1 Stochastic differential equations

1.1 Definitions

We call stochastic differential equation (SDE) an equation of the form

\[ dX_t = b(t, X_t) \, dt + \sigma(t, X_t) \, dB_t, \quad X_{t=0} = X_0 \]  

(1)

where \((B_t)_{t \geq 0}\) is a \(d\)-dimensional Brownian motion on a filtered probability space \((\Omega, \mathcal{F}, \mathcal{F}_t, P)\), \(X_0\) is \(\mathcal{F}_0\)-measurable, \(b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d\) and \(\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times d}\) have some regularity specified case by case and the solution \((X_t)_{t \geq 0}\) is a \(d\)-dimensional continuous adapted process. The meaning of the equation is the identity

\[ X_t = X_0 + \int_0^t b(s, X_s) \, ds + \int_0^t \sigma(s, X_s) \, dB_s \]  

(2)

where we have to assume conditions on \(b\) and \(\sigma\) which guarantees that \(s \mapsto b(s, X_s)\) is integrable, \(s \mapsto \sigma(s, X_s)\) is square integrable, with probability one; \(\int_0^t \sigma(s, X_s) \, dB_s\) is an Itô integral, and more precisely it is its continuous version in \(t\); and the identity has to hold uniformly in \(t\), with probability one. The generalization to different dimensions of \(B\) and \(X\) is obvious; we take the same dimension to have less notations.

Even if less would be sufficient with more arguments, let us assume that \(b\) and \(\sigma\) are at least continuous, so that the above mentioned conditions of integrability of \(s \mapsto b(s, X_s)\) and \(s \mapsto \sigma(s, X_s)\) are fulfilled.

In most cases, if \(X_0 = x_0\) is deterministic, when we prove that a solution exists, we can also prove that it is adapted not only to the filtration \((\mathcal{F}_t)\) but also to \((\mathcal{F}_t^B)\), the filtration associated to the Brownian motion; more precisely, its completion. This is just natural, because the input of the equation is only the Brownian motion. However, it is so natural if implicitly we think to have a suitable uniqueness. Otherwise, in principle, it is difficult to exclude that one can construct, maybe in some artificial ways, a solution which is not \(B\)-adapted. Indeed it happens that there are relevant examples of stochastic equations where solutions exist which are not \(B\)-adapted. This is the origin of the following definitions.

**Definition 1 (strong solutions)** We have strong existence for equation (1) if, given any filtered probability space \((\Omega, \mathcal{F}, \mathcal{F}_t, P)\) with a Brownian motion \((B_t)_{t \geq 0}\), given any deterministic initial condition \(X_0 = x_0\), there is a continuous \(\mathcal{F}_t\)-adapted process \((X_t)_{t \geq 0}\) satisfying (2) (in particular, we may choose \((\mathcal{F}_t) = (\mathcal{F}_t^B)\) and have a solution adapted to \(B\)). A strong solution is a solution adapted to \((\mathcal{F}_t^B)\).

**Definition 2 (weak solutions)** Given a deterministic initial condition \(X_0 = x_0\), a weak solution is the family composed of a filtered probability space \((\Omega, \mathcal{F}, \mathcal{F}_t, P)\) with a Brownian motion \((B_t)_{t \geq 0}\) and a continuous \(\mathcal{F}_t\)-adapted process \((X_t)_{t \geq 0}\) satisfying (2).
In the definition of weak solution, the filtered probability space and the Brownian motion are not specified a priori, they are part of the solution; hence we are not allowed to choose \((\mathcal{F}_t) = (\mathcal{F}_t^B)\).

When \(X_0\) is random, \(\mathcal{F}_0\)-measurable, the concept of weak solution is formally in trouble because the space where \(X_0\) has to be defined is not prescribed a priori. The concept of strong solution can be adapted for instance replacing \((\mathcal{F}_t^B)\) with \((\mathcal{F}_t^B \lor \mathcal{F}_0)\), or just saying that, if \((X_t)_{t \geq 0}\) is a solution on a prescribed space \((\Omega, \mathcal{F}, \mathcal{F}_t, P)\) where \(X_0\) and \(B\) are defined, then it is a strong solution. If we want to adapt the definition of weak solution to the case of random initial conditions, we have to prescribe only the law of \(X_0\) and put in the solution the existence of \(X_0\) with the given law.

Let us come to uniqueness. Similarly to existence, there are two concepts.

**Definition 3 (pathwise uniqueness)** We say that pathwise uniqueness holds for equation (1) if, given any filtered probability space \((\Omega, \mathcal{F}, \mathcal{F}_t, P)\) with a Brownian motion \((B_t)_{t \geq 0}\), given any deterministic initial condition \(X_0 = x_0\), if \((X_t^{(1)})_{t \geq 0}\) and \((X_t^{(2)})_{t \geq 0}\) are two continuous \(\mathcal{F}_t\)-adapted process which fulfill (2), then they are indistinguishable.

**Definition 4 (uniqueness in law)** We say that there is uniqueness in law for equation (1) if, given two weak solutions on any pair of spaces, their laws coincide.

### 1.2 Strong solutions

The most classical theorem about strong solutions and pathwise uniqueness holds, as in the deterministic case, under Lipschitz assumptions on the coefficients. Assume there are two constants \(L\) and \(C\) such that

\[
|b(t,x) - b(t,x')| \leq L|x - x'| \\
|\sigma(t,x) - \sigma(t,x')| \leq L|x - x'|
\]

\[
|b(t,x)| \leq C(1+|x|) \\
|\sigma(t,x)| \leq C(1+|x|)
\]

for all values of \(t\) and \(x\). The second condition on \(b\) and \(\sigma\) is written here for sake of generality, but as we assume, as said above, that \(b\) and \(\sigma\) are continuous, it follows from the first condition (the uniform in time Lipschitz property).

**Theorem 5** Under the previous assumptions on \(b\) and \(\sigma\), there is strong existence and pathwise uniqueness for equation (1). If, for some \(p \geq 2\), \(E[|X_0|^p] < \infty\), then

\[
E\left[\sup_{t \in [0,T]} |X_t|^p\right] < \infty.
\]

**Proof.** ... ■
1.3 Weak solutions

Let us see only a particular example of result about weak solutions. Assume that \( \sigma \) is constant and non-degenerate; for simplicity, assume it equal to the identity, namely consider the SDE with additive noise

\[
dX_t = b(t, X_t) \, dt + dB_t.
\]

Moreover, assume \( b \) only measurable and bounded (or continuous and bounded if we prefer to maintain the general assumption of continuity). The key features of these assumptions are: the noise is non-degenerate (hence more restrictive than above for strong solutions) but \( b \) is very weak, much weaker than the usual Lipschitz case. Under such assumption on \( b \), if we do not have the noise \( dB_t \) in the equation, it is easy to make examples without existence or without uniqueness.

**Theorem 6** Under these assumptions, for every \( x_0 \in \mathbb{R}^d \), there exists a weak solution and it is unique in law.

2 Links between SDEs and linear PDEs

2.1 Fokker-Planck equation

Along with the stochastic equation (1) defined by the coefficients \( b \) and \( \sigma \), we consider also the following parabolic PDE on \([0, T] \times \mathbb{R}^d\):

\[
\frac{\partial p}{\partial t} = \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j (a_{ij} p) - \text{div} (pb) , \quad p|_{t=0} = p_0
\]

called Fokker-Planck equation. Here

\[
a = \sigma \sigma^T.
\]

Although in many cases it has regular solutions, in order to minimize the theory it is convenient to introduce the concept of measure-valued solution \( \mu_t \); moreover we restrict to the case of probability measures. To give a meaning to certain integrals below, we assume (beside other assumptions depending on the result)

\( b, \sigma \) bounded continuous

but it will be clear that more cumbersome results can be done by little additional conceptual effort. We loosely write

\[
\frac{\partial \mu_t}{\partial t} = \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j (a_{ij} \mu_t) - \text{div} (\mu_t b) , \quad \mu|_{t=0} = \mu_0
\]
but we mean the following concept. By \( \langle \mu_t, \phi \rangle \) we mean \( \int_{\mathbb{R}^d} \phi (x) \mu_t (dx) \). By \( C_b ([0, T] \times \mathbb{R}^d) \) we denote the space of bounded continuous functions \( \varphi : [0, T] \times \mathbb{R}^d \to \mathbb{R} \) and by \( C_b^{1,2} ([0, T] \times \mathbb{R}^d) \) the space of functions \( \varphi \) such that \( \varphi, \frac{\partial \varphi}{\partial t}, \partial_i \varphi, \partial_i \partial_j \varphi \in C_b ([0, T] \times \mathbb{R}^d) \).

**Definition 7** A measure-valued solution of the Fokker-Planck equation (4) is a family of Borel probability measures \( (\mu_t)_{t \in [0,T]} \) on \( \mathbb{R}^d \) such that \( t \mapsto \langle \mu_t, \varphi (t, \cdot) \rangle \) is measurable for all \( \varphi \in C_b ([0, T] \times \mathbb{R}^d) \) and

\[
\langle \mu_t, \phi (t, \cdot) \rangle - \langle \mu_0, \phi (0, \cdot) \rangle = \int_0^t \left( \mu_s, \left( \frac{\partial \phi}{\partial s} + \frac{1}{2} \sum_{i,j=1}^d a_{ij} \partial_i \partial_j \phi + b \cdot \nabla \phi \right) (s, \cdot) \right) ds
\]

for every \( \phi \in C_b^{1,2} ([0, T] \times \mathbb{R}^d) \).

**Theorem 8 (existence)** The law \( \mu_t \) of \( X_t \) is a measure-valued solution of the the Fokker-Planck equation (4).

**Proof.** Let \( \phi \) be of class \( C_b^{1,2} ([0, T] \times \mathbb{R}^d) \). By Itô formula for \( \phi (t, X_t) \), we have

\[
d\phi (t, X_t) = \frac{\partial \phi}{\partial t} (t, X_t) dt + \nabla \phi (t, X_t) \cdot dX_t + \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j \phi (t, X_t) a_{ij} (t, X_t) dt
\]

\[
= \frac{\partial \phi}{\partial t} (t, X_t) dt + \nabla \phi (t, X_t) \cdot b (t, X_t) dt + \nabla \phi (t, X_t) \cdot \sigma (t, X_t) dB_t + \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j \phi (t, X_t) a_{ij} (t, X_t) dt.
\]

We have \( E \int_0^T |\nabla \phi (t, X_t) \cdot \sigma (t, X_t)|^2 dt < \infty \) (we use here that \( \sigma \) and \( \nabla \phi \) are bounded), hence \( E \int_0^T \nabla \phi (s, X_s) \cdot \sigma (s, X_s) dW_s = 0 \) and thus (all terms are finite by the boundedness assumptions)

\[
E [\phi (t, X_t)] - E [\phi (0, X_0)] = E \int_0^t \frac{\partial \phi}{\partial s} (s, X_s) ds + E \int_0^t \nabla \phi (s, X_s) \cdot b (s, X_s) ds
\]

\[
+ \frac{1}{2} \sum_{i,j=1}^d E \int_0^t \partial_i \partial_j \phi (s, X_s) a_{ij} (s, X_s) ds.
\]

Since \( E [\phi (t, X_t)] = \int_{\mathbb{R}^d} \phi (t, x) \mu_t (dx) \) (and similarly for the other terms) we get the weak formulation of equation (4). The preliminary property that \( t \mapsto \langle \mu_t, \varphi (t, \cdot) \rangle = E [\varphi (t, X_t)] \) is measurable for all \( \varphi \in C_b ([0, T] \times \mathbb{R}^d) \) is easy. \( \blacksquare \)

**Remark 9** Under suitable assumptions, like the simple case when \( a_{ij} \) is the identity matrix, if \( \mu_0 \) has a density \( p_0 \) then also \( \mu_t \) has a density \( p(t, \cdot) \), often with some regularity gained by the parabolic structures, and thus the Fokker-Planck equation in the differential form (3) holds.
Theorem 10 (uniqueness) Assume that the backward parabolic equation (called backward Kolmogorov equation)

$$\frac{\partial u}{\partial t} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij} \partial_i \partial_j u + b \cdot \nabla u = 0 \quad \text{on} \quad [0, T_0] \times \mathbb{R}^d$$

$$u|_{t=T_0} = \psi$$

has, for every $T_0 \in [0, T]$ and $\psi \in C^\infty_c(\mathbb{R}^d)$, at least one solution $u$ of class $C^{1,2}_b([0, T_0] \times \mathbb{R}^d)$. Then the Fokker-Planck equation (4) has at most one measure-valued solution.

Proof. If $\mu_t$ is a measure-valued solution of the Fokker-Planck equation and $u$ is a $C^{1,2}_b([0, T_0] \times \mathbb{R}^d)$ solution of the Kolmogorov equation, then (from the identity which defines measure-valued solutions and the identity of Kolmogorov equation)

$$\langle \mu_{T_0}, \psi \rangle - \langle \mu_0, \psi(0, \cdot) \rangle.$$  

Then, if $(i)$, $i = 1, 2$, are two measure-valued solutions of the Fokker-Planck equation with the same initial condition $\mu_0$, we have

$$\langle \mu_{T_0}^{(1)}, \psi \rangle = \langle \mu_{T_0}^{(2)}, \psi \rangle.$$  

This identity holds for every $\psi \in C^\infty_c(\mathbb{R}^d)$, hence $\mu_{T_0}^{(1)} = \mu_{T_0}^{(2)}$. The time $T_0 \in [0, T]$ is arbitrary, hence $\mu_{t_0}^{(1)} = \mu_{t_0}^{(2)}$. $\blacksquare$

Obviously the weak aspect of the previous uniqueness result is the assumption, not explicit in terms of the coefficients. The reason is that there are two main cases when such implicit assumption is satisfied. One is the case when $b$ and $\sigma$ are very regular; the other is when $\sigma$ is non-degenerate. As an example of the second case, let us mention the fundamental case of heat equation.

Example 11 Consider the case $b = 0$, $\sigma = \text{Id}$. The forward Kolmogorov equation

$$\frac{\partial v}{\partial t} = \frac{1}{2} \sum_{i=1}^{d} \partial_i^2 v \quad \text{on} \quad [0, T_0] \times \mathbb{R}^d$$

$$v|_{t=0} = \psi$$

has the explicit solution

$$v(t, x) = \int p_t(x - y) \psi(y) \, dy$$

which is infinitely differentiable with all bounded derivatives. The function $u(t, x) = v(T_0 - t, x)$ is then an explicit and regular solution of the backward Kolmogorov equation above. In this case, therefore, the Fokker-Planck equation has a unique measure-valued solution.
2.2 Backward Kolmogorov equation

Along with the stochastic equation (1) defined by the coefficients $b$ and $\sigma$, we consider also the following backward parabolic PDE on $[0,T] \times \mathbb{R}^d$, called backward Kolmogorov equation:

$$
\frac{\partial u}{\partial t} + \frac{1}{2} \sum_{i,j=1}^d a_{ij} \partial_i \partial_j u + b \cdot \nabla u = 0, \quad u|_{t=T} = \psi.
$$

To express in full generality the relation with the SDE we have to introduce the SDE on the time interval $[t_0, T]$, with any $t_0 \in [0, T]$:

$$
X_t = x + \int_{t_0}^t b(s, X_s) \, ds + \int_{t_0}^t \sigma(s, X_s) \, dB_s, \quad t \in [t_0, T].
$$

Obviously, on $[t_0, T]$, we have the same results as on $[0, T]$, in particular strong existence and pathwise uniqueness under Lipschitz assumptions. Assume these conditions and denote the unique solution, defined on some filtered probability space, by $X_{t_0}^{t_0,x}$. The relation with the backward Kolmogorov equation is

$$
u(t, x) = E \left[ \psi \left( X_T^{t_0,x} \right) \right] .$$

This relation holds under different assumptions and for solutions $u$ with different degrees of regularity. Let us start with the most elementary result.

**Proposition 12** If $u$ is a solution of the backward Kolmogorov equation of class $C^{1,2} \left( [0, T] \times \mathbb{R}^d \right)$, with bounded $\nabla u$ and $\sigma$, then $u(t, x) = E \left[ \psi \left( X_T^{t_0,x} \right) \right]$.

**Proof.** Given $t_0 \in [0, T]$, we apply Itô formula to $u \left( t, X_t^{t_0,x} \right)$ on $[t_0, T]$. The computation is the same done in the proof of Theorem 8. Since $u$ solves the backward Kolmogorov equation, we get

$$
\psi \left( T, X_T^{t_0,x} \right) = u(t_0, x) + \int_{t_0}^T \nabla u \left( t, X_t^{t_0,x} \right) \cdot \sigma \left( t, X_t^{t_0,x} \right) \, dB_t
$$

where we have used the identity $X_T^{t_0,x} = x$. Using the boundedness assumption, $\nabla u \left( t, X_t^{t_0,x} \right)$ is of class $M^2$, hence

$$
E \left[ \int_{t_0}^T \nabla u \left( t, X_t^{t_0,x} \right) \cdot \sigma \left( t, X_t^{t_0,x} \right) \, dB_t \right] = 0.
$$

The relation $u(t_0, x) = E \left[ \psi \left( X_T^{t_0,x} \right) \right]$ follows.  ■
2.3 Macroscopic limit

We may reformulate Theorem 8 as a macroscopic limit of a system of non-interacting particles.

Let \( W^n_t, n \in \mathbb{N}, \) be a sequence of independent Brownian motions in \( \mathbb{R}^k, \) defined on a probability space \( (\Omega, \mathcal{F}, P) . \) Let \( b, \sigma \) as above. Consider the sequence of SDEs in \( \mathbb{R}^d \)

\[
dX^n_t = b(t, X^n_t) \, dt + \sigma(t, X^n_t) \, dW^n_t
\]

with \( X^n_0 \) given independent \( \mathbb{R}^d \)-r.v.'s, \( \mathcal{F}_0 \)-measurable, with the same law \( \mu_0. \) Then, since weak uniqueness holds for the SDE, the processes \( X^n_t \) have the same law; in particular, with the notations above, the marginal at time \( t \) of \( X^n_t \) is \( \mu_t. \) Moreover, the processes \( X^n_t \) are independent, since each \( X^n_t \) is adapted to the corresponding Brownian motion \( W^n_t, \) which are independent.

Consider, for each \( N \in \mathbb{N}, \) the random probability measure, called empirical measure,

\[
S^N_t := \frac{1}{N} \sum_{n=1}^{N} \delta_{X^n_t}
\]

namely, for \( \phi \in C^0_c(\mathbb{R}^d), \)

\[
\langle S^N_t, \phi \rangle = \frac{1}{N} \sum_{n=1}^{N} \phi(X^n_t).
\]

It is a sort of discrete density of particles. The following simple theorem is our first example of macroscopic limit of a system of microscopic particles.

**Theorem 13** For every \( t \in [0, T] \) and \( \phi \in C^0_c(\mathbb{R}^d), \) a.s.

\[
\lim_{N \to \infty} \langle S^N_t, \phi \rangle = \langle \mu_t, \phi \rangle.
\]

In other words, \( S^N_t \) converges weakly, a.s., to a measure-valued solution of the Fokker-Planck equation (4).

**Proof.** Since, at each time \( t, \) the r.v. \( \phi(X^n_t) \) are bounded i.i.d., by the strong law of large numbers we have, a.s.,

\[
\lim_{N \to \infty} \langle S^N_t, \phi \rangle = E \left[ \phi \left( X^1_t \right) \right] = \langle \mu_t, \phi \rangle.
\]

And by Theorem 8 above, \( \mu_t \) is a measure-valued solution of the the Fokker-Planck equation (4). The only technical detail that we could discuss more deeply is the precise meaning of the sentence "converges weakly, a.s." The simplest meaning is the a.s. convergence \( \lim_{N \to \infty} \langle S^N_t, \phi \rangle = \langle \mu_t, \phi \rangle, \) for every a priori given \( \phi \in C^0_c(\mathbb{R}^d). \) But this implies the
stronger concept that, chosen a priori $\omega \in \Omega$ a.s., we have $\lim_{N \to \infty} \langle S_t^N, \phi \rangle = \langle \mu_t, \phi \rangle$ for every $\phi \in C_c^0(\mathbb{R}^d)$ (namely the null set of $\omega$’s where the convergence could fail is independent of $\phi$). To reach this result it is sufficient to notice first that it holds for any given countable set $\{\phi_n\}$. Then, it is possible to choose such set $\{\phi_n\} \subset C_c^0(\mathbb{R}^d)$ in a way that it is dense in $C_c^0(\mathbb{R}^d)$, in the topology of uniform convergence on the full space $\mathbb{R}^d$. Therefore, taken $\phi \in C_c^0(\mathbb{R}^d)$, taken a subsequence $\{\phi_{n_k}\}$ which converges uniformly to $\phi$, one has

$$\left| \langle S_t^N, \phi \rangle - \langle \mu_t, \phi \rangle \right| \leq \left| \langle S_t^N, \phi \rangle - \langle S_t^N, \phi_{n_k} \rangle \right| + \left| \langle S_t^N, \phi_{n_k} \rangle - \langle \mu_t, \phi_{n_k} \rangle \right| + \left| \langle \mu_t, \phi_{n_k} \rangle - \langle \mu_t, \phi \rangle \right|$$

$$\leq 2\|\phi_{n_k} - \phi\|_0 + \left| \langle S_t^N, \phi_{n_k} \rangle - \langle \mu_t, \phi_{n_k} \rangle \right|.$$ 

Choose $\omega \in \Omega$ such that $\lim_{N \to \infty} \langle S_t^N(\omega), \phi_{n_k} \rangle = \langle \mu_t, \phi_{n_k} \rangle$ for all $k \in \mathbb{N}$. Given $\varepsilon > 0$, we first take $k$ such that $\|\phi_{n_k} - \phi\|_0 < \frac{\varepsilon}{2}$; then we take $N_0$ such that $\left| \langle S_t^N(\omega), \phi_{n_k} \rangle - \langle \mu_t, \phi_{n_k} \rangle \right| < \frac{\varepsilon}{2}$ for all $N > N_0$. We get $\left| \langle S_t^N(\omega), \phi \rangle - \langle \mu_t, \phi \rangle \right| < \varepsilon$ for all $N > N_0$. Hence $\lim_{N \to \infty} \langle S_t^N, \phi \rangle = \langle \mu_t, \phi \rangle$ on the same set of $\omega$’s where the convergence was true for all $\phi_{n_k}, k \in \mathbb{N}$. The proof is complete. $\blacksquare$