University of Amsterdam

MSc Mathematical Physics

Master Thesis Freek Witteveen

The Mori-Zwanzig Formalism and Stochastic Modelling of Multiscale Dynamical Systems

Author:
Freek Witteveen

Supervisor:
prof. dr. Daan Crommelin
prof. dr. Jason Frank

Examination date:
June 27, 2016

Abstract

We investigate the Mori-Zwanzig formalism as a mathematical tool to study multi-scale systems and project the equations of motion onto a small set of relevant variables. This originates from statistical physics, and we explain how it can be used to derive a generalized Langevin equation. Typically, one obtains an equation of motion for the relevant degrees of freedom of the system that has a stochastic term, but also a memory kernel. Especially in systems which do not have a clear scale separation this memory kernel is important. We discuss approximations that start from the Mori-Zwanzig formalism to find (data-driven) numerical methods for simulation and prediction of a small number of selected degrees of freedom in high dimensional dynamical systems. To verify the usefulness of these methods we apply them to two toy models: the Kac-Zwanzig heat bath model and the Lorenz '96 model.

Title: The Mori-Zwanzig Formalism and Stochastic Modelling of Multiscale Dynamical Systems
Author: Freek Witteveen, freek.witteveen@student.uva.nl, 6065031
Supervisor: prof. dr. Daan Crommelin, prof. dr. Jason Frank
Examination date: June 27, 2016

Korteweg-de Vries Institute for Mathematics
University of Amsterdam
Science Park 105-107, 1098 XG Amsterdam
`http://kdvi.uva.nl`

Centrum voor Wiskunde en Informatica
Science Park 123, 1098 XG Amsterdam
`http://www.cwi.nl`

Acknowledgements

First and foremost I would like to thank Daan Crommelin for being an excellent and very kind advisor, always making time in his busy schedule for regular meetings. Also I am very grateful to Jason Frank, who also took part in the discussions with Daan and gave useful recommendations. More generally I have been lucky to be welcomed so warmly at the CWI by all the friendly colleagues. Finally my family has been very supportive, and of course Rasila and Sjors were always there for diversion and advice.

Contents

Observe what happens when sunbeams are admitted into a building and shed light on its shadowy places. You will see a multitude of tiny particles mingling in a multitude of ways... their dancing is an actual indication of underlying movements of matter that are hidden from our sight... It originates with the atoms which move of themselves. Then those small compound bodies that are least removed from the impetus of the atoms are set in motion by the impact of their invisible blows and in turn cannon against slightly larger bodies. So the movement mounts up from the atoms and gradually emerges to the level of our senses, so that those bodies are in motion that we see in sunbeams, moved by blows that remain invisible.

Lucretius, On the Nature of Things

1 Introduction

In nature one often finds complicated systems with many degrees of freedom. An example that immediately meets the eye is the weather, and its longer term behaviour, the climate. Its dynamics are derived from the equations for fluids and gases (the Navier-Stokes equations), thermodynamics, and gravitational/Coriolis effects. Another example is the behaviour of liquids and other materials themselves. Here we also know a description for the equations of motion of the individual molecules, which we can consider as an N-body problem, for very large N. However, in both cases even though a microscopic description is available, this does not mean we can easily understand the macroscopic phenomena we see. The dynamics depends very sensitively on initial conditions (which makes numerical simulation problematic). Besides this sensitivity we also often have incomplete knowledge of the state of the system, we can not always measure what happens at the microscopic level. For measuring the weather one always has to rely on only a finite number of measurement points, and similarly it is impossible to determine the positions and velocities of all molecules in a fluid at the same time. Finally, in practice these systems are often too large to simulate numerically for a long enough period of time (Thijsen, 1999).

However, an exact solution of all degrees of freedom is often not that relevant if we are interested in understanding macroscopic phenomena. For instance, in the example of climate, we do not really care about the particular movement of each tiny volume of the ocean or atmosphere, but we rather want to know how long term temperature averages will evolve (under influence of external perturbations). Or maybe we are interested in the periodic behaviour of certain oceanic currents. Also for the N-body problem, we do not need to know the trajectory of each particle, but we rather want to know the thermodynamic properties of the material. For example in a fluid, we might

want to be able to derive the transport coëfficients such as the diffusion constant from the microscopic laws, or try to understand phase transitions (sudden changes in some order parameter).

In both examples we have a very large system, but in the end we are interested only in specific functions of all the variables that describe the system. To what extent is it possible to derive from the microscopic description a reliable macroscopic description, without having to solve all of the dynamics? In statistical physics many techniques have been invented to deal with this problem, and in this thesis we will discuss one particular method, called the Mori-Zwanzig formalism. In this procedure, one first chooses some degrees of freedom that are thought to be of interest, and that have behaviour on a relatively long time scale or large spatial scale. Then starting from the equations of motion for the complete system, one projects (in a way that will be described later) onto the relevant variables. It turns out that through formal manipulations it is always possible to write the resulting equations in a particular form, where the time derivative of the relevant variables depends on 3 terms:

1. A self-interaction part, which is a function of the relevant variables,
2. A memory part, which depends on the value of the relevant variables at all previous times,
3. A part that depends on the remaining variables.

This will be made precise in chapter 2. The idea is that if the variables are chosen in the right way, one may hope that the third term consists of 'fast' variables, and it is reasonable to approximate it by a random noise, especially in the situation where the initial conditions of these fast variables are not known, and they should be considered as random variables. A situation where this approximation by a stochastic equation is exact in an appropriate limit is discussed in section 2.4. Such a stochastic differential equation (a generalized Langevin equation) may still be hard to solve analytically, but it is at least more amenable to computer simulation.

Let us make this idea more concrete with a physical example, of Brownian motion. Here the situation is that we have a pollen particle, contained in a liquid (which is a bath of molecules). We are interested in the dynamics of the pollen particle. We can give a microscopic description of the system that takes all the molecules into account, but this is too hard to solve. Thus we only consider the equation of motion of the pollen particle. Denote by v the velocity of the particle, then we want to obtain an equation for the time derivative $\dot{v}$. One can see experimentally that the part of the equation of motion that depends on v, which is the friction in the liquid, is linear. Thus we write

$$\frac{dv}{dt} = -\zeta v + F(t) \tag{1.1}$$

where $F(t)$ denote all the forces that come from collisions of the pollen particle with some molecule. In this case we have a clear separation of time scales: the pollen particle will move much slower than the molecules, and the molecules also collide

rapidly with each other. Because of this, we can make some assumptions on $F(t)$, namely that it is random, it has no preferred direction (so the expectation value $\mathbf{E}F(t)$ is zero) and it oscillates so rapidly that there is no time correlation of F with itself. This means that F can be seen as white noise. Moreover, in this case the memory is very short (which is related to the fact that F has no time correlations, we will discuss this in 2.3). With this approximation, equation (1.1) becomes the Langevin equation, a stochastic differential equation, that describes well the process of Brownian motion. We have dealt with the two problems: we focus on the variable of interest (the movement of the pollen) and we do not have to solve for the movement of the molecules. Also our ignorance of the (initial) state of the molecules is incorporated in the randomness of the noise.

The idea to use statistical processes to replace certain degrees of freedom is an old idea in nonequilibrium statistical physics (Einstein, 1905), and has much later been realized to also be of use in climate dynamics (Hasselman, 1976):

> The essential feature of stochastic climate models is that the non-averaged "weather" components are also retained. They appear formally as random forcing terms. The climate system, acting as an integrator of this short-period excitation, exhibits the same random-walk response characteristics as large particles interacting with an ensemble of much smaller particles in the analogous Brownian motion problem.

This is a line of research continuing to this day (Palmer and Williams, 2010), (Imkeller and Storch, 2001). In the case of Brownian motion we saw that there was not really any memory. This is due to the large scale separation between the pollen particle and the molecules in the liquid. Such a process without memory is called a *Markov process*. The Mori-Zwanzig method of dimension reduction however leads in general to systems that do have a non-trivial memory term which need not be Markovian. This is especially true in the context of climate dynamics, where there is typically not a large scale separation between the different processes (Nastrom and Gage, 1985). This suggests that investigating these memory terms is very relevant to the project of improving weather and climate models by including stochastic processes. This procedure is often called *stochastic parametrization* and in chapter 3 we will suggest some methods to estimate memory effects in estimation procedures for stochastic parametrization, and in chapter 4 we apply these ideas to a simple heat bath model from statistical physics and to the Lorenz '96 model, a toy model for the dynamics of atmospheric phenomena.

The Mori-Zwanzig formalism already dates back from the 1960s (Zwanzig, 1961), (Mori, 1965a), but using it in order to design numerical schemes for multi-scale dynamical systems is a recent development (Chorin et al., 2000). Most of the work in this thesis is not original, but an overview of existing literature. The application to the Lorenz '96 model however is to our best knowledge not present in the current literature.

Non-Markov is the rule, Markov is the exception.

Van Kampen, Remarks on non-Markov Processes

2

The Mori-Zwanzig formalism

To get an idea of what we want to achieve we will first consider an easy example, a linear ordinary differential equation:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}$$

with A an arbitrary 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Of course we know the solution of the system, but we now try to find it in a different way. Consider the equation for y which is $\dot{y} = a_{21}x + a_{22}y$, which we can solve if we assume x to be some given function:

$$y(t) = e^{a_{22}t}y(0) + \int_0^t e^{a_{22}(t-s)}a_{21}x(s)ds.$$

Inserting this in the equation for x we obtain

$$\begin{aligned} \dot{x}(t) &= a_{11}x(t) + \int_0^t a_{12}e^{a_{22}(t-s)}a_{21}x(s)ds + a_{12}e^{a_{22}t}y(0) \\ &= a_{11}x(t) + \int_0^t K(t-s)x(s)ds + F(t) \end{aligned}$$

so we obtain a single equation for x, which consists of three terms, a 'Markovian' term which just depends on $x(t)$, a term which depends on the values of $x(s)$ for s

at previous times (here $K(s) = a_{12}e^{a_{22}s}a_{21}$ is called the 'memory kernel') and finally an 'external' term $F(t) = a_{12}e^{a_{22}t}y(0)$ which does not depend on x but only on the initial condition of y. What we have done in this way is that we have taken a system of differential equations for some variables (x and y), and projected onto an equation for only a subset of the variables (x), so we have fewer equations. The price we pay for doing this is that we get a memory term in our equation, and a term that depends on the 'orthogonal dynamics' (the dynamics of y). We managed to get from 2 equations to 1, but it is a much more complicated looking one! In this chapter we will generalize this procedure to arbitrary ODE's, and explain how this procedure can be used as a starting point to approximate the dynamics of the system. If the system is not as easy to solve as our example (where we can explicitly describe the solutions of the full system) this may be a valuable way to approximate the original problem by solving a simpler one with fewer variables.

2.1 Derivation of the Mori-Zwanzig formalism

In this section, we will describe the Mori-Zwanzig procedure, in which we take a system of differential equations and rewrite it in a form that resembles a generalized Langevin equation, as formulated by Zwanzig (Zwanzig, 1973a) and Mori (Mori, 1965a). The derivation is standard, we more or less follow (Chorin et al., 2000). Thus, we start to consider the following problem on $M = \mathbf{R}^n$:

$$\frac{dx(t)}{dt} = R(x(t)) \tag{2.1}$$

where $x \in \mathbf{R}^n$ and $R : \mathbf{R}^n \to \mathbf{R}^n$ is a function which we assume to be such that the problem is well posed, and has a unique solution on some time interval given initial conditions (so R should for instance be uniformly Lipschitz continuous). This gives rise to a solution for every initial condition. We capture this with a flow $x(t)$ which is such that $x(0) = x$ (so the variable x represents the initial condition), and for fixed x is a solution to (2.1).

Next we want to know how a function evolves along this flow. For this we define a time evolution operator S^t which is such that for any function f we have $S^t f(x) = f(x(t))$. The equation of motion for this time evolution of functions is given by

$$\frac{\partial}{\partial t} S^t f(x) = (R(x) \cdot \nabla) S^t f(x).$$

Writing $L = R(x) \cdot \nabla$ we thus get the Liouville equation

$$\begin{aligned} \frac{\partial}{\partial t} S^t f &= L S^t f \\ S^0 f &= f \end{aligned} \tag{2.2}$$

Now S^t is a semigroup with generator L, and we may write $S^t = e^{tL}$ (but it is good to keep in mind that this is only notation). Since L is the generator of the semigroup we know that $S^t L = L S^t$.

We can make the above formulation a bit more general, by considering other manifolds than $\mathbf{R}^n$. Let M be any manifold, then corresponding to 2.1 we have a vector field X that defines a flow equation on M:

$$\begin{aligned} \frac{\partial}{\partial t}\phi_t(x) &= X(\phi_t(x)) \\ \phi_0(x) &= x \end{aligned}$$

The corresponding Liouville operator will just be the Lie derivative, so the evolution of functions along the flow satisfies the Liouville equation

$$\begin{aligned} \frac{\partial}{\partial t} S^t f &= \mathcal{L}_X S^t f \\ S^0 f &= f. \end{aligned}$$

In the particular case of Hamiltonian dynamics we have a symplectic manifold (M, ω) and a Hamiltonian function $H : M \to \mathbf{R}$ with Hamiltonian vector field X_H. In this situation the Lie derivative is given by a Poisson bracket:

$$\frac{\partial}{\partial t} S^t f = \mathcal{L}_{X_H} S^t f = \{H, f\}.$$

Next we use an idea from statistical physics. We want to be able to deal with a situation where we do not know the initial conditions of all variables. Thus we suppose that the initial conditions are drawn from some probability distribution μ. This allows us to consider the Hilbert space $L^2(M, \mu)$ with inner product given by

$$(f, g) = \int_M f(x) g(x) \mu(dx).$$

We want to project the dynamics onto some subspace of $L^2(M, \mu)$, so we choose an orthogonal projection $\mathcal{P} : L^2(M, \mu) \to L^2(M, \mu)$. We can split the operator L in the following way:

$$L = \mathcal{P}L + (1 - \mathcal{P})L$$

We let $S^t_\perp$ be the time evolution operator corresponding to the orthogonal dynamics, that is, $S^t_\perp$ is the solution of the orthogonal dynamics equation

$$\begin{aligned} \frac{\partial}{\partial t} S^t_\perp g &= (1 - \mathcal{P}) L S^t_\perp g \\ S^0_\perp g &= g. \end{aligned} \tag{2.3}$$

As a remark, the existence of the solution of these equations is nontrivial, even if the original problem is well-posed. In (Givon et al., 2005) the existence (or weak existence)

is shown for some specific projection operators. In semigroup notation, one would write $S^t_\perp = e^{(1-P)L}$, and from the equation it is easy to see that if g starts out in the null space of $\mathcal{P}$, it will remain in this null space. We can now derive the Dyson formula:

$$S^t = S^t_\perp + \int_0^t S^{t-s}\mathcal{P}LS^s_\perp ds. \tag{2.4}$$

Let us check this by differentiation of the right hand side:

$$\begin{aligned}
&\frac{\partial}{\partial t}(S^t_\perp + \int_0^t S^{t-s}\mathcal{P}LS^s_\perp ds) \\
&\qquad = \frac{\partial}{\partial t}S^t_\perp + S^{t-s}\mathcal{P}LS^s_\perp|_{s=t} + \int_0^t \frac{\partial}{\partial t}S^{t-s}\mathcal{P}LS^s_\perp ds \\
&\qquad = (1-\mathcal{P})LS^t_\perp + \mathcal{P}LS^t_\perp + \int_0^t LS^{t-s}\mathcal{P}LS^s_\perp ds \\
&\qquad = L(S^t_\perp + \int_0^t S^{t-s}\mathcal{P}LS^s_\perp ds)
\end{aligned}$$

so we see that $S^t_\perp + \int_0^t S^{t-s}\mathcal{P}LS^s_\perp ds$ satisfies the Liouville equation (2.2) and since at $t = 0$ we clearly have that both sides are the identity operator, we conclude equality by the uniqueness of the solution of (2.2).

With the projection we split the time evolution of a function in 2 parts:

$$S^t f = S^t\mathcal{P}f + S^t(1-\mathcal{P})f$$

We can use the Dyson formula (2.4) to rewrite the second term, and for notational convenience we let $F = (1-\mathcal{P})f$ (and we will often use the notation $F(t) = S^t_\perp F$):

$$S^t(1-\mathcal{P})f = S^t F = S^t_\perp F + \int_0^t S^{t-s}\mathcal{P}LS^s_\perp F ds.$$

Taking this together we get

$$S^t f = S^t\mathcal{P}f + \int_0^t S^{t-s}\mathcal{P}LS^s_\perp F ds + S^t_\perp F. \tag{2.5}$$

The main interest of this equation is in the case where we consider the function Lf in which case $S^t Lf = \frac{\partial}{\partial t}S^t f$ so (2.5) becomes

$$\frac{\partial}{\partial t}S^t f = S^t\mathcal{P}Lf + \int_0^t S^{t-s}\mathcal{P}LS^s_\perp F ds + S^t_\perp F \tag{2.6}$$

now with $F = (1-\mathcal{P})Lf$. How to interpret this equation? The first term is the Markovian part of the equation, which just depends on f at time t. The second term is a memory term, which depends on f throughout the interval $[0,t]$, we will often denote $\mathcal{P}LS^s_\perp F$ by $K(f,s)$ and call it the memory kernel. It is in the image of the projection $\mathcal{P}$. Finally the third term describes the dynamics orthogonal to the image

of $\mathcal{P}$. Indeed, it is easy to see that $\mathcal{P}S^t F = 0$ from the definition of F. Often in physical systems (we will see some examples later on), we do not know the initial conditions for the orthogonal term, and we can approximate F by a stochastic process (see (Kubo et al., 1985) for discussions about the legitimacy of such an approximation in a statistical physics context). If one indeed does so, the above derivation can be seen as a derivation of a generalized Langevin equation.

The issue is now to choose the projection $\mathcal{P}$ in such way that (2.6) becomes an interesting formula that allows making good approximations. Notice that (2.6) is still exact and is only a reformulation of the Liouville equation (2.2). Also, at this point we can still plug in any function f, but mostly we will want to use this formalism to study the behaviour of functions that are in the image of the projection $\mathcal{P}$.

Choice of projection operator

There are two choices of projection operator that are most common:

1. One way to get a projection in $L^2(M,\mu)$ is by using conditional expectations. Given functions $f, g \in L^2(M,\mu)$ the conditional expectation value

$$\mathcal{P}f = \mathbf{E}(f|g)$$

is defined to be the orthogonal projection of f onto $L^2(M,\sigma(g),\mu)$. Here $\sigma(g)$ denotes the σ-algebra generated by g, and $L^2(M,\sigma(g),\mu)$ denotes the square-integrable functions on M that are measurable with respect to the measure space $(M,\sigma(g),\mu)$ (where μ of course should be read as the restriction of μ to $\sigma(g)$). This definition maybe does not really give too much insight in the nature of the projection, but rewriting the definitions one finds that $L^2(M,\sigma(g),\mu)$ can be identified with the functions of g, so $L^2(M,\sigma(g),\mu) \cong \{h \in L^2(M,\mu) | h = k \circ g, k : \mathbf{R} \to \mathbf{R}\}$. In the physics literature this projection is called the 'nonlinear projection' (the projector itself is of course linear, but it projects onto a space of nonlinear functions)(Chorin et al., 2000). We can also do this for multiple functions and define $\mathbf{E}(f|g_1,\ldots,g_m)$ to be the orthogonal projection onto the space of all functions of the g_i. For notational convenience we will consider in this case $g = (g_1,\ldots,g_m)$ as a vector valued function and write $\mathbf{E}(f|g)$.

Let us give a more explicit formula for $\mathbf{E}(f|g)$, denote by $C_x = \cap_i g_i^{-1}(g_i(x))$ for $x \in M$, then for all x such that $\mu(C_x) \neq 0$ we have

$$\mathcal{P}f(x) = \mathbf{E}(f|g)(x) = \frac{\int_{C_x} f(x)\mu(dx)}{\mu(C_x)}.$$

Notice that the value of $\mathcal{P}f$ on sets of measure 0 is not very important since we work in $L^2(M,\mu)$. A (formal) way to write this is using delta functions:

$$\mathcal{P}f(x) = \mathbf{E}(f|g)(x) = \frac{\int f(\tilde{x})\delta(g(\tilde{x}) - g(x))\mu(d\tilde{x})}{\int \delta(g(\tilde{x}) - g(x))\mu(d\tilde{x})}. \tag{2.7}$$

2. Another way to construct a projection is by taking a set of functions $h_1, \ldots, h_m$ and projecting onto the linear span of these functions (this is known as the 'linear projection'). Explicitly the projection is then given by

$$\mathcal{P}f = \sum_{i,j=1}^{m} h_{ij}^{-1}(f, h_i)h_j$$

where h^{-1} is the inverse of the matrix with entries $h_{ij} = (h_i, h_j)$ (with respect to the inner product on $L^2(M, \mu)$). From the construction it is immediate that this is an orthogonal projection. If we choose the functions h_i orthonormally the expression becomes conveniently

$$\mathcal{P}f = \sum_{i=1}^{m}(f, h_i)h_i.$$

We can also use this projection to approximate the nonlinear projection, by taking the h_i to be a such that for m going to infinity, the h_i span $L^2(M, \sigma(g), \mu)$. This can be useful in numerical simulations.

Both these projections depend on the choice of measure μ. What measure should be chosen? The best situation is if the measure is time invariant, which means that S^t is a unitary operator with respect to the L^2 inner product for all times $t \geq 0$. Equivalently the measure μ_t defined by

$$\int f(x)\mu_t(dx) = \int S^t f(x)\mu(dx)$$

is constant in time, so $\mu_t = \mu$. Unitarity of S^t is equivalent with skew-adjointness of L by Stone's theorem. In general if U is unitary and $\mathcal{P}_X$ is a projection onto X, then $U\mathcal{P}_X = \mathcal{P}_{UX}U$. Thus if we consider the conditional expectation with respect to an invariant measure we have

$$S^t\mathbf{E}(f|g) = \mathbf{E}(S^t f|S^t g)$$

which means that we can rewrite the Markovian term in (2.6) as $S^t\mathbf{E}(Lf|g) = \mathbf{E}(LS^t f|S^t g)$.

For the first choice, the most common situation is where we let $M = \mathbf{R}^n$. We write $z = (x_1, \ldots, x_m, y_1, \ldots, y_{n-m})$ for the coordinates and we take $\mu(dz) = \rho(x, y)dx_1 \ldots dx_m dy_1 \ldots dy_{n-m}$ as a measure. We project onto the space of functions of coordinate functions $x = (x_1, \ldots, x_m)$, and we obtain

$$(\mathcal{P}f)(x) = \frac{\int f(x, y)\rho(x, y)dy_1 \ldots dy_{n-m}}{\int \rho(x, y)dy_1 \ldots dy_{n-m}}.$$

In this specific case, where the projection $\mathcal{P}$ should be thought of as projecting onto the functions of the first m variables, we can consider the Mori-Zwanzig equation (2.6) for the coordinate function x_i to find equations for the flow (denoting $S^t x_i = x_i(t)$):

$$\frac{\partial}{\partial t}x_i(t) = \tilde{R}_i(x(t)) + \int_0^t K_i(x(t-s), s)ds + F_i(t) \tag{2.8}$$

where we let $\tilde{R} = \mathbf{E}(R|x)$ and F and K are defined in the usual way. We have assumed invariance of the measure to commute S^t with $\tilde{R}$. Now we can apply the projection $\mathcal{P}$ once more to both sides of the equation, and using that F_i will be in the null space of $\mathcal{P}$ we obtain

$$\frac{\partial}{\partial t}\mathcal{P}x_i(t) = \mathcal{P}\tilde{R}_i(x(t)) + \int_0^t \mathcal{P}S^{t-s}K_i(x,s)ds. \tag{2.9}$$

This equation tells us the expected evolution of the relevant variables x given initial conditions for x and initial conditions for y drawn from a distribution (with respect to the invariant measure, and conditioned on the initial condition of x). Equation (2.9) is known as *optimal prediction with memory* (Chorin et al., 2002).

The most crude approximation of the Mori-Zwanzig equation is to throw away the memory and the orthogonal part, leaving only the Markovian part. If we use the conditional expectation as projection operator this approximation is called *optimal prediction*. By definition of the projection operator $\tilde{R}(x(t)) = \mathbf{E}(Lx(t)|x(t))$ is the best approximation of $Lx(t)$ by a function of $x(t)$. A discussion of the error of optimal prediction with respect to the original system can be found in (Hald and Kupferman, 2001). The term $\tilde{R}$ can be found from a time series, provided that the measure μ is ergodic (with respect to S^t):

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T S^t f(x)dt = \int_M f(y)\mu(dy) \tag{2.10}$$

for almost all x and for any $f \in L^1(M,\mu)$.

Lemma 2.1.1. *Assume that the system* (2.1) *has an invariant ergodic measure μ, and we project onto a set of functions $g = (g_1,\ldots,g_m)$, which are such that $g, Lg \in L^2(M,\mu)$. Then for almost all initial data x we get*

$$\underset{f\in L^2(M,\sigma(g),\mu)}{\arg\min}\ (\lim_{T\to\infty}\frac{1}{T}\int_0^T ||\frac{dg(x(t))}{dt} - f(g(x(t)))||^2) = \mathbf{E}(Lg|g). \tag{2.11}$$

Proof. Define ϕ to be the function $\phi(x) = ||Lg(x) - f(g(x))||^2$. Then $\phi \in L^1(M,\mu)$ since $f \in L^2(M,\sigma(g),\mu)$ and $Lg \in L^2(M,\mu)$. That means we can use the ergodicity to see that for almost all initial conditions x

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T ||LS^t g(x) - f(S^t g(x))||^2 dt = \int ||Lg(x) - f(g(y))||^2\mu(dy) \tag{2.12}$$

and $g(x) = \mathbf{E}(Lx|x)$ is precisely the minimizer of this last expression. □

The implication is that if we have a time series for the projected variables g, we can find the optimal prediction term $\mathbf{E}(Lg|g)$ by making a least squares fit to the data for the derivatives of g.

2.2 Hamiltonian systems

Suppose the equation (2.1) we start with is Hamiltonian, for some Hamiltonian function $H : \mathbf{R}^{2n} \to \mathbf{R}$. Then we have a natural choice of measure to draw initial conditions from, the canonical ensemble with density function $\rho(x) = \frac{1}{Z} e^{-\beta H(x)}$ (with Z a normalization constant). Here we remark that this choice of measure means we are in an equilibrium situation (S^t is unitary for all t). In this case the Markovian part of the Mori-Zwanzig equation will be Hamiltonian again if we project onto a canonical set of variables. To see this, let us write the coordinates on $\mathbf{R}^{2n}$ as $(q_1, \ldots, q_m, p_1, \ldots, p_m, q_{m+1}, \ldots, q_n, p_{m+1}, \ldots, p_n)$ such that the equations of motion are

$$\begin{aligned} \dot{q}_i &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i}. \end{aligned}$$

For notational convenience we let $Q = (q_1, \ldots, q_m)$, $q = (q_{m+1}, \ldots, q_n)$ and similarly for P and p. We let the projection operator $\mathcal{P}$ be the conditional expectation with respect to the first $2m$ variables Q and P. Because we are in an equilibrium state we have unitary S^t and hence $S^t \mathbf{E}(f|Q,P) = \mathbf{E}(S^t f | S^t Q, S^t P)$. If we ignore the memory and orthogonal term for the moment we get

$$\frac{\partial}{\partial t} S^t Q_i = S^t \mathcal{P} L Q_i = S^t E(LQ_i|Q,P) = E(S^t LQ_i | S^t Q, S^t P)$$

Inserting the definitions and writing $Q_i(t)$ and $P_i(t)$ for $S^t Q_i$ and $S^t P_i$ we find

$$\begin{aligned} \dot{Q}_i(t) &= E(\frac{\partial H}{\partial P_i} | Q(t), P(t)) \\ &= \frac{\int \frac{\partial H}{\partial p_i} e^{-H} dq dp}{\int e^{-H} dq dp} \\ &= \int \frac{\partial H}{\partial P_i} e^{\tilde{H} - H} dq dp = \frac{\partial \tilde{H}}{\partial P_i} \end{aligned}$$

where $\tilde{H} = -\log \int e^{-H} dq dp$. Similarly we can derive the equation for $\dot{P}_i$ and we find that the optimal prediction is Hamiltonian with reduced Hamiltonian function $\tilde{H}$.

Another option is to use a microcanonical ensemble, that is, the model is restricted to a fixed energy surface (Zwanzig, 1961). Keeping the notation as above, the corresponding projection operator $\mathcal{P}$ can (formally) be written as

$$\begin{aligned} \mathcal{P} f(Q,P) &= \frac{\int f(Q,P,q,p) \delta(H(Q,P,q,p) - E) dq dp}{\int \delta(H(Q,P,q,p) - E) dq dp} \\ &= \frac{\int_{H^{-1}_{P,Q}(E)} f(q,p) d\sigma}{\int_{H^{-1}_{P,Q}(E)} d\sigma} \end{aligned}$$

with $H_{P,Q}(p,q) = H(P,Q,p,q)$ and with $d\sigma$ the area element on $H_{P,Q}^{-1}(E)$, provided that these hypersurfaces are sufficiently smooth.

2.3 Fluctuation-dissipation theorem

In physics, one often encounters relations between fluctuations (perturbations) of a system and dissipation (response). In our derivation of a generalized Langevin equation we also encounter such a relation, if we take the term F to represent the fluctuations, and the memory kernel $K(f,s)$ measuring the dissipation (the memory term will generally have a dampening effect, as the motion 'leaks' out of the subspace we are projecting upon. We have the following formula for K:

$$K(t) = \mathcal{P}LS_{\perp}^{t}F = \mathcal{P}LF(t)$$

If we take the projection $\mathcal{P}$ to be the linear projection onto a function f, we have an equilibrium measure (so $L = -L^*$) and we consider the time evolution of f we find

$$\begin{aligned} K(f,t) &= \frac{(LF(t),f)}{(f,f)}f = -\frac{(F(t),Lf)}{(f,f)}f \\ &= \frac{((1-\mathcal{P})F(t),Lf)}{(f,f)}f = \frac{(F(t),(1-\mathcal{P})Lf)}{(f,f)}f \\ &= \frac{(F(t),F(0))}{(f,f)}f \end{aligned}$$

using that $F(t) = (1-\mathcal{P})F(t)$ and the fact that $(1-\mathcal{P})$ is an orthogonal projection. Thus we conclude

$$K(f,t) = (F(t),F(0))\frac{f}{(f,f)} \tag{2.13}$$

which is a fluctuation-dissipation relation where we see that the memory kernel (the dissipation) is related to the time correlation of the fluctuations. In this view it is not surprising that we did not have a memory kernel in the example of Brownian motion in equation (1.1): we assumed that the noise was not time-correlated.

As a remark, there are quite a few statements that go under the name 'fluctuation-dissipation theorem' in statistical physics, and most of them are related to linear response theory, where the reaction to a weak external perturbation of the system is expressed in term of the autocorrelation of the system. For some low order perturbative results on the relation between linear response theory and the Mori-Zwanzig formalism, see (Lucarini and Wouters, 2012).

Another case that allows a fluctuation-dissipation theorem is the projection of a Hamiltonian system onto a canonical set of variables (as in the previous section). We then have:

Theorem 2.3.1. *Suppose we have a Hamiltonian system where we use the canonical measure to project onto a set of canonical variables P, Q, and with remaining canonical variables p, q, with Hamiltonian H and reduced Hamiltonian $\tilde{H}$. Then the memory kernel allows the following expression:*

$$K_X(P,Q,t) = (\nabla_{P,Q} - \nabla_{P,Q}\tilde{H}) \cdot \mathcal{P}(F_{P,Q}(0)F_X(t)) \tag{2.14}$$

where X is a function of P and Q.

This theorem is discussed in a more general form in (Darve et al., 2009) for the case of conservative dynamics (so $\nabla \cdot R = 0$). Our proof is an adaptation of their arguments.

Proof. Writing $L = R \cdot \nabla$, we know that $\nabla \cdot R = 0$, so

$$\begin{aligned}
K_X(P,Q,t) = \mathcal{P}LF_X(t) &= \int R \cdot \nabla F_X R(t) e^{\beta(\tilde{H}-H)} dqdp \\
&= \int \nabla \cdot (RF_X(t)) e^{\beta(\tilde{H}-H)} dqdp \\
&= \int \nabla_{P,Q} \cdot (R_{P,Q}F_X(t)) e^{\beta(\tilde{H}-H)} dqdp \\
&= \nabla_{P,Q} \cdot \int R_{P,Q}F_X(t) e^{\beta(\tilde{H}-H)} dqdp - \int R_{P,Q}F_X(t) \cdot \nabla_{P,Q} e^{\beta(\tilde{H}-H)} dqdp
\end{aligned}$$

where in the third equality we have written $\nabla = (\nabla_{P,Q}, \nabla_{p,q})$ and integrated by parts

$$\begin{aligned}
\int \nabla_{p,q} \cdot (R_{p,q}F_X(t)) e^{\beta(\tilde{H}-H)} dqdp &= \int R_{p,q}F_X(t) \cdot \nabla_{p,q} e^{\beta(\tilde{H}-H)} dqdp \\
&= \int F_X(t) R_{p,q} \cdot (\nabla_{p,q} H) \beta e^{\beta(\tilde{H}-H)} dqdp = 0
\end{aligned}$$

since $R_{p,q} \cdot \nabla_{p,q} H = 0$ (as $R_q = \frac{\partial H}{\partial p}$ and $R_p = -\frac{\partial H}{\partial q}$). Finally we write

$$\begin{aligned}
\mathcal{P}LF_X(t) &= (\nabla_{P,Q} - \beta \nabla_{P,Q}\tilde{H}) \cdot \int R_{P,Q}F_X(t) e^{\beta(\tilde{H}-H)} dqdp \\
&= (\nabla_{P,Q} - \beta \nabla_{P,Q}\tilde{H}) \cdot \mathcal{P}(R_{P,Q}F_X(t)) \\
&= (\nabla_{P,Q} - \beta \nabla_{P,Q}\tilde{H}) \cdot \mathcal{P}(((1-\mathcal{P})R_{P,Q})F_X(t)) \\
&= (\nabla_{P,Q} - \beta \nabla_{P,Q}\tilde{H}) \cdot \mathcal{P}(F_{P,Q}(0)F_X(t))
\end{aligned}$$

where we used that $\mathcal{P}(fF(t)) = \mathcal{P}(((1-\mathcal{P})f)F(t))$ for any function f since $\mathcal{P}F(t) = 0$ (an easy consequence of the definition of the conditional expectation). □

2.4 The Kac-Zwanzig heat bath model

We will now consider a classical model in which the Mori-Zwanzig equation can be worked out completely explicitly. It is a mechanical model for a heavy particle in a heat bath (Ford et al., 1965) (Zwanzig, 1980), in which a particle in some potential

has a linear interaction with a system of harmonic oscillators. The idea is that this represents an approximation of a very general class of systems, where a particle has an interaction with a large system, and the interaction has been approximated to linear order. The Hamiltonian is given by

$$H = \frac{1}{2}v^2 + U(x) + \sum_j \frac{p_j^2}{2m_j} + \frac{k_j}{2}(q_j - x)^2 \tag{2.15}$$

where (x, v) are the position and momentum of the distinguished particle, and (q_i, p_i) are the positions and momenta of the harmonic oscillators that form the heat bath, with frequencies defined by $\omega_i^2 = \frac{k_i}{m_i}$. This leads to the equations of motion:

$$\begin{aligned}
\dot{x} &= v \\
\dot{v} &= -U'(x) + \sum_j k_j(q_j - x) \\
\dot{q_i} &= \frac{p_i}{m_i} \\
\dot{p_i} &= -k_i(q_i - x)
\end{aligned}$$

and we see that the Liouvillian is given by

$$L = v\frac{\partial}{\partial x} - (U'(x) - \sum_j k_j(q_j - x))\frac{\partial}{\partial v} + \sum_j [\frac{p_j}{m_j}\frac{\partial}{\partial q_j} - (k_j(q_j - x))\frac{\partial}{\partial p_j}]$$

In this case we are interested in the movement of the distinguished particle, and we consider the heat bath to consist of unresolved variables, which start out in an equilibrium distribution. From this it is clear which projection we should take: we want to project onto the functions of x and v, and hence we use the conditional expectation value with respect to x and v, using the canonical distribution as measure, so we are in the situation of 2.2, and the projection is given by

$$\mathcal{P}f = \frac{\int f e^{-H} \prod dq_j dp_j}{\int e^{-H} \prod dq_j dp_j}.$$

It is easy to check that $\mathcal{P}x = x$, $\mathcal{P}U'(x) = U'(x)$, $\mathcal{P}v = v$, $\mathcal{P}q_i = x$ and $\mathcal{P}p_i = 0$. Thus we can write down the equations for the orthogonal dynamics (2.3)

$$\begin{aligned}
\frac{\partial}{\partial t}S_\perp^t x &= 0 \\
\frac{\partial}{\partial t}S_\perp^t v &= \sum k_j(S_\perp^t q_j - S_\perp^t x) \\
\frac{\partial}{\partial t}S_\perp^t q_i &= \frac{S_\perp^t p_i}{m_i} \\
\frac{\partial}{\partial t}S_\perp^t p_i &= -k_i(S_\perp^t q_i - S_\perp^t x)
\end{aligned}$$

in addition to the initial condition $S^0 f = f$ for all variables. In this particular case we can actually solve these equations (which is very fortunate, in general this is almost never the case) to find

$$\begin{aligned} S_\perp^t x &= x \\ S_\perp^t q_i &= (q_i - x)\cos(\omega_i t) + \frac{p_i}{\omega_i m_i}\sin(\omega_i t) + x \end{aligned}$$

and the solutions for v and p_i can also be found easily from this (but we do not need them explicitly). We want to get the Mori-Zwanzig equations for x and v. For x the orthogonal dynamics are trivial (since the equation was already completely formulated in terms of functions of x and v), and hence the Mori-Zwanzig equation is just the original equation for x, $\frac{\partial}{\partial t} S^t x = v$. For v we first find the orthogonal fluctuation function $F(t) = S_\perp^t (1-\mathcal{P})Lv$ (or $e^{t(1-\mathcal{P})L}(1-\mathcal{P})Lv$ in semigroup notation). From the above explicit solution we see that

$$\begin{aligned} F(t) &= S_\perp^t (1-\mathcal{P})Lv \\ &= S_\perp^t (1-\mathcal{P})(U'(x) - \sum_j k_j(q_j - x)) = S_\perp^t (\sum_j k_j(q_j - x)) \\ &= \sum_j k_j[(q_j - x)\cos(\omega_j t) + \frac{p_j}{\omega_j m_j}\sin(\omega_j t)] \end{aligned}$$

Next we may use this to compute the memory kernel, which is given by

$$\begin{aligned} K(v(t-s), s) &= S^{t-s}\mathcal{P}LS_\perp^s (1-\mathcal{P})Lv = S^{t-s}\mathcal{P}LF(s) \\ &= S^{t-s}\mathcal{P}L\sum_j k_j[(q_j - x)\cos(\omega_j s) + \frac{p_j}{\omega_j m_j}\sin(\omega_j s)] \\ &= S^{t-s}\mathcal{P}(\sum_j k_j[(\frac{p_j}{m_j} - v)\cos(\omega_j s) - \frac{k_j}{\omega_j m_j}(q_j - x)\sin(\omega_j s)]) \\ &= -S^{t-s}\sum k_j \cos(\omega_j s)v \\ &= -\sum k_j \cos(\omega_j s)v(t-s) \end{aligned}$$

where in the second line the definition of L is used, in the third line we just apply the projection $\mathcal{P}$, using that $\mathcal{P}v = v$, $\mathcal{P}p_j = 0$ and $\mathcal{P}(q_j - x) = 0$ and then rearrange terms and apply S^{t-s}. From all this we conclude the following equations for x and v:

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -U'(x) - \int_0^t \tilde{K}(s)v(t-s)ds + F(t) \end{aligned}$$

or, rewriting to one equation:

$$\ddot{x} = -U'(x) - \int_0^t \tilde{K}(s)\dot{x}(t-s)ds + F(t) \tag{2.16}$$

with $\tilde{K}(s) = \sum k_j \cos(\omega_j s)$ and $F(t) = \sum_j k_j[(q_j - x)\cos(\omega_j t) + \frac{p_j}{\omega_j m_i}\sin(\omega_j t)]$ as above.

At this point, we should think about the meaning of equation (2.16). As suggested in the introduction, we may hope that the term $F(t)$ represents a noise if the time scales on which the (unresolved) orthogonal dynamics occur are much shorter than the time scales of the resolved variables. We are interested in the situation where we do not know the initial conditions of all the variables p_j and q_j, so we draw them from the canonical distribution (given the initial values of x and v), with inverse temperature β which means that the initial value of p_j is distributed as $\mathcal{N}(0, \frac{1}{\sqrt{\beta m_j}})$ (Gaussian with mean zero and standard deviation $\frac{1}{\sqrt{\beta m_j}}$), while q_j has a Gaussian distribution $\mathcal{N}(x, \frac{1}{\sqrt{\beta k_j}})$, as we easily read off from the Hamiltonian. We can verify the fluctuation-dissipation theorem explicitly here and we find

$$\beta \mathbf{E}(F(t)F(s)) = K(s-t) \tag{2.17}$$

which is the same expression one finds from (2.14).

A limit theorem for the heat bath model

The original idea of the Mori-Zwanzig formalism in statistical mechanics, is that the Mori-Zwanzig equation may be replaced by a generalized Langevin equation. That is, the orthogonal dynamics may be replaced by a stochastic process. In most situations it is very hard to show that the orthogonal dynamics in the MZ formalism indeed lead to a process that is stochastic. In the particular case of the heat bath model however, limit theorems for an infinite number of heat bath modes, the *thermodynamic limit*, have been proven. We will discuss such a theorem, and closely follow the exposition of (Kupferman et al., 2002). The situation is as above where the initial conditions of the heat bath particles are drawn from the canonical distribution (that is, Gaussian). We start with a heat bath consisting of N modes, and then take a limit $N \to \infty$.

To understand the limit as a stochastic equation we need to reformulate the problem (2.16) in integral notation:

$$\dot{x^N} = \dot{x}^N(0) + \int_0^t U'(X^N(s))ds - \int_0^t \kappa^N(t-s)\dot{x}^N(s)ds + \phi^N(t)$$

where

$$\kappa^N(t) = \int_0^t K^N(s)ds$$
$$\phi^N(t) = \int_0^t F^N(s)ds.$$

The subscript N is used to keep track of the number of heat bath modes. From this definition and using (2.17) we find that

$$\mathbf{E}\phi^N(t)\phi^N(s) = \beta^{-1}(\rho^N(t) + \rho^N(s) - \rho^N(t-s))$$
$$\rho^N(t) = \int_0^t \kappa^N(s)ds.$$

Now the idea is the following: we select the frequencies ω_j such that they have a broad and dense spectrum, and then choose an appropriate function f to let

$$k_j = f(\omega_j)\Delta\omega$$

with $\Delta\omega^{-1}$ the mean density of the frequencies. Then K^N will approximate an integral (the Fourier cosine transform of f) in the limit $N \to \infty$

$$K^N(t) = \sum_{i=1}^{N} f(\omega_i)\cos(\omega_i t)\Delta\omega \approx \int_0^\infty f(\omega)\cos(\omega t)d\omega.$$

To get a good spectrum one could take for instance evenly spaced $\omega_j = j^a$ with $a \in (0,1)$(Hald and Kupferman, 2002), but we will consider the case where the ω_j are uniform random variables on $[0, N^a]$ so $\omega_j = N^a\xi_i$ (Kupferman et al., 2002), where the ξ_i are IID $\mathcal{U}(0,1)$ random variables. Denote

$$K(t) = \int_0^\infty f(\omega)\cos(\omega t)d\omega$$
$$\kappa(t) = \int_0^t K(s)ds$$
$$\rho(t) = \int_0^t \kappa(s)ds.$$

and let ϕ be the Gaussian process with autocorrelation

$$\mathbf{E}\phi(t)\phi(s) = \beta^{-1}(\rho(t) + \rho(s) - \rho(t-s)).$$

In the remainder of this section $\mathbf{E}$ denotes the expectation with respect to the variables ξ. We must assume a convergence condition on the function f, so we assume that f is bounded (by c_1) and also bounded by $c_2\omega^{-\lambda}$, with $\lambda > 1$ (so $f \in L^2[0,\infty)$). With our choice of frequencies K^N becomes a Monte Carlo approximation of the integral, which is expressed by the following lemma:

Lemma 2.4.1. *Almost surely with respect to the choice of frequencies κ^N converges to κ in $L^2[0,T]$ (and hence also converges weakly) for any time $T > 0$.*

This lemma can be used to prove convergence of the orthogonal dynamics to a stochastic process:

Lemma 2.4.2. *Almost surely with respect to the choice of frequencies the processes ϕ^N converge weakly to ϕ in $C[0,T]$ for any time $T > 0$.*

The resulting kernel and stochastic process of course depend on the choice of function f. For example,

$$f(\omega) = \frac{2\alpha}{\pi(\alpha^2 + \omega^2)}$$

leads to $K(t) = e^{-\alpha t}$ and ϕ is in that case an OU process (Kupferman et al., 2002). A slight modification (introducing a cut-off near o for the spectrum) allows to take

$$f(\omega) = \frac{2}{\pi}\Gamma(1-\gamma)\sin(\frac{\gamma\pi}{2})\frac{1}{\omega^{1-\gamma}}$$

with Γ the Gamma function and $\gamma \in (0,1)$, which leads to $K(t) = \frac{1}{t^\gamma}$ and ϕ fractional Brownian motion with Hurst parameter γ (Kupferman, 2004). From the two lemmas the following theorem follows (Kupferman et al., 2002):

Theorem 2.4.3. *Let x^N be the solution to* (2.16)*. Assume that V' is globally Lipschitz. Then almost surely with respect to the frequencies x^N converges weakly in $C^1[0,T]$ for any time $T > 0$ to the solution x of*

$$\dot{x} = \dot{x}(0) + \int_0^t U'(x(s))ds - \int_0^t \kappa(t-s)\dot{x}(s)ds + \phi(t) \tag{2.18}$$

Proof. First of all, the assumption that V' is globally Lipschitz ensures existence and uniqueness of a $C^1[0,T]$ solution x given $(\kappa, \phi) \in L^1[0,T] \times C[0,T]$. Thus we get a map from $L^1[0,T] \times C[0,T]$ to $C^1[0,T]$, sending (κ, ϕ) to x (as defined by (2.18)). This mapping is continuous, as the defining equation is an equation of Volterra type (see (Kupferman et al., 2002) for details on this). From the lemmas 2.4.1 and 2.4.2 we know that (κ^N, ϕ^N) converges weakly in $L^1[0,T] \times C[0,T]$ to (κ, ϕ) (since κ^N converges in $L^2[0,T]$ it also converges in $L^1[0,T]$, and weak convergence is preserved under products), and because continuous mappings preserve weak convergence, this implies that x^N converges weakly to x in $C^1[0,T]$. □

The main work is in proving the lemmas 2.4.1 and 2.4.2.

Proof of lemma 2.4.1. Firstly, we know that

$$\mathbf{E}\kappa^N(t) = \int_0^{N^a} f(\omega)\frac{\sin(\omega t)}{\omega}d\omega$$

As N goes to infinity, $\mathbf{E}\kappa^N$ converges in $L^2[0,T]$ to κ. So it suffices to show that almost surely κ^N converges to $\mathbf{E}\kappa^N$. Using Borel-Cantelli we know that it suffices to show that $\mathbf{P}(||\kappa^N - \mathbf{E}\kappa^N||^2_{L^2[0,T]} > \epsilon)$ is summable (with respect to N) to get almost sure convergence. Using the Chebyshev inequality it is thus sufficient to show that $\sigma^N = \mathbf{E}||\kappa^N - \mathbf{E}\kappa^N||^{2p}_{L^2[0,T]}$ is summable for some positive integer p.

Write

$$\sigma^N = \int_0^T \cdots \int_0^T \mathbf{E}(\prod_{i=1}^{p} |\kappa^N(t_i) - \mathbf{E}\kappa^N(t_i)|^2) dt_1 \dots dt_p$$

(exchanging integration order using Tonelli) and let $g(\omega, t) = \omega^{-1} f(\omega) \sin(\omega t)$ (so $\kappa^N(t) = \sum_{j=1}^{N} g(\omega_j, t)\Delta\omega$), and let $\mu^N(t) = \mathbf{E}g(\omega, t)$. Then $\kappa^N(t) - \mathbf{E}\kappa^N(t) = \Delta\omega \sum_{j=1}^{N} h(\omega_j, t) - \mu(t)$ and we may rearrange

$$\sigma^N = \Delta\omega^{2p} \int_0^T \cdots \int_0^T \sum_J V_J(t_1, \dots, t_p) dt_1 \dots dt_p$$

with $J = (j_1, \dots, j_{2p})_{j_i=1}^{N}$ a multi-index, and

$$V_J(t_1, \dots, t_p) = \mathbf{E}(\prod_{i=1}^{p} |h(\omega_{j_{2i-1}}, t_i) - \mu(t_i)||h(\omega_{j_{2i}}, t_i) - \mu(t_i)|).$$

To estimate the sum over these terms, we regroup them by the number of distinct indices k in J. If an index occurs, it must occur at least twice (since the moments are centered around the mean). This implies that $k \leq p$. The number of terms with k distinct indices can be bounded by CN^k with k a constant C that counts the number of partitions of p and does not depend on N or k. Finally we notice that we may estimate any of the V_J with k distinct indices as

$$V_J(t_1, \dots, t_p) = \prod_{r=1}^{k} N^{-a} \int_0^{N^a} (g^*(\omega) + \mu^{N*})^{m_r}$$

where m_r is the number of times the index corresponding to r occurs in V_J, and $g^*(\omega)$ is an upper bound for $g(\omega, t)$, which we may choose to be integrable by our assumptions on f, and $\mu^{N*} = \mathbf{E}g^*(\omega)$. Notice that this means that μ^{N*} scales N^{-a}. These upper bounds exist and are integrable by our assumptions on f. Finally we estimate the last expression by

$$\int_0^{N^a} (g^*(\omega) + \mu^{N*})^{m_r} = \sum_{l_r=0}^{m_r} \binom{m_r}{l_r} (\mu^{N*})^{m_r - l_r} \int_0^{N^a} (g^*(\omega))^{l_r} \leq D$$

with D a constant that does not depend on N (μ^{N*} scales as N^{-a}, so if $l_r = 0$ the factor N^a is compensated by this, and if $l_r > 0$ the integral is bounded). Thus we conclude

$$V_J(t_1, \dots, t_p) \leq CN^{-ka}$$

so

$$\sigma^N \leq \Delta\omega^{2p} T^p \sum_{k=1}^{p} CN^k * N^{-ka} \leq CN^{2p(a-1)} N^{p(1-a)} = CN^{p(a-1)}$$

and choosing $p > (1-a)^{-1}$ ensures that σ^N is summable, concluding the proof. □

Proof of lemma 2.4.2. The proof has two ingredients:

- Firstly we use the previous lemma 2.4.1 to show that the autocorrelation function ρ^N converges to ρ pointwise. Indeed:

$$|\rho^N(t) - \rho(t)| = |\int_0^t (\kappa^N(s) - \kappa(s))ds| \leq ||\kappa^N - \kappa||_{L^1[0,T]}$$

  which goes to zero as κ^N converges to κ in the $L^2[0,T]$ norm and hence also in the $L^1[0,T]$ norm. Since ϕ^N and ϕ are Gaussian processes, this implies that for all finite collections of times $\{t_1, \ldots, t_k\}$ the joint probability density functions of $(\phi^N(t_1), \ldots, \phi^N(t_k))$ converge pointwise to the density functions of $(\phi(t_1), \ldots, \phi(t_k))$.

- Using Cauchy-Schwarz

$$|\rho^N(t) - \rho^N(s)| = |\int_s^t \kappa^N(s)ds| \leq |t-s|^{\frac{1}{2}} ||\kappa^N||_{L^2[0,T]}$$

  so $\rho^N(t)$ (and by a similar argument $\rho(t)$) are Hölder continuous with exponent $\frac{1}{2}$, and this is uniformly in N (since $||\kappa^N||_{L^2[0,T]}$ is bounded). We conclude that also the autocorrelation functions of ϕ^N and ϕ are Hölder continuous, and the Kolmogorov continuity theorem tells us that this means that ϕ^N and ϕ are almost surely continuous.

We combine these ingredients by using a theorem (Gikhman and Skorokhod, 1969, theorem 1 in section 9.2) that derives from Prokhorov's theorem on weak convergence, which states that if ϕ^N is a collection of real valued continuous stochastic processes, for which the finite dimensional distributions converge to that of an almost surely continuous process ϕ, and which satisfy the following tightness condition

$$\lim_{u \to 0} \sup_N \mathbf{P}(\sup_{\substack{t,s \in [0,T] \\ |t-s| \leq u}} |\phi^N(t) - \phi^N(s)| > \epsilon) = 0 \tag{2.19}$$

for all $\epsilon > 0$, then ϕ^N converges weakly to ϕ. Our first ingredient shows the convergence of the finite dimensional distributions, and the second ingredient shows that ϕ is almost surely continuous. The tightness condition also follows from the uniform Hölder continuity: since $\rho^N(0) = 0$ we may write $\rho^N(u) \leq C|u|^{\frac{1}{2}}$ and since $|\phi^N(t) - \phi^N(s)|$ has a normal distribution with variance $\beta^1 \rho^N(t-s)$ we may estimate (2.19) by

$$\lim_{u \to 0} \mathbf{P}(|X_u| > \epsilon)$$

where X_u is a $\mathcal{N}(0, \beta^{-1} C|u|^{\frac{1}{2}})$ normal variable, and this limit is indeed zero (follows from Chebyshev inequality). □

We thus see that for the heat bath model, the project of reducing a large deterministic system to a small stochastic system can be fully performed. In this section we have

proven weak convergence, in some cases it is actually also possible to prove strong pathwise convergence, that is

$$\sup_{t\leq T}\lim_{N\to\infty}\mathbf{E}(|x^N(t)-x(t)|^2+|\dot{x}^N(t)-\dot{x}(t)|^2)=0$$

see (Hald and Kupferman, 2002) for the case $\omega_j = j^{-a}$ and (Ariel and Vanden-Eijnden, 2008) for the case with randomly distributed ω (but constant k_j).

2.5 Approximations for the memory kernel

Since the Mori-Zwanzig formalism really consists in a *rewriting* of the original equations, the result will still be just as complicated as the problem we started with. The two main issues are that we still have to solve the orthogonal dynamics to find the 'noise' term, and that it is very hard in general to come up with an explicit expression for the memory kernel without solving the orthogonal dynamics. However, we get a good opportunity to start making approximations. With respect to the first problem we can try to approximate the orthogonal term by a noise (which may indeed be reasonable if the variable we projected on are slow). However, this does not immediately help us in finding the memory kernel. There are two situations for which method to approximate the kernel are known which are very general. The first is when the memory is very short, and the second one if the memory is very long and the kernel varies slowly. These approximations are given in (Chorin et al., 2002).

Short-memory approximation

Consider the expression for $K(f,s)$

$$\int_0^t S^{t-s}\mathcal{P}LS_\perp^s(1-\mathcal{P})Lfds=\int_0^t S^{t-s}\mathcal{P}L(1-\mathcal{P})S_\perp^s(1-\mathcal{P})Lfds$$

where we used that $S_\perp^t$ maps the null space of $\mathcal{P}$ onto itself. We write

$$\mathcal{P}L(1-\mathcal{P})S_\perp^s(1-\mathcal{P})Lf=\mathcal{P}L(1-\mathcal{P})S^s(1-\mathcal{P})Lf+\mathcal{P}L(1-\mathcal{P})(S_\perp^s-S^s)(1-\mathcal{P})Lf.$$

By differentiation one sees that $(1-\mathcal{P})(S_\perp^s-S^s)=\mathcal{O}(s^2)$ as $S_\perp^0-S^0=0$ and and $\frac{\partial}{\partial s}(1-\mathcal{P})(S_\perp^s-S^s)|_{s=0}=(1-\mathcal{P})((1-\mathcal{P})L-L)=0$ and thus we conclude

$$K(f,s)=\int_0^t S^{t-s}\mathcal{P}L(1-\mathcal{P})S^s(1-\mathcal{P})Lfds+\mathcal{O}(t^3). \tag{2.20}$$

In principle this is now just a short *time* approximation, but it may also be valid over longer times if the term $\mathcal{P}LS_\perp^s(1-\mathcal{P})Lf$ decays fast, in that case we have a short-memory approximation. We see that in (2.20) we do not have to solve the orthogonal dynamics equation (as $S_\perp^s$ does not appear).

Long-memory approximation

Another situation where we can actually make a reasonable approximation for the memory kernel which does not demand the solution of the orthogonal dynamics is in the case where the memory decays only slowly. In this case we write

$$\begin{aligned}\int_0^t S^{t-s}\mathcal{P}LS_\perp^s(1-\mathcal{P})Lfds &= \int_0^t S^{t-s}LS_\perp^s(1-\mathcal{P})Lfds - \int_0^t S^{t-s}(1-\mathcal{P})LS_\perp^s(1-\mathcal{P})Lfds \\ &= \int_0^t LS^{t-s}S_\perp^s(1-\mathcal{P})Lfds - \int_0^t S^{t-s}S_\perp^s(1-\mathcal{P})L(1-\mathcal{P})Lfds\end{aligned}$$

using that L commutes with S^t and $(1-\mathcal{P})L$ commutes with $S_\perp^t$. Now we approximate by letting $S_\perp^s \sim S^s$ (so $e^{(1-\mathcal{P})Ls} \sim e^{Ls}$) to get

$$\int_0^t LS^t(1-\mathcal{P})Lfds - \int_0^t S^t(1-\mathcal{P})L(1-\mathcal{P})Lfds = tS^t\mathcal{P}L(1-\mathcal{P})Lf.$$

Alternatively, one could also expand the expression for the integrand in the memory kernel around $s=0$ to find

$$\begin{aligned}\int_0^t S^{t-s}\mathcal{P}LS_\perp^s(1-\mathcal{P})Lfds &= \int_0^t S^t\mathcal{P}L(1-\mathcal{P})Lf + \mathcal{O}(s)ds \\ &= tS^t\mathcal{P}L(1-\mathcal{P})Lf + \mathcal{O}(t^2).\end{aligned}$$

When is the error small in this case? This happens if the integrand stays closely to its initial value at $s=0$, so the memory kernel is 'slowly decaying'. This is the opposite situation of the previous approximation, where we assumed that the memory kernel decayed very fast. This approximation is also known as the *t-model*. It is easy to extend to higher orders in t. Let us give an example from (Chorin et al., 2002) and (Chorin and Hald, 2014), where we consider a nonlinear Hamiltonian system with

$$H(q_1,q_2,p_1,p_2) = \frac{1}{2}(p_1^2+p_2^2+q_1^2+q_2^2+q_1^2q_2^2) \tag{2.21}$$

and we use the canonical projection operator to project onto (q_1,p_1), with $\beta=1$. The relevant projections are easy to compute (they are quadratic Gaussian integrals), and after an easy computation we find for the Markovian part that

$$\begin{aligned}\mathcal{P}Lq_1 &= p_1 \\ \mathcal{P}Lp_1 &= -q_1(1+\frac{1}{1+q_1^2})\end{aligned}$$

and for the t-model approximation of the memory

$$\begin{aligned}\mathcal{P}L(1-\mathcal{P})Lq_1 &= 0 \\ \mathcal{P}L(1-\mathcal{P})Lp_1 &= -\frac{2q_1^2p_1}{(1+q_1^2)^2}.\end{aligned}$$

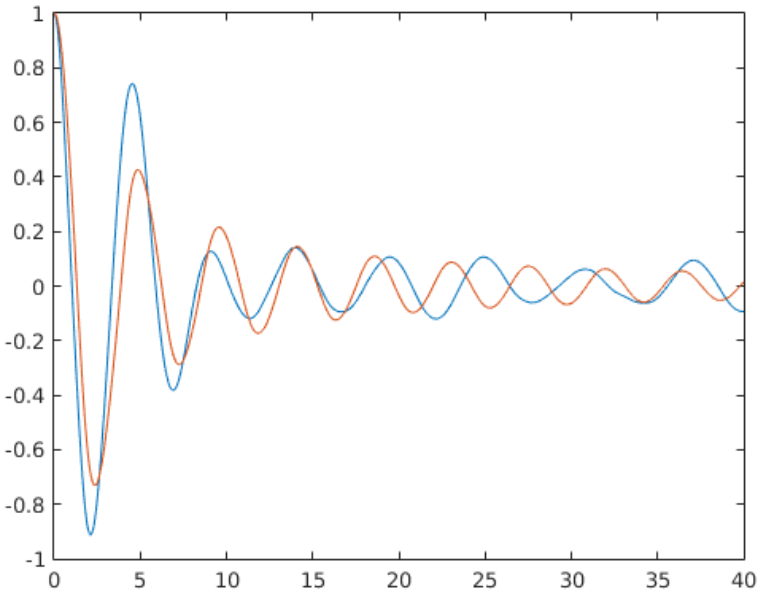


Figure 2.1: The (canonical) ensemble average over 5000 oscillators with fixed q_1, p_1 initial conditions (blue) compared to the t-model prediction (red).

This leads to the optimal prediction equation with memory (2.9) for the evolution of the expected value of (q_1, p_1) given initial conditions (so the evolution of $(\bar{q}_1, \bar{p}_1) = \mathcal{P}(q_1(t), p_2(t)) = \mathbf{E}(q_1(t), p_1(t)|p_1(0), q_1(0))$). We make another (crude) approximation, by commuting conditional expectations with functions of $q_1(t)$: $\mathcal{P}(f(q_1(t))) = f(\bar{q}_1(t))$ to obtain

$$\begin{aligned} \frac{d}{dt}\bar{q}_1 &= \bar{p}_1 \\ \frac{d}{dt}\bar{p}_1 &= -\bar{q}_1(1 + \frac{1}{(1+\bar{q}_1^2)}) - t\frac{2\bar{q}_1^2\bar{p}_1}{(1+\bar{q}_1^2)^2} \end{aligned} \tag{2.22}$$

Comparing the solution to this equation with the average over an ensemble average leads to a reasonably good result as can be seen in figure 2.1. In case just using the optimal prediction without memory leads to an oscillating solution, and the memory term has a dissipative effect, which models the 'leaking' of information to the unresolved degrees of freedom. It is not hard to show that for general mechanical Hamiltonians the memory term in the t-model is a dissipative term (meaning that it causes the value of the reduced Hamiltonian to decrease) (Chorin et al., 2002).

2.6 Markovian approximation of the generalized Langevin equation

In this section we will discuss a way to approximate the memory kernel in the stituation where we have a *linear* memory term. We will use the following theorem, closely following the proof in (Pavliotis, 2014):

Theorem 2.6.1. *Consider the generalized Langevin equation*

$$\dot{x} = f(x) - \int_0^t K(t-s)x(s)ds + F(t)$$

where $K(t-s) = \lambda \cdot e^{-A(t-s)}\lambda$ for $\lambda \in R^n$ and A a positive definite $n \times n$ matrix, is the autocorrelation for the process F. Then this equation is equivalent to the following SDE

$$\begin{aligned} \dot{x} &= f(x) + \lambda \cdot y \\ \dot{y} &= -\lambda x - Ay + \Sigma \dot{W} \end{aligned}$$

where $\Sigma\Sigma^\top = A + A^\top$ and the initial conditions for y are $\mathcal{N}(0, I)$ distributed.

Notice that for GLE's obtained from the Mori-Zwanzig procedure, with the linear projection we indeed know by the fluctuation dissipation theorem that K is the autocorrelation for F.

Proof. Let us start from the equation

$$\begin{aligned} \dot{x} &= f(x) + \lambda \cdot y \\ \dot{y} &= -\lambda x - Ay + \Sigma \dot{W} \end{aligned} \tag{2.23}$$

We can solve y for given x to get

$$y(t) = e^{-At}y(0) - \int_0^t e^{-A(t-s)}\lambda x(s)ds + \int_0^t e^{-As}\Sigma^\top \Sigma dW(s).$$

Inserting this in the equation for x we find

$$\dot{x} = f(x) - \int_0^t K(t-s)x(s)ds + F(t) \tag{2.24}$$

where $K(t) = \lambda \cdot e^{-At}\lambda$ and

$$F(t) = \lambda \cdot (e^{-At}y(0) + \int_0^t e^{-As}\Sigma^\top \Sigma dW(s)).$$

From this we may deduce that, if we assume $y(0)$ is $\mathcal{N}(0, I)$ distributed, $F(t)$ is a stationary Gaussian stochastic process with mean zero and covariance

$$\mathbf{E}(F(t)F(t')) = \lambda \cdot e^{-At}(I + \int_0^{\min(t,t')} e^{As}\Sigma^\top \Sigma e^{A^\top s}ds)e^{-A^\top t'}\lambda$$

Now the result follows from the assumption that $\Sigma\Sigma^\top = A + A^\top$ and hence

$$\mathbf{E}(F(t)F(t')) = \lambda \cdot e^{-At}(I + \int_0^{\min(t,t')} e^{As}(A + A^\top)e^{A^\top s}ds)e^{-A^\top t'}\lambda.$$

Let us assume $t' \leq t$ (the case $t \leq t'$ is similar), then we have

$$I + \int_0^{\min(t,t')} e^{As}(A + A^\top)e^{A^\top s}ds = e^{At'}e^{A^\top t'}$$

which follows by differentiation. Thus we conclude that

$$\mathbf{E}(F(t)F(t')) = \lambda \cdot e^{-At}e^{At'}\lambda = \gamma(t - t')$$

as desired. □

This is a very useful theorem. If we can approximate the memory kernel of an arbitrary GLE satisfying the fluctuation-dissipation theorem by a function of the form $\lambda \cdot e^{-A(t-s)}\lambda$, we can approximate the full system by a finite dimensional SDE.

Laplace transforms

One way to look at generalized Langevin equations with a linear memory is by considering Laplace transforms. The Laplace transform of a function f is the complex function $\hat{f}$ defined by

$$\hat{f}(z) = \int_0^\infty f(t)e^{-zt}dt$$

and for a memory kernel of the form given in theorem 2.6.1, we get

$$\hat{K}(z) = \lambda \cdot (A + zI)^{-1}\lambda \tag{2.25}$$

Starting with the full system, we can approximate the Laplace transform by some rational function, and then try to find A and λ such that it is of the form (2.25).

In the particular case of the particle in a heat bath, there is a close relation between the autocorrelation functions and the memory kernel, that can be best expressed in terms of Laplace transforms. Consider the system

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= \int K(t-s)v(s)ds + F(t) \end{aligned} \tag{2.26}$$

(a free particle in a heat bath) and let $C_{vv} = \mathbf{E}(v(t)v(0))$ be the velocity autocorrelation function. If the system is ergodic this autocorrelation may be obtained by a time

average rather than an ensemble average (so it can be estimated from a time series). Multiplying equation (2.26) by $v(0)$ and taking the ensemble average we find

$$\dot{C}_{vv}(t) = \int K(t-s)C_{vv}(s)ds$$

using that $\mathbf{E}(v(0)F(t)) = 0$. The Laplace transform of this equation is

$$z\hat{C}_{vv}(z) - C_{vv}(0) = \hat{K}\hat{C}_{vv}$$

so from the Laplace transform of the velocity autocorrelation (which can be estimated from available data) one obtains the Laplace transform of the memory kernel. This makes it interesting to investigate simplifications of $\hat{K}$ that allow a finite dimensional representation as a Markovian system as in theorem 2.6.1.

The easiest example is when the matrix A is diagonal:
Example 2.6.2. A diagonal matrix A with $A_{ii} = \alpha_i$ corresponds to a memory kernel

$$K(t) = \sum_i \lambda_i^2 e^{-\alpha_i t}.$$

This corresponds to a Laplace transform

$$\hat{K}(z) = \sum_i \frac{\lambda_i^2}{z+\alpha_i}.$$

Continued fraction expansion

Another approximation that was suggested by Mori is by making a continued fraction approximation of the Laplace transform of the memory kernel

$$\hat{K}(z) = \cfrac{\Delta_1^2}{z+\omega_1 + \cfrac{\Delta_2^2}{z+\omega_2 + \cfrac{\Delta_3^2}{z+\omega_3+\cdots}}}$$

and truncate this expansion at some finite order. Truncating is only appropriate if the coefficients Δ_i do not grow very fast (if they do all orders are needed for a good approximation) (Muller and Viswanath, 1994). This approximation can be physically motivated by the following procedure due to (Mori, 1965b) where we start with the Liouville equation for some variable x, and project onto the linear functions of x, leading to

$$\dot{x} = Lx = \omega_0 x + \int K(t-s)x(s)ds + F_1(t) \tag{2.27}$$

We can apply the Mori-Zwanzig formalism to $F_1(t)$, projecting onto F_1, and keep repeating this to get

$$\dot{F}_i = \omega_i F_i + \int_0^t K_1(t-s)F_i(s)ds + F_{i+1} \tag{2.28}$$

We may truncate by assuming that for some n the term F_n can be treated as a white noise term. To solve we introduce the Laplace transform. Using the properties of the Laplace transform with respect to convolution and differentiation we see that

$$z\hat{F}_i(z) + F_i(0) = \omega_i \hat{F}_i(z) + \hat{K}_i(z)\hat{F}_i(z) + \hat{F}_{i+1}(z)$$

which is solved by

$$\hat{F}_i(z) = \frac{F_i(0) + \hat{F}_{i+1}(z)}{z - \omega_i - \hat{K}_i(z)}.$$

Next we use the fluctuation dissipation relation (2.13) which tells us that the Laplace transform of the kernels K and K_i is given by

$$\hat{K}(z) = \frac{(\hat{F_1}(z)F(0)}{(x(0)^2)}$$
$$\hat{K_i}(z) = \frac{(\hat{F_{i+1}}(z)F_{i+1}(0)}{(F_i(0)^2)}$$

Using that $(F_{i+1}(t)F_i(0)) = 0$ it follows that

$$\hat{K}(z) = \cfrac{\Delta_1^2}{z + \omega_1 + \cfrac{\Delta_2^2}{z + \omega_2 + \cfrac{\Delta_3^2}{z + \omega_3 + \cdots}}}$$

where

$$\Delta_1^2 = \frac{(F_1(0)^2)}{(x(0)^2)}$$
$$\Delta_i^2 = \frac{(F_i(0)^2)}{(F_{i-1}(0)^2)}.$$

If we truncate the continued fraction the system can be realized in the form of theorem 2.6.1 using a matrix

$$A = \begin{pmatrix} \omega_1 & -\Delta_2 & & & \\ \Delta_2 & \omega_2 & -\Delta_3 & & \\ & \Delta_3 & \omega_3 & -\Delta_4 & \\ & & \ddots & \ddots & \ddots \\ & & & \Delta_n & \omega_n \end{pmatrix}$$

and $\lambda = (\Delta_1, 0, \ldots, 0)^\top$.

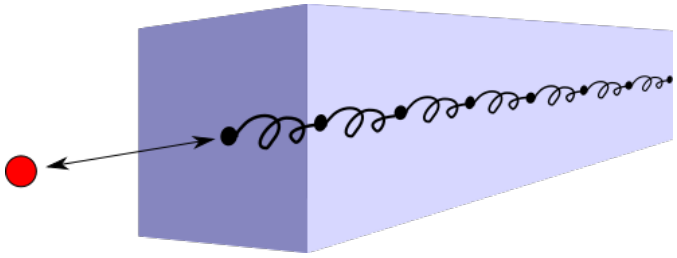


Figure 2.2: A model for the interaction of a gas atom (red) and a solid (the chain of linear springs).

Example: a gas particle and a solid

In this section we consider an example, due to (Zwanzig, 1960) and (Adelman and Doll, 1974), which can be treated analytically, and where the approximation by continued fractions is completely explicit. The Hamiltonian is given by

$$H(q,p) = \frac{p_0^2}{2M} + V(q_0 - q_1) + \sum_{i=1}^{\infty} \frac{p_i^2}{2m} + \kappa(q_i - q_{i+1})^2 \tag{2.29}$$

where q_0, p_0 denote position and momentum of a gas particle, and q_i, p_i form a one-dimensional harmonic lattice (the q_i represent the deviations from from the lattice positions). The idea is that the system is a highly simplified model for the collision of a gas particle with a solid, where the solid is represented by a harmonic lattice (see figure 2.2) and the gas particle only interacts with the outermost atom in the lattice through some prescribed potential V. This model is also very closely related to a model developed by Rubin to understand the behaviour of a defect in a harmonic lattice (Rubin, 1963). The equations of motion are

$$\begin{aligned}
\dot{q}_0 &= \frac{p_0}{M} \\
\dot{q}_i &= \frac{p_i}{m} \\
\dot{p}_0 &= -V'(q_0 - q_1) \\
\dot{p}_1 &= V'(q_0 - q_1) - \kappa(p_1 - p_2) \\
\dot{p}_i &= \kappa(q_{i-1} - 2q_i + q_{i+1})
\end{aligned} \tag{2.30}$$

To be able to use the projection formalism we make a change of variables: first we rescale time to $\tau = 2\sqrt{\frac{\kappa}{m}}t$ and then we map $(p_i, q_i)_{i=0}^{\infty}$ to $(u_i)_{i=0}^{\infty}$ by letting

$$\begin{aligned}
u_{2i} &= 2\frac{dq_j}{d\tau} \\
u_{2i+1} &= q_i - q_{i+1}
\end{aligned}$$

The new equations of motion are

$$\begin{aligned}
\dot{u}_0 &= -\frac{m}{2\kappa M} V'(u_1) \\
\dot{u}_1 &= \frac{1}{2}(u_0 - u_2) \\
\dot{u}_2 &= \frac{1}{2\kappa} V'(u_1) - \frac{1}{2} u_3 \\
\dot{u}_i &= \frac{1}{2}(u_{i-1} - u_{i+1}) \text{ for } i > 2.
\end{aligned}$$

In these new variables the Hamiltonian becomes a quadratic function (but notice that the transformation does not preserve the Hamiltonian structure). We project onto u_0, u_1 and u_2 using the canonical ensemble, which is easy because of the quadratic Hamiltonian. Denoting by $f_i(\tau)$ the term $S_\perp^\tau u_i$ we find the following equations for the orthogonal dynamics:

$$\begin{aligned}
\dot{f}_0 &= \dot{f}_1 = 0 \\
\dot{f}_2 &= -\frac{1}{2} f_3 \\
\dot{f}_3 &= -\frac{1}{2} f_4 \\
\dot{f}_i &= -\frac{1}{2}(f_{i-1} - f_{i+1}) \text{ for } i > 3.
\end{aligned}$$

Notice that this is precisely the form we investigated in the previous section with ω_i all zero and $\Delta_i = \frac{1}{2}$. Thus we know the Laplace transform of the memory kernel to be the continued fraction

$$\hat{K}(z) = \cfrac{\frac{1}{4}}{z + \cfrac{\frac{1}{4}}{z + \cfrac{\frac{1}{4}}{z + \cdots}}}$$

and the corresponding reduced equation of motion is

$$\begin{aligned}
\dot{u}_0 &= -\frac{m}{2\kappa M} V'(u_1) \\
\dot{u}_1 &= \frac{1}{2}(u_0 - u_2) \\
\dot{u}_2 &= \frac{1}{2\kappa} V'(u_1) + \int_0^t K(\tau - s) u_2(s) ds - \frac{1}{2} f_3(\tau)
\end{aligned}$$

which we may rewrite in a single equation using the original variables as

$$\ddot{q}_1 - \ddot{q}_0 = (\frac{1}{\kappa} - \frac{m}{\kappa M}) V'(q_0 - q_1) - \int_0^\tau K(\tau - s) p_1(s) ds - \frac{1}{2} f_3(\tau). \tag{2.31}$$

A finite truncation corresponds to a chain with a finite number of lattice particles. In this model we happen to be able to find the inverse of this continued fraction

for an infinite number of particles. This is done by explicitly solving the orthogonal dynamics, using a generating function trick. We introduce

$$F(z,\tau) = \sum_{n=2}^{\infty} f_n(\tau) z^n$$

which, using the equations for the orthogonal dynamics, has time derivative

$$\begin{aligned}\frac{\partial F(z,\tau)}{\partial \tau} &= \sum_{n=4}^{\infty} \frac{1}{2}(f_{n-1}(\tau) - f_{n+1}(\tau)) z^n - \frac{1}{2} f_4(\tau) z^3 - \frac{1}{2} f_3(\tau) z^2 \\ &= \frac{1}{2}(z - z^{-1}) F(z,\tau) + \frac{1}{2}(z - z^3) f_2(\tau).\end{aligned}$$

This can (formally) be solved to find

$$F(z,\tau) = e^{\frac{1}{2}(z-z^{-1})\tau} F(z,0) + \frac{1}{2}\int_0^\tau e^{\frac{1}{2}(z-z^{-1})(\tau-s)}(z - z^3) f_2(s) ds.$$

Next we use the expansion

$$e^{\frac{1}{2}(z-z^{-1})\tau} = \sum_{n=-\infty}^{\infty} J_n(\tau) z^n$$

where the J_n are Bessel functions of the first kind. We can now get f_3 by extracting the coefficient of z^3 to find

$$f_3(\tau) = \sum_{n=2}^{\infty} J_{-n+3}(\tau) f_n(0) + \frac{1}{2}\int_0^\tau (J_2(\tau - s) - J_0(\tau - s)) f_2(s) ds$$

We can get rid of the second term by looking at the coefficient of z (which must equal o by our definition of F). Using the relation $J_{-n} = (-1)^n J_n$ we arrive at

$$\begin{aligned} f_3(\tau) &= -\sum_{n=2}^{\infty} (-1)^n (J_{n-3}(\tau) + J_{n-1}(\tau)) f_n(0) \\ &= \frac{2J_1(\tau)}{\tau} f_3(0) - \sum_{n=4}^{\infty} (-1)^n (J_{n-3}(\tau) + J_{n-1}(\tau)) f_n(0)\end{aligned}$$

where in the last line we used $J_1 = -J_{-1}$ so the $n = 2$ term vanishes, and $J_0(\tau) + J_2(\tau) = \frac{2J_1(\tau)}{\tau}$. We conclude that the memory kernel in (2.31) is given by

$$\begin{aligned}\mathcal{P}LF_2(s) &= -\frac{1}{2}\mathcal{P} f_3(s) \\ &= \frac{2J_1(s)}{s} u_2\end{aligned}$$

and we conclude that inverting the continued fraction should lead to $K = \frac{2J_1(s)}{s}$. The result should physically be interpreted in the following way: the gas atom interacts with the outer molecule of the solid, but the energy that is transferred to the solid then dissipates away through the chain by phonons. The dissipation is described by the memory kernel. Thus (depending on the potential, see (Zwanzig, 1960) which has a slightly different derivation, for details) the atom may get trapped on the surface of the solid (at least if the solid is 'cold', that is, if the chain is initially inert.

2.7 An example: conservative Lotka-Volterra

From the shape of the generalized Langevin equation that results from the Mori-Zwanzig formalism, it is clear that if we are able to approximate the dynamics by an SDE, it will always have additive noise (since the orthogonal 'noisy' term does not depend on the relevant variables by construction). In this section we will nonetheless give a heuristic derivation of an SDE with multiplicative noise, and at the same time illustrate how to use perturbation theory for the orthogonal dynamics. We do so in the context of a model for the evolution of populations of multiple species that interact. The basic model for this type of dynamics is the Lotka-Volterra model. Different interactions between the different species are possible; we will only take into account a predator-prey relation, in which biomass is conserved. In that case, the system is actually Hamiltonian which greatly simplifies the analysis. We will try to track the evolution of the population of a single species, and it will turn out that it satisfies a generalized Verhulst equation, and in the limit with infinitely many species that have appropriate interaction we actually get the stochastic Verhulst equation, the standard equation for logistic growth. The analogy with the previous examples is that we interpret all the other species in the system as a sort of heat bath. In this case the equations for the 'heat bath' are nonlinear and can not be solved explicitly. Near equilibrium we can use a linear approximation, and we will also compute higher order correction terms. The linearized model was first considered in (Zwanzig, 1973b), and (Roerdink and Weyland, 1981) is the inspiration for computing the higher order approximations.

We consider the following Lotka-Volterra equations

$$\dot{N}_i = k_i N_i + \beta_i^{-1} \sum_{j=0}^{n} a_{ij} N_i N_j \tag{2.32}$$

where N_i represents the population of species i and runs from 0 to n. The assumption that the equations are conservative (biologically: total biomass is preserved in the interaction) is expressed by the condition that $a_{ij} = -a_{ji}$ so the matrix A with $A_{ij} = a_{ij}$ is anti-symmetric. Notice that from a modelling perspective, this is a problematic assumption, because this means that the system is structurally unstable, perturbing the parameters a little bit will destroy the conservative nature, and hence the following analysis no longer holds (May, 1973), but see (Goel et al., 1971) for a defence. In any case, if we drop this assumption, the behaviour of the system will be very different, as it will in most cases not be oscillatory. It is known that a very large class of dynamical system can be realized as (a subsystem of) a Lotka-Volterra model. We assume that an equilibrium population Q_i exists, so

$$k_i Q_i + \beta_i^{-1} \sum_{j=0}^{n} a_{ij} Q_i Q_j = 0$$

Moreover, if we assume this equilibrium state to be unique, $n+1$ must be even (a rather disquieting assumption from a biological perspective as there seems to be

no good reason to distinguish between ecosystems with an even or odd number of species), and the system (2.32) is Hamiltonian. This is seen by introducing variables $v_i = \ln(N_i/Q_i)$, so

$$\dot{v}_i = \beta_i^{-1} \sum_{j=0}^{n} a_{ij} Q_j (e^{v_j} - 1)$$

and the function

$$G(v) = \sum_{i=0}^{n} \beta_i Q_i (e^{v_i} - v_i)$$

is conserved. The equations of motion are Hamiltonian for G with bracket

$$\{f, g\} = \nabla f \cdot A \nabla g.$$

For convenience we introduce one more set of variables

$$x_i = \sqrt{\beta_i Q_i}(e^{v_i} - 1) = \sqrt{\frac{\beta_i}{Q_i}}(N_i - Q_i)$$

and we introduce parameters α_i and a matrix C with coefficients c_{ij} given by

$$\alpha_i = \frac{1}{\sqrt{\beta_i Q_i}} c_{ij} = \sqrt{\frac{Q_i}{\beta_i}} a_{ij} \sqrt{\frac{Q_j}{\beta_j}}.$$

The Liouvillian then takes the form

$$L = \sum_{i,j=0}^{n} c_{ij}(1 + \alpha_i x_i) x_j \frac{\partial}{\partial x_i}.$$

We want to find a reduced equation for the evolution of the zeroth species. To do so, we project using the conditional expectation with respect to v_0 with a 'canonical ensemble' with 'temperature' θ given by $e^{-G/\theta}$. For a justification and investigations with respect to ergodicity, see (Goel et al., 1971). One checks that for $i \neq 0$

$$\mathcal{P}(x_i) = \mathcal{P}(\frac{1}{\alpha_i}(e^{v_i} - 1)) = 0$$

since $x_i = \alpha \frac{\partial G}{\partial v_i}$. We apply the Mori-Zwanzig formalism to the equation for v_0 and see that $\mathcal{P} L v_0 = 0$. Thus

$$\begin{aligned}(1 - \mathcal{P}) L v_0 = L v_0 &= \beta_0^{-1} \sum_{i=0}^{n} a_{0i} Q_i (e^{v_i} - 1) \\ &= \sum_{i=1}^{n} \alpha_0 c_{0i} x_i\end{aligned}$$

and we want to solve the orthogonal dynamics for $x_i^\perp$. These are not readily analytically solvable, but we can use perturbation theory to get an approximation. Let us write $L = L_0 + L_1$ with

$$L_0 = \sum_{i,j=1}^{n} c_{ij} x_j \frac{\partial}{\partial x_i}$$

This is chosen in such a way that L_0 represents precisely the linearized orthogonal dynamics around the equilibrium Q_i (so around $x_i = 0$), notice that $\mathcal{P}L_0 = 0$. We will first consider the approximation $e^{(1-\mathcal{P})Lt} \sim e^{L_0 t}$, and later we will also compute second order effects of the nonlinear contribution $(1-\mathcal{P})L_1$.

Because L_0 is linear, the evolution of $e^{L_0 t}$ can be written in the following way:

$$e^{L_0 t} x_i = \sum_{j=1}^{n} U_{ij}(t) x_j$$

with $U_{ij}(t) = \exp(Ct)_{ij}$. We can thus write in first order approximation

$$\begin{aligned} F(t) &= e^{(1-\mathcal{P})Lt}(1-\mathcal{P})Lv_0 \sim e^{L_0 t}(\sum_{i=1}^{n} \alpha_0 c_{0i} x_i) \\ &= \sum_{i,j=1}^{n} \alpha_0 c_{0i} U_{ij}(t) x_j. \end{aligned}$$

The memory is then given first order by

$$\begin{aligned} \mathcal{P}LF(s) &\sim \mathcal{P}L \sum_{i,j=1}^{n} \alpha_0 c_{0i} U_{ij}(s) x_j \\ &= \sum_{i,j=1}^{n} \alpha_0 c_{0i} c_{j0} U_{ij}(s) x_0 \end{aligned}$$

and we conclude that

$$\dot{v} = \int_0^t K(t-s) x_0(s) ds + F(t)$$

with $K(s) = \sum_{i,j=1}^{n} \alpha_0 c_{0i} c_{j0} U_{ij}(s)$. If the initial conditions for N_i are drawn from the canonical distribution for $i > 0$

$$\mathbf{E}(F(t)F(s)) = \alpha_0 \theta K(t-s).$$

Finally, rewriting back to an equation for N_0 we get

$$\dot{N_0} = N_0 \int_0^t K(t-s) \sqrt{\frac{\beta_0}{Q_0}} (N_0(s) - Q_0) ds + N_0 F(t).$$

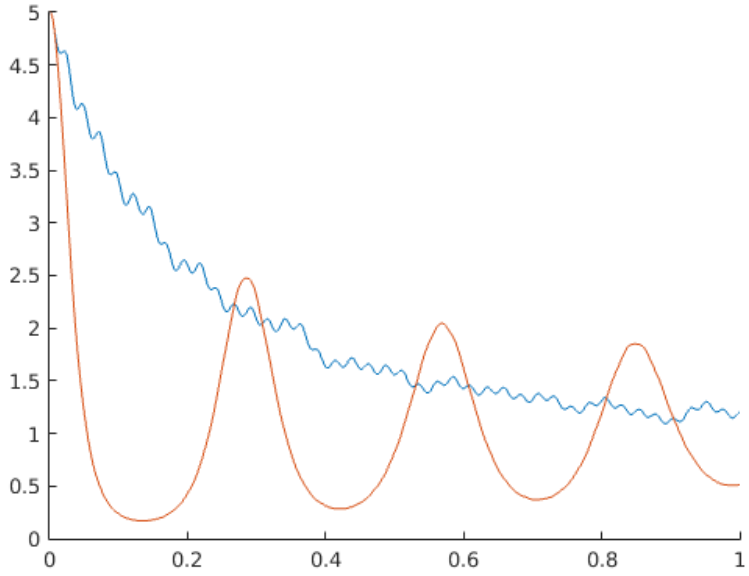


Figure 2.3: Plot of N_0 averaged over 100 simulations, with $n = 500$, $Q_i = 1$ for all i and $a_{ij} = 1$ for $i < j$, $i, j \neq 1$ with probability $p = 0.8$. The blue graph corresponds to the case where all a_{0i} are equal to 1 (short memory), and the red graph to the situation where a_{0i} is 1 or -1 with equal probability (long memory).

In the particular case where F is approximated by white noise, and hence $K(s) = K_0\delta(s)$ we get the stochastic Verhulst equation

$$\dot{N}_0 = K_0\sqrt{\frac{\beta_0}{Q_0}}N_0(N_0 - Q_0) + N_0F(t).$$

Can we indeed find predator-prey interactions which give rise to this type of noisy behaviour? Van Kampen (Van Kampen, 1974) indeed gives an example, and also discusses a case with long (exponential) memory, we have plotted the different types of behaviour in figure2.3.

Thus we have obtained a (heuristic) derivation of an SDE with multiplicative noise using the Mori-Zwanzig formalism. Notice however that it is actually an SDE with additive noise for v_0 (what happens here is more or less the Lamperti tranformation that maps an SDE with multiplicative noise to one with additive noise in one dimension (Pavliotis, 2014)).

Perturbation theory

To compute nonlinear contributions of the other species (the analogy to the heat bath), we use to Dyson formula to approximate

$$e^{(L_0+(1-\mathcal{P})L_1)t} = e^{L_0 t}(1 + \int_0^t e^{-L_0 s}(1-\mathcal{P})L_1 e^{L_0 s} ds + \ldots) = F_1(t) + F_2(t) + \ldots$$

So, we want to find $e^{-L_0(t-s)}(1-\mathcal{P})L_1 e^{L_0 s}(1-\mathcal{P})Lv_0$. We compute

$$\begin{aligned}
&(1-\mathcal{P})L_1 e^{L_0 s}\sum_{i=1}^n \alpha_0 c_{0i} x_i \\
&= (1-\mathcal{P})L_1 \sum_{i,j=1}^n \alpha_0 c_{0i} U_{ij}(s) x_j \\
&= (1-\mathcal{P})\sum_{i,j=1}^n \alpha_0 c_{0i} U_{ij}(s)(\sum_{k=1}^n c_{jk}(1+\alpha_j x_j)x_k + c_{j0}(1+\alpha_j x_j)x_0) \\
&= \sum_{i,j=1}^n \alpha_0 c_{0i} U_{ij}(s)(\sum_{k=1}^n c_{jk}\alpha_j x_j x_k + c_{j0}\alpha_j x_j x_0)
\end{aligned}$$

so

$$\begin{aligned}
&e^{-L_0(t-s)}(1-\mathcal{P})L_1 e^{L_0 s}(1-\mathcal{P})Lv_0 \\
&= \sum_{i,j,l=1}^n \alpha_0\alpha_j c_{0i} U_{ij}(s) U_{jl}(t-s)(\sum_{k,m=1}^n c_{jk} U_{km}(t-s)x_l x_m + c_{j0}x_l x_0).
\end{aligned}$$

We conclude that

$$F_2(t) = \int_0^t \sum_{i,j,l=1}^n \alpha_0\alpha_j c_{0i} U_{ij}(s) U_{jl}(t-s)(\sum_{k,m=1}^n c_{jk} U_{km}(t-s)x_l x_m + c_{j0}x_l x_0)ds.$$

Using this we can also compute the second order contribution to the memory term, $K_2(t,s,x_0) = e^{L(t-s)}\mathcal{P}LF_2(s)$. From computing some Gaussian integrals we find $\mathcal{P}x_i x_j = \delta_{ij}\theta$ and $\mathcal{P}x_i x_j x_k = \delta_{ij}\delta_{jk}\theta^2\alpha_i$ for $i,j,k \neq 0$ and with δ_{ij} the Kronecker delta. From this we easily check that

$$\begin{aligned}
\mathcal{P}Lx_l x_m &= \delta_{lm} 2c_{l0}\alpha_l \theta x_0 \\
\mathcal{P}Lx_l x_0 &= c_{l0}(x_0^2 - \theta(1+\alpha_0 x_0)).
\end{aligned}$$

We conclude that

$$K_2(t,s,x_0) = \int_0^t \sum_{i,j,l=1}^n \alpha_0\alpha_j c_{0i}c_{l0} U_{ij}(s) U_{jl}(t-s)(\sum_{k=1}^n c_{jk} U_{kl}(t-s)2\alpha_l\theta x_0 + c_{j0}(x_0^2 - \theta(1+\alpha_0 x_0)))ds$$

2.8 Averaging

Often we will look at systems with some form of scale separation, which take for example the form

$$\begin{aligned} \dot{x} &= R(x,y) \\ \dot{y} &= \frac{1}{\epsilon}S(x,y) + \frac{1}{\sqrt{\epsilon}}\gamma(x,y)\dot{W} \end{aligned} \tag{2.33}$$

with $\dot{W}$ white noise, and (x,y) defined on some space $X \times Y$. What can we say about the dynamics for x if ϵ is very small? The idea is the following: if ϵ is very small, the dynamics of y tend to an equilibrium distribution (which depends on x in general) very fast, so we can draw y in the equation of motion for x from this distribution. We will make this idea a bit more precise and discuss the relation to the Mori-Zwanzig formalism. The exposition is based on (Pavliotis and Stuart, 2008).

Consider the problem (2.33). We write $L = \frac{1}{\epsilon}L_0 + L_1$ for the Liouvillian where

$$\begin{aligned} L_1 &= R(x,y) \cdot \nabla_x \\ L_0 &= S(x,y) \cdot \nabla_y + \frac{1}{2}\Gamma(x,y) : \nabla_y \nabla_y \end{aligned}$$

with $\Gamma(x,y) = \gamma(x,y)\gamma(x,y)^\top$, and we assume that $\Gamma(x,y)$ is nondegenerate. We assume that for fixed x, the dynamics of L_0 are ergodic, in this case meaning that for each x the only solutions of $L_0 v = 0$ are the constant functions v. This means that there exists an invariant measure, with unique density function ρ_x satisfying $L_0^* \rho_x = 0$ such that ρ_x integrates to 1, and because we need ρ to decay sufficiently fast we assume that $\rho_x \in H_0^1(Y)$ (the closure of compactly supported smooth functions in the Sobolev space of Y). Define the measure $\mu_x(dy) = \rho_x(y)dy$, this represents the invariant measure for the situation where we keep x fixed and let y evolve in time. For times which are of smaller order than $\mathcal{O}(1)$ (which is the order on which x changes, so we assume that x is more or less fixed) but of larger order than $\mathcal{O}(\epsilon)$ (which is the order on which y changes, so we may assume that y has had enough time to behave according to the invariant measure) this should be a good approximation of the statistics of y given x. That motivates us to define the projection $\mathcal{P}$ onto functions of x which is given by

$$\mathcal{P}h(x) = \int h(x,y)\mu_x(dy)$$

and to approximate the dynamics of x by

$$\dot{x} = (\mathcal{P}f)(x). \tag{2.34}$$

Of course this will in general only be a good approximation if ϵ is very small. To see that in that case it is indeed a good approximation we look at the backward Kolmogorov equation. We define

$$v(x,y,t) = \mathbf{E}(x(t), y(t) | x(0) = x, y(0) = y)$$

which then satisfies

$$\frac{\partial v}{\partial t} = \frac{1}{\epsilon} L_0 v + L_1 v.$$

To solve this equation we expand v as $v = v_0 + \epsilon v_1 + \mathcal{O}(\epsilon^2)$ and on the level $\mathcal{O}(\frac{1}{\epsilon})$ we obtain

$$L_0 v_0 = 0 \tag{2.35}$$

and on the $\mathcal{O}(1)$ level we get

$$L_0 v_1 = \frac{\partial v_0}{\partial t} - L_1 v_0. \tag{2.36}$$

From equation (2.35) and our ergodicity assumption we know that v_0 is constant in y. Since we assume that $\Gamma(x, y)$ is nondegenerate, the PDE (2.36) is elliptic. This means that (given the appropriate boundary conditions and functional setting) we can use the Fredholm alternative for elliptic PDE's (Evans, 2010) on this equation (fixing x and t) to find that since $L_0^* \rho_x = 0$ has a nonzero solution (again the ergodicity assumption), that if $\frac{\partial v_0}{\partial t} - L_1 v_0 \in L^2(Y)$ (2.36) has a (weak) solution if and only if

$$\frac{\partial v_0}{\partial t} - L_1 v_0 \perp \text{Null}(L_0^*)$$

If we assume that such a solution v_0 indeed exists we thus get

$$\mathcal{P}(\frac{\partial v_0}{\partial t} - L_1 v_0) = 0$$

since the projection $\mathcal{P}$ is precisely the L^2 inner product with $\rho_x \in \text{Null}(L_0^*)$ and since v_0 does not depend on y, $\mathcal{P}(\frac{\partial v_0}{\partial t}) = \frac{\partial v_0}{\partial t}$ we find

$$\frac{\partial v_0}{\partial t} = \mathcal{P} L_1 v_0$$

which corresponds to (2.34). This justifies the approximation (2.34) as we let ϵ go to 0. In fact it is not hard to show that we have (strong) convergence in L^p for $p > 1$ (Pavliotis and Stuart, 2008). However, we did have to assume that the fast dynamics were stochastic, in order for (2.36) to be elliptic. For deterministic systems similar results are possible, but much harder to prove. One of the reasons for this is that the invariant measure of the invariant dynamics need not be continuous in x, and need also not have a density function.

At this point it is good to remark that for nonzero ϵ the projection $\mathcal{P}$ we defined above is not the same as the conditional expectation often used in the Mori-Zwanzig formalism, which depends on an invariant measure for L and not just for L_1. It turns out that in the $\epsilon \to 0$ limit they yield the same result under some smoothness assumptions as is made precise in the following lemma from (Stinis, 2005):

Lemma 2.8.1. *Assume that system* (2.33) *allows an invariant measure* $\tilde{\mu}^\epsilon$ *with density function* $\tilde{\rho}^\epsilon$*, and that in the limit* $\epsilon \to 0$ *this measure still admits a density* $\tilde{\rho}$*. Then the conditional measure*

$$\tilde{\rho}_x^\epsilon(y) = \frac{\tilde{\rho}^\epsilon(x,y)}{\int \tilde{\rho}^\epsilon(x,y)dy}$$

is in the limit $\epsilon \to 0$ *an invariant measure* $\tilde{\rho}_x$ *for the* L_0 *dynamics.*

Proof. Since $\tilde{\rho}^\epsilon$ is an invariant measure for the system (2.33) we know that

$$(\frac{1}{\epsilon}L_0^* + L_1^*)\tilde{\rho}^\epsilon = L^*\tilde{\rho}^\epsilon = 0.$$

In the limit $\epsilon \to 0$ the first term will dominate since we assumed that the limit of $\tilde{\rho}^\epsilon$ is smooth enough to allow a density function

$$\lim_{\epsilon\to 0} L_0^*\tilde{\rho}^\epsilon = 0$$

and hence

$$\lim_{\epsilon\to 0} L_0^*(\tilde{\rho}_x^\epsilon \int \tilde{\rho}^\epsilon(x,y)dy) = \lim_{\epsilon\to 0} \int \tilde{\rho}^\epsilon(x,y)dy L_0^*(\tilde{\rho}_x^\epsilon)$$

since $\tilde{\rho}^\epsilon(x,y)dy$ does not depend on y and L_0^* only has derivatives with respect to the y variables. We conclude that

$$\lim_{\epsilon\to 0} L_0^*\tilde{\rho}_x^\epsilon = 0$$

which is what we needed to show. □

From this we conclude that the Mori-Zwanzig formalism may be seen as a way to provide the correction terms for finite ϵ in the averaging procedure for deterministic systems.

Averaging Hamiltonian systems

The idea of averaging is also often used in the context of Hamiltonian systems to average out 'fast angles'. To see how this works we take an example Hamiltonian system with Hamiltonian

$$H(q_1,q_2,p_1,p_2) = \frac{1}{2}(p_1^2+p_2^2) + U(q_1) + \frac{V(q_1)}{2\epsilon^2}q_2^2. \tag{2.37}$$

with U and V some potentials. This example and the link to averaging are described in (Pavliotis and Stuart, 2008). Rescaling q_2 and p_2 shows that the corresponding dynamics are of the form (2.33). In this case the dynamics for fixed (q_1,p_1) for the fast variables (q_2,p_2) behave as a harmonic oscillator with energy $E - \frac{1}{2}p_1^2 - U(q_1)$ where

E is the total energy and frequency $\frac{V(q_1)}{2\epsilon^2}$. If we average $\frac{V'(q_1)}{2\epsilon^2}q_2^2$ over this oscillation we find $\frac{V'(q_1)}{2V(q_1)}(E - \frac{1}{2}p_1^2 - U(q_1))$ and hence the averaged equations are

$$\begin{aligned}\dot{q_1} &= p_1 \\ \dot{p_1} &= -U'(q_1) - \frac{V'(q_1)}{2V(q_1)}(E - \frac{1}{2}p_1^2 - U(q_1)).\end{aligned}$$

This result corresponds precisely with the optimal prediction $S^t\mathcal{P}L$ when using the microcanonical ensemble (which is easily seen to be the case in general). Thus the Mori-Zwanzig formalism provides correction terms for the averaging procedure for nonzero ϵ. It may be interesting to use this perspective to study low-dimensional Hamiltonian systems with chaotic regions, such as the Henon-Heiles model, which is of a similar form, but without the scale separation. A first attempt has been made by (Ishizaki et al., 2006) using the linear projection to analyse the autocorrelation and spectrum of these systems. Another interesting situation occurs for systems with more dimensions, where the averaging procedure does not give good results due to resonances.

I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'Law of Frequency of Error'. The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement,amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

Galton, Adress to the Anthropological Institute

3

Stochastic parametrization

Let us consider a very practical problem that was already alluded to in the introduction: predicting the weather (or climate). In principle we have partial differential equations that model the evolution of the atmosphere, and if we can solve these equations with the current state of the atmosphere as initial condition we might expect to get a reliable prediction of the weather. However, multiple problems arise that prevent us from obtaining (long term) predictions. Firstly we can only approximate the solution numerically, and our computers are only able to solve for relatively large scale components (large wave numbers or coarse grid points). Let us call these large scale components the resolved degrees of freedom, and the other components the unresolved degrees of freedom. Most of the time we only know the (estimates of) the initial conditions of the resolved degrees of freedom. To approach the problem we could discretize the equations, and only keep the finite set of resolved degrees of freedom and use this to numerically approximate a solution of the original system. The numerical cost is reduced in two ways: we need to update only a few variables each time step, and because we only keep track of the relatively slowly evolving components of the system we can use a larger time step. This procedure does however introduce an error, and in the case of atmospheric models (and more general for any chaotic system with positive Lyapunov exponents) this error can be very serious since events that take place on small scales may have strong effects on the large scale behaviour, a phenomenon first remarked upon by (Lorenz, 1963). Eliminating this error is not realistic, and there is a ' predictability horizon' beyond which we cannot predict the state of the system (Lorenz, 1995), so we have to settle for two less ambitious goals. The first goal would be to minimize this error (in some sense to be made precise) in order to get the best/most likely prediction (for relatively short times). The second

goal would be to get an idea of how large the error is, so we do not only know what is going to happen most likely, but also how good the agreement of our prediction with reality will probably be. One way to achieve the first goal in our model is by gathering data and estimate a term in the model that depends only on the resolved degrees of freedom. This term should correct for the presence of the unresolved degrees of freedom and should minimize the predictive error. Such a procedure is (in the context of climate and weather) called *parametrization* (Wilks, 2005), (Arnold et al., 2013). In order to achieve the second goal, we will however need a stochastic model, that takes into account the uncertainty of the initial conditions, by representing the unresolved degrees of freedom as some stochastic process.

The Mori-Zwanzig formalism can be used as a starting point to formulate such a procedure mathematically, by projecting on the resolved degrees of freedom. The main insight it gives is that, especially in systems without a very strong scale separation (such as the atmosphere), it may also be important to account for memory effects in the stochastic parametrization. In this chapter we will discuss some methods based on the Mori-Zwanzig formalism to approximate high-dimensional systems by low-dimensional (non-Markovian) stochastic systems. The analytic formulation that was presented in the previous chapter can in practice often not be performed. One reason may be that we do not even know the governing equations for the unresolved processes. Even if we do know a description of these processes, in most cases it is not possible to directly analyze the orthogonal dynamics, for instance because we do not have an expression for the invariant measure and hence for the projection operator. This means that we have to find ways to obtain estimates for the stochastic and memory terms using available data of the system. Since we will mostly be interested in systems that are not analytically tractable we have to rely on observations and numerical simulations to obtain these data, which will thus be discrete, and we have to find a way to obtain estimates from a time series.

3.1 General form of the discretization

Let us write x for the resolved variables and y for the variables that are unresolved. We assume that we have a system of differential equations similar to (2.33) for x and y which we write as

$$\dot{x} = R(x, y)$$
$$\dot{y} = S(x, y)$$

We consider the MZ-formalism for the projection onto functions of x, to get the following equation for the evolution of x, given initial condition $x(0)$:

$$\frac{\partial}{\partial t}x = \tilde{R}(x) + \int_0^t S^{t-s}K(x, s)ds + F(t) \tag{3.1}$$

where $\tilde{R} = \mathbf{E}(R|x(0))$ (so this is just equation (2.8) with slightly simplified notation).

We assume we want to solve for a time series $x^i = x(t_i)$ with fixed time step Δt, so $t_i = \Delta t \cdot i$. So $x^i = S^{t_i}x$ where x is the initial condition. We can derive a formula for a discrete operator in the same spirit as formula (2.5) (Beck et al., 2009). For this let M be a linear operator (the case that interests us is $S^{\Delta t} = e^{\Delta t L}$ on the space of L^2-functions) and let $\mathcal{P}$ be a projection. We want to study the behaviour of $M^k f$ for some f. Then we may define F^i by $F^0 = (1-\mathcal{P})f$ and $F^{i+1} = (1-\mathcal{P})MF^i$. Then we can prove by induction that

$$M^k f = M^k \mathcal{P} f + \sum_{i=1}^{k} M^{k-i}\mathcal{P}MF^{i-1} + F^k. \tag{3.2}$$

If we indeed take $M = S^{\Delta t}$ then this gives an exact expression for x_k, but notice that $F_k \neq F(\Delta t k)$ as defined in the continuous version. Now if we let in particular $f = Lx = R = \frac{\partial}{\partial t}x$, then we can use a numerical scheme (for example Runge-Kutta)

$$x^{n+1} = x^n + \Delta t R_\Delta(x^n, y^n).$$

From (3.2) it follows that $R(x^n, y^n) = \tilde{R}(x^n) + \sum_{i=1}^{n} M^{n-i}\mathcal{P}MF^{i-1} + F^n$ where $\tilde{R} = \mathcal{P}R$. If we write z^n for $\sum_{i=1}^{n} M^{n-i}\mathcal{P}MF^{i-1} + F^n$ we get an integration scheme

$$x^{n+1} = x^n + \Delta t \tilde{R}_\Delta(x^n) + \Delta t z^n. \tag{3.3}$$

The z^n can be determined from the x^n since they form a discrepancy $z^n = (x^{n+1} - x^n)/\Delta t - \tilde{R}_\Delta(x^n)$. This means that if we have data of x we can also determine a time series for z (notice that this is in general not stable for larger Δt). The problem we want to look at, is whether it is also possible from the data to get a good idea of how z^n behaves. As we see it has two parts, a 'noisy' part and a part that has memory (and depends on previous values of x).

In what follows we will assume that the original system is such that the time series x^n is stationary and ergodic. Thus it makes sense to try to find a way to sample z_n conditioned on $x^n, x^{n-1}, \ldots$ in such a way that the resulting process has statistics that are very similar to the statistics of the original process. We are generally not interested in the actual predictions we make for z^n, but only in the behaviour of the x^n, so this means we can be satisfied if we know the right statistical behaviour of z^n rather than following a (complicated) evolution of the orthogonal dynamics. In an abuse of notation we will also write z^n for this stochastic process, and x^n for the corresponding approximation to the original system.

3.2 Different procedures for stochastic parametrization

The NARMAX representation

From (3.3), and approximating the orthogonal dynamics by a stochastic process, it follows that $z^n = \Phi^n + \Psi^n$ where Ψ^n is a random variable (but the Ψ^n need not be

independent) and $\Phi^n = \Phi^n(x^n, x^{n-1}, \ldots)$ is some nonlinear memory. We will represent Ψ^n by a process that is a function of a series of IID Gaussian variable ζ^n, so $\Psi^n = \Psi^n(\zeta^n, \zeta^{n-1}, \ldots)$. This is still very general, and can considered as a NARMAX (nonlinear autoregressive moving averages exogeneous) process (Chorin and Lu, 2015). The exogeneous input here is the time series x^n, so this is slightly different from what is usual since the x^n depend on z^n. To be able to say something about such a model, one needs to make choices about which functions are allowed to occur and how deep the memory is. We will use the Mori-Zwanzig formalism to motivate the structure of this NARMAX process.

In many cases the Mori-Zwanzig formalism leads to a generalised Langevin equation of the form

$$\frac{\partial}{\partial t}x(t) = R(x(t)) + \int_0^t K(t-s)h(x(s))ds + F_\alpha(t)$$

for a (possibly vector valued) relevant variable x. This equation can be exact, or it can be obtained by perturbation theory for the orthogonal dynamics. In particular, if we use the linear projection operator we get this equation with R and h linear operators. However, for stochastic parametrizations it is useful to also consider the case where R and h need not be linear, but some other (known) function. In this case the discretization leads to a system (Schmitt and Schulz, 2006), (Horenko et al., 2007), (Chorin and Lu, 2015) as we will show now.

$$x^{n+1} = x^n + R_\Delta(x^n) - \sum_{i=1}^{k} K^{n-i}h(x^i) + \Psi^n.$$

where we assume the memory to be zero after some finite time k. The noise term Ψ can for instance be modeled by an AR process, leading to

$$x^{k+1} = x^k + R_\Delta(x^k) + \sum_{i=1}^{k} K^{k-i}h(x^i) + \zeta_k + \sum_{i=0}^{m} L^{k-i}\zeta^i.$$

again with ζ^i IID Gaussian variables. Also we recall the fluctuation dissipation theorem, that suggests that in certain situations the autocorrelation of the noise is proportional to the memory kernel, i.e $\beta L = K$. Another easy case is the situation where we assume that Ψ is white noise so the L_i are zero for $i > 0$.

Moment estimators

For a stochastic parametrization we want to estimate values for K^i and L^i using data from a time series for x^n with N data points available. Because we assume we know R_Δ explicitly, we also have data for z^n available. There is an extensive literature on this type of problem, and we will only discuss a very limited set of methods which is based on the autocorrelation functions. There is a standard least-squares estimation

procedure using autocorrelation functions, the Yule-Walker estimators (Kneller and Hinsen, 2001). Let us write $C_{f,g}(t)$ for the crosscorrelation of two functions f and g. If the time series is ergodic its discrete analog $C^k_{f,g}$ can be estimated from the data by

$$G^k_{f,g} = \frac{1}{N-k}\sum_{i=1}^{N-k}(f(x^i)-\bar{f})(g(x^{i+k})-\bar{g})$$

with $\bar{f}$ the average of $f(x^i)$. We find z^n from the data using

$$z_n = x^{k+1} - x^k - R_\Delta(x^k) = \sum_{i=1}^{k} K^{k-i}h(x^i) + \zeta_k + \sum_{i=1}^{k} L^{k-i}\zeta^i$$

If we consider the autocorrelation of z with x we find

$$C^n_{z,x} = \sum_{i=1}^{k} K^i C^{n-i}_{h(x),x} + \sum_{i=1}^{m} L^i C^{n-i}_{\zeta,x}.$$

We know that $C^n_{\zeta,x} = 0$ for $n > 0$ so we find that for $n > m$ we have

$$C^n_{z,x} = \sum_{i=1}^{k} K^i C^{n-i}_{h(x),x}.$$

This is a set of linear equations for K, similar to the Yule-Walker equations for ARMA processes. These equations can be recast in matrix form

$$\begin{pmatrix} C^m_{z,x} \\ \vdots \\ C^{m+k-1}_{z,x} \end{pmatrix} = \begin{pmatrix} C^m_{x,x} & \dots & C^{m-k+1}_{x,x} \\ \vdots & \ddots & \vdots \\ C^{m+k-1}_{x,x} & \dots & C^m_{x,x} \end{pmatrix} \begin{pmatrix} K^1 \\ \vdots \\ K^k \end{pmatrix}.$$

With the data for x and z available, in principle it is possible to invert this linear problem to obtain K. However, these systems tend to be unstable, especially for larger k (Haykin, 1986). So we will look for approximations to the problem that have better stability properties.

3.3 Markovian approximations

In section 2.6 the Laplace transform was used to introduce methods to approximate a generalized Langevin equation by a finite dimensional Markovian system. We will discuss some estimation procedures based on these approximations

Discrete Laplace transform methods

Since we work with a discrete time series, we use the discrete analog of the Laplace transform, the (one-sided) z-transform, in which a time series $\{x^i\}_{i=0}^{\infty}$ is transformed to a complex formal power series

$$\hat{x}(z) = \sum_{i=0}^{\infty} \frac{x^i}{z^i}$$

It has the nice property that if time series x and y are related by

$$x^k = \sum K^i y^{k-i}$$

we get that

$$\hat{x} = \hat{K}\hat{y}.$$

This suggests that one can obtain K by taking the inverse transform of $\hat{x}/\hat{y}$, where in our problem $x = C_{z,x}$ and $y = C_{x,x}$. Unfortunately, as for Laplace transforms, it is in general not so easy to compute inverse z-transforms (in principle it can be done by contour integration), and numerical algorithms are not stable (Epstein and Schotland, 2008). As already discussed in section 2.6 there are some possibilities to approximate the Laplace transform (and hence the z-transform) of the kernel, in terms of expressions for which we do know an explicit inverse transform. The easiest such way is by approximating the fraction $\hat{x}/\hat{y}$, which is a rational function, by using a partial fraction expansion, writing

$$\hat{x}/\hat{y} = \sum_{i=1}^{k} \frac{r_i}{p_i - z^{-i}}$$

where k is chosen to get a good approximation and is at most the length of x and y. We use that $\frac{r}{p-z^{-i}}$ is the transform of the series $x_{r,p}$ with

$$x_{r,p}^i = rp^i.$$

The partial fraction expansion approximation thus comes down to approximating the kernel K by a sum of (complex) exponentials. A version of this approach applied to simulations of an argon liquid can be found in (Kneller and Hinsen, 2001).

Parametrizing with Ornstein-Uhlenbeck processes

The Laplace transform and z-transform suggest that it is easiest to work with a memory kernel that has some prescribed parametrization, such as an exponential one. The form suggested by the Laplace transform is that of a sum of complex exponential functions.

$$K(t) = \sum_i r_i \exp p_i t$$

Because the memory has to be real, this means that we approximate the kernel by a function of the form

$$K(t) = \sum_i r_i e^{-\alpha_i t} \cos(\omega t + \beta).$$

This approach has two distinct advantages: it is much easier to estimate the kernel (because now only these parameters have to be estimated), and simulating the parametrized system is more efficient, because we can realize this memory kernel by means of a system where we have added additional OU processes with an x dependent mean

$$\begin{aligned} \dot{x} &= R(x) + \sum z_i \\ \dot{z_i} &= p_i z_i - r_i x + \Sigma_i \dot{W}. \end{aligned}$$

This approach has been worked out for a nonlinear Schrödinger equation by (Berry and Harlim, 2015), where a Kalman filter based algorithm is used to compute coefficients. In the next chapter we work out this approach for the heat bath model and for another toy model: the Lorenz '96 model.

Continued fractions and a multilevel algorithm

Finally there is a method based on the continued fraction expansion of Mori, which essentially comes down to add OU processes level by level (Kondrashov et al., 2015). As discussed in section 2.6 the continued fraction expansion of the Laplace transform corresponds to a hierarchy of projection operators. As in equation (2.28) we obtain a system of differential equations of the form

$$\begin{aligned} \dot{x} &= R(x) + f_1 \\ \dot{f_i} &= L_i f_i + f_{i+1} \text{ for } i = 1, 2, \ldots \end{aligned}$$

The algorithm proceeds as follows:

1. We obtain $R(x) = PLx$ by using a least squares fitting of a function (for instance a polynomial, but this may depend on what we already know about the structure of the x-variables) to data for x, motivated by lemma 2.1.1. We obtain a time series for the residue f_1
2. Then we compute for each f_i the term L_i, which is obtain by linear projection on f_i so we look at a time series for $\dot{f_i}$ (for instance $(f_i^{n+1} - f_i^n)/\Delta t$) and fit a least squares term $L_i f_i$ to this. Then we proceed to f_{i+1}
3. We truncate the procedure at some level and model the last f_i as white noise. The criterion for truncating at a given level may for instance be that the f_i have negligible autocorrelation in time (Kondrashov et al., 2015).

A variation on this is to let the projection on the i-th level be the projection onto $f_1, \ldots, f_i$. This corresponds to a Markovian approximation as in theorem 2.6.1 with lower triangular A. This kind of system has as a problem that it may be unstable, an issue that is also adressed in (Kondrashov et al., 2015) using energy conservation methods.

I of dice posses the science, and of numbers I am skilled.
Vyasa, Story of Nala in the Mahabharata

4 Numerical experiments

We will apply the usage of a Markovian approximation to the memory in two typical problems, firstly to the heat bath model (which is typical for statistical mechanics and has a Hamiltonian structure), and to the Lorenz 96 model, which is a toy model for a climate system.

4.1 Application to the heat bath model

We will apply some of these methods to the heat bath model previously discussed with equations of motion

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -U'(x) + \sum_j k_j(q_j - x) \\ \dot{q}_i &= \frac{p_i}{m_i} \\ \dot{p}_i &= -k_i(q_i - x). \end{aligned}$$

We will first consider the situation with $N = 100$ oscillators, and choose the parameters guided by theorem 2.4.3. That is, we draw ω_j from a uniform distribution on $[0, N^a]$. Then we take

$$k_j = \frac{2\alpha}{\pi(\alpha^2 + \omega^2)} N^{a-1}$$

where we choose $\alpha = 2$ and $a = \frac{1}{3}$. This also determines the mass parameters. The initial conditions for p_i and q_i are sampled from the canonical distribution with inverse temperature β, which we choose to be 10. For finite N we have a memory term

$$K(s) = \sum_{j=1}^{N} \cos(\omega_j s)$$

but we know that in the limit $N \to \infty$ we should get the system

$$\begin{aligned} \ddot{x} &= -U'(x) + y \\ \dot{y} &= -2y - \dot{x} + \sigma \dot{W} \end{aligned} \tag{4.1}$$

with W Brownian motion, so y is an OU process with mean $\dot{x}$. The parameter σ is given by

$$\sigma = \sqrt{\frac{2\alpha}{\beta}}$$

as follows from the FD theorem and the standard deviation of an OU process. We take a double well potential

$$U(x) = \frac{x^4}{4} - \frac{x^2}{2}.$$

If the heat bath is at a sufficiently high temperature, the particle will be able to 'jump' between the two wells, as can be seen in figure 4.1 for a typical trajectory. In this case the system is what is called metastable: it remains for a relatively long time stably in one of the wells, and then suddenly jumps to to the other. For even higher temperatures the barrier is no longer a serious obstacle and the particle freely jumps between the two basins.

For this model it is very natural to parametrize using a v dependent OU process (since we know that the $N \to \infty$ limit is of this form). For different parameter settings this has already been studied in (Verheul and Crommelin, 2016), indeed leading to good results. To simulate the full system we use a Verlet scheme, and we must take the step size to be of the order of the inverse of the smallest frequency (in this case that means we can take $\Delta t = 10^{-2}$), as argued in (Stuart and Warren, 1999). The full system was simulated for a time $T = 10^5$. From the time series x^n and v^n we computed

$$z^n = \frac{v^{n+1} - v^n}{\Delta t} - U'(x^n) \tag{4.2}$$

With this time series we computed the correlation functions C_{zv} and C_{vv}. Next we find a least squares approximation to the solution of (the discrete version of)

$$C_{zv}(t) = \int_0^t K(t-s) C_{vv}(s) ds \tag{4.3}$$

with $K(t-s) = re^{-p(t-s)}$. The parametrized system is then given by

$$\begin{aligned} \ddot{x} &= U'(x) + y \\ \dot{z} &= -pz + r\dot{x} + \tilde{\sigma}\dot{W} \end{aligned} \tag{4.4}$$

where we need to estimate $\tilde{\sigma}$. One way of doing this is by

$$\tilde{\sigma} = \sqrt{2p}\text{std}(\tilde{z}) \tag{4.5}$$

where $\tilde{z}$ is the OU process with mean 0 which replaces the orthogonal dynamics, and which we obtain from z by

$$\tilde{y}(t) = z(t) - \int K(t-s)v(s)ds$$

and which we can estimate from the data. In figure 4.3 we have compared the autocorrelation of $\tilde{z}$ resulting from 4.5 and from the real system. This suggests that it may be better to find σ by a least squares fit of these autocorrelations. Indeed in figure 4.4 the results of the autocorrelation of v for these parametrizations are compared, and the result obtained by fitting σ to the autocorrelation gives a superior result. From figure 4.4 it becomes clear that our method gives a good parametrization for this system. Of course we anticipated this result (knowing the $N \to \infty$ limit), but it is still interesting. Firstly it shows that also for finite N the approximation using an OU process is very good, and secondly it shows that in this case the estimation procedure works. In the next section we will apply our methods to a case where we do not have analytic knowledge about the memory kernel.

4.2 Application to the Lorenz 96 model

The dynamics of the atmosphere and the ocean are governed by partial differential equations. To solve these numerically, one would need to discretize space in some way, in order to get a set of ordinary differential equations. One way of doing this is by using Fourier analysis and truncating to keep only the large scale modes. The result of such a procedure often is a model of the form

$$\frac{dz}{dt} = Lz + B(z,z) + F \tag{4.6}$$

with z some multi-dimensional variable, L a linear operator and B a quadratic form, and F representing a constant forcing (Majda and Wang, 2006). One example of such a system is the Lorenz 96 model (Lorenz, 1995), which does not arise directly from a discretisation of a real geophysical system, but which has been suggested to mimic the behaviour of mid-latitude weather and which has properties that are analogous to real-world systems. It consists of the following differential equations:

$$\dot{x}_k = x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F$$

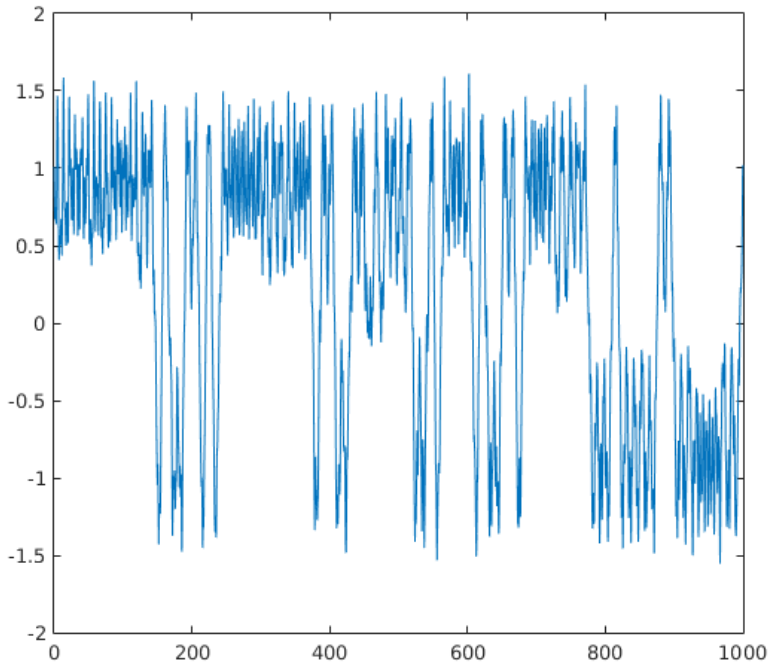


Figure 4.1: A typical solution for x at inverse temperature $\beta = 0.5$.

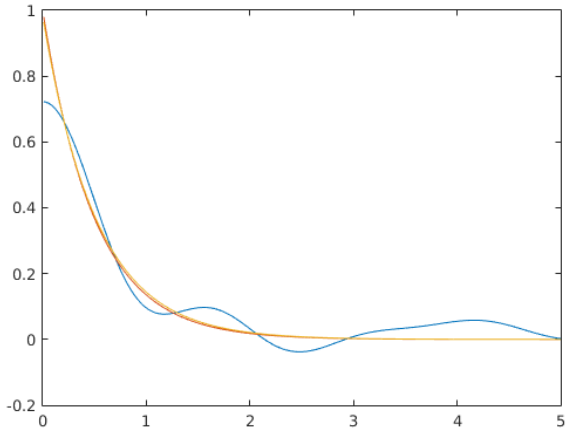


Figure 4.2: A comparison of the real memory kernel (blue), the estimated kernel (red) and the limit prediction (yellow).

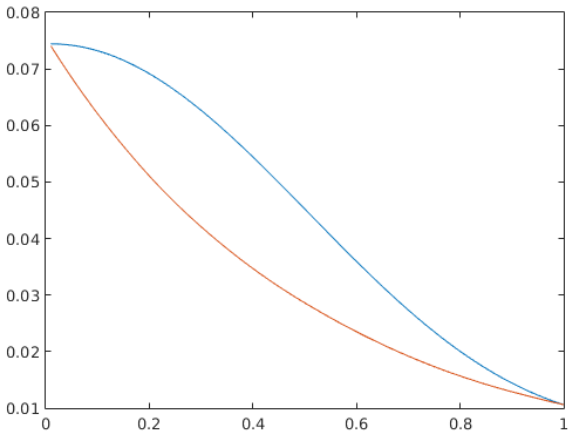


Figure 4.3: Comparison of the autocorrelation of the orthogonal dynamics (blue) and that of the noise $\tilde{z}$ in the parametrized system (red).

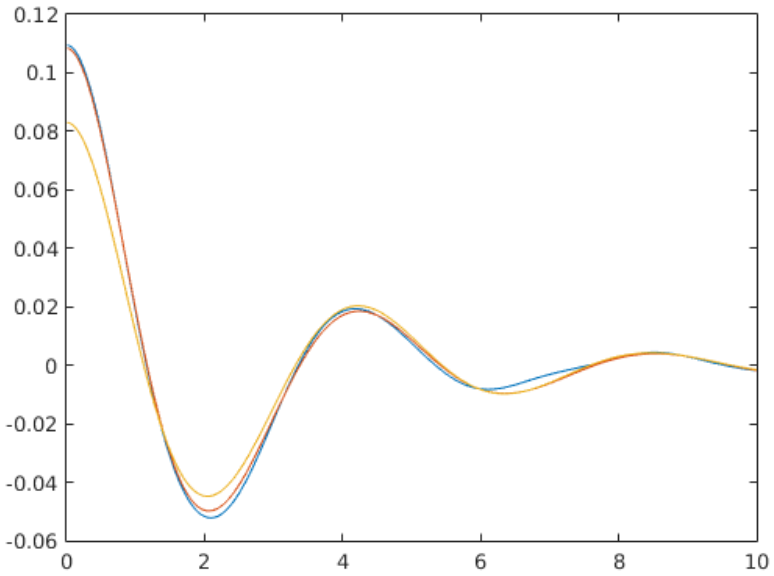


Figure 4.4: A comparison of the autocorrelation of v for the original system (blue) and the estimated parametrization using the improved estimator for $\tilde{\sigma}$ (red) compared to the result using 4.5 (yellow).

with periodic indices $k = 1, \ldots, K$ (so $x_k = x_{K+k}$). One can show by Fourier analysis that the linearised system has wave solutions, and the stability of these wave modes depends on the value of the forcing F. Larger forcing implies more unstable modes, and thus a more chaotic system.

In this section we will consider a version of the model (also suggested by Lorenz), where there is also a coupling to a smaller scale set of variables:

$$\begin{aligned} \dot{x}_k &= x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F + z_k \\ \dot{y}_{j,k} &= \frac{1}{\epsilon}(y_{j+1,k}(y_{j-1,k} - y_{j+2,k}) - y_{j,k} + h_y x_k) \end{aligned} \tag{4.7}$$

where $z_k = \frac{h_x}{J}\sum_j y_{j,k}$ and the indices run over $k = 1, \ldots, K$ and $j = 1, \ldots, J$ cyclically, so $y_{j+J,k} = y_{j,k+1}$ and $y_{j,k+K} = y_{j,k}$. The parameter ϵ determines how much smaller the time scale on which the y-variables evolve is compared to the time scale of the x-variables. It is not so straightforward to apply the Mori-Zwanzig formalism in an analytical way, since we do not have a good expression for the invariant measure.

The behaviour of (4.7) depends of course strongly on the choice of parameters. For very small ϵ the y-variables change very rapidly and one could use averaging techniques (which neglect memory effects). An analysis and numerical implementation of this approach can be found in (Fatkullin and Vanden-Eijnden, 2004). However, in real geophysical systems the scale separation between different processes is often not very small (Nastrom and Gage, 1985), and thus it is also interesting to look at situations with larger ϵ, where memory effects will be important. We will take $\epsilon = \frac{1}{2}$. Furthermore we choose all parameters as in (Chorin and Lu, 2015) and (Crommelin and Vanden-Eijnden, 2008), with $K = 18$, $J = 20$, $F = 10$, $h_x = -1$ and $h_y = 1$. The full system was simulated using a time step $\Delta t = 10^{-3}$ over a time $T = 10^4$ using a fourth order Runge-Kutta method. In figure 4.5 some snapshots for the time evolution of the system for these parameters are shown.

The Lorenz 96 model has been used as a test bed for stochastic parametrization schemes, in which the goal is to replace z_k by an (x-dependent) stochastic process (Crommelin and Vanden-Eijnden, 2008), (Fatkullin and Vanden-Eijnden, 2004), (Arnold et al., 2013), (Wilks, 2005). From the equations (4.7) we guess that it is not an unreasonable approximation that z_k does not depend on x_l for $l \neq k$. One way to achieve this approximation is by setting $y_{j+J,k} = y_{j,k}$ instead of $y_{j+J,k} = y_{j,k+1}$. Also the equations (4.7) are invariant under rotation of the indices, so all z_k will have the same (statistical) behaviour. A deterministic way to parametrize the system would then be to replace z_k by some function of x_k

$$z_k = G(x_k).$$

In figure 4.10a a scatter plot for z_k against x_k can be seen, with a fifth order polynomial fitted to the data. If we use this function to parametrize G we get approximations for the statistics as shown in figure 4.7.

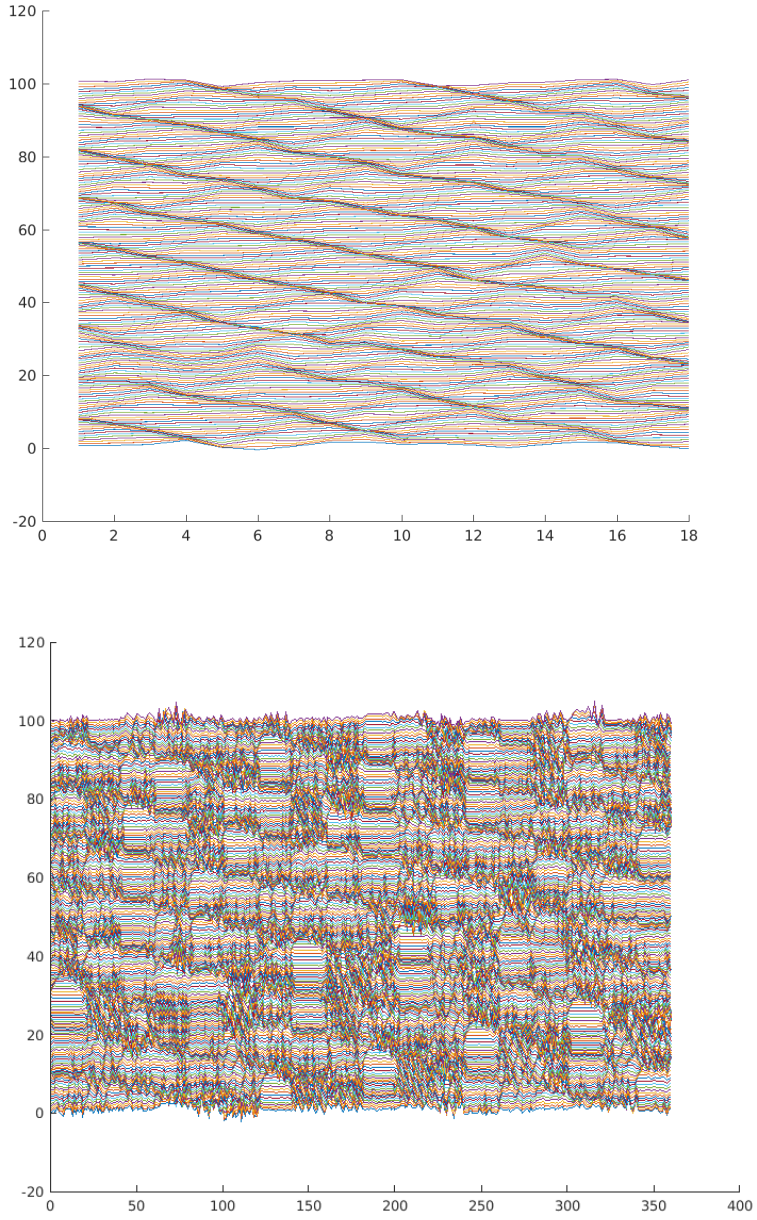


Figure 4.5: Snapshots of the evolution of the x and y variables in (4.7), vertical axis denotes time.

The statistics are not yet very accurately reproduced in this way, and it has already been shown that stochastic methods can improve these results. One of the approaches is to add an AR(1) process on top of the polynomial regression (POLYAR) (Wilks, 2005). This approach works well for small ϵ but breaks down for larger ϵ, which is not surprising since it does not take into account memory effects. For small ϵ these results can be improved using multiplicative noise (Arnold et al., 2013). In (Crommelin and Vanden-Eijnden, 2008) conditional Markov chains have been used, and in (Chorin and Lu, 2015) a NARMAX method has been used, which both implement memory effects. Both obtain very close reproductions of the statistics of the original problem for our choice of parameter settings.

Here we want to investigate the use of a parametrized memory, that is, adding an OU process with an x_k-dependent mean, which is essentially a linear continuous version of the conditional Markov chains used by (Crommelin and Vanden-Eijnden, 2008). We may also compare it to the MTV approach for stochastic parametrization outlined by (Majda et al., 2001) for systems of the form (4.6), in which case the quadratic interaction of the y-variables is approximated by an OU-process, and then a homogenization procedure is followed. In the particular case of the Lorenz '96 model this approach is easy to work out, changing the equations for the y-variables to

$$\dot{y}_{\cdot,k} = Ay_{\cdot,k} + \Sigma \dot{W}_k + \frac{h_y}{\epsilon} x_k \tag{4.8}$$

grouping together all the linear terms in $Ay_{\cdot,k}$. Notice that here we have neglected the interaction between the different 'blocks' of y-variables. As before, this leads to a solution

$$y_{j,k} = e_{ji}^{-At} y_{i,k}(0) - \int_0^t \sum_i (e^{-A(t-s)})_{ji} \frac{h_y}{\epsilon} x_k(s) ds + \int_0^t e^{-As} \Sigma^\top \Sigma dW_k(s)$$

and thus we get an expression for z as an OU-process plus an x_k-dependent memory term

$$z_k = \frac{h_x}{J} \sum_j y_{j,k} = \int_0^t K(t-s) x_k(s) ds + \frac{h_x}{J} \sum_j y_{j,k}^0$$

where

$$K(s) = \frac{h_x h_y}{J\epsilon} \sum_{i,j} (e^{-As})_{ij}$$

and y^0 is an OU-process satisfying

$$\dot{y}^0 = Ay^0 + \Sigma \dot{W}_k.$$

In (Majda et al., 2001) the next step is to assume ϵ to be small and approximate the memory term by a Markovian one, but we will not do so since we are interested in the case without strong scale separation. There is one major problem in the derivation above: the validity of replacing the dynamics for $y_{j,k}$ in (4.7) by stochastic dynamics as

in (4.8). If x_k is small, the equation for y becomes strongly dissipative and tends to zero quickly. This is demonstrated in figure 4.5 where we see that the $y_{j,k}$ are essentially constantly zero in regions where the x_k are small. One way to fix this is to add a stochastic white noise term to the original system and consider

$$\dot{y}_{j,k} = \frac{1}{\epsilon}(y_{j+1,k}(y_{j-1,k} - y_{j+2,k}) - y_{j,k} + h_y x_k + \sigma \dot{W}_{j,k}) \tag{4.9}$$

where $W_{j,k}$ is Brownian motion. In any case, the linear coupling suggests that it is reasonable to approximate the memory term by a linear memory term.

As in section 3.1 we try to estimate a stochastic process for a time series z_k^n, where we have discretised the system as

$$x_k^{n+1} = \Delta t R_\Delta(x^n) + \Delta t z_k^n$$

where $R_\Delta(x^n)$ has been obtained from a fourth order Runge Kutta discretization of the term $x_{k-1}(x_{k+1} - x_{k-2}) - x_k + F$. We will discuss two approaches to model z_k:

1. A parametrization inspired by the MTV approach, where we approximate z_k by an OU process that has a mean that depends (linearly) on x_k. In this case z_k^n satisfies
$$z_k^{n+1} = z_k^n - \Delta t(pz_k^n - rz_k^n) + \sqrt{\Delta t}\sigma\zeta_k^n \tag{4.10}$$
with p, r and σ parameters, and ζ_k^n IID $\mathcal{N}(0,1)$ random variables.

2. A parametrization inspired by the Mori-Zwanzig formalism, where we use the conditional expectation with respect to x as projection operator and estimate $\mathcal{P}Lz_k = \mathbf{E}(Lz_k|x) \approx \mathbf{E}(Lz_k|x_k)$ by $G(x_k)$ (based on lemma 2.1.1), and approximate $(1-\mathcal{P})z_k$ by an OU process that has a mean depending on x_k (as in section 3.3). Thus we write
$$\begin{aligned} z_k^n &= G(x_k^n) + \tilde{z}_k^n \\ \tilde{z}_k^{n+1} &= \tilde{z}_k^n - \Delta t(p\tilde{z}_k^n - rz_k^n) + \sqrt{\Delta t}\sigma\zeta_k^n \end{aligned} \tag{4.11}$$
again with parameters p, r and σ and ζ_k^n IID $\mathcal{N}(0,1)$ random variables.

The easiest approximation is to take a 1-dimensional OU process. We have used a moment estimator to obtain the parameters (but of course in this case a maximum likelihood estimator, as in (Verheul and Crommelin, 2016) is also possible). For the MTV procedure we start from the equation for the correlation functions

$$C_{z_k x_k}(t) = \int_0^t K(t-s)C_{x_k x_k}(s)ds + G_{F_k x_k}(t)$$

where F_k is the remaining stochastic process. We assume that the noise F_k has a small decorrelation time t_0, so that for $t > t_0$

$$C_{z_k x_k}(t) = \int_0^t K(t-s)C_{x_k x_k}(s)ds.$$

First the time series z_k^n is computed from the data for the original Lorenz '96 model using the discrepancy. The correlation functions are estimated from these data, and we use a Newton method to find a least squares solution $K(t) = re^{-pt}$ to the discrete version of this equation which can be seen in figure 4.6. A scatter plot of the convolution of x_k with this kernel is shown in figure 4.10b. Finally the parameter σ is estimated, just like in the heat bath case, from the residual time series

$$\tilde{F}_k^n = z_k^n - \sum K^{n-i} x_k^i. \tag{4.12}$$

The MZ procedure is the same, except we use $\tilde{z}_k = z_k - G(x_k)$ instead of z_k.

In order to see whether this stochastic procedure allows to use larger time steps we have performed the estimation procedure for different sampling intervals of $\Delta t = 0.001$, $\Delta t = 0.01$ and $\Delta t = 0.05$, and used these estimates to simulate the parametrized system with time step Δt. The statistics of the resulting simulations are shown in figure 4.8 for the first approach and in figure 4.9 for the second. Both results strongly improve on the deterministic parametrization and are comparable to the results in (Crommelin and Vanden-Eijnden, 2008) and (Chorin and Lu, 2015). It appears that in the literature these methods have not previously been used for the Lorenz '96 model. The difference between the two different approaches is not very large, although the MZ method appears to be more robust to increasing Δt. It should be noticed that these approaches are very similar to the one based on conditional Markov chains studied in (Crommelin and Vanden-Eijnden, 2008), which models z_k^n as a Markov chain that depends on the state of x_k^n (and $x_k^n - x_k^{n-1}$), and computes the transition probabilities from data. The difference is that our approach starts from simple continuous models, and only a small number of parameters needs to be estimated.

To summarize, we have learned from studying the Lorenz '96 model that in order to capture important statistics of the model it is crucial not only to use stochastic terms in a parametrization, but also that it is important to look at 'memory effects' due to a lack of scale separation. This can be incorporated by very simple approximations, where the memory is parametrized by adding a single layer of OU processes, and which can be reconstructed from data. An interesting question that should be investigated further is whether this simple form also works in systems which have nonlinear coupling.

A way to potentially improve the results is by using multiple levels as discussed in section 3.3 to take into account that the process z_k is not really a one-dimensional Gaussian process, and this may especially be relevant in other systems where the coupling is nonlinear. Finally it may be interesting to use these parametrizations to study a Fokker-Planck equation for the system and compare to ensemble predictions of the Lorenz '96 model (Venturi et al., 2016).

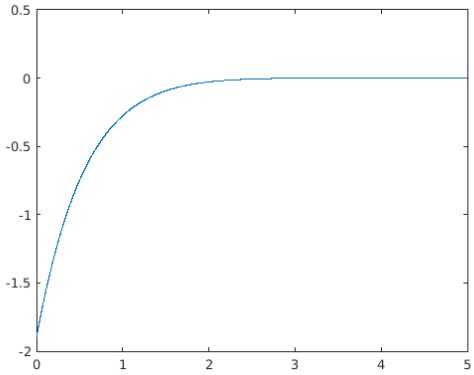


Figure 4.6: Plot of the estimate of the parametrized kernel against time.

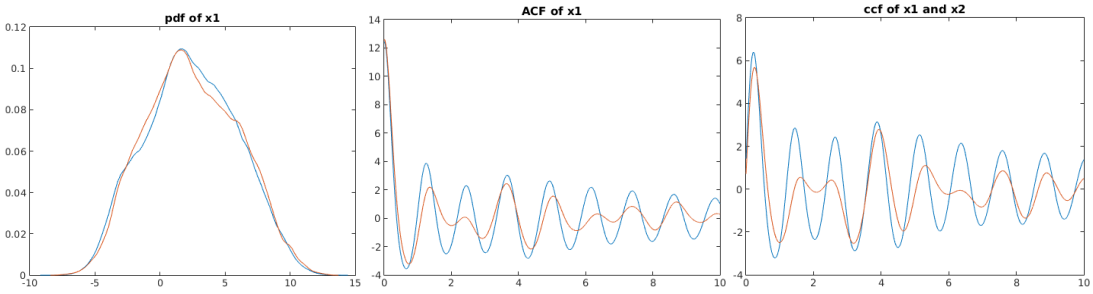


Figure 4.7: Comparison statistics of the Lorenz '96 model (blue) with the statistics of its deterministic parametrization.

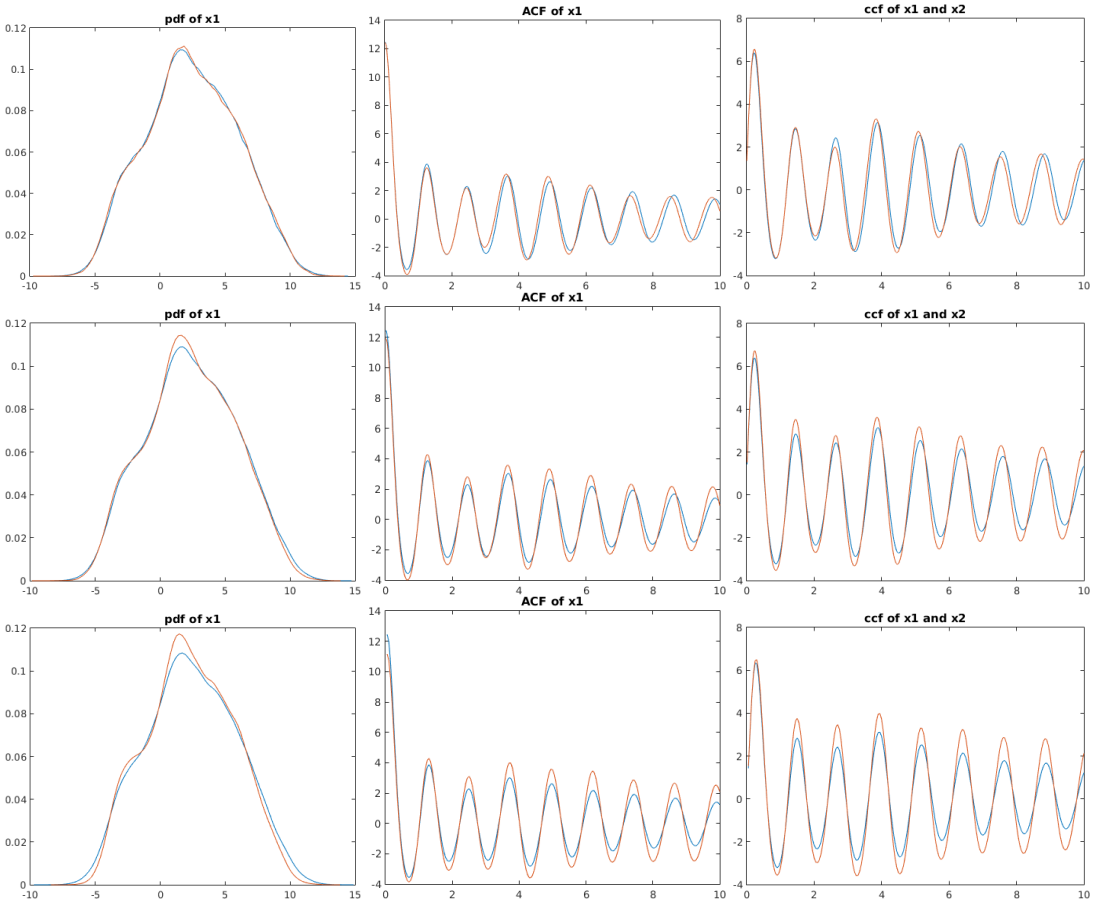


Figure 4.8: Comparison statistics of the Lorenz '96 model (blue) with the statistics of its stochastic parametrization by an OU process using the MTV method. The rows are for different timesteps, from upper to lower Δt is respectively 0.001, 0.01 and 0.05.

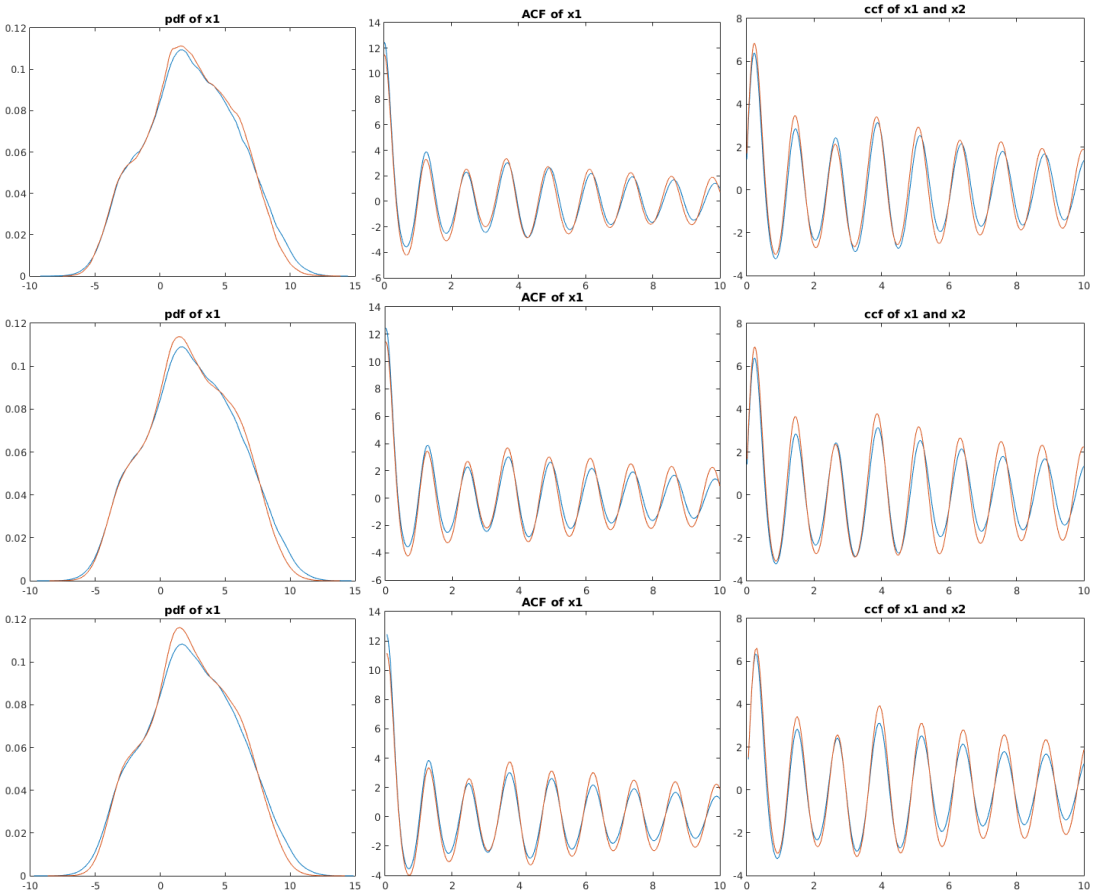
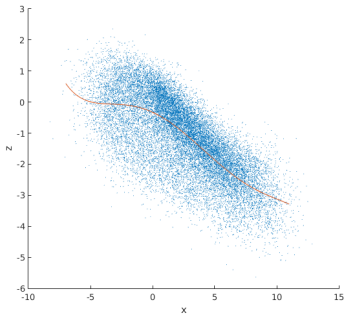
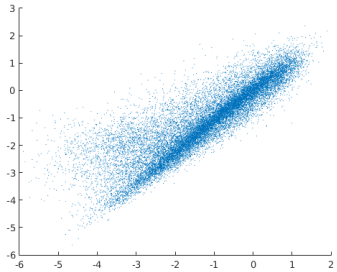


Figure 4.9: Comparison statistics of the Lorenz '96 model (blue) with the statistics of its stochastic parametrization by an OU process using the MZ method. The rows are for different timesteps, from upper to lower Δt is respectively 0.001, 0.01 and 0.05.



(a) A scatter plot of x_1 versus z_1, and the fifth order polynomial G fitted to the data.



(b) A scatter plot of z_1 versus the convolution of x_1 with K.

Often think of the rapidity with which things pass by and disappear, both the things which are and the things which are produced. For substance is like a river in a continual flow, and the activities of things are in constant change, and the causes work in infinite varieties; and there is hardly anything which stands still. And consider this which is near to thee, this boundless abyss of the past and of the future in which all things disappear. How then is he not a fool who is puffed up with such things or plagued about them and makes himself miserable? For they vex him only for a time, and a short time.

Marcus Aurelius, Meditations

5

Conclusion and discussion

In this thesis we have investigated model reduction of multi-scale systems through the use of projection operators. The Mori-Zwanzig formalism shows that in such a procedure memory and stochastic effects occur.

We investigated dimension reduction in two contexts. The first has been statistical mechanics, where the Kac-Zwanzig heat bath model is a paradigmatic model. This model represents a linearisation of a system consisting of a distinguished particle and a large number of degrees of freedom. Because the system is linear all these degrees of freedom (essentially the orthogonal dynamics in the Mori-Zwanzig formalism) can be solved and the program of dimension reduction can be performed explicitly. It can even be shown that if the frequencies and interactions of the heat bath are distributed in an appropriate way, the solutions of the heat bath model converge (weakly) to those of a stochastic integro-differential equation in the limit of infinitely many heat bath modes as shown in theorem 2.4.3.

The second context which has been discussed in which dimension reduction is of importance is in the context of the geophysics of atmosphere and ocean. A toy model for this that incorporates some features that are crucial for model reduction in this context is the Lorenz '96 model. Representing unresolved degrees of freedom in a model is called (stochastic) parametrization in this context. For the Lorenz '96 model we have shown that the memory effects are crucial in the (realistic) case where the scale separation in not very large, and that already a relatively simple approximation of the stochastic and memory terms (using Ornstein-Uhlenbeck processes) leads to good reproductions of some of the crucial statistics of the model.

Of course there many interesting issues remaining, which we have not covered due to

lack of time or understanding. Let us at least mention a few of these:

- An important aspect of the Mori-Zwanzig formalism is the choice on which variables one projects. The idea of approximating the orthogonal dynamics by a noise only works if one has indeed selected the slow/large scale variables to project onto. It is not always clear what these variables are, and methods to determine for a selection procedure are crucial.
- How well do the data-based approaches discussed here work when applied to real world data?
- Can the Mori-Zwanzig formalism be used to perform an error analysis of the approximation of a system by a stochastic system? In the case of the heat bath model we have been able to show that this can indeed be done. For general (nonlinear) systems this is much harder, and it is not clear what is possible.
- We have emphasized a Langevin equation approach (evolution of individual trajectories) to the problem. It is however also possible to look from a Fokker-Planck equation (evolution of densities) perspective, corresponding to ensemble predictions. How to do this is only clear when the system can be respresented by a finite dimensional Markovian system (as in section 2.6, but not necessarily linear). What can the Mori-Zwanzig formalism teach us about evolution of densities for distinctly non-Markovian systems?
- How is it possible to detect and determine nonlinearities in the memory kernel from data? And secondly, can it be motivated more thoroughly for chaotic systems using linear response theory that linear approximations to the kernel tend to be reasonably good?

Populaire samenvatting

Misschien ben je wel bekend met het 'vlinder effect', een metafoor voor onvoorspelbaarheid van bepaalde fenomenen, in dit geval het weer: een fladderende vlinder kan het verschil maken tussen het wel en niet ontstaan van een tornado aan de andere kant van de wereld. In eerste instantie werden deze krachten niet aan een vlinder maar een meeuw toegechreven:

> One meteorologist remarked that if the theory were correct, one flap of a sea gull's wings would be enough to alter the course of the weather forever. The controversy has not yet been settled, but the most recent evidence seems to favor the sea gulls. (Lorenz)

Deze metafoor betekent niet zozeer dat als we het weer willen voorspellen we vooral aan de andere kant van de wereld naar vlinders moeten gaan zoeken, maar dat het weer op een fundamentele wijze niet voorspelbaar is. Toch zien we elke dag in de krant voorspellingen voor het weer over een paar dagen, en die blijken meestal best goed te kloppen. Daar blijft het niet bij: we kunnen zelfs lezen over voorspellingen over veranderingen in het klimaat over meer dan vijftig jaar! Laten we nadenken wat we precies willen voorspellen als we 'het weer' (of het klimaat) willen voorspellen. Als we heel ambitieus zijn zouden we kunnen proberen om elk stukje atmosfeer en oceaan te beschrijven met een differentiaalvergelijking en (met een computer) op te lossen hoe deze zich in de toekomst gaat gedragen. Dit is te ambitieus, om twee redenen. De eerste reden is dat we niet weten in welke toestand het systeem precies begint. Om met het vlinder-effect te spreken: we zullen nooit kunnen bepalen op een enkel moment hoe alle vlinders ter wereld precies aan het fladderen zijn en dit maakt een groot verschil voor de precieze toestand na enige tijd. Ten tweede, zelfs al zouden we dit weten, kost het veel te veel tijd op een computer om dit uit te rekenen, zelfs als je de allerbeste supercomputer van het moment hebt. Twee grote beperkingen bij het simuleren van het weer zijn dus de onbekendheid van de begintoestand en het beperkte (computer)vermogen om een een grote hoeveelheid variabelen op te lossen.

Een manier om deze twee problemen in een keer aan te pakken, is door het aantal variabelen dat je wilt weten flink te beperken. Je zou bijvoorbeeld als variabelen alleen de luchtdruk, de wind en de temperatuur op een rooster kunnen nemen. De vraag is dan: hoe voer ik deze reductie precies het beste uit?

Een suggestie over hoe je hierover na kunt denken is gedaan in de (statistische) fysica. Hier spelen vergelijkbare problemen: een vloeistof bestaat uit een enorm aantal

deeltjes en hoewel je een idee hebt van de vergelijkingen die de tijdsontwikkeling beschrijven heb je dezelfde twee problemen die hierboven al werden genoemd: je kent de begintoestand niet precies en je hebt in de praktijk niet genoeg rekenkracht om het systeem te simuleren. Ook hier ben je niet echt geïnteresseerd in wat elk deeltje precies doet, maar vooral in globale eigenschappen van de vloeistof, zoals de temperatuur of de gemiddelde snelheid van de vloeistof. In deze context is er een wiskundige afleiding (het *Mori-Zwanzig formalisme*) bedacht die het mogelijk maakt om het aantal variabelen te reduceren. Om uit te leggen wat het resultaat inhoudt denken we aan de situatie dat we de beweging van een enkel deeltje in een vloeistof willen simuleren. De netto kracht die op het deeltje uitgeoefend wordt kan in drie termen onderscheiden worden:

1. De kracht die het deeltje ervaart door externe invloeden, en die alleen afhangt van de snelheid en positie van het deeltje, dit is dus de kracht die overblijft als je de rest van de vloeistof weg zou laten.
2. De kracht die de vloeistof zou uitoefenen, als de vloeistof niet beïnvloed werd door het deeltje.
3. De kracht die het deeltje ondervindt doordat het energie verliest bij botsingen met de vloeistofdeeltjes. Deze term kun je zien als een wrijvingsterm, en blijkt af te hangen van de toestand van deeltje op eerdere tijdstippen.

In de juiste omstandigheden kun je de tweede term vervangen door toevalsvariabelen: aangezien we de begintoestand van de vloeistof niet weten, gedraagt deze zich praktisch op een willekeurige manier (waarbij bepaalde statistische eigenschappen zoals het gemiddelde en standaard afwijkingen vastliggen), en we benaderen deze term door een *toevalsproces*, je kunt hierbij denken aan een soort ruis. De derde term die afhangt van de toestand van je deeltje in de verleden tijd, die een *geheugen* heeft, is vaak moeilijker te vatten. In de natuurkunde komt het vaak voor dat het verschil in schaalgrootte tussen het deeltje dat je volgt en de deeltjes waaruit de vloeistof bestaat heel groot is, en dit impliceert dat het geheugen heel 'ondiep' is, en we deze term kunnen benaderen met een wrijvingsterm die alleen van de toestand op het huidige tijdstip afhangt.

Om weer terug te komen op het problem van het voorspellen van het weer, hebben we nu geleerd dat als we maar een klein deel van de weervariabelen willen simuleren, we ook rekening moeten hoeden met een ruis en dat we de eerdere toestand van het systeem mee moeten nemen (geheugeneffecten). In klimaat en weermodellen zijn de schaalverschillen over het algemeen niet zo heel erg groot, en dat betekent dat geheugeneffecten minder ondiep zijn dan vaak in de statistische fysica wordt aangenomen. Het is zeer moeilijk, zo niet onmogelijk, om uit het oorspronkelijke weermodel een precieze vergelijking voor het gereduceerde model te vinden inclusief geheugen en ruis. Daarom is het ook belangrijk om methodes te ontwikkelen die op basis van beschikbare, empirische gegevens kunnen schatten hoe deze termen eruit zien.

Om terug te komen op de vlinder: we zagen dat er ergens een grote onzekerheid in het voorspellen van het weer zit. Het proces om een model te maken dat toeval meeneemt zorgt ervoor dat je deze onzekerheid expliciet meerekent bij je voorspellingen. Door met het toevalsmodel een hele serie voorspellingen te doen, kun je inschatten hoe waarschijnlijk bepaalde uitkomsten zijn, en hoe lang je voorspellingen betrouwbaar zijn.

Bibliography

S. Adelman and J.. Doll. Generalized Langevin equation approach for atom/solid-surface scattering: Collinear atom/harmonic chain model. *J. Chem. Phys.*, 1974.

G. Ariel and E. Vanden-Eijnden. A strong limit theorem in Kac-Zwanzig model. *Nonlinearity*, 2008.

H. Arnold, I. Moroz, and T. Palmer. Stochastic parametrizations and model uncertainty in the Lorenz '96 system. *Phil. Trans. R. Soc.*, 2013.

C. Beck, S. Lall, Liang T., and M. West. Model reduction, optimal prediction, and the Mori-Zwanzig representation of Markov chains. In *Proceedings of the IEEE Conference on Decision and Control*, 2009.

T. Berry and J. Harlim. Parametric reduced models for the nonlinear Schrödinger equation. *Phys. Rev. E*, 2015.

A. Chorin and O. Hald. Estimating the uncertainty in underresolved nonlinear dynamics. *Math Mech Solids.*, 2014.

A. Chorin and F. Lu. A discrete approach to stochastic parametrization and dimensional reduction in nonlinear dynamics. *Proc. Nat. Acad. Sci*, 2015.

A. Chorin, A. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proc. Nat. Acad. Sci*, 2000.

A. Chorin, O. Hald, and R. Kupferman. Optimal prediction with memory. *Physica D*, 2002.

D. Crommelin and E. Vanden-Eijnden. Subgrid scale parameterization with conditional Markov chains. *J. Atmos. Sci.*, 2008.

E. Darve, J. Solomon, and A. Kia. Computing generalized langevin equations and generalized Fokker-Planck equations. *Proc. Nat. Acad. Sci*, 2009.

A. Einstein. Uber die von der molekularkinetischen Theorie der Warme geforderte Bewegung von in ruhenden Flussigkeiten suspendierten Teilchen. *Ann. Phys.*, 1905.

C. Epstein and J. Schotland. The bad truth about Laplace's transform. *SIAM Rev.*, 2008.

L. Evans. *Partial Differential Equations*. Number 19 in Graduate Studies in Mathematics. American Mathematical Society, 2010.

I. Fatkullin and E. Vanden-Eijnden. A computational strategy for multi-scale systems with applications to Lorenz'96 model. *J. Comput. Phys.*, 2004.

G. Ford, M. Kac, and P. Mazur. Statistical mechanics of assemblies of coupled oscillators. *J. Math. Phys.*, 1965.

I. Gikhman and A. Skorokhod. *Introduction to the Theory of Random Processes*. W. B. Saunders Company, 1969.

D. Givon, O. Hald, and R. Kupferman. Existence proof for orthogonal dynamics and the Mori-Zwanzig formalism. *Israel Journal of Mathematics*, 2005.

N. Goel, S. Maitra, and E. Montroll. On the Volterra and other nonlinear models of interacting populations. *Rev. Mod. Phys.*, 1971.

O. Hald and R. Kupferman. Convergence of optimal prediction for nonlinear Hamiltonian systems. *Siam J. Numer. Anal.*, 2001.

O. Hald and R. Kupferman. Asymptotic and numerical analyses for mechanical models of heat baths. *J. Stat. Phys.*, 2002.

K. Hasselman. Stochastic climate models Part I. Theory. *Tellus*, 1976.

S. Haykin. *Adaptive Filter Theory*. Information and System Sciences Series. Prentice-Hall, 1986.

I. Horenko, C. Hartmann, C. Schutte, and F. Noe. Data-based parameter estimation of generalized multidimensional Langevin processes. *Phys. Rev. E*, 2007.

P. Imkeller and J. Storch. *Stochastic Climate Models*. Birkhauser, 2001.

R. Ishizaki, H. Mori, H. Tominaga, S. Kuroki, and N. Mori. The memory function and chaos-induced friction in the chaotic Henon-Heiles system. *Progr. Theoret. Phys.*, 2006.

G Kneller and K. Hinsen. Computing memory functions from molecular dynamics simulations. *J. Chem. Phys.*, 2001.

D. Kondrashov, M. Chekroun, and M. Ghil. Data-driven non-Markovian closure models. *Physica D*, 2015.

R. Kubo, H. Ichimura, and N. Hashitsume. *Statistical Physics II: Nonequilibrium Statistical Mechanics*. Springer-Verlag, 1985.

R. Kupferman. Fractional kinetics in Kac-Zwanzig heat bath models. *J. Stat. Phys.*, 2004.

R. Kupferman, A. Stuart, J. Terry, and P. Tupper. Long term behaviour of large mechanical systems with random initial data. *Stoch. Dynam.*, 2002.

E. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 1963.

E. Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, pages 1–18, 1995.

V. Lucarini and J. Wouters. Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach. *J. Stat. Phys.*, 2012.

A. Majda and X. Wang. *Nonlinear Dynamics and Statistical Theories for Geophysical Flows*. Cambridge University Press, 2006.

A. Majda, I. Timofeyev, and E. Vanden-Eijnden. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.*, 2001.

R. May. *Stability and Complexity in Model Ecosystems*. Princeton University Press, 1973.

H. Mori. Transport, collective motion, and Brownian motion. *Progr. Theoret. Phys.*, 1965a.

H. Mori. A continued-fraction representation of the time-correlation functions. *Progr. Theoret. Phys.*, 1965b.

G. Muller and V. Viswanath. *The Recursion Method: Applications to Many-Body Dynamics*. Number 23 in Lecture Notes in Physics. Springer, 1994.

G. Nastrom and K. Gage. A climatology of atmospheric wavenumber spectra of wind and temperature observed by a commercial aircraft. *J. Atmos. Sci.*, 1985.

T. Palmer and P. Williams. *Stochastic Physics and Climate Modelling*. Cambridge University Press, 2010.

G. Pavliotis. *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Texts in Applied Mathematics. Springer-Verlag New York, 2014.

G. Pavliotis and A. Stuart. *Multiscale methods: Averaging and Homogenization*. Texts in Applied Mathematics. Springer-Verlag New York, 2008.

J. Roerdink and A. Weyland. A generalized Fokker-Planck equation in the case of the Volterra Model. *Bull. Math. Biol.*, 1981.

R. Rubin. Momentum autocorrelation functions and energy transport in harmonic crystals containing isotopic defects. *Phys. Rev.*, 1963.

D. Schmitt and M. Schulz. Analyzing memory effects of complex systems from time series. *Phys. Rev. E*, 2006.

P. Stinis. A comparative study of two stochastic mode reduction models. *Phys. D*, 2005.

A. Stuart and J. Warren. Analysis and experiments for a computational model of a heat bath. *Journal of Statistical Physics*, 1999.

J. Thijsen. *Computational Physics*. Cambridge University Press, 1999.

N. Van Kampen. Remark on the Lotka-Volterra model. *J. Stat. Phys.*, 1974.

D. Venturi, H. Cho, and G. Karniadakis. The Mori-Zwanzig approach to uncertainty quantification. In *Handbook of Uncertainty Quantification*, 2016.

N. Verheul and D. Crommelin. Data-driven stochastic representations of unresolved features in multiscale models. *Comm. Math. Sci*, 2016.

D. Wilks. Effects of stochastic parametrizations in the Lorenz '96 system. *Q. J. Roy. Meteorol. Soc.*, 2005.

R. Zwanzig. Collision of a gas atom with a cold surface. *J. Chem. Phys*, 1960.

R. Zwanzig. Memory effects in irreversible thermodynamics. *Phys. Rev.*, 1961.

R. Zwanzig. Ensemble method in the theory of irreversibility. *J. Chem. Phys.*, 1973a.

R. Zwanzig. Generalized Verhulst laws for population growth. *Proc. Nat. Acad. Sci.*, 1973b.

R. Zwanzig. Problems in nonlinear transport theory. In L. Garrido, editor, *Systems far from Equilibrium*, pages 198–225. Springer, 1980.