Chapter 2

Dynamics under the Influence of Stochastic Forces

Contents

2.1 Newton's Equation and Langevin's Equation

In this section we assume that the constituents of matter can be described classically. We are interested in reaction processes occuring in the bulk, either in physiological liquids, membranes or proteins. The atomic motion of these materials is described by the Newtonian equation of motion

$$m_i \, \frac{d^2}{dt^2} \, r_i \;\;=\;\; -\frac{\partial}{\partial r_i} \, V(r_1, \dots, r_N) \tag{2.1}$$

where $r_i \quad (i = 1, 2, \dots, N)$ describes the position of the i-th atom. The number N of atoms is, of course, so large that solutions of Eq. (2.1) for macroscopic systems are impossible. In microscopic systems like proteins the number of atoms ranges between 10^3 to 10^5, i.e., even in this case the solution is extremely time consuming.

However, most often only a few of the degrees of freedom are involved in a particular biochemical reaction and warrant an explicit theoretical description or observation. For example, in the case of transport one is solely interested in the position of the center of mass of a molecule. It is well known that molecular transport in condensed media can be described by phenomenological equations much simpler than Eq. (2.1), e.g., by the Einstein diffusion equation. The same holds true for reaction

processes in condensed media. In this case one likes to focus onto the reaction coordinate, e.g., on a torsional angle.

In fact, there exist successful descriptions of a small subset of degrees of freedom by means of Newtonian equations of motion with effective force fields and added frictional as well as (time dependent) fluctuating forces. Let us assume we like to consider motion along a small subset of the whole coordinate space defined by the coordinates $q_1, \ldots, q_M$ for $M \ll N$. The equations which model the dynamics in this subspace are then ($j = 1, 2, \ldots, M$)

$$\mu_j \, \frac{d^2}{dt^2} \, q_j \;\; = \;\; -\frac{\partial}{\partial q_j} \, W(q_1, \ldots, q_M) \; - \; \gamma_j \, \frac{d}{dt} \, q_j \; + \; \sigma_j \, \xi_j(t). \tag{2.2}$$

The first term on the r.h.s. of this equation describes the force field derived from an effective potential $W(q_1, \ldots, q_M)$, the second term describes the velocity ($\frac{d}{dt} q_j$) dependent frictional forces, and the third term the fluctuating forces $\xi_j(t)$ with coupling constants σ_j. $W(q_1, \ldots, q_M)$ includes the effect of the thermal motion of the remaining $n - M$ degrees of freedom on the motion along the coordinates $q_1, \ldots, q_M$.

Equations of type (2.2) will be studied in detail further below. We will not "derive" these equations from the Newtonian equations (2.1) of the bulk material, but rather show by comparision of the predictions of Eq. (2.1) and Eq. (2.2) to what extent the suggested phenomenological descriptions apply. To do so and also to study further the consequences of Eq. (2.2) we need to investigate systematically the solutions of stochastic differential equations.

2.2 Stochastic Differential Equations

We consider stochastic differential equations in the form of a first order differential equation

$$\partial_t \, \boldsymbol{x}(t) \;\; = \;\; \boldsymbol{A}[\boldsymbol{x}(t), t] \; + \; \mathbf{B}[\boldsymbol{x}(\mathbf{t}), \mathbf{t}] \cdot \boldsymbol{\eta}(\mathbf{t}) \tag{2.3}$$

subject to the initial condition

$$\boldsymbol{x}(0) \;\; = \;\; \boldsymbol{x}_0. \tag{2.4}$$

In this equation $\boldsymbol{A}[\boldsymbol{x}(t), t]$ represents the so-called drift term and $\mathbf{B}[\boldsymbol{x}(\mathbf{t}), \mathbf{t}] \cdot \boldsymbol{\eta}(\mathbf{t})$ the noise term which will be properly characterized further below. Without the noise term, the resulting equation

$$\partial_t \boldsymbol{x}(t) \;\; = \;\; \boldsymbol{A}[\boldsymbol{x}(t), t]. \tag{2.5}$$

describes a deterministic drift of particles along the field $\boldsymbol{A}[\boldsymbol{x}(t), t]$.

Equations like (2.5) can actually describe a wide variety of phenomena, like chemical kinetics or the firing of neurons. Since such systems are often subject to random perturbations, noise is added to the deterministic equations to yield associated stochastic differential equations. In such cases as well as in the case of classical Brownian particles, the noise term $\mathbf{B}[\boldsymbol{x}(\mathbf{t}), \mathbf{t}] \cdot \boldsymbol{\eta}(\mathbf{t})$ needs to be specified on the basis of the underlying origins of noise. We will introduce further below several mathematical models of noise and will consider the issue of constructing suitable noise terms throughout this book. For this purpose, one often adopts a heuristic approach, analysing the noise from observation or from a numerical simulation and selecting a noise model with matching characteristics. These characteristics are introduced below.

Before we consider characteristics of the noise term $\boldsymbol{\eta}(t)$ in (2.3) we like to demonstrate that the one-dimensional Langevin equation (2.2) of a classical particle, written here in the form

$$\mu \, \ddot{q} \;\; = \;\; f(q) \; - \; \gamma \, \dot{q} \; + \; \sigma \, \xi(t) \tag{2.6}$$

is a special case of (2.3). In fact, defining $\boldsymbol{x} \in \mathbb{R}^2$ with components $x_1 = m$, and $\dot{q}$, $x_2 = m\,q$ reproduces Eq. (2.3) if one defines

$$\boldsymbol{A}[\boldsymbol{x}(t), t] = \begin{pmatrix} f(x_2/m) - \gamma\, x_1/m \\ x_1 \end{pmatrix}, \quad \mathbf{B}[\boldsymbol{x}(\mathbf{t}), \mathbf{t}] = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}, \text{ and } \quad \boldsymbol{\eta}(\mathbf{t}) = \begin{pmatrix} \xi(t) \\ 0 \end{pmatrix}. \quad (2.7)$$

The noise term represents a stochastic process. We consider only the factor $\boldsymbol{\eta}(t)$ which describes the essential time dependence of the noise source in the different degrees of freedom. The matrix $\mathbf{B}[\boldsymbol{x}(\mathbf{t}), \mathbf{t}]$ describes the amplitude and the correlation of noise between the different degrees of freedom.

2.3 How to Describe Noise

We are now embarking on an essential aspect of our description, namely, how stochastic aspects of noise $\boldsymbol{\eta}(t)$ are properly accounted for. Obviously, a particular realization of the time-dependent process $\boldsymbol{\eta}(t)$ does not provide much information. Rather, one needs to consider the *probability* of observing a certain sequence of noise values $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots$ at times $t_1, t_2, \dots$. The essential information is entailed in the *conditional probabilities*

$$p(\boldsymbol{\eta}_1, t_1; \boldsymbol{\eta}_2, t_2; \dots | \boldsymbol{\eta}_0, t_0; \boldsymbol{\eta}_{-1}, t_{-1}; \dots) \quad (2.8)$$

when the process is assumed to generate noise at fixed times t_i, $\quad t_i < t_j$ for $i < j$. Here $p(\,|\,)$ is the probability that the random variable $\boldsymbol{\eta}(t)$ assumes the values $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots$ at times $t_1, t_2, \dots$, if it had previously assumed the values $\boldsymbol{\eta}_0, \boldsymbol{\eta}_{-1}, \dots$ at times $t_0, t_{-1}, \dots$.

An important class of random processes are so-called *Markov processes* for which the conditional probabilities depend only on $\boldsymbol{\eta}_0$ and t_0 and not on earlier occurrences of noise values. In this case holds

$$p(\boldsymbol{\eta}_1, t_1; \boldsymbol{\eta}_2, t_2; \dots | \boldsymbol{\eta}_0, t_0; \boldsymbol{\eta}_{-1}, t_{-1}; \dots) = p(\boldsymbol{\eta}_1, t_1; \boldsymbol{\eta}_2, t_2; \dots | \boldsymbol{\eta}_0, t_0). \quad (2.9)$$

This property allows one to factorize $p(\,|\,)$ into a sequence of consecutive conditional probabilities.

$$\begin{aligned} p(\boldsymbol{\eta}_1, t_1; \boldsymbol{\eta}_2, t_2; \dots | \boldsymbol{\eta}_0, t_0) &= p(\boldsymbol{\eta}_2, t_2; \boldsymbol{\eta}_3, t_3; \dots | \boldsymbol{\eta}_1, t_1)\, p(\boldsymbol{\eta}_1, t_1 | \boldsymbol{\eta}_0, t_0) \\ &= p(\boldsymbol{\eta}_3, t_3; \boldsymbol{\eta}_4, t_4; \dots | \boldsymbol{\eta}_2, t_2)\, p(\boldsymbol{\eta}_2, t_2 | \boldsymbol{\eta}_1, t_1)\, p(\boldsymbol{\eta}_1, t_1 | \boldsymbol{\eta}_0, t_0) \\ &\vdots \end{aligned} \quad (2.10)$$

The unconditional probability for the realization of $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots$ at times $t_1, t_2, \dots$ is

$$p(\boldsymbol{\eta}_1, t_1; \boldsymbol{\eta}_2, t_2; \dots) = \sum_{\boldsymbol{\eta}_0} p(\boldsymbol{\eta}_0, t_0)\quad p(\boldsymbol{\eta}_1, t_1 | \boldsymbol{\eta}_0, t_0)\quad p(\boldsymbol{\eta}_2, t_2 | \boldsymbol{\eta}_1, t_1)\quad \cdots \quad (2.11)$$

where $p(\boldsymbol{\eta}_0, t_0)$ is the unconditional probability for the appearence of $\boldsymbol{\eta}_0$ at time t_0. One can conclude from Eq. (2.11) that a knowledge of $p(\boldsymbol{\eta}_0, t_0)$ and $p(\boldsymbol{\eta}_i, t_i | \boldsymbol{\eta}_{i-1}, t_{i-1})$ is sufficient for a complete characterization of a Markov process.

Before we proceed with three important examples of Markov processes we will take a short detour and give a quick introduction on mathematical tools that will be useful in handling probability distributions like $p(\boldsymbol{\eta}_0, t_0)$ and $p(\boldsymbol{\eta}_i, t_i | \boldsymbol{\eta}_{i-1}, t_{i-1})$.

Characteristics of Probability Distributions

In case of a one-dimensional random process $\boldsymbol{\eta}$, denoted by $\eta(t)$, $p(\eta, t)\, d\eta$ gives the probability that $\eta(t)$ assumes a value in the interval $[\eta, \eta + d\eta]$ at time t.

Let $f[\eta(t)]$ denote some function of $\eta(t)$. $f[\eta(t)]$ could represent some observable of interest, e.g., $f[\eta(t)] = \eta^2(t)$. The average value measured for this observable at time t is then

$$\Big\langle f[\eta(t)] \Big\rangle \;\; = \;\; \int_\Omega d\eta \; f[\eta] \; p(\eta, t) \; . \tag{2.12}$$

Here Ω denotes the interval in which random values of $\eta(t)$ arise. The notation $\langle \cdots \rangle$ on the left side of (2.12) representing the average value is slightly problematic. The notation of the average should include the probability distribution $p(\eta, t)$ that is used to obtain the average. Misunderstandings can occur,

- if $f[\eta(t)] = 1$ and hence any reference to η and $p(\eta, t)$ is lost,

- if dealing with more than one random variable, and if thus it becomes unclear over which variable the average is taken and,

- if more than one probability distribution $p(\eta, t)$ are under consideration and have to be distinguished.

We will circumvent all of these ambiguities by attaching an index to the average $\langle \cdots \rangle$ denoting the corresponding random variable(s) and probability distribution(s), if needed. In general however, the simple notation adopted poses no danger since in most contexts the random variable and distribution underlying the average are self-evident.

For simplicity we now deal with a one-dimensional random variable η with values on the complete real axis, hence $\Omega = \mathbb{R}$. In probability theory the Fourier-transform $G(s, t)$ of $p(\eta, t)$ is referred to as the *characteristic function* of $p(\eta, t)$.

$$G(s, t) \;\; = \;\; \int_{-\infty}^{+\infty} d\eta \; p(\eta, t) \, e^{i\, s\, \eta} \; . \tag{2.13}$$

Since the Fourier transform can be inverted to yield $p(\tilde{\eta}, t)$

$$p(\tilde{\eta}, t) \;\; = \;\; \frac{1}{2\pi} \int_{-\infty}^{+\infty} ds \; G(s, t) \, e^{-i\, s\, \tilde{\eta}} \; , \tag{2.14}$$

$G(s, t)$ contains all information on $p(\eta, t)$.

The characteristic function can be interpreted as an average of $f[\eta(t)] = e^{i\, s\, \eta(t)}$, and denoted by

$$G(s, t) \;\; = \;\; \Big\langle e^{i\, s\, \eta(t)} \Big\rangle \; . \tag{2.15}$$

Equation (2.15) prompts one to consider the Taylor expansion of (2.15) for (is) around 0:

$$G(s, t) \;\; = \;\; \sum_{n=0}^{\infty} \Big\langle \eta^n(t) \Big\rangle \frac{(i\, s)^n}{n!} \tag{2.16}$$

where

$$\Big\langle \eta^n(t) \Big\rangle \;\; = \;\; \int d\eta \; \eta^n \; p(\eta, t) \tag{2.17}$$

are the so-called *moments* of $p(\eta, t)$. One can conclude from (2.14, 2.16, 2.17) that the moments $\langle \eta^n(t) \rangle$ completely characterize $p(\eta, t)$.

The moments $\langle \eta^n(t) \rangle$ can be gathered in a statistical analysis as averages of powers of the stochastic variable $\eta(t)$. Obviously, it is of interest to employ averages which characterize a distribution $p(\eta, t)$ as succinctly as possible, i.e., through the smallest number of averages. Unfortunately moments $\langle \eta^n(t) \rangle$ of all orders of n contain significant information about $p(\eta, t)$.

There is another, similar, but more useful scheme to describe probability distributions $p(\eta, t)$; the *cumulants* $\langle\!\langle \eta^n(t) \rangle\!\rangle$. As moments are generated by the characteristic function $G(s, t)$, cumulants are generated by the logarithm of the characteristic function $\log[G(s, t)]$

$$\log[G(s, t)] \quad = \quad \sum_{n=1}^{\infty} \langle\!\langle \eta^n(t) \rangle\!\rangle \, \frac{(i\,s)^n}{n!} \; . \tag{2.18}$$

Cumulants can be expressed in terms of $\langle \eta^n(t) \rangle$ by taking the logarithm of equation (2.16) and comparing the result with (2.18). The first three cumulants are

$$\langle\!\langle \eta^1(t) \rangle\!\rangle \quad = \quad \langle \eta^1(t) \rangle \; , \tag{2.19}$$

$$\langle\!\langle \eta^2(t) \rangle\!\rangle \quad = \quad \langle \eta^2(t) \rangle \; - \; \langle \eta^1(t) \rangle^2 \; , \tag{2.20}$$

$$\langle\!\langle \eta^3(t) \rangle\!\rangle \quad = \quad \langle \eta^3(t) \rangle \; - \; 3 \langle \eta^2(t) \rangle \langle \eta^1(t) \rangle \; + \; 2 \langle \eta^1(t) \rangle^3 \; . \tag{2.21}$$

These expressions reveal that the first cumulant is equal to the average of the stochastic variable and the second cumulant is equal to the variance[1] . The higher orders of cumulants contain less information about $p(\eta, t)$ than lower ones. In fact it can be shown, that in the frequently arising case of probabilities described by *Gaussian* distributions (the corresponding random processes are called Gaussian) all, but the first and second-order cumulants vanish. For non-Gaussian distributions, though, all cumulants are non-zero as stated in the theorem of Marcienkiewicz [23]). Nevertheless, cumulants give a more succint description of $p(\eta, t)$ than moments do, dramatically so in case of Gaussian processes. This is not the only benefit as we will see considering scenarios with more than one random variable $\eta(t)$.

We now proceed to probability distributions involving two random variables as they arise in case of $\boldsymbol{\eta}(t) \in \mathbb{R}^2$ or if one looks at single random process $\eta(t) \in \mathbb{R}$ at two different times. Both cases are treated by the same tools, however, names and notation differ. We will adopt a notation suitable for a single random process $\eta(t)$ observed at two different times t_0 and t_1, and governed by the unconditional probability distribution $p(\eta_0, t_0; \eta_1, t_1)$. $p(\eta_1, t_1; \eta_0, t_0)\, d\eta_1 d\eta_0$ gives the probability that $\eta(t)$ assumes a value in the interval $[\eta_0, \eta_0 + d\eta_0]$ at time t_0, and a value $[\eta_1, \eta_1 + d\eta_1]$ at time t_1.

As stated in equation (2.11) $p(\eta_0, t_0; \eta_1, t_1)$ can be factorized into the unconditional probability $p(\eta_0, t_0)$ and the conditional probability $p(\eta_0, t_0 | \eta_1, t_1)$. Finding η_0 *and* η_1 is just as probable as first obtaining η_0 and then finding η_1 under the condition of having found η_0 already. The probability of the later is given by the conditional probability $p(\eta_1, t_1 | \eta_0, t_0)$. Hence one can write,

$$p(\eta_0, t_0; \eta_1, t_1) \quad = \quad p(\eta_1, t_1 | \eta_0, t_0) \, p(\eta_0, t_0) \; . \tag{2.22}$$

In the case that η_1 is statistically independent of η_0 the conditional probability $p(\eta_1, t_1 | \eta_0, t_0)$ does not depend on η_0 or t_0 and we obtain

$$p(\eta_1, t_1 | \eta_0, t_0) \quad = \quad p(\eta_1, t_1) \; , \tag{2.23}$$

---

[1]The variance is often written as the average square deviation from the mean $\langle (\eta(t) - \langle \eta(t) \rangle)^2 \rangle$ which is equivalent to $\langle \eta^2(t) \rangle - \langle \eta(t) \rangle^2$.

and, hence,

$$p(\eta_0, t_0; \eta_1, t_1) \;=\; p(\eta_1, t_1)\, p(\eta_0, t_0) \;. \tag{2.24}$$

In order to characterize $p(\eta_0, t_0; \eta_1, t_1)$ and $p(\eta_0, t_0|\eta_1, t_1)$ one can adopt tools similar to those introduced to characterize $p(\eta_0, t_0)$. Again one basic tool is the average, now the average of a function $g[\eta(t_0), \eta(t_1)]$ with two random variables. Note, that $g[\eta(t_0), \eta(t_1)]$ depends on two random values η_0 and η_1 rendered by a single random process $\eta(t)$ at times t_0 and t_1.

$$\begin{aligned}
\left\langle g[\eta(t_0)\,\eta(t_1)] \right\rangle &\;=\; \int_\Omega d\eta_1 \int_\Omega d\eta_0 \; g[\eta_1, \eta_0]\, p(\eta_0, t_0; \eta_1, t_1) \\
&\;=\; \int d\eta_0 \; p(\eta_0, t_0) \int d\eta_1 \; g[\eta_1, \eta_0]\, p(\eta_1, t_1|\eta_0, t_0) \;.
\end{aligned} \tag{2.25}$$

The same advise of caution as for the average of one random variable applies here aswell. The characteristic function is the Fourier-transform of $p(\eta_0, t_0; \eta_1, t_1)$ in η_0 and η_1.

$$\begin{aligned}
G(s_0, t_0; s_1, t_1) &\;=\; \int d\eta_0 \; p(\eta_0, t_0) \int d\eta_1 \; p(\eta_1, t_1|\eta_0, t_0) \, \exp\!\big[i\,(s_0\,\eta_0 \,+\, s_1\,\eta_1)\big] \\
&\;=\; \left\langle e^{i\,(s_0\eta(t_0) + s_1\eta(t_1))} \right\rangle \;.
\end{aligned} \tag{2.26}$$

This can be written as the average [c.f. Eq. (2.25)]

$$G(s_0, t_0; s_1, t_1) \;=\; \left\langle e^{i(s_0\eta(t_0) + s_1\eta(t_1))} \right\rangle \;. \tag{2.27}$$

The coefficients of a Taylor expansion of $G(s_0, t_0; s_1, t_1)$ in (is_0) and (is_1), defined through

$$G(s_0, t_0; s_1, t_1) \;=\; \sum_{n_0, n_1 = 0}^{\infty} \left\langle \eta^{n_0}(t_0)\, \eta^{n_1}(t_1) \right\rangle \frac{(i\,s_0)^{n_0}}{n_0!} \frac{(i\,s_1)^{n_1}}{n_1!} \tag{2.28}$$

$$\left\langle \eta^{n_0}(t_0)\, \eta^{n_1}(t_1) \right\rangle \;=\; \int d\eta_0 \; \eta_0^{n_0}\, p(\eta_0, t_0) \int d\eta_1 \; \eta_1^{n_1}\, p(\eta_1, t_1|\eta_0, t_0) \;. \tag{2.29}$$

are called *correlations* or *correlation functions*; the later if one is interested in the time dependency. Cumulants are defined through the expansion

$$\log\big[G(s_0, t_0; s_1, t_1)\big] \;=\; \sum_{n_0, n_1 = 0}^{\infty} \left\langle\!\!\left\langle \eta^{n_0}(t_0)\, \eta^{n_1}(t_1) \right\rangle\!\!\right\rangle \frac{(i\,s_0)^{n_0}}{n_0!} \frac{(i\,s_1)^{n_1}}{n_1!} \;. \tag{2.30}$$

These multi-dimensional cumulants can also be expressed in terms of correlation functions and moments. For example, one can show

$$\left\langle\!\!\left\langle \eta(t_0)\,\eta(t_1) \right\rangle\!\!\right\rangle \;=\; \left\langle \eta(t_0)\,\eta(t_1) \right\rangle \,-\, \left\langle \eta(t_0) \right\rangle \left\langle \eta(t_1) \right\rangle \;. \tag{2.31}$$

Cumulants are particularly useful if one has to consider the sum of statistically independent random values, for example the sum

$$\sigma \;=\; \eta_0 + \eta_1 \;. \tag{2.32}$$

The probability $p(\sigma, t_0, t_1)$ for a certain value σ to occur is associated with the characteristic function

$$G_\sigma(r, t_0, t_1) \;=\; \int d\sigma \; p(\sigma, t_0, t_1)\, e^{i\,r\sigma} \;. \tag{2.33}$$

$p(\sigma, t_0, t_1)$ can be expressed as

$$p(\sigma, t_0, t_1) \;=\; \iint d\eta_0 \, d\eta_1 \; p(\eta_0, t_0; \eta_1, t_1) \; \delta(\eta_0 + \eta_1 - \sigma) \;. \tag{2.34}$$

Accordingly, one can write

$$G_\sigma(r, t_0, t_1) \;=\; \int d\sigma \iint d\eta_0 \, d\eta_1 \; p(\eta_0, t_0; \eta_1, t_1) \; \delta(\eta_0 + \eta_1 - \sigma) \; e^{i \, r \sigma} \;. \tag{2.35}$$

Integrating over σ results in

$$G_\sigma(r, t_0, t_1) \;=\; \iint d\eta_0 \, d\eta_1 \; p(\eta_0, t_0; \eta_1, t_1) \; e^{i \, r(\eta_0 + \eta_1)} \;. \tag{2.36}$$

This expression can be equated to the characteristic function $G_{\eta_0 \eta_1}(r, t_0; r, t_1)$ of the two summands η_0 and η_1, where

$$G_{\eta_0 \eta_1}(s_0, t_0; s_1, t_1) \;=\; \iint d\eta_0 \, d\eta_1 \; p(\eta_0, t_0; \eta_1, t_1) \; e^{i \, (s_0 \eta_0 + s_1 \eta_1)} \;. \tag{2.37}$$

The statistical independence of η_0 and η_1 in (2.32) is expressed by equation (2.24) as $p(\eta_0, t_0; \eta_1, t_1) = p(\eta_0, t_0) \, p(\eta_1, t_1)$ and one can write

$$G_{\eta_0 \eta_1}(s_0, t_0; s_1, t_1) \;=\; \int d\eta_0 \; p(\eta_0, t_0) \; e^{i \, s_0 \eta_0} \int d\eta_1 \; p(\eta_1, t_1) \; e^{i \, s_1 \eta_1} \tag{2.38}$$

from which follows

$$G_{\eta_0 \eta_1}(s_0, t_0; s_1, t_1) \;=\; G_{\eta_0}(s_0, t_0) \, G_{\eta_1}(s_1, t_1) \;, \tag{2.39}$$

and, hence,

$$\log[G_{\eta_0 \eta_1}(s_0, t_0; s_1, t_1)] \;=\; \log[G_{\eta_0}(s_0, t_0)] \;+\; \log[G_{\eta_1}(s_1, t_1)] \;. \tag{2.40}$$

Taylor-expansion leads to the cumulant identity

$$\left\langle\!\!\left\langle \eta^{n_0}(t_0) \; \eta^{n_1}(t_1) \right\rangle\!\!\right\rangle \;=\; 0 \;, \quad \forall \; n_0, n_1 \geq 1 \;. \tag{2.41}$$

One can finally apply $G_\sigma(r, t_0, t_1) = G_{\eta_0 \eta_1}(r, t_0; r, t_1)$, see (2.36) and (2.37) and compare the Taylor coefficients.

$$\left\langle\!\!\left\langle \big(\eta(t_0) + \eta(t_1)\big)^n \right\rangle\!\!\right\rangle \;=\; \sum_{n_0, n_1} \left\langle\!\!\left\langle \eta^{n_0}(t_0) \; \eta^{n_1}(t_1) \right\rangle\!\!\right\rangle \frac{n!}{n_0! \, n_1!} \; \delta(n_0 + n_1 - n) \;. \tag{2.42}$$

According to equation (2.41) all but the two summands with $(n_0 = n, n_1 = 0)$ and $(n_0 = 0, n_1 = n)$ disappear and we deduce

$$\left\langle\!\!\left\langle \big(\eta(t_0) + \eta(t_1)\big)^n \right\rangle\!\!\right\rangle \;=\; \left\langle\!\!\left\langle \eta^n(t_0) \right\rangle\!\!\right\rangle \;+\; \left\langle\!\!\left\langle \eta^n(t_1) \right\rangle\!\!\right\rangle \;. \tag{2.43}$$

This result implies that cumulants of any order are simply added if one accumulates the corresponding statistically independent random variables, hence the name *cumulant*. For an arbitrary number of statistically independent random variables η_j or even continuously many $\eta(t)$ one can write

$$\left\langle\!\!\left\langle \Big(\sum_j \eta_j\Big)^n \right\rangle\!\!\right\rangle \;=\; \sum_j \left\langle\!\!\left\langle \eta_j^n \right\rangle\!\!\right\rangle \qquad \text{and} \tag{2.44}$$

$$\left\langle\!\!\left\langle \Big(\int dt \; \eta(t)\Big)^n \right\rangle\!\!\right\rangle \;=\; \int dt \; \left\langle\!\!\left\langle \eta^n(t) \right\rangle\!\!\right\rangle \;, \tag{2.45}$$

properties, which will be utilized below.
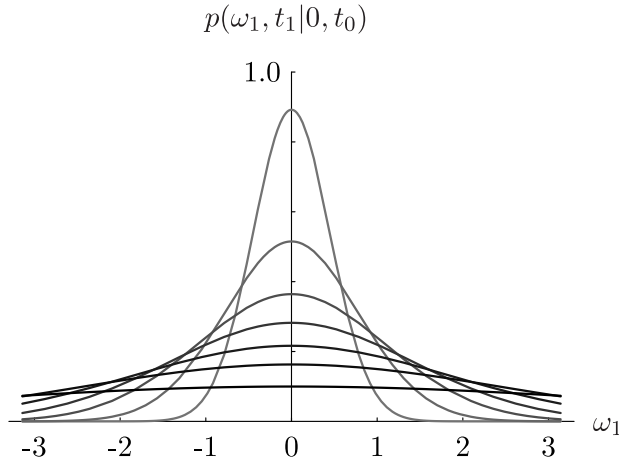
$$p(\omega_1, t_1 | 0, t_0)$$

Figure 2.1: The probability density distribution (2.47) of a Wiener process for $D = 1$ in arbitrary temporal and spatial units. The distribution (2.47) is shown for $\omega_0 = 0$ and $(t_1 - t_0) = 0.1, 0.3, 0.6, 1.0, 1.7, 3.0,$ and 8.0.

Wiener Process

We will now furnish concrete, analytical expressions for the probabilities characterizing three important Markov processes. We begin with the so-called *Wiener process*. This process, described by $\omega(t)$ for $t \geq 0$, is characterized by the probability distributions

$$p(\omega_0, t_0) \;=\; \frac{1}{\sqrt{4\pi D t_0}} \; \exp\left(- \frac{\omega_0^2}{4 D t_0} \right) , \tag{2.46}$$

$$p(\omega_1, t_1 | \omega_0, t_0) \;=\; \frac{1}{\sqrt{4\pi D \, \Delta t}} \; \exp\left(- \frac{(\Delta\omega)^2}{4 D \, \Delta t} \right) , \tag{2.47}$$

$$\text{with} \quad \Delta\omega \;=\; (\omega_1 - \omega_0) ,$$
$$\Delta t \;=\; t_1 - t_0 .$$

The probabilities (see Figure 2.1) are parameterized through the constant D, referred to as the diffusion constant, since the probability distributions $p(\omega_0, t_0)$ and $p(\omega_1, t_1 | \omega_0, t_0)$ are solutions of the diffusion equation (3.13) discussed extensively below. The Wiener process is homogeneous in time and space, which implies that the conditional transition probability $p(\omega_1, t_1 | \omega_0, t_0)$ depends only on the relative variables $\Delta\omega$ and Δt. Put differently, the probability $p(\Delta\omega, \Delta t)$ for an increment $\Delta\omega$ to occur is independent of the current state of the Wiener process $\omega(t)$. The probability is

$$p(\Delta\omega, \Delta t) \;=\; p(\omega_0 + \Delta\omega, t_0 + \Delta t | \omega_0, t_0) \;=\; \frac{1}{\sqrt{4\pi D \, \Delta t}} \; \exp\left(- \frac{(\Delta\omega)^2}{4 D \, \Delta t} \right) . \tag{2.48}$$

Characteristic Functions, Moments, Correlation Functions and Cumulants for the Wiener Process

In case of the Wiener process simple expressions can be provided for the characteristics introduced above, i.e., for the characteristic function, moments and cumulants. Combining (2.48) and (2.13) one obtains for the characteristic function

$$G(s, t) \;=\; e^{-D t s^2} . \tag{2.49}$$

A Taylor expansion allows one to identify the moments[2]

$$\left\langle \omega^n(t) \right\rangle \;\; = \;\; \begin{cases} 0 & \text{for odd } n, \\ (n-1)!! \, (2\,D\,t)^{n/2} & \text{for even } n, \end{cases} \tag{2.50}$$

The definition (2.18) and (2.49) leads to the expression for the cumulants

$$\left\langle\!\left\langle \omega^n(t) \right\rangle\!\right\rangle \;\; = \;\; \begin{cases} 2\,D\,t & \text{for } n = 2, \\ 0 & \text{otherwise .} \end{cases} \tag{2.51}$$

For the two-dimensional characteristic functions one can derive, using (2.47) and (2.26)

$$G(s_0, t_0; s_1, t_1) \;\; = \;\; \exp\left[-D \left(s_0^2\, t_0 + s_1^2\, t_1 + 2\, s_0\, s_1 \min(t_0, t_1) \right) \right]. \tag{2.52}$$

From this follow the correlation functions

$$\left\langle \omega^{n_1}(t_1)\, \omega^{n_0}(t_0) \right\rangle \;\; = \;\; \begin{cases} 0 & \text{for odd } (n_0 + n_1), \\ 2\,D\,\min(t_1, t_0) & \text{for } n_0 = 1 \text{ and } n_1 = 1, \\ 12\,D^2\, t_0 \min(t_1, t_0) & \text{for } n_0 = 1 \text{ and } n_1 = 3, \\ 4\,D^2 \left(t_0\, t_1 + 2 \min^2(t_1, t_0) \right) & \text{for } n_0 = 2 \text{ and } n_1 = 2, \\ \dots & \end{cases} \tag{2.53}$$

and, using the definition (2.30), the cumulants

$$\left\langle\!\left\langle \omega^{n_1}(t_1)\, \omega^{n_0}(t_0) \right\rangle\!\right\rangle \;\; = \;\; \begin{cases} 2\,D\,\min(t_1, t_0) & \text{for } n_0 = n_1 = 1, \\ 0 & \text{otherwise for } n_0, n_1 \neq 0 . \end{cases} \tag{2.54}$$

The Wiener Process as the Continuum Limit of a Random Walk on a Lattice

The Wiener process is closely related to a *random walk* on a one-dimensional lattice with lattice constant a. A n-step walk on a lattice is performed in discrete times steps $t_j = j\tau$, with $j = 0, 1, 2, \dots, n$. The walk may start at an arbitrary lattice site x_0. One can choose this starting position as the origin of the coordinate system so that one can set $x_0 = 0$. The lattice sites are then located at $x_i = ia$, $i \in \mathbb{Z}$.

At each time step the random walker moves with equal probability to the neighboring right or left lattice site. Thus, after the first step with $t = \tau$ one will find the random walker at $x = \pm a$, i.e. at site $x_{\pm 1}$ with probabitlity $P(\pm a, \tau) = \frac{1}{2}$. For a two-step walk the following pathes are possible:

path 1 :   two steps to the left,
path 2 :   one step to the left and then one step to the right,
path 3 :   one step to the right and then one step to the left,
path 4 :   two steps to the right.

Each path has a probability of $\frac{1}{4}$, a factor $\frac{1}{2}$ for each step. Pathes 2 and 3 both terminate at lattice site x_0. The probability to find a random walker after two step at position $x_0 = 0$ is therefore $P(0, 2\tau) = \frac{1}{2}$. The probabilities for lattice sites $x_{\pm 2}$ reached via path 1 and 4 respectively are simply $P(\pm 2a, 2\tau) = \frac{1}{4}$.

---

[2]The double factorial $n!!$ for positive $n \in \mathbb{N}$ denotes the product $n(n-2)(n-4)\dots 1$ for odd n and $n(n-2)(n-4)\dots 2$ for even n.

For an n-step walk one can proceed like this suming over all possible pathes that terminate at a given lattice site x_i. Such a summation yields the probability $P(ia, n\tau)$. However, to do so effectively a more elegant mathematical description is appropriate. We denote a step to the right by an operator $\mathcal{R}$, and a step to the left by an operator $\mathcal{L}$. Consequently a single step of a random walker is given by $\frac{1}{2}(\mathcal{L} + \mathcal{R})$, the factor $\frac{1}{2}$ denoting the probabilty for each direction. To obtain a n-step walk the above operator $\frac{1}{2}(\mathcal{L} + \mathcal{R})$ has to be iterated n times. For a two-step walk one gets $\frac{1}{4}(\mathcal{L} + \mathcal{R}) \circ (\mathcal{L} + \mathcal{R})$. Expanding this expression results in $\frac{1}{4}(\mathcal{L}^2 + \mathcal{L} \circ \mathcal{R} + \mathcal{R} \circ \mathcal{L} + \mathcal{R}^2)$. Since a step to the right and then to the left amounts to the same as a step first to the left and then to the right, it is safe to assume that $\mathcal{R}$ and $\mathcal{L}$ commute. Hence one can write $\frac{1}{4}\mathcal{L}^2 + \frac{1}{2}\mathcal{L} \circ \mathcal{R} + \frac{1}{4}\mathcal{R}^2$. As the operator expression $\mathcal{L}^p \circ \mathcal{R}^q$ stands for p steps to the left and q steps to the right one can deduce that $\mathcal{L}^p \circ \mathcal{R}^q$ represents the lattice site x_{q-p}. The coefficients are the corresponding probabilties $P\big((q-p)a, (q+p)\tau\big)$.

The algebraic approach above proofs useful, since one can utilize the well known binomial formula

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \, y^{n-k} \ . \tag{2.55}$$

One can write

$$\left[\frac{1}{2}(\mathcal{L} + \mathcal{R}) \right]^n = \left(\frac{1}{2}\right)^n \sum_{k=0}^{n} \binom{n}{k} \underbrace{\mathcal{L}^k \, \mathcal{R}^{n-k}}_{=x_{2k-n}} , \tag{2.56}$$

and obtains as coefficients of x_i the probabilities

$$P(i\,a, n\,\tau) = \frac{1}{2^n} \binom{n}{\frac{n+i}{2}} \ . \tag{2.57}$$

One can express (2.57) as

$$P(x, t) = \frac{1}{2^{t/\tau}} \binom{t/\tau}{\frac{t}{2\tau} + \frac{x}{2a}} \ . \tag{2.58}$$

The moments of the discrete probability distribution $P(x, t)$ are

$$
\begin{aligned}
\left\langle x^n(t) \right\rangle &= \sum_{x=-\infty}^{\infty} x^n \, P(x, t) \\
&= \begin{cases}
0 & \text{for odd } n, \\
a^2 \, \frac{t}{\tau} & \text{for } n = 2, \\
a^4 \, \frac{t}{\tau} \left(3\frac{t}{\tau} - 2 \right) & \text{for } n = 4, \\
a^6 \, \frac{t}{\tau} \left(15 \left(\frac{t}{\tau}\right)^2 - 30 \frac{t}{\tau} + 16 \right) & \text{for } n = 6, \\
\cdots \ .
\end{cases}
\end{aligned}
\tag{2.59}
$$

We now want to demonstrate that in the continuum limit a random walk reproduces a Wiener process. For this purpose we show that the unconditional probability distributions of both processes match. We do not consider conditional probabilities $p(x_1, t_1 | x_0, t_0)$ as they equal unconditional probabilities $p(x_1 - x_0, t_1 - t_0)$ in both cases; in a Wiener process aswell as in a random walk.

To turn the discrete probability distribution (2.58) into a continuous probability density distribution one considers adjacent bins centered on every lattice site that may be occupied by a random walker.
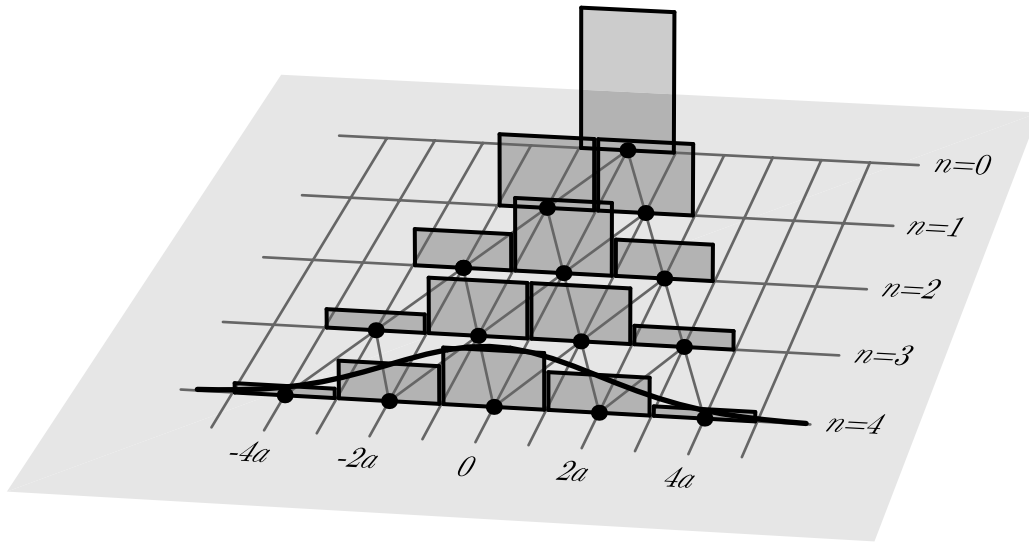
Figure 2.2: The probability density distributions (2.60) for the first four steps of a random walk on a discrete lattice with lattice constant a are shown. In the fourth step the continuous approximation (2.63) is superimposed.

Note, that only every second lattice site can be reached after a particular number of steps. Thus, these adjacent bins have a base length of $2a$ by which we have to divide $P(x,t)$ to obtain the probability density distribution $p(x,t)$ in these bins (see Figure 2.2).

$$p(x,t)\ dx\ \ =\ \ \frac{1}{2a}\ \frac{1}{2^{t/\tau}}\ \binom{t/\tau}{\frac{t}{2\tau}+\frac{x}{2a}}\ dx\ . \tag{2.60}$$

We then rescale the lattice constant a and the length τ of the time intervals to obtain a continuous description in time and space. However, τ and a need to be rescaled differently, since the spatial extension of the probability distribution $p(x,t)$, characterized by it's standard deviation

$$\sqrt{\left\langle\!\left\langle x^2(t)\right\rangle\!\right\rangle}\ \ =\ \ \sqrt{\left\langle x^2(t)\right\rangle-\left\langle x(t)\right\rangle^2}\ \ =\ \ a\sqrt{\frac{t}{\tau}}\ , \tag{2.61}$$

is not proportional to t, but to $\sqrt{t}$. This is a profound fact and a common feature for all processes accumulating uncorrelated values of random variables in time. Thus, to conserve the temporal-spatial proportions of the Wiener process one rescales the time step τ by a factor ε and the lattice constant a by a factor $\sqrt{\varepsilon}$:

$$\tau\ \longmapsto\ \varepsilon\,\tau\quad\text{and}\quad a\ \longmapsto\ \sqrt{\varepsilon}\,a\ . \tag{2.62}$$

A continuous description of the binomial density distribution (2.60) is then approached by taking the limit $\varepsilon\to 0$. When ε approaches 0 the number of steps $n=\frac{t}{\tau\varepsilon}$ in the random walk goes to

infinity and one observes the following identity derived in appendix 2.7 of this chapter.

$$
\begin{aligned}
p(x,t)\,dx \quad &= \quad \frac{1}{2\,\varepsilon\,a}\,2^{-\frac{t}{\varepsilon\tau}}\left(\begin{array}{c}\frac{t}{\varepsilon\tau}\\[2pt]\frac{t}{2\varepsilon\tau}+\frac{x}{2\sqrt{\varepsilon a}}\end{array}\right)\,dx\\[10pt]
&= \quad \sqrt{\frac{n\tau}{4\,a^2\,t}}\,2^{-n}\left(\begin{array}{c}n\\[2pt]\frac{n}{2}+\frac{x}{a}\sqrt{\frac{n\,\tau}{4t}}\end{array}\right)\,dx\\[10pt]
&\stackrel{(2.165)}{=}\quad \sqrt{\frac{\tau}{2\pi\,a^2\,t}}\,\exp\left(-\frac{x^2\,\tau}{2\,a^2\,t}\right)\,dx\,\left(1+\mathcal{O}\!\left(\frac{1}{n}\right)\right)\,.
\end{aligned}
\tag{2.63}
$$

The fraction τ/a^2 is invariant under rescaling (2.62) and, hence, this quantity remains in the continuous description (2.63) of the probability density distribution $p(x,t)$. Comparing equations (2.63) and (2.48) one identifies $D = a^2/2\tau$. The relation between random step length a and time unit τ obviously determines the rate of diffusion embodied in the diffusion constant D: the larger the steps a and the more rapidly these are performed, i.e., the smaller τ, the quicker the diffusion process and the faster the broadening of the probability density distribution $p(x,t)$. According to (2.61) this broadening is then $\sqrt{2Dt}$ as expected for a diffusion process.

Computer Simulation of a Wiener Process

The random walk on a lattice can be readily simulated on a computer. For this purpose one considers an ensemble of particles labeled by k, $k = 1, 2, \ldots N$, the positions $x^{(k)}(j\tau)$ of which are generated at time steps $j = 1, 2, \ldots$ by means of a random number generator. The latter is a routine that produces quasi-random numbers r, $r \in [0,1]$ which are homogenously distributed in the stated interval. The particles are assumed to all start at position $x^{(k)}(0) = 0$. Before every displacement one generates a new r. One then executes a displacement to the left in case of $r < \frac{1}{2}$ and a displacement to the right in case of $r \geq \frac{1}{2}$.

In order to characterize the resulting displacements $x^{(k)}(t)$ one can determine the mean, i.e. the first moment or first cumulant,

$$
\left\langle x(t) \right\rangle \quad = \quad \frac{1}{N}\sum_{k=1}^{N} x^{(k)}(t)
\tag{2.64}
$$

and the variance, i.e. the second cumulant,

$$
\left\langle\!\left\langle x^2(t) \right\rangle\!\right\rangle \quad = \quad \left\langle x^2(t) \right\rangle - \left\langle x(t) \right\rangle^2 \quad = \quad \frac{1}{N}\sum_{k=1}^{N}\left(x^{(k)}(t)\right)^2 \; - \; \left\langle x(t) \right\rangle^2
\tag{2.65}
$$

for $t = \tau, 2\tau, \ldots$. In case of $x^{(k)}(0) = 0$ one obtains $\langle x(t) \rangle \approx 0$. The resulting variance (2.65) is presented for an actual simulation of 1000 walkers in Figure 2.3.

A Wiener Process can be Integrated, but not Differentiated

We want to demonstrate that the path of a Wiener process cannot be differentiated. For this purpose we consider the differential defined through the limit

$$
\frac{d\omega(t)}{dt} \quad := \quad \lim_{\Delta t \to 0}\frac{\omega(t+\Delta t)-\omega(t)}{\Delta t} \quad = \quad \lim_{\Delta t \to 0}\frac{\Delta\omega(t)}{\Delta t}\,.
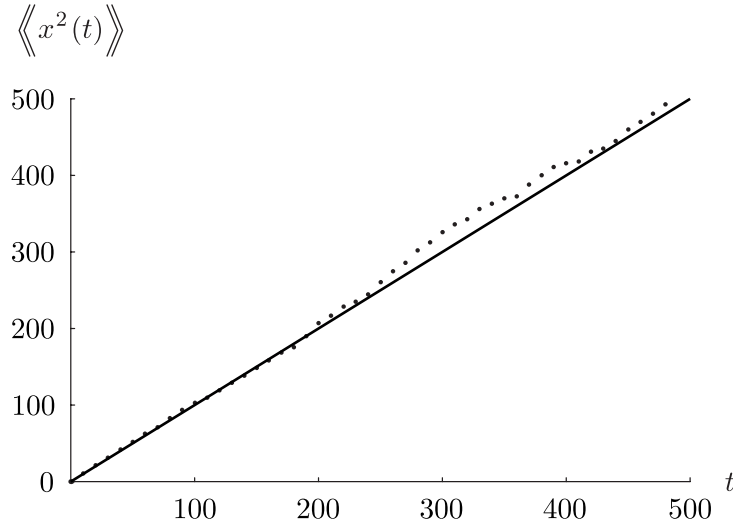\tag{2.66}
$$

Figure 2.3: $\langle\!\langle x^2(t)\rangle\!\rangle$ resulting from a simulated random walk of 1000 particles on a lattice for $\tau = 1$ and $a = 1$. The simulation is represented by dots, the expected [c.f., Eq. (2.61)] result $\langle\!\langle x^2(t)\rangle\!\rangle = t$ is represented by a solid line.

What is the probability for the above limit to render a finite absolute value for the derivative smaller or equal an arbitrary constant v? For this to be the case $|\Delta\omega(t)|$ has to be smaller or equal $v\,\Delta t$. The probability for that is

$$
\begin{aligned}
\int_{-v\,\Delta t}^{v\,\Delta t} d(\Delta\omega)\; p(\Delta\omega, \Delta t) &= \frac{1}{\sqrt{4\pi D\,\Delta t}} \int_{-v\,\Delta t}^{v\,\Delta t} d(\Delta\omega)\; \exp\left[-\frac{(\Delta\omega)^2}{4D\,\Delta t}\right] \\
&= \mathrm{erf}\left[\sqrt{\frac{\Delta t}{D}}\frac{v}{2}\right].
\end{aligned}
\tag{2.67}
$$

The above expression vanishes for $\Delta t \to 0$. Hence, taking the differential as proposed in equation (2.66) we would almost never obtain a finite value for the derivative. This implies that the velocity corresponding to a Wiener process is almost always plus or minus infinity.

It is instructive to consider this calamity for the random walk on a lattice as well. The scaling (2.62) renders the associated velocities like $\pm\frac{a}{\tau}$ infinite and the random walker seems to be infinitely fast as well. Nevertheless, for the random walk on a lattice with non-zero ϵ one can describe the velocity through a discrete stochastic process $\dot{x}(t)$ with the two possible values $\pm\frac{a}{\tau}$ for each time interval $]j\,\tau, (j+1)\,\tau]$, $j \in \mathbb{N}$. Since every random step is completely independent of the previous one, $\dot{x}_i = \dot{x}(t_i)$ with $t_i \in]i\,\tau, (i+1)\,\tau]$ is completely uncorrelated to $\dot{x}_{i-1} = \dot{x}(t_{i-1})$ with $t_{i-1} \in [(i-1)\,\tau, i\,\tau]$, and $x(t)$ with $t \leq i\,\tau$. Thus, we have

$$
P(\dot{x}_i, t_i) = \begin{cases} \frac{1}{2} & \text{for } \dot{x}_i = \pm\frac{a}{\tau}\,, \\ 0 & \text{otherwise,} \end{cases}
\tag{2.68}
$$

$$
P(\dot{x}_j, t_j | \dot{x}_i, t_i) = \begin{cases} \begin{cases} 1 & \text{for } \dot{x}_j = \dot{x}_i \\ 0 & \text{for } \dot{x}_j \neq \dot{x}_i \end{cases} & \text{, for } i = j\,, \\ P(\dot{x}_j, t_j) & \text{, for } i \neq j\,. \end{cases}
\tag{2.69}
$$

The velocity of a random walk on a lattice is characterized by the following statistical moments

$$\left\langle \dot{x}^n(t) \right\rangle = \begin{cases} \left(\frac{a}{\tau}\right)^n & \text{, for even } n \text{ ,} \\ 0 & \text{, for odd } n \text{ ,} \end{cases} \tag{2.70}$$

and correlation functions

$$\left\langle \dot{x}^n(t_j)\,\dot{x}^m(t_i) \right\rangle = \begin{cases} \begin{cases} \left(\frac{a}{\tau}\right)^{m+n} & \text{for even } (m+n) \\ 0 & \text{otherwise} \end{cases} & \text{for } i = j \text{ ,} \\ \begin{cases} \left(\frac{a}{\tau}\right)^{m+n} & \text{for even } m \text{ and even } n \\ 0 & \text{otherwise} \end{cases} & \text{for } i \neq j \text{ .} \end{cases} \tag{2.71}$$

If we proceed to a continuous description with probability density distributions as in equation (2.60), we obtain

$$p(\dot{x}_i, t_i) = \frac{1}{2}\left(\delta\left(\dot{x}_i + \frac{a}{\tau}\right) + \delta\left(\dot{x}_i - \frac{a}{\tau}\right)\right) , \tag{2.72}$$

$$p(\dot{x}_j, t_j | \dot{x}_i, t_i) = \begin{cases} \delta(\dot{x}_j - \dot{x}_i) & \text{, for } i = j \text{ ,} \\ p(\dot{x}_j, t_j) & \text{, for } i \neq j \text{ ,} \end{cases} \tag{2.73}$$

and we derive the same statistical momtents

$$\left\langle \dot{x}^n(t) \right\rangle = \begin{cases} \left(\frac{a}{\tau}\right)^n & \text{, for even } n \text{ ,} \\ 0 & \text{, for odd } n \text{ ,} \end{cases} \tag{2.74}$$

and correlation functions defined for continuous $\dot{x}$ range

$$\left\langle \dot{x}^n(t_j)\,\dot{x}^m(t_i) \right\rangle = \begin{cases} \begin{cases} \left(\frac{a}{\tau}\right)^{m+n} & \text{, for even } (m+n), \\ 0 & \text{, otherwise.} \end{cases} & \text{, for } i = j, \\ \begin{cases} \left(\frac{a}{\tau}\right)^{m+n} & \text{, for even } m \text{ and even } n, \\ 0 & \text{, otherwise,} \end{cases} & \text{, for } i \neq j \text{ .} \end{cases} \tag{2.75}$$

One encounters difficulties when trying to rescale the discrete stochastic process $\dot{x}(t_i)$ according to (2.62). The positions of the delta-functions $\pm\frac{a}{\tau}$ in the probability density distributions (2.72) wander to $\pm\infty$. Accordingly, the statistical moments and correlation functions of even powers move to infinity as well. Nevertheless, these correlation functions can still be treated as distributions in time. If one views the correlation function $\langle \dot{x}(t_1)\,\dot{x}(t_0) \rangle$ as a rectangle distribution in time t_1 (see Figure 2.4), one obtains for the limit $\varepsilon \to 0$

$$\left\langle \dot{x}(t_1)\,\dot{x}(t_0) \right\rangle dt_1 = \lim_{\varepsilon \to 0} \left(\frac{\sqrt{\varepsilon}\,a}{\varepsilon\tau}\right)^2 \varepsilon\, dt_1$$

$$= \frac{a^2}{\tau}\,\delta(t_1 - t_0)\,dt_1 . \tag{2.76}$$

Even though the probability distributions of the stochastic process $\dot{x}(t)$ exhibit some unusual features, $\dot{x}(t)$ is still an admissable Markov process. Thus, one has a paradox. Since
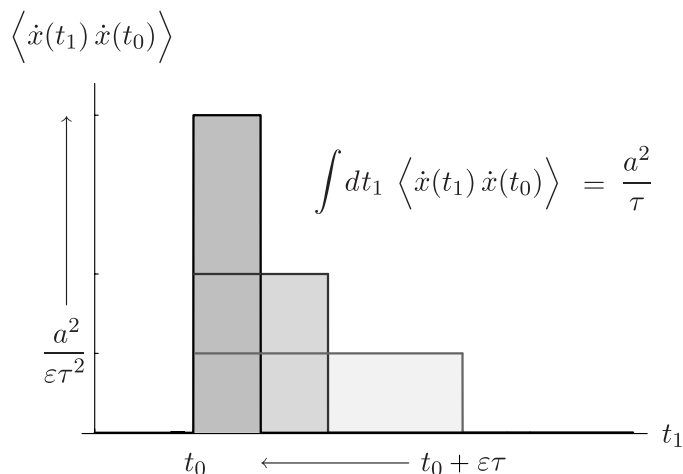
$$x(t_j) = \sum_{i=0}^{j} \tau\,\dot{x}(t_i) , \tag{2.77}$$

Figure 2.4: Cumulant (2.76) shown as a function of t_1. As ε is chosen smaller and smaller (2.76) approaches a Dirca delta distribution at t_0, since the area under the graph remains $\tau \frac{a^2}{\tau^2}$.

it is fair to claim that for the limit $\varepsilon \to 0$ the following integral equation holds.

$$x(t) \quad = \quad \int dt \; \dot{x}(t) \; . \tag{2.78}$$

The converse, however,

$$\frac{dx(t)}{dt} \quad = \quad \dot{x}(t) \tag{2.79}$$

is ill-defined as has been shown in equation (2.66).

Two questions come to mind. First, do stochastic equations like (2.3) and (2.79) make any sense? Second, is $\dot{x}(t)$ unique or are there other stochastic processes that sum up to $x(t)$?

The first question is quickly answered. Stochastic differential equations are only well defined by the integral equations they imply. Even then the integrals in these integral equations have to be handled carefully as will be shown below. Therefore, equation (2.79) should be read as equation (2.78).

We answer the second question by simply introducing another stochastic processes, the Ornstein-Uhlenbeck process, the integral over which also yields the Wiener process. Nevertheless, all processes that yield the Wiener process by integration over time do exhibit certain common properties that are used to define one encompassing, idealized Markov processes, the so-called Gaussian white noise. This process may be viewed as the time-derivative of the Wiener process. Gaussian white noise will be our third example of a stochastic process.
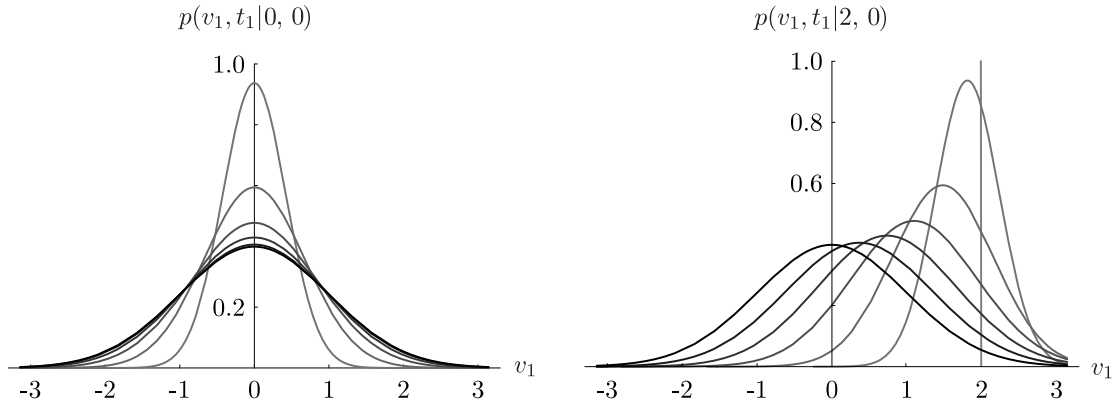
Figure 2.5: The probability density distribution (2.81) of a Ornstein-Uhlenbeck process for $\sigma = \sqrt{2}$ and $\gamma = 1$ in arbitrary temporal and spatial units. The distribution (2.81) is shown for $v_0 = 0$ and $v_0 = 2$ for $t_1 = 0.0, 0.1, 0.3, 0.6, 1.0, 1.7,$ and ∞.

Ornstein-Uhlenbeck Process

Our second example for a Markov process is the *Ornstein-Uhlenbeck process*. The *Ornstein-Uhlenbeck process*, describing a random variable $v(t)$, is defined through the probabilities

$$p(v_0, t_0) \;=\; \frac{1}{\sqrt{\pi \gamma \sigma^2}} \, \exp\!\left(-\frac{v_0^2}{\gamma \sigma^2}\right) \tag{2.80}$$

$$p(v_1, t_1 | v_0, t_0) \;=\; \frac{1}{\sqrt{\pi S}} \, \exp\!\left(-\frac{1}{S}\left(v_1 - v_0 \, e^{-\gamma \Delta t}\right)^2\right) \;, \tag{2.81}$$

$$\text{with} \quad \Delta t \;=\; |t_1 - t_0| \;, \tag{2.82}$$
$$S \;=\; \gamma \sigma^2 \left(1 - e^{-2\gamma \Delta t}\right) \;.$$

The probabilities (see Figure 2.5) are characterized through two parameters σ and γ. Their significance will be explained further below. The process is homogeneous in time, since (2.81) depends solely on Δt, but is not homogeneous in v. Furthermore, the Ornstein-Uhlenbeck Process is stationary, i.e., $p(v_0, t_0)$ does not change in time.

The characteristic function, associated with the unconditional probability distribution $p(v_0, t_0)$ in (2.80) is also independent of time and given by

$$G(s) \;=\; e^{-\gamma\left(\frac{\sigma s^2}{2}\right)} \;. \tag{2.83}$$

The associated moments and cumulants are

$$\left\langle v^n(t) \right\rangle \;=\; \begin{cases} 0 & \text{for odd } n, \\ (n-1)!! \, \left(\tfrac{1}{2}\gamma\sigma^2\right)^{n/2} & \text{for even } n, \end{cases} \tag{2.84}$$

and

$$\left\langle\!\!\left\langle v^n(t) \right\rangle\!\!\right\rangle \;=\; \begin{cases} \frac{\gamma\sigma^2}{2} & \text{for } n = 2, \\ 0 & \text{otherwise} \;. \end{cases} \tag{2.85}$$

The characteristic function for the conditional probability (2.81) is

$$G_v(s_0, t_0; s_1, t_1) \;=\; \exp\!\left[-\frac{1}{4}\gamma\sigma^2 \left(s_0^2 + s_1^2 + 2\, s_0 \, s_1 \, e^{-\gamma|t_1 - t_0|}\right)\right] \;. \tag{2.86}$$

The corresponding correlation functions, defined according to (2.28, 2.29) are

$$\left\langle v_1^{n_1}(t_1)\, v_0^{n_0}(t_0) \right\rangle = \int dv_0\; v_0^{n_0}\; p(v_0, t_0) \int dv_1\; v_1^{n_1}\; p(v_1, t_1 | v_0, t_0) \tag{2.87}$$

$$= \begin{cases} 0 & \text{, for odd } (n_0 + n_1), \\ \frac{1}{2}\gamma\sigma^2\; e^{-\gamma |t_1 - t_0|} & \text{, for } n_0 = 1 \text{ and } n_1 = 1, \\ \frac{3}{4}\gamma^2\sigma^4\; e^{-\gamma |t_1 - t_0|} & \text{, for } n_0 = 1 \text{ and } n_1 = 3, \\ \frac{1}{4}\gamma^2\sigma^4 \left(1 + 2\, e^{-2\gamma |t_1 - t_0|}\right) & \text{, for } n_0 = 2 \text{ and } n_1 = 2, \\ \dots \end{cases} \tag{2.88}$$

This implies that the correlation of $v(t_1)$ and $v(t_0)$ decays exponentially. As for the Wiener process, the most compact description of the unconditional probability is given by the cumulants

$$\left\langle\!\!\left\langle v^{n_1}(t_1)\, v^{n_0}(t_0) \right\rangle\!\!\right\rangle = \begin{cases} \frac{\gamma\sigma^2}{2}\, e^{-\gamma |t_1 - t_0|} & \text{for } n_0 = n_1 = 1, \\ 0 & \text{otherwise, for } n_0 \text{ and } n_1 \neq 0 \;. \end{cases} \tag{2.89}$$

We want to demonstrate now that integration of the Ornstein-Uhlenbeck process $v(t)$ yields the Wiener process. One expects the formal relationship

$$\tilde{\omega}(t) = \int_0^t ds\; v(s) \tag{2.90}$$

to hold where $\tilde{\omega}(t)$ is a Wiener process. In order to test this supposition one needs to relate the cumulants (2.51, 2.54) and (2.85, 2.89) for these processes according to

$$\left\langle\!\!\left\langle \tilde{\omega}(t)\, \tilde{\omega}(t') \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \int_0^t ds\, v(s) \int_0^{t'} ds'\, v(s') \right\rangle\!\!\right\rangle = \int_0^t ds \int_0^{t'} ds' \left\langle\!\!\left\langle v(s)\, v(s') \right\rangle\!\!\right\rangle \tag{2.91}$$

assuming $t \geq t'$. By means of (2.89) and according to the integration depicted in Figure 2.6 follows

$$\left\langle\!\!\left\langle \tilde{\omega}(t)\, \tilde{\omega}(t') \right\rangle\!\!\right\rangle$$
$$= \frac{1}{2}\gamma\sigma^2 \left[\underbrace{\int_0^{t'} ds' \int_0^{s'} ds\; e^{-\gamma(s'-s)}}_{①} + \underbrace{\int_0^{t'} ds \int_0^{s} ds'\; e^{-\gamma(s-s')}}_{②} + \underbrace{\int_{t'}^{t} ds \int_0^{t'} ds'\; e^{-\gamma(s-s')}}_{③} \right] \tag{2.92}$$
$$= \sigma^2\, t' + \frac{\sigma^2}{2\gamma} \left[-1 + e^{-\gamma t'} + e^{-\gamma t} - e^{-\gamma(t-t')} \right] \;.$$

For times long compared to the time scale of velocity relaxation γ^{-1} one reproduces Eq. (2.54) (we don't treat explicitly the case $t \leq t'$), and for $t = t'$ Eq. (2.54), where $D = \sigma^2/2$.
The relationship between the Ornstein-Uhlenbeck and Wiener processes defined through (2.90) holds for all cumulants, not just for the cumulants of second order. We only had to proof relation (2.90), since all other cumulants of both processes are simply 0. This allows one to state for $D = \sigma^2/2$

$$\omega(t) = \lim_{\gamma \to \infty} \int dt\; v(t) \;. \tag{2.93}$$

With respect to their probability distributions the Ornstein-Uhlenbeck process $v(t)$ and the velocity of a random walker $\dot{x}(t)$ are different stochastic processes. However, in the limit $\gamma \to \infty$ and $\varepsilon \to 0$
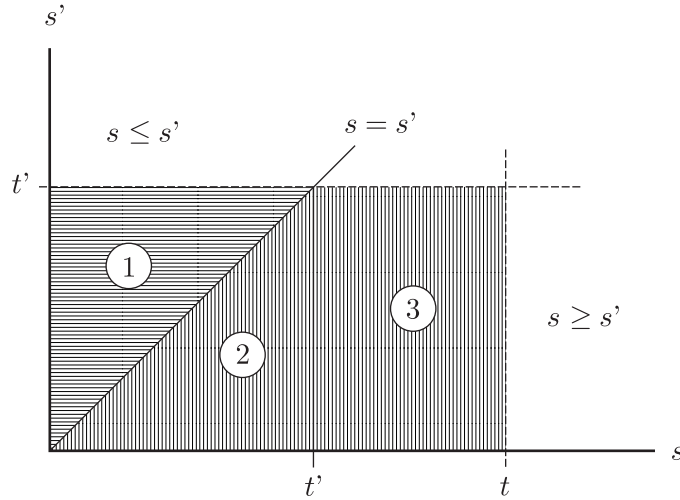
Figure 2.6: The three areas of integration as used in equation 2.92 are shown in the coordinate plane of the two integration variables s and s'.

the following moment and correlation function turn out to be the same for both processes, if $D = \frac{\sigma^2}{2} = \frac{a^2}{2\tau}$.

$$\Big\langle v(t) \Big\rangle \;\; = \;\; \Big\langle \dot{x}(t) \Big\rangle \;\; = \;\; 0 \;, \tag{2.94}$$

$$\lim_{\gamma \to \infty} \Big\langle v(t_1)\, v(t_0) \Big\rangle \;\; = \;\; \lim_{\varepsilon \to 0} \Big\langle \dot{x}(t_1)\, \dot{x}(t_0) \Big\rangle \;\; = \;\; 2\, D \;\, \delta(t_1 - t_0) \;. \tag{2.95}$$

Hence, one uses these later properties to define an idealized Markov process, the so-called *Gaussian white noise*.

White Noise Process

An important idealized stochastic process is the so-called '*Gaussian white noise*'. This process, denoted by $\xi(t)$, is not characterized through conditional and unconditional probabilities, but through the following statistical moment and correlation function

$$\Big\langle \xi(t) \Big\rangle \;\; = \;\; 0 \;, \tag{2.96}$$

$$\Big\langle \xi(t_1)\, \xi(t_0) \Big\rangle \;\; = \;\; \zeta^2 \;\, \delta(t_1 - t_0) \;. \tag{2.97}$$

The attribute *Gaussian* implies that all cumulants higher than of second order are 0.

$$\Big\langle\!\!\Big\langle \xi^{n_1}(t_1)\, \xi^{n_0}(t_0) \Big\rangle\!\!\Big\rangle \;\; = \;\; \begin{cases} \zeta^2 \;\, \delta(t_1 - t_0) & \text{, for } n_0, n_1 = 1 \;, \\ 0 & \text{, otherwise} \;. \end{cases} \tag{2.98}$$

The reason why this process is termed 'white' is connected with its correlation function (2.97), the Fourier transform of which is constant, i.e., entails all frequencies with equal amplitude just as white radiation. The importance of the process $\xi(t)$ stems from the fact that many other stochastic processes are described through stochastic differential equations with a (white) noise term $\xi(t)$. We will show this for the Wiener process below and for the Ornstein-Uhlenbeck processes later in the script.

As hinted by the examples in this section we can show that the integral of Gaussian white noise happens to be a Wiener process. We prove this in the same fashion as above by deriving the cumulants for $\int dt\, \xi(t)$. Again the task is simplified by the fact that only one cumulant is non-zero, namely,

$$
\begin{aligned}
\left\langle\!\!\!\left\langle \int_0^{t_1} ds_1\, \xi(s_1) \int_0^{t_0} ds_0\, v(s_0) \right\rangle\!\!\!\right\rangle
&= \int_0^{t_1} ds_1 \int_0^{t_0} ds_0 \left\langle\!\!\!\left\langle v(s_1)\, v(s_0) \right\rangle\!\!\!\right\rangle \\
&= \int_0^{t_1} ds_1 \int_0^{t_0} ds_0\, \zeta^2\, \delta(s_1 - s_0) \\
&= \zeta^2 \, \min(t_1, t_0) \ .
\end{aligned}
\tag{2.99}
$$

We demonstrated, thereby, the important relationship between white noise $\xi(t)$ and the Wiener process $\omega(t)$.

$$
\omega(t) = \int_0^t ds\, \xi(s) \ ,
\tag{2.100}
$$

for $2D = \zeta^2 = 1$.

2.4 Ito calculus

The introduction to Ito's calculus in this section is based on [13] which we recommend for further reading as well as [29, 32].

We return to the stochastic differential equation of section 2.2, which we will now express as an integral equation. We will model the noise term $\boldsymbol{\eta}(t)$ by a tupel of normalized Gaussian white noise processes $\boldsymbol{\xi}(t)$ with $\zeta = 1$.

$$
\mathbf{x}(t) = \int_0^t ds\, \boldsymbol{A}\left[\mathbf{x}(s), s\right] + \int_0^t ds\, \boldsymbol{\xi}(s)\cdot\mathbf{B}^T\left[\mathbf{x}(s), s\right] \ .
\tag{2.101}
$$

Since $\mathbf{x}(t)$ is continuous, the first integral on the r.h.s. is well defined, e.g., in the sense of a Riemann integral. However, the second integral poses problems. Let us consider the simple one-dimensional case with an arbitrary function or stochastic process $G[t]$.

$$
I(t) = \int_0^t ds\, \xi(s)\, G[s] \ .
\tag{2.102}
$$

One can rewrite the integral (2.102) in terms of a normalized Wiener process $\omega(t)$ with $D = 1/2$. One subsitutes $d\omega(s)$ for $ds\xi(s)$, since $\int_0^t ds\xi(s) = \omega(t) = \int_0^t d\omega(s)$ and obtains

$$
I(t) = \int_0^t d\omega(s)\, G[s] \ .
\tag{2.103}
$$

The kind of Riemann-Stieltjes integral (2.103) can be approximated by the sums $I_n^{(\alpha)}(t)$ which evaluate the integral at n discrete times steps $t_j = j\frac{t}{n}$.

$$
I^{(\alpha)} = \operatorname*{ms\text{-}lim}_{n\to\infty}\, I_n^{(\alpha)}(t) \ , \text{ with}
\tag{2.104}
$$

$$
I_n^{(\alpha)}(t) = \sum_{j=1}^n G\left[(1-\alpha)\, t_{j-1} + \alpha\, t_j\right] \left(\omega(t_j) - \omega(t_{j-1})\right) \ .
\tag{2.105}
$$

Two remarks about equation (2.104) and (2.105) are due. First, one has to specify the meaning of approximation, since one is dealing with random variables. To approximate the integral $I^{(\alpha)}$ one takes the so-called *mean square limit*. Such a limt has to satisfy the following condition of convergence.

$$X \;=\; \underset{n\to\infty}{\text{ms-lim}} \; X_n \;,\; \text{if} \tag{2.106}$$

$$\lim_{n\to\infty} \left\langle (X_n - X)^2 \right\rangle \;=\; 0 \;. \tag{2.107}$$

Second, note that the sum $I_n^{(\alpha)}(t)$ is parameterized by α. This allows one to choose the position where to evaluated $G[t]$ within the time step intervals $[t_{j-1}, t_j]$. In the limit $n \to \infty$ as the intervals $[t_{j-1}, t_j]$ become infinitely small, non-stochastic integrals become independent of α, not so $I_n^{(\alpha)}(t)$!

We will demonstrate the dependence of $I^{(\alpha)}$ on α in two ways. We first derive a closed form of a simple stochastic integral and thereby put forward an explicit example of α-dependence. In addition we will determine an estimates for $I_n^{(\alpha)}(t)$ on the orders of $\mathcal{O}\left(\frac{1}{n}\right)$ and demonstrate the persistence of the α-dependence as n goes to infinity.

The simple integral we want to solve explicitly is $G[t] = \omega(t)$.

$$\int_0^t d\omega(s)\, \omega(s) \tag{2.108}$$

$$= \; \underset{n\to\infty}{\text{ms-lim}} \; \sum_{j=1}^n \omega(\tau)\, \big(\omega(t_j) - \omega(t_{j-1})\big) \qquad , \text{ with } \tau = (1-\alpha)\, t_{j-1} + \alpha\, t_j$$

$$= \; \underset{n\to\infty}{\text{ms-lim}} \; \sum_{j=1}^n \Big[\omega(\tau)\, \big(\omega(\tau) - \omega(t_{j-1})\big) \;+\; \omega(\tau)\, \big(\omega(t_j) - \omega(\tau)\big) \Big]$$

$$= \; \underset{n\to\infty}{\text{ms-lim}} \; \sum_{j=1}^n \frac{1}{2} \Big[-\big[\omega(\tau) - \big(\omega(\tau) - \omega(t_{j-1})\big)\big]^2 + \omega^2(\tau) \;+\; \big(\omega(\tau) - \omega(t_{j-1})\big)^2$$

$$\qquad\qquad + \big[\omega(\tau) + \big(\omega(t_j) - \omega(\tau)\big)\big]^2 - \omega^2(\tau) \;-\; \big(\omega(t_j) - \omega(\tau)\big)^2 \Big]$$

$$= \; \underset{n\to\infty}{\text{ms-lim}} \; \sum_{j=1}^n \frac{1}{2} \Big[\omega^2(t_j) \;-\; \omega^2(t_{j-1}) \;+\; \big(\omega(\tau) - \omega(t_{j-1})\big)^2 - \big(\omega(t_j) - \omega(\tau)\big)^2 \Big] \;.$$

The first term of the jth summand cancels the second term in the $(j+1)^{\text{th}}$ summand, hence only $\omega^2(t)$ and $-\omega^2(0)$ remain.

$$\int_0^t d\omega(s)\, \omega(s) \tag{2.109}$$

$$= \; \frac{1}{2} \Big[\omega^2(t) \;-\; \omega^2(0) \;+\; \underset{n\to\infty}{\text{ms-lim}} \; \sum_{j=1}^n \Big[\big(\omega(\tau) - \omega(t_{j-1})\big)^2 - \big(\omega(t_j) - \omega(\tau)\big)^2 \Big] \Big] \;.$$

To determine the mean square limit of a sequence one can first calculate the limit of the average sequence and than verify if the result satisfies the mean square convergence condition (2.107). For the first step the average of the summands in (2.109) has to be derived. These summands consist of two squared increments of the normailzed Wiener process. These increments, we refer to both of

them with $\Delta\omega(\Delta t)$, exhibit the same unconditional probability distribution (2.48) as the Wiener process $\omega(t)$ itself und thus have the same statistical moments as displayed in (2.50), namely

$$
\left\langle \Delta\omega^n(\Delta t) \right\rangle \;=\; \begin{cases} 0 & \text{for odd } n, \\ (n-1)!! \,(2D\,\Delta t)^{n/2} & \text{for even } n, \text{ with } D = \tfrac{1}{2}\,. \end{cases} \tag{2.110}
$$

These moments applied to the increments $\omega(\tau) - \omega(t_{j-1})$ and $\omega(t_j) - \omega(\tau)$ resolve the limit in equation (2.109).

$$
\lim_{n\to\infty} \left\langle \sum_{j=1}^{n} \left[\big(\omega(\tau) - \omega(t_{j-1})\big)^2 - \big(\omega(t_j) - \omega(\tau)\big)^2 \right] \right\rangle
$$

$$
= \lim_{n\to\infty} \sum_{j=1}^{n} \left[\left\langle \big(\omega(\tau) - \omega(t_{j-1})\big)^2 \right\rangle - \left\langle \big(\omega(t_j) - \omega(\tau)\big)^2 \right\rangle \right]
$$

$$
= \lim_{n\to\infty} \sum_{j=1}^{n} \left[(\tau - t_{j-1}) - (t_j - \tau) \right]
$$

$$
= \lim_{n\to\infty} \sum_{j=1}^{n} \left[(2\alpha - 1)\,(t_j - t_{j-1}) \right]
$$

$$
= (2\alpha - 1)\,(t - 0)\,. \tag{2.111}
$$

Now the mean square convergence of the limit (2.111) has to be checked.

$$
\lim_{n\to\infty} \left\langle \left[\sum_{j=1}^{n} \Big[\underbrace{\big(\omega(\tau) - \omega(t_{j-1})\big)^2}_{=:\,\Delta\omega_{1-\alpha}^2(t_j)} - \underbrace{\big(\omega(t_j) - \omega(\tau)\big)^2}_{=:\,\Delta\omega_{\alpha}^2(t_j)} \Big] - (2\alpha - 1)\,t \right]^2 \right\rangle \tag{2.112}
$$

$$
= \lim_{n\to\infty} \left\langle \sum_{j=1}^{n} \left[\Delta\omega_{1-\alpha}^4(t_j) - 2\,\Delta\omega_{1-\alpha}^2(t_j)\,\Delta\omega_{\alpha}^2(t_j) + \Delta\omega_{\alpha}^4(t_j) \right] \right.
$$

$$
+ \,2 \sum_{j>k=1}^{n} \left[\Delta\omega_{1-\alpha}^2(t_j) - \Delta\omega_{\alpha}^2(t_j) \right] \left[\Delta\omega_{1-\alpha}^2(t_k) - \Delta\omega_{\alpha}^2(t_k) \right]
$$

$$
\left. - \,2\,(2\alpha - 1)\,t \sum_{j=1}^{n} \left[\Delta\omega_{1-\alpha}^2(t_j) - \Delta\omega_{\alpha}^2(t_j) \right] + (2\alpha - 1)^2\,t^2 \right\rangle .
$$

For a Wiener process one can verify that increments $\Delta\omega(t)$ of non-over lapping time intervals $[t, t + \Delta t]$ are statistically independent. Hence we can spread the moment brakets $\langle\ldots\rangle$ as follows

$$
= \lim_{n\to\infty} \left(\sum_{j=1}^{n} \left[\left\langle \Delta\omega_{1-\alpha}^4(t_j) \right\rangle - 2\left\langle \Delta\omega_{1-\alpha}^2(t_j) \right\rangle \left\langle \Delta\omega_{\alpha}^2(t_j) \right\rangle + \left\langle \Delta\omega_{\alpha}^4(t_j) \right\rangle \right] \right.
$$

$$
+ \,2 \sum_{j>k=1}^{n} \left[\left\langle \Delta\omega_{1-\alpha}^2(t_j) \right\rangle - \left\langle \Delta\omega_{\alpha}^2(t_j) \right\rangle \right] \left[\left\langle \Delta\omega_{1-\alpha}^2(t_k) \right\rangle - \left\langle \Delta\omega_{\alpha}^2(t_k) \right\rangle \right]
$$

$$
\left. - \,2\,(2\alpha - 1)\,t \sum_{j=1}^{n} \left[\left\langle \Delta\omega_{1-\alpha}^2(t_j) \right\rangle - \left\langle \Delta\omega_{\alpha}^2(t_j) \right\rangle \right] + (2\alpha - 1)^2\,t^2 \right).
$$

Applying equation (2.110) to every moment and then splitting the first sum gives

$$
\begin{aligned}
= \quad & \lim_{n\to\infty} \left(\sum_{j=1}^{n} \left[3\left(\tau - t_{j-1}\right)^2 - 2\left(\tau - t_{j-1}\right)\left(t_j - \tau\right) + 3\left(t_j - \tau\right)^2 \right] \right. \\
& + 2 \sum_{j>k=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right]\left[\left(\tau - t_{k-1}\right) - \left(t_k - \tau\right)\right] \\
& \left. - 2\left(2\alpha - 1\right) t \sum_{j=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right] + \left(2\alpha - 1\right)^2 t^2 \right) \\
= \quad & \lim_{n\to\infty} \left(\sum_{j=1}^{n} \left[2\left(\tau - t_{j-1}\right)^2 + 2\left(t_j - \tau\right)^2 \right] \right. \\
& + \sum_{j=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right]\left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right] \\
& + 2 \sum_{j>k=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right]\left[\left(\tau - t_{k-1}\right) - \left(t_k - \tau\right)\right] \\
& \left. - 2\left(2\alpha - 1\right) t \sum_{j=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right] + \left(2\alpha - 1\right)^2 t^2 \right).
\end{aligned}
$$

The first sum is of order $\mathcal{O}\left(\frac{1}{n}\right)$ and apporaches 0 for $n \to \infty$, since each summand is proportional to a time interval length squared and thus is of order $\mathcal{O}\left(\frac{1}{n^2}\right)$. The second sum and the following double sum combine to a single double sum without the limitation $j > k$. This resulting double sum can then be written as a product of two sums.

$$
\begin{aligned}
= \quad & \lim_{n\to\infty} \left(\sum_{j,k=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right]\left[\left(\tau - t_{k-1}\right) - \left(t_k - \tau\right)\right] \right. \\
& \left. - 2\left(2\alpha - 1\right) t \sum_{j=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right] + \left(2\alpha - 1\right)^2 t^2 \right). \\
= \quad & \lim_{n\to\infty} \left(\left[\sum_{j=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right]\right]\left[\sum_{k=1}^{n} \left[\left(\tau - t_{k-1}\right) - \left(t_k - \tau\right)\right]\right] \right. \\
& \left. - 2\left(2\alpha - 1\right) t \sum_{j=1}^{n} \left[\left(\tau - t_{j-1}\right) - \left(t_j - \tau\right)\right] + \left(2\alpha - 1\right)^2 t^2 \right).
\end{aligned}
$$

All sums are now equivalent to the one in equation (2.111) and one finally obtains

$$
\begin{aligned}
= \quad & \lim_{n\to\infty} \left(\left(2\alpha - 1\right)^2 t^2 - 2\left(2\alpha - 1\right) t \left(2\alpha - 1\right) t + \left(2\alpha - 1\right)^2 t^2 \right) \\
= \quad & 0 \, .
\end{aligned}
\tag{2.113}
$$

It is thus proven that the defining mean square limit of the stochastic integral (2.108) renders

$$
\int_0^t d\omega(s)\,\omega(s) \;=\; \frac{1}{2}\left[\omega^2(t) - \omega^2(0) + \left(2\alpha - 1\right) t \right] \, .
\tag{2.114}
$$

The observed dependence on α can be made plausible. Consider a time interval $[t_{i-1}, t_i]$ in the summation $I_n^{(\alpha)}(t)$ for $G(t) = \omega(t)$. The mean absolute difference of the Wiener process $\omega(t)$ at the left and the right side of the interval $[t_{i-1}, t_i]$ is given by the standard deviation of the difference $\omega(t_i) - \omega(t_{i-1})$

$$\sqrt{\left\langle (\omega(t_i) - \omega(t_{i-1}))^2 \right\rangle} = \sqrt{t_i - t_{i-1}} \, . \tag{2.115}$$

The difference in summing over all values on the left side of the intervals as opposed to the right side is then on average given by

$$\sum_{i=1}^{n} \underbrace{\sqrt{t_i - t_{i-1}}}_{\mathcal{O}\left(\sqrt{1/n}\right)} \underbrace{(\omega(t_i) - \omega(t_{i-1}))}_{\mathcal{O}\left(\sqrt{1/n}\right)} \, . \tag{2.116}$$

If we sum over all terms we obtain n-times an expression of order $\mathcal{O}(1/n)$ and consequently a finite value; a finite difference!

If we compare this observation with ordinary calculus we see the essential discrepancy. Consider the integral $\int_0^t dt \, f(t)$. The difference between evaluating the left and right side of interval $[t_{i-1}, t_i]$ is given by $f'(t_{i-1})(t_i - t_{i-1})$. Again the difference of summing over all values $f(t)$ on the left side of the intervals as opposed to the right side is

$$\sum_{i=1}^{n} \underbrace{f'(t_{i-1})(t_i - t_{i-1})}_{\mathcal{O}(1/n)} \underbrace{(t_i - t_{i-1})}_{\mathcal{O}(1/n)} \, . \tag{2.117}$$

This sum is of order $\mathcal{O}(1/n)$ and approaches 0 for $n \to \infty$. It is consequently irrelevant which side of the interval $[t_{i-1}, t_i]$ we choose to evaluate $f(t)$.

The underlying cause for the α-dependence of stochastic integrals is evident. It is the $\sqrt{1/n}$-scaling property of stochastic processes! The α-dependence is here to stay. Before we proceed we return once more to our random walker example to gain more insight in the derivations given above.

Gambling on a Random Walk

As in section 2.3 we begin with a random walk on a lattice, this time with $a = \tau^2$ for simplicity. The random walker will be joined by a gambler. Due to his nature the gambler will make a bet at each step trying to forecast the direction the random walker will take. The gambler's strategy is the following. At time t_{j-1} he will take the distance $x(t_{j-1})$ from the origin $x(0)$ as an indication for the direction in which the random walker will proceed. He will bet an amount proportional to $|x(t_{j-1})|$ and claim that the random walker will further increase the distance from the starting point. The investment for each bet will be proportional to the step size $x(t_j) - x(t_{j-1})$; the smaller the steps, the more bets to make, the smaller the amount available for each bet. The pay off or loss $dF(t_j)$ will be proportional to the amount put forward. Hence, $dF(t_{j-1}) \propto (x(t_j) - x(t_{j-1})) \, x(t_{j-1})$. Note, that $dF(t_{j-1})$ is positive or negative, if the forecast is true or false respectively. To see if this strategy pays we derive the total loss or gain of the gambler by suming over all bets and taking the mean square limit of $n \to \infty$. One determines

$$F(t) \quad \propto \quad \operatorname*{ms\text{-}lim}_{n \to \infty} \sum_{j=1}^{n} \left(x(t_j) - x(t_{j-1}) \right) x(t_{j-1}) \tag{2.118}$$

As the random walk $x(t)$ describes a Wiener process $\omega(t)$ in the limit $n \to \infty$, we can write the above as a stochastic integral with $\alpha = 0$ and obtain

$$
\begin{aligned}
F(t) &\propto \int_0^t d\omega(s)\, \omega(s) \qquad , \text{ with } \alpha = 0 \\
&\propto \frac{1}{2}\left(x^2(t) - x^2(0) - t\right) .
\end{aligned}
\tag{2.119}
$$

Except for the extra term t, Eq. (2.119) resembles the result of ordinary calculus. The term t, however, is essential. It prevents the gambler to run a winning strategy. The mean for the overall loss or gain $F(t)$ is

$$
\left\langle F(t)\right\rangle \;\propto\; \left\langle \frac{1}{2}\left(x^2(t) - x^2(0) - t\right)\right\rangle \;=\; \frac{1}{2}\left(\langle x^2(t)\rangle - t\right) \;=\; 0 .
\tag{2.120}
$$

as one might have expected all along.

To consider a case for which the above integral exhibits $\alpha = 1$ we turn to a cheating gambler. Assume that the cheating gambler has a way to tell which direction the random walker will turn next. Thus, at time t_{j-1} he will base his bet on the subsequent position $x(t_j)$ and not on $x(t_{j-1})$, a small, but decisive advantage. If he obeys the same strategy $d\tilde{F}(t_{j-1}) \propto (x(t_j) - x(t_{j-1}))\, x(t_j)$ as above, however based on $x(t_j)$ and not $x(t_{j-1})$, he will mimic a similar betting behavior as the honest gambler, especially, when τ goes to 0. Then his insight into the future seems to vanish. Surprisingly as time goes by he will win a fortune $\tilde{F}(t)$ as the following result shows.

$$
\begin{aligned}
\tilde{F}(t) &\propto \;\underset{n\to\infty}{\text{ms-lim}} \sum_{j=1}^n (x(t_j) - x(t_{j-1}))\; x(t_j) \\
&\propto \int_0^t d\omega(s)\, \omega(s) \qquad , \text{ with } \alpha = 1 \\
&\propto \frac{1}{2}\left(x^2(t) - x^2(0) + t\right) .
\end{aligned}
\tag{2.121}
$$

The mean gain is

$$
\left\langle \tilde{F}(t)\right\rangle \;\propto\; \left\langle \frac{1}{2}\left(x^2(t) - x^2(0) + t\right)\right\rangle \;=\; t .
\tag{2.122}
$$

A statistical analysis can detect the cheating gambler. The correlation between a random step $d\omega(t) = x(t_j) - x(t_{j-1})$ and the integrand functions $G[t] = x(t_{j-1})$ and $G[t] = x(t_j)$ reveals for the honest gambler

$$
\begin{aligned}
\left\langle x(t_{j-1})\, (x(t_j) - x(t_{j-1}))\right\rangle &= \left\langle x(t_{j-1})\right\rangle \left\langle x(t_j) - x(t_{j-1})\right\rangle \\
&= 0 ,
\end{aligned}
\tag{2.123}
$$

and for the cheating colleague

$$
\begin{aligned}
\left\langle x(t_j)\, (x(t_j) - x(t_{j-1}))\right\rangle &= \left\langle (x(t_{j-1}) + x(t_j) - x(t_{j-1}))\, (x(t_j) - x(t_{j-1}))\right\rangle \\
&= \left\langle x(t_{j-1})\, (x(t_j) - x(t_{j-1}))\right\rangle + \left\langle (x(t_j) - x(t_{j-1}))^2\right\rangle \\
&= (t_j - t_{j-1}).
\end{aligned}
\tag{2.124}
$$

The honest gambling scheme is not correlated to the imminent random step, the cheating scheme however is. One therefore distinguishes between so-called *non-anticipating* and *anticipating* functions. It is obvious that correlations between an integrand $G(t)$ and the integration steps $d\omega(t)$ accumulate and that they have an overall effect on the integral as seen in this example.

We will no longer pursuit this exciting money making scheme, since it has one unfortunate draw back; one has to bet infinitely fast!

Ito's Rules

We have seen that it is not admissable to neglect the α-dependence. Nevertheless it is possible to develope a consistent calculus by assuming a fixed value for parameter α. There are two popular approaches, each with distinct benefits and disadvantages:

$$
\begin{aligned}
\alpha &= 0 \quad \text{Ito calculus,} \\
\alpha &= \frac{1}{2} \quad \text{Stratonovich calculus.}
\end{aligned}
$$

The Stratonovich calculus with $\alpha = 1/2$ exhibits the same integration rules as ordinary calculus. It also models processes with finite correlation time correctly. However, the rules for the Ito calculus are easier to derive. In many instances corresponding derivations are impossible in the Stratonovich case. Hence we begin with an introduction to Ito calculus. Later, in section 2.6 we will compare the Ito and Stratonovich approach. In any case, we have to keep in mind, that the choice of $\alpha = 0$ or $\alpha = 1/2$ is not arbitrary and has to be justified when modeling physical processes with stochastic differential and corresponding integral equations. For now we set $\alpha = 0$.

The foundation of Ito calculus are the four rules

$$
\begin{aligned}
d\omega_i(t)\, d\omega_j(t) &= \delta_{ij}\, dt\ , & (2.125) \\
[d\omega(t)]^N &= 0\ , \text{ for } N > 2\ , & (2.126) \\
d\omega(t)^N\, dt &= 0\ , \text{ for } N \geq 1\ , & (2.127) \\
dt^N &= 0\ , \text{ for } N > 1\ . & (2.128)
\end{aligned}
$$

As with distributions, like the Dirac delta function $\delta(x)$, these rules (2.125-2.128) have to be seen in the context of integration. Furthermore the integration has to be over so-called non-anticipating functions or processes $G(t)$. This will become clear as we proof rule (2.125) for the one-dimensional case.

Rules (2.125) and (2.126) have to be read as

$$
\int_0^t [d\omega(s)]^N \ G(s) = \underset{n\to\infty}{\text{ms-lim}} \sum_{i=1}^n G(t_{i-1})\, [\Delta\omega(t_i)]^N \tag{2.129}
$$

$$
= \begin{cases} \int_0^t ds\ G(s) & \text{, for } N = 2 \\ 0 & \text{, for } N > 2\ , \end{cases} \tag{2.130}
$$

for a non-anticipating function $G(t)$ that is statistcally independent of $(\omega(s) - \omega(t))$ for any $s > t$.

$$
\Big\langle G[t]\, \big(\omega(s) - \omega(t)\big) \Big\rangle = 0\ , \quad \text{for } t < s\ . \tag{2.131}
$$

To prove rule (2.125) we have to show that the following mean square limit vanishes:

$$\left\langle \left[\int_0^t [d\omega(s)]^2 \ G(s) \ - \ \int_0^t ds \ G(s) \right]^2 \right\rangle \tag{2.132}$$

$$= \ \underset{n \to \infty}{\text{ms-lim}} \ \left\langle \left[\sum_{i=1}^n \left(\Delta\omega^2(\Delta t_i) \ - \ \Delta t_i \right) \ G(t_{i-1}) \right]^2 \right\rangle$$

$$= \ \underset{n \to \infty}{\text{ms-lim}} \ \left\langle \sum_{i=1}^n \underbrace{G^2(t_{i-1})}_{\text{stat. indep.}} \ \underbrace{\left(\Delta\omega^2(\Delta t_i) - \Delta t_i \right)^2}_{\text{stat. independent}} \right\rangle$$

$$+ \ \underset{n \to \infty}{\text{ms-lim}} \ \left\langle \sum_{i>j=1}^n \underbrace{G(t_{i-1})\,G(t_{j-1}) \left(\Delta\omega^2(\Delta t_j) - \Delta t_j \right)}_{\text{stat. independent}} \ \underbrace{\left(\Delta\omega^2(\Delta t_i) - \Delta t_i \right)}_{\text{stat. independent}} \right\rangle \ . \tag{2.133}$$

Each of the above underbraced terms is statistically independent of the other underbraced factor. Here the non-anticipation property (2.131) of $G(t)$ comes into play! We obtain

$$I = \ \underset{n \to \infty}{\text{ms-lim}} \ \sum_{i=1}^n \left\langle G^2(t_{i-1}) \right\rangle \ \underbrace{\left\langle \left(\Delta\omega^2(\Delta t_i) - \Delta t_i \right)^2 \right\rangle}_{= \, 2\Delta t_i^2, \text{ due to (2.110)}}$$

$$+ \ \underset{n \to \infty}{\text{ms-lim}} \ \sum_{i>j=1}^n \left\langle G(t_{i-1})\,G(t_{j-1}) \left(\Delta\omega^2(\Delta t_j) - \Delta t_j \right) \right\rangle \ \underbrace{\left\langle \left(\Delta\omega^2(\Delta t_i) - \Delta t_i \right) \right\rangle}_{= \, 0, \text{ due to (2.110)}}$$

$$= \ \underset{n \to \infty}{\text{ms-lim}} \ 2 \sum_{i>j=1}^n \left\langle G^2(t_{i-1}) \right\rangle \ \Delta t_i^2 \ . \tag{2.134}$$

As Δt_i^2 is of order $\mathcal{O}(1/n^2)$ and as long as $G(s)$ is a bounded function, the above sum vanishes as $n \to \infty$. Thus, we have proven Ito's first rule (2.125). All the other rules are shown in a similar fashion.

Ito's Formula

Combining the stochastic differential equation (2.3) and Ito's rules we can derive another important equation, the so-called Ito's formula. Let $f[x(t)]$ be an arbitrary function of a process $x(t)$ that satisfies the one-dimensional sochastic differential equation

$$dx(t) \ = \ \Big(a[x(t),t] \ + \ b[x(t),t] \ \xi(t) \Big) \, dt \ = \ a[x(t),t] \ dt \ + \ b[x(t),t] \ d\omega(t) \ . \tag{2.135}$$

To determine the change of $f[x(t)]$ with respect to dx and dt we perform a Taylor expansion

$$df[x(t)] = \ f[x(t) + dx(t)] \ - \ f[x(t)]$$

$$= \ f'[x(t)] \ dx(t) \ + \ \frac{1}{2} \, f''[x(t)] \ dx^2(t) \ + \ \mathcal{O}\big(dx^3(t)\big) \ .$$

Substituting equation (2.135) for dx we can write

$$df[x(t)] = \ f'[x(t)] \ a[x(t),t] \ dt \ + \ f'[x(t)] \ b[x(t),t] \ d\omega(t)$$

$$+ \ \frac{1}{2} \, f''[x(t)] \ \Big(b[x(t),t] \ d\omega(t) \Big)^2 \ + \ \mathcal{O}\big(d\omega^3(t)\big) \ \mathcal{O}\big(dt^2\big) \ .$$

We can neglect higher orders of $d\omega(t)$ and dt due to Ito's rules (2.126 - 2.128). We can also substitute $d\omega^2(t)$ by dt due to (2.125) and obatin

$$
\begin{aligned}
df[x(t)] = \quad & f'[x(t)] \, a[x(t), t] \, dt \; + \; f'[x(t)] \, b[x(t), t] \, d\omega(t) \\
& + \; \frac{1}{2} \, f''[x(t)] \, \Big(b[x(t), t] \Big)^2 \, dt \; .
\end{aligned}
\tag{2.136}
$$

The resulting Ito's formula, now in more than one-dimension, reads

$$
\begin{aligned}
df[\mathbf{x}(t)] \;\; = \;\; & \sum_i A_i \Big(\partial_i \, f[\mathbf{x}(t)] \Big) \, dt \;\; + \;\; \sum_{i,j} \mathsf{B}_{ij} \Big(\partial_i \, \mathsf{f}[\mathbf{x}(t)] \Big) \, d\omega_j(t) \\
& + \;\; \frac{1}{2} \sum_{i,j,k} \mathsf{B}_{ik} \, \mathsf{B}_{jk} \Big(\partial_i \partial_j \, \mathsf{f}[\mathbf{x}(t)] \Big) \, dt \; .
\end{aligned}
\tag{2.137}
$$

This formula is most helpful when one has to find a relation between the stochastic differential equation (2.3) of $\mathbf{x}(t)$ and a distribution function $f[\mathbf{x}(t)]$. We will utilize Ito's formula in the next section where we will derive the Fokker-Planck equation.

2.5 Fokker-Planck Equations

Again we consider the stochastic differential equation (2.3) with a noise term characterized through white noise, i.e., Eq. (2.97)

$$
\partial_t \, \boldsymbol{x}(t) = \boldsymbol{A}[\boldsymbol{x}(t), t] + \mathbf{B}[\mathbf{x}(t), t] \cdot \boldsymbol{\eta}(t)
\tag{2.138}
$$

assuming $\boldsymbol{\eta}(t) = \boldsymbol{\xi}(t)$ with

$$
\Big\langle \xi_i(t) \Big\rangle \;\; = \;\; 0
\tag{2.139}
$$

$$
\Big\langle \xi_i(t_1) \, \xi_j(t_0) \Big\rangle \, (dt)^2 \;\; = \;\; \delta_{ij} \, \delta(t_1 - t_0) \, dt \; .
\tag{2.140}
$$

For the sake of Ito's calculus one has to assume that coefficient $\mathbf{B}[\boldsymbol{x}(t), t]$ is a non-anticipating function. With this in mind we neglect the arguments $\boldsymbol{x}(t)$ and t of $\boldsymbol{A}$ and $\mathbf{B}$ for easier reading in the rest of this section.

We utilize the result of the section 2.4 and exploit the properties of white noise (2.96, 2.97) by considering the average of Ito's formula (2.137).

$$
\begin{aligned}
\Big\langle df[\boldsymbol{x}(t)] \Big\rangle \;\; = \;\; & \sum_i \Big\langle A_i \left(\partial_i \, f[\boldsymbol{x}(t)] \right) dt \Big\rangle \;\; + \;\; \sum_{i,j} \Big\langle \mathsf{B}_{ij} \left(\partial_i \, \mathsf{f}[\boldsymbol{x}(t)] \right) d\omega_j(t) \Big\rangle \\
& + \;\; \frac{1}{2} \sum_{i,j,k} \Big\langle \mathsf{B}_{ik} \, \mathsf{B}_{jk} \left(\partial_i \partial_j \, \mathsf{f}[\boldsymbol{x}(t)] \right) dt \Big\rangle \; .
\end{aligned}
\tag{2.141}
$$

The second sum on the r.h.s. vanishes, since $\mathbf{B}[\boldsymbol{x}(t), t]$ and $\partial_i f[\boldsymbol{x}(t)]$ are non-anticipating functions and therefore statistically independent of $d\omega_j(t)$, and because of equation (2.139) considering that $d\omega_j(t) = \xi_j(t) \, dt$.

$$
\Big\langle \mathsf{B}_{ij} \left(\partial_i \, \mathsf{f}[\boldsymbol{x}(t)] \right) d\omega_j(t) \Big\rangle \;\; = \;\; \Big\langle \mathsf{B}_{ij} \left(\partial_i \, \mathsf{f}[\boldsymbol{x}(t)] \right) \Big\rangle \underbrace{\Big\langle \xi_j(t) \Big\rangle}_{= \, 0} dt \;\; = \;\; 0 \; .
\tag{2.142}
$$

One is left with equation

$$\left\langle \frac{d}{dt} f[\boldsymbol{x}(t)] \right\rangle = \sum_i \left\langle A_i (\partial_i f[\boldsymbol{x}(t)]) \right\rangle + \frac{1}{2} \sum_{i,j} \left\langle [\mathbf{B} \cdot \mathbf{B}^{\mathsf{T}}]_{ij} (\partial_i \partial_j \mathbf{f}[\boldsymbol{x}(\mathbf{t})]) \right\rangle . \tag{2.143}$$

According to definition (2.12) $\langle f[\boldsymbol{x}(t)] \rangle$ can be expressed as

$$\left\langle f[\boldsymbol{x}(t)] \right\rangle = \int d\boldsymbol{x} \; f[\boldsymbol{x}] \; p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) . \tag{2.144}$$

The reader should note that the initial value of $\langle f[\boldsymbol{x}(t)] \rangle$ defined through (2.144) is $f[\boldsymbol{x}_0]$ in accordance with the initial condition assumed for Eq. (2.3). Applying the time derivative to the r.h.s. of (2.144) and comparing with (2.143) yields

$$\int d\boldsymbol{x} \; f[\boldsymbol{x}] \; \partial_t \, p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) = \tag{2.145}$$

$$\int d\boldsymbol{x} \left(\sum_i A_i \left(\partial_i f[\boldsymbol{x}] \right) + \frac{1}{2} \sum_{i,j} [\mathbf{B} \cdot \mathbf{B}^{\mathsf{T}}]_{ij} \left(\partial_i \partial_j \mathbf{f}[\boldsymbol{x}] \right) \right) \mathbf{p}(\boldsymbol{x}, \mathbf{t} | \boldsymbol{x}_0, \mathbf{t}_0) .$$

Partial integration assuming a volume Ω with a surface $\partial\Omega$ allows one to change the order of the partial differential operators. For example, the first sum becomes

$$\int_\Omega d\boldsymbol{x} \sum_i A_i \left(\partial_i f[\boldsymbol{x}] \right) p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) = - \int_\Omega d\boldsymbol{x} \; f[\boldsymbol{x}] \left(\sum_i \partial_i A_i \; p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) \right)$$

$$+ \int_\Omega d\boldsymbol{x} \left(\sum_i \partial_i A_i \; f[\boldsymbol{x}] \; p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) \right)$$

$$= - \int_\Omega d\boldsymbol{x} \; f[\boldsymbol{x}] \left(\sum_i \partial_i A_i \; p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) \right)$$

$$+ \int_{\partial\Omega} d\boldsymbol{a} \cdot \boldsymbol{A} \; f[\boldsymbol{x}] \; p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) . \tag{2.146}$$

Assuming a $p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0)$ of finite spatial extent, such that it vanishes on the boundary $\partial\Omega$, we can neglect the surface term. Applying the same calculation twice to the second term in (2.145) leads to

$$\int d\boldsymbol{x} \; f[\boldsymbol{x}] \; \partial_t \, p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) = \tag{2.147}$$

$$\int d\boldsymbol{x} \; f[\boldsymbol{x}] \left(- \sum_i \partial_i A_i \; p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) + \frac{1}{2} \sum_{i,j} \partial_i \partial_j \, [\mathbf{B} \cdot \mathbf{B}^{\mathsf{T}}]_{ij} \; \mathbf{p}(\boldsymbol{x}, \mathbf{t} | \boldsymbol{x}_0, \mathbf{t}_0) \right) .$$

Since $f[\boldsymbol{x}(t)]$ is arbitrary we can conclude

$$\partial_t \, p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) = - \sum_i \partial_i A_i \; p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) + \frac{1}{2} \sum_{i,j} \partial_i \partial_j \, [\mathbf{B} \cdot \mathbf{B}^{\mathsf{T}}]_{ij} \; \mathbf{p}(\boldsymbol{x}, \mathbf{t} | \boldsymbol{x}_0, \mathbf{t}_0) . \tag{2.148}$$

This is the celebrated Fokker-Planck equation which describes the time evolution of the probability that the stochastic process determined by (2.3) assumes the value $\boldsymbol{x}$ at time t when it had assumed the value $\boldsymbol{x}_o$ at time t_0.

Note, that the above Fokker-Planck equation holds for the stochastic differential equation (2.3) only within the framework of Ito calculus. The relation between SDE (2.3) and the Fokker-Planck equation (2.148) is slightly different when Stratonovitch calculus is applied!

2.6   Stratonovich Calculus

We take a quick look at Stratonovich calculus mentioned in section 2.4. We want to clarify the α-dependence of our results in sections 2.4 and 2.5. For this purpose it is sufficient to focus on processes satisfying the stochastic differential equation (2.3).
It is possible to show that a solution $\boldsymbol{x}(t)$ of the stochastic differential equation

$$\text{Ito}\ \ \alpha = 0: \qquad\qquad \partial_t\,\boldsymbol{x}(t) \ = \ \boldsymbol{A}[\boldsymbol{x}(t),t] \ + \ \mathbf{B}[\boldsymbol{x(t)},\mathbf{t}]\cdot\boldsymbol{\xi}(\mathbf{t}) \qquad\qquad (2.149)$$

solves a stochastic differential equation of the same form with different coefficients, this time however according to Stratonovich's calculus.

$$\text{Stratonovich}\ \ \alpha = \frac{1}{2}: \qquad \partial_t\,\boldsymbol{x}(t) \ = \ \overset{\mathcal{S}}{\boldsymbol{A}}[\boldsymbol{x}(t),t] \ + \ \overset{\mathcal{S}}{\mathbf{B}}[\boldsymbol{x}(t),t]\cdot\boldsymbol{\xi}(t) \ . \qquad (2.150)$$

We give a derivation for $\overset{\mathcal{S}}{\boldsymbol{A}}[\boldsymbol{x}(t),t]$ and $\overset{\mathcal{S}}{\mathbf{B}}[\boldsymbol{x}(t),t]$ in the one-dimensional case as the lower case coefficients $\overset{\mathcal{S}}{a}[x(t),t]$ and $\overset{\mathcal{S}}{b}[x(t),t]$ indicate. As a first step we solve the integral corresponding to equation (2.150).

$$x(t) \ = \ x(t_0) \ + \ \int_{t_0}^{t} ds \ \overset{\mathcal{S}}{a}[x(s),s] \ + \ \oint_{t_0}^{t} d\omega(s) \ \overset{\mathcal{S}}{b}[x(s),s] \ . \qquad (2.151)$$

The $\mathcal{S}$ on the second integral sign denotes a Stratonovich integral which has to be solved like a Riemann-Stieltjes integral as in equation (2.105) with $\alpha = 1/2$. The last term of equation (2.151) is the only one that differs from Ito's calculus and thus it is the only term that needs to be investigated. One can rewrite the last term as an Ito integral. We do so neglecting the mean square limit notation in the defintion of a Stratonovich integral and write

$$\oint_{t_0}^{t} d\omega(s) \ \overset{\mathcal{S}}{b}[x(s),s] \ \simeq \ \sum_i \left(\omega(t_i) - \omega(t_{i-1})\right) \ \overset{\mathcal{S}}{b}[x(\tau),\tau] \ , \qquad \text{with } \tau := \frac{1}{2}(t_i + t_{i-1})$$

$$= \ \sum_i \left(\omega(t_i) - \omega(\tau)\right) \ \overset{\mathcal{S}}{b}[x(\tau),\tau] \ + \ \sum_i \left(\omega(\tau) - \omega(t_{i-1})\right) \ \overset{\mathcal{S}}{b}[x(\tau),\tau] . \ (2.152)$$

$\overset{\mathcal{S}}{b}[x(\tau),\tau]$ can be approximated by extrapolation starting with $\overset{\mathcal{S}}{b}[x(t_{i-1}),t_{i-1}]$ at the left side of interval $[t_{i-1},t_i]$.

$$\overset{\mathcal{S}}{b}[x(\tau),\tau] \ = \ \overset{\mathcal{S}}{b}[x(t_{i-1}),t_{i-1}] \ + \ \left(\partial_x\,\overset{\mathcal{S}}{b}[x(t_{i-1}),t_{i-1}]\right)\left(x(\tau) - x(t_{i-1})\right)$$

$$+ \ \left(\partial_t\,\overset{\mathcal{S}}{b}[x(t_{i-1}),t_{i-1}]\right)\left(\tau - t_{i-1}\right) \qquad\qquad (2.153)$$

$$+ \ \frac{1}{2}\left(\partial_x^2\,\overset{\mathcal{S}}{b}[x(t_{i-1}),t_{i-1}]\right)\left(x(\tau) - x(t_{i-1})\right)^2 \ + \ \ldots.$$

Since $x(t)$ is a solution of Ito's stochastic equation (2.149) one can apply (2.149) to determine the infinitesimal displacement $x(\tau) - x(t_{i-1})$.

$$x(\tau) - x(t_{i-1}) \ = \ a[x(t_{i-1}),t_{i-1}]\left(\tau - t_{i-1}\right) \ + \ b[x(t_{i-1}),t_{i-1}]\left(\omega(\tau) - \omega(t_{i-1})\right) \ . \qquad (2.154)$$

Filling equation (2.154) into (2.153) and applying Ito's rules (2.128) to the infinitesimal displacements $dt = (\tau - t_{i-1})$ and $d\omega(t) = (\omega(\tau) - \omega(t_{i-1}))$ one obtains

$$
\begin{aligned}
\overset{s}{b}[x(\tau),\tau] \;=\; & \overset{s}{b}[x(t_{i-1}),t_{i-1}] \;+\; \left(\partial_t \overset{s}{b}[x(t_{i-1}),t_{i-1}]\right)(\tau - t_{i-1}) \\
& +\; a[x(t_{i-1}),t_{i-1}]\left(\partial_x \overset{s}{b}[x(t_{i-1}),t_{i-1}]\right)(\tau - t_{i-1}) \\
& +\; b[x(t_{i-1}),t_{i-1}]\left(\partial_x \overset{s}{b}[x(t_{i-1}),t_{i-1}]\right)\left(\omega(\tau) - \omega(t_{i-1})\right) \\
& +\; \frac{1}{2}\, b^2[x(t_{i-1}),t_{i-1}]\left(\partial_x^2 \overset{s}{b}[x(t_{i-1}),t_{i-1}]\right)(\tau - t_{i-1}) \qquad (2.155)
\end{aligned}
$$

Substituting the above result (2.155) into the second sum of equation (2.152) one derives

$$
\begin{aligned}
\oint_{t_0}^{t} d\omega(s)\, \overset{s}{b}[x(s),s] \;\simeq\; & \sum_i \left(\omega(t_i) - \omega(\tau)\right) \overset{s}{b}[x(\tau),\tau] \\
& +\; \sum_i \left(\omega(\tau) - \omega(t_{i-1})\right) \overset{s}{b}[x(t_{i-1}),t_{i-1}] \\
& +\; \sum_i \underbrace{\left(\omega(\tau) - \omega(t_{i-1})\right)(\tau - t_{i-1})}_{=\,0,\ \text{due to (2.127)}}\left(\partial_t \overset{s}{b}[x(t_{i-1}),t_{i-1}]\right) \\
& +\; \sum_i \underbrace{\left(\omega(\tau) - \omega(t_{i-1})\right)(\tau - t_{i-1})}_{=\,0,\ \text{due to (2.127)}}\, a[x(t_{i-1})]\left(\partial_x \overset{s}{b}[x(t_{i-1}),t_{i-1}]\right) \\
& +\; \sum_i \underbrace{\left(\omega(\tau) - \omega(t_{i-1})\right)^2}_{=\,(\tau - t_{i-1}),\ \text{due to (2.125)}}\, b[x(t_{i-1}),t_{i-1}]\left(\partial_x \overset{s}{b}[x(t_{i-1}),t_{i-1}]\right) \\
& +\; \frac{1}{2}\sum_i \underbrace{\left(\omega(\tau) - \omega(t_{i-1})\right)(\tau - t_{i-1})}_{=\,0,\ \text{due to (2.127)}}\, b^2[x(t_{i-1}),t_{i-1}]\left(\partial_x^2 \overset{s}{b}[x(t_{i-1}),t_{i-1}]\right) . \qquad (2.156)
\end{aligned}
$$

The first two terms on the r.h.s. of equation (2.156) make up a sum that approximates an Ito integral with time steps just half the size. In the fifth term one can replace $(\tau - t_{i-1})$ by $\frac{1}{2}(t_i - t_{i-1})$. The result, again written for the multi-dimensional case, is

$$
\oint_{t_0}^{t} d\boldsymbol{\omega}(s)\cdot\overset{s\mathsf{T}}{\mathbf{B}}[\boldsymbol{x}(s),s] \;=\; \int_{t_0}^{t} d\boldsymbol{\omega}(s)\cdot\overset{s\mathsf{T}}{\mathbf{B}}[\boldsymbol{x}(s),s] \;+\; \frac{1}{2}\int_{t_0}^{t} ds \sum_{i,j} \mathsf{B}_{ij}[\boldsymbol{x}(s),s]\left(\partial_i \overset{s\mathsf{T}}{\mathsf{B}}_{jk}[\boldsymbol{x}(s),s]\right). \qquad (2.157)
$$

Note, that the above connection (2.157) between Ito and Stratonovich integrals only holds for $\boldsymbol{x}(t)$ satisfying Ito's SDE (2.149) or Stratonovich's SDE (2.150). There is no *general* connection between Ito and Stratonovich integrals!

Substituting (2.157) into Stratonovich's integral equation (2.151) and comparing the coefficients with the integral solving Ito's stochastic differential equation (2.149) we obtain the following relations

$$
A_k \;=\; \overset{s}{A}_k \;+\; \frac{1}{2}\sum_{i,j} \overset{s}{\mathsf{B}}_{ij}\left(\partial_i \overset{s}{\mathsf{B}}_{kj}\right), \qquad (2.158)
$$

$$
\mathsf{B}_{jk} \;=\; \overset{s}{\mathsf{B}}_{jk} , \qquad (2.159)
$$

and conversely

$$\overset{s}{A}_k = A_k - \frac{1}{2} \sum_{i,j} \mathsf{B}_{ij} \left(\partial_i \mathsf{B}_{kj} \right), \tag{2.160}$$

$$\overset{s}{\mathsf{B}}_{jk} = \mathsf{B}_{jk} . \tag{2.161}$$

We see that a difference between Ito and Stratonovich calculus only occurs, if $\mathbf{B}$ depends on $\boldsymbol{x}(t)$, that is if $\partial_i \mathsf{B}_{kj} \neq 0$.

To conclude this section we write down the Fokker-Planck equation in Stratonovich's terms. One simply substitutes the coefficients $\boldsymbol{A}$ and $\mathbf{B}$ according to equations (2.158) and (2.159), and applies the product rule for differential operations to simplify the expression.

$$\partial_t\, p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) = - \sum_i \partial_i \overset{s}{A}_i\, p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) + \frac{1}{2} \sum_{i,j,k} \partial_i \overset{s}{\mathsf{B}}_{ik} \left(\partial_j \overset{s}{\mathsf{B}}_{jk}\, p(\boldsymbol{x}, t | \boldsymbol{x}_0, t_0) \right) . \tag{2.162}$$

2.7 Appendix: Normal Distribution Approximation

2.7.1 Stirling's Formula

We need Stirling's formula (2.163, 2.164) to prove Gauß's asymptotic approximation (2.165) of the binomial distribution. A derivation of Stirling's formula is outside the scope of this book. See [42] for a derivation based on Euler's summation formula.

$$n! \;=\; \sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n\,\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{2.163}$$

$$\ln n! \;=\; \ln\left(\sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n\right) \;+\; \ln\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

$$\ln n! \;=\; \frac{1}{2}\ln(2\pi) \;+\; (n+\frac{1}{2})\ln n \;-\; n \;+\; \mathcal{O}\left(\frac{1}{n}\right) \tag{2.164}$$

2.7.2 Binomial Distribution

We set forth to prove Eq. (2.165), i.e.,

$$\sqrt{\frac{n}{2}}\,2^{-n}\,\binom{n}{\frac{n}{2}+x\sqrt{\frac{n}{2}}} \;=\; \frac{1}{\sqrt{\pi}}\,\exp{-x^2}\,\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)\;. \tag{2.165}$$

Applying the natural logarithm on both sides of this equation we obtain

$$\ln\left[\sqrt{\frac{n}{2}}\,2^{-n}\,\binom{n}{\frac{n}{2}+x\sqrt{\frac{n}{2}}}\right] \;=\; \ln\left(\frac{1}{\sqrt{\pi}}\,\exp{-x^2}\right) \;+\; \ln\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$$

$$\ln\left[\sqrt{\frac{n}{2}}\,2^{-n}\,\binom{n}{\frac{n}{2}+x\sqrt{\frac{n}{2}}}\right] \;=\; -\frac{1}{2}\ln\pi \;-\; x^2 \;+\; \mathcal{O}\left(\frac{1}{n}\right)\;. \tag{2.166}$$

We will prove equation (2.166) by transforming the left hand side step by step. First, we utilize the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ for binomial coefficients.

$$\ln\left(\sqrt{\frac{n}{2}}\,2^{-n}\,\binom{n}{\frac{n}{2}+x\sqrt{\frac{n}{2}}}\right) \tag{2.167}$$

$$= \frac{1}{2}\,(\ln n - \ln 2) \;-\; n\ln 2 \;+\; \ln\left(\frac{n!}{(\frac{n}{2}+x\sqrt{\frac{n}{2}})!\,(\frac{n}{2}-x\sqrt{\frac{n}{2}})!}\right)$$

$$= \frac{1}{2}\ln n \;-\; \left(n+\frac{1}{2}\right)\ln 2 \;+\; \ln n! \;-\; \ln\left[\left(\frac{n}{2}+x\sqrt{\frac{n}{2}}\right)!\right] \;-\; \ln\left[\left(\frac{n}{2}-x\sqrt{\frac{n}{2}}\right)!\right]\;.$$

Applying Stirling's formula (2.164) we derive furthermore

$$= \frac{1}{2}\ln n \;-\; \left(n+\frac{1}{2}\right)\ln 2 \;+\; \frac{1}{2}\ln(2\pi) \;+\; (n+\frac{1}{2})\ln n \;-\; n \;+$$

$$-\; \frac{1}{2}\ln(2\pi) \;-\; \left(\frac{n}{2}+x\sqrt{\frac{n}{2}}+\frac{1}{2}\right)\ln\left(\frac{n}{2}+x\sqrt{\frac{n}{2}}\right) \;+\; \frac{n}{2}+x\sqrt{\frac{n}{2}} \;+$$

$$-\; \frac{1}{2}\ln(2\pi) \;-\; \left(\frac{n}{2}-x\sqrt{\frac{n}{2}}+\frac{1}{2}\right)\ln\left(\frac{n}{2}-x\sqrt{\frac{n}{2}}\right) \;+\; \frac{n}{2}-x\sqrt{\frac{n}{2}}$$

$$= \frac{1}{2}\ln n \; - \; \left(n+\frac{1}{2}\right)\ln 2 \; - \; \frac{1}{2}\ln(2\pi) \; + \; \left(n+\frac{1}{2}\right)\ln n \; +$$

$$- \; \left(\frac{n}{2} + x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\ln\left[\frac{n}{2}\left(1 + x\sqrt{\frac{2}{n}}\right)\right] \; +$$

$$- \; \left(\frac{n}{2} - x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\ln\left[\frac{n}{2}\left(1 - x\sqrt{\frac{2}{n}}\right)\right]$$

$$= \; (n+1)\ln n \; - \; \left(n+\frac{1}{2}\right)\ln 2 \; - \; \frac{1}{2}\ln 2 \; - \; \frac{1}{2}\ln\pi \; +$$

$$- \; \left(\frac{n}{2} + x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\left(\ln\frac{n}{2} + \ln\left(1 + x\sqrt{\frac{2}{n}}\right)\right) \; +$$

$$- \; \left(\frac{n}{2} - x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\left(\ln\frac{n}{2} + \ln\left(1 - x\sqrt{\frac{2}{n}}\right)\right) \; .$$

Performing a Taylor expansion of $\ln(1 \pm z)$ with respect to z we obtain,

$$= \; (n+1)\ln n \; - \; (n+1)\ln 2 \; - \; \frac{1}{2}\ln\pi \; +$$

$$- \; \left(\frac{n}{2} + x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\left(\ln\frac{n}{2} + x\sqrt{\frac{2}{n}} - \frac{x^2}{n} + \frac{x^3}{3}\sqrt{\frac{8}{n^3}} + \mathcal{O}\left(\frac{1}{n^2}\right)\right) \; +$$

$$- \; \left(\frac{n}{2} - x\sqrt{\frac{n}{2}} + \frac{1}{2}\right)\left(\ln\frac{n}{2} - x\sqrt{\frac{2}{n}} - \frac{x^2}{n} - \frac{x^3}{3}\sqrt{\frac{8}{n^3}} + \mathcal{O}\left(\frac{1}{n^2}\right)\right) \; ,$$

and expanding the products up to order $\mathcal{O}(1/n)$ we acquire the result, the right hand side of equation (2.166), results in

$$= \; (n+1)\left(\ln n - \ln 2\right) \; - \; \frac{1}{2}\ln\pi \; +$$

$$- \; \frac{n}{2}\ln\frac{n}{2} \; - \; x\sqrt{\frac{n}{2}} \; + \; \frac{x^2}{2} \; - \; \frac{x^3}{3}\sqrt{\frac{2}{n}} \; - \; x\sqrt{\frac{n}{2}}\ln\frac{n}{2} \; +$$

$$- \; x^2 \; + \; \frac{x^3}{\sqrt{2n}} \; - \; \frac{1}{2}\ln\frac{n}{2} \; - \; \frac{x}{\sqrt{2n}} \; + \; \mathcal{O}\left(\frac{1}{n}\right) \; +$$

$$- \; \frac{n}{2}\ln\frac{n}{2} \; + \; x\sqrt{\frac{n}{2}} \; + \; \frac{x^2}{2} \; + \; \frac{x^3}{3}\sqrt{\frac{2}{n}} \; + \; x\sqrt{\frac{n}{2}}\ln\frac{n}{2} \; +$$

$$- \; x^2 \; - \; \frac{x^3}{\sqrt{2n}} \; - \; \frac{1}{2}\ln\frac{n}{2} \; + \; \frac{x}{\sqrt{2n}} \; + \; \mathcal{O}\left(\frac{1}{n}\right)$$

$$= \; (n+1)\ln\frac{n}{2} \; - \; \frac{1}{2}\ln\pi \; +$$

$$- \; n\ln\frac{n}{2} \; - \; x^2 \; - \; \ln\frac{n}{2} \; + \; \mathcal{O}\left(\frac{1}{n}\right)$$

$$= \; -\frac{1}{2}\ln\pi \; - \; x^2 \; + \; \mathcal{O}\left(\frac{1}{n}\right) \qquad \text{q.e.d.}$$