Probab. Theory Relat. Fields 126, 395–420 (2003)
Digital Object Identifier (DOI) 10.1007/s00440-003-0270-6

P. Del Moral · M. Ledoux · L. Miclo

On contraction properties of Markov kernels

Received: 31 October 2000 / Revised version: 21 February 2003 /
Published online: 12 May 2003 – © Springer-Verlag 2003

Abstract. We study Lipschitz contraction properties of general Markov kernels seen as operators on spaces of probability measures equipped with entropy-like "distances". Universal quantitative bounds on the associated ergodic constants are deduced from Dobrushin's ergodic coefficient. Strong contraction properties in Orlicz spaces for relative densities are proved under more restrictive mixing assumptions. We also describe contraction bounds in the entropy sense around arbitrary probability measures by introducing a suitable Dirichlet form and the corresponding modified logarithmic Sobolev constants. The interest in these bounds is illustrated on the example of inhomogeneous Gaussian chains. In particular, the existence of an invariant measure is not required in general.

1. Introduction and results

The purpose of this paper is to study general properties of contractions of Markov kernels without assumptions on the existence of an invariant probability measure. Our main motivation for this investigation comes from nonlinear filtering(cf [6]), where one would like to show that the system is forgetting its initialization (especially when the latter is not the true one) without nevertheless converging in large times to an invariant distribution. Actually, an invariant measure may not even exist, such as in noncompact settings (for annealed results) or due the time-inhomogeneity of the underlying signal (from the quenched point of view). The results presented here may also be used to study contraction properties of Feynman-Kac semigroups [7] or even the speed of convergence in the central limit theorem [13] These applications are however not developed here, and we actually only concentrate on the "abstract" aspects and results.

Start with a measurable space $(S, \mathcal{S})$ and denote by $\mathcal{M}$ (respectively $\mathcal{P}$, $\mathcal{M}_+$ and $\mathcal{M}_0$) the set of associated signed bounded measures (resp.probability measures,

P. Del Moral, M. Ledoux, L. Miclo: Institut de Mathématiques, Laboratoire de Statistique et Probabilités, UMR 5583 CNRS et Université Paul Sabatier 31062 Toulouse cedex 4, France. e-mail: {delmoral,ledoux,miclo}@math.ups-tlse.fr

L. Miclo also thanks the hospitality and support of the Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brasil, where part of this work was done.

Mathematics Subject Classification (2000): 60J05, 60J22, 37A30, 37A25, 39A11, 39A12, 46E39, 28A33, 47D07

Key words or phrases: Lipschitz contraction – Generalized relative entropy – Markov kernel – Dobrushin's ergodic coefficient – Orlicz norm – Dirichlet form – Spectral gap – Modified logarithmic Sobolev inequality – Inhomogeneous Gaussian chains – Loose of memory property

nonnegative bounded measures and elements of $\mathcal{M}$ whose total mass is equal to 0). These sets are naturally endowed with the total variation metric $\|\cdot\|$. However, one may consider other ways to measure the "distance" between, say, elements of $\mathcal{M}_+$. Let for example $\Phi : \mathbb{R}_+^2 \to \mathbb{R} \sqcup \{+\infty\}$ be a convex function satisfying $\Phi(1, 1) = 0$ and homogeneous in the sense that

$$\forall\, a,\ x,\ y \in \mathbb{R}_+, \qquad \Phi(ax, ay) = a\Phi(x, y).$$

Then for any $\mu, \nu \in \mathcal{M}_+$, we may consider the so-called Φ-relative entropy between μ and ν defined by

$$H_\Phi(\mu, \nu) = \int \Phi\left(\frac{d\mu}{d\lambda}, \frac{d\nu}{d\lambda}\right) d\lambda \tag{1}$$

where λ stands for any measure in $\mathcal{M}_+$ such that $\mu \ll \lambda$ and $\nu \ll \lambda$. As is classical and easy to see, due to the homogeneity property of Φ, the above definition does not depend on the choice of λ. This definition generalizes the classical choice of $\Phi(x, y) = x \ln(x/y)$ giving rise to the usual relative entropy (see below). To see that (1) is well-defined, let $(b_1, b_2) \in \mathbb{R}^2$ be a subderivative of Φ at $(1, 1)$, so that

$$\forall\, (x, y) \in \mathbb{R}_+^2, \qquad \Psi(x, y) = \Phi(x, y) - b_1(x - 1) - b_2(y - 1) \ \geq\ 0.$$

One thus sees that the integral in (1) is always well-defined with values in $\mathbb{R} \sqcup \{+\infty\}$ since it can be written as

$$H_\Phi(\mu, \nu) = \int \ \Psi\left(\frac{d\mu}{d\lambda}, \frac{d\nu}{d\lambda}\right) d\lambda + b_1(\mu(S) - 1) + b_2(\nu(S) - 1).$$

(In particular, H_Φ is nonnegative on $\mathcal{P}^2$ and would even be positive outside the diagonal if Φ were assumed to be strictly convex).

Consider now Markov kernel K on $(S, \mathcal{S})$, i.e. a map $S \times \mathcal{S} \to [0, 1]$ satisfying

- $\forall\, A \in \mathcal{S}$, $S \ni x \mapsto K(x, A)$ is measurable
- $\forall\, x \in S$, $\mathcal{S} \ni A \mapsto K(x, A)$ is a probability measure.

The kernel K may also be seen as a left linear transformation on $\mathcal{M}$, via the formula

$$\forall\, \mu \in \mathcal{M},\ \forall\, A \in \mathcal{S}, \qquad (\mu K)(A) = \int K(x, A)\, \mu(dx).$$

Note furthermore that the subsets $\mathcal{P}$, $\mathcal{M}_+$ and $\mathcal{M}_0$ are left invariant by this mapping.

Our main objective in this work will be to investigate, for a given kernel K, the contraction properties of the map $\mathcal{P} \ni \mu \mapsto \mu K \in \mathcal{P}$ with respect to the Φ-relative entropy H_Φ. (It should be emphasized that, in practice, it is often more rewarding to look at the iterated kernel K^n, for some $n \geq 2$.) To this task, define, for a fixed element ν in $\mathcal{P}$, $\alpha(\Phi, K, \nu)$ to be the best constant α such that for all $\mu \in \mathcal{P}$,

$$H_\Phi(\mu K, \nu K) \leq (1 - \alpha) H_\Phi(\mu, \nu).$$

Therefore, $1 - \alpha(\Phi, K, \nu)$ can be seen as a "moving" coefficient of contraction around ν for K relative to H_Φ. Our first result indicates that Dobrushin's ergodic coefficient

$$a(K) = 1 - \frac{1}{2} \sup_{x,y \in S} \|K(x, \cdot) - K(y, \cdot)\| \; \geq \; 0 \tag{2}$$

is a universal lower bound for $\alpha(\Phi, K, \nu)$.

Proposition 1.1. *For Φ, K and ν as before,*

$$\alpha(\Phi, K, \nu) \geq a(K).$$

In particular, $\alpha(\Phi, K, \nu) \geq 0$, which means that K is at least contractive:

$$\forall\, \mu,\ \nu \in \mathcal{P}, \qquad H_\Phi(\mu K, \nu K) \leq H_\Phi(\mu, \nu).$$

These lower bounds were first obtained by Cohen, Iwasa, Răuţu, Ruskai, Seneta and Zbăganu [4] in the case of a finite state space, from which the general case actually follows. (See also [17], where Zaharopol and Zbăganu present another extension of a key inequality involving Dobrushin's ergodic coefficient to study stability properties of $\mathbb{L}^1$-operators. These authors do not discuss however Lipschitz-type contraction inequalities.) With respect to the matrix-algebraic (and rather difficult to read) argument of [4], our proof, presented here in Section 3, is measure-theoretic, and actually easier and much more general in both the hypotheses and conclusions.

Dobrushin's coefficient thus appears through Proposition 1.1 as a very coarse ergodic constant and it is a strong hypothesis to ask for its positivity. One of the simplest way for bounding it away from 0 is to resort to the even more restrictive ultra-mixing condition

(H) *For any $x, y \in S$, $K(x, \cdot) \sim K(y, \cdot)$ and there exists $\epsilon > 0$ such that*

$$\forall\, x, y \in S \qquad \frac{dK(x, \cdot)}{dK(y, \cdot)} \geq \epsilon \qquad K(y, \cdot)\text{–a.s.}$$

This hypothesis is well-known to imply that $a(K) \geq \epsilon$ ([9]). Our second objective in this work will actually be to describe this result (and more general ones) as a special occurrence of contraction properties for relative densities in Orlicz spaces valid under (H). More precisely, let $\psi : \mathbb{R}_+ \to \mathbb{R}$ be a Young function ($\psi(0) = 0$, increasing and convex). For $\nu \in \mathcal{P}$, denote by $\mathbb{L}^\psi(\nu)$ the associated Orlicz space (cf [11] as a general reference), that is the set of measurable functions $f : S \to \mathbb{R}$ for which there exists $a > 0$ such that $\int \psi(f/a)\, d\nu < +\infty$. Equipped with the norm

$$\|f\|_{\mathbb{L}^\psi(\nu)} = \inf \left\{ a > 0 \; : \; \int \psi(f/a)\, d\nu \leq 1 \right\},$$

$\mathbb{L}^\psi(\nu)$ is a Banach space.

Proposition 1.2. *Let K be a Markov kernel on $(S, \mathcal{S})$ and let $\mu, \nu \in \mathcal{P}$ be such that $\mu \ll \nu$ and $d\mu/d\nu \in \mathbb{L}^{\psi}(\nu)$. Then $\mu K \ll \nu K$ and $d\mu K/d\nu K \in \mathbb{L}^{\psi}(\nu K)$. Furthermore, if (H) is satisfied, we have*

$$\left\| \frac{d\mu K}{d\nu K} - 1 \right\|_{\mathbb{L}^{\psi}(\nu K)} \leq (1-\epsilon) \left\| \frac{d\mu}{d\nu} - 1 \right\|_{\mathbb{L}^{\psi}(\nu)}$$

(while the inequality with $\epsilon = 0$ always holds).

The proof of Proposition 1.2 is presented in Section 4. As we will see, the case $\psi(\cdot) = |\cdot|$ just amounts to the bound $a(K) \geq \epsilon$.

In the same spirit, let us mention that if we take $\Phi = \Phi_0$ in Proposition 1.1, where

$$\forall\, (x, y) \in \mathbb{R}_+^2, \qquad \Phi_0(x, y) = |x - y| \tag{3}$$

then we may show that

$$a(K) = \inf_{\nu \in \mathcal{P}} \alpha(\Phi_0, K, \nu). \tag{4}$$

Our final task in this work will be to look more closely at two other interesting choices of Φ, namely

$$\forall\, x,\ y \in \mathbb{R}_+, \qquad \begin{array}{l} \Phi_1(x, y) = (x - y)(x/y - 1) \\ \Phi_2(x, y) = x \ln(x/y) \end{array} \tag{5}$$

(with the usual convention $0 \cdot \infty = 0$ adopted throughout the paper). These choices are indeed associated to the important notions of relative $\mathbb{L}^2$-norm and relative entropy. In particular, as a main result, we will produce lower bounds on $\alpha(\Phi_1, K, \nu)$ and $\alpha(\Phi_2, K, \nu)$ respectively in terms of spectral gaps and modified Sobolev logarithmic constants. These results extend well-known ones in the case where ν is assumed to be invariant, cf [10]). It is worthwhile emphasizing that the usual logarithmic Sobolev constant is of no use in general to describe $\alpha(\Phi_2, K, \nu)$ and this is why we emphasize here modified Sobolev logarithmic constants. We postpone until Section 5 the details about the relevant underlying Dirichlet forms, that involve a number of operators described in next preliminary section. In the last part finally, a Gaussian example is discussed showing the advantage in this setting of the modified logarithmic Sobolev constant with respect to the traditional one. (See for instance [8] as a general reference or [12] for the corresponding considerations on entropy dissipation in the case of an invariant probability measure ν). We underline through this example the gain one can obtain not working with an invariant measure (even if one exists), emphasizing thus the importance of the Lipschitz point of view.

2. A few reminders

In this section, we recall some useful facts about total variation and entropy distances, and general Markov operator theory. General references for this section are [9], [14].

Let us first come back to the definition of Φ-entropy. Recall the homogeneous convex function $\Phi : \mathbb{R}_+^2 \to \mathbb{R} \sqcup \{+\infty\}$ with $\Phi(1,1) = 0$. Denote by $\phi : \mathbb{R}_+ \to \mathbb{R} \sqcup \{+\infty\}$ the convex function $\phi(x) = \Phi(x, 1)$, $x \in \mathbb{R}_+$. By the homogeneity hypothesis, Φ is almost determined by ϕ, in fact only the values $\Phi(x, 0)$ for $x > 0$ (or equivalently just $\Phi(1, 0)$), are missing. In most applications, the natural convention is that $\Phi(1, 0) = +\infty$. This somewhat innocent looking assumption admits interesting consequences.

Lemma 2.1. *Assume that* $\Phi(1, 0) = +\infty$. *Then,*

$$\forall\, \mu,\ \nu \in \mathcal{M}_+, \qquad H_\Phi(\mu, \nu) = \begin{cases} \int \phi\left(\frac{d\mu}{d\nu}\right) d\nu & \text{if } \mu \ll \nu \\ +\infty & \textit{otherwise.} \end{cases}$$

The right-hand side in this lemma is sometimes called ϕ-divergence in the sense of Csiszár [5].

Proof. Let $\mu,\ \nu \in \mathcal{M}_+$ be given and denote by $\mu = \mu_1 + \mu_2$ the Lebesgue decomposition of μ with respect to ν, that is $\mu_1 \ll \nu$ and $\mu_2 \perp \nu$. Consider $A \in \mathcal{S}$ such that $\nu(A^{\mathrm{c}}) = 0 = \mu_2(A)$. To compute $H_\Phi(\mu, \nu)$, choose $\lambda = \nu + \mu_2$ to get

$$H_\Phi(\mu, \nu) = \int_A \Phi\left(\frac{d\mu_1}{d\nu}, 1\right) d\nu + \int_{A^{\mathrm{c}}} \Phi(1, 0)\, d\mu_2 = \Phi(\mu_1, \nu) + \Phi(1, 0)\mu_2(S).$$

In particular, if $\mu_2(S) > 0$ we deduce that $H_\Phi(\mu, \nu) = +\infty$. Otherwise, choosing $\lambda = \nu$ we get

$$H_\Phi(\mu, \nu) = \int \phi\left(\frac{d\mu}{d\nu}\right) d\nu.$$

The lemma is established. □

Note that conversely, if $\phi : \mathbb{R}_+ \to \mathbb{R} \sqcup \{+\infty\}$ is a given convex function, then $l_0 = \lim_{t\to+\infty} \phi(t)/t$ exists in $\mathbb{R} \sqcup \{+\infty\}$. Then, for any $l \in [l_0, +\infty]$, the function

$$\forall\, (x, y) \in \mathbb{R}_+^2, \qquad \Phi(x, y) = \begin{cases} y\phi(x/y) & \text{if } y > 0 \\ lx & \text{if } y = 0 \end{cases} \tag{6}$$

is convex. In what follows, we will always take $l = l_0$.

Here are a few examples of classical Φ-entropies arising in the literature and which illustrate our work. For the cases described in (5), we obtain respectively, the $\mathbb{L}^2$-relative norm

$$\forall\, \mu, \nu \in \mathcal{P}, \qquad H_{\Phi_1}(\mu, \nu) = \begin{cases} \int \left(\frac{d\mu}{d\nu} - 1\right)^2 d\nu & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}$$

and the usual (Boltzmann-Shannon-Kullback)-entropy

$$\forall\, \mu, \nu \in \mathcal{P}, \qquad H_{\Phi_2}(\mu, \nu) = \mathrm{Ent}(\mu|\nu) = \begin{cases} \int \ln\left(\frac{d\mu}{d\nu}\right) d\mu & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise.} \end{cases}$$

These examples can be generalized in two directions: the Havrda-Charvat entropy of order $p > 1$ corresponds to the function $\phi : \mathbb{R}_+ \ni t \mapsto \frac{1}{p-1}(t^p - 1)$, i.e.

$$\forall\, \mu \ll \nu, \qquad H_\Phi(\mu, \nu) = C_p(\mu|\nu) = \frac{1}{p-1}\left[\int \left(\frac{d\mu}{d\nu}\right)^p d\nu - 1\right]$$

(which is converging to $\mathrm{Ent}(\mu|\nu)$ as p goes to 1_+). For $p \geq 1$, the $\mathbb{L}^p$-relative norm is associated to $\phi : \mathbb{R}_+ \ni t \mapsto |t-1|^p$, i.e.

$$\forall\, \mu \ll \nu, \qquad H_\Phi(\mu, \nu) = \left\| \frac{d\mu}{d\nu} - 1 \right\|_{\mathbb{L}^p(\nu)}^p = \int \left| \frac{d\mu}{d\nu} - 1 \right|^p d\nu$$

(the case $p = 1$ corresponding to Φ_0). Observe that for $\phi(t) = \frac{1}{p-1}(t^{1-p} - 1)$ with $p > 1$ (resp. $\phi(t) = \ln(1/t)$), we also have the identities $H_\Phi(\mu, \nu) = C_p(\nu|\mu)$ (resp. $H_\Phi(\mu, \nu) = \mathrm{Ent}(\nu|\mu)$) for $\mu \ll \nu$).

The two preceding examples are occurrences where $\Phi(1, 0) = +\infty$ is not satisfied. The Hellinger integrals of order $\theta \in (0, 1)$ are some other instances. They arise with the choice of $\phi(t) = t - t^\theta$, $t \in \mathbb{R}_+$. Hence, for any $\mu,\ \nu \in \mathcal{P}$ and λ such that $\mu \ll \lambda$, $\nu \ll \lambda$,

$$H_\Phi(\mu, \nu)11 - \int \left(\frac{d\mu}{d\lambda}\right)^\theta \left(\frac{d\nu}{d\lambda}\right)^{1-\theta} d\lambda.$$

(Alternatively, $H_\Phi(\mu, \nu) = 1 - \int \left(\frac{d\mu}{d\nu}\right)^\theta d\nu$ if $\mu \ll \nu$). In the special case $\theta = 1/2$, this definition coincides with the classical Kakutani-Hellinger integral.

The most important example of interest where $\Phi(1, 0) \neq +\infty$ is probably $\Phi = \Phi_0$ defined in (3). In this case, for any $\mu,\ \nu \in \mathcal{M}_+$,

$$H_{\Phi_0}(\mu, \nu) = \|\mu - \nu\| = \sup_{f \in \mathcal{B}_\mathrm{b},\ \|f\|_\infty = 1} (\mu - \nu)(f)$$

where $\mathcal{B}_\mathrm{b}$ is the set of bounded measurable functions on $(S, \mathcal{S})$ endowed with the supremum norm $\|\cdot\|_\infty$. This is a very basic result of measure theory which can be shown via the Hahn-Jordan decomposition of $\mu - \nu \in \mathcal{M}$ in terms of nonnegative measures. For this particular choice of $\Phi = \Phi_0$, Proposition 1.1 immediately follows from the seminal paper [9] of Dobrushin who proved that for any Markov kernel K,

$$1 - a(K) = \sup_{x \neq y \in S} \frac{\left\|\delta_x K - \delta_y K\right\|}{\left\|\delta_x - \delta_y\right\|} = \sup_{\mu \neq \nu \in \mathcal{P}} \frac{\|\mu K - \nu K\|}{\|\mu - \nu\|}. \tag{7}$$

The same argument also proves (4). To deduce that $a(K) \geq \epsilon$ from the Proposition 1.2 for $\varphi = |\cdot|$ is quite similar.

Next we present some further operator interpretations of a Markov kernel K. Writing

$$\forall\, x \in S, \qquad K(f)(x) = \int f(y)\, K(x, dy)$$

the kernel K can be seen as a right-acting operator on $f \in \mathcal{B}_{\mathrm{b}}$ or $\mathcal{B}_+$ (the set of $\bar{\mathbb{R}}_+$-valued measurable functions defined on $(S, \mathcal{S})$). Given a probability measure $\nu \in \mathcal{P}$, for $f \in \mathcal{B}_{\mathrm{b}}$ and $1 \leq p < \infty$, Hölder's inequality shows that

$$\nu((Kf)^p) \leq \nu(K(f^p)) = (\nu K)(f^p).$$

Hence K can also be extended as a linear map from $\mathbb{L}^p(\nu K)$ into $\mathbb{L}^p(\nu)$. The dual operator K_ν^* from $\mathbb{L}^q(\nu)$ to $\mathbb{L}^q(\nu K)$, where $q = p/(p-1) \in (1, +\infty)$ is the conjugate exponent of $1 < p < +\infty$, satisfies by definition

$$\forall\, f \in \mathbb{L}^q(\nu),\ \forall\, g \in \mathbb{L}^p(\nu K), \qquad (\nu K)(g K_\nu^*(f)) = \nu(K(g) f).$$

Since the dual of $\mathbb{L}^\infty(\nu)$ cannot be identified with $\mathbb{L}^1(\nu)$, this approach does not allow however for the construction of K_ν^* on the latter space. We have to work it out directly as is classical in the field of Markov kernels (see for instance [14] and its bibliography). First note that if $\mu \in \mathcal{M}$ satisfies $\mu \ll \nu$, then $\mu K \ll \nu K$. Indeed, if $A \in \mathcal{S}$ is such that $(\nu K)(A) = 0$, the function $K(\mathbb{1}_A)$ is ν-negligible, hence μ-negligible and finally $(\mu K)(A) = 0$. Set then

$$\forall\, f \in \mathbb{L}^1(\nu), \qquad K_\nu^*(f) = \frac{d(f\nu)K}{d\nu K}$$

so that we still have

$$\forall\, f \in \mathbb{L}^1(\nu),\ \forall\, g \in \mathbb{L}^\infty(\nu K), \qquad (\nu K)(g K_\nu^*(f)) = \nu(K(g) f). \tag{8}$$

Notice that the dual operators K_ν^* on $\mathbb{L}^p(\nu)$, for $1 < p < +\infty$, can be seen as the restrictions to these sets of the ones on $\mathbb{L}^1(\nu)$.

Actually, K_ν^* is almost a Markov kernel. Recall that for any two fixed probability measures $\mu,\ \nu$, a linear operator R from $\mathbb{L}^1(\nu)$ into $\mathbb{L}^1(\mu)$ is said to be (generalized) Markovian if

- $R(f) \geq 0$, μ-a.s. for any $f \in \mathbb{L}^1(\nu)$ such that $f \geq 0$ ν-a.s.
- $R(\mathbb{1}) = \mathbb{1}$ μ-a.s.

Lemma 2.2. *The operators $K_\nu^* : \mathbb{L}^1(\nu) \to \mathbb{L}^1(\nu K)$ and $K K_\nu^* : \mathbb{L}^1(\nu) \to \mathbb{L}^1(\nu)$ are Markovian.*

Proof. In view of (8), $(\nu K)(g K_\nu^*(f)) \geq 0$ for any nonnegative $f \in \mathbb{L}^1(\nu)$ and any nonnegative $g \in \mathbb{L}^\infty(\nu K)$. Furthermore, for any $g \in \mathbb{L}^\infty(\nu K)$,

$$(\nu K)(g K_\nu^*(\mathbb{1})) = \nu(K(g)) = (\nu K)(g)$$

thus implying that $K_\nu^*(\mathbb{1}) = \mathbb{1}$, νK-a.s. Finally, since $K : \mathbb{L}^1(\nu K) \to \mathbb{L}^1(\nu)$ is Markovian, the same remains true by composition for $K K_\nu^* : \mathbb{L}^1(\nu) \to \mathbb{L}^1(\nu)$. □

As a consequence of the preceding lemma, we are allowed to use Jensen's inequality for K_ν^*. Hence, for any function $f \in \mathcal{B}_+$ and any convex function $\varphi : \mathbb{R}_+ \to \mathbb{R}$, $K_\nu^*(\varphi(f)) \geq \varphi(K_\nu^*(f))$ νK a.s. Another interesting feature is that ν is invariant with respect to $K K_\nu^*$, meaning that for all $f \in \mathbb{L}^1(\nu)$, $\nu(K K_\nu^*(f)) = \nu(f)$. This is indeed an immediate consequence of the fact that

$$\nu(K K_\nu^*(f)) = (\nu K)(\mathbb{1} K_\nu^*(f)) = \nu(f K(\mathbb{1})) = \nu(f).$$

The probability measure ν is even reversible for the Markov operator KK_ν^* in the sense that

$$\forall\, f, g \in \mathbb{L}^2(\nu), \qquad \nu(fKK_\nu^*(g)) = \nu(gKK_\nu^*(f)).$$

Indeed, we have in the same way

$$\begin{aligned} \nu[fKK_\nu^*(g)] &= ((f\nu)K)[K_\nu^*(g)] \\ &= (K_\nu^*(f)(\nu K))[K_\nu^*(g)] \\ &= (\nu K)[K_\nu^*(f)K_\nu^*(g)] \\ &= \nu[gKK_\nu^*(f)]. \end{aligned}$$

It follows that the Dirichlet form given by

$$\forall\, f \in \mathbb{L}^2(\nu), \qquad \mathcal{E}_{\nu,KK_\nu^*}(f, f) = \nu[f(\mathrm{Id} - KK_\nu^*)(f)] \tag{9}$$

is symmetric. As announced in the introduction, this observation will be our starting point for defining spectral gaps and modified logarithmic Sobolev constants with respect to this Dirichlet form (see Section 5). Note that if ν is assumed to be invariant for K (i.e. $\nu K = \nu$), then K_ν^* coincides on $\mathbb{L}^2(\nu)$ with the adjoint of K on $\mathbb{L}^2(\nu)$. We recover thus in this case the Dirichlet form considered in [10] and [12].

To conclude this preliminary section, let us briefly describe K_ν^* in cases where some absolute continuity assumption is available. More precisely, assume that

$$\forall\, x \in S, \qquad K(x, \cdot) \ll \nu K.$$

As a consequence, the probability measure $\nu(dx)K(x, dy)$ on $S \times S$ is absolutely continuous with respect to $\nu(dx)\nu K(dy)$. By a slight abuse of notation, denote its density by

$$\frac{dK(x,\cdot)}{d\nu K}(y).$$

Expanding the right-hand side of (8) shows that

$$K_\nu^*(x, dy) = \frac{dK(y,\cdot)}{d\nu K}(x)\nu(dy).$$

In the same way, one can check that

$$KK_\nu^*(x, dy) = G(x, y)\nu(dy)$$

with, $\nu \otimes \nu$-a.s. in $(x, y) \in S \times S$,

$$G(x, y) = \int \frac{dK(x,\cdot)}{d\nu K}(z)\frac{dK(y,\cdot)}{d\nu K}(z)\, \nu K(dz).$$

3. On Dobrushin's ergodic coefficient

In this section, we present the proof of Proposition 1.1. According to (7), and via an obvious homogeneity property, we already know that for any Markov kernel K,

$$\forall\, \eta \in \mathcal{M}_0, \qquad \|\eta K\| \leq (1 - a(K))\, \|\eta\|\,.$$

Our first task will be to extend this inequality to the whole space $\mathcal{M}$.

Lemma 3.1. *For any $\eta \in \mathcal{M}$,*

$$\|\eta K\| \leq (1 - a(K))\, \|\eta\| + a(K)\, |\eta(S)|\,.$$

Proof. Again by homogeneity, it is sufficient to check this inequality in the case $\eta(S) \geq 0$. Next we note that if $\eta \in \mathcal{M}_+$, then $\|\eta\| = \eta(S)$ and $\|\eta K\| = |\eta K(S)| = \eta(S)$. Hence the lemma also holds in this case.

In the general case, let $\eta = \eta_+ - \eta_-$ be the Hahn-Jordan decomposition of η. We can also write $\eta = \eta_1 + \eta_2$, with

$$\eta_1 = \frac{\eta_+(S) - \eta_-(S)}{\eta_+(S)} \eta_+ \in \mathcal{M}_+, \qquad \eta_2 = \frac{\eta_-(S)}{\eta_+(S)} \eta_+ - \eta_- \in \mathcal{M}_0.$$

Therefore,

$$\begin{aligned} \|\eta K\| &\leq \|\eta_1 K\| + \|\eta_2 K\| \\ &\leq (1 - a(K))(\|\eta_1\| + \|\eta_2\|) + a(K)(|\eta_1(S)| + |\eta_2(S)|). \end{aligned}$$

Since $|\eta_1(S)| + |\eta_2(S)| = \eta_1(S) = \eta(S)$ and

$$\begin{aligned} \|\eta\| = \eta_+(S) + \eta_-(S) = {} & \frac{\eta_+(S) - \eta_-(S)}{\eta_+(S)}\, \eta_+(S) + \frac{\eta_-(S)}{\eta_+(S)}\, \eta_+(S) \\ & + \eta_-(S) = \|\eta_1\| + \|\eta_2\|\,, \end{aligned}$$

the conclusion follows. □

Remark 3.2. Recalling that for any signed measure $\eta \in \mathcal{M}$,

$$2\eta_+(S) \wedge \eta_-(S) = \|\eta\|_{\mathrm{tv}} - |\eta(S)|$$

and taking into account that for any Markov kernel K, we have $\eta K(S) = \eta(S)$, Lemma 3.1 may also be written as

$$(\eta K)_+(S) \wedge (\eta K)_-(S) \leq (1 - a(K))\, (\eta_+(S) \wedge \eta_-(S))\,.$$

In particular, if $a(K) > 0$, using that both sequences $((\eta K^n)_+(S))_{n\geq 0}$ and $((\eta K^n)_-(S))_{n\geq 0}$ are nonincreasing, either $\lim_{n\to\infty}(\eta K^n)_+(S) = 0$ or $\lim_{n\to\infty}(\eta K^n)_-(S)$=0, depending upon the sign of $\eta(S)$. If $\eta(S) = 0$, both are converging to zero, hence $\lim_{n\to+\infty} \|\eta K^n\| = 0$. There is thus a loss of the initial condition η_+ or η_-, which was the original result of Dobrushin [9].

The next result is essentially Lemma 3.3 of J.E. Cohen, Y. Iwasa, G. Răuţu, M.B. Ruskai, E. Seneta and G. Zbăganu [4], except that there the measures are assumed to have compact support.

Lemma 3.3. *Let m_1, m_2 be two bounded measures on the Borel sets of $\mathbb{R}$ admitting a first moment and such that*

- *m_1 and m_2 are acting in the same manner on affine maps:*

$$m_1(\mathbb{R}) = m_2(\mathbb{R}) \quad \textit{and} \quad \int t\, m_1(dt) = \int t\, m_2(dt).$$

- *For any $s \in \mathbb{R}$,*

$$\int |t-s|\, m_1(dt) \le \int |t-s|\, m_2(dt).$$

Then for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$, $m_1(\varphi) \le m_2(\varphi)$ (the value $+\infty$ being allowed).

Proof. The main modification with respect to the approach of [4] is that we replace uniform convergence by monotone convergence. Let φ be a given convex function on $\mathbb{R}$. One can find two sequences $(x_i)_{i\in\mathbb{Z}^*}$ and $(k_i)_{i\in\mathbb{Z}^*}$ of nonnegative real numbers such that if we set for $n \ge 1$,

$$\forall\, t \in \mathbb{R}, \qquad \varphi_n(t) = \varphi(0) + \varphi'_{\mathrm{r}}(0)t + \sum_{0<i\le n} k_i(t-x_i)_+ + \sum_{-n\le i<0} k_i(t+x_i)_-$$

where φ'_{r} is the right derivative of φ and $(\cdot)_+$ and $(\cdot)_-$ are respectively the nonnegative and nonpositive parts, then $(\varphi_n)_{n\ge1}$ is an increasing sequence converging to φ. Indeed,

$$\forall\, t \in \mathbb{R}, \qquad \varphi(t) = \varphi(0) + \varphi'_{\mathrm{r}}(0)t + \int_0^t \varphi'_{\mathrm{r}}(s) - \varphi'_{\mathrm{r}}(0)\, ds.$$

Since $\mathbb{R} \ni t \mapsto \varphi'_{\mathrm{r}}(t) - \varphi'_{\mathrm{r}}(0)$ is nondecreasing and equal to 0 at $t = 0$, it suffices to approximate from below this function by nondecreasing step functions (for example constant on appropriate dyadic intervals).

Next coming back to m_1 and m_2, for any $i \ge 1$,

$$\begin{aligned}\int k_i(t-x_i)_+\, m_1(dt) &= \frac{1}{2}\int k_i\,|t-x_i|\; m_1(dt) + \frac{1}{2}\int k_i(t-x_i)\, m_1(dt)\\ &\le \frac{1}{2}\int k_i\,|t-x_i|\; m_2(dt) + \frac{1}{2}\int k_i(t-x_i)\, m_2(dt)\\ &= \int k_i(t-x_i)_+\, m_2(dt).\end{aligned}$$

Similarly for the nonpositive parts,

$$\int k_{-i}(t+x_{-i})_-\, m_1(dt) \le \int k_{-i}(t+x_{-i})_-\, m_2(dt).$$

As a consequence, for any $n \geq 1$,

$$\int \varphi_n(t)\, m_1(dt) \leq \int \varphi_n(t)\, m_2(dt)$$

and letting n grow to infinity, we conclude by the monotone convergence theorem. The lemma is established. □

In its present form, the preceding lemma would only imply Proposition 1.1 for probability measures μ absolutely continuous with respect to ν. In order to reach the full conclusion, we need to slightly modify it.

Lemma 3.4. *Let m_1, m_2 be two bounded measures on the quadrant $(\mathbb{R}_+^2, \mathcal{R}_+^{\otimes 2})$ admitting a first moment and such that*

- *m_1 and m_2 are acting in the same manner on affine mappings:*

$$m_1(\mathbb{R}_+^2) = m_2(\mathbb{R}_+^2), \quad \int s\, m_1(ds, dt) = \int s\, m_2(ds, dt)$$
$$\text{and} \quad \int t\, m_1(ds, dt) = \int t\, m_2(ds, dt).$$

- *For any $a, b \in \mathbb{R}$,*

$$\int |as - bt|\, m_1(ds, dt) \leq \int |as - bt|\, m_2(ds, dt).$$

Then for any homogeneous convex function Φ on $\mathbb{R}_+^2$, $m_1(\Phi) \leq m_2(\Phi)$ (the value $+\infty$ being again allowed).

Proof. We use the representation (6) of Φ by a convex function ϕ together with the approximation procedure described in Lemma 3.3 for ϕ. Then Lemma 3.4 holds as soon as for all $a, b \in \mathbb{R}$,

$$\int (as - bt)_+\, m_1(ds, dt) \leq \int (as - bt)_+\, m_2(ds, dt) \tag{10}$$
$$\int \mathbb{1}_{\{t=0\}} s\, m_1(ds, dt) \leq \int \mathbb{1}_{\{t=0\}} s\, m_2(ds, dt)$$

which are immediate consequences of the hypotheses. (Note that the last inequality is needed in case $l > l_0$. It can actually be deduced from the first condition by letting there b going to infinity with $a = 1$.) □

Proposition 1.1 now follows quite easily from the preceding. Let $\mu,\ \nu \in \mathcal{P}$ be given and let $\lambda \in \mathcal{P}$ be such that $\mu \ll \lambda$, $\nu \ll \lambda$. We make use of Lemma 3.4 for m_1 and m_2 on $(\mathbb{R}_+^2, \mathcal{R}_+^{\otimes 2})$ acting on bounded measurable functions by $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ by

$$m_1(\varphi) = \int \varphi\left(\frac{d\mu K}{d\lambda K}, \frac{d\nu K}{d\lambda K}\right) d\lambda K,$$
$$m_2(\varphi) = (1 - a(K)) \int \varphi\left(\frac{d\mu}{d\lambda}, \frac{d\mu}{d\lambda}\right) d\lambda + a(K)\varphi(1, 1).$$

For these measures, the first condition of Lemma 3.4 is immediate. The second hypothesis amounts to check that for all $a, b \in \mathbb{R}$,

$$\int \left| a\frac{d\mu K}{d\lambda K} - b\frac{d\nu K}{d\lambda K} \right| \, d\lambda K \le (1 - a(K)) \int \left| a\frac{d\mu}{d\lambda} - b\frac{d\nu}{d\lambda} \right| \, d\nu + a(K) \, |a - b| \, .$$

In terms of the total variation distance,

$$\|(a\mu - b\nu)K\| \le (1 - a(K))\|a\mu - b\nu\| + a(K) \, |a - b| \, .$$

But this clear from Lemma 3.1 since $(a\mu - b\nu)(S) = a - b$. Therefore, Proposition 1.1 is proved.

4. On Orlicz norms

The spirit of Proposition 1.2 is quite similar to that of Proposition 1.1. For example, if $\psi(t) = t^p, t \ge 0$, for a fixed $p \ge 1$, Proposition 1.1 indicates that for any Markov kernel K and any two probability measures $\mu \ll \nu$,

$$\begin{aligned} \left\| \frac{d\mu K}{d\nu K} - 1 \right\|_{\mathbb{L}^p(\nu K)} &\le (1 - \alpha(\Phi, K, \nu))^{\frac{1}{p}} \left\| \frac{d\mu}{d\nu} - 1 \right\|_{\mathbb{L}^p(\nu)} \\ &\le (1 - a(K))^{\frac{1}{p}} \left\| \frac{d\mu}{d\nu} - 1 \right\|_{\mathbb{L}^p(\nu)} \end{aligned}$$

where $\Phi(x, y) = (x/y^{1-1/p} - y^{1/p})^p$, $(x, y) \in \mathbb{R}_+^2$. Since as we have seen (H) implies $a(K) \ge \epsilon$, we already have that

$$\left\| \frac{d\mu K}{d\nu K} - 1 \right\|_{\mathbb{L}^p(\nu K)} \le (1 - \epsilon)^{\frac{1}{p}} \left\| \frac{d\mu}{d\nu} - 1 \right\|_{\mathbb{L}^p(\nu)} .$$

This is just a little bit less precise than what is expressed by Proposition 1.2. The proof of Proposition 1.2 we will present is however completely different and in some respect aspects more direct.

More generally, let us also mention that qualitatively speaking, at least under the doubling condition for the Orlicz function ψ (i.e. there exists a constant $k > 1$ such that for all $x \ge 0$ large enough, $\psi(2x) \le k\psi(x)$), the convergences

$$\lim_{n\to\infty} H_\Phi(\mu K^n, \nu K^n) = 0$$

where Φ is the smallest convex function on $\mathbb{R}_+^2$ upper bounding $\mathbb{R}_+ \times \mathbb{R}_+^* \ni (x, y) \mapsto y\psi(x/y - 1)$, and

$$\lim_{n\to\infty} \left\| \frac{d\mu K^n}{d\nu K^n} - 1 \right\|_{\mathbb{L}^\psi(\nu K^n)} = 0$$

are equivalent (see for instance the proof of Theorem 9.4 p. 83 of [11]). Thus $\sum_{n\ge 0} \alpha(\Phi, K, \nu K^n) = +\infty$, and in particular $a(K) > 0$ is sufficient to insure this behavior. However, this type of argument does not produce any kind of quantitative exponential bound for this convergence. On the other hand, quantitative bounds will require the more restrictive hypothesis (H).

We start by rewriting hypothesis (H) in terms of the dual operators K_ν^*. Below, ϵ will always refer to the strictly positive constant of (H).

Lemma 4.1. *Condition (H) is equivalent to*

$$\forall\, \nu \in \mathcal{P}, \qquad K_\nu^* \geq \epsilon\, \nu \quad \nu K - a.s.$$

Proof. We aim to show that (H) is equivalent to

$$\forall\, f \in \mathbb{L}^1_+(\nu),\ \forall\, g \in \mathbb{L}^\infty_+(\nu K), \qquad (\nu K)[(K_\nu^*(f) - \epsilon\nu(f))g] \geq 0$$

(where the $+$ in subscripts indicate that we are only considering nonnegative functions). Now, by definition of K_ν^*,

$$(\nu K)[K_\nu^*(f - \epsilon\nu(f))g] = \nu[K(g)(f - \epsilon\nu(f))] = \nu[f(K(g) - \epsilon\nu(Kg))].$$

In other words, this is equivalent to saying that $f \in \mathbb{L}^1_+(\nu)$ and $g \in \mathbb{L}^\infty_+(\nu K)$,

$$\forall\, A \in \mathcal{S}, \qquad K(A) \geq \epsilon(\nu K)(A) \quad \nu\text{-a.s.} \tag{11}$$

To see that 11 implies (H), take $\nu = \eta\delta_x + (1-\eta)\delta_y$ where $0 < \eta < 1$ and $x, y \in S$ to get that, for any $A \in \mathcal{S}$,

$$K(x, A) \geq \epsilon(\eta K(x, A) + (1 - \eta)K(y, A))$$

that is

$$K(x, A) \geq \frac{\epsilon(1-\eta)}{1-\epsilon\eta} K(y, A).$$

Condition (H) follows by letting η tend to 0_+. Conversely, (H) indicates that for any $x, y \in S$ and $A \in \mathcal{S}$, $K(x, A) \geq \epsilon K(y, A)$. Integrating with respect to $\nu(dy)$, we end up with (11), which finally has to be satisfied for all $x \in S$ and not only $\nu(dx)$-a.s. Lemma 4.1 is thus established. □

The preceding characterization leads to the following result, which is the main step needed for the proof of Proposition 1.2.

Proposition 4.2. *Under assumption (H), for any probability measure $\nu \in \mathcal{P}$, any Markov kernel K, any convex function $\varphi : \mathbb{R} \to \mathbb{R}$ and any function $f \in \mathbb{L}^1(\nu)$,*

$$\int \varphi(K_\nu^*[f - \nu(f)])\, d\nu K \leq \int \varphi((1-\epsilon)(f - \nu(f)))\, d\nu.$$

Proof. Let ν, K and φ be as given above. Due to Lemma 4.1, the operator $R = (1-\epsilon)^{-1}(K_\nu^* - \epsilon\nu)$ is Markovian. Now, ν is again KR-invariant since

$$\nu K R = \nu K \frac{K_\nu^* - \epsilon\nu}{1-\epsilon} = \frac{\nu K K_\nu^*}{1-\epsilon} - \frac{\epsilon\nu}{1-\epsilon} = \frac{\nu}{1-\epsilon} - \frac{\epsilon\nu}{1-\epsilon} = \nu.$$

By convexity, for any $x, y \in S$ and $f \in \mathbb{L}^1(\nu)$

$$\begin{aligned}
&\varphi((1-\epsilon)(f(x)-\nu(f))) \\
&\quad \geq \varphi(K_\nu^*(f-\nu(f))(y)) + \varphi'_{\mathrm{d}}(K_\nu^*(f-\nu(f))(y)) \\
&\quad\quad \times [(1-\epsilon)(f(x)-\nu(f)) - K_\nu^*(f-\nu(f))(y)] \\
&\quad = \varphi(K_\nu^*(f-\nu(f))(y)) + (1-\epsilon)\varphi'_{\mathrm{d}}(K_\nu^*(f-\nu(f))(y)) \\
&\quad\quad \times [(f(x)-\nu(f)) - R(f-\nu(f))(y)].
\end{aligned}$$

(Actually, this is only satisfied $\nu(dx)$-a.s. and $\nu K(dy)$-a.s. since $f(x)$ and $K_\nu^*(f-\nu(f))(y)$ are merely defined this way). Integrating with respect to $R(y, dx)$ (which rigorously means letting act R in that way), we get

$$R[\varphi((1-\epsilon)(f(\cdot)-\nu(f)))](y) \geq \varphi(K_\nu^*(f-\nu(f))(y)).$$

The conclusion follows by integrating one more time with respect to $(\nu K)(dy)$:

$$\begin{aligned}
\int \varphi((1-\epsilon)(f-\nu(f)))\, d\nu &= \int R[\varphi((1-\epsilon)(f-\nu(f)))]\, d(\nu K) \\
&\geq \int \varphi(K_\nu^*(f-\nu(f))(y))\, \nu K(dy).
\end{aligned}$$

The proof is complete. □

We are now ready to prove Proposition 1.2 that follows rather easily from the preceding. Let ψ be the Young function in Proposition 1.2 and let $\mu,\ \nu$ be such that $d\mu/d\nu \in \mathbb{L}^\psi(\nu)$. Apply the previous proposition to $f = a^{-1}d\mu/d\nu$ for $a > 0$ (so that $K_\nu^*(f-\nu(f)) = a^{-1}((d\mu K)/(d\nu K) - 1))$. The result immediately follows from the definition of the norm $\|\cdot\|_{\mathbb{L}^\psi(\nu)}$.

It is worthwhile mentioning that Proposition 1.2 actually still holds even when $\|d\mu/d\nu\|_{\mathbb{L}^\psi(\nu)} = +\infty$. This is clear when $\epsilon < 1$. Whenever $\epsilon = 1$, due to our conventions, we have to show that $(d\mu K)/(d\nu K) = 1$ (in $\mathbb{L}^\psi(\nu K)$), i.e. $\mu K = \nu K$. Since $K_\nu^* = \nu$ for any $\nu \in \mathcal{P}$, by duality as $K = \nu K$. That means in fact that K is a probability measure, say $m \in \mathcal{P}$, and thus for any $\mu,\ \nu \in \mathcal{P}$, $\mu K = m = \nu K$. Finally, the above computations are still valid when $\epsilon = 0$ for which we just have the simple contraction property.

5. Modified logarithmic Sobolev constant

In this section, we investigate the situation corresponding to Φ_1 and Φ_2 of (5) to obtain precise bounds on $\alpha(\Phi_1, K, \nu)$ and $\alpha(\Phi_2, K, \nu)$ in terms of a natural Dirichlet form, which generalize and extend well-known estimates in the invariant situation.

We start with the simpler function Φ_1 and recall a result first due to Fill [10] in the special case where the probability measure ν is assumed to be invariant for the Markov kernel K. Note that for any $\mu, \nu \in \mathcal{P}$, $H_{\Phi_1}(\mu, \nu)$ is finite if and only if $\mu \ll \nu$ and $d\mu/d\nu \in \mathbb{L}^2(\nu)$. We always assume these conditions below. Then denoting by f the function $d\mu/d\nu - \mathbb{1}$,

$$\begin{aligned}\int \left(\frac{d\mu K}{d\nu K} - 1\right)^2 d\nu K &= \int \left(K_\nu^* f\right)^2 d\nu K \\ &= \int f K K_\nu^*(f)\, d\nu \\ &= \nu(f^2) - \mathcal{E}_{\nu, KK_\nu^*}(f, f)\end{aligned}$$

where the Dirichlet form $\mathcal{E}_{\nu, KK_\nu^*}$ was defined in (9). Since $\nu(f) = 0$ and $\nu(f^2) = H_{\Phi_1}(\mu, \nu)$, it easily follows that

$$\alpha(\Phi_1, K, \nu) = \lambda(\nu, KK_\nu^*)$$

where

$$\lambda(\nu, KK_\nu^*) = \inf_{f \in \mathbb{L}^2(\nu) \setminus \mathrm{Vect}(\mathbb{1})} \frac{\mathcal{E}_{\nu, KK_\nu^*}(f, f)}{\nu((f - \nu(f))^2)}$$

is the spectral gap of KK_ν^* in $\mathbb{L}^2(\nu)$. As the constant function $\mathbb{1}$ is an eigenfunction associated to the eigenvalue 0 of the self-adjoint operator $\mathrm{Id} - KK_\nu^*$ in the Hilbert space $\mathbb{L}^2(\nu)$, the quantity $\lambda(\nu, KK_\nu^*)$ can be viewed as the bottom of the spectrum of this operator restricted to the orthogonal complement of $\mathrm{Vect}(\mathbb{1})$ (some authors prefer to speak of spectral gap only if the above quantity is positive, which is not necessarily true under our hypotheses here.)

Next we turn to the function Φ_2. In quite a similar way, we define respectively the logarithmic Sobolev constant

$$\widetilde{l}(\nu, KK_\nu^*) = \inf_{f \in \mathbb{L}^\Psi(\nu) \setminus \mathrm{Vect}(\mathbb{1})} \frac{\mathcal{E}_{\nu, KK_\nu^*}(f, f)}{\mathrm{Ent}(f^2, \nu)}$$

and modified logarithmic Sobolev constant

$$l(\nu, KK_\nu^*) = \inf_{f \in \mathbb{L}^\Psi(\nu) \setminus \mathrm{Vect}(\mathbb{1})} \frac{\mathcal{E}_{\nu, KK_\nu^*}(f^2, \ln(f^2))}{\mathrm{Ent}(f^2, \nu)}\,. \tag{12}$$

Here and in what follows Ψ will always denote a fixed Young function such that $\Psi(x) \sim x^2 \ln(x^2)$ for large $x \in \mathbb{R}_+$ and the entropy of a function $f \in \mathbb{L}^\Psi(\nu)$ is given by

$$\mathrm{Ent}(f^2, \nu) = \int f^2 \ln(f^2/\nu(f^2))\, d\nu.$$

The modified logarithmic Sobolev constant $l(\nu, KK_\nu^*)$ has not been much studied explicitly for itself in the literature, with the notable exception of a paper by Wu [16]. Nevertheless it has often be used implicitly (see for instance [8] or [15]) through the bound

$$l(\nu, KK_\nu^*) \geq \widetilde{l}(\nu, KK_\nu^*). \tag{13}$$

This bound basically follows from the convexity inequality

$$\forall\, y \geq 0,\ \forall\, z > 0, \qquad \ln(z) - \ln(y) \geq \frac{2}{\sqrt{z}}(\sqrt{z} - \sqrt{y})$$

which implies that for any $f \in \mathbb{L}^{\Psi}(\nu)$, $\nu(dx)$-a.s.,

$$\begin{aligned}
|f|(x)(\mathrm{Id} - KK_\nu^*)[\ln(f^2)](x) &= |f|(x)KK_\nu^*[\ln(f^2)(x) - \ln(f^2)](x)\\
&\geq |f|(x)\frac{2}{|f|(x)}KK_\nu^*[|f|(x) - |f|](x)\\
&= 2KK_\nu^*[|f|(x) - |f|](x)\\
&= 2(\mathrm{Id} - KK_\nu^*)[|f|](x)
\end{aligned}$$

(at least if $|f|(x) > 0$). Integrating with respect to $|f|(x)\nu(dx)$ shows that

$$\mathcal{E}_{\nu,KK_\nu^*}(f^2, \ln(f^2)) \geq 2\,\mathcal{E}_{\nu,KK_\nu^*}(|f|\,, |f|) \geq 2\,\mathcal{E}_{\nu,KK_\nu^*}(f, f)$$

which in turn leads to (13).

Another general fact is that

$$2\lambda(\nu, KK_\nu^*) \geq l(\nu, KK_\nu^*). \tag{14}$$

This is proved by applying the definition of the modified logarithmic Sobolev constant to functions of the form $f = \mathbb{1} + \epsilon g$ with $\epsilon \to 0$. The same argument shows that if the infimum in (12) can be approximated by a sequence of functions converging to $\mathbb{1}$, then $l(\nu, KK_\nu^*) = 2\lambda(\nu, KK_\nu^*)$.

The logarithmic Sobolev constant is classically used to bound the decay of entropy (see e.g. [8]). It might well occur however that the classical logarithmic Sobolev constant is of no use (equal to 0) while the modified logarithmic Sobolev constant is strictly positive. The modified logarithmic Sobolev constant may actually be better suited for this purpose (as will be demonstrated by examples below). The following main result namely characterizes the ergodic coefficient $\alpha(\Phi_2, K, \nu)$ in terms of the modified logarithmic Sobolev constant $l(\nu, KK_\nu^*)$. (By (13), the classical logarithmic Sobolev constant $\widetilde{l}(\nu, KK_\nu^*)$ only appears as a lower bound, possibly equal to zero in examples of interest.)

Proposition 5.1. *There exists an universal constant* $0 < \rho < 1$ *such that for any Markov kernel* K *and any probability measure* ν,

$$\rho\, l(\nu, KK_\nu^*) \leq \alpha(\Phi_2, K, \nu) \leq l(\nu, KK_\nu^*)$$

Proof. In order to simplify the exposition, we will work as if the operator K_ν^* were given by a Markov kernel $K_\nu^*(x, dy)$, $x, y \in S$. This can be insured by taking appropriate topological assumptions on $(S, \mathcal{S})$. Nevertheless the above proposition is true without such assumptions, and everything may actually be justified from the operator point of view. However, to adopt the latter would hide the probabilistic intuition. This is why we restrict ourselves to this case and leave it to the reader to rewrite the argument in the operator theory language.

As a consequence, $\nu(dx)K(x,dy) = (\nu K)(dy)K_\nu^*(y,dx)$. This fact is convenient to evaluate $\triangle = \mathrm{Ent}(\mu|\nu) - \mathrm{Ent}(\mu K|\nu K)$. We already know that $\triangle \geq 0$ when $\mu \in \mathcal{P}$ is such that $\mu \ll \nu$ with $\sqrt{f} = \sqrt{d\mu/d\nu} \in \mathbb{L}^\Psi(\nu)$. We can write

$$\begin{aligned}
\triangle &= \int f(x)\ln(f(x))\,\nu(dx) - \int K_\nu^*[f](y)\ln(K_\nu^*[f](y))\,(\nu K)(dy) \\
&= \int f(x)\ln(f(x))\,\nu(dx) - \int f(x)\ln(K_\nu^*[f](y))\,\nu(dx)K(x,dy) \\
&= \int f(x)(\ln(f(x)) - \ln(K_\nu^*[f](y)))\,\nu(dx)K(x,dy) \\
&= \int (\nu K)(dy)\int f(x)(\ln(f(x)) - \ln(K_\nu^*[f](y)))\,K_\nu^*(y,dx).
\end{aligned}$$

For $y \in S$ fixed, the last integral is the entropy of the nonnegative measure $f(x)K_\nu^*(y,dx)$ with respect to the probability measure $K_\nu^*(y,dx)$. To control this entropy, we apply the general Lemma 5.2 presented at the end of this proof, which leads to the equality

$$\begin{aligned}
&\int f(x)\ln\left(\frac{f(x)}{K_\nu^*[f](y)}\right)\,K_\nu^*(y,dx) \\
&= \frac{1}{2}\int_0^\infty dt\int (f_t(x) - f_t(z))(\ln(f_t(x)) - \ln(f_t(z)))\,K_\nu^*(y,dx)K_\nu^*(y,dz)
\end{aligned}$$

where for any $t \geq 0$ and $x \in S$, we have $f_t(x) = e^{-t}f(x) + (1-e^{-t})$. Thus we end up with

$$\begin{aligned}
\triangle &= \frac{1}{2}\int_0^\infty dt\int (f_t(x) - f_t(z))(\ln(f_t(x)) - \ln(f_t(z)))\,(\nu K)(dy) \\
&\quad\times K_\nu^*(y,dx)K_\nu^*(y,dz) \\
&= \frac{1}{2}\int_0^\infty dt\int (f_t(x) - f_t(z))(\ln(f_t(x)) - \ln(f_t(z)))\,\nu(dx)K(x,dy)K_\nu^*(y,dz) \\
&= \int_0^\infty \mathcal{E}_{\nu,KK_\nu^*}(f_t,\ln(f_t))\,dt.
\end{aligned}$$

We are therefore in a position to use the definition of $l(\nu, KK_\nu^*)$ to get that

$$\begin{aligned}
\triangle &\geq l(\nu,KK_\nu^*)\int_0^\infty dt\int f_t\ln(f_t)\,d\nu = l(\nu,KK_\nu^*)\int_0^\infty dt \\
&\quad\times\int f_t\ln(f_t) - f_t + 1\,d\nu
\end{aligned}$$

since $\int f_t\,d\nu = 1$. We are thus led to introduce the functions

$$\begin{aligned}
\forall\, x,\ x \geq 0, \qquad & \varphi_+(x) = (1+x)\ln(1+x) - x \\
\forall\, x,\ -1 \leq x \leq 0, \qquad & \varphi_-(x) = (1+x)\ln(1+x) - x.
\end{aligned}$$

These functions satisfy "doubling property": there exist a constant $0 < \rho < 1$ such that

$$\begin{aligned}
\forall\, x,\ x \geq 0, \qquad & 0 \leq \varphi_+(x) \leq \rho^{-1}\varphi_+(x/2) \\
\forall\, x,\ -1 \leq x \leq 0, \qquad & 0 \leq \varphi_-(x) \leq \rho^{-1}\varphi_-(x/2).
\end{aligned}$$

Since φ_+ is nondecreasing and φ_- nonincreasing, the preceding "doubling property" is still satisfied if in the right-hand side $x/2$ is replaced by any gx with $1/2 \leq g \leq 1$. Next we observe that

$$\begin{aligned}
&\int_0^\infty dt \int f_t \ln(f_t) - f_t + 1\, d\nu \\
&\quad = \int_0^\infty dt \int \varphi_-((f_t - 1)_-) + \varphi_+((f_t - 1)_+)\, d\nu \\
&\quad = \int_0^\infty dt \int \varphi_-(\exp(-t)(f - 1)_-) + \varphi_+(\exp(-t)(f - 1)_+)\, d\nu \\
&\quad \geq \int_0^{\ln(2)} dt \int \varphi_-(\exp(-t)(f - 1)_-) + \varphi_+(\exp(-t)(f - 1)_+)\, d\nu \\
&\quad \geq \rho \int_0^{\ln(2)} dt \int \varphi_-((f - 1)_-) + \varphi_+((f - 1)_+)\, d\nu \\
&\quad = \ln(2)\rho \int \varphi_-((f - 1)_-) + \varphi_+((f - 1)_+)\, d\nu \\
&\quad = \ln(2)\rho \int f \ln(f)\, d\nu.
\end{aligned}$$

In summary, we have shown that

$$\mathrm{Ent}(\mu|\nu) - \mathrm{Ent}(\mu K|\nu K) \geq \ln(2)\rho\, l(\nu, KK_\nu^*)\mathrm{Ent}(\mu|\nu)$$

from which it easily follows that

$$\alpha(\Phi_2, K, \nu) \geq \ln(2)\rho\, l(\nu, KK_\nu^*).$$

The converse inequality is a simple consequence of Jensen's inequality. Let $f \in \mathbb{L}^\Psi(\nu) \setminus \mathrm{Vect}(\mathbb{1})$ be given. We would like want minimize the quotient appearing in the definition (12). To this task we can assume that $\int f^2\, d\nu = 1$. Consider then the probability measure $\mu(x) = f^2(x)\nu(dx)$, $x \in S$. We have

$$
\begin{aligned}
&\mathcal{E}_{\nu,KK_\nu^*}(f^2,\ln(f^2))\\
&\quad=\int f^2(x)KK_\nu^*[\ln(f^2(x))-\ln(f^2)](x)\,\nu(dx)\\
&\quad=\int f^2(x)\ln(f^2(x))\,\nu(dx)-\int f^2(x)KK_\nu^*[\ln(f^2)](x)\,\nu(dx)\\
&\quad\geq\int f^2(x)\ln(f^2(x))\,\nu(dx)-\int f^2(x)K[\ln(K_\nu^*[f^2])](x)\,\nu(dx)\\
&\quad=\int f^2(x)\ln(f^2(x))\,\nu(dx)-\int K_\nu^*[f^2](y)\ln(K_\nu^*[f^2](y))\,(\nu K)(dy)\\
&\quad=\mathrm{Ent}(\mu|\nu)-\mathrm{Ent}(\mu K|\nu K)\\
&\quad\geq\alpha(\Phi_2,K,\nu)\mathrm{Ent}(\mu|\nu)\\
&\quad=\alpha(\Phi_2,K,\nu)\mathrm{Ent}(f^2,\nu)
\end{aligned}
$$

from which the expected lower bound on $l(\nu, KK_\nu^*)$ follows. The proof of Proposition 5.1 is complete. □

To complete the above proof, we still have to establish the following auxiliary result that will be obtained via continuous time arguments. This is rather usual in this context (cf [8]).

Lemma 5.2. *Let $m, n \in \mathcal{P}$ be probability measures such that $m \ll n$ with $\sqrt{f} = \sqrt{dm/dn} \in \mathbb{L}^\Psi(n)$. Then*

$$\mathrm{Ent}(m|n) = \frac{1}{2}\int_0^\infty dt \int (f_t(x) - f_t(y))(\ln(f_t(x)) - \ln(f_t(y)))\, n(dx)n(dy)$$

where for all $t \geq 0$ and for n-a.s. all $x \in S$,

$$f_t(x) = e^{-t}\,\frac{dm}{dn}(x) + 1 - e^{-t}.$$

By homogeneity, this equality can also be extended to the case where m is only assumed to be a non-negative measure, at least if we still have $\sqrt{f} = \sqrt{dm/dn} \in \mathbb{L}^\Psi(n)$ and if we then adopt the convention $\mathrm{Ent}(m|n) = \mathrm{Ent}(f, n)$.

Proof. As announced, consider the Markov semigroup $(P_t)_{t\geq 0}$ with generator $n - \mathrm{Id}$. (From an operator point of view, $P_t = \exp(t(n - \mathrm{Id}))$ and the corresponding stochastic process just waits exponential times of parameter 1 before choosing new positions according to n.) Denote by $m_t = mP_t = e^{-t}m + (1 - e^{-t})n$ the image measure at time $t \geq 0$ of the initial measure m. Next we compute the derivative with respect to time of $\mathrm{Ent}(m_t|n)$ to obtain

$$
\begin{aligned}
-\int f_t(\mathrm{Id}-n)[\ln(f_t)]\,dn &= -\int f_t(x)(\ln(f_t(x))-\ln(f_t(y)))\,n(dx)n(dy)\\
&= -\frac{1}{2}\int (f_t(x)-f_t(y))(\ln(f_t(x))\\
&\qquad -\ln(f_t(y)))\,n(dx)n(dy)
\end{aligned}
$$

where $f_t = dm_t/dn = e^{-t}f + (1 - e^{-t})$. The lemma then follows by integration since $\lim_{t\to+\infty} \mathrm{Ent}(m_t|n) = 0$. □

Remarks 5.3. **a)** Despite the constant ρ, Proposition 5.1 really improves upon the previous result [12] in that direction expressing that $\widetilde{l}(\nu, KK^*_\nu) \leq \alpha(\Phi_2, K, \nu)$ in the case of an invariant ν. Note also that this relation can be extended to the general case, following the same type of considerations we used here.

b) For $\mu, \nu \in \mathcal{P}$ such that $\mu \ll \nu$ with $\sqrt{f} = \sqrt{d\mu/d\nu} \in \mathbb{L}^\Psi(\nu)$, the preceding considerations involve the quantity

$$\mathcal{E}_{\nu,\nu}(f^2, \ln(f^2)) = \frac{1}{2}\int (f^2(x) - f^2(y))(\ln(f^2(x)) - \ln(f^2(y)))\,\nu(dx)\nu(dy).$$

Along with this quantity, we could have introduced a new coefficient associated to K, namely

$$\bar{l}(\nu, KK^*_\nu) = \inf_{f\in\mathbb{L}^\Psi(\nu)\setminus\mathrm{Vect}(\mathbb{1})\,:\,\mathcal{E}_{\nu,\nu}(f^2,\ln(f^2))<+\infty} \frac{\mathcal{E}_{\nu,KK^*_\nu}(f^2, \ln(f^2))}{\mathcal{E}_{\nu,\nu}(f^2, \ln(f^2))}.$$

Following the same approach as above (removing actually the doubling property considerations), we would get $\bar{l}(\nu, KK^*_\nu) \geq \alpha(\Phi_2, K, \nu)$.. In particular,

$$\bar{l}(\nu, KK^*_\nu) \leq l(\nu, KK^*_\nu). \tag{15}$$

This may actually be obtained directly from the general inequality $\mathcal{E}_{\nu,\nu}(f^2, \ln(f^2)) \geq \mathrm{Ent}(f^2, \nu)$ for all $f \in \mathbb{L}^\Psi(\nu)$. However, there is no reverse inequality to the latter at least as soon as ν is not a Dirac mass (consider a function f such that $\nu(\{f = 0\} \sqcup \{f = 1\}) = 1$ and $0 < \nu(\{f = 0\}) < 1$. One may nevertheless wonder for a possible converse inequality to (15).

c) We also observe that one may push the argument of Lemma 5.2 a bit further. Keeping the same notation, we have seen that

$$\begin{aligned}&\mathrm{Ent}(\mu|\nu) - \mathrm{Ent}(\mu K|\nu K)\\ &= \frac{1}{2}\int_0^\infty dt\,\exp(-t)\int (f(x) - f(y))(\ln(f_t(x)) - \ln(f_t(y)))\,\nu(dx)KK^*_\nu(x, dy).\end{aligned}$$

Introduce then the function

$$\forall\, x > 0, \quad G(x) = \int_0^\infty \exp(-t)\ln[1 + \exp(-t)(x - 1)]\,dt = \frac{x\ln(x) - x + 1}{x - 1}.$$

The former right-hand side is equal to

$$\frac{1}{2}\int (f(x) - f(y))(G(f(x)) - G(f(y)))\,\nu(dx)KK^*_\nu(x, dy) = \mathcal{E}_{\nu,KK^*_\nu}(f, G(f)).$$

Therefore, we obtain a "variational characterization" of $\alpha(\Phi_2, K, \nu)$ as

$$\alpha(\Phi_2, K, \nu) = \inf_{f\in\mathbb{L}^\Psi(\nu)\setminus\mathrm{Vect}(\mathbb{1})} \frac{\mathcal{E}_{\nu,KK^*_\nu}(f^2, G(f^2/\nu(f^2)))}{\mathrm{Ent}(f^2, \nu)}.$$

d) Finally, let us return back to hypothesis (H). Using Lemma 4.1, that also implies that $KK^*_\nu \geq \epsilon\nu$, we get directly that under (H), $\lambda(\nu, KK^*_\nu) \geq \epsilon$ and $\bar{l}(\nu, KK^*_\nu) \geq \epsilon$. A fortiori $l(\nu, KK^*_\nu) \geq \epsilon$. However, it was shown in [12] that

$\widetilde{l}(\nu, KK_\nu^*) = 0$ as soon as ν is not a finite combination of Dirac masses. This simple observation already shows the advantage one has to consider $l(\nu, KK_\nu^*)$ instead of $\widetilde{l}(\nu, KK_\nu^*)$ for infinite state spaces.

Nevertheless, the interest of the modified logarithmic Sobolev constant can already be seen in the simplest situation one can imagine. Let S be the two point set $\{0, 1\}$ endowed with its total σ-field $\mathcal{S}$ and assume that $K_\nu^* = \nu$ (i.e. K is a probability measure) with ν charging both points. Then $1 \leq l(\nu, \nu) \leq 2$. In fact, these bounds even hold without any restriction on $(S, \mathcal{S})$, because they can be deduced from the trivial observation that $\lambda(\nu, \nu) = a(K) = 1$, in the cases where $K_\nu^* = \nu$, and from the general inequality $a(K) \leq l(\nu, KK_\nu^*) \leq 2\lambda(\nu, KK_\nu^*)$.

This result should be compared with a computation of Diaconis and Saloff-Coste [8] (see also [1]) showing that in this Bernoulli distribution context

$$\widetilde{l}(\nu, \nu) = \frac{1 - 2\nu_*}{\ln(1/\nu_* - 1)}$$

quantity which goes to zero with $\nu_* = \nu(0) \wedge \nu(1)$. The exact value of $l(\nu, \nu)$ does not seem however to be known. The only informations we have been able to compute are that $l(\nu, \nu) = 2$ for the symmetric law $\nu(0) = \nu(1) = 1/2$ and that otherwise $1 < l(\nu, \nu) < 2$ if $0 < \nu_* < 1/2$, with $\lim_{\nu_* \to 0_+} l(\nu, \nu) = 1$.

Using on one hand a tensorization property of the modified logarithmic Sobolev constant and on the other hand approximations of Poisson distributions by Bernoulli variables, it appears that this ergodic coefficient is larger than $a^{-1} > 0$ for the Metropolis birth and death algorithm associated to the Poisson law on $\mathbb{N}$ of parameter a (while its logarithmic Sobolev constant has to be zero). By appropriate exponential test functions, it can be shown that it is exactly a^{-1} (cf the first chapter of the collective book [1]). This fact was already used by Wu [16] to study Poisson point processes.

6. A Gaussian example

In order to illustrate the potential usefulness of the "Lipschitz point of view" developed in the preceding sections, we examine in this final part some simple Gaussian kernels. This short study will also provide us with one more opportunity to be convinced of the relevance of the modified logarithmic Sobolev constant in this context.

We start with some elementary continuous time computations. Let $\widetilde{a}$, $\widetilde{b}$ and $\widetilde{c}$ be fixed real numbers, and consider the one-dimensional diffusion process $X = (X_t)_{0 \leq t \leq 1}$ strong solution of the s.d.e.

$$dX_t = (\widetilde{a} + \widetilde{b}X_t)\,dt + \widetilde{c}\,dB_t.$$

(As usual, X_0 is assumed to be given independently of the standard Brownian motion $(B_t)_{0 \leq t \leq 1}$). It is a classical exercise to solve this equation to obtain that if $\widetilde{b} \neq 0$,

$$X_1 = \exp(\widetilde{b}) \left[X_0 + \frac{\widetilde{a}}{\widetilde{b}}(1 - \exp(-\widetilde{b})) + \widetilde{c} \int_0^1 \exp(-\widetilde{b}t)\,dB_t \right].$$

We thus introduce the Gaussian kernel K defined by

$$\forall\, x \in \mathbb{R},\ \forall\, A \in \mathcal{R}, \qquad \begin{aligned} K(x,A) &= \mathcal{N}(a+bx,c)[A] \\ &= \frac{1}{\sqrt{2\pi c}} \int_A \exp(-(y-ax-b)^2/(2c))\,dy \end{aligned}$$

where $\mathcal{R}$ is the Borel σ-field of $\mathbb{R}$ and

$$a\frac{\widetilde{a}}{\widetilde{b}}(\exp(\widetilde{b})-1), \quad b=\exp(\widetilde{b}), \quad c=\frac{\widetilde{c}^2}{2\widetilde{b}}(\exp(2\widetilde{b})-1).$$

For any $0 \le t \le 1$, denote by ν_t the law of X_t, if ν was the initial distribution of X_0. In particular, we have that $\nu_1 = \nu K$. If μ is another initial probability measure on the real line, we would like to evaluate, as in the previous section, the evolution of the relative entropy $\mathrm{Ent}(\mu_t|\nu_t)$ for $0 \le t \le 1$. To begin with, let us assume that ν and μ are equivalent to Lebesgue measure λ (we still denote by ν and μ their respective densities) and that the map μ/ν is of $\mathcal{C}^2$ class and bounded above and below by positive constants. Note that the same properties will then also be satisfied by $f_t = \mu_t/\nu_t$ at any time $0 \le t \le 1$. Under these restrictive hypotheses, the following steps are easily justified:

$$\begin{aligned} \partial_t \mathrm{Ent}(\mu_t|\nu_t) &= \int \ln(f_t)\partial_t\mu_t\,d\lambda + \int \partial_t\mu_t\,d\lambda - \int f_t\partial_t\nu_t\,d\lambda \\ &= \int L[\ln(f_t)]\mu_t\,d\lambda - \int L[f_t]\nu_t\,d\lambda \\ &= \int f_t L[\ln(f_t)] - L[f_t]\,d\nu_t \end{aligned}$$

where L is the generator of the diffusion X, which in particular acts on $\mathcal{C}_\mathrm{b}^2$ functions f by

$$\forall\, x \in \mathbb{R}, \qquad L[f](x) = \frac{\widetilde{c}^2}{2}\partial^2 f(x) + (\widetilde{a}+\widetilde{b}x)\partial f(x).$$

It is well-known [3] that by continuity of the paths of X, the operator L satisfies the change of variables formula: if φ and f are $\mathcal{C}_\mathrm{b}^2$ functions,

$$L[\varphi\circ f] = \varphi'(f)L[f] + \frac{\varphi''(f)}{2}\,\Gamma(f,f)$$

where Γ is the "carré du champ" associated to L (defined with $\varphi(x) = x^2$). In our example, it is simply given on appropriate test functions f by $\Gamma(f,f) = \widetilde{c}^2(\partial f)^2$. Since then, for any $0 \le t \le 1$, $f_t L[\ln(f_t)] - L[f_t] = 2\Gamma(\sqrt{f_t}, \sqrt{f_t})$, it is natural to introduce the (diffusive) logarithmic Sobolev constant of ν_t defined by

$$l(\nu_t) = \inf_{f\in\mathcal{C}_\mathrm{c}^2(\mathbb{R})\setminus\{0\}} \frac{\int(\partial f)^2\,d\nu_t}{\int f^2\ln(f^2/\nu_t(f^2))\,d\nu_t}\,.$$

In that way we obtain (noting that in the infimum, $\mathcal{C}^2_c(\mathbb{R})\setminus\{0\}$ can easily be replaced by the set of $\mathcal{C}^1$ functions f for which the denominator of the latter quotient is finite and positive) the simple differential inequality

$$\partial_t \mathrm{Ent}(\mu_t|\nu_t) \leq -2\widetilde{c}^2 l(\nu_t)\mathrm{Ent}(\mu_t|\nu_t).$$

Therefore,

$$\mathrm{Ent}(\mu K|\nu K) \leq \exp\left(-2\widetilde{c}^2\int_0^1 l(\nu_t)\,dt\right)\mathrm{Ent}(\mu|\nu).$$

Without much difficulty, this relation can be extended, first to any probability measure μ by standard approximations and next to all $\nu \in \mathcal{P}(\mathbb{R},\mathcal{R})$, using the fact that for arbitrary small $t > 0$, ν_t meets the above absolute continuity and regularity requirements.

We now evaluate the integral $\int_0^1 l(\nu_t)\,dt$. Let us consider the simple case where ν itself is a Gaussian distribution $\mathcal{N}(\widehat{a},\widehat{c})$. Traditional computations show that in such a nice situation, for any $0\leq t\leq 1$, $\nu_t = \mathcal{N}(\widehat{a}_t,\widehat{c}_t)$ with

$$\begin{aligned}
\widehat{a}_t &= \exp(\widetilde{b}t)\widehat{a} + \frac{\widetilde{a}}{\widetilde{b}}(\exp(\widetilde{b}t)-1),\\
\widehat{c}_t &= \exp(2\widetilde{b}t)\widehat{c} + \frac{\widetilde{c}^2}{2\widetilde{b}}(\exp(2\widetilde{b}t)-1).
\end{aligned}$$

Next we use another well-known result (cf for instance [1] and the references therein) which makes a direct link between the logarithmic Sobolev constant of a Gaussian distribution and its variance, namely

$$l(\nu_t) = \frac{1}{2\widehat{c}_t}\,.$$

Therefore, with the change of variable $s = \exp(2\widetilde{b}t)$, we get that

$$\begin{aligned}
2\widetilde{c}^2\int_0^1 l(\nu_t)\,dt &= \int_1^{\exp(2\widetilde{b})} \frac{1}{\widehat{c}s + (s-1)\widetilde{c}^2/(2\widetilde{b})}\,\frac{\widetilde{c}^2 ds}{2\widetilde{b}s}\\
&= \int_1^{\exp(2\widetilde{b})} \frac{1}{As+s-1}\,\frac{ds}{s}\\
&= \int_1^{\exp(2\widetilde{b})} \frac{A+1}{As+s-1} - \frac{1}{s}\,ds\\
&= [\ln(As+s-1)-\ln(s)]_1^{\exp(2\widetilde{b})}\\
&= \ln\left(\frac{(A+1)\exp(2\widetilde{b})-1}{A\exp(2\widetilde{b})}\right)
\end{aligned}$$

where $A = 2\widetilde{b}\widehat{c}/\widetilde{c}^2$. Rewriting everything in terms of K and ν, with in particular $\chi = c/(b^2\widehat{c}) \geq 0$, we have thus proved that

$$\mathrm{Ent}(\mu K|\nu K) \leq (1+\chi)^{-1}\,\mathrm{Ent}(\mu|\nu).$$

In other words,

$$\alpha(\Phi_2, K, \nu) \geq 1 - (1+\chi)^{-1}\, \frac{\chi}{(1+\chi)}\,.$$

This lower estimate is actually an equality. Taking indeed taking $\mu = \mathcal{N}(\bar{a}, \widehat{c})$, with $\bar{a} \neq \widehat{a}$, we see that

$$\begin{aligned} \mathrm{Ent}(\mu|\nu) &= \frac{(\bar{a}-\widehat{a})^2}{2\widehat{c}} \\ \mathrm{Ent}(\mu|\nu) - \mathrm{Ent}(\mu K|\nu K) &= \frac{\chi}{1+\chi}\mathrm{Ent}(\mu|\nu) \end{aligned}$$

and the claim follows.

Note that this example is another occurrence of $\widetilde{l}(\nu, KK_\nu^*) = 0$. From the results of Section 4, it follows that $\chi/(1+\chi)$ is of the order of $l(\nu, KK_\nu^*)$ up to an universal factor. This result does not seem to follow directly from the expression of the Gaussian kernel

$$\forall\, x \in \mathbb{R}, \qquad KK_\nu^*(x,\cdot) = \mathcal{N}\left(\frac{c\widehat{a} + b^2\widehat{c}x}{b^2\widehat{c}+c}, \frac{2b^2c\widehat{c}^2 + c^2\widehat{c}}{b^2\widehat{c}+c)^2}\right).$$

We are now in position to present an example of application of the above considerations. Let $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$ and $(c_n)_{n\geq 0}$ be three sequences of real numbers with $c_n \geq 0$ for all $n \geq 0$. Define the Gaussian kernels

$$\forall\, n \geq 0,\ \forall\, x \in \mathbb{R}, \qquad K_n(x,\cdot) = \mathcal{N}(a_n + b_n x, c_n).$$

Define furthermore a new sequence of positive numbers by the induction relation $\widehat{c}_{n+1} = b_n^2\widehat{c}_n + c_n$, $n \geq 0$, starting with $\widehat{c}_0 = 1$.

Set $\chi_n = \frac{c_n}{b_n^2\widehat{c}_n}$ for $n \geq 1$ (alternatively, these numbers can be given directly by the iteration $\chi_{n+1}^{-1} = (\chi_n^{-1}+1)b_{n+1}^2 c_n/c_{n+1})$. Assume that

$$\sum_{n\geq 0} \chi_n = +\infty. \tag{16}$$

This hypothesis insures that for any probability measures $\mu, \widetilde{\mu} \in \mathcal{P}(\mathbb{R}, \mathcal{R})$,

$$\lim_{n\to\infty} \|(\mu - \widetilde{\mu})K_0K_1\cdots K_n\| = 0. \tag{17}$$

Indeed, let $\nu = \mathcal{N}(0,1)$ and observe that $\widehat{c}_n$ is the variance of $\nu K_0K_1\cdots K_n$. Therefore, for every $n \geq 0$,

$$\begin{aligned} &\mathrm{Ent}(\mu K_0K_1\cdots K_n|\nu K_0K_1\cdots K_n) \\ &\quad \leq \frac{1}{1+\chi_n}\mathrm{Ent}(\mu K_0K_1\cdots K_{n-1}|\nu K_0K_1\cdots K_{n-1}). \end{aligned}$$

Thus under (16), we get that

$$\lim_{n\to\infty} \mathrm{Ent}(\mu K_0K_1\cdots K_n|\nu K_0K_1\cdots K_n) = 0.$$

This implies

$$\lim_{n\to\infty} \|\mu K_0 K_1 \cdots K_n - \nu K_0 K_1 \cdots K_n\| = 0$$

due to the general Pinsker-Csiszár-Kullback inequality $\|\mu - \nu\| \leq \sqrt{2\mathrm{Ent}(\mu|\nu)}$. The claim (17) follows. Note finally that (16) is in particular implied by $\sup_{n\geq 0} c_n < +\infty$ and $\limsup_{n\to\infty} b_n < 1$, under which $\sup_{n\geq 0} \widehat{c_n} < +\infty$.

Remarks 6.1. **a)** For $n \geq 0$, let us assume that $|b_n| < 1$ and denote by $\nu_n = \mathcal{N}(a_n/(1-b_n), c_n/(1-b_n^2))$ the reversible probability measure relative to K_n. A classical way to prove a result such as (17) is to show that for any probability measure $\mu \in \mathcal{P}(\mathbb{R}, \mathcal{R})$,

$$\lim_{n\to\infty} \|\mu K_0 K_1 \cdots K_n - \nu_n\| = 0 \tag{18}$$

(or sometimes more conveniently $\lim_{n\to\infty} \mathrm{Ent}(\mu K_0 K_1 \cdots K_n|\nu_n) = 0$). However, it should be noticed that (18) is not always satisfied under (16). Indeed, it is not difficult to choose a "wild" sequence $(a_n)_{n\geq 0}$ which will not allow for this convergence (and such an occurrence is not only theoretical, as it often appears in the context of linear and nonlinear filtering [6]).

Thus even if there exists an invariant probability, it may not always be interesting to consider it. It was this simple observation which convinced us that the Lipschitz point of view adopted in this work is a pertinent tool in the study of loose of memory properties for Markov chains. We expect this approach to be fruitful in situations where the system is far away from being at equilibrium (and thus the latter is not so meaningful to compute ergodic constants), for instance to study metastability properties in statistical mechanics.

b) Note that the above inhomogeneous example and the conclusions drawn from it can be easily rewritten in a continuous time context.

c) One can believe that the above computation of the diffusive logarithmic Sobolev constant along the trajectory of the time marginals of the process for a nice initial distribution is quite particular and restricted to the Gaussian case. This is not the case and Bakry [2] has developed a whole technology to derive such estimates in continuous time, which furthermore are easier to obtain in a compact state space context through comparison of Dirichlet forms and measure theoretic techniques.

References

1. Ané, C., Blachère, S., Chafaï, D., Fougères, P., Gentil, I., Malrieu, F., Roberto, C., Scheffer, G.: Sur les inégalités de Sobolev logarithmiques. Société Mathématique de France, Paris, 2000. With a preface by D. Bakry and M. Ledoux
2. Bakry, D.: On Sobolev and logarithmic Sobolev inequalities for Markov semigroups. In: K.D. Elworthy, S. Kusuoka, I. Shigekawa, editors, New Trends in Stochastic Analysis, pp. 43–75. World Scientific, 1997
3. Bakry, D., Émery, M.: Diffusions hypercontractives. In Séminaire de probabilités, XIX, 1983/84, volume 1123 of Lecture Notes in Math., pp. 177–206. Springer, Berlin, 1985

4. Cohen, J.E., Iwasa, Y., Răuţu, G., Ruskai, M.B., Seneta, E., Zbăganu, G.: Relative entropy under mappings by stochastic matrices. Linear Algebra Appl. **179**, 211–235 (1993)
5. Csiszár, I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Magyar Tud. Akad. Mat. Kutató Int. Közl. **8**, 85–108 (1963)
6. Del Moral, P., et Miclo, L.: Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In Séminaire de Probabilités, XXXIV, pp. 1–145. Springer, Berlin, 2000
7. Del Moral, P., Miclo, L.: Feynman-Kac formulae. In preparation, to be published by Springer-Verlag N.Y., 2004
8. Diaconis, P., Saloff-Coste, L.: Logarithmic Sobolev inequalities for finite Markov chains. The Annals of Applied Probability **6**(3), 695–750 (1996)
9. Dobrushin, R.L.: Central limit theorem for nonstationary Markov chains, I, II. Theory of Probability and its Applications. **1**(1 and 4), 66–80 and 330–385 (1956)
10. Fill, J.A.: Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. The Annals of Applied Probability **1**(1), 62–87 (1991)
11. Kranosel'skii, M.A., Rutickii, Y.B.: Convex functions and Orlicz spaces. P. Noordhoff, 1961
12. Miclo, L.: Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In: J. Azéma, M. Emery, and M. Yor, editors, Séminaire de Probabilités XXXI, Lecture Notes in Mathematics 1655, pp. 136–167. Springer-Verlag, Berlin, 1997
13. Miclo, L.: Notes on the speed of entropic convergence in the central limit theorem. Preprint, submitted to the Proceedings of the Conference on Stochastic Inequalities, Barcelona, 2002
14. Revuz, D.: Markov chains. North-Holland, 1975
15. Saloff-Coste, L.: Lectures on finite Markov chains. In P. Bernard, editor, Lectures on Probability Theory and Statistics. Ecole d'Eté de Probabilités de Saint-Flour XXVI-1996, Lecture Notes in Mathematics 1665. Springer-Verlag, Berlin, 1997
16. Wu, L.: A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. Probab. Theory Related Fields **118**(3), 427–438 (2000)
17. Zaharopol, R., Zbaganu, G.: Dobrushin coefficients of ergodicity and asymptotically stable L^1-contractions. J. Theoret. Probab. **12**(4), 885–902 (1999)