About supercontraction properties of Markov kernels

P. Del Moral M. Ledoux L. Miclo*

UMR C5583, CNRS and Université Toulouse III

Abstract

We study Lipschitz's contraction properties of general Markovian kernels seen as operators on spaces of probabilities endowed with entropy-like "distances". Universal quantitative bounds on the associated ergodic constants are deduced from Dobrushin's coefficient and strong contraction properties in Orlicz's spaces for relative densities are proved under restrictive mixing assumptions. Next we obtain contraction estimates in the entropy sense around an arbitrary probability by introducing a particular Dirichlet form and the corresponding modified logarithmic Sobolev inequalities. The interest of these bounds will be illustred by inhomogeneous Gaussian examples, emphasizing the irrelevence of the existence of an invariant measure assumption.

Keywords: Lipschitz's contraction, generalized relative entropy, Markov kernel, Dobrushin's coefficient, Orlicz's norm for densities, appropriate Dirichlet form, spectral gap, modified logarithmic Sobolev inequality, loose of memory property for inhomogeneous Gaussian chains.

AMS 2000 Mathematics Subject Classifications: 60J05, 60J22, 37A30, 37A25, 39A11, 39A12, 46E39, 28A33, 47D07.

1 Introduction and results

The purpose of this paper is to present general properties of supercontraction for Markov kernels without assumption on the existence of an invariant probability. Our principal motivation comes from nonlinear filtering, where one would like to show that the system is forgetting its initialisation (especially when it is wrong!) without nevertheless converging in large times to an invariant distribution, which may not exist in noncompact contexts (for annealed results) or due the randomness of the underlying signal (from the quentched point of view). But this application will not be developed here, as we prefer to work on a "abstract" setting.

So we start from a measurable space $(S, \mathcal{S})$ and we denote by $\mathcal{M}$ (respectively $\mathcal{P}$, $\mathcal{M}_+$ and $\mathcal{M}_0$) the set of associated signed bounded measures (resp. probabilities, nonnegative bounded measures and elements of $\mathcal{M}$ whose total mass is equal to 0). These sets will be naturally endowed with the total variation metric $\|\cdot\|$. But one can consider other ways to measure the "distance" between for instance elements of $\mathcal{M}_+$: let $\Phi : \mathbb{R}_+^2 \to \mathbb{R} \sqcup \{+\infty\}$ be a convex function verifying $\Phi(1,1) = 0$ and which is homogeneous:

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

Then for any $\mu, \nu \in \mathcal{M}_+$, we can consider their so-called Φ-relative entropy

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

where λ stands for any measure of $\mathcal{M}_+$ satisfying $\mu \ll \lambda$ and $\nu \ll \lambda$. Such an element always exists, it is sufficient to consider $\lambda = \mu + \nu$, and the above definition does not depend on its choice, mostly due to the homogeneity property of Φ. To see there is no real problem of integrability, let $(b_1, b_2) \in \mathbb{R}^2$ be a subderivative of the convex mapping Φ in $(1,1)$, ie a couple of reals satisfying

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

then it appears that the above integral is always well-defined with value in $\mathbb{R} \sqcup \{+\infty\}$, since it can be written

$$\forall\, \mu,\, \nu \in \mathcal{M}_+, \qquad H_\Phi(\mu,\nu) \;=\; \int \Psi\left(\frac{d\mu}{d\lambda}, \frac{d\nu}{d\lambda}\right) d\lambda + b_1(\mu(S)-1) + b_2(\nu(S)-1)$$

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

Now let be given a Markov kernel K on $(S, \mathcal{S})$, ie a mapping $S \times \mathcal{S} \to [0,1]$ 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

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

Then of primary interest for us is to study supercontraction properties of $\mathcal{P} \ni \mu \mapsto \mu K \in \mathcal{P}$ (nevertheless, as it is classical for this kind of questions, instead of considering the one-step kernel K, in practice it is often more rewarding to look at an iterated kernel K^n, for some $n \geq 2$), with respect of H_Φ for example.

Thus for furthermore fixed $\nu \in \mathcal{P}$, let us define $\alpha(\Phi, K, \nu)$ as the best constant such that for all $\mu \in \mathcal{P}$,

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

ie $1 - \alpha(\Phi, K, \nu)$ can be seen as a "moving" coefficient of contraction around ν for K relatively to H_Φ.

The first result says that the Dobrushin's constant given by

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

is an universal lower bound for them:

Proposition 1.1 *For Φ, K and ν as before, we are always assured of*

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

In particular we have $\alpha(\Phi, K, \nu) \geq 0$, 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 [3] in the case of a finite space (see also [13], 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, but they do not discuss Lipschitz-type contraction inequalities). Indeed our proof will mainly be a simplification and extension of their approach (as customary, the general situation could be deduced from the finite case, nevertheless we prefer to work directly, by modifying their arguments).

Thus the Dobrushin's coefficient appears 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 utra-mixing condition

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

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

because it is well-known to implies that $a(K) \geq \epsilon$.

Our second objective is to see this inequality as a special occurence of a kind of supercontraction property for relative densities in Orlicz spaces valid under (H). More precisely, let $\psi : \mathbb{R}_+ \to \mathbb{R}$ be a Young function (null in 0, increasing and convex) and for $\nu \in \mathcal{P}$, denote by $\mathbb{L}^\psi(\nu)$ the associated Orlicz space (cf [8] as a general reference): it is the set of measurable mappings $f : S \to \mathbb{R}$ for which there exists $a > 0$ such that

$$\int \psi(f/a)\, d\nu \;<\; +\infty$$

The smallest such constant a is then denoted $\|f\|_{\mathbb{L}^\psi(\nu)}$ and $\|\cdot\|_{\mathbb{L}^\psi(\nu)}$ is a norm on $\mathbb{L}^\psi(\nu)$ transforming it into a Banach space.

Proposition 1.2 *Let $\mu, \nu \in \mathcal{P}$ verifying $\mu \ll \nu$ and $d\mu/d\nu \in \mathbb{L}^{\psi}(\nu)$. Then we have $\mu K \ll \nu K$, $d\mu K/d\nu K \in \mathbb{L}^{\psi}(\nu K)$ and furthermore under hypothesis (H) we get*

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

(more generally this inequality with $\epsilon = 0$ is always satisfied).

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 the proposition 1.1, where

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

then we get

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

Our last aim is to look more closely at the two very interesting cases associated to the mappings

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

(the usual convention $0 \cdot \infty = 0$ is assumed enforced in the whole paper), because they are relative to the important notions of relative $\mathbb{L}^2$-norm and relative entropy.

So we will give lower bounds of $\alpha(\Phi_1, K, \nu)$ and $\alpha(\Phi_2, K, \nu)$ respectively in terms of certain spectral gaps (thus extending well-known results in the situation where ν is assumed to be invariant, cf [7]) and modified Sobolev logarithmic constants, but we postpone until section 5 the details about the relevent underlying Dirichlet forms, which envolve the operators described in next section.

Finally, a Gaussian example will show the advantage of the modified constant with respect to the traditional Sobolev logarithmic one (see for instance [5] as a general reference or [9] for corresponding considerations on the decreasing evolution of entropy in the case of an invariant probability ν) and will underline the gain one can obtain by not working with an invariant measure, even if one does exist, thus stressing the importance of the Lipschitz point of view.

2 A few reminders

We will recall here some useful facts about total variation and Markov operators theory.

First we come back to the definition of Φ-entropy. Let us denote by $\phi : \mathbb{R}_+ \to \mathbb{R} \sqcup \{+\infty\}$ the convex function given by

$$\forall\, x \in \mathbb{R}_+, \qquad \phi(x) = \Phi(x,1)$$

Due to its homogeneity property, 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 $\Phi(1,0) = +\infty$, since that innocent looking assumption admits interesting consequences:

Lemma 2.1 *Assume that $\Phi(1,0) = +\infty$, then we have*

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

The above rhs is sometimes called ϕ-divergence in the sense of Csiszar [4].

Proof:

So let μ, $\nu \in \mathcal{M}_+$ be given, denote $\mu = \mu_1 + \mu_2$ the Lebesgue decomposition of μ with respect to ν, ie $\mu_1 \ll \nu$ and $\mu_2 \perp \nu$, and consider $A \in \mathcal{S}$ be such that $\nu(A^c) = 0 = \mu_2(A)$. To compute $H_\Phi(\mu,\nu)$ we can take $\lambda \stackrel{\text{def.}}{=} \nu + \mu_2$ and we get

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

thus if $\mu_2(S) > 0$ we deduce that $H_\Phi(\mu,\nu) = +\infty$ and otherwise we obtain

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

because we can choose $\lambda = \nu$.

■

Note that reciproquely, 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 (use for instance the increasing caracter of $(1,+\infty) \ni t \mapsto (\phi(t)-\phi(1))/(t-1)$) and it can be shown that any function defined by

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

is convex, where we have chosen an element $l \in [l_0, +\infty]$ (implicitly in the sequel, we will only consider the case $l = l_0$).

Here are a few examples of classical Φ-entropies arising in the literature and which will 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

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

The former quantity can be generalized in two directions: the Havrda-Charvat entropy of order $p > 1$ corresponding to the function $\phi : \mathbb{R}_+ \ni t \mapsto \frac{1}{p-1}(t^p - 1)$, ie

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

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

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

($p = 1$ corresponds to Φ_0).

Observe that for $\phi : \mathbb{R}_+ \ni t \mapsto \frac{1}{p-1}(t^{1-p} - 1)$, with $p > 1$ (resp. $\phi : \mathbb{R}_+ \ni t \mapsto \ln(1/t)$), we get the reciprocal

$$\forall\, \mu \ll \nu, \qquad H_\Phi(\mu,\nu) \;=\; C_p(\nu|\mu)$$

(resp. $H_\Phi(\mu,\nu) = \mathrm{Ent}(\nu|\mu)$, also for $\mu \ll \nu$).

The latter examples are occurences where $\Phi(1,0) = +\infty$ is not naturally verified. The Hellinger integrals of order $\alpha \in (0,1)$ are some other instances: they are obtained with $\phi : \mathbb{R}_+ \ni t \mapsto t - t^\alpha$, thus for any μ, $\nu \in \mathcal{P}$ and λ dominating them,

$$H_\Phi(\mu,\nu) \;=\; 1 - \int \left(\frac{d\mu}{d\lambda}\right)^\alpha \left(\frac{d\nu}{d\lambda}\right)^{1-\alpha} d\lambda$$

(it can be written more simply $1 - \int \left(\frac{d\mu}{d\nu}\right)^\alpha d\nu$ if $\mu \ll \nu$), which in the special case $\alpha = 1/2$, coincide with the Kakutani-Hellinger distance defined by

$$H(\mu|\nu) \;\stackrel{\text{def.}}{=}\; \frac{1}{2}\int \left(\sqrt{\frac{d\mu}{d\lambda}} - \sqrt{\frac{d\nu}{d\lambda}}\right)^2 d\lambda$$

But maybe the most important exception of interest where $\Phi(1,0) \neq +\infty$ is $\Phi = \Phi_0$ defined in (3). Then it appears that for any μ, $\nu \in \mathcal{M}_+$,

$$\begin{aligned} H_{\Phi_0}(\mu,\nu) &= \|\mu - \nu\| \\ &\stackrel{\text{def.}}{=} \sup_{f\in\mathcal{B}_\mathrm{b},\,\|f\|_\infty=1} (\mu-\nu)(f) \end{aligned}$$

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.

Then the proposition 1.1 is clear for this particular mapping Φ_0, since in a seminal paper [6] Dobrushin proved that for any Markov kernel K,

$$\begin{aligned} 1 - a(K) &= \sup_{x\neq y\in S} \frac{\|\delta_x K - \delta_y K\|}{\|\delta_x - \delta_y\|} \\ &= \sup_{\mu\neq\nu\in\mathcal{P}} \frac{\|\mu K - \nu K\|}{\|\mu - \nu\|} \end{aligned} \tag{7}$$

and this also justifies (4). The deduction of $a(K) \geq \epsilon$ from the proposition 1.2 for $\varphi = |\cdot|$ is quite similar.

It is time now to give some further precisions about operator interpretations of a Markov kernel K: via the formula

$$\forall\, x \in S, \qquad K(f)(x) \;\stackrel{\text{def.}}{=}\; \int f(y)\, K(x,dy)$$

K can be seen as right acting on $f \in \mathcal{B}_\mathrm{b}$ or $\mathcal{B}_+$ (the set of $\bar{\mathbb{R}}_+$-valued measurable functions defined on $(S,\mathcal{S})$).

Let a probability $\nu \in \mathcal{P}$ be given, for $f \in \mathcal{B}_{\mathrm{b}}$ and $1 \leq p < \infty$, using Hölder inequality, we get that

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

thus K can also be extended as a linear mapping from $\mathbb{L}^p(\nu K)$ to $\mathbb{L}^p(\nu)$.

By duality, we obtain an 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 $1 < p < +\infty$. By definition, it satisfies

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

Unfortunately, due to the fact that the dual of $\mathbb{L}^\infty(\nu)$ cannot entirely be identified with $\mathbb{L}^1(\nu)$, this approach does not allow for the construction of K_ν^* on the latter space, rather we have to work it out directly: we begin by noting that if $\mu \in \mathcal{M}$ verifies $\mu \ll \nu$, 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, so μ-negligible and finally $(\mu K)(A) = 0$. Thus as we can see any $f \in \mathbb{L}^1(\nu)$ as the corresponding measure $\mu = f\nu \ll \nu$, we take by definition

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

and we still have

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

(this manipulation is classical in the field of Markov kernels, see for instance [10] and its bibliography, note also that the dual operators K_ν^* on $\mathbb{L}^p(\nu)$, for $1 < p < +\infty$, can be viewed as restrictions to these sets of the one on $\mathbb{L}^1(\nu)$).

In fact K_ν^* is almost a Markov kernel: recall that for any two fixed probabilities μ, ν, a linear operator R from $\mathbb{L}^1(\nu)$ to $\mathbb{L}^1(\mu)$ is said (generalized) Markovian if

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

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

Proof:

For the preservation of nonnegativity, it is sufficient to see that for any nonnegative $f \in \mathbb{L}^1(\nu)$ and any nonnegative $g \in \mathbb{L}^\infty(\nu K)$, $(\nu K)(gK_\nu^*(f)) \geq 0$, which is obvious in view of (8). Furthermore for any $g \in \mathbb{L}^\infty(\nu K)$,

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

thus implying that $K_\nu^*(\mathbb{1}) = \mathbb{1}$, νK-a.s.

As $K : \mathbb{L}^1(\nu K) \to \mathbb{L}^1(\nu)$ is clearly Markovian, by composition the same remains true for $KK_\nu^* : \mathbb{L}^1(\nu) \to \mathbb{L}^1(\nu)$.

■

In particular we are allowed to use Jensen's inequality, which in this context reads: for any function $f \in \mathcal{B}_+$ and any convex mapping $\varphi : \mathbb{R}_+ \to \mathbb{R}$, we have

$$K_\nu^*(\varphi(f)) \;\geq\; \varphi(K_\nu^*(f)) \qquad , \nu K\text{-a.s.}$$

Another interesting feature is that ν is invariant with respect to KK_ν^*, meaning that

$$\forall\, f \in \mathbb{L}^1(\nu), \qquad \nu(KK_\nu^*(f)) \;=\; \nu(f)$$

which is an immediate consequence of

$$\nu(KK_\nu^*(f)) \;=\; (\nu K)(\mathbb{1}K_\nu^*(f)) \;=\; \nu(fK(\mathbb{1})) \;=\; \nu(f)$$

The probability ν is even reversible for this Markovian operator KK_ν^*, in the sense that

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

since we have in the same manner

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

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 and it will be starting from this object that we will define spectral gaps and modified logarithmic Sobolev constants in section 5. Note that if ν was assumed to be invariant for K (ie $\nu K = \nu$), then K_ν^* coincides on $\mathbb{L}^2(\nu)$ with the adjoint operator of K on $\mathbb{L}^2(\nu)$, thus we recover in this situation the Dirichlet form considered in [7] and [9].

To finish this section, let us "compute" K_ν^* in cases where some absolutely continuity is available: more precisely, we make the hypothesis that

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

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

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

Then expanding the rhs of the formula (8), we obtain

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

In a similar , one verifies that

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

with

$$\forall\, (x,y) \in S \times S, \qquad G(x,y) \;=\; \int \frac{dK(x,\cdot)}{d\nu K}(z)\frac{dK(y,\cdot)}{d\nu K}(z)\,\nu K(dz)$$

3 On Dobrushin's coefficient

We will show in the present section that this constant is a lower bound for the other natural ergodic coefficients, as expressed by the 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 is to give an extension of this inequality to the whole space $\mathcal{M}$:

Lemma 3.1 *For any $\eta \in \mathcal{M}$, we are assured of*

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

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 $\|\eta K\| = |\eta K(S)| = \eta(S)$, thus the lemma is also clear in this situation.

But 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

$$\begin{aligned} \eta_1 &= \frac{\eta_+(S) - \eta_-(S)}{\eta_+(S)}\eta_+ \qquad \in \mathcal{M}_+ \\ \eta_2 &= \frac{\eta_-(S)}{\eta_+(S)}\eta_+ - \eta_- \qquad \in \mathcal{M}_0 \end{aligned}$$

Then we get

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

because $|\eta_1(S)| + |\eta_2(S)| = \eta_1(S) = \eta(S)$ and one verifies that

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

■

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 $\eta K(S) = \eta(S)$, the above estimate can also be written

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

thus 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, we are assured that either $\lim_{n\to\infty}(\eta K^n)_+(S) = 0$ or $\lim_{n\to\infty}(\eta K^n)_-(S)=0$, depending on the sign of $\eta(S)$. If it is null, both are converging to zero, $\lim_{n\to+\infty} \|\eta K^n\| = 0$, ie there is a lost of initial condition η_+ or η_-, which is the original result of Dobrushin.

The next result is essentially the lemma 3.3 of J.E. Cohen, Y. Iwasa, G. Răuţu, M.B. Ruskai, E. Seneta, and G. Zbăganu [3], except that there the measures were assumed to have compact support.

Lemma 3.3 *Let m_1, m_2 be two bounded measures on the Borelian real line $(\mathbb{R}, \mathcal{R})$ admitting a first moment and such that*

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

$$m_1(\mathbb{R}) = m_2(\mathbb{R}) \qquad \text{and} \qquad \int t\, m_1(dt) = \int t\, m_2(dt)$$

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

$$\int \ |t-s| \ m_1(dt) \ \ \leq \ \ \int \ |t-s| \ m_2(dt)$$

Then for any convex function φ, we have $m_1(\varphi) \leq m_2(\varphi)$ (the value $+\infty$ is permitted).

Proof:

The main modification with respect to the approach of [3] is replacing its uniform convergence by a monotonous one, better suited for measure theory.

So let φ be a given convex function on $\mathbb{R}$. One can find two two-sided sequences $(x_i)_{i\in\mathbb{Z}^*}$ and $(k_i)_{i\in\mathbb{Z}^*}$ of nonnegative reals such that if we note for $n \geq 1$,

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

where φ'_{r} is the right derivative of φ and $(\cdot)_+$ and $(\cdot)_-$ designate respectively the nonnegative and nonpositive parts, then $(\varphi_n)_{n\geq 1}$ is an increasing sequence converging to φ.

Indeed, writing

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

and remarking that the mapping

$$\mathbb{R} \ni t \ \mapsto \ \varphi'_{\mathrm{r}}(t) - \varphi'_{\mathrm{r}}(0)$$

is nondecreasing and null in zero, it is sufficient to approximate from below the latter function by nondecreasing step functions (for instance constant on appropriate dyadic intervals) to conclude to the expected convergence.

Next coming back to m_1 and m_2, we note that for any $i \geq 1$,

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

and idem for the nonpositive parts

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

In consequence, for any $n \geq 1$,

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

and letting n grow to infinity, we obtain via the monotonous convergence theorem the announced result.

■

In the present form, the previous lemma would only imply the proposition 1.1 for probabilities satisfying $\mu \ll \nu$, so let us modify it a little:

Lemma 3.4 *Let m_1, m_2 be two bounded measures on the Borelian 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:*

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

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

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

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

Proof:

We begin by coming back to the formula (6) and by reconsidering its notations. Using for the convex function ϕ appearing there the approximation presented in the proof of the lemma 3.3, we realize that our above affirmation will be shown if for all $a, b \in \mathbb{R}$,

$$\int (as - bt)_+\; m_1(ds, dt) \quad \leq \quad \int (as - bt)_+\; m_2(ds, dt) \tag{10}$$

$$\int \mathbb{1}_{\{t=0\}} s\, m_1(ds, dt) \quad \leq \quad \int \mathbb{1}_{\{t=0\}} s\; m_2(ds, dt)$$

The last inequality is needed for the cases where $l > l_0$, but note that it can be deduced from the first condition by letting there b going to infinity, with $a = 1$.

Finally, we also remark that (10) is insured by the hypotheses of the lemma, as before via a mere substraction.

■

Now the proof of proposition 1.1 follows quite straightfowardly:

Indeed, let μ, $\nu \in \mathcal{P}$ be given and $\lambda \in \mathcal{P}$ dominating them, it is sufficient to consider the measures m_1 and m_2 on $(\mathbb{R}_+^2, \mathcal{R}_+^{\otimes 2})$ acting on any bounded measurable function $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ by

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

Then the first condition of lemma 3.4 is immediate and the second hypothesis amounts to verify that for all $a, b \in \mathbb{R}$,

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

or in other words, in terms of the total variation distance, we need to check that

$$\|(a\mu - b\nu)K\| \;\leq\; (1-a(K)) \, \|a\mu - b\nu\| + a(K) \; |a-b|$$

which is clear from lemma 3.1, since $(a\mu - b\nu)(S) = a - b$.

4 On Orlicz's norms

The spirit of proposition 1.2 is quite similar to that of proposition 1.1; for instance if ψ is the typical mapping $\mathbb{R}_+ \ni t \mapsto t^p$ for a fixed $p \geq 1$, from the latter result we obtain that for any Markovian kernel K and any two probabilities $\mu \ll \nu$,

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

where Φ is the mapping $\mathbb{R}_+^2 \ni (x,y) \mapsto (x/y^{1-1/p} - y^{1/p})^p$.

Thus, as it is well-known that (H) implies $a(K) \geq \epsilon$, we already have

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

which is just a little less precise than what is deduced by the proposition 1.2. Nevertheless, its proof will be completely different and in some aspects more direct.

More generally, let us also mention that qualitatively speaking, at least under the doubling condition for ψ (ie there exists a constant $k > 1$ such that for all $x \geq 0$ large enough, $\psi(2x) \leq k\psi(x)$), 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 the theorem 9.4 p. 83 of [8]). Thus $\sum_{n\geq 0}\alpha(\Phi,K,\nu K^n)=+\infty$, and in particular $a(K)>0$, is sufficient to insure the latter behavior. But this not allows to deduce a nice quantitative exponential bound for this convergence (which in turn will unfortunately require the more restrictive hypothesis (H)).

We start by rewritting (H) in terms of the dual operators K_ν^*.

Lemma 4.1 *The condition (H) is equivalent to*

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

where $\epsilon>0$ will always designate the constant appearing in (H).

Proof:

Indeed, the above formulation means that

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

(the $+$ in subscripts indicate that we are only considering nonnegative functions) but we compute that the lhs is equal to

$$\begin{aligned}(\nu K)[K_\nu^*(f-\epsilon\nu(f))g] &= \nu[K(g)(f-\epsilon\nu(f))]\\ &= \nu[f(K(g)-\epsilon\nu(Kg))]\end{aligned}$$

Imposing that the latter quantity is nonnegative for all $f\in\mathbb{L}^1_+(\nu)$ and $g\in\mathbb{L}^\infty_+(\nu K)$ is just asking for

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

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

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

ie

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

from which (H) follows, by letting η tend to 0_+.

Reciproquely, the condition (H) meaning that for any $x,y\in S$ and $A\in\mathcal{S}$, $K(x,A)\geq\epsilon K(y,A)$, if we integrate this we 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.

■

This caracterisation leads to the following result, which is the main estimation needed for the proof of proposition 1.2.

Proposition 4.2 *Under the assumption (H), for any probability $\nu\in\mathcal{P}$, any Markovian kernel K, any convex mapping $\varphi:\mathbb{R}\to\mathbb{R}$ and any function $f\in\mathbb{L}^1(\nu)$, we are assured of*

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

Proof:

Let nu, K and φ be given as above. Due to lemma 4.1, the operator $R\stackrel{\text{def.}}{=}(1-\epsilon)^{-1}(K_\nu^*-\epsilon\nu)$ is Markovian. We remark that the probability ν is again KR-invariant:

$$\begin{aligned}
\nu KR &= \nu K\frac{K_\nu^*-\epsilon\nu}{1-\epsilon}\\
&= \frac{\nu KK_\nu^*}{1-\epsilon}-\frac{\epsilon\nu}{1-\epsilon}\\
&= \frac{\nu}{1-\epsilon}-\frac{\epsilon\nu}{1-\epsilon} \quad = \quad \nu
\end{aligned}$$

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

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

(in fact 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 in such a manner), thus integrating with respect to $R(y,dx)$ (which rigorously means letting acting R in that way), we obtain

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

In order to conclude, it is sufficient to integrate one more time with respect to $(\nu K)(dy)$:

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

■

Now the proof of proposition 1.2 is immediate:

Let ψ be the Young function appearing in the statement of that result, if μ, ν are such that $d\mu/d\nu\in\mathbb{L}^\psi(\nu)$, from the definition of the norm $\|\cdot\|_{\mathbb{L}^\psi(\nu)}$ and the previous proposition, applied with $f=a^{-1}d\mu/d\nu$, for $a>0$ (thus $K_\nu^*(f-\nu(f))=a^{-1}((d\mu K)/(d\nu K)-1)$, the expected bound follows at once.

But this inequality is even true if $\|d\mu/d\nu\|_{\mathbb{L}^\psi(\nu)}=+\infty$. Indeed, if $\epsilon<1$, it is clear and for $\epsilon=1$, due to our conventions, we have to show that $(d\mu K)/(d\nu K)=1$ (in $\mathbb{L}^\psi(\nu K)$), ie $\mu K=\nu K$.

But in this situation, we are assured of $K_\nu^*=\nu$ for any $\nu\in\mathcal{P}$, which can be rewritten by duality as $K=\nu K$. That means in fact that K is a probability, say $m\in\mathcal{P}$ and thus for any μ, $\nu\in\mathcal{P}$, $\mu K=m=\nu K$.

Remark 4.3: The above computations are still valid in the always verified case $\epsilon=0$, then we just get a contraction property which with the usual notations can be formulated as

$$\forall\ \mu\ll\nu\in\mathcal{P},\qquad \left\|\frac{d\mu K}{d\nu K}-1\right\|_{\mathbb{L}^\psi(\nu K)} \quad\leq\quad \left\|\frac{d\mu}{d\nu}-1\right\|_{\mathbb{L}^\psi(\nu)}$$

5 Modified logarithmic Sobolev constant

The purpose of this section is to consider more particularly the situation corresponding to the mappings Φ_1 and Φ_2 described in (5) and to obtain precise bounds for $\alpha(\Phi_1, K, \nu)$ 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 let us recall a result which was due to Fill [7] in the special case where the probability ν was assumed to be invariant for the Markovian kernel K: noting that for any $\mu, \nu \in \mathcal{P}$, the quantity $H_{\Phi_1}(\mu, \nu)$ is finite if and only if $\mu \ll \nu$ and $d\mu/d\nu \in \mathbb{L}^2(\nu)$, we will merely work under this condition. Then denoting by f the function $d\mu/d\nu - \mathbb{1}$, we compute that

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

where the Dirichlet form $\mathcal{E}_{\nu, KK_\nu^*}$ was defined in (9). If we take into account that $\nu(f) = 0$ and the fact that $\nu(f^2) = H_{\Phi_1}(\mu, \nu)$, we easily end up with

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

where the rhs is the spectral gap of KK_ν^* in $\mathbb{L}^2(\nu)$, ie

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

As the constant mapping $\mathbb{1}$ 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 restriction of this operator 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).

Now let us treat the Φ_2 situation: in quite a similar fashion, we define respectively the logarithmic Sobolev and the modified logarithmic Sobolev constants as

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

where in the sequel Ψ will always designate a fixed Young function chosen such that for large $x \in \mathbb{R}_+$, we have $\Psi(x) \sim x^2 \ln(x^2)$, and where 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 has not been much studied explicitly for itself in the literature, the notable exception being an article of Wu [12], nevertheless it has often be used implicitly (see for instance [5] or [11]) through the bound

$$l(\nu, KK_\nu^*) \geq 2\widetilde{l}(\nu, KK_\nu^*) \tag{13}$$

Basically, it comes 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 $|f|(x) > 0$, but it appears that this result is also trivially satisfied in the other case where $f(x) = 0$, due to the traditional conventions. Next by integration with respect to $|f|(x)\nu(dx)$, we get

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

which in turn leads to (13).

Another general fact is that we are also always assured of the inequality

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

which is proved by considering in the definition of the modified logarithmic Sobolev constant mappings of the form $f = \mathbb{1} + \epsilon g$ with $\epsilon > 0$ going to zero and g designating a generic function. The same argument shows that if indeed the infimum in (12) can be approximated through a sequence of functions converging to $\mathbb{1}$, then $l(\nu, KK_\nu^*) = 2\lambda(\nu, KK_\nu^*)$.

The logarithmic Sobolev constant is the usual object enabling to evaluate the quantitative decay of entropy, but the following result shows that indeed the modified logarithmic Sobolev constant is better suited for this purpose. And as we will see it below on natural examples, these coefficients can be very different, the former one being often null while this is not so true for the latter one.

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

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

Proof:

In order to simplify the presentation, we will work as if the operator K_ν^* was given by a Markovian kernel $K_\nu^*(x, dy)$, for $x, y \in S$. One could insure this property by taking appropriate topological assumptions on $(S, \mathcal{S})$, nevertheless the above proposition is true without them, since everything which follows can be justified from a operator point of view. But to adopt directly the latter approach would hide the probabilistic intuition, that is why we have prefered leave to the reader the not so difficult rewritting of the arguments below in an operator theory language.

Thus for instance we can now count on the equality of the measures $\nu(dx)K(x, dy)$ and $(\nu K)(dy)$ $K_\nu^*(y, dx)$, for $(x, y) \in S^2$. This fact is convenient to evaluate quantitatively the difference of interest, which is $\triangle \stackrel{\text{def.}}{=} \mathrm{Ent}(\mu|\nu) - \mathrm{Ent}(\mu K|\nu K)$ and which we already know to be nonnegative, when $\mu \in \mathcal{P}$ is satisfying $\mu \ll \nu$ with $\sqrt{f} \stackrel{\text{def.}}{=} \sqrt{d\mu/d\nu} \in \mathbb{L}^{\Psi}(\nu)$. Indeed, we compute that

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

But $y \in S$ being fixed, we recognize in the last integral the entropy of the nonnegative measure $f(x)K_\nu^*(y,dx)$ with respect to the probability $K_\nu^*(y,dx)$. In order to estimate it, 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 $f_t(x) = \exp(-t)f(x) + (1 - \exp(-t))$, and thus we end up with

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

This puts us in position to use the definition of $l(\nu, KK_\nu^*)$; the rhs is bounded below by

$$l(\nu,KK_\nu^*)\int_0^\infty dt \int f_t\ln(f_t)\,d\nu \;=\; l(\nu,KK_\nu^*)\int_0^\infty dt \int f_t\ln(f_t) - f_t + 1\,d\nu$$

since $\int f_t\,d\nu = 1$.

This leads us to introduce the next mappings

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

and we note that they admit the "doubling property": there exist a constant $0 < \rho < 1$ such that

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

(to be convinced of that, it is sufficient to study the behavior of the associated quotients respectively in $0, +\infty$ and 0).

By clear increasing and decreasing property of φ_+ and φ_-, these inequalities are still satisfied if in the rhs we replace $x/2$ by any gx with $1/2 \leq g \leq 1$. Next we observe that

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

In conclusion, 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 follows easily that

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

A similar inequality but in the other direction is a simple consequence of a Jensen inequality:

Let $f \in \mathbb{L}^\Psi(\nu) \setminus \mathrm{Vect}(\mathbb{1})$ be given, we want to minimize the quotient appearing in the definition (12), and for that purpose we can assume furthermore that $\int f^2\, d\nu = 1$, hypothesis which also leads us to introduce the probability $\mu(x) \stackrel{\text{def.}}{=} f^2(x)\nu(dx)$, for $x \in S$. Then we compute that

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

from which the expected lower bound of $l(\nu, KK_\nu^*)$ follows.

■

To complete the above proof, we still have to show an auxiliary result we have used, which will be obtained via continuous time arguments. The resort to them to treat a discrete time situation can seem strange at a first view, but it is dictated mainly from the fact that the form of the definition of the modified logarithmic Sobolev constant was itself inspired by continuous time considerations: if one is looking at a corresponding continuous time process of generator $KK_\nu^* - \mathrm{Id}$, then one is led to quotients of the type appearing in (12) by taking the temporal derivative of the entropy of the law of the time marginals of this process with respect to the reversible probability ν (the fact that the "natural" generator has to be $KK_\nu^* - \mathrm{Id}$ can already be understood by studying the $\mathbb{L}^2$ decay, ie the Φ_1 situation described previously).

Lemma 5.2 *Let $m, n \in \mathcal{P}$ be probabilities verifying $m \ll n$ with $\sqrt{f} \stackrel{def.}{=} \sqrt{dm/dn} \in \mathbb{L}^\Psi(n)$, then their relative entropy can be computed by the formula*

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

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

$$f_t(x) = \exp(-t)\frac{dm}{dn}(x) + 1 - \exp(-t)$$

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

Proof:

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

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

where $f_t = dm_t/dn = \exp(-t)f + (1 - \exp(-t))$.

The result then can be deduced by integration, due to the convergence

$$\lim_{t\to+\infty} \mathrm{Ent}(m_t|n) = 0$$

■

Remarks 5.3: a) Despite the presence of ρ, the proposition 5.1 really improves the previous result in that direction, which was that $\widetilde{l}(\nu, KK_\nu^*) \leq \alpha(\Phi_2, K, \nu)$ in the case of an invariant ν (cf [9]). Note also that this relation can be extended to the general case, following the same type of considerations we have used here.

b) For $\mu, \nu \in \mathcal{P}$ satisfying $\mu \ll \nu$ with $\sqrt{f} \stackrel{\text{def.}}{=} \sqrt{d\mu/d\nu} \in \mathbb{L}^\Psi(\nu)$, the next quantity has appeared in the previous considerations:

$$\begin{aligned} \mathcal{L}(\mu, \nu) &\stackrel{\text{def.}}{=} \mathcal{E}_{\nu,\nu}(f^2, \ln(f^2)) \\ &= \frac{1}{2}\int (f^2(x) - f^2(y))(\ln(f^2(x)) - \ln(f^2(y)))\, \nu(dx)\nu(dy) \end{aligned}$$

so we could have associated a new coefficient to the Markovian kernel K:

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

Indeed, following the same approach as above and removing the doubling property considerations, we are assured of

$$\bar{l}(\nu, KK_\nu^*) \leq \alpha(\Phi_2, K, \nu)$$

In particular, we deduce the bound

$$\bar{l}(\nu, KK_\nu^*) \leq l(\nu, KK_\nu^*) \tag{15}$$

which can also be obtained directly through the general inequality

$$\forall\, f\in\mathbb{L}^\Psi(\nu), \qquad \mathcal{E}_{\nu,\nu}(f^2,\ln(f^2)) \geq \mathrm{Ent}(f^2,\nu)$$

Unfortunately, there is no reverse inequality for the latter one, at least as soon as ν is not a Dirac mass, as can be seen by considering a function f verifying $\nu(\{f=0\}\sqcup\{f=1\})=1$ and $0<\nu(\{f=0\})<1$, because then the lhs is $+\infty$ while the rhs is finite.

Nevertheless, we are wondering if there would not exist a bound on the other direction for (15).

c) We also remark that an above computation can still be continued: keeping the same notations, we have seen that

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

and that leads us to introduce the function

$$\begin{aligned}\forall\, x>0, \qquad G(x) &= \int_0^\infty \exp(-t)\ln[1+\exp(-t)(x-1)]\,dt\\ &= \frac{x\ln(x)-x+1}{x-1}\end{aligned}$$

because we get that the former rhs is equal to

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

and thus we obtain a "variational functional caracterization" of $\alpha(\Phi_2,K,\nu)$:

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

d) Finally, let us return to the condition (H). Using the lemma 4.1, which also implies that $KK_\nu^*\geq \epsilon\nu$, we get directly from their definition that under (H), $\lambda(\nu,KK_\nu^*)\geq\epsilon$ and $\bar{l}(\nu,KK_\nu^*)\geq\epsilon$, and thus a fortiori $l(\nu,KK_\nu^*)\geq\epsilon$.

But as it was noted in [9], $\widetilde{l}(\nu,KK_\nu^*)=0$ as soon as ν is not a finite combination of Dirac masses. This simple constatation already shows the advantage one has to consider $l(\nu,KK_\nu^*)$ instead of $\widetilde{l}(\nu,KK_\nu^*)$ for nonfinite 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 points set $\{0,1\}$ endowed with its total σ-field $\mathcal{S}$ and assume that $K_\nu^* = \nu$ (ie K is a probability) with ν charging both points. Then we have

$$1 \quad \leq \quad l(\nu,\nu) \quad \leq \quad 2$$

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 [5], showing that in the same Bernoulli's distribution context

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

quantity which goes to zero with $\nu_* \stackrel{\text{def.}}{=} \nu(0) \wedge \nu(1)$.

On the contrary, the exact value of $l(\nu,\nu)$ does not seem to be known, and 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 $\lim_{\nu_* \to 0_+} l(\nu,\nu) = 1$.

But, using on one hand a tensorization property of the modified logarithmic Sobolev constant and on the other hand approximations of Poisson distributions by Bernoulli's 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 null), and by using appropriate exponential test functions, it can be shown that indeed it is exactly a^{-1} (cf the first chapter of the collective book [1]). This fact was already used by Wu [12] to study Poisson point processes.

6 A Gaussian example

In order to illustrate the potential usefulness of the "Lipschitz's point of view", we will consider in this section simple Gaussian kernels. This short study will also provide us with one more opportunity to be convinced of the relevence of the modified logarithmic Sobolev constant in this context.

We start by presenting continuous time elementary computations. Despite that we don't want to develop here a theory in that direction, let us mention that Lipschitz contraction properties hold in the same manner for general Markov processes and to obtain "rates of supercontraction" for them can also be very fruitful. Indeed, this kind of estimates are then essentially easier to get, because we are allowed to differentiate, as we will have a glimpse of it below.

So let $\widetilde{a}$, $\widetilde{b}$ and $\widetilde{c}$ be fixed real numbers, we consider the one dimensional diffusion $X \stackrel{\text{def.}}{=} (X_t)_{0\leq t\leq 1}$ strong solution of the sde

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

(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 exercice to solve this equation to obtain that if $\widetilde{b} \neq 0$,

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

and this equality leads us to introduce the Gaussian kernel K defined by

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

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

$$\begin{aligned} a &= \frac{\widetilde{a}}{\widetilde{b}}(\exp(\widetilde{b})-1) \\ b &= \exp(\widetilde{b}) \\ c &= \frac{\widetilde{c}^2}{2\widetilde{b}}(\exp(2\widetilde{b})-1) \end{aligned}$$

For any $0 \le t \le 1$, we denote by ν_t the law of X_t, if ν was the initial distribution of X_0. In particular, we have $\nu_1 = \nu K$. If μ is another initial probability on the real line, we would like to evaluate, as in the previous section, the evolution of the relative entropy $\mathrm{Ent}(\mu_t|\nu_t)$, now for $0 \le t \le 1$. To begin with, let us assume that ν and μ are equivalent to the Lebesgue measure λ (we will still designate by ν and μ their respective densities) and that the mapping 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 $f_t \stackrel{\text{def.}}{=} \mu_t/\nu_t$ at any time $0 \le t \le 1$. Thus these restrictive hypotheses enable at least to justify the following operations

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

where L is the generator of the diffusion X, which in particular acts on $\mathcal{C}^2_{\mathrm{b}}$ 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 that due to the continuity of the trajectories of X, this operator admits the next rule of composition: for instance if φ and f are $\mathcal{C}^2_{\mathrm{b}}$ mappings,

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

where Γ is the "carré du champ" associated to L, which is defined by this same equation but when φ is the particular mapping $\mathbb{R} \ni x \mapsto x^2$. We compute at once that in our case, it is given on appropriate test functions f by $\Gamma(f,f) = \widetilde{c}^2(\partial f)^2$. Then taking into account that we can deduce from these observations that 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

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

because in that way we obtain (remarking that in the infimum, $\mathcal{C}_c^2(\mathbb{R})\setminus\{0\}$ can be easily 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

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

which implies the estimate

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

Without much difficulty, this relation can be extended, first to any probability 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.

Now the problematic point is to evaluate the integral $\int_0^1 l(\nu_t)\,dt$. So let us consider the simple case where ν is itself a Gaussian distribution $\mathcal{N}(\widehat{a},\widehat{c})$. Traditional computations shows 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 bibliography given there) which makes a direct link between the logarithmic Sobolev constant of a Gaussian distribution and its variance

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

to be led to the following integration, via the variable change $s=\exp(2\widetilde{b}t)$,

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

with $A=2\widetilde{b}\widehat{c}/\widetilde{c}^2$

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

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

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

which means that

$$\begin{aligned}\alpha(\Phi_2, K, \nu) &\geq 1 - (1+\chi)^{-1} \\ &= \frac{\chi}{(1+\chi)}\end{aligned}$$

This estimate is in fact an equality, since by taking $\mu = \mathcal{N}(\bar{a}, \widehat{c})$, with $\bar{a} \neq \widehat{a}$, we find that

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

Note that this case is another occurence of $\widetilde{l}(\nu, KK_\nu^*) = 0$. But from the bounds of the previous section, it follows that $\chi/(1+\chi)$ is also the order of $l(\nu, KK_\nu^*)$, up to an universal factor. This result does not seem so trivial to derive 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 be given three sequences of real numbers $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$ and $(c_n)_{n\geq 0}$, with $c_n \geq 0$ for all $n \geq 0$. From them we define the Gaussian kernels

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

and a new sequence of positive numbers through the iteration

$$\forall\, n \geq 0, \qquad \widehat{c}_{n+1} \;=\; b_n^2\widehat{c}_n + c_n$$

starting with $\widehat{c}_0 = 1$.

Next we consider the assumption (H) saying that

$$\sum_{n\geq 0} \chi_n \;=\; +\infty$$

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

This hypothesis insures that for any probabilities $\mu, \widetilde{\mu} \in \mathcal{P}(\mathbb{R}, \mathcal{R})$, we have for instance

$$\lim_{n\to\infty} \|(\mu - \widetilde{\mu})K_0K_1\cdots K_n\| \;=\; 0 \tag{16}$$

To see this result, let $\nu = \mathcal{N}(0, 1)$, noting that $\widehat{c}_n$ is the variance of $\nu K_0K_1\cdots K_n$, we obtain that for $n \geq 0$,

$$\mathrm{Ent}(\mu K_0K_1\cdots K_n|\nu K_0K_1\cdots K_n) \;\leq\; \frac{1}{1+\chi_n}\mathrm{Ent}(\mu K_0K_1\cdots K_{n-1}|\nu K_0K_1\cdots K_{n-1})$$

and thus under (H), we get

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

which 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 inequality

$$\|\mu - \nu\| \leq \sqrt{2\mathrm{Ent}(\mu|\nu)}$$

The expected behavior (16) follows at once.

Note that (H) is in particular implied by the assumption that

$$\begin{aligned} \sup_{n\geq 0} c_n &< +\infty \\ \limsup_{n\to\infty} b_n &< 1 \end{aligned}$$

which insures that $\sup_{n\geq 0} \widehat{c}_n < +\infty$.

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 relative to K_n. A classical way to prove a result such as (16), is to show before that for any probability $\mu \in \mathcal{P}(\mathbb{R}, \mathcal{R})$,

$$\lim_{n\to\infty} \|\mu K_0 K_1 \cdots K_n - \nu_n\| = 0 \qquad (17)$$

(or sometimes more conveniently $\lim_{n\to\infty} \mathrm{Ent}(\mu K_0 K_1 \cdots K_n|\nu_n) = 0$).

Nevertheless, it should be noticed that (17) is not always satisfied under (H), 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 occurence is not only theorical, as it often appears in the context of linear and nonlinear filtering).

Thus even if there exists an invariant probability, it may be not interesting to consider it. It was this simple constatation which convinced us that the Lipschitz's point of view is a pertinent tool in the study of loose of memory property in the field of Markov chains theory. Indeed 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 mechanic. But a priori its first interest is to enable the treatment of settings where there is no invariant probability, as when we assume in the above example that $|b_n| > 1$ for all $n \geq 1$; (H) can still be fulfilled by considering c_n large enough.

b) Note that the above inhomogeneous example and the conclusions drawn from it can be easily rewritten in the 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, but this is not entirely true, since Bakry [2] has developed a whole technology to derive such estimates in continuous time, which furthermore are easier to obtain in compact state space context, through comparaison of Dirichlet forms and measures techniques.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. Autour de l'inégalité de Sobolev logarithmique. Preprint, 2000.

[2] D. Bakry. On sobolev and logarithmic sobolev inequalities for markov semigroups. In K.D. Elworthy, S. Kusuoka, and I. Shigekawa, editors, *New Trends in Stochastic Analysis*, pages 43–75. World Scientific, 1997.

[3] J.E. Cohen, Y. Iwasa, G. Răuţu, M.B. Ruskai, E. Seneta, and G. Zbăganu. Relative entropy under mappings by stochastic matrices. *Linear Algebra Appl.*, 179:211–235, 1993.

[4] I. Csiszár. 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.

[5] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 6(3):695–750, 1996.

[6] R.L. Dobrushin. Central limit theorem for nonstationnary Markov chains, I, II. *Theory of Probability and its Applications*, 1(1 and 4):66–80 and 330–385, 1956.

[7] J.A. Fill. 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.

[8] M.A. Kranosel'skii and Y.B. Rutickii. *Convex functions and Orlicz spaces.* P. Noordhoff, 1961.

[9] L. Miclo. 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, pages 136–167. Springer-Verlag, Berlin, 1997.

[10] D. Revuz. *Markov chains.* North-Holland, 1975.

[11] L. Saloff-Coste. 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.

[12] L. Wu. A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. Preprint, 1999.

[13] R. Zaharopol and G. Zbaganu. Dobrushin coefficients of ergodicity and asymptotically stable L^1-contractions. *J. Theoret. Probab.*, 12(4):885–902, 1999.