Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Physica A 365 (2006) 360–382

PHYSICA A

www.elsevier.com/locate/physa

Noise and conditional entropy evolution

Michael C. Mackey[a], Marta Tyran-Kamińska[b,*]

[a]*Departments of Physiology, Physics & Mathematics and Centre for Nonlinear Dynamics, McGill University, 3655 Promenade Sir William Osler, Montreal, QC, H3G 1Y6, Canada*
[b]*Institute of Mathematics, Silesian University, ul. Bankowa 14, 40-007 Katowice, Poland*

Received 27 July 2005; received in revised form 27 September 2005
Available online 25 October 2005

Abstract

We study the convergence properties of the conditional (Kullback–Leibler) entropy in stochastic systems. We have proved general results showing that asymptotic stability is a necessary and sufficient condition for the monotone convergence of the conditional entropy to its maximal value of zero. Additionally we have made specific calculations of the rate of convergence of this entropy to zero in a one-dimensional situation, illustrated by Ornstein–Uhlenbeck and Rayleigh processes, higher dimensional situations, and a two-dimensional Ornstein–Uhlenbeck process with a stochastically perturbed harmonic oscillator and colored noise as examples. We also apply our general results to the problem of conditional entropy convergence in the presence of dichotomous noise. In both the one-dimensional and multidimensional cases we show that the convergence of the conditional entropy to zero is monotone and at least exponential. In the specific cases of the Ornstein–Uhlenbeck and Rayleigh processes, as well as the stochastically perturbed harmonic oscillator and colored noise examples, we obtain exact formulae for the temporal evolution of the conditional entropy starting from a concrete initial distribution. The rather surprising result in this case is that the rate of convergence of the entropy to zero is independent of the noise amplitude.

Keywords: Markov operator; Asymptotic stability; Conditional entropy

1. Introduction

This paper examines the role of noise in the evolution of the conditional (or Kullback–Leibler or relative) entropy to a maximum. We were led to examine this problem because it is known that in invertible systems (e.g. measure preserving systems of differential equations or invertible maps) the conditional entropy is fixed at the value with which the system is prepared [1–4], but that the addition of noise can reverse this invertibility property and lead to an evolution of the conditional entropy to a maximum value of zero. Here, we make both general and concrete specific calculations to examine the entropy convergence. We carry this out by studying the convergence properties of the Fokker–Planck equation using 'entropy methods' [5], which have been known for some time to be useful for problems involving questions related to convergence of solutions in

*Corresponding author.
E-mail addresses: mackey@cnd.mcgill.ca (M.C. Mackey), mtyran@us.edu.pl (M. Tyran-Kamińska).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2005.10.001

partial differential [6–12]. Their utility can be traced, in some instances, to the fact that entropy may serve as a Liapunov functional or play a role in proving Sobolev-type inequalities [13,25]. Studies of the convergence properties of entropy have attracted a large number of investigators in a variety of fields other than partial differential equations, e.g. in the dynamic behavior of Markov chains. A partial survey of some of these results can be found in Refs. [14,15].

There are a variety of different definitions of 'entropy' in the physics, information theory, and probability theory literature which often leads to confusion. For an illuminating discussion of these different definitions, [14, Chapter 2] should be consulted. Here, we use the term 'conditional entropy' as defined in Eq. (2.1) to be consistent with our previous work [3,16].

The outline of the paper is as follows. Section 2 introduces the dynamic concept of asymptotic stability and the notion of conditional entropy. (Asymptotic stability is a strong convergence property of ensembles which implies mixing. Mixing, in turn, implies ergodicity.) This is followed by two main results connecting the convergence of the conditional entropy with asymptotic stability (Theorem 1), and the existence of unique stationary densities with the convergence of the conditional entropy (Theorem 2). Section 3 shows that asymptotic stability is a property that cannot be found in an invertible deterministic system (e.g. a system of ordinary differential equations), and consequently that the conditional entropy does not have a dynamic (time dependent) character for this type of invertible dynamics. Section 4 considers the stochastic extension where a system of ordinary equations is perturbed by Gaussian white noise (thus becoming non-invertible) and gives general results showing that in this stochastic case asymptotic stability holds. Section 4.1 considers a one-dimensional situation, and we show that the conditional entropy convergence to zero is at least exponential. We consider the Ornstein–Uhlenbeck process in Section 4.1.1 and a Rayleigh process in Section 4.1.2. We consider multidimensional stochastic systems with nondegenerate noise in Section 4.2 and indicate when the exponential convergence of the entropy holds. Examples of higher dimensional situations with degenerate noise are considered within the context of a two-dimensional Ornstein–Uhlenbeck process in Section 4.3 with specific examples of a stochastically perturbed harmonic oscillator (Section 4.3.1) and colored noise (Section 4.3.2) as examples. Section 5 applies our general results to the problem of conditional entropy convergence in the presence of dichotomous noise. The paper concludes with a short discussion.

2. Asymptotic stability and conditional entropy

Let $(X, \mathscr{B}, \mu)$ be a σ-finite measure space. Let $\{P^t\}_{t\geqslant 0}$ be a semigroup of Markov operators on $L^1(X,\mu)$, i.e., $P^t f \geqslant 0$ for $f \geqslant 0$, $\int P^t f(x)\mu(\mathrm{d}x) = \int f(x)\mu(\mathrm{d}x)$, and $P^{t+s}f = P^t(P^s f)$. If the group property holds for $t, s \in \mathbf{R}$, then we say that P is invertible, while if it holds only for $t, s \in \mathbf{R}^+$ we say that P is non-invertible. We denote the corresponding set of densities by $\mathscr{D}(X,\mu)$, or $\mathscr{D}$ when there will be no ambiguity, so $f \in \mathscr{D}$ means $f \geqslant 0$ and $\|f\|_1 = \int_X f(x)\mu(\mathrm{d}x) = 1$. We call a semigroup of Markov operators P^t on $L^1(X,\mu)$ *asymptotically stable* if there is a density f_* such that $P^t f_* = f_*$ for all $t > 0$ and for all densities f

$$\lim_{t\to\infty} \|P^t f - f_*\|_1 = 0.$$

The density f_* is called a stationary density of P^t.

We define the conditional entropy (also known as the Kullback–Leibler or relative entropy) of two densities $f, g \in \mathscr{D}(X,\mu)$ as

$$H_c(f|g) = -\int_X f(x) \log \frac{f(x)}{g(x)} \mu(\mathrm{d}x). \tag{2.1}$$

Observe that $H_c(f|g) \leqslant 0$ for any two densities f, g. Our first result connects the temporal convergence properties of H_c with those of P^t.

The proof of the following theorem is much simpler than found in [3, Theorem 7.7], which in turn generalizes Theorems 9.3.2 and 9.4.2 of Ref. [16].

Theorem 1. *Let P^t be a semigroup of Markov operators on $L^1(X,\mu)$ and f_* be a positive density. If*

$$\lim_{t\to\infty} H_c(P^t f|f_*) = 0 \tag{2.2}$$

for a density f then

$$\lim_{t\to\infty}\|P^t f - f_*\|_1 = 0. \tag{2.3}$$

Conversely, if $P^t f_* = f_*$ *for all* $t \geqslant 0$ *and Condition* 2.3 *holds for all f with* $H_c(f|f_*) > -\infty$, *then Condition* 2.2 *holds as well.*

Proof. To prove the first part, note that for densities $f, g \in \mathcal{D}$, we have from the Csiszár–Kullback inequality

$$-\|f-g\|_1^2 \geqslant 2H_c(f|g).$$

In particular,

$$-\|P^t f - f_*\|_1^2 \geqslant 2H_c(P^t f|f_*).$$

Thus if

$$\lim_{t\to\infty} H_c(P^t f|f_*) = 0,$$

then

$$\lim_{t\to\infty}\|P^t f - f_*\|_1 = 0.$$

Proof of the converse portion is not so straightforward. Assume initially that f/f_* is bounded, so $0 \leqslant f \leqslant af_*$. Since $P^t f_* = f_*$ and P^t is a positivity preserving operator, we have $0 \leqslant P^t f \leqslant af_*$. Making use of the inequality

$$-H_c(f|g) \leqslant \int \left(\frac{f}{g} - 1\right)^2 g\mu(\mathrm{d}x), \tag{2.4}$$

valid for all densities f, g (for the proof of inequality (2.4) see 4.1), we arrive at

$$\begin{aligned}-H_c(P^t f|g) &\leqslant \int \left|\frac{P^t f}{f_*} - 1\right| |P^t f - f_*|\, \mu(\mathrm{d}x)\\ &\leqslant \max\{1, |a-1|\}\|P^t f - f_*\|_{L^1}.\end{aligned}$$

Consequently, $H_c(P^t f|f_*) \to 0$ as $t \to \infty$ for every density f such that f/f_* is bounded.

Now assume that $H_c(f|f_*) > -\infty$ and write f in the form $f = g_a + f_a$ where $g_a = f - f_a$ and

$$f_a(x) = \begin{cases} 0, & f(x) > af_*(x),\\ f(x), & 0 \leqslant f(x) \leqslant af_*(x).\end{cases}$$

We have $1 = \|f\|_1 = \|g_a\|_1 + \|f_a\|_1$ and $\|g_a\|_1 = \int_{\{f>af_*\}} f\mu(\mathrm{d}x) \to 0$ as $a \to \infty$. Writing $P^t f$ in the form

$$P^t f = \|g_a\|_1 P^t\left(\frac{g_a}{\|g_a\|_1}\right) + \|f_a\|_1 P^t\left(\frac{f_a}{\|f_a\|_1}\right)$$

we obtain

$$\begin{aligned}H_c(P^t f|f_*) &\geqslant \|g_a\|_1 H_c\left(P^t\left(\frac{g_a}{\|g_a\|_1}\right)\bigg|f_*\right) + \|f_a\|_1 H_c\left(P^t\left(\frac{f_a}{\|f_a\|_1}\right)\bigg|f_*\right)\\ &\geqslant \|g_a\|_1 H_c\left(\frac{g_a}{\|g_a\|_1}\bigg|f_*\right) + \|f_a\|_1 H_c\left(P^t\left(\frac{f_a}{\|f_a\|_1}\right)\bigg|f_*\right)\end{aligned} \tag{2.5}$$

for every a and t, where the first inequality follows from the concavity of $H_c(\cdot|f_*)$ while the second is a consequence of $H_c(P^t h|f_*) \geqslant H_c(h|f_*)$ valid for any density h. Since $\frac{f_a}{f_*\|f_a\|_1}$ is bounded, we have $H_c(\frac{f_a}{\|f_a\|_1}|f_*) > -\infty$ and

$$\lim_{t\to\infty} H_c\left(P^t\left(\frac{f_a}{\|f_a\|_1}\right)\bigg|f_*\right) = 0$$

for every a by the first part of our proof. Finally, observe that

$$\|g_a\|_1 H_c\left(\frac{g_a}{\|g_a\|_1}\middle| f_*\right) = -\int_{\{f>af_*\}} f \log\frac{f}{f_*}\,\mu(\mathrm{d}x) + \|g_a\|_1 \log\|g_a\|_1,$$

which is convergent to 0 as $a\to\infty$, because of the condition $H_c(f|f_*) > -\infty$ and the fact that $\|g_a\|_1$ converges to 0. Letting $t\to\infty$ in Eq. (2.5) yields

$$0 \geqslant \liminf_{t\to\infty} H_c(P^t f|f_*) \geqslant \|g_a\|_1 H_c\left(\frac{g_a}{\|g_a\|_1}\middle| f_*\right)$$

for all a. Now taking the limit $a\to\infty$ completes the proof of the *if* part, and of the theorem.

The semigroup P^t on $L^1(X,\mu)$ is called *partially integral* if there exists a measurable function $q: X\times X\to [0,\infty)$ and $t_0>0$ such that

$$P^{t_0}f(x) \geqslant \int_X q(x,y)f(y)\,\mu(\mathrm{d}y)$$

for every density f and

$$\int_X\int_X q(x,y)\,\mu(\mathrm{d}y)\,\mu(\mathrm{d}x) > 0. \qquad \square$$

Our next result draws a connection between the existence of a unique stationary density f_* of P^t and the convergence of the conditional entropy H_c in the case of continuous time systems.

Theorem 2. *Let P^t be a partially integral continuous semigroup of Markov operators. If there is a unique stationary density f_* for P^t and $f_*>0$, then*

$$\lim_{t\to\infty} H_c(P^t f_0|f_*) = 0$$

for all f_0 with $H_c(f_0|f_*) > -\infty$.

Proof. From Pichór and Rudnicki [18, Theorem 2] the semigroup P^t is asymptotically stable if there is a unique stationary density f_* for P^t and $f_*>0$. Thus the result follows from Theorem 1. $\square$

3. Conditional entropy in invertible deterministic systems

In this section we briefly consider the behavior of the conditional entropy in situations where the dynamics are invertible in the sense that they can be run forward or backward in time without ambiguity. To make this clearer, consider a phase space X and a dynamics $S_t: X\to X$. For every initial point x^0, the sequence of successive points $S_t(x^0)$, considered as a function of time t, is called a *trajectory*. In the phase space X, if the trajectory $S_t(x^0)$ is nonintersecting with itself, or intersecting but periodic, then at any given final time t_f such that $x^f = S_{t_f}(x^0)$ we could change the sign of time by replacing t by $-t$, and run the trajectory backward using x^f as a new initial point in X. Then our new trajectory $S_{-t}(x^f)$ would arrive precisely back at x^0 after a time t_f had elapsed: $x^0 = S_{-t_f}(x^f)$. Thus in this case we have a dynamics that may be reversed in time *completely unambiguously*.

We formalize this by introducing the concept of a *dynamical system* $\{S_t\}_{t\in\mathbf{R}}$, which is simply any group of transformations $S_t: X\to X$ having the two properties: 1. $S_0(x)=x$; and 2. $S_t(S_{t'}(x)) = S_{t+t'}(x)$ for $t,t'\in\mathbf{R}$ or $\mathbf{Z}$. Since, from the definition, for any $t\in\mathbf{R}$, we have $S_t(S_{-t}(x)) = x = S_{-t}(S_t(x))$, it is clear that dynamical systems are invertible in the sense discussed above since they may be run either forward or backward in time. Systems of ordinary differential equations are examples of dynamical systems as are invertible maps.

Our first result is very general, and shows that the conditional entropy of any invertible system is constant and uniquely determined by the method of system preparation. This is formalized in

Theorem 3. *If P^t is an invertible Markov operator and has a stationary density f_*, then the conditional entropy is constant and equal to the value determined by f_* and the choice of the initial density f_0 for all time t.*

That is,

$$H_c(P^t f_0|f_*) \equiv H_c(f_0|f_*) \tag{3.1}$$

for all t.

Proof. Since P is invertible, by Voigt's theorem[1] with $g = f_*$ it follows that

$$\begin{aligned} H_c(P^{t+t'} f_0|f_*) &= H_c(P^{t'} P^t f_0|f_*) \\ &\geqslant H_c(P^t f_0|f_*) \geqslant H_c(f_0|f_*) \end{aligned}$$

for all times t and t'. Pick $t' = -t$ so

$$H_c(f_0|f_*) \geqslant H_c(P^t f_0|f_*) \geqslant H_c(f_0|f_*)$$

and therefore

$$H_c(P^t f_0|f_*) = H_c(f_0|f_*)$$

for all t, which finishes the proof. □

In the case where we are considering a deterministic dynamics $S^t : \mathcal{X} \to \mathcal{X}$ where $\mathcal{X} \subset X$, then the corresponding Markov operator is also known as the Frobenius Perron operator [16], and is given explicitly by

$$P^t f_0(x) = f_0(S^{-t}(x))|J^{-t}(x)|, \tag{3.2}$$

where $J^{-t}(x)$ denotes the Jacobian of $S^{-t}(x)$. A simple calculation shows

$$\begin{aligned} H_c(P^t f_0|f_*) &= -\int_{\mathcal{X}} P^t f_0(x) \log \left[\frac{P^t f_0(x)}{f_*(x)}\right] \mathrm{d}x \\ &= -\int_{\mathcal{X}} f_0(S^{-t}(x))|J^{-t}(x)| \log\left[\frac{f_0(S^{-t}(x))}{f_*(S^{-t}(x))}\right] \mathrm{d}x \\ &= -\int_{\mathcal{X}} f_0(y) \log\left[\frac{f_0(y)}{f_*(y)}\right] \mathrm{d}y \\ &\equiv H_c(f_0|f_*) \end{aligned}$$

as expected from Theorem 3.

More specifically, if the dynamics corresponding to our invertible Markov operator are described by the system of ordinary differential equations

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = F_i(x), \quad i = 1, \ldots, d \tag{3.3}$$

operating in a region of $\mathbf{R}^d$ with initial conditions $x_i(0) = x_{i,0}$, then [16] the evolution of $f(t,x) \equiv P^t f_0(x)$ is governed by the generalized Liouville equation

$$\frac{\partial f}{\partial t} = -\sum_i \frac{\partial (fF_i)}{\partial x_i}. \tag{3.4}$$

The corresponding stationary density f_* is given by the solution of

$$\sum_i \frac{\partial (f_* F_i)}{\partial x_i} = 0. \tag{3.5}$$

Note that the uniform density $f_* \equiv 1$, meaning that the flow defined by Eq. (3.3) preserves the Lebesgue measure, is a stationary density of Eq. (3.4) if and only if

$$\sum_i \frac{\partial F_i}{\partial x_i} = 0. \tag{3.6}$$

[1] Voigt's theorem [19] says that if P is a Markov operator, then $H_c(Pf|Pg) \geqslant H_c(f|g)$ for all $f, g \in \mathcal{D}$.

In particular, for the system of ordinary differential equations (3.3) whose density evolves according to the Liouville equation (3.4) we can assert that the conditional entropy of the density $P^t f_0$ with respect to the stationary density f_* will be constant for all time and will have the value determined by the initial density f_0 with which the system is prepared. This result can also be proved directly by noting that from the definition of the conditional entropy we may write

$$H_c(f|f_*) = -\int_{\mathbf{R}^d} f(x)\left[\log\left(\frac{f}{f_*}\right) + \frac{f_*}{f} - 1\right]\mathrm{d}x$$

when the stationary density is f_*. Differentiating with respect to time gives

$$\frac{\mathrm{d}H_c}{\mathrm{d}t} = -\int_{\mathbf{R}^d} \frac{\mathrm{d}f}{\mathrm{d}t}\log\left[\frac{f}{f_*}\right]\mathrm{d}x$$

or, after substituting from (3.4) for $(\partial f/\partial t)$, and integrating by parts under the assumption that f has compact support,

$$\frac{\mathrm{d}H_c}{\mathrm{d}t} = \int_{\mathbf{R}^d} \frac{f}{f_*}\sum_i \frac{\partial(f_* F_i)}{\partial x_i}\,\mathrm{d}x.$$

However, since f_* is a stationary density of P^t, it is clear from (3.4) that

$$\frac{\mathrm{d}H_c}{\mathrm{d}t} = 0,$$

and we conclude that the conditional entropy $H_c(P^t f_0|f_*)$ does not change from its initial value when the dynamics evolve in this manner.

4. Effects of Gaussian noise

In this section, we turn our attention to the behavior of the stochastically perturbed system

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = F_i(x) + \sum_{j=1}^{d} \sigma_{ij}(x)\xi_j, \quad i = 1, \dots, d \tag{4.1}$$

with the initial conditions $x_i(0) = x_{i,0}$, where $\sigma_{ij}(x)$ is the amplitude of the stochastic perturbation and $\xi_j = \mathrm{d}w_j/\mathrm{d}t$ is a 'white noise' term that is the derivative of a Wiener process. In matrix notation we can rewrite Eq. (4.1) as

$$\mathrm{d}x(t) = F(x(t))\,\mathrm{d}t + \Sigma(x(t))\,\mathrm{d}w(t), \tag{4.2}$$

where $\Sigma(x) = [\sigma_{ij}(x)]_{i,j=1,\dots,d}$. Here it is always assumed that the Itô, rather that the Stratonovich, calculus, is used. For a discussion of the differences see Refs. [20,16,21]. In particular, if the σ_{ij} are independent of x then the Itô and the Stratonovich approaches yield identical results.

To fully interpret (4.2) we briefly review the properties of the ξ when they are derived from a Wiener process. We say that a continuous process $\{w(t)\}_{t>0}$ is a one-dimensional *Wiener process* if:

1. $w(0) = 0$; and
2. for all values of s and t, $0 \leqslant s \leqslant t$ the random variable $w(t) - w(s)$ has the *Gaussian density*

$$g(t-s, x) = \frac{1}{\sqrt{2\pi(t-s)}}\exp\left[-\frac{x^2}{2(t-s)}\right]. \tag{4.3}$$

In a completely natural manner this definition can be extended to say that the d-dimensional vector

$$w(t) = \{w_1(t), \dots, w_d(t)\}_{t>0}$$

is a *d-dimensional Wiener process* if its components are one-dimensional Wiener processes. Because of the independence of the increments, it is elementary that the joint density of $w(t)$ is

$$g(t, x_1, \ldots, x_d) = g(t, x_1) \cdots g(t, x_d) \tag{4.4}$$

and thus

$$\int_{R^d} g(t, x)\, dx = 1, \tag{4.5}$$

that

$$\int_{R^d} x_i g(t, x)\, dx = 0, \quad i = 1, \ldots, d \tag{4.6}$$

and

$$\int_{R^d} x_i x_j g(t, x)\, dx = \delta_{ij} t, \quad i,j = 1, \ldots, d. \tag{4.7}$$

In (4.7),

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j \end{cases} \tag{4.8}$$

is the Kronecker delta. Therefore, the average of a Wiener variable is zero by Eq. (4.6), while the variance increases linearly with time according to (4.7).

The *Fokker–Planck equation* that governs the evolution of the density function $f(t, x)$ of the process $x(t)$ generated by the solution to the stochastic differential equation (4.2) is given by

$$\frac{\partial f}{\partial t} = -\sum_{i=1}^{d} \frac{\partial [F_i(x) f]}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 [a_{ij}(x) f]}{\partial x_i \partial x_j}, \tag{4.9}$$

where

$$a_{ij}(x) = \sum_{k=1}^{d} \sigma_{ik}(x) \sigma_{jk}(x).$$

If $k(t, x, x_0)$ is the fundamental solution of the Fokker–Planck equation, i.e., for every x_0 the function $(t, x) \mapsto k(t, x, x_0)$ is a solution of the Fokker–Planck equation with the initial condition $\delta(x - x_0)$, then the general solution $f(t, x)$ of the Fokker–Planck equation (4.9) with the initial condition

$$f(x, 0) = f_0(x)$$

is given by

$$f(t, x) = \int k(t, x, x_0) f_0(x_0)\, dx_0. \tag{4.10}$$

From a probabilistic point of view $k(t, x, x_0)$ is a stochastic kernel (transition density) and describes the probability of passing from the state x_0 at time $t = 0$ to the state x at a time t. Define the Markov operators P^t by

$$P^t f_0(x) = \int k(t, x, x_0) f_0(x_0)\, dx_0, \quad f_0 \in L^1. \tag{4.11}$$

Then $P^t f_0$ is the density of the solution $x(t)$ of Eq. (4.2) provided that f_0 is the density of $x(0)$.

The steady state density $f_*(x)$ is the stationary solution of the Fokker–Planck equation (4.9):

$$-\sum_{i=1}^{d}\frac{\partial[F_i(x)f]}{\partial x_i}+\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2[a_{ij}(x)f]}{\partial x_i\partial x_j}=0. \tag{4.12}$$

In the specific case of $X=\mathbf{R}^d$ and μ equal to the Lebesgue measure, we recover from Eq. (2.1) the conditional entropy

$$H_c(f(t)|f_*)=-\int_{\mathbf{R}^d}f(t,x)\ln\left[\frac{f(t,x)}{f_*(x)}\right]\mathrm{d}x. \tag{4.13}$$

If the coefficients a_{ij} and F_i are sufficiently regular so that a fundamental solution k exists, and $\int_X k(t,x,y)\,\mathrm{d}x=1$, then the unique generalized solution (4.10) to the Fokker–Planck equation (4.9) is given by Eq. (4.11). One such set of conditions is the following: (1) the F_i are of class C^2 with bounded derivatives; (2) the a_{ij} are of class C^3 and bounded with all derivatives bounded; and (3) the uniform parabolicity condition holds, i.e., there exists a strictly positive constant $\rho>0$ such that

$$\sum_{i,j=1}^{d}a_{ij}(x)\lambda_i\lambda_j\geqslant\rho\sum_{i=1}^{d}\lambda_i^2,\quad \lambda_i,\lambda_j\in\mathbf{R},\ \ x\in\mathbf{R}^d.$$

The uniform parabolicity condition implies that $k(t,x,y)>0$, and thus $P^tf(x)>0$ for every density, which implies that there can be at most one stationary density, and that if it exists then $f_*>0$. In this setting, the corresponding conditional entropy $H_c(P^tf_0|f_*)$ approaches its maximal value of zero, as $t\to\infty$, if and only if there is a stationary density f_* that satisfies (4.12).

4.1. The one-dimensional case

If we are dealing with a one-dimensional system, $d=1$, then the stochastic differential Eq. (4.1) simply becomes

$$\frac{\mathrm{d}x}{\mathrm{d}t}=F(x)+\sigma(x)\xi, \tag{4.14}$$

where again ξ is a (Gaussian distributed) perturbation with zero mean and unit variance, and $\sigma(x)$ is the amplitude of the perturbation. The corresponding Fokker–Planck equation (4.9) becomes

$$\frac{\partial f}{\partial t}=-\frac{\partial[F(x)f]}{\partial x}+\frac{1}{2}\frac{\partial^2[\sigma^2(x)f]}{\partial x^2}. \tag{4.15}$$

The Fokker–Planck equation can also be written in the equivalent form

$$\frac{\partial f}{\partial t}=-\frac{\partial S}{\partial x}, \tag{4.16}$$

where

$$S=-\frac{1}{2}\frac{\partial[\sigma^2(x)f]}{\partial x}+F(x)f \tag{4.17}$$

is called the *probability current.*

When stationary solutions of (4.15), denoted by $f_*(x)$ and defined by $P_tf_*=f_*$ for all t, exist they are given as the generally unique (up to a multiplicative constant) solution of (4.12) in the case $d=1$:

$$-\frac{\partial[F(x)f_*]}{\partial x}+\frac{1}{2}\frac{\partial^2[\sigma^2(x)f_*]}{\partial x^2}=0. \tag{4.18}$$

Integration of Eq. (4.18) by parts with the assumption that the probability current S vanishes at the integration limits, followed by a second integration, yields the solution

$$f_*(x) = \frac{K}{\sigma^2(x)} \exp\left[\int_{x_0}^{x} \frac{2F(z)}{\sigma^2(z)}\,\mathrm{d}z\right]. \tag{4.19}$$

This stationary solution f_* will be a density if and only if there exists a positive constant $K>0$ such that f_* can be normalized.

We now discuss rigorous results concerning the one-dimensional case. Let $a(x) = \sigma^2(x)$. Assume that a, a', and F are continuous on $X = (\alpha, \beta)$, where $\infty \leqslant \alpha < \beta \leqslant \infty$. Let $a(x)>0$ for all $x \in (\alpha, \beta)$ and x_0 be any point in (α, β). If we want to study the long term behavior, as $t \to \infty$, of the process $x(t)$ given by Eq. (4.14), we need to know that the process exists for all $t>0$; in other words that there is no explosion in finite time, and that it lives in the interval (α, β). If, for example, α were absorbing so that with positive probability we could reach α in finite time, then a solution f of Eq. (4.15) would show a decrease in norm in $L^1(\alpha, \beta)$.

There is a relation between the behavior of $x(t)$ at the boundary points α and β, and the existence and uniqueness of solutions of the corresponding Fokker–Planck equation as described by [22,23]. In particular, if the condition

$$\int_{\alpha}^{x_0} \exp\left[-\int_{x_0}^{x} \frac{2F(z)}{\sigma^2(z)}\,\mathrm{d}z\right]\mathrm{d}x = \int_{x_0}^{\beta} \exp\left[-\int_{x_0}^{x} \frac{2F(z)}{\sigma^2(z)}\,\mathrm{d}z\right]\mathrm{d}x = \infty \tag{4.20}$$

holds, then the generalized solutions $P^t f$ of the Fokker–Planck equation exist and constitute a semigroup of Markov operators on $L^1(\alpha, \beta)$. If only one of the integrals is finite then this conclusion holds as well, but we restrict ourselves to Condition (4.20) because it is also necessary for the existence of a stationary solution of Eq. (4.14), cf. Pinsky [24]. Thus, we conclude that there is a stationary density f_* if and only if Condition (4.20) holds and

$$\int_{\alpha}^{\beta} \frac{1}{\sigma^2(x)} \exp\left[\int_{x_0}^{x} \frac{2F(z)}{\sigma^2(z)}\,\mathrm{d}z\right]\mathrm{d}x < \infty. \tag{4.21}$$

Define

$$D(x) = \frac{\sigma^2(x)}{2} \quad \text{and} \quad B(x) = -\ln \frac{K}{\sigma^2(x)} - \int_{x_0}^{x} \frac{2F(z)}{\sigma^2(z)}\,\mathrm{d}z,$$

where K is such that $f_*(x) = \mathrm{e}^{-B(x)}$ is normalized. Then the Fokker–Planck equation (4.15) can be rewritten as

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}\left(D(x)\left(\frac{\partial f}{\partial x} + B'(x) f\right)\right) \tag{4.22}$$

and

$$f_*(x) = \mathrm{e}^{-B(x)}.$$

A sufficient condition for an exponential decay of the conditional entropy is the Bakry–Emery condition [25]

$$\frac{1}{2} D''(x) + \frac{D'(x)F(x)}{2D(x)} - F'(x) - \frac{(D'(x))^2}{4D(x)} \geqslant \lambda, \quad x \in X, \tag{4.23}$$

where $\lambda > 0$ is a positive constant. Consequently, if Eq. (4.23) holds then [9,25]

$$H_c(P^t f_0 | f_*) \geqslant \mathrm{e}^{-2\lambda t} H_c(f_0 | f_*) \tag{4.24}$$

for all initial densities f_0 with $H_c(f_0|f_*) > -\infty$. In the following example of an Ornstein–Uhlenbeck process the choice of a Gaussian initial density will result in equality in Eq. (4.24).

4.1.1. Example of an Ornstein– Uhlenbeck process

Trying to find specific examples of $\sigma(x)$ and $F(x)$ for which one can determine the *time dependent* solution $f(t, x)$ of Eq. (4.15) is not easy. One solution that is known is the one for an Ornstein–Uhlenbeck process. Since it is an Ornstein–Uhlenbeck process, which was historically developed in thinking about perturbations to the velocity of a Brownian particle, we denote the dependent variable by v so we have $\sigma(v) \equiv \sigma$ a constant, and $F(v) = -\gamma v$ with $\gamma \geqslant 0$. In this case, Eq. (4.14) becomes

$$\frac{dv}{dt} = -\gamma v + \sigma\xi \tag{4.25}$$

with the corresponding Fokker–Planck equation

$$\frac{\partial f}{\partial t} = \frac{\partial[\gamma v f]}{\partial v} + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial v^2}. \tag{4.26}$$

The unique stationary solution is

$$f_*(v) = \frac{e^{-\gamma v^2/\sigma^2}}{\int_{-\infty}^{+\infty} e^{-\gamma v^2/\sigma^2}\, dv} = \sqrt{\frac{\gamma}{\pi\sigma^2}}\, e^{-\gamma v^2/\sigma^2}. \tag{4.27}$$

Further, from Risken [21, Eq. (5.28)] the fundamental solution $k(t, v, v_0)$ is given by

$$k(t, v, v_0) = \frac{1}{\sqrt{2\pi b(t)}} \exp\left\{-\frac{(v - \exp(-\gamma t)v_0)^2}{2b(t)}\right\}, \tag{4.28}$$

where

$$b(t) = \frac{\sigma^2}{2\gamma}[1 - e^{-2\gamma t}]. \tag{4.29}$$

Note that for a given y and t the function $k(t, \cdot, y)$ is a Gaussian density with mean $v_0 \exp(-\gamma t)$ and variance $b(t)$. Since $k(t, v, v_0) > 0$ for all $v, v_0 \in \mathbf{R}$, the semigroup of Markov operators

$$P^t f_0(v) = \int_{\mathbf{R}} k(t, v, v_0) f_0(v_0)\, dv_0$$

satisfies all of the conditions of Theorem 2. Thus we can assert from Theorem 2 that $\lim_{t\to\infty} H_c(P^t f_0 | f_*) = 0$ for all f_0 with $H_c(f_0|f_*) > -\infty$.

In the case of the Ornstein–Uhlenbeck process, the sufficient condition (4.23) for the exponential lower bound on the conditional entropy reduces to

$$-F'(v) \geqslant \lambda, \quad v \in \mathbf{R}.$$

Thus, $\lambda = \gamma$ and Eq. (4.24) becomes

$$H_c(P^t f_0|f_*) \geqslant e^{-2\gamma t} H_c(f_0|f_*)$$

for all initial densities f_0 with $H_c(f_0|f_*) > -\infty$.

We will now show that this lower bound is optimal. Let us first calculate the conditional entropy of two Gaussian densities. Let $q_1, q_2 > 0$ and let $z_1, z_2 \in \mathbf{R}$. Consider densities of the form

$$g_i(x, z_i) = \sqrt{\frac{q_i}{\pi}} \exp\{-q_i(x - z_i)^2\}, \quad x \in \mathbf{R}, \; i = 1, 2.$$

Then

$$\log\frac{g_1(x, z_1)}{g_2(x, z_2)} = \log\sqrt{\frac{q_1}{q_2}} - q_1(x - z_1)^2 + q_2(x - z_2)^2.$$

Since

$$\int_{\mathbf{R}} g_1(x, z_1)x^2\, dx = \frac{1}{2q_1} + z_1^2 \quad \text{and} \quad \int_{\mathbf{R}} g_1(x, z_1)x\, dx = z_1,$$

we arrive at

$$H_c(g_1(\cdot, z_1)|g_2(\cdot, z_2)) = \frac{1}{2}\log\frac{q_2}{q_1} + \frac{1}{2}\left(1 - \frac{q_2}{q_1}\right) - q_2(z_1 - z_2)^2. \tag{4.30}$$

Now let f_0 be a Gaussian density of the form

$$f_0(v) = \sqrt{\frac{c_1}{\pi}} \exp\{-c_1(v-c_2)^2\},$$

where $c_1 > 0$ and $c_2 \in \mathbf{R}$. Since

$$P^t f_0(v) = \int_{\mathbf{R}} k(t, v, v_0) f_0(v_0)\,\mathrm{d}v_0,$$

we obtain by direct calculation using Eq. (4.28)

$$P^t f_0(v) = \sqrt{\frac{c_1}{\pi(\mathrm{e}^{-2\gamma t} + 2c_1 b(t))}} \exp\left\{-\frac{c_1}{\mathrm{e}^{-2\gamma t} + 2c_1 b(t)}\,(v - c_2\mathrm{e}^{-\gamma t})^2\right\}.$$

Consequently, from Eq. (4.30), with $q_1 = c_1/(\mathrm{e}^{-2\gamma t} + 2c_1 b(t))$, $q_2 = \gamma/\sigma^2$, $z_1 = c_2\mathrm{e}^{-\gamma t}$, and $z_2 = 0$, it follows that

$$H_c(P^t f_0|f_*) = \frac{1}{2}\log\left[1 - \left(1 - \frac{\gamma}{\sigma^2 c_1}\right)\mathrm{e}^{-2\gamma t}\right] + \frac{1}{2}\left(1 - \frac{\gamma}{\sigma^2 c_1} - \frac{2\gamma c_2^2}{\sigma^2}\right)\mathrm{e}^{-2\gamma t}. \tag{4.31}$$

In particular, if $c_1 = \gamma/\sigma^2$ then $H_c(f_0|f_*) = -(\gamma c_2^2/\sigma^2)$ and

$$H_c(P^t f_0|f_*) = \mathrm{e}^{-2\gamma t} H_c(f_0|f_*), \quad t \geqslant 0.$$

4.1.2. A Rayleigh process

Another example for which we have an analytic form for the density $f(t, x)$ is that of

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\gamma v + \frac{A}{2v} + \sigma\xi, \quad v \in [0, \infty) \tag{4.32}$$

and the associated Fokker–Planck equation is

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial v}\left[\left(\gamma v - \frac{A}{2v}\right)f\right] + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial v^2}. \tag{4.33}$$

The unique equilibrium solution, with the proper normalization, is given by

$$f_*(v) = \frac{2}{\Gamma(\beta)}\left(\frac{\gamma}{\sigma^2}\right)^{\beta} v^{2\beta-1}\mathrm{e}^{-\gamma v^2/\sigma^2}, \tag{4.34}$$

where

$$\beta = \frac{1}{2}\left(\frac{A}{\sigma^2} + 1\right) > 0 \tag{4.35}$$

and Γ is the gamma function. The fundamental solution is [26]

$$k(t, v, v_0) = f_*(v)\tilde{k}(t, v, v_0), \tag{4.36}$$

where $\tilde{k}(t, v, v_0) > 0$ for all $t, v, v_0 > 0$. Again from Theorem 2 it follows that $\lim_{t\to\infty} H_c(P^t f_0|f_*) = 0$ for all f_0 with $H_c(f_0|f_*) > -\infty$.

In the case of $A = \sigma^2$ the solution of Eq. (4.32) is called a *Rayleigh process* [27, pp. 135–136] because the stationary density $f_*(v)$ is the Rayleigh distribution. The sufficient Condition (4.23) for the exponential lower bound on the conditional entropy reduces to

$$-F'(v) \geqslant \lambda, \quad v \in (0, \infty).$$

Since $F'(v) = -\gamma - (\sigma^2/2v^2)$, we obtain $\lambda = \gamma$. Then Eq. (4.24) becomes

$$H_c(P^t f_0|f_*) \geqslant \mathrm{e}^{-2\gamma t} H_c(f_0|f_*)$$

for all initial densities f_0 with $H_c(f_0|f_*) > -\infty$.

Let $c \in (0,1)$ be a constant and β be as in Eq. (4.35). By a direct calculation it is easily checked that

$$P^t f_0(v) = \frac{2}{\Gamma(\beta)} \alpha(t)^\beta v^{2\beta-1} \mathrm{e}^{-\alpha(t)v^2}, \quad v \in (0,\infty), \ \ t \geqslant 0,$$

is a solution of Eq. (4.33), where

$$\alpha(t) = \frac{\gamma}{\sigma^2(1 - c\mathrm{e}^{-2\gamma t})}.$$

Now observe that

$$H_c(P^t f_0|f_*) = \log\left(\frac{\gamma}{\sigma^2\alpha(t)}\right)^\beta + \beta\left(1 - \frac{\gamma}{\sigma^2\alpha(t)}\right) = \beta(\log(1 - c\mathrm{e}^{-2\gamma t}) + c\mathrm{e}^{-2\gamma t})$$

for every $t \geqslant 0$. Hence

$$H_c(P^t f_0|f_*) = \frac{\log(1 - c\mathrm{e}^{-2\gamma t}) + c\mathrm{e}^{-2\gamma t}}{\log(1-c) + c} H_c(f_0|f_*).$$

4.2. *The multidimensional case*

In this section first we turn our attention to the existence of a stationary density in the case of multidimensional diffusion when the matrix $\Sigma(x)$ is nonsingular at every point x. This case is much more involved than the one-dimensional case and does not yield simple necessary and sufficient conditions for the existence of a stationary density.

Let us assume that F and a are of class C^2 and C^3, respectively, and that

$$\sum_{i,j=1}^{d} a_{ij}(x)\lambda_i\lambda_j > 0 \quad \text{for } x \in \mathbf{R}^d \text{ and } \lambda \in \mathbf{R}^d - \{0\}.$$

Under these assumptions, the so-called Liapunov method [24] implies that if there exists a nonnegative function V of class C^2 on $\mathbf{R}^d$ and a positive constant $r > 0$ such that

$$\sup_{|x| \geqslant r} L^- V(x) < 0, \tag{4.37}$$

where the operator

$$L^- = \sum_{i=1}^{d} F_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}, \tag{4.38}$$

then there is a unique stationary density f_* and $f_*(x) > 0$ for all x. Again, from Theorem 2 $\lim_{t\to\infty} H_c(P^t f_0|f_*) = 0$ for all f_0 with $H_c(f_0|f_*) > -\infty$. If we simply take $V(x) = |x|^2$, then Condition (4.37) becomes

$$\sup_{|x| \geqslant r} \left(\sum_{i,j=1}^{d} a_{ij}(x) + 2 \sum_{i=1}^{d} x_i F_i(x) \right) < 0.$$

One can also write an integral test [28] which reduces to Condition (4.21) in the one-dimensional case. Let

$$A_1(x) = \frac{1}{|x|^2} x^{\mathrm{T}} a(x) x = \frac{1}{|x|^2} \sum_{i,j=1}^{d} a_{ij}(x) x_i x_j,$$

and

$$A_2(x) = \mathrm{Tr}(a(x)) + 2\langle x, F(x) \rangle = \sum_{i=1}^{d} a_{ii}(x) + 2 \sum_{i=1}^{d} x_i F_i(x).$$

Now define

$$\beta(r) = \sup_{|x|=r} \frac{A_2(x) - A_1(x)}{A_1(x)} \quad \text{and} \quad \alpha(r) = \inf_{|x|=r} A_1(x).$$

Finally, for $r \geqslant 1$, let

$$J(r) = \int_1^r \frac{\beta(s)}{s}\,\mathrm{d}s$$

and assume that there exists $r_0 > 0$ such that the following two conditions hold:

$$\int_{r_0}^{\infty} \mathrm{e}^{-J(r)}\,\mathrm{d}r = \infty \quad \text{and} \quad \int_{r_0}^{\infty} \frac{1}{\alpha(r)}\,\mathrm{e}^{J(r)}\,\mathrm{d}r < \infty. \tag{4.39}$$

We claim that Condition (4.39) implies Condition (4.37). To see this, for $r \geqslant r_0$ define

$$\tilde{J}(r) = \int_1^r \mathrm{e}^{-J(s)} \int_s^{\infty} \frac{1}{\alpha(q)}\,\mathrm{e}^{J(q)}\,\mathrm{d}q\,\mathrm{d}s,$$

and take V to be of class C^2 on $\mathbf{R}^d$ such that $V(x) = \tilde{J}(|x|)$ for $|x| \geqslant r_0$. It is easy to check that

$$2L^- V(x) = A_1(x)\tilde{J}''(|x|) + \frac{\tilde{J}'(|x|)}{|x|}(A_2(x) - A_1(x))$$

and that $2L^- V(x) \leqslant -1$ for $|x| \geqslant r_0$.

We now discuss the speed of convergence of the conditional entropy. Refs. [9,25] provide a general sufficient condition for such convergence in the case of multidimensional diffusions for which the Fokker–Planck equation can be written in a divergence form:

$$\frac{\partial f}{\partial t} = \mathrm{div}(D(x)(\nabla f + f\nabla B(x))), \tag{4.40}$$

where $D(x)$ is locally uniformly positive and $B(x)$ is a real-valued function such that the stationary solution

$$f_*(x) = \mathrm{e}^{-B(x)}$$

is a density. For simplicity assume that $D(x) = D$ is a constant. Then the Fokker–Planck equation (4.40) describes the evolution of densities for the system

$$\mathrm{d}x(t) = -D\nabla B(x(t))\,\mathrm{d}t + \sqrt{2D}I\,\mathrm{d}w(t), \tag{4.41}$$

where I is the identity matrix. If there is $\lambda > 0$ such that

$$D\left(\frac{\partial^2 B(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\ldots,d} \geqslant \lambda I, \tag{4.42}$$

then we have

$$H_c(P^t f_0 | f_*) \geqslant \mathrm{e}^{-2\lambda t} H_c(f_0 | f_*) \tag{4.43}$$

for all initial densities f_0 with $H_c(f_0|f_*) > -\infty$.

4.3. Multidimensional Ornstein– Uhlenbeck process

Consider the multidimensional Ornstein–Uhlenbeck process

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Fx + \Sigma\xi, \tag{4.44}$$

where F is a $d \times d$ matrix, Σ is a $d \times d$ matrix an ξ is d-dimensional vector. The formal solution to Eq. (4.44) is given by

$$x(t) = \mathrm{e}^{tF}x(0) + \int_0^t \mathrm{e}^{(t-s)F}\Sigma\,\mathrm{d}w(t) \tag{4.45}$$

where $\mathrm{e}^{tF} = \sum_{n=0}^{\infty}(t^n/n!)F^n$ is the fundamental solution to $\dot{X}(t) = FX(t)$ with $X(0) = I$, and $w(t)$ is the standard d-dimensional Wiener process. From the properties of stochastic integrals it follows that

$$\eta(t) = \int_0^t \mathrm{e}^{(t-s)F}\Sigma\,\mathrm{d}w(t)$$

has mean 0 and covariance

$$R(t) = E\eta(t)\eta(t)^{\mathrm{T}} = \int_0^t \mathrm{e}^{sF}\Sigma\Sigma^{\mathrm{T}}\mathrm{e}^{sF^{\mathrm{T}}}\,\mathrm{d}s,$$

where B^{T} is the transpose of the matrix B. The matrix $R(t)$ is nonnegative definite but not necessarily positive definite. We follow the presentation of [29,30]. For each $t>0$ the matrix $R(t)$ has constant rank equal to the dimension of the space

$$[F,\Sigma] := \{F^{l-1}\Sigma\varepsilon_j : l,j = 1,\dots,d,\ \varepsilon_j = (\delta_{j1},\dots,\delta_{jp})^{\mathrm{T}}\}.$$

If $m = \operatorname{rank} R(t)$ then $d-m$ coordinates of the process $\eta(t)$ are equal to 0 and the remaining m coordinates constitute an m-dimensional Gaussian process. Thus if $m<d$ there is no stationary density. If $\operatorname{rank} R(t) = d$ then the transition probability function of $x(t)$ is given by the Gaussian density

$$k(t,x,x_0) = \frac{1}{(2\pi)^{d/2}(\det R(t))^{1/2}}\exp\left\{-\frac{1}{2}(x-\mathrm{e}^{tF}x_0)^{\mathrm{T}}R(t)^{-1}(x-\mathrm{e}^{tF}x_0)\right\}, \tag{4.46}$$

where $R(t)^{-1}$ is the inverse matrix of $R(t)$. An invariant density f_* exists if and only if all eigenvalues of F have negative real parts, and in this case the unique stationary density f_* has the form

$$f_*(x) = \frac{1}{(2\pi)^{d/2}(\det R_*)^{1/2}}\exp\left\{-\frac{1}{2}x^{\mathrm{T}}R_*^{-1}x\right\}, \tag{4.47}$$

where R_* is a positive definite matrix given by

$$R_* = \int_0^\infty \mathrm{e}^{sF}\Sigma\Sigma^{\mathrm{T}}\mathrm{e}^{sF^{\mathrm{T}}}\,\mathrm{d}s,$$

and is a unique symmetric matrix satisfying

$$FR_* + R_*F^{\mathrm{T}} = -\Sigma\Sigma^{\mathrm{T}}. \tag{4.48}$$

We conclude that if $[F,\Sigma]$ contains d linearly independent vectors and all eigenvalues of F have negative real parts, then from Theorem 2 it follows that $\lim_{t\to\infty} H_c(P^t f_0|f_*) = 0$ for all f_0 with $H_c(f_0|f_*) > -\infty$.

To simplify the following we first recall some properties of multivariate Gaussian distributions. Let Q_1, Q_2 be positive definite matrices and let

$$g_i(x,z) = \frac{1}{(2\pi)^{d/2}(\det Q_i)^{1/2}}\exp\left\{-\frac{1}{2}(x-z)^{\mathrm{T}}Q_i^{-1}(x-z)\right\}.$$

Then

$$\log\frac{g_1(x,z_1)}{g_2(x,z_2)} = \frac{1}{2}\log\frac{\det Q_2}{\det Q_1} - \frac{1}{2}(x-z_1)^{\mathrm{T}}Q_1^{-1}(x-z_1) + \frac{1}{2}(x-z_2)^{\mathrm{T}}Q_2^{-1}(x-z_2). \tag{4.49}$$

Since $\int g_1(x,0)xx^{\mathrm{T}}\,\mathrm{d}x = Q_1$, we have

$$\int g_1(x,z_1)xx^{\mathrm{T}}\,\mathrm{d}x = Q_1 + z_1z_1^{\mathrm{T}} \quad \text{and} \quad \int g_1(x,z_1)x\,\mathrm{d}x = z_1.$$

Note also that $z^{\mathrm{T}}Qz$ can be written with the help of the trace of a matrix as $\mathrm{Tr}[Qzz^{\mathrm{T}}]$ for any matrix Q and any vector z. Consequently,

$$H_c(g_1(\cdot,z_1)|g_2(\cdot,z_2)) = \frac{1}{2}\log\frac{\det Q_1}{\det Q_2} + \frac{1}{2}\mathrm{Tr}[(Q_1^{-1}-Q_2^{-1})Q_1] - \frac{1}{2}\mathrm{Tr}[Q_2^{-1}(z_1-z_2)(z_1-z_2)^{\mathrm{T}}]. \tag{4.50}$$

Now let f_0 be a Gaussian density of the form

$$f_0(x) = \frac{1}{(2\pi)^{d/2}(\det V(0))^{1/2}}\exp\left\{-\frac{1}{2}(x-m(0))^{\mathrm{T}}V(0)^{-1}(x-m(0))\right\}, \tag{4.51}$$

where $V(0)$ is a positive definite matrix and $m(0) \in \mathbf{R}^d$. From Eq. (4.45) it follows that $x(t)$ is Gaussian with the following mean vector $m(t) = e^{tF}m(0)$ and covariance matrix $V(t) = e^{tF}V(0)e^{tF^T} + R(t)$. Hence

$$P^t f_0(x) = \frac{1}{(2\pi)^{d/2}(\det V(t))^{1/2}} \exp\left\{-\frac{1}{2}(x - m(t))^T V(t)^{-1}(x - m(t))\right\}.$$

Consequently, from Eq. (4.50), with $Q_1 = V(t)$, $z_1 = m(t)$, $Q_2 = R_*$, and $z_2 = 0$, we obtain

$$H_c(P^t f_0|f_*) = \frac{1}{2}\log\frac{\det V(t)}{\det R_*} + \frac{1}{2}\operatorname{Tr}(I - R_*^{-1}V(t)) - \frac{1}{2}\operatorname{Tr}(R_*^{-1}m(t)m(t)^T). \tag{4.52}$$

In particular, if $V(0) = R_*$, then $V(t) = R_*$ and

$$H_c(P^t f_0|f_*) = -\tfrac{1}{2}\operatorname{Tr}(R_*^{-1}m(t)m(t)^T) \tag{4.53}$$

for all $t \geqslant 0$ and every f_0 of the form given by Eq. (4.51).

As a specific example of the multidimensional Ornstein–Uhlenbeck process, to which Eqs. (4.43) and (4.53) can be applied, consider the case when $\Sigma = \sigma I$ and F is a diagonal matrix, $F = -\lambda I$ with $\lambda > 0$. Then $R_*^{-1} = (2\lambda/\sigma^2)I$ and $f_*(x) = e^{-B(x)}$, where

$$B(x) = \frac{1}{2}\log((2\pi)^d \det R_*) + \frac{\lambda}{\sigma^2}x^T x.$$

Thus Condition (4.42) becomes

$$\frac{\sigma^2}{2}\left(\frac{\partial^2 B(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\ldots,d} = \lambda I.$$

Since $e^{tF} = e^{-\lambda t}I$, we conclude from Eq. (4.53) that $H_c(P^t f_0|f_*) = e^{-2\lambda t}H_c(f_0|f_*)$, so the estimate in Eq. (4.43) is optimal. We will show in the next section that in the case of a non-invertible matrix Σ a slower speed of convergence might occur.

To obtain a lower bound on the conditional entropy for the case of a not necessarily invertible matrix Σ and a general f_0, we make use of the inequality

$$H_c(P^t f_0|f_*) \geqslant \int\int f_0(y_1)f_*(y_2)H_c(k(t,\cdot,y_1)|k(t,\cdot,y_2))\,dy_1\,dy_2. \tag{4.54}$$

(For the proof of inequality (4.54) see Appendix B.) From Eq. (4.50) with $Q_1 = R(t)$, $Q_2 = R(t)$, $z_1 = e^{tF}y_1$, and $z_2 = e^{tF}y_1$ it follows that

$$H_c(k(t,\cdot,y_1)|k(t,\cdot,y_2)) = -\tfrac{1}{2}\operatorname{Tr}[R(t)^{-1}e^{tF}(y_1 - y_2)(y_1 - y_2)^T e^{tF^T}].$$

Since $\operatorname{Tr}[R(t)^{-1}e^{tF}yy^T e^{tF^T}] \leqslant \|y\|^2\|e^{tF^T}R(t)^{-1}e^{tF}\|$ for any y by the Schwartz inequality, we obtain from Eq. (4.54)

$$H_c(P^t f_1|f_*) \geqslant -\frac{1}{2}\|e^{tF}\|^2\|R(t)^{-1}\|\int\int\|y_1 - y_2\|^2 f_0(y_1)f_*(y_2)\,dy_1\,dy_2. \tag{4.55}$$

Finally, observe that the norm of $R(t)^{-1}$ is bounded, because $R(t)^{-1}$ converges to R_*^{-1} as $t \to \infty$. Thus, for sufficiently large t we have

$$H_c(P^t f_1|f_*) \geqslant -\|e^{tF}\|^2\|R_*^{-1}\|\int\int\|y_1 - y_2\|^2 f_0(y_1)f_*(y_2)\,dy_1\,dy_2. \tag{4.56}$$

The considerations of the previous paragraphs have been quite general and resulted in the lower bound estimate of inequality (4.54). There are two concrete examples for which it is relatively easy to place a precise lower bound on the conditional entropy evolution. These are the noisy harmonic oscillator and the colored noise which we illustrate in the following:

4.3.1. Harmonic oscillator

Consider the second order system

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}+\gamma\frac{\mathrm{d}x}{\mathrm{d}t}+\omega^2x=\sigma\xi \tag{4.57}$$

with constant positive coefficients m, γ and σ. Introduce the velocity $v=\frac{\mathrm{d}x}{\mathrm{d}t}$ as a new variable. Then Eq. (4.57) is equivalent to the system

$$\frac{\mathrm{d}x}{\mathrm{d}t}=v, \tag{4.58}$$

$$m\frac{\mathrm{d}v}{\mathrm{d}t}=-\gamma v-\omega^2x+\sigma\xi \tag{4.59}$$

and the corresponding Fokker–Planck equation is

$$\frac{\partial f}{\partial t}=-\frac{\partial[vf]}{\partial x}+\frac{1}{m}\frac{\partial[(\gamma v+\omega^2x)f]}{\partial v}+\frac{\sigma^2}{2m^2}\frac{\partial^2 f}{\partial v^2}.$$

We can use the results of Section 4.3 in the two-dimensional setting with

$$F=\begin{pmatrix}0 & 1\\ -\dfrac{\omega^2}{m} & -\dfrac{\gamma}{m}\end{pmatrix}\quad\text{and}\quad \Sigma=\begin{pmatrix}0 & 0\\ 0 & \dfrac{\sigma}{m}\end{pmatrix}.$$

Since

$$[F,\Sigma]=\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}0\\ \dfrac{\sigma}{m}\end{pmatrix},\frac{\sigma}{m}\begin{pmatrix}1\\ -\dfrac{\gamma}{m}\end{pmatrix}\right\},$$

the transition density function is given by Eq. (4.46). The eigenvalues of F are equal to

$$\lambda_1=\frac{-\gamma+\sqrt{\gamma^2-4m\omega^2}}{2m}\quad\text{and}\quad\lambda_2=\frac{-\gamma-\sqrt{\gamma^2-4m\omega^2}}{2m},$$

and are either negative real numbers when $\gamma^2\geqslant 4m\omega^2$ or complex conjugate numbers with negative real parts when $\gamma^2<4m\omega^2$. Thus the stationary density is given by Eq. (4.47). As is easily seen R_*, being a solution to Eq. (4.48), is given by

$$R_*=\begin{pmatrix}\dfrac{\sigma^2}{2\gamma\omega^2} & 0\\ 0 & \dfrac{\sigma^2}{2m\gamma}\end{pmatrix}.$$

The inverse of the matrix R_* is

$$R_*^{-1}=\frac{2\gamma}{\sigma^2}\begin{pmatrix}\omega^2 & 0\\ 0 & m\end{pmatrix}$$

and the unique stationary density becomes

$$f_*(x,v)=\frac{\gamma\omega\sqrt{m}}{\pi\sigma^2}\,\mathrm{e}^{-(\gamma/\sigma^2)[\omega^2x^2+mv^2]}.$$

As in Section 4.3 we conclude that $\lim_{t\to\infty}H_c(P^tf_0|f_*)=0$ for all f_0 with $H_c(f_0|f_*)>-\infty$.

The bound on the temporal convergence of $H_c(P^tf_0|f_*)$ to zero, as given by Eq. (4.56), is determined by $\|\mathrm{e}^{tF}\|^2$. Thus, we are going to calculate $\|\mathrm{e}^{tF}\|^2$ to see the nature of the general formula (4.56) and to compute expression (4.53) for the conditional entropy in the case of initial Gaussian densities.

First consider the overdamped case when $\lambda_1 \neq \lambda_2$ are real. Define, for $t \geqslant 0$,

$$c_1(t) = \frac{\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}}{\lambda_2 - \lambda_1} \quad \text{and} \quad c_2(t) = \frac{e^{\lambda_2 t} - e^{\lambda_1 t}}{\lambda_2 - \lambda_1}.$$

Then

$$e^{tF} = \begin{pmatrix} c_1(t) & c_2(t) \\ c_1'(t) & c_2'(t) \end{pmatrix}$$

and

$$\|e^{tF}\|^2 = e^{2\lambda_1 t} + e^{2\lambda_2 t} + (\lambda_1\lambda_2 + 1)^2 c_2^2(t).$$

If we take $m(0) = (1, \lambda_i)^{\mathrm{T}}$, $i = 1, 2$ then $m(t) = e^{\lambda_i t} m(0)$ in Eq. (4.53), and

$$H_c(P^t f_0 | f_*) = e^{2\lambda_i t} H_c(f_0 | f_*).$$

Next consider the underdamped case with complex λ_1, λ_2, and let

$$\alpha = -\frac{\gamma}{2m} \quad \text{and} \quad \beta = \frac{\sqrt{4m\omega^2 - \gamma^2}}{2m}.$$

Then

$$e^{tF} = \frac{e^{\alpha t}}{\beta} \begin{pmatrix} \beta\cos(\beta t) - \alpha\sin(\beta t) & \sin(\beta t) \\ -(\alpha^2 + \beta^2)\sin(\beta t) & \beta\cos(\beta t) + \alpha\sin(\beta t) \end{pmatrix}$$

and we have

$$\|e^{tF}\|^2 = \frac{e^{2\alpha t}}{\beta^2}(2\beta^2\cos^2(\beta t) + (2\alpha^2 + (\alpha^2 + \beta^2)^2 + 1)\sin^2(\beta t)).$$

If we take $m(0) = (0, \beta)^{\mathrm{T}}$ then $m(t) = e^{\alpha t}(\sin(\beta t), \beta\cos(\beta t) + \alpha\sin(\beta t))^{\mathrm{T}}$ in Eq. (4.53), and

$$H_c(P^t f_0 | f_*) = \frac{e^{2\alpha t}}{\beta^2}((\beta\cos(\beta t) + \alpha\sin(\beta t))^2 + (\alpha^2 + \beta^2)\sin^2(\beta t)) H_c(f_0 | f_*).$$

Finally, consider the critically damped case when $\lambda_1 = \lambda_2$, that is $\gamma^2 = 4m\omega^2$, and set

$$\lambda = -\frac{\gamma}{2m}.$$

Then we have

$$F = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & 2\lambda \end{pmatrix} \quad \text{and} \quad e^{tF} = e^{\lambda t}\begin{pmatrix} 1 - \lambda t & t \\ -\lambda^2 t & 1 + \lambda t \end{pmatrix}$$

and

$$\|e^{tF}\|^2 = e^{2\lambda t}(2 + (\lambda^2 + 1)^2 t^2).$$

If we take $m(0) = (1, \lambda)^{\mathrm{T}}$ then $m(t) = e^{\lambda t} m(0)$ and Eq. (4.53) becomes

$$H_c(P^t f_0 | f_*) = e^{2\lambda t} H_c(f_0 | f_*).$$

But, if we take $m(0) = (1, 0)^{\mathrm{T}}$, then $m(t) = e^{\lambda t}(1 - \lambda t, -\lambda^2 t)^{\mathrm{T}}$ and

$$H_c(P^t f_0 | f_*) = e^{2\lambda t}((1 - \lambda t)^2 + \lambda^2 t^2) H_c(f_0 | f_*).$$

In all these three cases the speed of convergence of the conditional entropy is determined by $\|e^{tF}\|^2$. A straightforward argument based on our expressions for $\|e^{tF}\|^2$ shows that the convergence is most rapid for the overdamped case, intermediate for the underdamped, and slowest for the critically damped case. Note that these conclusions are independent of the initial density, and specifically independent of the noise amplitude.

4.3.2. Colored noise

Consider the system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\alpha x + \eta, \tag{4.60}$$

where $\alpha > 0$ and η is the one-dimensional Ornstein–Uhlenbeck process with parameters $\gamma, \sigma > 0$

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = -\gamma\eta + \sigma\xi.$$

Then Eq. (4.60) is equivalent to the system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\alpha x + v, \tag{4.61}$$

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\gamma v + \sigma\xi, \tag{4.62}$$

and the corresponding Fokker–Planck equation is

$$\frac{\partial f}{\partial t} = \frac{\partial[(\alpha x - v)f]}{\partial x} + \frac{\partial[\gamma v f]}{\partial v} + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial v^2}.$$

We can use the results of Section 4.3 in the two-dimensional setting with

$$F = \begin{pmatrix} -\alpha & 1 \\ 0 & -\gamma \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0 & 0 \\ 0 & \sigma \end{pmatrix}.$$

Observe that

$$[F, \Sigma] = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ \sigma \end{pmatrix}, \begin{pmatrix} \sigma \\ -\gamma\sigma \end{pmatrix} \right\}.$$

The eigenvalues of F are given by

$$\lambda_1 = -\alpha \quad \text{and} \quad \lambda_2 = -\gamma,$$

and are evidently negative. Finally,

$$R_* = \frac{\sigma^2}{2\gamma(\alpha+\gamma)} \begin{pmatrix} \frac{1}{\alpha} & 1 \\ 1 & \alpha+\gamma \end{pmatrix}$$

and

$$R_*^{-1} = \frac{2\alpha(\alpha+\gamma)}{\sigma^2} \begin{pmatrix} \alpha+\gamma & -1 \\ -1 & \frac{1}{\alpha} \end{pmatrix}.$$

Thus, the unique stationary density is equal to

$$f_*(x, v) = \frac{\sqrt{\alpha\gamma}(\alpha+\gamma)}{\pi\sigma^2} \exp\left\{ -\frac{\alpha+\gamma}{\sigma^2}(\alpha(\alpha+\gamma)x^2 - 2\alpha x v + v^2) \right\}.$$

As in Section 4.3 we conclude that $\lim_{t\to\infty} H_c(P^t f_0 | f_*) = 0$ for all f_0 with $H_c(f_0|f_*) > -\infty$.

As in the case of the harmonic oscillator the bound on the temporal convergence of $H_c(P^t f_0|f_*)$ to zero, as given by Eq. (4.56), is determined by $\|\mathrm{e}^{tF}\|^2$. Again we calculate $\|\mathrm{e}^{tF}\|^2$ to see the nature of the formula (4.56) and compute expression (4.53) for the conditional entropy in the case of initial Gaussian densities.

First, consider the case when $\alpha \neq \gamma$. The fundamental matrix is given by

$$\mathrm{e}^{tF} = \begin{pmatrix} \mathrm{e}^{-\alpha t} & \beta(t) \\ 0 & \mathrm{e}^{-\gamma t} \end{pmatrix} \quad \text{with} \quad \beta(t) = \frac{1}{\alpha-\gamma}(\mathrm{e}^{-\gamma t} - \mathrm{e}^{-\alpha t})$$

and

$$\|e^{tF}\|^2 = e^{-2\gamma t} + e^{-2\alpha t} + \beta^2(t).$$

If we take $m(0) = (1, 0)^T$, then $m(t) = e^{-\alpha t} m(0)$ in Eq. (4.53) and

$$H_c(P^t f_0 | f_*) = e^{-2\alpha t} H_c(f_0 | f_*).$$

Similarly, for $m(0) = (1, \alpha - \gamma)^T$ we have $m(t) = e^{-\gamma t} m(0)$ and

$$H_c(P^t f_0 | f_*) = e^{-2\gamma t} H_c(f_0 | f_*).$$

Now consider the case when $\alpha = \gamma$. We have

$$e^{tF} = \begin{pmatrix} e^{-\gamma t} & t e^{-\gamma t} \\ 0 & e^{-\gamma t} \end{pmatrix} \quad \text{and} \quad \|e^{tF}\|^2 = e^{-2\gamma t}(2 + t^2).$$

Eq. (4.53) becomes, for $m(0) = (0, 1)^T$ and $m(t) = e^{-\gamma t}(t, 1)^T$,

$$H_c(P^t f_0 | f_*) = e^{-2\gamma t}(2\gamma^2 t^2 - 2\gamma t + 1) H_c(f_0 | f_*).$$

In both cases the speed of convergence of the conditional entropy is determined by $\|e^{tF}\|^2$. Again note that these conclusions are independent of the initial density, and specifically independent of the noise amplitude.

5. Markovian dichotomous noise

Another example where we can use our results is the case of dichotomous noise [20, Section 8]. The state space of the Markovian dichotomous noise $\xi(t)$ consists of two states $\{c_+, c_-\}$ and is characterized by a transition probability from the state c_+ to c_- in the small time interval Δt given by $\alpha \Delta t + o(\Delta t)$, and from the state c_- to c_+ given by $\beta \Delta t + o(\Delta t)$, where $\alpha, \beta > 0$. It has the correlation function

$$\langle \xi(t) \xi(s) \rangle = \frac{\alpha \beta (c_+ - c_-)^2}{(\alpha + \beta)^2} \exp(-(\alpha + \beta) t).$$

A system subject to this type of noise is described by the equation

$$\frac{dx}{dt} = F(x) + \sigma(x) \xi.$$

The pair $(x(t), \xi(t))$ is Markovian and writing $a(x, c_\pm) = F(x) + c_\pm \sigma(x)$ we arrive at

$$dx(t) = a(x(t), \xi(t))\, dt.$$

The process $\xi(t)$ determines which deterministic system,

$$\frac{dx_+}{dt} = a(x, c_+) \quad \text{or} \quad \frac{dx_-}{dt} = a(x, c_-),$$

to choose. We assume that given the initial condition $x_\pm(0) = x$, each of these equations has a solution $x_+(t)$ and $x_-(t)$, respectively, defined and finite for all $t \geqslant 0$, and we write $\pi_+^t(x) = x_+(t)$ and $\pi_-^t(x) = x_-(t)$. Furthermore, we assume that there is a minimal open set X such that $\pi_\pm^t(X) \subseteq X$ for all $t > 0$. Let $\mathcal{B}(X \times \{c_+, c_-\})$ be the sigma algebra of Borel subsets of $X \times \{c_+, c_-\}$ and let μ be the product measure on $\mathcal{B}(X \times \{c_+, c_-\})$ which, on every set of the form $B \times \{c_\pm\}$, is equal to the length of the set B. The norm of any element f of the space $L^1(X \times \{c_+, c_-\}, \mu)$ is equal to

$$\|f\|_1 = \int_X |f_+(x)|\, dx + \int_X |f_-(x)|\, dx,$$

and we now define a semigroup of Markov operators on this space which describes the temporal evolution of the densities of the process $(x(t), \xi(t))$.

The evolution equation for the densities $f_\pm(t, x) = f(t, x, c_\pm)$ of the process $(x(t), \xi(t))$ is of the form

$$\frac{\partial f_+}{\partial t} = -\frac{\partial [a(x, c_+) f_+]}{\partial x} - \alpha f_+ + \beta f_- \tag{5.1}$$

$$\frac{\partial f_-}{\partial t} = -\frac{\partial[a(x,c_-)f_-]}{\partial x} + \alpha f_+ - \beta f_-. \tag{5.2}$$

Formally writing $f = (f_+, f_-)^{\mathrm{T}}$, we arrive at the equation

$$\frac{\partial f}{\partial t} = Af + Mf, \tag{5.3}$$

where

$$M = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix} \quad \text{and} \quad Af = \begin{pmatrix} -\dfrac{\partial[a(x,c_+)f_+]}{\partial x} \\ -\dfrac{\partial[a(x,c_-)f_-]}{\partial x} \end{pmatrix}.$$

Then the operator A generates a semigroup of Markov operators T^t on $L^1(X \times \{c_+, c_-\}, \mu)$. Since M is a bounded operator, $A + M$ also generates a semigroup of Markov operators P^t, and

$$P^t f_0 = T^t f_0 + \int_0^t T^{t-s} M P^s f_0 \,\mathrm{d}s$$

holds for every $f_0 \in L^1(X \times \{c_+, c_-\}, \mu)$. The semigroup P^t gives the generalized solution $f(t,\cdot)$ of Eq. (5.3) with the initial condition $f(0,\cdot) = f_0$. Using the results of Pichór and Rudnicki [18, Proposition 2] we infer that if there is an $x_0 \in X$ such that $a(x_0, c_+) \neq a(x_0, c_-)$, then the semigroup P^t is partially integral, and if X is the minimal set such that $\pi^t_\pm(X) \subseteq X$, then this semigroup can have at most one stationary density. Consequently, if a stationary density f_* exists, Theorem 2 implies that

$$\lim_{t\to\infty} H_c(P^t f_0 | f_*) = 0 \tag{5.4}$$

for all f_0 with $H_c(f_0|f_*) > -\infty$, where the conditional entropy for P^t on $L^1(X \times \{c_+, c_-\}, \mu)$ is equal to

$$H_c(P^t f_0 | f_*) = -\int_X P^t f(x,c_+) \log \frac{P^t f(x,c_+)}{f_*(x,c_+)} \,\mathrm{d}x - \int_X P^t f(x,c_-) \log \frac{P^t f(x,c_-)}{f_*(x,c_-)} \,\mathrm{d}x$$

by Eq. (2.1).

The density of the state variable $x(t)$ is an element of the space $L^1(X, m)$, where m is the Lebesgue measure on X. It is given by

$$p(t,x) = P^t f_0(x,c_+) + P^t f_0(x,c_-). \tag{5.5}$$

Thus, the stationary density f_* of P^t gives us the stationary density of $x(t)$

$$p_*(x) = f_*(x,c_+) + f_*(x,c_-).$$

Since

$$H_c(p(t)|p_*) = -\int_X p(t,x) \log \frac{p(t,x)}{p_*(x)} \,\mathrm{d}x \geqslant H_c(P^t f_0 | f_*), \tag{5.6}$$

we conclude that

$$\lim_{t\to\infty} H_c(p(t)|p_*) = 0. \tag{5.7}$$

For a general one-dimensional system with dichotomous noise, one can derive a formula for the stationary density. For the sake of clarity, we follow [31] and restrict our discussion to the case of symmetric dichotomous noise where

$$c_+ = -c_- = c \quad \text{and} \quad \alpha = \beta,$$

and

$$a(x,c_+)a(x,c_-) < 0,$$

where the zeros of $a(x, c_+)a(x, c_-)$ are the boundaries of X. Then the unique stationary density of P^t is given by

$$f_*(x, c_\pm) = K \frac{1}{|a(x, c_\pm)|} \exp\left\{ -\alpha \int^x \left(\frac{1}{a(z, c_+)} + \frac{1}{a(z, c_-)} \right) \mathrm{d}z \right\},$$

where K is a normalizing constant.

As a specific example, consider the linear dichotomous flow

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\gamma x + \xi,$$

where $\gamma > 0$. Then $c_\pm = \pm c$ and $a(x, c_\pm) = -\gamma x \pm c$ for $x \in \mathbf{R}$. Thus

$$\pi^t_\pm(x) = x\mathrm{e}^{-\gamma t} \pm \frac{c}{\gamma}(1 - \mathrm{e}^{-\gamma t})$$

and the state space is

$$X = \left(-\frac{c}{\gamma}, \frac{c}{\gamma} \right).$$

The stationary density f_* in this case is given by

$$f_*(x, \pm c) = \frac{\gamma}{2(\mp \gamma x + c)B(1/2, \alpha/\gamma)} \left(1 - \frac{\gamma^2}{c^2} x^2 \right)^{\alpha/\gamma},$$

where B is the beta function. Since $a(x, +c) > 0 > a(x, -c)$ for all $x \in X$, Conditions (5.4) and (5.7) hold, where the stationary density p_* of the state variable $x(t)$ is equal to

$$p_*(x) = \frac{\gamma}{cB(1/2, \alpha/\gamma)} \left(1 - \frac{\gamma^2}{c^2} x^2 \right)^{\alpha/\gamma - 1}.$$

6. Discussion

Here we have examined the evolution of the conditional (or Kullback–Leibler or relative) entropy to a maximum in stochastic systems. We were motivated by a desire to understand the role of noise in the evolution of the conditional entropy to a maximum since in invertible systems (e.g. measure preserving systems of differential equations or invertible maps) the conditional entropy is fixed at the value with which the system is prepared. However, the addition of noise can reverse this property and lead to an evolution of the conditional entropy to a maximum value of zero. We have made concrete calculations to see how the entropy converges, and shown that it is monotone and at least exponential in several situations.

Specifically, in Section 2 we introduced the dynamic concept of asymptotic stability and the notion of conditional entropy, and gave two main results connecting the convergence of the conditional entropy with asymptotic stability (Theorem 1), and the existence of unique stationary densities with the convergence of the conditional entropy (Theorem 2). In Section 3 we illustrated the well-known fact that asymptotic stability is a property that cannot be found in an invertible deterministic system, such as a system of ordinary differential equations. Consequently the conditional entropy cannot have a temporal dependence for deterministic invertible dynamics, and will always have a value determined by the system preparation. Section 4 introduced a stochastic extension to this invertible and constant entropy situation in which a system of ordinary equations is perturbed by Gaussian white noise (thus becoming non-invertible). We summarized some general results showing that in this stochastic case asymptotic stability holds. Then in Section 4.1 we considered specific one-dimensional examples, and showed that the conditional entropy convergence to zero is monotone and at least exponential, considering the specific examples of an Ornstein–Uhlenbeck process in Section 4.1.1 and a Rayleigh process in Section 4.1.2. We went on to look at multidimensional stochastic systems with nondegenerate noise in Section 4.2, showing that the exponential convergence of the entropy still holds. Examples of higher dimensional situations with degenerate noise were considered within the context of a two-dimensional Ornstein–Uhlenbeck process in Section 4.3 with specific examples of a stochastically perturbed harmonic oscillator (Section 4.3.1) and colored noise (Section 4.3.2) as examples. In the cases of the Ornstein–Uhlenbeck and Rayleigh processes as well as the stochastically perturbed harmonic oscillator and colored noise examples, we obtained exact formulae for the temporal evolution of the conditional entropy starting from a concrete initial distribution. The rather surprising result is that the rate of convergence of the entropy to zero is independent of the noise amplitude. The final Section 5 applied our general results to the problem of conditional entropy convergence in the presence of dichotomous noise.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council (NSERC Grant OGP-0036920, Canada) and MITACS. This research was carried out while MT-K was visiting McGill University in 2004.

Appendix A. Proof of inequality (2.4)

Jensen's inequality states that when v is a normalized measure and ϕ is a concave function then

$$\phi\left(\int g(z)v(\mathrm{d}z)\right) \geqslant \int \phi(g(z))v(\mathrm{d}z).$$

Since log is a concave function, Jensen's inequality implies

$$\int f \log\frac{f}{g}\,\mu(\mathrm{d}x) \leqslant \log \int f\frac{f}{g}\,\mu(\mathrm{d}x).$$

Now note that

$$\int\left(\frac{f}{g}-1\right)^2 g\mu(\mathrm{d}x) = \int\left[\left(\frac{f}{g}\right)^2 g - 2\frac{f}{g}\,g + 1\right]\mu(\mathrm{d}x) = \int\left(\frac{f}{g}\right)^2 g\mu(\mathrm{d}x) - 1.$$

Thus,

$$-H_c(f|g) \leqslant \log\left[1 + \int\left(\frac{f}{g}-1\right)^2 g\mu(\mathrm{d}x)\right] \leqslant \int\left(\frac{f}{g}-1\right)^2 g\mu(\mathrm{d}x).$$

Appendix B. Proof of inequality (4.54)

To derive inequality (4.54), first write

$$H_c(P^t f_1|P^t f_2) = -\int P^t f_1(x)\log\frac{P^t f_1(x)}{P^t f_2(x)}\,\mathrm{d}x \equiv -\int \varphi(P^t f_1(x), P^t f_2(x)),$$

where $\varphi(u,v) = v\log(u/v)$ is convex. From the properties of convex functions there always exist sequences of real numbers $\{a_n\}$ and $\{b_n\}$ such that

$$\varphi(u,v) = \sup\{a_n u + b_n v : n \in \mathbf{N}\}.$$

Remembering (4.11) we can then write

$$\begin{aligned}
a_n P^t f_1(x) + b_n P^t f_2(x) &= a_n \int k(t,x,y_1) f_1(y_1)\,\mathrm{d}y_1 + b_n \int k(t,x,y_2) f_2(y_2)\,\mathrm{d}y_2 \\
&= \int\int \{a_n k(t,x,y_1) f_1(y_1) f_2(y_2) + b_n k(t,x,y_2) f_1(y_1) f_2(y_2)\}\,\mathrm{d}y_1\,\mathrm{d}y_2 \\
&= \int\int f_1(y_1) f_2(y_2)\{a_n k(t,x,y_1) + b_n k(t,x,y_2)\}\,\mathrm{d}y_1\,\mathrm{d}y_2 \\
&\leqslant \int\int f_1(y_1) f_2(y_2)\varphi(k(t,x,y_1), k(t,x,y_2))\,\mathrm{d}y_1\,\mathrm{d}y_2.
\end{aligned}$$

Thus,

$$\sup_{n\in\mathbf{N}}\{a_nP^tf_1(x)+b_nP^tf_2(x)\}=\varphi(P^tf_1(x),P^tf_2(x))=\int\int f_1(y_1)f_2(y_2)\varphi(k(t,x,y_1),k(t,x,y_2))\,\mathrm{d}y_1\,\mathrm{d}y_2,$$

so

$$\begin{aligned}H_c(P^tf_1(x)|P^tf_2(x))&\geqslant-\int\int f_1(y_1)f_2(y_2)\left[\int\varphi(k(t,x,y_1),k(t,x,y_2))\,\mathrm{d}x\right]\mathrm{d}y_1\,\mathrm{d}y_2\\&=-\int\int f_1(y_1)f_2(y_2)H_c(k(t,x,y_1),k(t,x,y_2))\,\mathrm{d}y_1\,\mathrm{d}y_2.\end{aligned}$$

References

[1] L. Andrey, Phys. Lett. A 111 (1985) 45.
[2] W.-H. Steeb, Physica A 95 (1979) 181.
[3] M.C. Mackey, Time's Arrow: The Origins of Thermodynamic Behaviour, Springer, Berlin, New York, Heidelberg, 1992.
[4] M.C. Mackey, in: C. Mugnai, A. Ranfagni, L. Schulman (Eds.), Time's Arrows, Quantum Measurement and Superluminal Behavior, Consiglio Nazionale Delle Richerche, Roma, 2001, pp. 49–65.
[5] L. Evans, Bull. Am. Math. Soc. 41 (2004) 409.
[6] K. Loskot, R. Rudnicki, Ann. Pol. Math. 52 (1991) 140.
[7] A. Abbondandolo, Stoch. Anal. Appl. 17 (1999) 131.
[8] G. Toscani, C. Villani, J. Stat. Phys. 98 (2000) 1279.
[9] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, Commun. Partial Differential Equations 26 (2001) 43.
[10] H. Qian, Phys. Rev. E 63 (2001) 042103.
[11] H. Qian, M. Qian, X. Tang, J. Stat. Phys. 107 (2002) 1129.
[12] A. Arnold, J.A. Carrillo, L. Desvillettes, J. Dolbeault, A. Jüngel, C. Lederman, P.A. Markowich, G. Toscani, C. Villani, Monatsh. Math. 142 (2004) 35.
[13] P. Markowich, C. Villani, Mat. Contemp. 19 (2000) 1.
[14] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley Series in Telecommunications, Wiley, New York, 1991.
[15] P. Diaconis, L. Saloff-Coste, Ann. Appl. Probab. 6 (1996) 695 (ISSN 1050–5164).
[16] A. Lasota, M.C. Mackey, Chaos, Fractals and Noise: Stochastic Aspects of Dynamics, Springer, Berlin, New York, Heidelberg, 1994.
[18] K. Pichór, R. Rudnicki, J. Math. Anal. Appl. 249 (2000) 668.
[19] J. Voigt, Commun. Math. Phys. 81 (1981) 31.
[20] W. Horsthemke, R. Lefever, Noise Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology, Springer, Berlin, New York, Heidelberg, 1984.
[21] H. Risken, The Fokker–Planck Equation, Springer, Berlin, New York, Heidelberg, 1984.
[22] W. Feller, Ann. Math. (2) 55 (1952) 468.
[23] W. Feller, Trans. Am. Math. Soc. 77 (1954) 1.
[24] R.G. Pinsky, Positive Harmonic Functions and Diffusion, Cambridge Studies in Advanced Mathematics, vol. 45, Cambridge University Press, Cambridge, 1995.
[25] D. Bakry, M. Émery, in: Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Mathematics, vol. 1123, Springer, Berlin, 1985, pp. 177–206.
[26] A.N. Borodin, P. Salminen, Handbook of Brownian motion—facts and formulae, Probability and its Applications, Birkhäuser, Basel, 2002.
[27] C. Gardiner, Handbook of Stochastic Methods, Springer, Berlin, Heidelberg, 1983 (ISBN 0-387-15607-0).
[28] R.N. Bhattacharya, Ann. Probab. 6 (1978) 541.
[29] M. Zakai, J. Snyders, J. Differential Equations 8 (1970) 27.
[30] R.V. Erickson, Ann. Math. Statist. 42 (1971) 820.
[31] K. Kitahara, W. Horsthemke, R. Lefever, Phys. Lett. A 70 (1979) 377.