LOGARITHMIC SOBOLEV INEQUALITIES FOR INHOMOGENEOUS MARKOV SEMIGROUPS

JEAN-FRANÇOIS COLLET[1] AND FLORENT MALRIEU[2]

Abstract. We investigate the dissipativity properties of a class of scalar second order parabolic partial differential equations with time-dependent coefficients. We provide explicit condition on the drift term which ensure that the relative entropy of one particular orbit with respect to some other one decreases to zero. The decay rate is obtained explicitly by the use of a Sobolev logarithmic inequality for the associated semigroup, which is derived by an adaptation of Bakry's Γ-calculus. As a byproduct, the systematic method for constructing entropies which we propose here also yields the well-known intermediate asymptotics for the heat equation in a very quick way, and without having to rescale the original equation.

1. INTRODUCTION

In this paper, we investigate a class of linear parabolic equations (or a class of stochastic differential equations) which due to the presence of time-dependent coefficients have no stationary solution, but for which orbits still do come together in the "metric" given by relative entropy or, more generally, by Φ-entropies. More precisely, we provide explicit conditions on the coefficients, adapting Bakry-Émery's criterion, which ensure that the entropies (1) or (2) decrease to zero for large time, with quantitative bounds, for any two orbits u and v, that is, **even when v is a non-stationary solution**. A similar work is done in [9] in the framework of discrete time Markov chains by the mean of Dobrushin's coefficient. This method no longer works for diffusions on non compact sets but its last remark suggest that Bakry and Émery technology can be used.

1.1. The Kullback-Leibler distance as a particular Φ-entropy

Given two probability densities u, v on $\mathbb{R}^d$, the entropy of u relative to v (also known in information theory as their Kullback-Leibler distance (see [12]), although it is not a distance) is defined by

$$H(u|v) := \int_{\mathbb{R}^d} u(x) \log \left(\frac{u(x)}{v(x)} \right) \mathrm{d}x. \tag{1}$$

Although this quantity does not satisfy the triangle inequality, it is always non-negative and vanishes only when $u = v$. These two facts are immediate consequences of the well-known Pinsker inequality (see [14]):

$$H(u|v) \geq \frac{1}{2}|u - v|_{L^1}^2.$$

Thus the quantity $H(u|v)$ may provide some notion of "distance" between u and v. In Partial Differential Equations $H(u|v)$ may be useful in studying the asymptotic behavior of a dissipative system. In this context v is in general a stationary solution (or in physical terminology a detailed balance equilibrium), u is the orbit of some Kolmogorov (or some other parabolic) equation, and $H(u|v)$ is a non increasing function of time. In many cases this fact can be combined with some clever inequalities to show that v in fact attracts u in some appropriate metric. This is the basis of the well-known entropy dissipation method (see [16]), which has been used to good advantage in many examples such as (linear or nonlinear) parabolic equations, kinetic equations, etc.

As is well-known from information theory and Statistical Physics (1) is a particular instance of a more general class of entropies:

$$H^\Phi(u|v) := \int_{\mathbb{R}^d} v(x)\Phi\left(\frac{u(x)}{v(x)}\right) dx, \tag{2}$$

where Φ is any convex function defined on $[0, \infty[$ and dx stands for the d-dimensional Lebesgue measure. The convexity of Φ implies a lower bound for $H^\Phi(u|v)$ as follows: using Jensen's inequality with the probability measure $\frac{v dx}{\int v \, dx}$ we obtain the inequality

$$\int_{\mathbb{R}^d} v\Phi\left(\frac{u}{v}\right) dx \geq \int_{\mathbb{R}^d} v \, dx \Phi\left(\frac{\int_{\mathbb{R}^d} u \, dx}{\int_{\mathbb{R}^d} v \, dx}\right).$$

If we assume that $\Phi(1) = 0$, then $H^\Phi(u|v)$ remains non-negative.

Remark 1.1. Formula (1) corresponds to the particular choice $\Phi(z) = z \log z$ which has an interesting extensivity property [8], but as a general rule the dissipation of entropy is a convexity property which has little to do with the specific properties of the $z \log z$ function.

1.2. **The entropy production for linear scalar advection-diffusion equations**

Consider a general linear second-order scalar advection-diffusion equation:

$$\frac{\partial u}{\partial t} + \text{div}(b(t,x)u - a(t,x)\nabla u) = 0, \tag{3}$$

where b is a vector field and a is a diffusion matrix (symmetric and positive) which are smooth enough to ensure the strong existence and uniqueness for (3). For any two positive solutions u, v to this equation, one can get:

$$\frac{d}{dt}H^\Phi(u|v) = -\int_{\mathbb{R}^d} \Phi''\left(\frac{u}{v}\right)\left[\nabla\left(\frac{u}{v}\right) \cdot a\nabla\left(\frac{u}{v}\right)\right] v \, dx. \tag{4}$$

As expected, this formula shows that linear transport does not play any role in the entropy production (the expression does not involve the velocity field b): in physical parlance, diffusion here is the only irreversible process. Let us emphasize that, from the convexity of Φ and the positive definiteness of the diffusion matrix a, we obtain that the relative entropy $H^\Phi(u|v)$ is a time-decreasing quantity, **whatever the solutions** u, v **are**, and **whatever the coefficients** b, a **are**. In particular for arbitrary time-dependent coefficients b and a the system will have no detailed balance equilibrium, *i.e.* the problem

$$b(t,x)u(t,x) = a(t,x)\nabla u(t,x)$$

will have no solution at all. Note that in the context of Markov processes, a similar dissipation property was exhibited by Yosida, and Kubo (see [11]).

The natural question then arises to investigate under which (sufficient) conditions on the coefficients of the equation does the entropy decrease to zero. In the classical setup where one investigates the trend toward a stationary solution, it is well-known that for a large class of such stationary solutions, some Logarithmic Sobolev Inequality is available [13]. This fact can be used to obtain a Gronwall-type inequality for the entropy, thereby yielding exponential decay. The supplementary difficulty here is that the measure relative to which the entropy is computed moves along the flow, in such a way that classical conditions which ensure that a logarithmic Sobolev inequality will hold cannot be checked *a priori*.

In the next section we revisit a well-known prototype, the Ornstein-Uhlenbeck equation with constant drift. This example shows that depending on the nature of the drift the entropy may decay to zero in an exponential or algebraic fashion, or converge to a nonzero value. Section 3 collects the technical tools needed to show that for the case where the diffusion matrix is the identity matrix, the solution of the evolution problem will satisfy the logarithmic Sobolev inequality at all positive times. The asymptotic behavior of the entropy is obtained as an easy corollary in Section 4. In Section 5 we show that at least for the heat equation (but we believe for a much larger class of parabolic problems), the choice of the fundamental solution for v provides a very quick proof of the classical Gaussian intermediate scaling. At least, in Section 6, we provide an application to genetics: we show that the Wright-Fisher diffusion with great mutations rates forget its initial condition exponentially fast if the mutations rates, which can depend on time, are greater than $1/2$.

2. The fundamental example: the Ornstein-Uhlenbeck process

Let us consider the simplest case. Denote by $(X)_{t \geq 0}$ the solution of

$$\mathrm{d}X_t = \sqrt{2}B_t - \lambda X_t\, \mathrm{d}t,$$

where $(B)_{t \geq 0}$ is a standard Brownian motion on $\mathbb{R}$ and $\lambda \in \mathbb{R}$ is a constant. This equation can be solved as follows:

$$X_t = X_0 \mathrm{e}^{-\lambda t} + \sqrt{2} \int_0^t \mathrm{e}^{\lambda(s-t)}\, \mathrm{d}B_s.$$

As a conclusion, the measure $P_t(\cdot)(x)$ which is defined as the law of X_t knowing that $X_0 = x$ is the Gaussian measure with mean $x\mathrm{e}^{-\lambda t}$ and variance $(1 - \mathrm{e}^{-2\lambda t})/\lambda$. One can then compute the relative entropy of $P_t(\cdot)(y)$ with respect to $P_t(\cdot)(x)$:

$$\alpha(t) := H(P_t(\cdot)(y)|P_t(\cdot)(x)) = \frac{\lambda(x-y)^2}{2(\mathrm{e}^{2\lambda t} - 1)},$$

since $P_t(\cdot)(y)$ and $P_t(\cdot)(x)$ have the same variance. Of course, in the case when $\lambda = 0$, the above formula has to be understood as

$$H(P_t(\cdot)(y)|P_t(\cdot)(x)) = \frac{(x-y)^2}{4t}.$$

As a conclusion, three different behaviors can occur:

- if $\lambda > 0$, then α decreases exponentially fast to 0, which is natural since $P_t(\cdot)(x)$ converges exponentially fast to its invariant measure $\mathcal{N}(0, 1/\lambda)$.
- if $\lambda = 0$, then α still goes to zero although $P_t(\cdot)$ does not converge to a probability measure,
- if $\lambda < 0$, then α converges exponentially fast to a nonzero limit:

$$\alpha(t) = \frac{-\lambda(x-y)^2}{2(1 - \mathrm{e}^{2\lambda t})} = -\frac{\lambda}{2}(x-y)^2 - \frac{\lambda(x-y)^2}{2(1 - \mathrm{e}^{2\lambda t})}\mathrm{e}^{2\lambda t} \xrightarrow[t \to \infty]{} -\frac{\lambda}{2}(x-y)^2.$$

3. The local Φ-Sobolev inequality for inhomogeneous diffusion Semigroups

3.1. **Notations**

In this section we consider the family of formal elliptic partial differential operators $(L_t)_{t>0}$ defined by

$$L_t f(x) := \sum_{i,j=1}^{d} a_{ij}(t,x)\partial_{ij} f(x) + \sum_{i=1}^{d} b_i(t,x)\partial_i f(x), \tag{5}$$

where $(a_{ij}(t,\cdot))_{1\leq i,j\leq d}$ is a symmetric definite positive diffusion matrix and $b(t,\cdot)$ is a given vector field on $\mathbb{R}^d$, defined for all $t > 0$. Let us suppose that the coefficients are smooth functions of (t,x). This family of operators $(L_t)_{t\geq 0}$ generates an inhomogeneous Markov semigroup which we will denote by $(P_{s,t})_{0\leq s\leq t}$ in the following sense. Writing as usual a as $a = \sigma\sigma^T$, one can associate to $(L_t)_{t\geq 0}$ the solution of the following SDE:

$$\forall t \geq r, \quad X_t^{x,r} = x + \int_r^t b(s, X_s^{x,r})\,\mathrm{d}s + \sqrt{2}\int_r^t \sigma(s, X_s^{x,r})\,\mathrm{d}B_s,$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion on $\mathbb{R}^d$. Semigroup and probabilistic approaches are linked by the fundamental relation

$$P_{s,t}f(x) := \mathbb{E}f(X_t^{s,x}).$$

The Markov property of X can be translated into a composition rule for the semigroup: for every $s \leq t \leq u$,

$$P_{s,u}f(x) = \mathbb{E}[f(X_u^{s,x})] = \mathbb{E}\left[f\left(X_u^{t,X_t^{s,x}}\right)\right] = \mathbb{E}[P_{t,u}f(X_t^{s,x})] = P_{s,t}P_{t,u}f(x).$$

This semigroup satisfies the well-known Kolmogorov equations:

$$\partial_s P_{s,t}f = -L_s P_{s,t}f, \quad \partial_t P_{s,t}f = P_{s,t}L_t f. \tag{6}$$

Let μ be a probability measure on $\mathbb{R}^d$ and u the density function of the law of X_t knowing that $\mathcal{L}(X_0) = \mu$. Then, for every smooth function f,

$$\mathbb{E}(f(X_t)) = \mathbb{E}(\mathbb{E}(f(X_t)|X_0)) = \int P_{0,t}f(x)\,\mu(\mathrm{d}x).$$

The Itô Formula ensures that, for every smooth function f and $s \leq t$,

$$f(X_t) - f(X_s) - \int_s^t L_r f(X_r)\,\mathrm{d}r$$

is a martingale. In other words,

$$\mathbb{E}f(X_t) - \mathbb{E}f(X_s) - \int_s^t \mathbb{E}L_r f(X_r)\,\mathrm{d}r = 0.$$

As a consequence, for any smooth function f,

$$\int \left[u(t,x) - u(s,x) - \int_s^t L_r^* u(r,x)\,\mathrm{d}r\right]f(x)\,\mathrm{d}x = 0,$$

and u satisfies

$$\begin{cases} \partial_t u = \operatorname{div}(a\nabla u + (\partial a - b)u) \\ u(t,x)\,\mathrm{d}x \xrightarrow[t\to 0]{} \mu(\mathrm{d}x), \end{cases}$$

where ∂a is the vector defined by $(\partial a)_j = \sum_i \partial_i a_{ij}$ and then it is of the form (3) (once b has been relabeled).

Following [2,3], let us associate to L_t the two bilinear forms $\Gamma(t)$ and $\mathrm{E}_2(t)$ defined by:

$$\Gamma(t)(f,g) := \frac{1}{2}[L_t(fg) - gL_tf - fL_tg],$$

$$\mathrm{E}_2(t)(f,g) := \frac{1}{2}[L_t\Gamma(f,g) - \Gamma(g,L_tf) - \Gamma(f,L_tg)].$$

We will write $\Gamma(t)(f)$ instead of $\Gamma(t)(f,f)$ and $\mathrm{E}_2(t)(f)$ instead of $\mathrm{E}_2(t)(f,f)$.

Remark 3.1. One can check that

$$\Gamma(t)(f,g)(x) = \nabla f(x)\cdot a(t,x)\nabla g(x) = \sum_{i,j=1}^{d} a_{ij}(t,x)\partial_i f(x)\partial_j g(x).$$

Remark 3.2. The expression of E_2 is much more complicated in the general case. In the simple (but informative) case when $a(t,\cdot)$ is the identity matrix, it is very easy to check the following formula:

$$\mathrm{E}_2(t)(f) := ||\operatorname{Hess}(f)||_2^2 - \nabla f \cdot \operatorname{Jac}(b(t))\nabla f, \tag{7}$$

where $\operatorname{Hess}(\cdot)$ (resp. $\operatorname{Jac}(\cdot)$) stands for the Hessian (resp. Jacobian) matrix, and $||B||_2$ denotes the Hilbert-Schmidt norm.

Remark 3.3. Notice that

$$\nabla f \cdot \operatorname{Jac}(b)\nabla f = \nabla f \cdot \operatorname{SJac}(b)\nabla f,$$

where $\operatorname{SJac}(\cdot)$ stands for the symmetric part of the Jacobian matrix *i.e.*,

$$\operatorname{SJac}(b)_{ij} = \frac{\operatorname{Jac}(b)_{ij} + \operatorname{Jac}(b)_{ji}}{2}.$$

The antisymmetric part of the Jacobian of b brings no contribution in our study. One can think about the following explicit example: consider the 2-dimensional process X solution of the following SDE:

$$\mathrm{d}X_t = \mathrm{d}B_t - \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} X_t\,\mathrm{d}t.$$

The antisymmetric part of the drift induces a rotation whereas the symmetric part ensures the convergence to equilibrium.

Let $\Phi\colon \mathcal{I} \to \mathbb{R}$ be a smooth convex function defined on a closed interval $\mathcal{I}$ of $\mathbb{R}$ not necessarily bounded. Let μ be a positive measure on a Borel space $(\Omega, \mathcal{F})$. The Φ-entropy functional $\operatorname{Ent}_\mu^\Phi$ is defined on the set of μ-integrable functions $f\colon (\Omega, \mathcal{F}) \to (\mathcal{I}, \mathcal{B}(\mathcal{I}))$ by

$$\operatorname{Ent}_\mu^\Phi(f) = \int_\Omega \Phi(f)\,\mathrm{d}\mu - \Phi\left(\int_\Omega f\,\mathrm{d}\mu\right).$$

In what follows, μ is a probability measure. As a consequence, $\int_\Omega f\,\mathrm{d}\mu \in \mathcal{I}$ and the definition make sense. in the sequel, one has to make an extra assumption in order to derive interesting functional inequalities:

$$(u,v) \mapsto \Phi''(u)v^2 \text{ is non negative and convex on } \mathcal{I} \times \mathcal{I}. \tag{8}$$

Remark 3.4. The classical variance and entropy are Φ-entropy functionals respectively associated to $x \mapsto x^2$ on $\mathcal{I} = \mathbb{R}$ and $x \mapsto x \log x$ on $\mathcal{I} = [0, +\infty)$.

Definition 3.5. The semigroup $(P_{s,t})_{0 \le s \le t}$ is said to satisfy a local Φ-Sobolev inequality with constants $(C_{s,t})_{0 \le s \le t}$ if for all $s \le t$ and smooth function f,

$$\mathrm{Ent}^{\Phi}_{P_{s,t}}(f) \le C_{s,t} P_{s,t}(\Phi''(f)\Gamma(t)(f)).$$

Remark 3.6. Under the so-called Bakry-Émery criterion,

$$\exists \rho \in \mathbb{R}, \quad \forall f \text{ smooth}, \quad \Gamma_2(f) \ge \rho \Gamma(f),$$

homogeneous diffusion semigroups satisfy a Poincaré and a logarithmic Sobolev inequality (see [4]). As a generalization, Φ-Sobolev inequalities can also be established (see [7]).

Our aim is to take into account the time dependence of the coefficients of the diffusion process. We will show that the appropriate adaptation of the Bakry-Émery criterion to that situation is as follows:

$$\exists \rho \; : \; t \mapsto \rho(t) \in \mathbb{R}, \quad \forall f \text{ smooth}, \quad \Gamma_2(t)(f) + \frac{1}{2}\partial_t \Gamma(t)(f) \ge \rho(t)\Gamma(t)(f), \tag{9}$$

where $\partial_t \Gamma(t)$ is defined as

$$\partial_t \Gamma(t)(f, g)(x) := \sum_{i,j=1}^{d} \partial_t a_{ij}(t, x)\partial_i f(x)\partial_j g(x).$$

Remark 3.7. As pointed by an anonymous referee, one can consider the process $(\tilde{X}_t)_{t \ge 0} = ((t, X_t))_{t \ge 0}$ whose generator is $\tilde{L} = \partial_t + L_t$ and recover (9) thanks to homogeneous Bakry-Émery criterion.

The key point in the homogeneous and diffusive case is to get the following commutation relation (which turns out to be equivalent to Bakry-Émery criterion):

$$\sqrt{\Gamma P_t f} \le \mathrm{e}^{-\rho t} P_t\left(\sqrt{\Gamma f}\right).$$

In the following subsection we derive such a commutation relation in the inhomogeneous case.

3.2. **The commutation relation**

Let s and t be two fixed times, with $0 \le s \le t$. The key point is the following lemma, which describes how the dissipative mechanism tends to flatten gradients:

Lemma 3.8. *Suppose that the family of operators $(L_t)_{t \ge 0}$ defined in (5) satisfies (9). For any τ between 0 and t, the following inequality holds true:*

$$\sqrt{\Gamma(\tau)(P_{\tau,t}g)} \le \exp\left(-\int_{\tau}^{t} \rho(u)\,\mathrm{d}u\right)P_{\tau,t}\left(\sqrt{\Gamma(t)(g)}\right). \tag{10}$$

Proof. For all $u \in [\tau, t]$, we define $\beta(u)$ by

$$\beta(u) = P_{\tau,u}\left(\sqrt{\Gamma(u)(P_{u,t}g)}\right),$$

and compute its derivative by using (6). The crucial assumption that L_u is a diffusion operator ensures that

$$L_u(\Phi(g)) = \Phi'(g)L_u(g) + \Phi''(g)\Gamma(u)(g). \tag{11}$$

In order to make the exposition clearer we denote $P_{u,t}g$ as h and Γ stands for $\Gamma(u)$. A straightforward computation leads to:

$$
\begin{aligned}
\beta'(u) &= P_{\tau,u}\left(L_u\sqrt{\Gamma P_{u,t}g} + \frac{1}{2\sqrt{\Gamma(P_{u,t}g)}}\{-2\Gamma(P_{u,t}g, L_u P_{u,t}g) + \partial_u\Gamma(P_{u,t}g)\}\right) \\
&= P_{\tau,u}\left[\frac{1}{2\sqrt{\Gamma h}}L_u\Gamma h - \frac{1}{4(\Gamma h)^{\frac{3}{2}}}\Gamma\Gamma h - \frac{\Gamma(h, L_u h)}{\sqrt{\Gamma h}} + \frac{\partial_u\Gamma(h)}{2\sqrt{\Gamma h}}\right] \\
&= P_{\tau,u}\left[\frac{2(\Gamma h)L_u\Gamma h - 4(\Gamma h)\Gamma(h, L_u h) - \Gamma\Gamma h + 2(\Gamma h)(\partial_u\Gamma h)}{4(\Gamma h)^{\frac{3}{2}}}\right] \\
&= P_{\tau,u}\left[\frac{4(\Gamma h)\Gamma_2(u)(h) - \Gamma\Gamma h + 2(\Gamma h)\partial_u\Gamma h}{4(\Gamma h)^{\frac{3}{2}}}\right].
\end{aligned}
$$

Therefore we obtain:

$$
\beta'(u) - \rho(u)\beta(u) = P_{\tau,u}\left[\frac{4(\Gamma h)(\Gamma_2(u)(h) + (1/2)\partial_u\Gamma h - \rho(u)\Gamma h) - \Gamma\Gamma h}{4(\Gamma h)^{\frac{3}{2}}}\right].
$$

Following [4], one can show, thanks to the diffusion assumption, that the criterion (9) implies that, for all smooth functions f,

$$
\Gamma_2(u)(f) + (1/2)\partial_u\Gamma f - \rho(u)\Gamma f \geq \frac{\Gamma\Gamma f}{4\Gamma f}.
$$

Remark 3.9. In the case when a is the identity matrix, one can easily get derive this inequality from the criterion (9) and the Cauchy-Schwarz inequality.

As a conclusion, β satisfy the following differential inequality:

$$
\beta'(u) \geq \rho(u)\beta(u).
$$

In other words, the function

$$
u \mapsto \beta(u)\exp\left(-\int_\tau^u \rho(v)\,\mathrm{d}v\right)
$$

is an increasing function on the interval $[\tau, t]$ which implies that

$$
\beta(\tau) \leq \beta(t)\exp\left(-\int_\tau^t \rho(u)\,\mathrm{d}u\right).
$$

This is precisely the desired inequality.   □

3.3. Local Φ-Sobolev inequalities

Theorem 3.10. *Suppose that the family of operators $(L_t)_{t\geq 0}$ defined in (5) satisfies (9). Then for any times s, t with $0 \leq s \leq t$ and any positive function g, $P_{s,t}$ satisfies the following Φ-Sobolev inequality:*

$$
\mathrm{Ent}^\Phi_{P_{s,t}}(g) := P_{s,t}(\Phi(g)) - \Phi(P_{s,t}g) \leq c(s,t)P_{s,t}(\Phi''(g)\Gamma(t)(g)),
$$

where the constant $c(s,t)$ can be chosen as:

$$
c(s,t) = \int_s^t \exp\left(-2\int_\tau^t \rho(u)\,\mathrm{d}u\right)\mathrm{d}\tau.
$$

Proof. Consider the function $\alpha : [s,t] \to \mathbb{R}$ defined by:

$$\alpha(\tau) := P_{s,\tau}(\Phi(P_{\tau,t}g)).$$

Let us compute the derivative of α:

$$\alpha'(\tau) = P_{s,\tau}(L_\tau(\Phi(P_{\tau,t}g)) - \Phi'(P_{\tau,t}g)L_\tau P_{\tau,t}g).$$

Thanks to the fact the diffusion assumption, (11) ensures that

$$\alpha'(\tau) = P_{s,\tau}(\Phi''(P_{\tau,t}g)\Gamma(\tau)(P_{\tau,t}g)).$$

The commutation relation (10) ensures that

$$\Phi''(P_{\tau,t}g)\Gamma(\tau)(P_{\tau,t}g) = \Phi''(P_{\tau,t}g)\left(\sqrt{\Gamma(\tau)(P_{\tau,t}g)}\right)^2 \leq \exp\left(-2\int_\tau^t \rho(u)\,\mathrm{d}u\right)\Phi''(P_{\tau,t}g)P_{\tau,t}\left(\sqrt{\Gamma(t)(g)}\right)^2.$$

Jensen inequality with the bivariate function $(u,v) \mapsto \Phi''(u)v^2$ (which is assumed to be convex according to (8)) ensures that

$$\Phi''(P_{\tau,t}g)P_{\tau,t}\left(\sqrt{\Gamma(t)(g)}\right)^2 \leq P_{\tau,t}(\Phi''(g)\Gamma(t)(g)).$$

As a conclusion,

$$\alpha'(\tau) \leq \exp\left(-2\int_\tau^t \rho(u)\,\mathrm{d}u\right)P_{s,\tau}P_{\tau,t}(\Phi''(g)\Gamma(t)(g)) = \exp\left(-2\int_\tau^t \rho(u)\,\mathrm{d}u\right)P_{s,t}(\Phi''(g)\Gamma(t)(g)).$$

Since $\alpha(t) = P_{s,t}(\Phi(g))$ and $\alpha(s) = \Phi(P_{s,t}(g))$, the result follows upon integration of this inequality between s and t. $\qquad\square$

In the special case of $\Phi : x \mapsto x\log x$, the local logarithmic Sobolev inequality for the semigroup $(P_{s,t})_{0\leq s\leq t}$ can be stated as follows:

Corollary 3.11. *Suppose that the family of operators $(L_t)_{t\geq 0}$ defined in (5) satisfies (9). Then for any times s,t with $0 \leq s \leq t$ and any positive function g, $P_{s,t}$ satisfies the following logarithmic Sobolev inequality:*

$$\mathrm{Ent}_{P_{s,t}}(g) := P_{s,t}(g\log g) - (P_{s,t}g)\log(P_{s,t}g) \leq c(s,t)P_{s,t}\left(\frac{\Gamma(t)(g)}{g}\right), \tag{12}$$

where the constant $c(s,t)$ can be chosen as:

$$c(s,t) = \int_s^t \exp\left(-2\int_\tau^t \rho(u)\,\mathrm{d}u\right)\mathrm{d}\tau.$$

Remark 3.12. If for every $x \in \mathbb{R}^d$, the matrix $(a_{ij}(t,x))_{i,j}$ is bounded by the identity matrix (as symmetric bilinear forms), then $\Gamma(t)(g) \leq |\nabla g|^2$ and $P_{s,t}$ satisfies the classical logarithmic Sobolev inequality:

$$\mathrm{Ent}_{P_{s,t}}(g) \leq c(s,t)P_{s,t}\left(\frac{|\nabla g|^2}{g}\right).$$

Remark 3.13. In the case where $\rho(t) = \rho$ for all $t > 0$, we recover the constant

$$c(s,t) = \frac{1}{2\rho}\left(1 - e^{-2\rho(t-s)}\right)$$

provided by [4] in the homogeneous case.

4. THE KULLBACK-LEIBLER DISTANCE OF TWO ORBITS OF A PARABOLIC PROBLEM

We now consider the following parabolic equation (3):

$$\frac{\partial v}{\partial t}(t,x) = L_t^* v(t,x), \quad t > 0, \ x \in \mathbb{R}^d, \tag{13}$$

$$v(0,x) = v_0(x), \quad x \in \mathbb{R}^d, \tag{14}$$

where L_t^* stands for the adjoint of L_t with respect to the Lebesgue measure on $\mathbb{R}^d$.

The aim of this section is to get a control of $H^\Phi(u|v)$ in terms of the curvature. We will consider in the following the particular case of the classical relative entropy. The (straightforward) generalization to Φ-entropies is left as an exercise to the reader...

If we assume that the initial data v_0 satisfies a logarithmic Sobolev inequality, the result of the previous section may be used to show that the inequality is propagated in time:

Theorem 4.1. *Assume the initial data v_0 satisfies the following logarithmic Sobolev inequality: for all smooth functions f,*

$$\int f \log f v_0 - \int f v_0 \log \int f v_0 \le d_0 \int \Gamma(0)(f) v_0.$$

If the family $(L_t)_{t \ge 0}$ satisfies the criterion (9). Then for any positive time, the solution $v(t,\cdot)$ of (13), (14) satisfies the following logarithmic Sobolev inequality

$$\int f \log f v(t,\cdot) - \int f v(t,\cdot) \log \int f v(t,\cdot) \le \mathrm{d}(t) \int \Gamma(t)(f) v(t,\cdot),$$

where

$$\mathrm{d}(t) := d_0 \exp\left(-2 \int_0^t \rho(r)\,\mathrm{d}r\right) + \int_0^t \exp\left(-2 \int_\tau^t \rho(r)\,\mathrm{d}r\right)\mathrm{d}\tau. \tag{15}$$

Proof. For any positive function g we have:

$$\int g(x) \log g(x) v(t,x)\,\mathrm{d}x = \int \mathbb{E}\left[f\left(X_t^{0,x}\right)\right] v_0(x)\,\mathrm{d}x = \int P_{0,t}(g \log g)(x) v(0,x)\,\mathrm{d}x.$$

The integrand may be bounded from above by using the local inequality (12) with $s = 0$, and the logarithmic Sobolev inequality for v_0:

$$
\begin{aligned}
\int P_{0,t}(g \log g) v(x,0)\,\mathrm{d}x &\le \int P_{0,t}g \log P_{0,t}g v_0(x)\,\mathrm{d}x + c(0,t)\int P_{0,t}\left(\frac{\Gamma(t)(g)}{g}\right)v_0(x)\,\mathrm{d}x \\
&\le d_0 \int \frac{\Gamma(0)(P_{0,t}g)}{P_{0,t}g}v_0(x)\,\mathrm{d}x + c(0,t)\int P_{0,t}\left(\frac{\Gamma(t)(g)}{g}\right)v_0(x)\,\mathrm{d}x.
\end{aligned}
$$

The first integral in the last inequality may be estimated by (10), and this completes the proof. $\qquad\square$

We are now in a position to estimate the Kullback-Leibler distance between two arbitrary orbits of (13):

Theorem 4.2. *Under the same assumptions as in previous theorem, let u be another solution of (13) (i.e., corresponding to different initial data u_0). Assume that u_0 and v_0 are positive. Define the relative entropy of u with respect to v at any positive time t by*

$$H(u(t)|v(t)) := \int u(t,x) \log \frac{u(t,x)}{v(t,x)} \, dx.$$

This quantity is then bounded as follows for all positive times:

$$H(u(t)|v(t)) \le H(u(0)|v(0))c(t),$$

where

$$c(t) = \exp\left(-\int_0^t \frac{1}{d(s)} \, ds\right), \tag{16}$$

and $d(t)$ is the constant defined in (15).

Proof. The proof is a straightforward application of Gronwall's lemma. Let us set $g := \frac{u}{v}$; from (4) with $\Phi(z) = z \log z$, we obtain

$$\frac{d}{dt} H(u(t)|v(t)) = -\int \frac{\Gamma(t)(g)}{g} v \, dx.$$

Therefore, Theorem 4.1 gives the following control:

$$\frac{d}{dt} H(u(t)|v(t)) \le -\frac{1}{d(t)} H(u(t)|v(t)),$$

which gives the result. $\square$

Let us conclude this section by indicating what the obtained rate is when the quantity $\rho(t)$ may be taken to be a fixed constant ρ. If $\rho = 0$, then (15) gives $d(t) = d_0 + t$, therefore we get the algebraic decay

$$c(t) = \frac{1}{1 + \frac{t}{d_0}}.$$

For $\rho \ne 0$, we have

$$d(t) = d_0 e^{-2\rho t} + \frac{1}{2\rho}(1 - e^{-2\rho t}),$$

and the integral in (16) may be computed to yield

$$c(t) = \frac{2\rho d_0 e^{-2\rho t}}{1 + (2\rho d_0 - 1)e^{-2\rho t}},$$

and then, we recover the exponential rate of convergence of the relative entropy.

5. An application to intermediate asymptotics

In the case where a Kolmogorov equation has a rather trivial (*e.g.* constant) asymptotic state, it is often the case that (due to the self similarity of the underlying Markov process) some appropriate rescaling of the orbit shows structure (*e.g.* Gaussian), a phenomenon termed *intermediate asymptotics* by Barenblatt [5]. To illustrate this point let us consider the one dimensional linear heat equation on the entire line:

$$\forall t > 0, \ \forall x \in \mathbb{R}, \quad \frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t,x).$$

It is very easy to check that for large times u converges point wise to zero, but it is also very easy to read off from the explicit form of the solution (as given in terms of the heat kernel) the following point wise convergence:

$$\sqrt{t}u(t, y\sqrt{t}) \xrightarrow[t \to \infty]{} C \exp\left(-\frac{y^2}{2}\right), \tag{17}$$

where the constant C may be determined from mass conservation. As is well-known, this convergence may be obtained by rescaling the equation and constructing an entropy functional for the rescaled equation. Let us briefly recall this argument. The first step consists in rescaling the function u by setting:

$$u(t, x) = \alpha(t)v\left(\tau(t), \frac{x}{\beta(t)}\right),$$

where the scaling functions $\alpha(t), \beta(t), \tau(t)$ are to be chosen so as to make the equation for v as simple as possible, while preserving the mass constraint:

$$\int_{\mathbb{R}^d} u(t, x)\,\mathrm{d}x = \text{const.}$$

These two requirements lead to the choice

$$v(\log t, y) = \sqrt{t}u(t, y\sqrt{t}), \tag{18}$$

thus to the equation

$$\frac{\partial v}{\partial \tau} = \frac{1}{2}\frac{\partial}{\partial y}\left(yv + \frac{\partial v}{\partial y}\right).$$

The detailed balance equilibria for this equation are exactly all multiples of the standard Gaussian density, *i.e.* take the form:

$$v_\infty(y) = C \exp\left(-\frac{y^2}{2}\right), \tag{19}$$

for some constant C. By the result of previous section, the relative entropy

$$H(v|v_\infty) = \int_{\mathbb{R}} v \log \frac{v}{v_\infty}\,\mathrm{d}y \tag{20}$$

decreases to zero for large times, which implies that v converges to a fixed point of the form (19) for some C.

This result, when rephrased in terms of u, is exactly the intermediate asymptotics (17). Note that in this case the constant C is uniquely determined from the mass conservation relation:

$$\int_{\mathbb{R}} u(t, x)\,\mathrm{d}x = \int_{\mathbb{R}} u(0, x)\,\mathrm{d}x.$$

The key point is now that by using the change of variable $x = y\sqrt{t}$ in (20) one obtains:

$$H(v|v_\infty) = \int_{\mathbb{R}} u(t, x) \ln \frac{u(t, x)}{C\frac{1}{\sqrt{t}}\exp\left(-\frac{x^2}{2t}\right)}\,\mathrm{d}x,$$

which is exactly the relative entropy of u with respect to the fundamental solution of the heat equation. The fact that this entropy is dissipated combined with Pinsker's inequality now immediately leads to the intermediate asymptotic (17), without having to resort to any rescaling of the function u.

In other words, the fundamental solution encodes the intermediate asymptotics.

6. An application to Wright-Fisher diffusion

Consider the diffusion process on $(0,1)$ solution of

$$\mathrm{d}X_t = \sqrt{X_t(1-X_t)}\,\mathrm{d}B_t + (-\alpha X_t + \beta(1-X_t))\,\mathrm{d}t, \tag{21}$$

where α and β are two positive numbers. This process describes the proportion of an allele A in a large haploid population for a gene with two alleles A and a when mutations from A to a (resp. from a to A) occur with rate α (resp. β), see [10] for the biological setting. We assume that α and β depend on time but are greater than $1/2$.

Proposition 6.1. *The process* (21) *satisfies the curvature criterion* (9) *with the function ρ defined by*

$$\rho(t) = \begin{cases} \dfrac{1}{2}\dfrac{(\alpha(t)-\beta(t))^2}{\alpha(t)+\beta(t)-1-\sqrt{(2\alpha(t)-1)(2\beta(t)-1)}} & \text{if } \alpha(t) \neq \beta(t), \\ 2\alpha(t)-1 & \text{if } \alpha(t) = \beta(t). \end{cases} \tag{22}$$

Proof. Since the diffusion X belongs to $(0,1)$ and the diffusion coefficient does not depend on time, the curvature criterion can be written as

$$\rho(t) = \inf_{x\in(0,1)} \left(\sigma(x)\sigma''(x) + b(t,x)\frac{\sigma'(x)}{\sigma(x)} - \partial_x b(t,x) \right),$$

where $\sigma(x) = \sqrt{x(1-x)}$ and $b(t,x) = -\alpha(t)x + \beta(t)(1-x)$, and then

$$\rho(t) = \inf_{x\in(0,1)} \frac{2(\alpha(t)-\beta(t))x - 1 + 2\beta(t)}{4x(1-x)}.$$

One can notice that if $\alpha(t) < 1/2$ or $\beta(t) < 1/2$ then ρ is $-\infty$. A straightforward computation leads to (22). $\quad\square$

As a conclusion, if the mutation rate are sufficiently great *i.e.* greater than $1/2$, the relative entropy between two distributions of allele A goes to zero (exponentially fast if $\alpha(t)$ and $\beta(t)$ are greater than $1/2 + \delta$ for all t) whatever the initial distributions are.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*. Collection "Panoramas et Synthèses", SMF(2000) No. 10.

[2] D. Bakry, M. Émery, *Hypercontractivité de semi-groupes de diffusion*. CRAS Ser. I **299** (1984) 775–778.

[3] D. Bakry, L'hypercontractivité et son utilisation en théorie des semigroupes. *Lect. Notes Math.* **1581** (1994) 1–114.

[4] D. Bakry, On Sobolev and logarithmic Sobolev inequalities for Markov semigroups, in *New trends in stochastic analysis (Charingworth, 1994)*, River Edge, Taniguchi symposium, World Sci. Publishing, NJ (1997) 43–75.

[5] G.I. Barenblatt, *Scaling, self-similarity, and intermediate asymptotics.* Cambridge Texts in Applied Mathematics **14** Cambridge University Press (1996).

[6] J. Bricmont, A. Kupiainen and G. Lin, Renormalization group and asymptotics of solutions of nonlinear parabolic equations. *Comm. Pure Appl. Math.* **47** (1994) 893–922.

[7] D.Chafaï, Entropies, Convexity en Functional Inequalities. *Journal of Mathematics of Kyoto University* **44** (2004) 325–363.

[8] J.F. Collet, Extensive Lyapounov functionals for moment-preserving evolution equations. *C.R.A.S. Ser. I* **334** (2002) 429–434.

[9] P. Del Moral, M. Ledoux and L. Miclo, On contraction properties of Markov kernels. *Probab. Theory Related Fields* **126** (2003) 395–420.

[10] W.J. Ewens, *Mathematical population genetics. I*, Interdisciplinary Applied Mathematics, Vol 27. Springer-Verlag (2004).

[11] R. Kubo, *H-Theorems for Markoffian Processes*, in *Perspectives in Statistical Physics*, H.J. Raveché Ed., North Holland Publishing (1981).

[12] S. Kullback and R.A. Leibler, On Information and Sufficiency. *Ann. Math. Stat.* **22** (1951) 79–86.

[13] F. Otto and C. Villani, Generalization of an inequality by Talagrand, and links with the Sobolev Logarihmic Inequality. *J. Func. Anal.* **173** (2000) 361–400.

[14] M.S. Pinsker, *Information and Information Stability of Random Variables and Processes.* Holden-Day Inc. (1964).

[15] G. Toscani, Remarks on entropy and equilibrium states. *Appl. Math. Lett.* **12** (1999) 19–25.

[16] G. Toscani and C. Villani, On the trend to equilibrium for some dissipative systems with slowly increasing *a priori* bounds. *J. Stat. Phys.* **98** (2000) 1279–1309.