Time scales and exponential trend to equilibrium: Gaussian model problems

Lara Neureither and Carsten Hartmann

Brandenburgische Technische Universität Cottbus-Senftenberg, Institut für Mathematik, Konrad-Wachsmann-Allee 1, D-03046 Cottbus, Germany,
neurelar@b-tu.de, hartmanc@b-tu.de

Abstract. We review results on the exponential convergence of multidimensional Ornstein-Uhlenbeck processes and discuss notions of characteristic time scales by means of concrete model systems. We focus, on the one hand, on exit time distributions and provide explicit expressions for the exponential rate of the distribution in the small-noise limit. On the other hand, we consider relaxation time scales of the process to its equilibrium measure in terms of relative entropy and discuss the connection with exit probabilities. Along these lines, we study examples which illustrate specific properties of the relaxation and discuss the possibility of deriving a simulation-based, empirical definition of slow and fast degrees of freedom which builds upon a partitioning of the relative entropy functional in connection with the observed relaxation behaviour.

Keywords: Multidimensional Ornstein-Uhlenbeck process, exponential convergence, relative entropy, large deviations, small noise asymptotics

1 Introduction

The characteristic time scales of a random dynamical system, e.g. a diffusion process are often associated with the speed at which the dynamics reaches an equilibrium state or samples its invariant measure. For dynamical systems with several metastable equilibria (also: "metastable states"), such as molecular systems [SS13], chemical reaction networks [HTB90], or earth and climate systems [IVS01,LB17], the speed of convergence is often related to the characteristic time scale of transitions between these equilibria. Here, specifically, we are concerned with the speed of convergence of linear ergodic diffusion processes to their unique stationary probability distribution, and we review different concepts of such characteristic time scales in terms of exponential estimates for entropy decay, exit probabilities and mean hitting times. In doing so, we focus on the relations between the different concepts from a computational perspective and on the basis of concrete case studies. Understanding the characteristic time scales of a random dynamical system is not only important in statistical physics (to which the aforementioned applications belong), but it is also relevant to assess the asymptotic properties of Markov chain Monte Carlo (MCMC) algorithms [CC96], or failure probabilities in system reliability and risk analysis [MSB14], to mention just two more examples.

Related work The analysis of characteristic time scales and exponential convergence to equilibrium is system specific, and various different approaches have been developed in the past. We refrain from giving a complete list of references (which would be difficult anyway), but focus on approaches that are most relevant for statistical mechanics applications. Specifically, for reversible and metastable Markov chains and diffusion processes that are relevant for the modelling and the simulation of many-particle systems and critical phase transitions, the analysis of the eigenvalues of the infinitesimal generator become a standard tool; see e.g. [BdH15,HS06] and the references therein. For a certain class of non-reversible diffusions, spectral properties have furthermore been analysed in connection with small-noise limits; see e.g. [Ven73,NH05] or [Ber13] for a tutorial review. Despite the limitation to reversible systems (satisfying detailed balance) or perturbations of such systems, the spectral approach is appealing, since it allows for an hierarchical decomposition of the dynamics, based on the eigenvalues and eigenfunctions of the associated semigroup or its generator [SS13]. The key observation here is that the eigenvalues close to the principal eigenvalue $\lambda = 0$ represent characteristic time scales (sometimes called "implied time scales") that can be associated with the local transitions between metastable sets [BGK05,HMS04]; related results for small-noise diffusions that establish a link between dominant eigenvalues of a diffusion and exit times are discussed in, e.g., [EK87,Day87].

A more global perspective to the relaxation dynamics is provided by entropy estimates that can be used to prove exponential convergence to the stationary distribution. These approaches are based on certain functional inequalities like the Poincaré and the (logarithmic) Sobolev inequality, and provide bounds for the convergence to the stationary distribution in the L^1 norm in terms of relative entropy. These bounds utilize the celebrated Csiszár-Kullback-Pinsker inequality, and for reversible systems with potentials that are growing quadratically at infinity, the use of logarithmic Sobolev constants and relative entropy (or: Kullback-Leibler divergence) can be attributed to Bakry and Émery [BE85]. These results have then been generalized to nonlinear [MV00], non-reversible [ACJ08] and linear diffusions with degenerate noise [AE14], including generalizations of the Csiszár-Kullback-Pinsker inequality to relative entropies beyond the Kullback-Leibler divergence (see e.g. [AMTU01, Ch. 2.2]). For a survey of entropy techniques, functional inequalities and exponential convergence estimates, with a special focus on applications in molecular dynamics, we refer to [LS16]. An attempt to make the entropy estimates hierarchical, like it is done in spectral approaches, has been undertaken in [MS14], but a truly hierarchical approach is, to our knowledge, yet missing.

One motivation for studying exponential rates for the convergence to equilibrium is to devise MCMC methods that either sample the stationary distribution at a higher exponential rate (e.g. [KJZ17])) or reduce the variance of certain statistical estimators (e.g. [HSZ16]). Importance sampling and related variance reduction methods are naturally connected to large deviations principles, in that they are often applied in the context of small-noise diffusions [DW04,VEW12], for which MCMC methods are known to converge poorly, or ergodic sampling problems [CT15,RBS15] that can benefit from faster convergence to equilibrium; see also the seminal articles [FS97,FM95] for a discussion of the theoretical connection between large deviations and stochastic control from the viewpoint of viscosity solutions, and [JS14,SWH12] for applications in molecular dynamics.

Outline The rest of the article is structured as follows: In Section 2 we introduce the multidimensional Ornstein-Uhlenbeck (OU) process and briefly discuss its asymptotic properties for large times. Section 3 is devoted to a review of relevant entropy estimates for reversible, non-reversible and degenerate OU processes, which is contrasted with the Donsker-Varadhan and Freidlin-Wentzell large deviations approaches to the corresponding exit problem in Section 4. Numerical examples that illustrate the theoretical results are shown in Section 5, with a special focus on degenerate diffusions of Langevin type and slow-fast systems. The discussion is summarised in Section 6, and an outlook to possible future research directions is given.

2 Linear Systems

In this section we review some known results about linear stochastic differential equations based on the works [AE14] and [VF70] and discuss some concrete specifications for the problem at hand. For this we assume the following setting. Consider an Ornstein-Uhlenbeck process $(X_t)_{(t\geq 0)}$ where for each t, $X_t \in \mathbb{R}^n$ which is described by the SDE

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + C\,\mathrm{d}W_t. \tag{1}$$

Here $A \in \mathbb{R}^{n\times n}$ is referred to as drift matrix, $C \in \mathbb{R}^{n\times m}$ as diffusion matrix and W_t is standard m-dimensional Brownian motion and $m \leq n$.

The corresponding Fokker-Planck equation, which describes the time evolution of the probability density function ρ_t according to the dynamics given by (1), reads

$$\partial_t \rho_t = -\nabla\cdot(A\rho_t) + \frac{1}{2}\nabla^2 : (CC^T\rho_t), \tag{2}$$

where ∇^2 is the Hessian matrix and $A : B$ is the Frobenius inner product of two matrices A and B.
In order to guarantee a unique invariant distribution for our process, we impose the following assumptions on the matrices A and C.

Assumption 1 *The drift matrix A is stable (Hurwitz), i.e. all eigenvalues of A have strictly negative real part.*

Assumption 2 *No eigenvector v of A^T is in the kernel of C^T.*

Assumption 1 guarantees asymptotic stability of the dynamics and entails positive recurrence, whereas Assumption 2 guarantees spreading of the noise in every direction of state space (irreducibility). Assumption 2 has many equivalent formulations, one of which is the complete controllability of the matrix pair (A, C); given a controlled ODE $\dot{x}(t) = Ax(t) + Cu(t)$ with A, C as before, controllability means that there exists a bounded and measurable control u such that the origin $x = 0$ can be reached from any point $y \in \mathbb{R}^n$ in finite time. For further details we refer to [Zab09, Thm. 1.2] or [Ant05, Thm. 4.5]. With these assumptions at hand we are now ready to characterize the unique invariant distribution of (1).

Proposition 1. *Assume that Assumptions 1 and 2 hold true. Then the Fokker-Planck equation* (2) *has a unique stationary state ρ_∞, given by the probability density function of the normal distribution $\mathcal{N}(0, \Sigma_\infty)$, with mean zero and covariance Σ_∞. Furthermore the covariance matrix Σ_∞ is the unique positive definite solution to the Lyapunov equation*

$$A\Sigma_\infty + \Sigma_\infty A^T = -CC^T. \tag{3}$$

We will also refer to ρ_∞ as the invariant distribution of the process $(X_t)_{(t\geq 0)}$ described by (1). For a detailed proof we refer the reader to [AE14, Thm. 3.1].

Here, we aim at giving an intuitive explanation of the result. Consider the analytic solution to the SDE (1) for a deterministic initial condition $X_0 = x_0$, i.e. the initial covariance Σ_0 fulfills $\Sigma_0 = 0$, which reads

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

The mean μ_t and covariance Σ_t of the process at time t are easily calculated. They fully characterize the distribution of the process at time t, since the dynamics are linear and hence the distributions will be gaussian at all times. Mean and covariance are calculated to be

$$\mu_t = \mathbb{E}(X_t) = e^{At}x_0,$$
$$\Sigma_t = \mathbb{E}\left((X_t - \mu_t)(X_t - \mu_t)^T\right) = \int_0^t e^{A(t-s)}CC^T e^{A^T(t-s)}\,\mathrm{d}s.$$

Clearly, $\mu_t \to 0$ as $t \to \infty$, since all eigenvalues of A have negative real part according to Assumption 1. For Σ_t we first note that, again by Assumption 1, $\lim_{t\to\infty} \Sigma_t = \Sigma_\infty$ is well-defined. Further we observe that $\lim_{t\to\infty} \Sigma_t = \Sigma_\infty$ is equivalent to $\lim_{t\to\infty} \dot{\Sigma}_t = 0$. Now,

$$\dot{\Sigma}_t = A\Sigma_t + \Sigma_t A^T + CC^T$$

and thus

$$\lim_{t\to\infty} \Sigma_t = \Sigma_\infty \;\Leftrightarrow\; A\Sigma_\infty + \Sigma_\infty A^T + CC^T = 0.$$

The same calculation goes through when the initial condition is not deterministic (i.e. when $\Sigma_0 \neq 0$), but follows a absolutely continuous probability distribution. Uniqueness of the solution is not hard to prove either, and we refer to [Zab09, Thm. 2.7] for the details.

3 Entropy decay

In this section we mainly review results by [AE14] who proved exponential convergence to equilibrium for densities evolving according to (2) under Assumptions 1 and 2. To this end, we first sketch the general procedure how to prove such results for the case of non-degenerate Fokker-Planck equations, which means $rank(C) = n$ in our case. Afterwards we explain how the procedure is modified for the degenerate case, i.e. $rank(C) < n$ and state the final result at the end of the section.

3.1 Non-degenerate case

We start with the non-degenerate case which corresponds to an SDE of the form

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + C\,\mathrm{d}W_t. \tag{4}$$

where A fulfills Assumption 1 and $rank(C) = n$. We assume the Bakry-Émery criterion

$$\Sigma_\infty^{-1} \geq 2\lambda\ D^{-1} \tag{5}$$

to hold, where we introduced the shorthand notation $D = CC^T$.

Exponential decay to equilibrium in relative entropy then follows from the steps described in the sequel. In the first step the time derivative of relative entropy is computed

$$-I(\rho_t|\rho_\infty) := \frac{\mathrm{d}}{\mathrm{d}t}H(\rho_t|\rho_\infty) = -\frac{1}{2}\int \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right)^T D \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right)\ \rho_t\,\mathrm{d}x \leq 0. \tag{6}$$

The functional I is called *Fisher information*. One would like to find an estimate of form $-I(\rho_t|\rho_\infty) \leq -\lambda H(\rho_t|\rho_\infty)$ since integration of this inequality yields exponential convergence of $H(\rho_t|\rho_\infty)$ to zero with rate $\lambda > 0$. In the second step - aiming at finding such an estimate - the time derivative of the Fisher information is computed (for details see e.g. [AMTU01,ACJ08])

$$\frac{\mathrm{d}}{\mathrm{d}t}I(\rho_t|\rho_\infty) = -\frac{1}{2}\int \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right)^T D\Sigma_\infty^{-1}D \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right)\ \rho_t\,\mathrm{d}x - F, \tag{7}$$

where $F \geq 0$. The third step consists of applying the Bakry-Émery condition (5) to (7) which yields

$$\frac{\mathrm{d}}{\mathrm{d}t}I(\rho_t|\rho_\infty) \leq -\lambda \int \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right)^T D \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right)\ \rho_t = -2\lambda I(\rho_t|\rho_\infty). \tag{8}$$

Integrating the last inequality in time from 0 to t and using Gronwall's Lemma, we get exponential decay of the Fisher information $I(\rho_t|\rho_\infty) \leq e^{-2\lambda t} I(\rho_0|\rho_\infty)$. Integrating instead from t to ∞, using $-I = \mathrm{d}H/\,\mathrm{d}t$, we find

$$-I(\rho_t|\rho_\infty) \leq -2\lambda H(\rho_t|\rho_\infty)\,, \tag{9}$$

which is the sought inequality. Integration of (9) from 0 to t, Gronwall's Lemma then yields

$$H(\rho_t|\rho_\infty) \leq H(\rho_0|\rho_\infty)e^{-2\lambda t}\,. \tag{10}$$

3.2 Degenerate case

In the degenerate case, i.e. $rank(D) < n$ the usual Bakry-Émery condition (5) cannot hold since D is not invertible. Also, due to the rank deficiency of D, I is not strictly positive anymore but only positive semidefinite. Hence the decay in relative entropy may not be strictly monotonous, but can also exhibit plateaus.

In order to achieve strict monotonicity in the decay of relative entropy, the Fisher information I is replaced by a modified Fisher information S where the degenerate diffusion matrix D is replaced by a non-degenerate matrix $P > 0$

$$S(\rho_t|\rho_\infty) = \int \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right)^T P \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right) \rho_t \, \mathrm{d}x.$$

The key ingredients in order to obtain exponential decay in relative entropy are then to prove exponential decay of the functional $S(\rho_t|\rho_\infty)$ and to see that $P \geq \frac{c_P}{2} D$ for some positive constant c_P, by which the exponential decay of the Fisher information follows and hence (8).

In order to establish exponential decay of $S(\rho_t|\rho_\infty)$ its time derivative is computed, yielding (cf. [AE14, Prop. 4.5])

$$\frac{\mathrm{d}}{\mathrm{d}t} S(\rho_t|\rho_\infty) = -\int \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right)^T \left[QP + PQ^T\right] \left(\nabla \log \frac{\rho_t}{\rho_\infty}\right) \rho_t \, \mathrm{d}x - F_P, \quad (11)$$

where $Q := \Sigma_\infty A^T \Sigma_\infty^{-1}$ and $F_P \geq 0$. The result which replaces the Bakry-Émery criterion (5) is given in [AE14, Lem. 4.3] and is indispensable for the proof. It yields the existence of a positive definite matrix P such that

$$QP + PQ^T \geq 2\lambda P \quad (12)$$

where either $\lambda = \nu = \min\{\Re(\lambda) : -Av = \lambda v\} > 0$ if A is diagonalizable (i.e. if all eigenvalues have the same geometric and algebraic multiplicity) or $\lambda = \nu - \varepsilon$ for some $\varepsilon > 0$ if at least one eigenvalue is defective (i.e. if geometric and algebraic multiplicity do not agree). Equations (11) and (12) then take the role of (7) and (5), which yields the exponential decay of the functional S. Noting that we can find a constant c_P such that $P \geq \frac{c_P}{2} D$, it follows that the Fisher information decays exponentially, which entails the exponential decay of the relative entropy. The results are summarized in the following Theorem (cf. [AE14, Thm. 4.9]).

Theorem 1. *Consider the SDE* (1) *with associated Fokker-Planck equation* (2), *and let Assumptions 1 and 2 hold. Define* $\nu = \min\{\Re(\lambda) : -Av = \lambda v\} > 0$ *to be the smallest eigenvalue of* $-A$ *and suppose that* $H(\rho_0|\rho_\infty) < \infty$.

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

$$H(\rho_t|\rho_\infty) \leq cH(\rho_0|\rho_\infty)e^{-2\nu t} \quad \forall t \geq 0\,.$$

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

$$H(\rho_t|\rho_\infty) \leq c_\varepsilon H(\rho_0|\rho_\infty)e^{-2(\nu-\varepsilon)t} \quad \forall t \geq 0\,.$$

The actually observed relaxation behaviour is explored in section 5 where we investigate the influence of temperature and the choice of initial conditions. Further, we study the occurence of plateaus in the decay and processes with multiple time scales.

4 Small noise

If the principal eigenvalue of the drift matrix A in the SDE (1) is simple, then the Csiszár-Kullback-Pinsker inequality together with the entropy estimate in Theorem 1 implies that (see e.g. [BV05] and the references therein)

$$\|\rho_t - \rho_\infty\|_{L^1(\mathbb{R}^n)} \leq \sqrt{2cH(\rho_0|\rho_\infty)}\, e^{-\nu t}\,, \tag{13}$$

where ν is minus the real part of the principal eigenvalue of A and $c \geq 1$ is a constant. Note, however, that at low temperature (i.e. for small noise) the stationary distribution ρ_∞ of (1) shrinks to a point mass δ_0 concentrated at the origin $x = 0$, and as a consequence the upper bound in (13) degrades for most initial data.

In this case the above estimate may not be so informative, and spectral estimates come into play. Under the Assumptions 1 and 2 it is a known result from [MPP02] that the infinitesimal generator

$$L = \frac{1}{2}CC^T : \nabla^2 + (Ax)\cdot\nabla \tag{14}$$

associated with (1) has a compact resolvent and therefore a discrete spectrum in $L^p(\mathbb{R}^n, \rho_\infty)$ for $1 < p < \infty$ that can be completely characterized in terms of the eigenvalues of the matrix A; in particular, all eigenvalues of L have multiplicity 1 if and only if A is diagonalizable, and the eigenvalues are independent of p for $1 < p < \infty$. (For $p = 1$, the spectrum of L is the closed left-half plane [MPP02].)

For reversible systems with $A = A^T$ and C being a scalar multiple of the identity $I_{n\times n}$, the spectral properties of L imply exponential convergence of the weighted density $\eta_t = \rho_t/\rho_\infty$ to $\eta_\infty = \mathbf{1}$ in $L^2(\mathbb{R}^n, \rho_\infty)$. For this reason time scales in molecular dynamics are conventionally associated with the dominant eigenvalues of L (see e.g. [SS13]), but it is easy to see that this setting requires that $\eta_0 \in L^2(\mathbb{R}^n, \rho_\infty)$ which is equivalent to the assumption that ρ_0 is in $L^2(\mathbb{R}^n, \rho_\infty^{-1})$; even in the simple situation at hand, this is quite restrictive in that it excludes many standard cases, such as sharp Gaussian or point-like initial conditions.

4.1 Large deviations: exit from a set

Here we describe an alternative characterization of the speed of convergence that is based on large deviations arguments and that includes non-reversible systems. To this end, we scale the diffusion matrix according to

$$C \mapsto \sqrt{\beta^{-1}}C\,, \quad \beta > 0\,. \tag{15}$$

We are interested in the situation $\beta \gg 1$, and, specifically, we want to study the probability that the process $X_t = X_t^\beta$ leaves a bounded and open set O that contains the unique stable fixed point $x = 0$. Looking at exit probabilities and exit rates is motivated by the observation that the convergence of MCMC methods for sampling general multimodal distributions depends on the probability that the process leaves a basin of attraction [Liu04]; other applications in which characteristic time scales are associated with exit events are reliability analysis or risk assessment [MSB14]. Now let

$$\tau = \inf\{t > 0\colon X_t \notin O\} \tag{16}$$

be the first exit time from the set O. As a consequence of the Donsker-Varadhan large deviations principle (see, e.g., [DV75,Ell85]), the quantity

$$\gamma = -\lim_{t\to\infty} \frac{1}{t} \log P(\tau > t \,|\, X_0 = x) \tag{17}$$

is the principal eigenvalue of $-L$ equipped with homogeneous Dirichlet boundary values on ∂O that we assume to be smooth. That is, γ is the eigenvalue with the smallest real part that solves the eigenvalue problem

$$\begin{aligned} -L\varphi(x) &= \gamma\varphi(x)\,,\ x \in O \\ \varphi(x) &= 0\,, \qquad x \in \partial O\,. \end{aligned} \tag{18}$$

Note that $\varphi > 0$ in the interior of the domain as a consequence of the Perron-Frobenius theorem and Assumptions 1 and 2. Formally, the relationship (18) can be derived using the Feynman-Kac theorem for parabolic boundary value problems (e.g. [Øks03, Chapters 8–9]), with the separation ansatz

$$P(\tau > t|X_0 = x) \simeq \varphi(x)\exp(-\gamma t) \quad \text{as} \quad t \to \infty \tag{19}$$

that can be justified by the ergodicity of (1) under Assumptions 1 and 2; see also [SWH12]. This implies that for large t we have the asymptotics

$$P(\tau > t) \simeq e^{-\gamma t} \quad \text{as} \quad t \to \infty\,, \tag{20}$$

independent of the initial condition $X_0 = x$. The interpretation of the principal eigenvalue of reversible metastable systems is straighforward: the closer the principal eigenvalue $\gamma > 0$ is to zero, the smaller is the probability to observe an exit from the set O before time t, where the dependence is exponential in γ.

Remark 1. It can be shown that the exit probability is *exactly* exponential when the initial probability density for X_0 is the solution of the eigenvalue equation, with L being replaced by its formal L^2 adjoint L^*. The corresponding eigenfunction is called the *quasi-stationary distribution*, and it has the property that exit times are exponentially distributed, which is relevant in the context of parallelized molecular sampling algorithms [LBLLP12].

4.2 Small-noise approximation of the principal eigenvalue

We seek a computable and easily interpretable expression for γ, and we will argue that γ can be computed from the stationary covariance matrix Σ_∞. To this end, we exploit a specific stochastic control interpretation of the principal eigenvalue that is due to Fleming and co-workers (see [Fle77,FM95,FS97]). Specifically, using that the strong maximum principle implies that the function φ in (18) is strictly positive in the interior of the domain, it follows that $v = -\beta^{-1}\log\varphi$ solves the nonlinear boundary value problem

$$Lv - \frac{1}{2}|\nabla v|^2_{CC^T} = \gamma/\beta \tag{21}$$

with $|w|_{CC^T} = \sqrt{w^T CC^T w}$ denoting a weighted Euclidean pseudo-norm with weight $CC^T \geq 0$ and the specification

$$v(x) \to \infty \quad \text{as} \quad \mathrm{dist}(x, \partial O) \to 0 \tag{22}$$

for the function v when its arguments approach the boundary of O. Noting that

$$-\frac{1}{2}|w|^2_{CC^T} = \min_{a\in\mathbb{R}^n}\left\{\frac{1}{2}|a|^2 + (Ca)\cdot w\right\}, \tag{23}$$

we observe that (21) is the dynamic programming equation of an ergodic stochastic control problem, which implies the following result (cf. [SWH12]):

Proposition 2. *Under the previous assumptions, it holds that a.s.*

$$\gamma = \min_{u\in\mathcal{U}}\lim_{T\to\infty}\frac{\beta}{T}\mathbb{E}\left(\frac{1}{2}\int_0^T |u_t|^2\,\mathrm{d}t - \log \mathbf{1}_{\{\tau>T\}}\right) \tag{24}$$

where $\tau = \tau^u$ is the first exit time of the set O under the controlled process

$$\mathrm{d}X_t^u = (Cu_t + AX_t^u)\mathrm{d}t + \sqrt{\beta^{-1}}C\,\mathrm{d}W_t. \tag{25}$$

and the minimization is over all Markovian controls $u \in \mathcal{U}$ such that (25) *has a unique strong solution. Furthermore the minimum is unique and attained at $u_t^* = \beta^{-1}C^T\nabla\log\varphi(X_t^{u^*})$ with $\varphi \in C^2(O)\cap C(\overline{O})$ being the solution of* (18).

In the limit $\beta \to \infty$, the corresponding dynamic programming equation (21) can be explicitly solved and it can be shown [FW12,EK87] that

$$\gamma \simeq e^{-\beta R} \quad \text{as} \quad \beta \to \infty\,. \tag{26}$$

Here $R = \min\{\Phi(x)\colon x \in \partial O\}$ is the large deviations rate function, with

$$\Phi(x) = \lim_{T\to\infty}\min_u\left\{\frac{1}{2}\int_0^T |u(t)|^2\,\mathrm{d}t : y(0) = 0,\, y(T) = x\right\} \tag{27}$$

being the associated quasi-potential and $y(t) = y(t; t_0, y_0)$ being the solution of

$$\dot{y}(t) = Ay(t) + Cu(t)\,, \quad y(t_0) = y_0\,. \tag{28}$$

Note that (27)–(28) is the deterministic counterpart to the stochastic control problem Proposition 2 with $\beta = 1$. (The large parameter β has been absorbed in the exponent in (26). The dynamic programming equation corresponding to (27) then reads

$$(Ax)\cdot\nabla\Phi - \frac{1}{2}|\nabla\Phi|^2_{CC^T} = 0\,, \quad \Phi(0) = 0\,, \tag{29}$$

where Φ is—in contrast to the solution of the dynamic programming equation (21)—bounded on $\overline{O}$, as a consequence of the complete controllability of the control system (28). A simple calculation shows that

$$\Phi(x) = \frac{1}{2}x^T\Sigma_\infty^{-1}x\,, \tag{30}$$

with $\Sigma_\infty \in \mathbb{R}^{n\times n}$ being the unique symmetric and positive definite solution of the Lyapunov equation $A\Sigma_\infty + \Sigma_\infty A^T + CC^T = 0$. The next statement is a straight consequence of the previous considerations (cf. [Zab85]):

Corollary 1. *If $O = \{x \in \mathbb{R}^n : |x| < 1\}$, then*

$$\lim_{\beta\to\infty} \beta^{-1}\log\gamma = -(2\Lambda)^{-1}\,, \tag{31}$$

where $\Lambda > 0$ is the largest eigenvalue of the asymptotic covariance matrix Σ_∞.

We can interpret this result as follows: Recalling the large deviations principle (20) for large t, we can conclude that the probability of observing an exit from the n-dimensional unit sphere before time t behaves like

$$P(\tau \le t) \simeq 1 - \exp\left(-t\exp\left(-\frac{\beta}{2\Lambda}\right)\right)\,, \quad t,\beta\to\infty \tag{32}$$

in the low-temperature regime. In other words, the probability of observing an exit before time t is large whenever the system is "easily controllable" (i.e. has large variance in some direction), whereas it is small if the system is "hardly controllable" (i.e. has uniformly small variance in all degrees of freedom).

4.3 Mean first exit times

Even though our asymptotics is not very precise, Corollary 1 has an interesting interpretation in terms of mean first exit times that is due to Zabczyk. In [Zab85, Thm. 6] it is shown that the mean first exit time satisfies

$$\mathbb{E}[\tau] \simeq \exp\left(\frac{\beta}{2\Lambda}\right)\,, \quad \beta\to\infty\,, \tag{33}$$

which implies that

$$P(\tau \le t) \simeq 1 - \exp\left(-\frac{t}{\mathbb{E}(\tau)}\right), \quad t, \beta \to \infty\,. \tag{34}$$

For reversible diffusions with $A = A^T$ and $C = I_{n\times n}$, the largest non-zero eigenvalue λ_1 of the operator L with Dirichlet boundary data satisfies $\lambda_1 \simeq -(\mathbb{E}(\tau))^{-1}$ as $\beta \to \infty$, and thus the last equation is consistent with the well-known interpretation of the principal eigenvalue of a reversible diffusion as an exponential decay rate; for details we refer to [BGK05].

The following Proposition establishes a relation between the two exponential time scales that are determined by $\nu = \min\{\Re(\lambda) : -Av = \lambda v\}$ and $\Lambda = \max\{\lambda : \Sigma_\infty v = \lambda v\}$; cf. [DSC96, Lem. 3.1] and Remark 2 below.

Proposition 3. *Let A and C fulfill Assumptions 1 and 2, and let Σ_∞ solve the corresponding Lyapunov equation* (3) *for $\beta = 1$. Let w be the normalized eigenvector of $-A$ which corresponds to ν, i.e. $-Aw = \nu w$. Introduce the splitting of $w = w_{\mathrm{Ker}} + w_{\mathrm{Im}}$, where $w_{\mathrm{Ker}} \in ker(D)$, $D = CC^T$ and $w_{\mathrm{Im}} \in Im(D)$. Further denote by $\lambda_{\min}(D)$ the smallest non-zero eigenvalue of D. Then*

$$\nu \ge \frac{\lambda_{\min}(D)}{2\Lambda}|w_{\mathrm{Im}}|. \tag{35}$$

Proof. First note, that due to Assumption 2 we have $w_{\mathrm{Im}} \neq 0$, but $w_{\mathrm{Ker}} = 0$ is possible. Multiplying the Lyapunov equation $A\Sigma + \Sigma A^T = -D$ from the left and right by w^T and w we find that $2\nu w^T \Sigma w = w^T D w$.
Now, $w^T \Sigma w \le \Lambda$ and $w^T D w = w_{\mathrm{Im}}^T D w_{\mathrm{Im}} \ge \lambda_{\min}(D)|w_{\mathrm{Im}}|$ which together yields the assertion.

Remark 2. The inequality (35) is sharp. In the reversible case with $A = A^T$ negative definite and $C = I_{n\times n}$, the matrix $A = -\nabla^2 V$ can be interpreted as the Hessian of a quadratic potential

$$V(x) = -\frac{1}{2}x^T A x\,, \tag{36}$$

such that the stationary distribution ρ_∞ of (1) has a density proportional to $\exp(-V/2)$. As a consequence, the Lyapunov equation $A\Sigma + \Sigma A + I_{n\times n} = 0$ has an explicit solution Σ with $2\Sigma = -A^{-1}$ and $2\nu = 1/\Lambda$. In view of the exponential bound in Theorem 1 and the relation $\lambda_1 \simeq -(\mathbb{E}(\tau))^{-1}$ for reversible diffusions, (35) furnishes the known connection between the principal eigenvalue of a finite Markov chain and the logarithmic Sobolev constant of the corresponding invariant measure (see [DSC96, Lem. 3.1]).

5 Numerical Examples

We will restrict ourselves to two and three dimensional examples, since the intention here is to only illustrate certain characteristics of the convergence behaviour of the system. As in the previous section we consider processes described by a SDE of the general form

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + \sqrt{\beta^{-1}}C\,\mathrm{d}W_t \tag{37}$$

where A and C fulfill the necessary Assumptions 1 and 2.

We will first discuss general dependencies for a given system (A, C) with respect to the temperature and initial conditions. Further, we shortly discuss the occurence of plateaus and the case where the system has multiple time scales.

5.1 Dependencies on the temperature and initial conditions

In order to study the influence which temperature has on our system we split up the relative entropy into three different terms of which the first two correspond to the relaxation of the covariance and the last one to the relaxation of the mean.

$$\begin{aligned} H(t) &:= \int \log\left(\frac{\rho_t}{\rho_\infty}\right)\rho_t\,\mathrm{d}x \\ &= \frac{1}{2}\left[\underbrace{\mathrm{Tr}(\Sigma_t\Sigma_\infty^{-1}) - n}_{=a(t)} + \underbrace{-(\log\det(\Sigma_t\Sigma_\infty^{-1}))}_{=b(t)} + \underbrace{\mu_t^T\Sigma_\infty^{-1}\mu_t}_{=c(t)}\right]. \end{aligned}$$

Specifically, we can interpret these terms in the following sense. The term a embodies the relaxation of the covariance Σ_t to Σ_∞, whereas b ensures the normalization of the densities and finally c comprises the relaxation of the mean μ_t to 0.

We consider the following example where drift and diffusion are given by

$$A = -\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix},\ C = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

In Figure 1 we illustrate temperature-effects. Recall that for a process given by (37) and $\Sigma_0 = 0$, the covariance Σ_t at time t is given by

$$\Sigma_t = \beta^{-1}\int_0^t e^{A(t-s)}CC^Te^{A^T(t-s)}\,\mathrm{d}s. \tag{38}$$

Hence, the only term which is temperature dependence is $c(t)$, due to Σ_∞^{-1}. Furthermore $c(t)$ grows as $\beta \to \infty$, i.e. for small temperatures $c(t)$ dominates the relaxation behaviour. The terms $a(t)$ and $b(t)$ do not have any temperature dependence because of the multiplication by another temperature dependent term Σ_t, such that β and β^{-1} cancel. This means that for larger temperatures the relaxation is governed by a and b, which describe the equilibration of the covariance (see Figure 1 right panel and Figure 2 upper right panel). Note that a and b always have opposite signs, and they change sign depending on whether

$\Sigma_t > \Sigma_\infty$ or $\Sigma_t < \Sigma_\infty$. In the first case, i.e. when $\Sigma_t > \Sigma_\infty$, $a(t)$ is strictly positive and contributes the most, in the second case it is $b(t)$ that dominates. For small temperature $c(t)$ plays this role (see Figure 1 left panel), and the overall relaxation is mainly determined by the relaxation of the mean.

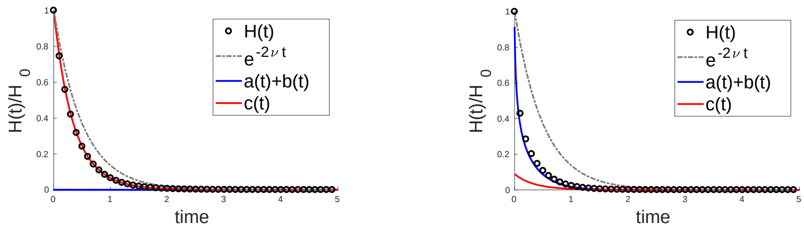

Fig. 1. Temperature-effects: the initial conditions are fixed. Left: low temperature ($\beta = 10^3$), right: high temperature ($\beta = 10^{-2}$).

Figure 2 shows the strong influence of the initial conditions on the relaxation behaviour, which is the only parameter varied in this figure. If we choose an eigenvector of the drift matrix A as a deterministic initial condition, this will yield exponential decay with the corresponding eigenvalue (left panel). The initial conditions can also be chosen such that one observes a plateau where $\dot{H}(t) = 0$ (upper right and lower panel). This also leads to the constant c of Theorem 1 being strictly greater than 1. Which term contributes the most to the total relaxation behaviour then depends on the choice of the initial covariance Σ_0. If $\Sigma_0 > \Sigma_\infty$ then $a(t)$ is the governing term (if the temperature is not too low), otherwise $c(t)$ will take this role.

Remark on the occurence of plateaus We shortly discuss the occurence of plateaus here. Relaxation to the equilibrium state pauses if and only if the time derivative of $H(t)$ vanishes (i.e. when $\dot{H}(t) = 0$), which is equivalent to the Fisher information being zero. Indeed, this can only happen if the diffusion is degenerate, for otherwise the Fisher information is strictly positive due to (6). Still, we can find regimes that are almost plateaus, even though the diffusion is non-degenerate.

Let us consider the example given by

$$A = -\begin{pmatrix} 1 & 18 \\ 0 & 10 \end{pmatrix}, C_{\text{deg}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We can choose the initial conditions such that in the degenerate case $\dot{H}(t)|_{t=0.25} = 0$ and in the non-degenerate case $\dot{H}(t)|_{t=0.25} \approx 0$, and we find that we almost cannot distinguish between the two cases, see Figure 3.

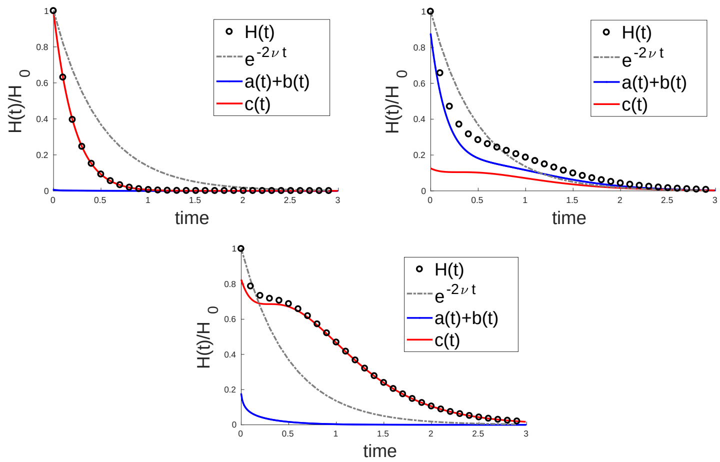

Fig. 2. Influence of the initial conditions when the temperature is fixed $\beta = 20$. Upper left: $x_0 = (\frac{20}{3}, \frac{10}{3})^T$ (eigenvector of A corresponding to $\lambda = 2$); upper right: $x_0 \sim \mathcal{N}((1.2840, -0.9023)^T, \Sigma_0)$, $\Sigma_0 > \Sigma_\infty$; lower: $x_0 = (1.2840, -0.9023)^T, \Sigma_0 = 0$.

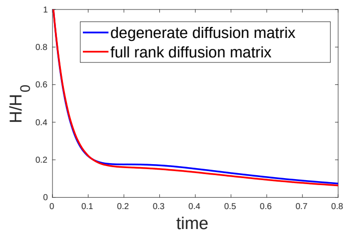

Fig. 3. Comparison of the relaxation behaviour for two processes with the same drift and degenerate or non-degenerate diffusion.

5.2 Multiple time scales: partitioning into slow and fast

We now split up the relative entropy into two terms, where one depends on conditional distributions and the other on marginal ones. More specifically, consider a process $(Z_t)_{(t\geq 0)}$ which consists of two components $Z = (X, Y)$. We will think of X being the slow component and Y the fast one. Denote by $\rho(z)$ the density of the joint process, by $\bar{\rho}(x)$ the marginal density of X and by $\hat{\rho}(y; x)$ the conditional density of Y where $X = x$ is given. We can always do the following computation which yields a partition of the relative entropy into conditional and marginal terms:

$$\begin{aligned}H_Z(t) :&= H(\rho_t|\rho_\infty)\\&= \int\int \bar\rho_t\ \hat\rho_t\ \log\left(\frac{\bar\rho_t}{\bar\rho_\infty}\right)\mathrm{d}y\,\mathrm{d}x + \int\int \bar\rho_t\ \hat\rho_t\ \log\left(\frac{\hat\rho_t}{\hat\rho_\infty}\right)\mathrm{d}y\,\mathrm{d}x\\&= H(\bar\rho_t|\bar\rho_\infty) + \int H(\hat\rho_t|\hat\rho_\infty)\bar\rho_t\,\mathrm{d}x\\&= H_X(t) + \mathbb{E}_{\bar\rho_t}(H_{Y|X=x}(t)).\end{aligned}$$

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.

From Langevin to overdamped Langevin The example we consider here is given by a Langevin equation with a time scale parameter $0 < \varepsilon \le 1$ (and all constants set equal to one). The coefficients read

$$A = \begin{pmatrix} 0 & -\varepsilon^{-1}\\ \varepsilon^{-1} & \varepsilon^{-2}\end{pmatrix}, C = \begin{pmatrix}0\\ \varepsilon^{-1}\end{pmatrix}.$$

Note that for $\varepsilon \to 0$, the first component of the dynamics approaches the overdamped Langevin equation [Pav14].

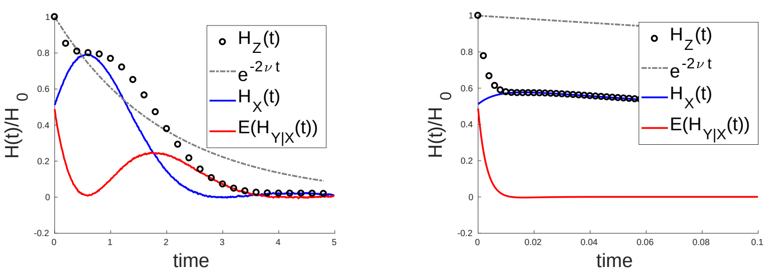

Fig. 4. Decay of the relative entropy for Langevin dynamics and no time scale separation, i.e. $\varepsilon = 1$ (left) and $\varepsilon = 0.08$ (right) and initial condition $z_0 = -(5,5)^T$.

No time scale separation This case is depicted in Figure 4 in the left panel and we see that $c > 1$ again. Additionally, the conditional and marginal terms are not monotonically decreasing in time, but can in fact increase. This is due to the fact that when computing the time derivative of e.g. $H(\bar\rho_t|\bar\rho_\infty)$ one finds as usually the Fisher information, but furthermore another term appears which can be estimated by the *empirical measure large deviations rate functional*; see [Sha17, Ch. 2]. The empirical measure large deviations rate functional of $\bar\rho_t$ and $\bar\rho_\infty$ will in general be non-zero, since the time evolution of $\bar\rho_t$ is still described

by the Fokker-Planck equation of the full process which differs from the Fokker-Planck equation which has $\bar{\rho}_\infty$ as equilibrium.
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.

Time scale separation We introduce a time scale separation by setting $\varepsilon = 0.08$ (see Figure 4 right panel) ands 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: for $\varepsilon \to 0$ the conditional term $H_{Y|X=x}(t)$ relaxes almost instantaneously to its equilibrium. Accordingly, the marginal term $H_X(t)$ governs the long term behaviour of the overall 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.
Furthermore, we observe that as $\varepsilon \to 0$ both terms, conditional and marginal, become monotonically decreasing.

6 Outlook and Discussion

In the previous section we have seen that relative entropy may be used as a tool to define fast and slow degrees of freedom. That is, the fast variable is defined via an almost immediate relaxation of its conditional density $\hat{\rho}_t$, with the slow variable being fixed. At the same time the slow variable is defined via the comparatively slow relaxation of its marginal density $\bar{\rho}_t$. Furthermore, the slow variable will govern the collective relaxation after very short time once the fast one has relaxed. This definition of fast and slow agrees with the coarse-graining concepts of averaging and homogenization in the reversible case and the conditional expectation (cf. [LL10]) in the general case. These methods seek low dimensional effective dynamics which are built by computing expectations of the slow variables' dynamics with respect to the conditional invariant distributions of the fast variables given the slow ones. The underlying idea is that the fast variables relax almost instantaneously such that their force on the dynamics is well captured by the statistics of their invariant distribution.

Moreover, being able to identify fast and slow degrees of freedom in the linear case brings up the question whether this can be extended to a more general, i.e. non-linear setting. In the case of a reversible diffusion described by

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{\beta^{-1}}\,\mathrm{d}W_t \tag{39}$$

with V being a confinement potential which grows sufficiently fast at infinity, it is known that, in the small temperature limit, the slowest processes—given by mean first passage times across the highest energy barriers—can be associated with the eigenvalues of the generator in a hierarchical manner [BGK05]. Furthermore, the eigenvalues describe the convergence to equilibrium in the ρ_∞^{-1} weighted L^2 norm. This analysis is inherent to the reversible case but a natural generalization of convergence to equilibrium in L^2 is given by convergence in relative entropy. Hence, the question of whether one is able to formulate a hierarchical ordering of the systems processes according to the relaxation time scales determined by the convergence in relative entropy—for reversible as well as irreversible processes—is reasonable (cf. [MS14, Remark 2.16])

Another topic that we have not touched here is the acceleration of the convergence to equilibrium by changing the drift and diffusion. Improving the convergence speed by making a reversible process irreversible, such that the equilibrium distribution is preserved has been intensively discussed in the last few years, see e.g. [LNP13], [GM16] or [KJZ17]. In [GM16] the eigenvalues of the drift matrix are optimized by changing both drift and diffusion while keeping the total amount of randomness, i.e. $\mathrm{Tr}(D)$ fixed. In [LNP13], in contrast, a reversible diffusion of the form (39) with V being quadratic is altered by adding a skew-symmetric drift term. We want to emphasize that the skew-symmetric term can be interpreted as a control as in (25) where accelerating the dynamics by minimizing the real part of the eigenvalue of the drift matrix then corresponds to what is called a *stabilization problem* in control theory (see e.g. [Zab09, Ch. 2]). An alternative is to apply a control so as to maximise the probability of making large excursions or an exit from a metastable set, without changing the stationary distribution. Similar ideas have been proposed by Pavon and Ticozzi [PT06], and we leave the application to numerical sampling problems for future work.

Acknowledgements. This research has been partially funded by Deutsche Forschungsgemeinschaft (DFG) through the grant CRC 1114 "Scaling Cascades in Complex Systems", Projects A05 "Probing scales in equilibrated systems by optimal nonequilibrium forcing" and B05 "Origin of the scaling cascades in protein dynamics".

References

[ACJ08] Anton Arnold, Eric Carlen, and Qiangchang Ju. Large-time behavior of non-symmetric Fokker-Planck type equations. *Communications on Stochastic Analysis*, 2(1):153–175, 2008.

[AE14] 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.

[AMTU01] 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.

[Ant05] Athanasios C. Antoulas. *Approximation of large-scale dynamical systems.* SIAM, 2005.

[BdH15] A. Bovier and F. den Hollander. *Metastability: A Potential-Theoretic Approach*, volume 351 of *Grundlehren der mathematischen Wissenschaften.* Springer International, 2015.

[BE85] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In J. Azéma and M. Yor, editors, *Séminaire de Probabilités XIX 1983/84*, volume 1123 of *Lecture Notes in Mathematics*, pages 177–206. Springer, Berlin, Heidelberg, 1985.

[Ber13] N. Berglund. Kramers' law: Validity, derivations and generalisations. *Markov Process Relat.*, 19(3):459–490, 2013.

[BGK05] Anton Bovier, Véronique Gayrard, 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.

[BV05] François Bolley and Cédric Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des Sciences de Toulouse: Mathématiques*, 14(3):331–352, 2005.

[CC96] Mary Kathryn Cowles and Bradley P. Carlin. Markov chain monte carlo convergence diagnostics: A comparative review. *J. Am. Stat. Assoc.*, 91(434):883–904, 1996.

[CT15] Raphaël Chetrite and Hugo Touchette. Nonequilibrium markov processes conditioned on large deviations. *Annales Henri Poincaré*, 16(9):2005–2057, 2015.

[Day87] Martin V. Day. Recent progress on the small parameter exit problem. *Stochastics: An International Journal of Probability and Stochastic Processes*, 20(2):121–150, 1987.

[DSC96] P. Diaconis and L. Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

[DV75] Monroe D. Donsker and S.R. Srinivasa Varadhan. On a variational formula for the principal eigenvalue for operators with maximum principle. *Proceedings of the National Academy of Sciences*, 72(3):780–783, 1975.

[DW04] P. Dupuis and H. Wang. Importance sampling, large deviations, and differential games. *Stochastics and Stochastic Rep.*, 76(6):481–508, 2004.

[EK87] Alexander Eizenberg and Yuri Kifer. The asymptotic behavior of the principal eigenvalue in a singular perturbation problem with invariant boundaries. *Probability theory and related fields*, 76(4):439–476, 1987.

[Ell85] Richard S. Ellis. *Entropy, Large Deviations and Statistical Mechanics.* Springer, Berlin, 1985.

[Fle77] Wendell H. Fleming. Exit probabilities and optimal stochastic control. *Applied Mathematics & Optimization*, 4:329–346, 1977.

[FM95] Wendell H. Fleming and William M. McEneaney. Risk-sensitive control on an infinite time horizon. *SIAM Journal on Control and Optimization*, 33(6):1881–1915, 1995.

[FS97] Wendell H Fleming and Shuenn-Jyi Sheu. Asymptotics for the principal eigenvalue and eigenfunction of a nearly first-order operator with large potential. *Ann. Probab.*, pages 1953–1994, 1997.

[FW12] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260. Springer, 2012.

[GM16] Arnaud Guillin and Pierre Monmarché. Optimal linear drift for the speed of convergence of an hypoelliptic diffusion. *Electron. Commun. Probab.*, 21, 2016.

[HMS04] Wilhelm Huisinga, Sean Meyn, and Christof Schütte. Phase transitions and metastability in Markovian and molecular systems. *Ann. Appl. Probab.*, 14(1):419–458, 2004.

[HS06] Wilhelm Huisinga and Bernd Schmidt. Metastability and dominant eigenvalues of transfer operators. In Benedict Leimkuhler, Christophe Chipot, Ron Elber, Aatto Laaksonen, Alan Mark, Tamar Schlick, Christoph Schütte, and Robert Skeel, editors, *New Algorithms for Macromolecular Simulation*, pages 167–182. Springer, Berlin, Heidelberg, 2006.

[HSZ16] Carsten Hartmann, Christof Schtte, and Wei Zhang. Model reduction algorithms for optimal control and importance sampling of diffusions. *Nonlinearity*, 29(8):2298–2326, 2016.

[HTB90] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251–341, 1990.

[IVS01] Peter Imkeller and Jin-Song Von Storch. *Stochastic Climate Models*. Progress in Probability. Springer, New York, 2001.

[JS14] Robert L. Jack and Peter Sollich. Large deviations of the dynamical activity in the east model: analysing structure in biased trajectories. *J. Phys. A*, 47(1):015003, 2014.

[KJZ17] Marcus Kaiser, Robert L. Jack, and Johannes Zimmer. Acceleration of convergence to equilibrium in Markov chains by breaking detailed balance. *J. Stat. Phys.*, 168(2):259–287, 2017.

[LB17] Valerio Lucarini and Tamás Bódai. Edge states in the climate system: exploring global instabilities and critical transitions. *Nonlinearity*, 30(7):R32, 2017.

[LBLLP12] Claude Le Bris, Tony Lelièvre, Mitchell Luskin, and Danny Perez. A mathematical formalization of the parallel replica dynamics. *Monte Carlo Methods Appl.*, 18(2):119–146, 2012.

[Liu04] Jun S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2004.

[LL10] Frédéric Legoll and Tony Lelièvre. Effective dynamics using conditional expectations. *Nonlinearity*, 23(9):2131–2163, September 2010.

[LNP13] Tony Lelièvre, Francis Nier, and Grigorios A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.*, 152(2):237–274, 2013.

[LS16] Tony Lelièvre and Gabriel Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016.

[MPP02] 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.

[MS14] Georg Menz and André Schlichting. Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.*, 42(5):1809–1884, 2014.

[MSB14] Sebastián Martorell, Carlos Guedes Soares, and Julie Barnett. *Safety, Reliability and Risk Analysis: Theory, Methods and Applications*. CRC Press, 2014.

[MV00] Peter A. Markowich and Cédric Villani. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp*, 19:1–29, 2000.

[NH05] F. Nier and B. Helffer. *Hypoelliptic Estimates and Spectral Theory for Fokker-Planck Operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer, Berlin, Heidelberg, 2005.

[Øks03] Bernt K. Øksendal. *Stochastic Differential Equations: An Introduction With Applications*. Springer, 2003.

[Pav14] Grigorios A. Pavliotis. *Stochastic processes and applications*, volume 60 of *Texts in Applied Mathematics*. Springer, 2014.

[PT06] Michele Pavon and Francesco Ticozzi. On entropy production for controlled Markovian evolution. *J. Math. Phys.*, 47(6):063301, 2006.

[RBS15] Luc Rey-Bellet and Konstantinos Spiliopoulos. Irreversible langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7):2081–2104, 2015.

[Sha17] Upanshu Sharma. *Coarse-graining of Fokker-Planck equations*. PhD thesis, Technische Universiteit Eindhoven, 2017.

[SS13] Chrisof Schütte and Marco Sarich. *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*, volume 24 of *Courant Lecture Notes*. American Mathematical Society, 2013.

[SWH12] Christof Schütte, Stefanie Winkelmann, and Carsten Hartmann. Optimal control of molecular dynamics using Markov state models. *Math. Prog. Ser. B*, 134:259–282, 2012.

[Ven73] Alexander D. Ventsel. Formulae for Eigenfunctions and Eigenmeasures Associated with a Markov Process. *Theory Probab. Appl.*, 18(1):1–26, 1973.

[VEW12] Eric Vanden-Eijnden and Jonathan Weare. Rare event simulation of small noise diffusions. *Commun. Pure Appl. Math.*, 65(12):1770–1803, 2012.

[VF70] Alexander D. Ventsel' and Mark I. Freidlin. On small random perturbations of dynamical systems. *Russian Mathematical Surveys*, 25(1):1–55, 1970.

[Zab85] Jerzy Zabczyk. Exit problem and control theory. *Systems & control letters*, 6(3):165–172, 1985.

[Zab09] Jerzy Zabczyk. *Mathematical control theory: an introduction*. Springer Science & Business Media, 2009.