universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Generalized Langevin Equations“

Verfasser

Maximilian Sadilek

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2012

Studienkennzahl lt. Studienblatt:	A 411
Studienrichtung lt. Studienblatt:	Diplomstudium Physik
Betreuerin / Betreuer:	Ao. Univ.Prof. Dr. Helmuth Hüffel

Abstract

This thesis is organized as follows: In chapter 1 generalized Langevin equations are presented as a class of stochastic differential-integral equations obtained by an application of the Zwanzig projection formalism.

The starting point is a statistical mechanical model describing a macroscopic system that interacts with a heat bath consisting of a large number of independent harmonic oscillators. The coupling between system and bath is assumed to be linear in the bath variables, but arbitrary in the system variables. This allows for an explicit integration of the bath equations of motion yielding an effective equation for the system. Stochasticity is introduced by the assumption that at some initial time, the bath variables are drawn at random from a canonical distribution.

In chapter 2 thermodynamic- and Markovian limits of generalized Langevin equations are studied. This investigation makes it possible to draw some interesting connections between generalized Langevin equations and stochastic differential equations and points in the direction of future research.

The last two chapters constituting the appendix are dedicated to the formal development of convergence theorems for stochastic processes that are necessary to handle the thermodynamic limits discussed in chapter 2. Apart from very few exceptions, all proofs are explicitly carried out.

Contents

Chapter 1

Generalized Langevin Equations

In 1908, Paul Langevin proposed the following heuristic equation in order to describe *Brownian motion*, i.e. the motion of a particle in a fluid due to collisions with the molecules of the fluid ([1]):

$$\dot{P}_t = -6\pi a\eta/m\, P_t + \xi(t). \tag{1.1}$$

Here P_t is the momentum of the particle at time t, a the radius of the particle and η the viscosity of the fluid. Langevin himself explained $\xi(t)$ to be a rapidly fluctuating random function which is required to obey

$$\langle P_t \xi(t)\rangle = 0 \quad \forall t,$$

where $\langle . \rangle$ denotes an "ensemble average“ not further defined. From a modern viewpoint, $\langle . \rangle$ denotes expectation values with respect to the Wiener measure (see below) and $\xi(t)$ can be identified with Gaussian white noise

$$\langle \xi(t)\rangle = 0 \quad \forall t$$
$$\langle \xi(t)\xi(t')\rangle = 12\pi a\eta\; k_B T\; \delta(t-t')$$

where k_B is the Boltzmann constant and T is the temperature of the fluid.

Equation (1.1) successfully provided a complementary picture to Einstein's and Smoluchovsky's description of Brownian motion in terms of diffusion equations (Einstein 1905, Smoluchovsky 1906), but it took quite some time before the underlying notions of *stochastic processes* (Wiener 1923, Kolmogorov 1938) and especially of *stochastic differential equations* (Itō, Stratonovich) were well understood mathematically.

As a first step, in 1923 Norbert Wiener showed the existence of a stochastic process W with the following properties:

(i) At some initial time $t_0 = 0$, the process is zero with probability one:

$$W_0 = 0;$$

(ii) The *sample paths* $t \to W_t$ are continuous with probability one;

(iii) W_t has independent increments with

$$W_t - W_s \sim \mathcal{N}(0, t-s) \text{ for } 0 \leq s < t.$$

This *Wiener process* proved very useful in the theory of Brownian motion in the sense that its sample paths constitute asymptotic solutions to the Langevin equation:

If the Brownian particle is at rest at time zero (i.e. $P_0 = 0$), then after $t >> m/6\pi a\eta$ it approaches a stationary state described by the equation

$$P_t = \int_0^t \xi(t')dt' \tag{1.2}$$

which has the solution ([7], p.80f)

$$P_t = \sqrt{12\pi a\eta\, k_B T}\, W_t.$$

Notice that we did not write

$$\dot{P}_t = \xi(t)$$

instead of (1.2) because the Wiener process is nowhere differentiable ([7], p.68). For the same reason, (1.1) is not well-defined mathematically. But if we write symbolically

$$\xi(t)dt = \sqrt{12\pi a\eta\, k_B T}\, dW_t$$

in the sense of (1.2), we may recast (1.1) into the following integral equation:

$$P_t - P_0 = -s \int_0^t P_{t'}dt' + \sqrt{2smk_BT} \int_0^t dW_{t'}. \tag{1.3}$$

with $s = 6\pi a\eta/m$. The second integral in (1.3) is understood as Riemann-Stieltjes integral over the Wiener measure dW associated with the Wiener process and a trivial example of an *Itō*- or *Stratonovich stochastic integral*. Consequently, (1.3) is a prototypical case of a *stochastic differential equation (SDE)* - its solution is called *Ornstein-Uhlenbeck process*.

Remark (Notation of (generalized) Langevin equations). In the following, we adopt the convention of symbolically writing (generalized) Langevin equations like (1.1) instead of the associated, more rigorous integral equations like (1.3).

As mentioned at the beginning, the Langevin equation (1.1) is a heuristic equation. It took quite a long time until it was given a solid foundation in Hamiltonian mechanics. The main idea of the works of Ford, Kac and Mazur ([2]) as well as Zwanzig ([4]) and much later Caldeira and Leggett ([6]) that addressed this problem is contained in the following program, which is attributed to Josiah Willard Gibbs, and taken from [2]:

(i) Solve the equations of motion of the mechanical system consisting of a Brownian particle coupled to heat bath. The solution consists of expressions for the coordinates and momenta at time t in terms of the initial coordinates and momenta.

(ii) Assume the initial coordinates and momenta of the heat bath to be distributed according to some statistical distribution, e.g. that of the canonical ensemble.

(iii) Show that the coordinate and momentum of the Brownian particle, as functions of time, will then represent stochastic processes (whose statistical properties arise from the initial distribution of the heat bath) of the kind familiar from standard theories.

While in [2] and [6] one of the primary aims was to develop a theory of *quantum Brownian motion* by studying a *quantum Langevin equation*, the scope of [4] is somewhat different:

By considering more general couplings between system and heat bath than in [2], [6], Zwanzig obtained equations that show many features of the Langevin equation (such as the fluctuation-dissipation relation) and indeed contain it as a special case, but have a much richer structure. Therefore he called them *generalized Langevin equations*.

Assumption A.1 (Classicality)**.** *The framework of this thesis is non-relativistic classical statistical mechanics. That is the reason why you will not find any $\hbar$ or c apart from the ones you just read.*

This chapter is dedicated to the derivation of generalized Langevin equations loosely following [4]. The starting point is the following situation: An arbitrary *system S* with characteristic timescale τ_S is brought in contact with a *heat bath B* (timescale τ_B). If $\tau_S >> \tau >> \tau_B$, then after a time τ the bath is in approximate equilibrium with respect to the system which effectively did not change in the meantime. This state will be denoted *initial state*, and the *initial time* is set to 0.

1.1 Hamiltonian and Equations of Motion

Remark (Generalized dynamics). Due to compactness of notation and greater generality, coordinates and momenta will be treated on equal footing and masses will be set equal to 1 which amounts to a suitable rescaling of coordinates. That is, the *state of the system* is collectively described by a vector X, and similarly the *state of the bath* is described by a vector Y. If necessary, the standard form of Hamiltonian dynamics can be regained by setting

$$X = \big(Q,P\big)^T$$
$$Y = \big(q_1,\dots q_n,p_1,\dots p_n\big)^T,$$

where $Q = (Q_1,\dots Q_d)$ etc. if d is the physical dimension and n is the number of components constituting the heat bath. This setting corresponds to the following representation of *symplectic matrices*:

$$C_X = \begin{pmatrix} 0 & \mathbb{1}_d \\ -\mathbb{1}_d & 0 \end{pmatrix} \tag{1.4}$$
$$C_Y = \begin{pmatrix} 0 & \mathbb{1}_{nd} \\ -\mathbb{1}_{nd} & 0 \end{pmatrix}, \tag{1.5}$$

where $\mathbb{1}_z$ is the unit matrix of size z.

Remark (Dimension). To simplify the following calculations, the dimension d will be set equal to 1 - it is straightforward to regain the full expressions if necessary.

We start with writing the *total Hamiltonian* H as a sum of the *system Hamiltonian* H_S and a *remainder Hamiltonian* H_R which collects bath-, interaction- and counterterms:

$$H(X,Y) =: H_S(X) + H_R(X,Y). \tag{1.6}$$

Assumption A.2 (Hamiltonian structure)**.**

- *$H_S(X)$ is an arbitrary Hamiltonian.*
- *H_R is given by the following quadratic form:*

$$H_R(X,Y) := \frac{1}{2}[Y - a(X)]^T\, K\, [Y - a(X)], \tag{1.7}$$

where K is a symmetric positive definite matrix and $a(X)$ is a vector-valued function.

By performing a suitable orthogonal transformation in the bath space and subsequent rescaling, K may always be written in the following form:

$$K := \begin{pmatrix} \Omega^2 & 0 \\ 0 & \mathbb{1}_n \end{pmatrix} \tag{1.8}$$

with $\Omega := diag(\omega_1, \dots \omega_n)$. Consequently, the *bath Hamiltonian* (which is the X-independent part of H_R) has the following standard representation:

$$H_B(Y) := \frac{1}{2} Y^T K Y = \frac{1}{2} \sum_{j=1}^{n} (p_j^2 + \omega_j^2 q_j^2). \tag{1.9}$$

We see that the heat bath consists of n independent harmonic oscillators with eigenfrequencies $\omega_1, \dots \omega_n$.

Assumption A.3 (Structure of the coupling)**.** *The coupling $a(X)$ can be written as a product in the following way:*

$$a(X) := \gamma_0 \, \Gamma(\Omega) \, G(X) \tag{1.10}$$

where γ_0 is a function of n,

$$\Gamma(\Omega) := \begin{pmatrix} \Gamma_q(\Omega) & 0 \\ 0 & \Gamma_p(\Omega) \end{pmatrix} \tag{1.11}$$

is a matrix-valued function of the eigenfrequencies and

$$G(X) := \begin{pmatrix} G_1(X) \\ G_2(X) \end{pmatrix} \tag{1.12}$$

is an arbitrary function of the system variables X.

Identifying the *interaction Hamiltonian* H_I as that term of H_R that depends both on X and Y we have

$$\begin{aligned} H_I(X,Y) &:= -\frac{1}{2}[a(X)^T K Y + Y^T K a(X)] \\ &= -Y^T K a(X) \\ &= -\gamma_0 \sum_{j=1}^{n} [\omega_j^2 \Gamma_q(\omega_j) q_j G_1(X) + \Gamma_p(\omega_j) p_j G_2(X)]. \end{aligned} \tag{1.13}$$

We notice that each oscillator couples to the system in the same way. The remaining term

$$H_{SI}(X) := \frac{1}{2} a(X)^T K a(X) = \frac{\gamma_0^2}{2} \sum_{j=1}^{n} [\omega_j^2 \Gamma_q(\omega_j)^2 G_1(X)^2 + \Gamma_p(\omega_j)^2 G_2(X)^2] \tag{1.14}$$

can be understood as *self-interaction Hamiltonian* and serves as a counterterm: If H_{SI} was not present,

- one can immediately see that H_R would not be positive;
- an unphysical renormalization of the *generalized force* would arise (see below).

Finally, we write down the *equations of motion*:

$$\dot{X} = C_X \nabla_X (H_S + H_R) = V(X) - C_X \tilde{W}(X) K[Y - a(X)] \tag{1.15}$$

$$\dot{Y} = C_Y \nabla_Y (H_R) = C_Y K[Y - a(X)], \tag{1.16}$$

where ∇_X and ∇_Y denote gradients with respect to X and Y and the following abbreviations are used:

- $V(X) := C_X \nabla_X H_S$ (the *generalized force*);
- $\tilde{W}(X) := \nabla_X a^T(X)$.

The time evolution of the states X and Y described by the equations of motion will be considered from 0 to $0 < T < \infty$ where T is arbitrary.

1.2 Zwanzig Projection Formalism

We start with solving the equation of motion for the heat bath (1.16) while treating $a(X)$ as an external field. Since different times are involved, every expression is labelled by the time it has to be evaluated at. By the method of variation of parameters we find

$$Y_t = \exp(t\, C_Y K)\, Y_0 - \int_0^t dt' \exp(t'\, C_Y K)\, C_Y\, K\, a(X_{t-t'})$$

and via integration by parts

$$Y_t - a(X_t) = \exp(t\, C_Y K)[Y_0 - a(X_0)] + \int_0^t dt' \exp(t'\, C_Y K) \frac{d}{dt'} a(X_{t-t'}).$$

Next we insert this result into the equation of motion for the system (1.15) and use the chain rule

$$\frac{d}{dt'} a(X_{t-t'}) = -\tilde{W}^T(X_{t-t'}) \dot{X}_{t-t'}$$

to obtain

$$\dot{X}_t = V(X_t) + \int_0^t dt' C_X \tilde{W}(X_t) K \exp(t' C_Y K) \tilde{W}^T(X_{t-t'}) \dot{X}_{t-t'} \\ - C_X \tilde{W}(X_t) K \exp(t C_Y K)[Y_0 - a(X_0)]. \tag{1.17}$$

This integral-differential equation has the following interesting structure:

- The first term on the r.h. side only depends on the actual state of the system (i.e. is local in time) and is the usual generalized force term (it is the only term remaining if the coupling is set to zero).
- The second term depends on the history of the state of the system X from 0 to t and is therefore called *memory function*.
- The third term depends on the actual state of the system and the initial values of bath and system.

Remark (Self-interaction). If the self-interaction term H_{SI} is disregarded , (1.17) becomes

$$\dot{X}_t = \tilde{V}(X_t) + \int_0^t dt' C_X \tilde{W}(X_t) K \exp(t' C_Y K) \tilde{W}^T(X_{t-t'}) \dot{X}_{t-t'} \\ - C_X \tilde{W}(X_t) K \exp(t C_Y K)[Y_0 - a(X_0)] \tag{1.18}$$

with

$$\tilde{V}(X_t) := V(X_t) - C_X \tilde{W}(X_t) K a(X_t),$$

so neglecting H_{SI} amounts to a renormalization of the generalized force. Especially, considering the important case of an *affine coupling*

$$a(X_t) = AX_t + b, \tag{1.19}$$

where A is an arbitrary, but nonzero matrix and b is an arbitrary vector in the bath space, we have

$$\tilde{W}(X_t) = \nabla_X a^T(X_t) = A^T.$$

Consequently, the only remaining term in (1.18) that depends both on the state of the system X_t and the coupling is $C_X A^T K A X_t$. If self-interaction is included, this term would not be present thus making the coupling *translationally invariant*.

Remark (Orthogonal dynamics). If we first solve the Hamiltonian equation for the system (1.15) and then plug the solution into the Hamiltonian equation of the heat bath (1.16), we obtain an equation that describes the so-called *orthogonal dynamics* of the total system (consisting of system and heat bath). This orthogonal dynamics equation together with equation (1.17) is completely equivalent to the original Hamiltonian equations.

However, the point is that because we assume the bath Hamiltonian to be quadratic and the interaction Hamiltonian to be linear in the bath variables (looking at the equations of motion, this corresponds to a *linear response* of the bath to the system), equation (1.17) becomes quite simple.

On the other hand, since both the system Hamiltonian H_S and the part of the coupling that depends on the state of the system $G(X)$ are (almost) arbitrary, it is not clear whether there even *exists* an analytical solution to equation (1.15), and consequently to the resulting orthogonal dynamics equation.

This is the reason why orthogonal dynamics will not be considered here any further.

Remark (Operator language). It is possible to generalize our discussion by introducing *projection operators* acting on the Liouville operator associated with the total Hamiltonian; equation (1.17) and the orthogonal dynamics equation then appear as projections of the Liouville equation onto the system and bath subspaces, respectively, and instead of a partial integration, the Dyson decomposition (Dyson 1949) is used to obtain the typical structure of the resulting equations.

But here we are following the original approach by Zwanzig [4], since it allows for explicit calculations.

To put the following into a more transparent form, we now define

- $\overline{W}(X) := \nabla_X G^T(X) = \gamma_0^{-1} \tilde{W}(X) \Gamma^{-1}$,
- $L(t) := \gamma_0^2 \Gamma K \exp(t C_Y K) \Gamma$,
- $F(t) := \gamma_0 \Gamma K \exp(t C_Y K)[Y_0 - a(X_0)]$.

Then we have ($\Gamma^T = \Gamma$)

$$\dot{X}_t = V(X_t) + \int_0^t dt' C_X \overline{W}(X_t) L(t') \overline{W}^T(X_{t-t'}) \dot{X}_{t-t'} - C_X \overline{W}(X_t) F(t). \quad (1.20)$$

1.3 Stochasticity

The aim of this chapter is to derive generalized Langevin equations, that is a class of equations of motion describing the time evolution of the state of a macroscopic

system that is subject to diffusive and randomly fluctuating forces with microscopic origin. Consequently, they should especially contain *stochastic terms*, which we have not encountered in our derivation so far.

Assumption A.4 (Initial distribution of the bath variables)**.** *At $t = 0$ the heat bath is in thermal equilibrium with respect to the initial state of the system.*

Hence if the bath temperature is given by $(k_B\beta)^{-1}$, where k_B is the Boltzmann constant, the values of the collective bath variable Y_0 are drawn from the following canonical distribution (X_0 is held fixed):

$$Y_0 \sim \exp[-\beta H(X_0, Y_0)] \sim \exp[-\beta H_R(X_0, Y_0)]. \tag{1.21}$$

This allows us to define *expectation values* in the following way:

$$\langle A \rangle := \frac{\int_{\mathbb{R}^{2n}} A \, \exp[-\beta H_R(X_0, Y_0)] \, dY_0}{\int_{\mathbb{R}^{2n}} \exp[-\beta H_R(X_0, Y_0)] \, dY_0}. \tag{1.22}$$

By a short calculation we obtain the mean and autocovariance of the process $F(t)$, which we call from now on *noise source*:

$$\langle F(t) \rangle = 0, \tag{1.23}$$

$$\langle F(t) F^T(t') \rangle = \beta^{-1} L(t - t'). \tag{1.24}$$

In order to obtain (1.24), the following identity was used:

$$K^{-1}[\exp(t\, C_Y K)]^T = \exp(-t\, C_Y K) K^{-1},$$

which can be verified by a series expansion of the exponential. Since

- $F(t)$ is linear in $[Y_0 - a(X_0)]$ which has a Gaussian distribution and
- its autocovariance $\langle F(t) F(t') \rangle$ is time-translation invariant,

it is a (mean zero) *stationary Gaussian process*; therefore it is completely described by equations (1.23) and (1.24).

Especially interesting is equation (1.24) relating the noise source $F(t)$ with $L(t)$: viewed as a general form of the *fluctuation-dissipation relation*, it gives $L(t)$ the meaning of a *generalized friction coefficient*.

Remark (Components of Y_0). Recall that the heat bath is modelled by n independent harmonic oscillators represented by the components of Y and each of them couples to the system individually; as a consequence, the bath and interaction Hamiltonians can be expressed as sums over the bath index $j = 1 \ldots n$, where

each summand corresponds to the bath or interaction Hamiltonian of one distinct oscillator (compare (1.9), (1.13)).

The same is true for the self-interaction Hamiltonian (compare (1.14)); because of (1.21) and $H_R = H_B + H_I + H_{SI}$, this especially means that *the components of Y_0 are statistically independent and obey a Gaussian distribution*:

$$q_{0,j} \sim \exp[-\frac{\beta}{2}\omega_j^2(q_{0,j} - \gamma_0\Gamma_q(\omega_j)G_1(X_0))^2] \quad \text{and}$$
$$p_{0,j} \sim \exp[-\frac{\beta}{2}(p_{0,j} - \gamma_0\Gamma_p(\omega_j)G_2(X_0))^2]$$

for all $j = 1 \ldots n$. We can put this result into the following, more useful form:

$$q_{0,j} = \gamma_0\Gamma_q(\omega_j)G_1(X_0) + \beta^{-1/2}\omega_j^{-1}\xi_j \quad \text{and} \tag{1.25}$$
$$p_{0,j} = \gamma_0\Gamma_p(\omega_j)G_2(X_0) + \beta^{-1/2}\eta_j, \tag{1.26}$$

where ξ_j and η_j are *mutually independent sequences of i.i.d. (independent, identically distributed) random variables* distributed according to a standard normal distribution. We write

$$\xi_j, \eta_j \sim \mathscr{N}[0,1].$$

Since the initial coordinates and momenta of the oscillators composing the heat bath are statistically independent stochastic variables, it is reasonable to assume that the oscillator frequencies are also independent random variables.

Assumption A.5 (Frequency distribution)**.** *The oscillator frequencies ω_j ($j = 1 \ldots n$) are uniformly distributed in $[0, n^a]$,*

$$\omega_j = n^a \nu_j, \quad \nu_j \ \textit{i.i.d.}, \quad \nu_1 \sim \mathscr{U}[0,1] \tag{1.27}$$

for some $a \in (0,1)$.

This means that with increasing number of oscillators n, the oscillator spectrum becomes increasingly broad and dense (since $a < 1$). As a consequence of the assumed distribution, the *mean spectral density* is defined as

$$\Delta\omega := \frac{n^a}{n} = n^{a-1}. \tag{1.28}$$

Remark (Notation of probabilities and expectation values). The process X_t is defined on a $3n$-dimensional probability space that is either generated by the $2n$

components of Y_0 together with $(\omega_1, \dots \omega_n)$ or by $(\xi_1, \dots \xi_n, \eta_1, \dots \eta_n, \nu_1, \dots \nu_n)$ (compare equations (1.25),(1.26) and (1.27)).

In the following we will generally refer to the second set of generating random variables. Especially, probabilities and expectation values with respect to the ν-, ξ- or η-components only or some combination of them (like ξ- and η-components) will be denoted $\mathbb{P}$ and $\mathbb{E}$ with subscript, for instance $\mathbb{P}_\nu$, $\mathbb{E}_\nu$ or $\mathbb{E}_{\xi\eta}$. Expectation values with respect to Y_0 (denoted $\langle . \rangle$) will not be used any further.

1.4 Generalized Langevin Equations

In fact, (1.20) together with (1.23), (1.24) already has the structure we expect from generalized Langevin equations. To put the result into a more transparent form, we are going to determine the representation of the noise source $F(t)$ and friction coefficient $L(t)$ in terms of ξ, η and ω and plug the result into (1.20). Recall that the friction coefficient is given by

$$L(t) := \gamma_0^2 \Gamma K \exp(t C_Y K) \Gamma$$

and the noise source by

$$F(t) := \gamma_0 \Gamma K \exp(t C_Y K)[Y_0 - a(X_0)],$$

where the matrices K, C_Y and Γ have been defined as (see also (1.8), (1.5) and (1.11)):

$$C_Y = \begin{pmatrix} 0 & \mathbb{1}_n \\ -\mathbb{1}_n & 0 \end{pmatrix} \qquad K = \begin{pmatrix} \Omega^2 & 0 \\ 0 & \mathbb{1}_n \end{pmatrix} \qquad \Gamma = \begin{pmatrix} \Gamma_q & 0 \\ 0 & \Gamma_p \end{pmatrix}$$

with $\Omega := diag(\omega_1, \dots \omega_n)$. Let us first look at the expression $\exp(t C_Y K)$. Noting that

$$(C_Y K)^2 = -\begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix}^2,$$

we have for any $l \in \mathbb{N}$

$$(C_Y K)^{2l} = (-1)^l \begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix}^{2l} \quad \text{and}$$

$$(C_Y K)^{2l+1} = (-1)^l \begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix}^{2l+1} \begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix}^{-1} C_Y K$$

and consequently

$$\begin{aligned}
\exp(t\,C_Y K) &= \sum_{l=0}^{\infty} \frac{t^l}{l!}(C_Y K)^l \\
&= \sum_{l=0}^{\infty} [\frac{t^{2l}}{(2l)!}(C_Y K)^{2l} + \frac{t^{2l+1}}{(2l+1)!}(C_Y K)^{2l+1}] \\
&= \sum_{l=0}^{\infty} [\frac{(-1)^l t^{2l}}{(2l)!} \begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix}^{2n} \\
&+ \frac{(-1)^l t^{2l+1}}{(2l+1)!} \begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix}^{2l+1} \begin{pmatrix} 0 & \Omega^{-1} \\ -\Omega & 0 \end{pmatrix}] \\
&= \begin{pmatrix} \cos(\Omega t) & \Omega^{-1}\sin(\Omega t) \\ -\Omega \sin(\Omega t) & \cos(\Omega t) \end{pmatrix}.
\end{aligned}$$

Therefore

$$K \exp(t\,C_Y K) = \begin{pmatrix} \Omega^2 \cos(\Omega t) & \Omega \sin(\Omega t) \\ -\Omega \sin(\Omega t) & \cos(\Omega t) \end{pmatrix}.$$

We arrive at the following expression for the *friction coefficient*:

$$L(t) = \gamma_0^2 \begin{pmatrix} \Gamma_q^2 \Omega^2 \cos(\Omega t) & \Gamma_q \Gamma_p \Omega \sin(\Omega t) \\ -\Gamma_q \Gamma_p \Omega \sin(\Omega t) & \Gamma_p^2 \cos(\Omega t) \end{pmatrix}. \tag{1.29}$$

Recall that (trigonometric) functions of Ω are defined by their power series; Consequently, since Ω is diagonal, this is also true for any functions of Ω. For instance,

$$\Gamma_q^2 \Omega^2 \cos(\Omega t) = diag(\Gamma_q^2(\omega_1)\omega_1^2 \cos(\omega_1 t), \ldots \Gamma_q^2(\omega_n)\omega_n^2 \cos(\omega_n t)).$$

In a similar way we obtain the *noise source*; it is given by

$$F(t) = \gamma_0 \beta^{-1/2} \begin{pmatrix} \Gamma_q(\omega_1)\omega_1[\cos(\omega_1 t)\xi_1 + \sin(\omega_1 t)\eta_1] \\ \vdots \\ \Gamma_q(\omega_n)\omega_n[\cos(\omega_n t)\xi_n + \sin(\omega_n t)\eta_n] \\ \Gamma_p(\omega_1)[\cos(\omega_1 t)\xi_1 - \sin(\omega_1 t)\eta_1] \\ \vdots \\ \Gamma_p(\omega_n)[\cos(\omega_n t)\xi_n - \sin(\omega_n t)\eta_n] \end{pmatrix}. \tag{1.30}$$

By using the trigonometric identities

$$\begin{aligned}
\cos(\alpha \pm \beta) &= \cos(\alpha)\cos(\beta) \mp \sin(\alpha)\sin(\beta) \\
\sin(\alpha \pm \beta) &= \sin(\alpha)\cos(\beta) \pm \cos(\alpha)\sin(\beta),
\end{aligned}$$

one may check the validity of the fluctuation-dissipation relation

$$\mathbb{E}_{\xi\eta}[F(t)F^T(t')] = \beta^{-1}L(t-t'), \tag{1.31}$$

which is true for every choice of oscillator frequencies. Let us next make use of the fact that although $\overline{W}(X) = \nabla_X G(X)^T$ is a $2 \times 2n$-matrix, the first n columns respectively the second n columns are identical; this allows us to write

$$C_X \overline{W}(X_t)F(t) = CW(X_t)Z_n(t)$$

with the 2×2-matrix

$$W(X) := \nabla_X \left(G_1(X), G_2(X)\right) \tag{1.32}$$

and the 2-vector

$$Z_n(t) := \gamma_0 \beta^{-1/2} \begin{pmatrix} \sum_{j=1}^n \Gamma_q(\omega_j)\omega_j[\cos(\omega_j t)\xi_j + \sin(\omega_j t)\eta_j] \\ \sum_{j=1}^n \Gamma_p(\omega_j)[\cos(\omega_j t)\xi_j - \sin(\omega_j t)\eta_j] \end{pmatrix} \tag{1.33}$$

which we call *forcing*. Analogously, rewrite the integrand of the memory function in (1.20):

$$C_X \overline{W}(X_t)L(t)\overline{W}^T(X_{t-t'})\dot{X}_{t-t'} = C_X W(X_t)K_n(t)W^T(X_{t-t'})\dot{X}_{t-t'}$$

with the *memory kernel*

$$K_n(t) := \tag{1.34}$$
$$\gamma_0^2 \begin{pmatrix} \sum_{j=1}^n \Gamma_q^2(\omega_j)\omega_j^2 \cos(\omega_j t) & \sum_{j=1}^n \Gamma_q(\omega_j)\Gamma_p(\omega_j)\omega_j \sin(\omega_j t) \\ -\sum_{j=1}^n \Gamma_q(\omega_j)\Gamma_p(\omega_j)\omega_j \sin(\omega_j t) & \sum_{j=1}^n \Gamma_p^2(\omega_j)\cos(\omega_j t) \end{pmatrix}.$$

Summarizing, generalized Langevin equations obtained by the Zwanzig projection formalism have the following structure:

$$\dot{X}_t = V(X_t) + \int_0^t dt' C_X W(X_t)K_n(t')W^T(X_{t-t'})\dot{X}_{t-t'} - C_X W(X_t)Z_n(t). \tag{1.35}$$

Chapter 2

Thermodynamic- and Markovian Limits

In the last chapter we applied the Zwanzig projection formalism to the equations of motion of a classical Hamiltonian system consisting of a distinguished particle (the *system*) and n harmonic oscillators (the *heat bath*). The coupling between system and heat bath was assumed to be linear in the bath variables Y, but rather general with respect to the system variables X (assumptions A.2 and A.3).

Further we assumed that at some initial time $t = 0$, the bath variables Y_0 are distributed according to a canonical distribution with temperature $(k_B\beta)^{-1}$ (assumption A.4) and that the oscillator frequencies $\omega_j, j = 1 \ldots n$ are uniformly distributed over the interval $[0, n^a]$ where $a \in (0,1)$ (assumption A.5).

This procedure (compare with the first two steps of the Gibbs-program mentioned at the beginning of the last chapter) yielded a class of stochastic differential-integral equations for the particle's position and momentum which we called *generalized Langevin equations*. The following structure was obtained:

$$\dot{X}_t = V(X_t) + \int_0^t dt' C_X W(X_t) K_n(t') W^T(X_{t-t'}) \dot{X}_{t-t'} - C_X W(X_t) Z_n(t) \qquad (2.1)$$

where

$$C_X = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

is the symplectic matrix in X-space,

$$V(X) = C_X \nabla_X H_S(X) = \begin{pmatrix} \nabla_P H_S(X) \\ -\nabla_Q H_S(X) \end{pmatrix}$$

is the generalized force or Hamiltonian vector field generated by H_S and

$$W(X) = \nabla_X G(X) = \begin{pmatrix} \nabla_Q G_1(X) & \nabla_Q G_2(X) \\ \nabla_P G_1(X) & \nabla_P G_2(X) \end{pmatrix} \tag{2.2}$$

contains information about the coupling between system and bath (compare A.3). The memory kernel

$$K_n(t) = \\ \gamma_0^2 \begin{pmatrix} \sum_{j=1}^n \Gamma_q^2(\omega_j)\omega_j^2 \cos(\omega_j t) & \sum_{j=1}^n \Gamma_q(\omega_j)\Gamma_p(\omega_j)\omega_j \sin(\omega_j t) \\ -\sum_{j=1}^n \Gamma_q(\omega_j)\Gamma_p(\omega_j)\omega_j \sin(\omega_j t) & \sum_{j=1}^n \Gamma_p^2(\omega_j)\cos(\omega_j t) \end{pmatrix}$$

is a 2×2-matrix-valued stochastic process defined on the probability space generated by the random variables

$$\nu_j \text{ i.i.d.}, \quad \nu_1 \sim \mathscr{U}[0,1].$$

Recall that $\omega_j = n^a \nu_j \; \forall j$ and $a \in (0,1)$. $K_n(t)$ further depends on the size of the heat bath n and the frequency-dependent part of the coupling $\Gamma(\Omega)$ (compare A.3). The forcing

$$Z_n(t) := \gamma_0 \beta^{-1/2} \begin{pmatrix} \sum_{j=1}^n \Gamma_q(\omega_j)\omega_j[\cos(\omega_j t)\xi_j + \sin(\omega_j t)\eta_j] \\ \sum_{j=1}^n \Gamma_p(\omega_j)[\cos(\omega_j t)\xi_j - \sin(\omega_j t)\eta_j] \end{pmatrix}.$$

is a 2-dimensional zero-mean stationary Gaussian process defined on the probability space generated by the mutually independent random variables

$$\begin{aligned} &\nu_j \text{ i.i.d.}, \quad \nu_1 \sim \mathscr{U}[0,1], \\ &\xi_j \text{ i.i.d.}, \quad \xi_1 \sim \mathscr{N}(0,1), \\ &\eta_j \text{ i.i.d.}, \quad \eta_1 \sim \mathscr{N}(0,1). \end{aligned}$$

$Z_n(t)$ also depends on the size of the heat bath n and the frequency-dependent part of the coupling $\Gamma(\Omega)$. The memory kernel and the forcing are related through the *fluctuation-dissipation relation*

$$\mathbb{E}_{\xi\eta}[Z_n(t)Z_n^T(t')] = \beta^{-1} K_n(t-t') \tag{2.3}$$

which follows directly from (1.31).

In this chapter, we will study the limiting behaviour of (2.1) with respect to two situations:

(i) The size of the heat bath n grows to infinity (*thermodynamic limit*). This affects the memory kernel $K_n(t)$ and the forcing $Z_n(t)$ whose thermodynamic limits will be denoted $K(t)$ and $Z(t)$. Consequently 2.1 becomes

$$\dot{X}_t = V(X_t) + \int_0^t dt' C_X W(X_t) K(t') W^T(X_{t-t'}) \dot{X}_{t-t'} - C_X W(X_t) Z(t). \tag{2.4}$$

(ii) Assuming (i), the coupling between system and heat bath (more precisely, the components of $\Gamma(\Omega)$ and $G(X)$) is chosen in such a way that the *memory function*

$$\int_0^t dt' C_X W(X_t) K(t') W^T(X_{t-t'}) \dot{X}_{t-t'}$$

becomes local in time. This corresponds to a *Markovian limit* of the stochastic process described by (2.4).

2.1 Thermodynamic Limits of Memory Kernel and Forcing

In order to simplify notation, we introduce the following quantities:

$$g_q(\omega) := \omega \Gamma_q(\omega) \tag{2.5}$$
$$g_p(\omega) := \Gamma_p(\omega). \tag{2.6}$$

Next we define an *effective coupling constant*:

$$\gamma := n^{\frac{1-a}{2}} \gamma_0 = (\Delta\omega)^{-1/2} \gamma_0. \tag{2.7}$$

where the second equality is just (1.28). The reason for this choice will become apparent below. Using (2.5), (2.6) and (2.7), the memory kernel and the forcing become

$$Z_n(t) = \gamma \beta^{-1/2} (\Delta\omega)^{1/2} \begin{pmatrix} \sum_{j=1}^n g_q(\omega_j)[\cos(\omega_j t)\xi_j + \sin(\omega_j t)\eta_j] \\ \sum_{j=1}^n g_p(\omega_j)[\cos(\omega_j t)\xi_j - \sin(\omega_j t)\eta_j] \end{pmatrix} \tag{2.8}$$

and

$$K_n(t) = \tag{2.9}$$
$$\gamma^2 \Delta\omega \begin{pmatrix} \sum_{j=1}^n g_q^2(\omega_j)\cos(\omega_j t) & \sum_{j=1}^n g_q(\omega_j) g_p(\omega_j) \sin(\omega_j t) \\ -\sum_{j=1}^n g_q(\omega_j) g_p(\omega_j) \sin(\omega_j t) & \sum_{j=1}^n g_p^2(\omega_j)\cos(\omega_j t) \end{pmatrix}.$$

In the following, we will study the limiting behaviour of K_n and Z_n for $n \to \infty$. We could do this separately for any component. However, since the components have a very similar structure, it is very convenient to restrict our investigations to the (1,1)-component of K_n respectively the 1-component of Z_n: in that sense, set for the moment

$$g_p(\omega_j) \equiv 0$$

and consider the 1-dimensional problem

$$Z_n(t) = \gamma\,\beta^{-1/2}(\Delta\omega)^{1/2}\sum_{j=1}^{n} g_q(\omega_j)[\cos(\omega_j t)\xi_j + \sin(\omega_j t)\eta_j]$$

$$K_n(t) = \gamma^2\Delta\omega\sum_{j=1}^{n} g_q^2(\omega_j)\cos(\omega_j t)$$

where, by a slight abuse of notation, Z_n and K_n have been identified with their non-zero components. One finds that in the limit $n\to\infty$, if g_q is positive and bounded with bound c_1 such that

$$g_q(\omega) \le c_2\;\omega^{-\lambda} \text{ for some } c_2 > 0 \text{ and } \lambda > 1/2,$$

then $K_n(t)$ converges to the integral

$$K(t) := \gamma^2\int_0^\infty g_q^2(\omega)\cos(\omega t)\,d\omega$$

ν-almost surely and $Z_n(t)$ converges to the stationary Gaussian process $Z(t)$ defined by

$$\mathbb{E}_{\xi\eta}Z(t) = 0$$
$$\mathbb{E}_{\xi\eta}[Z(t+s)\,Z(s)] = \beta^{-1}K(t)$$

ν-almost surely and ξ,η-weakly in $\mathscr{C}[0,T]$ (T is arbitrary, but finite). The detailed proof can be found in the appendix, where γ has been set to 1 and g_q is denoted by g, for simplicity (cf. theorems A.2.1 and B.4.1).

Remark (Renormalization of the coupling constant). Now we can see that the introduction of an effective coupling constant γ was necessary to avoid divergencies in the limit $n\to\infty$. This is a simple case of renormalization.

As mentioned above, very similar calculations and proofs can be carried through for the remaining components of K_n and Z_n. Therefore it is natural to impose the following assumption:

Assumption A.6 (Bounds on frequency coupling)**.** *g_q and g_p are positive and bounded with bound c_1 respectively c_2 such that*

$$g_q(\omega) \le c_3\;\omega^{-\lambda} \textit{ for some } c_3 > 0 \textit{ and } \lambda > 1/2 \tag{2.10}$$
$$g_p(\omega) \le c_4\;\omega^{-\mu} \textit{ for some } c_4 > 0 \textit{ and } \mu > 1/2. \tag{2.11}$$

As a result, in the limit $n \to \infty$ the memory kernel (2.9) converges to

$$K(t) = \gamma^2 \begin{pmatrix} \int_0^\infty g_q^2(\omega)\cos(\omega t)d\omega & \int_0^\infty g_q(\omega)g_p(\omega)\sin(\omega t)d\omega \\ -\int_0^\infty g_q(\omega)g_p(\omega)\sin(\omega t)d\omega & \int_0^\infty g_p^2(\omega)\cos(\omega t)d\omega \end{pmatrix} \tag{2.12}$$

ν-almost surely and in $L^2[0,T]$ or pointwise. Also, the forcing (2.8) converges to a stationary Gaussian process $Z(t)$ with zero mean and autocovariance

$$\mathbb{E}_{\xi\eta}[Z(t+s)\,Z^T(s)] = \beta^{-1}K(t) \tag{2.13}$$

ν-almost surely and ξ,η-weakly in $\mathscr{C}[0,T]$.

2.2 Markovian Limit

In this section, we are looking for conditions such that (2.4) becomes local in time. Obviously, this can only be the case for all t if *each component* of

$$\Lambda(t,t') := W(X_t)K(t')W^T(X_{t-t'}) \tag{2.14}$$

is either identically zero or proportional to the Dirac delta distribution $\delta(t')$, since then the integral term in (2.4) only contributes at time t. Calculating the components of $\Lambda(t,t')$ by inserting (2.2) and (2.12) into (2.14) shows that it is sufficient to require

Assumption A.7 (Markov condition on coupling)**.**

$$g_q(\omega) \equiv d_q \tag{2.15}$$
$$g_p(\omega) \equiv d_p \tag{2.16}$$

where d_q, d_p are constants, and

$$\{G_1(X), G_2(X)\}_{t,t'} = 0 \quad \forall t,t' \tag{2.17}$$

where

$$\{A(X),B(X)\}_{t,t'} = \nabla_Q A(X)\,|_{X=X_t}\,\nabla_P B(X)\,|_{X=X_{t'}} - \nabla_Q B(X)\,|_{X=X_t}\,\nabla_P A(X)\,|_{X=X_{t'}} \tag{2.18}$$

is a generalization of the Poisson bracket (which corresponds to $t = t'$). In order to see that this is true we will first calculate the components of $K(t)$ assuming (2.15) and (2.16). We have

$$\int_0^\infty \exp(i\omega t)d\omega = \int_{-\infty}^\infty \theta(\omega)\exp(i\omega t)d\omega = \sqrt{\frac{\pi}{2}}\Big(\delta(t) + \frac{i}{\pi}\mathrm{pv}\frac{1}{t}\Big) \tag{2.19}$$

where $\theta(t)$ is the Heaviside step function and pv denotes the principal value. Equality in (2.19) has to be understood in a distributional sense; in the last step the symmetric convention of the Fourier transform is adopted. By taking real and imaginary parts

$$\int_0^\infty \cos(\omega t)d\omega = \Re[\int_0^\infty \exp(i\omega t)d\omega] = \sqrt{\frac{\pi}{2}}\,\delta(t)$$
$$\int_0^\infty \sin(\omega t)d\omega = \Im[\int_0^\infty \exp(i\omega t)d\omega] = \frac{1}{\sqrt{2\pi}}\,\mathrm{pv}\frac{1}{t}$$

and setting for the sake of simplicity $d_q = d_p = (\frac{\pi}{2})^{1/4}$ we obtain

$$K(t) = \gamma^2 \begin{pmatrix} \delta(t) & \mathrm{pv}\frac{1}{\pi t} \\ -\mathrm{pv}\frac{1}{\pi t} & \delta(t) \end{pmatrix}. \tag{2.20}$$

Because of (2.13) we see that the components of the forcing $Z^1(t)$ and $Z^2(t)$ become white noise processes with nonvanishing correlation ($i, j = 1, 2$):

$$\mathbb{E}_{\xi\eta} Z^i(t) = 0 \qquad \forall i \tag{2.21}$$
$$\mathbb{E}_{\xi\eta}[Z^i(t)Z^i(t')] = \beta^{-1}\gamma^2\delta(t-t') \qquad \forall i \tag{2.22}$$
$$\mathbb{E}_{\xi\eta}[Z^i(t)Z^j(t')] = \beta^{-1}\gamma^2\mathrm{pv}\frac{1}{\pi(t-t')} \qquad \forall i \neq j \tag{2.23}$$

ν-almost surely. If we use (2.20) and (2.17) in (2.14) we obtain

$$\Lambda(t,t') = \gamma^2\delta(t')W(X_t)W(X_{t-t'})^T = \gamma^2\delta(t')W(X_t)W(X_t)^T. \tag{2.24}$$

This eventually proves the validity of assumption A.7: It was only due to (2.17) that the off-diagonal terms of (2.20) cancelled out. Finally we can determine the *Markovian generalized Langevin equations*

$$\dot{X}_t = V(X_t) + \gamma^2 C_X W(X_t)W^T(X_t)\dot{X}_t - C_X W(X_t)Z(t). \tag{2.25}$$

There are four major possibilities of choosing $G_1(X)$ and $G_2(X)$ such that (2.17) is fulfilled:

(i) $G_1(X) = G_2(X)$.

(ii) $G_1(X) = e_1$ where e_1 is a constant.

(iii) $G_2(X) = e_2$ where e_2 is a constant.

(iv) $G(X) = AX + b$ where A is a 2×2-matrix with $\det(A) = 0$ and b is a constant 2-vector.

Remark (Interpretation as Stochastic Differential Equations). Since the components of the forcing $Z^1(t)$ and $Z^2(t)$ are correlated, (2.25) cannot be associated with a 2-*dimensional* stochastic differential equation in the Itō- or Stratonovich sense for this would require $Z(t) = (Z^1(t), Z^2(t))^T$ to be a 2-dimensional Gaussian white noise process *with stochastically independent components*.

The only way to associate (2.25) with a 1-*dimensional* stochastic differential equation is to set all but one components of $W(X_t)$ to zero. This corresponds to the cases (ii) and (iii) with the additional requirement that the non-constant component of $G(X)$ only depends on either Q or P.

Consider for instance the situation $G_2(X) = e_2, G_1(X) = S(Q)$ where e_2 is a constant and $S(Q)$ is an arbitrary function of Q. The two components of (2.25) become

$$\begin{aligned}\dot{Q} &= \nabla_P H_S(X) \\ \dot{P} &= -\nabla_Q H_S(X) - \gamma^2 (\nabla_Q S(Q))^2 \dot{Q} - \nabla_Q S(Q) Z^1(t)\end{aligned}$$

while we have from (2.21) and (2.22)

$$\begin{aligned}&\mathbb{E}_{\xi\eta} Z^1(t) = 0 \\ &\mathbb{E}_{\xi\eta}[Z^1(t) Z^1(t')] = \beta^{-1}\gamma^2 \delta(t-t').\end{aligned}$$

For the special case $S(Q) = Q$ and $H_S(X) = \frac{P^2}{2}$ we obtain an equation which is up to constant factors similar to the original Langevin equation (1.1).

Appendix A

Thermodynamic Limit of the Memory Kernel

The memory kernel was defined as

$$K_n(t) := \sum_{j=1}^{n} g^2(\omega_j)\cos(\omega_j t)\,\Delta\omega \tag{A.1}$$
$$\text{with } \Delta\omega \equiv n^{a-1}$$

on the probability space generated by the random variables

$$\nu_j := n^{-a}\omega_j, \quad \nu_j \text{ i.i.d.}, \quad \nu_1 \sim \mathscr{U}[0,1] \tag{A.2}$$

for some $a \in (0,1)$.

In this chapter of the appendix, I will study the limit $n \to \infty$ of the memory kernel on a finite time interval $[0,T]$ which turns out to exist ν-almost surely and which is given by

$$K(t) := \int_0^\infty g^2(\omega)\cos(\omega t)\,d\omega. \tag{A.3}$$

While the first section is dedicated to the derivation of a general proving tool related to almost sure convergence and includes rather simple textbook-proofs (compare for instance [9]), the second section carries out the actual proof of $K_n(t) \xrightarrow{\text{a.s.}} K(t)\ \forall t$ following closely [13] and concludes with some interesting remarks.

A.1 Criterion for Almost Sure Convergence

Based on the Borel-Cantelli-lemma and Chebyshev's inequality, a useful criterion for almost sure convergence of a sequence of random variables is derived.

A.1.1 Almost sure convergence

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a *probability space* and $\{X_n : n \in \mathbb{N}\}$ a *sequence of random variables*. *Almost sure convergence* of the sequence to the random variable X is defined as follows:

Definition A.1.1 (Almost sure convergence)**.**

$$\begin{aligned} &X_n \to X \text{ almost surely } (X_n \xrightarrow{\text{a.s.}} X) \text{ for } n \to \infty \\ &\iff \mathbb{P}(\omega \in \Omega \mid X_n(\omega) \to X(\omega) \text{ for } n \to \infty) = 1, \end{aligned} \tag{A.4}$$

where "$\to$" on the right hand side means convergence in the sense of ordinary calculus:

Definition A.1.2 (Convergence in $\mathbb{R}^n$)**.**

$$\begin{aligned} &X_n(\omega) \to X(\omega) \text{ for } n \to \infty \\ &\iff \forall \varepsilon > 0 \; \exists N = N(\varepsilon, \omega) : \forall n > N \, |X_n(\omega) - X(\omega)| < \varepsilon \\ &(\omega \text{ fixed}). \end{aligned} \tag{A.5}$$

Now define

$$A_n(\varepsilon) := \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| \geq \varepsilon\}$$

and its *limes superior*

$$A(\varepsilon) := A_n(\varepsilon) \; i.o. \equiv \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n(\varepsilon)$$

where "i.o." means "infinitely often". This allows for an equivalent definition of a.s. convergence:

Lemma A.1.1 (Almost sure convergence 2)**.**

$$X_n \xrightarrow{a.s.} X \; for \; n \to \infty \iff \mathbb{P}(A(\varepsilon)) = 0 \quad \forall \varepsilon > 0. \tag{A.6}$$

Proof. Convert the ε/δ-statement into set language using $\bigcap$ and $\bigcup$. Look at the complementary set and replace ε by m^{-1}. Since its measure $\mathbb{P}(\bigcap_{m=1}^{\infty} A(m^{-1}))$ is bounded from below by $\mathbb{P}(A(m^{-1}))$ $\forall m > 0$ and from above by $\sum_{m=1}^{\infty} \mathbb{P}(A(m^{-1}))$ due to σ-additivity, the result follows. □

A.1.2 Borel-Cantelli-Lemma and Chebyshev's inequality

To estimate the r.h. side of (A.6), the first part of the *Borel-Cantelli-Lemma* is very useful, especially in combination with a version of *Chebyshev's inequality*.

Lemma A.1.2 (Borel-Cantelli 1)**.**

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \Rightarrow \mathbb{P}(A_n \; i.o.) = 0 \tag{A.7}$$

Proof. First notice that $\bigcup_{n=N}^{\infty} A_n$ is a decreasing function in N and therefore $A = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n$ its limit for $N \to \infty$. By continuity of the measure we have $\mathbb{P}(A) = \lim_{N\to\infty} \mathbb{P}(\bigcup_{n=N}^{\infty} A_n) \leq \lim_{N\to\infty} \sum_{n=N}^{\infty} \mathbb{P}(A_n) = 0$, where the last step is possible due to the assumption of the lemma. □

Lemma A.1.3 (Generalized Chebyshev's inequality)**.** *Let $g\colon \mathbb{R}^+ \to \mathbb{R}^+$ be strictly increasing. Then*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(g(|X|))}{g(a)} \; \forall a > 0. \tag{A.8}$$

where $\mathbb{E}$ denotes the expectation value associated with $\mathbb{P}$.

Proof. Define $A := \{h(X) \geq a\}$, $h\colon \mathbb{R} \to \mathbb{R}^+$, $a > 0$;
$\Rightarrow h(X) \geq a\, I_A$ with I_A being the index function of A. Taking the expectation value on both sides and setting $h(X) = g(|X|)$ yields the result. □

A.1.3 A useful criterion for almost sure convergence

Putting the pieces together, we arrive at a very useful *criterion for a.s. convergence*:

Theorem A.1.1 (Almost sure convergence criterion)**.**

$$\exists b \in \mathbb{N}\colon \sum_{n=1}^{\infty} \mathbb{E}(|X_n - X|^b) < \infty \Rightarrow X_n \xrightarrow{a.s.} X \tag{A.9}$$

Proof. First bound the summands in Lemma A.1.2 for $A_n = A_n(\varepsilon)$ by virtue of the generalized Chebyshev's inequality with the expectation of $g(|X_n - X|) = |X_n - X|^b$, $b \in \mathbb{N}$; then if there exists a b s.t. the sum converges, due to the comparison test and the lemmata A.1.2 and A.1.1 the result follows. □

Remark. For *stochastic processes* $X(t)$ in time t, it is necessary to specify how $|X_n(t) - X(t)|$ in the definition of a.s. convergence shall be understood. Later on, I will consider the cases of pointwise convergence in time and convergence in L^2 on a finite interval $[0, T]$.

A.2 Thermodynamic Limit of the Memory Kernel

In this section, I will use the criterion introduced in the preceding section to show that for a sufficiently bounded coupling $g(\omega)$, in the thermodynamic limit the memory kernel converges almost surely to a simple integral. The proof follows closely Kupferman *et al.* [13]. Moreover, the optimal rate of convergence is estimated and a heuristic introduction to the Monte-Carlo-approximation of integrals is given (following [10]), thus providing a different view on the result.

A.2.1 Almost sure convergence of the memory kernel

First I will derive upper bounds for the individual summands and their expectation values from quite general assumptions on $g(\omega)$.

Lemma A.2.1 (Upper bounds on the summands and their expectations)**.** *Suppose that g is a positive bounded real-valued function with bound c_1 such that*

$$g(\omega) \leq c_2\, \omega^{-\lambda} \text{ for some } c_2 > 0 \text{ and } \lambda > 1/2. \tag{A.10}$$

Let $t \in \mathbb{R}$. Define

$$h(\omega,t) := g^2(\omega)\cos(\omega t), \tag{A.11}$$

$$\mu_n(t) := n^{-a}\int_0^{n^a} h(\omega,t)\,d\omega, \quad a \in (0,1). \tag{A.12}$$

Then h and μ_n are bounded from above by

$$|h(\omega,t)| \leq \min(c_1^2, c_2^2\omega^{-2\lambda}) \equiv h^*(\omega), \tag{A.13}$$

$$|\mu_n(t)| \leq [c_1^2 + \frac{c_2^2}{2\lambda-1}]n^{-a} \equiv \mu_n^*, \tag{A.14}$$

$\forall\, t \in \mathbb{R}$.

Proof. (A.13) is clear since g is positive and $|\cos(.)|$ is bounded by 1. To prove (A.14), first use the triangle inequality and (A.13). Next, define $\omega_0 := (\frac{c_2}{c_1})^{1/\lambda}$ (this is the point where the two bounds on g coincide) and split the integral in the

following manner:

$$
\begin{aligned}
&\int_0^{n^a} \min(c_1^2, c_2^2\omega^{-2\lambda})\, d\omega = \int_0^{\omega_0} c_1^2\, d\omega + \int_{\omega_0}^{n^a} c_2^2\omega^{-2\lambda}\, d\omega \\
&\leq \int_0^{\omega_0} c_1^2\, d\omega + \int_{\omega_0}^{\infty} c_2^2\omega^{-2\lambda}\, d\omega \\
&= \begin{cases} \int_0^{\omega_0} c_1^2\, d\omega + \int_1^{\infty} c_2^2\omega^{-2\lambda}\, d\omega + \boxed{\int_{\omega_0}^{1} c_2^2\omega^{-2\lambda}\, d\omega} & c_1 \geq c_2 \\ \int_{\omega_0}^{\infty} c_2^2\omega^{-2\lambda}\, d\omega + \int_0^{1} c_1^2\, d\omega + \boxed{\int_{1}^{\omega_0} c_1^2\, d\omega} & c_1 < c_2 \end{cases} \\
&\leq \begin{cases} \int_0^{\omega_0} c_1^2\, d\omega + \int_1^{\infty} c_2^2\omega^{-2\lambda}\, d\omega + \boxed{\int_{\omega_0}^{1} c_1^2\, d\omega} & c_1 \geq c_2 \\ \int_{\omega_0}^{\infty} c_2^2\omega^{-2\lambda}\, d\omega + \int_0^{1} c_1^2\, d\omega + \boxed{\int_{1}^{\omega_0} c_2^2\omega^{-2\lambda}\, d\omega} & c_1 < c_2 \end{cases} \\
&= c_1^2 + \frac{c_2^2}{2\lambda - 1}
\end{aligned}
$$

using that $\omega^{-2\lambda}$ is strictly decreasing and the definition of ω_0. □

Theorem A.2.1 (Convergence of the memory kernel)**.** *Let ν_j be i.i.d. (independent and identically distributed) and $\nu_1 \sim \mathscr{U}[0,1]$ (uniformly distributed on [0,1]) and let $\omega_j = n^a\nu_j$, $a \in (0,1)$. Under the assumptions of Lemma A.2.1, the memory kernel* (A.1) *converges for $n \to \infty$ to the function* (A.3) *ν-almost surely and* $\begin{cases} \text{in } L^2[0,T] \\ \text{pointwise on } [0,\infty) \end{cases}$ *with respect to t.*

Proof. Because

$$
\mathbb{E}_\nu K_n(t) = \sum_{j=1}^{n} n^{-a}\Delta\omega \int_0^{n^a} g^2(\omega_j)\cos(\omega_j t)\, d\omega_j = \int_0^{n^a} g^2(\omega)\cos(\omega t)\, d\omega \quad \text{(A.15)}
$$

implies $\mathbb{E}_\nu K_n(t) \to K(t)$ uniformly with respect to $t \in [0,\infty)$, it is sufficient to show that ν-almost surely

$$
K_n(t) - \mathbb{E}_\nu K_n(t) \to 0 \begin{cases} \text{in } L^2[0,T] \\ \text{pointwise on } [0,\infty) \end{cases}. \quad \text{(A.16)}
$$

Here, $\mathbb{E}_\nu$ denotes the expectation value with respect to the probability measure

induced by $(\nu_1, \dots \nu_n)$. For instance,

$$\mathbb{E}_\nu f(\nu_1, \dots \nu_n) = \int_0^1 \cdots \int_0^1 f(\nu_1, \dots \nu_n)\, d\nu_1 \dots d\nu_n$$
$$= n^{-na} \int_0^{n^a} \cdots \int_0^{n^a} f(n^{-a}\omega_1, \dots n^{-a}\nu_n)\, d\omega_1 \dots d\omega_n.$$

In the following, I will restrict myself to the second case, thus omitting the t-dependence of the occurring functions which makes the proof more transparent. The first case can be obtained quite easily *mutatis mutandis*.

In order to show (A.16), according to Theorem A.1.1 it is sufficient to find an integer b such that the sequence

$$\sigma_n := \mathbb{E}_\nu \left| K_n - \mathbb{E}_\nu K_n \right|^{2b}$$

is summable (the replacement $b \to 2b$ is due to convenience). Using the definitions of $h(\omega)$ and μ_n as given in Lemma A.2.1 and omitting the absolute value because there is an even number of factors, one can rewrite the sequence as

$$\sigma_n = \Delta\omega^{2b}\, \mathbb{E}_\nu \left| \sum_{j=1}^{n} [h(\omega_j) - \mu_n] \right|^{2b} = \Delta\omega^{2b}\, \mathbb{E}_\nu \prod_{i=1}^{2b} \sum_{j_i=1}^{n} [h(\omega_{j_i}) - \mu_n]$$

or equivalently, defining the *centered joint moments of degree 2b* as

$$V_{j_1, j_2, \dots j_{2b}} := \mathbb{E}_\nu \prod_{i=1}^{2b} [h(\omega_{j_i}) - \mu_n], \tag{A.17}$$

$$\sigma_n = \Delta\omega^{2b} \sum_{j_1=1}^{n} \cdots \sum_{j_{2b}=1}^{n} V_{j_1, j_2, \dots j_{2b}}. \tag{A.18}$$

Now look at the (multi-)index set $J := \{j_1, j_2, \dots j_{2b}\}$: Each (trivial) equivalence relation $j_i \sim j_k \iff j_i = j_k$ on J induces a partition $J/\sim$. In total, there exist B_{2b} partitions of J where B_{2b} is the *Bell number* which only depends on the cardinality of J which is $2b$. Quite obviously, the multiple sum $\sum_{j_1=1}^{n} \cdots \sum_{j_{2b}=1}^{n}$ can be split into B_{2b} terms each of which contains only those elements with certain indices equal to each other and not equal to the rest (this is exactly the aforementioned partition).

For example, take $2b = 3$ (for the sake of the argument, here b is no integer):

$$\sum_{a,b,c} = \underbrace{\sum_{a=b=c}}_{k=1} + \underbrace{\sum_{a=b\neq c} + \sum_{a=c\neq b} + \sum_{b=c\neq a}}_{k=2} + \underbrace{\sum_{a\neq b\neq c}}_{k=3}$$

where the terms were arranged by introducing the index k as the *number of distinct indices* in J (here, $J = \{a,b,c\}$). Notice that the number of terms corresponding to a given k does only depend on $2b$ and especially not on the cardinality of the index set of any index.

Now, regroup the sum in (A.18) similar to the example by the number of distinct indices k (which runs from 1 to b because each index has to occur at least twice due to the independence of $h(\omega_i)$ and $h(\omega_j)$ for $i \neq j$). Because of the total symmetry of $V_{j_1,j_2,\ldots j_{2b}}$, there are n^k equal terms for each k (recall that each index runs from 1 to n) which I call V_k and which look like

$$V_k = \prod_{r=1}^{k} \frac{1}{n^a} \int_0^{n^a} [h(\omega) - \mu_n]^{m_r} \, d\omega$$

where $m_1, m_2, \ldots, m_k \geq 2$ and $m_1 + m_2 + \cdots + m_k = 2b$. Now use (A.13) and (A.14) to bound V_k in the following way:

$$\begin{aligned} |V_k| &\leq \prod_{r=1}^{k} \left(\frac{1}{n^a} \int_0^{n^a} [h^*(\omega) - \mu_n^*]^{m_r} \, d\omega\right) \\ &= \prod_{r=1}^{k} \sum_{l_r=0}^{m_r} \binom{m_r}{l_r} (\mu_n^*)^{m_r - l_r} n^{-a} \int_0^{n^a} [h^*(\omega)]^{l_r} \, d\omega) \\ &\leq \prod_{r=1}^{k} \left(\frac{M}{n^a}\right) = \left(\frac{M}{n^a}\right)^k \end{aligned} \tag{A.19}$$

where M is a constant. Going from the second to the third line is possible since when l_r is zero, the summand goes like $(n^{-a})^{m_r}$ and $m_r \geq 2$. When $l_r > 0$ the integral in the summand is convergent (for $n \to \infty$), so the summand goes to zero at least as fast as n^{-a}.

In total one arrives at

$$\begin{aligned} \sigma_n &\leq D(\Delta\omega)^{2b} \sum_{k=1}^{b} n^k n^{-ak} \\ &\leq E n^{2b(a-1)} n^{b(1-a)} = E n^{b(a-1)} \end{aligned} \tag{A.20}$$

where D and E are constants and the sum could be bounded because the summand is increasing with k. Now, for $b > (1-a)^{-1}$, σ_n is summable. □

A.2.2 Remarks

Remark (Optimal convergence). To estimate a rate of convergence of K_n to K (t-dependence omitted), look at the mean squared convergence. Because of the

triangle inequality we have

$$\mathbb{E}_\nu \left|K_n - K\right|^2 \leq \mathbb{E}_\nu \left|K_n - \mathbb{E}_\nu K_n\right|^2 + \left|K - \mathbb{E}_\nu K_n\right|^2$$

(notice that the second term on the r.h. side is non-random). The first term on the r.h. side, which is nothing but the variance of K_n, can be estimated in the following way:

$$Var_\nu K_n = n\Delta\omega^2 Var_\nu[g^2(\omega)\cos(\omega)] \leq \frac{n^{2a}}{n}\frac{1}{n^a}\int_0^{n^a} g^4(\omega)\, d\omega \leq C_1\Delta\omega$$

where C_1 is the value of the integral which is finite due to the assumptions on g (cf. equation (A.10)).

The second term can be bounded using (A.15) and again (A.10):

$$\left|K - \mathbb{E}_\nu K_n\right| \leq \int_{n^a}^{\infty} g^2(\omega)\, d\omega \leq \int_{n^a}^{\infty} \left(\frac{c_2}{\omega^\lambda}\right)^2 d\omega = \frac{c_2^2}{2\lambda - 1}\frac{1}{n^{a(2\lambda-1)}}.$$

Together we have

$$\mathbb{E}_\nu \left|K_n - K\right|^2 \leq \frac{C_1}{n^{1-a}} + \frac{C_2}{n^{2a(2\lambda-1)}}. \tag{A.21}$$

Notice that both terms of the asymptotic bound have a nonpositive exponent but while the first exponent becomes less negative with increasing a and fixed λ, the opposite is the case with the second term. Therefore, an estimate for optimal convergence is given by equating the two exponents which yields

$$a = [1 + 2(2\lambda - 1)]^{-1}. \tag{A.22}$$

Remark (Connection with Monte-Carlo-integral approximation). The result of the preceding subsection may be shown in a different way which is intuitionally more satisfying: The *memory kernel*

$$K_n(t) = \sum_{j=1}^{n} g^2(\omega_j)\cos(\omega_j t)\,\Delta\omega$$

can be viewed as a *Monte-Carlo-approximation* of the integral

$$K(t) := \int_0^{\infty} g^2(\omega)\cos(\omega t)\, d\omega.$$

To make that clear, I will give a short introduction to the Monte-Carlo-integral approximation (cf. [10], where also *Mathematica*®-implementations are available).

- In ordinary calculus, the *Mean Value Theorem for Integrals* states that the mean value of a function f over an interval $[a,b] \in \mathbb{R}$ can be expressed as the function value at an intermediate point c given that f is continuous on $[a,b]$:

$$f \in \mathscr{C}[a,b] \Rightarrow \exists c \in [a,b] : \frac{1}{b-a}\int_a^b f(x)\,dx = f(c)$$

- Now replace the r.h. side by an average of f over n points equally spaced over $[a,b]$:

$$\frac{1}{b-a}\int_a^b f(x)\,dx \approx \frac{1}{n}\sum_{k=1}^{n} f(c_k) \quad \text{with}$$
$$c_k = a + \frac{b-a}{n}(k-\frac{1}{2}) \quad \text{and} \quad k = 1,2,\ldots n.$$

  The error of this *Composite Midpoint Rule* decreases (for $f \in \mathscr{C}^2[a,b]$) with n^{-2} (*without proof*).

- Finally, averaging f over *n randomly distributed* points yields the so called *Monte-Carlo-approximation* of the integral, which however only converges with $n^{-1/2}$ (the proof relies on the *central limit theorem* and will not be given here).

Appendix B

Thermodynamic Limit of the Forcing

In this chapter of the appendix, the limit $n \to \infty$ of the forcing

$$Z_n(t) := \beta^{-1/2} \sum_{j=1}^{n} g(\omega_j)[\xi_j \cos(\omega_j t) + \eta_j \sin(\omega_j t)](\Delta\omega)^{1/2} \tag{B.1}$$
$$\text{with } \Delta\omega \equiv n^{a-1}, \quad \omega_j = n^a \nu_j, \quad a \in (0,1),$$

defined on the probability space generated by the mutually independent random variables

$$\nu_j \text{ i.i.d.}, \quad \nu_1 \sim \mathscr{U}[0,1], \tag{B.2}$$
$$\xi_j \text{ i.i.d.}, \quad \xi_1 \sim \mathscr{N}(0,1), \tag{B.3}$$
$$\eta_j \text{ i.i.d.}, \quad \eta_1 \sim \mathscr{N}(0,1) \tag{B.4}$$

is studied.

In that respect, let $\mathbb{E}_{\xi\eta}$ denote expectation values with respect to the Gaussian measure associated with $(\xi,\eta) = (\xi_1, \xi_2, \ldots, \eta_1, \eta_2, \ldots)$. For processes f that only depend on finitely many components $(\xi_1, \ldots \xi_n, \eta_1 \ldots \eta_n)$ this reduces to

$$\begin{aligned}
&\mathbb{E}_{\xi\eta} f(\xi_1, \ldots \xi_n, \eta_1 \ldots \eta_n) \\
&= (2\pi)^{-n} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(\xi_1, \ldots \xi_n, \eta_1 \ldots \eta_n) \\
&\exp[-\frac{1}{2}(\xi_1^2 + \cdots + \xi_n^2 + \eta_1^2 + \cdots + \eta_n^2)]\, d\xi_1 \ldots d\xi_n\, d\eta_1 \ldots d\eta_n.
\end{aligned}$$

$Z_n(t)$, or more precisely its *continuous version* ($\to$ next section), is shown to converge to a *stationary Gaussian process* $Z(t)$ with mean zero and autocorrelation function

$$\mathbb{E}_{\xi\eta}[Z(t+s)Z(s)] = \beta^{-1} \int_0^\infty g^2(\omega) \cos(\omega t)\, d\omega = \beta^{-1} K(t), \tag{B.5}$$

where $K(t)$ is the limit of the memory kernel (derived in the previous section) in accordance with the fluctuation-dissipation theorem, *almost surely in* ν, *weakly in* ξ *and* η and ξ, η*-a.s. uniformly on* $[0,T]$ *with finite* T. The proof of the above-mentioned statement (as given in [13]) heavily relies on a theorem for weak convergence of (sample path-)continuous stochastic processes supposedly first published by Gikhman and Skorokhod ([3] p. 450, see also [8] p. 98); this suggests the following structure of this chapter:

Starting with the theorem of *Kolmogorov-Chentsov* (with minor modifications, the proof is taken from [14]) to provide a necessary criterion for the existence of continuous sample paths (section B.1), I will turn to the characterization of stochastic processes in terms of their *(finite-dimensional) distributions* and *laws* in order to introduce the notion of weak convergence (section B.2; compare [5] and [12], although the latter is only available in German editions). The theorems of *Prokhorov*, *Arzelà-Ascoli* and finally *Gikhman-Skorokhod* (compare [12], [8] and [3]) will provide a criterion for weak convergence of continuous stochastic processes (section B.3).

Only then I will give the proof of the introductory statement (section B.4).

B.1 Kolmogorov-Chentsov theorem

Consider a real-valued stochastic process $X = \{X_t : t \in I\}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ with a continuous index set $I = [0,T] \in \mathbb{R}$. Let $\mathbb{E}$ denote the expectation value associated with $\mathbb{P}$.

Definition B.1.1 (Kolmogorov's continuity condition)**.** If there exist positive constants α, β and C such that

$$\mathbb{E}\,|X_t - X_s|^\alpha \leq C\,|t-s|^{1+\beta} \quad \forall\, s,t \in I, \tag{B.6}$$

then X is said to fulfill Kolmogorov's continuity condition.

This definition almost always comes together with the following theorem:

Theorem B.1.1 (Kolmogorov-Chentsov)**.** *A real-valued stochastic process* X *on* $I = [0,T]$ *that fulfills* (B.6) *has a* continuous version $\overline{X}$ *on* I*, i.e.*

$$t \mapsto \overline{X}_t \quad \textit{is continuous on } I \textit{ almost surely,} \tag{B.7}$$

$$\forall t \in I,\ X_t = \overline{X}_t \textit{ almost surely.} \tag{B.8}$$

Moreover, if $\gamma < \beta/\alpha$*, then for every compact subset* $K \subseteq I$

$$\max_{s \neq t \in K} \frac{\left|\overline{X}_t - \overline{X}_s\right|}{|t-s|^\gamma} < \infty \textit{ almost surely.} \tag{B.9}$$

Remark. Of course, (B.9) (*Hölder-continuity* of the extended sample paths $\overline{X_t}$) implies (B.7) (*continuity* of the extended sample paths $\overline{X}_t$) for $K = I$.

Remark. Actually it is sufficient to consider only an arbitrary countable dense subset $D \subset I$ (see proof); for similar reasons, I need not be closed. In that case, the index set of $\overline{X}$ has to be substituted by the closure $\overline{D} = \bar{I}$ (this does not change the almost sure statements (B.7) and (B.8)). Similarly, I in (B.9) will become $\bar{I}$.

Proof. The main idea of the proof is quite frequently used, but here I will basically follow the argumentation line of [14]. Without loss of generality, I will set $T = 1$ to simplify matters.

Consider the set D of all rationals in $I = [0,T] = [0,1]$ which is obviously countable and dense in I (for general countable dense subsets $D' \in I$ the following is true *mutatis mutandis*). Each point in D can be described by its binary coordinate of the form $k/2^m$ ($0 < k < 2^m$) and therefore $D = \bigcup_{m=1}^{\infty} L_m$ where $L_m = \bigcup_{0<k<2^m} k/2^m$ is the *dyadic level*. Let further $x(t)$ be a real-valued function on D and suppose that for $\gamma > 0$ and $C < \infty$,

$$|x(s) - x(t)| \leq C\,|s-t|^{\gamma} \tag{B.10}$$

for all neighboring pairs in the same dyadic level L_m.

Now for each point $t \in \overline{D} = \bar{I} = I$ there is a point $s_m(t) \in L_m$ within distance 2^{-m} (if there are two, choose one). Clearly, $s_m(t) \to t$ for $m \to \infty$ and since $|s_{m+1}(t) - s_m(t)| \leq 2 \cdot 2^{-m}$ we have due to (B.10)

$$|x(s_{m+1}(t)) - x(s_m(t))| \leq 2C2^{-\gamma m},$$

so for any t, $x(s_m(t))$ is a convergent Cauchy-sequence with the limit

$$\bar{x}(t) := \lim_{m\to\infty} x(s_m(t)). \tag{B.11}$$

For $t \in D$, $\bar{x}(t)$ coincides with $x(t)$ by construction. Moreover, using the triangle inequality one can estimate the distance of the m$^{\text{th}}$ element of the sequence to the limit:

$$|\bar{x}(t) - x(s_m(t))| \leq \sum_{k=m}^{\infty} |x(s_{k+1}(t)) - x(s_k(t))| \leq 2C \sum_{l=0}^{\infty} \left(2^{-\gamma}\right)^{l+m} = C''\,2^{-\gamma m}$$

where C'' is only a function of γ (and especially not a function of t).

Now choose distinct points $s,t \in \overline{D} = \bar{I} = I$; obviously there exists some $\hat{m} = m(s,t)$ such that

$$2^{-\hat{m}-1} \leq |s-t| < 2^{-\hat{m}}.$$

Since the distance between u and $s_{\hat{m}}(u)$ is itself bounded by $2 \cdot 2^{-\hat{m}}\ \forall u$, again due to the triangle inequality we have

$$|s_{\hat{m}}(s) - s_{\hat{m}}(t)| \leq C''' 2^{-\hat{m}}$$

and because of (B.10)

$$|\overline{x}(s_{\hat{m}}(s)) - \overline{x}(s_{\hat{m}}(t))| \leq C'''' 2^{-\gamma m}$$

since $s_{\hat{m}}(s) \in D\ \forall s$ and so $\overline{x}(s_{\hat{m}}(s)) = x(s_{\hat{m}}(s))\ \forall s$. In total, one gets

$$\begin{aligned}
&|\overline{x}(s) - \overline{x}(t)| \\
&\leq \underbrace{|\overline{x}(s) - \overline{x}(s_{\hat{m}}(s))|}_{\leq C'' 2^{-\gamma \hat{m}}} \\
&+ \underbrace{|\overline{x}(s_{\hat{m}}(s)) - \overline{x}(s_{\hat{m}}(t))|}_{\leq C'''' 2^{-\gamma \hat{m}}} \\
&+ \underbrace{|\overline{x}(s_{\hat{m}}(t)) - \overline{x}(t)|}_{\leq C'' 2^{-\gamma \hat{m}}} \\
&\leq (2C'' + C'''') 2^{-\gamma \hat{m}} \leq C' |s-t|^{\gamma},
\end{aligned} \tag{B.12}$$

so $x(t)$ on D extends to a Hölder-continuous function $\overline{x}(t)$ on I with the same Hölder-exponent γ.

Now, the cardinality of L_m grows like

$$\#L_m \sim 2^m$$

(for I closed this is exact). Take two neighboring points s,t in L_m (that is, $|s-t| = 2^{-m}$); due to (B.6) and Markov's-inequality we have

$$\begin{aligned}
&\mathbb{P}(|X_t - X_s| \geq C|t-s|^{\gamma}) \leq \frac{\mathbb{E}(X_t - X_s)^{\alpha}}{C|t-s|^{\gamma\alpha}} \\
&\leq |t-s|^{1+\beta-\gamma\alpha} \\
&= 2^{-m-m\beta+m\gamma\alpha}
\end{aligned}$$

Since the number of neighboring pairs grows like 2^m, we have

$$\begin{aligned}
&\mathbb{P}(B_m) \leq C' 2^{-m\beta+m\gamma\alpha} \quad \text{with} \\
&B_m := \{|X_t - X_s| \geq C|t-s|^{\gamma} \text{ for some neighboring } s,t \in L_m\}.
\end{aligned} \tag{B.13}$$

Since $\alpha > 0$ and $\beta > 0$ by assumption, one can choose γ so small that $-\beta + \gamma\alpha < 0$. Then due to the Borel-Cantelli-lemma we have

$$\sum_{m=1}^{\infty} \mathbb{P}(B_m) < \infty \to \mathbb{P}(B_m \text{ i.o.}) = 0$$

which means that with probability one, B_m occurs for only finitely many m; hence with probability one

$$|X_t - X_s| \leq C\,|t-s|^{\gamma}$$

for neighboring points s,t in the same dyadic level (due to $\mathbb{P}(X = x) = 0$ for continuous random variables, $\leq$ can be written instead of $<$). This, in turn, completes the proof because from the last equation (which is nothing but (B.10)) (B.12) follows. □

B.2 Stochastic Processes and Weak Convergence

Sample space measures associated with stochastic processes are introduced and discussed within the context of weak convergence.

B.2.1 Stochastic processes and measures

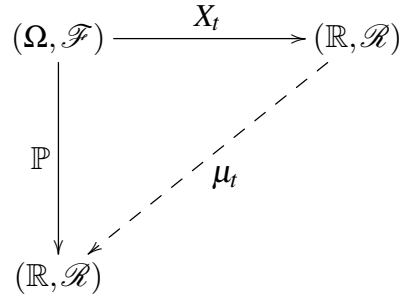
Given a real-valued stochastic process X on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with index set I (i.e. a map $X\colon I \times \Omega \to \mathbb{R}$ s.t. $\forall t \in I$, the projection $X_t\colon \Omega \to \mathbb{R}$ is $\mathscr{F}/\mathscr{R}$-measurable, where $\mathscr{R}$ is the Borel-σ-algebra over $\mathbb{R}$), several measures are associated with X.

(i) For each $t \in I$, the *projection* or *coordinate variable* $X_t\colon \Omega \to \mathbb{R}$ induces a measure called *distribution at t* on $\mathbb{R}$:

Definition B.2.1 (Distribution at t)**.** Given a stochastic process $X = \{X_t\colon t \in I\}$ on an index set I, the distribution of X at a point $t \in I$ is the pushforward-measure

$$\mu_t(A) := \mathbb{P}(\omega \in \Omega \mid X_t(\omega) \in A), \quad A \in \mathscr{R} \tag{B.14}$$

where $\mathscr{R}$ is the Borel-σ-algebra over $\mathbb{R}$.

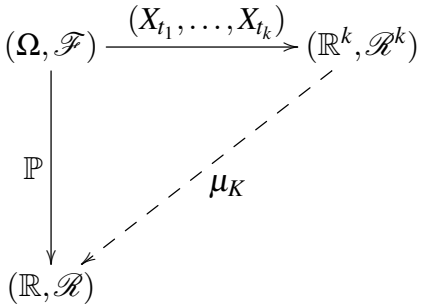


(ii) If *finitely* many points in I are considered, the previous definition generalizes in the following way:

Definition B.2.2 (Finite-dimensional distributions)**.** Given a stochastic process $X = \{X_t : t \in I\}$ on an index set I, the *finite-dimensional distributions* or *joint distributions* are the pushforward-measures

$$\mu_K(H) := \mathbb{P}(\omega \in \Omega \mid (X_{t_1}(\omega), \ldots, X_{t_k}(\omega)) \in H), \quad H \in \mathscr{R}^k, \tag{B.15}$$

where $K = (t_1, \ldots, t_k)$ is a k-tuple of distinct points in I with $k \in \mathbb{N}$ finite and $\mathscr{R}^k$ is the Borel-σ-algebra over $\mathbb{R}^k$.



Remark. The finite-dimensional distributions do not characterize the process completely:

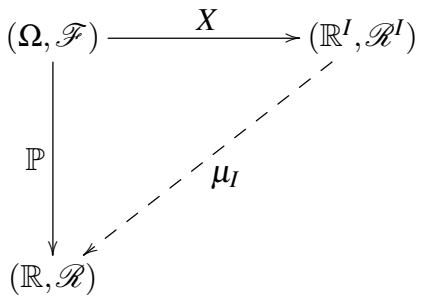
Given $\{\mu_K : K \subset_{finite} I\}$ such that two consistency conditions are met - namely, invariance of μ_K under permutations of the set K and $\mu_K(H) = \mu_{K \cup \{t_{k+1}\}}(H \times \mathbb{R})$ -, then *Kolmogorov's existence theorem* guarantees the *existence* of a probability space and a stochastic process associated with the given finite-dimensional distributions, but not their *uniqueness*.

(iii) Also, regarding X as random function $X : \Omega \to \mathbb{R}^I$, one can define the *law* of X assigning a probability to each sample path:

Definition B.2.3 (Law of a stochastic process)**.** The law of a real-valued process X on $(\Omega, \mathscr{F}, \mathbb{P})$ with index set I is the pushforward-measure

$$\mu_I(B) := \mathbb{P}(\omega \in \Omega \mid X(\omega) \in B), \quad B \in \mathscr{R}^I \tag{B.16}$$

where $\mathscr{R}^I$ is the *cylindrical* σ*-algebra* (explanation see below).

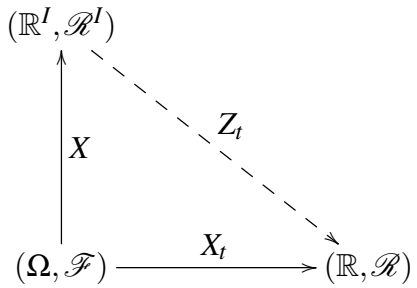


Remark. Remember that $X\colon I\times\Omega\to\mathbb{R}$ can be rewritten (by the means of *currying*) in two ways:

* On the one hand, it is defined as a map $I\to\mathbb{R}^{\Omega}$, where $\mathbb{R}^{\Omega}$ consists of the *projections* $X_t\colon\Omega\to\mathbb{R}$.
* On the other hand, X can be understood as a map $\Omega\to\mathbb{R}^I$, where $\mathbb{R}^I$ is the space of *sample paths* $X(\omega)\colon I\to\mathbb{R}$.

Now, instead of the projections or coordinate variables $X_t\colon\Omega\to\mathbb{R}$, one often uses their pushforwards by $X\colon\Omega\to\mathbb{R}^I$, which are functionals on $\mathbb{R}^I$:

$$\begin{aligned} Z_t\colon\ \ &\times_{i\in I}\mathbb{R}^i\equiv\mathbb{R}^I\to\mathbb{R}\\ &X(\omega)\mapsto X_t(\omega)\end{aligned}$$



This is useful because the original process $\{X_t\colon t\in I\}$ and the *coordinate-variable process* $\{Z_t\colon t\in I\}$ obviously have the same finite-dimensional distributions (see definition). Accordingly, identifying the event space Ω with the space of sample paths $\mathbb{R}^I$ defines the *canonical representation* of a process.

Pulling back the domain of the finite-dimensional distributions by Z_t yields the cylinder sets (ω omitted):

Definition B.2.4 (Cylinder sets)**.**

$$\begin{aligned} C(H,K)&=\{X\in\mathbb{R}^I\mid(Z_{t_1}(X),\dots,Z_{t_k}(X))\in H\}\\ &=\{X\in\mathbb{R}^I\mid(X_{t_1},\dots,X_{t_k})\in H\},\end{aligned}\tag{B.17}$$

where again $K=(t_1,\dots,t_k)$ is a k-tuple of distinct points in I (with $k\in\mathbb{N}$ finite) and $H\in\mathscr{R}^k$.

Let now

$$\mathscr{R}_0^I:=\{C(H,K)\colon K\subset_{finite}I\}\tag{B.18}$$

be the class of all cylinder sets which is an algebra (easily proved) but not a σ-algebra. Accordingly, the finite-dimensional distributions define a *premeasure* on $\mathbb{R}^I$ which is not an actual measure but is only finitely additive. Let now

$$\mathscr{R}^I := \sigma(\mathscr{R}_0^I) \tag{B.19}$$

be the σ-algebra generated by $\mathscr{R}_0^I$ (i.e. the minimal σ-algebra over the product state space $\mathbb{R}^I$ containing $\mathscr{R}_0^I$). Since $\mathscr{R}^I$ - which is called *cylindrical σ-algebra* - obviously is also generated by the sets

$$\begin{aligned} C(A) &= \{X \in \mathbb{R}^I \mid Z_t(X) \in A\} \\ &= \{X \in \mathbb{R}^I \mid X_t \in A\}, \end{aligned} \tag{B.20}$$

the coordinate-variable process $\{Z_t : t \in I\}$ is μ_I-measurable on $(\mathbb{R}^I, \mathscr{R}^I)$ and $\mathscr{R}^I$ is identical with the Borel σ-algebra of $\mathbb{R}^I$ equipped with the topology of pointwise convergence.

(iv) Finally, in many cases the sample paths of stochastic processes are required to be continuous (the previous section provides a sufficient criterion for that) - this condition is a special case of *separability*:

Definition B.2.5 (Separability of a stochastic process)**.** A real-valued process $\{X_t : t \in I\}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ is called *separable* if and only if there exist a countable dense subset $T \subset I$ and a set $D \in \mathscr{F}$ with measure zero ($\mathbb{P}(D) = 0$) such that

$$\forall t \in I \, \forall \omega \notin D \, \exists t_n \in T : t_n \to t \wedge X_{t_n}(\omega) \to X_t(\omega). \tag{B.21}$$

Explicitly, almost all sample paths of a separable process are determined by a countably infinite amount of points which are dense in I (and dense in $\mathbb{R}$ for continuous processes).

Therefore it is natural to consider only a *separable subspace* (in the following denoted by S) of the sample space $\mathbb{R}^I$ together with the cylindrical σ-algebra $\mathscr{R}^I$ restricted to S. Very often, S is also a metric space, thus demanding the *consistency condition* that the Borel-σ-algebra on S (denoted $\mathscr{B}(S)$) is measurable with respect to $\mathscr{R}^I|_S$.

Especially, if $\mathbb{R}^I$ is restricted to the space of continuous functions on the interval I $\mathscr{C}(I)$ with metric $d_\infty(X,Y) := \sup_{t \in I} |X_t - Y_t|$ (also known as *classical Wiener space*), it turns out that since $\mathscr{C}(I)$ is also complete, the cylindrical σ-algebra on $\mathscr{C}(I)$ (generated by the cylinder sets restricted to $\mathscr{C}(I)$) coincides with the Borel-σ-algebra generated by the topology of uniform convergence on $\mathscr{C}(I)$:

Due to completeness every sphere belongs to the cylindrical σ-algebra:

$$\{x; \sup_t |x(t) - \alpha(t)| \leq r\} = \bigcap_{k=1}^{\infty} \{x; |x(t_k) - \alpha(t_k)| \leq r\},$$

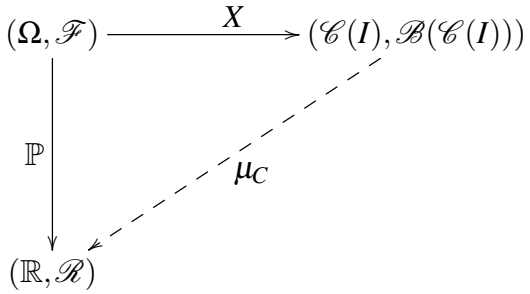
where $\alpha(t)$ is an arbitrary continuous function and $\{t_k\}$ is an arbitrary sequence everywhere dense on I ([3], p. 448f). But due to separability, the spheres are a basis for the Borel-σ-algebra over $\mathscr{C}(I)$.

Accordingly, the law of a *continuous* stochastic process is defined as follows:

Definition B.2.6 (Law of a continuous stochastic process)**.** The law of a real-valued process X on $(\Omega, \mathscr{F}, \mathbb{P})$ with index set I and a.s.-continuous sample paths is the pushforward-measure

$$\mu_C(B) := \mathbb{P}(\omega \in \Omega \mid X(\omega) \in B), \quad B \in \mathscr{B}(\mathscr{C}(I)) \tag{B.22}$$

where $\mathscr{B}(\mathscr{C}(I))$ is the Borel-σ-algebra on $\mathscr{C}(I)$.

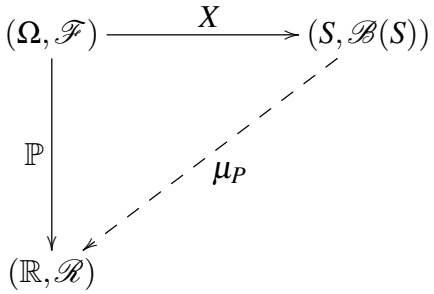


For real-valued stochastic process with sample paths in a general complete separable metric space (i.e. a *Polish space*) the following obvious modifications apply:

Definition B.2.7 (Law of a stochastic process with a Polish sample space)**.** The law of a real-valued process X on $(\Omega, \mathscr{F}, \mathbb{P})$ with index set I and sample paths in a Polish space S is the pushforward-measure

$$\mu_P(B) := \mathbb{P}(\omega \in \Omega \mid X(\omega) \in B), \quad B \in \mathscr{B}(S) \tag{B.23}$$

where $\mathscr{B}(S)$ is the Borel-σ-algebra on S.



Remark. Due to the additional structure of the sample space S (especially metric and separability), *any* functional on S that is continuous w.r.t. the metric d_S is also a *random variable* on $(S, \mathscr{B}(S))$.

B.2.2 Weak convergence of measures

Basic to all notions of weak convergence is the following definition:

Definition B.2.8 (Weak convergence of measures)**.** A sequence of finite measures μ_n is said to converge weakly to a finite measure μ on the Borel-σ-algebra $\mathscr{B}(S)$ over a metric space S if for any continuous bounded function $f \in \mathscr{C}_b(S)$

$$\lim_{n\to\infty} \int_S f \, d\mu_n = \int_S f \, d\mu. \tag{B.24}$$

Then the following equivalent notions of weak convergence can be given:

Lemma B.2.1 (Portmanteau-lemma)**.** (B.24) *is equivalent to:*

$$\limsup_{n\to\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}) \qquad \forall \text{ closed sets } \overline{A} \subset S \tag{B.25}$$

$$\liminf_{n\to\infty} \mu_n(A^0) \geq \mu(A^0) \qquad \forall \text{ open sets } A^0 \subset S \tag{B.26}$$

$$\lim_{n\to\infty} \mu_n(A) = \mu(A) \qquad \forall A \in \mathscr{B}(S)\colon \mu(\partial A) = 0, \tag{B.27}$$

where $\partial A := \overline{A} \backslash A^0$ *denotes the boundary of* A.

Proof. The proof is based on [3], p. 446f, [12], p. 385f and [5], p. 289; d_S denotes the metric on S.

- (B.24) $\rightarrow$ (B.25) Let $\overline{A} \subset_{closed} S$ be arbitrary but non-zero (otherwise the proof is trivial) and let $\varepsilon > 0$. Then there exist a number $k \in \mathbb{N}$ and a corresponding open superset of $\overline{A}$

  $$U_k := \{x \in S\colon d_S(x, \overline{A}) < \frac{1}{k}\}$$

  such that $\mu(U_k) < \mu(\overline{A}) + \varepsilon$. Further define a function $f\colon S \to \mathbb{R}$ by

  $$f(x) := \max(1 - k\, d_S(x, \overline{A}), 0),$$

  which is bounded and continuous (check triangle inequality) and satisfies $\chi_{\overline{A}} \leq f \leq \chi_{U_k}$ (where χ is the characteristic function or index function). Hence

  $$\limsup_{n\to\infty} \mu_n(\overline{A}) \leq \limsup_{n\to\infty} \int_S f \, d\mu_n = \int_S f \, d\mu \leq \mu(U_k) < \mu(\overline{A}) + \varepsilon$$

for all ε and therefore

$$\limsup_{n\to\infty} \mu_n(\overline{A}) \le \mu(\overline{A}).$$

- (B.25) $\leftrightarrow$ (B.26) Take complements.
- (B.26) $\rightarrow$ (B.27) Let $A \in \mathscr{B}(S)$ and $\overline{A}$ (respectively A^0) denote its closure (respectively interior). Due to (B.26) (and consequently, (B.25)) we have

$$\begin{aligned}&\mu(A^0) \le \liminf_{n\to\infty} \mu_n(A^0) \le \liminf_{n\to\infty} \mu_n(A)\\ &\le \limsup_{n\to\infty} \mu_n(A) \le \limsup_{n\to\infty} \mu_n(\overline{A}) \le \mu(\overline{A}),\end{aligned}$$

  thus for all *sets of continuity of the measure* μ (i.e. $\{A\colon \mu(\partial A) = 0\}$), we have $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.
- (B.27) $\rightarrow$ (B.25) Let B be an arbitrary closed Borel-set. The open supersets

$$B_\delta := \{x \in S\colon d_S(x,B) < \delta\}$$

  have disjoint boundaries for different values of $\delta > 0$, so due to σ-additivity at most countably many of the boundary sets can have a strictly positive measure. Therefore one can find a sequence δ_k such that $B_{\delta_k} \to B$ for $k \to \infty$ and the sets B_{δ_k} are sets of continuity of the measure μ. Consequently,

$$\limsup_{n\to\infty} \mu_n(B) \le \limsup_{n\to\infty} \mu_n(B_{\delta_k}) = \mu(B_{\delta_k}).$$

  For $k \to \infty$, this yields for any closed Borel-set B

$$\limsup_{n\to\infty} \mu_n(B) \le \mu(B).$$

- (B.25) $\rightarrow$ (B.24) Now for any $f \in \mathscr{C}_b(\mathbb{R})$, assume without loss of generality that $0 < f(x) \le 1\ \forall x \in S$. For $k \ge 1$ fixed, define

$$B_j := \{x \in S\colon \frac{j}{k} \le f(x)\} \quad \text{for } j = 0,\dots,k.$$

  Then the following integral approximation by simple functions holds:

$$\sum_{j=1}^{k} \frac{j-1}{k} \mu(B_{j-1} \cap B_j^c) \le \int f\, d\mu \le \sum_{j=1}^{k} \frac{j}{k} \mu(B_{j-1} \cap B_j^c).$$

Using

$$\mu(B_{j-1} \cap B_j^c) = \mu((B_{j-1}^c)^c \cap B_j^c) = \mu((B_{j-1}^c \cup B_j)^c)$$
$$= 1 - \mu(B_{j-1}^c) - \mu(B_j) = \mu(B_{j-1}) - \mu(B_j)$$

and $\mu(B_k) = 0$, $\mu(B_0) = 1$ after some standard sum manipulations (splitting, redefining indices) yields

$$\frac{1}{k}\sum_{j=1}^{k} \mu(B_j) \leq \int f \, d\mu \leq \frac{1}{k} + \frac{1}{k}\sum_{j=1}^{k} \mu(B_j).$$

Since the B_j are closed, we have

$$\limsup_{n\to\infty} \int f \, d\mu_n \leq \limsup_{n\to\infty} \{\frac{1}{k} + \frac{1}{k}\sum_{j=1}^{k} \mu_n(B_j)\}$$
$$\leq \frac{1}{k} + \frac{1}{k}\sum_{j=1}^{k} \mu(B_j) \leq \frac{1}{k} + \int f \, d\mu.$$

Taking the limit $k \to \infty$ and combining with the analog statement for $-f$ instead of f (recall that $\limsup(-x) = -\liminf(x)$) finally yields (B.24).

□

If the metric space S is assumed to be a Polish space (recall that this means that S is complete and separable), there is an additional characterization of weak convergence in terms of *weak compactness*.

Definition B.2.9 (Weak compactness of measures)**.** A sequence of finite measures μ_n on S is said to be *weakly compact* if every subsequence of it contains a weakly convergent subsequence of finite measures.

Remark. Actually it should read *weakly sequentially compact*, but it turns out (this is a consequence of the after next theorem) that if S is Polish, there is a natural metric on the space of finite measures on S (the so-called *Prokhorov-metric*) and hence the notions are equivalent.

The formulation of the following theorem as well as the proof are to some extent taken from [3], p. 446f.

Theorem B.2.1 (Weak convergence and weak compactness)**.** *For a sequence of finite measures μ_n to converge weakly to some finite measure μ on S, it is necessary and sufficient that the sequence is weakly compact and that $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all A belonging to some algebra $\mathscr{B}_0$ such that $\sigma(\mathscr{B}_0) = \mathscr{B}(S)$.*

Proof. Besides [3], compare also with the preceding proof (lemma B.2.1).

- *Necessity*: A weakly convergent sequence is clearly weakly compact. Because of (B.27), it remains to be shown that the sets of continuity of the measure μ (which form an algebra $\mathscr{A}$) generate the algebra $\mathscr{B}(S)$. The argument is very similar to the one given in the proof of (B.27) $\rightarrow$ (B.25): Only countably many spheres around a given point can have positive boundary measure (otherwise since the boundaries are disjoint and any sphere contains infinitely many boundary sets of smaller spheres, due to σ-additivity the latter would have infinite measure), so $\sigma(\mathscr{A})$ contains all spheres; since it is itself contained in $\mathscr{B}(S)$ by assumption and $\mathscr{B}(S)$ is the *minimal* σ-algebra containing all spheres, $\sigma(\mathscr{A}) = \mathscr{B}(S)$ follows.

- *Sufficiency*: Choose an arbitrary weakly convergent subsequence μ_{n_k} of μ_n having the limit $\overline{\mu}$. Like in the proof of the Portmanteau-lemma, (B.26) $\rightarrow$ (B.27), we have

  $$\overline{\mu}(A^0) \leq \liminf_{n_k \to \infty} \mu_{n_k}(A) \leq \limsup_{n_k \to \infty} \mu_{n_k}(A) \leq \overline{\mu}(\overline{A}),$$

  for $A \in \mathscr{B}(S)$ and by hypothesis, $\lim_{n_k \to \infty} \mu_{n_k}(A) = \mu(A)$ for $A \in \mathscr{B}_0$. Therefore, for all sets A in $\mathscr{B}_0$,

  $$\overline{\mu}(A^0) \leq \mu(A) \leq \overline{\mu}(\overline{A}). \tag{B.28}$$

  Since (B.28) is also satisfied for the limit of a monotonic sequence whose elements individually satisfy (B.28) (easily proven), the collection of sets that satisfy (B.28) is a monotone class containing $\mathscr{B}_0$ and hence especially contains $\sigma(\mathscr{B}_0) = \mathscr{B}(S)$. Especially for the sets of continuity $A \in \mathscr{A}$ we have

  $$\mu(A) = \overline{\mu}(A)$$

  by definition; but since $\mathscr{A}$ generates $\mathscr{B}(S)$, this result is general.

□

The following definition introduces a property of measures called *tightness*; its importance mainly lies in the fact that for finite measures on Polish spaces, tightness is equivalent to weak compactness (theorem of *Prokhorov*), thus providing a premise for weak convergence according to theorem B.2.1.

Definition B.2.10 (Tightness of measures)**.** A sequence of measures μ_n on S is called *tight* if

(i) for all $x \in S$

$$\sup_n \mu_n(x) < \infty, \tag{B.29}$$

(ii) for any $\varepsilon > 0$ there exists a compact subset K_ε of S such that

$$\sup_n \mu_n(S \setminus K_\varepsilon) < \varepsilon. \tag{B.30}$$

Remark. For probability measures, (B.30) is obviously equivalent to

$$\sup_n \mu_n(K_\varepsilon) > 1 - \varepsilon. \tag{B.31}$$

Furthermore, in that case (B.29) can be omitted since probability measures are finite per definition.

Theorem B.2.2 (Prokhorov)**.** *Let μ_n be a tight sequence of probability measures on the metric space S with Borel-σ-algebra $\mathscr{B}(S)$. Then μ_n is weakly compact (in the sense of B.2.9). Moreover, if S is complete and separable, then* (B.31) *is also necessary.*

Proof. Since the proof is quite involved, it is omitted here. It can be found for instance in [12], p.393ff or [3], p.441ff. □

B.2.3 Weak convergence of stochastic processes

Consider a sequence X_n of real-valued stochastic processes and a real-valued process X with index set I having sample paths in a Polish space S with Borel-σ-algebra $\mathscr{B}(S)$. Their laws will be denoted by $\mu_{P,n}$ and μ_P, their finite-dimensional distributions by $\mu_{K,n}$ and μ_K respectively. The actual definition of weak convergence of stochastic processes is very simple:

Definition B.2.11 (Weak convergence of stochastic processes)**.** The processes X_n converge weakly to the process X if and only if the laws $\mu_{P,n}$ converge weakly to μ_P according to definition B.2.8.

Now, making use of the theorems formulated in the preceding subsection yields the following result:

Theorem B.2.3 (Weak convergence of stochastic processes and tightness)**.** *With X_n and X defined as above, X_n converges weakly to X if and only if*

- *the finite-dimensional distributions $\mu_{K,n}$ converge to μ_K and*

- *the sequence of laws $\mu_{P,n}$ is tight.*

Proof. Because of the theorems B.2.1 and B.2.2 it remains to be shown that convergence of the finite-dimensional distributions is equivalent to $\lim_{n\to\infty}\mu_{P,n}(A) = \mu_P(A)$ for all A belonging to some algebra $\mathscr{B}_0$ such that $\sigma(\mathscr{B}_0) = \mathscr{B}(S)$. Take $\mathscr{B}_0 = \mathscr{R}_0^I$; $\sigma(\mathscr{B}_0) = \sigma(\mathscr{R}_0^I) = \mathscr{R}^I = \mathscr{B}(S)$ is just the consistency condition mentioned after definition B.2.5, and $\mu_{P,n}$ and $\mu_{K,n}$ (as well as μ_P and μ_K) coincide on $\mathscr{R}_0^I$ per definition. □

B.3 Tightness and Continuity

In this section, a theorem for weak convergence of stochastic processes in classical Wiener-space (cf. [3], p. 450f) will be proven (without loss of generality, $S = \mathscr{C}[0,1]$). The idea is to reexpress the tightness condition of theorem B.2.3 in terms of continuity properties of the associated stochastic processes.

Definition B.3.1 (Pre-compactness)**.** A subset $A \subset X$ of a complete metric space X is called *pre-compact* if all sequences in A contain a convergent subsequence.

Remark. The closure of a pre-compact set is of course compact.

Recall the following theorem for functions in $\mathscr{C}[0,1]$ relating *uniform boundedness* and *equicontinuity* with *pre-compactness*:

Theorem B.3.1 (Arzelà-Ascoli)**.** *A subset A of functions in $\mathscr{C}[0,1]$ is pre-compact if and only if*

(i) $\exists M < \infty\colon \sup_{t\in[0,1]} |\omega_t| \le M \quad \forall \omega \in A$ *(uniform boundedness),*

(ii) $\lim_{\delta\to 0} \sup_{\omega\in A} \nu_\omega(\delta) = 0$ *(equicontinuity),*

where $\nu_\omega(\delta)$ is the modulus of continuity *in $\omega \in C[0,1]$. In classical Wiener-space ($\mathscr{C}[0,1]$ equipped with the metric of uniform convergence d_∞), $\nu_\omega(\delta)$ is defined by*

$$\nu_\omega(\delta) := \sup_{|s-t|<\delta} |\omega_s - \omega_t| \tag{B.32}$$

and (i) (uniform boundedness) can be replaced by

(iii) $\sup_{\omega\in A} |\omega_0| < \infty$ *(boundedness at 0).*

Proof. The second statement is easily proven ([8], p. 97): Since $\nu_\omega(\delta) \to 0$ for $\delta \to 0$ (a consequence of equicontinuity), there is an $N \in \mathbb{R}$ such that

$$\sup_{\omega \in A} \nu_\omega(\frac{1}{N}) < 1$$

and therefore for all $0 \leq t \leq 1$ we have

$$|\omega_t| \leq |\omega_0| + \sum_{i=1}^{N} \left|\omega_{it/N} - \omega_{(i-1)t/N}\right| \leq \sup_{\omega \in A} |\omega_0| + N \sup_{\omega \in A} \nu_\omega(\frac{1}{N}) = M.$$

The converse direction (uniform boundedness to boundedness at 0) is of course trivial.

Proving the sufficiency of uniform boundedness and equicontinuity for precompactness requires a diagonalization argument (compare [5], p. 292f):

Take an enumeration $\{t_i, i \in \mathbb{N}\}$ of $\mathbb{Q}_{[0,1]}$ and an arbitrary sequence $\omega^j, j \in \mathbb{N}$ in A. Set $\omega_{i,j} := \omega^j_{t_i}$. Due to uniform boundedness of A, the sequence $\omega_{1,j}$ is bounded and therefore contains a convergent subsequence denoted by $\omega_{1,n_{1,j}}$ (*Bolzano-Weierstraß-theorem*). Next, look at $\omega_{2,n_{1,j}}$; by the same argument, it contains a convergent subsequence which will be denoted $\omega_{2,n_{2,j}}$.

Now, continuing inductively, one arrives at $\omega_{i,n_{i,j}}$ with the properties that it is convergent for each i and that $n_{i,j}$ is a subsequence of $n_{i-1,j}$. Set $n_k := n_{k,k}$; obviously, the sequence $n_i, n_{i+1}, \ldots$ is contained in $n_{i,j}$ and hence ω_{i,n_k} is convergent for all i since finitely many elements do not change the limit. For simplicity, put $\omega_{\cdot,k} := \omega_{\cdot,n_k}$.

To prove that A has compact closure it remains to be shown that $\omega_{\cdot,k}$ converges for *any* $t \in [0,1]$, not just for $t \in \mathbb{Q}_{[0,1]}$. This is achieved in a standard way: By equicontinuity, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\left|\omega_{t,k} - \omega_{t',k}\right| \leq \frac{\varepsilon}{3} \text{ whenever } \left|t - t'\right| < \delta;$$

because $\mathbb{Q}_{[0,1]}$ is dense in $[0,1]$, for any point $t' \in [0,1]$ and any $\delta > 0$ there exists a point $t \in \mathbb{Q}_{[0,1]}$ such that $|t - t'| < \delta$. Thus

$$\left|\omega_{t',k} - \omega_{t',k'}\right| \leq \left|\omega_{t',k} - \omega_{t,k}\right| + \left|\omega_{t,k} - \omega_{t,k'}\right| + \left|\omega_{t,k'} - \omega_{t',k'}\right| \leq \frac{2\varepsilon}{3} + \left|\omega_{t,k} - \omega_{t,k'}\right|;$$

since $\omega_{t,k}$ is Cauchy, for k, k' sufficiently large we have

$$\left|\omega_{t,k} - \omega_{t,k'}\right| < \frac{\varepsilon}{3}$$

and together

$$\left|\omega_{t',k} - \omega_{t',k'}\right| \leq \varepsilon.$$

Therefore $\omega_{t',k}$ is Cauchy for all $t \in [0,1]$ (thus converging pointwise) and equicontinuous, together implying uniform convergence. Summarizing, every sequence in A contains a convergent subsequence, making A pre-compact in the topology of uniform convergence.

Conversely, suppose that A is pre-compact. If A is unbounded (in the uniform norm), it contains a sequence with monotonely increasing and unbounded norm; but since norm convergence is necessary for convergence, this sequence cannot have convergent subsequences.

If A is not equicontinuous, there is an $\varepsilon > 0$ such that for all δ there exist $t, t' \in [0,1]$ and $\omega \in A$ such that $|t - t'| < \delta$, but $|\omega_t - \omega'_t| > \varepsilon$. Especially, for the sequence $\delta_n := \frac{1}{n}$ there are $t_n, t'_n \in [0,1]$ and $\omega_{\cdot,n} \in A$ such that $|t_n - t'_n| < \frac{1}{n}$, but $\left|\omega_{t,n} - \omega_{t',n}\right| > \varepsilon$. Take one such sequence $\omega_{\cdot,n}$; by assumption, it is not equicontinuous and therefore not convergent in the uniform metric, the same is especially true for its subsequences. This is a contradiction to the pre-compactness of A. □

The characterization of compact sets in $\mathscr{C}[0,1]$ by uniform boundedness and equicontinuity (Arzelà-Ascoli theorem) now allows for a reformulation of tightness of measures on $\mathscr{C}[0,1]$ (cf. [3] p. 449f, [8] p. 97f):

Theorem B.3.2 (Tightness of probability measures on the space of continuous functions)**.** *A sequence of probability measures μ_n on $\mathscr{C}[0,1]$ is* tight *if and only if the following two conditions hold for all n:*

(i) for each $\eta > 0$ there is a number B such that

$$\mu_n(\omega \in \mathscr{C}[0,1] \colon \sup_{t \in [0,1]} |\omega_t| > B) \leq \eta, \tag{B.33}$$

(ii) for each $\varepsilon > 0, \eta > 0$, there exists a $\delta > 0$ such that

$$\mu_n(\omega \in \mathscr{C}[0,1] \colon \nu_\omega(\delta) \geq \varepsilon) \leq \eta. \tag{B.34}$$

Proof. If μ_n is tight, then for each $\eta > 0$ there is a compact set K such that $\mu_n(K) > 1 - \eta$ for all n. But, by the Arzelà-Ascoli theorem, for all $\omega \in K$ there is a $B \geq \sup_{t \in [0,1]} |\omega_t|$ and hence

$$\mu_n(\omega \in \mathscr{C}[0,1] \colon \sup_{t \in [0,1]} |\omega_t| > B) \leq \mu_n(K^C) \leq 1 - (1 - \eta) = \eta.$$

Similarly, given $\varepsilon > 0, \eta > 0$ select $\delta > 0$ such that $\sup_{\omega \in K} \nu_\omega(\delta) < \varepsilon$. Then

$$\mu_n(\omega \in \mathscr{C}[0,1] \colon \nu_\omega(\delta) \geq \varepsilon) \leq \mu_n(K^C) \leq \eta.$$

Conversely, given $\eta > 0$, choose B such that

$$\mu_n(\omega \in \mathscr{C}[0,1] \colon \sup_{t\in[0,1]} |\omega_t| \leq B) \geq 1 - \frac{\eta}{2}$$

and δ_ε such that

$$\mu_n(\omega \in \mathscr{C}[0,1] \colon \nu_\omega(\delta_\varepsilon) < \varepsilon) \geq 1 - 2^{-(1+\frac{1}{\varepsilon})}\eta.$$

For K being the closure of

$$\{\omega \colon \sup_{t\in[0,1]} |\omega_t| \leq B\} \cap \bigcap_{\frac{1}{\varepsilon}=1}^{\infty} \{\omega \colon \nu_\omega(\delta_\varepsilon) < \varepsilon\}$$

we have $\mu_n(K) \geq 1 - \eta$ for all n. □

Together with theorem B.2.3 we have:

Theorem B.3.3 (Weak convergence of continuous stochastic processes)**.** *Let $X_n = \{X_t^{(n)} \colon t \in [0,1]\}$ and $X = \{X_t \colon t \in [0,1]\}$ be stochastic processes defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with a.s. continuous sample paths and suppose that the finite dimensional distributions of X_n converge to those of X. X_n converges weakly to X if and only if for each $\varepsilon > 0$*

$$\lim_{\delta\to 0} \sup_n \mathbb{P}(\omega \in \Omega \mid \nu_{X_n(\omega)}(\delta) \geq \varepsilon) = 0. \tag{B.35}$$

Proof. Since δ in (B.35) is independent of n (this turns out to be a consequence of equicontinuity), the limit can be taken after the supremum and (B.35) is equivalent to

$$\forall n, \lim_{\delta\to 0} \mathbb{P}(\omega \in \Omega \mid \nu_{X_n(\omega)}(\delta) \geq \varepsilon) = 0 \tag{B.36}$$

since $\mathbb{P}$ of course cannot be negative.

- *Necessity*: If X_n converges weakly to X, the sequence $\mu_{P,n}$ is tight (theorem B.2.3) and consequently (B.34) holds. Take the limit $\delta \to 0$.
- *Sufficiency*: Suppose that (B.36) holds. Given an arbitrary $\eta > 0$, it is always possible to find a $\delta > 0$ such that

  $$\mathbb{P}(\omega \in \Omega \mid \nu_{X_n(\omega)} \geq \varepsilon) \leq \eta$$

  for all n. But this is just (B.34).

On the other hand, since the finite-dimensional distributions of X_n converge by assumption, they must be bounded; especially, for any $\eta > 0$ there exists a B such that

$$\mathbb{P}(\omega \in \Omega \mid \sup_{t\in[0,1]} \left|X_t^{(n)}(\omega)\right| > B) \leq \eta$$

for all n which is (B.33).

□

If the metric on $\mathscr{C}[0,1]$ is specified to d_∞ (classical Wiener-space), one finally arrives at the following theorem:

Theorem B.3.4 (Gikhman-Skorokhod)**.** *Let X_n and X be stochastic processes defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with sample paths in classical Wiener-space. Suppose that the finite dimensional distributions of X_n converge to those of X and that the processes X_n fulfill Kolmogorov's continuity condition* (B.6) *individually. Then X_n converges weakly to X.*

Proof. The proof (first given by [3], p. 450f) is in some points related to the one given for Kolmogorov-Chentsov's theorem. Since $\mathscr{C}[0,1]$ with d_∞ is separable, switch with advantage from $[0,1]$ to the set of rationals on $[0,1]$. Also, since the proceeding is identical, the index n will be suppressed until the final result.

Using again the *dyadic representation*, the oscillation over $\delta \geq |t-s|$

$$v_X(\delta) = \sup_{|t-s|\leq\delta} |X_t - X_s|$$

can be estimated for $\delta = 2^{1-k}$ (k fixed) by

$$v_X(\delta) \leq 2 \sup_{j2^{-k}<i2^{-m}<(j+1)2^{-k}} \left|X_{i2^{-m}} - X_{j2^{-k}}\right|,$$

where suprema over $i, j \in \mathbb{N}$ and $m > k$ are taken. For $j2^{-k} < i2^{-m} < (j+1)2^{-k}$, write

$$i2^{-m} = j2^{-k} + \sum_{\nu=1}^{r} 2^{-m_\nu} \text{ where } k < m_1 < m_2 < \cdots < m_r \leq m;$$

further define $\alpha(\mu) := \sum_{\nu=1}^{\mu} 2^{-m_\nu}$. Then

$$X_{i2^{-m}} - X_{j2^{-k}} = \sum_{\mu=1}^{r} X_{j2^{-k}+\alpha(\mu)} - X_{j2^{-k}+\alpha(\mu-1)}$$

and therefore, since suprema over i and m (respectively μ and r) are taken,

$$\nu_X(\delta) \leq 2 \sum_{m=k+1}^{\infty} \sup_{0\leq h\leq 2^m-1} \left| X_{(h+1)2^{-m}} - X_{h2^{-m}} \right|.$$

For a given $\varepsilon > 0$, take $\delta = 2^{-k+1}$ so small (i.e. k so large) that $\sum_{m=k+1}^{\infty} \frac{1}{m^2} < \frac{\varepsilon}{2}$; then

$$\begin{aligned}
\mathbb{P}(\nu_X(\delta) > \varepsilon) &\leq \sum_{m=k+1}^{\infty} \mathbb{P}\Big(\sup_{0\leq h\leq 2^m-1} \left| X_{(h+1)2^{-m}} - X_{h2^{-m}} \right| > \frac{1}{m^2} \Big) \\
&\leq \sum_{m=k+1}^{\infty} \sum_{h=0}^{2^m-1} \mathbb{P}\big(\left| X_{(h+1)2^{-m}} - X_{h2^{-m}} \right| > \frac{1}{m^2} \big) \\
&\leq \sum_{m=k+1}^{\infty} m^{2\alpha} \sum_{h=0}^{2^m-1} \mathbb{E} \left| X_{(h+1)2^{-m}} - X_{h2^{-m}} \right|^{\alpha} \\
&\leq \sum_{m=k+1}^{\infty} m^{2\alpha} 2^m C 2^{-m(1+\beta)} = C \sum_{m=k+1}^{\infty} \frac{m^{2\alpha}}{2^{m\beta}}
\end{aligned}$$

where Chebyshev's inequality and (B.6) were used. This bound goes to zero with $\delta \to 0$ uniformly with respect to n yielding (B.35). □

B.4 Thermodynamic Limit of the Forcing

The theorem of Gikhman and Skorokhod introduced in the preceding section will be used to prove an important limit theorem (cf. [13]) for the forcing term of the Mori-Zwanzig-Langevin equation.

Theorem B.4.1 (Weak convergence of the forcing)**.** *Suppose that g is a positive bounded real-valued function with bound c_1 such that*

$$g(\omega) \leq c_2\, \omega^{-\lambda} \text{ for some } c_2 > 0 \text{ and } \lambda > 1/2. \tag{B.37}$$

Recall that under those assumptions, the memory kernel

$$\begin{aligned}
&K_n(t) := \sum_{j=1}^{n} g^2(\omega_j) \cos(\omega_j t)\, \Delta\omega \qquad (B.38)\\
&\text{with } \Delta\omega \equiv n^{a-1} \text{ and } \omega_j := n^a \nu_j, \quad \nu_j \text{ i.i.d.}, \quad \nu_1 \sim \mathscr{U}[0,1]
\end{aligned}$$

converges to

$$K(t) := \int_0^{\infty} g^2(\omega) \cos(\omega t)\, d\omega \tag{B.39}$$

ν-almost surely in the limit $n \to \infty$ for some $a \in (0,1)$ (see preceding chapter). But also, the forcing term

$$Z_n(t) := \beta^{-1/2} \sum_{j=1}^{n} g(\omega_j)[\xi_j \cos(\omega_j t) + \eta_j \sin(\omega_j t)](\Delta\omega)^{1/2}, \tag{B.40}$$

$$\text{where} \quad \xi_j \text{ i.i.d.}, \quad \xi_1 \sim \mathcal{N}(0,1), \quad \eta_j \text{ i.i.d.}, \quad \eta_1 \sim \mathcal{N}(0,1),$$

which is a zero-mean stationary Gaussian process with autocovariance function $\mathbb{E}_{\xi\eta}[Z_n(t+s)\, Z_n(s)] = \beta^{-1} K_n(t)$ ν-almost surely, converges to the zero-mean stationary Gaussian process $Z(t)$ with autocovariance function $\mathbb{E}_{\xi\eta}[Z(t+s)\, Z(s)] = \beta^{-1} K(t)$ ν-almost surely and ξ, η-weakly in $\mathscr{C}[0,T]$.

Proof. First check if $Z(t)$ has a.s.-continuous sample paths. On that purpose, use $\alpha = 2$ in (B.6): then there should exist $\gamma, C > 0$ such that

$$C\,|t|^{1+\gamma} \overset{!}{\geq} \mathbb{E}_{\xi\eta}\,|Z(t+s) - Z(s)|^2 = 2\,|K(0) - K(t)|\,;$$

but this follows from the assumptions on $g(\omega)$.

Next, check if the finite-dimensional distributions of Z_n converge to those of Z ν-almost surely. This is clear since, due to the Gaussian nature of Z_n and Z, they are completely characterized by the means $\mathbb{E}_{\xi\eta}[Z_n] = 0$ respectively $\mathbb{E}_{\xi\eta}[Z] = 0$ (which holds ν-almost surely) and the autocovariance functions K_n respectively K, which converge ν-almost surely as proven in the preceding chapter.

According to theorem B.3.4, it remains to be shown that Kolmogorov's continuity condition (B.6) holds for Z_n ν-almost surely. To that end, choose $\theta \in (0, 2\lambda - 1] \cap (0,2]$ and $b \in \mathbb{N}$ such that $b\theta = 1 + \gamma$ for some $\gamma > 0$. Then

$$\begin{aligned}
\mathbb{E}_{\xi\eta}\,|Z_n(t+s) - Z_n(s)|^{2b} &= (2b-1)!!(\mathbb{E}_{\xi\eta}\,|Z_n(t+s) - Z_n(s)|^2)^b \\
&= 2^b (2b-1)!!\,|K_n(t) - K_n(0)|^b \\
&= 2^{2b}(2b-1)!! \left| \sum_{j=1}^{n} g^2(\omega_j) \sin^2(\frac{1}{2}\omega_j t)\Delta\omega \right|^b \\
&\leq 2^{2b}(2b-1)!!(\sum_{j=1}^{n} g^2(\omega_j) \left|\frac{1}{2}\omega_j t\right| \Delta\omega)^b \\
&= 2^{b(2-\theta)}(2b-1)!!\,|t|^{b\theta}\,(\sum_{j=1}^{n} g^2(\omega_j)\omega_j^\theta \Delta\omega)^b \\
&\leq [2^{b(2-\theta)}(2b-1)!!M^b]\,|t|^{1+\gamma}.
\end{aligned}$$

Going from the first to the second line uses the Gaussian nature of the process and its stationarity; further, in the forth line the estimate $\sin^2(x) \leq |x|^\theta$ was used. Finally, since $\theta \leq \lambda - \frac{1}{2}$, $\tilde{g}(\omega_j) := \omega^{\frac{\theta}{2}} g(\omega_j) \leq c_2\, \omega^{\frac{\theta}{2} - \lambda}$ meets the condition (B.37),

and hence $K_n(0)$ with g replaced by $\tilde{g}$ converges ν-almost surely to

$$\int_0^\infty \tilde{g}^2(\omega)\, d\omega \leq M < \infty.$$

This completes the proof since $\tilde{g}^2$ and consequently M is positive. □

Bibliography

[1] P. Langevin. *Sur la théorie du mouvement brownien.* C. R. Acad. Sci. 146, 530-533 (1908).

[2] G.W. Ford, M. Kac and P. Mazur. *Statistical Mechanics of Assemblies of Coupled Oscillators.* Journal of Mathematical Physics 6, 504 (1965).

[3] I.I. Gikhman and A.V. Skorokhod. *Introduction to the Theory of Random Processes.* W. B. Saunders, 1969.

[4] Robert Zwanzig. *Nonlinear Generalized Langevin Equations.* Journal of Statistical Physics, Vol. 9, No. 3 (1973).

[5] P. Billingsley. *Probability and Measure.* John Wiley & Sons, 1979.

[6] A. Caldeira and A.J. Leggett. *Influence of dissipation on quantum tunneling in macroscopic systems.* Physical Review Letters 46,p. 211 (1981).

[7] C.W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Springer, 2nd edition, 1985.

[8] R. N. Bhattacharya and E. C. Waymire. *Stochastic Processes with Applications.* John Wiley & Sons, 1990.

[9] G. Grimmett and D. Stirzaker. *Probability and random processes.* Oxford University Press, 3rd edition, 2001.

[10] `http://math.fullerton.edu/mathews/n2003/montecarlomod.html`

[11] D. Givon, R. Kupferman and A.M. Stuart. *Extracting macroscopic dynamics: model problems and algorithms.* Nonlinearity, Vol. 17 (2004) R55-R127.

[12] J. Elstrodt. *Maß-und Integrationstheorie.* Springer, 4th edition, 2005.

[13] R. Kupferman, A.M. Stuart, J.R. Terry and P.F. Tupper. *Long Term Behaviour of Large Mechanical Systems with Random Initial Data.* Stochastics and Dynamics, Vol. 2, No. 4 (2002) 1-30, Errata corrected February 7, 2005.

[14] `http://galton.uchicago.edu/~lalley/Courses/385/GaussianProcesses.pdf`

Curriculum Vitae

Personal Data

Name	Maximilian Sadilek
Date of Birth	August 29, 1985
Place of Birth	Vienna
Address	Hernalser Hauptstraße 24-26/11, 1170 Vienna

Education

June 2003	Matura (school leaving examination) with honors
2004-2012	Physics studies at University of Vienna

Physics Qualifications

Studies	Awarded several performance scholarships; Special lectures/courses on particle physics, renormalization, relativity, cosmology, aerosol physics, nuclear physics, radiation protection, quantum decoherence, differential geometry etc.; June 2012 Intensive course "Quantum field theory in and out of equilibrium" in Zagreb
Organization	2010/2011 Organization assistent at the Vienna Seminar on Particle Physics and Quantum Field Theory and at the Vienna Seminar of Nobel Laureates
Didactics	2011/12 Participation at AYPT (Austrian Young Physicists' Tournament)

Miscellaneous

Chess	2003 Austrian Youth Champion U18
Music	Numerous gigs as pianist and choir singer (Burgtheater, Staatsoper, Konzerthaus); Choirmaster at the Vienna University Choir
Sports	2009-2012 Participation at the Boulder Worldcup in Vienna; Dedicated alpinist

Vienna, December 2012