



Brandenburgische  
Technische Universität  
Cottbus - Senftenberg

# Irreversible multi-scale diffusions: time scales and model reduction

von der Fakultät 1 - MINT -  
Mathematik, Informatik, Physik, Elektro- und  
Informationstechnik der Brandenburgischen Technischen  
Universität Cottbus-Senftenberg  
genehmigte Dissertation zur Erlangung des akademischen Grades  
eines Dr. rer. nat.

vorgelegt von

**Lara Neureither**

geboren am 17.10.1986 in Heidelberg.

**Vorsitzende:** Prof. Dr. Sabine Pickenhain

**Gutachter:** Prof. Dr. Carsten Hartmann

**Gutachter:** Prof. Dr. Tony Lelièvre

**Gutachter:** Prof. Dr. Christof Schütte

Tag der mündlichen Prüfung: 06.11.2019



# Summary

In this work, we consider non-reversible multi-scale stochastic processes, described by stochastic differential equations (SDEs), for which we review theory on the convergence behaviour to equilibrium and mean first exit times. Relations between these time scales for non-reversible processes are established, and, by resorting to a control theoretic formulation of the large deviations action functional, even the consideration of hypo-elliptic processes is permitted. The convergence behaviour of the processes is studied in a lot of detail, in particular with respect to initial conditions and temperature. Moreover, the behaviour of the conditional and marginal distributions during the relaxation phase is monitored and discussed as we encounter unexpected behaviour. In the end, this results in the proposal of a data-based partitioning into slow and fast degrees of freedom. In addition, recently proposed techniques promising accelerated convergence to equilibrium are examined and a connection to appropriate model reduction approaches is made. For specific examples this leads to either an interesting alternative formulation of the acceleration procedure or structural insight into the acceleration mechanism.

For the model order reduction technique of effective dynamics, introduced in [43], error bounds for non-reversible slow-fast stochastic processes are obtained. A comparison with the reduction method of averaging is undertaken, which, for non-reversible processes, possibly yields different reduced equations. For Ornstein-Uhlenbeck processes sufficient conditions are derived for the two methods (effective dynamics and averaging) to agree in the infinite time scale separation regime. Additionally, we provide oblique projections which allow for the sampling of conditional distributions of non-reversible Ornstein-Uhlenbeck processes.



# Zusammenfassung

In dieser Arbeit betrachten wir nicht-reversible mehrskalige stochastische Prozesse, beschrieben durch stochastische Differentialgleichungen (SDEs), für die wir die Theorie über das Konvergenzverhalten zum Gleichgewichtszustand und mittlere erste Austrittszeiten (MFETs) darlegen. Es werden Beziehungen zwischen diesen Zeitskalen für nicht-reversible Prozesse hergestellt, und durch Zurückgreifen auf eine kontrolltheoretische Formulierung des action functionals in der Theorie großer Abweichungen (large deviations theory) wird sogar die Betrachtung von hypoelliptischen Prozessen ermöglicht. Das Konvergenzverhalten der Prozesse wird sehr genau, insbesondere in Bezug auf Anfangsbedingungen und Temperatur, untersucht. Darüber hinaus wird das Verhalten der bedingten und marginalen Verteilungen während der Konvergenzphase beobachtet und diskutiert, da wir hierbei auf unerwartetes Verhalten stoßen. Aus der Aufteilung in bedingte und marginale Verteilung ergibt sich letztlich der Vorschlag einer datenbasierten Aufteilung in langsame und schnelle Freiheitsgrade. Darüber hinaus werden kürzlich vorgeschlagene Techniken, die eine beschleunigte Konvergenz zum Gleichgewicht versprechen, untersucht und es wird eine Verbindung zu geeigneten Modellreduktionsansätzen hergestellt. Für spezifische Beispiele führt dies entweder zu einer interessanten alternativen Formulierung der Beschleunigung oder zu struktureller Einsicht in den Beschleunigungsmechanismus.

Für die in [43] vorgestellte Modellreduktionstechnik der effektiven Dynamik werden Fehlerschranken für nicht-reversible schnell-langsame stochastische Prozesse hergeleitet. Es wird ein Vergleich mit der Reduktionsmethode averaging durchgeführt, die für nicht-reversible Prozesse möglicherweise andere reduzierte Gleichungen ergibt. Für Ornstein-Uhlenbeck-Prozesse werden hinreichende Bedingungen hergeleitet, unter denen die beiden Verfahren (effektive Dynamik und averaging) im Grenzwert unendlicher Zeitskalen-Trennung übereinstimmen. Zusätzlich geben wir schiefe Projektionen, die das Sampling von bedingten Verteilungen für nicht-reversible Ornstein-Uhlenbeck-Prozesse ermöglichen, an.



# Acknowledgements

First of all, I want to thank my supervisor Carsten Hartmann. Thank you for sharing your wisdom and knowledge with me, for being so open minded and allowing me to pursue all my interests, for having so much trust in me and giving me all the freedom I needed. Thank you for all the support and your kindness, I am very grateful for having been able to write a thesis in such a comfortable environment.

I also want to thank the commission members, Tony Lelièvre and Christof Schütte, very much for their willingness to review this thesis.

Moreover, let me thank the SFB1114 “Scaling Cascades in Complex Systems” for financial support.

I am also very grateful to Upanshu Sharma for teaching me many new things, being a great host and making research more fun. My thanks also go to Gabriel Stolz for the warm welcome to Paris, offering a great environment at the IHP and fruitful discussions. Further, I would like to thank my office mates at the FU Berlin, Stefanie Winkelmann and Sebastian Krumscheid for many discussions, sharing your viewpoints and making time much more enjoyable. This also applies to the other colleagues at FU Berlin and colleagues of the SFB1114, in particular, Nikki Vercauteren and Péter Koltai with whom I enjoyed many nice lunches, good wine and fun squash games. Also thanks to all the colleagues at the BTU Cottbus-Senftenberg, to my PhD fellows Lorenz Richter, Markus Strehlau and Jannes Quer. Further, I want to thank Nikolas Nüsken for various valuable discussions and comments at different stages during the PhD.

Besides all the colleagues I would also like to thank my friends and family. I am very grateful for having you. Thank you Chrissie for always being there whenever I needed it, especially during the hard times, and for always trusting in me. Thank you Uli for your unconditional support and Martin, Valle and Micha for naturally including me in your family. Thank you Oma Gretl for showing me that math is fun, you lay the foundations for all that came. Thank you Henriette for believing in me and encouraging me to always go one step further. Thank you all, especially, Merle, Marina, Jeannette, Leo and Josi, you are wonderful. Last but not least, thank you to my parents.





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Mathematical background</b>	<b>7</b>
2.1	Notation . . . . .	7
2.2	Itô stochastic processes . . . . .	9
2.3	Transfer operator, generator and evolution equations . . . . .	13
2.4	Invariant measures and (hypo-)ellipticity of the generator . . . . .	15
2.5	Reversibility and irreversibility . . . . .	20
<b>3</b>	<b>Different characterizations of time scales</b>	<b>25</b>
3.1	Exponential convergence to equilibrium for elliptic processes . . . . .	26
3.1.1	The Poincaré Inequality – convergence in $L^2_{\mu^{-1}}$ . . . . .	26
3.1.2	The Logarithmic Sobolev Inequality – convergence in relative entropy . . . . .	31
3.1.3	The Talagrand Inequality – convergence in Wasserstein distance . . . . .	34
3.1.4	The Bakry-Émery criterion . . . . .	35
3.2	Exponential convergence to equilibrium for hypo-elliptic OU-processes . . . . .	36
3.3	Mean first exit times for elliptic processes . . . . .	39
3.3.1	Diffusion in potential energy landscape . . . . .	40
3.4	Mean first exit times for hypo-elliptic processes . . . . .	42
3.4.1	The action functional and the controllability gramian . . . . .	42
3.4.2	OU-processes . . . . .	44
3.4.3	Underdamped Langevin dynamics . . . . .	46
3.4.4	A relation between the spectral gap and a MFET for underdamped Langevin dynamics . . . . .	48
<b>4</b>	<b>Model reduction</b>	<b>51</b>
4.1	Conditional expectations . . . . .	54
4.2	Error estimates . . . . .	55
4.2.1	Relative Entropy . . . . .	57
4.2.2	Wasserstein distance . . . . .	63
4.2.3	Estimates for the path measures . . . . .	66
4.3	Comparison of averaging and conditional expectations . . . . .	70
4.3.1	Review of averaging . . . . .	70

4.3.2	Some examples . . . . .	72
4.3.3	Sufficient conditions for the conformity of averaging and conditional expectations . . . . .	74
<b>5</b>	<b>Analytical results for OU-processes</b>	<b>77</b>
5.1	Comparison of averaging and conditional expectations . . . . .	77
5.1.1	Numerical comparison for $\varepsilon > 0$ . . . . .	81
5.2	Sampling from conditional distributions . . . . .	84
5.2.1	A covariance preserving projection . . . . .	85
5.2.2	A mean and variance preserving projection . . . . .	86
5.2.3	A scalar product with respect to which the projection becomes orthogonal . . . . .	91
<b>6</b>	<b>Applications and examples</b>	<b>93</b>
6.1	Convergence to equilibrium in relative entropy for OU-processes . . . . .	93
6.1.1	Dependence on temperature and initial conditions . . . . .	95
6.1.2	Structural properties of $a, b$ and $c$ . . . . .	97
6.1.3	Convergence of marginal and conditional distributions . . . . .	101
6.1.4	Multiple time scales: partitioning into slow and fast degrees of freedom . . . . .	104
6.2	Accelerating the convergence to equilibrium . . . . .	106
6.2.1	A general idea . . . . .	107
6.2.2	Hamiltonian systems with energy as coarse graining map . . . . .	109
6.2.3	Hamiltonian systems with soft constraints and coordinate projection as coarse graining map . . . . .	112
6.3	Mean first exit times for underdamped Langevin dynamics . . . . .	114
<b>7</b>	<b>Outlook</b>	<b>117</b>
	<b>Appendix A Background and theory</b>	<b>119</b>
A.1	Definition of the Itô integral . . . . .	119
A.2	PI and LSI . . . . .	120
	<b>Appendix B Applications and examples</b>	<b>123</b>
B.1	Convergence behaviour . . . . .	123
B.2	Accelerations . . . . .	124
	<b>Bibliography</b>	<b>127</b>

# INTRODUCTION

Stochastic processes play a prominent role in a large area of fields, such as statistical mechanics, financial models and climate models. The field of statistical mechanics, in particular, benefits from a well-developed mathematical theory. Thermodynamic averages therein constitute a quantity of great interest. On the one hand, they can be measured in experiments (and are thus also referred to as observables) and on the other hand, they can be computed from simulation data. Hence, they build a bridge between the model and reality. Simulations of complex systems, such as molecular dynamics, climate models, or socio-economic dynamics, have become increasingly important as they can provide insight and understanding of the underlying mechanisms. Understanding these intrinsic structures enables the researcher to manipulate the system's parameters systematically and thus to reach beyond the status quo in terms of the system's properties.

There are two main goals in the simulation of statistical mechanics applications. The first is the computation of thermodynamic averages, mentioned above, which, in the context of stochastic processes, amounts to computing averages of some function with respect to the stationary (equilibrium) measure of the process. This is, for example, the task of Markov Chain Monte Carlo (MCMC) algorithms and variants of these, leading to improved sampling speed of the stationary measure.

The second topic concerns the computation of so-called dynamic averages. These are averages over trajectories of the dynamics of a certain length. This length does not have to be a deterministic quantity, but can also be defined as the first exit time from a specified region of the state space. Usually, the interest here lies in determining exit times from so-called metastable regions, constituting parts of the state space which are almost stable, i.e., the process spends a lot of time in the region, but due to the noise eventually leaves it. Examples for such sudden transitions between metastable regions are climate transitions, e.g., the change from a glacial period to an interglacial period, or conformational changes in a molecule, possibly involving a change of the molecules properties.

For reversible processes, there are various well-developed methods towards realising these goals, dealing with challenges such as the high-dimensionality of the state space (for molecular dynamics  $\mathbb{R}^{3N}$  or  $\mathbb{R}^{6N}$ , where  $N$  is the number of atoms)

and complexity of the dynamics. Reversibility can intuitively be understood as the invariance of the process under time-reversal once it has reached its equilibrium. For the objective of sampling from high-dimensional probability distributions, there are many extensions of the MCMC algorithms (see e.g. [16]) building on the idea of finding the best possible transition function. In order to guarantee that the correct distribution is sampled, the transition function is usually required to satisfy a *detailed balance* condition with respect to the equilibrium distribution, meaning the process is reversible. Recently, various works have proposed to add a non-reversible perturbation to the reversible dynamics, under which the stationary measure is preserved, resulting in an acceleration of the convergence to equilibrium (see [27, 38, 47]) and also reduced variance (see [63]). The variance of the estimators requires attention due to the high-dimensionality and importance sampling provides another approach for dealing with this issue. Also when computing dynamic quantities, importance sampling finds its application (see e.g. [29]).

In molecular dynamics, if some prior knowledge about the reaction coordinate, i.e., the coordinate along which the molecule usually changes its conformation, is available, there are well-established tools such, as umbrella or blue moon sampling, artificially constraining the system to the reaction coordinate, enabling the computation of so-called free-energy differences, which are believed to provide fundamental insight of the system at hand (see e.g. [39], [18]). Both sampling methods rely on the reversibility of the underlying dynamics: umbrella sampling, which limits the dynamics to a certain region by adding a bias potential (which is often quadratic and therefore umbrella-shaped), accounts for the bias by appropriate reweighing which is straightforward for reversible dynamics opposed to non-reversible dynamics. Blue moon sampling, in turn, requires the computation of conditional expectations. There are algorithms sampling conditional distributions for reversible processes as well as for underdamped Langevin dynamics (see e.g. [48]) but general non-reversible processes have so far been unexplored. Other mentionable examples for non-reversible processes are the Lorentz-96 or stochastic triad system that mimic turbulent flows (see e.g. [50]).

Apart from the high-dimensionality, the multi-scale nature of the processes causes a huge computational effort for solving these problems through simulation. In molecular dynamics, for example, the movement of a molecule involves the smallest time scale of femtoseconds ( $10^{-15}$  seconds), whereas conformational changes occur on the time scale of seconds. In order to resolve all scales appropriately, one needs to choose a suitable time step for the fastest degrees of freedom, resulting in an infeasible computational effort for solving the high-dimensional system up to orders of seconds. Here, model reduction techniques come into play, equipping the user with a low-dimensional approximation of the dynamics. Also in this context, there are well-established methods for reversible processes, such as Markov state models (see e.g. [66]) and effective dynamics [43]. Markov state models comprise a discrete model capturing the slowest time scales of the process by exploiting the relation between metastability and the eigenvalues of the process' generator and this relation relies reversibility. For systems that admit a time scale separation parameter, the

theory of averaging and homogenization finds application (see e.g. [62]). Moreover, for reversible processes, there are results relating mean first exit times (MFETs) and eigenvalues of the generator, which correspond to relaxation time scales of the process, in a hierarchical manner (cf. [13]). Specifically, larger exit times correspond to a smaller absolute value of the eigenvalue, unfolding a lot of the intrinsic structure of the process.

### Contributions of this thesis and related work

In this thesis, we aim at gaining insight on these topics for *non-reversible* processes. We will often consider Ornstein-Uhlenbeck processes (OU-processes), which are linear stochastic differential equations (SDEs). This should be seen as a first step towards a refined analysis for general non-reversible processes, since OU-processes can be interpreted as a first order approximation (linearisation) of more complex dynamics in a metastable region.

This thesis is structured as follows. In order to make this thesis mostly self-contained, we provide the required mathematical framework in Chapter 2, which consists of basic theory of stochastic processes described by stochastic differential equations driven by Brownian motion.

In Chapter 3, we first review results on the equilibration time scale, i.e., the convergence to the stationary distribution. Herein, convergence takes place with respect to different metrics, which are encoded via appropriate functional inequalities (for a detailed overview on this topic see e.g. the textbook [9]). Even though these inequalities are usually formulated for reversible processes only, we will argue that they naturally extend to non-reversible processes as long as hypo-ellipticity, that is degeneracy of the diffusion, is excluded. For hypo-elliptic processes, we review a result by Arnold and co-authors [5] that yields exponential convergence to the stationary distribution for OU-processes. The study of processes with degenerate noise and their convergence to equilibrium has been promoted by Villani who also introduced the corresponding terminology of hypocoercivity, see e.g. [68]. In this regard, let us also mention the works [2], [20], [33] that are concerned with determining the rate of convergence to equilibrium for underdamped Langevin dynamics.

In the second part of Chapter 3, we report results on MFETs in the small noise limit by means of large deviations theory as introduced by Freidlin and Wentzel, see [26] for a textbook on this topic. Similar to the first part of Chapter 3, only elliptic processes, i.e., processes with non-degenerate noise, are allowed. We will bypass this limitation by turning to a control theoretic formulation of the problem, which has been introduced by Zabczyk in [69] (ideas in a similar direction were put forward e.g. in [25]). This enables us to link the MFET and equilibration time scale for hypo-elliptic OU-processes. Furthermore, we employ the control theoretic formulation of the underdamped Langevin equation, resulting in a description of MFETs which can be expressed in terms of the associated Hamiltonian by employing a result on the associated controllability function given in [56]. We show that, under

certain conditions, there is a relation between the spectral gap and the MFET from an appropriately chosen domain for the underdamped Langevin equations bearing the same flavour as the results of Bovier [13] for the overdamped case that we review in Section 3.3.1.

Chapter 4 is devoted to the topic of model reduction for non-reversible slow-fast processes. Within the class of model reduction techniques, we focus on the method of effective dynamics or conditional expectations, introduced in [43], for which pathwise error estimates for non-reversible dynamics were recently derived in [45]. We obtain estimates for the time  $t$  distributions of the processes, in relative entropy as well as in Wasserstein distance. Furthermore, we provide error estimates for the path-measures in relative entropy, which, to our knowledge, are the first results in this direction. Acquiring error estimates for path-measures is of great importance for assessing the performance of the reduced model to estimate dynamic averages, such as exit times.

In Section 4.3 we conduct a comparison of the previously considered method of effective dynamics and the averaging principle that is also applicable to slow-fast systems. Interestingly, averaging and conditional expectations agree for reversible processes which is not true for non-reversible processes. In particular, we show that for hypo-elliptic processes, the two methods can lead to qualitatively opposite behaviour. In Section 4.3.3 we derive a sufficient condition for the conformity of the two methods which is supported by formal arguments.

Chapter 5 offers analytical results for OU-processes, and its first part, Section 5.1, continues the comparison of averaging and conditional expectations. We obtain sufficient conditions for hypo-elliptic OU-processes in terms of drift and diffusion matrices for the two methods to agree in the limit as the slow and the fast scales become perfectly separated. We pursue the comparison in Section 5.1.1 by means of numerical experiments.

The last part of the chapter, Section 5.2, is devoted to the sampling of conditional distributions, which is required for the computation of conditional expectations. We construct an oblique projection – opposed to the orthogonal projection used for reversible dynamics – enabling for the sampling of conditional distributions for non-reversible OU-processes, when the condition consists of keeping certain degrees of freedom fixed. This is – to our knowledge – the first result in the direction of sampling conditional distributions for general non-reversible processes. The sampling of constrained underdamped Langevin dynamics, constituting a special case of constrained sampling for non-reversible processes, has been considered e.g. in [48].

In Chapter 6 we discuss some applications and examples, in particular related to the theory of Chapter 3. In the first part, Section 6.1, we begin with studying the convergence behaviour of OU-processes in relative entropy in dependence on initial conditions and temperature. To this end, we introduce a splitting into terms contributing to the relaxation of the mean and the covariance respectively and give a geometric interpretation of the plateaus which possibly occur in the decay towards equilibrium due to hypo-ellipticity. Moreover, we consider the factorization of the time  $t$  distribution into marginal and conditional distribution, revealing that

marginal and conditional distributions do not inherit the monotonicity property of the convergence behaviour. We give a mathematical justification for this observation. For slow-fast systems, we propose an identification of slow and fast degrees of freedom, based on the aforementioned splitting into conditional and marginal distributions, in a data driven fashion.

In Section 6.2 we continue the discussion about convergence, and investigate different acceleration approaches. In particular, the newly proposed acceleration techniques, based on adding a non-reversible perturbation, are examined (c.f. [47], [27], [38]). These techniques usually rely on a parameter, characterising the strength of the perturbation which can also be interpreted as a time scale separation parameter. Exploiting this interpretation we derive, for certain examples, the limit equations corresponding to an infinite acceleration. The limit equation leads in one case to an alternative proposal of a possible acceleration and in the other unveils structural insight into the acceleration. We conclude the chapter with an example for MFETs of underdamped Langevin equations in Section 6.3, as introduced in Chapter 3.

**Summary of contributions**

## Chapter 3

- Application of the control theoretic approach of Zabczyk to large deviations theory to underdamped Langevin equations, yielding a relation between the spectral gap and a MFET.
- Relation of the relaxation time scale and MFET for possibly degenerate OU-processes.

## Chapter 4 (joint work with Upanshu Sharma)

- Error estimates for effective dynamics applied to non-reversible processes with a time scale separation
  - \* for the time  $t$  distributions in relative entropy,
  - \* for the time  $t$  distributions in Wasserstein distance,
  - \* for the path measures in relative entropy.
- Comparison of averaging and effective dynamics.

A publication containing these results is in preparation.

## Chapter 5

- Comparison of averaging and effective dynamics for OU-processes. Derivation of sufficient criteria for the conformity of averaging and conditional expectations. This is joint work with Upanshu Sharma. A publication containing these results is in preparation.
- Derivation of an oblique projection enabling for sampling from conditional distributions for non-reversible OU-processes.

## Chapter 6

- Improvement of the understanding of the convergence behaviour for OU-processes, its dependence on initial condition and temperature using a splitting into contributions of mean and variance (contained in the publication [54]).
- Assessment of non-inheritance of the monotonicity property for convergences of conditional and marginal distributions to their invariant distributions (contained in the publication [54]).
- Study of different acceleration techniques for OU-process with respect to the acceleration parameter. This admits the derivation of an associated averaged dynamics leading to either an alternative proposal how to accelerate the convergence or insight on the effect of the applied acceleration.



# MATHEMATICAL BACKGROUND

## 2.1 Notation

We will use the following notational conventions: a vector  $x \in \mathbb{R}^n$  will always be column vector, i.e.,  $x \in \mathbb{R}^{n \times 1}$ . We denote by  $^T$  the (conjugate) transpose, such that  $x^T \in \mathbb{R}^{1 \times n}$  is the corresponding row vector. We will not make a notational distinction between the transpose and the conjugate transpose in case of complex valued vectors or matrices. For the standard euclidean scalar product we use the equivalent notations for  $x, y \in \mathbb{R}^n$

$$x \cdot y = x^T y.$$

For the euclidean norm we use the notation  $|x| = \sqrt{x \cdot x} = \sqrt{x^T x}$ , and its weighted version is  $|x|_A = \sqrt{x^T A x}$  for  $x \in \mathbb{R}^n, A \in \mathbb{R}^{n \times n}$ . For the  $L^2$  scalar product we write

$$\langle u, v \rangle = \int u \cdot v \, dx$$

for functions  $u, v : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and the  $L^2$  norm is denoted by  $|u|_{L^2} = \sqrt{\int u^2 \, dx} = \sqrt{\langle u, u \rangle}$ , where  $dx$  refers to Lebesgue measure, and its weighted version by  $|u|_{L^2_A} = \sqrt{\int u^2 \, d\nu} = \sqrt{\langle u, u \rangle_\nu}$ .

The gradient  $\nabla \in \mathbb{R}^{n \times 1}$  contains all partial derivatives and is also assumed to be a column vector. Accordingly, the divergence of a function  $u : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined by  $\nabla \cdot u = \sum_{i=1}^n \partial_{x_i} u_i$  assuming the partial derivatives exist. For matrix valued functions  $A : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$  we deviate from the convention, and  $\nabla \cdot A = (\nabla \cdot A_j)_{1 \leq j \leq m} \in \mathbb{R}^{m \times 1}$  is a column vector, where  $A_j$  is the  $j^{\text{th}}$  column. The Hessian matrix, containing all second derivatives, is denoted by  $\nabla^2 \in \mathbb{R}^{n \times n}$ . The Frobenius inner product of two matrices  $A, B \in \mathbb{R}^{n \times n}$  is denoted by  $A : B = \text{Tr}(A^T B)$ , where  $\text{Tr}(\cdot)$  is the trace of the matrix, and the Frobenius norm by  $|A|_F = \sqrt{A : A}$ . In general, we indicate the used norm and inner product by the corresponding subscript.

We use the same notation for a density and the corresponding measure, i.e., we write  $d\rho(x) = \rho(x) \, dx$ , where  $dx$  refers to Lebesgue measure.

We need the following classes of functions: denote by  $C^k(\mathcal{X}, \mathcal{Y})$  the class of  $k$  times continuously differentiable functions  $f : \mathcal{X} \rightarrow \mathcal{Y}$  and if  $\mathcal{Y} = \mathbb{R}$  we usually omit it. Accordingly denote by  $C_0$  continuous functions vanishing at infinity, by  $C_b$  continuous bounded functions, by  $L^2$  functions  $f$  with  $\int f^2 dx < \infty$  and by  $L^2_\nu$  functions  $f$  with  $\int f^2 d\nu < \infty$ .

Moreover, we use the following special symbols:

$\mathcal{H}(\nu \mu)$	relative entropy of $\nu$ with respect to $\mu$ ; see Definition 3.7
$\mathcal{R}(\nu \mu)$	relative Fisher information of $\nu$ with respect to $\mu$ ; see Definition 3.8
$\mathcal{W}(\nu \mu)$	second Wasserstein distance; see Definition 3.18
$\mathcal{L}$	infinitesimal generator of some dynamics
$\mathcal{L}_D, \mathcal{L}_{-D}$	infinitesimal generator of some dynamics, being of the form $\mathcal{L}_{\pm D} = -\Gamma \nabla V + \nabla \cdot \Gamma \pm D + \Gamma : \nabla^2$
$\rho_t$	time $t$ measure of the dynamics
$\mu$	invariant measure of the dynamics
$\mathcal{E}$	Dirichlet form; see Definition 3.1
$\text{Dom}(\mathcal{L})$	domain of definition for $\mathcal{L}$
$\mathcal{P}(\mathcal{X})$	space of probability measures on $\mathcal{X}$
$\mathcal{N}(m, \Sigma)$	normal distribution with mean $m$ and covariance $\Sigma$
$\Sigma_t, \Sigma_\infty$	covariance of an OU-process at time $t$ and in equilibrium respectively
$\text{Tr}(\cdot)$	trace operator
$\text{rk}(A)$	rank of a matrix $A$
$\lambda_{\min}(A), \lambda_{\max}(A)$	minimal and maximal eigenvalue of a matrix $A$

## 2.2 Itô stochastic processes

This chapter is devoted to give the basic notions of the formal language we need in this thesis. The main references we use are [58, 61], and we start with the definition of a stochastic process which is the mathematical object that describes the dynamics of interest.

**Definition 2.1.** Let  $(\Omega, \mathcal{F}, \mathbb{Q})$  be a probability space and  $\mathcal{T} \subset \mathbb{R}$ . A *stochastic process* is a function  $X : \Omega \times \mathcal{T} \rightarrow \mathcal{X}$ ,  $(\omega, t) \mapsto X_t(\omega)$ , where for each fixed  $t \in \mathcal{T}$  the function  $X_t : \Omega \rightarrow \mathcal{X}$  is a random variable. We call  $\Omega$  (equipped with the  $\sigma$ -algebra  $\mathcal{F}$ ) the *sample space* and  $\mathcal{X}$  (equipped with a  $\sigma$ -algebra  $\mathcal{E}$ ) the *state space*. The state space becomes a probability space via the measure  $\mathbb{P}$  induced by the random variable  $X_t$ , i.e.,  $\mathbb{P}(X_t \in E) = \mathbb{Q}(X_t^{-1}(E))$ ,  $\forall E \in \mathcal{E}$ .

Mostly, we will use the notation  $(X_t)_{t \in \mathcal{T}}$  or in short  $X_t$  for the stochastic process at hand. The state space  $\mathcal{X}$  is usually  $\mathcal{X} = \mathbb{R}^n$ . Note that, while the definition builds up on  $X_t$  being a random variable for each fixed  $t \in \mathcal{T}$ , it is also possible to fix  $\omega \in \Omega$  which yields a function  $X(\omega) : \mathcal{T} \rightarrow \mathcal{X}$  known as the *path* of the process  $(X_t)_{t \in \mathcal{T}}$  associated to  $\omega \in \Omega$ . We will also refer to a path as a *realisation* or *trajectory*.

Next, we introduce a specific stochastic process, that is the Brownian motion also known as the Wiener process. Brownian motion will be the stochastic driver of the dynamics we consider in this work. In molecular dynamics, Brownian motion can be thought of as the solvent surrounding the molecule leading to random motion of the molecule's atoms through interaction with the solution. The two names originate from, firstly, the botanist Robert Brown (1773-1858), who was the first to describe the specific behaviour of the diffusive motion of pollen in a solution, which he was studying under the microscope. The second contributor is Norbert Wiener (1894-1964), who first described this motion in a mathematical way and gave proofs for many of its properties.

**Definition 2.2** (Brownian motion).

- The *one-dimensional standard Brownian motion* is a stochastic process  $(B_t)_{t \in \mathbb{R}_+}$  taking values in  $\mathbb{R}$  with the following properties:
  - (i) The process starts almost surely at zero, i.e.  $\mathbb{P}(B_0 = 0) = 1$ .
  - (ii) The paths  $B_t(\omega)$  are continuous functions in  $t$  for almost all  $\omega \in \Omega$ .
  - (iii) For  $0 \leq t_0 < t_1 < \dots < t_n$ , the increments  $\Delta_i = B_{t_{i+1}} - B_{t_i}$ , for  $i = 0, \dots, n-1$ , are independent and normally distributed with  $\Delta_i \sim \mathcal{N}(0, t_{i+1} - t_i)$ .
- The *n-dimensional standard Brownian motion*  $(B_t)_{t \in \mathbb{R}_+}$  takes values in  $\mathbb{R}^n$  and is a collection of  $n$  independent one-dimensional standard Brownian motions  $(B_t^i)_{t \in \mathbb{R}_+}$ ,  $i = 1, \dots, n$ , i.e.,

$$B_t = (B_t^1, \dots, B_t^n)^T.$$

The existence of the standard Brownian motion can be shown by Kolmogorov's extension theorem together with Kolmogorov's continuity theorem, see e.g., Theorems 2.1.5 and 2.23 in [58]. There exist different constructions of the Brownian motion, such as the Karhunen-Loève expansion or random walk approximations, and also equivalent characterizations to the above definition such as the Lévy characterization of Brownian motion.

**Remark 2.3.** According to the second property the paths of the Brownian motion are continuous functions in time, i.e., the paths are elements of  $C([0, \infty), \mathbb{R}^n)$ . Hence, we can introduce the notion of the *path space* given by the space  $C([0, \infty), \mathbb{R}^n)$ . The measure on  $C([0, \infty), \mathbb{R}^n)$  induced by the Brownian motion is called the *Wiener measure*.

Brownian motion will be the stochastic driver of our dynamics, and hence we will need to describe integration with respect to the Wiener measure, i.e., define

$$\int_0^T g(\omega, s) dB_s(\omega). \quad (2.1)$$

The problem one encounters here is the irregularity of the paths of the Brownian motion such that one has to reach beyond the usual integral construction. Specifically, the paths are not differentiable, in fact they are not even of bounded variation, and hence we cannot employ the notion of a Riemann-Stieltjes integral directly for an integration against paths of Brownian motion<sup>1</sup>. In order to give an intuitive idea of how the Itô integral is constructed we proceed as follows. Assume that  $f \in C^2(\mathbb{R})$  is only a function of  $B_t$ , and define  $\Delta_n = \{t_0^n, \dots, t_{k_n}^n\}$  to be a partition of the interval  $[0, T]$ , i.e.,  $0 = t_0^n < \dots < t_{k_n}^n = T$ , which becomes finer for increasing  $n$ , i.e.  $\Delta_n \subset \Delta_{n+1}$ . By Taylor's Theorem there exists a  $\xi_i^n \in [B_{t_i^n}, B_{t_{i+1}^n}]$  such that

$$f(B_{t_{i+1}^n}) = f(B_{t_i^n}) + f'(B_{t_i^n})(B_{t_{i+1}^n} - B_{t_i^n}) + \frac{1}{2}f''(\xi_i^n)(B_{t_{i+1}^n} - B_{t_i^n})^2,$$

which, summing up from  $i = 0$  to  $i = k_n - 1$ , yields that

$$f(B_T) - f(B_0) = \sum_{i=0}^{k_n-1} f'(B_{t_i^n})(B_{t_{i+1}^n} - B_{t_i^n}) + \frac{1}{2} \sum_{i=0}^{k_n-1} f''(\xi_i^n)(B_{t_{i+1}^n} - B_{t_i^n})^2. \quad (2.2)$$

It can be shown – see Lemma A.3 – that the quadratic variation of the Brownian motion  $\langle B_t \rangle_t$  is

$$\langle B_t \rangle_t := \sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} (B_{t_i^n} - B_{t_{i+1}^n})^2 = t$$

almost surely, which in particular means that  $\langle B_t \rangle_t$  is of bounded variation, and together with  $f''$  being continuous this yields convergence to the Riemann-Stieltjes integral

$$\sum_{i=0}^{k_n-1} f''(\xi_i^n, t_i^n)(B_{t_{i+1}^n} - B_{t_i^n})^2 \xrightarrow{n \rightarrow \infty} \int_0^T f''(B_s) ds$$

<sup>1</sup>For more details see the Appendix A.1.

in  $L^2(\mathbb{P})$ . This implies that also the first term on the right hand side

$$\sum_{i=0}^{k_n-1} f'(B_{t_i^n})(B_{t_{i+1}^n} - B_{t_i^n})$$

has to converge, since the left hand side of the equation (2.2) is fixed and independent of  $n$ . In fact, it is the discretized version of the Itô integral  $\int_0^T f'(B_s)dB_s$  and yields the definition of (2.1) for  $g(\omega, s) = f'(B_s)$ .

**Definition 2.4.** Let  $f : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}$ ,  $(\omega, t) \mapsto f(\omega, t)$  be  $\mathcal{F} \times \mathcal{B}(\mathbb{R}^+)$  measurable and adapted to the filtration generated by  $B_t$ . Further, let  $\mathbb{E}(\int_0^t f(\cdot, s)^2 ds) < \infty$ . The *Itô stochastic integral*

$$I(t, \omega) = \int_0^t f(\omega, s) dB_s(\omega)$$

is defined as the limit of the Riemannian sum approximation

$$I(t, \omega) = \lim_{N \rightarrow \infty} \sum_{n=0}^{N-2} f(\omega, t_n)(B_{t_{n+1}}(\omega) - B_{t_n}(\omega))$$

in  $L^2$ , where  $t_n = n \delta t$  and  $\delta t$  is such that  $N \delta t = t$ . The definition analogously carries over to the  $n$ -dimensional case when  $f : \Omega \times [0, T] \rightarrow \mathbb{R}^{n \times m}$  and  $B_t$  is a  $m$ -dimensional standard Brownian motion.

The construction of the Itô integral builds on simple functions  $\phi : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ , that is, step-functions in time. The definition of the Itô integral for simple functions follows promptly, by noting that one can resort to the properties of the increments of the Brownian motion in this case. Making use of the fact that any square-integrable function can be approximated by simple functions then yields the general definition of the Itô integral. Note that  $I(t, \omega)$  is itself a stochastic process. In the following – for notational simplicity – we omit the explicit dependence on  $\omega$ . The Itô integral has many useful properties, such as *Itô isometry*, which states that

$$\mathbb{E} \left( \left( \int_0^t f(\omega, s) dB_s \right)^2 \right) = \mathbb{E} \left( \int_0^t f(\omega, s)^2 ds \right).$$

Also, we should note that due to the third property of the Brownian increments, it holds that

$$\mathbb{E} \left( \int_0^t f(\omega, s) dB_s \right) = 0.$$

**Remark 2.5.** There are other versions of stochastic integrals, such as, e.g., the *Stratonovich integral* defined by

$$I^S(t, \omega) = \int_0^t f(\omega, s) \circ dB_s(\omega) = \lim_{N \rightarrow \infty} \sum_{n=0}^{N-2} f \left( \omega, \frac{t_n + t_{n+1}}{2} \right) (B_{t_{n+1}}(\omega) - B_{t_n}(\omega)),$$

where  $t_n = n \delta t$ , and  $\delta t$  such that  $N \delta t = t$ . Here, the midpoint of each interval – instead of the left endpoint as in the case of the Itô integral – is used for the evaluation of  $f$ . This leads to different properties of the integral.

We can now give the definition of a stochastic differential equation (SDE) which describes the time evolution of a stochastic process and is the main object of interest throughout this thesis.

**Definition 2.6.** Let  $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $\gamma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$  be smooth functions and  $B_t$  a  $m$ -dimensional standard Brownian motion. We call

$$dX_t = b(X_t) dt + \gamma(X_t) dB_t, \quad X_0 = x_0 \quad (2.3)$$

an *Itô stochastic differential equation (SDE)* with *drift*  $b$  and *diffusion*  $\gamma$ . This stochastic differential equation should be understood as the differential form of the Itô stochastic process given by

$$X_t = x_0 + \int_0^t b(X_s) ds + \int_0^t \gamma(X_s) dB_s, \quad (2.4)$$

for which we stick to the “incremental” form (2.3) since the time derivative of the Brownian motion does not exist.

Assuming that, there exists  $C > 0$  such that

- $|b(x)| + |\gamma(x)|_F \leq C(1 + |x|), \quad \forall x \in \mathbb{R}^n,$
- $|b(x) - b(y)| + |\gamma(x) - \gamma(y)|_F \leq C|x - y|, \quad \forall x, y \in \mathbb{R}^n,$  and
- $\mathbb{E}(X_0^2) < \infty$

hold true, there exists a *strong* and *unique solution* to (2.3). The solution is called *strong*, if  $X_t$  is almost surely continuous and adapted to the filtration generated by  $B_t$ . There exists the notion of a weak solution, in which case the Brownian motion is not given a priori, but is part of the solution. A strong solution  $X_t$  of (2.3) is called *unique* if

$$\mathbb{P}(X_0 = \tilde{X}_0) = 1 \Rightarrow \mathbb{P}(X_t = \tilde{X}_t) = 1, \quad \forall t > 0,$$

for any other strong solution  $\tilde{X}_t$  of (2.3). From now on, we will assume that we fulfil the above assumptions such that we are dealing with processes given by unique and strong solutions of some SDE.

Associated to the SDE (2.3) introduce its *natural filtration*

$$\mathcal{F}_t = \sigma(\{X_s^{-1}(E) : s \leq t, E \in \mathcal{E}\}) \subset \mathcal{F},$$

where  $\sigma(\cdot)$  denotes the associated  $\sigma$ -algebra, which intuitively speaking contains all the information of the process  $X_t$  up to time  $t$ , given by all the pre-images of  $X_s$  for  $s \leq t$ . A process defined by (2.3) is *time-homogeneous*, since drift and diffusion

only depend on the state and not on time. Furthermore, it defines a *Markov process*, that is, its behaviour in the future  $t > s$  only depends on its current state  $X_s$  and not on all of its history contained in  $\mathcal{F}_s$ , i.e.,

$$\mathbb{P}(X_t \in E | \mathcal{F}_s) = \mathbb{P}(X_t \in E | X_s) \quad \forall 0 \leq s < t, \quad E \in \mathcal{E}.$$

Next, we state *Itô's formula*, which allows to compute functionals of an Itô process:

**Theorem 2.7** (Itô formula). *Let  $X_t \in \mathbb{R}^n$  be an Itô stochastic process,  $g(t, x) \in C^2([0, \infty) \times \mathbb{R}^n; \mathbb{R}^m)$  and define  $Y(t) = g(t, X_t)$ . Then  $Y(t)$  is an  $m$ -dimensional Itô process and its  $k^{\text{th}}$  component follows the SDE*

$$dY_k(t) = \frac{\partial g_k}{\partial t}(t, X_t) dt + \nabla_x g_k(t, X_t) \cdot dX_t + \frac{1}{2} (dX_t)^\top \nabla^2 g_k(t, X_t) dX_t,$$

where the rules  $dt \cdot dt = dt \cdot dB_t^i = dB_t^i \cdot dt = 0$  and  $dB_t^i dB_t^j = \delta_{ij} dt$  apply, with  $\delta_{ij}$  being the Kronecker delta and  $B_t^i$  being the  $i^{\text{th}}$  component of  $B_t$ .

**Example 2.8** (Ornstein-Uhlenbeck process). Employing Itô's formula and choosing the right function  $g$ , makes it easy to compute the solution of linear SDEs, also known as *Ornstein-Uhlenbeck processes* (*OU-processes*). That is, given the SDE

$$dX_t = AX_t dt + C dB_t, \quad X_0 = x_0,$$

where  $X_t \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{n \times n}$ ,  $C \in \mathbb{R}^{n \times m}$  and  $B_t$  is a standard  $m$ -dimensional Brownian motion, we use  $Y_t = g(t, X_t) = e^{-At} X_t$  which gives according to Itô's formula that

$$dY_t = -e^{-At} AX_t dt + e^{-At} dX_t = e^{-At} C dB_t.$$

Thus we find

$$Y_t - Y_0 = \int_0^t dY_s = \int_0^t e^{-As} C dB_s$$

and a multiplication from the left by  $e^{At}$  yields

$$X_t = e^{At} x_0 + \int_0^t e^{A(t-s)} C dB_s.$$

### 2.3 Transfer operator, generator and evolution equations

We now introduce the operators describing the propagation of probabilistic quantities - expectations and probability density functions - associated to the SDE (2.3) at hand. For this, we first introduce the *transition function*  $p: \mathcal{T} \times \mathcal{X} \times \mathcal{E} \rightarrow [0, 1]$  of the process  $X_t$  which gives the probability to transition from an initial state  $x$  to a part of the state space  $E$  in time  $t$ , that is,

$$p(t, x, E) = \mathbb{P}(X_t \in E | X_0 = x), \quad \forall t \in \mathbb{R}^+.$$

Since the process  $X_t$  defined by (2.3) is time-homogeneous, i.e.,  $\mathbb{P}(X_{t+s} \in E | X_s = x) = \mathbb{P}(X_t \in E | X_0 = x)$ , the transition function fully characterises the time-evolution of the process. Furthermore, we mention its fundamental properties: as a function of  $x$  alone,  $p(t, \cdot, E)$  is measurable  $\forall t \in \mathbb{R}^+, E \in \mathcal{E}$ , whereas as a function of  $E$  alone,  $p(t, x, \cdot)$  is a probability measure  $\forall t \in \mathbb{R}^+, x \in \mathbb{R}^n$ ; finally, the Chapman-Kolmogorov equation holds, i.e.,  $p(t+s, x, E) = \int p(s, y, E)p(t, x, dy)$ .

The transition function can be used to describe the time-evolution of quantities related to  $X_t$ . Define the *transfer operator*  $\mathcal{T}_t : C_b(\mathbb{R}^n) \rightarrow C_b(\mathbb{R}^n)$  by

$$(\mathcal{T}_t f)(x) := \int f(y)p(t, x, dy) = \mathbb{E}(f(X_t) | X_0 = x) \quad (2.5)$$

which has the interpretation of propagating the expected value  $f(x) = \mathbb{E}(f(X_0) | X_0 = x)$  in time  $t$  to  $(\mathcal{T}_t f)(x) = \mathbb{E}(f(X_t) | X_0 = x)$ . Associated to the transfer operator, we introduce the *infinitesimal generator*  $\mathcal{L}$  by

$$\begin{aligned} \mathcal{L}f(x) &:= \lim_{t \searrow 0} \frac{\mathbb{E}(f(X_t) | X_0 = x) - f(x)}{t} \\ &= b(x) \cdot \nabla f(x) + \frac{1}{2} \gamma(x) \gamma(x)^T : \nabla^2 f(x) \end{aligned} \quad (2.6)$$

which is defined to act on the set of functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  for which the limit exists and will be denoted by  $\text{Dom}(\mathcal{L})$ , and  $f \in C_0^2(\mathbb{R}^n)$  gives a sufficient criterion for this. Note also that  $\text{Dom}(\mathcal{L}) \subset L_\mu^2$  (cf. Section 1.7 in [9]). Clearly, by its definition it characterizes the time evolution of the conditional expectation  $\mathbb{E}(f(X_t) | X_0 = x)$  at time  $t = 0$ . Formula (2.6) follows by applying Itô's formula to the function  $f(X_t)$ , taking the conditional expectation for  $X_0 = x$ , and letting  $t \searrow 0$ .

Taking the above definition of the generator and employing it to express the time-evolution of the expectation of a function of the process (and not the conditional expectation), we find the expression for the operator describing the time-evolution of the associated probability density function. More specifically, let  $\rho_t$  be the probability distribution of the process  $X_t$  at time  $t$  starting from some initial distribution  $\rho_0$ , i.e.,  $x_0 \sim \rho_0$  meaning  $x_0$  is drawn from the distribution  $\rho_0$ . We will always assume the distributions to admit a smooth density with respect to Lebesgue measure and abuse notation in the sense that we write  $d\rho_t(x) = \rho_t(x) dx$ , i.e., we do not distinguish between densities and measures explicitly. Observe that, due to Chapman-Kolmogorov, there are two equivalent expressions for  $\mathbb{E}(f(X_t) | x_0 \sim \rho_0)$ , namely,

$$\begin{aligned} \mathbb{E}(f(X_t) | x_0 \sim \rho_0) &= \int \mathbb{E}(f(X_t) | X_s = x) \rho_s(x) dx \\ \text{and } \mathbb{E}(f(X_t) | x_0 \sim \rho_0) &= \int f(x) \rho_t(x) dx. \end{aligned}$$



With this we calculate for the first expression

$$\frac{d}{dt} \mathbb{E}(f(X_t)|x_0 \sim \rho_0) \Big|_{t=s} = \int \frac{\partial}{\partial t} \mathbb{E}(f(X_t)|X_s = x) \Big|_{t=s} \rho_s(x) dx = \int \rho_s(x) \mathcal{L}f(x) dx$$

whereas the second yields

$$\frac{d}{dt} \mathbb{E}(f(X_t)|x_0 \sim \rho_0) \Big|_{t=s} = \int f(x) \frac{\partial}{\partial t} \rho_t(x) \Big|_{t=s} dx =: \int f(x) \mathcal{A} \rho_s(x) dx.$$

Now, note that  $\mathcal{A}$  is in fact the  $L^2$ -adjoint of  $\mathcal{L}$  which we denote by  $\mathcal{L}^*$  and is thus computable. For the generator  $\mathcal{L}$  given by (2.6) the  $L^2$ -adjoint  $\mathcal{L}^*$  describing the propagation of probability densities in time reads

$$\mathcal{L}^* \rho_t(x) = -\nabla \cdot (b(x) \rho_t(x)) + \frac{1}{2} \nabla^2 : (\gamma(x) \gamma(x)^T \rho_t(x)). \quad (2.7)$$

The above observations are captured in the famous Kolmogorov-backward and forward (also known as Fokker-Planck) equations. The Kolmogorov-backward equation hereby comprises the action of the operator  $\mathcal{L}$ , that is, the time-evolution of expectations of the process for some given initial condition  $u(t, x) = \mathbb{E}(f(X_t)|X_0 = x)$ .

**Theorem 2.9** (Kolmogorov backward equation). *Let  $X_t$  be an Itô stochastic process in  $\mathbb{R}^n$ ,  $\mathcal{L}$  the associated generator,  $f \in C_0^2(\mathbb{R}^n)$  and  $u(t, x) := \mathbb{E}(f(X_t)|X_0 = x)$ . Then*

$$\begin{aligned} \frac{\partial u}{\partial t}(t, x) &= \mathcal{L}u(t, x), & (t, x) &\in \mathbb{R}^+ \times \mathbb{R}^n, \\ u(0, x) &= f(x), & (t, x) &\in \{0\} \times \mathbb{R}^n. \end{aligned}$$

For the time evolution of the density that gives the distribution of  $X_t$  at each time  $t$  we state the celebrated Kolmogorov forward/Fokker-Planck equation.

**Theorem 2.10** (Kolmogorov forward/Fokker-Planck equation). *Let  $X_t$  be an Itô stochastic process in  $\mathbb{R}^n$ , let the initial condition be distributed according to the density  $\nu_0$ , i.e.  $x_0 \sim \nu_0$ , assume that the distribution of  $X_t$  at time  $t$  has a density  $\rho(t, x) = \rho_t(x) \in C^{1,2}(\mathbb{R}^+ \times \mathbb{R}^n)$ . Then  $\rho_t$  solves*

$$\begin{aligned} \frac{\partial \rho_t}{\partial t}(x) &= \mathcal{L}^* \rho_t(x), & (t, x) &\in \mathbb{R}^+ \times \mathbb{R}^n, \\ \rho_0(x) &= \nu_0(x), & (t, x) &\in \{0\} \times \mathbb{R}^n. \end{aligned}$$

## 2.4 Invariant measures and (hypo-)ellipticity of the generator

Throughout this thesis, we will assume that there is a unique invariant state of the process at hand. Working with stochastic processes, the invariant state is given by an invariant density  $\mu$  – or invariant measure  $d\mu(x) = \mu(x) dx$  – i.e., a density

which remains unchanged under the dynamics of the process. Mathematically, this is rephrased as

$$\frac{\partial \mu}{\partial t} = 0 \quad \Leftrightarrow \quad \mathcal{L}^* \mu = 0.$$

We will use the names *invariant measure*, *equilibrium*, *stationary distribution* for  $\mu$  synonymously.

The assumption of a unique invariant measure equips us with the following structure for drift and diffusion:

**Proposition 2.11** (See Assumption 2.1 in [45]). *Assume that the SDE*

$$dX_t = b(X_t) dt + \sqrt{2}\gamma(X_t) dB_t \quad X_t \in \mathbb{R}^n$$

*admits a unique invariant measure  $d\mu = \frac{1}{Z} e^{-V(x)} dx$ , where  $Z^{-1} = \int d\mu$  is the normalization constant. Then there exists a function  $D : \mathbb{R}^n \rightarrow \mathbb{R}^n$  with  $\nabla \cdot (\mu D) = 0$  such that the drift  $b$  can be rewritten as*

$$b = -\Gamma(X_t) \nabla V(X_t) + \nabla \cdot \Gamma(X_t) + D(X_t), \quad \text{with } \Gamma = \gamma \gamma^T. \quad (2.8)$$

*Proof.* The proof is taken from [45] and the statement follows by considering the corresponding Fokker-Planck equation applied to the stationary density  $\mu$  which reads

$$0 = \nabla \cdot (-\mu b + \nabla \cdot (\Gamma \mu)) = \nabla \cdot (\mu(-b + \Gamma \cdot \nabla \ln(\mu) + \nabla \cdot \Gamma)).$$

Inserting (2.8) for  $b$  together with  $\nabla \ln(\mu) = -\nabla V$  we conclude by noting that

$$0 = -\nabla \cdot (\mu D).$$

□

On the other hand we might as well start from some given SDE

$$dX_t = (-\Gamma(X_t) \nabla V(X_t) + \nabla \cdot \Gamma(X_t) + D(X_t)) dt + \sqrt{2}\gamma(X_t) dB_t,$$

and ask under which conditions on the functions  $\Gamma$ ,  $V$  and  $D$  one finds a unique invariant measure. To this end, introduce the following assumptions.

**Assumption A1.**

- (i)  $V : \mathbb{R}^n \rightarrow \mathbb{R}$  is a *confining potential*, that is, it is smooth and grows sufficiently fast at infinity, such that  $e^{-V(x)} \in L^1$ ,
- (ii)  $\gamma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$  is smooth and  $\Gamma(x) = \gamma(x)\gamma(x)^T$  is uniformly bounded, i.e.,  $\exists 0 < m \leq M < \infty \in \mathbb{R}$  such that

$$\forall x_1, x_2 \in \mathbb{R}^n : m x_1^T x_1 < x_1^T \Gamma(x_2) x_1 < M x_1^T x_1,$$

- (iii)  $D$  is such that  $\nabla \cdot (D(x)e^{-V(x)}) = 0$ .

Rephrasing the Fokker-Planck equation again and writing  $\mu(x) = e^{-V(x)}/Z$  we find

$$\begin{aligned}
 \mathcal{L}^* f &= \nabla \cdot ((\Gamma \nabla V - \nabla \cdot \Gamma - D) f) + \nabla^2 : (\Gamma f) \\
 &= \nabla \cdot ((\Gamma \nabla V - \nabla \cdot \Gamma - D) f + \nabla \cdot (\Gamma f)) \\
 &= \nabla \cdot (f [\Gamma \nabla V - D + \Gamma \nabla \ln f]) \\
 &= \nabla \cdot \left( f \left[ -D + \Gamma \nabla \ln \frac{f}{\mu} \right] \right). \tag{2.9}
 \end{aligned}$$

Clearly,  $\mu = e^{-V(x)}/Z$ ,  $Z = \int e^{-V(x)} dx$  is an invariant measure of the process. Furthermore, under the given assumption that  $\Gamma > 0$  it is also unique as can be seen from the last equation and is also referred to as *ellipticity* of the corresponding generator. If  $\Gamma \geq 0$  the above equation might have several or non-smooth solutions due to the lack of irreducibility of the state space which in turn is due to the lack of smoothing properties of the generator  $\mathcal{L}$ . This can be circumvented by the concept of *hypo-ellipticity* which translates the question of smooth solutions of the above equation to dynamics with possibly degenerate noise.

Let us give the precise definitions of ellipticity and hypo-ellipticity:

**Definition 2.12.** Let  $b, \gamma \in C^\infty$ . The generator

$$\mathcal{L} = b(x) \cdot \nabla + \frac{1}{2} \Gamma(x) : \nabla^2, \quad \Gamma(x) = \gamma(x) \gamma^T(x)$$

is *elliptic* if

$$y^T \Gamma(x) y \geq 0 \quad \text{and} \quad y^T \Gamma(x) y = 0 \Leftrightarrow y = 0.$$

It is called *hypo-elliptic* if  $\mathcal{L}u \in C^\infty$  implies that  $u \in C^\infty$ .

In other words, the operator  $\mathcal{L}$  is hypo-elliptic if all solutions  $u$  to the equation  $\mathcal{L}u = f$  are smooth whenever  $f$  is smooth. Note, that we refer to a elliptic or hypo-elliptic process, if the corresponding generator is elliptic or hypo-elliptic. Now, by the seminal work of Lars Hörmander [35] there is a criterion which guarantees hypo-ellipticity. To this end, introduce the *Hörmander form* of the generator which reads

$$\mathcal{L} = \sum_i^n X_i^2 + X_0 + c,$$

where  $c \in \mathbb{R}$ ,  $X_i = \sum_{j=1}^n f_{ij} \partial_{x_j}$  consists of first order partial derivatives and  $X_i^2 = X_i(X_i)$  is the composition of the same, i.e.,

$$X_i^2 = \sum_{k=1}^n f_{ik} \partial_{x_k} \left( \sum_{j=1}^n f_{ij} \partial_{x_j} \right).$$

Identify  $X_i$  by the corresponding vector field  $X_i = (f_{i1}, f_{i2}, \dots, f_{in})^T$ . Writing the generator  $\mathcal{L} = b(x) \cdot \nabla + \frac{1}{2} \Gamma(x) : \nabla^2$  in this form would, e.g., yield

$$X_0 = b \cdot \nabla - \frac{1}{2} \sum_{k,j,i=1}^n \gamma_{jk} (\partial_{x_j} \gamma_{ik}) \partial_{x_i}, \quad X_i = \frac{1}{\sqrt{2}} \sum_{j=1}^n \gamma_{ji} \partial_{x_i}, \quad c = 0.$$

Furthermore, introduce the *commutator* of step 2 of  $X_i$  and  $X_j$  as

$$[X_i, X_j] = X_i X_j - X_j X_i,$$

the commutator of step 3 as  $[[X_i, X_j], X_k]$  and proceed analogously for any step  $l \in \mathbb{N}$ . Define the *Lie algebra* generated by  $X_i$  as the set of vector fields obtained as linear combination of the fields  $X_i$  and their commutators of any finite step. The condition for hypo-ellipticity is formulated in the following theorem.

**Theorem 2.13** (Hörmanders Theorem, see e.g. Theorem 12 in [14]). *Let*

$$\mathcal{L} = \sum_i^n X_i^2 + X_0 + c$$

*and assume that the coefficients  $f_{ij}$  of the vectorfields  $X_i$  are smooth. If the Lie algebra generated by the fields  $X_i$  has dimension  $n$  at every point of an open set  $\mathcal{O} \subset \mathbb{R}^n$ , then  $\mathcal{L}$  is hypo-elliptic in  $\mathcal{O}$ .*

Let us cite the following remark of the book [14] giving an assessment of the importance of the theorem for stochastic analysis: “The link between Hörmander’s operators and Kolmogorov equations is so relevant from the standpoint of stochastic analysis that a probabilistic proof of Hörmander’s theorem has been developed by Malliavin, [51], 1976, using the stochastic calculus of variations, also called Malliavin calculus. In this field, the hypo-ellipticity of Kolmogorov operator is interpreted as a result of regularity for the probability law of a process. An exposition of Malliavin calculus, with application to the proof of Hörmander’s theorem, can be found in the book [57].”

For the case of an OU-process the invariant measure is explicitly computable and also the Hörmander condition takes a concrete form, as the following example shows.

**Example 2.14** (OU-process continued). Consider the OU-process

$$dX_t = AX_t dt + C dB_t \quad X_0 = x_0 \in \mathbb{R}^n$$

$$\text{with solution} \quad X_t = e^{At} x_0 + \int_0^t e^{A(t-s)} C dB_s = e^{At} x_0 + \int_0^t e^{As} C dB_s.$$

Assume that  $x_0 \sim \mathcal{N}(m_0, \Sigma_0)$  where  $\Sigma_0 = 0$  would refer to the deterministic initial value  $x_0 = m_0$ . Computing the mean and covariance of  $X_t$  yields

$$\begin{aligned} m_t &= \mathbb{E}(X_t) = \mathbb{E}\left(e^{At} x_0 + \int_0^t e^{As} C dB_s\right) = e^{At} m_0, \quad \text{and} \\ \Sigma_t &= \text{Cov}(X_t) = \mathbb{E}\left(\left(e^{At}(x_0 - m_0) + \int_0^t e^{As} C dB_s\right)\left(e^{At}(x_0 - m_0) + \int_0^t e^{As} C dB_s\right)^T\right) \\ &= \mathbb{E}\left(e^{At}(x_0 - m_0)(x_0 - m_0)^T e^{A^T t}\right) + \mathbb{E}\left(\int_0^t e^{As} C C^T e^{A^T s} ds\right) \\ &= e^{At} \Sigma_0 e^{A^T t} + \int_0^t e^{As} C C^T e^{A^T s} ds. \end{aligned}$$

Aiming at a process with a well-defined invariant measure, it is immediately clear that the matrix  $A$  has to be Hurwitz, i.e., all eigenvalues should have strictly negative real part, since else both mean and variance explode for  $t \rightarrow \infty$ . Under the assumption that  $A$  is Hurwitz, it follows that the mean and the covariance of the invariant measure are well-defined and uniquely determined by the limit

$$\lim_{t \rightarrow \infty} m_t = 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} \Sigma_t = \int_0^\infty e^{As} C C^T e^{A^T s} ds =: \Sigma_\infty \geq 0.$$

There is an alternative equivalent characterization of the covariance matrix  $\Sigma_\infty$ , derived from noting that

$$\lim_{t \rightarrow \infty} \Sigma_t = \Sigma_\infty \quad \Leftrightarrow \quad \lim_{t \rightarrow \infty} (\dot{\Sigma}_t) = 0.$$

Now the time derivative can be expressed as

$$\begin{aligned} \dot{\Sigma}_t &= \frac{d}{dt} \left( e^{At} \Sigma_0 e^{A^T t} \right) + e^{At} C C^T e^{A^T t} \\ &= A e^{At} \Sigma_0 e^{A^T t} + e^{At} \Sigma_0 e^{A^T t} A^T + \int_0^t \frac{d}{ds} \left( e^{As} C C^T e^{A^T s} \right) ds + C C^T \\ &= A e^{At} \Sigma_0 e^{A^T t} + e^{At} \Sigma_0 e^{A^T t} A^T + \int_0^t A \left( e^{As} C C^T e^{A^T s} \right) + \left( e^{As} C C^T e^{A^T s} \right) A^T ds + C C^T \\ &= A \Sigma_t + \Sigma_t A^T + C C^T \end{aligned}$$

and thus the equivalent characterization of  $\Sigma_\infty$  is given by  $\Sigma_\infty$  being a solution to the *Lyapunov equation*

$$A \Sigma_\infty + \Sigma_\infty A^T = -C C^T.$$

The remaining question is how to ensure positive definiteness of  $\Sigma_\infty$  such that the invariant measure is not degenerate. This is handled by noting that positive definiteness of  $\Sigma_\infty$ , or equivalently hypo-ellipticity of the corresponding Fokker-Planck equation, is equivalent to either of the following conditions (cf. Theorem 1.2 in [70] and Lemma 2.3 in [5]).

#### Hypo-ellipticity conditions for OU-processes

- (i) The matrix pair  $(A, C)$  is controllable
- (ii) No eigenvector of  $A^T$  is in the kernel of  $C^T$ .
- (iii)  $\text{rk}(C, AC, A^2C, \dots, A^{n-1}C) = n$ ,  
where the matrix  $(C, AC, A^2C, \dots, A^{n-1}C) \in \mathbb{R}^{n \times nm}$ .

Without giving the proofs – which can be found in the above references – let us say some intuitive words about the conditions. To this end, consider the controlled ODE

$$\dot{x}^u(t) = Ax^u(t) + Cu(t), \quad x^u(0) = x_0^u, \quad A \in \mathbb{R}^{n \times n}, \quad C \in \mathbb{R}^{n \times m}, \quad m < n,$$

where  $u : \mathbb{R}^+ \rightarrow \mathbb{R}^m$  is referred to as control. Denote by  $x^u(t)$  the solution for a given control  $u$ . Controllability of the above dynamics means – we refer to the definition given in Chapter 1 of [70] – that there is a control  $u \in L^2([0, T], \mathbb{R}^m)$  such that starting from an arbitrary state  $x^u(0) = y \in \mathbb{R}^n$  any other point  $z \in \mathbb{R}^n$  can be reached in finite time, i.e.,  $x^u(T) = z$ ,  $T < \infty$ . This is equivalently referred to as controllability of the matrix pair  $(A, C)$  as given by (i). Indeed, controllability is equivalent to the other two conditions: they are rephrasing that, even though  $C$  may have rank  $m < n$ , the consecutive execution of  $A$  enables the system to reach any direction in  $\mathbb{R}^n$  and thus rephrases the property of irreducibility for Markov chains. For the corresponding SDE this means that even though  $C$  may be degenerate ( $m < n$ ), the noise will spread into the full phase space via the drift  $A$ , which is indeed what makes the invariant measure non-degenerate. This is precisely Hörmander's condition.

We summarise the discussed conditions for a unique invariant measure for OU-processes in the following assumption.

**Assumption A2.** The drift matrix  $A$  is Hurwitz and drift and diffusion  $(A, C)$  are controllable.

## 2.5 Reversibility and irreversibility

In this section we give the definitions of reversibility and irreversibility of which there exist many equivalent formulations. Just like in the previous section we will make things explicit for OU-processes.

Let  $X_t$  be a stochastic process with invariant measure  $d\mu(x) = \mu(x) dx$  and let  $X_t^- = X_{T-t}$ ,  $T \in [0, \infty)$  be the *time reversed process*. We say  $X_t$  is *reversible* iff starting from the invariant state, the statistics for going forward or backward in time are the same, i.e.,

$$\begin{aligned} X_0 \sim \mu \quad \Rightarrow \quad \mathbb{P}(X_t \in B, X_0 \in A) &= \mathbb{P}(X_{-t} \in B, X_0 \in A) \\ &= \mathbb{P}(X_0 \in B, X_t \in A), \quad \forall A, B \in \mathcal{E}, t > 0, \end{aligned}$$

where the last equality is due to the time homogeneity of the process. Making use of the transition function of  $X_t$  and noting that  $\mathbb{P}(X_t \in B, X_0 \in A) = \int_A p(t, x, B) d\mu(x)$  the above equation can be rephrased as what is known as the *detailed balance* condition

$$\int_A p(t, x, B) d\mu(x) = \int_B p(t, y, A) d\mu(y), \quad \forall A, B \in \mathcal{E}$$

or equivalently

$$p(t, x, dy) d\mu(x) = p(t, y, dx) d\mu(y), \quad \forall x, y \in \mathbb{R}^n.$$

In terms of the transfer operator the reversibility condition can be expressed as self-adjointness<sup>2</sup> in  $L_\mu^2$

$$\begin{aligned}\langle \mathcal{T}_t f, g \rangle_\mu &= \int (\mathcal{T}_t f)(x) g(x) d\mu(x) = \int \int f(y) p(t, x, dy) g(x) d\mu(x) \\ &= \int \int f(y) p(t, y, dx) g(x) d\mu(y) = \int f(y) (\mathcal{T}_t g)(y) d\mu(y) = \langle f, \mathcal{T}_t g \rangle_\mu.\end{aligned}$$

This can directly be translated to the generator  $\mathcal{L}$ , such that the reversibility condition then reads

$$\langle \mathcal{L} f, g \rangle_\mu = \langle f, \mathcal{L} g \rangle_\mu,$$

i.e., self-adjointness of  $\mathcal{L}$  in  $L_\mu^2$ . In the instance of a non-reversible process, for which the above equality does not hold, we can introduce the symmetric and anti-symmetric part of the generator as follows. Denote by  $\mathcal{L}^\dagger$  the adjoint of  $\mathcal{L}$  in  $L_\mu^2$  and decompose  $\mathcal{L}$  in its *symmetric* and *anti-symmetric* part via

$$\mathcal{L} = \mathcal{L}^s + \mathcal{L}^a, \quad \text{where } \mathcal{L}^s = \frac{\mathcal{L} + \mathcal{L}^\dagger}{2} \text{ and } \mathcal{L}^a = \frac{\mathcal{L} - \mathcal{L}^\dagger}{2}. \quad (2.10)$$

**Example 2.15.** Let us make the decomposition precise for a general SDE of the form

$$dX_t = (-\Gamma(X_t) \nabla V(X_t) + \nabla \cdot \Gamma(X_t) + D(X_t)) dt + \sqrt{2} \gamma(X_t) dB_t, \quad (2.11)$$

where  $\Gamma = \gamma \gamma^\top$ , for which we assume Assumption A1 to hold such that a unique invariant measure  $d\mu(x) = e^{-V(x)} / Z dx$  exists. Recall from (2.9) that we can write the Fokker-Planck operator as

$$\mathcal{L}_D^* f = \nabla \cdot \left( f \left[ -D + \Gamma \nabla \ln \frac{f}{\mu} \right] \right),$$

where we have introduced the subscript  $D$  for  $\mathcal{L}_D^*$  indicating that the function  $D$  enters the SDE with a “+” sign. Now, observe that

$$\begin{aligned}\mathcal{L}_D^*(f\mu) &= -\nabla \cdot (f\mu D) + \nabla \cdot (f\mu \Gamma \nabla \ln f) \\ &= -\mu D \nabla f - f \underbrace{\nabla \cdot (\mu D)}_{=0} + \nabla \cdot (\mu \Gamma \nabla f) \\ &= -\mu D \nabla f - \mu \Gamma \nabla V \cdot \nabla f + \mu (\nabla \cdot \Gamma) \cdot \nabla f + \mu \Gamma : \nabla^2 f \\ &= \mu \mathcal{L}_{-D} f,\end{aligned}$$

where  $\mathcal{L}_{-D}$  is the generator of the above SDE (2.11) with  $D$  being replaced by  $-D$ . In fact, the calculation is the proof to the following theorem:

---

<sup>2</sup>In fact we only show that  $\mathcal{T}_t$  is symmetric here. Self-adjointness requires the agreement of domains of definition for  $\mathcal{T}_t$  and its adjoint which requires a more careful analysis.

**Theorem 2.16** (Propositions 4.5 and 4.7 in [61]). *For an SDE of the form (2.11) fulfilling Assumption A1 it holds that*

$$\mu^{-1}\mathcal{L}_D^*(\mu f) = \mathcal{L}_{-D}f$$

or equivalently

$$\mathcal{L}_D^*(f) = \mu \mathcal{L}_{-D}(\mu^{-1}f).$$

Furthermore,  $X_t$  is reversible iff  $D = 0$  and hence

$$\mathcal{L}^s = (-\Gamma\nabla V + \nabla \cdot \Gamma) \cdot \nabla + \Gamma : \nabla^2 \quad \text{and} \quad \mathcal{L}^a = D \cdot \nabla.$$

The reversibility can be seen as follows:

$$\langle \mathcal{L}_D f, g \rangle_\mu = \langle \mathcal{L}_D f, \mu g \rangle = \langle f, \mathcal{L}_D^*(\mu g) \rangle = \langle f, \mu \mathcal{L}_{-D}g \rangle = \langle f, \mathcal{L}_{-D}g \rangle_\mu$$

and hence  $\mathcal{L}_D$  is self-adjoint in  $L_\mu^2$  if  $\mathcal{L}_D = \mathcal{L}_{-D}$  which is equivalent to  $D = 0$ . Summarising, the term  $D \neq 0$  alone makes the process non-reversible. We thus have the following equivalent conditions for reversibility:

$$X_t \text{ is reversible} \Leftrightarrow \mathcal{L} = (-\Gamma\nabla V + \nabla \cdot \Gamma) \cdot \nabla + \Gamma : \nabla^2 \Leftrightarrow \mathcal{L} \text{ is self-adjoint in } L_\mu^2.$$

Another way of defining reversibility is via the law of paths associated to the forward and the time-reversed process. That is, let  $X_t$  be the forward process,  $X_t^- = X_{-t}$  the time-reversed process, and  $X_0 \sim \mu$ . Given a time interval  $[0, T]$ , denote the path measure on  $C([0, T])$  of the forward and time-reversed process by  $\nu_{[0, T]}$  and  $\nu_{[0, T]}^-$ , respectively. Reversibility then means that the law of the forward process coincides with the law of the time-reversed process, i.e.,  $\nu_{[0, T]} = \nu_{[0, T]}^-$  almost surely. To put it differently, the *degree of irreversibility* can be measured by the *entropy-production*  $e_P$ , which is defined as the relative entropy between  $\nu_{[0, T]}$  and  $\nu_{[0, T]}^-$ , i.e.,

$$e_P(\nu_{[0, T]} | \nu_{[0, T]}^-) = \begin{cases} \int \log \left( \frac{d\nu_{[0, T]}}{d\nu_{[0, T]}^-} \right) d\nu_{[0, T]}, & \text{if } \nu_{[0, T]} \ll \nu_{[0, T]}^- \\ \infty, & \text{else.} \end{cases}$$

When starting from the invariant measure  $\mu$ , the *time-reversed process*  $X_t^- = X_{-t}$  is described by the SDE (cf. [30])

$$dX_t^- = b^-(X_t^-) dt + \gamma(X_t^-) dB_t, \quad (2.12)$$

where  $b^-(x) = -b(x) + \frac{2}{\mu(x)} \nabla \cdot (\gamma(x)\gamma(x)^T \mu(x))$ .

So yet another way of telling whether a process is reversible or not is by checking whether  $b^- = b$ .

**Example 2.17.** Let us calculate  $b^-$  for the general class of SDEs given by

$$dX_t = -\Gamma(X_t)\nabla V(X_t) + \nabla \cdot \Gamma(X_t) + D(X_t) dt + \sqrt{2}\gamma(X_t) dB_t,$$



where  $\Gamma(X_t) = \gamma(X_t)\gamma(X_t)^\top$ , as always suppose Assumption A1 holds and denote the invariant measure by  $d\mu(x) = 1/Z e^{-V(x)} dx$ . Let us first calculate

$$\begin{aligned} (\nabla \cdot (\Gamma(x)\mu(x)))_i &= \sum_{j=1}^n \frac{d}{dx_j} (\Gamma_{ij}(x)\mu(x)) \\ &= \mu(x) \sum_{j=1}^n \left( \frac{d}{dx_j} \Gamma_{ij}(x) - \Gamma_{ij}(x) \frac{d}{dx_j} V(x) \right) \\ &= \mu \left( (\nabla \cdot \Gamma(x))_i - (\Gamma(x)\nabla V(x))_i \right), \end{aligned}$$

and hence

$$\begin{aligned} b^-(x) &= -(-\Gamma(x)\nabla V(x) + \nabla \cdot \Gamma(x) + D(x)) + 2\nabla \cdot \Gamma(x) - 2\Gamma(x)\nabla V(x) \\ &= -\Gamma(x)\nabla V(x) + \nabla \cdot \Gamma(x) - D(x) \\ &= b(x) - 2D(x), \end{aligned}$$

that is, in agreement with the previous result of Theorem 2.16 we find that  $D$  is the quantity that contains the non-reversibility.

For the OU-processes we can derive explicit expression for  $D$  in terms of the drift matrix  $A$  and the invariant measure, which is fully characterised by its covariance  $\Sigma_\infty$ .

**Example 2.18** (OU-process continued). First note, that the drift matrix  $A$  can be rewritten using the Lyapunov equation as

$$\begin{aligned} A\Sigma_\infty + \Sigma_\infty A^\top &= -CC^\top \\ \Leftrightarrow A &= -(\Sigma_\infty A^\top + CC^\top)\Sigma_\infty^{-1} = -(\Sigma_\infty A^\top + \frac{1}{2}CC^\top + \frac{1}{2}CC^\top)\Sigma_\infty^{-1} \end{aligned}$$

Further, introduce the skew symmetric matrix  $J$  as

$$J = \frac{1}{2}(-\Sigma_\infty A^\top + A\Sigma_\infty) = -\Sigma_\infty A^\top + \frac{1}{2}(A\Sigma_\infty + \Sigma_\infty A^\top) = -\Sigma_\infty A^\top - \frac{1}{2}CC^\top.$$

Thus, the drift can be written as

$$AX = -\frac{1}{2}(2J + CC^\top)\Sigma_\infty^{-1}X = -\frac{1}{2}CC^\top\nabla V(X) + J\nabla V(X)$$

where  $V(X) = \frac{1}{2}X^\top\Sigma_\infty^{-1}X$  and  $D(X) = -J\Sigma_\infty^{-1}X$ .



## DIFFERENT CHARACTERIZATIONS OF TIME SCALES

Being concerned with multi-scale processes in this work we should first clarify how we define different scales of a process. Let us consider an  $\mathbb{R}^n$ -valued stochastic process  $(X_t)_{t \geq 0}$ . We are interested in two different notions of time scales, the first being the *equilibration time scale* and the second being the *mean first exit time (MFET)*.

We define the equilibration time scale by the exponential rate which describes the relaxation of the process to its equilibrium or invariant measure. More precisely, let  $\rho_t$  be the distribution of  $X_t$  at time  $t$  corresponding to some initial datum  $\rho_0 = \nu_0$  and denote the invariant measure by  $\mu$ . The relaxation behaviour of  $\rho_t$  to  $\mu$  can be phrased as the question asking whether there exists a  $\lambda > 0$  such that

$$\|\rho_t - \mu\| \leq e^{-\lambda t} \|\rho_0 - \mu\|$$

holds. The apparent questions are

- In which norm  $\|\cdot\|$  does the above hold (if, at all)?
- How can the largest possible  $\lambda > 0$  be determined?

We will address these questions later in more detail, but first introduce the second notion of time scale that is of interest to us. Assume for now that we start from some deterministic initial condition, i.e.,  $X_0 = x_0 \in \mathcal{D} \subset \mathbb{R}^n$  and we are interested in how long on average it takes for the process to leave the domain  $\mathcal{D}$ , that is, we ask, what is the *mean first exit time (MFET)*

$$\tau = \mathbb{E}(\inf t > 0 : X_t \notin \mathcal{D} | X_0 = x_0).$$

The mean first exit time can directly be linked to the metastability of the associated process.  $X_t$  being metastable intuitively means that there are parts of the state space in which  $X_t$  remains for a very long time. Here “very long” is relative to the time scale of the fluctuations of the process, and hence the respective mean first exit time from these domains will be very large. At the same time, if  $\tau$  is large for some region  $\mathcal{D}$ , then  $X_t$  spends a long time in this region on average, and hence

we have discovered a metastable region of the process.

Clearly, there is an intuitive connection between the two time scales. If the process is metastable, there will be regions for which the mean first exit time is large and at the same time the relaxation to the invariant measure is also slow as it takes very long for the process to explore the whole state space.

This chapter is meant to give an overview of existing theory on these time scales, the connections between them and the problems which arise due to non-reversibility of the process and in particular non-ellipticity of the generator. It is organised as follows. The first section reviews results on equilibration time scales for elliptic processes corresponding to different norms. We discuss their relation and conditions to guarantee exponential convergence. Further, we state results by Arnold et. al [5] for hypo-elliptic OU-processes. In the second section we review results on MFETs for elliptic processes and their relation to the equilibration time scale. Furthermore, we state results for hypo-elliptic OU-processes by Zabczyk [70] who employs a formulation of control theory. This technique enables us to relate MFETs and the equilibration time scale for OU-processes and for underdamped Langevin equation in a new fashion which we present in the last part of the chapter.

### 3.1 Exponential convergence to equilibrium for elliptic processes

In this section we give the definitions of functional inequalities, namely the Poincaré and Logarithmic Sobolev Inequality, which are commonly used tools to deduce exponential convergence to equilibrium. For both inequalities we make a distinction between what we call the dynamic version and static version. The dynamic version is formulated for a pair  $(\mu, \mathcal{L})$  with  $\mu$  being the unique invariant measure of a dynamics with generator  $\mathcal{L}$ . The fact that  $(\mu, \mathcal{L})$  satisfies one of the inequalities with a given constant  $\alpha$  can then be shown to be equivalent to exponential convergence with rate  $2\alpha$  in some sense.

Alternatively, one can formulate the inequalities as a property of the measure  $\mu$  alone, which in fact corresponds to the dynamic version with gradient dynamics ( $\Gamma = I_{n \times n}$ ) unveiling an interpretation in terms of concentration of measures.

#### 3.1.1 The Poincaré Inequality – convergence in $L^2_{\mu^1}$

We begin this section by giving the definition of a *Dirichlet form* which is pivotal for the Poincaré inequality. To this end, assume we are given a general diffusion described by (2.11), with  $\Gamma = \Gamma(x) \geq \lambda_{\min}(\Gamma) > 0$ , which admits a unique invariant measure  $d\mu(x) = \mu(x) dx$  such that  $\mu(x) \in L^1$ . Denote as usual the corresponding generator by  $\mathcal{L}$  given by

$$\mathcal{L} = (-\Gamma \nabla \nabla + \nabla \cdot \Gamma + D) \cdot \nabla + \Gamma : \nabla^2. \quad (3.1)$$

**Definition 3.1.** The *Dirichlet form* associated to  $\mathcal{L}$  is defined by

$$\mathcal{E}_{\mathcal{L}}(f) := \langle -\mathcal{L}f, f \rangle_{\mu}, \quad f \in \text{Dom}(\mathcal{L}). \quad (3.2)$$

Actually, the Dirichlet form has a larger domain of definition, see Sections 1.7 and 3.1.4 in [9] being defined by the corresponding limit, but we keep it as simple as possible here.

First note that by the decomposition of the generator (2.10) in its symmetric and anti-symmetric part (see (2.10)) with respect to  $\mu$  we have

$$\langle \mathcal{L}f, f \rangle_{\mu} = \langle \mathcal{L}^s f, f \rangle_{\mu} + \langle \mathcal{L}^a f, f \rangle_{\mu} = \langle f, \mathcal{L}^s f \rangle_{\mu} - \langle f, \mathcal{L}^a f \rangle_{\mu}$$

which implies that  $\langle \mathcal{L}^a f, f \rangle_{\mu} = 0$ . Hence, the Dirichlet form only depends on the symmetric part of the generator, i.e.,

$$\mathcal{E}_{\mathcal{L}}(f) = \mathcal{E}_{\mathcal{L}^s}(f)$$

which can be rewritten using integration by parts and  $d\mu(x) = e^{V(x)}/Z dx$  as

$$\begin{aligned} \mathcal{E}_{\mathcal{L}}(f) &= \int f(x) (\Gamma(x) \nabla V(x) \cdot \nabla f(x) - (\nabla \cdot \Gamma(x)) \cdot \nabla f(x) - \Gamma(x) : \nabla^2 f(x)) e^{-V(x)} dx \\ &= - \int f(x) \nabla \cdot (e^{-V(x)} \Gamma(x) \nabla f(x)) dx \\ &= \int \nabla f(x) \cdot \Gamma(x) \nabla f(x) e^{-V(x)} dx = \langle \nabla f, \Gamma \nabla f \rangle_{\mu}. \end{aligned} \quad (3.3)$$

The subsequent lemma reveals an even more appealing property of the Dirichlet form, namely its relation to the time derivative of the variance with respect to the measure  $\mu$ . To this end, introduce the *variance of  $f$  with respect to  $\mu$*

$$\mathbb{V}ar_{\mu}(f) = \int \left( f - \int f d\mu \right)^2 d\mu,$$

for any function  $f \in L^2_{\mu}$ .

**Lemma 3.2.** *Let  $f \in L^2_{\mu}$  and define  $f_t = \mathcal{T}_t f$ , i.e.,  $f_t$  solves  $\frac{\partial}{\partial t} f_t = \mathcal{L} f_t$ ,  $f_0 = f$ . Then*

$$\frac{d}{dt} \mathbb{V}ar_{\mu}(f_t) = -2\mathcal{E}_{\mathcal{L}}(f_t).$$

*Proof.* Since  $\mu$  is the invariant measure associated to  $\mathcal{L}$  we have, recalling that  $p(t, x, y)$  is the transition function,

$$\int f_t(x) d\mu(x) = \int \int f(y) p(t, x, y) \mu(x) dy dx = \int f(y) d\mu(y),$$

i.e., the mean of  $f_t$  with respect to  $\mu$  is constant in time. Computing the time derivative of the variance immediately yields the result:

$$\begin{aligned} \frac{d}{dt} \mathbb{V}ar_{\mu}(f_t) &= \int 2 \left( f_t - \int f d\mu \right) \frac{\partial}{\partial t} \left( f_t - \int f d\mu \right) d\mu, \\ &= 2 \int \left( f_t - \int f d\mu \right) \mathcal{L} \left( f_t - \int f d\mu \right) d\mu, \\ &= -2\mathcal{E}_{\mathcal{L}}(f_t), \end{aligned}$$

using that  $\mathcal{E}_{\mathcal{L}}(f_t + c) = \mathcal{E}_{\mathcal{L}}(f_t)$  for any constant  $c$  according to (3.3).  $\square$

**Remark 3.3.** Note that the very same result hold true if  $f_t$  was instead the solution to  $\frac{\partial}{\partial t}f_t = \mathcal{L}_{-D}f_t$ ,  $f_0 = f$ , where  $\mathcal{L}_{-D}$  is (3.1) with  $+D$  being replaced by  $-D$ . This is due to  $\mathcal{E}_{\mathcal{L}}(f)$  only depending on the symmetric part of  $\mathcal{L}$  as discussed above. In particular, the time derivative of the variance is determined by the reversible part of the generator only.

Let us now state the definition of a Poincaré inequality.

**Definition 3.4.** Let  $\mu \in \mathcal{P}(\mathbb{R}^n)$  be the unique invariant measure associated to the dynamics with generator  $\mathcal{L}$ . We say  $\mu$  satisfies a *dynamic Poincaré Inequality (PI)* with constant  $\alpha_{\text{PI}} > 0$  if and only if for all  $f \in \text{Dom}(\mathcal{L})$  it holds that

$$\alpha_{\text{PI}} \mathbb{V}ar_{\mu}(f) \leq \mathcal{E}_{\mathcal{L}}(f). \quad (3.4)$$

Due to the relation of the variance and the Dirichlet form established in Lemma 3.2 the assumption of a dynamic PI instantly yields exponential convergence of the variance to zero for any observable  $f \in L^2_{\mu}$ . Furthermore, this also entails exponential convergence of the time  $t$  density  $\rho_t$  to its equilibrium  $\mu$  in  $L^2_{\mu^{-1}}$  as the next theorem shows.

**Theorem 3.5** (Theorem 4.2.5 in [9]). *Let  $\mu \in \mathcal{P}(\mathbb{R}^n)$  be the unique invariant measure of the dynamics associated to  $\mathcal{L}$  and let  $\rho_t$  be the corresponding time  $t$  density, i.e.,  $\rho_t$  solves  $\frac{\partial}{\partial t}\rho_t = \mathcal{L}^*\rho_t$  for some given initial datum  $\rho_{t=0} = \rho_0$ . Then the following statements are equivalent:*

- (i)  $\mu$  satisfies a dynamic PI.
- (ii)  $\forall f \in L^2_{\mu}, \forall t > 0$  it holds that  $\mathbb{V}ar_{\mu}(f_t) \leq e^{-2\alpha_{\text{PI}}t} \mathbb{V}ar_{\mu}(f)$ , where  $f_t = \mathcal{T}_t f$ .
- (iii)  $\forall \rho_0 \in L^2_{\mu^{-1}}, \forall t > 0$  it holds that  $|\rho_t - \mu|_{L^2_{\mu^{-1}}}^2 \leq e^{-2\alpha_{\text{PI}}t} |\rho_0 - \mu|_{L^2_{\mu^{-1}}}^2$ .

*Proof.* The proof of (i)  $\Leftrightarrow$  (ii) is very easy:

- (i) $\Rightarrow$ (ii) This follows directly by Lemma 3.2 together with the Lemma of Gronwall.
- (ii) $\Rightarrow$ (i) Subtracting  $\mathbb{V}ar_{\mu}(f)$ , dividing by  $t$  and finally letting  $t \rightarrow 0$  gives the result:

$$\begin{aligned} \frac{\mathbb{V}ar_{\mu}(f_t) - \mathbb{V}ar_{\mu}(f)}{t} &\leq \mathbb{V}ar_{\mu}(f) \frac{e^{-2\alpha_{\text{PI}}t} - 1}{t} \\ \xrightarrow{t \rightarrow 0} & -2\mathcal{E}_{\mathcal{L}}(f) \leq -2\alpha_{\text{PI}}\mathbb{V}ar_{\mu}(f). \end{aligned}$$

Proving the equivalence of (ii) and (iii) is now almost as easy: First note that by defining  $f_t = \frac{\rho_t}{\mu}$ , where  $\rho_t$  is the time  $t$  density of the dynamics for an initial datum  $\rho_0$ , i.e.,  $\rho_t$  solves  $\frac{\partial}{\partial t}\rho_t = \mathcal{L}^*\rho_t$ ,  $\rho_{t=0} = \rho_0$ , we find by Theorem 2.16 that

$$\frac{\partial}{\partial t}\rho_t = \mathcal{L}_D^*(\rho_t) = \mathcal{L}_D^*(f_t\mu) = \mu\mathcal{L}_{-D}f_t,$$

i.e.,  $f_t$  is the solution to  $\frac{\partial}{\partial t} f_t = \mathcal{L}_D f_t$ ,  $f_0 = \frac{\rho_0}{\mu}$ . Secondly, note that  $f_0 \in L^2_\mu \Leftrightarrow \rho_0 \in L^2_{\mu^{-1}}$ . By Remark 3.3 we also have equivalence of (i) and (ii) for  $f_t$  given by  $f_t = \frac{\rho_t}{\mu}$ .

It remains to show (ii)  $\Leftrightarrow$  (iii):

Inserting  $f_t = \frac{\rho_t}{\mu}$  into (ii) directly yields the equivalence:

$$\mathbb{V}ar_\mu(f_t) = \int \left( \frac{\rho_t}{\mu} - 1 \right)^2 d\mu = \int (\rho_t - \mu)^2 d\mu^{-1} = |\rho_t - \mu|_{L^2_{\mu^{-1}}}^2.$$

□

Let us remark that since the right hand side of (3.4) only depends on the reversible part of the dynamics, adding any non-reversibility cannot effect the PI negatively. In other words, adding non-reversibility can only improve convergence thus making it faster.

Instead of stating a dynamic PI it is also possible to state a PI as a property of a measure alone and we refer to this as the *static PI*.

**Definition 3.6.** A measure  $\mu \in \mathcal{P}(\mathbb{R}^n)$  satisfies a *static PI* with constant  $\alpha_{\text{PI}} > 0$  if  $\forall f \in L^2_\mu$  it holds that

$$\alpha_{\text{PI}} \mathbb{V}ar_\mu(f) \leq |\nabla f|_{L^2_\mu}^2.$$

The static PI can also be interpreted as a dynamic PI, with  $\Gamma = I_{n \times n}$ , i.e., it is equivalent to exponential convergence for the associated gradient dynamics with generator  $\mathcal{L} = -\nabla V \cdot \nabla + \Delta$ .

If a measure  $\mu$  satisfies a static PI but convergence to  $\mu$  under the dynamics associated to some other generator  $\tilde{\mathcal{L}}$  of general form (3.1) is considered, we find – due to  $\mathcal{E}_{\mathcal{L}}(f) = \langle \nabla f, \Gamma \nabla f \rangle_\mu \geq \lambda_{\min}(\Gamma) \langle \nabla f, \nabla f \rangle_\mu$  – that the associated time  $t$  density  $\rho_t$  converges at least with rate given by  $2\alpha_{\text{PI}} \lambda_{\min}(\Gamma)$ , i.e.,

$$|\rho_t - \mu|_{L^2_{\mu^{-1}}} \leq e^{-2\alpha_{\text{PI}} \lambda_{\min}(\Gamma)t} |\rho_0 - \mu|_{L^2_{\mu^{-1}}}.$$

Conversely, if one observes exponential convergence for any initial condition  $\rho_0 \in L^2_{\mu^{-1}}$ , i.e., the above holds true for all  $\rho_0 \in L^2_{\mu^{-1}}$  and  $\Gamma < \lambda_{\max}(\Gamma) < \infty$  then  $\mu$  satisfies a dynamic PI with constant  $\alpha_{\text{PI}} \lambda_{\min}(\Gamma) (\lambda_{\max}(\Gamma))^{-1}$  which can easily be deduced by the inverse estimate  $\mathcal{E}_{\mathcal{L}}(f) = \langle \nabla f, \Gamma \nabla f \rangle_\mu \leq \lambda_{\max}(\Gamma) \langle \nabla f, \nabla f \rangle_\mu$ .

Often, the PI inequality is also referred to as spectral gap inequality. This is due to the fact that in certain cases, the spectral gap of the generator equals the PI constant. For example in the case that the generator admits a discrete spectrum and the corresponding eigenfunctions form a basis of  $L^2_\mu$ , this is an easy calculation (see the Appendix A.2). In general, one can resort to properties of the associated semigroup and we refer the interested reader to the book [24], see in particular Section 3 and Corollary 3.12 for conditions guaranteeing equality of the two constants. Let us now discuss the relation between the two quantities for non-reversible but elliptic OU-processes.

**Relation of PI and spectral gap for non-reversible OU-processes**

For non-reversible processes the relation of the PI constant and the spectral gap is not straightforward. Let us discuss the simple example of an  $n$ -dimensional OU-process

$$dX_t = AX_t dt + C dB_t, \quad (3.5)$$

with  $A, C \in \mathbb{R}^{n \times n}$  and  $A$  being Hurwitz and  $C$  having full rank. For degenerate  $C$  a PI with constant  $\alpha_{\text{PI}} > 0$  cannot be derived, since the right hand side of the PI (3.4) can become 0. Under these assumptions on  $A$  and  $C$  (see also Assumption A2 and Example 2.14) the process admits a unique invariant measure  $\mu \sim \mathcal{N}(0, \Sigma_\infty)$  with  $\Sigma_\infty$  being the solution to

$$\Sigma_\infty A + A^T \Sigma_\infty = -CC^T. \quad (3.6)$$

The associated reversible dynamics (see Example 2.18)

$$d\tilde{X}_t = -\frac{1}{2}CC^T\Sigma_\infty^{-1}\tilde{X}_t dt + C dB_t,$$

possess a PI constant that – due to the reversibility – equals the spectral gap and is in this case given by the smallest eigenvalue of the drift  $\frac{1}{2}CC^T\Sigma_\infty^{-1}$  (see [53]). This means, we have

$$\frac{1}{2}\lambda_{\min}(CC^T\Sigma_\infty^{-1})\text{Var}_\mu(f) \leq \mathcal{E}_\mathcal{L}(f), \quad \forall f \in \text{Dom}(\mathcal{L}).$$

Since  $CC^T$  and  $\Sigma_\infty^{-1}$  are both symmetric positive definite it follows that  $\lambda_{\min}(CC^T\Sigma_\infty^{-1}) \geq \lambda_{\min}(CC^T)\lambda_{\min}(\Sigma_\infty^{-1})$  and we have an estimate for the dynamic PI constant associated to the dynamics (3.5):  $\alpha_{\text{PI}} \geq \frac{1}{2}\lambda_{\min}(\Sigma_\infty^{-1})\lambda_{\min}(CC^T) =: \underline{\alpha}_{\text{PI}}$ . Now, referring to the same reference [53] as above, we know that for (3.5) the spectral gap is given by the smallest absolute value of the eigenvalues of the drift, i.e.,  $\lambda_1(-\mathcal{L}) = \lambda_{\min}(-A)$ . Employing the Lyapunov equation (3.6) and letting  $v$  be the eigenvector associated to  $\lambda_{\min}(-A)$  we find

$$\lambda_{\min}(CC^T)v^T v \leq v^T \Sigma_\infty A v + v^T A^T \Sigma_\infty v = 2\lambda_{\min}(-A)v^T \Sigma_\infty v \leq 2\lambda_{\min}(-A)\lambda_{\max}(\Sigma_\infty)$$

which implies by  $\lambda_{\max}(\Sigma_\infty) = 1/\lambda_{\min}(\Sigma_\infty^{-1})$  that

$$\alpha_{\text{PI}} = \frac{1}{2}\lambda_{\min}(\Sigma_\infty^{-1})\lambda_{\min}(CC^T) \leq \lambda_{\min}(-A) = \lambda_1(-\mathcal{L}).$$

Of course this is not a direct relation between  $\alpha_{\text{PI}}$  and  $\lambda_1(-\mathcal{L})$ , but only of the lower bound  $\underline{\alpha}_{\text{PI}}$  and  $\lambda_1(-\mathcal{L})$ . In fact, this simple example reveals already that for non-reversible processes the relation is not straight-forward.

We should remark again that, even though the Dirichlet form only depends on the reversible part of the generator, all of the theory presented in this section does not depend on reversibility or non-reversibility of the underlying dynamics as long



as  $\Gamma > 0$ . It does heavily rely on the fact that  $\Gamma > 0$ , though, which appears in the Dirichlet form. If  $\Gamma$  was only positive semi-definite it wouldn't be possible to derive exponential convergence by the tools presented in this section, as the right hand side of the PI – and thus the time derivative of the variance – can become zero. In this case, one has to rely on the theory of hypocoercivity developed by Villani (see e.g. [68]) which leads to an additional prefactor  $c \geq 1$  on the right hand side of the estimate which describes exponential convergence.

### 3.1.2 The Logarithmic Sobolev Inequality – convergence in relative entropy

The equivalence of a PI and convergence in  $L^2_{\mu^{-1}}$  is a nice result, but it also has several drawbacks. First of all, we should note that for finiteness of the right hand side of the estimate, we require that  $\rho_0 \in L^2_{\mu^{-1}}$ , i.e.,  $\int \rho_0^2(x) e^{V(x)} dx < \infty$ . This is a severe restriction in the admissible initial states, since  $V$  is assumed to grow sufficiently fast such that  $\int e^{-V(x)} dx < \infty$ . In addition, convergence in  $L^2_{\mu^{-1}}$  is physically not well-interpretable and a convergence statement in  $L^1$  – being the natural space for probability densities – would be preferable.

In this section we review results which yield exponential decay to equilibrium in relative entropy and thus in  $L^1$ .

We start by giving the definitions needed for this section.

**Definition 3.7.** Let  $\nu, \mu \in \mathcal{P}(\mathbb{R}^n)$ . The *relative entropy* (or *Kullback-Leibler divergence*) of  $\nu$  with respect to  $\mu$  is defined as

$$\mathcal{H}(\nu|\mu) = \begin{cases} \int \ln \left( \frac{d\nu}{d\mu} \right) d\nu, & \text{if } \nu \ll \mu \\ \infty, & \text{else.} \end{cases}$$

Note, that relative entropy is not a metric, since it is lacking the properties of symmetry and subadditivity. Yet, it satisfies  $\mathcal{H}(\nu|\mu) \geq 0$  by Jensen's inequality with equality iff  $\nu = \mu$   $\nu$ -almost surely.

**Definition 3.8.** Let  $\nu, \mu \in \mathcal{P}(\mathbb{R}^n)$ ,  $\nu \ll \mu$  and  $A \in \mathbb{R}^{n \times n}$ . The *A-weighted relative Fisher information* of  $\nu$  with respect to  $\mu$  is defined as

$$\mathcal{R}_A(\nu|\mu) = \int \left| \nabla \ln \frac{d\nu}{d\mu} \right|_A^2 d\nu.$$

If  $A = I_{n \times n}$  the subscript is omitted.

The following lemma reveals the important connection between relative entropy and the relative Fisher information for our purposes and constitutes the analogue of Lemma 3.2.

**Lemma 3.9.** Let  $\mu \in \mathcal{P}(\mathbb{R}^n)$  be the unique invariant measure associated to  $\mathcal{L}$  given by (3.1) and let  $\rho_t$  be the solution to  $\frac{\partial \rho_t}{\partial t} = \mathcal{L}^* \rho_t$  for some initial datum  $\rho_0$ . Then

$$\frac{d}{dt} \mathcal{H}(\rho_t|\mu) = -\mathcal{R}_\Gamma(\rho_t|\mu).$$

*Proof.* Performing integration by parts twice, the computation is straightforward:

$$\begin{aligned}
\frac{d}{dt} \int \ln \left( \frac{\rho_t}{\mu} \right) d\rho_t &= \int \partial_t(\rho_t) \ln \left( \frac{\rho_t}{\mu} \right) dz + \int \partial_t(\rho_t) dz \\
&= \int \nabla \cdot \left( \rho_t \left[ -D + \Gamma \nabla \ln \left( \frac{\rho_t}{\mu} \right) \right] \right) \ln \left( \frac{\rho_t}{\mu} \right) dz + 0 \\
&= - \int \left( \rho_t \left[ -D + \Gamma \nabla \ln \left( \frac{\rho_t}{\mu} \right) \right] \right) \cdot \nabla \ln \left( \frac{\rho_t}{\mu} \right) dz \\
&= \int \rho_t D \cdot \nabla \ln \left( \frac{\rho_t}{\mu} \right) dz - \int \nabla \ln \frac{\rho_t}{\mu} \cdot \Gamma \nabla \ln \left( \frac{\rho_t}{\mu} \right) d\rho_t(z) \\
&= - \int \left| \nabla \ln \left( \frac{\rho_t}{\mu} \right) \right|_{\Gamma}^2 d\rho_t(z).
\end{aligned}$$

The last equality follows by the divergence-free property of the function  $D$  (see (A1)(iii)) which gives

$$\int \rho_t D \cdot \nabla \left( \ln \frac{\rho_t}{\mu} \right) dz = \int \mu D \cdot \nabla \left( \frac{\rho_t}{\mu} \right) dz = - \int \nabla \cdot (\mu D) \left( \frac{\rho_t}{\mu} \right) dz = 0.$$

□

The Logarithmic Sobolev Inequality (LSI) introduced below plays the role of the PI in this section. As it connects relative entropy and its time derivative via an inequality, we can hope for exponential convergence in relative entropy. Equivalently to the PI, we can also state the LSI as a property of the measure alone or alternatively in its dynamics version by directly associating the measure to some dynamics generated by  $\mathcal{L}$ . Let us first state the dynamic LSI and deduce the exponential convergence in the subsequent corollary.

**Definition 3.10.** Let  $\mu \in \mathcal{P}(\mathbb{R}^n)$  be the unique invariant measure associated to the dynamics with generator  $\mathcal{L}$ . We say  $\mu$  satisfies a *dynamic Logarithmic Sobolev Inequality (LSI)* with constant  $\alpha_{\text{LSI}}$  if for all  $\nu \in \mathcal{P}(\mathbb{R}^n)$  with  $\nu \ll \mu$  it holds that

$$\mathcal{H}(\nu|\mu) \leq \frac{1}{2\alpha_{\text{LSI}}} \mathcal{R}_{\Gamma}(\nu|\mu).$$

**Remark 3.11.** An equivalent formulation of the LSI, which is also common in the literature, is as follows: Let  $\mu \in \mathcal{P}(\mathbb{R}^n)$  be the unique invariant measure associated to  $\mathcal{L}$ . Then  $\mu$  satisfies a LSI with constant  $\alpha_{\text{LSI}}$  if for all  $f \in L_{\mu}^2$  it holds that

$$\int f^2 \ln \left( \frac{f^2}{\int f^2 d\mu} \right) d\mu \leq \frac{2}{\alpha_{\text{LSI}}} \mathcal{E}_{\mathcal{L}}(f).$$

The equivalence can be seen as follows: let  $f = \sqrt{\frac{d\nu}{d\mu}}$  such that  $\nu \ll \mu$ . Then  $f \in L_{\mu}^2$  since  $\int f^2 d\mu = \int d\nu = 1$ . Conversely  $f \in L_{\mu}^2$  implies that  $\nu \in \mathcal{P}(\mathbb{R}^n)$  since we

can replace  $f$  by its normalized version  $f/\|f\|_{L^2_\mu}$  which does not affect the above inequality. Plugging  $f = \sqrt{\frac{d\nu}{d\mu}}$  into the left hand side we find

$$\int f^2 \ln \left( \frac{f^2}{\int f^2 d\mu} \right) d\mu = \int \ln \left( \frac{d\nu}{d\mu} \right) d\nu = \mathcal{H}(\nu|\mu).$$

Noticing that  $\mathcal{E}_{\mathcal{L}}(f) = \int |\nabla f|_\Gamma^2 d\mu$  and  $\nabla f = \frac{1}{2} \sqrt{\frac{d\mu}{d\nu}} \left( \frac{1}{d\mu} \nabla d\nu - \frac{d\nu}{d\mu^2} \nabla d\mu \right)$  and at the same time  $\nabla \ln \frac{d\nu}{d\mu} = \frac{d\nu}{d\nu} \left( \frac{1}{d\mu} \nabla d\nu - \frac{d\nu}{d\mu^2} \nabla d\mu \right)$  it follows that

$$\mathcal{E}_{\mathcal{L}}(f) = 4 \mathcal{R}_\Gamma(\nu|\mu)$$

which yields the equivalence of the two formulations.

**Corollary 3.12.** *Let  $\mu \in \mathcal{P}(\mathbb{R}^n)$  satisfy a LSI with constant  $\alpha_{\text{LSI}}$  and let  $\rho_0 \ll \mu$ . Then  $\rho_t$ , the solution to  $\frac{\partial \rho_t}{\partial t} = \mathcal{L}^* \rho_t$  with initial datum  $\rho_0$ , converges exponentially fast to  $\mu$  in relative entropy with rate  $\alpha_{\text{LSI}}$ , that is,*

$$\mathcal{H}(\rho_t|\mu) \leq e^{-2\alpha_{\text{LSI}} t} \mathcal{H}(\rho_0|\mu).$$

*Conversely, if the above holds true for all  $\rho_0$  with  $\mathcal{H}(\rho_0|\mu) < \infty$  then  $\mu$  satisfies a LSI with constant  $\alpha_{\text{LSI}}$ .*

*Proof.* The exponential convergence follows directly by Lemma 3.9 and the Gronwall Lemma.

The other direction follows by subtracting  $\mathcal{H}(\rho_0|\mu)$  on both sides, dividing by  $t$  and letting  $t \rightarrow 0$

$$\begin{aligned} \frac{\mathcal{H}(\rho_t|\mu) - \mathcal{H}(\rho_0|\mu)}{t} &\leq \mathcal{H}(\rho_0|\mu) \frac{e^{-2\alpha_{\text{LSI}} t} - 1}{t} \\ \xrightarrow{t \rightarrow 0} &\quad -\mathcal{R}_\Gamma(\rho_0|\mu) \leq -2\alpha_{\text{LSI}} \mathcal{H}(\rho_0|\mu). \end{aligned}$$

□

We can also formulate a LSI as a property of a measure alone.

**Definition 3.13** (Static LSI). Given a measure  $\mu \in \mathcal{P}(\mathbb{R}^n)$  we say  $\mu$  satisfies a static LSI with constant  $\alpha_{\text{LSI}}$  if for all measures  $\nu \ll \mu$  it holds that

$$\mathcal{H}(\nu|\mu) \leq \frac{1}{2\alpha_{\text{LSI}}} \mathcal{R}(\nu|\mu). \tag{3.7}$$

Clearly, if  $\mu$  is the unique invariant measure associated to a generator with  $\Gamma = I_{n \times n}$  then the above is equivalent to exponential convergence in relative entropy with rate  $2\alpha$ . If instead  $\mu$  satisfies a static LSI and is the invariant measure associated to some dynamics with diffusion matrix  $\Gamma \neq I_{n \times n}$ , it follows by  $|\cdot| \leq \frac{1}{\lambda_{\min}(\Gamma)} |\cdot|_\Gamma$  that exponential convergence in relative entropy takes place at least with rate  $2\lambda_{\min}(\Gamma)\alpha$ . Conversely, if we observe exponential convergence to equilibrium in relative entropy with rate  $2\alpha$  for dynamics with diffusion matrix  $\Gamma$  and invariant measure  $\mu$  then by  $|\cdot|_\Gamma \leq \lambda_{\max}(\Gamma) |\cdot|$  we find a static LSI holds for  $\mu$  with constant  $\alpha/\lambda_{\max}(\Gamma)$ .

By means of the static LSI we also find a convenient interpretation for large static LSI constants.

**Remark 3.14.** Given a measure  $\mu$  which satisfies a static LSI with a large constant  $\alpha \gg 0$  we know that the corresponding gradient dynamics converges fast – exponentially with rate  $2\alpha$  – to its equilibrium. This means that the gradient dynamics cannot exhibit strong metastability. If there was metastability in the gradient dynamics then  $\alpha$  would not be “as large”. Indeed, the larger  $\alpha$  the faster the corresponding gradient dynamics relax to their equilibrium and by the Bakry-Émery condition (see Theorem 3.21 below) the steeper is the potential ( $\nabla^2 V \geq \alpha$ ). At the same time, a lack of metastability also means that there are no barriers to overcome in the potential  $V$ . Concluding, this means the larger the LSI the more peaked or concentrated the measure should be. So here, the LSI equips us with a property of the measure  $\mu$  alone. See also [46] for a discussion of the connection between the LSI and metastability for overdamped Langevin dynamics.

### Other properties of the LSI

Another appealing property of convergence in relative entropy is that it entails convergence in  $L^1$  (assuming as usual that  $d\nu(x) = \nu(x) dx$  and  $d\mu(x) = \mu(x) dx$ ) via the *Csiszár - Kullback inequality* (see e.g. [64])

$$\int |\nu(x) - \mu(x)| dx \leq \sqrt{2\mathcal{H}(\nu|\mu)}.$$

Thus, even though relative entropy is not a metric itself, we can deduce convergence in  $L^1$ .

Moreover, a LSI is stronger than a PI as the following lemma states.

**Lemma 3.15.** *Let  $\mu \in \mathcal{P}(\mathbb{R}^n)$  satisfy a LSI with constant  $\alpha_{\text{LSI}}$ . Then  $\mu$  satisfies a PI with constant  $\alpha_{\text{PI}} = \alpha_{\text{LSI}}$ .*

*Proof.* For a proof see the Appendix A.2 Lemma A.5. □

The next theorem states, that once we have found a measure which satisfies a LSI it is possible to modify it by adding a bounded perturbation to the potential and we still get an estimate on the LSI constant.

**Theorem 3.16** (Holley-Stroock Lemma [34]). *Let  $(\mu, \mathcal{L}^V)$  satisfy a LSI with constant  $\alpha_{\text{LSI}}$ , where  $\mu \sim e^{-V}$  and the superscript in  $\mathcal{L}^V$  clarifies the dependence of the generator on the potential  $V$ . Let  $\tilde{V} = V + U$ , where  $U$  is a smooth and bounded potential, i.e.,  $m \leq U(x) \leq M \forall x \in \mathbb{R}^n$ . Then  $(\tilde{\mu}, \mathcal{L}^{\tilde{V}})$ , where  $\tilde{\mu} \sim e^{-\tilde{V}}$ , also satisfy a LSI with constant  $\tilde{\alpha}_{\text{LSI}} = \alpha_{\text{LSI}} e^{M-m}$ .*

### 3.1.3 The Talagrand Inequality – convergence in Wasserstein distance

In this section relative entropy will be related to the second *Wasserstein* distance by the so-called *Talagrand inequality*. We will employ the Talagrand inequality in Chapter 4 deriving error estimates for the reduced model. To this end, recall the definition of a coupling of two measures and the Wasserstein distance.

**Definition 3.17.** Let  $\nu, \mu \in \mathcal{P}(\mathbb{R}^n)$ . A measure  $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$  is called *coupling* of  $\nu$  and  $\mu$  if its marginals are  $\nu$  and  $\mu$  respectively, that is, for any bounded continuous test functions  $\phi, \psi$  it holds that

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \phi(x_1) + \psi(x_2) d\pi(x_1, x_2) = \int_{\mathbb{R}^n} \phi(x_1) d\nu(x_1) + \int_{\mathbb{R}^n} \psi(x_2) d\mu(x_2).$$

**Definition 3.18.** Let  $\nu, \mu \in \mathcal{P}(\mathbb{R}^n)$ . The second *Wasserstein distance* of  $\nu$  with respect to  $\mu$  is given by

$$\mathcal{W}(\nu|\mu) = \inf_{\pi \in \Pi(\nu, \mu)} \sqrt{\int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\pi(x, y)},$$

where  $\Pi$  is the set of all couplings of  $\nu$  and  $\mu$ .

We will omit the supplement second and will simply talk about the Wasserstein-distance from now on. The Wasserstein distance admits a vivid interpretation: assuming that  $\nu, \mu \in \mathcal{P}(\mathbb{R}^2)$ , we can picture both measures as sand piles in the plane whose height is the value of the associated density. Being asked to transfer the sand pile associated to  $\nu$  to the one belonging to  $\mu$ , the set  $\Pi$  captures all possible ways of achieving this goal. The transfer which costs the least among all possible transfers with respect to some metric – in our case  $L^2$  and else e.g.  $L^p$  – yields the corresponding Wasserstein distance.

**Definition 3.19.** A measure  $\mu \in \mathcal{P}(\mathbb{R}^n)$  satisfies as *Talagrand inequality (TI)* with constant  $\alpha_{\text{TI}} > 0$  if for all  $\nu \in \mathcal{P}(\mathbb{R}^n)$  with  $\nu \ll \mu$  it holds that

$$\mathcal{W}(\nu|\mu)^2 \leq \frac{2}{\alpha_{\text{TI}}} \mathcal{H}(\nu|\mu).$$

**Lemma 3.20.** Let  $\mu$  satisfy a static LSI (i.e.  $\Gamma = I_{n \times n}$ ) with constant  $\alpha_{\text{LSI}}$ . Then  $\mu$  also satisfies a TI with constant  $\alpha_{\text{TI}} = \alpha_{\text{LSI}}$ .

*Proof.* For the proof we refer to [59]. □

### 3.1.4 The Bakry-Émery criterion

Being mostly concerned with convergence in relative entropy together with the fact that a LSI implies a PI, we only present the famous *Bakry-Émery* condition here, constituting a sufficient condition to guarantee a LSI. A sufficient criterion for a PI only – building on Lyapunov functions – can be found in the appendix in Theorem A.6.

**Theorem 3.21** (The Bakry-Émery criterion). Let  $\Gamma(x) \equiv \Gamma > 0$  be constant and  $d\mu(x) = \frac{1}{Z} e^{-V(x)} dx$  the invariant measure associated to  $\mathcal{L} = -\Gamma \nabla V \cdot \nabla + \Gamma : \nabla^2$ . If

$$y^T \nabla^2 V(x) y \geq \alpha y^T \Gamma^{-1} y, \quad \forall x, y \in \mathbb{R}^n \tag{3.8}$$

holds true, then  $(\mu, \mathcal{L})$  fulfils a dynamic LSI with constant  $\alpha$ .

*Sketch of proof.* We outline the idea of the proof and refer to it as the *Bakry-Émery procedure* as it is reused in the subsequent results.

1. Show:  $\nabla^2 V \geq \alpha \Gamma^{-1} \Rightarrow \mathcal{R}_\Gamma(\rho_t|\mu) \leq -\frac{1}{2\alpha} \frac{d}{dt} \mathcal{R}_\Gamma(\rho_t|\mu)$

This is the main step of the proof and the difficulty lies within computing the time derivative of the relative fisher information (for the computations see e.g. [4, 6]) which in the end yields

$$-\frac{d}{dt} \mathcal{R}_\Gamma(\rho_t|\mu) = 2 \int \left| \nabla \ln \frac{\rho_t}{\mu} \right|_{\Gamma \nabla^2 V \Gamma}^2 d\rho_t + r,$$

where  $r \geq 0$  is a positive remainder. By assumption (3.8) it readily follows that  $\left| \nabla \ln \frac{\rho_t}{\mu} \right|_{\Gamma \nabla^2 V \Gamma}^2 \geq \alpha \left| \nabla \ln \frac{\rho_t}{\mu} \right|_\Gamma^2$  which yields the result of the first step.

2. Show:  $\mathcal{R}_\Gamma(\rho_t|\mu) \leq -\frac{1}{2\alpha} \frac{d}{dt} \mathcal{R}_\Gamma(\rho_t|\mu) \Rightarrow \mathcal{R}_\Gamma(\rho_t|\mu) \leq e^{-2\alpha t} \mathcal{R}_\Gamma(\rho_0|\mu)$ .

This follows directly by integrating the above inequality in time from 0 to  $t$ .

3. Show:  $\mathcal{R}_\Gamma(\rho_t|\mu) \leq e^{-2\alpha t} \mathcal{R}_\Gamma(\rho_0|\mu) \Rightarrow 2\alpha \mathcal{H}(\rho_0|\mu) \leq \mathcal{R}_\Gamma(\rho_0|\mu)$

This follows directly by integrating the first inequality in time from 0 to  $\infty$  together with noting that  $\int_0^\infty \mathcal{R}_\Gamma(\rho_t|\mu) dt = -\mathcal{H}(\mu|\mu) + \mathcal{H}(\rho_0|\mu) = \mathcal{H}(\rho_0|\mu)$ .

□

### Multiplicative non-degenerate noise

If  $\Gamma = \Gamma(x) > 0$  the procedure is exactly as above, only condition (3.8) is replaced by  $\Gamma(x) Ric(x) \geq \alpha I_{n \times n}$ , where  $Ric(x)$  is the Ricci-tensor of the Fokker-Planck operator induced by  $\Gamma(x)$  (cf. Section 6 in [7] or Section 2.3 in [6]), which is interpreted as a metric tensor. The term on the left hand side will show up in the calculation of the time derivative of the relative Fisher information and hereby replace  $\nabla^2 V$  (step 1 above). Calculating the time derivative of the relative Fisher information as well as the remainder becomes a lot harder in this setting, but in general the procedure does not change.

## 3.2 Exponential convergence to equilibrium for hypo-elliptic OU-processes

We have seen in the previous section that the Bakry-Émery condition for the potential  $V$  plus bounded perturbations suffice to guarantee exponential convergence to equilibrium. This heavily relies on the positive definiteness of the diffusion matrix. If the diffusion  $\Gamma$  was only positive-semidefinite, i.e., there if there existed a  $v \in \ker \Gamma$ , the relaxation towards equilibrium in relative entropy can lose the property of strict monotonicity which is replaced by monotonicity. To make this point clear, assume that the initial condition  $\rho_0$  is such that  $\rho_t = (v \cdot x)\mu$ , where  $v, x \in \mathbb{R}^n$ ,  $v \in \ker \Gamma$ . Then

$$\mathcal{R}_\Gamma(v \cdot x \mu) = \int |\nabla \ln v \cdot x|_\Gamma^2 d\mu = \int (v \cdot x)^{-2} v^T \Gamma v d\mu = 0.$$

In particular, this prohibits the first step in the Bakry-Émery procedure ( $\Gamma$  is not invertible), such that proving exponential convergence of the relative Fisher information with the usual tools is not possible.

### OU-processes

In [5] Arnold et al. have modified the Bakry-Émery procedure to be able to deal with OU-processes with degenerate noise. To this end, consider an OU-process given by

$$dX_t = AX_t dt + \sqrt{2}C dW_t, \quad (3.9)$$

where  $(A, C)$  fulfill assumption A2 such that we have a unique invariant measure  $\mu \sim \mathcal{N}(0, \Sigma_\infty)$ . Instead of working directly with the relative Fisher information  $\mathcal{R}_\Gamma$ , where  $\Gamma = CC^T$ , they introduce a modified Fisher information  $\mathcal{R}_S$  which does decay exponentially and at the same time yields an upper bound for  $\mathcal{R}_\Gamma$  such that exponential decay for  $\mathcal{R}_\Gamma$  follows. To be more precise, they prove existence of a positive definite matrix  $S$  such that  $S \geq c\Gamma$ ,  $c > 0$  and

$$S\tilde{A} + \tilde{A}^T S \geq 2\lambda^* S, \quad (3.10)$$

where  $\tilde{A} = \Sigma_\infty A^T \Sigma_\infty^{-1}$  and  $\lambda^* = \min \{\operatorname{Re}(\lambda) : -Av = \lambda v, v \neq 0\}$ . If  $A$  and thus also  $\tilde{A}$  is diagonalisable the matrix  $S$  is constructed from the eigenvectors of  $\tilde{A}$ , else it is constructed from the generalized eigenvectors of  $\tilde{A}$  which leads to the replacement of  $\lambda^*$  by  $\lambda^* - \varepsilon$ ,  $\varepsilon > 0$ . Step 1 in the Bakry-Émery procedure is then replaced by the computation of the time derivative of the modified Fisher information which gives

$$-\frac{d}{dt} \mathcal{R}_S(\rho_t | \mu) = \int \nabla \left( \ln \frac{\rho_t}{\mu} \right) (S\tilde{A} + \tilde{A}^T S) \nabla \ln \left( \frac{\rho_t}{\mu} \right) d\rho_t + r,$$

where  $r \geq 0$  is a positive remainder and equation (3.10) plays the role of the Bakry-Émery condition (3.8). This implies exponential decay of  $\mathcal{R}_S$  with rate  $2\lambda^*$  which – after some more involved estimates – also entails exponential convergence of the relative entropy. The statement of their theorem is as follows.

**Theorem 3.22** (Theorem 4.9 in [5]). *Consider the SDE (3.9) where the coefficients satisfy assumption A2 and let  $\rho_t$  be the solution to the associated Fokker-Planck equation with initial condition  $\rho_0$ . Define  $\lambda^* = \min \{\operatorname{Re}(\lambda) : -Av = \lambda v, v \neq 0\} > 0$  as the smallest eigenvalue of  $-A$  and suppose that  $H(\rho_0 | \mu) < \infty$ .*

(i) *If all eigenvalues of  $A$  are non-defective, then there exists a constant  $c \geq 1$  such that*

$$H(\rho_t | \mu) \leq c H(\rho_0 | \mu) e^{-2\lambda^* t} \quad \forall t \geq 0.$$

(ii) *If one or more eigenvalues are defective, then there exists a constant  $c_\varepsilon > 1$  for all  $\varepsilon \in (0, \lambda^*)$ , such that*

$$H(\rho_t | \mu) \leq c_\varepsilon H(\rho_0 | \mu) e^{-2(\lambda^* - \varepsilon)t} \quad \forall t \geq 0.$$

The actually observed relaxation behaviour is explored in Section 6.1 where we investigate the influence of temperature and the choice of initial conditions. Further, we study the occurrence of plateaus in the decay and processes with multiple time scales.

**Remark 3.23.** Observe that even though Theorem 3.22 establishes exponential convergence to equilibrium this does not imply a LSI for the measure: If we proceed as in the proof of Theorem 3.12 subtracting  $\mathcal{H}(\rho_0|\mu)$  from both sides and dividing by  $t$  we find

$$\frac{\mathcal{H}(\rho_t|\mu) - \mathcal{H}(\rho_0|\mu)}{t} \leq c \mathcal{H}(\rho_0|\mu) \frac{e^{-2\lambda^* t} - 1/c}{t}$$

and while the left-hand side tends to  $-\mathcal{R}(\rho_0|\mu)$  as  $t \rightarrow 0$  the right-hand side does not have a well-defined limit except if  $c = 1$  which amounts to  $\Gamma > 0$  and hence we are back in the non-degenerate case.

This is precisely the constant  $c \geq 1$  mentioned before, which is due to the hypo-ellipticity and thus hypocoercivity of process.



### 3.3 Mean first exit times in the small diffusion limit for elliptic processes

In this section, results on the second time scale of interest – given by mean first exit times – will briefly be reviewed. Similar to the previous section, we first treat the non-degenerate case  $\Gamma > 0$  and in the second part dedicate ourselves to the degenerate case  $\Gamma \geq 0$ . Mathematically, the degenerate case needs different tools than the standard ones from the first part of the chapter, and control theory will turn out to be a useful tool.

We start by reviewing the approach of determining the mean first exit time by means of large deviation theory as introduced by Freidlin and Wentzell (for more details we refer to [26]). To this end, consider the  $\mathbb{R}^n$ -valued SDE

$$dX_t = b(X_t) dt + \varepsilon \gamma dW_t, \quad (3.11)$$

where  $b$  is assumed to be Lipschitz,  $\Gamma = \gamma\gamma^T \in \mathbb{R}^{n \times n}$  is constant and has full rank and  $\varepsilon > 0$  determines the strength of the noise. Let  $\mathcal{D} \subset \mathbb{R}^n$  be a bounded domain with smooth boundary  $\partial\mathcal{D}$  and consider the mean of the first exit time  $\tau$  from the domain  $\mathcal{D}$  given the dynamics start in  $x_0 \in \mathcal{D}$ , that is,

$$\mathbb{E}_{x_0}(\tau) := \mathbb{E}(\tau | X_0 = x_0), x_0 \in \mathcal{D}, \quad \text{where } \tau = \inf \{t > 0 : X_t \notin \mathcal{D}\}.$$

Note, that  $\tau$  depends on  $\varepsilon$ , which we will omit in the following. We will now touch the theory of large deviations, which could also be named theory of improbable or rare events. We start with the main result which the theory offers for this problem.

**Theorem 3.24** (Theorem 4.1, Chapter 4 in [26]). *Let  $z \in \mathcal{D}$  be an asymptotically stable equilibrium for the deterministic dynamics  $\dot{x}_t = b(x_t)$  and assume that  $\mathcal{D}$  is attracted to  $z$ . Further assume that  $n(x) \cdot b(x) < 0 \ \forall x \in \partial\mathcal{D}$ , where  $n(x)$  is the outward normal vector of the boundary  $\partial\mathcal{D}$  at the point  $x$  (that is  $\mathcal{D}$  is invariant under the dynamics). Then, for any  $x \in \mathcal{D}$  it holds that*

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \ln \mathbb{E}_x(\tau) = \min_{y \in \partial\mathcal{D}} V^q(z, y), \quad \text{where}$$

$$V^q(x, y) = \inf_{\phi} \left\{ \frac{1}{2} \int_0^T \left| \dot{\phi}(t) - b(\phi(t)) \right|_{\Gamma^{-1}}^2 dt, \phi(0) = x, \phi(T) = y, T > 0 \right\},$$

where the infimum is taken over all functions  $\phi$  that are absolutely continuous and  $V^q$  is called the quasipotential of the dynamics (3.11).

The theorem states that in the small-noise limit the mean first exit time scales exponentially in  $1/\varepsilon^2$  with rate given by the quasipotential. The functional

$$S_T \phi = \frac{1}{2} \int_0^T \left| \dot{\phi}(t) - b(\phi(t)) \right|_{\Gamma^{-1}}^2 dt$$

is referred to as *rate function* or *action functional*, where the first denomination is due to the fact that in the small noise limit

$$\mathbb{P}(X_t \approx \phi(t), t \in [0, T]) \simeq e^{-S_T \phi / \varepsilon^2}.$$

This means the probability that a trajectory  $X_t$  of (3.11) follows some prescribed path  $\phi(t)$  for small  $\varepsilon$ , scales exponentially in  $1/\varepsilon^2$  and the rate is given by  $S_T\phi$ . Observe that  $S_T\phi \geq 0$  and  $S_T\phi = 0 \Leftrightarrow \dot{\phi}(t) = b(\phi(t))$ , i.e., in the small noise limit the dynamics follow the corresponding deterministic dynamics with probability 1. The second name originates from physics nomenclature, where  $\left| \dot{\phi}(t) - b(\phi(t)) \right|_{\Gamma^{-1}}^2$  would be termed Lagrangian. We can further interpret the rate function as follows: if we wanted to follow a prescribed path  $\phi(t)$  with the stochastic dynamics (3.11), then  $S_T\phi$  measures the amount of noise which makes it possible to do so, i.e., it measures the amount of energy needed such that  $\dot{X}_t = \dot{\phi}(t)$  could hold true.

### 3.3.1 Diffusion in potential energy landscape

Consider now a diffusion in a potential energy landscape, i.e.,

$$dX_t = -\nabla V(X_t) dt + \varepsilon dW_t,$$

denote by  $V_0 = V(x_0)$  a local minimum of  $V$  such that  $x_0 \in \mathcal{D} \subset \mathbb{R}^n$ , where  $\mathcal{D}$  is a domain fulfilling the assumptions of Theorem 3.24. We will first give explicit expressions for the quasipotential in this situation – which turns out to be a scalar multiple of the true potential  $V$  – and close the section by making the connection to the first part of the chapter relating mean first exit times and equilibration time scales by the seminal work of Bovier [13].

**Theorem 3.25** (Theorem 3.1, Chapter 4 in [26]). *Let  $V \in C^2$ . The rate function  $S_T\phi$  subject to  $\phi(0) = x_0$ ,  $\phi(T) = y$ ,  $y \in \partial\mathcal{D}$  is minimised by  $\phi = \phi^*$  solution to  $\dot{\phi}^*(t) = \nabla V(\phi^*(t))$  and for any  $y \in \partial\mathcal{D}$  the quasipotential takes the form  $V^q(x_0, y) = 2(V(y) - V_0)$ .*

*Proof.* The short proof can be found in the appendix in Theorem A.7.  $\square$

This result is known as Arrhenius law by physicists and chemists stating that the mean time it takes for a chemical reaction to take place is exponential in the activation energy scaled by the inverse temperature. Kramer's law in turn also contains the prefactor of the exponential thus giving more precise asymptotics and there are various works concerning the calculation of these. We are not particularly interested in the prefactors, but the relation to the eigenvalues of the generator. To this end, assume that the potential  $V$  fulfils  $\liminf_{x \rightarrow \infty} V(x) = \infty$ ,  $\liminf_{x \rightarrow \infty} |\nabla V(x)| = \infty$ ,  $\liminf_{x \rightarrow \infty} (|\nabla V(x)| - 2\Delta V(x)) = \infty$ . These conditions bear a close resemblance of the conditions for a PI (see Theorem A.6) and in fact yield that the resolvent  $(\mathcal{L} - \lambda I)^{-1}$  is compact for small enough  $\varepsilon$  and hence also  $\mathcal{L}$  is compact and admits a discrete spectrum (cf. [31]).

Further assume that  $V$  possesses  $n$  local minima at  $M = \{x_1^*, \dots, x_n^*\}$  which will be ordered according to the barrier height that needs to be overcome in order to reach any minima with a lower index. To be more precise, denote by  $p(a, b) =$

$\{\gamma : [0, 1] \rightarrow \mathbb{R}, \gamma(0) \in a, \gamma(1) \in b\}$  the set of paths from  $a \subset \mathbb{R}^n$  to  $b \subset \mathbb{R}^n$ . Also introduce

$$\hat{V}(a, b) = \inf_{\gamma \in \mathcal{P}(a, b)} \sup_{t \in [0, 1]} V(\gamma(t))$$

which is the minimal height (or minimal saddle) that has to be crossed on the way from  $a$  to  $b$ . It will be assumed that the minimal saddle between any two minima is unique. The unique argument corresponding to the minimal saddle is denoted by  $z(a, b) = \{z : V(z) = \hat{V}(a, b)\}$ . The ordering is done as follows: the first minimum is the global minimum, i.e.,

$$x_1 = \arg \min V(x)$$

and the next ones are listed according to barrier height – starting with the highest – for reaching any minima with lower index, i.e.,

$$x_j = \arg \max_{x_k^* \in M \setminus \{x_1, \dots, x_{j-1}\}} \left( V(z(x_k^*, \cup_{i=1}^{j-1} x_i)) - V(x_k^*) \right), \quad j \in \{2, \dots, n\}.$$

Introduce  $S_j = \cup_{i=1}^{j-1} B_\varepsilon(x_i)$  as the union of  $\varepsilon$ -balls around the minima with lower index than  $j$ , i.e., the union of minima corresponding to the  $j$  highest energy barriers, and  $\tau(S_j) = \inf \{t > 0 : X_t \in S_j\}$  which is the associated first passage time into the set  $S_j$  and also introduce the shorthand  $z_k = z(x_{k+1}, \cup_{i=1}^k x_i)$  for the  $k^{\text{th}}$  barrier height.

**Theorem 3.26** (Proposition 4.10 in [13]). *Let the above assumptions hold true. Then the first  $n$  eigenvalues  $\lambda_1 = 0 > \lambda_2 > \dots > \lambda_n$  are simple and satisfy, for small enough  $\varepsilon$  and  $k = 2, \dots, n$ ,*

$$\begin{aligned} \lambda_k &\approx -\frac{1}{\mathbb{E}_{x_k}(\tau(S_k))} \\ &\approx -\frac{|\lambda^-(\nabla^2 V(z_{k-1}))|}{2\pi} \sqrt{\frac{|\det(\nabla^2 V(x_k))|}{|\det(\nabla^2 V(z_{k-1}))|}} \exp\left(-2\frac{V(z_{k-1}) - V(x_k)}{\varepsilon}\right), \end{aligned}$$

where  $\lambda^-(\nabla^2 V(z_k))$  is the unique negative eigenvalue of the Hessian of  $V$  at the minimal saddle point.

This theorem provides well-interpretable insight to the problem at hand: it establishes a connection between the eigenvalues of the generator and the slowest time scales of the dynamics – given by transitions from metastable states to states being energetically even more stable – in a hierarchical manner. That is, a larger eigenvalue corresponds to a larger transition time. At the same time, larger transition times correspond to overcoming higher energy barriers, as given by large deviations theory, but here we learn something about the prefactor which depends on the curvature of the potential at the minimum and at the saddle.

The proof of the above result is based on potential theory and the analysis of Witten Laplacians. There are also other approaches, such as e.g. the work by Schlichting and Menz [52], who give an estimate for  $\lambda_1$  and the LSI constant

$\alpha_{\text{LSI}}$  in the same setting as above (Langevin dynamics) but with other techniques based on mean-difference estimates, leading to the same result as above for  $\lambda_1$ . Generalizations of the result have been proven by Nectoux and others in [19] where they derive results on exit times from a metastable state with boundary that does not necessarily admit a unique minimal saddle. In the context of non-reversible elliptic diffusions there is work by Bouchet and Reygner (cf. [12]) concerning also the prefactors of the associated MFET. In this case the quasipotential takes the role of the potential above and there is an extra term accounting for the non-reversibility of the dynamics. Even though all approaches are more general in the sense that they either allow for non-unique saddles or non-reversibility, the non-degeneracy of the noise – guaranteeing uniform ellipticity of the generator – is still most essential. Surprisingly enough, this fundamental assumption becomes redundant in the next section by employing a control theoretic formulation of the action functional.

### 3.4 Mean first exit times in the small diffusion limit for hypo-elliptic processes

#### 3.4.1 The action functional and the controllability gramian

In this section we review results obtained by Zabczyk in 1985 in [69] which are based on rephrasing the large deviations action functional in the language of control theory. To this end, recall the definition of the action functional

$$S_T\phi = \frac{1}{2} \int_0^T \left| \dot{\phi}(t) - b(\phi(t)) \right|_{\Gamma^{-1}}^2 dt, \quad \Gamma = \gamma\gamma^T > 0.$$

associated to the SDE

$$dX_t = b(X_t) dt + \varepsilon\gamma dW_t, \quad (3.12)$$

where  $X_t \in \mathbb{R}^n$ ,  $b: \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\varepsilon > 0$ ,  $\gamma \in \mathbb{R}^{n \times n}$ . Further introduce the corresponding controlled ODE (for  $\varepsilon = 1$ )

$$\dot{x}_t^u = b(x_t^u) + \gamma u(t), \quad (3.13)$$

where  $x_t^u \in \mathbb{R}^n$ ,  $b$  as above,  $\gamma \in \mathbb{R}^{n \times m}$  and  $u \in L^2([0, T], \mathbb{R}^m)$  is the control which replaces the noise. For a given control  $u$  denote by  $x_t^u$  the associated solution to (3.13). The *cost functional* is defined as the total energy provided to the system by the control

$$\tilde{S}_T x^u = \frac{1}{2} \int_0^T |u(t)|^2 dt.$$

Observe that if  $\Gamma = \gamma\gamma^T$  has full rank then by (3.13)

$$\tilde{S}_T x^u = \frac{1}{2} \int_0^T |\dot{x}_t^u - b(x_t^u)|_{\Gamma^{-1}}^2 dt = S_T x^u,$$

i.e., the cost function equals the rate function. Taking the infimum over the controls such that the cost functional is minimised with  $x^u(0) = 0$  and  $x^u(T) = b$  for some  $T > 0$  (or equivalently  $x^u(-\infty) = 0$  and  $x^u(0) = b$ ) yields the definition of the *controllability function*, i.e.,

$$L_c(b) = \inf_u \left\{ \frac{1}{2} \int_0^T |u(t)|^2 dt, x_0^u = 0, x_T^u = b, T > 0 \right\}.$$

In fact, the controllability function does not only agree with the action functional in the case when the diffusion is non-degenerate, but yields a generalisation of the action functional in the sense that the admissible systems for which a large deviations principle can be derived is augmented in several aspects. Most importantly, it allows for treating systems with degenerate noise. Other assumptions of Theorem 3.24 concerning the domain  $\mathcal{D}$  and the quasipotential

$$\tilde{V}^q(a, b) = \inf_{\phi} \left\{ \tilde{S}_T \phi, \phi(0) = a, \phi(T) = b, T > 0 \right\}$$

also change:

**Assumption A3.** We say assumption A3 holds if the following conditions are satisfied.

- (i) Let  $\dot{z}(t) = b(z(t))$  and  $z(t, z_0)$  the solution to the initial condition  $z(0) = z_0$ . Assume that there is a domain  $G \supset \mathcal{D}$  such that  $z^* \in \mathcal{D}$  is an asymptotically stable fixed point for this domain. That is,  $\forall z_0 \in G, \forall \delta > 0, \exists t_0 > 0$  such that  $|z(t, z_0) - z^*| < \delta, \forall t \geq t_0$ .
- (ii) Assume that the linearization of  $b$  around  $z^*$  given by the Jacobian  $A = \nabla b(z^*)$  together with  $\gamma$  fulfils the Hörmander or Kalman rank condition  $\text{rk}[\gamma, A\gamma, \dots, A^{n-1}\gamma] = n$ .
- (iii) There exist neighbourhoods  $N_1, N_2$  of  $z^*$  and  $\partial\mathcal{D}$  respectively such that the quasipotential  $\tilde{V}^q$  is continuous on these neighbourhoods.

Further introduce the notation

$$\mathcal{D}_0 = \left\{ x \in \mathcal{D} : \inf_{y \in \partial\mathcal{D}} |z(t, x) - y| > 0 \forall t > 0 \right\}. \quad (3.14)$$

Let us comment on these assumptions: the first one relaxes the previous condition, which asked for  $z^*$  being asymptotically stable as well as for  $\mathcal{D}$  being invariant and the latter is omitted in the new assumptions. The second assumption basically compensates for the possible degeneracy of the noise: it guarantees that through interaction of drift and noise, the noise will eventually spread into all space directions (compare also with the Hörmander condition in Theorem 2.13 and the conditions for existence and uniqueness of an invariant measure for OU-processes in Example 2.14).

We have the following theorem which yields that the novel quasipotential  $\tilde{V}^q$  determines the corresponding rate for the exit from the domain.

**Theorem 3.27** (Theorem 4 and 5 in [69]). *Let A3 hold true. Then, for all  $x \in \mathcal{D}_0$  it holds that*

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \ln \mathbb{E}_x(\tau) = \inf_{y \in \partial \mathcal{D}} \tilde{V}^q(z^*, y) =: \Delta \tilde{V}^q.$$

Further, denoting by  $E = \{y \in \partial \mathcal{D} : \tilde{V}^q(z^*, y) = \Delta \tilde{V}^q\}$  the set of points for which the infimum on the boundary  $\partial \mathcal{D}$  for  $\tilde{V}^q$  is attained, it holds that for any  $\delta > 0$

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P}_x(|X_\tau - E| < \delta) = 1.$$

### 3.4.2 OU-processes

We now come to the nice result, that for OU-processes the controllability gramian – and thus also the rate function as well as  $E$  the exit direction – is explicitly computable.

**Theorem 3.28** (Proposition 1.1 in [70] or [42]). *Consider the linear control system*

$$\dot{x}^u = Ax^u + \gamma u, \quad (3.15)$$

where  $A \in \mathbb{R}^{n \times n}$ ,  $\gamma \in \mathbb{R}^{n \times m}$ ,  $u \in \mathbb{R}^m$  and assume that  $\text{rk}[\gamma, A\gamma, \dots, A^{n-1}\gamma] = n$ . Then, the control

$$\hat{u}(s) = -\gamma^T B_{T-s}^T Q_T^{-1} (B_T a - b), \quad \text{where } B_t = e^{At}, \quad Q_t = \int_0^t B_r \gamma \gamma^T B_r^T dr$$

transfers any state  $a \in \mathbb{R}^n$  to  $b \in \mathbb{R}^n$  in time  $T$ , i.e.,  $x^{\hat{u}}(0) = a$ ,  $x^{\hat{u}}(T) = b$ .

Furthermore,

$$Q_{t_1} > Q_{t_2}, \quad \text{for } t_1 > t_2$$

and

$$\hat{u}(s) = \arg \min_{u \in L^2([0, T], \mathbb{R}^m)} \int_0^T |u(s)|^2 ds$$

is the unique minimiser of the cost function.

*Proof.* By the variation of constants formula the solution to (3.15) is

$$x^u(t) = e^{tA} x^u(0) + \int_0^t e^{(t-s)A} \gamma u(s) ds = B_t x^u(0) + \int_0^t B_{t-s} \gamma u(s) ds. \quad (3.16)$$

Further note that by the rank condition  $Q_t$  is invertible for all  $t > 0$ . In fact  $Q_t = Q_t^T$  and  $Q_t > 0$ ,  $t > 0$ . Let  $x^{\hat{u}}(0) = a$ , then by the definition of  $Q_t$

$$x^{\hat{u}}(T) = B_T a - \int_0^T B_{T-s} \gamma \gamma^T B_{T-s}^T ds Q_T^{-1} (B_T a - b) = b.$$

Also, for  $t_1 > t_2$  we have  $Q_{t_1} - Q_{t_2} = \int_{t_2}^{t_1} B_r \gamma \gamma^T B_r^T dr > 0$  by the Kalman rank condition.

Now compute the cost associated to  $\hat{u}$

$$\begin{aligned} \int_0^T |\hat{u}(s)|^2 ds &= \int_0^T |\gamma^T B_{T-s}^T Q_T^{-1} (B_T a - b)|^2 ds \\ &= \int_0^T (B_{T-s} \gamma \gamma^T B_{T-s}^T Q_T^{-1} (B_T a - b)) \cdot (Q_T^{-1} (B_T a - b)) ds \\ &= \left( \int_0^T B_{T-s} \gamma \gamma^T B_{T-s}^T ds Q_T^{-1} (B_T a - b) \right) \cdot (Q_T^{-1} (B_T a - b)) \\ &= (B_T a - b)^T Q_T^{-1} (B_T a - b) \end{aligned}$$

Let  $u$  be any other control such that  $x^u(0) = a$ ,  $x^u(T) = b$  which implies by (3.16)  $-\int_0^T B_{T-s} \gamma u(s) ds = B_T a - b$ . Then

$$\begin{aligned} \int_0^T u(s)^T \hat{u}(s) ds &= - \int_0^T u(s)^T \gamma^T B_{T-s}^T Q_T^{-1} (B_T a - b) ds \\ &= - \left( \int_0^T B_{T-s} \gamma u(s) ds \right)^T Q_T^{-1} (B_T a - b) \\ &= (B_T a - b)^T Q_T^{-1} (B_T a - b) = \int_0^T |\hat{u}(s)|^2 ds \end{aligned}$$

which implies that

$$0 \leq |u - \hat{u}|^2 = u^2 - 2u\hat{u} + \hat{u}^2 = u^2 - \hat{u}^2$$

and thus  $u^2 \geq \hat{u}^2$  which concludes the proof.  $\square$

By means of this theorem, the upcoming theorem, which will provide results for MFET for OU-processes, is an easy corollary.

**Theorem 3.29** (Theorem 6 in [69]). *Consider the SDE*

$$dX_t = AX_t dt + \varepsilon \gamma dW_t, \quad X_0 = x_0.$$

Let  $A, \gamma$  satisfy assumptions A2, define  $\mathcal{D} = \{x \in \mathbb{R}^n : |x| \leq r\}$  to be the ball with radius  $r$  around the origin and

$$\tau = \inf \{t > 0 : X_t \notin \mathcal{D}\}$$

to be the first exit time from the ball. It then holds that  $\forall x_0 \in D_0$  (with  $D_0$  as in Theorem 3.27)

$$\lim_{\varepsilon \searrow 0} \varepsilon^2 \log \mathbb{E}_{x_0}(\tau) = \frac{1}{2\lambda_{\max}(\Sigma_\infty)} r^2, \quad \text{with } \Sigma_\infty = Q_\infty.$$

Furthermore, the set of most probable exit points (as defined in Theorem 3.27) is given by the corresponding eigenvectors

$$E = \{x \in \mathbb{R}^n : \Sigma_\infty x = \lambda_{\max}(\Sigma_\infty)x, |x| = r\}.$$

*Proof.* This follows by the definition of the quasipotential  $\tilde{V}^q$ , Theorems 3.27 and 3.28 which gives  $\tilde{V}^q(0, a) = \frac{1}{2}a^T \Sigma_\infty^{-1} a$ . This implies  $\tilde{V}^q(0, a) \geq \frac{1}{2}\lambda_{\min}(\Sigma_\infty^{-1})|a|^2$  such that  $\inf_{a:|a|=r} \tilde{V}^q(0, a) = \frac{1}{2}\lambda_{\min}(\Sigma_\infty^{-1})r^2 = \frac{1}{2\lambda_{\max}(\Sigma_\infty)}r^2$ .  $\square$

This in turn enables us to establish a relation between MFETs and the relaxation time scale for possibly hypo-elliptic OU-processes. The following result can be found in our publication [54].

**Proposition 3.30.** *Let  $v \in \mathbb{R}^n$  be the normalised eigenvector of  $-A^T$  corresponding to the eigenvalue with the minimum real part given by  $\lambda^* = \min \{\operatorname{Re}(\lambda) : -Ax = \lambda x, x \neq 0\}$ . Introduce the splitting of  $v = v_{\ker} + v_{\text{Im}}$ , where  $v_{\ker} \in \ker(\Gamma)$ ,  $\Gamma = \gamma\gamma^T$ ,  $v \in \text{Im}(\Gamma)$  and let  $\lambda_{\min}(\Gamma) = \{\lambda > 0 : \Gamma x = \lambda x\}$ . Then*

$$\lambda^* \geq \frac{\lambda_{\min}(\Gamma)}{2\lambda_{\max}(\Sigma_\infty)} |v_{\text{Im}}|$$

*Proof.* Note that by Assumption A2  $v_{\text{Im}} \neq 0$ . We multiply the Lyapunov equation  $A\Sigma_\infty + \Sigma_\infty A^T = \Gamma$  from left and right by  $v^T$  and  $v$  respectively which gives  $2\lambda^* v^T \Sigma_\infty v = v^T \Gamma v$ . Now,  $v^T \Sigma_\infty v \leq \lambda_{\max}(\Sigma_\infty)$  and  $v^T \Gamma v = v_{\text{Im}}^T \Gamma v_{\text{Im}} \geq \lambda_{\min}(\Gamma) |v_{\text{Im}}|$  which yields the assertion.  $\square$

Thus, for hypo-elliptic OU-processes, we can determine the rate of convergence to equilibrium, which is given by the smallest realpart of the eigenvalue  $\lambda^*$  of  $-A$ , as well as the rate with which the MFET scales as  $\varepsilon \rightarrow 0$  being proportional to  $1/(2\lambda_{\max}(\Sigma_\infty))$ . Even the spectral gap is known, which is also given by  $\lambda^*$  (see [53]).

### 3.4.3 Underdamped Langevin dynamics

In this paragraph we derive concrete expressions for MFETs for underdamped Langevin dynamics by means of the previously reviewed theory. To this end, we rely on results by Newman and Krishnaprasad [56], who computed the associated controllability function.

Consider the *underdamped Langevin dynamics* in  $\mathbb{R}^{2n}$

$$dZ_t = \begin{pmatrix} 0 & I_{n \times n} \\ -I_{n \times n} & -\gamma \end{pmatrix} \nabla H(Z_t) dt + \varepsilon \begin{pmatrix} 0 \\ \sigma \end{pmatrix} dB_t, \quad (3.17)$$



where  $Z = (q, p)$ , with  $q, p \in \mathbb{R}^n$  and  $H : \mathbb{R}^{2n} \rightarrow \mathbb{R}$  being the system's Hamiltonian,  $\gamma, \sigma \in \mathbb{R}^{n \times n}$  comprising the friction and forcing coefficients respectively and let the following fluctuation-dissipation relation hold

$$\frac{\gamma_{ij}}{(\sigma\sigma^T)_{ij}} = l, \quad \forall i, j, \quad \text{where } l > 0. \quad (3.18)$$

We begin with stating the result which provides us with the controllability function for underdamped Langevin dynamics. To this end, introduce the to (3.17) associated controlled ODE

$$\dot{Z}_t^u = \begin{pmatrix} 0 & I_{n \times n} \\ -I_{n \times n} & -\gamma \end{pmatrix} \nabla H(Z_t) + \varepsilon \begin{pmatrix} 0 \\ \sigma \end{pmatrix} u_t, \quad (3.19)$$

for which the noise is again replaced by the control  $u_t$ .

Different to the case of linear SDEs one cannot hope to find an explicit expression for the optimal control, steering the system from  $a$  to  $b$  within time  $T$ . Instead one has to resort to solving the associated *Hamilton-Jacobi-Bellman (HJB) equation* by which we obtain the minimum of the associated cost directly (for a proof of this result see, e.g., Theorem 3.2 in [65]): the HJB equation corresponding to (3.19) with  $\varepsilon = 1$ , seeking the optimal control  $u$  such that  $\int_{-\infty}^0 |u_t|^2 dt$  subject to  $z^u(-\infty) = 0$ ,  $z^u(0) = x$  is minimised, reads

$$\begin{pmatrix} 0 & I_{n \times n} \\ -I_{n \times n} & -\gamma \end{pmatrix} \nabla H \cdot \nabla L_c + \frac{1}{2} (\nabla L_c)^T \begin{pmatrix} 0 & 0 \\ 0 & \sigma\sigma^T \end{pmatrix} \nabla L_c = 0, \quad L_c(0) = 0$$

and its solution  $L_c$  satisfies indeed

$$L_c(x) = \inf_u \left\{ \frac{1}{2} \int_{-\infty}^0 |u_t|^2 dt, z^u(-\infty) = 0, z^u(0) = x \right\},$$

i.e.,  $L_c(z)$  gives the minimum energy needed to steer from w.l.o.g. 0 to any other state  $x$  in arbitrary time. Now, this result has been used in [56] and equips us with the following convenient form of  $L_c$ .

**Theorem 3.31** (Theorem 3.2 in [65], Theorem 2.5 in [56]). *Assume that (3.18) and  $\nabla H(0) = 0$  hold true. Further assume there exists a domain  $\mathcal{D}$  with  $0 \in \mathcal{D}$  such that  $\nabla^2 H(z) > 0 \quad \forall z \in \mathcal{D}$  holds. Then the unique controllability function of system (3.17) with  $\varepsilon = 1$  is given by*

$$L_c(z) = 2lH(z) + C \quad \forall z \in \mathcal{D},$$

where  $C$  is chosen such that  $L_c(0) = 0$ .

The combination of Theorem 3.31 and Theorem 3.27 directly establishes the following corollary.

**Corollary 3.32.** *Let the assumptions of Theorem 3.27 for  $b(z) = \begin{pmatrix} 0 & I_{n \times n} \\ -I_{n \times n} & -\gamma \end{pmatrix} \nabla H(z)$  together with the fluctuation-dissipation relation (3.18) hold true and choose the domain  $\mathcal{D} \subset \mathbb{R}^{2n}$  accordingly. Assume that w.l.o.g.  $0 \in \mathcal{D}$  is the asymptotically stable fixed point.*

*Then, for any  $z \in \mathcal{D}_0$ , with  $\mathcal{D}_0$  as in (3.14), it holds that*

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log \mathbb{E}_z(\tau) = \inf_{y \in \partial \mathcal{D}} 2LH(y) + C. \quad (3.20)$$

*Proof.* This follows directly by employing Theorem 3.31 and Theorem 3.27.  $\square$

**Remark 3.33.** If  $\text{rk}(\sigma) = n$  and  $\nabla_q^2 H(0) \neq 0$  the rank condition (ii) of A3 is fulfilled.

We illustrate the results of Corollary 3.32 in Section 6.3 on the example of a double pendulum.

### 3.4.4 A relation between the spectral gap and a MFET for underdamped Langevin dynamics

Now, let the Hamiltonian  $H$  be given by  $H(q, p) = \frac{1}{2}|p|^2 + V(q)$ . Provided that  $V$  satisfies certain assumption, we can apply our previous result, Corollary 3.32, and establish a relation between a MFET and the spectral gap, bearing the same structure known from Theorem 3.26 that relates MFETs and eigenvalues of the generator for overdamped Langevin dynamics.

To this end, we state a result by [32], giving a description of the generators spectrum for underdamped Langevin.

**Theorem 3.34** ([32], also see Theorem 4.9 in [66]). *Let the Hamiltonian  $H$  be given by  $H(q, p) = \frac{1}{2}|p|^2 + V(q)$ , where  $V$  is smooth and satisfies the conditions of Theorem 3.26. Then  $\exists \varepsilon^* > 0$  such that for all  $0 < \varepsilon < \varepsilon^*$  the associated generator  $\mathcal{L}$  has  $n$  simple eigenvalues  $\lambda_1 = 0 > \lambda_2 > \dots > \lambda_n$  which satisfy for  $k = 2, \dots, n$ ,*

$$\lambda_k \approx -\frac{|\lambda^-|}{2\pi} \sqrt{\frac{|\det(\nabla^2 V(x_k))|}{|\det(\nabla^2 V(z_{k-1}))|}} \exp\left(-2\frac{V(z_{k-1}) - V(x_k)}{\varepsilon^2}\right),$$

where  $\lambda^-$  is the unique negative eigenvalue of the matrix  $\begin{pmatrix} 0 & I_{n \times n} \\ \nabla^2 V(z_{k-1}) & \gamma \end{pmatrix}$  and we use the same notation as in Theorem 3.26.

The following corollary establishes a connection between the spectral gap of the generator and a MFET for underdamped Langevin dynamics, which – to our knowledge – is new, dealing with an instance of a hypo-elliptic process here.

We use the same notational conventions as in Theorem 3.26 and we recall that  $z_1$  is the maximum of all minimal saddles, where each minimal saddle connects a minimum of  $V$  with the global minimum  $x_1$  of  $V$ . The minimum which corresponds to the minimal saddle  $z_1$  in order to reach the global minimum  $x_1$  is denoted by  $x_2$ . For a more detailed definition of these quantities see Section 3.3.1.

**Corollary 3.35.** *Consider the Langevin dynamics (3.17), let w.o.l.g.  $x_2 = 0, V(x_2) = 0$  and assume that  $x_2$  is an asymptotically stable fixed point for the domain  $\mathcal{D} = \{z \in \mathbb{R}^{2n} : |z| \leq |z_1|\}$ . Assume that  $\mathcal{D}$  satisfies conditions A3 and that the fluctuation-dissipation relation (3.18) holds with  $l = 1$ . If  $\inf_{y \in \partial \mathcal{D}} H(y) = V(z_1)$  then, for all  $z_0 \in \mathcal{D}_0$  (with  $\mathcal{D}_0$  as in (3.14)) and small enough  $\varepsilon$  it holds that*

$$\lambda_2 \simeq -1/\mathbb{E}_{z_0}(\tau),$$

where  $\tau = \inf \{t > 0 : Z_t \notin \mathcal{D}\}$ .

Here  $\simeq$  is used for the logarithmic equivalence of the two quantities (also see the proof below).

Let us comment on the condition  $\inf_{y \in \partial \mathcal{D}} H(y) = V(z_1)$ . Note that

$$\{z \in \mathbb{R}^{2n} : z = (q, 0), |q| = |z_1|\} \subset \left\{z \in \mathbb{R}^{2n} : H(z) = \inf_{y \in \partial \mathcal{D}} H(y)\right\},$$

for  $H(q, p) = \frac{1}{2}|p|^2 + V(q)$ . We do not need the minimum of  $H$  in  $\partial \mathcal{D}$  to be unique, hence the required condition  $\inf_{y \in \partial \mathcal{D}} H(y) = V(z_1)$  can be rephrased as  $V(q) \geq V(z_1)$  for all  $q \in \mathbb{R}^n$  with  $|q| = |z_1|$ .

*Proof of Corollary 3.35.* The proof is a mere combination of the results of Theorem 3.34 and Corollary 3.32 which hold under the assumptions: the first one states that for small enough  $\varepsilon$  we have  $\lambda_2 \approx -c e^{-2\frac{V(z_1)}{\varepsilon^2}}$  such that  $\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log(-\lambda_2) = \lim_{\varepsilon \rightarrow 0} \varepsilon^2 \left( \log(c) - 2\frac{V(z_1)}{\varepsilon^2} \right) = -2V(z_1)$  and the second one gives  $\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log \mathbb{E}_{z_0}(\tau) = 2V(z_1)$ . Thus, for small enough  $\varepsilon$  we have  $-\lambda_2 \approx 1/\mathbb{E}_{z_0}(\tau)$ .  $\square$

**Remark 3.36.** This result can straightforwardly be extended to relate the  $k^{\text{th}}$  eigenvalue with the exit from the  $k^{\text{th}}$  minimum, where  $k = 2, \dots, n$ , as long as the corresponding domain fulfils the required assumptions.

**Remark 3.37.** Note that even though this result has the same flavour as the one for the overdamped Langevin dynamics, Theorem 3.26, there is the difference of the first exit time from the domain corresponding to the minimum  $x_2$  versus the first passage time into the domain around the global minimum  $x_1$ . Intuitively speaking, the exit and passage time should differ by a factor of 2, since once the process has left the domain around  $x_2$ , meaning it has reached the minimal saddle  $z_1$ , the dynamics will quickly either reach the domain around global minimum  $x_1$  or return to the local minimum  $x_2$  (also see [19] for a discussion).

Admittedly, we are not providing a relation to the relaxation behaviour here. There are various works concerned with the rate of convergence for underdamped Langevin dynamics: in [2] convergence in relative entropy is derived, whereas convergence in  $L^2$  has been proven in [33] and [20].



## MODEL REDUCTION

Working with complex multi-scale processes requires a lot of computational power in order to treat these problems simulation-wise. For example, the simulation of the motion of a single protein involves three spatial variables – and sometimes, additionally, three momentum variables – for thousands of atoms. Moreover, the smallest time scale involved has usually the order of femtoseconds, that is,  $10^{-15}$  seconds. Aiming at simulating the dynamics of the protein up to the order of a second, which is the time scale on which conformational changes occur, this is hardly feasible. Since the main interest usually lies exactly in the simulation of the slow processes, in this case the conformational changes of the molecule, some help is needed. This is where model reduction techniques come into play, providing us with models that live on a much lower dimensional space, e.g., in the case when the conformational change is described by a change in some dihedral angle this can even be a one-dimensional torus. Furthermore, the simulation step can now be increased, since the fast degrees of freedom, such as quickly vibrating bonds, have been eliminated. Hence, model reduction techniques are important tools to study long time properties of complex multi-scale dynamics. Obviously, this postulates that the model reduction yields a good approximation of the original dynamics in some sense.

In this chapter the focus lies on the model reduction technique of *effective dynamics* also referred to as *conditional expectations* which has been introduced in [43]. In the first works on this technique, error bounds for reversible gradient systems were proven (see [43, 44]). These have recently been augmented in [45] to non-reversible processes for which the authors prove pathwise, i.e. trajectorial error bounds if the reduced models variable is an affine function of the full dynamics. Here, we extend the results of [43], by considering non-reversible two-scale processes for which effective dynamics yields a reduced model for the slow degrees of freedom only. After introducing the concept of effective dynamics, Section 4.2 is devoted to the derivation of different error bounds for the reduced dynamics. Since two-scale processes are the natural objects for the model reduction technique named *averaging*, we will compare the two methods in Section 4.3. Again, we find that irreversibility plays a crucial role in the analysis and we derive sufficient conditions which guarantee conformity of the two approaches for OU-processes in the limit of infinite time scale

separation. In the last section of the chapter, Section 5.2, we devote ourselves to the issue of sampling conditional distributions for non-reversible processes, which are for example needed for computing effective dynamics coefficients. Once more the usual approach for reversible processes fails for non-reversible ones and we design a projection for possibly non-reversible dynamics, which yields the correct distribution.

Let us now start by giving the precise definition of a two-scale process followed by a teaser for Section 4.3 dealing with the comparison.

We refer to a *two-scale process* if the corresponding SDE is of the form

$$\begin{aligned} dX_t &= f_1(X_t, Y_t) dt + \gamma_1(X_t, Y_t) dW_t, \quad X_{t=0} = X_0 \\ dY_t &= \frac{1}{\varepsilon} f_2(X_t, Y_t) dt + \frac{1}{\sqrt{\varepsilon}} \gamma_2(X_t, Y_t) dU_t, \quad Y_{t=0} = Y_0, \end{aligned} \quad (4.1)$$

where  $0 < \varepsilon \ll 1$  is the time scale separation parameter,  $X \in \mathbb{R}^{n_x}$ ,  $Y \in \mathbb{R}^{n_y}$ ,  $f_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{n_x}$ ,  $\gamma_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{n_x \times n_x}$ ,  $f_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{n_y}$ ,  $\gamma_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{n_y \times n_y}$ ,  $n = n_x + n_y$  and  $W, U$  are standard  $n_x, n_y$ -dimensional Brownian motions respectively. Let us also introduce the more compact form of (4.1) by writing  $Z = (X, Y) \in \mathbb{R}^n$  and

$$\begin{aligned} dZ_t &= f^\varepsilon(Z_t) dt + \gamma^\varepsilon(Z_t) dB_t, \quad (4.2) \\ \text{with } f^\varepsilon &= \begin{pmatrix} f_1 \\ f_2/\varepsilon \end{pmatrix}, \quad \gamma^\varepsilon = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2/\sqrt{\varepsilon} \end{pmatrix}, \quad B = \begin{pmatrix} W \\ U \end{pmatrix}. \end{aligned}$$

Note, that the time of the fast variable  $Y$  is indeed accelerated by a factor of  $1/\varepsilon$  compared to the slow dynamics  $X$ .

### A reversible example

Before we learn the precise methodologies of what we refer to as *averaging* and *conditional expectations* let us present a simple example given by the reversible dynamics

$$\begin{aligned} dX_t &= \nabla_x V(X_t, Y_t) dt + dW_t, \quad X_{t=0} = X_0 \\ dY_t &= \frac{1}{\varepsilon} \nabla_y V(X_t, Y_t) dt + \frac{1}{\sqrt{\varepsilon}} dU_t, \quad Y_{t=0} = Y_0, \end{aligned}$$

assuming that the potential  $V$  grows sufficiently fast such that we have a unique invariant measure  $\mu \sim e^{-V}$ . The method of conditional expectations computes the effective dynamics

$$d\bar{X}_t = \nabla \bar{V}(\bar{X}_t) dt + dW_t,$$

where

$$\bar{V}(x) = \int V(x, y) d\mu_x(y) \quad \text{and} \quad \mu_x(y) = \frac{\mu(x, y)}{\int \mu(x, y) dy}$$

is the conditional of  $\mu$  for given  $X = x$ . Clearly,  $\mu_x \sim e^{-V_x(y)}$  where the subscript  $x$  indicates that  $x$  is not a variable any more, but fixed. The method of conditional

expectations is based on the idea of *optimal prediction* for the variable of interest in an  $L^2$  sense (see [17]) together with the idea that the conditional measure of the fast variable, given the slow variable, will quickly relax to its equilibrium and thus computes averages of the slow process' coefficients with respect to the conditional of the invariant measure of the full process.

On the other hand, the averaging principle, yields the equation

$$dX_t^{\text{av}} = V^{\text{av}}(X_t^{\text{av}}) dt + dW_t, \text{ where } V^{\text{av}}(x) = \int V(x, y) d\nu_x(y)$$

and  $\nu_x$  is the invariant measure of the fast dynamics  $Y_t$  in which the slow  $X_t$  is treated as constant. This method is based on the decomposition of the generator into orders of  $1/\varepsilon$  and deducing the action of the lowest order or slowest time scale by solution methods for the associated PDE. For our example, the invariant measure of  $Y_t$  for fixed  $X_t = x$  is  $\nu_x \sim e^{-V_x(y)}$ , where again the subscript  $x$  indicates that  $x$  is not a variable any more, but fixed.

Thus, for reversible dynamics, the two methods agree, which is not necessarily the case for non-reversible dynamics, as the upcoming example shows.

### A non-reversible example

In order to make the differences in averaging and conditional expectations obvious at the same time motivating the relevance of our work, we next give a simple non-reversible example.

To this end, consider the two-dimensional SDE

$$\begin{aligned} dX_t &= (-X_t + Y_t) dt \\ dY_t &= -\frac{1}{\varepsilon} Y_t dt + \frac{1}{\sqrt{\varepsilon}} dB_t, \quad \varepsilon > 0, \end{aligned}$$

which is an OU-process fulfilling assumption A2 for any fixed  $\varepsilon > 0$  and its invariant measure is  $\mu \sim \mathcal{N}(0, \Sigma)$ , where  $\Sigma = \begin{pmatrix} \frac{\varepsilon}{2(\varepsilon+1)} & \frac{\varepsilon}{2(\varepsilon+1)} \\ \frac{\varepsilon}{2(\varepsilon+1)} & \frac{1}{2} \end{pmatrix}$ .

Averaging will give that  $X_t^\varepsilon \rightarrow X_t^{\text{av}}$  as  $\varepsilon \rightarrow 0$ , with  $X_t^{\text{av}}$  being the solution to the ODE

$$\dot{X}_t^{\text{av}} = -X_t^{\text{av}}.$$

On the other hand, the method of conditional expectations yields, for any  $\varepsilon > 0$ , the approximation  $X_t^{\varepsilon^e}$  solution to the ODE

$$\dot{X}_t^{\varepsilon^e} = 0.$$

Clearly, the behaviour of the two equations is qualitatively very different: the solution of the averaged equation will - independent of the initial value - always decay towards 0, whereas the effective dynamics will constantly stay at the initial value. This example is meant to serve as an appetizer and will be discussed in more detail in Section 4.3.2.

## 4.1 Conditional expectations

We now derive the first model reduction technique, which we refer to as *conditional expectation* or *effective dynamics*. The idea of using conditional expectations in order to formulate equations for certain variables of interest has first been put forward by Chorin and Kupferman in the article [17], and they refer to the method as *optimal prediction*. The idea is to resolve only certain variables of interest, while using available statistical information of the other unresolved variables in order to make an optimal prediction for the dynamics of the resolved variables in an  $L^2$ -sense. That is, the functions in  $x$  and  $y$  describing the full dynamics are approximated by the corresponding functions in  $x$  which minimise the  $L^2$  distance between the two. In other words, the approximation is performed by computing conditional expectations for which the available statistical information enters.

These ideas have been formalised and further developed by Legoll and Lelièvre in [43].

Throughout this chapter we assume that drift and diffusion coefficients of (4.1) are such that  $(Z_t)_{t \geq 0}$  admits a unique invariant measure  $d\mu(z) = \frac{1}{\tilde{Z}} e^{-V(z)} dz$ , where  $\tilde{Z}$  is the normalisation constant, which entails by Proposition 2.11 that  $f^\varepsilon$  can be recast as

$$\begin{aligned} f^\varepsilon(z) &= -\Gamma^\varepsilon(Z_t) \nabla V^\varepsilon(Z_t) + \nabla \cdot \Gamma^\varepsilon(Z_t) + D^\varepsilon(Z_t), \\ \text{with } \Gamma^\varepsilon &= \gamma^\varepsilon (\gamma^\varepsilon)^\text{T} = \begin{pmatrix} \gamma_1 \gamma_1^\text{T} & 0 \\ 0 & \gamma_2 \gamma_2^\text{T} / \varepsilon \end{pmatrix}, \quad \nabla \cdot (D^\varepsilon \mu) = 0. \end{aligned} \quad (4.3)$$

### Derivation of the effective dynamics

Looking at (4.1) the first equation describes the dynamics for  $X_t$  alone but it clearly depends on  $Y_t$ . Following the work of Gyöngy [28] we introduce the closed dynamics by taking expectations with respect to the conditional time marginal  $\rho_{t,x}$  - the subscript  $x$  indicating the condition  $X_t = x$  - that is,

$$\begin{aligned} d\hat{X}_t &= \hat{f}(t, \hat{X}_t) dt + \hat{\gamma}(t, \hat{X}_t) dW_t, \quad \text{where} \\ \hat{f}(t, x) &:= \mathbb{E}[f_1(X_t, Y_t) | X_t = x] = \int_{\mathbb{R}^{n_y}} f_1(x, y) d\rho_{t,x}(y), \quad \text{and} \\ \hat{\gamma}(t, x) &:= \mathbb{E}[\gamma_1(X_t, Y_t) \gamma_1(X_t, Y_t)^\text{T} | X_t = x]^{1/2} \\ &= \left( \int_{\mathbb{R}^{n_y}} \gamma_1(X_t, Y_t) \gamma_1(X_t, Y_t)^\text{T} d\rho_{t,x}(y) \right)^{1/2}. \end{aligned} \quad (4.4)$$

By the work of Gyöngy it is known that the marginals (i.e. push-forward under the coarse graining map  $\xi(z) = x$ ) of  $X_t$  and  $\hat{X}_t$  coincide for each  $t$ , i.e.,

$$\hat{\rho}_t := \text{law}(X_t) = \text{law}(\hat{X}_t).$$



In particular this means that  $\hat{\rho}_t$  solves the Fokker-Planck equation that corresponds to (4.4)

$$\partial_t \hat{\rho}_t = -\nabla \cdot (\hat{f} \hat{\rho}_t) + \nabla^2 : (\hat{\Gamma} \hat{\rho}_t), \quad \hat{\rho}_{t=0} = \hat{\rho}_0, \quad \text{where } \hat{\Gamma} = \hat{\gamma} \hat{\gamma}^T. \quad (4.5)$$

For a proof see [43, Lemma 2.3] or [21, Proposition 2.8].

Now, computationally (4.4) is still not a big improvement to the original equation (4.1), as the functions  $\hat{f}$  and  $\hat{\gamma}$  are time dependent and hence have to be recomputed at each time  $t$ . In particular, they depend on  $\rho_{t,x}$  which is the conditional of  $\rho_t$ , the solution of the original Fokker-Planck equation, and computing  $\rho_t$  and  $\rho_{t,x}$  (or averages with respect to it) is not an easy task itself. Assuming that for each fixed  $x$  the conditional density  $\rho_{t,x}$  converges fast to its equilibrium  $\mu_x$  – where again the subscript  $x$  indicated the condition  $X = x$  – the next approximation unfolds and equips us with the time independent SDE, the so-called *effective dynamics*

$$\begin{aligned} d\bar{X}_t &= \bar{f}(\bar{X}_t) dt + \bar{\gamma}(\bar{X}) dW_t, \quad \text{where} \\ \bar{f}(x) &= \int_{\mathbb{R}^{n_y}} f_1(x, y) d\mu_x(y), \quad \text{and} \\ \bar{\gamma}(x) &= \left( \int_{\mathbb{R}^{n_y}} \gamma_1(x, y) \gamma_1(x, y)^T d\mu_x(y) \right)^{1/2}. \end{aligned} \quad (4.6)$$

Along with the dynamics, we introduce the notation

$$\bar{\rho}_t := \text{law}(\bar{X}_t).$$

In the following, we will compare  $\bar{\rho}_t$  to  $\hat{\rho}_t$ , that is, the distribution of the effective to the distribution of the original process.

**Remark 4.1** (Optimal prediction). The functions  $\bar{f}, \bar{\gamma}$  are optimal approximations of  $f_1, \gamma_1$  in the  $L^2_\mu$ -sense, i.e.,

$$\begin{aligned} \bar{f}(X) &= \mathbb{E}_\mu(f_1(Z)|X) = \arg \min_{f \in L^2_\mu} \mathbb{E}_\mu((f_1(Z) - f(X))^2) \\ \text{and } \bar{\gamma}(X) &= \mathbb{E}_\mu(\gamma_1(Z)|X) = \arg \min_{\gamma \in L^2_\mu} \mathbb{E}_\mu((\gamma_1(Z) - \gamma(X))^2). \end{aligned}$$

## 4.2 Error estimates

Aiming at deriving reduced models that approximate the original dynamics well, this section furnishes the reader with error estimates in different norms for model reduction, conducted via conditional expectations, as explained in the previous section. We will start comparing the time  $t$  marginals of the original and effective dynamics, given by  $\hat{\rho}_t$  and  $\bar{\rho}_t$  respectively, which are useful when computing averages with respect to the time  $t$  distribution of the process. We will derive the estimates first in relative entropy in Section 4.2.1 and second in Wasserstein distance in Section 4.2.2. The latter distance is weaker by the Talagrand Inequality (see Definition 3.19), but

the estimate has one favourable property, which is that, even for diffusion matrices  $\Gamma = \Gamma(X, Y)$  depending on both slow and fast variables, the errorbound is linear in the time scale separation parameter  $\varepsilon$  and thus the error vanishes as  $\varepsilon \rightarrow 0$ . In Section 4.2.3 we finally state results on the error between the law of the paths which is useful when computing averages over the paths, as done e.g. for mean first exit times or other dynamic properties of the dynamics.

Apart from using the definitions of Chapter 3, we need one more inequality that will be of central importance to the proofs. It is given by the following proposition.

**Proposition 4.2** (Theorem 1.1 in [11]). *Let  $T > 0$  be fixed. Let  $\nu_t, \mu_t \in \mathcal{P}(\mathbb{R}^n)$  be smooth solutions to the Fokker-Planck equations*

$$\begin{aligned}\partial_t \nu_t &= -\nabla \cdot (b_1 \nu_t) + \nabla^2 : (A_1 \nu_t), \\ \partial_t \mu_t &= -\nabla \cdot (b_2 \mu_t) + \nabla^2 : (A_2 \mu_t),\end{aligned}$$

where  $b_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $A_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $i = 1, 2$ . Define

$$h_t = (b_1 - b_2) + (\nabla \cdot A_1 - \nabla \cdot A_2) + (A_1 - A_2) \nabla \ln \mu_t$$

and assume that  $|h_t|_{A_1^{-1}} \in L^2_{\mu_t}(\mathbb{R}^n \times [0, T])$ . Further let the coefficients fulfil the following conditions:

$(1 + |z|)^{-2} |(A_1)_{ij}|$ ,  $(1 + |z|)^{-1} |b_1|$ ,  $(1 + |z|)^{-1} |h_t| \in L^1_{\mu_t}(\mathbb{R}^n \times [0, T])$ . Then

$$\mathcal{H}(\mu_T | \nu_T) \leq \mathcal{H}(\mu_0 | \nu_0) + \int_0^T \int_{\mathbb{R}^n} |h_t|_{A_1^{-1}}^2 d\mu_t dt.$$

Due to the considerations of the thesis [67], we will refer to the last inequality as the *Free-energy-rate-functional-relative-Fisher-Information Inequality* or in short *FIR* inequality.

Since we want to compare conditional expectations and averaging, we keep  $\varepsilon$ -dependency of the functions at hand. Yet, for notational simplicity we drop the  $\varepsilon$ -superscript throughout the section, since possibly all functions of interest - that is  $f$ ,  $D$ ,  $V$ ,  $\rho_t$ ,  $\rho_{t,x}$ ,  $\mu$ ,  $\mu_x$  - depend on  $\varepsilon$ . Only the diffusion matrix retains the previous notation, i.e., we write

$$\Gamma^\varepsilon = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2^\varepsilon \end{pmatrix}, \quad \Gamma_2^\varepsilon = \frac{1}{\varepsilon} \Gamma_2.$$

Further, introduce the notation  $D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}$ ,  $D_1 : \mathbb{R}^n \rightarrow \mathbb{R}^{n_x}$ ,  $D_2 : \mathbb{R}^n \rightarrow \mathbb{R}^{n_y}$ .

The following assumptions are needed in both estimates, the relative entropy estimate as well as the Wasserstein estimate.

(B1) The diffusion coefficients  $\Gamma_1, \Gamma_2$  are uniformly bounded from below, i.e.,

$$\exists \lambda_{\min}(\Gamma_1), \lambda_{\min}(\Gamma_2) \in \mathbb{R}^+ : \forall z \in \mathbb{R}^n \quad \Gamma_1 \geq \lambda_{\min}(\Gamma_1), \quad \Gamma_2 \geq \lambda_{\min}(\Gamma_2).$$

(B2) The conditional measure  $\mu_x$  satisfies a static Logarithmic-Sobolev-Inequality (LSI) uniformly for all  $x \in \mathbb{R}^{n_x}$  with constant  $\alpha_{\text{LSI}}$ . In particular it holds that

$$\mathcal{H}(\rho_{t,x}|\mu_x) \leq \frac{1}{2\alpha_{\text{LSI}}} \mathcal{R}(\rho_{t,x}|\mu_x), \quad \forall x \in \mathbb{R}^{n_x}.$$

**Assumption A4.** We say that Assumption A4 holds true if (B1)-(B2) together with the assumptions of Proposition 4.2 hold.

### 4.2.1 Relative Entropy

In this section we state and prove estimates for the time  $t$  marginals of the law of the effective (4.6) and original dynamics (4.4) given by  $\bar{\rho}_t$  and  $\hat{\rho}_t$  respectively.

We begin by stating the assumptions of this section. In the case that the diffusion coefficient of the slow process is constant these assumptions boil down to the drift  $f_1$  being Lipschitz continuous in  $y$  and the diffusion coefficients of the slow process being non-degenerate (see Corollary 4.6). In the general case, when  $\Gamma_1 = \Gamma_1(x, y)$ , the assumptions read:

(H1) The function  $C(x, y) := \bar{\Gamma}(x)^{-1/2} D_1(x, y) - \bar{\Gamma}(x)^{1/2} \nabla_x V(x, y)$  is Lipschitz in the second argument, i.e.  $|C(x, y_1) - C(x, y_2)| \leq L_C |y_1 - y_2|$

(H2)  $\bar{\Gamma}^{-1/2}(\Gamma_1 - \bar{\Gamma})$  is uniformly bounded from above, i.e.,  $\exists \kappa < \infty$  such that  $\forall x \in \mathbb{R}^{n_x}, y \in \mathbb{R}^{n_y} : \bar{\Gamma}^{-1/2}(x)(\Gamma_1(x, y) - \bar{\Gamma}(x)) \leq \kappa I_{n \times n}$ .

**Remark 4.3.** If the diffusion is uniformly bounded from above and below, i.e.,  $\exists m, M$  such that  $0 < m < \Gamma(z) < M < \infty$ , then (H2) holds true.

**Remark 4.4.** The results also include generalised reversible processes, which correspond to  $C = 0$  in the above decomposition. The case  $C = 0$  and  $\Gamma = I_{n \times n}$  corresponds to the overdamped Langevin dynamics and has also been treated in [21].

Let us now state the main result of the section.

**Theorem 4.5.** *Consider the dynamics given by (4.1) and the approximation for the slow dynamics given by (4.6). Let A4, (H1) and (H2) hold true. If the initial distributions agree, i.e.,  $\hat{\rho}_0 = \bar{\rho}_0$  then for any  $t > 0$  it holds that*

$$\mathcal{H}(\hat{\rho}_t|\bar{\rho}_t) \leq \left( \varepsilon \frac{L_C^2}{\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_2)} + \frac{\kappa^2}{\lambda_{\min}(\Gamma_1)} \right) \mathcal{H}(\rho_0|\mu).$$

If  $\Gamma_1 = \Gamma_1(x)$ , i.e., the diffusion only depends on the slow variable, the result simplifies to:

**Corollary 4.6.** *Let  $\Gamma_1 = \Gamma_1(x)$  and assume the conditions of Theorem 4.5 hold true. Then*

$$\mathcal{H}(\hat{\rho}_t|\bar{\rho}_t) \leq \varepsilon \frac{L^2}{\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_2) \lambda_{\min}(\Gamma_1)} \mathcal{H}(\rho_0|\mu),$$

where  $L$  is the Lipschitz constant of  $f_1 - \nabla_x \cdot \Gamma_1 = -\Gamma_1 \nabla_x V + D_1$  in  $y$ .

Note that in particular  $L$  is independent of  $\varepsilon$ , since  $f_1$  and  $\Gamma_1$  are independent of  $\varepsilon$  by the structure of the equation.

For the proof we will make use of the following useful lemma.

**Lemma 4.7.** *Let (B1)-(B2) hold true, then*

$$\int_{\mathbb{R}^{n_x}} \mathcal{H}(\rho_{t,x}|\mu_x)\hat{\rho}_t(x)dx \leq -\frac{\varepsilon}{2\alpha_{\text{LSI}}\lambda_{\min}(\Gamma_2)}\frac{d}{dt}\mathcal{H}(\rho_t|\mu).$$

*Proof of Lemma 4.7.* According to assumption (B1), (B2) and using the notation  $\Gamma_2^\varepsilon = \frac{1}{\varepsilon}\Gamma_2$  we have

$$\begin{aligned} \mathcal{H}(\rho_{t,x}|\mu_x) &\leq \frac{1}{2\alpha_{\text{LSI}}}\mathcal{R}(\rho_{t,x}|\mu_x) = \frac{1}{2\alpha_{\text{LSI}}}\int_{\mathbb{R}^{n_y}} \left| \nabla_y \ln \frac{\rho_{t,x}}{\mu_x} \right|^2 d\rho_{t,x}(y) \\ &\leq \frac{\varepsilon}{2\alpha_{\text{LSI}}\lambda_{\min}(\Gamma_2)}\int_{\mathbb{R}^{n_y}} \left| \nabla_y \ln \frac{\rho_{t,x}\hat{\rho}_t}{\mu_x\hat{\mu}} \right|_{\Gamma_2^\varepsilon}^2 d\rho_{t,x}(y) \\ &\leq \frac{\varepsilon}{2\alpha_{\text{LSI}}\lambda_{\min}(\Gamma_2)}\int_{\mathbb{R}^{n_y}} \left| \nabla_y \ln \frac{\rho_t}{\mu} \right|_{\Gamma_2^\varepsilon}^2 + \left| \nabla_x \ln \frac{\rho_t}{\mu} \right|_{\Gamma_1}^2 d\rho_{t,x}(y) \\ &= \frac{\varepsilon}{2\alpha_{\text{LSI}}\lambda_{\min}(\Gamma_2)}\int_{\mathbb{R}^{n_y}} \left| \nabla \ln \frac{\rho_t}{\mu} \right|_{\Gamma^\varepsilon}^2 d\rho_{t,x}(y). \end{aligned}$$

Integrating the inequality with respect to  $\hat{\rho}_t$  and using Lemma 3.9 yields

$$\begin{aligned} \int_{\mathbb{R}^{n_x}} \mathcal{H}(\rho_{t,x}|\mu_x)\hat{\rho}_t(x) dx &\leq \frac{\varepsilon}{2\alpha_{\text{LSI}}\lambda_{\min}(\Gamma_2)}\int_{\mathbb{R}^n} \left| \nabla \ln \frac{\rho_t}{\mu} \right|_{\Gamma^\varepsilon}^2 d\rho_t(z) \\ &= -\frac{\varepsilon}{2\alpha_{\text{LSI}}\lambda_{\min}(\Gamma_2)}\frac{d}{dt}\mathcal{H}(\rho_t|\mu). \end{aligned}$$

□

*Proof of Theorem 4.5 and Corollary 4.6.* The densities  $\hat{\rho}_t$  and  $\bar{\rho}_t$  of the Gyöngy and effective dynamics solve the Fokker-Planck equations corresponding to (4.4) and (4.6), i.e.

$$\begin{aligned} \partial_t \hat{\rho}_t &= -\nabla_x \cdot (\hat{f}\hat{\rho}_t) + \nabla_x^2 : (\hat{\Gamma}\hat{\rho}_t), & \hat{\rho}_{t=0} &= \hat{\rho}_0, & \text{where } \hat{\Gamma} &= \hat{\gamma}\hat{\gamma}^\text{T} \\ \partial_t \bar{\rho}_t &= -\nabla_x \cdot (\bar{f}\bar{\rho}_t) + \nabla_x^2 : (\bar{\Gamma}\bar{\rho}_t), & \bar{\rho}_{t=0} &= \bar{\rho}_0, & \text{where } \bar{\Gamma} &= \bar{\gamma}\bar{\gamma}^\text{T}. \end{aligned}$$

We will mostly omit the arguments of the functions for compactness of notation. Whenever a function has a  $\bar{\cdot}$  or  $\hat{\cdot}$  on it, it is a function of  $x$ . If there is subscript  $x$  it is a function of  $y$  and if there is no marking it will in general be a function of both  $x$  and  $y$  unless stated differently.

According to Proposition 4.2 it holds that

$$\mathcal{H}(\hat{\rho}_T|\bar{\rho}_T) \leq \mathcal{H}(\hat{\rho}_0|\bar{\rho}_0) + \mathcal{I}_{\mathcal{L}}(\hat{\rho}_T) = \int_0^T \int_{\mathbb{R}^{n_x}} |h_t(x)|_{\mathbb{R}^{-1}}^2 d\hat{\rho}_t(x) dt, \text{ where}$$

$$h_t(x) = -\bar{f}(x) + \nabla_x \cdot \bar{\Gamma}(x) - (-\hat{f}(x) + \nabla_x \cdot \hat{\Gamma}(x)) + (\bar{\Gamma}(x) - \hat{\Gamma}(x))\nabla_x(\ln \hat{\rho}_t(x)).$$

Hence, the aim is to estimate the right hand side of the inequality.

**Corollary 4.6**

Let us start with the calculation for the corollary since it is a lot easier but already reveals the structure of the proof. So  $\Gamma_1 = \Gamma_1(x)$  and thus  $\bar{\Gamma} = \hat{\Gamma}$  which implies

$$h_t = -\bar{f} + \hat{f} = \int g(x, y)(\rho_{t,x} - \mu_x) dy,$$

where we introduce  $g(x, y) = -\Gamma_1(x)\nabla_x V(x, y) + D_1(x, y)$ . Let  $\Pi$  be any coupling of  $\rho_{t,x}$  and  $\mu_x$  (see Definition 3.17) and use again the same notation for density and measure  $d\Pi(y_1, y_2) = \Pi(y_1, y_2) dy_1 dy_2$ . By applying (B1) in the first, Jensen's inequality in the second and the Lipschitz assumption for  $g$  in the last step, we find

$$\begin{aligned} |h_t|_{\bar{\Gamma}^{-1}}^2 &\leq \frac{1}{\lambda_{\min}(\Gamma_1)} \left| \int g(x, y_1) - g(x, y_2) d\Pi(y_1, y_2) \right|^2 \\ &\leq \int |g(x, y_1) - g(x, y_2)|^2 d\Pi(y_1, y_2) \\ &\leq \frac{L^2}{\lambda_{\min}(\Gamma_1)} \int |y_1 - y_2|^2 d\Pi(y_1, y_2), \end{aligned}$$

and taking  $\Pi$  to be the measure which minimises the last expression, we find by assumption (B2) and the TI (Lemma 3.20) that

$$|h_t|_{\bar{\Gamma}^{-1}}^2 \leq \frac{L^2}{\lambda_{\min}(\Gamma_1)} \mathcal{W}(\rho_{t,x} | \mu_x)^2 \leq \frac{2L^2}{\alpha_{\text{LSI}} \lambda_{\min}(\Gamma_1)} \mathcal{H}(\rho_{t,x} | \mu_x).$$

Employing this estimate together with Lemma 4.7 and the FIR inequality yields the result, since

$$\begin{aligned} \mathcal{H}(\hat{\rho}_T | \bar{\rho}_T) &\leq \int_0^T \int_{\mathbb{R}^{n_x}} |h_t(x)|_{\bar{\Gamma}^{-1}}^2 d\hat{\rho}_t(x) dt \\ &\leq \frac{2L^2}{\alpha_{\text{LSI}} \lambda_{\min}(\Gamma_1)} \int_0^T \int_{\mathbb{R}^{n_x}} \mathcal{H}(\rho_{t,x} | \mu_x) d\hat{\rho}_t(x) dt \\ &= -\frac{2L^2}{\alpha_{\text{LSI}} \lambda_{\min}(\Gamma_1)} \frac{\varepsilon}{2\alpha_{\text{LSI}} \lambda_{\min}(\Gamma_2)} \int_0^T \frac{d}{dt} \mathcal{H}(\rho_t | \mu) dt \\ &= -\varepsilon \frac{L^2}{\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_1) \lambda_{\min}(\Gamma_2)} (\mathcal{H}(\rho_t | \mu) - \mathcal{H}(\rho_0 | \mu)) \\ &\leq \varepsilon \frac{L^2}{\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_1) \lambda_{\min}(\Gamma_2)} \mathcal{H}(\rho_0 | \mu). \end{aligned}$$

The calculation for the general case follows the same line, only the estimation of  $h_t$  becomes more involved.

**The calculation for the general case: Theorem 4.5**

First, we calculate the terms appearing in  $h_t$  and start with  $\nabla_x \bar{\Gamma} = \left( \sum_{i=1}^{n_x} \partial_{x_i} (\bar{\Gamma})_{ji} \right)_{1 \leq j \leq n_x}$ , where  $(\bar{\Gamma})_{ji} = (\bar{\Gamma})_{ij}$  is the  $(j, i)$ -th entry of  $\bar{\Gamma}$ . To make the calculation more trans-

parent consider the  $i^{\text{th}}$  summand first:

$$\begin{aligned}\partial_{x_i}(\bar{\Gamma})_{ji} &= \partial_{x_i} \int_{\mathbb{R}^{n_y}} (\Gamma_1)_{ji} d\mu_x = \int_{\mathbb{R}^{n_y}} \partial_{x_i} \left( (\Gamma_1)_{ji} \mu_x(y) \right) dy \\ &= \int_{\mathbb{R}^{n_y}} \partial_{x_i} \left( (\Gamma_1)_{ji} \frac{\mu}{\hat{\mu}(x)} \right) dy \\ &= \int_{\mathbb{R}^{n_y}} \left( \partial_{x_i} (\Gamma_1)_{ji} + (\Gamma_1)_{ji} \partial_{x_i} (\ln \mu) - (\Gamma_1)_{ji} \partial_{x_i} (\ln \hat{\mu}) \right) \mu_x dy\end{aligned}$$

Now, noting that  $\sum_{i=1}^{n_x} \partial_{x_i} (\Gamma_1)_{ji} + (\Gamma_1)_{ji} \partial_{x_i} (\ln \mu) = (f_1 - D_1)_j$ , it follows that

$$\nabla_x \cdot \bar{\Gamma} = \int_{\mathbb{R}^{n_y}} f_1 - D_1 d\mu_x(y) - \bar{\Gamma} \nabla (\ln \hat{\mu}) = \bar{f} - \bar{\Gamma} \nabla_x (\ln \hat{\mu}) - \int_{\mathbb{R}^{n_y}} D_1 d\mu_x.$$

Similarly - exchanging  $\mu_x$  by  $\rho_{t,x}$  - we compute

$$\begin{aligned}\partial_{x_i}(\hat{\Gamma})_{ji} &= \int_{\mathbb{R}^{n_y}} \left( \partial_{x_i} (\Gamma_1)_{ji} + (\Gamma_1)_{ji} \partial_{x_i} (\ln \rho_t) - (\Gamma_1)_{ji} \partial_{x_i} (\ln \hat{\rho}_t) \right) d\rho_{t,x}(y) \\ &= \int_{\mathbb{R}^{n_y}} \left( \partial_{x_i} (\Gamma_1)_{ji} + (\Gamma_1)_{ji} \partial_{x_i} (\ln \mu) + (\Gamma_1)_{ji} \partial_{x_i} (\ln \frac{\rho_t}{\mu}) - (\Gamma_1)_{ji} \partial_{x_i} (\ln \hat{\rho}_t) \right) d\rho_{t,x}(y)\end{aligned}$$

and hence by the same reasoning as above

$$\nabla_x \cdot \hat{\Gamma} = \hat{f} - \hat{\Gamma} \nabla_x (\ln \hat{\rho}_t) - \int_{\mathbb{R}^{n_y}} D_1 + \Gamma_1 \nabla_x (\ln \frac{\rho_t}{\mu}) d\rho_{t,x}(y).$$

Combining the terms we arrive at

$$h_t = \int_{\mathbb{R}^{n_y}} D_1(\rho_{t,x} - \mu_x) dy + \bar{\Gamma} \nabla_x \ln \frac{\hat{\rho}_t}{\hat{\mu}} - \int_{\mathbb{R}^{n_y}} \Gamma_1 \nabla_x (\ln \frac{\rho_t}{\mu}) d\rho_{t,x}(y).$$

Let us take a closer look at the second term

$$\begin{aligned}\nabla_x \ln \frac{\hat{\rho}_t}{\hat{\mu}} &= \frac{1}{\hat{\rho}_t} \nabla_x \left( \int_{\mathbb{R}^{n_y}} \rho_t dy \right) - \frac{1}{\hat{\mu}} \nabla_x \left( \int_{\mathbb{R}^{n_y}} \mu dy \right) \\ &= \frac{1}{\hat{\rho}_t} \int_{\mathbb{R}^{n_y}} \nabla_x \rho_t dy - \frac{1}{\hat{\mu}} \int_{\mathbb{R}^{n_y}} \nabla_x \mu dy \\ &= \int_{\mathbb{R}^{n_y}} \frac{\rho_t}{\hat{\rho}_t} \nabla_x (\ln \rho_t) dy - \int_{\mathbb{R}^{n_y}} \frac{\mu}{\hat{\mu}} \nabla_x (\ln \mu) dy \\ &= \int_{\mathbb{R}^{n_y}} \nabla_x (\ln \rho_t) d\rho_{t,x}(y) - \int_{\mathbb{R}^{n_y}} \nabla_x (\ln \mu) \mu_x dy \\ &= \int_{\mathbb{R}^{n_y}} \nabla_x \left( \ln \frac{\rho_t}{\mu} \right) d\rho_{t,x}(y) - \int_{\mathbb{R}^{n_y}} \nabla_x (\ln \mu) (\mu_x - \rho_{t,x}) dy,\end{aligned}$$

and hence we rewrite  $h_t$  as

$$h_t = \int_{\mathbb{R}^{n_y}} (D_1 - \bar{\Gamma} \nabla_x V)(\rho_{t,x} - \mu_x) dy - \int_{\mathbb{R}^{n_y}} (\Gamma_1 - \bar{\Gamma}) \nabla_x \left( \ln \frac{\rho_t}{\mu} \right) d\rho_{t,x}(y),$$

and split it up accordingly into

$$\begin{aligned} |h_t|_{\Gamma^{-1}} &\leq \left| \int_{\mathbb{R}^{n_y}} (D_1 - \bar{\Gamma} \nabla_x V)(\rho_{t,x} - \mu_x) dy \right|_{\Gamma^{-1}}^2 + \left| \int_{\mathbb{R}^{n_y}} (\Gamma_1 - \bar{\Gamma}) \nabla_x (\ln \frac{\rho_t}{\mu}) d\rho_{t,x}(y) \right|_{\Gamma^{-1}}^2 \\ &=: I_1 + I_2. \end{aligned}$$

Recall the definition  $C(x, y) := \bar{\Gamma}(x)^{-1/2} D_1(x, y) - \bar{\Gamma}(x)^{1/2} \nabla_x V(x, y)$  which is assumed to be Lipschitz in the second argument (see (H1)). Let  $\Pi$  be any coupling of  $\rho_{t,x}$  and  $\mu_x$  (see Definition 3.17). By Jensen's inequality and (H1) we find

$$\begin{aligned} I_1 &= \left| \bar{\Gamma}^{-1/2} \int_{\mathbb{R}^{n_y}} (D_1 - \bar{\Gamma} \nabla_x V)(\rho_{t,x} - \mu_x) dy \right|^2 \\ &= \left| \int_{\mathbb{R}^{n_y} \times \mathbb{R}^{n_y}} C(x, y_1) - C(x, y_2) d\Pi(y_1, y_2) \right|^2 \\ &\leq L_C^2 \int_{\mathbb{R}^{n_y} \times \mathbb{R}^{n_y}} |y_1 - y_2|^2 d\Pi(y_1, y_2) \end{aligned}$$

and taking  $\Pi$  to be the measure which minimises the last expression, we find by assumption (B2) and Lemma 3.20 that

$$I_1 \leq L_C^2 \mathcal{W}(\rho_{t,x} | \mu_x)^2 \leq \frac{2L_C^2}{\alpha_{\text{LSI}}} \mathcal{H}(\rho_{t,x} | \mu_x).$$

We estimate the second term, using assumption (H2) in the first, Jensen's inequality in the second and (B1) in the third step, which gives

$$\begin{aligned} I_2 &= \left| \bar{\Gamma}^{-1/2} \int_{\mathbb{R}^{n_y}} (\Gamma_1 - \bar{\Gamma}) \nabla_x \left( \ln \frac{\rho_t}{\mu} \right) d\rho_{t,x}(y) \right|^2 \\ &\leq \left| \int_{\mathbb{R}^{n_y}} \kappa \nabla_x \left( \ln \frac{\rho_t}{\mu} \right) d\rho_{t,x}(y) \right|^2 \\ &\leq \kappa^2 \int_{\mathbb{R}^{n_y}} \left| \nabla_x \ln \frac{\rho_t}{\mu} \right|^2 d\rho_{t,x}(y) \\ &\leq \frac{\kappa^2}{\lambda_{\min}(\Gamma_1)} \int_{\mathbb{R}^{n_y}} \left| \nabla_x \ln \frac{\rho_t}{\mu} \right|_{\Gamma_1}^2 d\rho_{t,x}(y) \\ &\leq \frac{\kappa^2}{\lambda_{\min}(\Gamma_1)} \int_{\mathbb{R}^{n_y}} \left| \nabla \ln \frac{\rho_t}{\mu} \right|_{\Gamma_\varepsilon}^2 d\rho_{t,x}(y). \end{aligned}$$

Together with Lemma 4.7 this yields the final result

$$\begin{aligned} \mathcal{I}_{\bar{L}}(\hat{\rho}_T) &= \int_0^T \int_{\mathbb{R}^{n_x}} |h_t(x)|_{\bar{\Gamma}^{-1}} d\hat{\rho}_t(x) dt \\ &\leq - \left( \varepsilon \frac{L_C^2}{\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_2)} + \frac{\kappa^2}{\lambda_{\min}(\Gamma_1)} \right) \int_0^T \frac{d}{dt} \mathcal{H}(\rho_t | \mu) dt \\ &\leq \left( \varepsilon \frac{L_C^2}{\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_2)} + \frac{\kappa^2}{\lambda_{\min}(\Gamma_1)} \right) \mathcal{H}(\rho_0 | \mu). \end{aligned}$$

□

### Numerical examples

We numerically illustrate that the  $\varepsilon$ -scaling we find for the relative entropy estimate in Theorem 4.5 is correct. We consider two different OU-processes whose drift and diffusion matrices read

$$\begin{aligned} \text{Example 1: } A^\varepsilon &= \begin{pmatrix} -3/2 & 1/2 \\ 1/(2\varepsilon) & -3/(2\varepsilon) \end{pmatrix}, & C^\varepsilon &= \begin{pmatrix} 1 & 0 \\ 0 & 1/\sqrt{\varepsilon} \end{pmatrix} \\ \text{Example 2: } A^\varepsilon &= \begin{pmatrix} -4 & 7 & -4 \\ -1 & 3 & -3 \\ -1/\varepsilon & 5/\varepsilon & -5/\varepsilon \end{pmatrix}, & C^\varepsilon &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\sqrt{\varepsilon} \end{pmatrix}. \end{aligned}$$

For these two examples we compute  $\max_{t>0} \mathcal{H}(\hat{\rho}_t^\varepsilon | \bar{\rho}_t^\varepsilon)$  for  $\varepsilon \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$

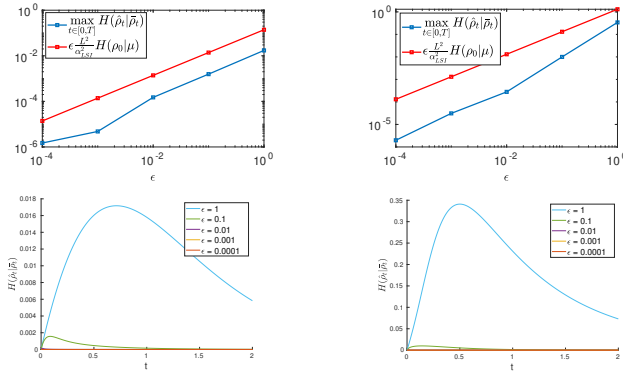


Figure 4.1: Top: Plot of  $\max_{t>0} \mathcal{H}(\hat{\rho}_t^\varepsilon | \bar{\rho}_t^\varepsilon)$  against  $\varepsilon$  for Example 1 (left) and Example 2 (right) which is compared to the right hand side of Corollary 4.6. Bottom: Plot of  $\mathcal{H}(\hat{\rho}_t^\varepsilon | \bar{\rho}_t^\varepsilon)$  over time for Example 1 (left) and Example 2 (right).

which we plot on a doubly-logarithmic scale and we find indeed that the error scales linearly in  $\varepsilon$ . We also depict the time evolution of  $\mathcal{H}(\hat{\rho}_t^\varepsilon | \bar{\rho}_t^\varepsilon)$  for different  $\varepsilon$  and observe that it first increases, but afterwards monotonously relaxes towards 0. Note that for the second example the effective dynamics is two-dimensional.

Computing the parameters in the estimate for the examples yields for the first example:  $L = 1/2$ ,  $\lambda_{\min}(\Gamma_1) = \lambda_{\min}(\Gamma_2) = 1$ ,  $\alpha_{LSI} = 3$ . For the second example  $L = 4$ ,  $\lambda_{\min}(\Gamma_1) = \lambda_{\min}(\Gamma_2) = 1$ , and the LSI constant becomes larger with decreasing  $\varepsilon$ :  $\alpha_{LSI}^\varepsilon \in \{3.8162, 8.7479, 9.8547, 9.9852, 9.9985\}$ .



### 4.2.2 Wasserstein distance

In this section we also derive estimates for the time  $t$  marginal of  $X_t$  and the effective  $\bar{X}_t$ , this time in Wasserstein distance. To this end, we replace the assumptions (H1) and (H2) by:

- (W1) The drift of the slow process  $f_1$  is Lipschitz continuous in the second argument with constant  $L_{f_1}$  uniformly in  $x$ , i.e.  $|f_1(x, y_1) - f_1(x, y_2)| \leq L_{f_1}|y_1 - y_2| \forall x, y_1, y_2$  and the effective drift  $\bar{f} = \mathbb{E}_{\mu_x}(f)$  is Lipschitz continuous with constant  $L_{\bar{f}}$ .
- (W2) The diffusion of the slow process is Lipschitz continuous in the second argument with constant  $L_{\gamma_1}$  uniformly in  $x$ , i.e.  $|\Gamma_1(x, y_1) - \Gamma_1(x, y_2)|_F \leq L_{\gamma_1}|y_1 - y_2| \forall x, y_1, y_2$  and the effective diffusion  $\bar{\gamma} = \mathbb{E}_{\mu_x}(\gamma_1 \gamma_1^T)^{1/2}$  is Lipschitz continuous with constant  $L_{\bar{\gamma}}$ .

The main result of this section then reads:

**Theorem 4.8.** *Consider the dynamics given by (4.1) and the approximation for the slow dynamics given by (4.6). Let A4, (W1) and (W2) hold true. If the initial distributions agree, i.e.,  $\hat{\rho}_0 = \bar{\rho}_0$  then for any  $t > 0$  it holds that*

$$\mathcal{W}^2(\hat{\rho}_t | \bar{\rho}_t) \leq \varepsilon \frac{L_{f_1}^2 + L_{\gamma_1}^2}{\alpha_{\text{LSI}}^2} \mathcal{H}(\rho_0 | \mu) e^{c_W t}, \quad \text{where } c_W = \left(1 + L_{\bar{\gamma}} + \frac{1}{2} L_{\bar{f}}\right).$$

Even though it seems like the estimate is weaker, since the right hand side is growing exponentially in time, it has one advantageous property compared to the estimate in relative entropy: even though the diffusion of the slow process  $\Gamma_1$  is a function of both, slow and fast, all terms on the right hand side contain the factor  $\varepsilon$  and hence for  $\varepsilon \rightarrow 0$ , that is in the infinite time scale separation regime, the error in Wasserstein distance vanishes. This is a desirable property since we want to provide a good reduced model for the slow dynamics which is in particular important when the time scale separation between slow and fast increases, as the treatment of the fast becomes more and more expensive.

Besides, the proof is based on a coupling method and we choose the simplest coupling here, namely the synchronous coupling. Employing instead other couplings, such as, e.g., a reflection coupling or other clever choices, might enable an improvement the estimate (see e.g. [22]).

*Proof.* The densities  $\hat{\rho}_t$  and  $\bar{\rho}_t$  solve the Fokker-Planck equations corresponding to (4.4) and (4.6), i.e.

$$\begin{aligned} \partial_t \hat{\rho}_t &= -\nabla_x \cdot (\hat{f} \hat{\rho}_t) + \nabla_x^2 : (\hat{\Gamma} \hat{\rho}_t), & \hat{\rho}_{t=0} &= \hat{\rho}_0, \\ \partial_t \bar{\rho}_t &= -\nabla_x \cdot (\bar{f} \bar{\rho}_t) + \nabla_x^2 : (\bar{\Gamma} \bar{\rho}_t), & \bar{\rho}_{t=0} &= \bar{\rho}_0. \end{aligned}$$

According to Lemma 2.25 in [21] there exists a time-dependent coupling  $\theta_t$  of  $\hat{\rho}_t$  and  $\bar{\rho}_t$ , which is a solution to the Fokker-Planck equation

$$\partial_t \theta_t = -\nabla \cdot \left[ \left( \frac{\hat{f}}{\bar{f}} \right) \theta_t \right] + \nabla^2 : \left[ \begin{pmatrix} \hat{\Gamma} & \hat{\gamma} \bar{\gamma}^T \\ \bar{\gamma} \hat{\gamma}^T & \bar{\Gamma} \end{pmatrix} \theta_t \right], \quad \theta_{t=0} = \theta_0, \quad (4.7)$$

where  $\theta_0$  is the optimal Wasserstein coupling of  $\hat{\rho}_0$  and  $\bar{\rho}_0$ . The main ingredient of the proof is to estimate  $\partial_t \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 \theta_t(\hat{x}, \bar{x}) d\hat{x} d\bar{x}$  and then apply Gronwall's lemma in the end. We begin by calculating

$$\partial_t \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 d\theta_t(\hat{x}, \bar{x}) = \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 \partial_t \theta_t(\hat{x}, \bar{x}) d\hat{x} d\bar{x}$$

In order to conduct this calculation we will substitute for  $\partial_t \theta_t$  the right-hand side of the Fokker-Planck equation and perform integration by parts for the two terms separately. For this purpose note in advance that

$$\nabla \frac{1}{2} |\hat{x} - \bar{x}|^2 = \begin{pmatrix} \hat{x} - \bar{x} \\ -(\hat{x} - \bar{x}) \end{pmatrix}, \quad \nabla^2 \frac{1}{2} |\hat{x} - \bar{x}|^2 = \begin{pmatrix} I_{n_x \times n_x} & -I_{n_x \times n_x} \\ -I_{n_x \times n_x} & I_{n_x \times n_x} \end{pmatrix}.$$

With this in mind and performing integration by parts twice, we find

$$\begin{aligned} & \partial_t \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 d\theta_t(\hat{x}, \bar{x}) \\ &= \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 \left( -\nabla \cdot \left[ \left( \frac{\hat{f}}{\bar{f}} \right) \theta_t \right] + \nabla^2 : \left[ \begin{pmatrix} \hat{\Gamma} & \hat{\gamma} \bar{\gamma}^T \\ \bar{\gamma} \hat{\gamma}^T & \bar{\Gamma} \end{pmatrix} \theta_t \right] \right) d\hat{x} d\bar{x} \\ &= \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \begin{pmatrix} \hat{x} - \bar{x} \\ -(\hat{x} - \bar{x}) \end{pmatrix} \cdot \begin{pmatrix} \hat{f} \\ \bar{f} \end{pmatrix} + \begin{pmatrix} I_{n_x \times n_x} & -I_{n_x \times n_x} \\ -I_{n_x \times n_x} & I_{n_x \times n_x} \end{pmatrix} : \begin{pmatrix} \hat{\Gamma} & \hat{\gamma} \bar{\gamma}^T \\ \bar{\gamma} \hat{\gamma}^T & \bar{\Gamma} \end{pmatrix} d\theta_t(\hat{x}, \bar{x}) \\ &= \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} (\hat{f} - \bar{f}) \cdot (\hat{x} - \bar{x}) d\theta_t(\hat{x}, \bar{x}) + \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} |\hat{\gamma} - \bar{\gamma}|_F^2 d\theta_t(\hat{x}, \bar{x}) =: I_1 + I_2. \end{aligned}$$

Let us consider these terms separately. For both terms, we first add a zero, afterwards apply Young's inequality, such that in the last step we can then employ (W1) and (W2) respectively. This gives

$$\begin{aligned} I_1 &= \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} (\hat{f}(t, \hat{x}) - \bar{f}(\bar{x})) \cdot (\hat{x} - \bar{x}) d\theta_t(\hat{x}, \bar{x}) \\ &= \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \left( (\hat{f}(t, \hat{x}) - \bar{f}(\hat{x})) \cdot (\hat{x} - \bar{x}) + (\bar{f}(\hat{x}) - \bar{f}(\bar{x})) \cdot (\hat{x} - \bar{x}) \right) d\theta_t(\hat{x}, \bar{x}) \\ &\leq \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \left( \frac{1}{2} |\hat{f}(t, \hat{x}) - \bar{f}(\hat{x})|^2 + |\hat{x} - \bar{x}|^2 + \frac{1}{2} |\bar{f}(\hat{x}) - \bar{f}(\bar{x})|^2 \right) d\theta_t(\hat{x}, \bar{x}) \\ &\leq \int_{\mathbb{R}^{n_x}} \frac{1}{2} |\hat{f}(t, \hat{x}) - \bar{f}(\hat{x})|^2 d\hat{\rho}_t(\hat{x}) + \left( 1 + \frac{L\bar{f}}{2} \right) \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} |\hat{x} - \bar{x}|^2 d\theta_t(\hat{x}, \bar{x}). \\ I_2 &= \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} |\hat{\gamma}(t, \hat{x}) - \bar{\gamma}(\bar{x})|_F^2 d\theta_t(\hat{x}, \bar{x}) \\ &\leq \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \left( |\hat{\gamma}(t, \hat{x}) - \bar{\gamma}(\hat{x})|_F^2 + |\bar{\gamma}(\hat{x}) - \bar{\gamma}(\bar{x})|_F^2 \right) d\theta_t(\hat{x}, \bar{x}) \\ &\leq \int_{\mathbb{R}^{n_x}} |\hat{\gamma}(t, \hat{x}) - \bar{\gamma}(\hat{x})|_F^2 d\hat{\rho}_t(\hat{x}) + L\bar{\gamma} \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} |\hat{x} - \bar{x}|^2 d\theta_t(\hat{x}, \bar{x}). \end{aligned}$$

Letting  $\Pi$  be any coupling of  $\rho_{t,x}$  with  $t > 0$  and  $\mu_x$ , we estimate the term  $|\hat{f}(t, x) - \bar{f}(x)|^2$  by employing Jensen's inequality and (W1)

$$\begin{aligned} |\hat{f}(t, x) - \bar{f}(x)|^2 &= \left| \int_{\mathbb{R}^{n_y} \times \mathbb{R}^{n_y}} f_1(x, y_1) - f_1(x, y_2) d\Pi(y_1, y_2) \right|^2 \\ &\leq \int_{\mathbb{R}^{n_y} \times \mathbb{R}^{n_y}} |f_1(x, y_1) - f_1(x, y_2)|^2 d\Pi(y_1, y_2) \\ &\leq L_{f_1}^2 \int_{\mathbb{R}^{n_y} \times \mathbb{R}^{n_y}} |y_1 - y_2|^2 d\Pi(y_1, y_2) \end{aligned}$$

Estimating the second term  $\int |\hat{\gamma}(t, x) - \bar{\gamma}(x)|_F^2 d\hat{\rho}_t(x)$  we make use of the fact that the function  $\phi : \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ ,  $(A, B) \mapsto |A^{1/2} - B^{1/2}|_F^2 = \text{Tr}(A - 2\sqrt{A}\sqrt{B} + B)$  is convex for  $A, B > 0$  (this is where we need (B1) to guarantee that  $\bar{\Gamma}, \hat{\Gamma} > 0$ ) by Lieb's concavity theorem (cf. Theorem IX.6.1 in [10]) and we can hence employ the two-sided Jensen's inequality in the first, together with (W2) in the second step, which yields

$$\begin{aligned} |\hat{\gamma}(t, x) - \bar{\gamma}(x)|_F^2 &= \left| \sqrt{\int \Gamma_1 d\rho_{t,x}(y)} - \sqrt{\int \Gamma_1 d\mu_x} \right|_F^2 \\ &\leq \int_{\mathbb{R}^{n_y} \times \mathbb{R}^{n_y}} |\Gamma_1(x, y_1) - \Gamma_1(x, y_2)|_F^2 d\Pi(y_1, y_2) \\ &\leq L_{\gamma_1} \int_{\mathbb{R}^{n_y} \times \mathbb{R}^{n_y}} |y_1 - y_2|^2 d\Pi(y_1, y_2). \end{aligned}$$

Since the coupling  $\Pi$  for  $t > 0$  was arbitrary, we can in particular choose  $\Pi$  to be the coupling, which minimises the last expression. By this choice it follows that

$$|\hat{f}(t, x) - \bar{f}(x)|^2 + |\hat{\gamma}(t, x) - \bar{\gamma}(x)|_F^2 \leq (L_{f_1}^2 + L_{\gamma_1}^2) \mathcal{W}(\rho_{t,x} | \mu_x) \leq 2 \frac{L_{f_1}^2 + L_{\gamma_1}^2}{\alpha_{\text{LSI}}} \mathcal{H}(\rho_{t,x} | \mu_x).$$

Putting all the terms together and employing Lemma 4.7 we arrive at

$$\begin{aligned} \partial_t \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 d\theta_t(\hat{x}, \bar{x}) \\ \leq -\varepsilon \frac{L_{f_1}^2 + L_{\gamma_1}^2}{\alpha_{\text{LSI}}^2} \frac{d}{dt} \mathcal{H}(\rho_t | \mu) + \left(1 + L_{\bar{\gamma}} + \frac{L_{\bar{f}}}{2}\right) \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 d\theta_t(\hat{x}, \bar{x}) \end{aligned}$$

and integration in time from 0 to  $t$  yields

$$\begin{aligned} \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 d\theta_t(\hat{x}, \bar{x}) \\ \leq \varepsilon \frac{L_{f_1}^2 + L_{\gamma_1}^2}{\alpha_{\text{LSI}}^2} \mathcal{H}(\rho_0 | \mu) + \left(1 + L_{\bar{\gamma}} + \frac{L_{\bar{f}}}{2}\right) \int_0^t \int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 d\theta_s(\hat{x}, \bar{x}) ds \end{aligned}$$

where we used that  $\hat{\rho}_0 = \bar{\rho}_0$ . Finally, applying Gronwall's inequality gives

$$\int_{\mathbb{R}^{n_x} \times \mathbb{R}^{n_x}} \frac{1}{2} |\hat{x} - \bar{x}|^2 d\theta_t(\hat{x}, \bar{x}) \leq \varepsilon \frac{L_{f_1}^2 + L_{\gamma_1}^2}{\alpha_{\text{LSI}}^2} \mathcal{H}(\rho_0 | \mu) e^{c_W t}, \quad c_W = \left(1 + L_{\bar{\gamma}} + \frac{L_{\bar{f}}}{2}\right)$$

which implies the result

$$\mathcal{W}^2(\hat{\rho}_t | \bar{\rho}_t) \leq \varepsilon \frac{L_{f_1}^2 + L_{\gamma_1}^2}{\alpha_{\text{LSI}}^2} \mathcal{H}(\rho_0 | \mu) e^{c_w t}.$$

□

### 4.2.3 Estimates for the path measures

In this section our aim is to compare  $\hat{\rho} := \text{law}((X_t)_{0 \leq t \leq T}) \in \mathcal{P}(C([0, T]; \mathbb{R}^{n_x}))$  with  $\hat{\nu} := \text{law}((\bar{X}_t)_{0 \leq t \leq T}) \in \mathcal{P}(C([0, T]; \mathbb{R}^{n_x}))$ , where  $T > 0$  is fixed. That is, we want to compare the law of the paths, contrary to the previous sections where we only considered the marginals at time  $t$  of these objects. This kind of estimate is important, aiming at making statements about the reproducibility of dynamic properties by the reduced model, such as transition rates.

Throughout this section let  $T > 0$  be fixed and assume that:

- (P1) The diffusion of the slow variable  $X_t$  is independent of the fast variable  $Y_t$ , i.e.,  $\Gamma_1 = \Gamma_1(x)$  and there is no coupling of the noise, i.e.,  $\gamma_{12} = 0$
- (P2) The first component of the drift  $f_1$  is Lipschitz in the second argument with constant  $L$ . Furthermore,  $\rho_{t,x}$  has finite second moments uniformly in  $x \in \mathbb{R}^{n_x}$ ,  $t \in [0, T]$ , i.e.,

$$\sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^{n_x}} \int_{\mathbb{R}^{n_y}} |y|^2 d\rho_{t,x}(y) < \infty.$$

This entails that

$$\text{Var}_{\rho_{t,x}}(f_1) := \sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^{n_x}} \int \left( f_1(x, y) - \int f_1(x, y') d\rho_{t,x}(y') \right)^2 d\rho_{t,x}(y) < \infty.$$

- (P3)  $f_1 - \bar{f}_1$  satisfies the Novikov's condition, i.e.,  $\mathbb{E}[\exp(\int_0^T |f_1 - \bar{f}_1|_{-1}^2 ds)] < \infty$ .

The main result of this section is the following theorem.

**Theorem 4.9.** *Suppose that assumptions A4 and (P1)-(P3) hold. Let  $\hat{\rho} := \text{law}((X_t)_{0 \leq t \leq T}) \in \mathcal{P}(C([0, T]; \mathbb{R}^{n_x}))$  and  $\hat{\nu} := \text{law}((\bar{X}_t)_{0 \leq t \leq T}) \in \mathcal{P}(C([0, T]; \mathbb{R}^{n_x}))$ , where  $X_t, \bar{X}_t$  are solutions to (4.1), (4.6) respectively. Then*

$$\mathcal{H}(\hat{\rho} | \hat{\nu}) \leq \mathcal{H}(\rho_0 | \nu_0) + T \frac{\text{Var}_{\rho_{t,x}}(f_1)}{2\lambda_{\min}(\Gamma_1)} + \varepsilon \frac{L^2}{2\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_1) \lambda_{\min}(\Gamma_2)} \mathcal{H}(\rho_0 | \mu).$$

*Proof.* First note that by assumption (P2) we have that

$$\begin{aligned}
\text{Var}_{\rho_{t,x}}(f_1) &= \sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^{n_x}} \int \left( \int f_1(x, y') - f_1(x, y) d\rho_{t,x}(y) \right)^2 d\rho_{t,x}(y') \\
&\leq \sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^{n_x}} \int \int (f_1(x, y') - f_1(x, y))^2 d\rho_{t,x}(y) d\rho_{t,x}(y') \\
&\leq L \sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^{n_x}} \int \int (y - y')^2 d\rho_{t,x}(y) d\rho_{t,x}(y') \\
&\leq 2L \sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^{n_x}} \int_{\mathbb{R}^{n_y}} |y|^2 d\rho_{t,x}(y).
\end{aligned}$$

Instead of directly working with  $\hat{\rho}$  and  $\hat{\nu}$ , we will compare (4.1), which we recall here for readability

$$\begin{aligned}
dX_t &= f_1(X_t, Y_t) dt + \gamma_1(X_t) dW_t, \\
dY_t &= \frac{1}{\varepsilon} f_2(X_t, Y_t) dt + \frac{1}{\sqrt{\varepsilon}} \gamma_2(X_t, Y_t) dU_t,
\end{aligned}$$

with

$$\begin{aligned}
d\bar{X}_t &= \bar{f}_1(\bar{X}_t) dt + \gamma_1(\bar{X}_t) dW_t, \\
d\bar{Y}_t &= \frac{1}{\varepsilon} \bar{f}_2(\bar{X}_t, \bar{Y}_t) dt + \frac{1}{\sqrt{\varepsilon}} \gamma_2(\bar{X}_t, \bar{Y}_t) dU_t.
\end{aligned}$$

Note that  $\hat{\rho}$  and  $\hat{\nu}$  are the marginals of  $\rho = \text{law}((X_t, Y_t)_{0 \leq t \leq T})$  and  $\nu = \text{law}((\bar{X}_t, \bar{Y}_t)_{0 \leq t \leq T})$  with respect to the coarse graining map  $\xi(x, y) = x$ . Denoting by  $\hat{\rho}$  and  $\hat{\nu}$  the corresponding conditionals, we will later make use of the so-called *chain rule*

$$\mathcal{H}(\rho|\nu) = \mathcal{H}(\hat{\rho}|\hat{\nu}) + \mathbb{E}_{\hat{\rho}}(\mathcal{H}(\hat{\rho}|\hat{\nu}))$$

which follows by  $\rho = \hat{\rho}\hat{\nu}$ ,  $\nu = \hat{\nu}\hat{\nu}$  (also see the computation (6.7) in Section 6.1.3) and implies that if we bound  $\mathcal{H}(\rho|\nu)$  we also find a bound for  $\mathcal{H}(\hat{\rho}|\hat{\nu})$  since by the positivity of all the terms

$$\mathcal{H}(\hat{\rho}|\hat{\nu}) \leq \mathcal{H}(\rho|\nu). \quad (4.8)$$

Writing  $f = \begin{pmatrix} f_1 \\ \frac{1}{\varepsilon} f_2 \end{pmatrix}$ ,  $\bar{f} = \begin{pmatrix} \bar{f}_1 \\ \frac{1}{\varepsilon} \bar{f}_2 \end{pmatrix}$  and using Girsanov's Theorem (see e.g. [58] or [23]) together with the notation  $Z = (X, Y)$  we find

$$\frac{d\rho}{d\nu}(Z_{[0, T]}) = \frac{d\rho_0}{d\nu_0}(Z_0) \exp \left( \int_0^T (f(Z_t) - \bar{f}(Z_t)) \gamma dW_t + \frac{1}{2} \int_0^T |f(Z_t) - \bar{f}(Z_t)|_{\Gamma^{-1}}^2 dt \right).$$

Noting that  $f(z) - \bar{f}(z) = \begin{pmatrix} f_1(z) - \bar{f}_1(z) \\ \frac{1}{\varepsilon} (f_2(z) - \bar{f}_2(z)) \end{pmatrix}$ , where we write  $\bar{f}_1(z)$  even though  $\bar{f}_1(z) =$

$\bar{f}_1(x)$  is effectively only a function of  $x$  and the same applies to  $\bar{f}$ , we have

$$\begin{aligned}
\mathcal{H}(\rho|\nu) &= \mathbb{E}_\rho \left( \ln \frac{d\rho}{d\nu} \right) \\
&= \mathcal{H}(\rho_0|\nu_0) + \mathbb{E}_\rho \left( \int_0^T (f(Z_t) - \bar{f}(Z_t))\gamma \, dW_t + \frac{1}{2} \int_0^T |f(Z_t) - \bar{f}(Z_t)|_{\Gamma^{-1}}^2 dt \right) \\
&= \mathcal{H}(\rho_0|\nu_0) + \frac{1}{2} \mathbb{E}_\rho \left( \int_0^T |f(Z_t) - \bar{f}(Z_t)|_{\Gamma^{-1}}^2 dt \right) \\
&= \mathcal{H}(\rho_0|\nu_0) + \frac{1}{2} \int_0^T \mathbb{E}_{\rho_t} (|f_1(Z_t) - \bar{f}_1(Z_t)|_{\Gamma^{-1}}^2) dt \\
&\leq \mathcal{H}(\rho_0|\nu_0) + \frac{1}{2} \int_0^T \mathbb{E}_{\rho_t} (|f_1(Z_t) - \hat{f}_1(Z_t)|_{\Gamma^{-1}}^2) + \mathbb{E}_{\rho_t} (|\hat{f}_1(Z_t) - \bar{f}_1(Z_t)|_{\Gamma^{-1}}^2) dt,
\end{aligned}$$

with  $\hat{f}_1 = \int f_1 d\rho_{t,x}$  as introduced in (4.4). Using (P2) and (B1), the first term in the integral can be estimated by

$$\mathbb{E}_{\rho_t} (|f_1(Z_t) - \hat{f}_1(Z_t)|_{\Gamma^{-1}}^2) \leq \frac{1}{\lambda_{\min}(\Gamma_1)} \mathbb{E}_{\rho_t} (\mathbb{E}_{\rho_{t,x}} |f_1(Z_t) - \hat{f}_1(Z_t)|^2) \leq \frac{\text{Var}_{\rho_{t,x}}(f_1)}{\lambda_{\min}(\Gamma_1)}.$$

Controlling the second term in the right hand side as in the proof of Corollary 4.6, we arrive at

$$\mathcal{H}(\rho|\nu) \leq \mathcal{H}(\rho_0|\nu_0) + T \frac{\text{Var}_{\rho_{t,x}}(f_1)}{2\lambda_{\min}(\Gamma_1)} + \varepsilon \frac{L^2}{2\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_2)} (\mathcal{H}(\rho_0|\mu) - \mathcal{H}(\rho_t|\mu)). \quad (4.9)$$

By (4.8) and  $\mathcal{H}(\rho_t|\mu) \geq 0$  we arrive at the final result

$$\mathcal{H}(\hat{\rho}|\hat{\nu}) \leq \mathcal{H}(\rho_0|\nu_0) + T \frac{\text{Var}_{\rho_{t,x}}(f_1)}{2\lambda_{\min}(\Gamma_1)} + \varepsilon \frac{L^2}{2\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_1) \lambda_{\min}(\Gamma_2)} \mathcal{H}(\rho_0|\mu).$$

□

The estimate suggests that the main contribution, especially for large times, is due to how much  $f_1$  varies in  $y$  for given  $x$ , i.e., how large  $\text{Var}_{\rho_{t,x}}(f_1)$  is. Let us investigate this conjecture with a simple numerical example.

### Numerical example

We present a numerical example for which we can explicitly compute the quantity  $\mathbb{E}_{\rho_t} (|f_1(Z_t) - \bar{f}_1(Z_t)|_{\Gamma^{-1}}^2)$ . Since this requires the knowledge of the times  $t$  distribution of the process, we once more consider OU-processes for which this is viable. Specifically, we employ the following lemma:

**Lemma 4.10.** *Let  $A, C \in \mathbb{R}^{n \times n}$ , assume  $(A, C)$  satisfy condition A2 and consider the linear SDE in  $\mathbb{R}^n$*

$$\begin{aligned}
dZ_t &= AZ_t \, dt + C \, dB_t, & A &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\
\text{and } d\bar{Z}_t &= \bar{A}\bar{Z}_t \, dt + C \, dB_t, & \bar{A} &= \begin{pmatrix} \bar{A}_{11} & 0 \\ \bar{A}_{21} & \bar{A}_{22} \end{pmatrix}
\end{aligned}$$

where  $\bar{A}_{11} = A_{11} + A_{12}\Sigma_{21}\Sigma_{11}^{-1}$ ,  $\bar{A}_{21} = A_{21}$ ,  $\bar{A}_{22} = A_{22}$  and  $\Sigma$  is the unique solution to

$$A\Sigma + \Sigma A^T = CC^T.$$

Then

$$\begin{aligned} \mathbb{E}_{\rho_t}(|f_1(Z_t) - \bar{f}_1(Z_t)|_{\Gamma_1^{-1}}^2) &= \text{Tr}(\text{Var}(Y_t)A_{12}^T\Gamma_1^{-1}A_{12}) - 2\text{Tr}(\text{Cov}(X_t, Y_t)A_{12}^T\Gamma_1^{-1}\Sigma_{21}\Sigma_{11}^{-1}) \\ &\quad + \text{Tr}(\text{Var}(X_t)\Sigma_{11}^{-1}\Sigma_{12}A_{12}^T\Gamma_1^{-1}A_{12}\Sigma_{21}\Sigma_{11}^{-1}) \\ &\quad + |m_t^y - \Sigma_{21}\Sigma_{11}^{-1}m_t^x|_{A_{12}^T\Gamma_1^{-1}A_{12}}^2. \end{aligned}$$

where  $\Gamma_1 = C_{11}C_{11}^T$ .

*Proof.* Note that  $f_1(Z_t) = A_{11}X_t + A_{12}Y_t$  and  $\bar{f}_1(Z_t) = (A_{11} + A_{12}\Sigma_{21}\Sigma_{11}^{-1})X_t$  and thus

$$\begin{aligned} \mathbb{E}_{\rho_t}(|f_1(Z_t) - \bar{f}_1(Z_t)|_{\Gamma_1^{-1}}^2) &= \mathbb{E}_{\rho_t} \left[ |(A_{12}Y_t - A_{12}\Sigma_{21}\Sigma_{11}^{-1}X_t)|_{\Gamma_1^{-1}}^2 \right] \\ &= \mathbb{E}_{\rho_t} \left[ |(Y_t - m_t^y + m_t^y - \Sigma_{21}\Sigma_{11}^{-1}(X_t - m_t^x + m_t^x))|_{A_{12}^T\Gamma_1^{-1}A_{12}}^2 \right] \\ &= \mathbb{E}_{\rho_t} \left[ (Y_t - m_t^y)^T A_{12}^T \Gamma_1^{-1} A_{12} (Y_t - m_t^y) \right] \\ &\quad + \mathbb{E}_{\rho_t} \left[ (X_t - m_t^x)^T \Sigma_{11}^{-1} \Sigma_{12} A_{12}^T \Gamma_1^{-1} A_{12} \Sigma_{21} \Sigma_{11}^{-1} (X_t - m_t^x) \right] \\ &\quad - 2\mathbb{E}_{\rho_t} \left[ (X_t - m_t^x)^T \Sigma_{11}^{-1} \Sigma_{12} A_{12}^T \Gamma_1^{-1} A_{12} (Y_t - m_t^y) \right] \\ &\quad + (m_t^y - \Sigma_{21}\Sigma_{11}^{-1}m_t^x)^T A_{12}^T \Gamma_1^{-1} A_{12} (m_t^y - \Sigma_{21}\Sigma_{11}^{-1}m_t^x) \\ &= \text{Tr}(\text{Var}(Y_t)A_{12}^T\Gamma_1^{-1}A_{12}) - 2\text{Tr}(\text{Cov}(X_t, Y_t)A_{12}^T\Gamma_1^{-1}A_{12}\Sigma_{21}\Sigma_{11}^{-1}) \\ &\quad + \text{Tr}(\text{Var}(X_t)\Sigma_{11}^{-1}\Sigma_{12}A_{12}^T\Gamma_1^{-1}A_{12}\Sigma_{21}\Sigma_{11}^{-1}) + |m_t^y - \Sigma_{21}\Sigma_{11}^{-1}m_t^x|_{A_{12}^T\Gamma_1^{-1}A_{12}}^2 \end{aligned}$$

□

Let us exemplify our result on a simple example, which is determined by the following drift and diffusion matrices

$$A = \begin{pmatrix} -3/2 & a \\ a/\varepsilon & -3/(2\varepsilon) \end{pmatrix} \quad \text{and} \quad C_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1/\sqrt{\varepsilon} \end{pmatrix}$$

for which we can choose the coupling constant  $a$  and time scale separation  $\varepsilon = 0.05$ . In figure 4.2, we illustrate the dependence of the estimate (4.9) on the variance term  $\kappa = \nabla \text{ar}_{\rho_t, x}(f_1)$  by varying the coupling  $a$  from  $a = 1$  in the left to  $a = 0.2$  in the right panel. We observe that the error  $\mathcal{H}(\rho|\nu)$  (in red) depends on the coupling as sensitive as the variance itself, which is contained in the right hand side of (4.9) and shown in blue, since the slopes of the lines change in the same way when varying the coupling parameter  $a$ . This means the influence of the variance of the fast degrees of freedom does have a strong influence on the error as suggested by our estimate. Note that we are only depicting the error  $\mathcal{H}(\rho|\nu)$  here and the ‘‘true’’ error satisfies  $\mathcal{H}(\hat{\rho}|\hat{\nu}) \leq \mathcal{H}(\rho|\nu)$  for which we would expect that  $\mathcal{H}(\hat{\rho}|\hat{\nu}) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ , as suggested by corresponding averaging results (see Section 4.3).

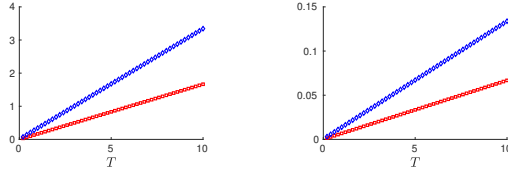


Figure 4.2: Plot of the error  $\mathcal{H}(\rho|\nu)$  (red), as in the left hand side of (4.9), and  $\mathcal{H}(\rho_0|\nu_0) + T \frac{\text{var}_{\rho_{t,x}}(f_1)}{2\lambda_{\min}(\Gamma_1)} + \varepsilon \frac{L^2}{2\alpha_{\text{LSI}}^2 \lambda_{\min}(\Gamma_2)} \mathcal{H}(\rho_0|\mu)$  (blue) for the dynamics described by  $A$  and  $C$ , where  $\varepsilon = 0.05$ . Left:  $a = 1$ . Right:  $a = 0.2$ .

### 4.3 Comparison of averaging and conditional expectations

#### 4.3.1 Review of averaging

We will shortly introduce the method of averaging here, more details on this topic can be found, e.g., in Chapter 7 of [26], the review [1], or in the book [62]. In the latter the analysis is mainly restricted to the case of compact domains.

Consider dynamics described by

$$\dot{x}_t^\varepsilon = b(x_t^\varepsilon, y_{t/\varepsilon}), \quad x_0 = x^0,$$

where  $x \in \mathbb{R}^{n_x}$ ,  $y \in \mathbb{R}^{n_y}$ ,  $\varepsilon > 0$  for which we assume that  $b : \mathbb{R}^{n_x+n_y} \rightarrow \mathbb{R}^{n_x}$  is bounded, continuous and Lipschitz in  $x$ . The basic idea is, intuitively speaking, that if

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T b(x, y_s) ds = b^{\text{av}}(x) \quad (4.10)$$

holds in some sense, it is possible to approximate the dynamics for  $x_t^\varepsilon$  by

$$\dot{x}_t^{\text{av}} = b^{\text{av}}(x_t^{\text{av}}), \quad x_0^{\text{av}} = x^0.$$

It is clear that assumption (4.10) plays a crucial part here. In the case of  $y_s$  being an ergodic stochastic process with unique invariant measure  $\nu_x$  the limit (4.10) can be rewritten as

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T b(x, y_s) ds = \int_{\mathbb{R}^{n_y}} b(x, y) d\nu_x(y) = b^{\text{av}}(x).$$

If this is true for each  $x \in \mathbb{R}^{n_x}$ , then (4.10) holds true with probability one  $\forall x \in \mathbb{R}^{n_x}$  and thus convergence of  $x_t^\varepsilon$  to  $x_t^{\text{av}}$  follows uniformly in time with probability one as  $\varepsilon \rightarrow 0$  (see [41]). Here, the limit  $\varepsilon \rightarrow 0$  justifies keeping  $x$  in (4.10) fixed, since  $\varepsilon \rightarrow 0$  is equivalent to an infinite time scale separation between the slow variable  $x$  and the fast variable  $y$ . This means, working with the natural time scale of  $x$ , since time



is accelerated by  $1/\varepsilon$  for the fast variable,  $y$  attains its equilibrium while  $x$  remains constant.

Also under milder assumptions, e.g. convergence in probability of (4.10), it is possible to deduce results for the convergence of trajectories  $x_t^\varepsilon \rightarrow x_t^{\text{av}}$  which then holds in probability (see e.g. Chapter 7, Theorem 2.1 in [26], or Theorem 1.1 in [41]).

When traversing from an ODE for  $x_t$  to an SDE, i.e., the right hand side is no longer deterministic but stochastic itself, convergence in distribution or weak convergence of  $(x_t^\varepsilon)_{t \in [0, T]} \rightarrow (x_t^{\text{av}})_{t \in [0, T]}$  in  $C([0, T], \mathbb{R}^{n_x})$  becomes the natural object. That is,  $x_t^\varepsilon$  converges weakly to  $x_t^{\text{av}}$  in  $C([0, T], \mathbb{R}^{n_x})$ , if for any bounded functional  $f : C([0, T], \mathbb{R}^{n_x}) \rightarrow \mathbb{R}$  it holds that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E}(f((x_t^\varepsilon)_{0 \leq t \leq T})) = \mathbb{E}(f((x_t^{\text{av}})_{0 \leq t \leq T})),$$

where the expectation  $\mathbb{E}$  should be understood as an expectation with respect to the path measure of  $(x_t^\varepsilon)_{(0 \leq t \leq T)}$  and  $(x_t^{\text{av}})_{(0 \leq t \leq T)}$  respectively.

Recall our setting which is

$$\begin{aligned} dX_t^\varepsilon &= f_1(X_t^\varepsilon, Y_t^\varepsilon) dt + \gamma_1(X_t^\varepsilon, Y_t^\varepsilon) dW_t, \quad X_{t=0}^\varepsilon = X_0 \\ dY_t^\varepsilon &= \frac{1}{\varepsilon} f_2(X_t^\varepsilon, Y_t^\varepsilon) dt + \frac{1}{\sqrt{\varepsilon}} \gamma_2(X_t^\varepsilon, Y_t^\varepsilon) dU_t, \quad Y_{t=0}^\varepsilon = Y_0, \end{aligned}$$

with  $f, \gamma$  satisfying A4. Assume now that for any fixed  $x \in \mathbb{R}^{n_x}$ , the fast dynamics admits a unique invariant measure  $\nu_x^{\text{av}} \in \mathcal{P}(\mathbb{R}^{n_y})$ , i.e., its density satisfies

$$\nabla_y \cdot (f_2(x, \cdot) \nu_x^{\text{av}}) + \nabla_y^2 : (\Gamma_2(x, \cdot) \nu_x^{\text{av}}) = 0$$

or equivalently

$$\mathcal{L}_2^* \nu_x^{\text{av}} = 0, \quad \text{where } \mathcal{L}_2 = f_2 \cdot \nabla_y + \Gamma_2 : \nabla_y^2$$

is the generator of the fast dynamics whose coefficients  $f_2, \Gamma_2$  are here evaluated at a fixed  $x$ . Further introduce

$$\begin{aligned} dX_t^{\text{av}} &= f_1^{\text{av}}(X_t^{\text{av}}) dt + \gamma_1^{\text{av}}(X_t, Y_t) dW_t, \quad X_{t=0}^{\text{av}} = X_0 \tag{4.11} \\ f_1^{\text{av}}(x) &= \int_{\mathbb{R}^{n_y}} f_1(x, y) d\nu_x^{\text{av}}(y), \quad \gamma_1^{\text{av}}(x) = \left( \int_{\mathbb{R}^{n_y}} \Gamma_1(x, y) d\nu_x^{\text{av}}(y) \right)^{1/2} \end{aligned}$$

which is assumed to admit a unique strong solution.

When working on compact domains the proofs showing convergence of  $X^\varepsilon \rightarrow X^{\text{av}}$  are simpler and less assumptions are needed. We still present the general result here, since also the results for conditional expectations are derived for  $X_t^\varepsilon \in \mathbb{R}^{n_x}$ .

Let us state the assumptions needed for the upcoming theorem stating convergence of  $X^\varepsilon$  to  $X^{\text{av}}$ , which are taken from [60].

(H<sub>a</sub>) There exist  $0 < m < M < \infty$  such that  $m \leq \gamma_2 \gamma_2^T \leq M$

$$(H_b) \quad \lim_{|y| \rightarrow \infty} \sup_x f_2(x, y) y = -\infty$$

$(H^{2+\alpha,1})$   $f_2, \gamma_2$  both have bounded first derivative in  $x$  and second derivatives in  $y$  and all derivatives  $\partial_y^i \partial_x^j$ ,  $i \in \{0, 1, 2\}$ ,  $j \in \{0, 1\}$  are Hölder continuous with respect to the  $y$  variable with exponent  $0 < \alpha < 1$  uniformly in  $x$ .

Here,  $(H_a)$  and  $(H_b)$  guarantee existence of a unique invariant measure for the fast variable. Together these assumptions are used to derive bounds on derivatives of the solution of the Poisson equation and the transition probability density. Finally the main result for our purpose in [60] is:

**Theorem 4.11** (Theorem 4 in [60]). *Let  $(H^{2+\alpha,1})$ ,  $(H_a)$ ,  $(H_b)$  hold. Further assume that  $f_1, \gamma_1$  are Lipschitz and they fulfil the growth conditions:  $\exists K, m_1, m_2 > 0$  such that  $f_1(x, y) \leq K(1 + |y|)(1 + |x|^{m_1})$  and  $\gamma_1(x, y) \leq K(1 + \sqrt{|y|})(1 + |x|^{m_2})$ . Then for any  $T > 0$*

$$\text{law}((X_t^\varepsilon)_{0 \leq t \leq T}) \xrightarrow{\varepsilon \rightarrow 0} \text{law}((X_t^{\text{av}})_{0 \leq t \leq T})$$

weakly in  $C([0, T], \mathbb{R}^{n_x})$ .

In what follows, we refer to the limit dynamics (4.11) as the *averaged dynamics*.

**Remark 4.12.** Methods for proving these kind of results rely on associated Poisson equations whose solutions yield the effective coefficients. Requiring bounds on the coefficients thus requires bounds on the Poisson equations, which is doable for compact state spaces but becomes a lot harder when working in  $\mathbb{R}^{n_y}$ . Still, under a lot of technical assumptions, Pardoux and Veretennikov prove in [60] convergence in distribution – they even prove it for a homogenization scaling, which we omit here.

### 4.3.2 Some examples

We first discuss the standard example of Langevin equation in gradient form for which the difference in the two approaches does not show up. Afterwards we return to the example from the very beginning of the chapter and discuss it in more detail.

#### Reversible dynamics

Consider the reversible dynamics

$$\begin{aligned} dX_t^\varepsilon &= -(\Gamma_1 \nabla_x V - \nabla_x \cdot \Gamma_1)(X_t^\varepsilon, Y_t^\varepsilon) dt + \sqrt{2} \gamma_1(X_t^\varepsilon, Y_t^\varepsilon) dW_t \\ dY_t^\varepsilon &= -\frac{1}{\varepsilon} (\Gamma_2 \nabla_y V - \nabla_y \cdot \Gamma_2)(X_t^\varepsilon, Y_t^\varepsilon) dt + \sqrt{\frac{2}{\varepsilon}} \gamma_2(X_t^\varepsilon, Y_t^\varepsilon) dU_t \end{aligned} \quad (4.12)$$

where  $X^\varepsilon \in \mathbb{R}^{n_x}$ ,  $Y^\varepsilon \in \mathbb{R}^{n_y}$ ,  $V : \mathbb{R}^{n_x+n_y} \rightarrow \mathbb{R}$ ,  $\gamma_1 : \mathbb{R}^{n_x+n_y} \rightarrow \mathbb{R}^{n_x \times n_x}$ ,  $\gamma_2 : \mathbb{R}^{n_x+n_y} \rightarrow \mathbb{R}^{n_y \times n_y}$  and  $\Gamma_1 = \gamma_1 \gamma_1^T$ ,  $\Gamma_2 = \gamma_2 \gamma_2^T$  and  $W_t, U_t$  are  $n_x, n_y$ -dimensional standard Brownian motions respectively. This dynamics possesses the generator

$$\mathcal{L} = \underbrace{(-\Gamma_1 \nabla_x V + \nabla_x \cdot \Gamma_1) \cdot \nabla_x + \Gamma_1 : \nabla_x^2}_{\mathcal{L}_1} + \frac{1}{\varepsilon} \underbrace{(-\Gamma_2 \nabla_y V + \nabla_y \cdot \Gamma_2) \cdot \nabla_y + \Gamma_2 : \nabla_y^2}_{\mathcal{L}_2}$$

which we decompose into  $\mathcal{L}_1$  and  $\mathcal{L}_2$  and further we assume the coefficients to be such that this dynamics admits a unique invariant measure  $\mu$ . Now, the averaging measure  $\nu_x$  will be the one satisfying

$$\mathcal{L}_2^* \nu_x = 0$$

where the coefficients of  $\mathcal{L}_2^*$ , i.e.,  $\Gamma_2$  and  $\nabla_y V$  are evaluated at a given  $x$ . This means  $\nu_x(y) \sim e^{-V_x(y)}$  where  $V_x(y)$  is supposed to express that the potential is evaluated at a fixed  $x$  thus treated as a parameter and the dependence is only in  $y$ . In turn, computing the conditional  $\mu_x$  from the invariant distribution yields exactly the same. Conditioning on  $X = x$  means fixing  $x$  in  $\mu \sim e^{-V(x,y)}$  thus turning it into a parameter of the conditional distribution  $\mu_x \sim e^{-V_x(y)}$ . Hence, in the instance of reversible dynamics we never observe any distinction between the two approaches. Note that when considering reversible dynamics of a slow fast structure, the above example (4.12) is in the most general form. If we had a diffusion matrix which coupled the noise of the slow and fast equation, the time scale separation parameter would inevitably also show up in the slow equation, hence giving a different kind of scaling such that a comparison to averaging would not be appropriate.

### Non-reversible dynamics

Let us discuss in more detail the example from the beginning of the chapter where we introduce an additional parameter  $\delta \geq 0$  which was set equal to 0 previously. More precisely, we consider the two-dimensional linear SDE given by

$$\begin{aligned} dX_t &= (-X_t + Y_t) dt + \sqrt{\delta} dW_t \\ dY_t &= -\frac{1}{\varepsilon} Y_t dt + \frac{1}{\sqrt{\varepsilon}} dU_t, \end{aligned} \tag{4.13}$$

with  $W, U$  being independent standard Brownian motions. Solving the corresponding Lyapunov equation

$$\begin{pmatrix} -1 & 1 \\ 0 & -1/\varepsilon \end{pmatrix} \Sigma(\varepsilon) + \Sigma(\varepsilon) \begin{pmatrix} -1 & 0 \\ 1 & -1/\varepsilon \end{pmatrix} = - \begin{pmatrix} \delta & 0 \\ 0 & 1/\varepsilon \end{pmatrix}$$

equips us with the invariant measure  $\mu^\varepsilon \sim \mathcal{N}(0, \Sigma(\varepsilon))$ , with

$$\Sigma(\varepsilon) = \frac{1}{2} \begin{pmatrix} [\varepsilon + \delta(1 + \varepsilon)] / (1 + \varepsilon) & \varepsilon / (1 + \varepsilon) \\ \varepsilon / (1 + \varepsilon) & 1 \end{pmatrix}.$$

Hence, the conditional measure for given  $x$  reads  $\mu_x^\varepsilon \sim \mathcal{N}(m^c(\varepsilon), \Sigma^c(\varepsilon))$ , with

$$m^c(\varepsilon) = \frac{\varepsilon}{\varepsilon + \delta(1 + \varepsilon)} x, \quad \Sigma^c(\varepsilon) = \frac{1}{2} - \frac{\varepsilon}{2(1 + \delta(1/\varepsilon + 2 + \varepsilon) + \varepsilon)}.$$

On the other hand, the invariant measure for the fast variable  $Y$  which in this example is independent of  $x$  reads  $\nu_x \sim \mathcal{N}(m^{\text{av}}, \Sigma^{\text{av}})$  with

$$m^{\text{av}} = 0, \quad \Sigma^{\text{av}} = \frac{1}{2}.$$

Observe that for  $\delta = 0$  we have  $m^c(\varepsilon) \equiv x$  and hence  $\mu_x^\varepsilon \rightarrow \nu_x$  as  $\varepsilon \rightarrow 0$  whereas for  $\delta = 1$  we find  $\mu_x^\varepsilon \rightarrow \nu_x$  as  $\varepsilon \rightarrow 0$ . This raises the idea to study the convergence of  $\mu_x^\varepsilon$  as  $\varepsilon \rightarrow 0$  as a function of  $\delta$ . More specifically, we let  $\delta = \varepsilon^p$ ,  $p \geq 0$  which implies that we can still do averaging and the relation to the averaging result holds up to a time of  $\mathcal{O}(1)$  (but not  $\mathcal{O}(\varepsilon^{-p})!$ ). Computing the limit of the mean  $m^c(\varepsilon)$  as  $\varepsilon \rightarrow 0$  yields

$$m^c(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} \begin{cases} x, & p > 1, \\ \frac{1}{2}x, & p = 1, \\ 0, & 0 \leq p < 1. \end{cases}$$

This means for  $0 \leq p < 1$  the reduced dynamics obtained by averaging and conditional expectations coincide in the infinite time scale separation limit. We will discuss this observation in a more systematic way for linear diffusions in the next section, where we also discuss the *degree of irreversibility* and its relation to the agreement of averaging and conditional expectation in the limit as  $\varepsilon \rightarrow 0$ .

Further note that  $\Sigma^c(\varepsilon) \rightarrow \frac{1}{2}$  for any  $p \geq 0$ , that is, the variance always agrees.

### 4.3.3 Sufficient conditions for the conformity of averaging and conditional expectations

In this section we aim at giving a general idea how differences in averaging and conditional expectations can arise in the limit  $\varepsilon \rightarrow 0$ . To this end, recall the general setting of this section given by (4.1). The associated generator  $\mathcal{L}$  can be decomposed into a slow and a fast part as follows

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_1 + \frac{1}{\varepsilon} \mathcal{L}_2, \quad \text{where} \quad \mathcal{L}_1 = f_1 \cdot \nabla_x + \Gamma_1 : \nabla_x^2 \\ &\quad \mathcal{L}_2 = f_2 \cdot \nabla_y + \Gamma_2 : \nabla_y^2. \end{aligned}$$

Assume that  $Z_t^\varepsilon$  admits a unique invariant distribution  $\mu^\varepsilon$ , i.e.,

$$\mathcal{L}^* \mu^\varepsilon = 0.$$

Decompose  $\mu^\varepsilon$  into its marginal and conditional part as follows

$$\mu^\varepsilon(x, y) = \mu_x^\varepsilon(y) \hat{\mu}^\varepsilon(x),$$

i.e.,  $\hat{\mu}$  is the marginal distribution of  $X$  and  $\mu_x$  is the conditional distribution of  $y$  for given  $X = x$ . Suppose that the fast process is ergodic, i.e.,  $\forall x \in \mathbb{R}^{n_x}$  there exists a unique  $\nu_x$  such that

$$\mathcal{L}_2^* \nu_x = 0.$$

Further assume that  $\hat{\mu}$  and  $\mu_x$  admit the perturbation expansions

$$\hat{\mu} = \hat{\mu}^0 + \varepsilon \hat{\mu}^1 + \mathcal{O}(\varepsilon^2) \quad \text{and} \quad \mu_x = \mu_x^0 + \varepsilon \mu_x^1 + \mathcal{O}(\varepsilon^2).$$

We can then state the following conjecture which is supported by formal calculations below.

**Conjecture 4.13.** If  $\hat{\mu}^0 = \lim_{\varepsilon \rightarrow 0} \hat{\mu}^\varepsilon$  has full support, that is,  $\hat{\mu}^0(x) > 0 \forall x \in \mathbb{R}^{n_x}$ , then

$$\forall x \in \mathbb{R}^{n_x} : \mu_x^\varepsilon \rightarrow \nu_x, \text{ as } \varepsilon \rightarrow 0.$$

*Proof.* We write

$$\mu = \mu_x^0 \hat{\mu}^0 + \varepsilon (\mu_x^1 \hat{\mu}^0 + \mu_x^0 \hat{\mu}^1) + \mathcal{O}(\varepsilon^2). \quad (4.14)$$

Since  $\mu$  is the invariant distribution of  $Z_t^\varepsilon$  it holds that

$$\mathcal{L}^* \mu = 0,$$

which we rewrite with the perturbation Ansatz. The leading order term, that is the  $\mathcal{O}(1/\varepsilon)$ -term, reads

$$\mathcal{L}_2^* (\mu_x^0 \hat{\mu}^0) = 0.$$

Since  $\mathcal{L}_2^*$  is a differential operator in  $y$  only, it does not act on  $\hat{\mu}^0$  which only depends on  $x$ . Hence, we have that

$$\mathcal{L}_2^* (\mu_x^0 \hat{\mu}^0) = 0 \quad \Leftrightarrow \quad \hat{\mu}^0 \mathcal{L}_2^* (\mu_x^0) = 0.$$

Finally, observe that  $\hat{\mu}^0 \mathcal{L}_2^* \mu_x^0 = 0, \forall x \in \mathbb{R}^{n_x}$  then either  $\mathcal{L}_2^* \mu_x^0 = 0$  or  $\hat{\mu}^0 = 0$ . Since by assumption  $\hat{\mu}^0(x) > 0, \forall x \in \mathbb{R}^{n_x}$  it follows that

$$\mathcal{L}_2^* \mu_x^0 = 0 \quad \forall x \in \mathbb{R}^{n_x},$$

which is precisely the statement that the  $\mathcal{O}(1)$ -term of  $\mu_x^\varepsilon$  agrees with  $\nu_x$ , i.e.,  $\mu_x^\varepsilon \rightarrow \nu_x$ .  $\square$

### Connection to the guiding example

Establishing a relation of the above result to the example (4.13) recall that here  $\mu^\varepsilon \sim \mathcal{N}(0, \Sigma(\varepsilon))$  where

$$\Sigma(\varepsilon) = \frac{1}{2} \begin{pmatrix} [\varepsilon + \varepsilon^p(1 + \varepsilon)] / (1 + \varepsilon) & \varepsilon / (1 + \varepsilon) \\ \varepsilon / (1 + \varepsilon) & 1 \end{pmatrix} \xrightarrow{\varepsilon \rightarrow 0} \begin{cases} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, & p = 0 \\ \begin{pmatrix} 0 & 0 \\ 0 & 1/2 \end{pmatrix}, & p > 0. \end{cases}$$

Hence the condition  $\mu^0(x) > 0$  is clearly fulfilled if  $p = 0$  and this includes the conformity of the two approaches. Yet, again we find that this can only be a sufficient condition, since for  $0 < p < 1$  due to the above considerations the assumption of the conjecture  $\mu^0(x) > 0$  are not met – in fact  $x$  and  $y$  decouple and in  $x$  we find a dirac at 0 – still  $\mu_x^\varepsilon \rightarrow \nu_x$  holds true.



# ANALYTICAL RESULTS FOR OU-PROCESSES

This chapter is devoted to OU-processes for which it is possible to make the analysis of the previous chapter precise. In particular, in the first part of the chapter, we discuss the conformity of averaging and conditional expectations in the limit  $\varepsilon \rightarrow 0$  and establish conditions on the drift and diffusion matrices  $A$  and  $C$  for the conformity to hold. Our analysis involves the study of the arising distributions with respect to which the averages of the coefficients are computed. Moreover, we consider a measure attributing the degree of irreversibility and summarise our conjectures at the end of Section 5.1. Furthermore, we extend the comparison of the two approaches by studying numerical examples for finite values of  $\varepsilon > 0$ .

The second part of the chapter deals with the issue of sampling conditional distributions for non-reversible processes, which is, for example, needed in order to compute the coefficients of the effective dynamics by MCMC. While this is an easy task when conditioning on certain degrees of freedom of a reversible process, which amounts to keeping these degrees of freedom fixed, things become involved when working with a non-reversible process instead. Our contribution should be seen as a first step towards sampling from conditional distributions of non-reversible processes. We find that an oblique projection – in contrast to a orthogonal projection for the reversible case – yields the correct sampling scheme for OU-processes. In Section 5.2.3 we propose a scalar product, provided the drift matrix fulfils certain conditions, with respect to which the considered projection becomes orthogonal and is further hoped to provide structural insight on the problem at hand.

## 5.1 Comparison of averaging and conditional expectations

In this section we make the comparison between averaging and conditional expectations explicit for the general class of Ornstein-Uhlenbeck processes described by

the following SDEs

$$dZ_t = \begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ \frac{1}{\varepsilon}A_{21} & \frac{1}{\varepsilon}A_{22} \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + \begin{pmatrix} \sqrt{\varepsilon^p}C_{11} & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}}C_{22} \end{pmatrix} \begin{pmatrix} dW_t \\ dU_t \end{pmatrix}, \quad (5.1)$$

here  $0 < \varepsilon \ll 1$  is the time scale separation parameter,  $p \geq 0$ ,  $A_{11}, C_{11} \in \mathbb{R}^{n_x \times n_x}$ ,  $A_{12}, C_{12} \in \mathbb{R}^{n_x \times n_y}$ ,  $A_{21}, C_{21} \in \mathbb{R}^{n_y \times n_x}$ ,  $A_{22}, C_{22} \in \mathbb{R}^{n_y \times n_y}$  and  $W_t, U_t$  are  $n_x, n_y$ -dimensional standard Brownian motions respectively.

We denote by

$$A^\varepsilon = \begin{pmatrix} A_{11} & A_{12} \\ \frac{1}{\varepsilon}A_{21} & \frac{1}{\varepsilon}A_{22} \end{pmatrix}, \quad C^\varepsilon = \begin{pmatrix} \sqrt{\varepsilon^p}C_{11} & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}}C_{22} \end{pmatrix}$$

the  $\varepsilon$ -dependent matrices. We make the usual assumptions which guarantee existence and uniqueness of the invariant measure of the full process (and hence of the conditional measure) as well as of the fast process  $Y$ :

**Assumption A5.** Assume that  $\forall \varepsilon > 0$  the matrices  $A^\varepsilon$  and  $A_{22}$  are Hurwitz and the matrix pairs  $(A^\varepsilon, C^\varepsilon)$  and  $(A_{22}, C_{22})$  are controllable.

The corresponding Lyapunov equation of the full process  $Z_t$  reads

$$A^\varepsilon \Sigma^\varepsilon + \Sigma^\varepsilon (A^\varepsilon)^\top = -C^\varepsilon (C^\varepsilon)^\top. \quad (5.2)$$

The measure conditioned on  $X = x$  is denoted by  $\mu_x^\varepsilon = \mathcal{N}(m^c(\varepsilon), \Sigma^c(\varepsilon))$  and is well defined for all  $\varepsilon > 0$ . The mean and covariance read

$$m^c(\varepsilon) = \Sigma_{21}^\varepsilon (\Sigma_{11}^\varepsilon)^{-1} x, \quad \Sigma^c(\varepsilon) = \Sigma_{22}^\varepsilon - \Sigma_{21}^\varepsilon (\Sigma_{11}^\varepsilon)^{-1} \Sigma_{12}^\varepsilon.$$

Similarly, under the above assumptions on  $A_{22}$  and  $(A_{22}, C_{22})$  we denote the invariant measure of the fast process  $(Y_t)_{t \geq 0}$  by  $\nu_x = \mathcal{N}(m^{\text{av}}, \Sigma^{\text{av}})$ , where

$$m^{\text{av}} = -A_{22}^{-1} A_{21} x, \quad A_{22} \Sigma^{\text{av}} + \Sigma^{\text{av}} A_{22}^\top = -C_{22} C_{22}^\top.$$

The following proposition states a sufficient condition to assure that the two measures agree as  $\varepsilon \rightarrow 0$ , which – since we are dealing with Gaussians here – is equivalent to  $m^c(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} m^{\text{av}}$  and  $\Sigma^c(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} \Sigma^{\text{av}}$  as  $\varepsilon \rightarrow 0$ .

**Proposition 5.1.** *Let A5 hold true. Further assume that  $A_{11} - A_{12} A_{22}^{-1} A_{21}$  is Hurwitz and the pair  $(A_{11} - A_{12} A_{22}^{-1} A_{21}, C_{11})$  is controllable and  $0 \leq p < 1$  then*

$$m^c(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} m^{\text{av}} \quad \text{and} \quad \Sigma^c(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} \Sigma^{\text{av}}.$$

As an immediate consequence, the LSI constant  $\alpha_{\text{LSI}}^\varepsilon$  remains bounded as  $\varepsilon \rightarrow 0$  and is given by the LSI constant of  $\nu_x$ :

**Corollary 5.2.** *Under the same assumptions of Proposition 5.1 the conditional measure  $\lim_{\varepsilon \rightarrow 0} \mu_x^\varepsilon =: \mu_x^0$  satisfies a static LSI with constant  $\alpha_{\text{LSI}}^0$  determined by*

$$(\Sigma^{\text{av}})^{-1} \geq \alpha_{\text{LSI}}^0 I_{n_y \times n_y}.$$



*Proof of Corollary 5.2.* The proof is an immediate consequence of the Bakry-Emery criterion (see Theorem 3.21): we know that  $\nu_x(y) \sim e^{-V(y)} dy$  with  $V(y) = \frac{1}{2}y^T(\Sigma^{\text{av}})^{-1}y$  and thus  $\nabla_y^2 V(y) = (\Sigma^{\text{av}})^{-1}$  and the corresponding LSI constant  $\alpha_{\text{LSI}}^0$  is given by the smallest eigenvalue of  $(\Sigma^{\text{av}})^{-1}$  or equivalently by the inverse of the largest eigenvalue.  $\square$

*Proof of Proposition 5.1.* Our intention is to compare  $\nu_x \sim \mathcal{N}(-A_{22}^{-1}A_{21}x, \Sigma^{\text{av}})$  to the conditional measure  $\mu_x = \mathcal{N}(\Sigma_{21}^\varepsilon(\Sigma_{11}^\varepsilon)^{-1}x, \Sigma_{22}^\varepsilon - \Sigma_{21}^\varepsilon(\Sigma_{11}^\varepsilon)^{-1}\Sigma_{12}^\varepsilon)$ . Here  $\Sigma_{ij}^\varepsilon$  refers to the  $(i, j)^{\text{th}}$  block of  $\Sigma^\varepsilon$  solution to (5.2).

Writing out the equation for  $\Sigma^\varepsilon$  in its different components yields the following three equations

$$A_{11}\Sigma_{11}^\varepsilon + A_{12}\Sigma_{21}^\varepsilon + \Sigma_{11}^\varepsilon A_{11}^\text{T} + \Sigma_{12}^\varepsilon A_{12}^\text{T} = -\varepsilon^p C_{11} C_{11}^\text{T} \quad (5.3)$$

$$A_{11}\Sigma_{12}^\varepsilon + A_{12}\Sigma_{22}^\varepsilon + \frac{1}{\varepsilon}(\Sigma_{11}^\varepsilon A_{21}^\text{T} + \Sigma_{12}^\varepsilon A_{22}^\text{T}) = 0 \quad (5.4)$$

$$A_{21}\Sigma_{12}^\varepsilon + A_{22}\Sigma_{22}^\varepsilon + \Sigma_{21}^\varepsilon A_{21}^\text{T} + \Sigma_{22}^\varepsilon A_{22}^\text{T} = -C_{22} C_{22}^\text{T}. \quad (5.5)$$

Due to the structure of the above equations we make the following perturbation Ansatz for  $\Sigma^\varepsilon$ :

$$\Sigma^\varepsilon = \Sigma^{(0)} + \varepsilon^p \Sigma^{(p)} + \varepsilon \Sigma^{(1)} + \mathcal{O}(\varepsilon^2).$$

**Case 1:**  $0 < p < 1$

Collecting the same powers of  $\varepsilon$  in (5.4) we find that

$$\mathcal{O}(\varepsilon^{-1}): \quad \Sigma_{12}^{(0)} = -\Sigma_{11}^{(0)} A_{21}^\text{T} A_{22}^\text{T} \quad (5.6)$$

$$\mathcal{O}(\varepsilon^{p-1}): \quad \Sigma_{12}^{(p)} = -\Sigma_{11}^{(p)} A_{21}^\text{T} A_{22}^\text{T}. \quad (5.7)$$

Observe that if  $A_{21} = 0$  then by (5.6)-(5.7)  $\Sigma_{12}^\varepsilon = \mathcal{O}(\varepsilon)$ . Furthermore, if  $A_{21} = 0$  and  $C_{11} = 0$  then by (5.3) also  $\Sigma_{11}^\varepsilon = \mathcal{O}(\varepsilon)$ .

Plugging (5.6) and (5.7) into (5.3) respectively and collecting the same powers of  $\varepsilon$  again we find that

$$\mathcal{O}(\varepsilon^p): \quad (A_{11} - A_{12}A_{22}^{-1}A_{21})\Sigma_{11}^{(p)} + \Sigma_{11}^{(p)}(A_{11} - A_{12}A_{22}^{-1}A_{21})^\text{T} = -C_{11}C_{11}^\text{T} \quad (5.8)$$

$$\mathcal{O}(1): \quad (A_{11} - A_{12}A_{22}^{-1}A_{21})\Sigma_{11}^{(0)} + \Sigma_{11}^{(0)}(A_{11} - A_{12}A_{22}^{-1}A_{21})^\text{T} = 0. \quad (5.9)$$

The last equation (5.9) implies that  $\Sigma_{11}^{(0)} = 0$  and hence by (5.6) also  $\Sigma_{12}^{(0)} = 0$ . This holds independent of specific choices for  $A$  or  $C$ , hence we conclude that

$$\Sigma_{11}^{(0)} = 0 \quad \text{and} \quad \Sigma_{12}^{(0)} = 0. \quad (5.10)$$

Equation (5.8) together with the assumption that  $((A_{11} - A_{12}A_{22}^{-1}A_{21}), C_{11})$  is controllable implies that  $\Sigma_{11}^{(p)}$  is invertible (cf. [70, Theorem 1.2]). Comparing the means and noting – since  $\Sigma_{11}^{(0)} = 0$  – we have, for small  $\varepsilon$ , that

$$\begin{aligned} (\Sigma_{11}^\varepsilon)^{-1} &= \varepsilon^{-p}(\Sigma_{11}^{(p)} + \varepsilon^{1-p}\Sigma_{11}^{(1)} + \mathcal{O}(\varepsilon^2))^{-1} \\ &\approx \varepsilon^{-p}(\Sigma_{11}^{(p)})^{-1} - \varepsilon^{1-p}(\Sigma_{11}^{(p)})^{-1}(\Sigma_{11}^{(1)})(\Sigma_{11}^{(p)})^{-1} + \mathcal{O}(\varepsilon^{2(1-p)}) \end{aligned}$$

and using (5.7) we find

$$m^c(\varepsilon) = \Sigma_{21}^\varepsilon (\Sigma_{11}^\varepsilon)^{-1} x = -A_{22}^{-1} A_{21} x + \mathcal{O}(\varepsilon^{1-p}) \xrightarrow{\varepsilon \rightarrow 0} m^{\text{av}} = -A_{22}^{-1} A_{21} x. \quad (5.11)$$

Note that the above consideration remain unchanged if  $A_{21} = 0$  and thus  $m^{\text{av}} = 0$ . Only in the case that  $C_{11} = 0$ , such that both  $\Sigma_{12}^{(p)} = \Sigma_{11}^{(p)} = 0$ , the  $\mathcal{O}(\varepsilon)$  terms can lead to a non-trivial change of the mean for the conditional measure and this is indeed what we observe in the example from the beginning of the chapter.

Regarding the variance we note that since

$$\Sigma^c(\varepsilon) = \Sigma_{22}^\varepsilon - \Sigma_{21}^\varepsilon (\Sigma_{11}^\varepsilon)^{-1} \Sigma_{12}^\varepsilon = \Sigma_{22}^{(0)} - \varepsilon^p \Sigma_{21}^{(p)} (\Sigma_{11}^{(p)})^{-1} \Sigma_{12}^{(p)} + \mathcal{O}(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} \Sigma_{22}^{(0)}$$

and the  $\mathcal{O}(1)$ - term equation of (5.5) reads

$$A_{22} \Sigma_{22}^{(0)} + \Sigma_{22}^{(0)} A_{22}^\top = -C_{22} C_{22}^\top,$$

which has a unique positive definite solution by our assumption, also the variances agree as  $\varepsilon \rightarrow 0$  since  $\Sigma_{22}^{(0)}$  and  $\Sigma^{\text{av}}$  are solutions to the same Lyapunov equation.

### Case 2: $p=0$

If  $p = 0$  then equations (5.6)-(5.7) and (5.8)-(5.9) collapse into one equation respectively, i.e., the corresponding equations read

$$\begin{aligned} \mathcal{O}(\varepsilon^{-1}) : \quad & \Sigma_{12}^{(0)} = -\Sigma_{11}^{(0)} A_{21}^\top A_{22}^{-\top} \\ \mathcal{O}(\varepsilon^p) : \quad & (A_{11} - A_{12} A_{22}^{-1} A_{21}) \Sigma_{11}^{(0)} + \Sigma_{11}^{(0)} (A_{11} - A_{12} A_{22}^{-1} A_{21})^\top = -C_{11} C_{11}^\top \end{aligned}$$

which means  $\Sigma_{11}^{(0)} > 0$  and the inverse can be expressed as  $(\Sigma_{11}^{(0)})^{-1} \approx (\Sigma_{11}^{(0)})^{-1} - \varepsilon (\Sigma_{11}^{(0)})^{-1} (\Sigma_{11}^{(1)}) (\Sigma_{11}^{(0)})^{-1} + \mathcal{O}(\varepsilon^2)$ . Thus the calculations for the mean remain valid:  $m^c(\varepsilon) = \Sigma_{21}^\varepsilon (\Sigma_{11}^\varepsilon)^{-1} x = -A_{22}^{-1} A_{21} x + \mathcal{O}(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} m^{\text{av}} = -A_{22}^{-1} A_{21} x$  and for the variance we have

$$\Sigma^c(\varepsilon) = \Sigma_{22}^{(0)} - \Sigma_{21}^{(0)} (\Sigma_{11}^{(0)})^{-1} \Sigma_{12}^{(0)} + \mathcal{O}(\varepsilon)$$

but also the  $\mathcal{O}(1)$ - term equation of (5.5) now reads

$$A_{22} (\Sigma_{22}^{(0)} - \Sigma_{21}^{(0)} (\Sigma_{11}^{(0)})^{-1} \Sigma_{12}^{(0)}) + (\Sigma_{22}^{(0)} - \Sigma_{21}^{(0)} (\Sigma_{11}^{(0)})^{-1} \Sigma_{12}^{(0)}) A_{22}^\top = -C_{22} C_{22}^\top,$$

i.e., our conclusions are unchanged.  $\square$

**Remark 5.3.** By Proposition 5.1 we find that a sufficient condition for averaging and conditional expectations to agree in the limit  $\varepsilon \rightarrow 0$  is the controllability assumption of the pair  $(A_{11} - A_{12} A_{22}^{-1} A_{21}, C_{11})$  which matches the observations of the example: For  $p = 0$  we have  $C_{11} = 1$  and the condition is fulfilled. Yet, even for  $p > 0$  we have  $C_{11} \rightarrow 0$ , i.e., the conditions are not met, still for  $0 < p < 1$  the approaches agree.

Other conditions, for example the *degree of irreversibility*, would also be reasonable candidates to find necessary or sufficient criterions for the (non-)conformity of the approaches, since only for non-reversible processes conditional expectations and averaging possibly differ (see Section 4.3.2). The degree of irreversibility can be measured, e.g., by the *entropy production rate* defined as the relative entropy between the path-measure  $\rho_{[0,T]}^+$  of the forward process described by (5.1) and the path-measure of the time-reversed process associated to (5.1)  $\rho_{[0,T]}^-$  (see e.g. [40]; for a definition of the time-reversed process see (2.12)). Computing the degree of irreversibility for the Example 4.13 and letting  $\delta = \varepsilon^p$ , we find that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathcal{H}(\rho_{[0,T]}^+ | \rho_{[0,T]}^-) = \frac{2}{\varepsilon^p(1 + \varepsilon)} \xrightarrow{\varepsilon \rightarrow 0} \begin{cases} \infty, & p > 0 \\ 2, & p = 0. \end{cases}$$

This means, also the degree of irreversibility being finite can only serve as a sufficient condition for the two approaches to agree. Conversely, taking the value infinity seems to be a necessary condition for them to disagree.

Summarising, we find the following **conjectures for a sufficient condition for the conformity of the approaches**. The first two offer formal calculations supporting them, the last one is a pure conjecture.

- for OU processes: controllability of drift and diffusion for all  $\varepsilon > 0$  together with controllability of the Schur complement of the drift and the diffusion of the slow variable  $(A_{11} - A_{12}A_{22}^{-1}A_{21}, C_{11})$  (see Proposition 5.1),
- hypo-ellipticity of the full generator, which guarantees strict positivity of the invariant measure (see Conjecture 4.13),
- finiteness of the degree of irreversibility given by the entropy production rate between the forward and backward process (see Remark 5.3).

### 5.1.1 Numerical comparison for $\varepsilon > 0$

In this section we compare averaging to effective dynamics for finite values of  $\varepsilon > 0$ . We will compute on the one hand the path-wise error of the averaged and effective dynamics, respectively. On the other hand, we compute the relative entropy of the time  $t$  marginals. The path-wise error requires the numerical solution of the respective SDE and will be approximated by Monte Carlo, i.e.,

$$\mathbb{E} \left( \sup_{t \in [0, T]} |X_t - X_t^{\text{av/eff}}| \right) \approx \frac{1}{N} \sum_{j=1}^N \max_{i \in \{1, \dots, M\}} |x_j(i) - x_j^{\text{av/eff}}(i)|$$

where  $x_j(i)$  corresponds to the  $j^{\text{th}}$  sample – with  $j \in \{1, \dots, N\}$  and  $N$  is the number of samples – at time step  $i \in \{1, \dots, M\}$  and  $M$  is the number of time steps, for which  $Mh = T$  where  $h$  the step size of the numerical method.

We consider the following examples with averaging scaling, choose the parameters

$T = 1, h = 5 \cdot 10^{-5}, N = 100$  and let the dynamics evolve from  $x(0) = (1, 1)^T$  for Examples 1 and 3,  $x(0) = (1, 1, 1)^T$  for Example 2:

Example 1:  $B^\varepsilon = \begin{pmatrix} -2 & -1 \\ -\frac{1}{\varepsilon} & -\frac{1}{\varepsilon} \end{pmatrix}, A^\varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & \frac{2}{\sqrt{\varepsilon}} \end{pmatrix},$

Example 2:  $B^\varepsilon = \begin{pmatrix} -2 & -1 & -1 \\ -\frac{1}{\varepsilon} & -\frac{3}{\varepsilon} & \frac{3}{\varepsilon} \\ -\frac{1}{\varepsilon} & -\frac{1}{2\varepsilon} & -\frac{1}{\varepsilon} \end{pmatrix}, A^\varepsilon = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\varepsilon}} \end{pmatrix},$

and additionally one example with a third time scale suitable for homogenization.

Example 3:  $B^\varepsilon = \begin{pmatrix} -2 & -\frac{1}{\sqrt{\varepsilon}} \\ \frac{1}{\sqrt{\varepsilon}} & -1 \end{pmatrix}, A^\varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}} \end{pmatrix}.$

Let us first discuss the results for for Examples 1 and 2, which correspond to the

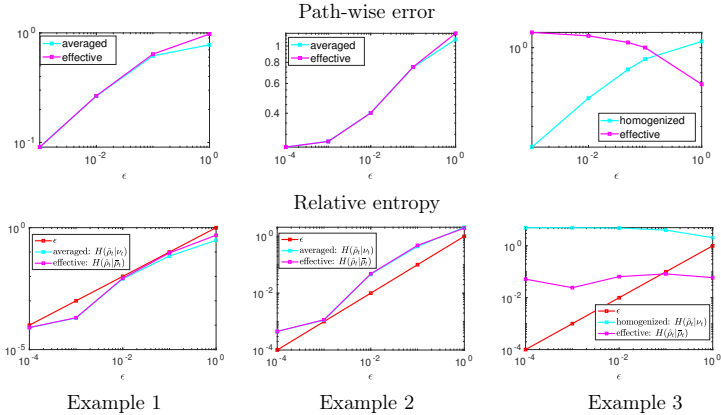


Figure 5.1: Top: Plot of the path-wise error of the averaged/effective dynamics against  $\varepsilon$  on a doubly-logarithmic scale. Bottom: Plot of  $\max_{t>0} \mathcal{H}(\hat{\rho}_t^\varepsilon | \nu_t)$  (averaged/homogenized) and  $\max_{t>0} \mathcal{H}(\hat{\rho}_t^\varepsilon | \bar{\rho}_t^\varepsilon)$  (effective). Left: Example 1, middle: Example 2, right: Example 3.

first and middle column of Figure 5.1. The path-wise error (top row in Figure 5.1) decreases for  $\varepsilon \rightarrow 0$  and the two methods become indistinguishable matching the theory of Section 5.1. For  $\varepsilon$  of order one, averaging performs slightly better in path-wise sense, but this can change if the initial conditions are altered. Measuring the error instead in relative entropy of the time  $t$  marginals yields a similar picture. Summarising, these results suggest that for an averaging scaling there is no clear judgement possible, which of the methods is preferable for finite  $\varepsilon$ . Let us remark that considering non-linear examples would be desirable, unfortunately

non-linearity together with an averaging scaling renders the computation of the effective dynamics impossible. We refrain from using approximation methods for the effective coefficients here, since they would introduce additional numerical errors hence making the comparison too involved.

Contrary to the averaging scaling, the third example, which suffices a homogenization scaling, gives a completely different picture: for the path-wise error the homogenized approximation clearly outperforms effective dynamics as  $\varepsilon \rightarrow 0$ . Interestingly, considering relative entropy of the time  $t$  marginals instead, the roles change. Even though there seems to be no  $\varepsilon$ -dependency of the error, effective dynamics yields a better approximation in this case. This is presumably due to the fact that by construction the effective dynamics approach the same invariant measure as the original process, i.e., for  $t \rightarrow \infty$  we have that  $\mathcal{H}(\hat{\rho}_t^\varepsilon | \bar{\rho}_t^\varepsilon) \rightarrow 0$  and this is not true for the homogenized dynamics, which admits a different stationary distribution.

## 5.2 Sampling from conditional distributions

We have already learned that for non-reversible processes averaging does not necessarily give the same result as conditional expectations. This involves in particular the fact that the fast process, while fixing the slow one, does not sample the conditional measure. In this section we will introduce projections which enable us to sample from conditional distributions for non-reversible OU-processes. To make the problem clearly visible and to fix the notation for the section, consider the process  $Z = (X, Y)^T$ ,  $X \in \mathbb{R}^{n_x}$ ,  $Y \in \mathbb{R}^{n_y}$ ,  $n = n_x + n_y$  whose dynamics written in blockform are given by

$$\begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} dt + \sqrt{2} \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix} \begin{pmatrix} dB_t^x \\ dB_t^y \end{pmatrix}, \quad (5.12)$$

where  $A_{11} \in \mathbb{R}^{n_x \times n_x}$ ,  $A_{12} \in \mathbb{R}^{n_x \times n_y}$ ,  $A_{21} \in \mathbb{R}^{n_y \times n_x}$ ,  $A_{22} \in \mathbb{R}^{n_y \times n_y}$ ,  $C_{11} \in \mathbb{R}^{n_x \times n_x}$ ,  $C_{12} \in \mathbb{R}^{n_x \times n_y}$ ,  $C_{21} \in \mathbb{R}^{n_y \times n_x}$ ,  $C_{22} \in \mathbb{R}^{n_y \times n_y}$ . Here  $B_t^x, B_t^y$  are  $n_x, n_y$ -dimensional standard Brownian motions respectively.

Under the usual assumptions A2 on  $A$  and  $C$  the process has a unique invariant measure  $\mu \sim \mathcal{N}(0, \Sigma)$  where  $\Sigma$  solves the *Lyapunov equation*

$$A\Sigma + \Sigma A^T = -2CC^T, \quad (5.13)$$

which can also be formulated in a block wise fashion giving three equations (since the off diagonal equations are the same) and we will exploit this formulation in the proofs of the consecutive sections.

### The conditional distribution

For a normally distributed random variable  $Z = (X, Y) \sim \mathcal{N}((\mu_1, \mu_2)^T, \Sigma)$  it is easy to compute the distribution of the conditioned variable  $(X|Y = y)$  which – employing completion of the square – turns out to be also normally distributed with mean  $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_2)$  and variance  $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ . Hence, the conditional distribution we want to sample from is given by  $\mathcal{N}(\Sigma_{12}\Sigma_{22}^{-1}y, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ . Note that, opposed to the previous section where we considered conditional expectations with respect to the condition  $X = x$ , we change the role of  $X$  and  $Y$  here and consider now the condition  $Y = y$ .

### The constrained dynamics

The aim is to constrain the dynamics in such a way that the invariant measure equals the target conditional distribution. Obviously  $y$  should be fixed, i.e.,  $\dot{Y}_t = 0$  and  $Y_0 = y$ . Employing a projection given by

$$P = \begin{pmatrix} I_{n_x \times n_x} & \alpha \\ 0 & 0 \end{pmatrix}$$

and thus replacing  $A$  by  $PA$  and  $C$  by  $PC$  in the dynamics (5.12) we achieve that that  $\dot{Y}_t = 0$ . The choice for the projection lies within  $\alpha$  and the question is how to

choose  $\alpha$  such that the projected dynamics produce the correct conditional distribution. If the dynamics are reversible and  $C$  is block diagonal it will turn out that the orthogonal projection, i.e.,  $\alpha = 0$  yields the correct the distribution.

Note that the replacement of  $A$  and  $C$  by its projected versions is not a transformation of the original dynamics (5.12), but defines a new process  $(\hat{Z}_t)_{t \geq 0} = (\hat{X}_t, \hat{Y}_t)_{t \geq 0}$  given by the original one plus constraints. The corresponding SDE reads

$$d\hat{Z}_t = PA\hat{Z}_t dt + PC dB_t,$$

or equivalently with  $\tilde{A}_1^{-1} = A_{11} + \alpha A_{21}$ ,  $C_1 = (C_{11}, C_{12})$ ,  $C_2 = (C_{21}, C_{22})$ ,  $B = ((B^x)^T, (B^y)^T)^T$ :

$$\begin{aligned} d\hat{X}_t &= \tilde{A}_1^{-1} \left( \hat{X}_t + \tilde{A}_1 (A_{12} + \alpha A_{22}) \hat{Y}_t \right) dt + \sqrt{2}(C_1 + \alpha C_2) dB_t & (5.14) \\ d\hat{Y}_t &= 0. \end{aligned}$$

### 5.2.1 A covariance preserving projection

Let us state the first result that yields the correct covariance but not yet the correct mean.

**Theorem 5.4.** *Let  $\alpha = -\Sigma_{12}\Sigma_{22}^{-1}$  and assume that  $(A_{11} + \alpha A_{21})$  is Hurwitz and  $((A_{11} + \alpha A_{21}), C_1 + \alpha C_2)$  is controllable.*

*Then  $\hat{X}$  defined by (5.14) has a unique positive stationary distribution given by a normal distribution  $\mathcal{N}(\hat{\mu}, \hat{\Sigma})$  where  $\hat{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  and  $\hat{\mu} = (A_{11} + \alpha A_{21})^{-1}(A_{12} + \alpha A_{22})y$ .*

*Proof.* The Lyapunov equation for the covariance of the invariant distribution of the process  $(\hat{X})_{t \geq 0}$  given by (5.14) is

$$(A_{11} + \alpha A_{21})\hat{\Sigma} + \hat{\Sigma}(A_{11} + \alpha A_{21})^T = -2(C_1 + \alpha C_2)(C_1 + \alpha C_2)^T, \quad (5.15)$$

and we note that the right hand side is the first block of the projected diffusion, i.e.,

$$(PCC^T P^T)_{11} = (C_1 + \alpha C_2)(C_1 + \alpha C_2)^T.$$

Thus, we can make use of the first block of the Lyapunov equation (5.13) to which is multiplied by  $P$  and  $P^T$  from the left and right respectively. Using that  $\Sigma_{12}^T + \Sigma_{22}\alpha^T = 0$  for  $\alpha = -\Sigma_{12}\Sigma_{22}^{-1}$  the left hand side of this equation yields

$$\begin{aligned} &(PA\Sigma P^T + P\Sigma A^T P^T)_{11} \\ &= (A_{11} + \alpha A_{21})(\Sigma_{11} + \Sigma_{12}\alpha^T) + (A_{12} + \alpha A_{22})(\Sigma_{12}^T + \Sigma_{22}\alpha^T) \\ &\quad + (\Sigma_{11} + \alpha^T \Sigma_{12}^T)(A_{11} + \alpha A_{21})^T + (\Sigma_{12} + \alpha^T \Sigma_{22})(A_{12} + \alpha A_{22})^T \\ &= (A_{11} + \alpha A_{21})(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T) + (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T)(A_{11} + \alpha A_{21})^T \end{aligned}$$

and so  $\hat{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$  is indeed the solution to (5.15).  $\square$

### 5.2.2 A mean and variance preserving projection

In this section we derive a projection which additionally yields the correct mean. To state the next result we introduce the block wise notation for the inverse of  $A$

$$A^{-1} = \begin{pmatrix} A_{11}^- & A_{12}^- \\ A_{21}^- & A_{22}^- \end{pmatrix}.$$

**Theorem 5.5.** *Let condition A2 hold and assume that either  $A_{11}$  or  $A_{22}$  is invertible. Further assume that*

$$\alpha = - (A_{11}\Sigma_{12}\Sigma_{22}^{-1} + A_{12}) (A_{22} + A_{21}\Sigma_{12}\Sigma_{22}^{-1})^{-1}$$

*exists, that  $(A_{11} + \alpha A_{21})$  is invertible and that  $C_1 C_1^T + \alpha C_2 C_2^T \alpha^T > 0$ .*

*Then  $(A_{11} + \alpha A_{21})$  is Hurwitz and the pair  $((A_{11} + \alpha A_{21}), C_1 + \alpha C_2)$  is controllable. Further  $\bar{X}$ , the solution to (5.14), has a unique positive stationary distribution given by a normal distribution  $\mathcal{N}(\bar{\mu}, \bar{\Sigma})$  where  $\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  and  $\bar{\mu} = -\Sigma_{12}\Sigma_{22}^{-1}y$ .*

*Proof.* We start with the calculation for the mean which is the easy part of the proof.

#### The mean

Recall that the aim is to find  $\bar{\mu} = -\Sigma_{12}\Sigma_{22}^{-1}y$ . By the virtue of the equation (5.14) we know that  $\bar{\mu} = (A_{11} + \alpha A_{21})^{-1}(A_{12} + \alpha A_{22})y$ . Hence we check whether  $(A_{11} + \alpha A_{21})^{-1}(A_{12} + \alpha A_{22}) = -\Sigma_{12}\Sigma_{22}^{-1}$ . This is indeed the case by our assumptions and we compute straightforwardly:

$$\begin{aligned} (A_{11} + \alpha A_{21})^{-1}(A_{12} + \alpha A_{22}) &= -\Sigma_{12}\Sigma_{22}^{-1} & | (A_{11} + \alpha A_{21}) \cdot \\ \Leftrightarrow A_{12} + \alpha A_{22} &= - (A_{11} + \alpha A_{21}) \Sigma_{12}\Sigma_{22}^{-1} & | + \alpha A_{21}\Sigma_{12}\Sigma_{22}^{-1} - A_{12} \\ \Leftrightarrow \alpha (A_{22} + A_{21}\Sigma_{12}\Sigma_{22}^{-1}) &= - (A_{11}\Sigma_{12}\Sigma_{22}^{-1} + A_{12}) & | \cdot (A_{22} + A_{21}\Sigma_{12}\Sigma_{22}^{-1})^{-1} \\ \Leftrightarrow \alpha &= - (A_{11}\Sigma_{12}\Sigma_{22}^{-1} + A_{12}) (A_{22} + A_{21}\Sigma_{12}\Sigma_{22}^{-1})^{-1}. \end{aligned} \quad (5.16)$$

#### The covariance

The harder part is to check that this choice of  $\alpha$  also gives the correct covariance  $\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ . For this we need to check that

$$(A_{11} + \alpha A_{21})\bar{\Sigma} + \bar{\Sigma}(A_{11} + \alpha A_{21})^T = -2(C_1 + \alpha C_2)(C_1 + \alpha C_2)^T \quad (5.17)$$

such that  $\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ .

First note that in general

$$A\Sigma + \Sigma A^T = -2CC^T \Leftrightarrow A^{-1}\Sigma + \Sigma A^{-T} = -2A^{-1}CC^T A^{-T}. \quad (5.18)$$

We profit from the calculation (5.16) in the proof of Theorem 1 which we exploit here for the last equation (5.18) to which we refer as the *reciprocal Lyapunov* equation (see e.g. [49]). In this auxiliary calculation we employ the previously chosen projection

$$\tilde{P} = \begin{pmatrix} I_{n_x \times n_x} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & 0 \end{pmatrix}$$



since we know that this gives the correct covariance. This yields

$$\left( \tilde{P}A^{-1}\Sigma\tilde{P}^T + \tilde{P}\Sigma A^{-T}\tilde{P}^T \right)_{11} = -2 \left( \tilde{P}A^{-1}CC^T A^{-T}\tilde{P}^T \right)_{11} \Leftrightarrow \tilde{A}_1\hat{\Sigma} + \hat{\Sigma}\tilde{A}_1^T = -2\tilde{C}\tilde{C}^T \quad (5.19)$$

where we introduce the auxiliary variables

$$\tilde{A}_1 = (A_{11}^- - \Sigma_{12}\Sigma_{22}^{-1}A_{21}^-), \quad \tilde{C} = \tilde{A}_1C_1 + \tilde{A}_2C_2, \quad \tilde{A}_2 = (A_{12}^- - \Sigma_{12}\Sigma_{22}^{-1}A_{22}^-) \quad (5.20)$$

and we know that  $\hat{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  is the correct variance. Hence our aim is to show that equations (5.19) and (5.17) are equivalent. To this end, we transform equation (5.19) to its equivalent reciprocal version, i.e.,

$$\tilde{A}_1\hat{\Sigma} + \hat{\Sigma}\tilde{A}_1^T = -2\tilde{C}\tilde{C}^T \Leftrightarrow \tilde{A}_1^{-1}\hat{\Sigma} + \hat{\Sigma}\tilde{A}_1^{-T} = -2 \left( C_1 + \tilde{A}_1^{-1}\tilde{A}_2C_2 \right) \left( C_1 + \tilde{A}_1^{-1}\tilde{A}_2C_2 \right)^T$$

and hence if we can show that

- (i)  $\tilde{A}_1^{-1} = A_{11} + \alpha A_{21}$
- (ii)  $\tilde{A}_1^{-1}\tilde{A}_2 = \alpha$

we know that  $\hat{\Sigma}$  is a solution to (5.17).

**Version 1: Assume that  $A_{22}$  is invertible**

First note, that if  $A_{22}$  is invertible and  $A$  is stable it follows that  $T = A_{11} - A_{12}A_{22}^{-1}A_{21}$  is also invertible. This follows from noting that

$$0 \neq \det \left( \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \right) = \det \left( \begin{pmatrix} T & A_{12}A_{22}^{-1} \\ 0_{n_y \times n_x} & I_{n_y \times n_y} \end{pmatrix} \begin{pmatrix} I_{n_x \times n_x} & 0_{n_x \times n_y} \\ A_{21} & A_{22} \end{pmatrix} \right) = \det(A_{22}) \det(T)$$

which implies  $\det(T) \neq 0$  which is equivalent to  $T$  being invertible.

**Step 1: (i) Show  $\tilde{A}_1^{-1} = A_{11} + \alpha A_{21}$ .**

By block inversion of matrices we have  $A_{11}^- = T^{-1}$ ,  $A_{21}^- = -A_{22}^{-1}A_{21}T^{-1}$  such that  $\tilde{A}$  defined in (5.20) can be rewritten as

$$\tilde{A}_1^{-1} = A_{11}^- - \Sigma_{12}\Sigma_{22}^{-1}A_{21}^- = (I_{n_x \times n_x} + \Sigma_{12}\Sigma_{22}^{-1}A_{22}^{-1}A_{21})T^{-1}. \quad (5.21)$$

Also  $\alpha$  can be rewritten as

$$\begin{aligned} \alpha &= -(A_{11}\Sigma_{12}\Sigma_{22}^{-1} + A_{12})(A_{22} + A_{21}\Sigma_{12}\Sigma_{22}^{-1})^{-1} \\ &= -(A_{11}\Sigma_{12} + A_{12}\Sigma_{22})(A_{22}\Sigma_{22} + A_{21}\Sigma_{12})^{-1}. \end{aligned}$$

With this we calculate - using the definition of  $\tilde{A}$  by (5.21) and  $T$  in the first step, rearranging terms in the second and third step, employing the definition of  $\alpha$  in the fourth and finally rearranging terms in the last step - we find

$$\begin{aligned}
\tilde{A}_1^{-1} &= A_{11} + \alpha A_{21} \\
&\Leftrightarrow (A_{11} - A_{12}A_{22}^{-1}A_{21}) (I_{n_x \times n_x} + \Sigma_{12}\Sigma_{22}^{-1}A_{22}^{-1}A_{21})^{-1} = A_{11} + \alpha A_{21} \\
&\Leftrightarrow A_{11} - A_{12}A_{22}^{-1}A_{21} = (A_{11} + \alpha A_{21}) (I_{n_x \times n_x} + \Sigma_{12}\Sigma_{22}^{-1}A_{22}^{-1}A_{21}) \\
&\Leftrightarrow -(A_{11}\Sigma_{12} + A_{12}\Sigma_{22})\Sigma_{22}^{-1}A_{22}^{-1}A_{21} = \alpha A_{21} (I_{n_x \times n_x} + \Sigma_{12}\Sigma_{22}^{-1}A_{22}^{-1}A_{21}) \\
&\Leftrightarrow \Sigma_{22}^{-1}A_{22}^{-1}A_{21} = (A_{22}\Sigma_{22} + A_{21}\Sigma_{12})^{-1}A_{21} (I_{n_x \times n_x} + \Sigma_{12}\Sigma_{22}^{-1}A_{22}^{-1}A_{21}) \\
&\Leftrightarrow A_{21} + A_{21}\Sigma_{12}\Sigma_{22}^{-1}A_{22}^{-1}A_{21} = A_{21} (I_{n_x \times n_x} + \Sigma_{12}\Sigma_{22}^{-1}A_{22}^{-1}A_{21}).
\end{aligned}$$

**Step 2: (ii) Show**  $\tilde{A}_1^{-1}\tilde{A}_2 = \alpha$ .

Again - using block inversion of matrices - we rewrite  $\tilde{A}_2$  defined in (5.20) as

$$\begin{aligned}
\tilde{A}_2 &= A_{12}^{-} - \Sigma_{12}\Sigma_{22}^{-1}A_{22}^{-} \\
&= -(T^{-1}A_{12} - \Sigma_{12}\Sigma_{22}^{-1}(I_{n_y \times n_y} + A_{22}^{-1}A_{21}T^{-1}A_{12}))A_{22}^{-1}. \quad (5.22)
\end{aligned}$$

Using (5.22) and (5.21) a straightforward calculation gives

$$\begin{aligned}
\tilde{A}_1^{-1}\tilde{A}_2 &= \alpha \\
&\Leftrightarrow -\tilde{A}_1^{-1}(T^{-1}A_{12} + \Sigma_{12}\Sigma_{22}^{-1}(I_{n_y \times n_y} + A_{22}^{-1}A_{21}T^{-1}A_{12})) = \alpha A_{22} \\
&\Leftrightarrow -\tilde{A}_1^{-1}(I_{n_y \times n_y} + A_{22}^{-1}A_{21})T^{-1}A_{12} - \tilde{A}_1^{-1}\Sigma_{12}\Sigma_{22}^{-1} = \alpha A_{22} \\
&\Leftrightarrow -A_{12} - (A_{11} + \alpha A_{21})\Sigma_{12}\Sigma_{22}^{-1} = \alpha A_{22} \\
&\Leftrightarrow -A_{12}\Sigma_{22} - A_{11}\Sigma_{12} = \alpha(A_{21}\Sigma_{12} + A_{22}\Sigma_{22})
\end{aligned}$$

which is true by the definition of  $\alpha$ .

**Version 2: Assume that  $A_{11}$  is invertible**

This proof is similar to version 1 given above. It mainly differs in replacing the Schur complement  $T = A_{11} - A_{12}A_{22}^{-1}A_{21}$  by the other Schur complement  $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$  since here we assume that  $A_{11}$  is invertible and modifying the computations appropriately.

First note that since  $A_{11}$  is invertible and  $A$  is stable it follows that  $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$  exists and is invertible. This follows from

$$0 \neq \det \left( \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \right) = \det \left( \begin{pmatrix} A_{11} & 0_{n_x \times n_y} \\ A_{21} & I_{n_y \times n_y} \end{pmatrix} \begin{pmatrix} I_{n_x \times n_x} & A_{11}^{-1}A_{12} \\ 0_{n_x \times n_x} & S \end{pmatrix} \right) = \det(A_{11}) \det(S)$$

which means again that  $\det(S) \neq 0$  such that  $S$  is invertible.

**Step 1: (i) Show**  $\tilde{A}_1^{-1} = A_{11} + \alpha A_{21}$ .

Let us rewrite  $\tilde{A}_1$  with the Schur complement  $S$  which gives

$$\tilde{A}_1 = A_{11}^- - \Sigma_{12} \Sigma_{22}^{-1} A_{21}^- = (I_{n_x \times n_x} + A_{11}^{-1} A_{12} S^{-1} A_{21} + \Sigma_{11} \Sigma_{22}^{-1} S^{-1} A_{21}) A_{11}^{-1} \quad (5.23)$$

Multiplying the fourth block of the Lyapunov equation (5.13) from the right by  $\Sigma_{22}^{-1}$  we have

$$A_{21} \Sigma_{12} \Sigma_{22}^{-1} = -A_{22} - (\Sigma_{12}^T A_{21}^T + \Sigma_{22} A_{22}^T + C_2 C_2^T) \Sigma_{22}^{-1}$$

and adding  $A_{21} A_{11}^{-1} A_{12}$  we find

$$\begin{aligned} A_{21} (A_{11}^{-1} A_{12} + \Sigma_{12} \Sigma_{22}^{-1}) &= -S - (\Sigma_{12}^T A_{21}^T + \Sigma_{22} A_{22}^T + C_2 C_2^T) \Sigma_{22}^{-1} \\ &= -S - (A_{21} \Sigma_{12} \Sigma_{22}^{-1} + A_{22}). \end{aligned} \quad (5.24)$$

With this - using (5.23) in the first step, rearranging terms in the second step, inserting (5.24) in the third and applying the definition of  $\alpha$  in the last step - we calculate

$$\begin{aligned} \tilde{A}_1^{-1} &= A_{11} + \alpha A_{21} \\ \Leftrightarrow A_{11} &= (A_{11} + \alpha A_{21}) [A_{11}^{-1} A_{12} + \Sigma_{12} \Sigma_{22}^{-1}] S^{-1} A_{21} + A_{11} + \alpha A_{21} \\ \Leftrightarrow 0 &= \alpha (A_{21} A_{11}^{-1} A_{12} + A_{21} \Sigma_{12} \Sigma_{22}^{-1} + S) S^{-1} A_{21} + (A_{12} + A_{11} \Sigma_{12} \Sigma_{22}^{-1}) S^{-1} A_{21} \\ \Leftrightarrow 0 &= \alpha (A_{22} + A_{21} \Sigma_{12} \Sigma_{22}^{-1}) S^{-1} A_{21} + (A_{12} + A_{11} \Sigma_{12} \Sigma_{22}^{-1}) S^{-1} A_{21} \\ \Leftrightarrow 0 &= -(A_{12} + A_{11} \Sigma_{12} \Sigma_{22}^{-1}) S^{-1} A_{21} + (A_{12} + A_{11} \Sigma_{12} \Sigma_{22}^{-1}) S^{-1} A_{21}. \end{aligned}$$

**Step 2: (ii) Show**  $\tilde{A}_1^{-1} \tilde{A}_2 = \alpha$ .

Rewriting  $\tilde{A}_2$  in terms of  $S$  gives

$$\tilde{A}_2 = A_{12}^- - \Sigma_{12} \Sigma_{22}^{-1} A_{22}^- = - (A_{11}^{-1} A_{12} - \Sigma_{12} \Sigma_{22}^{-1}) S^{-1}$$

which together with the previous step yields

$$\begin{aligned} \tilde{A}_1^{-1} \tilde{A}_2 &= \alpha \\ \Leftrightarrow (A_{11} + \alpha A_{21}) (-A_{11}^{-1} A_{12} - \Sigma_{12} \Sigma_{22}^{-1}) &= \alpha S \\ \Leftrightarrow -A_{12} - \alpha A_{21} A_{11}^{-1} A_{12} - A_{11} \Sigma_{12} \Sigma_{22}^{-1} - \alpha A_{21} \Sigma_{12} \Sigma_{22}^{-1} &= \alpha (A_{22} - A_{21} A_{11}^{-1} A_{12}) \\ \Leftrightarrow -A_{12} - A_{11} \Sigma_{12} \Sigma_{22}^{-1} - \alpha A_{21} \Sigma_{12} \Sigma_{22}^{-1} &= \alpha A_{22} \\ \Leftrightarrow -A_{12} \Sigma_{22} - A_{11} \Sigma_{12} &= \alpha (A_{21} \Sigma_{12} + A_{22} \Sigma_{22}), \end{aligned}$$

which is true by the definition of  $\alpha$ .

Now, since  $\hat{\Sigma} = \tilde{\Sigma} > 0$  and by assumption  $C_1 C_1^T + \alpha C_2 C_2^T \alpha^T > 0$  it follows by Theorem 6.17 in [3] that  $A_{11} + \alpha A_{21}$  is stable.

Hence  $(A_{11} + \alpha A_{21}, C_1 C_1^T + \alpha C_2 C_2^T)$  is controllable and the process  $\hat{X}_t$  has a unique invariant distribution given by  $\mathcal{N}(0, \tilde{\Sigma})$ . □

### Alternative representation for $\alpha$

Since we can always express  $A$  via

$$A = 2(-D + J)\Sigma^{-1}, \text{ where } D = CC^T, J = -J^T$$

we can also get an alternative representation of  $\alpha$  in terms of  $D$  and  $J$  alone. Using the blockwise notation for the inverse of  $\Sigma$  and the Schur complements  $S = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$  and  $T = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  we get the following representations of the blocks of  $A$ :

$$\begin{aligned} A_{11} &= (-D_{11} + J_{11})T^{-1} + (D_{12} - J_{12})\Sigma_{22}^{-1}\Sigma_{21}T^{-1} \\ A_{12} &= (D_{11} - J_{11})T^{-1}\Sigma_{12}\Sigma_{22}^{-1} + (-D_{12} + J_{12})(\Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}T^{-1}\Sigma_{12}\Sigma_{22}^{-1}) \\ A_{21} &= (-D_{21} + J_{21})(\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}S^{-1}\Sigma_{21}\Sigma_{11}^{-1}) + (D_{22} - J_{22})S^{-1}\Sigma_{21}\Sigma_{11}^{-1} \\ A_{22} &= (D_{21} - J_{21})\Sigma_{11}^{-1}\Sigma_{12}S^{-1} + (-D_{22} + J_{22})S^{-1}. \end{aligned}$$

With this we see that

$$\begin{aligned} A_{11}\Sigma_{12}\Sigma_{22}^{-1} + A_{12} &= [(-D_{11} + J_{11})T^{-1} + (D_{12} - J_{12})\Sigma_{22}^{-1}\Sigma_{21}T^{-1}] \Sigma_{12}\Sigma_{22}^{-1} \\ &\quad + (D_{11} - J_{11})T^{-1}\Sigma_{12}\Sigma_{22}^{-1} + (-D_{12} + J_{12})(\Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}T^{-1}\Sigma_{12}\Sigma_{22}^{-1}) \\ &= (-D_{12} + J_{12})\Sigma_{22}^{-1}, \end{aligned} \tag{5.25}$$

$$\begin{aligned} A_{22} + A_{21}\Sigma_{12}\Sigma_{22}^{-1} &= (D_{21} - J_{21})\Sigma_{11}^{-1}\Sigma_{12}S^{-1} + (-D_{22} + J_{22})S^{-1} \\ &\quad + [-(D_{21} - J_{21})\Sigma_{11}^{-1}\Sigma_{12} - D_{22} + J_{22})S^{-1}\Sigma_{21}\Sigma_{11}^{-1} + (-D_{21} + J_{21})\Sigma_{11}^{-1}] \Sigma_{12}\Sigma_{22}^{-1} \\ &= (D_{21} - J_{21})(\Sigma_{11}^{-1}\Sigma_{12}S^{-1} - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1} - S^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{11}^{-1}\Sigma_{12}) \\ &\quad + (-D_{22} + J_{22})S^{-1}(I_{n_y \times n_y} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}) \\ &= (D_{22} - J_{22})\Sigma_{22}^{-1}. \end{aligned} \tag{5.26}$$

In the last step we used that

$$\Sigma_{11}^{-1}\Sigma_{12}S^{-1} - S^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma_{11}^{-1}\Sigma_{12}S^{-1}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\Sigma_{22}^{-1} = \Sigma_{11}\Sigma_{12}\Sigma_{22}^{-1}$$

and

$$S^{-1}(I_{n_y \times n_y} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}) = S^{-1}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\Sigma_{22}^{-1} = \Sigma_{22}^{-1}.$$

Consequently

$$\begin{aligned} \alpha &= -(A_{11}\Sigma_{12}\Sigma_{22}^{-1} + A_{12})(A_{22} + A_{21}\Sigma_{12}\Sigma_{22}^{-1})^{-1} \\ &= -(-D_{12} + J_{12})(D_{22} - J_{22})^{-1}. \end{aligned} \tag{5.27}$$

**Remark 5.6.** We find that  $\alpha$  exists if  $D_{22} - J_{22}$  is invertible and that  $\alpha = 0$  if  $D_{12} = J_{12}$ . If additionally to  $D_{12} = J_{12}$  it holds that  $A_{11}$  is invertible then by (5.25)  $A_{11}^{-1}A_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$ . This means, in this case the canonical orthogonal projection will yield the correct conditional distribution.

If  $A = A^T$  (and hence can interpreted as the Hessian of a quadratic potential) and  $C = I_{n \times n}$  we are clearly in the case that  $\alpha = 0$  and still we arrive at the correct constrained distribution by our choice of projecting.

### 5.2.3 A scalar product with respect to which the projection becomes orthogonal

In this section we shortly discuss the introduction of an appropriately chosen scalar product with respect to which the projection

$$P = \begin{pmatrix} I_{n_x \times n_x} & \alpha \\ 0 & 0 \end{pmatrix}$$

is orthogonal. To this end, recall that a projection  $P : \mathbb{R}^n \rightarrow U \subset \mathbb{R}^n$  is orthogonal with respect to the scalar product  $(x, y)_R = x^T R y$  if

$$(Pv - v, u)_R = 0 \quad \forall v \in \mathbb{R}^n, u \in U.$$

We write  $v = \begin{pmatrix} v_x \\ v_y \end{pmatrix} \in \mathbb{R}^n$  with  $v_x \in \mathbb{R}^{n_x}$ ,  $v_y \in \mathbb{R}^{n_y}$ . Then orthogonality in our case requires that

$$\begin{pmatrix} \alpha v_y \\ -v_y \end{pmatrix}^T R \begin{pmatrix} u_x \\ 0 \end{pmatrix} = (v_y^T \alpha^T R_{11} - v_y^T R_{21}) u_x = 0 \quad \forall v_y \in \mathbb{R}^{n_y}, u_x \in \mathbb{R}^{n_x},$$

where we use the same notation for the blocks of the matrix  $R$  as in the previous section. This means we need  $R$  to be such that

$$\alpha^T R_{11} = R_{21} \Leftrightarrow \alpha = R_{11}^{-T} R_{21}^T.$$

Since  $\alpha$  is of the form

$$\alpha = -(A_{11} \Sigma_{12} \Sigma_{22}^{-1} + A_{12})(A_{22} + A_{21} \Sigma_{12} \Sigma_{22}^{-1})^{-1},$$

assuming that  $A_{11}$  is invertible, we can choose

$$\begin{aligned} R_{11} &= -A_{11}^{-T} \\ R_{21}^T &= (\Sigma_{12} \Sigma_{22}^{-1} + A_{11}^{-1} A_{12})(A_{22} + A_{21} \Sigma_{12} \Sigma_{22}^{-1})^{-1}. \end{aligned}$$

Now, in order for  $R$  to define a scalar product  $R = R^T > 0$  has to hold, which means that besides  $A_{11}$  being invertible also symmetry of  $A_{11}$  is required. Furthermore, we choose  $R_{12} = R_{21}^T$  and since the choice of  $R_{22}$  is free, it should be chosen such that  $R$  is indeed positive definite. The interpretation of this scalar product is – similar to the interpretation of the projection – not obvious.



## APPLICATIONS AND EXAMPLES

In this chapter we will probe the theoretical findings of the previous chapters, especially Chapter 3, by means of various examples and applications.

### 6.1 Convergence to equilibrium in relative entropy for OU-processes

This section comprises the practical counterpart to the theoretical considerations of Chapter 3 dealing with convergence to equilibrium in relative entropy for possibly degenerate OU-processes, see in particular Section 3.2. Theorem 3.22 constitutes the first result for exponential convergence in relative entropy for degenerate diffusions with a rate that solely depends on the drift. This is surprising, in the sense that it seems to imply that the temperature of the system does not play a role at all for convergence rate. This will be the starting point of our analysis of the convergence behaviour. A splitting into terms which contribute to the relaxation of the mean and the covariance will be suggested in order to put the numerical observations on a theoretical basis. Further, we will examine the influence of the initial condition on the convergence and therein the occurrence of plateaus. In Section 6.1.3 we investigate the splitting into marginal and conditional terms, which later on allows for a data-based division into slow and fast degrees of freedom.

The dynamics of interest for this section is described by the SDE

$$dX_t = AX_t dt + \sqrt{\beta^{-1}}C dB_t, \quad (6.1)$$

where  $A \in \mathbb{R}^{n \times n}$ ,  $C \in \mathbb{R}^{n \times m}$  are assumed to fulfil the usual Assumption A2, which guarantees a unique positive invariant measure which we denote again by  $\mu$ . Additionally, we introduce  $\beta \in \mathbb{R}^+$  which represents the inverse temperature, i.e., the strength of the noise. Recall that under Assumption A2 the unique invariant measure  $\mu$  is a mean zero Gaussian with covariance

$$\Sigma_\infty^\beta = \beta^{-1}\Sigma_\infty$$

where  $\Sigma_\infty$  is the unique solution to

$$A\Sigma_\infty + \Sigma_\infty A^T = -CC^T.$$

We use this notation to make the dependence on the temperature explicit. In the following we will study the effect of the initial conditions and the temperature on the convergence behaviour. Further we investigate in more detail the occurrence of plateaus and in later sections the possibility of accelerating the convergence.

To this end, we first compute the relative entropy between two Gaussian distributions, for which we employ the following useful lemma.

**Lemma 6.1.** *Let  $X \sim \mathcal{N}(m, \Sigma)$ ,  $m \in \mathbb{R}^n$ ,  $\Sigma \in \mathbb{R}^{n \times n}$  and let  $A \in \mathbb{R}^{n \times n}$ ,  $c \in \mathbb{R}^n$  be fixed. Then*

$$\mathbb{E}[(X - c)^T A (X - c)] = (m - c)^T A (m - c) + \text{Tr}(A \Sigma).$$

*Proof.* A straightforward calculation yields

$$\begin{aligned} \mathbb{E}[(X - c)^T A (X - c)] &= \mathbb{E}[(X - m) + (m - c)]^T A [(X - m) + (m - c)] \\ &= \mathbb{E}[(X - m)^T A (X - m)] + 2\mathbb{E}[(X - m)^T A (m - c)] \\ &\quad + (m - c)^T A (m - c) \\ &= \mathbb{E}[\text{Tr}((X - m)(X - m)^T A)] + (m - c)^T A (m - c) \\ &= \text{Tr}(\Sigma A) + (m - c)^T A (m - c) \end{aligned}$$

□

**Lemma 6.2.** *Let  $\rho_1, \rho_2$  be two Gaussian densities on  $\mathbb{R}^n$ ,*

$$\rho_i(x) = ((2\pi)^n \det(\Sigma_i))^{-1/2} e^{\frac{1}{2}(x - m_i)^T \Sigma_i^{-1} (x - m_i)}, \quad i = 1, 2.$$

*The relative entropy of  $\rho_1$  with respect to  $\rho_2$  is then given by*

$$\mathcal{H}(\rho_1 | \rho_2) = \frac{1}{2} [\text{Tr}(\Sigma_2^{-1} \Sigma_1) - \text{Tr}(\log(\Sigma_2^{-1} \Sigma_1)) + (m_1 - m_2)^T \Sigma_2^{-1} (m_1 - m_2) - n].$$

*Proof.* Denote by  $X_1$  the random variable with distribution  $\rho_1$

$$\begin{aligned} \mathcal{H}(\rho_1 | \rho_2) &= \int \rho_1(x) \log \left( \frac{\rho_1(x)}{\rho_2(x)} \right) dx \\ &= \int \rho_1(x) \log \left( \frac{e^{-0.5(x - m_1)^T \Sigma_1^{-1} (x - m_1)} \sqrt{\det \Sigma_2}}{e^{-0.5(x - m_2)^T \Sigma_2^{-1} (x - m_2)} \sqrt{\det \Sigma_1}} \right) dx \\ &= -\frac{1}{2} \mathbb{E}[(X_1 - m_1)^T \Sigma_1^{-1} (X_1 - m_1)] \\ &\quad + \frac{1}{2} \mathbb{E}[(X_1 - m_2)^T \Sigma_2^{-1} (X_1 - m_2)] + \log \left( \sqrt{\frac{\det \Sigma_2}{\det \Sigma_1}} \right) \\ &= -\frac{n}{2} + \frac{1}{2} (\text{Tr}(\Sigma_2^{-1} \Sigma_1) + (m_1 - m_2)^T \Sigma_2^{-1} (m_1 - m_2)) + \log \left( \sqrt{\frac{\det \Sigma_2}{\det \Sigma_1}} \right) \\ &= \frac{1}{2} [\text{Tr}(\Sigma_2^{-1} \Sigma_1) - \text{Tr}(\log(\Sigma_2^{-1} \Sigma_1)) + (m_1 - m_2)^T \Sigma_2^{-1} (m_1 - m_2) - n] \end{aligned}$$

where in the second to last step we used Lemma 6.2 and in the last step that by Jacobi's identity  $\log(\det(A) \det(B)) = \text{Tr}(\log(AB))$ . □



Let us use this result to investigate the convergence behaviour of the dynamics (6.1) with respect to initial conditions and temperature. To this end, we exploit the splitting into terms which cover the relaxation of the covariance and terms which cover the relaxation of the mean.

From Example 2.14, recall that for an initial value  $X_0$  with  $\mathbb{E}(X_0) = m_0$  and  $\text{Var}(X_0) = \Sigma_0$  we have that  $X_t \sim \mathcal{N}(m_t, \Sigma_t)$  with

$$\begin{aligned} m_t &= e^{At} m_0, \\ \Sigma_t &= e^{At} \Sigma_0 e^{A^T t} + \beta^{-1} \int_0^t e^{As} C C^T e^{A^T s} ds = \Sigma_\infty^\beta + e^{At} (\Sigma_0 - \Sigma_\infty^\beta) e^{A^T t}, \end{aligned}$$

The relative entropy of  $\rho_t$  with respect to  $\mu$  can thus be written in terms of these quantities as

$$\mathcal{H}(t) := \mathcal{H}(\rho_t | \mu) = \frac{1}{2} [\beta m_t^T \Sigma_\infty^{-1} m_t + \beta \text{Tr}(\Sigma_t \Sigma_\infty^{-1}) - \text{Tr}(\log(\beta \Sigma_t \Sigma_\infty^{-1})) - n]. \quad (6.2)$$

Note that we could also investigate the convergence behaviour by looking at  $\mathcal{H}(\mu | \rho_t) = \frac{1}{2} [m_t^T \Sigma_t^{-1} m_t + \beta^{-1} \text{Tr}(\Sigma_\infty \Sigma_t^{-1}) - \text{Tr}(\log(\beta^{-1} \Sigma_\infty \Sigma_t^{-1})) - n]$ , but we refrain from doing this, as the  $\mathcal{H}(\rho_t | \mu)$  is the quantity which is usually considered in the literature in this context.

### 6.1.1 Dependence on temperature and initial conditions

We know by the theory of Section 3.2 that convergence for a system of form (6.1) is exponential with rate given by the smallest real part of the eigenvalues of the drift matrix  $A$ . In particular the temperature does not show up in the theoretical results concerning the convergence behaviour. Whether the system's convergence behaviour is indeed independent of the temperature and how it can additionally be influenced by the choice of the initial conditions is the subject of this section.

To this end, we split up  $\mathcal{H}(t)$  given by (6.2) into three different terms

$$\mathcal{H}(t) = \frac{1}{2} (a(t) + b(t) + c(t))$$

for which we study the convergence behaviour in more detail. The terms are given by

$$\begin{aligned} (i) \quad a(t) &= \beta \text{Tr}(\Sigma_t \Sigma_\infty^{-1}) - n = \text{Tr}(e^{At} (\beta \Sigma_0 - \Sigma_\infty) e^{A^T t} \Sigma_\infty^{-1}) \\ &= \text{Tr}((\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}} e^{A^T t} \Sigma_\infty^{-1} e^{At} (\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}}) \\ (ii) \quad b(t) &= -\text{Tr}(\log(\beta \Sigma_t \Sigma_\infty^{-1})) \\ &= -\text{Tr}(\log(I_{n \times n} + (\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}} e^{A^T t} \Sigma_\infty^{-1} e^{At} (\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}})) \\ (iii) \quad c(t) &= \beta \mu_t^T \Sigma_\infty^{-1} \mu_t = \beta x_0^T e^{A^T t} \Sigma_\infty^{-1} e^{At} x_0 \end{aligned}$$

with  $a, b$  capturing the convergence of the covariance, whereas  $c$  describes the convergence of the mean. Clearly, we find the same structure in all of the terms  $a, b$  and  $c$  which is

$$z^T e^{A^T t} \Sigma_\infty^{-1} e^{At} z. \quad (6.3)$$

For the covariance terms  $a(t)$  and  $b(t)$ , we find that  $z$  is given by the differ-

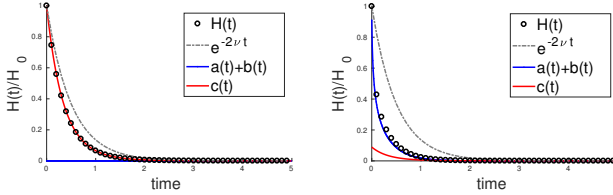


Figure 6.1: Temperature dependency of  $\mathcal{H}(t)$  for the dynamics given by  $A$  and  $C$  for the initial condition  $x_0 = (-7.5, -1)^T$ ,  $\Sigma_0 = 0$ . Left: low temperature regime,  $\beta = 10^3$ ,  $c(t)$  dominates. Right: high temperature regime,  $\beta = 10^{-2}$ ,  $a + b$  dominate.

ence of the initial condition for the covariance to the equilibrium covariance, i.e.,  $z = (\beta\Sigma_0 - \Sigma_\infty)^{\frac{1}{2}}$ . Similarly, for  $c(t)$  representing the mean, we have  $z = x_0$  which is the difference of the initial condition for the mean ( $x_0$ ) to the mean in equilibrium which is 0. This means (6.3) and its structure is the object of interest.

In general, we can say that if the temperature is low, i.e.  $\beta$  is large, the term  $c(t)$  contributes most to the relative entropy. For  $a(t)$  and  $b(t)$  there is no dependence on the temperature if  $\Sigma_0 = 0$ , i.e., deterministic initial conditions. Otherwise, both the terms  $a(t)$  and  $b(t)$  can have the largest contribution to the relative entropy, depending on how  $\Sigma_0$  is chosen.

Let us now discuss the possible scenarios more closely:

- If  $m_0 = x_0 = 0$  then  $c(t) \equiv 0$ .
- If  $\Sigma_0 = \Sigma_\infty^\beta$  then  $a(t) \equiv b(t) = 0$ .
- If  $\Sigma_0 < \Sigma_\infty^\beta$  then  $a(t) < 0$  and  $b(t) > 0$ , else if  $\Sigma_0 > \Sigma_\infty^\beta$  then  $a(t) < 0$  and  $b(t) > 0$ .
- If  $x_0 \neq 0$  and  $\Sigma_0 \neq \Sigma_\infty^\beta$  then the dominance of the single terms and the overall convergence behaviour depends on  $\beta$  and the initial conditions  $x_0, \Sigma_0$ .

Let us discuss the last case in more detail and demonstrate the expected behaviour exemplarily for the dynamics determined by the drift and diffusion matrices

$$A = -\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In the low temperature regime, i.e., when  $\beta$  is large, the convergence of the mean given by  $c(t)$  dominates, see the left panel of Figure 6.1. Contrary, in the high temperature regime, i.e., when  $\beta$  is small, the covariance terms  $a + b$  contribute most, see the right panel of Figure 6.1.

If we fix  $\beta$  and vary only the initial conditions we find that this has a great impact

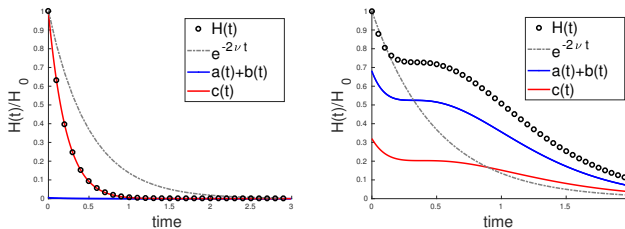


Figure 6.2: Dependence of  $\mathcal{H}(t)$  on the initial conditions for the dynamics given by  $A$  and  $C$  with  $\beta = 20$ . Left:  $x_0 = (3, 1)^T$ ,  $\Sigma_0 = 0$ ,  $x_0$  is eigenvector of  $A$  to the eigenvalue 2,  $\nu = 1$ . Right:  $x_0 \sim \mathcal{N}((0.6638, -0.6325)^T, \Sigma_0)$ ,  $\Sigma_0 = \begin{pmatrix} 1.5700 & -1.1353 \\ -1.1353 & 0.8573 \end{pmatrix}$ .

on the convergence as can be seen in Figure 6.2. In particular, if  $m_0$  is chosen as an eigenvector of  $A$  the observed convergence of  $c(t)$  to 0 is exponential with rate given by two times the corresponding eigenvalue, see the left panel of Figure 6.2. If other initial conditions are chosen it is possible to observe a plateau in the decay which can also lead to the prefactor in Theorem 3.22 being strictly greater than 1, compare right panel of Figure 6.2.

**Remark 6.3.** We should remark that our analysis is restrictive in the way that the invariant measure changes with  $\beta$ , that is  $\Sigma_\infty^\beta = \frac{1}{\beta}\Sigma_\infty$  with  $\Sigma_\infty$  being the covariance in equilibrium for  $\beta = 1$ . But since by Theorem 3.22 the convergence is only limited by the drift  $A$ , the aim here is to see whether the convergence behaviour is indeed independent of the temperature. We find that the temperature plays an important role in which terms contribute most, e.g., that  $a + b$  dominate for larger temperature. At the same time  $a + b$  admit faster convergence than  $c$  such that  $H(t)$  decreases faster with increasing temperature. Hence, we might argue that larger temperatures are favourable for convergence properties. If, contrary to our previous considerations, we constrained ourselves to keeping  $\Sigma_\infty$  fixed while playing with  $\beta$  we would have to rescale  $A$  by  $1/\beta$ . Indeed: for decreasing  $\beta$ , the eigenvalues of  $A/\beta$  increase, but doing this is nothing else than rescaling the time, and for  $\beta < 1$  this yields an acceleration.

### 6.1.2 Structural properties of $a, b$ and $c$

We aim at giving some structural insight on the properties of the terms  $a$ ,  $b$  and  $c$ . Since we have exponential convergence of  $\mathcal{H}(t)$ , we could also expect this to hold true for  $a$ ,  $b$  and  $c$ . Trying to derive this characteristic, we compute the corresponding

time derivatives, which yields the following result.

$$\begin{aligned}
c(t) &= m_0^\top e^{A^\top t} \Sigma_\infty^{-1} e^{At} x_0 \\
\Rightarrow \dot{c}(t) &= x_0^\top e^{A^\top t} (A^\top \Sigma_\infty^{-1} + \Sigma_\infty^{-1} A) e^{At} x_0 \\
a(t) &= \text{Tr}((\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}} e^{A^\top t} \Sigma_\infty^{-1} e^{At} (\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}}) = \beta \text{Tr}(\Sigma_t \Sigma_\infty^{-1}) - n \\
\Rightarrow \dot{a}(t) &= \text{Tr}((\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}} e^{A^\top t} (A^\top \Sigma_\infty^{-1} + \Sigma_\infty^{-1} A) e^{At} S^\top (\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}}) = \text{Tr}(\beta \dot{\Sigma}_t \Sigma_\infty^{-1}) \\
&= \text{Tr}((A^\top e^{A^\top t} (\beta \Sigma_0 - \Sigma_\infty) e^{At} + e^{A^\top t} (\beta \Sigma_0 - \Sigma_\infty) e^{At} A) \Sigma_\infty^{-1}) \\
&= 2 \text{Tr}((\beta \Sigma_t - \Sigma_\infty) A \Sigma_\infty^{-1}) = 2 \text{Tr}(A(-I_{n \times n} + \beta \Sigma_\infty^{-1} \Sigma_t)) \\
b(t) &= -\text{Tr}(\log(I_{n \times n} + (\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}} e^{A^\top t} \Sigma_\infty^{-1} e^{At} (\beta \Sigma_0 - \Sigma_\infty)^{\frac{1}{2}})) \\
&= \log(\det(\beta^{-1} \Sigma_\infty) \det(\Sigma_t^{-1})) \\
&= \log(\det(\beta^{-1} \Sigma_\infty)) - \log(\det(\Sigma_t)) \\
\Rightarrow \dot{b}(t) &= -\text{Tr}(\dot{\Sigma}_t \Sigma_t^{-1}) \\
&= -\text{Tr}((A^\top e^{A^\top t} (\beta \Sigma_0 - \Sigma_\infty) e^{At} + e^{A^\top t} (\beta \Sigma_0 - \Sigma_\infty) e^{At} A) \Sigma_t^{-1}) \\
&= -\text{Tr}((A^\top (\Sigma_t - \beta^{-1} \Sigma_\infty) + (\Sigma_t - \beta^{-1} \Sigma_\infty) A) \Sigma_t^{-1}) \\
&= -2(\text{Tr}(A^\top (I_{n \times n} - \beta^{-1} \Sigma_\infty \Sigma_t^{-1})))
\end{aligned}$$

Unfortunately, these calculations do not yield any insight on the behaviour of the terms. Only in the special case when  $A = A^\top$  it is possible to derive some estimates. Instead of pursuing this route for some special cases, we rather shed light on the general structure given by the expression (6.3).

### Studying the behaviour of $c(t)$ and a geometric interpretation of plateaus

We perform our investigations on the term  $c(t)$  but as explained above the arguments carry over to the other terms, since the structure is the same for all of them.

The simplest case is when  $A$  is diagonalizable, i.e., when there exists an invertible matrix  $S$  such that  $S^{-1}AS = -\Lambda$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ . We will assume that this is the case for the rest of this analysis. We can perform a basis transformation  $\tilde{X} = S^{-1}X$  which yields

$$d\tilde{X} = -\Lambda \tilde{X} dt + \sqrt{\beta^{-1}} \tilde{C} dB_t, \quad \text{where } \tilde{C} = S^{-1}C.$$

The Lyapunov equation transforms into

$$\Lambda \tilde{\Sigma}_\infty + \tilde{\Sigma}_\infty \Lambda = \tilde{C} \tilde{C}^\top, \quad \text{where } \tilde{\Sigma}_\infty = S^{-1} \Sigma_\infty S^{-\top}.$$

It is important to note that  $\Sigma_\infty$  does not undergo a similarity transformation when applying the basis transformation to the SDE. Still, the representation of  $c(t)$  in the eigenbasis of  $A$  reads

$$\begin{aligned}
\tilde{c}(t) &= \tilde{x}_0^\top e^{-\Lambda t} \tilde{\Sigma}_\infty^{-1} e^{-\Lambda t} \tilde{x}_0 \\
&= x_0^\top S^{-\top} S^\top e^{A^\top t} S^{-\top} S^\top \Sigma_\infty^{-1} S S^{-1} e^{At} S S^{-1} x_0 = c(t),
\end{aligned}$$

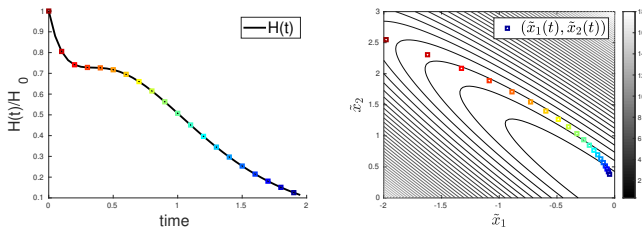


Figure 6.3: Left: Evolution of  $\mathcal{H}(t)$  over time for the same initial conditions as in the right panel of figure 6.2 with a coloured circle being drawn after a fixed time step of 0.005 being marked. Right: Contour lines of the potential  $p(x_1, x_2)$  in grey with  $\tilde{x}(t)$  plotted according to the time step colouring of the right panel.

i.e. the term  $c(t)$  does not change when changing the basis. Writing  $c(t)$  in the eigenbasis of  $A$  allows for better understanding and hence we stick to the representation

$$c(t) = \tilde{x}_0^T e^{-At} \tilde{\Sigma}_\infty^{-1} e^{-At} \tilde{x}_0 = \langle \tilde{x}_0^T e^{-At}, e^{-At} \tilde{x}_0 \rangle_{\tilde{\Sigma}_\infty^{-1}}.$$

In this notation we can see that each component  $(\tilde{x}_0)_i$ ,  $1 \leq i \leq n$  decays exponentially to 0 with the respective rate given by  $\lambda_i$  but is weighted according to  $\tilde{\Sigma}_\infty^{-1}$ . One could think of this as a process given by

$$\dot{\tilde{x}}_i = -\lambda_i \tilde{x}_i, \quad \tilde{x}_i(0) = (\tilde{x}_0)_i \quad (6.4)$$

for each component  $i = 1, \dots, n$ . This describes the evolution of the expected value  $m_i$  in each component in the eigenbasis of  $A$ . Even though the evolution is an exponential decay in each component, due to the movement within the potential induced by  $\Sigma_\infty^{-1}$ , the term  $c(t)$  can intermittently become constant if it moves on contour lines of the potential as we will discover in the following.

We make the computation for the two dimensional case first. We denote the two dimensions of the process (6.4) by  $(\tilde{x}_1(t), \tilde{x}_2(t))$ , the entries of the inducing potential and the potential itself by

$$\sigma_{ij} = \left( \tilde{\Sigma}_\infty^{-1} \right)_{ij}, \quad p(x, y) = (x, y) \tilde{\Sigma}_\infty^{-1} (x, y)^T, \quad (6.5)$$

respectively such that we can write

$$c(t) = p(\tilde{x}_1(t), \tilde{x}_2(t)).$$

### A necessary and sufficient condition to find a plateau

In Section 3.1.2 we have learned that for  $\mathcal{H}(t) = \mathcal{H}(\rho_t | \mu)$  we have

$$\dot{\mathcal{H}}(t) = -\mathcal{R}(t) = - \int \left| \nabla \ln \frac{\rho_t}{\mu} \right|_{CC^T}^2 d\rho_t.$$

If  $C$  has full rank, then  $\dot{\mathcal{H}}(t) < 0$ , else it is possible that  $\dot{\mathcal{H}}(t)$  becomes zero whenever  $\nabla \ln \frac{c}{\mu}$  is in the kernel of  $C^T$ . Thus,  $\dot{\mathcal{H}}(t) = 0$  is a necessary and sufficient condition for a plateau. Restricting ourselves to the term  $c$ , we compute the time derivative of  $c(t)$  which yields

$$\dot{c}(t) = \dot{p}(\tilde{x}_1(t), \tilde{x}_2(t)) = (\dot{\tilde{x}}_1(t), \dot{\tilde{x}}_2(t)) \cdot \nabla p(\tilde{x}(t), \tilde{y}(t)).$$

This implies that the observation of a plateau is equivalent to the trajectory moving perpendicular to the gradient of  $p$  or in other words moving on a contour line of the potential  $p$  at these times. This can be verified in Figure 6.3 in which the orange colouring correspond to  $c(t) \approx \text{const}$  and  $\tilde{x}(t)$  moving on a contour line of  $p$ . More explicitly, we compute, using the representation (6.5)

$$\dot{c}(t) = -2\lambda_1\sigma_{11}\tilde{x}_1(t)^2 - 2\lambda_2\sigma_{22}\tilde{x}_2(t)^2 - 2\sigma_{12}(\lambda_1 + \lambda_2)\tilde{x}_1(t)\tilde{x}_2(t)$$

such that for fixed  $t$  and  $\dot{c}(t) = 0$  we compute  $\tilde{x}_2(t)$  as the solution to

$$\begin{aligned} & \sigma_{11}\lambda_1\tilde{x}_1(t)^2 + \sigma_{12}(\lambda_1 + \lambda_2)\tilde{x}_1(t)\tilde{x}_2(t) + \sigma_{22}\lambda_2\tilde{x}_2(t)^2 = 0 \\ \Rightarrow \tilde{x}_2(t) &= -\frac{1}{2} \frac{\sigma_{12}(\lambda_1 + \lambda_2)}{\lambda_2\sigma_{22}} \tilde{x}_1(t) \pm \tilde{x}_1(t) \sqrt{\frac{1}{4} \left( \frac{\sigma_{12}(\lambda_1 + \lambda_2)}{\lambda_2\sigma_{22}} \right)^2 - \frac{\sigma_{11}\lambda_1}{\sigma_{22}\lambda_2}}. \end{aligned}$$

Looking for real initial conditions which result in  $\dot{c}(t) = 0$  this is only possible if  $\sigma_{12} \neq 0$ , i.e.,  $A$  and  $\Sigma_\infty$  must not be simultaneously diagonalizable. Additionally, we need to fulfil the following condition

$$\frac{1}{4} \left( \frac{\sigma_{12}(\lambda_1 + \lambda_2)}{\lambda_2\sigma_{22}} \right)^2 - \frac{\sigma_{11}\lambda_1}{\sigma_{22}\lambda_2} \geq 0 \Leftrightarrow \frac{(\lambda_1 + \lambda_2)^2}{\lambda_1\lambda_2} \geq 4 \frac{\sigma_{11}\sigma_{22}}{\sigma_{12}^2}.$$

At the same time we know that

$$A = -\frac{1}{2}(CC^T + J)\Sigma_\infty^{-1} \Leftrightarrow \Lambda = \frac{1}{2}(\tilde{C}\tilde{C}^T + \tilde{J})\tilde{\Sigma}_\infty^{-1}. \quad (6.6)$$

Solving for the entries of  $\tilde{C}\tilde{C}^T$  and  $\tilde{J}$  for given  $A, \Sigma_\infty$ <sup>3</sup> we find that the positive semi-definiteness of  $CC^T$  in the two-dimensional case, expressed by  $\Lambda, \tilde{\Sigma}_\infty^{-1}$ , reads

$$CC^T \geq 0 \Leftrightarrow \frac{(\lambda_1 + \lambda_2)^2}{\lambda_1\lambda_2} \leq 4 \frac{\sigma_{11}\sigma_{22}}{\sigma_{12}^2}.$$

Thus, only  $rk(CC^T) < 2$ , i.e.,  $\frac{(\lambda_1 + \lambda_2)^2}{\lambda_1\lambda_2} = 4 \frac{\sigma_{11}\sigma_{22}}{\sigma_{12}^2}$ , enables us to find real initial conditions such that  $\dot{c}(t) = 0$ .

<sup>3</sup>the computation can be found in the appendix Lemma B.1

**Remark 6.4.** Asking for  $\dot{c}(t) = 0$  in the  $n$ -dimensional case we can express  $x_1(t)$  by  $x_2(t), \dots, x_n(t)$  accordingly:

$$x_1(t) = -\frac{1}{2\lambda_1\sigma_{11}} \sum_{i=2}^n (\lambda_1 + \lambda_i) \sigma_{1i} x_i(t) \\ \pm \sqrt{\left( \frac{1}{2\lambda_1\sigma_{11}} \sum_{i=2}^n (\lambda_1 + \lambda_i) \sigma_{1i} x_i(t) \right)^2 - \frac{1}{\lambda_1\sigma_{11}} \sum_{i=2}^n \lambda_i \sigma_{ii} x_i^2(t)}.$$

**Remark 6.5.** If  $A$  is not diagonalizable but only admits a Jordan normal form, i.e.,  $\exists S : S^{-1}AS = -\Lambda + N$ , with  $N \in \mathbb{R}^{n \times n}$  being nilpotent. Exemplarily pick one Jordan block  $\Lambda^m + N^m$  where  $m$  refers to the size of the Jordan block, i.e., there is one eigenvector corresponding to the eigenvalue  $\lambda$  and  $m - 1$  corresponding generalized eigenvectors. Then  $(\tilde{x}_k, \dots, \tilde{x}_{k+1})e^{\Lambda^m t + N^m t} = (\tilde{x}_k e^{\lambda t}, \tilde{x}_k t e^{\lambda t} + \tilde{x}_{k+1} e^{\lambda t}, \dots, \sum_{i=1}^m \tilde{x}_{k+i} t^{m-i} e^{\lambda t})$ . This means, convergence does not take place exponentially in each component but depending on the order of the generalized eigenvector the corresponding component is slowed down by a multiplication with a polynomial in  $t$  of that order.

**Remark 6.6.** As long as  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  is such that  $\lambda_i \in \mathbb{R} \forall i = 1, \dots, n$  the above geometric interpretation is valid, even if  $A$  admits only a Jordan normal form (see the previous remark). However, if there is a complex eigenvalue, then even though  $\tilde{\Sigma}_\infty$  remains in  $\mathbb{R}^{n \times n}$ ,  $\tilde{x}_i$  will be complex. The interpretation of the movement of a complex-valued process in a potential is not clear to me.

### 6.1.3 Convergence of marginal and conditional distributions

In the previous sections a splitting into terms describing the relaxation of the mean and covariance respectively was considered and the contribution of each term was studied. In this section a different splitting, namely the one into conditional and marginal terms, will be analysed. To this end, we consider a process  $(Z_t)_{(t \geq 0)} \in \mathbb{R}^n$  which consists of two components  $Z = (X, Y)$ , where  $X \in \mathbb{R}^{n_x}$ ,  $Y \in \mathbb{R}^{n_y}$ ,  $n = n_x + n_y$ . We will later think of  $X$  as being the slow component and  $Y$  the fast one. Denote by  $\rho(z)$  the density of the joint process, by  $\hat{\rho}(x)$  the marginal density of  $X$  and by  $\rho_x$  the conditional density of  $Y$  where  $X = x$  is given. The following computation yields a partition of the relative entropy into conditional and marginal terms:

$$\begin{aligned} \mathcal{H}_{Z(t)} &:= \mathcal{H}(\rho_t | \mu) \\ &= \int \int \hat{\rho}_t \rho_{t,x} \log \left( \frac{\hat{\rho}_t}{\hat{\mu}} \right) dy dx + \int \int \hat{\rho}_t \rho_{t,x} \log \left( \frac{\rho_{t,x}}{\mu_x} \right) dy dx \\ &= \mathcal{H}(\hat{\rho}_t | \hat{\mu}) + \int \mathcal{H}(\rho_{t,x} | \mu_x) \hat{\rho}_t dx \\ &= \mathcal{H}_X(t) + \mathbb{E}_{\hat{\rho}_t}(\mathcal{H}_{Y|X=x}(t)). \end{aligned} \tag{6.7}$$

In the example of this section we investigate the contribution of the two terms, namely the conditional and the marginal term, to the overall decay in relative entropy. Note that a splitting into a marginal term  $\mathcal{H}_Y$  and a conditional term  $\mathcal{H}_{X|Y=y}$  is possible in the same way.

Consider the example from the previous section with  $A = -\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}$ ,  $C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ , for which some exemplary decay is shown in Figure 6.4.

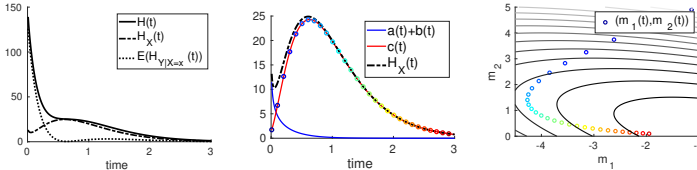


Figure 6.4: Left: Overall decay of the relative entropy  $\mathcal{H}(t)$  over time and contribution of the conditional  $\mathbb{E}(\mathcal{H}_{Y|X=x}(t))$  and marginal  $\mathcal{H}_X(t)$  respectively for the deterministic initial condition  $m_0 = (-1, 5)$ . Middle: plot of the marginal term  $\mathcal{H}_X(t)$  over time and the respective contribution of the covariance  $a(t) + b(t)$  and mean  $c(t)$  for which distinct times in the decay are marked by colored circles. Right: Evolution of the mean  $m_t = (m_1(t), m_2(t))$  over time.

### Monotonicity of $\mathcal{H}(\rho_t|\mu)$ is not inherited by the conditional and marginal terms

We provide a separate section for this subject as we believe this has not yet been discussed in the literature and also elucidates possible false interpretations of a LSI for measures (see Definitions 3.10 and 3.13). The main observation is that even though  $\mathcal{H}(t)$  is monotonically decreasing in time this is not necessarily true for either  $\mathcal{H}_X$  or  $\mathcal{H}_{Y|X=x}$  as can be seen in Figure 6.4. In fact, we witness an increase of  $\mathcal{H}_X(t)$  over a significant time interval and  $\mathbb{E}(\mathcal{H}_{Y|X=x}(t))$  also increases over some interval, but the increase has smaller magnitude. In order to explain this phenomenon, note that in the right panel of Figure 6.4 we monitor a monotonous convergence of the mean to zero with respect to the metric induced by  $\Sigma_\infty^{-1}$ . However, this does not entail monotonous convergence of, e.g., the mean of the marginal in  $X$  given by  $m_1(t)$  to zero, for which we first encounter an increase of the distance to zero. On a theoretical level this is due to the fact that when computing the time derivative of e.g.  $H(\hat{\rho}_t|\hat{\mu})$  one finds – as usual – the Fisher information, but additionally another term appears which can be estimated by the *empirical measure large deviations rate functional*. Let us shortly explain this for two families of measures  $(\nu_t)_{t \in [0, T]}, (\eta_t)_{t \in [0, T]} \in \mathcal{P}(\mathbb{R}^n)$



which are solutions to the Fokker-Planck equations

$$\begin{aligned}\frac{\partial \nu_t}{\partial t} &= \nabla \cdot (b_1 \nu_t) + \nabla^2 : (\Gamma_1 \nu_t), \quad \nu_{t=0} = \nu_0, \quad \Gamma_1 > 0 \\ \frac{\partial \eta_t}{\partial t} &= \nabla \cdot (b_2 \eta_t) + \nabla^2 : (\Gamma_2 \eta_t), \quad \eta_{t=0} = \eta_0,\end{aligned}$$

respectively. The time derivative of the relative entropy of  $\nu_t$  with respect to  $\eta_t$  can then be estimated by (see [67, Chapter 2])

$$\frac{d}{dt} \mathcal{H}(\nu_t | \eta_t) \leq -(1 - \tau) \mathcal{R}(\nu_t | \eta_t) + \frac{1}{4\tau} \frac{d}{dt} \mathcal{I}(\nu_t | \eta_t) \quad \tau > 0$$

where the *empirical measure large deviations rate functional*<sup>4</sup> reads

$$\mathcal{I}(\nu_t | \eta_t) = \int_0^T \int |b_2 - b_1| + \nabla \cdot (\Gamma_2 - \Gamma_1) + (\Gamma_2 - \Gamma_1) \nabla \nu_t |_{\Gamma_1}^2 d\nu_t dt.$$

If  $\nu_t$  and  $\eta_t$  are solutions to the same Fokker-Planck equation, i.e.  $b_1 = b_2$  and  $\Gamma_1 = \Gamma_2$  then  $\mathcal{I}(\nu_t | \eta_t) = 0$ . In fact,  $\mathcal{I}(\nu_t | \eta_t) = 0$  is equivalent to  $\nu_t$  and  $\eta_t$  solving the same Fokker-Planck equations which is also known as the *variational formulation of the Fokker-Planck equation*.

Clearly, if we are concerned with the converge to equilibrium of the time  $t$  measure  $\rho_t$  to the invariant measure  $\mu$  of the dynamics associated to the drift  $A$  and diffusion  $C$  then both  $\rho_t$  and  $\mu$  are solutions to the same Fokker-Planck equation and calculating the time derivative of  $\mathcal{H}(\rho_t | \mu)$  yields  $\frac{d}{dt} \mathcal{H}(\rho_t | \mu) = -\mathcal{R}_{CC^T}(\rho_t | \mu)$ .

On the contrary, if  $\hat{\rho}_t$  and  $\hat{\mu}$  do not solve the same Fokker-Planck equation then the empirical measure large deviations rate functional of  $\hat{\rho}_t$  and  $\hat{\mu}$  is in general non-zero. Also the time derivative of the relative entropy might experience a change of sign over time. To be more precise, we shortly repeat some of the derivations of Section 4.1. To this end, consider the process  $Z_t = (X_t, Y_t)$  described by the SDE

$$\begin{aligned}dX_t &= f_1(Z_t) dt + \sigma_1(Z_t) dW_t^1 \\ dY_t &= f_2(Z_t) dt + \sigma_2(Z_t) dW_t^2.\end{aligned}$$

Due to the work of Gyöngi [28] it is known that  $\hat{\rho}_t$ , the marginal in  $x$ , is solution to the Fokker-Planck with coefficients

$$b_1(x) = \mathbb{E}(f_1(X_t, Y_t) | X_t = x), \quad \Gamma_1(x) = \mathbb{E}(\sigma_1 \sigma_1^T(X_t, Y_t) | X_t = x),$$

that is, we compute conditional expectations with respect to the time  $t$  measure  $\rho_t$ , or, more precisely its conditional  $\rho_{t,x}(y)$ . On the other hand,  $\hat{\mu}$  is solution to the Fokker-Planck equation with coefficients

$$b_2(x) = \mathbb{E}_\mu(f_1(X, Y) | X = x), \quad \Gamma_2(x) = \mathbb{E}_\mu(\sigma_1 \sigma_1^T(X, Y) | X = x),$$

i.e., we compute expectations with respect to the conditional  $\mu_x$ . Clearly, as long as  $\rho_t \neq \mu$  the coefficients  $b_1$ ,  $b_2$  and  $\Gamma_1$ ,  $\Gamma_2$  will differ in general.

<sup>4</sup>Note that this is the same quantity appearing in Proposition 4.2.

The same reasoning holds for the evolution of the conditional  $\rho_{t,x}$  to  $\mu_x$  where not even an evolution equation for the time  $t$  measure  $\rho_{t,x}$  is available. This entails that a LSI assumption for a measure, being the conditional of the invariant measure of some process, does not imply exponential convergence to the equilibrium conditional measure with rate given by twice the LSI constant.

Note that the increase of the relative entropy in time cannot be traced back to the irreversibility of the process, but can also be observed for reversible processes with appropriate initial conditions.

We want to emphasize that in the case of the example displayed in the left panel of Figure 6.4, no time scale separation is present and thus no clear judgement is possible as to which of the terms displays a fast or slow relaxation behaviour. This is true for both splittings and will be contrasted with the case of a clear time scale separation below.

#### 6.1.4 Multiple time scales: partitioning into slow and fast degrees of freedom

In this section, we consider the same splitting into conditional and marginal terms as in the previous section with the difference that now the system admits a slow and a fast time scale and the splitting is conducted accordingly. This will lead to a significant change the of convergence behaviour of the individual contribution in the infinite time scale separation limit, which we discuss below. To this end, introduce the time scale separation parameter  $0 < \varepsilon \leq 1$  such that the coefficients now read

$$A = - \begin{pmatrix} 1 & 3\varepsilon^{-1} \\ 0 & 2\varepsilon^{-2} \end{pmatrix}, \quad C = \begin{pmatrix} 0 \\ \varepsilon^{-1} \end{pmatrix}.$$

This is an instance of a homogenization scaling (cf. [62]) and for  $\varepsilon \rightarrow 0$ , the solution of the first component of the dynamics converges weakly to the solution of an SDE with effective coefficients  $\bar{A} = 1$  and  $\bar{C} = \frac{3}{2}$ .

Let us also establish the connection to Chapter 2 and the relation between the constants corresponding to the convergence and exit time behaviour of the system, as described in Section 3.2 and 3.4.1 in Proposition 3.30, which gave

$$\lambda^* \geq \frac{\lambda_{\min}(CC^T)}{2\lambda_{\max}(\Sigma_{\infty})} |v_{\text{Im}}|,$$

where  $v_{\text{Im}}$  is the part of  $v$  which lies not in the kernel of  $CC^T$  and  $v$  being the eigenvector of  $-A^T v = \lambda^* v$ , with  $\lambda^* = 1$  here.  $v_{\text{Im}}$  is determined by the second component of  $v$  and reads  $v_{\text{Im}} = \frac{3\varepsilon}{\varepsilon^4 + 5\varepsilon^2 + 4}$ . The covariance matrix of the stationary solution to the corresponding Lyapunov equation is given by

$$\Sigma_{\infty} = \begin{pmatrix} \frac{9}{4(2+\varepsilon^2)} & -\frac{3\varepsilon}{4(2+\varepsilon^2)} \\ -\frac{3\varepsilon}{4(2+\varepsilon^2)} & \frac{1}{4} \end{pmatrix},$$

and its largest eigenvalue is  $\lambda_{\max}(\Sigma_{\infty}) = \frac{11+\varepsilon^2+\sqrt{49+22\varepsilon^2+\varepsilon^4}}{8(2+\varepsilon^2)}$ . Thus, the inequality for our example reads

$$1 \geq \frac{4(2+\varepsilon^2)}{11+\varepsilon^2+\sqrt{49+22\varepsilon^2+\varepsilon^4}} \cdot \frac{3}{\varepsilon^4+5\varepsilon^2+4}$$

and the right hand side can be shown to be monotonically increasing as  $\varepsilon \rightarrow 0$  with the well-defined limit  $2/3$  confirming the validity of the inequality.

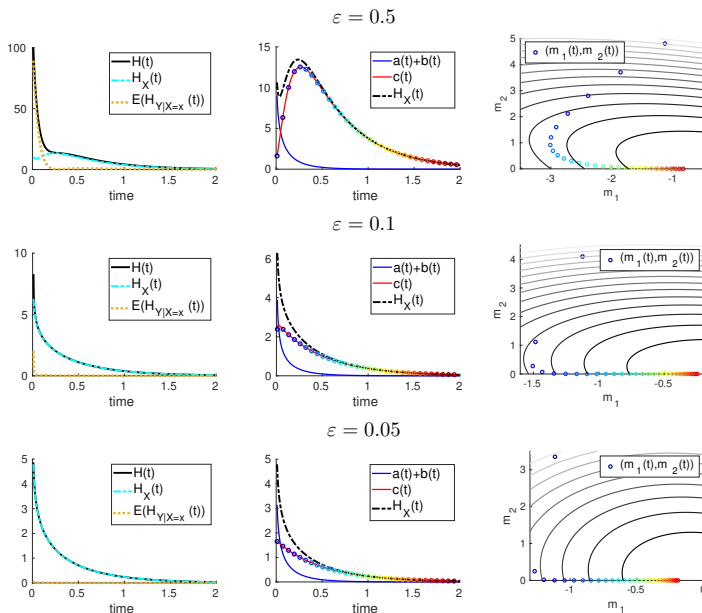


Figure 6.5: From top to bottom:  $\varepsilon$  decreases from  $\varepsilon = 0.5$  in the top,  $\varepsilon = 0.1$  in the middle to  $\varepsilon = 0.05$  in the lower row. Left: Overall decay of the relative entropy  $\mathcal{H}(t)$  over time and contribution of the conditional  $\mathbb{E}(\mathcal{H}_{Y|X=x}(t))$  and marginal  $\mathcal{H}_X(t)$  respectively for the deterministic initial condition  $m_0 = (-1, 5)$ . Middle: plot of the marginal term  $\mathcal{H}_X(t)$  over time and the respective contribution of the covariance  $a(t)+b(t)$  and mean  $c(t)$  for which distinct times in the decay are marked by coloured circles. Right: Evolution of the mean  $m_t = (m_1(t), m_2(t))$  over time.

**Time scale separation** We introduce a time scale separation by setting  $\varepsilon < 1$  and now refer to  $X$  as the slow process and  $Y$  as the fast one. The *a priori* assignment of slow and fast degrees of freedom agrees with the observation in the plots (see Figure 6.5): for  $\varepsilon \rightarrow 0$  the conditional term  $\mathcal{H}_{Y|X=x}(t)$  relaxes almost instantaneously to its

equilibrium. Accordingly, the marginal term  $\mathcal{H}_X(t)$  governs the long term behaviour of the overall relaxation. If we use the other splitting into a marginal term in  $Y$  and conditional in  $X$  for fixed  $y$  we observe the same behaviour. The marginal of  $Y$  converges very fast whereas the conditional of  $X$  dominates the long term relaxation. This observation suggests that we can use the partitioning of relative entropy into conditional and marginal terms as a definition for fast and slow degrees of freedom. Both splittings seem reasonable in our setting, but this is due to the linearity of our system. With the aim of applying this idea to non-linear diffusions we propose to use the marginal term for the slow and the conditional term for the fast variable. In order to explain this let us consider the non-linear example given by the SDE

$$dZ_t = -\nabla V(Z_t)dt + dB_t, \quad Z = (X, Y) \in \mathbb{R}^2, \\ V(x, y) = (x^2 - 1)^2 + \varepsilon^{-1} (1 + e^x)^{-1} y^2.$$

This SDE describes the diffusive motion of a particle in the potential energy landscape  $V$ . In  $x$  direction, there are two metastable states, given by the domains around the minima at  $x = \pm 1$  and in between there is a barrier to overcome. In  $y$  direction the motion is confined by a quadratic potential with varying slope. We expect that for each fixed  $x$  the conditional distribution of  $Y$  will quickly approach its equilibrium, contrary to the marginal of  $Y$  which needs the slow variable  $X$  to cross the barrier of the potential at  $x = 0$  in order to converge to its equilibrium distribution.

Furthermore, we observe that as  $\varepsilon \rightarrow 0$  all terms become monotonically decreasing which is due to the fact that the Fokker-Planck equation becomes decoupled as  $\varepsilon \rightarrow 0$  (see Section 4.3.3 for a formal derivation).

## 6.2 Accelerating the convergence to equilibrium

Having studied the convergence behaviour for OU-processes in a lot of detail in the previous section, we now come to the aspect of accelerating it. There are already various works in this direction and we will shortly review their ideas and results. Starting from some reversible dynamics the main ingredient will turn out to be the perturbation by a non-reversible Hamiltonian part effectively resulting in an equalisation of the potential energy landscape for the dynamics. Since this perturbation admits a parameter, which has to be large enough in order for the equalisation to work, we will relate this parameter to a time scale separation. This, in turn, enables us to resort to corresponding averaging or homogenization results. We examine this kind of relation in Section 6.2.2 for two-dimensional dynamics whose slow variable is the system's energy. Here, the averaged equation admits an interesting interpretation in terms of time rescaled versions of the original process. In Section 6.2.3 we employ the acceleration for a system which already has an intrinsic time scale separation and find that effectively this only leads to a time acceleration for the slow variable.

### 6.2.1 A general idea

We are still concerned with linear diffusions, that is we consider the SDE

$$dX_t = AX_t dt + C dB_t, \quad (6.8)$$

under the usual Assumptions A2 on  $A$  and  $C$ . Accelerating the convergence to equilibrium in the sense of relative entropy consists of the very basic idea which is to maximise the minimum of the absolute value of the real part of the eigenvalues of the drift  $A$  (cf. [27, 36, 37, 47])<sup>5</sup>. This is due to Theorem 3.22 which tells us that convergence is exponential with rate given by the real part of  $A$  being closest to zero, i.e.,

$$\operatorname{Re}(\lambda_{\min}(-A)).$$

There are different approaches in how to alter  $A$  and we discuss them below. For the ones which admit an interpretation as multi-scale dynamics we also establish the connection to the corresponding homogenized dynamics and the interpretation coming along with it in the subsequent sections.

Let us start with shedding some light on the possible ways of accelerating the convergence to the invariant measure  $\mu \sim \mathcal{N}(0, \Sigma_\infty)$  in the sense of relative entropy  $\mathcal{H}(\rho_t | \mu)$ . Here,  $\Sigma_\infty$  is, as usual, the unique positive definite solution to the Lyapunov equation

$$A\Sigma_\infty + \Sigma_\infty A^T = -CC^T.$$

Consider the special case of  $\operatorname{Re}(\lambda_{\max}(-A)) < 1$  and replace  $A$  by  $A^{-1}$ ,  $C$  by  $A^{-1}C$  in (6.8) which yields the minimal convergence rate

$$\operatorname{Re}(\lambda_{\min}(-A^{-1})) = (\operatorname{Re}(\lambda_{\max}(-A)))^{-1} > 1.$$

The resulting system still has the correct invariant measure since  $(A, C)$  controllable and  $A$  being Hurwitz is equivalent to  $(A^{-1}, A^{-1}C)$  controllable and  $A^{-1}$  being Hurwitz. Moreover, and more importantly,

$$A\Sigma_\infty + \Sigma_\infty A^T = -CC^T \Leftrightarrow A^{-1}\Sigma_\infty + \Sigma_\infty A^{-T} = -A^{-1}CC^T A^{-T},$$

hence the invariant measure is unchanged.

The most straightforward idea is to replace  $-A$  by  $-(A+B)$  such that  $\operatorname{Re}(\lambda_{\min}(-A)) < \operatorname{Re}(\lambda_{\min}(-(A+B)))$  and  $\Sigma_\infty$  is unchanged. Assuming that  $A$  is diagonalizable, we have  $\exists S : S^{-1}AS = \Lambda_A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ ,  $\operatorname{Re}(\lambda_i) < 0$  where w.l.o.g.  $\lambda_1, \dots, \lambda_k$  are the eigenvalues which need improvement. Hence, a matrix  $B$  is added such that  $S^{-1}BS = \Lambda_B = \operatorname{diag}(\nu_1, \dots, \nu_k, 0, \dots, 0)$ ,  $\operatorname{Re}(\nu_i) < 0$ . Clearly,  $\operatorname{Re}(\lambda_{\min}(-A)) < \operatorname{Re}(\lambda_{\min}(-(A+B)))$ , but since  $B$  and  $\Sigma_\infty$  do not commute in general, the right hand side of the Lyapunov equation, given by  $-(CC^T - B\Sigma_\infty - \Sigma_\infty B^T)$ , will not necessarily be positive semi-definite but can be negative definite or indefinite. Fortunately there are more systematic answers to this problem. We state them in the subsequent propositions.

<sup>5</sup>This is in fact also true for convergence in  $L^2_{\mu^{-1}}$ , but we will stay with relative entropy here.

**Proposition 6.7** (Proposition 4 in [47]). *Let  $\Sigma_\infty = \Sigma_\infty^T > 0 \in \mathbb{R}^{n \times n}$  be given. Then there exists  $J \in \mathbb{R}^{n \times n}$  with  $J^T = -J$  such that*

$$\operatorname{Re}(\lambda_{\min}((I_{n \times n} + J)\Sigma_\infty^{-1})) = \frac{\operatorname{Tr}(\Sigma_\infty^{-1})}{n}.$$

The interpretation of the above theorem is the following: starting from some reversible dynamics with  $-A = \Sigma_\infty^{-1}$  and  $C = \sqrt{2}I_{n \times n}$  it is possible to add a non reversible perturbation to the drift which is given by  $J\Sigma_\infty^{-1}$  with  $J^T = -J$ , which does not change the invariant measure. The work [47] is concerned with finding the optimal  $J^*$  which yields the optimal convergence rate of

$$\operatorname{Re}(\lambda_{\min}((I_{n \times n} + J^*)\Sigma_\infty^{-1})) = \operatorname{Re}(\lambda_{\max}((I_{n \times n} + J^*)\Sigma_\infty^{-1})) = \frac{\operatorname{Tr}(\Sigma_\infty^{-1})}{n}.$$

The construction uses a matrix equation similar to the Lyapunov equation which allows for an algorithmic search of  $J^*$ . Adding this non-reversible perturbation to the drift means that - instead of having some directions that are slow and some that are fast which entails  $\min \sigma(A) < \max \sigma(A)$  - all directions are evened by the skew symmetric term  $J$  which makes the system rotate on the level sets of the potential  $\frac{1}{2}x^T \Sigma_\infty^{-1} x$ . In order to truly achieve a balance between the slow and fast directions it has to be assured that the rotation is constructed according to the underlying geometry given by  $\Sigma_\infty^{-1}$  and is fast enough. We will come back to this point in Section 6.2.2. A similar idea has already been put forward by [36].

Instead of only looking for the best possible drift, it is also possible to optimise both drift *and* diffusion with respect to the convergence rate as done in [27]. To this end, recall the notation

$$dX_t = -(D + J)\Sigma_\infty^{-1}X_t dt + C dW_t, \quad \text{with } D = \frac{1}{2}CC^T, \quad J^T = -J$$

which - keeping  $\Sigma_\infty$  fixed and varying  $C$  and  $J$  - equips us with all possible choices for a linear process with invariant measure  $\mu \sim \mathcal{N}(0, \Sigma_\infty)$ .

**Proposition 6.8** (Theorem 1 in [27]). *Let  $\Sigma_\infty = \Sigma_\infty^T > 0 \in \mathbb{R}^{n \times n}$  be given. Then there exists  $J, D \in \mathbb{R}^{n \times n}$  with  $J^T = -J$ ,  $D = D^T \geq 0$  and  $\operatorname{Tr}(D) = n$  such that*

$$\operatorname{Re}(\lambda_{\max}((D + J)\Sigma_\infty^{-1})) = \operatorname{Re}(\lambda_{\max}(\Sigma_\infty^{-1})).$$

The constraint  $\operatorname{Tr}(D) = n$  with  $D = \frac{1}{2}CC^T$  is introduced in order to keep the total amount of noise injected to the system constant and comparable to the preceding approach. The idea builds up on the previous considerations which in the first step yield that

$$\max_{J: J^T = -J} \operatorname{Re}(\lambda_{\max}((D + J)\Sigma_\infty^{-1})) = \frac{\operatorname{Tr}(\Sigma_\infty^{-1}D)}{n}$$

Using that  $\Sigma_\infty$  is orthogonally diagonalizable, i.e.  $\exists Q: Q^{-1} = Q^T$  and  $Q\Sigma_\infty^{-1}Q^T = \Lambda_{\Sigma_\infty^{-1}} = \operatorname{diag}(\lambda_1(\Sigma_\infty^{-1}), \dots, \lambda_n(\Sigma_\infty^{-1}))$ , it follows that

$$\operatorname{Tr}(\Sigma_\infty^{-1}D) = \operatorname{Tr}(\Lambda_{\Sigma_\infty^{-1}}QDQ^T) \leq \operatorname{Re}(\lambda_{\max}(\Sigma_\infty^{-1}))\operatorname{Tr}(D).$$

The crucial point is to construct  $D$  such that  $\text{Tr}(\Sigma_\infty^{-1}D) = n\text{Re}(\lambda_{\max}(\Sigma_\infty^{-1}))$ . Assume that the  $k^{\text{th}}$  eigenvalue is the maximal eigenvalue, i.e.,  $\text{Re}(\lambda_k(\Sigma_\infty^{-1})) = \text{Re}(\lambda_{\max}(\Sigma_\infty^{-1}))$  and denote the associated normalised eigenvector by  $v_k$ . Let  $D = nv_kv_k^T$  which yields

$$\text{Tr}(\Sigma_\infty^{-1}D) = n\text{Tr}(\Sigma_\infty^{-1}v_kv_k^T) = n\text{Tr}(\text{Re}(\lambda_k(\Sigma_\infty^{-1}))v_kv_k^T) = n\text{Re}(\lambda_{\max}(\Sigma_\infty^{-1}))$$

since  $v_k$  being normalised implies  $\text{Tr}(v_kv_k^T) = 1$ .

This yields in fact an even larger exponential convergence rate, but due to the diffusion being highly degenerate (rank 1) also the prefactor becomes large. For details see [27].

### 6.2.2 Hamiltonian systems with energy as coarse graining map

Let us return to the idea of [47], i.e., acceleration via adding a non-reversible perturbation to the drift. We extend the idea by examining the proposed construction using the perspective of averaging. To this end consider the SDE

$$dX_t^\varepsilon = -(\nabla V(X_t^\varepsilon) + \frac{1}{\varepsilon}J\nabla V(X_t^\varepsilon)) dt + \sqrt{2\beta^{-1}} dB_t \quad (6.9)$$

$$\text{with } V(X) = \frac{1}{2}(\lambda_1 x_1^2 + \lambda_2 x_2^2), \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (6.10)$$

$X \in \mathbb{R}^2$ ,  $\lambda_1, \lambda_2 > 0$  and  $B_t$  is 2-dimensional Brownian motion. In our previous notation  $-A^\varepsilon = (I_{2 \times 2} + \frac{1}{\varepsilon}J)\Sigma_\infty^{-1}$  with  $\Sigma_\infty^{-1} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ , i.e., we have chosen the coordinates such that the reversible part of the drift is already in diagonal form. For a two dimensional example it is easy to compute the eigenvalues of the perturbed drift which are

$$\nu_{1,2} = \frac{\lambda_1 + \lambda_2}{2} \pm \sqrt{\frac{(\lambda_1 - \lambda_2)}{4} - \frac{1}{\varepsilon^2}}.$$

In order to have a pair of complex conjugate eigenvalues whose real part  $\text{Re}(\lambda_{\min}((I_{2 \times 2} + \varepsilon^{-1}J)\Sigma_\infty^{-1})) = \frac{\lambda_1 + \lambda_2}{2}$  the condition

$$\frac{(\lambda_1 - \lambda_2)}{4} < \frac{1}{\varepsilon^2} \quad (6.11)$$

has to hold true which is certainly the case for  $\varepsilon \rightarrow 0$ . At the same time, letting  $\varepsilon \rightarrow 0$ , we are in an averaging scenario.

Following the work of [63] and [26] it is known that the system's energy  $V(X_t^\varepsilon) =: Y_t^\varepsilon$  can be used to define a slow variable of the system, and the fast variable is the one moving on the level sets of  $V$ , i.e., the one described by  $J\nabla V$ . Employing homogenization theory [63] and [26] show that

$$(Y_t^\varepsilon)_{0 \leq t \leq T} \xrightarrow{\varepsilon \rightarrow 0} (Y_t)_{0 \leq t \leq T} \quad \text{weakly in } \mathcal{C}([0, T], \mathbb{R}).$$

Here,  $Y_t$  is the process which corresponds to the SDE

$$dY_t = \bar{\lambda}(\beta^{-1} - Y_t) dt + \sqrt{2\bar{\lambda}\beta^{-1}Y_t} dB_t, \quad \text{with } \bar{\lambda} = \lambda_1 + \lambda_2,$$

which is known as the square-root process (cf. e.g. [15]). For a calculation of the averaged equation see the appendix Proposition B.2. The interesting finding is the following:

**Proposition 6.9.** *The homogenized process  $Y_t$  can again be expressed as the energy of two OU-processes, i.e.,  $Y_t = V(X_t^1, X_t^2)$  with  $X_t$  being the solution to*

$$dX_t^i = -\frac{\bar{\lambda}}{2}X_t^i dt + \sqrt{\frac{\bar{\lambda}\beta^{-1}}{\lambda_i}} dB_t^i, \quad (6.12)$$

respectively. Furthermore, the invariant measures of  $(X^{1,\varepsilon}, X^{2,\varepsilon})$  and  $(X^1, X^2)$  agree and are given by  $\mu \sim \mathcal{N}\left(0, \beta^{-1} \begin{pmatrix} \lambda_1^{-1} & 0 \\ 0 & \lambda_2^{-1} \end{pmatrix}\right)$ .

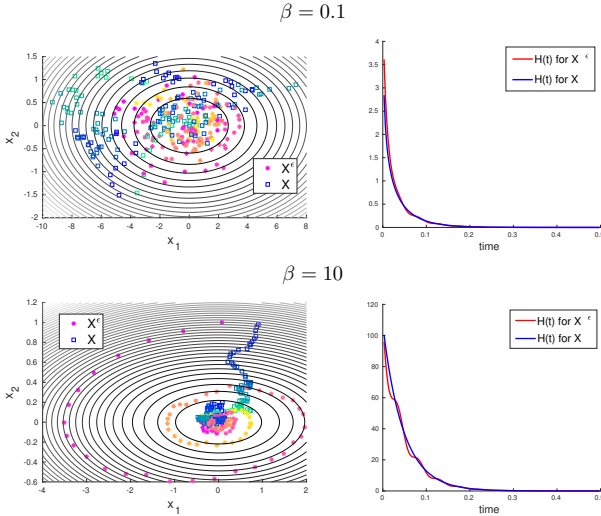


Figure 6.6: Left: plot of  $(X_t^{\varepsilon,1}, X_t^{\varepsilon,2})$  solution to (6.9) and  $(X_t^1, X_t^2)$  solution to (6.12) where the colouring corresponds to the evolution of time  $t \in (0, 0.5)$  and the contourlines correspond to the potential induced by  $\beta\Sigma_\infty^{-1}$ . Right: Plot of  $\mathcal{H}(t) = \mathcal{H}(\rho_t|\mu)$  for the dynamics  $(X_t^\varepsilon)_t$  (red) and  $(X_t)_t$  (blue). Parameter values:  $\lambda_1 = 1$ ,  $\lambda_2 = 20$ ,  $X_0 = (1, 1)^\top$ ,  $\beta = 0.1$ (upper panel)  $\beta = 10$  (lower panel)  $\varepsilon = 15$ .

Let us first discuss the interpretation offered by the proposition before providing the proof on the next page. Employing homogenization theory we find a new process  $X_t$  with the same invariant measure as the initial process  $X_t^\varepsilon$  but its drift  $A$  admits the real eigenvalue  $-\bar{\lambda}/2$ , whereas  $A^\varepsilon$  contains a pair of complex conjugate eigenvalues.



Furthermore, going from the original dynamics (6.9), which do not contain the acceleration term  $J\nabla U(X)$ , to the dynamics (6.12), which correspond to the homogenized energy, we effectively observe a *rescaling of time* by the factor  $\bar{\lambda}/2\lambda_i$  for the  $i^{\text{th}}$  variable. This means for one variable time is accelerated, whereas for the other variable time will be slowed down. The rescaling of time acts here on both drift and diffusion, opposed to the non-reversible perturbation, which only affects and balances the time scales of the drift. The numerical experiments show that both processes serve well to sample from the invariant measure, as can be seen in Figure 6.6, as they nearly converge in the same way. Still, it might be advantageous, from a numerical perspective, to choose the homogenized process (6.12), since for  $(X_t^\varepsilon)$  one has to choose  $\varepsilon$  small enough such that (6.11) is fulfilled. For the numerical example of Figure 6.6 with  $\lambda_1 = 1$ ,  $\lambda_2 = 20$  we have, that  $\varepsilon^{-1} > 9.5$  and choosing  $\varepsilon^{-1} = 10$  yields the drift  $A^\varepsilon = -\begin{pmatrix} 1 & -200 \\ 10 & 20 \end{pmatrix}$ . That is, the entries of  $A^\varepsilon$  become large, which becomes even more problematic as the time scale separation becomes larger. Contrary, for the homogenized dynamics we have  $A = -\begin{pmatrix} 10.5 & 0 \\ 0 & 10.5 \end{pmatrix}$  which admits a larger time step. Only if one of the time scales becomes very slow, e.g.  $\lambda_1$  close to 0, the diffusion coefficient can cause problems since the diffusion in this component becomes larger and larger as  $\lambda_1 \rightarrow 0$ .

*Proof of Proposition 6.9.* Let  $Y_t = V(X_t^1, X_t^2)$  with  $V$  as in (6.10). Itô's formula gives

$$\begin{aligned} dY_t &= \lambda_1 X_t^1 \left( -\frac{\bar{\lambda}}{2} X_t^1 dt + \sqrt{\frac{\bar{\lambda}\beta^{-1}}{\lambda_1}} dB_t^1 \right) + \bar{\lambda}\beta^{-1} dt + \lambda_2 X_t^2 \left( -\frac{\bar{\lambda}}{2} X_t^2 dt + \sqrt{\frac{\bar{\lambda}\beta^{-1}}{\lambda_2}} dB_t^2 \right) \\ &= \bar{\lambda}(\beta^{-1} - Y_t) dt + \sqrt{\bar{\lambda}\beta^{-1}} \left( \sqrt{\lambda_1} X_t^1 dB_t^1 + \sqrt{\lambda_2} X_t^2 dB_t^2 \right). \end{aligned}$$

Now define the process  $Z_t$  by

$$dZ_t = \int_0^t \sqrt{\lambda_1} X_s^1 dB_s^1 + \sqrt{\lambda_2} X_s^2 dB_s^2$$

which is a martingale.

Its quadratic variation is given by

$$\langle Z \rangle_t = \int_0^t \lambda_1 (X_s^1)^2 + \lambda_2 (X_s^2)^2 ds = \int_0^t 2Y_s ds.$$

Hence

$$\tilde{B}_t = \int_0^t \frac{1}{\sqrt{2Y_s}} (\sqrt{\lambda_1} X_s^1 dB_s^1 + \sqrt{\lambda_2} X_s^2 dB_s^2)$$

equips us with a new Brownian motion according to Lévy's characterization theorem.

Altogether this yields that

$$dY_t = \bar{\lambda}(\beta^{-1} - Y_t) dt + \sqrt{\bar{\lambda}\beta^{-1}} \sqrt{2Y_t} d\tilde{B}_t.$$

Furthermore, it is easy to check that the Lyapunov equations of the matrix pairs

$$A^\varepsilon = - \begin{pmatrix} \lambda_1 & \lambda_2/\varepsilon \\ -\lambda_1/\varepsilon & \lambda_2 \end{pmatrix}, \quad D^\varepsilon = 2\beta^{-1}I_{n \times n}$$

and

$$A = - \begin{pmatrix} \frac{\lambda_1 + \lambda_2}{2} & 0 \\ 0 & \frac{\lambda_1 + \lambda_2}{2} \end{pmatrix}, \quad D = \beta^{-1} \begin{pmatrix} \frac{\lambda_1 + \lambda_2}{\lambda_1} & 0 \\ 0 & \frac{\lambda_1 + \lambda_2}{\lambda_2} \end{pmatrix}$$

have the same unique solution  $\Sigma_\infty^\beta = \beta^{-1} \begin{pmatrix} 1/\lambda_1 & 0 \\ 0 & 1/\lambda_2 \end{pmatrix}$ .  $\square$

### 6.2.3 Hamiltonian systems with soft constraints and coordinate projection as coarse graining map

In the previous section we have learned that for a system described by the SDE

$$dZ_t^\varepsilon = - \left( \nabla V(Z_t^\varepsilon) + \frac{1}{\varepsilon} J \nabla V(Z_t^\varepsilon) \right) dt + dB_t$$

where  $V$  is a potential that is sufficiently fast growing at infinity,  $J$  is a skew symmetric matrix and  $B_t$  is standard Brownian motion, the energy  $V(Z)$  can be interpreted as the slow variable of the system. Furthermore, for  $\varepsilon \rightarrow 0$  we have that  $V(Z^\varepsilon) \rightarrow V(Z^0)$  weakly<sup>6</sup>, where  $V(Z^0)$  is the energy of a different process  $Z^0$ , consisting of two processes which are time rescaled versions of the original dynamics. The applied time rescaling yields in fact a balancing of the intrinsic time scales given by the eigenvalues of the drift, resulting in an accelerated convergence to the equilibrium distribution.

In this section the aim is to investigate whether the same kind of acceleration also works in the case when we have an intrinsic time scale separation in the dynamics. To be more precise, we will consider an instance of a slow fast dynamics with a time scale separation parameter  $0 < \varepsilon \ll 1$  and for  $\varepsilon \rightarrow 0$  we compare the effective dynamics of the unperturbed with the effective dynamics of the perturbed dynamics.

#### A system with intrinsic time scale separation

We introduce the process  $Z = (X, Y)$ , where  $X$  will be the slow and  $Y$  will be the fast variable, described by the SDE

$$dZ_t = -\nabla V(Z_t) dt + \sqrt{2} dB_t,$$

where we use the same notation as in the previous chapter and  $V$ , given by,

$$V(x, y) = U(x) + \frac{1}{\varepsilon^2} (y - m(x))^T A(x) (y - m(x)), \quad (6.13)$$

where  $A(x) = A(x)^T > 0$  is a positive definite symmetric matrix and  $m : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{d_y}$  determines the mean of the fast process, such that we are able to compute the

<sup>6</sup>We assume here that  $V$  is quadratic, else the limiting process also contains a variable which gives the segment of the graph the process is moving in.

effective dynamics for  $X$ . We can picture this process as  $(X, Y)$  moving slowly through the energy landscape given by  $U$  surrounded by a tube whose shape at the point  $x$  is determined by the quadratic potential  $y^T A(x)y$ . For  $\varepsilon \rightarrow 0$  the tube becomes more and more narrow.

Introduce the rescaled variable

$$\xi = \frac{Y}{\varepsilon}$$

in terms of which the dynamics translate to, using  $d\xi = \frac{1}{\varepsilon}dY$  and  $\partial_y = \frac{1}{\varepsilon}\partial_\xi$ ,

$$\begin{aligned} dX_t &= -\nabla_x V(X_t, \xi_t) dt + \sqrt{2} dB_t^x \\ d\xi_t &= -\frac{1}{\varepsilon^2} \nabla_\xi V(X_t, \xi_t) dt + \frac{\sqrt{2}}{\varepsilon} dB_t^\xi. \end{aligned}$$

For ease of notation we now restrict ourselves to the case that  $X$  and  $\xi$  are one dimensional variables and  $m(x) = 0$ , but the calculation proceeds analogously in the general case. For our choice of  $V$  given by (6.13) we have

$$\begin{aligned} dX_t &= -(U'(X_t) + \frac{1}{2}A'(X_t)\xi^2)dt + \sqrt{2}dB_t^x \\ d\xi_t &= -\frac{1}{\varepsilon^2}A(X_t)\xi_t dt + \frac{\sqrt{2}}{\varepsilon}dB_t^\xi. \end{aligned}$$

The fast variable  $\xi$  admits, for fixed  $X_t = x$ , the invariant distribution  $\nu_x \sim \mathcal{N}(0, A(x)^{-1})$ , with which we arrive at the averaged dynamics for the slow variable

$$d\bar{X}_t = -(U'(\bar{X}_t) + \frac{1}{2}\partial_{\bar{x}} \log(A(\bar{X}_t))) dt + \sqrt{2} dB_t^x. \quad (6.14)$$

This will be the reference solution to which we compare the solution accelerated by a non-reversible perturbation.

In order to accelerate the convergence of the dynamics we add a skew symmetric term again, which for our choice of  $V$  yields

$$\begin{aligned} d\tilde{X}_t &= -(U'(\tilde{X}_t) + \frac{1}{2}A'(\tilde{X}_t)\tilde{\xi}_t^2 + \frac{1}{\varepsilon}A(\tilde{X}_t)\tilde{\xi}_t) dt + \sqrt{2} dB_t^x \\ d\tilde{\xi}_t &= -(\frac{1}{\varepsilon^2}A(\tilde{X}_t)\tilde{\xi}_t - \frac{1}{\varepsilon}(U'(\tilde{X}_t) + \frac{1}{2}\tilde{\xi}_t^2 A'(\tilde{X}_t))) dt + \frac{\sqrt{2}}{\varepsilon} dB_t^\xi. \end{aligned} \quad (6.15)$$

**Proposition 6.10.** *Given a system of type (6.15), homogenization theory applies and the effective dynamics for the slow variable  $\tilde{X}_t$  are given by a time rescaled version of (6.14), that is, time is accelerated by a factor of two and the dynamics read*

$$d\hat{X}_t = -2(U'(\hat{X}_t) + \frac{1}{2}\partial_{\hat{x}} \log(A(\hat{X}_t))) dt + 2 dB_t^x. \quad (6.16)$$

*Proof.* The calculation is transferred to the appendix, see Proposition B.3.  $\square$

The proposition implies that adding a skew symmetric term in order to accelerate the convergence to equilibrium, in the case where one variable of the dynamics determines the slow process, does not actually speed up the relaxation. Instead, adding such a skew symmetric term, making the dynamics irreversible, leads to a time-rescaled version of the original effective dynamics which will not yield any computational improvement.

### 6.3 Mean first exit times for underdamped Langevin dynamics

In this section, we test the theoretical findings of Section 3.4.3 on the example of a double pendulum which we take from [56]. For more details on the example see Chapter 4.6 in [55].

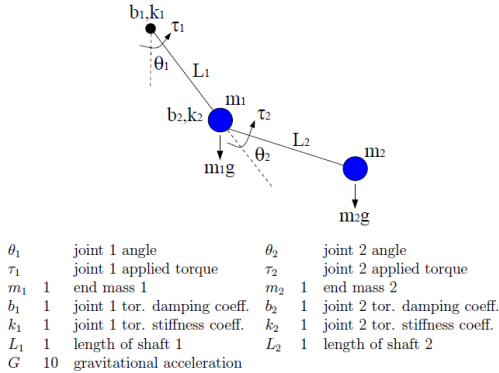


Figure 6.7: Illustration of a planar double pendulum system with massless shafts, linear torsional damping, linear torsional stiffness and torque input applied at the rotary points. Copied from [55], p.194.

The dynamics of the double pendulum, which are illustrated in Figure 6.7, are given in the variables  $q = (q_1, q_2)$  and  $\dot{q} = (\dot{q}_1, \dot{q}_2)$  which refer to the joint angles and the angular velocities respectively.

The kinetic, potential and dissipation energies for this system are given respectively by

$$K(q, \dot{q}) = \frac{1}{2} (m_1 L_1^2 \dot{q}_1^2 + m_2 L_1^2 \dot{q}_1^2 + m_2 L_2^2 (\dot{q}_1 + \dot{q}_2)^2 + 2m_2 L_1 L_2 \cos(q_2) \dot{q}_1 (\dot{q}_1 \dot{q}_2))$$

$$U(q, \dot{q}) = \frac{1}{2} (k_1 q_1^2 + k_2 q_2^2) - (m_1 + m_2) G L_1 \cos(q_1) - m_2 G L_2 \cos(q_1 + q_2)$$

$$R(q, \dot{q}) = \frac{1}{2} b_1 \dot{q}_1^2 + \frac{1}{2} b_2 \dot{q}_2^2,$$

and the system's Lagrangian is  $L = K - U$ . The equations of motion follow from the Euler-Lagrange equation

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = F - \frac{\partial R}{\partial \dot{q}}. \quad (6.17)$$

Here,  $F$  is the force exerted to the system, which in our case is randomized. That is,  $F$  is mimicked by a Brownian motion with prefactor chosen such that the fluctuation-dissipation relation (3.18) holds with parameter  $l = 1$ . We set all constants equal to 1 except for the gravitational force  $G = 10$ . The controllability function can be computed – using that the Hamiltonian is the Legendre transform of the Lagrangian – to be

$$\begin{aligned} L_c(q, \dot{q}) &= 2\dot{q}_1^2 + (\dot{q}_1 + \dot{q}_2)^2 + 2 \cos(q_2) \dot{q}_1 (\dot{q}_1 + \dot{q}_2) \\ &\quad + q_1^2 + q_2^2 - 40 \cos(q_1) - 20 \cos(q_1 + q_2) + 60. \end{aligned}$$

**MFETs from domains around the origin.** Clearly, by physical intuition, the origin yields a stable stationary point for the corresponding deterministic dynamics. We will thus compute MFETs from domains around the origin given by

$$\mathcal{D}_1 = \{(q, \dot{q}) \in \mathbb{R}^{2n} : |(q, \dot{q})| \leq 1\}$$

and

$$\mathcal{D}_2 = \{q \in \mathbb{R}^n : |q| \leq 0.5\}.$$

Note that  $\mathcal{D}_2$  is not bounded in  $\mathbb{R}^{2n}$  and thus the assumptions for Corollary 3.32 are not met in this case. Recall, that according to Corollary 3.32, the rate describing the MFET in the small noise limit, is given by the right hand side of

$$\lim_{\epsilon \rightarrow 0} \epsilon^2 \log \mathbb{E}_z(\tau) = \inf_{(q, \dot{q}) \in \partial \mathcal{D}} L_c(q, \dot{q}).$$

We compute the analytical values for the minimum on the boundaries of the domains, using the constrained minimisation function `fmincon` implemented in Matlab, which yields  $\inf \{L_c(q, \dot{q}) : (q, \dot{q}) \in \partial \mathcal{D}_1\} = 0.1715$  and  $\inf \{L_c(q, \dot{q}) : q \in \partial \mathcal{D}_2\} = 1.7075$ .

Let us now compute the rate, yielding the MFET in the small noise limit, by means of numerical simulations. For the estimation of the rate we generate  $n = 100$  trajectories according to the dynamics (6.17) using an Euler-Maruyama scheme. The mean first exit times are computed for different  $\beta > 0$ , ( $\epsilon = \sqrt{2\beta^{-1}}$ ) via a Monte Carlo estimation over the trajectories and we denote these estimates by  $\hat{\mathbb{E}}(\tau^\beta)$ . Let us remark, that we choose the range for  $\beta$  as large as possible for each domain, while already being in a regime for which the exits are rare, but not too rare, so we can still compute the MFETs by means of standard Monte Carlo. We perform 10 independent runs and average in the end over the results. A plot of  $\beta$  against the estimate  $\hat{\mathbb{E}}(\tau^\beta)$  of each run and the mean over the runs are shown in Figure 6.8.

Using these estimates, we perform a linear interpolation of  $\frac{\beta}{2}$  against  $\log(\hat{\mathbb{E}}(\tau^\beta))$  by which we obtain the numerical estimates of the corresponding rate.

The numerical results yield an estimate of 0.1547 for the exit rate from domain  $\mathcal{D}_1$ , which results in a relative error of 0.0980. For the domain  $\mathcal{D}_2$  we get the estimate 1.7288, and the relative error is 0.0125.

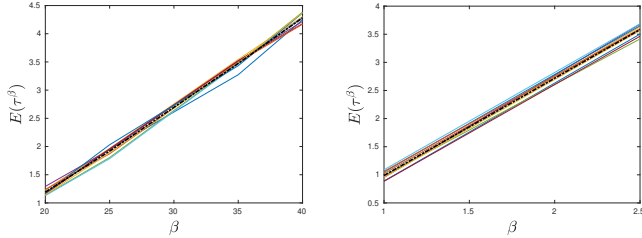


Figure 6.8: Illustration of the results for the estimation of MFET rates for the double pendulum. Each colored line corresponds to one of the 10 single runs and shows  $\log(\hat{\mathbb{E}}(\tau^\beta))$  as a function of  $\beta$ . The dashed black line is the mean of all the polynomial fits for this data. The left figure corresponds to the domain  $\mathcal{D}_1$  and the dashed black line has a slope of 0.1547, the right figure corresponds to the domain  $\mathcal{D}_2$  and the dashed black line has a slope of 1.7288.

**Remark 6.11.** Even though the assumptions of Corollary 3.32 are not met for the domain  $\mathcal{D}_2$  which is unbounded, the numerics suggest that the theory is still applicable. Presumably, this is due to the fact that the systems energy is approximately constant due to the fluctuation-dissipation relation and thus also the angular velocities  $\dot{q}$  remain in a bounded region. This suggests the following conjecture: as long as the potential grows sufficiently fast such that the dynamics dwell in a bounded domain, also exit times from unbounded domains can be treated.

## OUTLOOK

In this thesis we presented various results regarding time scales and model reduction for non-reversible SDEs. Let us now summarise the open questions emerging from this work.

Regarding MFETs for dynamics with degenerate noise, the use of the control theoretic formulation of the action functional seems promising and should be applied for determining MFETs of more general non-reversible dynamics than the ones considered here. Furthermore, our numerical example in Section 6.3 suggests that boundedness of the domain from which the exit is considered is not a necessary condition to hold as long as the dynamics is naturally bounded by its potential energy landscape. Hence the requirement of boundedness of the domain is an open question to be investigated.

Concerning the discussed model reduction of effective dynamics, the existing methods for proving error bounds rely on non-degeneracy of the diffusion. This raises the question whether error estimates can also be derived for dynamics with degenerate noise. We conjecture that this should indeed be possible, which is supported by our results for OU-processes and Section 4.3.3 providing formal calculations for effective dynamics and averaging to agree. Developing tools for proving corresponding error estimates is thus another open question.

Obtaining sharp error bounds for path measures is important when the effective dynamics are used for computing dynamic quantities, such as exit times and our result as given in Theorem 4.9 should be probed in more detail. Moreover, in order to improve the reduced model, a memory term or additional degrees of freedom could be incorporated.

With regard to sampling from conditional distributions, the possibility of deriving projections for non-linear dynamics which are based on our results should be investigated.





# BACKGROUND AND THEORY

## A.1 Definition of the Itô integral

In this chapter we provide some supplementary theory needed for the definition of the *Itô integral*.

**Definition A.1.** Introduce the partition  $\Delta_n = \{t_0^n, \dots, t_{k_n}^n\}$  such that  $0 = t_0^n < \dots < t_{k_n}^n = T$  for some  $T > 0$ . A *refinement of the partition* fulfils that

$$\max_i |t_{i+1}^n - t_i^n| \xrightarrow{n \rightarrow \infty} 0 \text{ and } \Delta_{n+1} \supset \Delta_n.$$

A function  $g : [0, T] \rightarrow \mathbb{R}$  is of

- *bounded variation* if  $\sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} |g(t_i^n) - g(t_{i+1}^n)| < \infty$  for all sequences of refinement partitions  $(\Delta_n)_{n \in \mathbb{N}}$ .
- *bounded quadratic variation* if  $\langle g \rangle_t = \sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} |g(t_i^n) - g(t_{i+1}^n)|^2 < \infty, \forall t \in [0, T]$  for all sequences of refinement partitions  $(\Delta_n)_{n \in \mathbb{N}}$ .

**Definition A.2.** For a continuous function  $f : [0, T] \rightarrow \mathbb{R}$  and a function  $g : [0, T] \rightarrow \mathbb{R}$  of bounded variation the *Riemann-Stieltjes integral* is defined as

$$\int_0^T f(s) dg(s) := \lim_{n \rightarrow \infty} \sum_{i=0}^{k_n-1} f(\xi_i^n) (g(t_{i+1}^n) - g(t_i^n))$$

which converges to a unique limit for any sequence of refinement partitions  $(\Delta_n)_{n \in \mathbb{N}}$  and independent of the choice of  $\xi_i^n \in [t_i^n, t_{i+1}^n]$  in  $L^2$ .

**Lemma A.3.** *Let  $B_t$  be a standard Brownian motion. Then  $\langle B_t \rangle_t = t$  almost surely.*

*Proof.* We have to show that  $\mathbb{E}(\langle B_t \rangle_t) = t$  and  $\text{Var}(\langle B_t \rangle_t) = 0$ .

Let  $(\Delta_n)_{n \in \mathbb{N}}$  be a sequence of refinement partitions of the interval  $[0, t]$ . Then

$$\mathbb{E}(\langle B_t \rangle_t) = \mathbb{E}(\sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} |B_{t_i^n} - B_{t_{i+1}^n}|^2) = \sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} \text{Var}(B_{t_i^n} - B_{t_{i+1}^n}) = t,$$

by the properties that the increments  $(B_{t_i} - B_{t_{i+1}})$  and  $(B_{t_j} - B_{t_{j+1}})$  are independent for  $i \neq j$  and  $(B_{t_i} - B_{t_{i+1}}) \sim \mathcal{N}(0, t_{i+1} - t_i)$ . Furthermore, using the independence of the increments again in the first step, the relation between squared standard normal random variables and the  $\chi^2$  distribution in the second step, and the fact that the quadratic variation of continuous functions is zero in the last step, we calculate

$$\begin{aligned} \mathbb{V}ar(\langle B \rangle_t) &= \mathbb{V}ar\left(\sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} |B_{t_i^n} - B_{t_{i+1}^n}|^2\right) \\ &= \sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} \mathbb{V}ar(|B_{t_i^n} - B_{t_{i+1}^n}|^2) \\ &= \sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} 2(t_{i+1}^n - t_i^n)^2 = 2\langle t \rangle_t = 0. \end{aligned}$$

□

**Corollary A.4.**  $B_t$  is not of bounded variation.

*Proof.* This is an immediate consequence following from the observation that if  $g : [0, T] \rightarrow \mathbb{R}$  is continuous and of bounded variation, then  $\langle g \rangle_t = 0$ . Indeed, due to continuity and the property of bounded variation, we have

$$\begin{aligned} 0 \leq \langle g \rangle_t &= \sup_{n \in \mathbb{N}} \sum_{i=0}^{k_n-1} |g(t_i^n) - g(t_{i+1}^n)|^2 \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^{k_n-1} |g(t_i^n) - g(t_{i+1}^n)|^2 \\ &\leq \lim_{n \rightarrow \infty} \sup_{i \in \{1, \dots, n\}} |g(t_i^n) - g(t_{i+1}^n)| \sum_{i=0}^{k_n-1} |g(t_i^n) - g(t_{i+1}^n)| \end{aligned}$$

□

## A.2 PI and LSI

### Equivalence of the dynamic PI and the existence of a spectral gap

Equation (3.3) shows that  $-\mathcal{L}$  is a positive operator on  $L_\mu^2$  and if  $D = 0$  then by (2.16)  $\mathcal{L}$  is self-adjoint on  $L_\mu^2$ . Assume now that  $-\mathcal{L}$  admits a discrete spectrum with eigenvalues  $\lambda_i > 0, i \in \mathbb{N}$  and eigenfunctions  $\phi_i$  providing a basis of  $L_\mu^2$ , such that the spectral decomposition reads

$$-\mathcal{L} \cdot = \sum_{i=0}^{\infty} \lambda_i \langle \phi_i, \cdot \rangle_\mu \phi_i,$$

By means of the spectral decomposition, we find that for any  $f \in \text{Dom}\mathcal{L}$  with  $\int f(x) d\mu(x) = 0$  we have, using  $f_i = \langle f, \phi_i \rangle_\mu$ ,

$$\mathcal{E}_{\mathcal{L}}(f) = - \int f \mathcal{L} f d\mu = - \int f \mathcal{L} \sum_{i=0}^{\infty} f_i \phi_i d\mu = \int f \sum_{i=0}^{\infty} \lambda_i f_i \phi_i d\mu \geq \lambda_1 \int f^2 d\mu.$$

The last inequality follows by noting that expressing  $f$  as  $f = \sum_{j \geq 0} f_j \phi_j$ , due to the orthogonality of the eigenvectors, only the quadratic terms remain.

Conversely, assume a PI holds with constant  $\alpha_{\text{PI}} > 0$ . Denote by  $\phi_0 \equiv 1$  the eigenvector corresponding to the eigenvalue 0 and by  $\phi_i, i > 0$  all eigenvectors corresponding to eigenvalues  $\lambda_i > 0$  and note that  $\int \phi_i d\mu = \langle 1, \phi_i \rangle_\mu = 0$  due to the orthogonality of the eigenvectors. Then

$$\mathcal{E}_{\mathcal{L}}(\phi_i) = \lambda_i \int \phi_i^2 d\mu$$

and since by assumption  $\mathcal{E}_{\mathcal{L}}(\phi_i) \geq \alpha_{\text{PI}} \int \phi_i^2 d\mu$  it follows that  $\lambda_i \geq \alpha_{\text{PI}}, \forall i \geq 1$  meaning there is a gap between  $\lambda_0 = 0$  and  $\lambda_1 = \alpha_{\text{PI}}$ .

**Lemma A.5.** *Let  $\mu \in \mathcal{P}(\mathbb{R}^n)$  satisfy a LSI with constant  $\alpha_{\text{LSI}}$ . Then  $\mu$  satisfies a PI with constant  $\alpha_{\text{PI}} = \alpha_{\text{LSI}}$ .*

*Proof.* Let  $f \in L^2_\mu$  with  $\int f d\mu = 0$ . Then  $(1 + \varepsilon f)\mu$  is still a probability density for small enough  $\varepsilon$ . By assumption

$$\begin{aligned} \mathcal{H}((1 + \varepsilon f)\mu | \mu) &\leq \frac{1}{2\alpha_{\text{LSI}}} \mathcal{R}_\Gamma((1 + \varepsilon f)\mu | \mu) \\ \Leftrightarrow \int (1 + \varepsilon f) \ln(1 + \varepsilon f) d\mu &\leq \frac{1}{2\alpha_{\text{LSI}}} \int (1 + \varepsilon f) |\nabla \ln(1 + \varepsilon f)|_\Gamma^2 d\mu = \frac{\varepsilon^2}{2\alpha_{\text{LSI}}} \int \frac{1}{1 + \varepsilon f} |\nabla f|_\Gamma^2 d\mu \end{aligned}$$

Expanding the left and right-hand side around  $\varepsilon = 0$  yields

$$\int \varepsilon f + \frac{\varepsilon^2}{2} f^2 + \mathcal{O}(\varepsilon^3) d\mu \leq \frac{\varepsilon^2}{2\alpha_{\text{LSI}}} \int |\nabla f|_\Gamma^2 (1 + \mathcal{O}(\varepsilon)) d\mu.$$

Noting that  $f$  is centered with respect to  $\mu$ , dividing the above equation by  $\varepsilon^2$  and letting  $\varepsilon \rightarrow 0$  leads to

$$\frac{1}{2} \int f^2 d\mu \leq \frac{1}{2\alpha_{\text{LSI}}} \int |\nabla f|_\Gamma^2 d\mu.$$

□

We state some results which guarantee existence of a spectral gap and are taken from [8].

**Theorem A.6.** *Assume there exists a Lyapunov function  $W$ , that is,  $W \in C^2, W \geq 1$  and  $\exists a, r > 0, b \geq 0$  such that*

$$\mathcal{L}W(x) \leq -aW(x) + b\mathbb{1}_{B_0(r)}(x).$$

*Then  $(\mu, \mathcal{L})$  satisfy a PI with constant  $\alpha_{\text{PI}} = (1 + b\alpha_r)/a$ , where  $\alpha_r$  is the PI constant of  $\mu$  restricted to  $B_0(r)$ , where  $B_0(r)$  is the open ball around 0 with radius  $r$  ( $\alpha_r$  exists since  $V$  is locally bounded).*

Sufficient criterions to apply the theorem are

- $\exists c > 0, r \geq 0$  such that  $\nabla V(x) \cdot x \geq r|x|$  for  $|x| \geq r$
- $\exists d \in (0, 1), e > 0, r \geq 0$  such that  $d|\nabla V(x)|^2 - \Delta V(x) \geq c$  for  $|x| \geq r$

The first criterion states that  $V$  has to grow sufficiently fast at infinity and the second is reminiscent of the spectral analysis for Schrödinger operators.

### Supplementaries on the rate function for gradient dynamics

**Theorem A.7** (Theorem 3.1 in [26]). *Let  $V \in C^2$ . The rate function  $S_T\phi$  is minimised by  $\phi = \phi^*$  solution to  $\dot{\phi}^*(t) = \nabla V(\phi^*(t))$  and the quasipotential takes the form  $V^q(x, y) = 2(V(y) - V_0)$ .*

*Proof.* Assume that  $\phi(t) \in \mathcal{D} \forall t \leq T$ . Else, if there was a  $\bar{t}$  with  $0 < \bar{t} < T$  such that  $\phi(\bar{t}) \notin \mathcal{D}$  then  $S_T\phi \geq S_{\bar{t}}\phi$ . Instead of considering  $S_T\phi$  one would then consider  $S_{\bar{t}}\phi$  for which  $\phi$  hasn't left the domain  $\forall t \leq \bar{t}$ . Note that  $\int_0^T \frac{d}{dt} \nabla V(\phi(t)) dt = \int_0^T \dot{\phi}(t) \cdot \nabla V(\phi(t)) dt$  and hence

$$\begin{aligned} S_T\phi &= \frac{1}{2} \int_0^T \left| \dot{\phi}(t) + \nabla V(\phi(t)) \right|^2 dt \\ &= \frac{1}{2} \int_0^T \left| \dot{\phi}(t) - \nabla V(\phi(t)) \right|^2 dt + 2 \int_0^T \dot{\phi}(t) \cdot \nabla V(\phi(t)) dt \\ &\geq 2(V(\phi(T)) - V(\phi(0))). \end{aligned}$$

If  $\dot{\phi}(t) = \nabla V(\phi(t))$  then  $S_T\phi = 2(V(\phi(T)) - V(\phi(0)))$ . Together with the constraint that  $\phi(0) = x_0$  and  $\phi(T) = y$  and noting that  $V \in C^2$  implies uniqueness of the solution, the claim follows.  $\square$

## APPLICATIONS AND EXAMPLES

### B.1 Convergence behaviour

**Lemma B.1.** *Let  $A, C \in \mathbb{R}^{2 \times 2}$  fulfil assumption A2 and let  $S \in \mathbb{R}^{2 \times 2}$  be such that  $S^{-1}AS = -\Lambda$ . Then  $\Lambda = \frac{1}{2}(\tilde{C}\tilde{C}^T + \tilde{J})\tilde{\Sigma}_\infty^{-1}$  with  $S^{-1}C = \tilde{C}$ ,  $\tilde{J} = S^{-1}JS^{-t}$ ,  $\tilde{\Sigma}_\infty = S^{-1}\Sigma_\infty S^{-T}$  and  $\Sigma_\infty$  being the unique solution to  $A\Sigma_\infty + \Sigma_\infty A^T = -CC^T$ . The entries of the matrices by  $d_{ij} = (\tilde{C}\tilde{C}^T)_{ij}$ ,  $j_{ij} = (\tilde{J})_{ij}$ ,  $(\tilde{\Sigma}_\infty^{-1})_{ij} = \sigma_{ij}$  are given by*

$$\begin{aligned} j_{12} &= \frac{1}{4}(\lambda_2 - \lambda_1) \frac{\sigma_{12}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \\ d_{11} &= \lambda_1 \frac{\sigma_{22}}{2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)} \\ d_{12} &= -\frac{1}{4}(\lambda_1 + \lambda_2) \frac{\sigma_{12}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \\ d_{22} &= \lambda_2 \frac{\sigma_{11}}{2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}. \end{aligned}$$

*Proof.* Diagonalizing  $A$  we have  $\Lambda = \frac{1}{2}(\tilde{C}\tilde{C}^T + \tilde{J})\tilde{\Sigma}_\infty^{-1}$ . Assuming that we are given  $\Lambda$  and  $\tilde{\Sigma}_\infty$  we can rewrite the diagonalized system of equations as a system which we solve for the unknowns  $d_{ij}, j_{ij}$ . In the two dimensional case we have four equations and the unknowns are  $d_{11}, d_{12}, d_{22}, j_{12}$  (the other entries can be computed by  $d_{21} = d_{12}, j_{11} = j_{22} = 0$  and  $j_{21} = -j_{12}$ ) altogether we arrive for the two-dimensional case at

$$\begin{pmatrix} \sigma_{11} & \sigma_{21} & 0 & \sigma_{21} \\ \sigma_{12} & \sigma_{22} & 0 & \sigma_{22} \\ 0 & \sigma_{11} & \sigma_{21} & -\sigma_{11} \\ 0 & \sigma_{12} & \sigma_{22} & -\sigma_{12} \end{pmatrix} \begin{pmatrix} d_{11} \\ d_{12} \\ d_{22} \\ j_{12} \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ 0 \\ 0 \\ \lambda_2 \end{pmatrix} \quad (\text{B.1})$$

**Solving for  $d_{11}, d_{12}, d_{22}$  and  $j_{12}$  in the 2-dimensional case:** We solve the system of linear equations (B.1) and use the numbering (I) – (IV) of the equations

which for each transformation gain a  $()'$ . Rewrite (B.1) as

$$\begin{aligned}
 (I) & \quad \left( \begin{array}{cccc|c} \sigma_{11} & \sigma_{21} & 0 & \sigma_{21} & \lambda_1 \\ (II) & \sigma_{12} & \sigma_{22} & 0 & \sigma_{22} & 0 \\ (III) & 0 & \sigma_{11} & \sigma_{21} & -\sigma_{11} & 0 \\ (IV) & 0 & \sigma_{12} & \sigma_{22} & -\sigma_{12} & \lambda_2 \end{array} \right) \\
 (II)' & = \left( (II) - \frac{\sigma_{12}}{\sigma_{11}}(I) \right) \sigma_{11} : \\
 & \quad \left( \begin{array}{cccc|c} \sigma_{11} & \sigma_{21} & 0 & \sigma_{21} & \lambda_1 \\ 0 & \sigma_{22}\sigma_{11} - \sigma_{12}^2 & 0 & \sigma_{22}\sigma_{11} - \sigma_{12}^2 & -\lambda_1\sigma_{12} \\ 0 & \sigma_{11} & \sigma_{21} & -\sigma_{11} & 0 \\ 0 & \sigma_{12} & \sigma_{22} & -\sigma_{12} & \lambda_2 \end{array} \right) \\
 (III)'' & = (III)' - \frac{\sigma_{11}}{\sigma_{22}\sigma_{11} - \sigma_{12}^2}(II)' \text{ and } (IV)'' = (IV)' - \frac{\sigma_{12}}{\sigma_{22}\sigma_{11} - \sigma_{12}^2} : \\
 & \quad \left( \begin{array}{cccc|c} \sigma_{11} & \sigma_{21} & 0 & \sigma_{21} & \lambda_1 \\ 0 & \sigma_{22}\sigma_{11} - \sigma_{12}^2 & 0 & \sigma_{22}\sigma_{11} - \sigma_{12}^2 & -\lambda_1\sigma_{12} \\ 0 & 0 & \sigma_{21} & -2\sigma_{11} & \frac{\lambda_1\sigma_{11}\sigma_{12}}{\sigma_{22}\sigma_{11} - \sigma_{12}^2} \\ 0 & 0 & \sigma_{22} & -2\sigma_{12} & \lambda_2 + \frac{\lambda_1\sigma_{12}^2}{\sigma_{22}\sigma_{11} - \sigma_{12}^2} \end{array} \right) \\
 (IV)''' & = (IV)'' - \frac{\sigma_{22}}{\sigma_{21}}(III)'' : \\
 & \quad \left( \begin{array}{cccc|c} \sigma_{11} & \sigma_{21} & 0 & \sigma_{21} & \lambda_1 \\ 0 & \sigma_{22}\sigma_{11} - \sigma_{12}^2 & 0 & \sigma_{22}\sigma_{11} - \sigma_{12}^2 & -\lambda_1\sigma_{12} \\ 0 & 0 & \sigma_{21} & -2\sigma_{11} & \frac{\lambda_1\sigma_{11}\sigma_{12}}{\sigma_{22}\sigma_{11} - \sigma_{12}^2} \\ 0 & 0 & 0 & 2\frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{12}} & \lambda_2 - \lambda_1 \end{array} \right)
 \end{aligned}$$

With this the obtained values follow. □

## B.2 Accelerations

**Proposition B.2.** *Consider*

$$\begin{aligned}
 dX_t^\varepsilon & = -(\nabla U(X_t^\varepsilon) + \frac{1}{\varepsilon} J \nabla U(X_t^\varepsilon)) dt + \sqrt{2\beta^{-1}} dB_t \\
 \text{with } U(X) & = \frac{1}{2} (\lambda_1 x_1^2 + \lambda_2 x_2^2), \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},
 \end{aligned}$$

$X \in \mathbb{R}^2$  and  $B_t$  is 2-dimensional Brownian motion. Let  $Y_t = U(X_t^\varepsilon)$ , then

$$Y^\varepsilon \xrightarrow{\varepsilon \rightarrow 0} Y \quad \text{weakly in } \mathcal{C}([0, T], \mathbb{R}).$$

with  $Y$  being the solution of the SDE

$$dY_t = \mu (\beta^{-1} - Y_t) dt + \sqrt{2\mu\beta^{-1}Y_t} dB_t, \quad \text{with } \mu = \lambda_1 + \lambda_2,$$

*Proof.* We omit the formal proof which can be found in [26, 63] and only present the calculation of the effective coefficients. According to [63] (see the calculations below Theorem 3.5) the effective drift and diffusion read

$$b(y) = \frac{-M(y) + \beta^{-1}M'(y)}{T(y)}, \quad \gamma(y) = \sqrt{\frac{2\beta^{-1}M(y)}{T(y)}}$$

respectively with  $M(y) = \int_{\text{int}(d(y))} \Delta U(x) dx$ ,  $T(y) = \oint_{d(y)} |J\nabla U(x)|^{-1} dx$ ,

and  $d(y) = \{x \in \mathbb{R}^2 : U(x) = y\}$  being the corresponding level set. For  $M(y)$  we have to integrate the constant function  $\Delta U(x) = \lambda_1 + \lambda_2$  over the interior of the ellipse described by

$$\frac{\lambda_1}{2}x_1^2 + \frac{\lambda_2}{2}x_2^2 = y. \tag{B.2}$$

In general, the interior area of a two-dimensional ellipse with semi-major and semi-minor axis  $a, b$  respectively, is given by  $\pi ab$ . In our case

$$a = \sqrt{\frac{2y}{\lambda_1}}, \quad b = \sqrt{\frac{2y}{\lambda_2}},$$

which implies

$$M(y) = \frac{2\pi}{\sqrt{\lambda_1\lambda_2}}(\lambda_1 + \lambda_2)y.$$

For the calculation of  $T(y)$  we have to parametrise the level set  $d(y)$ , i.e., the ellipse B.2 which can be done using  $f(t) = (a \cos(t), b \sin(t))$  with  $a, b$  as above. Then

$$\begin{aligned} T(y) &= \oint_{d(y)} |J\nabla U(x)|^{-1} dx = \int_0^{2\pi} \frac{|f'(t)|}{|J\nabla U(x)|} dt \\ &= \int_0^{2\pi} \sqrt{\frac{\lambda_1^{-1} \sin^2(t) + \lambda_2^{-1} \cos^2(t)}{\lambda_2 \sin^2(t) + \lambda_1 \cos^2(t)}} dt \\ &= \int_0^{2\pi} \sqrt{\frac{\lambda_1^{-1} \sin^2(t) + \lambda_2^{-1} \cos^2(t)}{\lambda_1 \lambda_2 (\lambda_1^{-1} \sin^2(t) + \lambda_2^{-1} \cos^2(t))}} dt = 2\pi \frac{1}{\sqrt{\lambda_1 \lambda_2}}, \end{aligned}$$

which results in the drift  $b(y) = \frac{-M(y) + \beta^{-1}M'(y)}{T(y)} = \mu(\beta^{-1} - y)$  and diffusion  $\gamma(y) = \sqrt{\frac{2\beta^{-1}M(y)}{T(y)}} = \sqrt{2\beta^{-1}\mu y}$ .  $\square$

**Proposition B.3.** *Given the dynamics*

$$\begin{aligned} d\tilde{X}_t &= -(U'(\tilde{X}_t) + \frac{1}{2}A'(\tilde{X}_t)\tilde{\xi}_t^2 + \frac{1}{\varepsilon}A(\tilde{X}_t)\tilde{\xi}_t) dt + \sqrt{2} dB_t^x \\ d\tilde{\xi}_t &= -(\frac{1}{\varepsilon^2}A(\tilde{X}_t)\tilde{\xi}_t - \frac{1}{\varepsilon}(U'(\tilde{X}_t) + \frac{1}{2}\tilde{\xi}_t^2 A'(\tilde{X}_t))) dt + \frac{\sqrt{2}}{\varepsilon} dB_t^\xi. \end{aligned} \tag{B.3}$$

homogenization theory applies and the effective dynamics for the slow variable  $\tilde{X}_t$  is given by a time rescaled version of (6.14), that is, time is accelerated by a factor of two and the dynamics read

$$d\hat{X}_t = -2(U'(\hat{X}_t) + \frac{1}{2}\partial_x \log(A(\hat{X}_t))) dt + 2 dB_t^x. \quad (\text{B.4})$$

*Proof.* We only do formal perturbation theory expansions and follow the presentation in [62]. In the following, for ease of notation, replace  $\tilde{x}, \tilde{\xi}$  by  $x, \xi$  respectively. Homogenization theory applies since the fastest process admits for each fixed  $x$  a unique invariant distribution given by  $\nu_x \sim \mathcal{N}(0, A(x)^{-1})$  and the centring condition holds, i.e., the  $\mathcal{O}(\frac{1}{\varepsilon})$  term in the slow dynamics averages out with respect to  $\nu_x$ :

$$\int \xi A(x) \nu_x(\xi) d\xi = 0. \quad (\text{B.5})$$

The generator of (B.3) is given by  $\mathcal{L} = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2$  where, recalling that  $V(x, \xi) = U(x) + \xi^T A(x)^{-1} \xi$ ,

$$\begin{aligned} \mathcal{L}_0 &= -\nabla_\xi V \cdot \nabla_\xi + \Delta_\xi \\ \mathcal{L}_1 &= \nabla_x V \cdot \nabla_\xi - \nabla_\xi V \cdot \nabla_x \\ \mathcal{L}_2 &= -\nabla_x V \cdot \nabla_x + \Delta_x. \end{aligned}$$

Also expanding the solution to the Kolmogorov backward equation  $\mathcal{L}v = \frac{\partial v}{\partial t}$  formally in terms of  $\varepsilon$  gives  $v = v_0 + \varepsilon v_1 + \varepsilon^2 v_2 + \mathcal{O}(\varepsilon^2)$ . Equating terms of same power yields

$$\mathcal{O}\left(\frac{1}{\varepsilon^2}\right): \quad \mathcal{L}_0 v_0 = 0 \quad (\text{B.6})$$

$$\mathcal{O}\left(\frac{1}{\varepsilon}\right): \quad \mathcal{L}_1 v_0 = -\mathcal{L}_0 v_1 \quad (\text{B.7})$$

The first equation (B.6) implies – noting that  $\mathcal{L}_0$  is a differential operator in  $\xi$  only – that  $v_0 = v_0(x)$ . The second equation (B.7) possesses – by the Fredholm alternative – a solution  $v_1$  iff  $\mathcal{L}_1 v_0 = -\nabla_\xi V \cdot \nabla_x v_0$  is orthogonal to the kernel of  $\mathcal{L}_0$  given by  $\nu_x$ , i.e.,

$$\int -\nabla_\xi V \cdot \nabla_x v_0 \nu_x(\xi) d\xi = 0$$

which is precisely the centring condition (B.5) The *cell problem* corresponding to (B.7) reads

$$\nabla_\xi V(X, \xi) \cdot \nabla_\xi \phi + \Delta_\xi \phi = -\nabla_\xi V(X, \xi)$$

with solution  $\phi = -\xi$  and thus  $v_1 = \phi \cdot \nabla_x v_0 = -\xi \cdot \nabla_x v_0$ .

On the  $\mathcal{O}(1)$  time scale we have

$$\mathcal{L}_0 v_2 = -\frac{\partial v_0}{\partial t} + \mathcal{L}_2 v_0 + \mathcal{L}_1 v_1.$$



In order to guarantee the existence of a solution  $v_2$ , again it is required that the right hand side averages to zero under  $\nu_x$ , that is we compute for the  $\mathcal{O}(1)$  time scale solution  $v_0$ :

$$\frac{\partial v_0}{\partial t} = \int (\mathcal{L}_2 v_0 + \mathcal{L}_1 v_1) d\nu_x(\xi) = I_1 + I_2.$$

We compute the terms separately,

$$\begin{aligned} I_1 &= \int \mathcal{L}_2 v_0 d\nu_x(\xi) \\ &= \int -U'(x) \cdot \nabla_x v_0 - \frac{1}{2} A'(x) \xi^2 \cdot \nabla_x v_0 + \Delta_x v_0 d\nu_x(\xi) \\ &= -U'(x) \cdot \nabla_x v_0 - \frac{1}{2} A(x)^{-1} A'(x) \cdot \nabla_x v_0 + \Delta_x v_0 \\ I_2 &= \int \mathcal{L}_1 v_1 d\nu_x(\xi) \\ &= \int (U'(x) \cdot \nabla_\xi (-\xi \cdot \nabla_x v_0) + \frac{1}{2} A'(x) \xi^2 \cdot \nabla_\xi (-\xi \cdot \nabla_x v_0) - A(x) \xi \cdot \nabla_x (-\xi \cdot \nabla_x v_0)) d\nu_x(\xi) \\ &= \int (-U'(x) \cdot \nabla_x v_0 - \frac{1}{2} A'(x) \xi^2 \cdot \nabla_x v_0 + A(x) \xi^2 \Delta_x v_0) d\nu_x(\xi) \\ &= -U'(x) \cdot \nabla_x v_0 - \frac{1}{2} A(x)^{-1} A'(x) \cdot \nabla_x v_0 + \Delta_x v_0. \end{aligned}$$

With this the Kolmogorov backward equation for  $v_0$  reads

$$\frac{\partial v_0}{\partial t} = -2U'(x) \cdot \nabla_x v_0 - \partial_x(\log(A(x))) \cdot \nabla_x v_0 + 2\Delta_x v_0$$

which in turn corresponds to the SDE

$$d\hat{X}_t = -2 \left( U'(\hat{X}_t) + \frac{1}{2} \partial_{\hat{x}} \log(A(\hat{X}_t)) \right) dt + 2 dB_t^x.$$

□



# Bibliography

- [1] Niloufar Abourashchi and Alexander Yur'evich Veretennikov. On stochastic averaging and mixing. *Theory of Stochastic Processes*, 16(32):1, 2010.
- [2] Franz Achleitner, Anton Arnold, and Dominik Stürzer. Large-time behavior in non-symmetric fokker-planck equations. *arXiv preprint arXiv:1506.02470*, 2015.
- [3] Athanasios C Antoulas. *Approximation of large-scale dynamical systems*, volume 6. Siam, 2005.
- [4] Anton Arnold, Eric Carlen, and Qiangchang Ju. Large-time behavior of non-symmetric fokker-planck type equations. *Communications on Stochastic Analysis*, 2(1):11, 2008.
- [5] Anton Arnold and Jan Erb. Sharp entropy decay for hypocoercive and non-symmetric fokker-planck equations with linear drift. *arXiv preprint arXiv:1409.5425*, 2014.
- [6] Anton Arnold, Peter Markowich, Giuseppe Toscani, and Andreas Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Commun. Part. Diff. Eq.*, 26(1-2):43–100, 2001.
- [7] Dominique Bakry. L'hypercontractivité et son utilisation en théorie des semi-groupes. In *Lectures on probability theory*, pages 1–114. Springer, 1994.
- [8] Dominique Bakry, Franck Barthe, Patrick Cattiaux, Arnaud Guillin, et al. A simple proof of the poincaré inequality for a large class of probability measures. *Electronic Communications in Probability*, 13:60–66, 2008.
- [9] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer Science & Business Media, 2013.
- [10] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.
- [11] VI Bogachev, M Röckner, and SV Shaposhnikov. Distances between transition probabilities of diffusions and applications to nonlinear fokker–planck–kolmogorov equations. *Journal of Functional Analysis*, 271(5):1262–1300, 2016.

- 
- [12] Freddy Bouchet and Julien Reygner. Generalisation of the eyring–kramers transition rate formula to irreversible diffusion processes. In *Annales Henri Poincaré*, volume 17, pages 3499–3532. Springer, 2016.
- [13] Anton Bovier, Véronique Gayraud, and Markus Klein. Metastability in reversible diffusion processes ii: Precise asymptotics for small eigenvalues. *Journal of the European Mathematical Society*, 7(1):69–99, 2005.
- [14] Marco Bramanti. *An invitation to hypoelliptic operators and Hörmander’s vector fields*. Springer, 2014.
- [15] Damiano Brigo and Fabio Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007.
- [16] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.
- [17] Alexandre J Chorin, Anton P Kast, and Raz Kupferman. Optimal prediction of underresolved dynamics. *Proceedings of the National Academy of Sciences*, 95(8):4094–4098, 1998.
- [18] Giovanni Ciccotti, Raymond Kapral, and Eric Vanden-Eijnden. Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *ChemPhysChem*, 6(9):1809–1814, 2005.
- [19] Giacomo Di Gesù, Tony Lelièvre, Dorian Le Peutrec, and Boris Nectoux. Sharp asymptotics of the first exit point density. *Annals of PDE*, 5(1):5, 2019.
- [20] Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Transactions of the American Mathematical Society*, 367(6):3807–3828, 2015.
- [21] MH Duong, A Lamacz, MA Peletier, A Schlichting, and U Sharma. Quantification of coarse-graining error in langevin and overdamped langevin dynamics. *Nonlinearity*, 31(10):4517, 2018.
- [22] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability theory and related fields*, 166(3-4):851–886, 2016.
- [23] Kenneth David Elworthy. *Stochastic differential equations on manifolds*, volume 70. Cambridge University Press, 1982.
- [24] Klaus-Jochen Engel and Rainer Nagel. One-parameter semigroups for linear evolution equations. In *Semigroup Forum*, volume 63, pages 278–280. Springer, 2001.
- [25] Wendell H Fleming. Stochastic control and large deviations. In *Conference Organized by INRIA, France*, pages 291–300. Springer, 1992.

- 
- [26] M. Freidlin and A. Wentzell. *Random perturbations of Hamiltonian systems*, volume 260. Springer Berlin Heidelberg, 2012.
- [27] Arnaud Guillin, Pierre Monmarché, et al. Optimal linear drift for the speed of convergence of an hypoelliptic diffusion. *Electronic Communications in Probability*, 21, 2016.
- [28] István Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an itô differential. *Probability theory and related fields*, 71(4):501–516, 1986.
- [29] Carsten Hartmann, Christof Schütte, Marcus Weber, and Wei Zhang. Importance sampling in path space for diffusion processes with slow-fast variables. *Probability Theory and Related Fields*, 170(1-2):177–228, 2018.
- [30] Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pages 1188–1205, 1986.
- [31] Bernard Helffer and Francis Nier. Quantitative analysis of metastability in reversible diffusion processes via a witten complex approach: the case with boundary. *Matemática contemporânea*, 26:41–85, 2004.
- [32] Frédéric Hérau, Michael Hitrik, and Johannes Sjöstrand. Tunnel effect for kramers-fokker-planck type operators. In *Annales Henri Poincaré*, volume 9, pages 209–274. Springer, 2008.
- [33] Frédéric Hérau and Francis Nier. Isotropic hypoellipticity and trend to equilibrium for the fokker-planck equation with a high-degree potential. *Archive for Rational Mechanics and Analysis*, 171(2):151–218, 2004.
- [34] Richard Holley and Daniel Stroock. Logarithmic sobolev inequalities and stochastic ising models. *Journal of statistical physics*, 46(5):1159–1194, 1987.
- [35] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Mathematica*, 119(1):147–171, 1967.
- [36] Chii-Ruey Hwang, Shu-Yin Hwang-Ma, and Shuenn-Jyi Sheu. Accelerating gaussian diffusions. *The Annals of Applied Probability*, pages 897–913, 1993.
- [37] Chii-Ruey Hwang, Shu-Yin Hwang-Ma, Shuenn-Jyi Sheu, et al. Accelerating diffusions. *The Annals of Applied Probability*, 15(2):1433–1444, 2005.
- [38] Marcus Kaiser, Robert L Jack, and Johannes Zimmer. Acceleration of convergence to equilibrium in markov chains by breaking detailed balance. *Journal of Statistical Physics*, 168(2):259–287, 2017.
- [39] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

- [40] Markos Katsoulakis, Yannis Pantazis, and Luc Rey-Bellet. Measuring the irreversibility of numerical schemes for reversible stochastic differential equations\*. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(5):1351–1379, 2014.
- [41] RZ Khas' minskii. On stochastic processes defined by differential equations with a small parameter. *Theory of Probability & Its Applications*, 11(2):211–228, 1966.
- [42] Ernest Bruce Lee and Lawrence Markus. Foundations of optimal control theory. Technical report, Minnesota Univ Minneapolis Center For Control Sciences, 1967.
- [43] Frédéric Legoll and Tony Lelièvre. Effective dynamics using conditional expectations. *Nonlinearity*, 23(9):2131, 2010.
- [44] Frédéric Legoll, Tony Lelièvre, and Stefano Olla. Pathwise estimates for an effective dynamics. *Stochastic Processes and their Applications*, 127(9):2841–2863, 2017.
- [45] Frédéric Legoll, Tony Lelièvre, and Upanshu Sharma. Effective dynamics for non-reversible stochastic differential equations: a quantitative study. *arXiv preprint arXiv:1809.10498*, 2018.
- [46] Tony Lelièvre. Two mathematical tools to analyze metastable stochastic processes. In *Numerical Mathematics and Advanced Applications 2011*, pages 791–810. Springer, 2013.
- [47] Tony Lelièvre, Francis Nier, and Grigorios A Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152(2):237–274, 2013.
- [48] Tony Lelièvre, Mathias Rousset, and Gabriel Stoltz. Langevin dynamics with constraints and computation of free energy differences. *Mathematics of computation*, 81(280):2071–2125, 2012.
- [49] Yi Liu and Brian DO Anderson. Singular perturbation approximation of balanced systems. *International Journal of Control*, 50(4):1379–1405, 1989.
- [50] Andrew J Majda. *Introduction to turbulent dynamical systems in complex systems*. Springer, 2016.
- [51] Paul Malliavin. Stochastic calculus of variations and hypoelliptic operators. In *Proc. Internat. Symposium on Stochastic Differential Equations, Kyoto Univ., Kyoto, 1976*. Wiley, 1978.
- [52] Georg Menz, André Schlichting, et al. Poincaré and logarithmic sobolev inequalities by decomposition of the energy landscape. *The Annals of Probability*, 42(5):1809–1884, 2014.

- [53] Giorgio Metafune, Diego Pallara, and Enrico Priola. Spectrum of ornstein-uhlenbeck operators in lp spaces with respect to invariant measures. *Journal of Functional Analysis*, 196(1):40–60, 2002.
- [54] Lara Neureither and Carsten Hartmann. Time scales and exponential trend to equilibrium: Gaussian model problems. In *International workshop on Stochastic Dynamics out of Equilibrium*, pages 391–410. Springer, 2017.
- [55] Andrew J Newman. *Modeling and reduction with applications to semiconductor processing*. PhD thesis, 1999.
- [56] Andrew J Newman and Perinkulam S Krishnaprasad. Computing balanced realizations for nonlinear systems. Technical report, ALPHATECH INC ARLINGTON VA, 2000.
- [57] David Nualart. *The Malliavin calculus and related topics*, volume 1995. Springer, 2006.
- [58] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.
- [59] Felix Otto and Cédric Villani. Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000.
- [60] E Pardoux, A Yu Veretennikov, et al. On poisson equation and diffusion approximation 2. *The Annals of Probability*, 31(3):1166–1192, 2003.
- [61] Grigoris A Pavliotis. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, volume 60. Springer, 2014.
- [62] Grigoris Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.
- [63] Luc Rey-Bellet, Konstantinos Spiliopoulos, et al. Variance reduction for irreversible langevin samplers and diffusion on graphs. *Electronic Communications in Probability*, 20, 2015.
- [64] Gilles Royer. *An initiation to logarithmic Sobolev inequalities*. Number 5. American Mathematical Soc., 2007.
- [65] Jacqueline Maria Aleida Scherpen. Balancing for nonlinear systems. *Systems & Control Letters*, 21(2):143–153, 1993.
- [66] Christof Schütte and Marco Sarich. *Metastability and Markov State Models in Molecular Dynamics*, volume 24. American Mathematical Soc., 2013.
- [67] Upanshu Sharma. *Coarse-graining of Fokker-Planck equations*. PhD thesis, 2017.

- [68] Cédric Villani. Hypocoercive diffusion operators. In *International Congress of Mathematicians*, volume 3, pages 473–498, 2006.
- [69] J Zabczyk. Exit problem and control theory. *Systems & control letters*, 6(3):165–172, 1985.
- [70] Jerzy Zabczyk. *Mathematical control theory: an introduction*. Springer Science & Business Media, 2009.