20. Stochastic differential equations

We will now move to the central subject of this course, which is the theory of *stochastic differential equations*, to be abbreviated as SDE.

20.1 Setup for strong solutions.

The SDEs are generally equations for a semimartingale X of the form

$$\mathrm{d}X_t = U_t\mathrm{d}t + Y_t\mathrm{d}B_t \tag{20.1}$$

whose driving processes U and Y are, at each time $t \geqslant 0$, assumed to be given by prescribed functions of t and X_t; namely,

$$U_t = a(t, X_t) \ \wedge \ Y_t = \sigma(t, X_t) \tag{20.2}$$

This is a *choice* made for fundamental reasons (good dynamical theories such as Newton's mechanics describe motions using local rules) as well as practical ones (anything more complicated will hardly be generally useful).

An example of such a problem is the equation

$$\mathrm{d}N_t = \alpha N_t\mathrm{d}t + \beta N_t\mathrm{d}B_t \tag{20.3}$$

that defines the dynamics of a *population growth* model. We already encounter the example of the d-dimensional Bessel process,

$$\mathrm{d}X_t = \frac{d-1}{2X_t}\mathrm{d}t + \mathrm{d}B_t. \tag{20.4}$$

Yet another example is the *Langevin equation* which takes the form

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \mathrm{d}B_t \tag{20.5}$$

This describes motion of a particle in a potential field V that, without noise, simply follows the steepest descent of V. (The Langevin equation in physics arose as a model of the velocity of actual physical Brownian motion. In this case V is just quadratic representing friction forces.)

We may at times consider a generalization of (20.1–20.2) to a family of SDEs for a vector valued process $X = (X^{(1)}, \dots, X^{(d)})$ depending on m-dimensional standard Brownian motion $B = (B^{(1)}, \dots, B^{(m)})$ as

$$\mathrm{d}X_t^{(i)} = a_i(t, X_t)\mathrm{d}t + \sum_{j=1}^{m} \sigma_{ij}(t, X_t)\mathrm{d}B_t^{(j)} \tag{20.6}$$

To make this notationally close to the one-dimensional case, we adopt a vector notation and write this as

$$\mathrm{d}X_t = a(t, X_t)\mathrm{d}t + \sigma(t, X_t) \cdot \mathrm{d}B_t \tag{20.7}$$

where "$\cdot$" denotes the Euclidean dot product, a is assumed $\mathbb{R}^d$-valued and σ is assumed $d \times m$-matrix valued.

In order to treat the above concepts mathematically, we now define a term that describes the context in which we will work with (20.7).

Definition 20.1 (Standard setting for SDE) *A* standard setting *for (strong form of) SDE* (20.7) *for $\mathbb{R}^d$-valued process X consists of a four-tuple of objects*

$$\Big((\Omega, \mathcal{F}, P), \{\mathcal{F}_t\}_{t\geqslant 0}, \{B_t\colon t \geqslant 0\}, X_0\Big) \tag{20.8}$$

where $(\Omega, \mathcal{F}, P)$ is a probability space,

(1) *$B = \{B_t\colon t \geqslant 0\}$ is an m-dimensional standard Brownian motion on $(\Omega, \mathcal{F}, P)$,*
(2) *$\{\mathcal{F}_t\}_{t\geqslant 0}$ is a Brownian filtration on $(\Omega, \mathcal{F}, P)$ such that*

$$\mathcal{F}_0 \text{ contains all } P\text{-null sets} \tag{20.9}$$

(3) *X_0 is an $\mathbb{R}^d$-valued random variable on $(\Omega, \mathcal{F}, P)$, to be called* initial value*, which is assumed $\mathcal{F}_0$-measurable (and thus independent of B),*

and a pair of functions

$$a\colon [0,\infty)\times\mathbb{R}^d \to \mathbb{R}^d \ \wedge\ \sigma\colon [0,\infty)\times\mathbb{R}^m \to \mathbb{R}^{dm} \tag{20.10}$$

which are assumed to be Borel measurable.

With this in hand, we can define what it means to solve (20.7).

Definition 20.2 (Strong solution) *Assume a standard setting for SDE* (20.7) *with notations as in Definition 20.1. A* strong solution *to* (20.7) *is then a stochastic process $\{X_t\colon t \geqslant 0\}$ defined on the underlying probability space $(\Omega, \mathcal{F}, P)$ such that*

(1) *X is continuous and adapted to $\{\mathcal{F}_t\}_{t\geqslant 0}$,*
(2) *for all $t \geqslant 0$,*

$$\int_0^t |a(s, X_s)|\mathrm{d}s < \infty \ \wedge\ \int_0^t |\sigma(s, X_s)|^2\mathrm{d}s < \infty \quad \textit{a.s.} \tag{20.11}$$

(3) *for all $t \geqslant 0$,*

$$X_t = X_0 + \int_0^t a(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\cdot \mathrm{d}B_s \quad \textit{a.s.} \tag{20.12}$$

where X_0 on the right denotes the initial value.

Note that (1) along with Definition 20.1 ensure that $t \mapsto a(t, X_t)$ and $t \mapsto \sigma(t, X_t)$ are adapted and jointly measurable processes. The conditions (20.11) in (2), where we lighten the notation by writing $|\cdot|$ for Euclidean norms if X is vector valued, then guarantee that the integrals in (20.12) are meaningful and finite a.s. We write a.s. in (20.12) to reflect on the stochastic integral being defined only up to a null set. Thanks to (20.9) the integrals admit continuous adapted versions for which the equality in (20.12) then holds for all $t \geqslant 0$ simultaneously on a set of full measure.

20.2 Finding a strong solution.

There is no good reason for calling the solution "strong" at this point except that we will later consider the concept of a weak solution — where, reader beware, "weak" means something rather different from what it designates in standard ODE/PDE theory. To elucidate the difference nonetheless, let us "solve" the equations (20.3) and (20.4).

Starting with the population growth equation (20.3), assuming that N is a solution with $N_0 > 0$, the Itô formula gives

$$\begin{aligned} \mathrm{d}\log N_t &= \frac{1}{N_t}\mathrm{d}N_t - \frac{1}{2N_t^2}\mathrm{d}\langle N\rangle_t \\ &= \frac{1}{N_t}\big(\alpha N_t \mathrm{d}t + \beta N_t \mathrm{d}B_t\big) - \frac{1}{2N_t^2}\beta^2\mathrm{d}t \\ &= \big(\alpha - \tfrac{1}{2}\beta^2\big)\mathrm{d}t + \beta \mathrm{d}B_t \end{aligned} \tag{20.13}$$

The right-hand side is the differential of the process $t \mapsto (\alpha - \frac{1}{2}\beta^2)t + \beta B_t$. Exponentiation thus gives

$$N_t = N_0\,\mathrm{e}^{(\alpha-\beta^2/2)t+\beta B_t} \tag{20.14}$$

While the above steps were not completely justified (for instance, we never worried about N_t hitting zero, which is a singularity of the logarithm) we can now forget all of it and check, using the Itô formula, that (20.14) is indeed a strong solution to (20.3).

Concerning the Bessel SDE (20.4), here we recall that, for d natural, the d-dimensional Bessel process was discovered as the radial process of d-dimensional standard Brownian motion. This suggests that we "solve" (20.4) by putting

$$X_t := |X_0 + B_t| \tag{20.15}$$

for $|\cdot|$ denoting the Euclidean norm in $\mathbb{R}^d$. However, a closer look shows that this is not what we aimed for. Indeed, the Itô formula yields

$$\mathrm{d}X_t = \frac{d-1}{2X_t} + \mathrm{d}\widetilde{B}_t \tag{20.16}$$

where

$$\widetilde{B}_t := \sum_{i=1}^{d}\int_0^t \frac{1}{X_s}B_s^{(i)}\mathrm{d}B_s^{(i)} \tag{20.17}$$

While, as noted earlier (thanks to Theorem 14.6), $\widetilde{B}$ is a standard Brownian motion it is *not* the one in (20.15) — after all, that Brownian motion is d-dimensional!

It thus appears that we did solve the Bessel SDE, but not using the prescribed Brownian motion. Instead, we allowed the structure of the "standard setting" for the SDE to become part of our solution and not something that is fixed beforehand. As it turns out, this is exactly what will mean that the solution (20.15) to SDE (20.4) is weak.

A natural question to answer now is under what general conditions on the standard setting and, in particular, the coefficients in the SDE, a strong solution exists. The following theorem, due to K. Itô in early 1940s, answers this as follows:

Theorem 20.3 *Assume the standard setting with initial data obeying $X_0 \in L^2$ and coefficients such that there is $K \in (0,\infty)$ for which*

$$\forall t \geqslant 0\,\forall x,y \in \mathbb{R}^d\colon \quad \big|a(t,x)-a(t,y)\big| + \big|\sigma(t,x)-\sigma(t,y)\big| \leqslant K|x-y| \tag{20.18}$$

and

$$\forall t \geqslant 0\colon \quad \big|a(t,0)\big| + \big|\sigma(t,0)\big| \leqslant K \tag{20.19}$$

hold true. Then a strong solution to SDE (20.7) *with initial data X_0 exists.*

We remark that the above applies equally well for vector valued processes; on thus needs to replace $|a(t,x)-a(t,y)|$ and $|a(t,0)|$ by the Euclidean norm of the corresponding dimension and $|\sigma(t,x)-\sigma(t,y)|$, resp., $|\sigma(t,0)|$ by the matrix Euclidean norm defined, for any $d\times m$-matrix Σ by $\|\Sigma\|_2 := [\mathrm{Tr}(\Sigma\Sigma^{\mathrm{T}})]^{1/2} = [\sum_{i,j}\Sigma_{ij}^2]^{1/2}$.

20.3 Proof of Theorem 20.3.

As usual in the theory of ordinary differential equations, we will construct a solution as the limit of so called *Picard-Lindelöf iterations*. These are processes $\{X^{(n)}\}_{n\geqslant 0}$ defined by setting, for all $t\geqslant 0$,

$$X_t^{(0)} := X_0 \tag{20.20}$$

and, for each $n\geqslant 0$,

$$X_t^{(n+1)} := X_0 + \int_0^t a(s, X_s^{(n)})\mathrm{d}s + \int_0^t \sigma(s, X_s^{(n)})\cdot \mathrm{d}B_s \tag{20.21}$$

In order to make sure that the integrals exist and the construction can proceed inductively for all $n\geqslant 0$, we need to prove:

Lemma 20.4 *Assume a standard setting for the SDE* (20.7)*. Then for any process $\{Y_t\colon t\geqslant 0\}$ on $(\Omega,\mathcal{F},P)$ which is continuous and adapted, the (vector valued) processes $\{a(t,Y_t)\colon t\geqslant 0\}$ and $\{\sigma(t,Y_t)\colon t\geqslant 0\}$ are jointly measurable and adapted. If, in addition,*

$$\forall t\geqslant 0\colon\quad \sup_{s\leqslant t} E\big(|Y_s|^2\big) < \infty \tag{20.22}$$

and (20.18–20.19) hold, then

$$\forall t\geqslant 0\colon\quad E\int_0^t \big|a(s,Y_s)\big|\mathrm{d}s < \infty \ \wedge\ E\int_0^t \big|\sigma(s,Y_s)\big|^2\mathrm{d}s < \infty \tag{20.23}$$

and, for a suitable version of the stochastic integral,

$$Z_t := \int_0^t a(s,Y_s)\mathrm{d}s + \int_0^t \sigma(s,Y_s)\cdot \mathrm{d}B_s \tag{20.24}$$

defines a continuous, adapted process that obeys

$$\forall t\geqslant 0\colon\quad \sup_{s\leqslant t} E\big(|Z_s|^2\big) < \infty \tag{20.25}$$

Here, throughout, we wrote $|\cdot|$ for Euclidean norms of appropriate dimension.

Proof. A continuous adapted process is automatically jointly measurable. The first part of the claim then follows by standard facts about composition of measurable functions. For the second part note that, under (20.18–20.19), the triangle inequality yields

$$\forall t\geqslant 0\,\forall x\in\mathbb{R}^d\colon\quad \big|\,a(t,x)\big| + \big|\sigma(t,x)\big| \leqslant K\big[1+|x|\big] \tag{20.26}$$

To get the bounds (20.23) we invoke (20.22) with the help of Fubini-Tonelli and, in the first part, Cauchy-Schwarz. Using that $(a+b)^2 \leqslant 2a^2+2b^2$, the latter also shows

$$E\left(\sup_{u\leqslant t}\left|\int_0^u a(s,Y_s)\mathrm{d}s\right|^2\right) \leqslant tE\left(\int_0^t \left|a(s,Y_s)\right|^2 \mathrm{d}s\right) \leqslant 2t^2K^2 \sup_{s\leqslant t}\left[1+E(|Y_s|^2)\right] \tag{20.27}$$

and, using Doob's L^2-martingale inequality for the stochastic integral on the right of (20.24) and then Itô isometry, we similarly get

$$\begin{aligned} E\left(\sup_{u\leqslant t}\left|\int_0^u \sigma(s,Y_s)\cdot \mathrm{d}B_s\right|^2\right) &\overset{\text{Doob}}{\leqslant} 4E\left(\left|\int_0^t \sigma(s,Y_s)\cdot \mathrm{d}B_s\right|^2\right) \\ &\overset{\text{Itô}}{=} 4E\int_0^t \left|\sigma(s,Y_s)\right|^2 \mathrm{d}s \leqslant 8tK^2 \sup_{s\leqslant t}\left[1+E(|Y_s|^2)\right] \end{aligned} \tag{20.28}$$

Using that $(a+b)^2 \leqslant 2a^2+2b^2$, (20.22) thus implies (20.25). □

Using Lemma 20.4 along with the construction of the iterates, we inductively check that, since $X^{(0)}$ satisfies the assumptions of continuity, adaptedness and (20.22), so does the process $X^{(n)}$ for all $n \geqslant 0$. Our next goal is to control the convergence of $X_t^{(n)}$ as $n \to \infty$. We will do this using (local) supremum norms, as our aim is to prove uniform convergence a.s. to ensure that the limit is continuous, and under second moments, as this is what the above is set up for. The key estimate comes in:

Lemma 20.5 *For each $t \geqslant 0$ and $n \geqslant 1$ set*

$$g_n(t) := E\Big(\sup_{s\leqslant t}\big|X_s^{(n)} - X_s^{(n-1)}\big|^2\Big) \tag{20.29}$$

Then g_n is continuous and, for K as in (20.22),

$$g_{n+1}(t) \leqslant 2K^2(t+4)\int_0^t g_n(s)\,\mathrm{d}s \tag{20.30}$$

holds for all all $t \geqslant 0$ and all $n \geqslant 0$ where (for $n := 0$) we set

$$g_0(t) := 4K^2(t+4)\big[1+E(X_0^2)\big] \tag{20.31}$$

Proof. The continuity of g_n (which will guarantee that the integral in (20.30) is meaningful) follows from the continuity and monotonicity of the supremum in t via the Monotone Convergence Theorem. For the main part of the claim, note that for $n \geqslant 1$,

$$X_t^{(n+1)} - X_t^{(n)} = A_t + M_t \tag{20.32}$$

where

$$A_t := \int_0^t \big[a(s,X_s^{(n)}) - a(s,X_s^{(n-1)})\big]\mathrm{d}s \tag{20.33}$$

and

$$M_t := \int_0^t \big[\sigma(s,X_s^{(n)}) - \sigma(s,X_s^{(n-1)})\big]\cdot \mathrm{d}B_s \tag{20.34}$$

Using the inequality $(a+b)^2 \leqslant 2a^2+2b^2$ we get

$$\sup_{s\leqslant t}\bigl|X_t^{(n)}-X_t^{(n-1)}\bigr|^2 \leqslant 2\sup_{s\leqslant t}A_s^2+2\sup_{s\leqslant t}M_s^2 \tag{20.35}$$

Similarly as in (20.27), for the first term on the right we employ the Cauchy-Schwarz inequality along with the bound (20.22) to obtain

$$\begin{aligned}E\bigl[\sup_{s\leqslant t}A_s^2\bigr] &\leqslant tE\int_0^t\bigl|a(s,X_s^{(n)})-a(s,X_s^{(n-1)})\bigr|^2\mathrm{d}s\\ &\leqslant K^2t\,E\int_0^t\bigl|X_s^{(n)}-X_s^{(n-1)}\bigr|^2\mathrm{d}s \leqslant K^2t\int_0^t g_n(s)\,\mathrm{d}s\end{aligned} \tag{20.36}$$

where Fubini-Tonelli's theorem and the definition of g_n were used in the last step. For second term on the right of (20.35), using that M is a continuous martingale, Doob's L^2-inequality and Itô isometry show

$$\begin{aligned}E\bigl[\sup_{s\leqslant t}M_s^2\bigr] \leqslant 4E\bigl[M_t^2\bigr] &= 4E\int_0^t\bigl|a(s,X_s^{(n)})-a(s,X_s^{(n-1)})\bigr|^2\mathrm{d}s\\ &\leqslant 4K^2E\int_0^t\bigl|X_s^{(n)}-X_s^{(n-1)}\bigr|^2\mathrm{d}s \leqslant 4K^2\int_0^t g_n(s)\,\mathrm{d}s\end{aligned} \tag{20.37}$$

Putting (20.36–20.37) together, the claim (20.30) for $n\geqslant 1$ follows via (20.35).

In order to prove the inequality (20.30) for $n:=0$, we note that in this case (20.32) still holds but with the argument of the integrals (20.33–20.34) using the convention $a(s,X^{(-1)}):=$ and $\sigma(s,x^{-1}):=0$. Instead of the Lipschitz bounds (20.18) we instead call upon (20.26) and the fact that $[1=|x|]^2\leqslant 2[1+|x|^2]$ to effectively produce $g_0(t)$ on the right-hand sides of (20.36–20.37). □

We are now ready to give:

Proof of Theorem 20.3. Denoting $C(t):=2K^2(t+4)$ and observing that $t\mapsto C(t)$ is non-decreasing, iterations of (20.30) show

$$\forall s\leqslant t\colon\quad g_n(s)\leqslant 2\frac{C(t)^n s^n}{n!}\bigl[1+E(X_0^2)\bigr] \tag{20.38}$$

Chebyshev's inequality now shows

$$P\Bigl(\sup_{s\leqslant t}\bigl|X_s^{(n)}-X_s^{(n-1)}\bigr|>2^{-n}\Bigr)\leqslant 4^n g_n(t)\leqslant\frac{[4C(t)t]^n}{n!}\bigl[1+E(X_0^2)\bigr] \tag{20.39}$$

and the fact that the right-hand side is summable for each integer $t\geqslant 1$ combined with Borel-Cantelli lemma gives

$$P\Biggl(\forall t\geqslant 0\colon\sum_{n\geqslant 1}\sup_{s\leqslant t}\bigl|X_s^{(n)}-X_s^{(n-1)}\bigr|<\infty\Biggr)=1 \tag{20.40}$$

It follows that the infinite series in

$$X_t:=X_0+\sum_{n\geqslant 1}\bigl[X_s^{(n)}-X_s^{(n-1)}\bigr] \tag{20.41}$$

converges locally uniformly a.s. and so X is continuous a.s.

Redefining X_t to equal X_0 on the complement of the event in (20.40) produces a continuous process which is adapted thanks to (20.9) and equals the locally-uniform pointwise limit of $X_t^{(n)}$ as $n \to \infty$. In addition, we have

$$\forall t \geqslant 0\colon \quad E\Big(\sup_{s\leqslant t} \big|X_s^{(n)} - X_s\big|^2\Big) \underset{n\to\infty}{\longrightarrow} 0, \tag{20.42}$$

which follows by noting that, with the help of Fatou's lemma and above locally uniform pointwise convergence, the square root of the expectation is dominated by $\sum_{m>n} g_m(t)^{1/2}$ and that $\{g_m(t)^{1/2}\}_{m\geqslant 1}$ is still summable.

It remains to check that X is indeed a desired strong solution to (20.7). We have already checked that X is adapted and continuous. Using (20.27) — and the fact that each $X^{(n)}$ obeys the following bound — we get

$$\forall t \geqslant 0\colon \quad \sup_{s\leqslant t} E(|X_s|^2) < \infty \tag{20.43}$$

and so, by Lemma 20.4, the stochastic integrals in (20.11) are in L^2. In order to show that X satisfies the integral form (20.12) of the SDE, we note one more time that estimates of the kind (20.36–20.37) show

$$\int_0^t a(s, X_s^{(n)})\mathrm{d}s \ \underset{n\to\infty}{\overset{L^2}{\longrightarrow}} \ \int_0^t a(s, X_s)\mathrm{d}s \tag{20.44}$$

and

$$\int_0^t \sigma(s, X_s^{(n)})\mathrm{d}B_s \ \underset{n\to\infty}{\overset{L^2}{\longrightarrow}} \ \int_0^t \sigma(s, X_s)\mathrm{d}B_s \tag{20.45}$$

This now implies (20.12) by taking the L^2-limit on the right-hand side of (20.21) and pointwise limit on the left-hand side. □

We have constructed a strong solution to a robust class of SDEs. Uniqueness of this solution as well as other ramifications will be addressed in the next lecture.

Further reading: Karatzas-Shreve, Sections 5.1-5.2