Chapter 1

Introduction to Conditional Averaging

In complex system modeling, one often finds mathematical models that consist of many differential equations with different temporal and spatial scales. One of the basic approaches to the reduction of complexity is based on elimination of fast degrees of freedom (DOF): Describe the effective dynamical behaviour in terms of some slow (or essential) DOF of the system, while the collective effect of the majority of fast (or inessential) DOF is represented implicitly.

Many different mathematical techniques are available as, e.g., averaging for deterministic [1, 15, 38] and stochastic dynamical systems [14, 15, 22, 23, 24], stochastic modeling [28, 30, 33, 45], adiabatic theories [3, 6, 17, 44], or homogenization in time [4, 5, 6]. These techniques have mathematically been considered in a variety of articles. However, surprisingly few articles deal with the question of how to extend the approach to the case where the approximation may be inappropriate for reproducing the effective behaviour of the original dynamical system. The present thesis is motivated by the observation that the usual averaging techniques do not account for fast scale effects which generatea time scale in the system which is comparable to the time scale of the slow motion or even longer. This phenomenon is related to the notion of metastability in the fast dynamics, i.e., to the existence of metastable subsets in the fast state space [11, 41, 20, 39]: There is some subset of the accessible state space from which the fast motion will most probably exit only on some time scale of order ord(1) or even larger. Consequently, the thesis is concerned with the effect of metastable subsets in the accessible state space of the fast DOF on the slow DOF. We basically consider stochastic differential equations where the fast mode of the system rarely switches from one almost invariant set in its state space to another one such that the time scale of the switching is as slow as the slow modes of the system. The basic idea is that the fast process then can be decomposed

into several 'almost irreducible' subprocesses, each of which corresponds to one metastable or almost invariant state. To quantify this principle, the rare transitions between these states are described by means of the expected exit times. They can be used to parametrize a Markov chain model mimicking the transitions between the states.

1.1 The System under Consideration

A prototype system for analyzing metastability is the Smoluchowski dynamics of a particle in two dimensions, with the motion along one direction (y) occurring on a fast time scale $\epsilon \ll 1$ and the motion along the other direction (x) occurring on a slower, order unity time scale ([43]):

$$\dot{x}^\epsilon = -D_x V(x,y) + \sigma \dot{W}_1 \tag{1.1}$$

$$\dot{y}^\epsilon = -\frac{1}{\epsilon} D_y V(x,y) + \frac{\sigma}{\sqrt{\epsilon}} \dot{W}_2, \tag{1.2}$$

with $0 < \epsilon \ll 1$, $D_x V$, $D_y V$ denoting the derivatives of the potential $V = V(x,y)$ wrt x, y, respectively, and W_j, $j = 1, 2$ standard Brownian motion. Such multiscale dynamics could arise in rescaled coordinates from a potential V which is steeper in the y direction than in the x direction.

We also can analyze the dynamics (1.1)&(1.2) through the equivalent Fokker-Planck equation

$$\partial_t \rho^\epsilon = \mathcal{L}^\epsilon \rho^\epsilon \tag{1.3}$$

$$\mathcal{L}^\epsilon = \frac{1}{\epsilon} \mathcal{L}_x + \mathcal{L}_y, \tag{1.4}$$

which describes the evolution of the probability density $\rho^\epsilon(t,x,y)\mu(x,y)$ of the state of the system, where $\mu(x,y)$ is the invariant probability density of the dynamics. $\mathcal{L}^\epsilon$ is a differential operator with $\mathcal{L}_x$ acting on the variable y but depending on x, and $\mathcal{L}_y$ acting as a differential operator on x and depending on y. For the Smoluchowski dynamics, we have

$$\mathcal{L}_x = \frac{\sigma^2}{2} \Delta_y - D_y V(x,\cdot) \cdot D_y \,, \tag{1.5}$$

$$\mathcal{L}_y = \frac{\sigma^2}{2} \Delta_x - D_x V(\cdot,y) \cdot D_x \,, \tag{1.6}$$

$$\mu(x,y) = \frac{1}{Z} \exp(-\beta V(x,y)), \quad Z = \int \exp(-\beta V(x,y)) \,\mathrm{d}x \,\mathrm{d}y. \tag{1.7}$$

The invariant density μ depends on the inverse temperature $\beta = 2/\sigma^2$. The Fokker-Planck equation is considered within the Hilbert space weighted with respect to the invariant probability density:

$$L^2(\mu) = \left\{ g = g(x,y) : \int |g(x,y)|^2 \mu(x,y) \,\mathrm{d}x \,\mathrm{d}y < \infty \right\}.$$

It is endowed with the natural inner product

$$\langle f, g\rangle_\mu \;=\; \int f(x,y)\bar{g}(x,y)\,\mu(x,y)\,\mathrm{d}x\mathrm{d}y, \qquad f, g \in L^2(\mu),$$

which renders the Fokker-Planck equation (1.3) self-adjoint (so that, in particular, it also serves as the backward Kolmogorov equation describing the evolution of the expectations of functions of the state of the system). Self-adjointness of the generator in $L^2(\mu)$ is equivalent to reversibility of the associated Markov process, as expressed by the so-called detailed-balance condition of the probability transition function $p(t, x, \mathrm{d}y)$:

$$p(t, x, \mathrm{d}y)\,\mu(\mathrm{d}x) \;=\; p(t, y, \mathrm{d}x)\,\mu(\mathrm{d}y).$$

1.2 The Averaging Principle

Averaging theorems describe the effective motion of the slow DOF in the limit $\epsilon \to 0$. Their statement is based on a decoupling of fast and slow motion:

> For any given state of the slow DOF, there is an accessible part of the state space that the fast DOF completely explore before the position of the slow ones changes effectively. The fast variables are eliminated from the original equation of motion by averaging according to the probability distribution corresponding to the exploration of the accessible state space. The effective motion of the slow variables is governed by an averaged equation of motion.

Under suitable conditions on the potential energy V (cf. [15]), averaging completely characterizes the limit x^0 of the slow dynamics x^ϵ for $\epsilon \to 0$ as obeying an averaged SDE

$$\dot{x}^0 \;=\; -\int_{\mathbf{R}^n} D_x V(x,y)\,\mu_x(y)\mathrm{d}y \;+\; \sigma\dot{W}_1, \tag{1.8}$$

where μ_x denotes the (presumed unique) invariant density of the fast dynamics for fixed x:

$$\mu_x(y) \;=\; \frac{1}{Z_x}\exp(-\frac{2}{\sigma^2}V(x,y)), \quad Z_x \;=\; \int_{\mathbf{R}^n}\exp(-\frac{2}{\sigma^2}V(x,y))\,\mathrm{d}y. \tag{1.9}$$

We expect the invariant probability density of (1.8) to be obtained by integrating μ over the fast variable y:

$$\bar{\mu}(x) \;=\; \int \mu(x,y)\,\mathrm{d}y. \tag{1.10}$$

In order to derive the effective x dynamics in terms of the ensemble description (1.3), we will restrict attention to an initial probability density

which modulates the invariant probability density only through a function of x: $\rho^\epsilon(t=0,x,y) = f(x)$ ([43]). For preparation of this discussion we define the weighted Hilbert space $L^2(\mu_x)$ acting on fibres of constant x, with inner product

$$\langle f, g\rangle_{\mu_x} \;=\; \int f(y)\bar{g}(y)\,\mu_x(y)\,\mathrm{d}y.$$

The restriction on the initial condition is equivalently expressed by means of the projection operator Π

$$(\Pi g)(x,y) \;=\; \int g(x,y)\,\mu_x(y)\,\mathrm{d}y, \tag{1.11}$$

projecting onto functions which do not depend on y. With it we obtain

$$\Pi\rho^\epsilon(t=0) \;=\; \rho^\epsilon(t=0).$$

Using the inner product $\langle\cdot,\cdot\rangle_{\mu_x}$, the projection operator Π can be expressed as

$$\Pi g \;=\; \langle g, \mathbf{1}\rangle_{\mu_x}\cdot\mathbf{1},$$

which immediately provides

$$\Pi\mathcal{L}_x \;=\; 0 \;=\; \mathcal{L}_x\Pi.$$

Here, we have used $\mathcal{L}_x\mathbf{1} = 0$ and $\mathcal{L}_x^*\mathbf{1} = 0$ with $\mathcal{L}_x^*$ denoting the formal adjoint of $\mathcal{L}_x$. The operator Π allows the averaged force along the x direction in the dynamics (1.8) as $\Pi D_x V$ and the averaged probability density $\bar{\mu}$ to be related to the averaged potential $\overline{V}$, also variously referred to as the*potential of mean force* or *conformational free energy landscape:*

$$\begin{aligned}\Pi D_x V &= \int D_x V(x,y)\,\mu_x(y)\,\mathrm{d}y \;=\; D_x\overline{V},\\ \overline{V}(x) &= -\frac{1}{\beta}\ln\bar{\mu}(x).\end{aligned}$$

The corresponding Fokker-Planck equation is considered in $L^2(\bar{\mu})$ and reads

$$\partial_t\rho^0 \;=\; \overline{\mathcal{L}}\rho^0, \tag{1.12}$$

where $\overline{\mathcal{L}}$ is the operator $\mathcal{L}^\epsilon$ averaged against the invariant measure of the y dynamics:

$$\overline{\mathcal{L}} \;=\; \Pi\mathcal{L}^\epsilon\Pi \;=\; \Pi\mathcal{L}_y\Pi \;=\; \frac{\sigma^2}{2}\Delta_x \;-\; D_x\overline{V}\cdot D_x\ .$$

1.3 Breakdown of Averaging Principle

We now consider the situation in which the fast dynamics exhibit *metastable states*. That is,the effective dynamics in the fast degrees of freedom (DOF) can be described by (rare) jumps between certain metastable subsets of state space, while in between the jumps the dynamics evolves on a fast time scale inside one of these metastable subsets. Under this condition, the standard averaging procedure may fail to reproduce the effective dynamics of the original system, mainly for the following reason: The averaging principle is based on the assumption that the fast DOF completely explore the accessible state space before any substantial changes in the slow DOF.This can fail to hold in the presence of metastability in the fast dynamics, in particular if there is some subset of the accessible state space from which the fast motion will most probably exit only on some time scale comparable or ever greater than the ord(1) time scale of the slow motion. We can make this statement rigorous by introducing the *mean exit time* for the process y_x^ϵ from one of the metastable subsets, where y_x^ϵ is governed by the SDE (1.2) for fixed x. If we assume the existence of two metastable sets $B_x \cup B_x^c = \mathbf{R}^n$ with $B_x \cap B_x^c = \emptyset$, the mean exit time $\bar{\tau}_x^\epsilon(y)$ from B_x is the expected value of the first exit time $\tau_x^\epsilon(y)$ of the process y_x^ϵ from B_x started at $y_x^\epsilon(t=0) = y$, which is defined as

$$\tau_x^\epsilon(y) \;=\; \inf\{t \in \mathbf{R}^+ :\, y_x^\epsilon(t) \notin B_x,\, y_x^\epsilon(0) = y\}.$$

Although we would expect that exit times depend on the starting point, i.e., $y_x^\epsilon(0) = y$, it can be shown that there do exist subsets D, for which the exit time is basically independent for all states $y \in D$. Especially for a metastable collection of sets D_i of the Smoluchowski dynamics, in the limit of vanishing noise intensity we are able to assign a first exit time to an entire subset D_i rather than to single points $y \in D_i$, see [20, 41, 40]. Thus, as a consequence, the mean exit time $\bar{\tau}_x^\epsilon$ from B_x can be quoted independently of the starting point $y \in B_x$. The question of the asymptotic behaviour of the mean exit time for vanishing noise term σ has been discussed in detail by, for example, Freidlin and Wentzell in [15], from which the following result is taken (up to some slight modifications tailored to (1.1)&(1.2)) :

Theorem 1.3.1 ([15, Thm. 4.1 of Chap. 4], [40]) *Let the potential $V(x,\cdot)$ be twice continuously differentiable, let $y_{\min}$ be one of its local minima, and B_x a metastable subset with sufficiently smooth boundary ∂B_x containing $y_{\min}$ in its interior, but containing no other local minimum of $V(x,\cdot)$ within its interior. Without loss of generality we may assume that $V(x, y_{\min}) = 0$. Suppose that y_0 is the unique point on the boundary ∂B_x with*

$$V_{\mathrm{bar}}^x \;=\; V(x,y_0) \;=\; \min\{V(x,y) :\, y \in \partial B_x\}.$$

The mean exit time $\bar{\tau}_x^\epsilon$ for the process y_x^ϵ with $y_x^\epsilon(0) \in B_x$ then satisfies

$$\lim_{\sigma\to 0} \sigma^2 \ln \frac{\bar{\tau}_x^\epsilon}{\epsilon} \;=\; 2\, V_{\rm bar}^x.$$

As we are interested in the case where the averaging principle fails, let us have a closer look at the relation between the time scale of the fast motion and the exit times from metastable subsets in the fast DOF. The result of the above theorem tells us two things: First, the separation of time scales $\bar{\tau}_x^\epsilon \ll 1$ can be realized by fixing σ and the potential energy function; then we are always able to find an ϵ small enough such that averaging yields a good approximation of the effective dynamics; second, if we decrease σ or increase the potential energy barrier, the smallness parameter ϵ has to be chosen exponentially small such that the averaged system still is a satisfactory approximation. If we want to study the effect of metastabilities in the fast motion, it is natural to relate $V_{\rm bar}^x/\sigma^2$ to ϵ so that the exit times from metastable sets vary on a timescale of order ord(1), that is,

$$\bar{\tau}_x^\epsilon \;\simeq\; C(x)\,\epsilon\, \exp(\frac{2}{\sigma^2} V_{\rm bar}^x) \;=\; \text{ord}(1), \tag{1.13}$$

where $C(x)$ denotes the subexponential pre-factor that necessarily depends on x. The symbols $\simeq$ and ord(1) are defined at the end of the introduction.

Metastable Fast Dynamics

The effect of metastability on every fibre

$$\Phi(x) \;=\; \{(x,y) \;:\; y \in \mathbf{R}\}$$

of the fast state space is related to the dominant spectrum of the generator $\mathcal{L}_x$. In case of two metastable subsets $B_x^{(1)}$, $B_x^{(2)}$ for every x, the generator $\mathcal{L}_x$ has two dominant eigenvalues, $\lambda_0 = 0$ and a small negative eigenvalue $\lambda_1(x) \approx 0$, whereas the remainder $\{\lambda_k \;:\; k \geq 2\} \subset \mathbf{R}^-$ is bounded away from the origin by an order unity spectral gap. The eigenfunction u_0 associated to λ_0 is given exactly by $u_0 = \mathbf{1}$, whereas the eigenfunction $u_1(x,\cdot)$ associated to $\lambda_1(x)$ is an approximate stepfunction that is almost constant on the metastable subsets ([11, 19, 20, 39, 41]):

$$u_1(x,\cdot) \;\;\approx\;\; \sqrt{\frac{\mu_x(B_x^{(2)})}{\mu_x(B_x^{(1)})}}\,\mathbf{1}_{B_x^{(1)}} \;-\; \sqrt{\frac{\mu_x(B_x^{(1)})}{\mu_x(B_x^{(2)})}}\,\mathbf{1}_{B_x^{(2)}}\,. \tag{1.14}$$

It is shown in [43], that the averaging principle according to (1.8) is inappropriate for representing the effective dynamics if the size of the most

dominant eigenvalue $\lambda_1(x) \neq 0$ of $\mathcal{L}_x$ is comparable to the size of the smallness parameter ϵ, that is, if we explicitly couple $\lambda_1(x)$ to ϵ and assume for every x

$$\lambda_1(x) \;=\; \epsilon\,\tilde{\lambda}_1(x), \qquad \tilde{\lambda}_1(x) \,=\, \mathrm{ord}(1). \tag{1.15}$$

This establishes a relation between the time scale of the fast motion and the exit rate/time from the metastable subsets in the fast DOF since $\bar{\tau}_x^\epsilon$ can be expressed in terms of λ_1; compare (1.13).

As a consequence of (1.15) we can express $\mathcal{L}_x$ in the following form:

$$\mathcal{L}_x \;=\; \epsilon\,\tilde{\lambda}_1(x)\left\langle \cdot, u_1(x,\cdot)\right\rangle_{\mu_x} u_1(x,\cdot) \,+\, \mathcal{R}_x \,=\, \epsilon\,\mathcal{L}_x^{\mathrm{act}} \,+\, \mathcal{R}_x, \tag{1.16}$$

with $\mathcal{R}_x u_k = 0$ for $k = 0, 1$.

1.4 Conditional Averaging

The scaling assumption (1.15) represents a modeling step which will lead tothe derivation of the *principle of conditional averaging* that may yield an appropriate reduced model in cases where the ordinary averaging scheme fails: Since we observe rapid sampling of the invariant density μ_x in each of the metastable subsets, we propose averaging over each of these sets alone and to couple the resulting systems by a Markovian switching process which describes the flipping behaviour between the metastable sets.

In [43], these limit dynamics are derived in terms of multiscale analysis of the Fokker-Planck equation (1.3) by *projecting the ensemble dynamics onto the subspace spanned by the dominant spectrum* of $\mathcal{L}_x$. To this end, we define the projection

$$(\tilde{\Pi} f)(x,y) \;=\; \left\langle f, \mathbf{1}\right\rangle_{\mu_x} \mathbf{1} \,+\, \left\langle f, u_1\right\rangle_{\mu_x} u_1(x,\cdot),$$

and exploit $\tilde{\Pi}\mathcal{R}_x = 0$. For the limit dynamics $\rho^\epsilon \to \rho^0$ we have $\tilde{\Pi}\rho^0 = \rho^0$ and, similarly

$$\partial_t \rho^0 \;=\; \tilde{\Pi}\mathcal{L}^\epsilon \tilde{\Pi}\rho^0 \,=\, (\tilde{\Pi}\mathcal{L}_y\tilde{\Pi} \,+\, \mathcal{L}_x^{\mathrm{act}})\rho^0, \tag{1.17}$$

where we have used (1.16). The explicit form is derived by reformulating for $\rho^0 = \tilde{\Pi}\rho^0$ the ansatz

$$\rho^0(t,x,y) \,=\, A_0(t,x)\mathbf{1}(x,y) \,+\, A_1(t,x)u_1(x,y).$$

Assuming equality in (1.14), we obtain the equivalent formulation

$$\rho^0(t,x,y) \,=\, c_1(t,x)\mathbf{1}_{B_x^{(1)}}(y) \,+\, c_2(t,x)\mathbf{1}_{B_x^{(2)}}(y),$$

with $\mathbf{1}_D$ denoting the indicator function of a set D. We proceed by assuming that the boundary between the metastable subsets $B_x^{(1)}$ and $B_x^{(2)}$ does not

depend on x. By applying the orthogonal projections $\left\langle \cdot, \mathbf{1}_{B_x^{(i)}} \right\rangle_{\mu_x}$ for $i = 1, 2$ to both sides of equation (1.17), we end up with the evolution system

$$\begin{pmatrix} \partial_t c_1 \\ \partial_t c_2 \end{pmatrix} = \begin{pmatrix} \overline{\mathcal{L}}^{(1)} & 0 \\ 0 & \overline{\mathcal{L}}^{(1)} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \mathcal{Q}_x \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \tag{1.18}$$

$$\mathcal{Q}_x = \frac{|\lambda_1|}{\epsilon} \begin{pmatrix} -\mu_x(B_x^{(2)}) & \mu_x(B_x^{(2)}) \\ \mu_x(B_x^{(1)}) & -\mu_x(B_x^{(1)}) \end{pmatrix}, \tag{1.19}$$

where the operators $\overline{\mathcal{L}}^{(i)}$ are given for $i = 1, 2$ by

$$\overline{\mathcal{L}}^{(i)} = \frac{\sigma^2}{2}\Delta_x - \left(\int D_x V(x, y)\, \mu_x^{(i)}(y)\, \mathrm{d}y \right) \cdot D_x \,.$$

The densities $\mu_x^{(1)}$ and $\mu_x^{(2)}$ are supported on $B_x^{(1)}$ and $B_x^{(2)}$ only, and are given explicitly by

$$\mu_x^{(i)}(y) = \frac{1}{\mu_x(B_x^{(i)})}\, \mu_x(y)\, \mathbf{1}_{B_x^{(i)}}, \qquad i = 1, 2. \tag{1.20}$$

The second term $\mathcal{Q}_x$ in (1.18) denotes a rate matrix[1] that generates a transition process between the levels $i = 1$ and $i = 2$ *for fixed slow variable* x.

Consequently, the effective x dynamics within the SDE formulation will be characterized as follows:

$$\dot{x}^0 = -\int D_x V(x, y) \mu_x^{(I(t,x))}(y)\, \mathrm{d}y + \sigma\, \dot{W}_1, \tag{1.21}$$

with $I(t, x)$ denoting the Markov jump process whose transition rates are given by $\mathcal{Q}_x$. It turns out that $\overline{\mathcal{L}}^{(i)}$, $i = 1, 2$ is an effectively computable Fokker-Planck generator corresponding to the averaged potential $\overline{V}^{(i)}$ with

$$\begin{aligned} \int D_x V(x, y)\, \mu_x^{(i)}(y)\, \mathrm{d}y &= D_x \overline{V}^{(i)}(x), \\ \overline{V}^{(i)}(x) &= -\frac{1}{\beta} \ln \int_{B_x^{(i)}} \exp(-\beta V(x, y)) \mathrm{d}y. \end{aligned} \tag{1.22}$$

1.5 Related Approaches

The conditionally averaged dynamics (1.21) may serve as the starting point for further examinations which arise from a variety of questions and concerns

[1] This is not given a priori as we have to consider the matrix $\mathcal{Q}_x$ in a weighted space. The function spaces under consideration are examined in detail in Section 3.5.1. For advance information, see Remark 3.5.1.

arising from a closer inspection of system (1.18) and its derivation[2]. In this direction, we may basically turn our attention to the derivation of a thorough understanding of the conditionally averaged system on the one hand, and, on the other hand, a reconsideration of the asymptotic procedure that was used to obtain the result. These considerations will motivate two fundamentally different approaches to retain the principle of 'Conditional Averaging', which we shortly introduce in the following. A completion of the introductory comments on the two approaches is given at the beginning of the corresponding chapters.

The first approach is dedicated to obtain a deeper insight into the nature of the conditionally averaged system. To this end, we take advantage of the methodology employed to extract the effective dynamics (1.21). The result tells us that each metastable subset of the fast dynamics is connected to one averaged equation. This motivates the idea to construct a system of fast-slow equations which allows for the incorporation of temporal fast scale effects in a natural way: the fast motion *within* one metastable subset is approximated by an irreducible subprocess that corresponds to a stochastic differential equation. We then parameterize a Markov chain model that controls the switches from one (sub)process to the other according to the transition rates between the metastable subsets of the original dynamics. A reduced system in the slow variable is then obtained by applying the well-known averaging results from [35, 28, 25, 15] to each of these stochastic differential equations.

For the second approach, we switch attention to the formal analysis of the Fokker-Planck equation (1.3) in the asymptotic limit $\epsilon \to 0$, and reconsider the derivation of the effective x dynamics over order unity time scale. Separating the time scale over which the x dynamics proceed and the time scale of the metastable transitions will enable us to obtain a categorization of the various kinds of long-time effective behaviour that can emerge from the system (1.3). The conditionally averaged system (1.18) is then retained as one of the different scenarios that now include metastable transitions that can also occur along the dynamics of the slow variable x.

Figure 1.1 visualises the structuring of the thesis into two independent parts corresponding to the two different approaches to 'Conditional Averaging'. The key idea of the ansatz illustrated in the left-hand side of the figure is to replace the fast dynamics within each of the metastable sets by irreducible processes that are specified as Ornstein-Uhlenbeck (OU) processes. The OU processes are coupled by a transition chain that recovers the transition rates between the metastable fast dynamics for fixed x. After applying simple averaging techniques to the approximated system with OU processes, we finally end up with the averaged dynamics that we call *OU-averaged dynamics* in the following. Consequently, the approach requires two different

[2]We will dwell on these questions in Sections 2.1 and 3.1

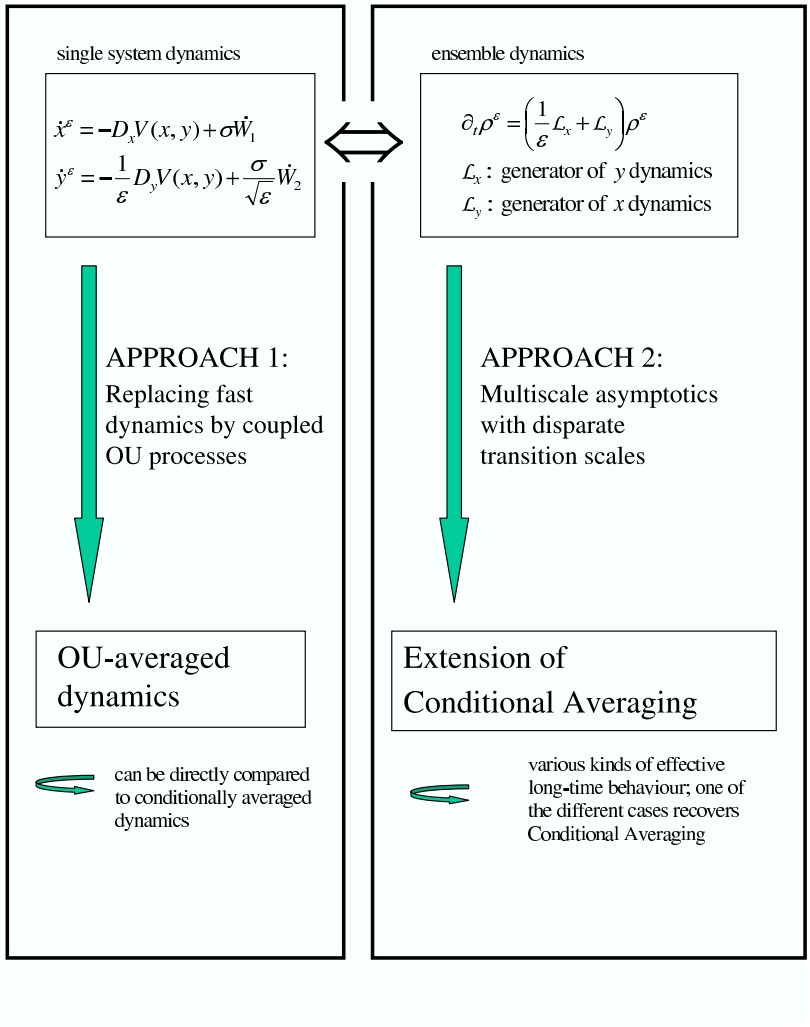

Figure 1.1: Two possible ansatzes to derive Conditional Averaging.

steps, "replacing of fast dynamics" and "averaging" for short, but the essential step is given by the replacement of the fast dynamics by coupled OU processes such that the approach as a whole will be called *"Replacing (metastable) fast dynamics by coupled OU processes"*.

The ansatz that is pursued in the second part of the thesis is shown at the right-hand side of Figure 1.1. Now, we start from the ensemble dynamics (1.3), and develop a systematic mathematical strategy for the derivation of appropriate reduced models. All of these models are derived from a self-consistent mathematical framework for eliminating the unresolved variables that is mathematically rigorous in a suitable asymptotic limit. In so doing, we distinguish between three different parameters: the smallness parameter ϵ which describes the time scale separation of the y and x motion, the pa-

rameter δ which defines the time scale of the metastable transitions that are induced by the y dynamics, and, finally, the parameter $\tilde{\epsilon}$ which describes the time scale of the metastable transitions that are induced by the x dynamics. The reduced model that is appropriate to describe the effective dynamics in the limit $\epsilon \to 0$ will depend on the relation of the smallness parameter ϵ to the parameters δ and $\tilde{\epsilon}$ that describe the order of the metastable transitions. Therefore, we will refer to the approach by calling it *"Multi-scale asymptotics with disparate transition scales"*. Assuming that the fast dynamics exhibits metastability (with two metastable sets), the procedure provides us with reduced model equations that can be considered as an extension of the principle of Conditional Averaging: The solution over time scale ord(1) consists in any case of the averaged generators $\overline{\mathcal{L}}^{(1)}$ and $\overline{\mathcal{L}}^{(2)}$; as for the dynamics (1.18), we again obtain some exchange term, but this time it can appear on a longer time scale and is not necessarily given by $\mathcal{Q}_x$.

Notation

Throughout the thesis we use the following relation symbols:

$\approx$	approximately equal
$\sim$	proportional
$\simeq$	asymptotically equal: $f(\epsilon) \simeq g(\epsilon)$ iff $\lim_{\epsilon\to 0} \frac{f(\epsilon)}{g(\epsilon)} = 1$
$\doteq$	approaches the limit
$\div$	$f(\epsilon) \div g(\epsilon)$ iff $\frac{f(\epsilon)}{g(\epsilon)} = 1 + \mathcal{O}(\epsilon)$
$\ll$	asymptotically smaller
$\gg$	asymptotically larger
$\lesssim$	asymptotically equal or smaller
$\gtrsim$	asymptotically equal or larger
ord	$f(\epsilon) = \mathrm{ord}(g(\epsilon))$ iff $\lim_{\epsilon\to 0} \frac{\lvert f(\epsilon)\rvert}{\lvert g(\epsilon)\rvert} = C,\ C > 0$
$\mathcal{O}$	$f(\epsilon) = \mathcal{O}(g(\epsilon))$ iff $\lim_{\epsilon\to 0} \frac{\lvert f(\epsilon)\rvert}{\lvert g(\epsilon)\rvert} = C,\ C \geq 0$
o	$f(\epsilon) = o(g(\epsilon))$ iff $\lim_{\epsilon\to 0} \frac{\lvert f(\epsilon)\rvert}{\lvert g(\epsilon)\rvert} = 0$

ACKNOWLEDGEMENT. Last but not least I want to say thank you to all the persons who contributed to the here presented results with their support and assistance. Special thanks I want to express to Christof Schütte for guiding me towards this field and offering permanent support.

I also thank Andrew Stuart and Grigoris Pavliotis for three fruitful months at the University of Warwick. Both of them never grew tired of answering my questions and encouraged me on the way to solutions.

Many thanks go to the biocomputing group at Free University for stimulating interdisciplinary atmosphere, and for enduring me in the last seven months. Special thanks to Carsten who agreed to read my drafts again and again.

Finally, I thank my family and my friends for permanent moral support.