Rend. Istit. Mat. Univ. Trieste
Vol. XXXII, 217–269 (2000)

An Analytical Introduction to Stochastic Differential Equations: Part I — The Langevin Equation

Ph. Clément and O.W. van Gaans (*)

Summary. - *We present an introduction to the theory of stochastic differential equations, motivating and explaining ideas from the point of view of analysis. First the notion of white noise is developed, introducing at the same time probabilistic tools. Then the one dimensional Langevin equation is formulated as a deterministic integral equation with a parameter. Its solution leads to stochastic convolution, which is defined as a Riemann-Stieltjes integral. It is shown that the parameter dependence yields a Gaussian system, of which the means and covariances are computed. We conclude by introducing briefly the notion of invariant measure and the associated Kolmogorov equations.*

Contents

1 Introduction **218**
1.1 Langevin Equation I 219
1.2 How to model noise? 219

2 Probability theory **222**
2.1 Random variables 222
2.2 Stieltjes integrals 225

(*) Authors' addresses: Ph. Clément, Faculty ITS/AMA, Delft University of Technology, Mekelweg 4, NL 2628 CD Delft, The Netherlands
O.W. van Gaans, Mathematical Institute, Leiden University, Niels Bohrweg 1, NL 2333 CA Leiden, The Netherlands

2.3 Independence 226
2.4 Product spaces 227
2.5 Sequences of independent random variables 230
2.6 Functionals of a sequence of random variables 232
2.7 Expectation and variance 233

3 Convergence of random vectors 234
3.1 Langevin Equation II 234
3.2 Convergence of random variables 235
3.3 Langevin Equation III 238
3.4 Convergence of random vectors 240
3.5 Langevin Equation IV 244

4 Gaussian systems 245
4.1 Multidimensional central limit theorem 248
4.2 Langevin Equation V 248

5 Solution of the Langevin Equation 252
5.1 Langevin Equation VI 252
5.2 Volterra integral equations 254
5.3 Langevin equation VII 256

6 Stochastic convolution 257
6.1 Langevin Equation VIII 261

7 Stochastic initial conditions and invariant measure 262

8 Kolmogorov equations 264

1. Introduction

The goal of this text is to give an introduction from an analytical point of view to the theory of differential equations *perturbed with noise.* As a first step we shall consider the *Langevin equation*, which is an equation of the form

$$\begin{cases} \frac{dX}{dt}(t) = f(X(t)) + \text{“noise”}, \quad t > 0, \\ X(0) = x, \end{cases} \tag{1}$$

where $f\colon \mathbb{R} \to \mathbb{R}$ is an affine function and $x \in \mathbb{R}$. The “noise” term will be a Gaussian process. The study of such equations inevitably needs concepts and results from probability theory. We start with the notion of noise and some tools to work with it. Then we will discuss existence, uniqueness, and some properties of solutions to the Langevin equation.

We aim at readers who are not familiar with probability theory, but who do have a background in analysis. Our focus is on explaining ideas and motivating the mathematical concepts to describe them. Accordingly, we give detailed statements but leave many of the proofs to the reader or refer to the literature.

1.1. Langevin Equation I

Let $c, \theta \in \mathbb{R}$, and let $f(u) = c(\theta - u)$ $(u \in \mathbb{R})$. Then problem (1) becomes:

$$\begin{cases} \frac{dX}{dt}(t) = c(\theta - X(t)) + \text{“noise”}, & t > 0, \\ X(0) = x. \end{cases} \tag{2}$$

We write $\bar{X}$ for the solution to the problem without noise:

$$\begin{cases} \frac{d\bar{X}}{dt}(t) = c(\theta - \bar{X}(t)), & t > 0, \\ \bar{X}(0) = x. \end{cases} \tag{3}$$

It can directly be verified that

$$\bar{X}(t) = e^{-ct}x + (1 - e^{-ct})\theta, \qquad t \geq 0, \tag{4}$$

is a solution to (3). We see that for $c > 0$, $\bar{X}(t)$ is a convex combination of x and θ, for $t = 0$ starting at x and transforming to θ as $t \to \infty$.

1.2. How to model noise?

Intuitively, the ”noise” in problem (2) is a random influence on the system, as if at every moment a coin is tossed to decide in which way the influence will be. A proper mathematical description of noise is rather involved. It brings us to the Brownian Motion and stochastic processes in general. Before we introduce the specialities of probability theory, we elaborate a little more at an intuitive level.

As there is no obvious direct way, a natural approach to introducing noise in problem (2) is to consider discretizations. Indeed, even by definition, the derivative term is a limit of difference quotients.

Divide the interval $[0,T]$ in n subintervals of length $h = T/n$. Consider the step function $\bar{X}^{(n)}$ as an approximation of $\bar{X}$ of (4), given by the *Euler implicit scheme:*

$$\begin{cases} \frac{\bar{X}^{(n)}(t+h)-\bar{X}^{(n)}(t)}{h} = c(\theta - \bar{X}^{(n)}(t+h)), & t \geq h, \\ \bar{X}^{(n)}(t) = x, & 0 \leq t < h, \end{cases} \tag{5}$$

for $n = 1, 2, \ldots$. From these expressions we can recollect the solution of (3) by rewriting and passing to the limit. We find

$$(1+ch)\bar{X}^{(n)}(t+h) = \bar{X}^{(n)}(t) + ch\theta, \qquad t \geq h,$$

so that for all $t \in [kh, (k+1)h)$:

$$\bar{X}^{(n)}(t) = (1+ch)^{-k}x + ch\theta \sum_{i=1}^{k} (1+ch)^{-i},$$

which leads to

$$\begin{cases} \bar{X}^{(n)}(t) = (1+ch)^{-k}x + \theta \frac{ch}{1+ch} \frac{1-(1+ch)^{-k}}{1-(1+ch)^{-1}}, & t \in [kh, (k+1)h), \\ \bar{X}^{(n)}(t) = x, & t \in [0, h). \end{cases}$$

In particular, for $t = T$:

$$\bar{X}^{(n)}(T) = \left(1 + c\frac{T}{n}\right)^{-n} x \ + \ \theta \frac{c\frac{T}{n}}{1+c\frac{T}{n}} \frac{1-(1+c\frac{T}{n})^{-n}}{1-(1+c\frac{T}{n})^{-1}}. \tag{6}$$

Letting $n \to \infty$, we find $\bar{X}^{(n)}(T) \to e^{-cT}x + (1-e^{-cT})\theta$, which is the solution to (3).

Our approach now is to add a noise term to (5), to rewrite in the same way as above, and to see what happens if n tends to infinity. Consider

$$\begin{cases} \frac{X^{(n)}(t+h)-X^{(n)}(t)}{h} = c(\theta - X^{(n)}(t+h)) + \alpha^{(n)}\xi^{(n)}(t), & t \geq h, \\ X^{(n)}(t) = x, & 0 \leq t < h, \end{cases} \tag{7}$$

where $\alpha^{(n)}$ are real numbers to be chosen later on for suitable scaling and where $\xi^{(n)}(t)$ represents the randomness. In view of the discretization, we take the functions $\xi^{(n)}$ constant on each interval $[kh,(k+1)h)$ and equal to $+1$ or -1 with equal probability:

$$\xi^{(n)}(t) = \begin{cases} 0, & t \in [0,h), \\ \eta_k, & t \in [kh,(k+1)h), \end{cases}$$

where the *random variables* η_k are, independently, $+1$ or -1 both with probability $1/2$.

Rewriting (7) yields:

$$(1+ch)X^{(n)}(t+h) = X^{(n)}(t) + ch\theta + h\alpha^{(n)}\xi^{(n)}(t), \qquad t \geq h,$$

so

$$X^{(n)}(t) = (1+ch)^{-k}x + ch\theta\sum_{i=1}^{k}(1+ch)^{-i} + \sum_{i=1}^{k}(1+ch)^{-i}h\alpha^{(n)}\eta_{(k+1)-i}$$

and thus

$$X^{(n)}(T) = \bar{X}^{(n)}(T) + \sum_{k=1}^{n}(1+c\frac{T}{n})^{-k}\frac{T}{n}\alpha^{(n)}\eta_{(n+1)-k}, \tag{8}$$

with $\bar{X}^{(n)}$ given by (6). Our objective is to give a meaning to problem (2) as a limit of (7), starting from (8) and its limit for $n \to \infty$. As we have seen, this approach will involve:

- the notion of random variable
- the notion of independence of random variables
- existence of a sequence $\{\eta_k\}_{k=1}^{\infty}$ of independent random variables such that each η_k takes values 1 or -1 both with probability $1/2$
- interpretation of $\lim_{n\to\infty}\sum_{k=1}^{n}\beta_k^{(n)}\eta_k$ for certain sequences of numbers $\{\beta_k^{(n)}\}_{k=1}^{n}$, $n = 1, 2, \ldots$.

We will now present the ideas and tools from probability theory that address the above issues.

2. Probability theory

We assume that the reader is familiar with elementary measure and integration theory. For reference we mention: [1], [5], [11], and [7]. Recall that a *σ-algebra* (*σ-field*) in a set S is a collection $\mathcal{F}$ of subsets of S such that (i) $\emptyset \in \mathcal{F}$, (ii) $A^c \in \mathcal{F}$ for all $A \in \mathcal{F}$, and (iii) $\cup_{k=1}^{\infty} A_k \in \mathcal{F}$ for all $A_1, A_2, \ldots \in \mathcal{F}$. If S is a topological space, for instance a metric space, then the *Borel σ-algebra* is the σ-algebra generated by the open sets, i.e. the smallest σ-algebra containing all open sets of S. The members of the Borel σ-algebra are called the *Borel sets* of S. The Borel σ-algebra of $\mathbb{R}$ is denoted by $\mathcal{B}$.

A *measure* on a σ-algebra $\mathcal{F}$ is a mapping $\mu\colon \mathcal{F} \to [0,\infty]$ that is 0 at $\emptyset$ and *σ-additive*: $\mu(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$ for every pairwise disjoint collection $\{A_1, A_2, \ldots\}$ in $\mathcal{F}$. If $\mu(S) = 1$, then μ is called a *(Borel) probability measure* on $\mathcal{F}$, or, less precisely, on S.

2.1. Random variables

Definition 2.1. *A* probability space *consists of a triple* $(\Omega, \mathcal{F}, \mathbb{P})$ *where*

(i) *Ω is a non empty set of points ω, called the* sample space *and* sample points,

(ii) *$\mathcal{F}$ is a σ-algebra of subsets of Ω; these subsets are called* events,

(iii) *$\mathbb{P}(.)$ is a* probability measure *or briefly* probability *on $\mathcal{F}$.*

Example 2.2. *Let $\Omega = \{a, b\}$, let $\mathcal{F}$ be the σ-algebra of all subsets of Ω and let $\mathbb{P}$ be defined by*

$$\begin{aligned} \mathbb{P}(\emptyset) &= 0, \\ \mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) &= 1/2, \\ \mathbb{P}(\{a,b\}) &= 1. \end{aligned}$$

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Definition 2.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X\colon \Omega \to \mathbb{R}$ is called a* random variable *if X is* measurable*, that means: for every Borel set B in $\mathbb{R}$ the set*

$$\{\omega \in \Omega\colon\ X(\omega) \in B\} \in \mathcal{F}.$$

We will denote

$$\{X \in B\} := \{\omega \in \Omega\colon\ X(\omega) \in B\}$$

and, accordingly,

$$\{X \le x\} := \{\omega \in \Omega\colon\ X(\omega) \le x\}, \qquad x \in \mathbb{R}.$$

Every random variable induces a σ-algebra in Ω and a measure on the Borel sets of $\mathbb{R}$:

DEFINITION 2.4. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $X\colon\ (\Omega, \mathcal{F}) \to \mathbb{R}$ *be a random variable. Then:*

(i) The collection of subsets of Ω *of the form* $\{X \in B\}$, $B \in \mathcal{B}$, *is a* σ*-algebra, denoted by* $\mathcal{F}(X)$ *or* $\sigma(X)$*, called the* σ*-algebra* generated by X*. Note that* $\mathcal{F}(X)$ *is a subcollection of* $\mathcal{F}$*, since* X *is Borel measurable.*

(ii) The image measure of $\mathbb{P}$ *under* X*, given by*

$$(X \circ \mathbb{P})(B) := \mathbb{P}(\{X \in B\}), \qquad B \in \mathcal{B},$$

is called the law *of the random variable* X *or the* probability distribution *of* X*.*

(iii) The function $F\colon\ \mathbb{R} \to [0,1]$ *defined by*

$$F(x) := \mathbb{P}(\{X \le x\})$$

is called the distribution function *of* X*.*

EXAMPLE 2.5. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be as in Example 2.2. Let* $\xi\colon\ \Omega \to \mathbb{R}$ *be defined by*

$$\begin{aligned} \xi(a) &= -1, \\ \xi(b) &= +1. \end{aligned}$$

Then ξ *is a random variable and its distribution function is given in figure 1.*

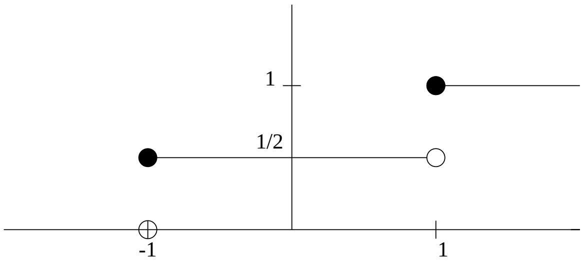

Figure 1: The distribution function of ξ.

Proposition 2.6. *A function $F\colon \mathbb{R} \to [0,1]$ is the distribution function of a random variable if and only if the following holds:*

(i) F is nondecreasing,

(ii) $F(-\infty) = 0$, $F(\infty) = 1$,

(iii) F is right continuous or, equivalently,

$$F(x) = \inf_{y>x} F(y), \ \textit{for all } x \in \mathbb{R}.$$

(See [1, Prop. 2.25 and Thm 2.26, p. 27–28].)

We shall denote by $\mathcal{N}$ the set of all distribution functions on $\mathbb{R}$. For $F \in \mathcal{N}$, its set of discontinuities is at most countable and hence the set of continuity $C(F)$ is dense in $\mathbb{R}$ ([11, Thm 9-1.I, p. 380]). A function $F \in \mathcal{N}$ is uniquely determined by its restriction to $C(F)$. More generally, if $F, G \in \mathcal{N}$ and $F(x) = G(x)$ for every x in a dense subset A of $\mathbb{R}$, then $F = G$ (compare with [1, Problem 8.2, p. 160]). We remark that given a function $F \in \mathcal{N}$ there exists a unique Borel probability measure μ_F on $\mathbb{R}$ such that

$$\mu_F((-\infty, x]) = F(x), \qquad x \in \mathbb{R}.$$

(See [7, Prop. 12.12, p. 301], [8, Thm 1, p. 150].) Let us denote by $\mathcal{P}(\mathbb{R})$ the set of Borel probability measures on $\mathbb{R}$.

2.2. Stieltjes integrals

Calculations with random variables can usually be done in terms of probability distributions or distribution functions. The former usually involve Lebesgue integrals, the latter Stieltjes integrals. Let us summarize some of the main properties of Stieltjes integrals.

Let $a, b \in \mathbb{R}$, $a < b$ and let $f, g\colon [a, b] \to \mathbb{R}$ (or $\mathbb{C}$) be arbitrary functions. For a partition $a = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n = b$ and intermediate points s_k with $t_{k-1} \leq s_k \leq t_k$, $k = 1, \ldots, n$, the corresponding *Riemann-Stieltjes sum* is defined as

$$\sum_{k=1}^{n} f(s_k)(g(t_k) - g(t_{k-1})).$$

If there exists a number I such that for every $\varepsilon > 0$ there is a $\delta > 0$ such that for any partition $a = t_0 \leq t_1 \leq \cdots \leq t_n = b$ with $\max_k(t_k - t_{k-1}) < \delta$ and any choice of intermediate points the corresponding Riemann-Stieltjes sum differs less than ε from I, then f is called *Stieltjes integrable* with respect to g and the number I, which is denoted by $\int_a^b f(t)dg(t)$, is called the *Stieltjes integral* of f with respect to g.

Let $BV[a, b]$ denote the space of all real (or complex) valued functions on $[a, b]$ with bounded variation. Recall that $BV[a, b]$ is the vector space generated by the nondecreasing functions, and that $C^1[a, b] \subset BV[a, b]$.

Proposition 2.7. *Let $f, g\colon [a, b] \to \mathbb{R}$ (or $\mathbb{C}$).*

(i) If f is Stieltjes integrable with respect to g, then g is Stieltjes integrable with respect to f and

$$\int_a^b f(t)dg(t) + \int_a^b g(t)df(t) = f(b)g(b) - f(a)g(a).$$

(Integration by parts)

(ii) If $f \in C[a, b]$ and $g \in BV[a, b]$, then f is Stieltjes integrable with respect to g and

$$|\int_a^b f(t)dg(t)| \leq \|f\|_\infty V_{[a,b]}(g),$$

where $V_{[a,b]}(g)$ denotes the total variation of g.

(iii) If $f \in C[a,b]$ *and* $g \in C^1[a,b]$*, then*

$$\int_a^b f(t)dg(t) = \int_a^b f(t)g'(t)dt.$$

(See [11, Thm 9-5.I,II, p. 394].)

Exercise 2.8: (A substitution rule) Let $T > 0$ and let $f, g\colon [0,T] \to \mathbb{R}$ be such that f is Stieltjes integrable with respect to g. Let $\tilde{f} := f(T-t)$ and $\tilde{g}(t) := g(T-t)$, $0 \le t \le T$. Show that $\tilde{f}$ is Stieltjes integrable with respect to $\tilde{g}$ and that

$$\int_0^T \tilde{f} d\tilde{g}(t) = -\int_0^T f(t)dg(t).$$

2.3. Independence

The idea of independence of random variables is not easily catched in terms of measure theory. Let us therefore say some words about the stochastic intuition behind it.

The probability of an event can be interpreted as the relative frequency of occurences in a (infinitely) large number of samples. Thinking of two events, A and B, and a large number of samples, we can consider the probabilistic relationship between A and B by comparing the relative frequency of occurences of B among all the samples and among only those where A occurs. One extreme case we can imagine is that B occurs then and only then when A occurs. An other extreme case is that B occurs with the same frequency among the samples where A occurs as among all samples. Whether A occurs or not seems then of no importance to B. In the latter case, A and B are called *independent*. Since the fraction of the samples where both A and B occur equals the fraction of occurence of A times the fraction of occurence of B among these where A occurs, we see that for independent A and B we have: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Thus we arrive at the common definition of independency.

Definition 2.9. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space.* σ*-algebras* $\mathcal{F}_1, \ldots, \mathcal{F}_n$ *contained in* $\mathcal{F}$ *are said to be* independent *if for any choice of sets* $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$ *one has that*

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \Pi_{k=1}^n \mathbb{P}(A_k).$$

Random variables $X_1, \dots, X_n$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *are said to be* independent *if the* σ*-algebras* $\mathcal{F}(X_1), \dots, \mathcal{F}(X_n)$ *are independent.*

Clearly, $X_1, \dots, X_n$ *are independent if and only if*

$$\mathbb{P}(\{X_1 \in B_1, \dots, X_n \in B_n\}) = \Pi_{k=1}^n \mathbb{P}(\{X_k \in B_k\})$$

for any sets $B_1, \dots, B_n \in \mathcal{B}$.

A sequence $X_1, X_2, \dots$ *of random variables is called independent if for every* $n \geq 2$, *the random variables* $X_1, \dots, X_n$ *are independent.*

PROPOSITION 2.10. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $\{X_n\}_{n=1}^\infty$ *be a sequence of random variables. A necessary and sufficient condition for these random variables to be* independent *is that for every* n, *and* n*-tuple* $(x_1, \dots, x_n) \in \mathbb{R}^n$ *one has that*

$$\mathbb{P}(\{X_1 \leq x_1, \dots, X_n \leq x_n\}) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \tag{9}$$

where F_{X_i} *denotes the distribution function of* X_i, $i = 1, 2, \dots, n$. (See [1, Thm 3.7, p. 38]).

COROLLARY 2.11. *If* $\{X_n\}_{n=1}^\infty$ *are independent random variables, then*

$$\mathbb{P}(\{X_1 \leq x_1, X_2 \leq x_2, \dots\}) = \Pi_{k=1}^\infty F_{X_k}(x_k), \tag{10}$$

where $x_1, x_2, \dots$ *is an arbitrary sequence in* $\mathbb{R}$.

2.4. Product spaces

Given a sequence of distribution functions $\{F_n\}_{n=1}^\infty$ in $\mathcal{N}$ one may ask whether there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $\{X_n\}_{n=1}^\infty$ such that (10) holds. In this respect the notion of countable product of probability spaces will be useful.

For probability spaces $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k)$, $k = 1, 2, \dots$, we will construct a product space as follows. Take $\Omega := \Pi_{k=1}^\infty \Omega_k = \Omega_1 \times \Omega_2 \times \cdots$, the cartesian product. A subset of Ω is called a *(rectangular) cylinder set* if it is of the form $A \times \Omega_{k+1} \times \Omega_{k+2} \times \cdots$, where $A = A_1 \times \cdots \times A_k$ with $A_1 \in \mathcal{F}_1 \dots, A_k \in \mathcal{F}_k$ and $k \geq 1$. Let $\mathcal{F}$ be the σ-algebra in Ω generated by the cylinder sets. A probability measure $\mathbb{P}$ on $\mathcal{F}$ is called a *product measure* if $\mathbb{P}(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \cdots) =$

$\mathbb{P}_1(A_1)\cdots\ \mathbb{P}_n(A_n)$ for every $n \geq 1$ and $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$. In that case, $(\Omega, \mathcal{F}, \mathbb{P})$ is called the *product* of the probability spaces $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k)$, $k = 1, 2, \ldots$. Does such a measure $\mathbb{P}$ always exist and, if so, is it unique?

THEOREM 2.12. *Let $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k)$, $k = 1, 2, \ldots$ be probability spaces, let $\Omega = \Pi_{k=1}^{\infty}\Omega_k$, and let $\mathcal{F}$ be the σ-algebra in Ω generated by the cylinder sets. Then there exists a unique probability measure $\mathbb{P}$, denoted by $\otimes_{k=1}^{\infty}\mathbb{P}_k$, on $(\Omega, \mathcal{F})$ such that*

$$\mathbb{P}(A_1 \times A_2 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots) = \Pi_{k=1}^{n}\mathbb{P}_k(A_k), \quad (11)$$

for every $n \geq 1$ and every sets $A_1 \in \mathcal{F}_1, \cdots, A_n \in \mathcal{F}_n$.
(See [5, §38 Thm B, p.157]; if $\Omega_k = \mathbb{R}$ for all k then special case of Kolmogorov's extension theorem, [1, Cor. 2.19, p. 24], [8, Thm 3, p. 161].)

EXAMPLE 2.13. *(sequel to Example 2.5) Take for every $n \geq 1$:*

$$\Omega_n = \{a, b\},$$
$$\mathcal{F}_n = \{\emptyset, \{a\}, \{b\}, \{a, b\}\},$$
$$\mathbb{P}_n \text{ such that } \mathbb{P}_n(\{a\}) = \mathbb{P}_n(\{b\}) = 1/2.$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the product of the spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, $n = 1, 2, \ldots$. For $\omega = (\omega_1, \omega_2, \ldots) \in \Omega$, set for $k \geq 1$

$$\xi_k(\omega) := \begin{cases} -1 & \text{if } \omega_k = a \\ 1 & \text{if } \omega_k = b. \end{cases}$$

Then $\xi_1, \xi_2, \ldots$ are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. They are independent and identically distributed, and

$$F_{\xi_k}(x) = \begin{cases} 0 & \text{if } x < -1, \\ 1/2 & \text{if } -1 \leq x < 1, \\ 1 & \text{if } x \geq 1. \end{cases}$$

Moreover, we have $F_{-\xi_k}(x) = F_{\xi_k}(x)$, $x \in \mathbb{R}$.

The above example may seem rather simple. However, it has strong consequences. In the next Exercise it is concluded that it entails existence of the Lebesgue measure on $[0, 1)$.

EXERCISE 2.14: Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{\xi_k\}_{k=1}^{\infty}$ be as in the previous example. Set

$$\Omega' := \{\omega \in \Omega \colon \ \omega \text{ contains an infinite number of } a\text{'s}\}.$$

1. Show that $\Omega' \in \mathcal{F}$ and $\mathbb{P}(\Omega') = 1$.

2. Show that the map j defined by $j(\omega) := \sum_{k=1}^{\infty} 2^{-1}(\xi_k(\omega) + 1)2^{-k}$ for $\omega \in \Omega'$ is a bijection from Ω' onto $[0,1)$.

3. Let $\mathcal{F}' := \{A' \subset \Omega' \colon \ A' = \Omega' \cap A \text{ for an } A \in \mathcal{F}\}$. Show that $\mathcal{F}'$ is a σ-algebra in Ω'.

4. Let $A_1, A_2 \in \mathcal{F}$ be such that $\Omega' \cap A_1 = \Omega' \cap A_2$. Show that $\mathbb{P}(A_1) = \mathbb{P}(A_2)$. Define $\mathbb{P}' : \mathcal{F}' \to [0,1]$ by

$$\mathbb{P}'(A') := \mathbb{P}(A)$$

   where $A \in \mathcal{F}$ such that $A' = \Omega' \cap A$ and show that $(\Omega', \mathcal{F}', \mathbb{P}')$ is a probability space.

5. Let B denote the Borel σ-algebra of $[0,1)$. Show that $j^{-1}([0,x)) \in \mathcal{F}'$ for every $x \in [0,1]$ and that $j^{-1}(B) \in \mathcal{F}'$ for every $B \in \mathcal{B}$.

6. Show that $\mathbb{P}'(j^{-1}([0,x))) = x$ for every $x \in [0,1]$.

7. Show that there exists at most one Borel measure μ on $[0,1)$ such that $\mu([0,x)) = x$ for every $x \in [0,1]$.

8. Show that $\mathbb{P}'(j^{-1}(B)) = m(B)$ for every $B \in \mathcal{B}$ where m is the Lebesgue measure on $[0,1)$.

EXAMPLE 2.15. The space $(\mathbb{R}^{\mathbb{N}}, B(\mathbb{R}^{\mathbb{N}}), \otimes_{k=1}^{\infty}\mu_k)$. *Let* $\{F_k\}_{k=1}^{\infty}$ *be a sequence of distribution functions in* $\mathcal{N}$. *Let* $\{\mu_k\}_{k=1}^{\infty}$ *be the corresponding Borel probability measures on* $(\mathbb{R}, B(\mathbb{R}))$. *Applying Theorem 2.12 with*

$$\begin{aligned} \Omega_k &= \mathbb{R}, \\ \mathcal{F}_k &= B(\mathbb{R}), \quad k \geq 1, \\ \mathbb{P}_k &= \mu_k, \end{aligned}$$

we obtain a probability space $(\mathbb{R}^{\mathbb{N}}, \mathcal{F}, \otimes_{k=1}^{\infty}\mu_k)$, *satisfying (11). The* σ*-algebra* $\mathcal{F}$, *which is the* σ*-algebra generated by the cylinder sets in*

$\mathbb{R}^{\mathbb{N}}$, turns out to be the Borel σ-algebra for the following metric on $\mathbb{R}^{\mathbb{N}}$:

$$\rho(\omega,\eta) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{|\omega_k - \eta_k|}{1 + |\omega_k - \eta_k|}, \quad \omega, \eta \in \mathbb{R}^{\mathbb{N}}.$$

Equipped with this metric, $\mathbb{R}^{\mathbb{N}}$ is complete and separable, in other words: a Polish space. *A sequence $\{\omega^{(n)}\}_{n=1}^{\infty}$ converges to ω with respect to ρ if and only if $\lim_{n\to\infty} \omega_k^{(n)} = \omega_k$ for every $k \geq 1$, which is pointwise convergence of $\{\omega^{(n)}\}_{n=1}^{\infty}$ to ω. It can be shown that $\mathcal{F}$ is precisely the σ-algebra of Borel sets of $(\mathbb{R}^{\mathbb{N}}, \rho)$, denoted by $B(\mathbb{R}^{\mathbb{N}})$ (see Exercise 2.16).*

Exercise 2.16: Show that $\mathcal{F}$ of the previous example equals the Borel σ-algebra on $(\mathbb{R}^{\mathbb{N}}, \rho)$, i.e. the σ-algebra generated by the open (or, equivalently, closed) sets of $(\mathbb{R}^{\mathbb{N}}, \rho)$. Show that $\mathcal{F}$ is also the σ-algebra generated by cylinders in $\mathbb{R}^{\mathbb{N}}$, i.e. the sets of the form $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \times \mathbb{R} \times \mathbb{R} \times \cdots$ where $n \geq 1$ and $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$ with $a_1 < b_1, \ldots, a_n < b_n$.

Observe that if $X_k(\omega) = \omega_k$, $\omega \in \mathbb{R}^{\mathbb{N}}$, $k = 1, 2, \ldots$, then the functions $X_k \colon \Omega \to \mathbb{R}$, $k \geq 1$, are continuous, hence Borel measurable, independent as random variables, and the corresponding distribution functions are F_k, $k = 1, 2, \ldots$.

2.5. Sequences of independent random variables

Similar to the definition of the law of a random variable, we will associate with a sequence of random variables a measure on $\mathbb{R}^{\mathbb{N}}$ if the sequence is infinite and on $\mathbb{R}^n$ if the sequence is finite and of length n.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Analogously to the definition of random variable (Definition 2.3), we will call a measurable map from Ω to $\mathbb{R}^n$ a *random vector* and a measurable map from Ω to $\mathbb{R}^{\mathbb{N}}$ a *random sequence*, where $\mathbb{R}^{\mathbb{N}}$ is equipped with the Borel-σ-algebra induced by the metric of Example 2.15.

If $X_1, \ldots, X_n$ are random variables on Ω, we can associate with the vector of random variables $(X_1, \ldots, X_n)$ the vector valued function X (i.e. a map from Ω to $\mathbb{R}^n$) given by $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$, $\omega \in \Omega$. If $A_1, \ldots, A_n \in B(\mathbb{R})$, then $\{A_1 \times$

$\cdots \times A_n\} \in \mathcal{F}$. Moreover the class of all $A \in B(\mathbb{R}^n)$ such that $\{X \in A\} \in \mathcal{F}$ is a σ-algebra in $B(\mathbb{R}^n)$. Since it contains the n-dimensional rectangles, it contains $B(\mathbb{R}^n)$. Therefore we have $\{X \in A\} \in \mathcal{F}$, for every $A \in B(\mathbb{R}^n)$. This means that X is measurable, so that it is a random vector.

Similarly, a sequence of random variables is a random sequence. Indeed, if $\{X_k\}_{k=1}^{\infty}$ is a sequence of random variables on Ω, then the map $X\colon \Omega \to \mathbb{R}^n$ given by $X(\omega) = (X_1(\omega), X_2(\omega), \dots)$, $\omega \in \Omega$, satisfies $\{X \in A\} \in \mathcal{F}$ for every $A \in B(\mathbb{R}^{\mathbb{N}})$, so that it is measurable.

Thus we are in a position to define the following.

Definition 2.17. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.*

(i) *Let $X = (X_1, \dots, X_n)$ be a random vector. The (image) measure $(X \circ \mathbb{P})(A) := \mathbb{P}(\{X \in A\})$, $A \in B(\mathbb{R}^n)$, on $(\mathbb{R}^n, B(\mathbb{R}^n))$ is called the* law *or* distribution *of X.*

(ii) *Let $\{X_n\}_{n \geq 1}$ be a sequence of random variables. The (image) measure $(X \circ \mathbb{P})(A) := \mathbb{P}(\{X \in A\})$, $A \in B(\mathbb{R}^{\mathbb{N}})$, on $(\mathbb{R}^{\mathbb{N}}, B(\mathbb{R}^{\mathbb{N}}))$ is called the* law *or* distribution *of the sequence $\{X_k\}_{k \geq 1}$.*

(iii) *Two sequences (possibly of the same finite length) $\{X_n\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\{X'_n\}$ on $(\Omega', \mathcal{F}', \mathbb{P}')$ are said to have* the same law *(or distribution) if*

$$(X \circ \mathbb{P})(A) = (X' \circ \mathbb{P}')(A), \quad \textit{for all } A \in B(\mathbb{R}^{\mathbb{N}}) \ (A \in B(\mathbb{R}^n)).$$

(iv) *The functions*

$$F_k(x_1, \dots, x_k) := \mathbb{P}(\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k\}), \tag{12}$$

for $k = 1, \dots, n$ if $X = (X_1, \dots, X_n)$ is a random vector and for $k = 1, 2, \dots$ if $\{X_k\}$ is a sequence of random variables, are called the k-dimensional (joint) distribution functions *of X or the sequence $\{X_k\}$, respectively.*

In the sequel we will sometimes use the notations of infinite sequences for finite sequences as well. They should then be interpreted in the above way.

The law of the sequence $\{X_n\}$ contains all the information which is relevant to probability theory. For instance, random variables $X_1, X_2, \ldots$ are independent if and only if the law of the sequence is a product measure. The sequence is i.i.d. (independent, identically distributed) if and only if the law is given by $\otimes_{k=1}^{\infty}\mu_k$, with $\mu_k = \mu_1$ for all $k \geq 1$. It follows from the uniqueness part of Theorem 2.12 that if the random variables are independent, then the law of the sequence is completely determined by its finite dimensional joint distribution functions.

2.6. Functionals of a sequence of random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{X_k\}_{k\geq 1}$ be a sequence of random variables and let $\phi\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ be Borel measurable, i.e. $\phi^{-1}(A) \in B(\mathbb{R}^{\mathbb{N}})$ for every $A \in B(\mathbb{R})$. Then $\phi \circ (X_1, X_2, \ldots)\colon \Omega \to \mathbb{R}$ is a random variable. If ϕ is nonnegative, then $\int_\Omega \phi(X) d\mathbb{P}$ is well-defined (possibly ∞). We can rewrite this integral by means of the law of the sequence:

PROPOSITION 2.18. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{X_n\}_{n\geq 1}$ a sequence of random variables, and $\phi\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ a Borel measurable nonnegative function. Let $\hat{\mathbb{P}}$ denote the law of the sequence. Then*

$$\int_\Omega \phi(X) d\mathbb{P} = \int_{\mathbb{R}^{\mathbb{N}}} \phi(x_1, x_2, \ldots) \hat{\mathbb{P}}(dx_1, dx_2, \ldots).$$

Consequently, if the random variables are independent:

$$\int_\Omega \phi(X) d\mathbb{P} = \int_{\mathbb{R}^{\mathbb{N}}} \phi(x_1, x_2, \ldots) \otimes_{k=1}^{\infty} \mu_k(dx_k),$$

where μ_k is the probability distribution of X_k, $k \geq 1$.
(See [1, Cor. 2.4.1, p. 32]).

COROLLARY 2.19. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $d \in \mathbb{N}$, and $X_1, \ldots, X_d$ be random variables with corresponding probability distributions $\mu_1, \ldots, \mu_d$. Let $\hat{\mathbb{P}}$ be the law of $(X_1, \ldots, X_n)$.*

1. *If $\phi\colon \mathbb{R}^d \to \mathbb{R}_+$ is Borel measurable, then*

$$\int_\Omega \phi(X_1, \ldots, X_d) d\mathbb{P} = \int_{\mathbb{R}^d} \phi(x_1, \ldots, x_d) \hat{\mathbb{P}}(dx_1, \ldots, dx_d).$$

2. If $X_1, \dots, X_d$ are independent and $\phi_1, \dots, \phi_d \colon \mathbb{R} \to \mathbb{R}_+$ are Borel measurable, then

$$\begin{aligned}
&\int_\Omega \phi_1(X_1) \cdots \phi_d(X_d) d\mathbb{P} \\
&\qquad = \int_{\mathbb{R}^2} \phi_1(x_1) \cdots \phi_d(x_d) \mu_1(dx_1) \otimes \cdots \otimes \mu_d(dx_d) \\
&\qquad = \left(\int_{\mathbb{R}} \phi_1(x_1) \mu_1(dx_1) \right) \cdots \left(\int_{\mathbb{R}} \phi_d(x_d) \mu_d(dx_d) \right) \\
&\qquad = \left(\int_{\mathbb{R}} \phi_1(x_1) dF_1(x_1) \right) \cdots \left(\int_{\mathbb{R}} \phi_d(x_d) dF_d(x_d) \right),
\end{aligned}$$

where F_k is the distribution function of X_k, $k = 1, \dots, d$.

3. If $X_1 \dots, X_d$ are independent, then

$$\int_\Omega e^{i(u_1 X_1 + \cdots + u_d X_d)} d\mathbb{P} = \Pi_{k=1}^d \int_{\mathbb{R}} e^{i u_k x_k} \mu_k(dx_k),$$

for every $u_1, \dots, u_d \in \mathbb{R}$.

4. If X_1 and X_2 are independent and F_1 and F_2 are the corresponding distribution functions, then for every $t \in \mathbb{R}$:

$$\begin{aligned}
F_{X_1+X_2}(t) &= \int_\Omega \mathbb{1}_{\{X_1+X_2 \le t\}} d\mathbb{P} \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{x_1+x_2 \le t\}} dF_1(x_1) dF_2(x_2) \\
&= \int_{\mathbb{R}} \int_{-\infty}^{t-x_2} dF_1(x_1) dF_2(x_2) = \int_{\mathbb{R}} F_1(t-x) dF_2(x).
\end{aligned}$$

2.7. Expectation and variance

Let X be a random variable such that X is nonnegative or $\int_\Omega |X| d\mathbb{P} < \infty$. Then the *expectation* of X is given by $\mathbb{E}(X) = \int_\Omega X d\mathbb{P}$. In view of Corollary 2.19, we have

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \mu_X(dx),$$

where μ_X is the law of X. If F is the distribution of X, then

$$\mathbb{E}(X) = \int_{\mathbb{R}} x dF(x) = \lim_{M,N\to\infty} \int_{-N}^{M} x dF(x),$$

where $\int_{-N}^{M} x dF(x)$ is a Stieltjes integral.

If $\mathbb{E}(X)$ is the expectation of X and $\mathbb{E}(|X|) < \infty$, then $\operatorname{var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2)$ is called the *variance* of X, also denoted by $\sigma^2(X)$. We have

$$\operatorname{var}(X) = \int_{\Omega} X^2 d\mathbb{P} - (\mathbb{E}(X))^2.$$

If $X_1, \dots, X_n$ are random variables, then the *covariance matrix* of $X_1, \dots, X_n$, denoted by $\operatorname{cov}(X_1, \dots, X_n)$ is defined by

$$\operatorname{cov}(X_1, \dots, X_n)_{ij} = \mathbb{E}\Big((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\Big), \quad i, j = 1, \dots, n,$$

provided that $\mathbb{E}(X_i^2) < \infty$, $i = 1, \dots, n$. If the random variables $\{X_k\}_{k=1}^n$ are independent, then the covariance matrix is diagonal with diagonal elements $d_i = \operatorname{var}(X_i)$, $i = 1, \dots, n$.

In Example 2.13, $\mathbb{E}(\xi_k) = 0$, $k = 1, 2, \dots$, and $\operatorname{cov}(\xi_1, \dots, \xi_n) = I$, the unit matrix.

3. Convergence of random vectors

3.1. Langevin Equation II

We now come back to our discretized Langevin equation (equation (7)) with $c = 0$ and $x = 0$:

$$\begin{cases} \frac{X^{(n)}(t+h) - X^{(n)}(t)}{h} = \alpha^{(n)} \xi^{(n)}(t), & t \geq h, \\ X^{(n)}(t) = 0, & 0 \leq t < h. \end{cases} \tag{13}$$

We have presented the notions to describe the random behaviour of $X^{(n)}(t)$, in particular of $X^{(n)}(T)$. We have

$$X^{(n)}(T) = \tfrac{T}{n} \alpha^{(n)} \sum_{k=1}^{n} \xi_k, \tag{14}$$

where the sequence $\{\xi_k\}_{k\geq 1}$ is the sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ of Example 2.13. In order to obtain a meaning of the non-discretized problem, we want to let the stepsize tend to zero, i.e. $n \to \infty$, and consider "the limit of $X^{(n)}(T)$". In view of the idea that all relevant information about the random behaviour is contained in the distribution, our primary interest is the determination of the distribution of $\{X^{(n)}(T)\}_{n\geq 1}$ and of its limitary behaviour as $n \to \infty$. We introduce some definitions and theorems which are useful for this purpose. After that, we are going to vary T and we will consider joint distributions of $\{X^{(n)}(t)\}_{t\geq 0}$ and their limiting behaviour.

3.2. Convergence of random variables

If X_n, X, $n \geq 1$, are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, a natural candidate for the definition of convergence of the sequence $\{X_n\}_{n\geq 1}$ to X "in law" or "in distribution" is that $\lim_{n\to\infty} \mathbb{P}(\{X_n \leq x\}) = \mathbb{P}(\{X \leq x\})$ for every $x \in \mathbb{R}$, or, equivalently, $F_{X_n}(x) \to F_X(x)$, for all $x \in \mathbb{R}$. It appears, however, that the requirement "for all $x \in \mathbb{R}$" is too restrictive. Indeed, if $X_n(\omega) = 1/n$ for all $\omega \in \Omega$ and all $n \geq 1$, then $F_{X_n}(x) \to F_X(x)$ for every $x \in \mathbb{R}$ *except* at 0. Observe that 0 is a point of discontinuity of F_X. If we denote by $\mathcal{N}$ the set of all distribution functions (see Definition 2.4 and Proposition 2.6) and for $F \in \mathcal{N}$ by $C(F)$ the set of continuity of F (which is dense in $\mathbb{R}$), we have:

DEFINITION 3.1. *A sequence of distribution functions $\{F_n\}_{n\geq 1}$ in $\mathcal{N}$ is said to* converge *to $F \in \mathcal{N}$ (notation $F_n \xrightarrow{\text{D}} F$ or $F_n \Rightarrow F$) if, as $n \to \infty$,*

$$F_n(x) \to F(x) \text{ for each } x \in C(F). \tag{15}$$

A sequence $\{X_n\}_{n\geq 1}$ of random variables is said to converge in law *(*in distribution*) to a random variable X (also denoted by $X_n \xrightarrow{\text{D}} X$ or $X_n \Rightarrow X$) if $F_n \xrightarrow{\text{D}} F$, where F_n and F are the distribution functions of X_n and X, respectively.*

REMARK 3.2. *(i) If $F_n \xrightarrow[\overline{n}]{} {}^{\text{D}} F$ and $F_n \xrightarrow[\overline{n}]{} {}^{\text{D}} G$, $F_n, F, G \in \mathcal{N}$, then $F = G$.*

(ii) If F is continuous, *i.e.* $C(F) = \mathbb{R}$ *and (15) holds, then the sequence $\{F_n\}$ converges to F not only pointwise, but even* uniformly *on* $\mathbb{R}$.

We now give equivalent forms of the convergence defined by (15). In order to do so, we introduce the following classes of functions on $\mathbb{R}$:

$$\begin{aligned} BC(\mathbb{R}) &:= \{u\colon \mathbb{R} \to \mathbb{R} \text{ (or } \mathbb{C})\colon\ u \text{ is bounded and continuous}\}, \\ BUC(\mathbb{R}) &:= \{u \in BC(\mathbb{R})\colon\ u \text{ is uniformly continuous}\}. \end{aligned}$$

The space $BC(\mathbb{R})$ equipped with the pointwise addition and multiplication by a scalar becomes a vector space over $\mathbb{R}$ (or $\mathbb{C}$). Equipped with the supremum norm $\|u\| := \sup_{x\in\mathbb{R}} |u(x)|$, $u \in BC(\mathbb{R})$, the space $BC(\mathbb{R})$ is a Banach space and $BUC(\mathbb{R})$ is a closed linear subspace of it.

Given $f \in BC(\mathbb{R})$ and $F \in \mathcal{N}$, for every $a < b$ the function f is Stieltjes integrable on $[a, b]$ with respect to F and

$$\lim_{\substack{a \to -\infty \\ b \to \infty}} \int_a^b f(x)dF(x) = \int_{-\infty}^{\infty} f(x)dF(x) \text{ exists.}$$

On the other hand, if μ denotes the unique Borel probability meaure on $\mathbb{R}$ ($\mu \in P(\mathbb{R})$) such that $\mu((-\infty, x]) = F(x)$ for all $x \in \mathbb{R}$, we have that $f \in \mathcal{L}^1(\mathbb{R}, B(\mathbb{R}), \mu)$, since $\mu(\mathbb{R}) = 1$ and f is bounded and Borel measurable. It follows that $\int_{\mathbb{R}} f d\mu$ is well-defined as a Lebesgue integral and we have:

$$\int_{\mathbb{R}} f d\mu = \int_{-\infty}^{\infty} f(x)dF(x), \text{ for all } f \in BC(\mathbb{R}). \tag{16}$$

Thus, given $F \in \mathcal{N}$, we can define a map $f \mapsto \phi_F(f) \in \mathbb{R}$ from $BC(\mathbb{R})$ to $\mathbb{R}$ by setting $\phi_F(f) := \int_{-\infty}^{\infty} f(x)dF(x)$. Clearly, ϕ_F is a linear functional on $BC(\mathbb{R})$ and from $|\phi_F(f)| \le \|f\|$, $f \in BC(\mathbb{R})$, ϕ_F is a bounded linear functional. Denoting by $(BC(\mathbb{R}))'$ the dual space of $(BC(\mathbb{R}), \|.\|)$, $F \mapsto \phi_F$ is a map from $\mathcal{N}$ into $(BC(\mathbb{R}))'$. It appears that this map is injective, i.e. if $F, G \in \mathcal{N}$ and $\phi_F = \phi_G$, then $F = G$. In other words $\int_{\mathbb{R}} f(x)dF(x) = \int_{\mathbb{R}} f(x)dG(x)$, $F, G \in \mathcal{N}$ for every $f \in BC(\mathbb{R})$ implies $F = G$.

A stronger result holds.

Proposition 3.3. *Let $F, G \in \mathcal{N}$ (resp. $\mu, \nu \in P(\mathbb{R})$) be such that*

$$\int_{-\infty}^{\infty} e^{iux} dF(x) = \int_{-\infty}^{\infty} e^{iux} dG(x) \text{ for all } u \in \mathbb{R}$$

(resp. $\int_{\mathbb{R}} e^{iux}\mu(dx) = \int_{\mathbb{R}} e^{iux}\nu(dx)$ for all $u \in \mathbb{R}$), then $F = G$ (resp. $\mu = \nu$).
(See [1, Thm 8.24, p. 170].)

Observe that the functions $\{e^{iux}\}_{x\in\mathbb{R}}$ belong to $BUC(\mathbb{R})$.

The next proposition gives equivalent forms of condition (15).

Proposition 3.4. *Let $F, F_1, F_2, \ldots \in \mathcal{N}$. The following statements are equivalent:*

(i) $F_n \xrightarrow{D} F$

(ii) $\int_{-\infty}^{\infty} f(x)dF_n(x) \to \int_{-\infty}^{\infty} f(x)dF(x)$ *for all* $f \in BC(\mathbb{R})$

(iii) $\int_{-\infty}^{\infty} e^{iux}dF_n(x) \to \int_{-\infty}^{\infty} e^{iux}dF(x)$ *for all* $u \in \mathbb{R}$.

(See [1, Prop. 8.19, p. 167 and Cor. 8.30, p. 172].)

It follows from Propositions 3.3 and 3.4 that the function $\hat{F}\colon \mathbb{R} \to \mathbb{C}$ defined by

$$\hat{F}(u) := \int_{\mathbb{R}} e^{iux} dF(x), \quad u \in \mathbb{R}, F \in \mathcal{N}, \tag{17}$$

is useful for studying convergence in distribution.

Definition 3.5. *Given a distribution function $F \in \mathcal{N}$, its* characteristic function *$\hat{F}\colon \mathbb{R} \to \mathbb{C}$ is the function defined by (17).*

We postpone mentioning important properties and characterizations of characteristic functions to § 3.4, where we state them in the more general context of random vectors.

3.3. Langevin Equation III

We return to the problem of convergence in distribution of $\{X^{(n)}(T)\}_{n\geq 1}$ as $n \to \infty$. We recall that

$$X^{(n)}(T) = \beta^{(n)} \sum_{k=1}^{n} \xi_k$$

with $\beta^{(n)} = h\alpha^{(n)} = \frac{T}{n}\alpha^{(n)}$, and $\xi_1, \xi_2, \ldots$ as in Example 2.13. Let F_n denote the distribution funtion of $X^{(n)}(T)$, let F denote the distribution function of ξ_k and let μ denote the corresponding Borel probability measure on $\mathbb{R}$ ($\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, where δ_a denotes the Dirac measure at $a \in \mathbb{R}$). For $n \geq 1$, the characteristic function of F_n is:

$$\begin{aligned}
\hat{F}_n(u) &= \mathbb{E}(e^{iu\beta^{(n)}\sum_{k=1}^{n}\xi_k}) \\
&= \int_{\mathbb{R}^n} e^{iu\beta^{(n)}\sum_{k=1}^{n}x_k} \otimes_{l=1}^{n} \mu(dx_l) \\
&= \Pi_{k=1}^{n} \int_{\mathbb{R}} e^{iu\beta^{(n)}x_k} dF(x_k) \\
&= \Pi_{k=1}^{n} \tfrac{1}{2}(e^{iu\beta^{(n)}} + e^{-iu\beta^{(n)}}) \\
&= (\cos(\beta^{(n)}u))^n, \qquad u \in \mathbb{R}.
\end{aligned}$$

What can we say about $\hat{F}_n$ as $n \to \infty$? If we take the scaling factors $\alpha^{(n)}$ such that $\beta^{(n)} \to 0$, then we have for an arbitrary $u \in \mathbb{R}$ and large n that

$$(\cos(\beta^{(n)}u))^n = \exp[n \log(1 + (\cos(\beta^{(n)}u) - 1))]$$

and

$$\log(1 + (\cos(\beta^{(n)}u) - 1)) \sim -\tfrac{1}{2}(\beta^{(n)}u)^2, \text{ as } n \to \infty.$$

Hence

$$\begin{aligned}
(\cos(\beta^{(n)}u))^n &\to 1 && \text{if } (\beta^{(n)})^2 n \to 0, \\
(\cos(\beta^{(n)}u))^n &\to 0 && \text{if } (\beta^{(n)})^2 n \to \infty, \\
(\cos(\beta^{(n)}u))^n &\to e^{-\frac{1}{2}u^2\sigma^2} && \text{if } (\beta^{(n)})^2 n \to \sigma^2 > 0.
\end{aligned}$$

So if we want a *nondegenerate* ("nondeterministic") limit distribution, we need $\beta^{(n)} \sim \frac{\sigma}{\sqrt{n}}$, as $n \to \infty$, for some $\sigma > 0$. Then we have $X^{(n)}(T) \xrightarrow{\mathrm{D}} X$, where X is a random variable with characteristic function $e^{-\frac{1}{2}\sigma^2 u^2}$. That means that X is Gaussian with mean 0 and variance σ^2, see Definition 4.1. In terms of $\alpha^{(n)}$, the choice of $\beta^{(n)}$ reads $\alpha^{(n)} = \frac{n}{T}\beta^{(n)} \sim \frac{\sigma}{T}\sqrt{n}$ as $n \to \infty$.

EXERCISE 3.6: Compute the limiting distribution as $n \to \infty$ of $X^{(n)}(T)$ with $\alpha^{(n)} = \frac{\sigma}{T}\sqrt{n}$, $c > 0$, and $x \neq 0$ in (14).

REMARK 3.7. *In the above considerations we have used Proposition 3.4 in an essential way. We also used the fact that for $F \in \mathcal{N}$ and $m \in \mathbb{R}$, $\sigma \geq 0$ one has that*

$$\hat{F}(u) = e^{imu} e^{-\frac{1}{2}\sigma^2 u^2}, \text{ for all } u \in \mathbb{R},$$

if and only if

$$F(x) = \begin{cases} \textit{case } \sigma = 0: & 0 \quad \textit{for } x < m, \\ & 1 \quad \textit{for } x \geq m; \\ \textit{case } \sigma > 0: & \\ \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{(y-m)^2}{2\sigma^2}}dy, & x \in \mathbb{R}. \end{cases} \tag{18}$$

From now on we assume that $\alpha^{(n)} = \frac{\sigma}{T}\sqrt{n}$, where $\sigma > 0$. Let us investigate the asymptotic distributions of $X^{(n)}(t)$, $t > 0$, as $n \to \infty$. Recall that we have (with $h = T/n$):

$$\begin{cases} X^{(n)}(t) = 0 & \text{if } t \in [0, h), \\ X^{(n)}(t) = (\frac{T}{n})\alpha^{(n)} \sum_{k=1}^{[\frac{nt}{T}]} \xi_k, & \text{if } t \geq h, \end{cases} \tag{19}$$

where $[r]$ denotes the greatest integer $\leq r$.

Let $t > 0$. For any $n_0 \geq T/t$ we have for every $n \geq n_0$ that

$$\begin{aligned} X^{(n)}(t) &= \sigma\frac{1}{\sqrt{n}}\sum_{k=1}^{[\frac{nt}{T}]}\xi_k \\ &= \frac{\sigma}{\sqrt{T}}\sqrt{t}\frac{1}{\sqrt{\frac{nt}{T}}}\sum_{k=1}^{[\frac{nt}{T}]}\xi_k \\ &= \frac{\sigma}{\sqrt{T}}\sqrt{t}\sqrt{\frac{[\frac{nt}{T}]}{\frac{nt}{T}}}\frac{1}{\sqrt{[\frac{nt}{T}]}}\sum_{k=1}^{[\frac{nt}{T}]}\xi_k. \end{aligned}$$

As above, we find $X^{(n)}(t) \xrightarrow{\mathrm{D}} X_t$, where X_t is a Gaussian variable with mean 0 and variance $\frac{\sigma^2}{T}t$, i.e. a random variable with a distribution function satisfying (18) with mean $m = 0$ and σ replaced by $\frac{\sigma}{\sqrt{T}}\sqrt{t}$. Before studying the joint distributions of $\{X^{(n)}(t)\}_{t\geq 0}$ we mention:

THEOREM 3.8. *1.* **Central limit theorem for i.i.d. random variables**
Let $\xi_1, \xi_2, \ldots,$ be a sequence of i.i.d. nondegenerate random variables with $\mathbb{E}\xi_1^2 < \infty$ and $\mathbb{E}\xi_1 = 0$. Set $S_n := \xi_1 + \cdots + \xi_n$. Then as $n \to \infty$:

$$\mathbb{P}(\{\tfrac{1}{\sigma}\tfrac{1}{\sqrt{n}}S_n \leq x\}) \to \Phi(x), \text{ for all } x \in \mathbb{R} \tag{20}$$

where $\sigma = (\mathbb{E}\xi_1^2)^{1/2}$ and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt, \quad x \in \mathbb{R}.$$

(See [8, Thm 3, p. 324].)

2. **Law of the iterated logarithm**
Under the same assumptions:

$$\limsup_{n\to\infty} \tfrac{1}{\sqrt{\log(\log n)}} \tfrac{1}{\sigma\sqrt{n}} |S_n| = \sqrt{2} \ a.s. \tag{21}$$

(See [1, Thm 3.52, p. 64, Thm 13.25, p. 291], [8, Thm 1, p.372 and Rem. 1, p. 374].)

3.4. Convergence of random vectors

In what follows we shall study the limiting behaviour of the joint distributions of $\{X^{(n)}(t)\}_{t\geq 0}$. We are interested in the "convergence in distribution" of the random vectors $(X^{(n)}(t_1), \ldots, X^{(n)}(t_m))$, where $0 < t_1 < \cdots < t_m$. The law of the random vector $(X^{(n)}(t_1), \ldots, X^{(n)}(t_m))$ is the Borel probability measure μ_n on $(\mathbb{R}^m, B(\mathbb{R}^m))$ (see Definition 2.17(i)) satisfying

$$\mu_n(A) = \mathbb{P}(\{(X^{(n)}(t_1), \ldots, X^{(n)}(t_m)) \in A\}), \quad A \in B(\mathbb{R}^m).$$

We are looking for a measure μ on $(\mathbb{R}^m, B(\mathbb{R}^m))$ which is a limit of μ_n in some appropriate sense. As for $m = 1$, it is in general to much to ask that $\mu_n(A) \to \mu(A)$ for every $A \in B(\mathbb{R}^m)$. We will consider a notion of convergence that is similar to (15) and use characteristic functions again as a major tool.

Let us denote by $P(\mathbb{R}^m)$ the set of Borel probability measures on $(\mathbb{R}^m, B(\mathbb{R}^m))$. There is the following analogue of Proposition 3.3.

PROPOSITION 3.9. *Let $\mu, \nu \in P(\mathbb{R}^m)$ be such that*

$$\int_{\mathbb{R}^m} e^{i\sum_{k=1}^m u_k x_k} \mu(dx) = \int_{\mathbb{R}^m} e^{i\sum_{k=1}^m u_k x_k} \nu(dx)$$

for every $u_1, \dots, u_m \in \mathbb{R}$, then $\mu = \nu$.
(See [1, Thm 11.4, p. 235].)

Let $BC(\mathbb{R}^m)$ denote the space of bounded, continuous functions from $\mathbb{R}^m$ to $\mathbb{C}$.

PROPOSITION 3.10. *Let $\mu, \mu_1, \mu_2, \dots \in P(\mathbb{R}^m)$. The following statements are equivalent:*

(i) $\int_{\mathbb{R}^m} f(x)\mu_n(dx) \to \int_{\mathbb{R}^m} f(x)\mu(dx)$, *for all* $f \in BC(\mathbb{R}^m)$,

(ii) $\int_{\mathbb{R}^m} e^{i\sum_{k=1}^m u_k x_k}\mu_n(dx) \to \int_{\mathbb{R}^m} e^{i\sum_{k=1}^m u_k x_k}\mu(dx)$, *for all* $u_1, \dots, u_m \in \mathbb{R}$.

(See [9, Cor. 2.8, p. 25].)

A sequence of random vectors is said to converge in distribution if their distribution functions converge in the sense of Proposition 3.10(i). There is a more direct way to state this. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random vector X on Ω taking values in $\mathbb{R}^m$ with law μ_X, we have that

$$\mathbb{E}(f(X)) = \int_\Omega f(X) d\mathbb{P} = \int_{\mathbb{R}^m} f(x_1, \dots, x_n)\mu_X(dx),$$

$$\text{for all } f \in BC(\mathbb{R}^m).$$

So we can rewrite the convergence of Proposition 3.10(i) in terms of random vectors in the following way.

Definition 3.11. *Let $X, X^{(1)}, X^{(2)}, \dots$ be random vectors taking values in $\mathbb{R}^m$. The sequence $\{X^{(n)}\}_{n\geq 1}$* converges in law *(in distribution) to X, notation $X^{(n)} \Rightarrow X$ or $X^{(n)} \xrightarrow{\mathrm{D}} X$ if*

$$\mathbb{E}(f(X^{(n)})) \to \mathbb{E}(f(X)), \text{ for every } f \in BC(\mathbb{R}^m). \tag{22}$$

Remark 3.12. *Proposition 3.10 implies* Levy's Theorem*: $X^{(n)} \xrightarrow{\mathrm{D}} X$ if and only if $\mathbb{E}(e^{i\sum_{k=1}^m u_k X_k^{(n)}}) \to \mathbb{E}(e^{i\sum_{k=1}^m u_k X_k})$ for all $u_1, \dots, u_m \in \mathbb{R}$, where $X^{(n)}$ and X are random vectors taking values in $\mathbb{R}^m$.*

Definition 3.13. *Given a Borel probability measure $\mu \in P(\mathbb{R}^m)$, its* characteristic function *$\hat{\mu}\colon \mathbb{R}^m \to \mathbb{C}$ is defined by*

$$\hat{\mu}(u_1, \dots, u_m) := \int_{\mathbb{R}^m} e^{i\sum_{k=1}^m u_k x_k} \mu(dx), \quad u_1, \dots, u_m \in \mathbb{R}. \tag{23}$$

The function $\hat{\mu}$ has a property that is called *positive definiteness*, wich means the following.

Definition 3.14. *A function $f\colon \mathbb{R}^m \to \mathbb{C}$ is called* positive definite *(sometimes* positive semi-definite*) if for every $n \geq 1$ and $\underline{u}_1, \dots, \underline{u}_n \in \mathbb{R}^m$*

$$\sum_{k=1}^n \sum_{l=1}^n f(\underline{u}_k - \underline{u}_l) z_k \bar{z}_l \in \mathbb{R}$$

and

$$\sum_{k=1}^n \sum_{l=1}^n f(\underline{u}_k - \underline{u}_l) z_k \bar{z}_l \geq 0$$

for all $z_1, \dots, z_n \in \mathbb{C}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X = (X_1, \dots, X_m)$ be a random vector. If $\mu_X \in P(\mathbb{R}^m)$ denotes the law of X, we have in view of Proposition 2.18 that

$$\begin{aligned}\mathbb{E}(e^{i\sum_{k=1}^m u_k X_k}) &= \int_{\mathbb{R}^m} e^{i\sum_{k=1}^m u_k x_k} \mu_X(dx) \\ &= \hat{\mu}_X(u_1, \dots, u_m), \quad u_1, \dots, u_m \in \mathbb{R}.\end{aligned}$$

Definition 3.15. *The function* $(u_1, \dots, u_m) \to \mathbb{E}(e^{i\sum_{k=1}^m u_k X_k})$ *is called the* characteristic function *of the random vector* X.

We are now in a position to state important properties of characteristic functions, convergence in distribution, and their relationship.

Theorem 3.16. *I. Let* X *be a random vector with values in* $\mathbb{R}^m$ *and let* $(u_1, \dots, u_m) \mapsto \phi_X(u_1, \dots, u_n)$ *be its characteristic function. Then:*

(i) ϕ_X *is positive definite, in particular*

(a) $|\phi_X(u)| \leq \phi_X(0) = 1$ *for all* $u \in \mathbb{R}^m$ *(hence* ϕ_X *is bounded),*

(b) $\phi_X(-u) = \overline{\phi_X(u)}$ *for all* $u \in \mathbb{R}^m$,

(ii) ϕ_X *is uniformly continuous on* $\mathbb{R}^m$.

(iii) If $\mathbb{E}(\|X\|^n) < \infty$ *for some* $n \geq 1$, *then* $\phi_X \in C^n(\mathbb{R}^m, \mathbb{C})$, *i.e.* ϕ_X *is* n *times continuously differentiable, and:*

$$\left(\frac{\partial}{\partial u_1}\right)^{r_1} \cdots \left(\frac{\partial}{\partial u_m}\right)^{r_m} \phi_X(u) = \mathbb{E}\left((iX^{(1)})^{r_1} \cdots (iX^{(n)})^{r_m} e^{i\sum_{k=1}^m u_k X^{(n)}}\right), \quad (24)$$

$u \in \mathbb{R}^m$, *with* $r_1 + \cdots + r_m \leq n$,

$$\phi_X(u) = 1 + i\sum_{k=1}^m \mathbb{E}(X^{(k)})u_k + o(\|u\|), \quad (25)$$

if $\mathbb{E}(\|X\|) < \infty$, *and*

$$\phi_X(u) = 1 + i\sum_{k=1}^m \mathbb{E}(X^{(k)})u_k - \tfrac{1}{2}\sum_{k=1}^m\sum_{l=1}^m \mathbb{E}(X^{(k)}X^{(l)})u_k u_l + o(\|u\|^2), \quad (26)$$

if $\mathbb{E}(\|X\|^2) < \infty$.

(See [6, Prop. 4.1.3,4, p. 201–202], [8, §II.12.8, p. 287–288]; 1-dim also [1, Prop. 8.27, p. 171, Prop. 8.44, p.180].)

II. (i) *(Bochner's Theorem) Let $f\colon \mathbb{R}^m \to \mathbb{C}$ be a positive definite function satisfying $f(0) = 1$. Then the following statements are equivalent:*

(a) *$f = \phi_X$ for some random vector X,*

(b) *f is continuous,*

(c) *f is uniformly continuous.*

(ii) *(Continuity Theorem) Let $X^{(1)}, X^{(2)}, \dots$ be random vectors with values in $\mathbb{R}^m$. Let $h\colon \mathbb{R}^m \to \mathbb{C}$ be such that*

$$\phi_{X^{(n)}}(u) \to h(u), \text{ for all } u \in \mathbb{R}^m.$$

Then there exists a random vector $X\colon \Omega \to \mathbb{R}^m$ with characteristic function h if and only if h is continuous at 0.

(See [1, Problem 17, p.173, Thm 11.6, p. 236].)

3.5. Langevin Equation IV

We proceed with our investigation of the Langevin equation. For $n \geq 1$ and $h = T/n$ we have (19):

$$\begin{cases} X^{(n)}(t) = 0 & \text{if } t \in [0, h), \\ X^{(n)}(t) = \beta^{(n)} \sum_{k=1}^{[\frac{nt}{T}]} \xi_k, & \text{if } t \geq h, \end{cases}$$

with $\beta^{(n)} = \frac{\sigma}{\sqrt{n}}$, $\sigma > 0$. Let $0 = t_0 < t_1 < \cdots < t_m$. In order to determine the limiting distribution of $\{(X^{(n)}(t_0), \dots, X^{(n)}(t_m))\}_n$, we will first consider the increments

$$\begin{aligned} Y_k^{(n)} &:= X^{(n)}(t_k) - X^{(n)}(t_{k-1}), \quad k = 1, \dots, m, \\ Y^{(n)} &:= (Y_1^{(n)}, \dots, Y_m^{(n)}). \end{aligned}$$

We have for the characteristic function of $Y^{(n)}$:

$$\begin{aligned} \phi_{Y^{(n)}}(u) &= \int_\Omega e^{i \sum_{k=1}^m u_k (X^{(n)}(t_k) - X^{(n)}(t_{k-1}))(\omega)} \mathbb{P}(d\omega) \\ &= \int_\Omega e^{i \sum_{k=1}^m u_k \beta^{(n)} (\sum_{l=1}^{[\frac{nt_k}{T}]} \xi_l - \sum_{j=1}^{[\frac{nt_{k-1}}{T}]} \xi_j)(\omega)} \mathbb{P}(d\omega) \\ &= \int_{\mathbb{R}^\mathbb{N}} e^{i \sum_{k=1}^m u_k \beta^{(n)} (\sum_{l=1}^{[\frac{nt_k}{T}]} x_l - \sum_{j=1}^{[\frac{nt_{k-1}}{T}]} x_j)} \mu(dx), \end{aligned}$$

where $\mu(dx) = \otimes_{k=1}^{\infty}(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)(dx_k)$.

Now we choose $n_1 \geq \max\{\frac{T}{t_1}, \frac{T}{t_2-t_1}, \ldots, \frac{T}{t_m-t_{m-1}}\}$ such that $[\frac{nt_{k-1}}{T}] < [\frac{nt_k}{T}]$ for all $1 \leq k \leq m$ and $n \geq n_1$. Then for $n \geq n_1$:

$$\begin{aligned}
\phi_{Y^{(n)}}(u) &= \int_{\mathbb{R}^{\mathbb{N}}} e^{i\sum_{k=1}^m u_k\beta^{(n)}\sum_{l=[\frac{nt_{k-1}}{T}]+1}^{[\frac{nt_k}{T}]} x_l}\mu(dx) \\
&= \Pi_{k=1}^m \int_{\mathbb{R}^{\mathbb{N}}} e^{iu_k\beta^{(n)}\sum_{l=[\frac{nt_{k-1}}{T}]+1}^{[\frac{nt_k}{T}]} x_l}\mu(dx) \\
&= \Pi_{k=1}^m \Pi_{l=[\frac{nt_{k-1}}{T}]+1}^{[\frac{nt_k}{T}]} \int_{\mathbb{R}} e^{iu_k\beta^{(n)}x_l}(\tfrac{1}{2}\delta_{-1} + \tfrac{1}{2}\delta_1)(dx_l) \\
&= \Pi_{k=1}^m \Pi_{l=[\frac{nt_{k-1}}{T}]+1}^{[\frac{nt_k}{T}]} \cos(u_k\beta^{(n)}) \\
&= \Pi_{k=1}^m [\cos(u_k\beta^{(n)})]^{[\frac{nt_k}{T}]-[\frac{nt_{k-1}}{T}]}.
\end{aligned}$$

Since $\beta^{(n)} = \frac{\sigma}{\sqrt{n}}$, we have

$$\lim_{n\to\infty}[\cos(u_k\beta^{(n)})]^{[\frac{nt_k}{T}]-[\frac{nt_{k-1}}{T}]} = e^{-\frac{1}{2}u_k^2\sigma^2(\frac{t_k-t_{k-1}}{T})}, \quad k = 1, \ldots, m.$$

Hence

$$\lim_{n\to\infty}\phi_{Y^{(n)}}(u) = \Pi_{k=1}^m e^{-\frac{1}{2}u_k^2\sigma^2(\frac{t_k-t_{k-1}}{T})}, \quad u \in \mathbb{R}^m. \tag{27}$$

We arrive at the conclusion that the joint distributions of $(X^{(n)}(t_1), X^{(n)}(t_2) - X^{(n)}(t_1), \ldots, X^{(n)}(t_m) - X^{(n)}(t_{m-1}))$ converge to the joint distribution of m independent Gaussian variables $Y_1, \ldots, Y_m$ with mean 0 and variance $\sigma_k^2 = \frac{\sigma^2}{T}(t_k - t_{k-1})$ for every $0 = t_0 < t_1 < \cdots < t_m$. This leads us to the important notion of Gaussian measure, Gaussian random variable, and Gaussian system (family).

4. Gaussian systems

DEFINITION 4.1. *(i) A* Gaussian measure μ on $\mathbb{R}^d$ *is a Borel probability measure on* $\mathbb{R}^d$ *(i.e.* $\mu \in P(\mathbb{R}^d)$*) with characteristic function* $\hat{\mu}\colon \mathbb{R}^d \to \mathbb{C}$ *of the form*

$$\begin{aligned}
\hat{\mu}(u_1, \ldots, u_d) &= e^{i\sum_{k=1}^d u_k m_k - \frac{1}{2}\sum_{k,l=1}^d C_{kl}u_k u_l} \\
&= e^{i<\underline{u},\underline{m}> - \frac{1}{2}<C\underline{u},\underline{u}>},
\end{aligned}$$

where $m = (m_1, \dots, m_d) \in \mathbb{R}^d$ and $\{C_{kl}\}$ is a symmetric non-negative definite $d \times d$ matrix. The measure μ is denoted by $N(m, C)(dx)$. (Observe that the function $\hat{\mu}$ is positive definite, uniformly continuous, C^∞, and satisfies $\hat{\mu}(0) = 1$).

(ii) *A* random vector $(X_1, \dots, X_d)$ *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called* Gaussian *(or* normally distributed*) if its distribution is a Gaussian measure on* $\mathbb{R}^d$.

(iii) *A collection of random variables* $\{X_\alpha\}_{\alpha \in \mathcal{I}}$ *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, indexed by a nonempty index set* $\mathcal{I}$*, is called a* Gaussian system *(or family) if the random vector* $(X_{\alpha_1}, \dots, X_{\alpha_n})$ *is Gaussian for every* $n \geq 1$ *and all indices* $\alpha_1, \dots, \alpha_n$ *chosen from* $\mathcal{I}$.

PROPOSITION 4.2. *If* $(X_1, \dots, X_d)$ *is a Gaussian vector with characteristic function* $\phi_{(X_1,\dots,X_d)}(u_1, \dots, u_d) = \exp(i < \underline{u}, \underline{m} > -\frac{1}{2}\underline{u}^T C \underline{u})$, $\underline{u} \in \mathbb{R}^d$*, then*

$$\begin{aligned} m_k &= \int_\Omega X_k d\mathbb{P}, \quad k = 1, \dots, d, \text{ and} \\ C_{kl} &= \int_\Omega (X_k - m_k)(X_l - m_l) d\mathbb{P}, \quad 1 \leq k, l \leq d. \end{aligned}$$

Hence $\underline{m} = (m_1, \dots, m_d)$ *is the mean value of* $(X_1, \dots, X_d)$ *and* C *is the covariance matrix of* $(X_1, \dots, X_d)$*. If* $\underline{m} = 0$*, then the random vector is called* centered.
(See [8, §II.13.3, p. 298–299].)

PROPOSITION 4.3. *Let* $\mu \in P(\mathbb{R}^d)$ *be a Gaussian measure with mean value* $m \in \mathbb{R}^d$ *and covariance (matrix)* C*. Then* μ *is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^d$ *if and only if* C *is regular. In that case the density of* μ *is given by*

$$p(x) = \frac{1}{(2\pi)^{d/2} (\det C)^{1/2}} \exp\{-\tfrac{1}{2}(x - m)^T C^{-1}(x - m)\}, \quad x \in \mathbb{R}^d.$$

(See [8, §II.13.2, p. 296–298].)

PROPOSITION 4.4. *Let $\mu \in P(\mathbb{R}^d)$ be a Gaussian measure with mean value $m \in \mathbb{R}^d$ and covariance C. Let $\phi\colon \mathbb{R}^d \to \mathbb{R}^n$ be defined by $\phi(x) = a + Ax$ where $a \in \mathbb{R}^n$ and A is an $n \times d$ matrix. Then the image measure $(\phi \circ \mu)$ on $\mathbb{R}^n$ is Gaussian with mean value $a + Am$ and covariance matrix ACA^T.*

Proof. Let $u \in \mathbb{R}^n$, then

$$\begin{aligned}
\widehat{(\phi \circ \mu)}(u) &= \int_{\mathbb{R}^n} e^{i<y,u>}(\phi \circ \mu)(dy) \\
&= \int_{\mathbb{R}^d} e^{i<\phi(x),u>}\mu(dx) \\
&= \int_{\mathbb{R}^d} e^{i<a,u>}e^{i<Ax,u>}\mu(dx) \\
&= e^{i<a,u>}e^{i<m,A^Tu>}e^{-\frac{1}{2}<CA^Tu,A^Tu>} \\
&= e^{i<a+Am,u>-\frac{1}{2}<ACA^Tu,u>}.
\end{aligned}$$

□

COROLLARY 4.5. *If we take in Proposition 4.4 $n = 1$, $a = 0$, and $Ax = \lambda_1 x_1 + \cdots + \lambda_d x_d$, $x \in \mathbb{R}^d$, then $(A \circ \mu)$ is Gaussian. As a consequence, if $(X_1, \ldots, X_d)$ is a Gaussian random vector, then $\lambda_1 X_1 + \cdots + \lambda_d X_d$ is a Gaussian random variable. In particular with $\lambda_i = \delta_{ij}$, X_i is a Gaussian random variable for all $i = 1, \ldots, d$.*

Conversely, if $X_1, \ldots, X_d$ are Gaussian random variables, then $\lambda_1 X_1 + \cdots + \lambda_d X_d$ may not be Gaussian.

EXAMPLE 4.6. *If X_1, X_2 are independent $N(0,1)$ and*

$$(X,Y) := \begin{cases} (X_1, |X_2|) & \text{if } X_1 \geq 0, \\ (X_1, -|X_2|) & \text{if } X_1 < 0, \end{cases}$$

then X and Y are both Gaussian but (X,Y) is not.

PROPOSITION 4.7. *(i) Let $(X_1, \ldots, X_d)$ be a random vector. Then $(X_1, \ldots, X_d)$ is a Gaussian random vector if and only if $\lambda_1 X_1 + \cdots + \lambda_d X_d$ is a Gaussian random variable for every $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$.*

(ii) *Let $(X_1, \dots, X_d)$ be Gaussian. Then $X_1, \dots, X_d$ are independent if and only if they are* uncorrelated, *i.e.*

$$\mathbb{E}((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)) = \delta_{ij}, \quad i \neq j,\ 1 \leq i, j \leq d.$$

(iii) *If $X_1, \dots, X_d$ are independent Gaussian random variables, then $(X_1, \dots, X_d)$ is Gaussian.*
(See [8, Thm 1, p.299].)

EXERCISE 4.8: (S.N. Bernstein) Let X and Y be independent identically distributed random variables with finite variance. Show that if $X + Y$ and $X - Y$ are independent, then X and Y are Gaussian. (See [9, Thm 5.23, p. 102].)

PROPOSITION 4.9. *Let $\{\mu_n\}_{n \geq 1}$ be a sequence of Gaussian measures on $\mathbb{R}^d$ with expectations $\{m_n\}_n$ and covariances $\{C_n\}_n$. Then $\mu_n \xrightarrow{\text{D}} \mu$ for some $\mu \in P(\mathbb{R}^d)$ if and only if $m_n \to m$ and $C_n \to C$ for some $m \in \mathbb{R}^d$ and some $d \times d$ matrix C. If this is the case, then μ is Gaussian with expectation m and covariance C.*
(See [8, §II.13.5, p. 302–303].)

4.1. Multidimensional central limit theorem

THEOREM 4.10. *Let $\{(X_1^{(n)}, X_2^{(n)}, \dots, X_d^{(n)})\}_{n \geq 1}$ be a sequence of independent random vectors, all having the same distribution with zero mean and covariance matrix C. Then*

$$\tfrac{1}{\sqrt{n}} \left(\sum_{k=1}^{n} X_1^{(k)}, \dots, \sum_{k=1}^{n} X_d^{(k)} \right) \xrightarrow{\text{D}} (X_1, \dots, X_d),$$

where $(X_1, \dots, X_d)$ is a Gaussian random vector with mean zero and covariance matrix C.
(See [1, Thm 11.10, p. 238].)

4.2. Langevin Equation V

We return to our sequence $\{X^{(n)}(t)\}_{n \geq 1}$, again. Let $0 = t_0 < t_1 < \dots < t_m$. Then the sequence of random vectors $\{(X^{(n)}(t_1), X^{(n)}(t_2) -$

$X^{(n)}(t_1), \dots , X^{(n)}(t_m)-X^{(n)}(t_{m-1}))\}_n$ converges in law to the Gaussian random vector $(Y_1, \dots , Y_m)$ with mean zero and covariance matrix

$$C = \frac{\sigma^2}{T}\begin{pmatrix} t_1 & & & \\ & t_2 - t_1 & & 0 \\ 0 & & \ddots & \\ & & & t_m - t_{m-1} \end{pmatrix}.$$

It follows that $\{(X^{(n)}(t_1), \dots , X^{(n)}(t_m))\}_n$ converges in law to the random vector $(X_1, X_2, \dots , X_m) = (Y_1, Y_1 + Y_2, \dots , Y_1 + Y_2 + \cdots + Y_m)$, which is also a Gaussian random vector, with mean zero. We have for $1 \le i \le j$:

$$\begin{aligned} \mathbb{E}(X_i X_j) &= \mathbb{E}(\sum_{k=1}^{i} Y_k \sum_{l=1}^{j} Y_l) = \sum_{k=1}^{i}\sum_{l=1}^{j} \mathbb{E}(Y_k Y_l) \\ &= \sum_{k=1}^{i}\sum_{l=1}^{i} \mathbb{E}(Y_k Y_l) = \sum_{k=1}^{i} \mathbb{E}(Y_k^2) \\ &= \tfrac{\sigma^2}{T}\sum_{k=1}^{i}(t_k - t_{k-1}) = \tfrac{\sigma^2}{T} t_i = \tfrac{\sigma^2}{T}\min\{t_i, t_j\}. \end{aligned}$$

We now want to vary the t_i and think of finer and finer partitions of the interval $[0, T]$. A natural question then is to know whether it is possible to view the limits of the finite dimensional distributions of $\{X^{(n)}(t)\}_{n\ge1}$ as the finite dimensional distributions of a Gaussian system $\{X(t)\}_{t\ge0}$. That means: is there a Gaussian system $\{X(t)\}_{t\ge0}$ such that $\{(X^{(n)}(t_1), \dots , X^{(n)}(t_m))\}_n$ converges in law to $(X(t_1), \dots , X(t_m))$ for each partition $0 = t_0 < t_1 < \cdots < t_m$ of $[0, T]$? By the above calculation, this system should have covariances $\mathbb{E}(X(t)X(s)) = \min\{t, s\}\frac{\sigma^2}{T}$, $t, s \ge 0$. Thus, we arrive at the question: *does there exist a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a centered Gaussian system* $\{X(t)\}_{t\ge0}$ *on it satisfying*

$$\mathbb{E}(X(t)X(s)) = \min\{t, s\}, \quad t, s \ge 0, \text{ together with } X(0) = 0?$$

The next theorem gives an answer to this question together with an additional and useful property of such a Gaussian system.

THEOREM 4.11. *Let $\Omega = C_0([0,\infty))$ be the space of continuous functions on $[0,\infty)$ vanishing at 0 equipped with the metric $d(\omega,\eta) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{d_k(\omega,\eta)}{1+d_k(\omega,\eta)}$, where $d_k(\omega,\eta) := \max_{0\le t\le k} |\omega(t)-\eta(t)|$, $\omega,\eta \in \Omega$. (Then (Ω,d) is a complete separable metric space). Let $B(\Omega)$ be the Borel σ-algebra on (Ω,d). Then there exists a* unique *Borel probability measure $\mathbb{P}$ on $(\Omega, B(\Omega))$ such that the system of random variables $\{X_t\}_{t\ge 0}$ defined by $X_t(\omega) = \omega(t)$, $t \ge 0$, $\omega \in \Omega$, is a centered Gaussian system satisfying $\int_\Omega X_t X_s d\mathbb{P} = \min\{t,s\}$, $t,s \ge 0$. Such a measure is called the* Wiener measure *and the corresponding system the* normalized Brownian motion. *We shall denote this system by $\{W_t\}_{t\ge 0}$.*
(One of many proofs is to apply Theorem 2.1.6, p. 51 of [10].)

REMARK 4.12. *The Brownian motion has a very important property: the independence of its increments. Let $0 \le t_0 < t_1 \le t_2 < t_3$. Then*

$$\begin{aligned}
&\mathbb{E}(X_{t_3} - X_{t_2})(X_{t_1} - X_{t_0}) \\
&= \mathbb{E}(X_{t_3}X_{t_1}) + \mathbb{E}(X_{t_2}X_{t_0}) - \mathbb{E}(X_{t_3}X_{t_0}) - \mathbb{E}(X_{t_2}X_{t_1}) \\
&= t_1 + t_0 - t_0 - t_1 = 0.
\end{aligned}$$

Since $\mathbb{E}(X_{t_k}) = 0$, $k = 1,2,3$, the variables $X_{t_3} - X_{t_2}$ and $X_{t_1} - X_{t_0}$ are uncorrelated and since the system is Gaussian, they are independent. Moreover, the law of $X_{t+h} - X_t$, $t,h \ge 0$, is Gaussian with mean 0 and variance:

$$\begin{aligned}
\mathbb{E}(X_{t+h} - X_t)^2 &= \mathbb{E}(X_{t+h})^2 + \mathbb{E}(X_t)^2 - 2\mathbb{E}(X_{t+h}X_t) \\
&= t + h + t - 2t = h.
\end{aligned}$$

Therefore the variance and consequently the law of the increments $X_{t+h} - X_t$ is independent of t. We say that the process $\{X_t\}_{t\ge 0}$ has stationary independent increments.

Let us compute the finite dimensional distributions of X_t. Let $0 = t_0 < t_1 < t_2 < \cdots < t_m$, $m \ge 1$, and let $A \in B(\mathbb{R}^m)$. Set $Y_1 := X_{t_1}$, $Y_2 := X_{t_2} - X_{t_1}, \ldots, Y_m := X_{t_m} - X_{t_m-1}$. Then

$$
\begin{aligned}
&\mathbb{P}\{(X_{t_1}, X_{t_2}, \dots, X_{t_m}) \in A\} \\
&= \mathbb{P}\{(Y_1, Y_1 + Y_2, \dots, Y_1 + Y_2 + \cdots + Y_m) \in A\} \\
&= \int_{\mathbb{R}^m} \mathbb{1}_A(y_1, y_1 + y_2, \dots, y_1 + \cdots + y_m) \mu_{Y_1}(dy_1) \cdots \mu_{Y_m}(dy_m) \\
&= \frac{1}{\sqrt{\Pi_{k=1}^m 2\pi(t_k - t_{k-1})}} \int_{\mathbb{R}^m} \mathbb{1}_A(y_1, y_1 + y_2, \dots, y_1 + \cdots + y_m) \\
&\qquad \cdot e^{-\frac{1}{2}\sum_{k=1}^m \frac{y_k^2}{2(t_k - t_{k-1})}} dy_1 \cdots dy_m \\
&= \frac{1}{\sqrt{\Pi_{k=1}^m 2\pi(t_k - t_{k-1})}} \int_{\mathbb{R}^m} \mathbb{1}_A(x_1, x_2, \dots, x_m) \\
&\qquad \cdot e^{-\frac{1}{2}\sum_{k=1}^m \frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}} dx_1 \cdots dx_m \\
&= \frac{1}{\sqrt{\Pi_{k=1}^m 2\pi(t_k - t_{k-1})}} \int_A e^{-\frac{1}{2}\sum_{k=1}^m \frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}} dx_1 \cdots dx_m.
\end{aligned}
$$

In particular, if $A = I_1 \times I_2 \times \cdots \times I_m$, we get:

$$
\begin{aligned}
&\mathbb{P}(\{X_{t_1} \in I_1, \dots, X_{t_m} \in I_m\}) \\
&\qquad = \int_{I_1} \cdots \int_{I_m} \Pi_{k=1}^m p(t_k - t_{k-1}; x_{k-1}, x_k) dx_1 \cdots dx_m,
\end{aligned}
$$

where

$$
p(s; x, y) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2}\frac{(y-x)^2}{s}}, \quad s > 0,\ x, y \in \mathbb{R}.
$$

This formula for $\mathbb{P}(\{X_{t_1} \in I_1, \dots, X_{t_m} \in I_m\})$ is called the *Einstein-Smoluchowski formula.*

Finally, we note that there may be other probability spaces then the one of Theorem 4.11 with systems of random variables on it that constitute centered Gaussian systems with the same covariances as $\{W_t\}_{t\geq 0}$. One can introduce the notion of 'stochastic equivalence' of stochastic systems, which yields a notion of uniqueness without specifying the underlying space. We will use the normalized Brownian motion $\{W_t\}_{t\geq 0}$ on Ω as defined in the theorem. The usual definition of Brownian motion is the following.

DEFINITION 4.13. *A* normalized Brownian motion *(or* Wiener process*) in $\mathbb{R}$ is a family of random variables $\{W_t\}_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that* *(i)* $W_0(\omega) = 0$ *for all* $\omega \in \Omega$,

(ii) *for every* $s, t \geq 0$, $W_{s+t} - W_s$ *is a Gaussian random variable with mean* 0 *and variance* t,

(iii) *for every* $0 = t_0 < t_1 < \cdots < t_n$ *the random variables* $W_{t_k} - W_{t_{k-1}}$, $k = 1, \ldots, n$, *are independent,*

(iv) *for every* $\omega \in \Omega$ *the* sample path $t \mapsto W_t(\omega)$ *from* $[0, \infty)$ *to* $\mathbb{R}$ *is continuous.*

(See [9, Def. 8.1, p. 220].)

5. Solution of the Langevin Equation

5.1. Langevin Equation VI

Let us collect the above presented notions and results in view of our initial problem: the Langevin equation. Consider first the case $c = 0$ and $x = 0$. Then the equation is

$$\frac{dX}{dt}(t) = \text{“noise”}, \quad t \geq 0. \tag{28}$$

To mathematically describe the randomness in the function X, we have introduced the notion of random variable. For every $t \geq 0$, $X(t)$ will be a random variable. We made a specific choice of randomness to model the ‘complete random influence’ expressed by the word ‘noise’. We have done so using discretizations and obtained

$$X^{(n)}(t) = \frac{t}{n}\alpha^{(n)} \sum_{k=1}^{n-1} \mathbb{1}_{[kh,(k+1)h)}(t)\xi_k, \quad t \geq 0,$$

where $\{\xi_k\}_{k=1}^{\infty}$ is a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(\{\xi_k = 1\}) = \mathbb{P}(\{\xi_k = -1\}) = 1/2$. To construct such a sequence, we used the notion of a product of a sequence of probability spaces. In order to get a non-degenerate limit as $n \to \infty$, we have chosen

$$\alpha^{(n)} = \tfrac{\sigma}{T}\sqrt{n}, \quad n \in \mathbb{N}.$$

As a limit of $\{X^{(n)}(t)\}_{t\geq 0, n\geq 1}$ we have the system $\{W(t)\}_{t\geq 0}$ from Theorem 4.11: for any partition $0 = t_0 < t_1 < \cdots < t_m \leq T$ the joint distributions of $(X^{(n)}(t_1), \ldots, X^{(n)}(t_m))$, $n \geq 1$, converge to the joint distribution of $(W(t_1), \ldots, W(t_m))$. Thus, we say that we have the normalized Brownian motion as a limit of $\{X^{(n)}(t)\}_{t\geq 0, n\geq 1}$. Accordingly, we find that the random function we meant to describe by (28) is the normalized Brownian motion $\{W(t)\}_{t\geq 0}$.

One may observe that we have given a meaning to the solution of (28) rather than to the equation itself. We could think of differentiating X to really obtain the "noise". For a fixed $\omega \in \Omega$, $t \mapsto X(t)(\omega)$ is a continuous function. If it is differentiable for sufficiently many ω, this would yield the noise term explicitly. However, it turns out that with probability 1 the function $t \mapsto X(t)(\omega)$ is not differentiable, even nowhere differentiable! On the one hand it is disappointing that we do not get a direct meaning of the equation, on the other it clearly expresses the irregularity of noise!

Let us now consider the general Langevin equation with arbitrary $c, x \in \mathbb{R}$:

$$\begin{cases} \frac{dX}{dt}(t) = c(\theta - X(t)) + \text{"noise"}, & t > 0, \\ X(0) = x. \end{cases} \tag{29}$$

We know now how to interpret this equation in an integrated form:

$$X(t) = x + c\int_0^t (\theta - X(s))ds + \sigma W(t), \quad t \geq 0. \tag{30}$$

(Note that we have introduced an extra parameter $\sigma \in \mathbb{R}$.) For fixed t, $W(t)$ and hence $X(t)$ are random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of the Brownian motion of Theorem 4.11, i.e. $W(t)$ and $X(t)$ are measurable functions from Ω to $\mathbb{R}$. So (30) is an equation of real valued functions on Ω. Explicitly,

$$X(t)(\omega) = x + c\int_0^t (\theta - X(s)(\omega))ds + \sigma W(t)(\omega), \quad \omega \in \Omega, t \geq 0,$$

or, in other words,

$$X(t)(\omega) = x + c\int_0^t (\theta - X(s)(\omega))ds + \sigma\omega(t), \; t \geq 0, \; \omega \in \Omega. \tag{31}$$

If we fix $\omega \in \Omega$, we can think of it as a parameter and (31) is then a deterministic integral equation of Volterra type, to which deterministic theory can be applied.

5.2. Volterra integral equations

Let $T > 0$. For $a, f \in C[0,T]$ consider the *Volterra integral equation*

$$u(t) + \int_0^t a(t-s)u(s)ds = f(t), \quad 0 \leq t \leq T. \tag{32}$$

For $a, b \in C[0,T]$ define the *convolution* of a with b as

$$(a * b)(t) = \int_0^t a(t-s)b(s)ds. \tag{33}$$

Observe that

(1) $a * b \in C[0,T]$

(2) $a * b = b * a$

(3) $(a * b) * c = a * (b * c)$

(4) $\|(a * b)\|_\infty \leq \|a\|_1 \|b\|_\infty$,

where $\|.\|_1$ and $\|.\|_\infty$ denote the integral norm and the supremum norm on $C[0,T]$, respectively. With this notation (32) becomes $u + a * u = f$.

Proposition 5.1. *Let $T > 0$ and let $a, f \in C[0,T]$. There exists one and only one function $u \in C[0,T]$ such that*

$$u + a * u = f. \tag{34}$$

Let $r \in C[0,T]$ be such that

$$r + a * r = a.$$

Then the solution to (34) is given by

$$u = f - r * f. \tag{35}$$

Proof. To prove existence and uniqueness, the idea is to use Banach's fixed point theorem. Observe that for any $\alpha \in \mathbb{R}$, (34) is equivalent to

$$e^{\alpha t}u(t) + \int_0^t e^{\alpha(t-s)}a(t-s)e^{\alpha s}u(s)ds = e^{\alpha t}f(t),\ t \in [0,T],$$

or

$$\tilde{u} + \tilde{a} * \tilde{u} = \tilde{f}, \tag{36}$$

with the notation $\tilde{v}(t) := e^{\alpha t}v(t)$, $t \in [0,T]$, for $v \in C[0,T]$.

Define $Tv := \tilde{f} - \tilde{a} * v$, $v \in C[0,T]$, then

$$\|Tv_1 - Tv_2\|_\infty \leq \|\tilde{a}\|_1 \|v_1 - v_2\|_\infty, \quad v_1, v_2 \in C[0,T].$$

Choose $\alpha < 0$ such that $\|\tilde{a}\|_1 < 1$. Then T is a strict contraction, so that (36) and therefore (34) have a unique solution in $C[0,T]$.

The last assertion follows from straightforward substitutions:

$$(f - r * f) + a * (f - r * f) = f - r * f + a * f - (a - r) * f = f.$$

□

The function r of the proposition is called the *resolvent* of equation (34). The second part of the proposition says that once the resolvent is solved from the equation, for any right hand side the solution can be computed from the resolvent by formula (35). For our purposes it is more convenient to use another function than r.

Proposition 5.2. *Let $T > 0$ and let $a, f \in C[0,T]$. Let $s \in C[0,T]$ be such that*

$$s + a * s = \mathbb{1}.$$

Then the solution $u \in C[0,T]$ of

$$u + a * u = f$$

is given by

$$u(t) = s(t)f(0) + \int_0^t s(t-\tau)df(\tau), \quad \tau \in [0,T],$$

where the integral is a Stieltjes integral.

Proof. According to Proposition 5.1, $s = \mathbb{1} - r * \mathbb{1}$, which is a C^1-function since $\mathbb{1} * r$ is a primitive of r. In particular, s is of bounded variation and the Stieltjes integral exists for all t. Moreover, $s' = -r$ and $s(0) = 1$.

By integration by parts and a substitution (see Exercise 2.8),

$$\begin{aligned} s(t)f(0) &+ \int_0^t s(t-\tau)df(\tau) \\ &= s(t)f(0) + s(0)f(t) - s(t)f(0) - \int_0^t f(\tau)ds(t-\tau) \\ &= f(t) + \int_0^t f(t-\tau)ds(\tau) = f(t) + \int_0^t f(t-\tau)s'(\tau)d\tau \\ &= f(t) - (f*r)(t) = u(t), \quad t \in [0,T]. \end{aligned}$$

□

5.3. Langevin equation VII

In the case of our Langevin equation (31), we have for any fixed $\omega \in \Omega$ a Volterra integral equation with kernel $a(t) = c$ and right hand side $f(t) = c\theta t + \sigma\omega(t)$, $t \geq 0$. The equation for the function s of Proposition 5.2 is

$$s + (c\mathbb{1}) * s = \mathbb{1},$$

or, equivalently, $s' + cs = 0$ and $s(0) = 1$, so that

$$s(t) = e^{-ct}, \quad t \in [0,T].$$

According to Propositions 5.1 and 5.2 we may conclude that equation (31) has a unique continuous solution, given by

$$\begin{aligned} X(t)(\omega) &= e^{-ct}x + \sigma\left\{\int_0^t e^{-c(t-s)}c\theta\,dt + \int_0^t e^{-c(t-s)}d\omega(s)\right\} \\ &= e^{-ct}x + (1-e^{-ct})\theta + \sigma\int_0^t e^{-c(t-s)}d\omega(s), \\ &\quad t \geq 0, \omega \in \Omega. \end{aligned} \tag{37}$$

For $\sigma = 0$ we find again the solution of the problem without noise (cf. (4)). For $\sigma \neq 0$, we get an additional term due to the noise.

To find out the probabilistic nature of this term, we want to fix a t and see $X(t)(\omega)$ as function of ω. Is it measurable? What can be said about its distribution? These questions will be dealt with in the next section.

6. Stochastic convolution

Let $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ be the probability space of Brownian motion as defined in Theorem 4.11. Every $\omega \in \Omega$ is a continuous function from $[0,\infty)$ to $\mathbb{R}$ and $\omega(0) = 0$. Therefore, for any $T > 0$ and any function $f\colon [0,T] \to \mathbb{R}$ that has bounded variation, ω is Stieltjes integrable with respect to f and, therefore, f is Stieltjes integrable with respect to ω on $[0,T]$. Define

$$W(f)(t)(\omega) := \int_0^t f(s)d\omega(s), \quad \omega \in \Omega,\ t \in [0,T],\ f \in BV[0,T],$$

where $BV[0,T]$ denotes the class of all real valued functions of bounded variation on $[0,T]$. (Recall that $C^1[0,T] \subset BV[0,T]$.) For $f \in BV[0,T]$, $W(f(t-\bullet))(t)(\omega) = \int_0^t f(t-s)d\omega(s)$, $\omega \in \Omega$, $t \in [0,T]$, is called *stochastic convolution* of f with Brownian motion.

The map $(\omega, t) \mapsto W(f)(t)(\omega)$ is continuous from $\Omega \times [0,T]$ to $\mathbb{R}$ —where Ω is equipped with the metric of Theorem 4.11— for any $f \in BV[0,T]$, because for $\omega_0 \in \Omega$ and $t_0 \in [0,T]$:

$$
\begin{aligned}
&|\int_0^t f(s)d\omega(s) - \int_0^{t_0} f(s)d\omega_0(s)| \\
&\leq \ |\int_0^t f(s)d\omega(s) - \int_0^t f(s)d\omega_0(s)| + |\int_{t_0}^t f(s)d\omega_0(s)| \\
&\leq \ |f(t)(\omega(t) - \omega_0(t)) - \int_0^t (\omega(s) - \omega_0(s))df(s)| \\
&\qquad + |f(t)\omega_0(t) - f(t_0)\omega_0(t_0) - \int_{t_0}^t \omega_0(s)df(s)| \\
&\leq \ |f(t)||\omega(t) - \omega_0(t)| + \max_{s\in[0,t]} |\omega(s) - \omega_0(s)| V_{[0,T]}(f) \\
&\qquad + |\int_{t_0}^t (\omega_0(t_0) - \omega_0(s))df(s) - \omega_0(t_0)f(t) + f(t)\omega_0(t)| \\
&\leq \ \max_{s\in[0,T]} |\omega(s) - \omega_0(s)| \Big(\|f\|_\infty + V_{[0,T]}(f) \Big) \\
&\qquad + \max_{s\in[t_0,t]} |\omega_0(s) - \omega_0(t_0)| \Big(V_{[0,T]}(f) + \|f\|_\infty \Big), \\
&\qquad \omega \in \Omega,\ 0 \leq t_0 \leq t \leq T,
\end{aligned}
$$

where $V_{[0,T]}(f)$ denotes the total variation of f on $[0, T]$, and a similar estimate holds for $t \leq t_0$.

Consequently, $\omega \mapsto W(f)(t)(\omega)$ is a random variable for any $f \in BV[0, T]$ and $t \in [0, T]$, and it is denoted by

$$\int_0^t f(s)dW(s) := W(f)(t).$$

What is its distribution? By definition, the Stieltjes integral is a limit of Riemann-Stieltjes sums. Let $f \in BV[0, T]$. Then we have for any $\omega \in \Omega$ and $n \geq 1$:

$$
\begin{aligned}
\int_0^T f(t)d\omega(t) &= \lim_{n\to\infty} \sum_{k=1}^n f(\tfrac{kT}{n}) \Big(\omega(\tfrac{k}{n}T) - \omega(\tfrac{k-1}{n}T) \Big) \\
&= \lim_{n\to\infty} Z_n(\omega),
\end{aligned}
$$

where

$$Z_n(\omega) := \sum_{k=1}^n f(\tfrac{kT}{n}) \Big(\omega(\tfrac{k}{n}T) - \omega(\tfrac{k-1}{n}T) \Big).$$

Since $Y_{n,k} := W(\frac{k}{n}T) - W(\frac{k-1}{n}T)$, $k = 1, \ldots, n$, are independent Gaussian random variables with mean 0 and variance T/n, then, by Proposition 4.7, each Z_n is a Gaussian random variable, $\mathbb{E}\,Z_n = 0$, and

$$\begin{aligned}\operatorname{var} Z_n &= \mathbb{E}\Big(\sum_{k=1}^{n} f(\tfrac{kT}{n})Y_{n,k}\Big)^2 = \sum_{k=1}^{n}\sum_{l=1}^{n} f(\tfrac{kT}{n})f(\tfrac{lT}{n})\mathbb{E}\,(Y_{n,k}Y_{n,l}) \\ &= \sum_{k=1}^{n} f(\tfrac{kT}{n})^2\tfrac{T}{n}.\end{aligned}$$

Since f is of bounded variation, so is f^2 and hence f^2 is Riemann integrable, and it follows that $\operatorname{var} Z_n \to \int_0^T f^2(t)dt$ as $n \to \infty$. Application of Proposition 4.9 yields that the sequence $\{Z_n\}_n$ converges in distribution to a Gaussian random variable with mean zero and variance $\int_0^T f(t)^2dt$. More explicitly, we can look at the characteristic function of Z_n:

$$\mathbb{E}\,e^{iZ_n u} = e^{-\frac{u^2}{2}\sum_{k=1}^{n} f(\frac{kT}{n})^2\frac{T}{n}}, \quad u \in \mathbb{R},$$

by Proposition 4.2. Since $Z_n(\omega) \to \int_0^T f(t)d\omega(t)$ for all $\omega \in \Omega$, Lebesgue's dominated convergence theorem yields that

$$\mathbb{E}\,e^{i\int_0^T f(t)dW(t)u} = \lim_{n\to\infty}\mathbb{E}\,e^{iZ_n u} = e^{-\frac{u^2}{2}\int_0^T f(t)^2dt}, \text{ for all } u \in \mathbb{R}.$$

Thus, $\int_0^T f(t)dW(t)$ is a Gaussian random variable with mean 0 and variance equal to $\int_0^T f(t)^2dt$.

Proposition 6.1. *Let $T > 0$ and $f \in BV[0,T]$. Let*

$$\int_0^t f(s)dW(s)(\omega) := \int_0^t f(s)d\omega(s), \quad \omega \in \Omega, t \in [0,T].$$

Then

(i) *$\int_0^t f(s)dW(s)$ is a Gaussian random variable with mean 0 and variance equal to $\int_0^t f(s)^2ds$, for every $t \in [0,T]$.*

(ii) *If $g \in BV[0,T]$, then*

$$\mathbb{E}\Big(\int_0^t f(\tau)dW(\tau)\int_0^t g(\tau)dW(\tau)\Big) = \int_0^t f(\tau)g(\tau)d\tau,$$

for every $t \in [0,T]$.

(iii) $\{\int_0^t f(s)dW(s)\}_{t\in[0,T]}$ *is a Gaussian system with*

$$\mathbb{E}\int_0^t f(\tau)dW(\tau) = 0 \quad \textit{for all } t \in [0,T]$$

and

$$\mathbb{E}\Big(\int_0^t f(\tau)dW(\tau)\int_0^s f(\sigma)dW(\sigma)\Big) = \int_0^{t\wedge s} f(\tau)^2 d\tau, \quad t,s \in [0,T].$$

Proof. (i) Apply the above results to the bounded variation function $s \mapsto f(s)\mathbb{1}_{[0,t]}(s)$ instead of f.

(ii) Note first that $g \mapsto \int_0^t g(s)dW(s)$ is linear and that $g^2 \in BV[0,T]$ for every $g \in BV[0,T]$. By polarization, we obtain from (i):

$$\begin{aligned}
&\mathbb{E}(\int_0^t f(\tau)dW(\tau)\int_0^t g(\tau)dW(\tau)) \\
&= \tfrac{1}{4}\mathbb{E}\Big(\int_0^t (f(\tau)+g(\tau))dW(\tau)\Big)^2 - \tfrac{1}{4}\mathbb{E}\Big(\int_0^t (f(\tau)-g(\tau))dW(\tau)\Big)^2 \\
&= \tfrac{1}{4}\int_0^t (f(\tau)+g(\tau))^2 d\tau - \tfrac{1}{4}\int_0^t (f(\tau)-g(\tau))^2 d\tau \\
&= \int_0^t f(\tau)g(\tau)d\tau.
\end{aligned}$$

(iii) Let $0 \le t_1 < t_2 < \cdots < t_n \le T$. We have to show that $(\int_0^{t_1} f(s)dW(s), \dots, \int_0^{t_n} f(s)dW(s))$ is a Gaussian random vector. This follows with Proposition 4.7 from the observation that

$$\sum_{k=1}^n \lambda_k \int_0^{t_k} f(s)dW(s) = \int_0^T (\sum_{k=1}^n \lambda_k f(s)\mathbb{1}_{[0,t_k]}(s))dW(s)$$

is a Gaussian random variable (by (i)) for every $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Moreover, its mean is 0. Similarly, by (ii):

$$\mathbb{E}\Big(\int_0^t f(\tau)dW(\tau)\int_0^s f(\sigma)dW(\sigma)\Big) =$$

$$\begin{aligned}
&= \mathbb{E}\Big(\int_0^T f(\tau)\mathbb{1}_{[0,t]}(\tau)d\tau\int_0^T f(\sigma)\mathbb{1}_{[0,s]}(\sigma)d\sigma\Big)\\
&= \int_0^T f(\tau)^2\mathbb{1}_{[0,t]}(\tau)\mathbb{1}_{[0,s]}(\tau)d\tau = \int_0^{t\wedge s} f(\tau)^2 d\tau, \quad t,s\in[0,T].
\end{aligned}$$

□

6.1. Langevin Equation VIII

Formula (37) gives the solution to the Langevin equation:

$$X(t)(\omega) = e^{-ct}x + (1-e^{-ct})\theta + \sigma\int_0^t e^{-c(t-s)}d\omega(s), \quad t\geq 0, \omega\in\Omega. \tag{38}$$

The last term is a stochastic convolution. According to Proposition 6.1, it is a Gaussian random variable with mean 0 and variance equal to $\int_0^t(\sigma e^{-c(t-s)})^2 ds = \frac{\sigma^2}{2c}(1-e^{-2ct})$, $t\geq 0$, if $c\neq 0$ and σt if $c=0$.

Proposition 6.2. *The Langevin equation (2), which is to be interpreted in integrated form (30):*

$$X(t)(\omega) = x + c\int_0^t(\theta - X(s)(\omega))ds + \sigma W(t)(\omega), \quad \omega\in\Omega,\ t\geq 0,$$

has solution $\{X(t)\}_{t\geq 0}$ given by (38), which is a Gaussian system defined on the probability space $(\Omega,\mathcal{B}(\Omega),\mathbb{P})$ of Theorem 4.11, if $c\neq 0$ with

$$\begin{aligned}
\mathbb{E}X(t) &= e^{-ct}x + (1-e^{-ct})\theta, \quad t\geq 0,\\
\text{cov}\,(X(t),X(s)) &= \tfrac{\sigma^2}{2c}\{e^{-c|t-s|} - e^{-c(t+s)}\}, \quad t,s\geq 0,
\end{aligned}$$

and if $c=0$ with

$$\begin{aligned}
\mathbb{E}X(t) &= x, \quad t\geq 0,\\
\text{cov}\,(X(t),X(s)) &= \sigma^2(t\wedge s), \quad t,s\geq 0.
\end{aligned}$$

In particular, if $c \neq 0$,

$$\operatorname{var} X(t) = \tfrac{\sigma^2}{2c}(1 - e^{-2ct}), \quad t \geq 0,$$

and if $c > 0$ then $\mathcal{L}(X(t)) \xrightarrow{\mathrm{D}} N(\theta, \frac{\sigma^2}{2c})$ as $t \to \infty$.

Proof. Assume $c \neq 0$. According to Proposition 6.1,

$$\begin{aligned}
\operatorname{cov}(X(t), X(s)) &= \mathbb{E}\Big((X(t) - \mathbb{E}X(t))(X(s) - \mathbb{E}X(s))\Big) \\
&= \mathbb{E}\Big(\int_0^t \sigma e^{-c(t-\tau)} dW(\tau) \int_0^s \sigma e^{-c(s-\tau)} dW(\tau)\Big) \\
&= \sigma^2 e^{-c(t+s)} \mathbb{E}\Big(\int_0^t e^{c\tau} dW(\tau) \int_0^s e^{c\tau} dW(\tau)\Big) \\
&= \sigma^2 e^{-c(t+s)} \int_0^{t \wedge s} e^{2c\tau} d\tau \\
&= \tfrac{\sigma^2}{2c}\{e^{-c|t-s|} - e^{-c(t+s)}\}, \quad t, s \geq 0.
\end{aligned}$$

If $c > 0$, then $\mathbb{E}X(t) \to \theta$ and $\operatorname{var} X(t) \to \frac{\sigma^2}{2c}$, so that Proposition 4.9 yields that $X(t)$ converges in distribution to a Gaussian with mean θ and variance $\frac{\sigma^2}{2c}$. □

7. Stochastic initial conditions and invariant measure

As the solution of the Langevin equation is at every t a random variable, we may as well want to start with an initial condition that is a random variable. To stay within our framework, we assume it Gaussian and stochastically independent of the Brownian motion involved. In view of our construction of independent random variables by means of products of probability spaces, we will construct a new probability space $\tilde{\Omega}$ (as a product of $\mathbb{R}$ and Ω) and consider the random variables in the Langevin equation as functions on this space.

Let $(\Omega, B(\Omega), \mathbb{P})$ be the probability space of normalized Brownian motion of Theorem 4.11 and let $\mathbb{R}$ be equipped with a Gaussian measure $N(x, \rho^2)$. Let $\tilde{\Omega} = \mathbb{R} \times \Omega$ be the product space, that means that it is equipped with the σ-algebra $B(\mathbb{R}) \otimes B(\Omega)$ and the measure

$N(x,\rho^2)\otimes\mathbb{P}$. Let

$$\begin{aligned} X_0(\gamma,\omega) &:= \gamma, \quad (\gamma,\omega)\in\mathbb{R}\times\Omega, \\ W(t)(\gamma,\omega) &:= \omega(t), \quad (\gamma,\omega)\in\mathbb{R}\times\Omega,\ t\ge 0; \end{aligned}$$

then X_0 is an $N(x,\rho^2)$ distributed random variable on $\tilde\Omega$ independent of $W(t)$ for every $t\ge 0$. We can now consider the Langevin equation with stochastic initial condition X_0 as an equation of functions on $\tilde\Omega$:

$$X(t) = X_0 + c\int_0^t (\theta - X(s))ds + \sigma W(t), \quad t\ge 0.$$

(Note that if $\rho = 0$ we have the deterministic initial value x.) As in § 5.1, the equation can be seen as a deterministic Volterra integral equation with parameter $(\gamma,\omega)\in\tilde\Omega$. It has a unique continuous solution given by

$$X(t)(\gamma,\omega) = e^{-ct}X_0(\gamma,\omega) + (1-e^{-ct})\theta + \sigma\int_0^t e^{-c(t-s)}d\omega(s), \tag{39}$$

$t\ge 0$, $(\gamma,\omega)\in\mathbb{R}\times\Omega$. The first and last term, seen as functions on $\tilde\Omega$, are independent Gaussian random variables, hence $X(t)$ is Gaussian for every $t\ge 0$. $\{X(t)\}_{t\ge 0}$ is even a Gaussian system. The mean and variance are given by

$$\begin{aligned} \mathbb{E}X(t) &= e^{-ct}x + (1-e^{-ct})\theta, & (40)\\ \operatorname{cov}(X(t),X(s)) &= \rho^2 e^{-(t+s)} + \tfrac{\sigma^2}{2c}\{e^{-c|t-s|} - e^{-c(t+s)}\}, \quad t,s\ge 0, \end{aligned}$$

in particular

$$\operatorname{var}X(t) = \rho^2 e^{-2ct} + \tfrac{\sigma^2}{2c}(1-e^{-2ct})$$

if $c\ne 0$ and

$$\begin{aligned} \operatorname{cov}(X(t),X(s)) &= \rho^2 + \sigma^2(t\wedge s), \\ \operatorname{var}X(t) &= \rho^2 + \sigma^2 t \end{aligned}$$

if $c = 0$. The Gaussian system given by (39), at least for $\theta = 0$, is called an *Ornstein-Uhlenbeck proces* (see [1, Def. 16.4, p. 349]). If

$c > 0$, we find as limit distribution for $t \to \infty$ (with aid of Proposition 4.9) a Gaussian distribution with mean θ and variance $\frac{\sigma^2}{2c}$, which does not depend on the initial condition. If we choose the limit distribution as initial condition, i.e. $x = \theta$ and $\rho^2 = \frac{\sigma^2}{2c}$, then the formulas (40) show that the distribution of $X(t)$ is the same for every $t \geq 0$. Such a distribution is called a *stationary distribution* or *invariant measure* for the equation and the system $\{X(t)\}_{t\geq 0}$ is called a *stationary process.*

8. Kolmogorov equations

There is a remarkable connection between stochastic differential equations and deterministic partial differential equations. We will show this for the Langevin equation:

$$X(t) = x + c\int_0^t (\theta - X(s))ds + \sigma W(t), \quad t \geq 0. \tag{41}$$

Let $\{X(t,x)\}_{t\geq 0}$ be the solution to this equation with initial condition $x(\in \mathbb{R})$. From Proposition 6.2 we know that $X(t,x)$ is Gaussian with mean

$$\mathbb{E}X(t,x) = (x-\theta)e^{-ct} + \theta$$

and variance

$$\operatorname{var} X(t,x) = \begin{cases} \frac{\sigma^2}{2c}(1 - e^{-2ct}) & \text{if } c \neq 0, \\ \sigma^2 t & \text{if } c = 0. \end{cases}$$

Fix a $\xi \in \mathbb{R}$ and denote the characteristic function of $X(t,x)$ at ξ by $u(t,x)$. With aid of Proposition 4.2 we have if $c \neq 0$:

$$u(t,x) := \mathbb{E}e^{iX(t,x)\xi} = e^{i(x-\theta)e^{-ct}\xi - \frac{1}{2}\frac{\sigma^2}{2c}(1-e^{-2ct})\xi^2}, \quad t \geq 0, x \in \mathbb{R}.$$

Differentiation of u yields:

$$\begin{aligned} u_t &= \{ic(x-\theta)e^{-ct}\xi - \tfrac{1}{2}\sigma^2 e^{-2ct}\xi^2\}\,u, \\ u_x &= ie^{-ct}\xi u, \\ u_{xx} &= -e^{-2ct}\xi^2 u. \end{aligned}$$

Further, $u(0,x) = e^{ix\xi}$, $x \in \mathbb{R}$. Hence we obtain that

$$\begin{cases} u_t(t,x) &= \frac{1}{2}\sigma^2 u_{xx}(t,x) + c(\theta - x)u_x(t,x), \quad t \geq 0,\ x \in \mathbb{R}, \\ u(0,x) &= e^{ix\xi}, \qquad x \in \mathbb{R}. \end{cases}$$

In other words, we have that the function

$$u(t,x) := \mathbb{E}\varphi(X(t,x)) = \int_{\mathbb{R}} \varphi(y)\mu_{X(t,x)}(dy) \tag{42}$$

$$= \begin{cases} \displaystyle\int_{\mathbb{R}} \varphi(e^{-ct}x + (1-e^{-ct})\theta \pm y) \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2c}(1-e^{-2ct})}} e^{-\frac{1}{2}\frac{y^2}{\frac{\sigma^2}{2c}(1-e^{-2ct})}} dy \\ \text{if } c \neq 0, \\ \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{1}{2}\frac{y^2}{\sigma^2 t}} \qquad \text{if } c = 0, \end{cases}$$

with $\varphi(y) = e^{iy\xi}$, $y \in \mathbb{R}$, satisfies

$$\begin{cases} u_t &= \frac{1}{2}\sigma^2 u_{xx} + c(\theta - x)u_x, \quad t > 0,\ x \in \mathbb{R}, \\ u(0,x) &= \varphi(x), \qquad x \in \mathbb{R}. \end{cases} \tag{43}$$

These equations are called the *Kolmogorov equations* associated with (41). By direct verification one can show that the function u given by (42) satisfies (43) for any $\varphi \in BC^2(\mathbb{R})$. Furthermore, it can be shown that as $t \to 0$, $u(t,x) \to \varphi(x)$ uniformly on bounded intervals and if in addition $\varphi \in BUC(\mathbb{R})$ and $c = 0$, then uniformly on all of $\mathbb{R}$.

If $\sigma = 0$, then problem (43) is a first order partial differential equation of hyperbolic type. The corresponding stochastic differential equation (41) is then in fact deterministic and for all $\omega \in \Omega$, $X(t,x)(\omega)$ satisfies

$$\begin{cases} \frac{dX}{dt}(t) &= c(\theta - X(t)), \quad t \geq 0, \\ X(0) &= x. \end{cases}$$

So in this case, the above considerations reduce to the conclusion that $u(t,x) = \mathbb{E}\varphi(X(t,x)) = \varphi(X(t,x))$ is a solution to (43), which is actually the method of characteristics. If $\sigma \neq 0$, (43) is a second order parabolic equation and we can use the genuinely stochastic equation (41) to find a solution $u(t,x) = \mathbb{E}\varphi(X(t,x))$. This approach may be interpreted as a method of characteristics for the stochastic flow.

There is also a connection of the invariant measure for the Langevin equation introduced in § 7 with the Kolmogorov equations. Let us introduce the class

$$\mathcal{U} := \{u \in C((0,\infty) \times \mathbb{R}) \colon\ u_t, u_x, u_{xx} \text{ exist and are continuous on } (0,\infty) \times \mathbb{R}\},$$

and denote the differential operator at the right hand side of the Kolmogorov equation (43) by

$$\mathcal{L}u := \tfrac{1}{2}\sigma^2 u_{xx} + c(\theta - x)u_x, \quad u \in \mathcal{U}.$$

Assume that $c > 0$. If we multiply $\mathcal{L}u$ by a test function $v \in C_c^1(\mathbb{R})$ (i.e. v is a C^1-function with compact support) and integrate with respect to the invariant measure $N(\theta, \frac{\sigma^2}{2c})$ of the Langevin equation, then we obtain, with the abbreviation $p(x) := \frac{1}{\sqrt{2\pi \frac{\sigma^2}{2c}}} e^{-\frac{1}{2}\frac{(x-\theta)^2}{\frac{\sigma^2}{2c}}}$:

$$\begin{aligned}
&\int_{\mathbb{R}} \mathcal{L}(u) v N(\theta, \tfrac{\sigma^2}{2c})(dx) \\
&= \lim_{M,N\to\infty} \int_{-M}^{N} (\tfrac{1}{2}\sigma^2 u_{xx} + c(\theta - x)u_x) vp\, dx \\
&= \lim_{M,N\to\infty} \Big\{ \tfrac{1}{2}\sigma^2 \int_{-M}^{N} u_{xx} vp dx + \int_{-M}^{N} c(\theta - x) vp\, dx \Big\} \\
&= \lim_{M,N\to\infty} \Big\{ -\tfrac{1}{2}\sigma^2 \int_{-M}^{N} u_x (vp)_x\, dx + \int_{-M}^{N} c(\theta - x) u_x vp\, dx \Big\} \\
&= \lim_{M,N\to\infty} \Big\{ -\tfrac{1}{2}\sigma^2 \int_{-M}^{N} u_x v_x p\, dx \\
&\qquad + \int_{-M}^{N} u_x v(-\tfrac{1}{2}\sigma^2 p_x + c(\theta - x)p)\, dx \Big\} \\
&= -\tfrac{1}{2}\sigma^2 \int_{\mathbb{R}} u_x v_x p\, dx,
\end{aligned}$$

because

$$-\tfrac{1}{2}\sigma^2 p_x + c(\theta - x)p = 0.$$

Hence with aid of the invariant measure we find a weak formulation of the Kolmogorov equation:

$$\int_{\mathbb{R}} u_t vp\, dx + \tfrac{1}{2}\sigma^2 \int_{\mathbb{R}} u_x v_x p\, dx = 0, \quad \forall v \in C_c^1(\mathbb{R}).$$

Moreover, if $u, v \in \mathcal{U} \cap C_c^1(\mathbb{R})$, we have

$$\int_{\mathbb{R}} \mathcal{L}(u) vp\, dx = -\tfrac{1}{2}\sigma^2 \int_{\mathbb{R}} u_x v_x p\, dx = \int_{\mathbb{R}} u\mathcal{L}(v) p\, dx,$$

which means that the invariant measure 'symmetrizes' the differential operator $\mathcal{L}$. These observation can be used to prove uniqueness of solutions of (43) in a certain class. It is, however, not our goal to go further in that direction. We conclude with the remark that exploitation of the above ideas together yields a probabilistic method to solve and prove properties of deterministic partial differential equations.

Finally, we state as an exercise that the previous results concerning the Langevin equation can be generalized to $\mathbb{R}^d$.

EXERCISE 8.1: Let $d, n \in \mathbb{N}$. Let $W_1, \dots, W_n$ be n independent Brownian Motions (on Ω^n), and denote $W(t) := (W_1(t), \dots, W_n(t))^T$, $t \geq 0$. Let C be a $d \times d$-matrix, σ a $d \times n$-matrix, and $x, \theta \in \mathbb{R}^d$. Consider the stochastic differential equation

$$\text{“ } X'(t) = -C(\theta - X(t)) + \sigma W'(t) \text{ ”}, \quad t \geq 0,$$

which means

$$X(t) = x + \int_0^t C(\theta - X(s))ds + \sigma W(t), \quad t \geq 0. \tag{44}$$

At every t, X has to be a random vector $(X_1, \dots, X_d)$ on Ω^n.

1. Show that

$$X(t) = e^{-tC}x + (I - e^{-tC})\theta + \int_0^t e^{-(t-s)C}\sigma dW(s), \quad t \geq 0,$$

where the integral is a matrix-vector Stieltjes integral:

$$\left(\int_0^t e^{-(t-s)C}\sigma\, dW(s)\right)_k = \sum_{j=1}^n \int_0^t (e^{-(t-s)C}\sigma)_{kj} dW_j(s), \qquad k = 1, \dots, d.$$

(See [4].)

2. Show that $\{X(t)\}_{t\geq 0}$ is a Gaussian system with
$$\mathbb{E}X(t) = e^{-tC}x + (I - e^{-tC})\theta$$
and
$$\operatorname{cov}(X_i(t), X_j(s)) = \left(\int_0^{t\wedge s} e^{-(t-\tau)C}\sigma\sigma^T e^{-(s-\tau)C^T}\, d\tau \right)_{ij}, \quad i,j = 1,\ldots,d,$$
in particular,
$$\operatorname{cov}(X_i(t), X_j(t)) = \left(\int_0^{t} e^{-\tau C}\sigma\sigma^T e^{-\tau C^T}\, d\tau \right)_{ij}.$$

3. Show that the Kolmogorov equation associated with (44) is
$$u_t(t,x) = \tfrac{1}{2}\mathrm{Trace}\,(\sigma\sigma^T D^2 u) + < C(\theta - x), Du > \tag{45}$$
where
$$Du = (\tfrac{\partial u}{\partial x_1}, \ldots, \tfrac{\partial u}{\partial x_d})^T, \text{ and } (D^2u)_{ij} = \tfrac{\partial^2 u}{\partial x_i \partial x_j},\ i,j = 1,\ldots,d.$$

4. Verify that the function u given by
$$\begin{aligned} u(t,x) &= \mathbb{E}\varphi(X(t,x)) = \int_{\mathbb{R}^d} \varphi(y)\mu_{X(t,X)}(dy) \\ &= \int_{\mathbb{R}^d} \varphi(e^{-tC}x + (I - e^{-tC})\theta \pm y) N(0, Q_t)(dy), \end{aligned}$$
where
$$Q_t = \int_0^t e^{-sC}\sigma\sigma^T e^{-sC^T}\, ds, \quad t \geq 0,$$
and $\varphi \in BC^2(\mathbb{R}^d)$ satisfies (45) for $t > 0$.

Acknowledgements The authors want to express their thanks to S. Cerrai for reading the manuscript and correcting some errors. The first author would like to thank professor E. Mitidieri for his invitation to present in May 2000 at the University of Trieste a series of lectures on stochastic differential equations based on these notes.

References

[1] L. Breiman, *Probability*, Addison-Wesley, Reading, Mass., 1968.
[2] G. Da Prato, *Introduction to differential stochastic equations*, Appunti, Scuola Normale Superiore, Pisa, 1998, 2nd ed.
[3] G. Da Prato and J. Zabczyk, *Stochastic equations in infinite dimensions*, Cambridge University Press, 1992, (Encyclopedia of mathematics and its applications 45).
[4] G. Gripenberg, S.O. Londen, and O.J. Staffans, *Volterra integral and functional equations*, Cambridge University Press, Cambridge, 1990.
[5] P.R. Halmos, *Measure theory*, Van Nostrand, New York, 1950.
[6] P. Malliavin, *Integration and probability*, vol. (GTM 157), Springer-Verlag, New York, 1995, transl. from the French.
[7] H.L. Royden, *Real analysis*, Macmillan, New York, 1987, 3rd ed.
[8] A.N. Shiryayev, *Probability*, Springer, New York, 1996, transl. from the Russian, 2nd ed.
[9] K.R. Stromberg, *Probability for analysts*, Chapman and Hall, New York, 1994.
[10] D.W. Stroock and V.S.R. Srinivasa, *Multidimensional diffusion processes*, Springer, Berlin, 1979.
[11] A.E. Taylor, *General theory of functions and integration*, Blaisdell, New York, 1965.

Received August 21, 2000.