy sequence in $L^2_{\mathscr{P}}(\mathbb{R}_+ \times \Omega)$ which is complete therefore there exists a unique limit $F = \lim_n F^n \in L^2_{\mathscr{P}}(\mathbb{R}_+ \times \Omega)$ and from this we get that

$$M_t = \mathbb{E}[\Phi] + \int_0^t F_s \mathrm{d}X_s.$$

By taking $t \to \infty$ and using L^2 convergence and $M_t \to \mathbb{E}[\Phi|\mathscr{F}_\infty] = \Phi$ in L^2 (because $\mathscr{F}_\infty = \mathscr{F}$) we obtain that there exists $F \in L^2_{\mathscr{P}}(\mathbb{R}_+ \times \Omega)$ such that

$$\Phi = \mathbb{E}[\Phi] + \int_0^\infty F_s \mathrm{d}X_s.$$

In general there is no easy formula for F. □

Corollary 3. *All local martingales in a Brownian probability space are continuous.*

Proof. Exercise. □

Applications of the martingale representation theorem

a) Mathematical finance: if you model the evolution of stock prices with the probability space generated by a multidimensional Brownian motion X then any "contract" Φ can be expressed as

$$\Phi = \mathbb{E}[\Phi] + \int_0^\infty F_s \mathrm{d}X_s$$

which means that we can replicate the contract by trading the underlying assets X using the strategy given by F (if we are able to compute or approximate F). The strategy F (which is a vector $(F^1, \ldots, F^d)$) has to be interpreted as follows: F^k is the number of stocks of the asset k which one has to acquire at the beginning of every "infinitesimal" trading round.

b) Study of the entropy $H(\mathbb{Q}|\mathbb{P})$ of two measures $\mathbb{P}, \mathbb{Q}$ on the Brownian probability space with application to the estimation of averages of functionals and to small noise large deviations of diffusion, i.e. investigate the behaviour of the law μ^ε of the solution of the SDE

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

as $\varepsilon \to 0$.

c) Backward SDEs (BSDE): this is a class of stochastic differential equations with final condition (instead of initial condition). Let Φ be a given random variable which is $\mathscr{F}_T$ measurable for given $T > 0$ (deterministic) the solution to a BSDE with *driver* $f(t, y, z)$ is a pair (Y, Z) of adapted processes such that

$$-\mathrm{d}Y_t = f(t, Y_t, Z_t)\mathrm{d}t + Z_t \mathrm{d}W_t, \qquad t \in [0, T]$$

and $Y_T = \Phi$, where $(W_t)_{t\geqslant 0}$ it is an adapted Brownian motion and $t\mapsto f(t,y,z)$ an adapted process. This kind of equations has application in finance but also applications in the representations of solutions to non-linear PDEs (very much like SDE can represent solutions to certain classes of linear PDEs, e.g. via Feynman-Kac formula).

---

2 Boué–Dupuis formula

We assume that $(\Omega, \mathscr{F}, (\mathscr{F}_t)_t, \mathbb{P})$ is the canonical d-dimensional Wiener space, i.e. $\Omega = \mathscr{C}^d = C(\mathbb{R}_+, \mathbb{R}^d)$, $X_t(\omega) = \omega(t)$, $\mathbb{P}$ is the law of the Brownian motion and $(\mathscr{F}_t)_{t\geqslant 0}$ is the right continuous $\mathbb{P}$-completed filtration generated by the canonical process $(X_t)_{t\geqslant 0}$ in particular we have $\mathscr{F}_\infty = \mathscr{F} = \overline{\mathscr{B}(\Omega)}^{\mathbb{P}}$. We will also use the notation μ for the Wiener measure $\mathbb{P}$.

In this section we are going to prove the following result.

Theorem 4. *(Boué–Dupuis formula) For any function $f: \Omega\to\mathbb{R}$ measurable and bounded from below we have*

$$\log \mathbb{E}_\mu[e^f] = \sup_{u\in\mathbb{H}} \mathbb{E}_\mu\left[f(X + I(u(X))) - \frac{1}{2}\|u(X)\|_{\mathbb{H}}^2\right]$$

where the supremum on the r.h.s. is taken wrt. all the predictable functions $u: \mathbb{R}_+\times\Omega\to\mathbb{R}$ such that

$$\|u\|_{\mathbb{H}}^2 = \int_0^\infty |u_s|^2 \mathrm{d}s < \infty, \qquad \mu - a.s. \tag{2}$$

and we write $u(\omega) = u(X(\omega))$ to stress the measurability wrt. the filtratrion $\mathscr{F}$ generated by X and where

$$I(u)(t) = \int_0^t u_s(X)\mathrm{d}s, \qquad t\geqslant 0.$$

We call a function u as above, a drift (wrt. μ).

Remark 5. This formula is useful because transform the problem of computing the average $\mathbb{E}_\mu[e^f]$ into a control problem: one has find a control u which does not cost much (the cost is measured by the norm $\|u\|_{\mathbb{H}}$) and which allows the Brownian motion X to reach regions where f is large.

Entropy of a probability measure

We consider the measure space $(\Omega, \mathscr{B}(\Omega))$ but the following definition makes sense for any Polish space. Denote $\Pi(\Omega)$ the (Polish) space of probability measures on $(\Omega, \mathscr{B}(\Omega))$ endowed with the weak topology.

Definition 6. *The relative entropy of a probability measure ν wrt. μ where $\mu,\nu\in\Pi(\Omega)$ is defined as*

$$H(\nu|\mu) = \sup_{\varphi\in L^\infty(\Omega)} (\nu(\varphi) - \log\mu(e^\varphi))$$

where $\nu(f) = \int_\Omega f(\omega)\nu(\mathrm{d}\omega)$ *denotes the average of f wrt. the measure* ν*.*

Remark 7. The supremum is taken over the set $L^\infty(\Omega)$ of bounded measurable functions. The following properties are true (but we will not prove them).

a) The supremum can also be taken wrt. all the continuous bounded functions on Ω

b) The function $\nu \mapsto H(\nu|\mu)$ is non-negative, convex, lower semi-continuous (wrt. the weak topology) and moreover

$$H(\nu|\mu) = \int_\Omega \log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\nu$$

if $\nu \ll \mu$ and $H(\nu|\mu) = +\infty$ otherwise. Note that $H(\nu|\mu) = 0$ iff $\nu = \mu$.

c) We have also the convex dual formula

$$\log \mu(e^\varphi) = \sup_{\nu \in \Pi(\Omega)} [\nu(\varphi) - H(\nu|\mu)]$$

This last formula will be important to prove the BD formula. And in general one has

$$\nu(\varphi) \leqslant \log \mu(e^\varphi) + H(\nu|\mu)$$

for any $\varphi \in L^\infty(\Omega)$ and $\nu, \mu \in \Pi(\Omega)$.

We need to prove several lemmas before being ready to prove the BD formula. In the following μ will stand always for the Wiener measure and all drifts will be taken wrt. the Wiener measure (i.e. $\|u\|_{\mathbb{H}} < \infty$ μ-a.s.).

Lemma 8. *Let u be a drift and let* ν *be the law of the process* $Y = X + I(u(X))$ *under* μ*. Then*

$$H(\nu|\mu) \leqslant \frac{1}{2}\mathbb{E}_\mu[\|u(X)\|_{\mathbb{H}}^2].$$

Proof. Assume for the moment that $\|u\|_{\mathbb{H}}$ is almost surely bounded by a finite deterministic number $K < \infty$. By Novikov's criterion we can define the probability measure $\rho \in \Pi(\Omega)$ with density

$$\frac{\mathrm{d}\rho}{\mathrm{d}\mu} = \mathscr{E}\left(-\int_0^\cdot u_s(X)\mathrm{d}X_s\right)_\infty = \exp\left(-\int_0^\infty u_s(X)\mathrm{d}X_s - \frac{1}{2}\int_0^\infty |u_s|^2 \mathrm{d}s\right)$$

with respect to μ. By Girsanov's theorem the process $Y = X + I(u(X))$ is a Brownian motion under ρ, that is it has law μ. This means that for any measurable bounded function $f \in L^\infty(\Omega)$ we have

$$\mathbb{E}_\nu[f(X)] = \mathbb{E}_\mu[f(Y)] = \mathbb{E}_\mu[f(X + I(u(X)))]$$

$$\mathbb{E}_\mu[f(X)] = \mathbb{E}_\rho[f(X + I(u(X)))]$$

Now, using the definition of the relative entropy $H(\nu|\mu)$ we have (by the above equalities)

$$H(\nu|\mu) = \sup_{\varphi\in L^\infty(\Omega)} (\nu(\varphi) - \log\mu(e^\varphi)) = \sup_{\varphi\in L^\infty(\Omega)} (\mathbb{E}_\nu[\varphi(X)] - \log\mathbb{E}_\mu[e^{\varphi(X)}])$$

$$= \sup_{\varphi\in L^\infty(\Omega)} \left(\mathbb{E}_\mu\left[\underbrace{\varphi(X+I(u(X)))}_{\psi(X)}\right] - \log\mathbb{E}_\rho\left[\underbrace{e^{\varphi(X+I(u(X)))}}_{e^{\psi(X)}}\right]\right)$$

$$\leqslant \sup_{\psi\in L^\infty(\Omega)} (\mathbb{E}_\mu[\psi(X)] - \log\mathbb{E}_\rho[e^{\psi(X)}]) = H(\mu|\rho) = \int_\Omega \log\frac{\mathrm{d}\mu}{\mathrm{d}\rho}\mathrm{d}\mu = -\mathbb{E}_\mu\left[\log\frac{\mathrm{d}\rho}{\mathrm{d}\mu}\right]$$

$$= \mathbb{E}_\mu\left[\int_0^\infty u_s(X)\mathrm{d}X_s + \frac{1}{2}\int_0^\infty |u_s|^2\mathrm{d}s\right] = \mathbb{E}_\mu\left[\frac{1}{2}\int_0^\infty |u_s|^2\mathrm{d}s\right]$$

since under μ X is a Brownian motion and $M_t = \int_0^t u_s(X)\mathrm{d}X_s$ a square integrable martingale up to $t=+\infty$. This proves the formula for $\|u\|_{\mathbb{H}}$ bounded. In general case one has to use stopping times τ_n and approximate drifts $u_s^n = 1_{\tau_n\leqslant s}u_s$ stopped as soon as $\int_0^{\tau_n}|u_s|^2\mathrm{d}s = n$ and then taking limits as $n\to\infty$. Moreover one has to consider also the possibility that $\mathbb{E}_\mu[\|u(X)\|_{\mathbb{H}}^2] = +\infty$. In order to pass to the limit one uses the lower semicontinuity of the entropy, i.e. if $\nu_n\to\nu$ weakly then $H(\nu|\mu)\leqslant \liminf_n H(\nu_n|\mu)$. Details are left to reader. (They are not necessary for the exam). □

Lemma 9. *Let ν be a probability measure which is absolutely continuous wrt. μ with density Z such that $Z\in\mathscr{C}$ (defined last week) and $Z\geqslant\delta$ for some $\delta>0$. Let us call $\mathscr{S}_\mu\subseteq\Pi(\Omega)$ the set of all such measures. Then under $\nu\in\mathscr{S}_\mu$ the canonical process X is a strong solution of the SDE*

$$\mathrm{d}X_t = u_t(X)\mathrm{d}t + \mathrm{d}W_t, \qquad t\geqslant 0$$

where W is a ν-Brownian motion and u a drift such that

$$\|u_t(x) - u_t(y)\| \leqslant L\|x-y\|_{C([0,t];\mathbb{R}^d)} \qquad x,y\in\Omega \tag{3}$$

for some finite constant L. Moreover

$$H(\nu|\mu) = \frac{1}{2}\mathbb{E}_\nu\|u(X)\|_{\mathbb{H}}^2.$$

Proof. Define the adapted process $Z_t(X) := \mathbb{E}[Z|\mathscr{F}_t]$ by the martingale representation theorem we have that

$$Z_t(X) = 1 + \int_0^t F_s(X)\mathrm{d}X_s, \qquad t\geqslant 0$$

where since $Z\in\mathscr{C}$ we can compute explicitly both $Z_t(x)$ and $F_t(x)$ as functions of $x\in\Omega$, respectively as linear combinations of random variables of the form

$$\sum_{k=0}^n \sum_{\sigma\in S_n} V_t^{\sigma,k}(x)e^{-\alpha(\sigma,k)t}U^{\alpha(\sigma,k)}(H^{\sigma,k})(x_t), \qquad \sum_{k=0}^n \sum_{\sigma\in S_n} V_t^{\sigma,k}(x)e^{-\alpha(\sigma,k)t}\nabla U^{\alpha(\sigma,k)}(H^{\sigma,k})(x_t) \tag{4}$$

where the important point is that the functions $V_t^{\sigma,k}(x)$ are smooth functionals of $x\in\Omega$ (a sequence of iterated integrals in time of nice smooth functions of values of the path x at various times) and where $U^{\alpha(\sigma,k)}(H^{\sigma,k})$ are smooth functions on $\mathbb{R}^d$.

Moreover we also have $Z_t(X) \geqslant \varepsilon$ since $Z \geqslant \varepsilon$ and conditional expectation preserves this inequality. We will assume that is also true that $Z_t(x) \geqslant \varepsilon$ for all $x \in \Omega$. So it is not difficult to prove that if we let

$$u_t(x) := \frac{F_t(x)}{Z_t(x)}, \qquad x \in \Omega$$

then it satisfies the Lipshitz bound (3) and moreover

$$Z_t(X) = 1 + \int_0^t Z_s(X) u_s(X) \mathrm{d}X_s,$$

which implies that

$$Z = \mathscr{E}\left(\int_0^{\cdot} u_s(X) \mathrm{d}X_s \right)_{\infty}.$$

So by Girsanov's theorem, under the measure $\mathrm{d}\nu = Z \mathrm{d}\mu$ the process $W = X - I(u)$ is a Brownian motion, namely X satisfies the SDE

$$\mathrm{d}X_t = u_t(X) \mathrm{d}t + \mathrm{d}W_t, \qquad t \geqslant 0.$$

Given the Lipschitz bound on u, this SDE has a pathwise unique solution which is strong by the Yamada-Watanabe theorem. We denote by $X = \Phi(W)$ the strong solution, where $\Phi: \Omega \to \Omega$ is the solution map which is adapted. Finally,

$$H(\nu|\mu) = \mathbb{E}_{\nu}\left[\log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right] = \mathbb{E}_{\nu}\left[\int_0^{\infty} u_s(X) \mathrm{d}X_s - \frac{1}{2} \int_0^{\infty} |u_s(X)|^2 \mathrm{d}s \right]$$

$$= \mathbb{E}_{\nu}\left[\int_0^{\infty} u_s(X) \mathrm{d}W_s + \frac{1}{2} \int_0^{\infty} |u_s(X)|^2 \mathrm{d}s \right] = \mathbb{E}_{\nu}\left[\frac{1}{2} \int_0^{\infty} |u_s(X)|^2 \mathrm{d}s \right].$$

The fact that the drift satisfies $\frac{1}{2} \int_0^{\infty} |u_s(X)|^2 \mathrm{d}s \leqslant K$ for some K is left as exercise (this needs to use the exponential decay in time of the contributions of the form (4). □

Recall that

$$\log \mu[e^f] = \sup_{\nu} [\nu(f) - H(\nu|\mu)]$$

Lemma 10. *Let $f: \Omega \to \mathbb{R}$ which is measurable and bounded from below. Assume $\mu(e^f) < \infty$. For every $\varepsilon > 0$ there exists $\nu \in \mathscr{S}_{\mu}$ such that*

$$\log \mu[e^f] \leqslant \nu(f) - H(\nu|\mu) + \varepsilon.$$

If $\mu(e^f) = +\infty$ then there exist a sequence $(\nu_n) \subseteq \mathscr{S}_{\mu}$ such that

$$+\infty = \log \mu[e^f] = \sup_n (\nu_n(f) - H(\nu_n|\mu)).$$

Proof. We start by assuming that $\log \mu[e^f] < \infty$. By monotone convergence it is enough to consider only bounded functions f and moreover such that $\mu[e^f] = 1$. Indeed if f is bounded below I can introduce $f_n = (f \wedge n)$ which is now a bounded function for any n and if we prove the claim for bounded functions then we have that for any n and $\varepsilon > 0$ we have

$$\log \mu[e^{f_n}] \leqslant \nu_n(f_n) - H(\nu_n|\mu) + \varepsilon / 2$$

for some ν_n. But then we observe that $f_n \leqslant f$ so

$$\log \mu[e^{f_n}] \leqslant \nu_n(f) - H(\nu_n|\mu) + \varepsilon/2.$$

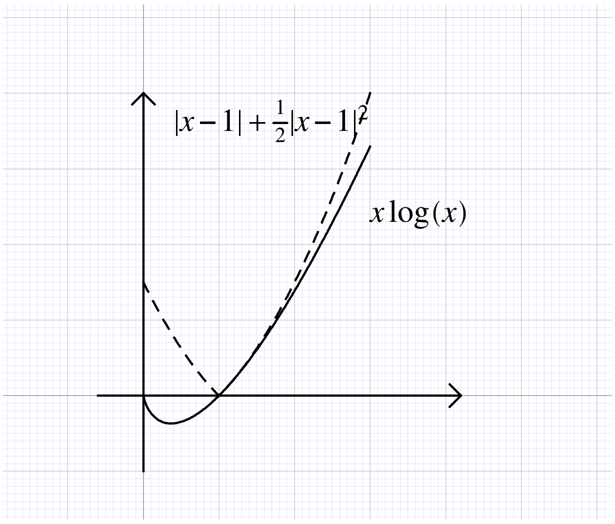
Moreover by monotone convergence we have $\log \mu[e^{f_n}] \to \log \mu[e^f]$. Then there exist n finite such that $\log \mu[e^f] \leqslant \log \mu[e^{f_n}] + \varepsilon/2$ and in this case we are done since

$$\log \mu[e^f] \leqslant \log \mu[e^{f_n}] + \varepsilon/2 \leqslant \nu_n(f) - H(\nu_n|\mu) + \varepsilon.$$

Note also that

$$\log \mu[e^{f-c}] - \nu(f-c) = \log \mu[e^f] - \nu(f)$$

so this shows that we can take c such that $\log \mu[e^{f-c}] = 0$, namely we can assume that f is such that $\mu[e^f] = 1$. Let $F = e^f$ and let ν be a probability measures on Ω. Note that



$$x \log(x) \leqslant |x-1| + \frac{1}{2}|x-1|^2, \qquad x \geqslant 0,$$

and using this we get

$$\mathrm{H}(\nu|\mu) - \nu(f) = \int_\Omega \left(\log\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)\right] - f(\omega) \right) \nu(\mathrm{d}\omega)$$

$$= \int_\Omega \left(\log\left[\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)\right] - \log F(\omega) \right) \nu(\mathrm{d}\omega) = \int_\Omega \left(\log\left[\frac{1}{F(\omega)}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)\right] \right) \nu(\mathrm{d}\omega)$$

$$= \int_\Omega \left(\log\left[\frac{1}{F(\omega)}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)\right] \right) \left(\frac{1}{F(\omega)}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega) \right) F(\omega)\, \mu(\mathrm{d}\omega)$$

$$= \int_\Omega \left(\log\left[\frac{G(\omega)}{F(\omega)}\right] \right) \left(\frac{G(\omega)}{F(\omega)} \right) F(\omega)\, \mu(\mathrm{d}\omega)$$

where $G = \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in \mathscr{C}$ since $\nu \in \mathscr{S}_\mu$. Using the inequality above we get

$$\mathrm{H}(\nu|\mu) - \nu(f) \leqslant \int_\Omega \left[\left|\frac{G}{F} - 1\right| + \frac{1}{2}\left|\frac{G}{F} - 1\right|^2 \right] F(\omega)\, \mu(\mathrm{d}\omega) \leqslant \|F - G\|_{L^1(\mu)} + C_f \|F - G\|_{L^2(\mu)}^2$$

where the constant C_f depends only on the lower bound on f. Moreover $\|F-G\|_{L^1(\mu)} \leqslant \|F-G\|_{L^2(\mu)}$. This proves that $\mathrm{H}(\nu|\mu)-\nu(f)$ can be made as small as we want since $\mathscr{C}$ is dense in $L^2(\mu)$ and we can always find $G \in \mathscr{C}$ such that $G \geqslant \delta$ and $\|e^f - G\|_{L^2(\mu)} \leqslant \varepsilon$.

If $\log \mu[e^f] = +\infty$ the above argument allows to conclude the existence of the claimed sequence by using f_n as lower bound of f. □

Now we are going to complete the proof of the Theorem 4

Proof of Theorem 4. We are going to prove that we have $\leqslant$ with an arbitrarily small loss ε and then that we have also the reverse inequality. Recall that we proved that if u is a drift and ν is the law of $X + I(u)$ then we have

$$H(\nu|\mu) \leqslant \frac{1}{2}\mathbb{E}_\mu[\|u(X)\|_{\mathbb{H}}^2]$$

then using this measure ν in the variational characterisation of $\log \mathbb{E}_\mu[e^f]$ we have

$$\log \mathbb{E}_\mu[e^f] = \sup_\rho (\rho(f) - H(\rho|\mu)) \geqslant \nu(f) - H(\nu|\mu)$$

$$\geqslant \nu(f) - \frac{1}{2}\mathbb{E}_\mu[\|u(X)\|_{\mathbb{H}}^2] = \mathbb{E}_\mu\left(f(X + I(u(X))) - \frac{1}{2}\|u\|_{\mathbb{H}}^2\right)$$

so we have one of the bounds because we can now optimize over all drifts u. In order to prove the reverse inequality we use the Lemma 10. Assume that $\log \mathbb{E}_\mu[e^f] < \infty$. For any $\varepsilon > 0$ there exists $\nu \in \mathscr{S}_\mu$ satisfying

$$\log \mathbb{E}_\mu[e^f] - \varepsilon \leqslant \nu(f) - H(\nu|\mu)$$

Now recall by Lemma 9 under ν the canonical process satisfies the SDE $\mathrm{d}X = z(X)\mathrm{d}t + \mathrm{d}W$ for a “nice” drift z (which is Lipshitz) and a process W which is a Brownian motion under ν. This SDE has a unique strong solution, so we can write $X = \Phi(W)$ with some adapted functional Φ. Therefore we concolude that

$$X = W + I(z(X)) = W + I(u(W))$$

where we let $u(x) = z(\Phi(x))$ for all $x \in \Omega$. With this new expression we have that

$$\nu(f) = \mathbb{E}_\nu(f(X)) = \mathbb{E}_\nu(f(W + I(z(X)))) = \mathbb{E}_\nu(f(W + I(u(W)))) = \mathbb{E}_\mu(f(X + I(u(X))))$$

since $\mathrm{Law}_\nu(W) = \mathrm{Law}_\mu(X)$. Moreover we have also (for similar reasons)

$$H(\nu|\mu) = \frac{1}{2}\mathbb{E}_\nu\|z(X)\|_{\mathbb{H}}^2 = \frac{1}{2}\mathbb{E}_\nu\|z(\Phi(W))\|_{\mathbb{H}}^2 = \frac{1}{2}\mathbb{E}_\nu\|u(W)\|_{\mathbb{H}}^2 = \frac{1}{2}\mathbb{E}_\mu\|u(X)\|_{\mathbb{H}}^2.$$

Therefore putting pieces together we have

$$\log \mathbb{E}_\mu[e^f] - \varepsilon \leqslant \nu(f) - H(\nu|\mu) = \mathbb{E}_\mu(f(X + I(u(X)))) - \frac{1}{2}\mathbb{E}_\mu \|u(X)\|^2_{\mathbb{H}}.$$

So, for any $\varepsilon > 0$ we have found a particular drift u such that

$$\log \mathbb{E}_\mu[e^f] \leqslant \mathbb{E}_\mu(f(X + I(u(X)))) - \frac{1}{2}\mathbb{E}_\mu \|u(X)\|^2_{\mathbb{H}} + \varepsilon.$$

While if $\log \mathbb{E}_\mu[e^f] = +\infty$ then by the same lemma one has that there exists a sequence of drifts $(u_n)_{n\geqslant 1}$ such that

$$+\infty = \log \mathbb{E}_\mu[e^f] = \sup_n \left[\mathbb{E}_\mu(f(X + I(u_n(X)))) - \frac{1}{2}\mathbb{E}_\mu \|u_n(X)\|^2_{\mathbb{H}}\right].$$

In both casesn putting together the two inequalities we conclude that

$$\log \mathbb{E}_\mu[e^f] = \sup_u \left[\mathbb{E}_\mu(f(X + I(u(X)))) - \frac{1}{2}\mathbb{E}_\mu \|u(X)\|^2_{\mathbb{H}}\right]$$

which is our claim. □

Applications to functional analysis

This formula and similar formulas can be used (amazingly) to prove functional inequalities for finite dimensional measures, see for example

- Lehec, Joseph. "Representation Formula for the Entropy and Functional Inequalities." *Annales de l'Institut Henri Poincaré Probabilités et Statistiques* 49, no. 3 (2013): 885–899.
- Lehec, Joseph. . "Short Probabilistic Proof of the Brascamp-Lieb and Barthe Theorems." *Canadian Mathematical Bulletin* 57, no. 3 (September 1, 2014): 585–97. https://doi.org/10.4153/CMB-2013-040-x.
- Borell, Christer. "Diffusion Equations and Geometric Inequalities." *Potential Analysis. An International Journal Devoted to the Interactions between Potential Theory, Probability Theory, Geometry and Functional Analysis* 12, no. 1 (2000): 49–71. https://doi.org/10.1023/A:1008641618547.
- Handel, Ramon van. "The Borell–Ehrhard Game." *Probability Theory and Related Fields* 170, no. 3–4 (April 2018): 555–85. https://doi.org/10.1007/s00440-017-0762-4.
- Hariya, Yuu, and Sou Watanabe. "The Bouè–Dupuis Formula and the Exponential Hypercontractivity in the Gaussian Space." *ArXiv:2110.14852 [Math]*, November 3, 2021. http://arxiv.org/abs/2110.14852.

We will not look into these, but they are very interesting.

3 Applications to probabilitistic problems

Gaussian bounds on functional of Brownian motion.

Theorem 11. *Let* (E, d) *a metric space and* $f: \Omega \to E$ *such that there an* $e \in E$ *for which*

$$d(f(x + I(h)), e) \leqslant c(x)(g(x) + \|h\|_{\mathbb{H}}), \qquad h \in \mathbb{H},$$

for μ*-almost every* $x \in \Omega$ *where* $\mu(cg) < \infty$ *and* $\mu(c^2) < \infty$. *Then for all* $\lambda > 0$ *we have*

$$\mathbb{E}_\mu[e^{\lambda d(f(X), e)}] \leqslant e^{\lambda^2 \mu(c^2) + \lambda \mu(cg)}.$$

In particular the r.v. $d(f(X), e)$ *has Gaussian tails, i.e.*

$$\mathbb{P}_\mu(d(f(X), e) > k) \lesssim C_1 e^{-C_2 k^2}$$

for some $C_1, C_2 > 0$.

Remark 12. Note that if we let $y = x + I(h)$ then $y(t) = x(t) + \int_0^t h(s) \mathrm{d}s$. Note that the natural norm on y is given by the sup norm, i.e.

$$\|y\|_{C([0,1], \mathbb{R}^d)} = \sup_{t \in [0,1]} \left| x(t) + \int_0^t h(s) \mathrm{d}s \right|$$

but on the r.h.s. of the inequality you have to control the L^2 norm of h which corresponds to the H^1 norm of $I(h)$, i.e.

$$\|h\|_{\mathbb{H}} = \|I(h)\|_{\dot{H}^1(\mathbb{R}_+, \mathbb{R}^d)} = \left\| \frac{\mathrm{d}}{\mathrm{d}t} I(h) \right\|_{L^2(\mathbb{R}_+, \mathbb{R}^d)}.$$

This is coherent with the fact that increments of Brownian motion are independent so formally the Wiener measure can be understood as given by

$$\mu(\mathrm{d}\omega) \propto \exp\left(-\frac{1}{2}\int_0^\infty |\dot{\omega}(s)|^2 \mathrm{d}s\right) \mathrm{D}\omega.$$

Proof. By Boué–Dupuis formula and the hypothesis on f

$$\begin{aligned}
\log \mathbb{E}_\mu[e^{\lambda d(f(X), e)}] &= \sup_u \mathbb{E}_\mu\left[\lambda d(f(X + I(u)), e) - \frac{1}{2}\|u\|_{\mathbb{H}}^2\right] \\
&\leqslant \sup_u \mathbb{E}_\mu\left[\lambda c(X)(g(X) + \|u\|_{\mathbb{H}}) - \frac{1}{2}\|u\|_{\mathbb{H}}^2\right]
\end{aligned}$$

We observe now that the polynomial $\lambda c(X)(g(X)+t)-\frac{1}{2}t^2$ is upperbounded by

$$\lambda c(X)g(X)+\lambda c(X)t-\frac{1}{2}t^2 \leqslant \lambda c(X)g(X)+\frac{1}{2}\lambda^2 c(X)^2-\frac{1}{2}\underbrace{(t-\lambda c(X))^2}_{\geqslant 0} \leqslant \lambda c(X)g(X)+\frac{1}{2}\lambda^2 c(X)^2$$

therefore

$$\log \mathbb{E}_\mu[e^{\lambda d(f(X),e)}] \leqslant \sup_u \mathbb{E}_\mu\left[\lambda c(X)g(X)+\frac{1}{2}\lambda^2 c(X)^2\right] = \mathbb{E}_\mu\left[\lambda c(X)g(X)+\frac{1}{2}\lambda^2 c(X)^2\right]$$

$$=\lambda\mu(cg)+\frac{1}{2}\lambda^2\mu(c^2).$$

□

Exercise 1. Take

$$f(x)=\sup_{t,s\in[0,1]}\frac{|x(t)-x(s)|}{|t-s|^\alpha}$$

and prove that is satisfies the hypothesis of the previous theorem. Conclude that

$$\mathbb{E}_\mu\left[\exp\left(\lambda \sup_{t,s\in[0,1]}\frac{|X(t)-X(s)|}{|t-s|^\alpha}\right)\right] \leqslant e^{C_1\lambda^2+C_2\lambda}$$

for any $\alpha\in(0,1/2)$ any $\lambda>0$. From this you can also conclude that

$$\mathbb{E}_\mu\left[\exp\left(\rho\left(\sup_{t,s\in[0,1]}\frac{|X(t)-X(s)|}{|t-s|^\alpha}\right)^2\right)\right]<\infty$$

for some $\rho>0$ small.

---

4 Large deviations of diffusion

The goal will be now to understand what happens when we have a family of SDEs in $\mathbb{R}^d$ of the form

$$\mathrm{d}X_t^\varepsilon = b(X_t^\varepsilon)\mathrm{d}t+\varepsilon^{1/2}\sigma(X_t^\varepsilon)\mathrm{d}B_t, \qquad X^\varepsilon = x_0\in\mathbb{R}^d$$

with ε a small parameter and B a d-dimensional BM. Let's assume the coefficient b: $\mathbb{R}^d\to\mathbb{R}^d$, σ: $\mathbb{R}^d\to\mathscr{L}(\mathbb{R}^d,\mathbb{R}^d)$ are nice (bounded and Lipshitz) so that we have a strong solution. We would like to understand how the law μ^ε of X^ε looks like as $\varepsilon\to 0$.

Is not difficult to prove that $(\mu^\varepsilon)_\varepsilon$ converges in law (as probability measures on $\mathscr{C}^d=C(\mathbb{R}_+;\mathbb{R}^d)$ with its Borel σ-field) to the Dirac mass μ^0 concentrated on the solution x^0 of the ODE

$$\dot{X}_t^0 = b(X_t^0), \qquad X_0^0 = x_0.$$

(for example proving that $\mathbb{E}[\sup_{t\in[0,T]}|X_t^\varepsilon - X_t^0|^2] \to 0$ and conclude from this).

In large deviations theory one is concerned with the *speed* with which $\mu^\varepsilon \to \mu^0$, namely one would like to quantify this convergence and usually it will happen that this convergence is exponential, in the sense that

$$\mathbb{P}(X^\varepsilon \in A) \approx e^{-r(\varepsilon)C(A)}$$

where $r(\varepsilon) \to \infty$ as $\varepsilon \to 0$ nd it is usually something like $\varepsilon^{-\alpha}$ and $C(A)$ is a constant which depends only on the particular set A.

For example we could ask $A_{\gamma,T,\delta} = \{\omega \in \mathscr{C}^d : \sup_{t\in[0,T]}|\omega(t) - \gamma(t)| < \delta\}$ for given $\gamma \in \mathscr{C}^d$, $\delta > 0$ and $T > 0$. In this case if $\sup_{t\in[0,T]}|\omega(t) - X^0(t)| > \delta$ then $X^0 \notin A_{\gamma,T,\delta}$ and $\mu^\varepsilon(A_{\gamma,T,\delta}) \to 0$. We are going to prove that what will happen is that

$$\varepsilon \log \mu^\varepsilon(A_{\gamma,T,\delta}) = \varepsilon \log \mathbb{P}(X^\varepsilon \in A_{\gamma,T,\delta}) \approx -\inf_{x\in A_{\gamma,T,\delta}} I(x)$$

where I is a function which is only depending on b, σ and on the original problem and is called a *rate function*. They are called large deviations because they happen on an exponential scale. Otherwise stated we have an explitic asymptotic formula for the probability which looks like

$$\mu^\varepsilon(B) \approx e^{-\frac{1}{\varepsilon}\inf_{x\in B} I(x)}$$

Large Deviation Theory is concerned in general in the study of such large fluctuations in a variety of contextes (deviations from the law of large numbers, deviations from the ergodic behaviour, deviations from small noise behaviour like in this case, deviations from the large sample behaviour in statistics).

In order to properly speak about large deviations for the SDEs above we need some standard definitions from large deviation theory.

Definition 13. *A function $I: \mathscr{E} \to [0, +\infty]$ is called a (good) rate function on a Polish space $\mathscr{E}$ if the sets $I^{-1}[0, M] = \{x \in \mathscr{E} : I(x) \leqslant M\} \subset \mathscr{E}$ are compact for all $M < +\infty$.*

In particular, a rate function is always lower semicontinuous.

Definition 14. *Let I be a rate function on a Polish space $\mathscr{E}$ and $(Y^\varepsilon)_{\varepsilon>0}$ a family of random variables with values in $\mathscr{E}$. The this family satisfies the Laplace principle on $\mathscr{E}$ with rate function I (and rate $1/\varepsilon$) if for any function $h \in C_b(\mathscr{E})$ (bounded and continuous) we have*

$$\lim_{\varepsilon\to 0} -\varepsilon \log \mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] = \inf_{x\in\mathscr{E}} [I(x) + h(x)]. \tag{5}$$

A Laplace principle is telling us that the law μ^{ε} of Y^{ε} is behaving like $e^{-I(x)/\varepsilon}$, in the sense that

$$\mathbb{E}[e^{-h(Y^{\varepsilon})/\varepsilon}] = \int e^{-h(x)/\varepsilon}\mu_{\varepsilon}(\mathrm{d}x) \approx \int e^{-h(x)/\varepsilon}e^{-I(x)/\varepsilon}\mathrm{d}x = \int e^{-(h(x)+I(x))/\varepsilon}\mathrm{d}x = e^{-\frac{1}{\varepsilon}\inf_{x\in\mathscr{E}}[I(x)+h(x)](1+o(1))}.$$

Definition 15. *A family $(Y^{\varepsilon})_{\varepsilon>0}$ satisfies the Large Deviation principle on $\mathscr{E}$ with rate function I (and rate $1/\varepsilon$) if for any open set $A\in\mathscr{E}$ and closed set $B\in\mathscr{E}$ we have*

$$\liminf_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(Y^{\varepsilon}\in A) \geqslant -\inf_{x\in A} I(x),$$

$$\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}(Y^{\varepsilon}\in B) \leqslant -\inf_{x\in B} I(x).$$

Remark 16. Recall that if $\mu^{\varepsilon}\to\mu$ weakly, then the Portmanteau theorem asserts that for any open set A and closed set B you have

$$\liminf_{\varepsilon\to 0}\mu^{\varepsilon}(A)\geqslant\mu(A), \qquad \limsup_{\varepsilon\to 0}\mu^{\varepsilon}(B)\leqslant\mu(B)$$

while if $f\in C_b(\mathscr{E})$ the of course

$$\lim_{\varepsilon\to 0}\int f(x)\,\mu^{\varepsilon}(\mathrm{d}x) = \int f(x)\,\mu(\mathrm{d}x).$$

There are very strong similarities between weak convergence and large deviations.

Theorem 17. *The Laplace principle is equivalent to the Large Deviation principle.*

Proof. Exercise. □

Now we are going to use the Boué–Dupuis formula to prove large deviations for a large class of problems which in particular include the small noise diffusion problem introduced above.

Let $(Y^{\varepsilon})_{\varepsilon>0}$ a family of random variables defined on a Wiener space $(\Omega,\mathscr{F},\mathbb{W},X)$ with $\mathbb{W}$ the Wiener measure and taking values in $\mathscr{E}$ which are obtained from X using a family of mappings $\mathscr{G}^{\varepsilon}:\Omega\to\mathscr{E}$ i.e. $Y^{\varepsilon}=\mathscr{G}^{\varepsilon}(X)$.

Let $\mathbb{U}_M\subseteq L^2(\mathbb{R}_{\geqslant 0};\mathbb{R}^d)$ the subset of elements $u\in L^2(\mathbb{R}_{\geqslant 0};\mathbb{R}^d)$ such that $\|u\|_{\mathbb{H}}\leqslant M$ and let $\mathscr{U}_M\subseteq L^2_{\mathscr{P}}(\mathbb{R}_{\geqslant 0}\times\Omega;\mathbb{R}^d)$ the subset of drifts $u\in L^2_{\mathscr{P}}(\mathbb{R}_{\geqslant 0}\times\Omega;\mathbb{R}^d)$ such that $\|u\|_{\mathbb{H}}\leqslant M$ holds μ-almost surely, i.e. $u(\cdot,\omega)\in\mathbb{U}_M$ for μ almost every $\omega\in\Omega$.

Note that $\mathbb{U}_M$ is a compact Polish space with respect to the weak topology of $L^2(\mathbb{R}_{\geqslant 0};\mathbb{R}^d)$ (by Banach-Alaoglu theorem).

We define $J(u)(t) := \int_0^t u(s)\mathrm{d}s$ for any $u\in\mathbb{H}=L^2(\mathbb{R}_{\geqslant 0};\mathbb{R}^d)$ and then $J: L^2(\mathbb{R}_{\geqslant 0};\mathbb{R}^d)\to C(\mathbb{R}_{\geqslant 0};\mathbb{R}^d)=\Omega$.

We will make the following assumptions on the family $(\mathscr{G}^\varepsilon)_{\varepsilon>0}$.

Hypothesis 18. *There exists a measurable mapping $\mathscr{G}^0: \Omega \to \mathscr{E}$ such that the following holds*

a) for every $M < \infty$ and any family $(u^\varepsilon)_\varepsilon \subseteq \mathscr{U}_M$ such that $(u^\varepsilon)_\varepsilon$ converges in law (as a random element of $\mathbb{U}_M$, and with the weak topology of $L^2(\mathbb{R}_{\geqslant 0}; \mathbb{R}^d)$) to u we have that

$$\mathscr{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^\varepsilon)) \to \mathscr{G}^0(J(u))$$

in law as random variables (on $(\Omega, \mathscr{F}, \mathbb{W})$) with values in $\mathscr{E}$ (of course as $\varepsilon \to 0$).

b) for every $M < \infty$ the set $\Gamma_M := \{\mathscr{G}^0(J(u)) : u \in \mathbb{U}_M\}$ is a compact subset of $\mathscr{E}$.

For each $x \in \mathscr{E}$ we define

$$I(x) := \frac{1}{2} \inf_{u \in \Gamma(x)} \|u\|_{\mathbb{H}}^2 \tag{6}$$

where the infimum is take over the set $\Gamma(x) \subseteq \mathbb{H} = L^2(\mathbb{R}_{\geqslant 0}; \mathbb{R}^d)$ such that $x = \mathscr{G}^0(J(u))$ and is taken to be $+\infty$ if this set is empty.

Lemma 19. *Under the Hypothesis 18 the function I is a rate function.*

Proof. (exercise) □

Theorem 20. *Under the Hypothesis 18 the family $(Y^\varepsilon = \mathscr{G}^\varepsilon(X))_{\varepsilon>0}$ satisfies the Laplace principle with rate function I as defined in (6) and speed $1/\varepsilon$.*

Proof. We need to show that

$$\lim_{\varepsilon \to 0} -\varepsilon \log \mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] = \inf_{x \in \mathscr{E}} [I(x) + h(x)]$$

holds for any $h \in C_b(\mathscr{E})$.

Lower bound. By Boué–Dupuis formula we have

$$-\varepsilon \log \mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] = -\varepsilon \log \mathbb{E}[e^{-h(\mathscr{G}^\varepsilon(X))/\varepsilon}] = \inf_u \mathbb{E}\left[h(\mathscr{G}^\varepsilon(X + J(u))) + \frac{1}{2}\|\varepsilon^{1/2} u\|_{\mathbb{H}}^2\right]$$

By renaming $u \to \varepsilon^{-1/2} u$ we have

$$-\varepsilon \log \mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] = \inf_u \mathbb{E}\left[h(\mathscr{G}^\varepsilon(X + \varepsilon^{-1/2} J(u))) + \frac{1}{2}\|u\|_{\mathbb{H}}^2\right].$$

Fix $\delta > 0$. Then for any $\varepsilon > 0$ there exists an approximate minimiser u^ε such that

$$-\varepsilon \log \mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] \geqslant \mathbb{E}\left[h(\mathcal{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^\varepsilon))) + \frac{1}{2}\|u^\varepsilon\|_{\mathbb{H}}^2\right] - \delta.$$

This implies in particular that

$$\mathbb{E}\left[\frac{1}{2}\|u^\varepsilon\|_{\mathbb{H}}^2\right] \leqslant \delta - \varepsilon \log \mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] + \|h\|_{C_b(\mathscr{E})} \leqslant \delta + 2\|h\|_{C_b(\mathscr{E})} < \infty,$$

and this bound is independent of ε.

Moreover taking N large enough we can replace u^ε by the stopped process $u_t^{\varepsilon,N} = u_t^\varepsilon \mathbb{1}_{t \leqslant \tau_{\varepsilon,N}}$ with

$$\tau_{\varepsilon,N} := \inf\{t \geqslant 0 : \|u^\varepsilon \mathbb{1}_{[0,t]}\|_{\mathbb{H}} \geqslant N\}.$$

In this case $u_t^{\varepsilon,N} \in \mathscr{U}_N$ and morever we have that

$$\mathbb{P}(u^\varepsilon \neq u^{\varepsilon,N}) \leqslant \mathbb{P}(\|u^\varepsilon\|_{\mathbb{H}} > N) \leqslant \frac{\mathbb{E}[\|u^\varepsilon\|_{\mathbb{H}}^2]}{N^2} \leqslant \frac{2\delta + 4\|h\|_{C_b(\mathscr{E})}}{N^2}$$

uniformly in ε. This implies that we can choose N uniformly in ε so that

$$|\mathbb{E}[h(\mathcal{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^\varepsilon)))] - \mathbb{E}[h(\mathcal{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^{\varepsilon,N})))]|$$

$$\leqslant \mathbb{E}|h(\mathcal{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^\varepsilon))) - h(\mathcal{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^{\varepsilon,N})))|$$

$$\leqslant 2\|h\|_{C_b(\mathscr{E})} \mathbb{P}(u^\varepsilon \neq u^{\varepsilon,N}) = 2\|h\|_{C_b(\mathscr{E})} \frac{2\delta + 4\|h\|_{C_b(\mathscr{E})}}{N^2} \leqslant \delta.$$

Of course we have also $\mathbb{E}\left[\frac{1}{2}\|u^\varepsilon\|_{\mathbb{H}}^2\right] \geqslant \mathbb{E}\left[\frac{1}{2}\|u^{\varepsilon,N}\|_{\mathbb{H}}^2\right]$ therefore we conclude that

$$-\varepsilon \log \mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] \geqslant \mathbb{E}\left[h(\mathcal{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^{\varepsilon,N}))) + \frac{1}{2}\|u^{\varepsilon,N}\|_{\mathbb{H}}^2\right] - 2\delta.$$

Now, we have $\|u^{\varepsilon,N}\|_{\mathbb{H}} \leqslant N$ by construction almost surely and for any $\varepsilon > 0$. Therefore from any subsequence of $(u^{\varepsilon,N})_\varepsilon$ we can extract a weakly converging subsequence $(u^{\varepsilon_n,N})_n$ and let $u \in \mathscr{U}_N$ be its limit. Using Hypothesis 18 we have that $\mathcal{G}^\varepsilon(X + \varepsilon_n^{-1/2} J(u^{\varepsilon_n,N}))$ converges in law to $\mathcal{G}^0(J(u))$ and moreover by Fatou $\liminf_{n\to\infty} \mathbb{E}[\|u^{\varepsilon_n,N}\|_{\mathbb{H}}^2] \geqslant \mathbb{E}[\|u\|_{\mathbb{H}}^2]$ therefore (we use that h is a continuous function)

$$\liminf_{n\to\infty} \mathbb{E}\left[h(\mathcal{G}^{\varepsilon_n}(X + \varepsilon_n^{-1/2} J(u^{\varepsilon_n,N}))) + \frac{1}{2}\|u^{\varepsilon_n,N}\|_{\mathbb{H}}^2\right] \geqslant \mathbb{E}\left[h(\mathcal{G}^0(J(u))) + \frac{1}{2}\|u\|_{\mathbb{H}}^2\right]$$

$$\geqslant \inf_{v \in \mathbb{H}} \mathbb{E}\left[h(\mathcal{G}^0(J(v))) + \frac{1}{2}\|v\|_{\mathbb{H}}^2\right]$$

$$= \inf_{x \in \mathscr{E}} \inf_{v \in \Gamma(x)} \mathbb{E}\left[h(x) + \frac{1}{2}\|v\|_{\mathbb{H}}^2\right] = \inf_{x \in \mathscr{E}} \left(h(x) + \underbrace{\inf_{v \in \Gamma(x)} \mathbb{E}\left[\frac{1}{2}\|v\|_{\mathbb{H}}^2\right]}_{I(x)}\right) = \inf_{x \in \mathscr{E}} [I(x) + h(x)].$$

From this we conclude that

$$\liminf_{\varepsilon\to 0} -\varepsilon\log\mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] \geqslant \liminf_{\varepsilon\to 0} \mathbb{E}\left[h(\mathcal{G}^\varepsilon(X+\varepsilon^{-1/2}J(u^{\varepsilon,N}))) + \frac{1}{2}\|u^{\varepsilon,N}\|_{\mathbb{H}}^2\right] - 2\delta$$

$$\geqslant \inf_{x\in\mathcal{E}} [I(x) + h(x)] - 2\delta$$

because from any sequence we can extract a subsequence for which the bound works. This establish the lower bound since now δ is arbitrary and can be taken to zero.

Upper bound. By Boué–Dupuis formula for any $v\in\mathbb{H}$ (deterministic) we have

$$\limsup_{\varepsilon\to 0} -\varepsilon\log\mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] = \limsup_{\varepsilon\to 0}\inf_u \mathbb{E}\left[h(\mathcal{G}^\varepsilon(X+\varepsilon^{-1/2}J(u))) + \frac{1}{2}\|u\|_{\mathbb{H}}^2\right]$$

$$\leqslant \limsup_{\varepsilon\to 0} \mathbb{E}\left[h(\mathcal{G}^\varepsilon(X+\varepsilon^{-1/2}J(v))) + \frac{1}{2}\|v\|_{\mathbb{H}}^2\right]$$

$$= \left(\limsup_{\varepsilon\to 0} \mathbb{E}[h(\mathcal{G}^\varepsilon(X+\varepsilon^{-1/2}J(v)))]\right) + \frac{1}{2}\|v\|_{\mathbb{H}}^2$$

By Hypothesis $\mathcal{G}^\varepsilon(X+\varepsilon^{-1/2}J(v)) \to \mathcal{G}^0(J(v)) =: x_0$ in law, and $v\in\Gamma(x_0)$, therefore by optimizing over v we have

$$\limsup_{\varepsilon\to 0} -\varepsilon\log\mathbb{E}[e^{-h(Y^\varepsilon)/\varepsilon}] \leqslant \inf_{v\in\mathbb{H}}\left[h(\mathcal{G}^0(J(v))) + \frac{1}{2}\|v\|_{\mathbb{H}}^2\right]$$

$$= \inf_{x\in\mathcal{E}}\inf_{v\in\Gamma(x)}\left[h(\mathcal{G}^0(J(v))) + \frac{1}{2}\|v\|_{\mathbb{H}}^2\right] = \inf_{x\in\mathcal{E}}[I(x)+h(x)]$$

so we proved the claim. □

Example 21. We can take $\mathcal{E}=\Omega$ and $Y^\varepsilon = \mathcal{G}^\varepsilon(X) = \varepsilon^{1/2}X$. In this case note that we have the following convergence in law

$$\mathcal{G}^\varepsilon(X+\varepsilon^{-1/2}J(u^\varepsilon)) = \varepsilon^{1/2}X + J(u^\varepsilon) \to J(u)$$

therefore we can take $\mathcal{G}^0(x) = x$ and check that we fullfill Hypothesis 18. The theorem gives as a consequence that the family $(\varepsilon^{1/2}X)_\varepsilon$ satisfies the Laplace principle with rate function

$$I(x) = \inf_{v\in\Gamma(x)} \frac{1}{2}\|v\|_{\mathbb{H}}^2 = \inf_{v\in\mathbb{H}:x=J(v)} \frac{1}{2}\|v\|_{\mathbb{H}}^2 = \frac{1}{2}\int_0^\infty |\dot{x}(s)|^2\mathrm{d}s$$

if $x\in H^1(\mathbb{R}_{\geqslant 0};\mathbb{R}^d)$ (Sobolev space of functions with L^2 derivative) and $I(x)=+\infty$ otherwise. This follows from the fact that $x=J(v)$ means really that $x(t)=\int_0^t v(s)\mathrm{d}s$ for some $v\in L^2$. In the formula $\dot{x}(s)=v(s)$ denotes the derivative of x.

And as consequence it satisfies also the Large Deviation principle with the same rate function. This is called Schilder's theorem.

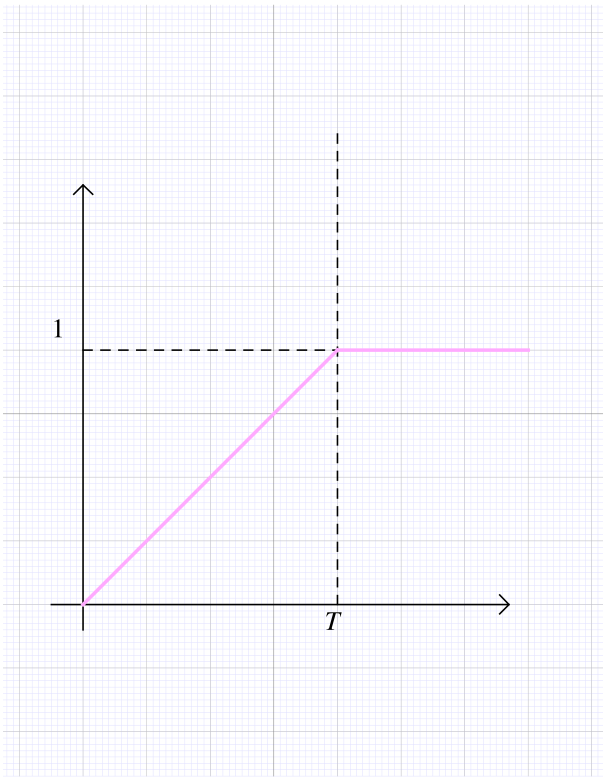
Theorem 22. *(Schilder's theorem) Let X be a Brownian motion, then $(\varepsilon^{1/2}X)_\varepsilon$ satisfies the large deviation principle on Ω with rate $1/\varepsilon$ and rate function given by*

$$I(x) = \begin{cases} \frac{1}{2}\int_0^\infty |\dot{x}(s)|^2 \mathrm{d}s & \text{if } x \in H^1, \\ +\infty & \text{otherwise.} \end{cases}$$

Example 23. This means in particular that if $L \to \infty$, by Schilder's theorem (using $\varepsilon^{1/2} = 1/L$)

$$\log \mathbb{P}\left(\sup_{t\in[0,T]} X_t \geqslant L\right) = \log \mathbb{P}\left(\sup_{t\in[0,T]} L^{-1}X_t \geqslant 1\right) = \log \mathbb{P}(L^{-1}X \in A) \approx -L^2 \inf_{x\in A} I(x)$$

where $A = \{\omega \in \Omega: \sup_{t\in[0,T]}\omega(t) \geqslant 1\}$ is a closed set. (here $\approx$ means appropriate upper and lower bounds for the closed set A and its interior).



Now the minimizer of the variational problem

$$\inf_{x\in A} I(x)$$

is easily seen to be (see image left) $x^*(t) = (1 \wedge (t/T))$ which gives

$$I(x^*) = \frac{1}{2}\left(\frac{1}{T}\right)^2 T = \frac{1}{2T}.$$

So we conclude that LD gives us the estimate

$$\log \mathbb{P}\left(\sup_{t\in[0,T]} X_t \geqslant L\right) \approx -\frac{L^2}{2T}$$

Exercise: let $f(t)$ be an arbitrary increasing function, try to estimate with Schilder's theorem for $L \to \infty$ the probability

$$\mathbb{P}\left(\sup_{t\geqslant 0} X_t - Lf(t) \geqslant 0\right)$$

for example when $f(t) = 1 + t^2$.

Let's now apply our large deviation statement to small noise diffusions. Let $\mathscr{E} = \Omega$. Assume $Y^\varepsilon = \mathscr{G}^\varepsilon(X)$ is the strong solution to the SDE

$$\mathrm{d}Y_t^\varepsilon = b(Y_t^\varepsilon)\mathrm{d}t + \varepsilon^{1/2}\mathrm{d}X_t, \qquad t \geqslant 0$$

for a Lipshitz drift $b: \mathbb{R}^d \to \mathbb{R}^d$ and a given initial condition $Y_0^\varepsilon = y_0 \in \mathbb{R}^d$. We have to identify $\mathscr{G}^0$. Recall that $\mathscr{G}^0$ is defined by having the property that we have the weak convergence

$$\mathscr{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^\varepsilon)) \to \mathscr{G}^0(J(u))$$

as soon as $u^\varepsilon \to u$ (in law, see above for precise conditions). Call $Z_t^\varepsilon = \mathscr{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^\varepsilon))$. Note that

$$\mathscr{G}^\varepsilon(X)(t) = Y_t^\varepsilon = y_0 + \int_0^t b(Y_s^\varepsilon) \mathrm{d}s + \varepsilon^{1/2} X_t, \qquad t \geqslant 0$$

so we can take $\mathscr{G}^\varepsilon: \Omega \to \mathscr{E}$ to be the unique mapping solving the integral equation

$$\mathscr{G}^\varepsilon(x) = y_0 + \int_0^t b(\mathscr{G}^\varepsilon(x)(s)) \mathrm{d}s + \varepsilon^{1/2} x(t).$$

Therefore

$$Z_t^\varepsilon = \mathscr{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^\varepsilon)) = y_0 + \int_0^t b\Big(\underbrace{\mathscr{G}^\varepsilon(X + \varepsilon^{-1/2} J(u^\varepsilon))(s)}_{Z_s^\varepsilon}\Big) \mathrm{d}s + \varepsilon^{1/2}(X(t) + \varepsilon^{-1/2} J(u^\varepsilon)(t))$$

$$= y_0 + \int_0^t b(Z_s^\varepsilon) \mathrm{d}s + \varepsilon^{1/2} X(t) + J(u^\varepsilon)(t)$$

so $(Z_t^\varepsilon)_{t \geqslant 0}$ is the solution to the SDE wih an additional drift term given by $J(u^\varepsilon)(t)$. One can then easily prove that $(Z^\varepsilon)_\varepsilon$ converges in $\mathscr{E}$ to the solution Z^0 of

$$Z_t^0 = y_0 + \int_0^t b(Z_s^0) \mathrm{d}s + J(u)(t)$$

Therefore we define $\mathscr{G}^0: \Omega \to \mathscr{E}$ to be the unique solution to

$$\mathscr{G}^0(x)(t) = y_0 + \int_0^t b(\mathscr{G}^0(x)(s)) \mathrm{d}s + x(t)$$

in such a way that $Z_t^0 = \mathscr{G}^0(J(u))(t)$ and the Hypothesis 18 can be then easily check. We conclude that the family of solutions $(Y^\varepsilon)_\varepsilon$ satisfies the Large Deviation principle with rate function

$$I(x) = \inf_{v \in \Gamma(x)} \frac{1}{2} \|v\|_{\mathbb{H}}^2$$

with $v \in \Gamma(x)$ iff $x = \mathscr{G}^0(J(u))$, that is x has to be the solution to the ODE

$$x(t) = y_0 + \int_0^t b(x(s)) \mathrm{d}s + J(v)(t)$$

meaning that

$$\dot{x}(t) = b(x(t)) + v(t)$$

and as a consequence there is at most one v such that $v \in \Gamma(x)$ and in this case

$$I(x) = \frac{1}{2}\|v\|_{\mathbb{H}}^2 = \frac{1}{2}\int_0^\infty |v(s)|^2 \mathrm{d}s = \frac{1}{2}\int_0^\infty |\dot{x}(s) - b(x(s))|^2 \mathrm{d}s$$

otherwise $I(x) = +\infty$. This is the rate function for small noise diffusion.

In the general case of a nondegenerate diffusion coefficients

$$\mathrm{d}Y_t^\varepsilon = b(Y_t^\varepsilon)\mathrm{d}t + \varepsilon^{1/2}\sigma(Y_t^\varepsilon)\mathrm{d}X_t, \qquad t \geqslant 0$$

one can prove that the LD rate function is in the form

$$I(x) = \frac{1}{2}\int_0^\infty |\sigma(x(s))^{-1}(\dot{x}(s) - b(x(s)))|^2 \mathrm{d}s.$$

---

5 Backward SDEs and non-linear PDEs

This section is based on the works:

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

Backward SDEs are a different kind of SDEs which have numerous applications:

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

Let's start by reminding the classical Feynman–Kac formula. Consider the first order differential operator

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

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

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

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

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

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

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

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

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

$$\partial_t u(t,x) + \mathscr{L}u(t,x) + f(t,x,u(t,x),\nabla u(t,x)) = 0 \qquad (7)$$

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

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

by using the PDE (7) we have

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

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

$$\partial_t u(t,x) + \mathscr{L}u(t,x) + f(t,x,u(t,x),\sigma(x)\nabla u(t,x)) = 0 \qquad (8)$$

It is clear that if σ is invertible then this PDE if equivalent to a PDE of the form (7), indeed we have

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

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

$$\mathrm{d}Y_s = -f(t,X_s^{t,x},Y_s,Z_s)\mathrm{d}s + Z_s\mathrm{d}W_s, \tag{9}$$

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

Exercise 2. Think about the theory we are going to develop below for the equations of the kind

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

in this case one would have $Z_s = \nabla u(t,X_s^{t,x})$ with the original formulation (7) of the PDE.

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

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

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

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

that is

$$\xi = Y_0 + \int_0^T Z_s\mathrm{d}W_s \tag{10}$$

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

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

with $Y_0 \hat{\in} \mathscr{F}_0$, let's assume that this is the trivial σ-field. Then $Y_0 = \mathbb{E}[\xi]$ and moreover if we are on Brownian filtration (i.e. the probability space is generated by the Brownian motion W) and $\xi \in L^2$, we deduce that there must exists a predictable process $Z \in L^2_{\mathscr{P}}(\mathbb{R}_+\times\Omega;\mathbb{R})$ such that (10) is statisfied. This by the martingale representation theorem. That the solution is unique is clear since if (Y',Z') is another solution with the same final condition then we have

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

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

5.1 Solution theory for BSDEs

In the following we consider BSDEs of the general form

$$-\mathrm{d}Y_s = f(t,\omega,Y_s,Z_s)\mathrm{d}s - Z_s\mathrm{d}W_s, \qquad Y_T = \xi \tag{11}$$

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

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

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

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

for almost every (t,ω).

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

Let us introduce the notations

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

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

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

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

for example.

Theorem 24. *Under these conditions the BSDE (14) has a unique solution* $(Y,Z) \in L^2_T(\mathbb{R}^n) \times L^2_T(\mathbb{R}^{n\times d})$.

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

$$-\mathrm{d}Y'_s = f(t,\omega,Y_s,Z_s)\mathrm{d}s - Z'_s\mathrm{d}W_s, \qquad Y'_T = \xi \tag{12}$$

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

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

but we have

$$Y_0' = \xi + \int_0^T f(t,\omega,Y_s,Z_s)\mathrm{d}s - \int_0^T Z_s'\mathrm{d}W_s,$$

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

$$Y_0' = \mathbb{E}\left[\xi + \int_0^T f(t,\omega,Y_s,Z_s)\mathrm{d}s\right].$$

As consequence

$$\mathbb{E}\left[\xi + \int_0^T f(t,\omega,Y_s,Z_s)\mathrm{d}s\middle|\mathscr{F}_t\right] = Y_0' + \int_0^t Z_s'\mathrm{d}W_s = Y_0' + \int_0^T Z_s'\mathrm{d}W_s - \int_t^T Z_s'\mathrm{d}W_s$$

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

so we concldue that we have

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

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

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

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

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

$$e^{\beta t}|\tilde{Y}_t^1 - \tilde{Y}_t^2|^2 + \int_t^T e^{\beta s}|\tilde{Z}_s^1 - \tilde{Z}_s^2|^2\mathrm{d}s + \beta\int_t^T e^{\beta s}|\tilde{Y}_s^1 - \tilde{Y}_s^2|^2\mathrm{d}s$$

$$= M_T - M_t + 2\int_t^T e^{\beta s}\langle\tilde{Y}_s^1 - \tilde{Y}_s^2, f(s,\omega,Y_s^1,Z_s^1) - f(s,\omega,Y_s^2,Z_s^2)\rangle\mathrm{d}s$$

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

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

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

5.2 Representation formula for non-linear PDEs

Recall the notation

$$\mathscr{L}_t f(t,x) = \sum_{i=1}^d b^i(t,x)\nabla^i f(t,x) + \sum_{i,j=1}^d a^{i,j}(t,x)\cdot\nabla^i\nabla^j f(t,x), \qquad t\geqslant 0, x\in\mathbb{R}^d,$$

where $f\in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^d;\mathbb{R})$ and $b\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^d$, $a\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^{d\times d}$ and b,a are sufficiently regular and $a=\frac{1}{2}\sigma\sigma^T$ for some $\sigma\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^{d\times d}$. We argued that if $(X_s^{t,x})_{s\geqslant t}$ is the solution to

$$\mathrm{d}X_s^{t,x} = b(s,X_s^{t,x})\mathrm{d}s + \sigma(s,X_s^{t,x})\mathrm{d}W_s, \qquad s\geqslant t, \tag{13}$$

with

$$X_t^{t,x} = x\in\mathbb{R}^d$$

and if we let $Y_s = u(s,X_s^{t,x})$, $Z_s=\sigma(X^{t,x})\nabla u(t,X_s^{t,x})$ for $s\geqslant t$ the the pair (Y,Z) satisfies the BSDE (9).

This was our motivation to look into the solution theory of a more general class of BSDEs of the form

$$-\mathrm{d}Y_s = f(s,\omega,Y_s,Z_s)\mathrm{d}s - Z_s\mathrm{d}W_s, \qquad Y_T=\xi \tag{14}$$

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

$$f(\cdot,\cdot,0,0)\in L^2_{\mathscr{P}}([0,T]\times\Omega;\mathbb{R}^n) \tag{15}$$

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

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

for almost every (t,ω).

We proved a theorem guarateeing that under these conditions the BSDE (14) has a unique solution

$$(Y, Z) \in L_T^2(\mathbb{R}^n) \times L_T^2(\mathbb{R}^{n\times d}).$$

We let now $(X_s^{t,x})_{s\geqslant 0}$ solving the (forward) SDE

$$\mathrm{d}X_s^{t,x} = b(s, X_s^{t,x})\mathrm{d}s + \sigma(s, X_s^{t,x})\mathrm{d}W_s, \qquad s \geqslant t, \tag{16}$$

for $s \geqslant t$ and such that $X_s^{t,x} = x$ for $s \leqslant t$. For given

$$f: \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \to \mathbb{R}^n$$

and

$$\Psi: \mathbb{R}^d \to \mathbb{R}^n,$$

let $(Y_s^{t,x}, Z_s^{t,x})_{s\in[0,T]}$ the solution of the BSDE ($s \in [0, T]$)

$$-\mathrm{d}Y_s^{t,x} = f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\mathrm{d}s - Z_s^{t,x}\mathrm{d}W_s, \qquad Y_T = \Psi(X_T^{t,x}) \tag{17}$$

This system of a forward SDE and a BSDE is called a (decoupled) forward-backward-SDE (FBSDE), is decoupled because the forward process $(X_s^{t,x})_s$ does not depends on $(Y^{t,x}, Z^{t,x})$ (otherwise is called fully-coupled).

We will assume that σ, b are *Lipshitz and of linear growth*, that f depends in a Lipschitz way on Y, Z (like in the general theory of the previous lecture) and moreover we have that

$$|f(t, x, 0, 0)| + |\Psi(x)| \leqslant C(1 + |x|^p),$$

for some $p \geqslant 1/2$. In this case the generator $f(t, X^{t,x}(\omega), y, z)$ satisfies the condition (15) and the final condition $\Psi(X_T^{t,x})$ is in L^2 because from the general theory of SDEs we can prove that solutions to (16) satisfy

$$\sup_{s\in[0,T]} \mathbb{E}[|X_s^{t,x}|^{2p}] \leqslant K(1 + |x|^{2p})$$

for some $K > 0$. This can be proven easily from a combination of BDG inequality (remember these are the L^p for the stochastic integral) and Grownwall's lemma, via the integral formulation of the SDE exploiting the linear growth of the coefficients b, σ.

From these assumptions it follows that the data of the BSDE satisfy the standard assumptions (those we introduced the last lecture) and therefore by the Theorem we proved it has a unique solution $(Y_s^{t,x}, Z_s^{t,x})_{s\in[0,T]}$.

Observe also that the process $(X_s^{t,x})_{s\in[0,T]}$ is a Markov process (exercise, it follows from the uniqueness of solutions to the SDE) and one has for all $t \leqslant u$

$$X_s^{t, X_t^{u,x}} = X_s^{t,x}, \qquad u \leqslant s$$

almost surely.

We want to prove now that we can express $Y_s^{t,x}, Z_s^{t,x}$ as deterministic functions of $X_s^{t,x}$. Namely that there exists two functions u, v such that $Y_s^{t,x} = u(s, X_s^{t,x})$ and $Z_s^{t,x} = \sigma(s, X_s^{t,x}) v(s, X_s^{t,x})$.

Introduce $(\mathscr{F}_{t,s})_{s\geqslant t}$ to be the completed right-continuous filtration generated by $(W_u - W_t)_{u\geqslant t}$, i.e. the future filtration of W after time t.

Proposition 25. *The solution $(Y_s^{t,x}, Z_s^{t,x})_{s\in[0,T]}$ is $(\mathscr{F}_{t,s})_{s\in[t,T]}$ adapted. In particular $\mathscr{F}_{t,s}$ is $\mathscr{F}_{t,t}$ measureable and therefore deterministic and $(Y_s^{t,x})_{s\in[0,t]}$ is also deterministic.*

Proof. Consider the new Brownian motion $\tilde{W}_s = W_{t+s} - W_t$ and let $\tilde{\mathscr{F}}$ its complected right-contrinuous filtration. Let (X', Y', Z') be the solution to the FBSDE:

$$\mathrm{d}X_s' = b(t+s, X_s')\mathrm{d}s + \sigma(t+s, X_s')\mathrm{d}W_s', \qquad s\geqslant 0, \qquad X_0' = x,$$

$$-\mathrm{d}Y_s' = f(t+s, X_s', Y_s', Z_s')\mathrm{d}s - Z_s'\mathrm{d}W_s, \qquad s\geqslant 0, \qquad Y_{T-t}' = \Psi(X_{T-t}').$$

By the general theory this FBSDE has a unique solution and then it is clear that $X_s' = X_{t+s}^{t,x}$ for $s\in[0, T-t]$ and similarly $(Y_s', Z_s') = (Y_{t+s}^{t,x}, Z_{t+s}^{t,x})$ for $s\in[0, T-t]$. However X', Y', Z' are adapted to $(\tilde{\mathscr{F}}_s)_{s\geqslant 0}$ which means that $(X_{t+s}^{t,x}, Y_{t+s}^{t,x}, Z_{t+s}^{t,x})_{s\geqslant 0}$ is adapted to $(\tilde{\mathscr{F}}_s)_{s\geqslant 0}$ and therefore $(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})_{s\in[t,T]}$ is adapted to $(\mathscr{F}_{t,s})_{s\in[s,T]}$ and therefore $(X_t^{t,x}, Y_t^{t,x}, Z_t^{t,x})$ is deterministic.

When $t'\leqslant t$ to see that $(Y_{t'}^{t,x}, Z_{t'}^{t,x})$ is deterministic one can just take $\tilde{W}_s = W_{t'+s} - W_{t'}$ and repeat the above argument by replacing there t with t'. Indeed the crucial remark is that $X_{t'}^{t,x} = x$ for any $t'\leqslant t$. □

Proposition 26. *There exists two deterministic measurable functions u, v such that $Y_s^{t,x} = u(s, X_s^{t,x})$ and $Z_s^{t,x} = \sigma(s, X_s^{t,x}) v(s, X_s^{t,x})$*

Proof. By induction, as follows. Assume first f does not depends on y, z. Then in this case

$$Y_s^{t,x} = \mathbb{E}\Big[\int_s^T f(r, X_r^{t,x})\mathrm{d}r + \Psi(X_T^{t,x})\Big|\mathscr{F}_s\Big] = \mathbb{E}\Big[\int_s^T f(r, X_r^{t,x})\mathrm{d}r + \Psi(X_T^{t,x})\Big|X_s^{t,x}\Big] = u(s, X_s^{t,x})$$

because $(X_s^{t,x})_{s\geqslant 0}$ is a Markov process and we can use the Markov property in the 2nd equality and the 3rd equality is just the statement that there exists a measurable function which represents the conditional expectation wrt. $\sigma(X_s^{t,x})$. Similarly one can show that $Z_s^{t,x} = \sigma(s, X_s^{t,x}) v(s, X_s^{t,x})$. (See Perkowski).

In the general case we introduce an iterative procedure. Define $Y^{(0)} = Z^{(0)} = 0$ then define $(Y^{(k+1)}, Z^{(k+1)})$ and the solution of the BSDE with driver $f(r, X_r^{t,x}, Y^{(k)}, Z^{(k)})$. We know from the proof of existence and uniqueness that there exists only one fixed point for this iteration and therefore $(Y^{(k)}, Z^{(k)}) \to (Y^{t,x}, Z^{t,x})$ (if you want this is the Picard iteration to construct the solution to the BSDE). From this we deduce that there exists functions u_k, v_k such that $Y_s^{(k)} = u_k(s, X_s^{t,x})$ and $Z_s^{(k)} = \sigma(s, X_s^{t,x}) v_k(s, X_s^{t,x})$, and the is not difficult to pass to the limit by letting $u^i(s,x) := \limsup_{k\to\infty}(u_k(s,x))^i$ (componentwise) and then $u^i(s, X_s^{t,x}) = \lim_{k\to\infty} Y_s^{(k)} = Y_s^{t,x}$ by convergence of the Picard iterations. Similarly one reason for the sequence $Z^{(k)}$ to deduce that

$$Z_s^{t,x} = \lim_{k\to\infty} Z_s^{(k)} = \sigma(s, X_s^{t,x}) \lim_{k\to\infty} v_k(s, X_s^{t,x}) = \sigma(s, X_s^{t,x})\, v(s, X_s^{t,x}).$$

This concludes the proof. □

Finally it remains to identify the functions u, v as associated to a nonlinear PDE.

We reason as follows: let u be the solution of the semilinar parabolic PDE

$$\partial_t u(t,x) + \mathscr{L}_t u(t,x) + f(t,x,u(t,x),\sigma(t,x)\nabla u(t,x)) = 0, \qquad t \in [0,T], x \in \mathbb{R}^d$$

with *final* condition $u(T,x) = \Psi(x)$.

Theorem 27. *(Generalised Feynman-Kac formula for quasilinear equations) Assume that* $u \in C^{1,2}([0,T] \times \mathbb{R}^d; \mathbb{R}^n)$ *is a solution to the PDE (26) such that*

$$|u(s,x)| + |\sigma(s,x)\nabla u(s,x)| \leqslant C(1+|x|^k)$$

for some $k \geqslant 1$. *Then if* $(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})_{s \in [0,T]}$ *is the unique solution to the FBSDE with final condition* Ψ *and driver f then we have*

$$Y_s^{t,x} = u(s, X_s^{t,x}), \qquad Z_s^{t,x} = \sigma(s, X_s^{t,x})\nabla u(s, X_s^{t,x}), \qquad s,t \in [0,T], x \in \mathbb{R}^d.$$

In particular

$$u(t,x) = Y_t^{t,x}, \qquad t \in [0,T], x \in \mathbb{R}^d,$$

and therefore the PDE has a unique solution.

Proof. We apply Ito formula

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

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

which means that the pair $(u(s, X_s^{t,x}), \sigma(s, X_s^{t,x}) \nabla u(s, X_s^{t,x}))$ is a solution to the BSDE, the final condition is ok since $u(T, X_T^{t,x}) = \Psi(X_T^{t,x})$ and by uniqueness we have $(u(s, X_s^{t,x}), \sigma(s, X_s^{t,x}) \nabla u(s, X_s^{t,x})) = (Y_s^{t,x}, Z_s^{t,x})$ for all $s \in [0,T]$. □

Remark 28. With stronger conditions on the coefficients of the PDE one can prove directly that given a solution to the BSDE which then, as we have seen can always be represented as $Y_s^{t,x} = u(s, X_s^{t,x})$ and $Z_s^{t,x} = \sigma(s, X_s^{t,x}) v(s, X_s^{t,x})$ for *some* functions u, v, then one necessarily have that $u \in C^{1,2}$ and $v = \nabla u$ and u solves the PDE. (see the notes of Perkowski for some literature on this).