Introduction to Stochastic Calculus and to the Resolution of PDEs Using Monte Carlo Simulations

Emmanuel Gobet

Abstract I give a pedagogical introduction to Brownian motion, stochastic calculus introduced by Itô in the fifties, following the elementary (at least not too technical) approach by Föllmer [Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French), pp. 143–150. Springer, Berlin, 1981]. Based on this, I develop the connection with linear and semi-linear parabolic PDEs. Then, I provide and analyze some Monte Carlo methods to approximate the solution to these PDEs. This course is aimed at master students, Ph.D. students and researchers interesting in the connection of stochastic processes with PDEs and their numerical counterpart. The reader is supposed to be familiar with basic concepts of probability (say first chapters of the book *Probability essentials* by Jacod and Protter [Probability Essentials, 2nd edn. Springer, Berlin, 2003]), but no a priori knowledge on martingales and stochastic processes is required.

1 The Brownian Motion and Related Processes

1.1 A Brief History of Brownian Motion

Historically, the Brownian motion (BM in short) is associated with the analysis of motions which time evolution is so disordered that it seems difficult to forecast their evolution, even in a very short time interval. It plays a central role in the theory of random processes, because in many theoretical and applied problems, the Brownian motion (or the diffusion processes that are built from Brownian motion) provides simple limit models on which many calculations can be made.

E. Gobet (✉)

Centre de Mathématiques Appliquées, Ecole Polytechnique and CNRS, Route de Saclay, 91128 Palaiseau Cedex, France

e-mail: emmanuel.gobet@polytechnique.edu

In 1827, the English botanist Robert Brown (1773–1858) first described the erratic motion of fine organic particles in suspension in a gas or a fluid. At the nineteenth century, after him, several physicists had admitted that this motion is very irregular and does not seem to admit a tangent; thus one could not speak of his speed, nor apply the laws of mechanics to him! In 1900 [4], Louis Bachelier (1870–1946) introduced the Brownian motion to model the dynamics of the stock prices, but his approach then is forgotten until the sixties... His Ph.D. thesis, *Théorie de la spéculation*, is the starting point of modern finance.

But Physics is the field at the beginning of the twentieth century which is at the origin of great interest for this process. In 1905, Albert Einstein (1879–1955) built a probabilistic model to describe the motion of a diffusive particle: he found that the law of the particle position at the time t, given the initial state x, admits a density which satisfies the *heat equation*, and actually it is *Gaussian*. Its theory is then quickly confirmed by experimental measurements of satisfactory diffusion constants. The same year as Einstein, a discrete version of the Brownian motion is proposed by the Polish physicist Smoluchowski using random walks.

In 1923, Norbert Wiener (1894–1964) built rigorously the random function that is called *Brownian motion*; he established in particular that the trajectories are continuous. By 1930, while following an idea of Paul Langevin, Ornstein and Uhlenbeck studied the Gaussian random function which bears their name and which seems to be the stationary or mean-reverting equivalent model associated to the Brownian motion.

It is the beginning of a very active theoretical research in Mathematics. Paul Lévy (1886–1971) discovered then, with other mathematicians, many properties of the Brownian motion [55] and introduced a first form of the stochastic differential equations, the study of which is later systematized by K. Itô (1915–2008). His work is gathered in a famous treaty published in 1948 [44] which is usually referred to as *Itô stochastic calculus*.

But History knows sometimes incredible bounces. Indeed in 2000, the French Academy of Science opened a manuscript remained sealed since 1940 pertaining to the young mathematician Doeblin (1915–1940), a French telegraphist died during the German offensive. Doeblin was already known for his remarkable achievements in the theory of probability due to his works on the stable laws and the Markov processes. This sealed manuscript gathered in fact his recent research, written between November 1939 and February 1940: it was actually related to his discovery (before Itô) of the stochastic differential equations and their relations with the Kolmogorov partial differential equations. Perhaps the Itô stochastic calculus could have been called Doeblin stochastic calculus...

1.2 The Brownian Motion and Its Paths

In the following, we study the basic properties of the Brownian motion and its paths.

1.2.1 Definition and Existence

The very erratic path which is a specific feature of the Brownian motion is in general associated with the observation that the phenomenon, although very disordered, has a certain time homogeneity, i.e. the origin date does not have importance to describe the time evolution. These properties underly the next definition.

Definition 1 (of Standard Brownian Motion). A *standard Brownian motion* is a random process $\{W_t; t \geq 0\}$ with continuous paths, such that

- $W_0 = 0$.
- The time increment $W_t - W_s$ with $0 \leq s < t$ has the Gaussian law,[1] with zero mean and variance equal $(t - s)$.
- For any $0 = t_0 < t_1 < t_2 \ldots .. < t_n$, the increments $\{W_{t_{i+1}} - W_{t_i}; 0 \leq i \leq n-1\}$ are independent[2] random variables.

There are important remarks following from the definition.

1. The state W_t of the system at time t is distributed as a Gaussian r.v. with mean 0 and variance t (increasing as time gets larger). Its probability density is

$$\mathbb{P}(W_t \in [x, x + dx]) = g(t, x)dx = \frac{1}{\sqrt{2\pi t}} \exp(-x^2/2t)dx. \qquad (1)$$

2. With probability 95 %, we have $|W_t| \leq 1.96\sqrt{t}$ (see Fig. 1) for a given time t. However, it may occur that W goes out this confidence interval.
3. The random variable W_t, as the sum of its increments, can be decomposed as a sum of independent Gaussian r.v.: this property serves as a basis from the further stochastic calculus.

Theorem 1. *The Brownian motion exists!*

Proof. There are different constructive ways to prove the existence of Brownian motion. Here, we use a Fourier based approach (proposed by Wiener), showing that W can be represented as a superposition of Gaussian signals. Also, we use a

---

[1]A Gaussian random variable X (see [46]) with mean μ and variance $\sigma^2 > 0$ (often denoted by $\mathcal{N}(\mu, \sigma^2)$) is the r.v. with density

$$g_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{(x-\mu)^2}{2\sigma^2}], \quad x \in \mathbb{R}.$$

If $\sigma^2 = 0$, $X = \mu$ with probability 1. Moreover, for any $u \in \mathbb{R}$, $\mathbb{E}(e^{uX}) = e^{u\mu + \frac{1}{2}u^2\sigma^2}$.

[2]Two random variables X_1 and X_2 are independent if and only if $\mathbb{E}(f(X_1)g(X_2)) = \mathbb{E}(f(X_1))\mathbb{E}(g(X_2))$ for any bounded functions f and g. This extends similarly to a vector.

Fig. 1 Simulation of a
Brownian motion with the
95 %-confidence interval
curves $f_{\pm}(t) = \pm 2\sqrt{t}$

equivalent characterization of Brownian motion as a Gaussian process[3] with zero
mean and covariance function $\mathbb{C}\mathrm{ov}(W_t, W_s) = \min(s, t) = s \wedge t$.

Let $(G_m)_{m \geq 0}$ be a sequence of independent Gaussian r.v. with zero mean and unit
variance and set

$$W_t = \frac{t}{\sqrt{\pi}} G_0 + \sqrt{\frac{2}{\pi}} \sum_{m \geq 1} \frac{\sin(mt)}{m} G_m.$$

We now show that W is a Brownian motion on $[0, \pi]$; then it is enough to
concatenate and sum up such independent processes to get finally a Brownian
motion defined on $\mathbb{R}^+$. We sketch the proof of our statement on W. First, the series
is *a.s.*[4] convergent since this is a Cauchy sequence in L_2: indeed, thanks to the
independence of the Gaussian random variables, we have

$$\| \sum_{m_1 \leq m \leq m_2} \frac{\sin(mt)}{m} G_m \|_{L_2}^2 = \sum_{m_1 \leq m \leq m_2} \frac{\sin^2(mt)}{m^2} \leq \sum_{m_1 \leq m} \frac{1}{m^2} \underset{m_1 \to +\infty}{\longrightarrow} 0.$$

---

[3]$(X_1, \ldots, X_n)$ is a Gaussian vector if and only if for any $(\lambda_i)_{1 \leq i \leq n} \in \mathbb{R}^n$, $\sum_{i=1}^n \lambda_i X_i$ has a
Gaussian distribution. Independent Gaussian random variables form a Gaussian vector. A process
$(X_t)_t$ is Gaussian if $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian vector for any times $(t_1, \ldots, t_n)$ and any n. A
Gaussian process is characterized by its mean $m(t) = \mathbb{E}(X_t)$ and its covariance function $K(s, t) = \mathbb{C}\mathrm{ov}(X_s, X_t)$.

[4]We recall that "an event A occurs *a.s.*" (*almost surely*) if $\mathbb{P}(\omega : \omega \in A) = 1$ or equivalently if
$\{w : w \notin A\}$ is a set of zero probability measure.

Fig. 2 Arithmetic Brownian motion with different drift parameters
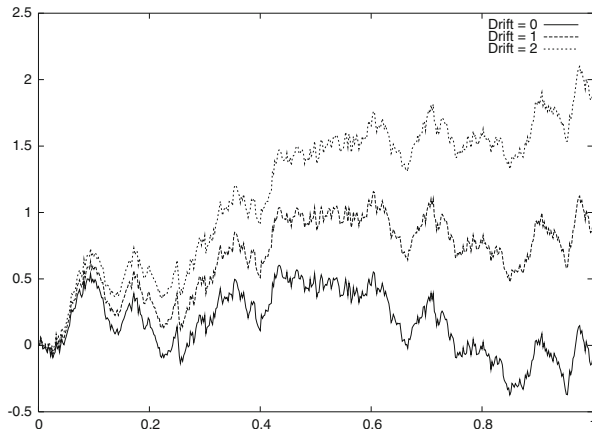
The partial sum has a Gaussian distribution, thus the *a.s.* limit[5] too. The same argument gives that W is a Gaussian process. It has zero mean and its covariance is the limit of the covariance of partial sums: thus

$$\mathbb{Cov}(W_t, W_s) = \frac{ts}{\pi} + \frac{2}{\pi} \sum_{m \geq 1} \frac{\sin(mt)}{m} \frac{\sin(ms)}{m}.$$

The above series is equal to $\min(s, t)$ for $(s, t) \in [0, \pi]^2$, by a standard computation of the Fourier coefficients of the function $t \in [-\pi, \pi] \mapsto \min(s, t)$ (for s fixed). The proof of continuity of W is based on the uniform convergence of the function series along some appropriate subsequences, which we do not detail (see [45, pp. 21–22]).

□

In many applications, it is useful to consider non standard Brownian motion.

Definition 2 (of Arithmetic Brownian Motion). An *arithmetic Brownian motion* (ABM in short) is a random process $\{X_t; t \geq 0\}$ where $X_t = x_0 + bt + \sigma W_t$ and

- W is a standard Brownian motion.
- $x_0 \in \mathbb{R}$ is the starting value of X.
- $b \in \mathbb{R}$ is the drift parameter.
- $\sigma \in \mathbb{R}$ is the diffusion parameter.

Usually, σ can be taken non-negative due to the symmetry of Brownian motion (see Proposition 1). X is still a Gaussian process, which position X_t at time t is distributed as $\mathcal{N}(x_0 + bt, \sigma^2 t)$ (Fig. 2).

---

[5]Here, we use the following standard result: let $(X_n)_{n \geq 1}$ be a sequence of random variables, each having the Gaussian distribution with mean μ_n and variance σ_n^2. If the distribution of X_n converges, then (μ_n, σ_n^2) converge to (μ, σ^2), and the limit distribution is Gaussian with mean μ and variance σ^2. We recall that if X_n converges *a.s.*, then it also converges in distribution.

1.2.2 First Easy Properties of the Brownian Path

Proposition 1. *Let $\{W_t; t \in \mathbb{R}^+\}$ a standard Brownian motion.*

i) SYMMETRY PROPERTY*: $\{-W_t; t \in \mathbb{R}^+\}$ is a standard Brownian motion.*

ii) SCALING PROPERTY*: for any $c > 0$, $\{W_t^c; t \in \mathbb{R}^+\}$ is a standard Brownian motion where*

$$W_t^c = c^{-1} W_{c^2 t}. \tag{2}$$

iii) TIME REVERSAL*: for any fixed T, $\hat{W}_t^T = W_T - W_{T-t}$ defines a standard Brownian motion on $[0, T]$.*

iv) TIME INVERSION*: $\{\hat{W}_t = tW_{1/t}, t > 0, \quad \hat{W}_0 = 0\}$ is a standard Brownian motion.*

The scaling property is important and illustrates the fractal feature of Brownian motion path: ε times W_t behaves like a Brownian motion at time $\varepsilon^2 t$.

Proof. It is a direct verification of the Brownian motion definition, related to independent, stationary and Gaussian increments. The continuity is also easy to verify, except for the case *iv)* at time 0. For this, we use that $\lim_{t \to 0^+} tW_{1/t} = \lim_{s \to +\infty} \frac{W_s}{s} = 0$, see Proposition 7. $\qquad\square$

1.3 Time-Shift Invariance and Markov Property

Previously, we have studied simple spatial transformation of Brownian motion. We now consider time-shifts, by first considering deterministic shifts.

Proposition 2 (Invariance by a Deterministic Time-Shift). *The Brownian Motion shifted by $h \geq 0$, given by $\{\bar{W}_t^h = W_{t+h} - W_h; t \in \mathbb{R}^+\}$, is another Brownian motion, independent of the Brownian Motion stopped at h, $\{W_s; s \leq h\}$.*

In other words, $\{W_{t+h} = W_h + \bar{W}_t^h; t \in \mathbb{R}^+\}$ is a Brownian motion starting from W_h. The above property is associated to the *weak Markov property* which states (possibly applicable to other processes) that *the distribution of W after h conditionally on the past up to time h depends only on the present value W_h.*

Proof. The Gaussian property of $\bar{W}^h$ is clear.

The independent increments of W induce those of $\bar{W}^h$.

It remains to show the independence w.r.t. the past up to h, i.e. the sigma-field generated by $\{W_s; s \leq h\}$, or equivalently w.r.t. the sigma-field generated by $\{W_{s_1}, \ldots W_{s_N}\}$ for any $0 \leq s_1 \leq \cdots \leq s_N \leq h$. The independence of increments of W ensures that $(\bar{W}_{t_1}^h, \bar{W}_{t_2}^h - \bar{W}_{t_1}^h, \cdots, \bar{W}_{t_k}^h - \bar{W}_{t_{k-1}}^h) = (W_{t_1+h} - W_h, \cdots, W_{t_k+h} - W_{t_{k-1}+h})$ is independent of $(W_{s_1}, W_{s_2} - W_{s_1}, \cdots, W_{s_j} - W_{s_{j-1}})$. Then $(\bar{W}_{t_1}^h, \bar{W}_{t_2}^h, \cdots, \bar{W}_{t_k}^h)$ is independent of $\{W_s; s \leq h\}$. $\qquad\square$

As a consequence, we can derive a nice symmetry result making the connection between the maximum of Brownian motion monitored along a finite time grid $t_0 = 0 < t_1 < \cdots < t_N = T$ and that of W_T only.

Proposition 3. *For any $y \geq 0$, we have*

$$\mathbb{P}[\sup_{i \leq N} W_{t_i} \geq y] \leq 2\mathbb{P}[W_T \geq y] = \mathbb{P}[|W_T| \geq y]. \tag{3}$$

Proof. The equality at the r.h.s. comes from the symmetric distribution of W_T. Now we show the inequality on the left. Denote by t_y^* the first time t_j when W reaches the level y. Notice that $\{\sup_{i \leq N} W_{t_i} \geq y\} = \{t_y^* \leq T\}$ and $\{t_y^* = t_j\} = \{W_{t_i} < y, \forall i < j, W_{t_j} \geq y\}$. For each $j < N$, the symmetry of Brownian increments gives $\mathbb{P}[W_T - W_{t_j} \geq 0] = \frac{1}{2}$. Since the shifted Brownian motion $(\bar{W}_t^{t_j} = \bar{W}_{t_j+t} - W_{t_j} : t \in \mathbb{R}^+)$ is independent of $(W_s : s \leq t_j)$, we have

$$\frac{1}{2} \mathbb{P}[\sup_{i \leq N} W_{t_i} \geq y] = \frac{1}{2}\mathbb{P}[t_y^* \leq T] = \frac{1}{2} \sum_{j=0}^{N} \mathbb{P}[t_y^* = t_j]$$

$$= \frac{1}{2}\mathbb{P}[W_{t_i} < y, \forall i < N, W_T \geq y] + \sum_{j=0}^{N-1} \mathbb{P}[W_{t_i} < y, \forall i < j, W_{t_j} \geq y]\mathbb{P}[W_T - W_{t_j} \geq 0]$$

$$= \frac{1}{2}\mathbb{P}[W_{t_i} < y, \forall i < N, W_T \geq y] + \sum_{j=0}^{N-1} \mathbb{P}[W_{t_i} < y, \forall i < j, W_{t_j} \geq y, W_T - W_{t_j} \geq 0]$$

$$\leq \mathbb{P}[W_{t_i} < y, \forall i < N, W_T \geq y] + \sum_{j=0}^{N-1} \mathbb{P}[W_{t_i} < y, \forall i < j, W_{t_j} \geq y, W_T \geq y]$$

$$= \mathbb{P}[t_y^* \leq T, W_T \geq y] = \mathbb{P}[W_T \geq y].$$

At the two last lines, we have used $\{W_{t_j} \geq y, W_T - W_{t_j} \geq 0\} \subset \{W_{t_j} \geq y, W_T \geq y\}$ and $\{W_T \geq y\} \subset \{t_y^* \leq T\}$. $\qquad\square$

Taking a grid with time step T/N with $N \to +\infty$, we have $\sup_{i \leq N} W_{t_i} \uparrow \sup_{0 \leq t \leq T} W_{t_i}$. Then, we can pass to the limit (up to some probabilistic convergence technicalities) in the inequality (3) to get

$$\mathbb{P}[\sup_{0 \leq t \leq T} W_t \geq y] \leq \mathbb{P}[|W_T| \geq y]. \tag{4}$$

Actually, the inequality (4) is an equality: it is proved later in Proposition 5.

Now, our aim is to extend Proposition 2 to the case of stochastic time-shifts h. Without extra assumption on h, the result is false in general: a counter-example is the last passage time of W at zero before the time 1 ($L = \sup\{t \leq 1 : W_t = 0\}$), which does not satisfy the property. Indeed, since $(W_{s+L} - W_s)_{s \geq 0}$ do not vanish

a.s. at short time (due to the definition of L), the marginal distribution can not be Gaussian and the time-shifted process can not be a Brownian motion.

The right class for extension is the class of stopping times, defined as follows.

Definition 3 (Stopping Time). A stopping time is non-negative random variable U (taking possibly the value $+\infty$), such that for any $t \geq 0$, the event $\{U \leq t\}$ depends only on the Brownian motion values $\{W_s; s \leq t\}$.

The stopping time is discrete if it takes only a countable set of values $(u_1, \cdots, u_n, \cdots)$.

In other words, it suffices to observe the Brownian motion until time t to know whether or not the event $\{U \leq t\}$ occurs. Of course, deterministic times are stopping times. A more interesting example is the first hitting time of a level $y > 0$

$$T_y = \inf\{t > 0; W_t \geq y\};$$

it is a stopping time, since $\{T_y \leq t\} = \{\exists s \leq t, W_s = y\}$ owing to the continuity of W. Observe that the counter-example of last passage time L is not a stopping time.

Proposition 4. *Let U be a stopping time. On the event $\{U < +\infty\}$, the Brownian motion shifted by $U \geq 0$, i.e. $\{\bar{W}_t^U = W_{t+U} - W_U; t \in \mathbb{R}^+\}$, is a Brownian motion independent of $\{W_t; t \leq U\}$.*

This result is usually referred to as the *strong Markov property*.

Proof. We show that for any $0 \leq t_1 < \cdots < t_k$, any $0 \leq s_1 < \cdots < s_l$, any $(x_1, \cdots, x_k)$ and any measurable sets $(B_1, \cdots, B_{l-1})$, we have

$$\mathbb{P}(\bar{W}_{t_1}^U < x_1, \cdots, \bar{W}_{t_k}^U < x_k, W_{s_1} \in B_1, \cdots, W_{s_{l-1}} \in B_{l-1}, s_l \leq U < +\infty)$$
$$= \mathbb{P}(W_{t_1}' < x_1, \cdots, W_{t_k}' < x_k)\mathbb{P}(W_{s_1} \in B_1, \cdots, W_{s_{l-1}} \in B_{l-1}, s_l \leq U < +\infty), \quad (5)$$

where W' is a Brownian motion independent of W. We begin with the easier case where U is a discrete stopping time valued in $(u_n)_{n \geq 1}$: then

$$\mathbb{P}(\bar{W}_{t_1}^U < x_1, \cdots, \bar{W}_{t_k}^U < x_k, W_{s_1} \in B_1, \cdots, W_{s_{l-1}} \in B_{l-1}, s_l \leq U < +\infty)$$
$$= \sum_n \mathbb{P}(\bar{W}_{t_1}^U < x_1, \cdots, \bar{W}_{t_k}^U < x_k, W_{s_1} \in B_1, \cdots, W_{s_{l-1}} \in B_{l-1}, s_l \leq U, U = u_n)$$
$$= \sum_n \mathbb{P}(\bar{W}_{t_1}^{u_n} < x_1, \cdots, \bar{W}_{t_k}^{u_n} < x_k, W_{s_1} \in B_1, \cdots, W_{s_{l-1}} \in B_{l-1}, s_l \leq U, U = u_n)$$
$$= \sum_n \mathbb{P}(W_{t_1}' < x_1, \cdots, W_{t_k}' < x_k)\mathbb{P}(W_{s_1} \in B_1, \cdots, W_{s_{l-1}} \in B_{l-1}, s_l \leq U, U = u_n)$$
$$= \mathbb{P}(W_{t_1}' < x_1, \cdots, W_{t_k}' < x_k)\mathbb{P}(W_{s_1} \in B_1, \cdots, W_{s_{l-1}} \in B_{l-1}, s_l \leq U < +\infty)$$

Fig. 3 Brownian motion $(W_{T_y+t} = \bar{W}_t^{T_y} + y : t \in \mathbb{R}^+)$ starting from y and its symmetric path
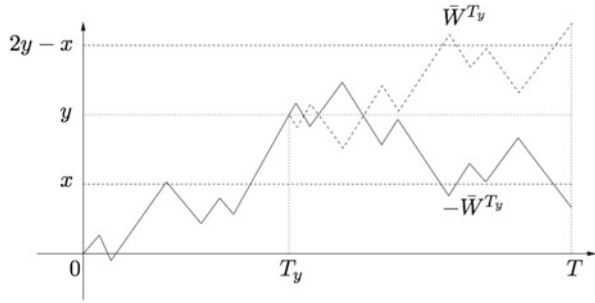
applying at the last equality but one the time-shift invariance with deterministic shift u_n. For the general case for U, we apply the result to the discrete stopping time $U_n = \frac{[nU]+1}{n}$, and then pass to the limit using the continuity of W. □

1.4 Maximum, Behavior at Infinity, Path Regularity

We apply the strong Markov property to identify the law of the Brownian motion maximum.

Proposition 5 (Symmetry Principle). *For any $y \geq 0$ and any $x \leq y$, we have*

$$\mathbb{P}[\sup_{t \leq T} W_t \geq y; W_T \leq x] = \mathbb{P}[W_T \geq 2y - x], \tag{6}$$

$$\mathbb{P}[\sup_{t \leq T} W_t \geq y] = \mathbb{P}[|W_T| \geq y] = 2 \int_{\frac{y}{\sqrt{T}}}^{+\infty} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \mathrm{d}x. \tag{7}$$

Proof. Denote by $T_y = \inf\{t > 0 : W_t \geq y\}$ and $+\infty$ if the set is empty. Observe that T_y is a stopping time and that $\{\sup_{t \leq T} W_t \geq y; W_T \leq x\} = \{T_y \leq T; W_T \leq x\}$. By Proposition 4, on $\{T_y \leq T\}$, $(W_{T_y+t} = \bar{W}_t^{T_y} + y : t \in \mathbb{R}^+)$ is a Brownian motion starting from y, independent of $(W_s : s \leq T_y)$. By symmetry (see Fig. 3), the events $\{T_y \leq T, W_T < x\}$ and $\{T_y \leq T, W_T > 2y-x\}$ has the same probability. But for $x \leq y$, we have $\{T_y \leq T, W_T > 2y - x\} = \{W_T > 2y - x\}$ and the first result is proved.

For the second result, take $y = x$ and write $\mathbb{P}[\sup_{t \leq T} W_t \geq y] = \mathbb{P}[\sup_{t \leq T} W_t \geq y, W_T > y] + \mathbb{P}[\sup_{t \leq T} W_t \geq y, W_T \leq y] = \mathbb{P}[W_T > y] + \mathbb{P}[W_T \geq y] = 2\mathbb{P}(W_T \geq y) = \mathbb{P}(|W_T| \geq y)$. □

As a consequence of the identification of the law of the maximum up to a fixed time, we prove that the range of Brownian motion becomes $\mathbb{R}$ at time goes to infinity.

Proposition 6. *With probability 1, we have*

$$\limsup_{t \to +\infty} W_t = +\infty, \qquad \liminf_{t \to +\infty} W_t = -\infty.$$

Proof. For $T \geq 0$, set $M_T = \sup_{t \leq T} W_t$. As $T \uparrow +\infty$, it defines a sequence of increasing r.v., thus converging *a.s.* to a limit r.v. M_∞. Applying twice the monotone convergence theorem, we obtain

$$\mathbb{P}[M_\infty = +\infty] = \lim_{y \uparrow +\infty} \mathbb{P}[M_\infty > y] = \lim_{y \uparrow +\infty} \left(\lim_{T \uparrow +\infty} \mathbb{P}[M_T > y] \right)$$

$$= \lim_{y \uparrow +\infty} \left(\lim_{T \uparrow +\infty} \mathbb{P}[|W_T| \geq y] \right) = 1$$

using (7). This proves that $\limsup_{t \to +\infty} W_t = +\infty$ *a.s.* and a symmetry argument gives the liminf. $\square$

However, the increasing rate of W is sublinear as time goes to infinity.

Proposition 7. *With probability 1, we have*

$$\lim_{t \to +\infty} \frac{W_t}{t} = 0.$$

Proof. The strong law of large numbers yields that $\frac{W_n}{n} = \frac{1}{n} \sum_{i=1}^{n} (W_i - W_{i-1})$ converges *a.s.* to $\mathbb{E}(W_1) = 0$. The announced result is thus proved along the sequence of integers. To fill the gaps between integers, set $\tilde{M}_n = \sup_{n < t \leq n+1} (W_t - W_n)$ and $\tilde{M}'_n = \sup_{n < t \leq n+1} (W_n - W_t)$: due to Proposition 5, $\tilde{M}_n$ and $\tilde{M}'_n$ have the same distribution as $|W_1|$. Then, the Chebyshev inequality writes

$$\mathbb{P}(|\tilde{M}_n| + |\tilde{M}'_n| \geq n^{3/4}) \leq 2 \frac{\mathbb{E}(|\tilde{M}_n|^2) + \mathbb{E}(|\tilde{M}'_n|^2)}{n^{3/2}} = 4n^{-3/2},$$

implying that $\sum_{n \geq 0} \mathbb{P}(|\tilde{M}_n| + |\tilde{M}_n| \geq n^{3/4}) < +\infty$. Thus, by Borel–Cantelli's lemma, we obtain that with probability 1, for n large enough $|\tilde{M}_n| + |\tilde{M}'_n| < n^{3/4}$, i.e. $\frac{\tilde{M}_n}{n}$ and $\frac{\tilde{M}'_n}{n}$ both converge *a.s.* to 0. $\square$

By time inversion, $\hat{W}_t = t W_{1/t}$ is another Brownian motion: the $\hat{W}$-growth in infinite time gives an estimate on W at 0, which writes

$$+\infty = \limsup_{t \to +\infty} |\hat{W}_t| = \limsup_{s \to 0^+} \frac{|W_s - W_0|}{s}$$

which shows that W is not differentiable at time 0. By time-shift invariance, this is also true at any given time t. The careful reader may notice that the set of full probability measure depends on t and it is unclear at this stage if a single full set is available for any t, i.e. if

$$\mathbb{P}(\exists t_0 \text{ such that } t \mapsto W_t \text{ is differentiable at } t_0) = 0.$$

Actually, the above result holds true and it is due to Paley–Wiener–Zygmund (1933). The following result is of comparable nature: we claim that *a.s.* there does not exist any interval on which W is monotone.

Proposition 8 (Nowhere Monotonicity). *We have*

$$\mathbb{P}(t \mapsto W_t \text{ is monotone on an interval}) = 0.$$

Proof. Define $M_{s,t}^{\uparrow} = \{\omega : u \mapsto W_u(\omega) \text{ is increasing on the interval}]s, t[\}$ and $M_{s,t}^{\downarrow}$ similarly. Observe that

$$M = \{t \mapsto W_t \text{ is monotone on the interval}\} = \bigcup_{s,t \in \mathbb{Q}, 0 \le s < t} (M_{s,t}^{\uparrow} \cup M_{s,t}^{\downarrow}),$$

and since this is a countable union, it is enough to show $\mathbb{P}(M_{s,t}^{\uparrow}) = \mathbb{P}(M_{s,t}^{\downarrow}) = 0$ to conclude $\mathbb{P}(M) \le \sum_{s,t \in \mathbb{Q}, 0 \le s < t} [\mathbb{P}(M_{s,t}^{\uparrow}) + \mathbb{P}(M_{s,t}^{\downarrow})] = 0$. For fixed n, set $t_i = s + i(t - s)/n$, then

$$\mathbb{P}(M_{s,t}^{\uparrow}) \le \mathbb{P}(W_{t_{i+1}} - W_{t_i} \ge 0, 0 \le i < n) = \prod_{i=0}^{n-1} \mathbb{P}(W_{t_{i+1}} - W_{t_i} \ge 0) = \frac{1}{2^n},$$

leveraging the symmetric distribution of the increments. Taking now n large gives $\mathbb{P}(M_{s,t}^{\uparrow}) = 0$. We argue similarly for $\mathbb{P}(M_{s,t}^{\downarrow}) = 0$. $\qquad\square$

In view of this lack of smoothness, it seems impossible to define differential calculus along the paths of Brownian motion. However, as it will be further developed, Brownian motion paths enjoy a nice property of finite quadratic variations, which serves to build an appropriate stochastic calculus.

There are much more to tell about the properties of Brownian motion. We mention few extra properties without proof:

- HOLDER REGULARITY: for any $\rho \in (0, \frac{1}{2})$ and any deterministic $T > 0$, there exists a *a.s.* finite r.v. $C_{\rho,T}$ such that

$$\forall\, 0 \le s, t \le T, \quad |W_t - W_s| \le C_{\rho,T} |t - s|^{\rho}.$$

- LAW OF ITERATED LOGARITHM: setting $h(t) = \sqrt{2t \log \log t^{-1}}$, we have

$$\limsup_{t \downarrow 0} \frac{W_t}{h(t)} = 1 \quad \text{a.s.} \quad \text{and} \quad \liminf_{t \downarrow 0} \frac{W_t}{h(t)} = -1 \quad \text{a.s.}$$

- ZEROS OF BROWNIAN MOTION: the set $\chi = \{t \geq 0 : W_t = 0\}$ of the zeros of W is closed, unbounded, with null Lebesgue measure and it has no isolated points.

1.5   The Random Walk Approximation

Another algorithmic way to build a Brownian motion consists in rescaling a random walk. This is very simple and very useful for numerics: it leads to the so-called tree methods and it has some connections with finite differences in PDEs.

Consider a sequence $(X_i)_i$ of independent random variables with Rademacher distribution: $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$. Then

$$S_n = \sum_{i=1}^{n} X_i$$

defines a random walk on $\mathbb{Z}$. Like Brownian motion, it is a process with stationary independent increments, but it is not Gaussian. Actually S_n has a binomial distribution:

$$\mathbb{P}(S_n = -n + 2k) = \mathbb{P}(k \text{ rises}) = 2^{-n} \binom{n}{k}.$$

A direct computation shows that $\mathbb{E}(S_n) = 0$ and $\mathbb{V}\text{ar}(S_n) = n$. When we rescale the walk and we let n go towards infinity, we observe however that due the Central Limit Theorem, the distribution of $\frac{S_n}{\sqrt{n}}$ converges to the Gaussian law with zero mean and unit variance. The fact that it is equal to the law of W_1 is not a coincidence, since it can be justified that the full trajectory of the suitably rescaled random walk converges towards that of a Brownian motion, see Fig. 4. This result is known as *Donsker theorem*, see for instance [12] for a proof.

Proposition 9. *Define* $(Y_t^n)_t$ *as the piecewise constant process*

$$Y_t^n = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i. \tag{8}$$

The distribution of the process $(Y_t^n)_t$ *converges to that of a Brownian motion* $(W_t)_t$ *as* $n \to +\infty$, *i.e. for any continuous functional*

$$\lim_{n \to \infty} \mathbb{E}(\Phi(Y_t^n : t \leq 1)) = \mathbb{E}(\Phi(W_t : t \leq 1)).$$
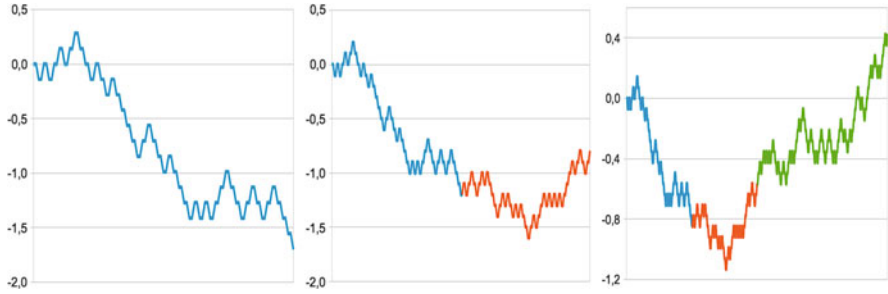
Fig. 4 The random walk rescaled in time and space. *From left to right*: the process Y^n for $n = 50,100,200$. The pieces of path with same color are built with the same X_i

The last result gives a simple way to evaluate numerically expectations of functionals of Brownian motion. It is the principle of the so-called *binomial tree methods*.

Link with Finite Difference Scheme. The random walk can be interpreted as a explicit FD scheme for the heat equation. We anticipate a bit on the following where the connection between Brownian motion and heat equation will be more detailed.

For $t = \frac{i}{n}$ ($i \in \{0,\ldots,n\}$) and $x \in \mathbb{R}$, set

$$u^n(t,x) = \mathbb{E}\Big(f\big(x + Y^n_{\frac{i}{n}}\big)\Big).$$

The independence of $(X_i)_i$ gives

$$u^n\Big(\frac{i}{n},x\Big) = \mathbb{E}\big(f(x + Y^n_{\frac{i-1}{n}} + \frac{X_i}{\sqrt{n}})\big)$$

$$= \frac{1}{2}u^n\Big(\frac{i-1}{n},x + \frac{1}{\sqrt{n}}\Big) + \frac{1}{2}u^n\Big(\frac{i-1}{n},x - \frac{1}{\sqrt{n}}\Big),$$

$$\frac{u^n\big(\frac{i}{n},x\big) - u^n\big(\frac{i-1}{n},x\big)}{\frac{1}{n}} = \frac{1}{2}\frac{u^n\big(\frac{i-1}{n},x + \frac{1}{\sqrt{n}}\big) - 2u^n\big(\frac{i-1}{n},x\big) + u^n\big(\frac{i-1}{n},x - \frac{1}{\sqrt{n}}\big)}{\big(\frac{1}{\sqrt{n}}\big)^2}.$$

Thus, u^n related to the expectation of the random walk can be read as an explicit FD scheme of the heat equation $\partial_t u(t,x) = \frac{1}{2}\partial^2_{xx} u(t,x)$ and $u(0,x) = f(x)$, with time step $\frac{1}{n}$ and space step $\frac{1}{\sqrt{n}}$.

1.6 Other Stochastic Processes

We present other one-dimensional processes, with continuous trajectories, which derive from the Brownian motion.

1. Geometric Brownian motion: this model is popular in finance to model stocks and other assets by a positive process.
2. Ornstein–Uhlenbeck process: it has important applications in physics, mechanics, economy and finance to model stochastic phenomena exhibiting mean-reverting features (like spring endowed with random forces, interest-rates or inflation, ...).
3. Stochastic differential equations: it gives the more general framework.

1.6.1 Geometric Brownian Motion

Definition 4. A Geometric Brownian Motion (GBM in short) with deterministic initial value $S_0 > 0$, drift coefficient μ and diffusion coefficient σ, is a process $(S_t)_{t \geq 0}$ defined by

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}, \tag{9}$$

where $\{W_t ; t \geq 0\}$ is a standard Brownian motion.

As the argument in the exponential has a Gaussian distribution, the random variable S_t (with t fixed) is known as Lognormal.

This is a process with continuous trajectories, which takes strictly positive values. The Geometric Brownian motion is often used as a model of asset price (see Samuelson [65]): this choice is justified on the one hand, by the positivity of S and on the other hand, by the simple Gaussian properties of its returns:

- The returns $\log(S_t) - \log(S_s)$ are Gaussian with mean $(\mu - \frac{1}{2}\sigma^2)(t - s)$ and variance $\sigma^2(t - s)$.
- For all $0 < t_1 < t_2 \ldots .. < t_n$, the relative increments $\{\frac{S_{t_{i+1}}}{S_{t_i}} ; 0 \leq i \leq n - 1\}$ are independent.

The assumption of Gaussian returns is not valid in practice but this model still serves as a proxy for more sophisticated models.

Naming μ the drift parameter may be surprising at first sight since it appears in the deterministic component as $(\mu - \frac{1}{2}\sigma^2)t$. Actually, a computation of expectation gives easily

$$\mathbb{E}(S_t) = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t} \mathbb{E}(e^{\sigma W_t}) = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t} e^{\frac{1}{2}\sigma^2 t} = S_0 e^{\mu t}.$$

The above equality gives the interpretation to μ as a mean drift term: $\mu = \frac{1}{t} \log[\mathbb{E}(S_t)/S_0]$.

1.6.2 Ornstein–Uhlenbeck Process

Let us return to physics and to the Brownian motion by Einstein in 1905. In order to propose a more adequate modeling of the phenomenon of particles diffusion, we introduce the process of Ornstein–Uhlenbeck and its principal properties.

So far we have built the Brownian motion like a model for a microscopic particle in suspension in a liquid subjected to thermal agitation. An important criticism made with this modeling concerns the assumption that displacement increments are independent and they do not take into account the effects of the particle speed due to particle inertia.

Let us denote by m the particle mass and by $\dot{X}(t)$ its speed. Owing to Newton's second law, the momentum change $m\dot{X}(t+\delta(t)) - m\dot{X}(t)$ is equal to the resistance $-k\dot{X}(t)\delta t$ of the medium during time δt, plus the momentum change due to molecular shocks, that we assume to be with stationary independent increments and thus associated with a Brownian motion. The process thus modeled is called sometimes the *physical Brownian motion*. The equation for the increments becomes

$$m\,\delta[\dot{X}(t)] = -k\dot{X}(t)\delta t + m\sigma\delta W_t.$$

Trajectories of the Brownian motion being not differentiable, the equation has to be read in an integral form

$$m\dot{X}(t) = m\dot{X}(0) - \int_0^t k\dot{X}(s)\mathrm{d}s + m\sigma W_t.$$

$\dot{X}(t)$ is thus solution of the linear stochastic differential equation (known as Langevin equation)

$$V_t = v_0 - a\int_0^t V_s\mathrm{d}s + \sigma W_t$$

where $a = \frac{k}{m}$. If $a = 0$, we recover an arithmetic Brownian motion and to avoid this trivial reduction, we assume $a \neq 0$ in the sequel. However, the existence of solution is not clear since W is not differentiable. To overcome this difficulty, set $Z_t = V_t - \sigma W_t$: that leads to the new equation

$$Z_t = v_0 - a\int_0^t (Z_s + \sigma W_s)\mathrm{d}s,$$

which is now a linear ordinary differential equation that can be solved path by path. The variation of parameter method gives the representation of the unique solution of this equation like

Fig. 5 Ornstein–Uhlenbeck
paths with $V_0 = 1$, $a = 2$ and
$\sigma = 0.1$
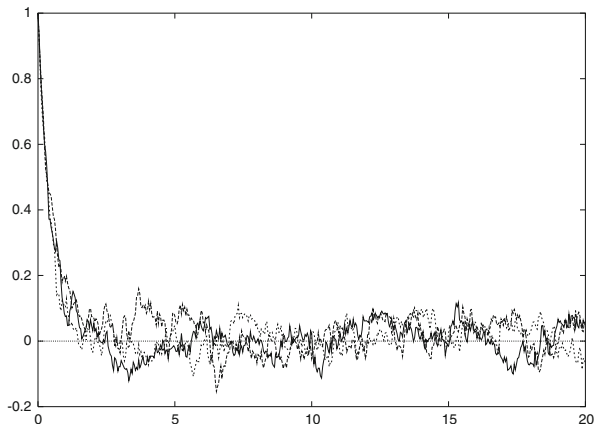
$$Z_t = v_0 e^{-at} - \sigma \int_0^t a e^{-a(t-s)} W_s \mathrm{d}s.$$

The initial solution is thus

$$V_t = v_0 e^{-at} + \sigma W_t - \sigma \int_0^t a e^{-a(t-s)} W_s \mathrm{d}s. \tag{10}$$

Using stochastic calculus, we will derive later (see Sect. 3.3) another convenient representation of V as follows:

$$V_t = v_0 e^{-at} + \sigma \int_0^t e^{-a(t-s)} \mathrm{d}W_s \tag{11}$$

using a stochastic integral not yet defined. From (10), assuming that v_0 is deterministic, we can show the following properties (see also Sect. 3.3).

- For a given t, V_t has a Gaussian distribution: indeed, as the limit of a Riemann sum, it is the *a.s.* limit of a Gaussian r.v., see footnote 5 page 111.
- More generally, V is a Gaussian process.
- Its mean is $v_0 e^{-at}$, its covariance function $\mathbb{Cov}(V_t, V_s) = e^{-a(t-s)} \frac{\sigma^2}{2a}(1 - e^{-2as})$ for $t > s$.

Observe that for $a > 0$, the Gaussian distribution of V_t converges to $\mathcal{N}(0, \frac{\sigma^2}{2a})$ as $t \to +\infty$: it does not depend anymore on v_0 and illustrates the mean-reverting feature of this model, see Fig. 5.

1.6.3 Stochastic Differential Equations and Euler Approximations

The previous example gives the generic form of a Stochastic Differential Equation, that generalizes the usual Ordinary Differential Equations $x_t' = b(x_t)$ or in integral form $x_t = x_0 + \int_0^t b(x_s)\mathrm{d}s$.

Definition 5. Let $b, \sigma : x \in \mathbb{R} \mapsto \mathbb{R}$ be two functions, respectively the drift and the diffusion coefficient. A Stochastic Differential Equation (SDE in short) with parameter (b, σ) and initial value x is a stochastic process $(X_t)_{t \geq 0}$ solution of

$$X_t = x + \int_0^t b(X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s, \quad t \geq 0,$$

where $(W_t)_t$ is a standard Brownian motion.

A slightly more general definition (not considered here) could include the case of time-dependent coefficients $b(t, x)$ and $\sigma(t, x)$, the subsequent analysis would be quite similar. In the definition above, we use a stochastic integral $\int_0^t \ldots \mathrm{d}W_s$ which has not yet been defined: it will be explained in the next section. For the moment, the reader needs to know that in the simplest case where σ is constant, we simply have $\int_0^t \sigma(X_s)\mathrm{d}W_s = \sigma W_t$. The previous examples fit this setting:

- The arithmetic Brownian motion corresponds to $b(x) = b$ et $\sigma(x) = \sigma$.
- The Ornstein–Uhlenbeck process corresponds to $b(x) = -ax$ et $\sigma(x) = \sigma$.

Taking σ to be non constant allows for more general situations and more flexible models. Instead of discussing now the important issues of existence and uniqueness to such SDE, we rather consider natural approximations of them, namely the Euler scheme (which is the direct extension of Euler scheme for ODEs).

Definition 6. Let (b, σ) be given drift and diffusion coefficients. The Euler scheme associated to the SDE with coefficients (b, σ), initial value x and time step h, is defined by

$$\begin{cases} X_0^h = x, \\ X_t^h = X_{ih}^h + b(X_{ih}^h)(t - ih) + \sigma(X_{ih}^h)(W_t - W_{ih}), \quad i \geq 0, t \in (ih, (i+1)h]. \end{cases} \tag{12}$$

In other words, X^h is a piecewise arithmetic Brownian motion with coefficients on the interval $(ih, (i+1)h]$ computed according to the functions (b, σ) evaluated at X_{ih}^h. In general, the law of X_t^h is not known analytically: at most, we can give explicit representations using an induction of the time-step. On the other hand, as it will be seen further, the random simulation of X^h at time $(ih)_{i \geq 0}$ is easily performed by simulating the independent Brownian increments $(W_{(i+1)h} - W_{ih})$. The accuracy of the approximation of X by X^h is expected to get improved as h goes to 0.

Complementary References. See [48, 57, 63].

2   Feynman–Kac Representations of PDE Solutions

Our purpose in this section is to make the connection between the expectations of
functionals of Brownian motion and the solution of second order linear parabolic
partial differential equations (PDE in short): this leads to the well-known Feynman–
Kac representations. We extend this point of view to other simple processes
introduced before.

2.1   The Heat Equations

2.1.1   Heat Equation in the Whole Space

Let us return to the law of $x + W_t$, the Gaussian density of which is

$$g(t, x, y) := g(t, y - x) = \frac{1}{\sqrt{2\pi t}} \exp(-(y - x)^2/2t),$$

often called in this context the fundamental solution of the heat equation. One of the
key properties is *the property of convolution*

$$g(t + s, x, y) = \int_{\mathbb{R}} g(t, x, z)g(s, z, y)\mathrm{d}z \tag{13}$$

which says in an analytical language that $x + W_{t+s}$ is the sum of the independent
Gaussian variables $x + W_t$ and $W_{t+s} - W_t$. A direct calculation on the density shows
that the Gaussian density is solution to the *heat equation* w.r.t. the two variables x
and y

$$\begin{cases} g'_t(t, x, y) = \frac{1}{2}g''_{yy}(t, x, y), \\ g'_t(t, x, y) = \frac{1}{2}g''_{xx}(t, x, y). \end{cases} \tag{14}$$

This property is extended to a large class of functions built from the Brownian
motion.

Theorem 2 (Heat Equation with Cauchy Initial Boundary Condition). *Let f
be a bounded[6] measurable function. Consider the function*

$$u(t, x, f) = \mathbb{E}[f(x + W_t)] = \int_{\mathbb{R}} g(t, x, y)f(y)\mathrm{d}y :$$

---

[6]This growth condition can be relaxed into $|f(x)| \leq C \exp\left(\frac{|x|^2}{2\alpha}\right)$ for any x, for some positive
constants C and α: in that case, the smoothness of the function u is satisfied for $t < \alpha$ only.

the function u is infinitely continuously differentiable in space and time for t > 0 and solves the heat equation

$$u'_t(t, x, f) = \frac{1}{2}u''_{xx}(t, x, f), \quad u(0, x, f) = f(x). \tag{15}$$

Equation (15) is the *heat equation with initial boundary condition* (Cauchy problem, see [22]).

Proof. Standard Gaussian estimates allow to differentiate u w.r.t. t or x by differentiating under the integral sign: then, we have

$$u'_t(t, x, f) = \int_{\mathbb{R}} g'_t(t, x, y) f(y) dy = \int_{\mathbb{R}} \frac{1}{2} g''_{xx}(t, x, y) f(y) dy = \frac{1}{2} u''_{xx}(t, x, f).$$

$\square$

When the function considered is regular, another formulation can be given to this relation, which will play a significant role in the following.

Proposition 10. *If f is of class $\mathscr{C}_b^2$ (bounded and twice continuously differentiable with bounded derivatives),[7] we have*

$$u'_t(t, x, f) = u(t, x, \frac{1}{2} f''_{xx}),$$

or equivalently using a probabilistic viewpoint

$$\mathbb{E}[f(x + W_t)] = f(x) + \int_0^t \mathbb{E}\Big[\frac{1}{2} f''_{xx}(x + W_s)\Big] ds. \tag{16}$$

Proof. Write $u(t, x, f) = \mathbb{E}[f(x + W_t)] = \int_{\mathbb{R}} g(t, 0, y) f(x + y) dy = \int_{\mathbb{R}} g(t, x, z) f(z) dz$ and differentiate under the integral sign: it gives

$$u''_{xx}(t, x, f) = \int_{\mathbb{R}} g(t, 0, y) f''_{xx}(x + y) dy = u(t, x, f''_{xx}) = \int_{\mathbb{R}} g''_{xx}(t, x, z) f(z) dz,$$

$$u'_t(t, x, f) = \int_{\mathbb{R}} g'_t(t, x, z) f(z) dz = \frac{1}{2} \int_{\mathbb{R}} g''_{xx}(t, x, z) f(z) dz = \frac{1}{2} u(t, x, f''_{xx}),$$

using at the first line two integration by parts and at the second line the heat equation satisfied by g. Then the probabilistic representation (16) easily follows:

---

[7] Here again, the boundedness could be relaxed to some exponential growth.

$$\mathbb{E}[f(x + W_t)] - f(x) = u(t, x, f) - u(0, x, f) = \int_0^t u_t'(s, x, f)\mathrm{d}s$$

$$= \int_0^t u(s, x, \frac{1}{2} f_{xx}'')\mathrm{d}s = \int_0^t \mathbb{E}[\frac{1}{2} f_{xx}''(x + W_s)]\mathrm{d}s.$$

$$\square$$

2.1.2   Heat Equation in an Interval

We now extend the previous results in two directions: first, we allow the function f to also depend smoothly on time and second, the final time t is replaced by a stopping time U. The first extension is straightforward and we state it without proof.

Proposition 11. *Let f be a function of class $\mathscr{C}_b^{1,2}$ (bounded, once continuously differentiable in time, twice in space, with bounded derivatives): we have*

$$\mathbb{E}[f(t, x + W_t)] = f(0, x) + \int_0^t \mathbb{E}[f_t'(s, x + W_s) + \frac{1}{2} f_{xx}''(s, x + W_s)]\mathrm{d}s$$

$$= f(0, x) + \mathbb{E}\left[\int_0^t (f_t'(s, x + W_s) + \frac{1}{2} f_{xx}''(s, x + W_s))\mathrm{d}s\right].$$

$$(17)$$

The second equality readily follows from Fubini's theorem to invert $\mathbb{E}$ and time integral: this second form is more suitable for an extension to stochastic times t.

Theorem 3. *Let f be a function of class $\mathscr{C}_b^{1,2}$, we have*

$$\mathbb{E}[f(U, x + W_U)] = f(0, x) + \mathbb{E}\left[\int_0^U (f_t'(s, x + W_s) + \frac{1}{2} f_{xx}''(s, x + W_s))\mathrm{d}s\right]$$

$$(18)$$

for any bounded[8] stopping time U.

The above identity between expectations is far to be obvious to establish by hand since the law of U is quite general and an analytical computation is out of reach. This level of generality on U is quite interesting for applications: it provides a powerful tool to determine the distribution of hitting times, to show how often multidimensional Brownian motion visits a given point or a given set. Regarding

---

[8]Meaning that for a deterministic positive constant C, $\mathbb{P}(U \leq C) = 1$.

this lecture, it gives a key tool to derive probabilistic representations of heat equation with Dirichlet boundary conditions.

Proof. Let us start by giving alternatives of the relation (17). We observe that it could have been written with a random initial condition X_0, like for instance

$$\mathbb{E}[\mathbf{1}_{A_0} f(t, X_0 + W_t)]$$
$$= \mathbb{E}\left[\mathbf{1}_{A_0} f(0, X_0) + \mathbf{1}_{A_0} \int_0^t (f'_t(s, X_0 + W_s) + \frac{1}{2} f''_{xx}(s, X_0 + W_s)) ds\right],$$

with W independent of X_0 and where the event A_0 depends on X_0. Similarly, using the time-shifted Brownian motion $\{\bar{W}_t^u = W_{t+u} - W_u; t \in \mathbb{R}^+\}$ that is independent of the initial condition $x + W_u$ (Proposition 2), it leads to

$$\mathbb{E}[\mathbf{1}_{A_u} f(t + u, x + W_u + \bar{W}_t^u)] = \mathbb{E}\left[\mathbf{1}_{A_u} f(u, x + W_u) + \right.$$
$$\left. \mathbf{1}_{A_u} \int_0^t (f'_t(u + s, x + W_u + \bar{W}_s^u) + \frac{1}{2} f''_{xx}(u + s, x + W_u + \bar{W}_s^u)) ds\right]$$

for any event A_u depending only of the values $\{W_s : s \leq u\}$, or equivalently

$$\mathbb{E}[\mathbf{1}_{A_u} f(t + u, x + W_{t+u})] = \mathbb{E}\left[\mathbf{1}_{A_u} f(u, x + W_u)\right.$$
$$\left. + \mathbf{1}_{A_u} \int_u^{t+u} (f'_t(s, x + W_s) + \frac{1}{2} f''_{xx}(s, x + W_s)) ds\right].$$

Set $M_t = f(t, x + W_t) - f(0, x) - \int_0^t (f'_t(s, x + W_s) + \frac{1}{2} f''_{xx}(s, x + W_s)) ds$: our aim is to prove $\mathbb{E}(M_U) = 0$. Observe that the preliminary computation has shown that

$$\mathbb{E}(\mathbf{1}_{A_u}(M_{t+u} - M_u)) = 0 \tag{19}$$

for $t \geq 0$. In particular, taking $A_u = \Omega$ we obtain that the expectation $\mathbb{E}(M_t)$ is constant[9] w.r.t. t.

Now, consider first that U is a discrete stopping time valued in $\{0 = u_0 < u_1 < \cdots < u_n = T\}$: then

$$\mathbb{E}(M_U) = \sum_{k=0}^{n-1} \mathbb{E}(M_{U \wedge u_{k+1}} - M_{U \wedge u_k}) = \sum_{k=0}^{n-1} \mathbb{E}(\mathbf{1}_{U > u_k}(M_{u_{k+1}} - M_{u_k})) = 0$$

---

[9]Actually, (19) proves that M is a martingale and the result to be proved is related to the *optional sampling theorem*.

by applying (19) since $\{U \le u_k\}$ does depend only of $\{W_s : s \le u_k\}$ (by definition of a stopping time). $\qquad\square$

Second, for a general stopping time (bounded by T), we take $U_n = \frac{[nU]+1}{n}$ which is a stopping time converging to U: since $(M_t)_{0 \le t \le T}$ is bounded and continuous, the dominated convergence theorem gives $0 = \mathbb{E}(M_{U_n}) \xrightarrow[n\to\infty]{} \mathbb{E}(M_U)$. $\qquad\square$

As a consequence, we now make explicit the solutions of the heat equation in an interval and with initial condition: it is a partial generalization[10] of Theorem 2, which characterized them in the whole space. The introduction of (non-homogeneous) boundary conditions of Dirichlet type is connected to the passage time of the Brownian motion.

Corollary 1 (Heat Equation with Cauchy–Dirichlet Boundary Condition).
Consider the PDE

$$\begin{cases} u'_t(t,x) = \frac{1}{2}u''_{xx}(t,x), & \text{for } t > 0 \text{ and } x \in]a,b[, \\ u(0,x) = f(0,x) & \text{for } t = 0 \text{ and } x \in [a,b], \\ u(t,x) = f(t,x) & \text{for } x = a \text{ or } b, \text{ with } t \ge 0. \end{cases}$$

If a solution u of class $C_b^{1,2}([0,T] \times [a,b])$ exists, then it is given by

$$u(t,x) = \mathbb{E}[f(t - U, x + W_U)]$$

where $U = T_a \wedge T_b \wedge t$ (using the previous notation for the first passage time T_y at the level y for the Brownian motion starting at x, i.e. $(x + W_t)_{t \ge 0}$).

Proof. First, extend smoothly the function u outside the interval $[a,b]$ in order to apply previous results. The way to extend is unimportant since u and its derivatives are only evaluated inside $[a,b]$. Clearly U is a bounded (by t) stopping time. Apply now the equality (18) to the function $(s,y) \mapsto u(t - s, y) = v(s,y)$ of class $C_b^{1,2}([0,t] \times \mathbb{R})$, satisfying $v'_s(s,y) + \frac{1}{2}v''_{yy}(s,y) = 0$ for $(s,y) \in [0,t] \times [a,b]$. We obtain

$$\mathbb{E}[v(U, x + W_U)] = v(0,x) + \mathbb{E}\left[\int_0^U (v'_s(s, x + W_s) + \frac{1}{2}v''_{yy}(s, x + W_s))ds\right] = v(0,x),$$

since for $s \le U$, $(s, x + W_s) \in [0,t] \times [a,b]$. To conclude, we easily check that $v(0,x) = u(t,x)$ and $v(U, x + W_U) = f(t - U, x + W_U)$. $\qquad\square$

---

[10]Indeed, the result gives uniqueness and not the existence.

2.1.3 A Probabilistic Algorithm to Solve the Heat Equation

To illustrate our purpose, we consider a toy example regarding the numerical evaluation of $u(t, x) = \mathbb{E}(f(x + W_t))$ using random simulations, in order to discuss main ideas underlying to Monte Carlo methods. Actually, the arguments below apply also to $u(t, x) = \mathbb{E}[f(t - U, x + W_U)]$ with $U = T_a \wedge T_b \wedge t$, although there are some extra significant issues in the simulation of (U, W_U).

For the notational simplicity, denote by X the random variable inside the expectation to compute, that is $X = f(x + W_t)$ in our toy example. As a difference with a PDE method (based on finite differences or finite elements), a standard Monte Carlo method provides an approximation of $u(t, x)$ at a given point (t, x), without evaluating the values at other points. Actually, this fact holds because the PDE u is linear; in Sect. 5 related to non-linear PDEs, the situation is different.

The Monte Carlo method is able to provide a convergent, tractable approximation of $u(t, x)$, with a priori error bounds, under two conditions.

1. An arbitrary large number of independent realizations of X can be generated (denote them by $(X_i)_{i \geq 1}$): in our toy example, this is straightforward since it requires only the simulation of W_t which is distributed as a Gaussian r.v. $\mathcal{N}(0, t)$ and then we have to compute $X = f(x + W_t)$. The independence of simulations is achieved by using a *good* generator of random numbers, like the excellent *Mersenne Twister*[11] generator.
2. Additionally, X which is already integrable ($\mathbb{E}|X| < +\infty$) is assumed to be square integrable: $\mathbb{V}\mathrm{ar}(X) < +\infty$.

Then, by the law of large numbers, we have

$$\overline{X}_M = \frac{1}{M} \sum_{i=1}^{M} X_i \underset{M \to +\infty}{\longrightarrow} \mathbb{E}(X); \tag{20}$$

hence the empirical mean of simulations of X provides a convergent approximation of the expectation $\mathbb{E}(X)$. As a difference with PDE methods where some stability conditions may be required (like the Courant–Friedrichs–Lewy condition), the above Monte Carlo method does not require any extra condition to converge: it is *unconditionally convergent*. The extra moment condition is used to define a priori error bounds on the *statistical error*: the approximation error is controlled by means of the Central Limit Theorem

$$\lim_{M \to +\infty} \mathbb{P}\left(\sqrt{\frac{M}{\mathbb{V}\mathrm{ar}(X)}} \left(\overline{X}_M - \mathbb{E}(X) \right) \in [a, b] \right) = \mathbb{P}(G \in [a, b]),$$

---

[11]http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html.

where G is a centered unit variance Gaussian r.v. Observe that the error bounds are stochastic: we can not do better than arguing that with probability $\mathbb{P}(G \in [a, b])$, the unknown expectation (asymptotically as $M \to +\infty$) belongs to the interval

$$\left[\overline{X}_M - b \sqrt{\frac{\mathbb{V}\mathrm{ar}(X)}{M}}, \overline{X}_M - a \sqrt{\frac{\mathbb{V}\mathrm{ar}(X)}{M}} \right].$$

This is known as a confidence interval at level $\mathbb{P}(G \in [a, b])$. The larger a and b, the larger the confidence interval, the higher the confidence probability.

To obtain a fully explicit confidence interval, one may replace $\mathbb{V}\mathrm{ar}(X)$ by its estimator using the same simulations:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \approx \frac{M}{M-1} \left(\frac{1}{M} \sum_{i=1}^{M} X_i^2 - \overline{X}_M^2 \right) := \sigma_M^2.$$

The factor $M/(M-1)$ plays the role of unbiasing[12] the value $\mathbb{V}\mathrm{ar}(X)$, although it is not a big deal for M large ($M \geq 100$). Anyway, we can prove that the above conditional intervals are asymptotically unchanged by taking the empirical variance σ_M^2 instead of $\mathbb{V}\mathrm{ar}(X)$. Gathering these different results and seeking a symmetric confidence interval $-a = b = 1.96$ and $\mathbb{P}(G \in [a, b]) \approx 95\%$, we obtain the following: with probability 95 %, approximatively for M large enough, we have

$$\mathbb{E}(X) \in \left[\overline{X}_M - 1.96 \frac{\sigma_M}{\sqrt{M}}, \overline{X}_M + 1.96 \frac{\sigma_M}{\sqrt{M}} \right]. \tag{21}$$

The symmetric confidence interval at level 99 % is given by $-a = b = 2.58$. Since a Monte Carlo method provides random evaluations of $\mathbb{E}(X)$, different program runs will give different results (as a difference with a deterministic method which systematically has the same output) which seems uncomfortable: that is why it is important to produce a confidence interval. This is also very powerful and useful to have at hand a numerical method able to state that the error is at most of xxx with high probability.

The confidence interval depends on

- The confidence level $\mathbb{P}(G \in [a, b])$, chosen by the user.
- The number of simulations: improving the accuracy by a factor 2 requires 4 times more simulations.
- The variance $\mathbb{V}\mathrm{ar}(X)$ or its estimator σ_M^2, which depends on the problem to handle (and not much on M as soon as M is large). This variance can be very different from one problem to another: on Fig. 6, see the width of confidence intervals for two similar computations. There exist variance reduction techniques

---

[12]Indeed, we can show that $\mathbb{E}(\sigma_M^2) = \mathbb{V}\mathrm{ar}(X)$.
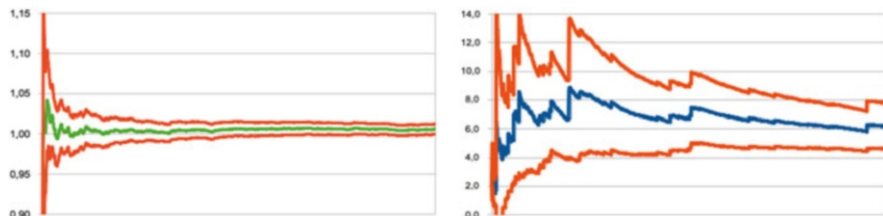
Fig. 6 Monte Carlo computations of $\mathbb{E}(e^{G/10}) = e^{\frac{1}{2}\frac{1}{10^2}} \approx 1.005$ on *the left* and $\mathbb{E}(e^{2G}) = e^{\frac{1}{2}2^2} \approx 7.389$ on *the right*, where G is a Gaussian r.v. with zero mean and unit variance. The empirical mean and the symmetric 95 %-confidence intervals are plotted w.r.t. the number of simulations

able to significantly reduce this factor in order to provide thinner confidence intervals while maintaining the same computational cost.

Another advantage of such a Monte Carlo algorithm is the simplicity of code, consisting of one loop on the number of simulations; within this loop, the empirical variance should be simultaneously computed. However, the simulation procedure of X can be delicate in some situations, see Sect. 4.

At last, we focus our discussion on the impact of the dimension of the underlying PDE, which has been equal to 1 so far. Consider now a state variable in $\mathbb{R}^d$ ($d \geq 1$) and a heat equation with Cauchy initial condition in dimension d; (15) becomes

$$u_t'(t, x, f) = \frac{1}{2}\Delta u(t, x, f), \quad u(0, x, f) = f(x), \quad t > 0, x \in \mathbb{R}^d, \qquad (22)$$

where $\Delta = \sum_{i=1}^d \partial_{x_i x_i}^2$ stands for the Laplacian in $\mathbb{R}^d$. Using similar arguments as in dimension 1, we check that

$$u(t, x, f) = \int_{\mathbb{R}^d} \frac{1}{(2\pi t)^{d/2}} \exp(-|y - x|^2/2t) f(y)\mathrm{d}y = \mathbb{E}[f(x + W_t)]$$

where $W = \begin{pmatrix} W_1 \\ \vdots \\ W_d \end{pmatrix}$ is a d-dimensional Brownian motion, i.e. each W_i is a one-

dimensional Brownian motion and the d components are independent (Fig. 7).

- The Monte Carlo computation of $u(t, x)$ is then achieved using independent simulations of $X = f(x + W_t)$: the accuracy is then of order $1/\sqrt{N}$ and the computational effort is $N \times d$. Thus, the dimension has a very low effect on the complexity of the algorithm.
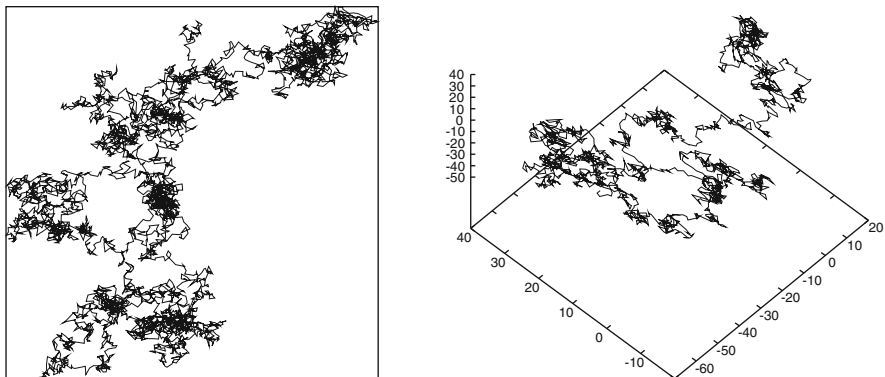
Fig. 7 Brownian motion in dimension 2 and 3

- As a comparison with a PDE discretization scheme, to achieve an accuracy of order $1/N$, we essentially[13] need N points in each spatial direction and it follows that the resulting linear system to invert is of size N^d: thus, without going into full details, it is clear that the computational cost to achieve a given accuracy depends much on the dimension d. And the situation becomes less and less favourable as the dimension increases. Also, the memory required to run a PDE algorithm increases exponentially with the dimension, as a difference with a Monte Carlo approach.

It is commonly admitted that a PDE approach is more suitable and efficient in dimension 1 and 2, whereas a Monte Carlo procedure is more adapted for higher dimensions. On the other hand, a PDE-based method computes a global approximation of u (at any point (t, x)), while a Monte Carlo scheme gives a pointwise approximation only. The probabilistic approach can be directly used for Parallel Computing, each processor being in charge of a bunch of simulations at a given point (t, x).

2.2 PDE Associated to Other Processes

We extend the Feynman–Kac representation for the Brownian motion to the Arithmetic Brownian Motion and the Ornstein–Uhlenbeck process.

---

[13]In fact, it generally depends on the regularity of u.

2.2.1 Arithmetic Brownian Motion

First consider the Arithmetic Brownian motion defined by $\{X_t^x = x + bt + \sigma W_t, t \geq 0\}$. The distribution of X_t is Gaussian with mean $x + bt$ and variance $\sigma^2 t$: we assume in the following that $\sigma \neq 0$ which ensures that its density exists and is given by

$$g_{b,\sigma^2}(t, x, y) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp -\frac{(y - x - bt)^2}{2\sigma^2 t} = g(\sigma^2 t, x + bt, y)$$
$$= g(\sigma^2 t, x, y - bt).$$

Denote by $L_{b,\sigma^2}^{\mathrm{ABM}}$ the second order operator

$$L_{b,\sigma^2}^{\mathrm{ABM}} = \frac{1}{2}\sigma^2 \partial_{xx}^2 + b\partial_x , \tag{23}$$

also called *infinitesimal generator*[14] of X. A direct computation using the heat equation for $g(t, x, y)$ gives

$$\partial_t g_{b,\sigma^2}(t, x, y) = \frac{1}{2}\sigma^2 g_{xx}''(\sigma^2 t, x+bt, y) + bg_x'(\sigma^2 t, x+bt, y) = L_{b,\sigma^2}^{\mathrm{ABM}} g_{b,\sigma^2}(t, x, y).$$

Hence, multiplying by $f(y)$ and integrating over $y \in \mathbb{R}$, we obtain the following representation that generalizes Theorem 2.

Theorem 4. *Let f be a bounded measurable function. The function*

$$u_{b,\sigma^2}(t, x, f) = \mathbb{E}[f(X_t^x)] = \int_{\mathbb{R}} g_{b,\sigma^2}(t, x, y) f(y)\mathrm{d}y \tag{24}$$

solves

$$\begin{cases} u_t'(t, x, f) = L_{b,\sigma^2}^{\mathrm{ABM}} u(t, x, f) = \frac{1}{2}\sigma^2 u_{xx}''(t, x, f) + bu_x'(t, x, f), \\ u(0, x, f) = f(x). \end{cases} \tag{25}$$

The extension of Propositions 10 and 11 follows the arguments used for the BM case.

Proposition 12. *If $f \in \mathscr{C}_b^{1,2}$ and U is a bounded stopping time (including deterministic time), then*

---

[14]This labeling comes from the infinitesimal decomposition of $\mathbb{E}(f(X_t))$ as time is small, $\partial_t \mathbb{E}(f(X_t))|_{t=0} = L_{b,\sigma^2}^{\mathrm{ABM}} f(x)$, see Proposition 12.

$$\mathbb{E}[f(U, X_U^x)] = f(0, x) + \mathbb{E}\Big[\int_0^U [L_{b,\sigma^2}^{ABM} f(s, X_s^x) + f_t'(s, X_s^x)]\mathrm{d}s\Big].$$

Theorem 4 gives the Feynman–Kac representation of the Cauchy problem written w.r.t. the second order operator L_{b,σ^2}^{ABM}. When Dirichlet boundary conditions are added, Corollary 1 extends as follows, using Proposition 12.

Corollary 2. *Assume the existence of a solution u of class $C_b^{1,2}([0, T] \times [a, b])$ to the PDE*

$$\begin{cases} u_t'(t, x, f) = L_{b,\sigma^2}^{ABM}u(t, x, f), & \text{for } t > 0 \text{ and } x \in]a, b[, \\ u(0, x, f) = f(0, x) & \text{for } t = 0 \text{ and } x \in [a, b], \\ u(t, x, f) = f(t, x) & \text{for } x = a \text{ or } b, \text{ with } t \geq 0. \end{cases}$$

Then it is given by

$$u(t, x) = \mathbb{E}[f(t - U^x, X_{U^x}^x)]$$

where $U^x = \inf\{s > 0 : X_s^x \notin]a, b[\} \wedge t$ is the first exit time from the interval $]a, b[$ by the process X^x before t.

As for the standard heat equation, this representation naturally leads to a probabilistic algorithm to compute the PDE solution, by empirical mean of independent simulation of $f(t - U^x, X_{U^x}^x)$.

2.2.2 Ornstein–Uhlenbeck Process

Now consider the process solution to $V_t^x = x - a \int_0^t V_s^x \mathrm{d}s + \sigma W_t$: we emphasize in our notation the dependence w.r.t. the initial value $V_0 = x$. We define an appropriate second order operator

$$L_{a,\sigma^2}^{OU} g(t, x) = \frac{1}{2}\sigma^2 g_{xx}''(t, x) - ax g_x'(t, x)$$

which plays the role of the infinitesimal generator for the Ornstein–Uhlenbeck process. We recall that the Gaussian distribution of V_t^x has mean xe^{-at} and variance $\frac{\sigma^2}{2a}(1 - e^{-2at})$, the density of which at y (assuming $\sigma \neq 0$ for the existence) is

$$p(t, x, y) = g(v_t, xe^{-at}, y).$$

Using the heat equation satisfied by g, we easily derive that

$$p_t'(t, x, y) = \frac{1}{2}\sigma^2 p_{xx}''(t, x, y) - ax p_x'(t, x, y) = L_{a,\sigma^2}^{OU} p(t, x, y),$$

from which we deduce the PDE satisfied by $u(t, x, f) = \mathbb{E}[f(V_t^x)]$. Incorporating Dirichlet boundary conditions is similar to the previous cases. We state the related results without extra details.

Theorem 5. *Let f be a bounded measurable function. The function*

$$u(t, x, f) = \mathbb{E}[f(V_t^x)] = \int_{\mathbb{R}} p(t, x, y) f(y) \mathrm{d}y$$

solves

$$\begin{cases} u_t'(t, x, f) = L_{a,\sigma^2}^{OU} u(t, x, f), \\ u(0, x, f) = f(x). \end{cases}$$

Proposition 13. *If $f \in \mathscr{C}_b^{1,2}$ and U is a bounded stopping time, then*

$$\mathbb{E}[f(U, V_U^x)] = f(0, x) + \mathbb{E}\left[\int_0^U [L_{a,\sigma^2}^{OU} f(s, V_s^x) + f_t'(s, V_s^x)] \mathrm{d}s \right].$$

Corollary 3. *Assume the existence of a solution u of class $C_b^{1,2}([0, T] \times [a, b])$ to the PDE*

$$\begin{cases} u_t'(t, x, f) = L_{a,\sigma^2}^{OU} u(t, x, f), & \text{for } t > 0 \text{ and } x \in]a, b[, \\ u(0, x, f) = f(0, x) & \text{for } t = 0 \text{ and } x \in [a, b], \\ u(t, x, f) = f(t, x) & \text{for } x = a \text{ or } b, \text{ with } t \geq 0. \end{cases}$$

Then u is given by

$$u(t, x) = \mathbb{E}[f(t - U^x, V_{U^x}^x)]$$

where $U^x = \inf\{s > 0 : V_s^x \notin]a, b[\} \wedge t$.

2.2.3 A Natural Conjecture for Stochastic Differential Equations

The previous examples serve as a preparation for more general results, relating the dynamics of a process and its Feynman–Kac representation. Denote X^x the solution (whenever it exists) to the Stochastic Differential Equation

$$X_t^x = x + \int_0^t b(X_s^x) \mathrm{d}s + \int_0^t \sigma(X_s^x) \mathrm{d}W_s, \quad t \geq 0.$$

In view of the results in simpler models, we announce the following facts.

1. Set $L^X_{b,\sigma^2}g = \frac{1}{2}\sigma^2(x)g''_{xx} + b(x)g'_x$.
2. $u(t,x) = \mathbb{E}(f(X^x_t))$ solves

$$u'_t(t,x) = L^X_{b,\sigma^2}u(t,x), \quad u(0,x) = f(x).$$

3. If $f \in \mathscr{C}^{1,2}_b$ and U is a bounded stopping time, then

$$\mathbb{E}[f(U,X^x_U)] = f(0,x) + \mathbb{E}\Big[\int_0^U [L^X_{b,\sigma^2}f(s,X^x_s) + f'_t(s,X^x_s)]\mathrm{d}s\Big].$$

4. If u of class $C^{1,2}_b([0,T]\times[a,b])$ solves the PDE

$$\begin{cases} u'_t(t,x) = L^X_{b,\sigma^2}u(t,x), & \text{for } t > 0 \text{ and } x \in]a,b[, \\ u(0,x) = f(0,x) & \text{for } t = 0 \text{ and } x \in [a,b], \\ u(t,x) = f(t,x) & \text{for } x = a \text{ or } b, \text{ with } t \geq 0, \end{cases}$$

then it is given by $u(t,x) = \mathbb{E}[f(t-U^x, X^x_{U^x})]$ where $U^x = \inf\{s > 0 : X^x_s \notin]a,b[\} \wedge t$.

The above result could be extended to PDE with a space variable in $\mathbb{R}^d$ ($d \geq 1$) by considering a $\mathbb{R}^d$-valued SDE: it would be achieved by replacing W by a d-dimensional standard Brownian motion, by having a drift coefficient $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ and a diffusion coefficient $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^d \otimes \mathbb{R}^d$, a reward function $f : [0,T] \times \mathbb{R}^d \mapsto \mathbb{R}$, by replacing the interval $[a,b]$ by a domain D in $\mathbb{R}^d$ and defining U^x as the first exit time by X^x from that domain. Then the operator L would be a linear parabolic second order operator of the form

$$L^X_{b,\sigma\sigma^\top}g = \frac{1}{2}\sum_{i,j=1}^d [\sigma\sigma^\top]_{i,j}(x)\partial^2_{x_ix_j}g + \sum_{i=1}^d b_i(x)\partial_{x_i}g,$$

where $\top$ denotes the transpose. We could also add a zero-order term in $L^X_{b,\sigma\sigma^\top}$, by considering a discounting factor for f; we do not develop further this extension.

The next section provides stochastic calculus tools, that allow to show the validity of these Feynman–Kac type results, under some appropriate smoothness and growth assumptions on b, σ, f. To allow non smooth f or Dirichlet boundary conditions, we may additionally assume a non-degeneracy condition on $L^X_{b,\sigma\sigma^\top}$ (like ellipticity condition $|\sigma\sigma^\top(x)| \geq \frac{1}{c}$ for some $c > 0$).

Complementary References. See [1, 15, 20, 22, 23, 48].

3 The Itô Formula

One achievement of Itô's formula is to go from an infinitesimal time-decomposition in expectation like

$$\mathbb{E}[f(t, x + W_t)] - f(0, x) = \int_0^t \mathbb{E}[f_t'(s, x + W_s) + \frac{1}{2} f_{xx}''(s, x + W_s)]\mathrm{d}s$$

(see (17)) to a pathwise infinitesimal time-decomposition of

$$f(t, x + W_t) - f(0, x).$$

Since Brownian motion paths are not differentiable, it is hopeless to apply standard differential calculus based on usual first order Taylor formula. Instead of this, we go up to the second order, taking advantage of the fact that W has a finite quadratic variation. The approach presented below is taken from the nice paper *Calcul d'Itô sans probabilité* by Föllmer [19], which does not lead to the most general and deepest approach but it has the advantage of light technicalities and straightforward arguments compared to the usual tough arguments using L_2-spaces and isometry (see for instance [48] or [63] among others).

3.1 Quadratic Variation

3.1.1 Notations and Definitions

Brownian increments in a small interval $[t, t + h]$ are centered Gaussian r.v. with variance h, which thus behave like $\sqrt{h}$. The total variation does not exist, because the trajectories are not differentiable, but the quadratic variation has interesting properties.

To avoid convergence technicalities, we consider particular time subdivisions.

Definition 7 (Dyadic Subdivision of Order n). Let n be an integer. The subdivision of $\mathbb{R}^+$ defined by $\mathbb{D}_n = \{t_0 < \cdots < t_i < \cdots\}$ where $t_i = i2^{-n}$ is called the dyadic subdivision of order n. The subdivision step is $\delta_n = 2^{-n}$.

Definition 8 (Quadratic Variation). The quadratic variation of a Brownian motion W associated with the dyadic subdivision of order n is defined, for $t \geq 0$, by

$$V_t^n = \sum_{t_i \leq t} (W_{t_{i+1}} - W_{t_i})^2. \tag{26}$$

3.1.2   Convergence

Then there is the following remarkable result.

Proposition 14 (Pointwise Convergence). *With probability 1, we have*

$$\lim_{n \to \infty} V_t^n = t$$

for any $t \in \mathbb{R}^+$.

Had W been differentiable, the limit of V^n would be equal to 0.

Proof. First let us show the *a.s.* convergence for a fixed time t, and denote by $n(t)$ the index of the dyadic subdivision of order n such that $t_{n(t)} \le t < t_{n(t)+1}$. Then observe that $V_t^n - t = \sum_{j=0}^{n(t)} Z_j + (t_{n(t)+1} - t)$ where $Z_j = (W_{t_{j+1}} - W_{t_j})^2 - (t_{j+1} - t_j)$. The term $t_{n(t)+1} - t$ converges to 0 as the subdivision step shrinks to 0. The random variables Z_j are independent, centered, square integrable (since the Gaussian law of $W_{t_{j+1}} - W_{t_j}$ has finite fourth moments): additionally, the scaling property of Proposition 1 ensures that $\mathbb{E}(Z_j^2) = C_2(t_{j+1} - t_j)^2$ for a positive constant C_2. Thus

$$\mathbb{E}\left(\sum_{j=0}^{n(t)} Z_j\right)^2 = \sum_{j=0}^{n(t)} \mathbb{E}\left(Z_j^2\right) = \sum_{j=0}^{n(t)} C_2(t_{j+1} - t_j)^2 \le C_2(T+1)\delta_n.$$

This proves the L_2-convergence of $\sum_{j=0}^{n(t)} Z_j$ towards 0.

  Moreover we obtain $\sum_{n \ge 1} \mathbb{E}\left(\sum_{j=0}^{n(t)} Z_j\right)^2 < \infty$, i.e. the random series $\sum_{n \ge 1} \left(\sum_{j=0}^{n(t)} Z_j\right)^2$ has a finite expectation, whence *a.s.* finite and consequently its general term converges *a.s.* to 0. This shows that for any fixed t, $V_t^n \to t$ except on a negligible set N_t.

  We now extend the result to any time: first the set $N = \cup_{t \in \mathbb{Q}^+} N_t$ is still negligible because the union of negligible sets is taken on a countable family. For an arbitrary t, take two monotone sequences of rational numbers $r_p \uparrow t$ and $s_p \downarrow t$ as $p \to +\infty$. Since $t \mapsto V_t^n$ is increasing for fixed n, we deduce, for any $\omega \notin N$

$$r_p = \lim_{n \to \infty} V_{r_p}^n(\omega) \le \liminf_{n \to \infty} V_t^n(\omega) \le \limsup_{n \to \infty} V_t^n(\omega) \le \lim_{n \to \infty} V_{s_p}^n(\omega) = s_p.$$

Passing to the limit in p gives the result.   □

  As a consequence, we obtain the formula giving the infinitesimal decomposition of W_t^2.

Proposition 15 (A First Itô Formula). *Let W be a standard Brownian motion. With probability 1, we have for any $t \ge 0$*

$$W_t^2 = 2 \int_0^t W_s \mathrm{d}W_s + t \tag{27}$$

where the stochastic integral $\int_0^t W_s dW_s$ is the a.s. *limit of $\sum_{t_i \le t} W_{t_i}(W_{t_{i+1}} - W_{t_i})$, along the dyadic subdivision.*

For usual C^1-function $f(t)$, we have $f^2(t) - f^2(0) = 2 \int_0^t f(s) \mathrm{d}f(s)$: the extra term t in (27) is intrinsically related to Brownian motion paths.

Proof. Adopting once again the notation with $n(t)$, we have

$$W_t^2 = W_t^2 - W_{t_{n(t)+1}}^2 + \sum_{t_i \le t}(W_{t_{i+1}}^2 - W_{t_i}^2)$$

$$= W_t^2 - W_{t_{n(t)+1}}^2 + \sum_{t_i \le t}(W_{t_{i+1}} - W_{t_i})^2 + 2\sum_{t_i \le t} W_{t_i}(W_{t_{i+1}} - W_{t_i}).$$

The first term at the r.h.s. tends towards 0 by continuity of the Brownian paths. The second term is equal to V_t^n and converges towards t. Consequently, the third term at the right-hand side *must* converge *a.s.* towards a term that we call *stochastic integral* and that we denote by $2 \int_0^t W_s \mathrm{d}W_s$. ☐

The random function V_t^n, as a function of t, is increasing and can be associated to the cumulative distribution function of the positive discrete measure

$$\sum_{i \ge 0}(W_{t_{i+1}} - W_{t_i})^2 \delta_{t_i}(.) = \mu^n(.)$$

satisfying $\mu_n(f) = \sum_{i \ge 0} f(t_i)(W_{t_{i+1}} - W_{t_i})^2$.

The convergence of cumulative distribution function of $\mu^n(.)$ (Proposition 14) can then be extended to integrals of continuous functions (possibly random as well). It is the purpose of the following result which is of deterministic nature.

Proposition 16 (Convergence as a Positive Measure). *For any continuous function f, with probability 1 we have*

$$\lim_{n \to \infty} \sum_{t_i \le t} f(t_i)(W_{t_{i+1}} - W_{t_i})^2 = \int_0^t f(s) \mathrm{d}s$$

for any $t \ge 0$.

The proof is standard: the result first holds for functions of the form $f(s) = \mathbf{1}_{]r_1, r_2]}(s)$, then for piecewise constant functions, at last for continuous functions by simple approximations.

3.2   The Itô Formula for Brownian Motion

Differential calculus extends to other functions than $x \to x^2$. To the usual classical formula with functions that are smooth in time, a term should be added, due to the non-zero quadratic variation.

Theorem 6 (Itô Formula). *Let $f \in \mathscr{C}^{1,2}(\mathbb{R}^+ \times \mathbb{R}, \mathbb{R})$. Then with probability 1, we have $t \geq 0$*

$$f(t, x + W_t) = f(0, x) + \int_0^t f_x'(s, x + W_s)\, dW_s$$

$$+ \int_0^t f_t'(s, x + W_s)\, ds + \frac{1}{2}\int_0^t f_{xx}''(s, x + W_s)\, ds. \quad (28)$$

The term $\mathscr{I}_t(f) = \int_0^t f_x'(s, x + W_s)\,dW_s$ is called the stochastic integral of $f_x'(s, x + W_s)$ w.r.t. W and it is the a.s. *limit of*

$$\mathscr{I}_t^n(f, W) = \sum_{t_i \leq t} f_x'(t_i, x + W_{t_i})(W_{t_{i+1}} - W_{t_i})$$

taken along the dyadic subdivision of order n.

The reader should compare the equality (28) with (17) to see that, under the extra assumptions that f is bounded with bounded derivatives, we have proved that the stochastic integral $\mathscr{I}_t(f)$ is centered:

$$\mathbb{E}\left(\int_0^t f_x'(s, x + W_s)\, dW_s\right) = 0. \quad (29)$$

This explains how we can expect to go from (28) to (17):

1. Apply Itô formula.
2. Take expectation.
3. Prove that the stochastic integral is centered.

This is an interesting alternative proof to the property satisfied by the Gaussian kernel, which is difficult to extend to more general (non Gaussian) process.

Proof. As before, let us introduce the index $n(t)$ such that $t_{n(t)} \leq t < t_{n(t)+1}$; then we can write

$$f(t, x + W_t) = f(0, x) + [f(t, x + W_t) - f(t_{n(t)+1}, x + W_{t_{n(t)+1}})]$$

$$+ \sum_{t_i \leq t}[f(t_{i+1}, x + W_{t_{i+1}}) - f(t_i, x + W_{t_{i+1}})]$$

$$+ \sum_{t_i \leq t}[f(t_i, x + W_{t_{i+1}}) - f(t_i, x + W_{t_i})].$$

- The second term of the r.h.s. $[f(t, x + W_t) - f(t_{n(t)+1}, x + W_{t_{n(t)+1}})]$ converges to 0 by continuity of $f(t, x + W_t)$.
- The third term is analyzed by means of the first order Taylor formula:

$$f(t_{i+1}, x + W_{t_{i+1}}) - f(t_i, x + W_{t_{i+1}}) = f_t'(\tau_i, x + W_{t_{i+1}})(t_{i+1} - t_i)$$

for $\tau_i \in]t_i, t_{i+1}[$. The uniform continuity of $(W_s)_{0 \le s \le t+1}$ ensures that $\sup_i |f_t'(\tau_i, x + W_{t_{i+1}}) - f_t'(\tau_i, x + W_{t_i})| \to 0$: thus $\lim_{n\to\infty} \sum_{t_i \le t} f_t'(\tau_i, x + W_{t_{i+1}})(t_{i+1} - t_i)$ equals to

$$\lim_{n\to\infty} \sum_{t_i \le t} f_t'(t_i, x + W_{t_i})(t_{i+1} - t_i) = \int_0^t f_t'(s, x + W_s)\mathrm{d}s.$$

- A second order Taylor formula allows to write the fourth term: $f(t_i, x + W_{t_{i+1}}) - f(t_i, x + W_{t_i})$ equals

$$f_x'(t_i, x + W_{t_i})(W_{t_{i+1}} - W_{t_i}) + \frac{1}{2} f_{xx}''(t_i, x + \xi_i)(W_{t_{i+1}} - W_{t_i})^2$$

where $\xi_i \in (W_{t_i}, W_{t_{i+1}})$. Similarly to before, $\sup_i |f_{xx}''(t_i, x + \xi_i) - f_{xx}''(t_i, x + W_{t_i})| = \epsilon_n \to 0$ and it leads to

$$\left| \sum_{t_i \le t} (f_{xx}''(t_i, x + \xi_i) - f_{xx}''(t_i, x + W_{t_i}))(W_{t_{i+1}} - W_{t_i})^2 \right| \le \epsilon_n V_t^n,$$

$$\lim_{n\to\infty} \sum_{t_i \le t} f_{xx}''(t_i, x + W_{t_i})(W_{t_{i+1}} - W_{t_i})^2 = \int_0^t f_{xx}''(s, x + W_s)\mathrm{d}s,$$

by applying Proposition 16.

Observe that in spite of the non-differentiability of W, $\sum_{t_i \le t} f_x'(t_i, x + W_{t_i})(W_{t_{i+1}} - W_{t_i})$ is necessarily convergent as a difference of convergent terms. □

Interestingly, we obtain a representation of the random variable $f(x + W_t)$ as a stochastic integral, in terms of the derivatives of solution u to the heat equation

$$u_t'(t, x) = \frac{1}{2} u_{xx}''(t, x), \quad u(0, x) = f(x).$$

Corollary 4. *Assume that $u \in \mathscr{C}_b^{1,2}([0, T] \times \mathbb{R})$. We have*

$$f(x + W_T) = u(T, x) + \int_0^T u_x'(T - s, x + W_s)\mathrm{d}W_s. \tag{30}$$

Proof. Apply the Itô formula to $v(t,x) = u(T-t,x)$ (which satisfies $v'_t(t,x) + \frac{1}{2}v''_{xx}(t,x) = 0$) at time T. This gives $f(x + W_T) = u(0, x + W_T) = u(T,x) + \int_0^T u'_x(T-s, x+W_s)\mathrm{d}W_s$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

This representation formula leads to important remarks.

- If the above stochastic integral has zero expectation (as for the examples presented before), taking the expectation shows that

$$u(T,x) = \mathbb{E}(f(x + W_T)),$$

  recovering the Feynman–Kac representation of Theorem 2.
- Then, the above representation writes, setting $\Psi = f(x + W_T)$,

$$\Psi = \mathbb{E}(\Psi) + \int_0^T h_s \mathrm{d}W_s.$$

  Actually, a similar stochastic integral representation theorem holds in a larger generality on the form of Ψ, since any bounded[15] functional of $(W_t)_{0 \le t \le T}$ can be represented as its expectation plus a stochastic integral: the process h is not tractable in general, whereas here it is explicitly related to the derivative of u along the Brownian path.
- Assume $u \in \mathscr{C}_b^{1,2}([0,T] \times \mathbb{R})$ imposes that $f \in \mathscr{C}_b^2(\mathbb{R})$ which is too strong for many applications: however, the assumptions on u can be relaxed to handle bounded measurable function f, because the heat equation is immediately smoothing out the initial condition. The proof of this extension involves extra stochastic calculus technicalities that we do not develop.

3.3 Wiener Integral

In general, it is not possible to make explicit the law of the stochastic integral $\int_0^t f'_x(s, x + W_s)\mathrm{d}W_s$, except in a situation where $f'_x(s,x) = h(s)$ is independent of x and square integrable. In that case, $\int_0^t h(s)\mathrm{d}W_s$ is distributed as a Gaussian r.v. The resulting stochastic integral is called *Wiener integral*. We sum up its important properties.

Proposition 17 (Wiener Integral and Integration by Parts). *Let $f : [0,T] \mapsto \mathbb{R}$ be a continuously differentiable function, with bounded derivatives on $[0,T]$.*

1. With probability 1, for any $t \in [0,T]$ we have

---

[15]Integrability is the right assumption.

$$\int_0^t f(s)\mathrm{d}W_s = f(t)W_t - \int_0^t W_s f'(s)\mathrm{d}s. \tag{31}$$

2. *The process $\{\int_0^t f(s)\mathrm{d}W_s\,; t \in [0,T]\}$ is a continuous Gaussian process, with zero mean and with a covariance function*

$$\mathbb{C}\mathrm{ov}\left(\int_0^t f(u)\mathrm{d}W_u, \int_0^s f(u)\mathrm{d}W_u\right) = \int_0^{s\wedge t} f^2(u)\mathrm{d}u. \tag{32}$$

3. *For another function g satisfying the same assumptions, we have*

$$\mathbb{C}\mathrm{ov}\left(\int_0^t f(u)\mathrm{d}W_u, \int_0^s g(u)\mathrm{d}W_u\right) = \int_0^{s\wedge t} f(u)g(u)\mathrm{d}u. \tag{33}$$

Proof. The first item is a direct application of Theorem 6 to the function $(t,x) \mapsto f(t)x$.

For any coefficients $(\alpha_i)_{1\le i\le N}$ and times $(T_i)_{1\le i\le N}$, $\sum_{i=1}^N \alpha_i \int_0^{T_i} f(u)\mathrm{d}W_u$ is a Gaussian r.v. since it can written as a limit of Gaussian r.v. of the form $\sum_j \beta_j (W_{t_{j+1}} - W_{t_j})$: thus, $\{\int_0^t f(s)\mathrm{d}W_s\,; t \in [0,T]\}$ is a Gaussian process. Its continuity is obvious in view of (31). Its expectation is the limit of the expectation of $\sum_{t_i\le t} f(t_i)[W_{t_{i+1}} - W_{t_i}]$, thus equal to 0. The covariance is the limit of the covariance

$$\mathbb{C}\mathrm{ov}\left(\sum_{t_i\le t} f(t_i)[W_{t_{i+1}} - W_{t_i}], \sum_{t_i\le s} f(t_i)[W_{t_{i+1}} - W_{t_i}]\right)$$

$$= \sum_{t_i\le t, t_j\le s} f(t_i)f(t_j)\mathbb{C}\mathrm{ov}(W_{t_{i+1}} - W_{t_i}, W_{t_{i+1}} - W_{t_i})$$

$$= \sum_{t_i\le t, t_j\le s} f(t_i)f(t_j)\delta_{i,j}(t_{i+1} - t_i) \xrightarrow[n\to+\infty]{} \int_0^{s\wedge t} f^2(u)\mathrm{d}u.$$

The second item is proved. The last item is proved similarly. □

As a consequence, going back to the Ornstein–Uhlenbeck process (Sect. 1.6.2), we can complete the proof of its representation (11) using a stochastic integral, starting from (10). For this apply the result below to the function $f(s) = e^{-a(t-s)}$ (t fixed): it gives $\int_0^t e^{-a(t-s)}\mathrm{d}W_s = W_t - a\int_0^t e^{-a(t-s)}W_s\mathrm{d}s$. It leads to

$$V_t = v_0 e^{-at} + \sigma \int_0^t e^{-a(t-s)}\mathrm{d}W_s. \tag{34}$$

Then the Gaussian property from Proposition 17 gives that the variance of V_t is equal to $\sigma^2 \int_0^t e^{-2a(t-s)}\mathrm{d}s = \frac{\sigma^2}{2a}(1 - e^{-2at})$.

3.4 Itô Formula for Other Processes

The reader should have noticed that the central property for the proof of Theorem 6 is that the Brownian motion has a finite quadratic variation. Thus, the Itô formula can directly be extended to processes X which enjoy the same property.

3.4.1 The One-Dimensional Case

In this paragraph, we first consider scalar processes. The multidimensional extension is made afterwards.

Definition 9 (Quadratic Variation of a Process). A continuous process X has a finite quadratic variation if for any $t \geq 0$, the limit

$$V_t^n = \sum_{t_i \leq t} (X_{t_{i+1}} - X_{t_i})^2 \tag{35}$$

along the dyadic subdivision of order n, exists *a.s.* and is finite. We denote this limit by $\langle X \rangle_t$ and it is usually called the *bracket* of X at time t.

If $X = W$ is a Brownian motion, we have $\langle X \rangle_t = t$. More generally, it is easy to check that $\langle X \rangle$ is increasing and continuous. We associate to it a positive measure and this extends Proposition 16 to X.

Proposition 18. *For any continuous function f, with probability 1 for any $t \geq 0$ we have*

$$\lim_{n \to \infty} \sum_{t_i \leq t} f(t_i)(X_{t_{i+1}} - X_{t_i})^2 = \int_0^t f(s) \mathrm{d}\langle X \rangle_s.$$

Theorem 6 becomes

Theorem 7 (Itô Formula for X). *Let $f \in \mathscr{C}^{1,2}(\mathbb{R}^+ \times \mathbb{R}, \mathbb{R})$ and X be with finite quadratic variation. With probability 1, for any $t \geq 0$ we have*

$$f(t, X_t) = f(0, X_0) + \int_0^t f_x'(s, X_s) \, \mathrm{d}X_s + \int_0^t f_t'(s, X_s) \, \mathrm{d}s$$

$$+ \frac{1}{2} \int_0^t f_{xx}''(s, X_s) \, \mathrm{d}\langle X \rangle_s, \tag{36}$$

where $\int_0^t f_x'(s, X_s)\mathrm{d}X_s$ is the stochastic integral of $f_x'(s, X_s)$ w.r.t. X and it is the a.s. limit of $\sum_{t_i \leq t} f_x'(t_i, X_{t_i})(X_{t_{i+1}} - X_{t_i})$ along dyadic subdivision of order n.

Often, the Itô formula is written formally under a differential form

$$\mathrm{d}f(t, X_t) = f_x'(t, X_t)\,\mathrm{d}X_t + f_t'(t, X_t)\,\mathrm{d}t + \frac{1}{2}f_{xx}''(t, X_t)\,\mathrm{d}\langle X\rangle_t.$$

We now provide hand-made tools to compute the bracket of X in practice.

Proposition 19 (Computation of the Bracket). *Let A and M two continuous processes such that A has a finite variation[16] and M has a finite quadratic variation:*

1. $\langle A\rangle_t = 0$.
2. *If $X_t = x + M_t$, then $\langle X\rangle_t = \langle M\rangle_t$.*
3. *If $X_t = \lambda M_t$, then $\langle X\rangle_t = \lambda^2\langle M\rangle_t$.*
4. *If $X_t = M_t + A_t$, then $\langle X\rangle_t = \langle M\rangle_t$.*
5. *If $X_t = f(A_t, M_t)$ with $f \in C^1$, then $\langle X\rangle_t = \int_0^t [f_m'(A_s, M_s)]^2\mathrm{d}\langle M\rangle_s$.*

The proof is easy and it uses deterministic arguments based on the definition of $\langle X\rangle$, we skip it. Item (5) shows that the class of processes with finite quadratic variation is stable by smooth composition. The following examples are important.

Example 1 (Arithmetic Brownian Motion). $(X_t = x + bt + \sigma W_t)$: we have

$$\langle X\rangle_t = \langle \sigma W\rangle_t = \sigma^2\langle W\rangle_t = \sigma^2 t.$$

Itô's formula becomes

$$\mathrm{d}f(t, X_t) = (f_t'(t, X_t) + f_x'(t, X_t)b + \frac{1}{2}f_{xx}''(t, X_t)\sigma^2)\mathrm{d}t + f_x'(t, X_t)\sigma\mathrm{d}W_t$$

$$:= (f_t'(t, X_t) + L_{b,\sigma^2}^{\mathrm{ABM}}f(t, X_t))\mathrm{d}t + f_x'(t, X_t)\sigma\mathrm{d}W_t. \tag{37}$$

An important example is associated to $f(x) = \exp(x)$:

$$\mathrm{d}[\exp(X_t)] = \exp(X_t)(b + \frac{1}{2}\sigma^2)\mathrm{d}t + \exp(X_t)\sigma\mathrm{d}W_t. \tag{38}$$

Example 2 (Geometric Brownian Motion). $(S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t})$: we have

$$\langle S\rangle_t = \int_0^t \sigma^2 S_s^2 \mathrm{d}s.$$

From (38), we obtain a linear equation for the dynamics of S,

---

[16]That is the sum of $\sum_{t_i \leq t}|A_{t_{i+1}} - A_{t_i}|$ exists and is finite, for instance A is continuously differentiable.

$$dS_t = S_t \mu dt + S_t \sigma dW_t$$

also written $\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$ putting an emphasize of the financial interpretation as returns. The Itô formula writes

$$df(t, S_t) = (f'_t(t, S_t) + f'_x(t, S_t)S_t\mu + \frac{1}{2}f''_{xx}(t, S_t)S_t^2\sigma^2)dt + f'_x(t, S_t)\sigma S_t dW_t$$

$$:= (f'_t(t, S_t) + L^{\text{GBM}}_{\mu,\sigma^2} f(t, S_t))dt + f'_x(t, S_t)\sigma S_t dW_t. \tag{39}$$

Example 3 (Ornstein–Uhlenbeck Process). $(V_t = v_0 - a\int_0^t V_s ds + \sigma W_t)$: we have

$$\langle V \rangle_t = \sigma^2 t.$$

The Itô formula follows

$$df(t, V_t) = (f'_t(t, V_t) - af'_x(t, V_t)V_t + \frac{1}{2}\sigma^2 f''_{xx}(t, V_t))dt + f'_x(t, V_t)\sigma dW_t$$

$$:= (f'_t(t, V_t) + L^{\text{OU}}_{a,\sigma^2} f(t, V_t))dt + f'_x(t, V_t)\sigma dW_t. \tag{40}$$

Example 4 (Euler Scheme Defined in (12)*).* $(X_t^h = X_{ih}^h + b(X_{ih}^h)(t - ih) + \sigma(X_{ih}^h)(W_t - W_{ih})$ for $i \geq 0, t \in (ih, (i+1)h]$). Since X^h is an arithmetic Brownian motion on each interval $(ih, (i+1)h]$, we easily obtain

$$\langle X^h \rangle_t = \int_0^t \sigma^2(\varphi(s), X_{\varphi(s)}^h)ds$$

where $\varphi(t) = ih$ for $t \in (ih, (i+1)h]$. The Itô formula writes

$$df(t, X_t^h) = (f'_t(t, X_t^h) + b(X_{\varphi(t)}^h)f'_x(t, X_t^h) + \frac{1}{2}\sigma^2(X_{\varphi(t)}^h)f''_{xx}(t, X_t^h))dt$$

$$+ f'_x(t, X_t^h)\sigma(X_{\varphi(t)}^h)dW_t. \tag{41}$$

3.4.2 The Multidimensional Case

We briefly expose the situation when $X = (X_1, \ldots, X_d)$ takes values in $\mathbb{R}^d$. The main novelty consists in considering the cross quadratic variation defined by the limit (assuming its existence, along dyadic subdivision) of

$$\langle X_k, X_l \rangle_t^n = \sum_{t_i \leq t}(X_{k,t_{i+1}} - X_{k,t_i})(X_{l,t_{i+1}} - X_{l,t_i}) \xrightarrow[n \to +\infty]{} \langle X_k, X_l \rangle_t. \tag{42}$$

We list basic properties.

Properties 8

1. SYMMETRY: $\langle X_k, X_l \rangle_t = \langle X_l, X_k \rangle_t$.
2. USUAL BRACKET: $\langle X_k, X_k \rangle_t = \langle X_k \rangle_t$.
3. POLARIZATION: $\langle X_k, X_l \rangle_t = \frac{1}{4} \left(\langle X_k + X_l \rangle_t - \langle X_k - X_l \rangle_t \right)$.
4. $\langle \cdot, \cdot \rangle_t$ is bilinear.
5. For any continuous function f, we have

$$\lim_{n \to \infty} \sum_{t_i \leq t} f(t_i)(X_{k,t_{i+1}} - X_{k,t_i})(X_{l,t_{i+1}} - X_{l,t_i}) = \int_0^t f(s) \mathrm{d}\langle X_k, X_l \rangle_s.$$

6. Let $X_{1,t} = f(A_{1,t}, M_{1,t})$ and $X_{2,t} = g(A_{2,t}, M_{2,t})$, where the variation (resp. quadratic variation) of $A = (A_1, A_2)$ (resp. $M = (M_1, M_2)$) is finite, and let f and g be two $\mathscr{C}^1$-functions: we have

$$\langle X_1, X_2 \rangle_t = \int_0^t f'_m(A_{1,s}, M_{1,s}) g'_m(A_{2,s}, M_{2,s}) \mathrm{d}\langle M_1, M_2 \rangle_s.$$

   In particular, $\langle A_1 + M_1, A_2 + M_2 \rangle_t = \langle M_1, M_2 \rangle_t$.
7. Let W_1 and W_2 be two independent Brownian motions: then

$$\langle W_1, W_2 \rangle_t = 0.$$

Proof. The statements (1)–(6) are easy to check from the definition or using previous arguments. The statement (7) is important and we give details: use the polarization identity

$$\langle W_1, W_2 \rangle_t = \frac{1}{4} \left(\langle W_1 + W_2 \rangle_t - \langle W_1 - W_2 \rangle_t \right).$$

We observe that both $\frac{1}{\sqrt{2}}(W_1 + W_2)$ and $\frac{1}{\sqrt{2}}(W_1 - W_2)$ are Brownian motions, since each one is a continuous Gaussian process with the right covariance function. Thus, $\langle \frac{1}{\sqrt{2}}(W_1 + W_2) \rangle_t = \langle \frac{1}{\sqrt{2}}(W_1 - W_2) \rangle_t = t$ and the result follows. $\qquad\square$

The Itô formula naturally extends to this setting.

Theorem 9 (Multidimensional Itô Formula). *Let $f \in \mathscr{C}^{1,2}(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R})$ and X be a continuous d-dimensional process with finite quadratic variation. Then, with probability 1, for any $t \geq 0$ we have*

$$f(t, X_t) = f(0, X_0) + \sum_{k=1}^{d} \int_0^t f'_{x_k}(s, X_s) \, \mathrm{d}X_{k,s}$$

$$+ \int_0^t f'_t(s, X_s) \, \mathrm{d}s + \frac{1}{2} \sum_{k,l=1}^{d} \int_0^t f''_{x_k, x_l}(s, X_s) \, \mathrm{d}\langle X_k, X_l \rangle_s$$

where the sum of stochastic integrals are defined as before.

In particular, the integration by parts formula writes

$$X_{1,t}X_{2,t} = X_{1,0}X_{2,0} + \int_0^t X_{1,s}\mathrm{d}X_{2,s} + \int_0^t X_{2,s}\mathrm{d}X_{1,s} + \langle X_1, X_2 \rangle_t.$$

For two independent Brownian motions, we recover the usual deterministic formula (because $\langle W_1, W_2 \rangle_t = 0$), but in general, formulas are different because of the quadratic variation.

3.5 More Properties on Stochastic Integrals

So far, we have defined some specific stochastic integrals, those appearing in deriving a Itô formula and which have the form

$$\int_0^t f'_x(s, X_s)\mathrm{d}X_s = \lim_{n\to+\infty} \sum_{t_i \leq t} f'_x(t_i, X_{t_i})(X_{t_{i+1}} - X_{t_i}), \qquad (43)$$

the limit being taken along dyadic subdivision. Also, we have proved that if f has bounded derivatives and $X = W$ is a Brownian motion, the above stochastic integral must have zero-expectation [see equality (29)]. Moreover, we also have established that in the case of deterministic integrand (Wiener integral), the second moment of the stochastic integral is explicit and given by

$$\mathbb{E}(\int_0^t h_s\mathrm{d}W_s)^2 = \int_0^t h_s^2\mathrm{d}s.$$

The aim of this paragraph is to provide extensions of the above properties on the two first moments to more general integrands, under some suitable boundedness or integrability conditions.

3.5.1 Heuristic Arguments

In view of the previous construction, there is a natural candidate to be the stochastic integral $\int_0^t h_s\mathrm{d}W_s$. When h is piecewise constant process (called *simple process*), that is $h_s = h_{t_i}$ if $s \in [t_i, t_{i+1}]$ for a given deterministic time grid $(t_i)_i$, we set

$$\int_0^t h_s\mathrm{d}W_s = \sum_{t_i \leq t} h_{t_i}(W_{t \wedge t_{i+1}} - W_{t_i}), \qquad (44)$$

Without extra assumptions on the stochasticity of h, it is not clear why its expectation equals 0. This property should come from the centered Brownian increments $W_{t \wedge t_{i+1}} - W_{t_i}$ and their independence to h_{t_i} so that

$$\mathbb{E}(\int_0^t h_s \mathrm{d}W_s) = \sum_{t_i \leq t} \mathbb{E}(h_{t_i})\mathbb{E}(W_{t \wedge t_{i+1}} - W_{t_i}) = 0.$$

To validate this computation, we shall assume that h_t depends only the Brownian Motion W before t and it is integrable. To go to the second moment, assume additionally that h is square integrable: then

$$\mathbb{E}|\int_0^t h_s \mathrm{d}W_s|^2$$

$$= 2 \sum_{t_i < t_j \leq t} \mathbb{E}(h_{t_i} h_{t_j} (W_{t \wedge t_{i+1}} - W_{t_i}))\mathbb{E}(W_{t \wedge t_{j+1}} - W_{t_j}) + \sum_{t_i \leq t} \mathbb{E}(h_{t_i}^2)\mathbb{E}|W_{t \wedge t_{i+1}} - W_{t_i}|^2$$

$$= \sum_{t_i \leq t} \mathbb{E}(h_{t_i}^2)(t \wedge t_{i+1} - t_i) = \mathbb{E}(\int_0^t h_s^2 \mathrm{d}s). \tag{45}$$

This equality should be read as an isometry property (usually referred to as *Itô isometry*), on which we can rely an extension of the stochastic integral of simple process to more general process. At this point, we should need to enter into measurability considerations to describe what "h_t *depends only the Brownian Motion W before t*" means at the most general level. It goes far beyond this introductory lecture: for the exposure of the general theory, see for instance [48] or [63].

For most of the examples considered in this lecture, we can restrict to *very good integrands*, in the sense that a integrand h is *very good* if

1. $(h_t)_t$ is continuous or piecewise continuous (as for simple processes).
2. For a given t, h_t is a continuous functional of $(W_s : s \leq t)$.
3. It is square integrable in time and ω: $\mathbb{E}(\int_0^t h_s^2 \mathrm{d}s) < +\infty$ for any t.

This setting ensures that we can define stochastic integrals for very good integrands as the L_2-limit of stochastic integrals for simple integrands: indeed, a Cauchy sequence $(h_n)_n$ in $L_2(\mathrm{d}t \otimes \mathrm{d}\mathbb{P})$ gives a Cauchy sequence $(\int_0^t h_{n,s}\mathrm{d}W_s)_n$ in $L_2(\mathbb{P})$ due to the isometry (45).

3.5.2 General Results

We collect here all the stochastic integration results needed in this lecture.

Theorem 10. *Let h be a very good integrand. Then the stochastic integral $\int_0^t h_s \mathrm{d}W_s$ is such that*

1. *It is the L_2 limit of $\sum_{t_i \leq t} h_{t_i}(W_{t \wedge t_{i+1}} - W_{t_i})$ along time subdivision which time step goes to 0.*
2. *It is centered: $\mathbb{E}(\int_0^t h_s \mathrm{d}W_s) = 0$.*

3. *It is square integrable:* $\mathbb{E}|\int_0^t h_s dW_s|^2 = \mathbb{E}(\int_0^t h_s^2 ds)$.
4. *For two very good integrands h_1 and h_2, we have*

$$\mathbb{E}\Big[(\int_0^t h_{1,s} dW_s)(\int_0^t h_{2,s} dW_s)\Big] = \mathbb{E}(\int_0^t h_{1,s} h_{2,s} ds).$$

Beyond the t-by-t construction, actually the full theory gives a construction for any t simultaneously, proving additionally time continuity property, general centering property (*martingale* property), tight L_p-estimates on the value at time t and the extrema until time t (Burkholder–Davis–Gundy inequalities) and so on... For multidimensional W and h, the construction should be understood componentwise. Another fruitful extension is to allow t to be a bounded stopping time, similarly to the discussion we have made in the proof of Theorem 3.

Another interesting part in the theory is devoted to the existence and uniqueness of solution to Stochastic Differential Equations (also known as diffusion processes). The easiest setting is to assume globally Lipschitz coefficients[17]: it is similar to the ODE framework, and the proof is also based on the Picard fixed-point argument. We state the results without proof.

Theorem 11. *Let W be a d-dimensional standard Brownian motion.*

Assume that the functions $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^d \otimes \mathbb{R}^d$ are globally Lipschitz. Then, for any initial condition $x \in \mathbb{R}^d$, there exists a unique[18] continuous solution $(X_t^x)_{t \geq 0}$ valued in $\mathbb{R}^d$ which satisfies

$$X_t^x = x + \int_0^t b(X_s^x) ds + \int_0^t \sigma(X_s^x) dW_s, \tag{46}$$

with $\sup_{0 \leq t \leq T} \mathbb{E}|X_t^x|^2 < +\infty$ for any given $T \in \mathbb{R}^+$.
The continuous process X^x has a finite quadratic variation given by

$$\langle X_k^x, X_l^x \rangle_t = \int_0^t [\sigma \sigma^\top]_{k,l}(X_s^x) ds, \quad 1 \leq k, l \leq d. \tag{47}$$

Observe that this general result includes all the model considered before, such as Arithmetic and Geometric Brownian Motion, Ornstein–Uhlenbeck processes, here stated in a possibly multidimensional framework.

Complementary References. See [48, 63].

---

[17]Leading to the notion of *strong* solution; the case of non-smooth coefficients is much more delicate and related to *weak* solutions, see [67].

[18]Up to a set of zero probability measure.

4   Monte Carlo Resolutions of Linear PDEs Related to SDEs

Probabilistic methods to solve PDEs have become very popular during the two last decades. They are usually not competitive compared to deterministic methods in low dimension, but for higher dimension they provide very good alternative schemes. In the sequel, we give a brief introduction to the topics, relying on the material presented in the previous sections. We start with linear parabolic PDEs, with Cauchy–Dirichlet boundary conditions. Next section is devoted to semi-linear PDEs.

4.1   Second Order Linear Parabolic PDEs with Cauchy Initial Condition

4.1.1   Feynman–Kac Formulas

We start with a verification Theorem generalizing Theorems 2, 4, 5 to the case of general SDEs. We incorporate a source term g.

Theorem 12. *Under the assumptions of Theorem 11, let X^x be the solution* (46) *starting from $x \in \mathbb{R}^d$ and set*

$$L^X_{b,\sigma\sigma^\top} = \frac{1}{2} \sum_{i,j=1}^{d} [\sigma\sigma^\top]_{i,j}(x)\partial^2_{x_i x_j} + \sum_{i=1}^{d} b_i(x)\partial_{x_i}.$$

Assume there is a solution $u \in \mathscr{C}^{1,2}_b(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R})$ to the PDE

$$\begin{cases} u'_t(t,x) = L^X_{b,\sigma\sigma^\top}u(t,x) + g(x), \\ u(0,x) = f(x) \end{cases} \tag{48}$$

for two given functions $f, g : \mathbb{R}^d \to \mathbb{R}$. Then u is given by

$$u(t,x) = \mathbb{E}[f(X^x_t) + \int_0^t g(X^x_s)\mathrm{d}s]. \tag{49}$$

Proof. Let t be fixed. We apply the general Itô formula (Theorem 9) to the process X^x and to the function $v : (s,y) \mapsto u(t-s,y)$: it gives

$$\mathrm{d}v(s, X^x_s) = \left[v'_s(s, X^x_s) + L^X_{b,\sigma\sigma^\top}v(s, X^x_s)\right]\mathrm{d}s + Dv(s, X^x_s)\sigma(X^x_s)\mathrm{d}W_s \tag{50}$$

$$= -g(X^x_s)\mathrm{d}s + Dv(s, X^x_s)\sigma(X^x_s)\mathrm{d}W_s \tag{51}$$

where $Dv := (\partial_{x_1}v, \ldots, \partial_{x_d}v)$. Observe that the integrand $h_s = Dv(s, X_s^x)\sigma(X_s^x)$ is very good, since v has bounded derivatives, σ has a linear growth, and X_s has bounded second moments, locally uniformly in s: thus, the stochastic integral $\int_0^t Dv(s, X_s^x)\sigma(X_s^x)dW_s$ has zero expectation. Hence, applying the above decomposition between $s = 0$ and $s = t$ and taking the expectation, it gives

$$\mathbb{E}(f(X_t^x)) = \mathbb{E}(v(t, X_t^x)) = v(0, x) - \mathbb{E}(\int_0^t g(X_s^x)ds) = u(t, x) - \mathbb{E}(\int_0^t g(X_s^x)ds).$$

We are done.   □

Smoothness assumptions on u are satisfied in f, g are smooth enough. If not, and if a uniform ellipticity condition is met on $\sigma\sigma^\top$, the fundamental solution of the PDE is smoothing the data and the result can be extended. However, the derivatives blow up as time goes to 0, and more technicalities are necessary to justify the same stochastic calculus computations. The fundamental solution $p(t, x, y)$ has a simple probabilistic interpretation: it is the density of X_t^x at y. Indeed, identify $\mathbb{E}[f(X_t^x) + \int_0^t g(X_s^x)ds]$ with

$$u(t, x) = \int_{\mathbb{R}^d} p(t, x, y)f(y)dy + \int_0^t \int_{\mathbb{R}^d} p(s, x, y)g(y)dy\,ds.$$

4.1.2 Monte Carlo Schemes

Since $u(t, x)$ is represented as an expectation, it allows the use of a Monte Carlo method to numerically compute the solution. The difficulty is that in general, X can not be simulated perfectly accurately, only an approximation on a finite time-grid can be simply and efficiently produced. Namely we use the Euler scheme with time step $h = t/N$:

$$\begin{cases} X_0^{x,h} = x, \\ X_s^{x,h} = X_{ih}^{x,h} + b(X_{ih}^{x,h})(s - ih) + \sigma(X_{ih}^{x,h})(W_s - W_{ih}), & i \geq 0, s \in (ih, (i+1)h]. \end{cases}$$

(52)

Observe that to get $X_t^{x,h}$, we do not need to sample the continuous path of $X^{x,h}$ (as difficult as having a continuous path of a Brownian motion): in fact, we only need to compute $X_{ih}^{x,h}$ iteratively for $i = 0$ to $i = N$. Each time iteration requires to sample d new independent Gaussian increments $W_{k,(i+1)h} - W_{k,ih}$, centered with variance h: it is straightforward. The computational cost is essentially equal to $C(d)N$ where the constant depends on the dimension (coming from d-dimensional vector and matrix computations).

As an approximation of the expectation of $\mathscr{E}(f, g, X^x) = f(X_t^x) + \int_0^t g(X_s^x)ds$, we take the expectation

$$\mathscr{E}(f, g, X^{x,h}) = f(X_{Nh}^{x,h}) + \sum_{i=0}^{N-1} g(X_{ih}^{x,h})h, \tag{53}$$

a random variable of which we sample M independent copies, that are denoted by $\{\mathscr{E}(f, g, X^{x,h,m}) : 1 \leq m \leq M\}$. Then, the Monte Carlo approximation, based on this sample of M Euler schemes with time step h, is

$$\frac{1}{M} \sum_{m=1}^{M} \mathscr{E}(f, g, X^{x,h,m}) = u(t, x) + \underbrace{\frac{1}{M} \sum_{m=1}^{M} \mathscr{E}(f, g, X^{x,h,m}) - \mathbb{E}(\mathscr{E}(f, g, X^{x,h}))}_{\text{statistical error Err.}_{\text{stat.}}(h, M)}$$

$$+ \underbrace{\mathbb{E}(\mathscr{E}(f, g, X^{x,h})) - u(t, x)}_{\text{discretization error Err.}_{\text{disc.}}(h)}. \tag{54}$$

The first error contribution is due to the sample of finite size: the larger M, the better the accuracy. As mentioned in Sect. 2.1.3, once renormalized by $\sqrt{M}$, this error is still random and its distribution is closed to the Gaussian distribution with zero mean and variance $\mathbb{V}\text{ar}(\mathscr{E}(f, g, X^{x,h}))$: the latter still depends on h but very little, since it is expected to be close to $\mathbb{V}\text{ar}(\mathscr{E}(f, g, X^x))$.

The second error contribution is related to the time discretization effect: the smaller the time h, the better the accuracy. In the sequel (Sect. 4.1.3), we theoretically estimate this error in terms of h, and proves that it is of order h (even equivalent to) under some reasonable and fairly general assumptions.

What Is the Optimal Tuning of $h \to 0$ and $M \to +\infty$? An easy complexity analysis shows that the computational effort is $\mathscr{C}_e = C(d)Mh^{-1}$. Observe that the rate does not depend on the dimension d, as a difference with a PDE method, but on the other hand, the solution is computed only at single point (t, x). The squared quadratic error is equal to

$$[\text{Err}_2(h, M)]^2 := \mathbb{E}\Big[\frac{1}{M} \sum_{m=1}^{M} \mathscr{E}(f, g, X^{x,h,m}) - u(t, x)\Big]^2$$

$$= \frac{\mathbb{V}\text{ar}(\mathscr{E}(f, g, X^{x,h}))}{M} + \Big[\mathbb{E}(\mathscr{E}(f, g, X^{x,h})) - u(t, x)\Big]^2.$$

Only the first factor $\mathbb{V}\text{ar}(\mathscr{E}(f, g, X^{x,h}))$ can be estimated with the same sample, for M large, and it depends little of h. Say that the second term is equivalent to $(Ch)^2$ as $h \to 0$, with $C \neq 0$. Then, three asymptotic situations occur:

1. If $M \gg h^{-2}$, the statistical error becomes negligible and $\frac{1}{M} \sum_{m=1}^{M} \mathscr{E}(f, g, X^{x,h,m}) - u(t, x) \sim Ch$. The computational effort is $\mathscr{C}_e \gg h^{-3}$ and thus $\text{Err}_2(h, M) \gg \mathscr{C}_e^{-1/3}$. Deriving a confidence interval as in Sect. 2.1.3 is meaningless, we face with the discretization error only.

2. If $M \ll h^{-2}$, the discretization error becomes negligible and the distribution of $\sqrt{M} \left(\frac{1}{M} \sum_{m=1}^{M} \mathscr{E}(f, g, X^{x,h,m}) - u(t, x) \right)$ converges to that a Gaussian r.v. *centered* with variance $\mathbb{V}\mathrm{ar}(\mathscr{E}(f, g, X^x))$ (that can be asymptotically computed using the M- sample). Thus, we can derive confidence intervals: setting $\sigma_{h,M}^2$ the empirical variance of $\mathscr{E}(f, g, X^{x,h})$, with probability 95 % we have

$$u(t, x) \in \left[\frac{1}{M} \sum_{m=1}^{M} \mathscr{E}(f, g, X^{x,h,m}) - 1.96 \frac{\sigma_{h,M}}{\sqrt{M}}, \frac{1}{M} \sum_{m=1}^{M} \mathscr{E}(f, g, X^{x,h,m}) + 1.96 \frac{\sigma_{h,M}}{\sqrt{M}} \right].$$

Regarding the computational effort, we have $\mathscr{C}_e \gg M^{3/2}$ and thus $\mathrm{Err}_2(h, M) \gg \mathscr{C}_e^{-1/3}$.

3. If $M \sim ch^{-2}$, both statistical and discretization errors have the same magnitude and one can still derive a asymptotic confidence interval, but it is no more centered (as in $M \ll h^{-2}$) and unfortunately, the bias is not easily estimated on the fly. The problem is that the bias is of same magnitude as the size of the confidence interval, thus it reduces the interest of having such a priori statistical error estimate. Here, $\mathrm{Err}_2(h, M) = O(\mathscr{C}_e^{-1/3})$.

Summing up by considering the ability of having or not on-line error estimates and by optimizing the final accuracy w.r.t. the computational effort, the second case $M = h^{-2+\varepsilon}$ (for a small $\varepsilon > 0$) may be the most attractive since it achieves (almost) the best accuracy w.r.t. the computational effort and gives a centered confidence interval (and therefore tractable and meaningful error bounds).

4.1.3   Convergence of the Euler Scheme

An important issue is to analyze the impact of time discretization of SDE. This dates back to the end of eighties, see [68] among others. The result below gives a mathematical justification of the use of the Euler scheme as an approximation for the distribution of the SDE.

Theorem 13. *Assume that b and σ are $\mathscr{C}_b^2$, let X^x be the solution* (46) *starting from $x \in \mathbb{R}^d$ and let $X^{h,x}$ be its Euler scheme defined in* (52). *Assume that $u(t, x) = \mathbb{E}[f(X_t^x) + \int_0^t g(X_s^x)\mathrm{d}s]$ is a $\mathscr{C}_b^{2,4}([0, T] \times \mathbb{R}^d, \mathbb{R})$-function solution of the PDE of Theorem* 12. *Then,*

$$\mathbb{E}\left[f(X_{Nh}^{x,h}) + \sum_{i=0}^{N-1} g(X_{ih}^{x,h})h \right] - \mathbb{E}\left[f(X_t^x) + \int_0^t g(X_s^x)\mathrm{d}s \right] = O(h).$$

Proof. Denote by $\mathrm{Err.}_{\mathrm{disc.}}(h)$ the above discretization error. As in Theorem 12, we use the function $v : (s, y) \mapsto u(t - s, y)$ (for a fixed t) and we apply the Itô formula to $X^{h,x}$ (Theorem 9): it gives (setting $Dv := (\partial_{x_1} v, \dots, \partial_{x_d} v)$)

$$dv(s, X_s^{h,x}) = \Big[v_s'(s, X_s^{h,x}) + \frac{1}{2} \sum_{i,j=1}^d [\sigma\sigma^\top]_{i,j}(X_{\varphi(s)}^{h,x}) \partial_{x_i x_j}^2 v(s, X_s^{h,x})$$

$$+ \sum_{i=1}^d b_i(X_{\varphi(s)}^{h,x}) \partial_{x_i} v(s, X_s^{h,x}) \Big] ds + Dv(s, X_s^{h,x})\sigma(X_{\varphi(s)}^{h,x}) dW_s.$$

$$= \Big[\frac{1}{2} \sum_{i,j=1}^d \big([\sigma\sigma^\top]_{i,j}(X_{\varphi(s)}^{h,x}) - [\sigma\sigma^\top]_{i,j}(X_s^{h,x}) \big) \partial_{x_i x_j}^2 v(s, X_s^{h,x})$$

$$+ \sum_{i=1}^d \big(b_i(X_{\varphi(s)}^{h,x}) - b_i(X_s^{h,x}) \big) \partial_{x_i} v(s, X_s^{h,x}) - g(X_s^{h,x}) \Big] ds$$

$$+ Dv(s, X_s^{h,x})\sigma(X_{\varphi(s)}^{h,x}) dW_s$$

where at the second equality, we have used the PDE solved by v at (s, X_s^x). Then, by taking the expectation (it removes the stochastic integral term because the integrand is very good), we obtain

$$\text{Err.}_{\text{disc.}}(h) = \mathbb{E}\Big[v(Nh, X_{Nh}^{x,h}) + \sum_{i=1}^N h g(X_{ih}^{x,h}) \Big] - v(0, x)$$

$$= \mathbb{E}\Big(\int_0^t \Big[\frac{1}{2} \sum_{i,j=1}^d \big([\sigma\sigma^\top]_{i,j}(X_{\varphi(s)}^{h,x}) - [\sigma\sigma^\top]_{i,j}(X_s^{h,x}) \big) \partial_{x_i x_j}^2 v(s, X_s^{h,x}) \Big] ds \Big)$$

$$+ \mathbb{E}\Big(\int_0^t \Big[\sum_{i=1}^d \big(b_i(X_{\varphi(s)}^{h,x}) - b_i(X_s^{h,x}) \big) \partial_{x_i} v(s, X_s^{h,x}) \Big] ds \Big)$$

$$+ \mathbb{E}\Big(\int_0^t \big[g(X_{\varphi(s)}^{h,x}) - g(X_s^{h,x}) \big] ds \Big).$$

The global error is represented as a summation of local errors. For instance, let us estimate the first term related to $\sigma\sigma^\top$: apply once again the Itô formula on the interval $[kh, s] \subset [kh, (k+1)h]$ and to the function $(s, y) \mapsto \big([\sigma\sigma^\top]_{i,j}(X_{\varphi(s)}^{h,x}) - [\sigma\sigma^\top]_{i,j}(y) \big) \partial_{x_i x_j}^2 v(s, y)$. It gives raise to a time integral between $kh = \varphi(s)$ and s and a stochastic integral that vanishes in expectation. Proceed similarly for the other contributions with b and g. Finally we obtain a representation formula of the form

$$\text{Err.}_{\text{disc.}}(h) = \sum_{\alpha: 0 \le |\alpha| \le 4} \mathbb{E}\Big(\int_0^t \int_{\varphi(s)}^s \partial_x^{|\alpha|} v(r, X_r^{h,x}) l_\alpha\big(X_{\varphi(r)}^{h,x}, X_r^{h,x} \big) dr ds \Big)$$

where the summation is made on differentiation multi-indices of length smaller than 4, where l_α are functions depending on b, σ, g and their derivatives up to order 2,

and where l_α has at most a linear growth w.r.t its two variables. Taking advantage of the boundedness of the derivatives of v, we easily complete the proof.

Observe that, by strengthening the assumptions and by going a bit further in the analysis, we could establish an expansion w.r.t. h.   □

The previous assumption on u implies that $f \in \mathscr{C}_b^4$ and $g \in \mathscr{C}_b^2$, which is too strong in practice. The extension to non smooth f is much more difficult and we have to take advantage of the smoothness coming from the non-degenerate distribution of X or X^h. We may follow the same types of computations, mixing PDE techniques and stochastic arguments, see [6]. But this is a pure stochastic analysis approach (Malliavin calculus) which provides the extension under the minimal non-degeneracy assumption (i.e. only stated at the initial point x), see [38]. We state the result without proofs.

Theorem 14. *Assume that b and σ are $\mathscr{C}_b^\infty$, let X^x be the solution* (46) *starting from $x \in \mathbb{R}^d$ and let $X^{h,x}$ be its Euler scheme defined in* (52). *Assume additionally that $\sigma\sigma^\top(x)$ is invertible. Then, for any bounded measurable function f, we have*

$$\mathbb{E}\Big[f(X_t^{x,h})\Big] - \mathbb{E}\Big[f(X_t^x)\Big] = O(h).$$

In the same reference [38], the result is also proved for hypoelliptic system, where the hypoellipticity holds only at the starting point x. On the other hand, without such a degeneracy condition and for non smooth f (like Heaviside function), the convergence may fail.

The case of coefficients b and σ with low regularity or exploding behavior is still an active fields of research.

4.1.4   Sensitivities

If in addition we are interested by computing derivatives of $u(t, x)$ w.r.t. x or other model parameters, this is still possible using Monte Carlo simulations. For the sake of simplicity, in our discussion we focus on the gradient of u w.r.t. x. Essentially, two approaches are known.

Resimulation Method.   The derivative is approximated using the finite difference method

$$\partial_{x_i} u(t, x) \approx \frac{u(t, x + \varepsilon e_i) - u(t, x - \varepsilon e_i)}{2\varepsilon}$$

where $e_i = (0, \ldots, 0, \underset{i^{th}}{1}, 0, \ldots)$, and ε is small. Then, each value function is approximated by its Monte Carlo approximation given in (54). However, we have to be careful in generating the Euler scheme starting from $x + \varepsilon e_i$ and $x - \varepsilon e_i$: its sampling should use the same Brownian motion increments, that is

$$\partial_{x_i} u(t, x) \approx \frac{1}{M} \sum_{m=1}^{M} \frac{\mathcal{E}(f, g, X^{x+\varepsilon e_i, h, m}) - \mathcal{E}(f, g, X^{x-\varepsilon e_i, h, m})}{2\varepsilon}. \tag{55}$$

Indeed, for an infinite sample ($M = +\infty$), it does not have any impact on the statistical error whether or not we use the *same driving noise*, but for finite M, this trick likely maintains a smaller statistical error. Furthermore, the optimal choice of h, M and ε is an important issue, but here results are different according to the regularity of f and g, we do not go into details.

Likelihood Method.   To avoid the latter problems of selecting the appropriate value of the finite difference parameter ε, we may prefer another Monte Carlo estimator of $\partial_{x_i} u(t, x)$, which consists in appropriately weighting the output. When g equals 0, it takes the following form

$$\partial_{x_i} u(t, x) \approx \frac{1}{M} \sum_{m=1}^{M} f(X_t^{x, h, m}) H_t^{x, h, m} \tag{56}$$

where $H_t^{x, h, m}$ is simultaneously generated with the Euler scheme and does not depend on f. The advantage of this approach is to avoid the possibly delicate choice of the perturbation parameter ε and it is valid for any function f: thus, it may reduce much the computational time, if many sensitivities are required for the same model. On the other hand, the confidence interval may be larger than that of the resimulation method.

We now provide the formula for the weight H (known as Bismut–Elworthy–Li formula). It uses the tangent process, which is the (well-defined, see [52]) derivative of $x \mapsto X_t^x$ w.r.t. x and which solves

$$DX_t^x := Y_t^x = \mathrm{Id} + \int_0^t Db(X_s^x)\, Y_s^x\, \mathrm{d}s + \sum_{j=1}^{d} \int_0^t D\sigma_j(X_s^x)\, Y_s^x\, \mathrm{d}W_{j,s} \tag{57}$$

where σ_j is the j-th column of the matrix σ.

Theorem 15. *Assume that b and σ are $\mathscr{C}_b^2$-functions, that $u \in \mathscr{C}^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$ solves the PDE* (48)*, and that σ is invertible with a uniformly bounded inverse σ^{-1}. We have*

$$Du(t, x) = \mathbb{E}\left(\frac{f(X_t^x)}{t} \left[\int_0^t [\sigma^{-1}(X_s^x) Y_s^x]^\top \mathrm{d}W_s \right]^\top \right).$$

Proof. First, we recall the decomposition (51) obtained from Itô formula, using $v(s, y) = u(t - s, y)$:

$$\begin{cases} v(r, X_r^x) = v(0, x) + \int_0^r Dv(s, X_s^x)\sigma(X_s^x)\mathrm{d}W_s, & \forall 0 \le r \le t, \\ f(X_t^x) = v(t, X_t^x). \end{cases} \tag{58}$$

Second, taking expectation, it gives $v(0, x) = u(t, x) = \mathbb{E}(v(r, X_r^x))$ for any $r \in [0, T]$. By differentiating w.r.t. x, we obtain a nice relation letting the expectation constant in time (actually deeply related to martingale property):

$$Dv(0, x) = \mathbb{E}(Dv(r, X_r^x)Y_r^x), \quad \forall 0 \leq r \leq t.$$

Thus, we deduce

$$\begin{aligned} Du(t, x) = Dv(0, x) &= \mathbb{E}\left(\frac{1}{t}\int_0^t Dv(s, X_s^x)Y_s^x \, ds\right) \\ &= \mathbb{E}\left(\frac{1}{t}\left[\int_0^t Dv(s, X_s^x)\sigma(X_s^x)dW_s\right]\left[\int_0^t [\sigma^{-1}(X_s^x)Y_s^x]^\top dW_s\right]^\top\right) \\ &= \mathbb{E}\left(\frac{v(t, X_t^x) - v(0, x)}{t}\left[\int_0^t [\sigma^{-1}(X_s^x)Y_s^x]^\top dW_s\right]^\top\right) \\ &= \mathbb{E}\left(\frac{f(X_t^x)}{t}\left[\int_0^t [\sigma^{-1}(X_s^x)Y_s^x]^\top dW_s\right]^\top\right) \end{aligned}$$

using Theorem 10 at the second and fourth equality, (58) at the third one. □

In view of the above assumptions of u, implicitly the function f is smooth. However, under the current ellipticity condition, u is still smooth even if f is not; since the formula does depend on f and not on its derivatives, it is standard to extend the formula to any bounded function f (without any regularity assumption).

The Monte Carlo evaluation of $Du(t, x)$ easily follows by independently sampling $\frac{f(X_t^x)}{t}\left[\int_0^t [\sigma^{-1}(X_s^x)Y_s^x]^\top dW_s\right]^\top$ and taking the empirical mean. The exact simulation is not possible and once again, we may use an Euler-type scheme, with time step h:

- The dimension-augmented Stochastic Differential Equation (X^x, Y^x) is approximated using the Euler scheme.
- We use a simple-approximation of the stochastic integral

$$\int_0^t [\sigma^{-1}(X_s^x)Y_s^x]^\top dW_s = \sum_{i=0}^{N-1} [\sigma^{-1}(X_{ih}^{x,h})Y_{ih}^{x,h}]^\top (W_{(i+1)h} - W_{ih}).$$

The analysis of discretization error is more intricate than for $\mathbb{E}(f(X_t^{x,h}) - f(X_t^x))$: nevertheless, the error is still of magnitude h (the convergence order is 1 w.r.t. h, as proved in [38]).

Theorem 16. *Under the setting of Theorem 14, for any bounded measurable function f, we have*

$$\mathbb{E}\left(\frac{f(X_t^{x,h})}{t}\sum_{i=0}^{N-1}\left[[\sigma^{-1}(X_{ih}^{x,h})Y_{ih}^{x,h}]^{\top}(W_{(i+1)h}-W_{ih})\right]^{\top}\right)-Du(t,x)=O(h).$$

4.1.5 Other Theoretical Estimates in Small Time

The representation formula of Theorem 15 is the starting point for getting accurate probabilistic estimates on the derivatives of the underlying PDE as time is small, in terms of the fractional smoothness of $f(X_t^x)$ which is related to the decay of

$$\|f(X_t^x)-\mathbb{E}(f(X_{t-s}^y))|_{y=X_s^x}\|_{L_2}\quad\text{as }s\to t.$$

The derivatives are measured in weighted L_2-norms and surprisingly, the above results are equivalence results [36]; we are not aware of such results using PDE arguments.

Theorem 17. *Under the setting[19] of Theorem 14, let t be fixed, for $0<\theta\le1$ and a bounded f, the following assertions are equivalent:*

i) *For some $c\ge0$, $\mathbb{E}|f(X_t^x)-\mathbb{E}(f(X_{t-s}^y))|_{y=X_s^x}|^2\le c(t-s)^{\theta}$ for $0\le s\le t$.*
ii) *For some $c\ge0$, $\mathbb{E}|Du(t-s,X_s^x)|^2\le\frac{c}{(t-s)^{1-\theta}}$ for $0\le s<t$.*
iii) *For some $c\ge0$, $\int_0^s\mathbb{E}|D^2u(t-r,X_r^x)|^2dr\le\frac{c}{(t-s)^{1-\theta}}$ for $0\le s<t$.*

If $0<\theta<1$, it is also equivalent to:

iv) *For some $c\ge0$, $\mathbb{E}|D^2u(t-s,X_s^x)|^2\le\frac{c}{(t-s)^{2-\theta}}$ for $0\le s<t$.*

Theorem 18. *Under the setting of Theorem 14, let t be fixed, for $0<\theta<1$ and a bounded f, the following assertions are equivalent:*

i) $\int_0^t(t-s)^{-\theta-1}\mathbb{E}|f(X_t^x)-\mathbb{E}(f(X_{t-s}^y))|_{y=X_s^x}|^2ds<+\infty.$
ii) $\int_0^t(t-s)^{-\theta}\mathbb{E}|Du(t-s,X_s^x)|^2ds<+\infty.$
iii) $\int_0^t(t-s)^{1-\theta}\mathbb{E}|D^2u(t-s,X_s^x)|^2ds<+\infty.$

4.2 The Case of Dirichlet Boundary Conditions and Stopped Processes

4.2.1 Feynman–Kac Formula

In view of Corollary 1, the natural extension of Theorem 12 in the case of Dirichlet boundary condition is the following. We state the result without source term to simplify. The proof is similar and we skip it.

---

[19]To simplify the exposure.

Theorem 19. *Let D be a bounded domain of* $\mathbb{R}^d$. *Under the setting of Theorem 12, assume there is a solution* $u \in \mathscr{C}_b^{1,2}([0,T] \times \overline{D}, \mathbb{R})$ *to the PDE*

$$\begin{cases} u_t'(t,x) = L_{b,\sigma\sigma^\top}^X u(t,x), & \text{for } (t,x) \in]0, +\infty[\times D, \\ u(0,x) = f(0,x), & \text{for } x \in D, \\ u(t,x) = f(t,x), & \text{for } (t,x) \in \mathbb{R}^+ \times \partial D, \end{cases} \tag{59}$$

for a given function $f : \mathbb{R}^+ \times \overline{D} \to \mathbb{R}$. *Then u is given by*

$$u(t,x) = \mathbb{E}[f(t - \tau^x \wedge t, X_{\tau^x \wedge t}^x)] \tag{60}$$

for $x \in D$, *where* $\tau^x = \inf\{s > 0 : X_s^x \notin D\}$ *is the first exit time from D by X.*

4.2.2   Monte Carlo Simulations

Performing a Monte Carlo algorithm in this context is less easy since we have to additionally simulate the exit time of X. A simple approach consists in discretizing X using the Euler scheme with time step h, and then taking for the exit time

$$\tau^{x,h} = \inf\{ih > 0 : X_{ih}^{x,h} \notin D\}.$$

It does not require any further computations than those needed to generate $(X_{ih}^{x,h}, 0 \le i \le N)$. But, the discretization error worsens much since it becomes of magnitude $\sqrt{h}$. Actually, even if the values of $(X_{ih}^{x,h}, 0 \le i \le N)$ are generated without error (like in Brownian motion case or other simple processes), the convergence order is still $\frac{1}{2}$ w.r.t. h [27]. The deterioration of the discretization error really comes from the high irregularity of Brownian motion paths (and SDE paths): even if two successive points $X_{ih}^{x,h}$ and $X_{(i+1)h}^{x,h}$ are close to the boundary but inside the domain, a discrete monitoring scheme does not detect the exit while a continuous Brownian motion-like path would likely exit from the domain between ih and $(i + 1)h$. Moreover, it gives a systematic (in mean) underestimation of the true exit time. To overcome this lack of accuracy, there are several improved schemes.

- The Brownian bridge technique consists in simulating the exit time of local arithmetic Brownian motion [corresponding to the local dynamics of Euler scheme, see (12)]. For simple domain like half-space, the procedure is explicit and tractable, this is related the explicit knowledge of the distribution of the Brownian maximum, see Proposition 5. For smooth domain, we can approximate locally the domain by half-spaces. This improvement allows to recover the order 1 for the convergence, see [27, 28]. For non smooth domains (including corners for instance) and general SDEs, providing an accurate scheme and performing

its error analysis is still an open issue; for heuristics and numerical experiments, see [29] for instance.

- The boundary shifting method consists in shrinking the domain to compensate the systematic bias in the simulation of the discrete exit time. Very remarkably, there is a universal elementary rule to make the domain smaller:

  locally at a point y close to the boundary, move the boundary inwards by a quantity proportional to $c_0\sqrt{h}$ times the norm of the diffusion coefficient in the normal direction.

  The constant c_0 is equal to the mean of the asymptotic overshoot of the Gaussian random walk as the ladder height goes to infinity: it can be expressed using the zeta function

  $$c_0 = -\frac{\zeta(\frac{1}{2})}{\sqrt{2\pi}} = 0.5826\ldots.$$

  This procedure strictly improves the order $\frac{1}{2}$ of the discrete procedure, but it is still an open question whether the convergence order is 1, although numerical experiments corroborates this fact.

The result is stated as follows, see [37].

Theorem 20. *Assume that the domain D is bounded and has a $\mathscr{C}^3$-boundary, that b, σ are $\mathscr{C}_b^2$ and $f \in \mathscr{C}_b^{1,2}$. Let $n(y)$ be the unit inward normal vector to the boundary ∂D at the closest[20] point to y on the boundary. Set*

$$\hat{\tau}^{x,h} = \inf\left\{ih > 0 : X_{ih}^{x,h} \notin D \text{ or } d(X_{ih}^{x,h}, \partial D) \le c_0\sqrt{h}\left|n^\top \sigma\right|(X_{ih}^{x,h})\right\}.$$

Then, we have

$$\mathbb{E}[f(t - \hat{\tau}^{x,h} \wedge t, X_{\hat{\tau}^{x,h} \wedge t}^x)] - \mathbb{E}[f(t - \tau^x \wedge t, X_{\tau^x \wedge t}^x)] = o(\sqrt{h}).$$

Observe that this improvement is very cheap regarding the computational cost. It can be extended (regarding to the numerical scheme and its mathematical analysis) to a source term, to time-dependent domain and to stationary problems (elliptic PDE).

Complementary References. See [2, 13, 26, 49, 53, 64] for general references. For reflected processes and Neumann boundary conditions, see [10, 28]. For variance reduction techniques, see [34, 47, 58]. For domain decomposition, see [35, 62]. This list is not exhaustive.

---

[20]Uniquely defined if y is close to the boundary.

5 Backward Stochastic Differential Equations and Semi-linear PDEs

The link between PDEs and stochastic processes have been developed since several decades and more recently, say in the last 20 years, researchers have paid attention to the probabilistic interpretation of non-linear PDEs, and in particular semi-linear PDEs. These PDEs are connected to non-linear processes, called Backward Stochastic Differential Equations (BSDE in short). In this section, we define these equations, firstly introduced by Pardoux and Peng [60], and give their connection with PDEs. Finally, we present a Monte Carlo algorithm to simulate them, using empirical regressions: it has the advantage to suit well the case of multidimensional problems, with a great generality on the type of semi-linearity.

These equations have many fruitful applications in stochastic control theory and mathematical finance, where they usually provide elegant proofs to characterize the solution to optimal investment problems for instance; for the related applications, we refer to reader to [17, 18]. Regarding the semi-linear PDE point of view, the applications are reaction-diffusion equations in chemistry [24], evolution of species in population biology [51,66], Hodgkin–Huxley model in neuroscience [43], Allen–Cahn equation for phase transition in physics... see the introductive course [30] and references therein. For other non-linear equations with connections with stochastic processes, see the aforementioned reference.

5.1 Existence of BSDE and Feynman–Kac Formula

5.1.1 Heuristics

As a difference with a Stochastic Differential Equation defined by (46) where the initial condition is given and the dynamics is imposed, a Backward SDE is defined through a random *terminal condition* ξ at a fixed terminal T and a dynamics imposed by a *driver g*. It takes the form

$$Y_t = \xi + \int_t^T g(s, Y_s, Z_s) \mathrm{d}s - \int_t^T Z_s \mathrm{d}W_s \qquad (61)$$

where we write the integrals between t and T to emphasize the backward point of view: ξ should be thought as a stochastic target to reach at time T. A solution to (61) is the couple (Y, Z): without extra conditions, the problem has an infinite number of solutions and thus is ill-posed. For instance, if $g \equiv 0$ and $\xi = f(W_T)$: taking $c \in \mathbb{R}$, a solution is $Z_t = c$ and $Y_t = \xi + c(W_T - W_t)$, thus uniqueness fails. In addition to integrability properties (appropriate L_2-spaces) that we do not detail, an important condition is that the solution does not anticipate the future of Brownian motion, i.e. the solution Y_t depends on the Brownian Motion W up to t, and similarly to Z: we

informally say that *the solution is adapted to W*. In a stochastic control problem, this adaptedness constraint is natural since it states that the value function or the decision can not be made in advance to the flow of information given by W. Observe that in the uniqueness counter-example, Y is not adapted to W since Y_t depends on the Brownian motion on $[0, T]$ and not only on $[0, t]$.

Taking the conditional expectation in (61) gives

$$Y_t = \mathbb{E}\Big(\xi + \int_t^T g(s, Y_s, Z_s)\mathrm{d}s \Big| W_s : s \le t\Big), \tag{62}$$

because the stochastic integral (built with Brownian increments after t) is centered conditionally on the Brownian motion up to time t. Of course, this rule is fully justified by the stochastic calculus theory. Since Y_t in (62) is adapted to W, it should be the right solution (if unique); then, Z serves as a control to make the equation (61) valid (with Y adapted).

5.1.2 Feynman–Kac Formula

The connection with PDE is possible when the terminal condition is a function of a (forward) SDE: this case is called *Markovian BSDE*. Additionally, the driver may depend also on this SDE as $g(s, X_s, Y_s, Z_s)$ for a deterministic function g. We now proceed by a verification theorem. To allow a more natural presentation as backward system, we choose to write the semi-linear PDE with a terminal condition at time T instead of an initial condition at time 0.

Theorem 21. *Let $T > 0$ be given. Under the assumptions of Theorem 11, let X^x be the solution (46) starting from $x \in \mathbb{R}^d$, assume there is a solution $v \in \mathscr{C}_b^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$ to the semi-linear PDE*

$$\begin{cases} v_t'(t, x) + L_{b,\sigma\sigma^\top}^X v(t, x) + g(t, x, v(t, x), Dv(t, (x)\sigma(x)) = 0, \\ v(T, x) = f(x), \end{cases} \tag{63}$$

for two given functions $f : \mathbb{R}^d \to \mathbb{R}$ and $g : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R} \otimes \mathbb{R}^d) \to \mathbb{R}$. Then, $Y_t^x = v(t, X_t^x)$ and $Z_t^x = [Dv\,\sigma](t, X_t^x)$ solves the BSDE

$$Y_t^x = f(X_T^x) + \int_t^T g(s, X_s^x, Y_s^x, Z_s^x)\mathrm{d}s - \int_t^T Z_s^x \mathrm{d}W_s. \tag{64}$$

Proof. The Itô formula (50) applied to v and X^x gives

$$\mathrm{d}v(s, X_s^x) = \big[v_s'(s, X_s^x) + L_{b,\sigma\sigma^\top}^X v(s, X_s^x)\big]\mathrm{d}s + Dv(s, X_s^x)\sigma(X_s^x)\mathrm{d}W_s$$
$$= -g(s, X_s^x, v(s, X_s^x), [Dv\,\sigma](s, X_s^x))\mathrm{d}s + Dv(s, X_s^x)\sigma(X_s^x)\mathrm{d}W_s,$$

which writes between $s = t$ and $s = T$:

$$v(T, X_T^x) = v(t, X_t^x) - \int_t^T g(s, X_s^x, v(s, X_s^x), [Dv\,\sigma](s, X_s^x))\mathrm{d}s$$

$$+ \int_t^T Dv(s, X_s^x)\sigma(X_s^x)\mathrm{d}W_s.$$

Since $v(T, .) = f(.)$, we complete the proof by identification.   □

In particular, at time 0 where $X_0^x = x$, we obtain $Y_0^x = v(0, x)$ and in view of (62), it gives a Feynman–Kac representation to v:

$$v(0, x) = \mathbb{E}\Big(f(X_T^x) + \int_t^T g(s, X_s^x, Y_s^x, Z_s^x)\mathrm{d}s \Big). \qquad (65)$$

As in case of linear PDEs, the assumption of uniform smoothness on v up to T is too strong to include the case of non-smooth terminal function f. But with an extra ellipticity condition, as for the heat equation, the solution becomes smooth immediately away from T (see [21]) and a similar verification could be checked under milder conditions.

The above Backward SDE (64) is coupled to a Forward SDE, but the latter is not coupled to the BSDE. Another interesting extension is to allow the coupling in both directions by having the coefficients of X dependent on v, i.e. $b(x)$ and $\sigma(x)$ become functions of $x, v(t, x), Dv(t, (x)$. The resulting process is called a Forward Backward Stochastic Differential Equations and is related to Quasi-Linear PDEs, where the operator $L_{b,\sigma\sigma^\top}^X$ also depends on v and Dv, see [56].

5.1.3   Other Existence Results Without PDE Framework

So far, only Markovian BSDEs are presented but from the probabilistic point of view, the Markovian structure is not required to define a solution: what is really crucial is the ability to represent a random variable built from $(W_s : s \leq T)$ as a stochastic integral w.r.t. the Brownian motion. This point has been discussed in Corollary 4. Then, in the simple case where g is Lipschitz w.r.t. y, z, (Y, Z) are built by means of an usual fixed point procedure in suitable L_2-norms and of this stochastic integral representation. We now state a more general existence and uniqueness result for BSDE, which is valid without any underlying (finite-dimensional) semi-linear PDE, we omit the proof.

Theorem 22. *Let $T > 0$ be fixed and assume the assumptions of Theorem 11 for the existence of X and that*

- *The terminal condition $\xi = f(X_s : s \leq T)$ is a square integrable functional of the stochastic process $(X_s : s \leq T)$.*

- *The measurable function $g : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R} \otimes \mathbb{R}^d)$ is uniformly Lipschitz in (y, z):*

$$|g(t, x, y_1, z_1) - g(t, x, y_2, z_2)| \leq C_g(|y_1 - y_2| + |z_1 - z_2|),$$

  uniformly in (t, x).
- *The driver is square integrable at $(y, z) = (0, 0)$: $\mathbb{E}(\int_0^T g^2(t, X_t, 0, 0)dt) < +\infty$.*

Then, there exists a unique solution (Y, Z), adapted and in L_2-spaces, to

$$Y_t = f(X_s : s \leq T) + \int_t^T g(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s.$$

Many works have been done in the last decade to go beyond the case of Lipschitz driver, which may be too stringent for some applications. In particular, having g with quadratic growth in Z is particularly interesting in exponential utility maximization problem (the non-linear PDE term is quadratic in $|Dv|$). This leads to quadratic BSDEs (see for instance [50]). A simple example of such BSDEs can be cooked up from heat equation and Brownian motion. Namely from Corollary 4, for a smooth function f with compact support, set $u(t, x) = \mathbb{E}(\exp(f(x + W_t)))$ and $v(t, y) = u(1 - t, y)$, so that

$$\exp(f(W_1)) = u(1, 0) + \int_0^1 u'_x(1 - s, W_s)dW_s,$$

$$u(1 - t, W_t) = u(1, 0) + \int_0^t u'_x(1 - s, W_s)dW_s,$$

$$v(t, W_t) = \exp(f(W_1)) - \int_t^1 v'_x(s, W_s)dW_s,$$

and by setting $Y_t = \log(v(t, W_t))$ and $Z_t = v'_x(t, W_t)/Y_t$, we obtain

$$Y_t = f(W_1) + \int_t^1 \frac{1}{2}Z_s^2 ds - \int_t^1 Z_s dW_s,$$

which is the simplest quadratic BSDE.

5.2 Time Discretization and Dynamic Programming Equation

5.2.1 Explicit and Implicit Schemes

To perform the simulation, a first stage may be the derivation of a discretization scheme, written backwardly in time (*backward dynamic programming equation*).

For the further analysis, assume that the terminal condition is of the form $\xi = f(X_T)$ where X is standard (forward) SDE.

Consider a time grid with N time steps $\pi = \{0 = t_0 < \cdots < t_i < \cdots < t_N = T\}$, with possibly non uniform time step, and set $|\pi| = \max_i(t_{i+1} - t_i)$. We will suppose later that $|\pi| \to 0$.

We write $\Delta_i = t_{i+1} - t_i$ and $\Delta W_i = W_{t_{i+1}} - W_{t_i}$. Writing the Eq. (64) between times t_i and t_{i+1}, we have

$$Y_{t_i} = Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} g(s, X_s, Y_s, Z_s)\mathrm{d}s - \int_{t_i}^{t_{i+1}} Z_s\mathrm{d}W_s.$$

Then, by applying simple approximations for $\mathrm{d}s$ and $\mathrm{d}W_s$ integrals and by replacing X by a Euler scheme computed along the grid π (and denoted X^π), we may define the discrete BSDE as

$$(Y_{t_i}^\pi, Z_{t_i}^\pi) = \underset{(Y,Z)\in L_2(\mathscr{F}_{t_i}^\pi)}{\arg\min} \mathbb{E}(Y_{t_{i+1}}^\pi + \Delta_i g(t_i, X_{t_i}^\pi, Y, Z) - Y - Z\Delta W_i)^2$$

with the initialization $Y_T^\pi = f(X_T^\pi)$ at $i = N$, where $L_2(\mathscr{F}_{t_i}^\pi)$ stands for the set of random variables (with appropriate dimension) that are square integrable and depend on the Brownian motion increments $(\Delta W_j : j \leq i - 1)$. The latter property is the measurability w.r.t. the sigma-field $\mathscr{F}_{t_i}^\pi$ generated by $(\Delta W_j : j \leq i - 1)$.

Then, a direct computation using the properties of Brownian increments gives

$$\begin{cases} Y_T^\pi = f(X_T^\pi), \\ Z_{t_i}^\pi = \dfrac{1}{\Delta_i}\mathbb{E}(Y_{t_{i+1}}^\pi \Delta W_i^\top | \mathscr{F}_{t_i}^\pi), \quad i < N \\ Y_{t_i}^\pi = \mathbb{E}(Y_{t_{i+1}}^\pi + \Delta_i g(t_i, X_{t_i}^\pi, Y_{t_i}^\pi, Z_{t_i}^\pi)|\mathscr{F}_{t_i}^\pi), \quad i < N. \end{cases} \tag{66}$$

This is the *implicit scheme* since the arguments of the function at the r.h.s. depend on the quantity $Y_{t_i}^\pi$ to compute on the l.h.s. Nevertheless, since g is uniformly Lipschitz in y, it is not difficult to show that the Dynamic Programming Equation (DPE in short) (66) is well-defined for $|\pi|$ small enough and that $Y_{t_i}^\pi$ can be computed using a Picard iteration procedure.

It is easy to turn the previous scheme into an explicit scheme and therefore, to avoid this extra Picard procedure. It writes

$$\begin{cases} Y_T^\pi = f(X_T^\pi), \\ Z_{t_i}^\pi = \dfrac{1}{\Delta_i}\mathbb{E}(Y_{t_{i+1}}^\pi \Delta W_i^\top | \mathscr{F}_{t_i}^\pi), \quad i < N \\ Y_{t_i}^\pi = \mathbb{E}(Y_{t_{i+1}}^\pi + \Delta_i g(t_i, X_{t_i}^\pi, Y_{t_{i+1}}^\pi, Z_{t_i}^\pi)|\mathscr{F}_{t_i}^\pi), \quad i < N. \end{cases} \tag{67}$$

In our personal experience on numerics, we have not observed a significant outperformance of one scheme on another. Moreover, from the theoretical point

of view, both schemes exhibit the same rates of convergence w.r.t. $|\pi|$, at least when the driver is Lipschitz.

The explicit scheme is the simplest one, and this is the one that we recommend in practice.

5.2.2 Time Discretization Error

Define the measure of the quadratic error

$$\mathscr{E}(Y^\pi - Y, Z^\pi - Z) = \max_{0 \leq i \leq N} \mathbb{E}|Y_{t_i}^\pi - Y_{t_i}|^2 + \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|Z_{t_i}^\pi - Z_t|^2 \mathrm{d}t.$$

Although not explicitly mentioned in the previous existence results on BSDE, this type of norm is appropriate to perform the fixed point argument in the proof of Theorem 22. We now state an error estimate [33], in order to show the convergence of the DPE to the BSDE.

Theorem 23. *For a Lipschitz driver w.r.t. (x, y, z) and $\frac{1}{2}$-Hölder w.r.t. t, there is a constant C independent on π such that we have*

$$\mathscr{E}(Y^\pi - Y, Z^\pi - Z) \leq C \Big(|\pi| + \sup_{i \leq N} \mathbb{E}|X_{t_i}^\pi - X_{t_i}|^2 + \mathbb{E}|f(X_T^\pi) - f(X_T)|^2$$

$$+ \sum_{i=0}^{N-1} \frac{1}{\Delta_i} \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_{i+1}} \mathbb{E}|Z_t - Z_s|^2 \mathrm{d}s \, \mathrm{d}t\Big).$$

Let us discuss on the nature and the magnitude of different error contributions.

- First, we face the *strong approximation error* of the forward SDE by its Euler scheme. Here we rather focus on convergence of paths (in L_2-norms), whereas in Sect. 4.1.3, we have studied the convergence of expectations of function of X_T^π towards those of X_T. Anyway, the problem is now well-understood: under a Lipschitz condition on b and σ, we can prove $\sup_{i \leq N} \mathbb{E}|X_{t_i}^\pi - X_{t_i}|^2 = O(|\pi|)$.
- Second, we should ensure a good *strong approximation* of the terminal conditions: if f is Lipschitz continuous, it readily follows from the previous term and $\mathbb{E}|f(X_T^\pi) - f(X_T)|^2 = O(|\pi|)$. For non Lipschitz f, there are partial answers, see [3].
- Finally, the last contribution $\sum_{i=0}^{N-1} \frac{1}{\Delta_i} \int_{t_i}^{t_{i+1}} \int_{t_i}^{t_{i+1}} \mathbb{E}|Z_t - Z_s|^2 \mathrm{d}s \, \mathrm{d}t$ is related to the L_2-*regularity of Z* (or equivalently of the gradient of the semi-linear PDE along the X-path) and it is intrinsic to the BSDE-solution. For smooth data, Z has the same regularity of Brownian paths and this error term is $O(|\pi|)$. For non smooth f (but under ellipticity condition on X), the L_2-norm of Z_t blows up as $t \to T$ and the rate $|\pi|$ usually worsens: for instance for $f(x) = 1_{x \geq 0}$, it becomes $N^{-\frac{1}{2}}$ for uniform time-grid.

Fig. 8 On the horizontal axis, uniform grid. On the vertical axis, the grid $(t_k^{\bar{\theta}} : 0 \le k \le N)$, with $T = 1$
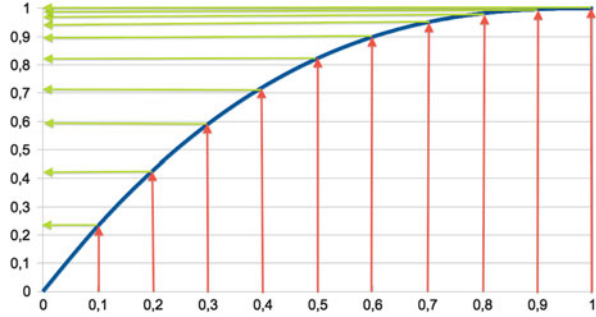
The analysis is very closely related to the fractional smoothness of $f(X_T)$ briefly discussed in Sect. 4.1.5, see also [25]. Choosing an appropriate grid of the form (see Fig. 8)

$$t_k^{\bar{\theta}} = T - T(1 - k/N)^{1/\bar{\theta}} \quad (\bar{\theta} \in (0, 1])$$

compensates this blow-up (for appropriate value of $\bar{\theta}$) and enables to retrieve the rate N^{-1}.

Actually in [31], it is shown that the upper bounds in Theorem 23 can be refined for smooth data, to finally obtain that the main error comes from *strong approximation error* on the forward component. This is an incentive to accurately approximate the SDE in L_2-sense.

5.2.3 Towards the Resolution of the Dynamic Programming Equation

The effective implementation of the explicit scheme (67) requires the iterative computations of conditional expectations: this is discussed in the next paragraphs.

Prior to this, we make some preliminary simplifications for the sake of conciseness. First, we consider the case of g independent of z,

$$g(t, x, y, z) = g(t, x, y),$$

therefore we only approximate Y^π; the general case is detailed in [39,54]. Second, it can be easily seen that it is enough to take the conditioning w.r.t. $X_{t_i}^\pi$ instead of $\mathscr{F}_{t_i}^\pi$, because of the Markovian property of X^π along the grid π and of the independent Brownian increments. Thus, (67) becomes

$$\begin{cases} Y_T^\pi = f(X_T^\pi), \\ Y_{t_i}^\pi = \mathbb{E}(Y_{t_{i+1}}^\pi + \Delta_i g(t_i, X_{t_i}^\pi, Y_{t_{i+1}}^\pi)|X_{t_i}^\pi), \quad i < N. \end{cases} \tag{68}$$

The same arguments apply to assert that for a (measurable) deterministic function y_i^π we have

$$y_i^\pi(X_{t_i}^\pi) = Y_{t_i}^\pi. \tag{69}$$

Therefore, simulating Y^π is equivalent to the computation of the functions y_i^π for any i and the simulation of the process X^π.

5.3 Approximation of Conditional Expectations Using Least-Squares Method

5.3.1 Empirical Least-Squares Problem

We adopt the point of view of conditional expectation as a *projection operator in* L_2. This is not the only possible approach, but it has the advantages (as it will be seen later)

1. To be much flexible w.r.t. the knowledge on the model for X (or X^π): only independent simulations of X^π are required (which is straightforward to perform).
2. To be little demanding on the assumptions on the underlying stochastic model: in particular, no ellipticity nor degeneracy condition are required, it could also include jumps (corresponding to PDE with a non-local Integro-Differential operator).
3. To provide robust theoretical error estimates, which allow to optimally tune the convergence parameters.
4. To be possibly adaptive to the data (*data-driven* scheme).

We recall that if a scalar random variable R (called the *response*) is square integrable, the conditional expectation of R given another possibly multidimensional r.v. O (called the *observation*) is given by

$$\mathbb{E}(R|O) = \underset{m(O) \text{ s.t. } m(.) \text{ is a meas. funct. with } \mathbb{E}|m(O)|^2 < +\infty}{\text{Arg min}} \mathbb{E}|R - m(O)|^2.$$

This is a least-squares problem in infinite dimension, also called *regression problem*. Usually in this context of BSDE simulation, none of the distributions of O, R or (O, R) is known in analytical and tractable form: thus an exact computation of $\mathbb{E}(R|O)$ is hopeless. The difficulty remains unchanged if we approximate the regression function

$$m(\cdot) = \mathbb{E}(R|O = \cdot)$$

on a finite dimensional functions basis. Alternatively, we can rely on independent simulations of (O, R) to compute an empirical version of m. This is the approach subsequently developed.

The basis functions are $(\phi_k(.))_{1 \leq k \leq K}$ and we assume that $\mathbb{E}|\phi_k(O)|^2 < +\infty$ for any k. We emphasize that

we can not assume that $(\phi_k(O))_{1 \leq k \leq K}$ forms an orthonormal basis in L_2,

since in our setting, the distribution of O is not explicit. Using this finite dimensional approximation, we anticipate to unfortunately retrieve the curse of dimensionality: the larger the dimension d of O, the larger the required K for a good accuracy of m, the larger the complexity.

We compute the coefficients on the basis by solving a *empirical least-squares problem*

$$(\alpha_k^M)_k = \arg\min_{\alpha \in \mathbb{R}^K} \frac{1}{M} \sum_{i=1}^{M} (R_i - \sum_{k=1}^{K} \alpha_k \phi_k(O_i))^2,$$

where $(O_i, R_i)_{1 \leq i \leq M}$ are independent simulations of the couple (O, R). Then, for the approximation of m, we set

$$\tilde{m}_M(.) = \sum_{k=1}^{K} \alpha_k^M \phi_k(.).$$

To efficiently compute the coefficients $(\alpha_k^M)_k$, we might use a SVD decomposition to account for instability issues, see [41].

5.3.2   Model-Free Error Estimates

Without extra assumptions on the model, we can derive model-free error estimates, see [42].

Theorem 24. *Assume that*

- $R = m(O) + \epsilon$ *with* $\mathbb{E}(\epsilon|O) = 0$.[21]
- $(O_1, R_1), \cdots, (O_M, R_M)$ *are independent copies of* (O, R).
- $\sigma^2 = \sup_x \mathbb{V}\mathrm{ar}(R|O = x) < +\infty$.
- *Let K be a finite positive integer and Φ be the linear vector space spanned by some functions $(\phi_1, \ldots \phi_K)$, with $\dim(\Phi) \leq K$.[22]*

---

[21]Meaning that $m(O) = \mathbb{E}(R|O)$.

[22]There may be some colinearities within $(\phi_j)_{1 \leq j \leq K}$.

Denote by μ^M the empirical measure associated to $(O_1, \cdots, O_M)$, μ the probability measure of O and by $|\phi|_M^2 = \frac{1}{M} \sum_{i=1}^{M} \phi^2(O_i)$ the empirical L_2-measure of ϕ w.r.t. μ^M, and set:

$$\tilde{m}_M(.) = \arg\min_{\phi \in \Phi} \frac{1}{M} \sum_{i=1}^{M} |\phi(O_i) - R_i|^2. \tag{70}$$

Then

$$\mathbb{E}(|\tilde{m}_M - m|_M^2) \leq \sigma^2 \frac{K}{M} + \min_{\phi \in \Phi} |\phi - m|_{L_2(\mu)}^2.$$

The first term in the r.h.s. above is interpreted as a *statistical error*[23] term (due to a finite sample to compute the empirical coefficients), while the second term is an *approximation error of the functions class*[24] (due to finite-dimensional vector space). The first term converges to 0 as $M \to +\infty$ but it blows up if $K \to +\infty$, while the second one converges to 0 as $K \to +\infty$ (as least if Φ asymptotically spans all the functions in $L_2(\mu)$). This bias-variance decomposition shows that there is a necessary trade-off between K and M to ensure a convergent approximation. Without this right balance, the approximation (70) may be not convergent. Furthermore, the parameter tuning can also be made optimally.

In the quoted reference [42], the space Φ could also depend on the simulations (data-driven approximation spaces).

Proof. Assume that

$$\mathbb{E}\left(|\tilde{m}_M - m|_M^2 \big| O_1, \cdots, O_M\right) \leq \sigma^2 \frac{K}{M} + \min_{\phi \in \Phi} |\phi - m|_M^2. \tag{71}$$

Then, the announced result directly follows by taking expectations and observing that

$$\mathbb{E}\left(\min_{\phi \in \Phi} |\phi - m|_M^2\right) \leq \min_{\phi \in \Phi} \mathbb{E}(|\phi - m|_M^2) = \min_{\phi \in \Phi} |\phi - m|_{L_2(\mu)}^2.$$

We now prove (71). As far as computations conditionally on $O_1, \cdots, O_M$ are concerned, without loss of generality we can assume that $(\phi_1, \ldots \phi_{K_M})$ is an orthonormal family in $L_2(\mu^M)$, with possibly $K_M \leq K$:

$$\frac{1}{M} \sum_{i=1}^{M} \phi_k(O_i)\phi_l(O_i) = \delta_{k,l}.$$

---

[23] Also called variance term.

[24] Squared bias term.

Consequently, the solution $\arg\min\limits_{\phi \in \Phi} \dfrac{1}{M} \sum\limits_{i=1}^{M} |\phi(O_i) - R_i|^2$ is given by

$$\tilde{m}_M(.) = \sum_{j=1}^{K_M} \alpha_j \phi_j(.) \quad \text{with} \quad \alpha_j = \frac{1}{M} \sum_{i=1}^{M} \phi_j(O_i) R_i.$$

Now, set $\mathbb{E}^*(.) = \mathbb{E}(.|O_1, \cdots, O_M)$. Then, observe that $\mathbb{E}^*(\tilde{m}_M(.))$ is the least-squares solution to $\min\limits_{\phi \in \Phi} \dfrac{1}{M} \sum_{i=1}^{M} |\phi(O_i) - m(O_i)|^2 = \min\limits_{\phi \in \Phi} |\phi - m|_M^2$. Indeed:

- On the one hand, the above least-squares solution is given by $\sum_{j=1}^{K_M} \alpha_j^* \phi_j(.)$ with $\alpha_j^* = \frac{1}{M} \sum_{i=1}^{M} \phi_j(O_i) m(O_i)$.
- On the other hand, $\mathbb{E}^*(\tilde{m}_M(.)) = \sum_{j=1}^{K_M} \mathbb{E}^*(\alpha_j) \phi_j(.)$ and $\mathbb{E}^*(\alpha_j) = \frac{1}{M} \sum_{i=1}^{M} \phi_j(O_i) \mathbb{E}^*(R_i) = \frac{1}{M} \sum_{i=1}^{M} \phi_j(O_i) \mathbb{E}(m(O_i) + \epsilon_i | O_1, \cdots, O_M) = \alpha_j^*$.

Thus, by the Pythagoras theorem, we obtain

$$|\tilde{m}_M - m|_M^2 = |\tilde{m}_M - \mathbb{E}^*(\tilde{m}_M)|_M^2 + |\mathbb{E}^*(\tilde{m}_M) - m|_M^2,$$

$$\mathbb{E}^*|\tilde{m}_M - m|_M^2 = \mathbb{E}^*|\tilde{m}_M - \mathbb{E}^*(\tilde{m}_M)|_M^2 + |\mathbb{E}^*(\tilde{m}_M) - m|_M^2$$

$$= \mathbb{E}^*|\tilde{m}_M - \mathbb{E}^*(\tilde{m}_M)|_M^2 + \min_{\phi \in \Phi} |\phi - m|_M^2.$$

Since $(\phi_j)_j$ is orthonormal in $L_2(\mu_M)$, we have $|\tilde{m}_M - \mathbb{E}^*(\tilde{m}_M)|_M^2 = \sum_{j=1}^{K_M} |\alpha_j - \mathbb{E}^*(\alpha_j)|^2$. Since $\alpha_j - \mathbb{E}^*(\alpha_j) = \frac{1}{M} \sum_{i=1}^{M} \phi_j(O_i)(R_i - m(O_i))$, we obtain

$$\mathbb{E}^*|\tilde{m}_M - \mathbb{E}^*(\tilde{m}_M)|_M^2 = \sum_{j=1}^{K_M} \frac{1}{M^2} \mathbb{E}^* \sum_{i,l=1}^{M} \phi_j(O_i)\phi_j(O_l)(R_i - m(O_i))(R_l - m(O_l))$$

$$= \sum_{j=1}^{K_M} \frac{1}{M^2} \sum_{i=1}^{M} \phi_j^2(O_i) \mathbb{V}\mathrm{ar}(R_i | O_i)$$

taking advantage that the $(\epsilon_i)_i$ conditionally on $(O_1, \cdots O_M)$ are centered. This proves

$$\mathbb{E}^*|\tilde{m}_M - \mathbb{E}^*(\tilde{m}_M)|_M^2 \le \sigma^2 \sum_{j=1}^{K_M} \frac{1}{M^2} \sum_{i=1}^{M} \phi_j^2(O_i) = \sigma^2 \frac{K_M}{M} \le \sigma^2 \frac{K}{M}.$$

The proof of (71) is complete.   □

5.3.3 Least-Squares Method for Solving Discrete BSDE

We now apply the previous empirical least-squares method to numerically solve the DPE (68). For simplicity of exposure, we consider here only uniform time grids with N time steps, $\Delta_i = T/N$. In addition to assumptions of Theorem 23, we assume that the terminal condition $f(.)$ is bounded: then, we can easily establish the following result.

Proposition 20. *Under these assumptions, the function $y_i^\pi(.)$ defined in (69) is bounded by a constant $C_\star$, which is independent on N and i.*

Actually, $C_\star$ can be given explicitly in terms of the data. To force the stability in the iterative computations of conditional expectations (68), we truncate the numerical solution at the level $C_\star$ using the soft thresholding

$$[\psi]_{C_\star} := -C_\star \vee \psi \wedge C_\star.$$

Algorithm for Approximating $y_k^\pi(\cdot)$. At each time index $0 \leq k \leq N-1$, we consider a vector space Φ_k spanned by basis functions $p_k(\cdot)$, which are understood as vectors of K_k functions. The final approximation of $y_k^\pi(\cdot)$ has the form

$$y_k^{\pi,M}(\cdot) = [\alpha_k^M \cdot p_k(\cdot)]_{C_\star}.$$

The coefficients α_k^M are computed with M independent simulations of $(X_{t_k}^\pi)_k$, that are denoted by $\{(X_{t_k}^{\pi,m})_k\}_{1\leq m\leq M}$: this single set of simulated paths are used to compute all the coefficients at once. This is done as follows:

- Initialization : for $k = N$, take $y_N^\pi(\cdot) = f(\cdot)$.
- Iteration : for $k = N-1, \cdots, 0$, solve the least-squares problem

$$\alpha_k^M = \arg\min_{\alpha\in\mathbb{R}^{K_k}} \sum_{m=1}^M |y_{k+1}^{\pi,M}(X_{t_{k+1}}^{\pi,m}) + \Delta_k g(t_k, X_{t_k}^{\pi,m}, y_{k+1}^{\pi,M}(X_{t_{k+1}}^{\pi,m})) - \alpha \cdot p_k(X_{t_k}^{\pi,m})|^2$$

and define $y_k^{\pi,M}(\cdot) = [\alpha_k^M \cdot p_k(\cdot)]_{C_\star}$.

Error Analysis. We now turn to the error estimates. The analysis combines the BSDE techniques (a priori estimates using stochastic calculus), regression tools as those exposed in Sect. 5.3.2, but there is a slight difference which actually requires a significant improvement in the arguments. Since we use a single set of independent paths, the "responses" $(y_{k+1}^{\pi,M}(X_{t_{k+1}}^{\pi,m}))_{1\leq m\leq M}$ are not independent, because of their dependence through the function $y_{k+1}^{\pi,M}$. To overcome this interdependence issue in the proof, we shall replace the random function $y_{k+1}^{\pi,M}$ by a deterministic neighbor: of course, there is a complexity cost to cover the different function spaces in order to provide close neighbors, and the covering numbers are well controlled using the Vapnik–Chervonenkis dimension, when the function spaces are *bounded*

(Proposition 20). This is the technical reason why we consider bounded functions. We now state a result regarding the global error, see [30, Theorem VIII.3.4] for full details.

Theorem 25. *Under the previous notations and assumptions, there is a constant $C > 0$ (independent on N) such that we have*

$$\max_{0 \le k \le N} \mathbb{E}|Y_{t_k}^{\pi} - y_k^{\pi,M}(X_{t_k}^{\pi})|^2 \le C \sum_{k=0}^{N-1} \left\{ N \underbrace{\frac{K_k}{M}}_{\text{statistical error}} + \underbrace{\min_{\phi \in \Phi_k} \mathbb{E}|y_k^{\pi}(X_{t_k}^{\pi}) - \phi(X_{t_k}^{\pi})|^2}_{\text{approximation error of functions class}} \right\}$$

$$+ C \max_{0 \le k \le N} \underbrace{\sqrt{\frac{K_k \log(M)}{M}}}_{\text{interdependence error}} .$$

When the Z-component has to be approximated as well, the estimates are slightly modified, see [54] for details.

Parameter Tuning. We conclude this analysis by providing an example of how to choose appropriately the parameters N, K_k and M. Assume that the value function y^{π} is Lipschitz continuous, uniformly in N (which usually follows from a Lipschitz terminal condition). Our objective is to achieve a global error of order $\varepsilon = \frac{1}{N}$ for $\max_{0 \le k \le N} \mathbb{E}|Y_{t_k}^{\pi} - y_k^{\pi,M}(X_{t_k}^{N})|^2$, i.e. the same error magnitude than the time-discretization error.

For the vector spaces Φ_k, we consider those generated by functions that are constant on disjoint hypercubes of small edge. Since X^{π} has exponential moments, we can restrict the partitioning to a compact set of $\mathbb{R}^d$ with size $c \log(N)$ in any direction, and the induced error is smaller than N^{-1} provided that c is large enough. If the edge of the hypercube is like N^{-1}, the vector spaces have dimension $K_k \sim N^d$ up to logarithmic factors: then, the terms from *approximation error of functions class* are $O(N^{-2})$ and they sum up to give a contribution $O(N^{-1})$ as required. A quick inspection of the upper bounds of Theorem 25 shows that the highest constraint on M comes from the statistical error: we obtain $M \sim cN^{3+d}$, up to logarithmic terms. The complexity of the scheme is of order NM (still neglecting the log terms), because the computation of all regression coefficients at a given date has a computational cost $O(M \log(N))$ due to our specific choice of function basis. Hence, the global complexity is

$$\mathscr{C} \sim \varepsilon^{-\frac{1}{4+d}}$$

up to logarithmic terms. Not surprisingly, the convergence order deteriorates as the dimension increases, this is the curse of dimensionality. Had the value function been smoother, we would have used local polynomials and the convergence order would have been improved: the smoother the functions, the better the convergence rate.

In practice, the algorithm has been performed on a computer up to dimension $d = 10$ with satisfactory results and rather short computational times (less than 1 min). There are several possible improvements to this basic version of the algorithm.

- We can use variance reduction techniques, see [8, 9].
- Instead of writing the DPE between t_i and t_{i+1}, it can be written between t_i and T: it has the surprising effect (mathematically justified) to reduce the propagation of errors in the DPE. This scheme is called MDP scheme (for Multi step forward Dynamic Programming equation) and it is studied in [39].

Complementary References. For theoretical aspects, see [16, 56, 59, 61]; for applications, see [17, 18]; for numerics, see [5, 7, 11, 14, 32, 40, 54, 69]. This list is not exhaustive.

References

1. Achdou, Y., Pironneau, O.: Computational Methods for Option Pricing. SIAM Series. Frontiers in Applied Mathematics, Philadelphia (2005)
2. Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis. Stochastic Modelling and Applied Probability, vol. 57. Springer, New York (2007)
3. Avikainen, R.: On irregular functionals of SDEs and the Euler scheme. Financ. Stoch. **13**, 381–401 (2009)
4. Bachelier, L.: Théorie de la spéculation. Ph.D. thesis, Ann. Sci. École Norm. Sup. (1900)
5. Bally, V., Pagès, G.: Error analysis of the optimal quantization algorithm for obstacle problems. Stoch. Process. Appl. **106**(1), 1–40 (2003)
6. Bally, V., Talay, D.: The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. Probab. Theory Relat. Fields **104**(1), 43–60 (1996)
7. Bender, C., Denk, R.: A forward scheme for backward SDEs. Stoch. Process. Their Appl. **117**(12), 1793–1823 (2007)
8. Bender, C., Steiner, J.: Least-squares Monte Carlo for BSDEs. In: Carmona, R., Del Moral, P., Hu, P., Oudjane, N. (eds.) Numerical Methods in Finance. Series: Springer Proceedings in Mathematics, vol. 12. Springer, Berlin (2012)
9. Ben Zineb, T., Gobet, E.: Preliminary control variates to improve empirical regression methods. Monte Carlo Methods Appl. **19**(4), 331–354 (2013)
10. Bossy, M., Gobet, E., Talay, D.: Symmetrized Euler scheme for an efficient approximation of reflected diffusions. J. Appl. Probab. **41**(3), 877–889 (2004)
11. Bouchard, B., Touzi, N.: Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations. Stoch. Process. Their Appl. **111**, 175–206 (2004)
12. Breiman, L.: Probability. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1992) [Corrected reprint of the 1968 original]
13. Cessenat, M., Dautray, R., Ledanois, G., Lions, P.L., Pardoux, E., Sentis, R.: Méthodes probabilistes pour les équations de la physique. Collection CEA, Eyrolles (1989)
14. Crisan, D., Manolarakis, K.: Solving backward stochastic differential equations using the cubature method. Preprint (2010)

15. Durrett, R.: Brownian Motion and Martingales in Analysis. Wadsworth Mathematics Series. Wadsworth International Group, Belmont (1984)
16. El Karoui, N., Kapoudjian, C., Pardoux, E., Peng, S., Quenez, M.C.: Reflected solutions of backward SDE's and related obstacle problems for PDE's. Ann. Probab. **25**(2), 702–737 (1997)
17. El Karoui, N., Peng, S.G., Quenez, M.C.: Backward stochastic differential equations in finance. Math. Financ. **7**(1), 1–71 (1997)
18. El Karoui, N., Hamadène, S., Matoussi, A.: Backward stochastic differential equations and applications. In: Carmona, R. (ed.) Indifference Pricing: Theory and Applications, Chap. 8, pp. 267–320. Princeton University Press, Princeton (2008)
19. Föllmer, H.: Calcul d'Itô sans probabilités. In: Seminar on Probability, XV (University of Strasbourg, Strasbourg, 1979/1980) (French), pp. 143–150. Springer, Berlin (1981)
20. Freidlin, M.I.: Functional Integration and Partial Differential Equations. Annals of Mathematics Studies. Princeton University Press, Princeton (1985)
21. Friedman, A.: Partial Differential Equations of Parabolic Type. Prentice-Hall, Eglewood Cliffs (1964)
22. Friedman, A.: Stochastic Differential Equations and Applications, vol. 1. Academic, New York (1975) [A Subsidiary of Harcourt Brace Jovanovich, Publishers, XIII]
23. Friedman, A.: Stochastic Differential Equations and Applications, vol. 2. Academic, New York (1976) [A Subsidiary of Harcourt Brace Jovanovich, Publishers, XIII]
24. Gavalas, G.R.: Nonlinear Differential Equations of Chemically Reacting Systems. Springer Tracts in Natural Philosophy, vol. 17. Springer, New York (1968)
25. Geiss, C., Geiss, S., Gobet, E.: Generalized fractional smoothness and L_p-variation of BSDEs with non-Lipschitz terminal condition. Stoch. Process. Their Appli. **122**(5), 2078–2116 (2012)
26. Glasserman, P.: Monte Carlo Methods in Financial Engineering. Springer, New York (2003)
27. Gobet, E.: Euler schemes for the weak approximation of killed diffusion. Stoch. Process. Their Appl. **87**, 167–197 (2000)
28. Gobet, E.: Euler schemes and half-space approximation for the simulation of diffusions in a domain. ESAIM Probab. Stat. **5**, 261–297 (2001)
29. Gobet, E.: Advanced Monte Carlo methods for barrier and related exotic options. In: Ciarlet, P.G., Bensoussan, A., Zhang, Q. (eds.) Handbook of Numerical Analysis, vol. XV. Special Volume: Mathematical Modeling and Numerical Methods in Finance, pp. 497–528. Elsevier/North-Holland, Netherlands (2009)
30. Gobet, E.: Méthodes de Monte-Carlo et processus stochastiques: du linéaire au non-linéaire. Editions de l'École Polytechnique, Palaiseau (2013)
31. Gobet, E., Labart, C.: Error expansion for the discretization of backward stochastic differential equations. Stoch. Process. Their Appl. **117**(7), 803–829 (2007)
32. Gobet, E., Labart, C.: Solving BSDE with adaptive control variate. SIAM Numer. Anal. **48**(1), 257–277 (2010)
33. Gobet, E., Lemor, J.P.: Numerical simulation of BSDEs using empirical regression methods: Theory and practice. In: Proceedings of the Fifth Colloquium on BSDEs, 29th May–1st June 2005, Shangai. Available at http://hal.archives-ouvertes.fr/hal-00291199/fr/ (2006)
34. Gobet, E., Maire, S.: Sequential control variates for functionals of Markov processes. SIAM J. Numer. Anal. **43**(3), 1256–1275 (2005)
35. Gobet, E., Maire, S.: Sequential Monte Carlo domain decomposition for the Poisson equation. In: Proceedings of the 17th IMACS World Congress, Scientific Computation, Applied Mathematics and Simulation, Paris, 11–15 July 2005
36. Gobet, E., Makhlouf, A.: L_2-time regularity of BSDEs with irregular terminal functions. Stoch. Proces. Their Appli. **120**, 1105–1132 (2010)
37. Gobet, E., Menozzi, S.: Stopped diffusion processes: Boundary corrections and overshoot. Stoch. Process. Their Appl. **120**, 130–162 (2010)
38. Gobet, E., Munos, R.: Sensitivity analysis using Itô-Malliavin calculus and martingales. Application to stochastic control problem. SIAM J. Control Optim. 43(5), 1676–1713 (2005)

39. Gobet, E., Turkedjiev, P.: Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions (2013) [In Revision for Mathematics of Computation]. Available at http://hal.archives-ouvertes.fr/hal-00642685

40. Gobet, E., Lemor, J.P., Warin, X.: A regression-based Monte Carlo method to solve backward stochastic differential equations. Ann. Appl. Probab. **15**(3), 2172–2202 (2005)

41. Golub, G., Van Loan, C.F.: Matrix Computations, 3rd edn. The Johns Hopkins University Press, Balimore (1996)

42. Gyorfi, L., Kohler, M., Krzyzak, A., Walk, H.: A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics. Springer, New York (2002)

43. Hodgkin, A.L., Huxley, A.F., Katz, B.: Measurement of current-voltage relations in the membrane of the giant axon of Loligo. J. Physiol. **116**, 424–448 (1952). Available at http://www.sfn.org/skins/main/pdf/HistoryofNeuroscience/hodgkin1.pdf

44. Ito, K.: On stochastic differential equations. Mem. Am. Math. Soc. (4), 1–51 (1951)

45. Itô, K., McKean, H.P.: Diffusion Processes and Their Sample Paths. Springer, Berlin (1965)

46. Jacod, J., Protter, P.: Probability Essentials, 2nd edn. Springer, Berlin (2003)

47. Jourdain, B., Lelong, J.: Robust adaptive importance sampling for normal random vectors. Ann. Appl. Probab. **19**(5), 1687–1718 (2009)

48. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus, 2nd edn. Springer, New York (1991)

49. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, Berlin (1995)

50. Kobylanski, M.: Backward stochastic differential equations and partial differential equations with quadratic growth. Ann. Probab. **28**(2), 558–602 (2000)

51. Kolmogorov, A.N., Petrovsky, I.G., Piskunov, N.S.: Etude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. Bulletin Université d'État à Moscou, Série internationale A 1, pp. 1–26 (1937)

52. Kunita, H.: Stochastic differential equations and stochastic flows of diffeomorphisms. In: Ecole d'Eté de Probabilités de St-Flour XII, 1982. Lecture Notes in Mathematics, vol. 1097, pp. 144–305. Springer, Berlin (1984)

53. Lapeyre, B., Pardoux, E., Sentis, R.: Methodes de Monte Carlo pour les processus de transport et de diffusion. Collection Mathématiques et Applications, vol. 29. Springer, Berlin (1998)

54. Lemor, J.P., Gobet, E., Warin, X.: Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. Bernoulli **12**(5), 889–916 (2006)

55. Lévy, P.: Sur certains processus stochastiques homogènes. Compos. Math. **7**, 283–339 (1939)

56. Ma, J., Yong, J.: Forward-Backward Stochastic Differential Equations. Lecture Notes in Mathematics, vol. 1702. Springer, Berlin (1999)

57. Nelson, E.: Dynamical Theories of Brownian Motion. Princeton University Press, Princeton (1967)

58. Newton, N.J.: Variance reduction for simulated diffusions. SIAM J. Appl. Math. **54**(6), 1780–1805 (1994)

59. Pardoux, E.: Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order. In: Stochastic Analysis and Related Topics, VI (Geilo, 1996). Progress in Probability, vol. 42, pp. 79–127. Birkhäuser, Boston (1998)

60. Pardoux, E., Peng, S.G.: Adapted solution of a backward stochastic differential equation. Syst. Control Lett. **14**(1), 55–61 (1990)

61. Pardoux, E., Peng, S.: Backward stochastic differential equations and quasilinear parabolic partial differential equations. In: Stochastic Partial Differential Equations and Their Applications. Proceedings of the IFIP International Conference, Charlotte/NC, 1991. Lecture Notes in Control and Information Sciences, vol. 176, pp. 200–217. Springer, Berlin (1992)

62. Peirano, É., Talay, D.: Domain decomposition by stochastic methods. In: Domain Decomposition Methods in Science and Engineering, pp. 131–147 (electronic). National Autonomous University of Mexico, México (2003)

63. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion, 3rd edn. Comprehensive Studies in Mathematics. Springer, Berlin (1999)
64. Sabelfeld, K.K.: Monte Carlo Methods in Boundary Value Problems. Springer Series in Computational Physics. Springer, Berlin (1991) [translated from Russian]
65. Samuelson, P.A.: Proof that properly anticipated prices fluctuate randomly. Ind. Manag. Rev. **6**, 42–49 (1965)
66. Shigesada, N., Kawasaki, K. (eds.): Biological Invasions: Theory and Practice. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford (1997)
67. Stroock, D.W., Varadhan, S.R.S.: Multidimensional Diffusion Processes, 2nd edn. Comprehensive Studies in Mathematics. Springer, Berlin (1997)
68. Talay, D., Tubaro, L.: Expansion of the global error for numerical schemes solving stochastic differential equations. Stoch. Anal. Appl. **8**(4), 94–120 (1990)
69. Zhang, J.: A numerical scheme for BSDEs. Ann. Appl. Probab. **14**(1), 459–488 (2004)