Around Nash Inequalities

Dominique Bakry, François Bolley, Ivan Gentil

Around Nash inequalities

Dominique Bakry*†, François Bolley‡ and Ivan Gentil§

September 10, 2010

Introduction

In the Euclidean space $\mathbb{R}^n$, the classical Nash inequality may be stated as

$$\|f\|_2^{1+n/2} \leq C_n \|f\|_1 \|\nabla f\|_2^{n/2} \tag{0.1}$$

for all smooth functions f (with compact support for instance) where the norms are computed with respect to the Lebesgue measure. This inequality has been introduced by J. Nash in 1958 (see [9]) to obtain regularity properties on the solutions to parabolic partial differential equations. The computation of the optimal constant C_n has been performed more recently in [6].

This inequality may be stated in the general framework of symmetric Markov semigroups, where it is a simple and powerful tool to obtain estimates on the associated heat kernel. In this context, one replaces $\|\nabla f\|_2^2$ by the Dirichlet form $\mathcal{E}(f,f)$ associated with the semigroup, and the Lebesgue measure by its reversible measure. Moreover, the power function x^n in the inequality is replaced by a more general convex function Φ, and under this form it can be valid (and useful) even in infinite dimensional situations such as those which appear in statistical mechanics. One can also give weighted forms of these inequalities : they also lead to precise estimates on the semigroup, or on the spectral decomposition of the generator.

The aim of this short note is to explain how Nash inequalities lead to such estimates in a general setting and also to show simple techniques used to establish the required Nash inequalities. There is no claim for originality, most of the material included here may be found in various papers such as [1, 2, 5, 7, 13].

*Institut de Mathématiques de Toulouse, UMR CNRS 5219, Université Paul-Sabatier
†Institut Universitaire de France.
‡Ceremade, UMR CNRS 7534, Université Paris-Dauphine.
§Institut Camille Jordan, UMR CNRS 5208, Université Claude Bernard Lyon 1.

Nash inequalities belong to the very large family of functional inequalities for symmetric Markov semigroups which have led to many recent works. Many of these inequalities compare $\mathcal{L}^p$ norms of functions to the $\mathcal{L}^2$ norms of their gradients, which in this context is called the Dirichlet form; this is the case of the simplest ones, the spectral gap (or Poincaré) inequalities. But one may also consider $\mathcal{L}^1$ norms of the gradients, in the area of isoperimetric inequalities, or $\mathcal{L}^p$ norms, even $\mathcal{L}^\infty$ norms, when one is concerned with estimates on Lipshitz functions, for instance in the area of concentration of measure phenomena.

Here, we shall concentrate only on $\mathcal{L}^2$ norms of gradients. Even in this setting, there exists a wide variety of inequalities, which are adapted to the kind of measure one wants to study on one side, and to the properties they describe on the other. For example, measures with polynomial decay are not covered by the same inequalities as measures with exponential, or square exponential decay.

The family of Nash type inequalities we present here belongs to the wide family of the Sobolev type inequalities. Their main interest is that they easily provide good (and sometimes almost optimal) control on heat kernels. Starting from the classical inequality, we shall show how to extend them first by the introduction of a rate function Φ, and then by the extra introduction of a weight function V (a Lyapunov function). As we shall see, the link between Nash inequalities and estimates on the semigroup spectrum is very simple and, as usual in the field, roughly relies on derivation along time and integration by parts. This is why it is tempting to use it in a wide range of situations.

Then, we shall show how to obtain these inequalities in the simplest models on the real line. Restricting ourselves to the real line may be thought as looking only at the easy case. In fact, by choosing various measures, one may produce a lot of different model cases which really illustrate what may or may not be expected from these inequalities. Then the extension to higher dimensional situations (like $\mathbb{R}^n$ or manifolds) is very often a pure matter of technicalities, extending in a direct way the one-dimensional methods.

The paper is organized as follows. In the first section, we briefly present the context of symmetric Markov semigroups, and particularly diffusion semigroups. Then, we show different variations of Nash inequalities and how to get estimates on heat kernels from them. Then, in fine, we show how to produce such Nash inequalities on the basic models on the real line we are interested in.

1 Symmetric Markov semigroups and difusions

To understand the general context of Markov semigroups, we first consider a measure space $(E, \mathcal{B}, \mu)$, where $\mathcal{B}$ is a σ-field and μ is a σ-finite measure on it. Although we shall always focus on examples where $(E, \mathcal{B})$ is $\mathbb{R}^n$ equipped with the usual Borel sets (or some open set in it, or a finite dimensional manifold with or without boundaries), it may be an infinite dimensional space, as we already mentioned, in which case one has to be careful

about the measurable structure of the space. In any case, one should always suppose that $(E, \mathcal{B}, \mu)$ is a "reasonable" measure space : we shall not say in details what we mean by "reasonable", but results such as the decomposition of measure theorems should be valid, which covers all cases one could look at in practise.

Given $(E, \mathcal{B}, \mu)$ a symmetric Markov semigroup is a family $(P_t)_{t\geq 0}$ of linear operators mapping the set of bounded measurable functions into itself with the following properties:

(i) Preservation of positivity : if $f \geq 0$, so is $P_t f$.

(ii) Preservation of constant functions : $P_t\, 1 = 1$.

(iii) Semigroup property : $P_t \circ P_s = P_{t+s}$.

(iv) Symmetry : P_t maps $\mathcal{L}^2(\mu)$ into itself and, for any pair $(f, g) \in \mathcal{L}^2(\mu)$, one has

$$\int_E P_t f g \, d\mu = \int_E f P_t g \, d\mu.$$

(v) Continuity at $t = 0$: $P_0 = Id$ and $P_t f \to f$ when $t \to 0$ in $\mathcal{L}^2(\mu)$.

Such semigroups naturally appear in probability theory as $P_t f(x) = \mathbb{E}(f(X_t)/X_0 = x)$ where $(X_t)_{t\geq 0}$ is a Markov process. The symmetry property does not always hold and it is equivalent to the reversibility of the process. They also appear in many situations when one solves a "heat equation" of the form

$$\partial_t F(x,t) = LF, \;\; F(x,0) = f(x);$$

here L is a second order sub-elliptic (or hypo-elliptic) differential operator, in which case $P_t f$ is the solution $F(x,t)$ at time t when the initial value is $F(x,0) = f(x)$.

Let us start with some elementary preliminary remarks.

(a) Since P_t is symmetric, and $P_t\, 1 = 1$, one gets $\int_E P_t f d\mu = \int_E f d\mu$ by taking $g = 1$ in the property (iv): μ is *invariant* under the semigroup.

(b) Since P_t is linear and positivity preserving, $|P_t f| \leq P_t |f|$. This implies that P_t is a contraction in $\mathcal{L}^1(\mu)$ by invariance of the measure.

(c) By the same argument, P_t is also a contraction in $\mathcal{L}^\infty(\mu)$ and therefore, by interpolation, P_t is a contraction in $\mathcal{L}^p(\mu)$ for any $p \in [1, \infty]$.

(d) Since $(P_t)_t$ is a semigroup of contractions in $\mathcal{L}^2(\mu)$, by the Hille-Yoshida theory, it admits an infinitesimal generator L, which is densely defined by $Lf = \partial_t P_t f$ at $t = 0$. Then $P_t f$ is the solution at time t of the heat equation $\partial_t F = LF$, given $F(x,0) = f(x)$ at time $t = 0$.

Formally, any property of the semigroup may be translated into a property of the generator L, and vice versa. For instance, the preservation of constants property translates into $L1 = 0$. Also, the symmetry translates into the fact that L is self-adjoint, that is,

$$\int_E fLg\, d\mu = \int_E gLf d\mu.$$

The positivity preserving property is more subtle. In general, it is translated into a maximum principle of the generator. But this requires a bit more than just a measurable structure on the space, and we prefer to translate this into the positivity of the *carré du champ* operator, see point (k) below.

(e) The measure μ being symmetric (or "reversible") is in general unique up to a normalizing constant (it is however a restrictive condition that such a measure exists : see formula (1.3) below). When the measure is finite, we may therefore normalize it as to be a probability measure, and we shall always do it. In this case, the constant function 1 is always a normalized eigenvector, associated with the eigenvalue 0 which is the smallest value of the spectrum of $-L$. In the infinite case, there is no canonical way of choosing a good normalization.

(f) Since the measure space $(E, \mathcal{B}, \mu)$ is a "reasonable" space, any such operator P_t which preserves the constants and positivity may be represented as

$$P_t f(x) = \int_E f(y) P_t(x, dy),$$

where $P_t(x, dy)$ is a kernel of probability measures, that is, a probability measure on E depending on the parameter $x \in E$ in a measurable way. This enables for example to apply to P_t any generic property of probability measures, such as the variance inequality $P_t(f^2) \geq (P_t f)^2$.

(g) Very often (and we shall see that Nash inequalities provide a useful criterium for this), this kernel has a *density* with respect to the reversible measure μ, that is

$$P_t(x, dy) = p_t(x, y)\mu(dy);$$

here $p_t(x, y)$ is a non negative function which is defined almost everywhere (with respect to $\mu \otimes \mu$) on $E \times E$. Then the symmetry property (iv) is equivalent to the symmetry of this kernel $p_t(x, y) = p_t(y, x)$. Much attention has been brought over the last decades to various estimates on this kernel density (in particular in Riemannian geometry, for heat kernels on Riemannian manifolds, using tools from geometry like curvature, Riemannian distance, etc). Once again, Nash inequalities may provide good such estimates, as we shall see later on.

(h) When we have such densities, the semigroup property translates into the Chapman-Kolmogorov equation

$$p_{t+s}(x, y) = \int_E p_t(x, z) p_s(z, y) \mu(dz).$$

Hence, by the Cauchy-Schwarz inequality,

$$p_{2t}(x,y)^2 \leq p_{2t}(x,x)\, p_{2t}(y,y).$$

As the consequence, the maximum of $p_t(x,y)$ is always obtained on the diagonal.

(i) The generator L being self-adjoint has a spectral decomposition with a spectrum in $(-\infty, 0]$ according to (1.2).

(j) It may be the case that the spectrum is discrete, and that we have a complete sequence of orthonormal eigenvectors (f_n) in $\mathcal{L}^2(\mu)$, with eigenvalues $-\lambda_n$ for L. In this situation, the kernel density $p_t(x,y)$ is given by

$$p_t(x,y) = \sum_n e^{-\lambda_n t} f_n(x) f_n(y).$$

Then we have the trace formula

$$\int_E p_t(x,x) d\mu(x) = \sum_n e^{-\lambda_n t}.$$

Once again, Nash inequalities will provide uniform (or non uniform) bounds on the densities, and therefore bounds on the counting function of the sequence (λ_n).

(k) By derivation at $t = 0$ the variance inequality $P_t(f^2) \geq (P_t f)^2$ gives the inequality

$$L(f^2) \geq 2fLf.$$

In particular

$$\int_E fLf d\mu \leq 0 \tag{1.2}$$

by invariance of μ. Of course, one has to take care about which functions these do apply. In general, we assume that there is an algebra of functions $\mathcal{A}$ dense in the domain of L, for which this is valid. In this case, one defines the *carré du champ* as the bilinear form

$$\Gamma(f,g) = \frac{1}{2}\big(L(fg) - fLg - gLf\big).$$

It satisfies $\Gamma(f,f) \geq 0$, and in some sense this characterizes the positivity preserving property of P_t.

The *Dirichlet form* associated to P_t is finally defined by

$$\mathcal{E}(f,g) = -\int gLf d\mu = -\int fLg\, d\mu = \int \Gamma(f,g) d\mu.$$

The last identity is based on the identity $\int_E L(fg)d\mu = 0$ and is called the *integration by parts formula*. The Dirichlet form is in general defined on a larger domain than the operator L itself (formally, it requires only one derivative of the function to be in $\mathcal{L}^2(\mu)$ instead of 2 for the generator).

The knowledge of the measure and of the *carré du champ* (or of the Dirichlet form) entirely describes the operator L (and therefore the semigroup), since L may be defined from Γ and μ through the integration by parts formula (see (1.3)).

The basic example of such semigroups is of course the standard heat kernel in the Euclidean space $\mathbb{R}^n$; for $t > 0$, its density $p_t(x, y)$ with respect to the Lebesgue measure dy is

$$p_t(x, y) = \frac{1}{(4\pi t)^{n/2}} \exp(-\frac{|x - y|^2}{4t}).$$

Here, $\mu(dy) = dy$, $L = \Delta$ and

$$\Gamma(f, f) = |\nabla f|^2.$$

This corresponds to the case studied by Nash in [9]. It is one of the very few examples where one explicitly knows P_t, since in general we only know L, and the issue is to deduce as much information as possible on P_t from the knowledge of L.

Another model case is the Ornstein-Uhlenbeck semigroup on $\mathbb{R}^n$, for which

$$Lf(x) = \Delta f(x) - x \cdot \nabla f(x),\ \Gamma(f, f) = |\nabla f|^2,\ \mu(dx) = \frac{1}{(2\pi)^{n/2}} \exp(-|x|^2/2)dx.$$

Its density with respect to the Gauss measure $\mu(dy)$ is

$$p_t(x, y) = (1 - e^{-2t})^{-n/2} \exp\left[-\frac{1}{2(1 - e^{-2t})}(|y|^2 e^{-2t} - 2\, x \cdot y e^{-t} + |x|^2 e^{-2t})\right]$$

and it behaves in a very different way from the previous example as long as functional inequalities are concerned.

In the two previous cases, the *carré du champ* is the same (and the semigroups only differ by the measure $\mu(dx)$). Observe that $\Gamma(f, g)$ is in both cases a first order differential operator in its two arguments. They both belong to the large class of *diffusion* Markov semigroups, which are semigroups such that for all smooth functions ϕ

$$\Gamma(\phi(f), g) = \phi'(f)\Gamma(f, g)$$

or equivalently

$$L\phi(f) = \phi'(f)Lf + \phi''(f)\Gamma(f, f).$$

This property is called the *change of variable formula* for L and is an intrinsic way of saying that L is a second order differential operator. The fact that $L(1) = 0$ says that there is no 0-order term in L. One may easily check that among all differential operators on $\mathbb{R}^n$ (or a manifold) with smooth coefficients, only the second order ones may satisfy

$\Gamma(f,f) \geq 0$, provided the matrix of the second order terms is positive-semidefinite at any point.

Non diffusion cases are of considerable interest since they are related to Markov processes with jumps and also naturally appear when one looks at subordinators. However we shall concentrate here on the diffusion case, even though the Nash techniques may be used in the same way in the general case.

In general, a second order differential operator without 0-order terms of the form

$$Lf = \sum_{ij} a^{ij}(x)\partial^2_{ij} f + \sum_i b^i(x)\partial_i f$$

has a *carré du champ* given by

$$\Gamma(f,g) = \sum_{ij} a^{ij}(x)\partial_i f \partial_j g = \nabla f \cdot A(x)\nabla g.$$

Therefore the positivity of Γ is equivalent to the fact that at any point x the matrix $A(x) = (a^{ij}(x))$ is positive-semidefinite. Conversely, when the *carré du champ* is given (on a open set in $\mathbb{R}^n$ or on a smooth manifold in local coordinates) by

$$\Gamma(f,g) = \sum_{ij} a^{ij}(x)\partial_i f \partial_j g,$$

with positive-semidefinite matrices $(a^{ij}(x))$ having smooth coefficients, and when the reference measure $\mu(dx)$ has a smooth positive density $\rho(x)$ with respect to the Lebesgue measure, it corresponds to a unique symmetric diffusion operator L which is

$$Lf = \frac{1}{\rho(x)} \sum_i \partial_i \Big(\rho(x) \sum_j a^{ij}(x)\partial_j f \Big). \tag{1.3}$$

In other words, Γ codes for the second order part of the operator while μ codes for the first order terms. Observe also that each (Γ, μ) leads to unique *symmetric* L, but possibly several non symmetric L.

A model case on which we shall focus is the case when $E = \mathbb{R}$ and $\Gamma(f,f) = f'^2$. We shall look at the measures

$$\mu(dx) = C \exp(-|x|^a)dx, \tag{1.4}$$

where $a > 0$ and C is a normalizing constant. In order to avoid irrelevant difficulties due to the non smoothness of $|x|$ at 0, we shall replace $|x|$ by $\sqrt{1+x^2}$. Depending on the value of a, the corresponding semigroups present diverse behaviours.

For $a = 2$, the celebrated Nelson theorem [10] asserts that the Ornstein-Uhlenbeck semigroup is "hypercontractive", which means that P_t is bounded from $\mathcal{L}^2(\mu)$ to $\mathcal{L}^{q(t)}(\mu)$

for all $t > 0$ and some $q(t) > 2$. This is equivalent to the also famous Gross logarithmic Sobolev inequality [8]

$$\int f^2 \log f^2 d\mu \leq \int f^2 d\mu \, \log\Big(\int f^2 d\mu\Big) + C\mathcal{E}(f,f). \tag{1.5}$$

When $a > 2$, the semigroup is "ultracontractive", which means that P_t maps $\mathcal{L}^1(\mu)$ into $\mathcal{L}^\infty(\mu)$ for any $t > 0$, while it is not even hypercontractive for $a < 2$. Nevertheless, for $1 < a < 2$, it has a discrete spectrum and it is Hilbert-Schmidt, and we shall see below how to get estimates on the spectrum through weighted Nash inequalities.

For $a = 1$, the spectrum is no longer discrete and the only property left is the existence of a spectral gap : the spectrum of $-L$ lies in $\{0\} \cup [A, \infty)$ for some $A > 0$, and this property is equivalent to a spectral gap (or Poincaré) inequality

$$\int f^2 d\mu \leq \Big(\int f d\mu\Big)^2 + \frac{1}{A}\mathcal{E}(f,f). \tag{1.6}$$

When $a < 1$, even the spectral gap property is lost.

Of course, one may look at similar models in $\mathbb{R}^n$, or on Riemannian manifolds with density measures (with respect to the Riemann measure) depending on the distance to some point. In this latter case, one would get more complicated results, since in general one has to take into account lower bounds on the Ricci curvature, and even more if one works with boundaries (with Neumann boundary conditions). We shall not develop this here.

2 Nash inequalities

In the context of Dirichlet forms associated to symmetric Markov semigroups as described above, a Nash inequality is an inequality of the form

$$\|f\|_2^{1+n/2} \leq \|f\|_1 \big\|\big[C_1\|f\|_2^2 + C_2\mathcal{E}(f,f)\big]^{n/4}; \tag{2.7}$$

here the norms $\mathcal{L}^p$ are of course computed with respect to the reversible measure μ and n is any positive parameter (that we call the dimension in the Nash inequality, since in the classical case the unique possible value for n is really the dimension of the space). This inequality should apply for any f in the Dirichlet domain, but it is enough to check it in a dense subspace of it which, in many examples, will be the set of smooth compactly supported functions.

It is worth mentioning that when μ is a probability measure, then $C_1 \geq 1$ (as can be seen by choosing $f = 1$), while for example in $\mathbb{R}^n$ with the Lebesgue measure, one may have $C_1 = 0$, as it is the case for the classical Nash inequality.

When μ is a probability measure and $C_1 = 1$, we say that the inequality is tight. It then implies a spectral gap inequality, as one may see by applying the inequality to $1+\epsilon f$ and letting ϵ go to 0.

Conversely, if (2.7) is valid with $C_1 > 1$ and with μ being a probability measure, together with a spectral gap inequality, then a tight Nash inequality holds (see [1] for example). In general, we say that a functional inequality is tight when one may deduce from the inequality that $\{\mathcal{E}(f,f) = 0\} \Longrightarrow \{f = \text{constant}\}$. Here, when $C_1 = 1$, this is ensured by the equality case in the inequality $\|f\|_1 \leq \|f\|_2$. As we shall see, tightness may be used to control the convergence to equilibrium, that is the asymptotic behaviour when $t \to \infty$, while the general inequality is useful to control the short time behaviour. Most of functional inequalities may be tightened in presence of a spectral gap inequality, as it is the case here.

In the case of an infinite measure, tightness corresponds to the case when $C_1 = 0$, as in the Euclidean case.

However, there is a strong difference between the forms that the Nash inequalities may take according to whether the measure is finite or not. We know that a tight Nash inequality holds true in the Euclidean space, but it can be proved that the tight Nash inequality (2.7) may not be valid on a finite measure space unless the space is compact. More precisely, when one has a tight Nash inequality (2.7) on a finite measure space, one may get a bound on the oscillation of Lipchitz functions, whence a bound on the diameter of the space (this diameter being measured in terms of an intrinsic distance associated with the *carré du champ* [4]). This explains why below we introduce the extended Nash inequalities (2.9) and (2.10), which may be valid on finite measure spaces with unbounded support, as we shall see.

When $n > 2$, the Nash inequality (2.7) is one of the many forms of the Sobolev inequality

$$\|f\|^2_{2n/(n-2)} \leq C_1\|f\|^2 + C_2\mathcal{E}(f,f).$$

Indeed, observe first that this and Hölder's inequalities lead to the Nash inequality (2.7) with the same "dimension" n and the same constants C_1 and C_2. The way back is a little more subtle: the argument in [3] is based on applying the Nash inequality to the sequence of functions $f_n = \min\{(f-2^n)_+, 2^n\}$, adding the obtained estimates and using the identity $\sum_n \mathcal{E}(f_n,f_n) = \mathcal{E}(f,f)$. This enables to keep the same dimension n, but not the constants C_1 and C_2.

In the context of Dirichlet forms, such Sobolev inequalities may appear under different forms such as Energy-Entropy, Gagliardo-Nirenberg, Faber-Krahn etc inequalities. We refer to [3] for full details. The connection between Sobolev (and Nash) inequalities and various bounds on heat kernels has been explored by many authors, see [1, 7, 11] for example.

We have the following

Theorem 2.1. *Assume that inequality* (2.7) *holds. Then,*

$$\|P_t f\|_2 \leq C(t)\|f\|_1, \tag{2.8}$$

where

$$C(t) = \Big(\max\{2C_1, \frac{2nC_2}{t}\}\Big)^{n/4}.$$

Conversely, if (2.8) *holds with* $C(t) \leq a + bt^{-n/4}$*, then a Nash inequality* (2.7) *holds with the same dimension* n *and constants* C_1 *and* C_2 *depending only on* n, a *and* b*.*

Proof — Let us rewrite the inequality under the form

$$\Big(\frac{\|f\|_2^2}{\|f\|_1^2}\Big)^{1+2/n} \leq C_1 \frac{\|f\|_2^2}{\|f\|_1^2} + C_2 \frac{\mathcal{E}(f,f)}{\|f\|_1^2}.$$

Now, choose a positive function f and apply the preceding bound to $P_t f$. We know from invariance of μ that $\int_E P_t f d\mu = \int_E f d\mu$.

Let us set $H(t) = \frac{\|P_t f\|_2^2}{\|P_t f\|_1^2} = \frac{\|P_t f\|_2^2}{\|f\|_1^2}$. We have

$$\partial_t \|P_t f\|_2^2 = 2\int P_t f L P_t f \, d\mu = -2\,\mathcal{E}(P_t f, P_t f).$$

Therefore, H is decreasing and

$$H'(t) = -2\frac{\mathcal{E}(P_t f, P_t f)}{\|P_t f\|_1^2},$$

and the Nash inequality (2.7) becomes

$$H^{1+2/n} \leq C_1 H - 2C_2 H'.$$

Now, as long as $H \geq (2C_1)^{n/2}$, one has $H^{1+2/n} \geq 2C_1 H$, and we get

$$H^{1+2/n} \leq -4C_2 H',$$

and this differential inequality (with the fact that H is decreasing) gives the result.

To see the reverse way, we may observe that, for a general symmetric Markov semigroup, the function $t \mapsto K(t) = \log \|P_t f\|_2^2$ is convex. Indeed, the derivative of $K_1(t) = \|P_t f\|_2^2$ is $-2\int P_t f L P_t f d\mu$, while the second derivative is $4\int (LP_t f)^2 d\mu$, and therefore by Cauchy-Schwarz inequality one has

$$K_1'^2 \leq K_1 K_1'',$$

which says that $\log K_1 = K$ is convex. Then so is the function $h(t) = \log \frac{\|P_t f\|_2^2}{\|P_t f\|_1^2} = \log H(t)$, so that

$$H'(0) \leq \frac{H(0)}{t} \log \frac{H(t)}{H(0)}.$$

Now, if we have a bound of the form $H(t) \leq a + bt^{-n/2}$, we may plug this upper bound in the previous inequality, and then optimise in t to get the result. ∎

In fact, having a bound for P_t as an operator from $\mathcal{L}^1$ into $\mathcal{L}^2$, we are very close from a uniform bound on the kernel p_t. Indeed, if P_t is bounded from $\mathcal{L}^1$ into $\mathcal{L}^2$ with norm $C(t)$, then by symmetry and duality, it is also bounded from $\mathcal{L}^2$ into $\mathcal{L}^\infty$ with norm $C(t)$, and therefore by composition and semigroup property P_{2t} is bounded from $\mathcal{L}^1$ into $\mathcal{L}^\infty$ with norm $C(t)^2$.

Conversely, by the Riesz-Thorin theorem, if P_t is bounded from $\mathcal{L}^1$ into $\mathcal{L}^\infty$ with norm $C_1(t)$, being bounded from $\mathcal{L}^1$ into itself with norm 1, it is also bounded from $\mathcal{L}^1$ into $\mathcal{L}^2$ with norm $C_1(t)^{1/2}$. In the end, we have obtained the following

Theorem 2.2. *A Nash inequality* (2.7) *holds with dimension n if and only if P_t is bounded from $\mathcal{L}^1$ into $\mathcal{L}^\infty$ with norm bounded above by $a + bt^{-n/2}$.*

Of course, in the case when $C_1 = 0$, which corresponds to the classical Euclidean Nash inequality, the equivalence is valid with a bound of the form $C(t) = at^{-n/2}$.

Moreover, a very general fact (valid on "reasonable measure spaces" $(E, \mathcal{B}, \mu)$) asserts that an operator K is bounded from $\mathcal{L}^1(\mu)$ into $\mathcal{L}^\infty(\mu)$ if and only if it may be represented by a bounded kernel density k : $K(f)(x) = \int_E k(x,y)f(y)\mu(dy)$. Moreover, the norm operator of K is exactly the $\mathcal{L}^\infty$ norm of k (on $E \times E$).

So we have seen that a Nash inequality is equivalent to a uniform bound on the kernel of P_t (and also carries the existence of such kernel), with very few assumptions on the space.

Observe that there is no reason why we should restrict ourselves to the case of power functions in Nash inequalities. One may consider extensions of the form

$$\Phi\Big(\frac{\|f\|_2^2}{\|f\|_1^2}\Big) \leq \frac{\mathcal{E}(f,f)}{\|f\|_1^2}, \tag{2.9}$$

valid say whenever $\|f\|_2 > M\|f\|_1$. Here Φ is a smooth convex increasing function defined on an interval (M, ∞). (It does not require formally to be convex increasing, but it is really useful only in this case).

Such inequalities have been introduced by F.-Y. Wang in [12] under the form

$$\|f\|_2^2 \leq a\mathcal{E}(f,f) + b(a)\|f\|_1^2,$$

called *super Poincaré inequalities*. These inequalities may be optimized under the parameter a to give

$$\frac{\|f\|_2^2}{\|f\|_1^2} \leq \Psi\Big(\frac{\mathcal{E}(f,f)}{\|f\|_1^2}\Big)$$

with some concave function Ψ, which is equivalent to inequality (2.9).

Then, one can write the argument of Theorem 2.1 with (2.9) instead of (2.7) and we see that the key assumption is $\int^\infty \frac{1}{\Phi(s)} ds < \infty$:

Theorem 2.3 (Wang). *Assume that an extended Nash inequality* (2.9) *is valid with a rate function* Φ *defined on some interval* (M, ∞) *and such that* $\int^{\infty} \frac{1}{\Phi(s)} ds < \infty$. *Then we have*

$$\|P_t f\|_2 \leq K(2t) \, \|f\|_1$$

for all $t > 0$ *and all functions* $f \in \mathcal{L}^2(\mu)$; *here the function* K *is defined by*

$$K(x) = \begin{cases} \sqrt{U^{-1}(x)} & \text{if } 0 < x < U(M), \\ \sqrt{M} & \text{if } x \geq U(M) \end{cases}$$

where U *denotes the (decreasing) function defined on* $(M, +\infty)$ *by*

$$U(x) = \int_x^{\infty} \frac{1}{\phi(u)} du.$$

In particular, the density $p_t(x, y)$ *is bounded from above by* $K(t)^2$.

Conversely, if there exists a positive function K *defined on* $(0, \infty)$ *such that*

$$\|P_t f\|_2 \leq K(t) \|f\|_1$$

for all $t > 0$, *then the Nash inequality* (2.9) *holds with* $M = 0$ *and function*

$$\Phi(x) = \sup_{t>0} \frac{x}{2t} \log \frac{x}{K(t)^2}, \quad x \geq 0.$$

With the techniques presented in the next section, we may see that such extended Nash inequalities with functions Φ of the form $x(\log x)^{\alpha}$ are adapted to the study of the the measures μ_a described in (1.4) for $a > 2$: as we already mentioned, because of non compactness, there is no hope in this case to have a classical Nash inequality (2.7) with a power function Φ.

In the case when the measure is finite (and therefore a probability measure), then we know that $\|f\|_2 / \|f\|_1 \geq 1$. For such a general inequality, tightness corresponds to the fact that $\Phi(x) \to 0$ when $x \to 1$. (Of course, this supposes that $M = 1$ in the previous theorem).

In this situation, assume that $\Phi(x) \sim \lambda(x - 1)$ when $x \to 1$ and $1/\Phi$ is integrable at infinity. This is the case in particular for the tight form of the classical Nash inequality (2.7). Then $K(t) \sim 1 + Ce^{-\lambda t}$ when $t \to \infty$. This shows that the kernel $p_t(x, y)$ is bounded from above by a quantity which converges exponentially fast to 1 as t goes to infinity. This is what may be expected, since $P_t f \to \mu(f)$ when $t \to \infty$. In the case of a classical tight Nash inequality (which can only occur when the measure has a bounded support), then one may also deduce a lower uniform bound on the kernel p_t which also goes exponentially fast to 1, but this requires some other techniques (see [1]).

The different Nash inequalities introduced so far may only carry information on the heat kernel in case of ultracontractivity, that is, when the kernel density is bounded. In

the general case when it is not bounded we may still use this method with the trick of introducing an auxiliary Lyapunov function V and *weighted* Nash inequalities.

For us, a *Lyapunov function* V is simply a positive function V on E such that $LV \leq cV$ for some constant c. We shall require those functions V to be in $\mathcal{L}^2(\mu)$ and in the domain to get interesting results, but it is not formally necessary.

Being a Lyapunov is not a very restrictive requirement for smooth functions in the examples below, as long as we do not ask $c < 0$ (in which case it cannot be true for any function V in the domain).

The *weighted Nash inequality* takes then the form

$$\Phi\left(\frac{\|f\|_2^2}{\|fV\|_1^2}\right) \leq \frac{\mathcal{E}(f,f)}{\|fV\|_1^2} \tag{2.10}$$

for all functions f in the domain of the Dirichlet form such that $\|f\|_2^2 > M\,\|fV\|_1^2$, where the rate function Φ is defined on (M,∞) and such that $\Phi(x)/x$ is increasing.

Theorem 2.4 (Wang)**.** *Assume that a weighted Nash inequality* (2.10) *holds with a rate function* Φ *defined on some interval* (M,∞) *such that* $\int^\infty \frac{1}{\Phi(s)} ds < \infty$. *Then*

$$\|P_t f\|_2 \leq K(2t) e^{ct}\,\|fV\|_1$$

for all $t > 0$ *and all functions* $f \in \mathcal{L}^2(\mu)$, *where* K *is defined as in Theorem 2.3. In particular, the kernel density* $p_t(x,y)$ *satisfies*

$$p_t(x,y) \leq K(t)^2 e^{ct} V(x) V(y).$$

Conversely, if there exists a positive function K *defined on* $(0,\infty)$ *such that*

$$\|P_t f\|_2 \leq K(t) \|fV\|_1$$

for all $t > 0$, *then the weighted Nash inequality* (2.10) *holds with* $M = 0$ *and rate function*

$$\Phi(x) = \sup_{t>0} \frac{x}{2t} \log \frac{x}{K(t)^2}, \quad x \geq 0.$$

Proof — It is given in detail in [2]. It follows the proof of Theorem 2.1 by replacing the function $K(t) = \dfrac{\|P_t f\|_2^2}{\|f\|_1^2}$ by $\hat{K}(t) = \dfrac{\|P_t f\|_2^2}{\|Vf\|_1^2}$. Now, the quantity $\int P_t f V d\mu$ is no longer invariant in time. But by properties of the Lyapunov function we have

$$\partial_t \int_E V P_t f d\mu = \partial_t \int P_t V f d\mu = \int P_t L V f d\mu \leq c \int_E V P_t f d\mu, \tag{2.11}$$

from which we get

$$\int_E P_t f V d\mu \leq e^{ct} \int V f d\mu.$$

Using this, we get again a differential inequality on $\hat{K}$ when we apply the Nash inequality (2.10) to $P_t f$, and the $\mathcal{L}^1 \to \mathcal{L}^2$ boundedness result follows.

To get the (non uniform) bound on the kernel, it remains to observe that if a symmetric operator K satisfies $\|Kf\|_2 \leq \|fV\|_1$, the norms being considered with respect to a measure μ, then the operator K_1 defined by

$$K_1(f) = \frac{1}{V} K(fV)$$

is a contraction from $\mathcal{L}^1(\nu)$ into $\mathcal{L}^2(\nu)$, where $d\nu = V^2 d\mu$. Moreover, K_1 is symmetric in $\mathcal{L}^2(\nu)$, and therefore $K_1 \circ K_1$ is a contraction from $\mathcal{L}^1(\nu)$ into $\mathcal{L}^\infty(\nu)$. It follows that it has a density kernel bounded by 1 with respect to ν; and this amounts to say that K has a density kernel with respect to μ bounded above by $V(x)V(y)$, since the kernel of K_1 with respect to ν is $\dfrac{k(x,y)}{V(x)V(y)}$, where k is the kernel of K with respect to μ. ∎

Observe that Theorem 2.4 produces non uniform bounds on the kernel. Moreover, when $V \in \mathcal{L}^2(\mu)$, then the operator P_{2t} is Hilbert-Schmidt so has a discrete spectrum and we get an estimate on the eigenvalues $-\lambda_n$ of L :

$$\sum_n e^{-\lambda_n t} \leq K^2(t) e^{ct} \|V\|_2^2.$$

3 Weighted Nash inequalities on the real line.

As already mentioned, we shall mainly concentrate on model examples on the real line, and show elementary techniques to obtain weighted Nash inequalities for measures with density ρ with respect to the Lebesgue measure and the usual *carré du champ* $\Gamma(f,f) = |\nabla f|^2 = f'^2$. These techniques may be easily extended to the n-dimensional Euclidean space, and with some extra work to Riemannian manifolds.

Let us first state a universal weighted Nash inequality in the Euclidean space. We consider the case when $\Gamma(f,f) = |\nabla f|^2$ and $\mu(dx) = \rho(x)dx$. We are mainly interested in the case when μ is a probability measure. Recall that in this situation, there may not exist any classical Nash inequality (classical means with a power function as rate function Φ) unless the measure is compactly supported.

Here, the symmetric operator associated with the corresponding Dirichlet form is

$$Lf = \Delta f + \nabla \log \rho \cdot \nabla f.$$

We may always choose $V = \rho^{-1/2}$: it is not hard to check that $LV \leq cV$ for some constant c to get the universal weighted Nash inequality (with respect to μ)

$$||f||_2^{2+\frac{4}{n}} \leq C_n^{\frac{4}{n}} \, ||f\,V||_1^{\frac{4}{n}} \left(\mathcal{E}(f,f) + c \int_{\mathbb{R}^n} f^2 \, d\mu \right).$$

Here C_n is the constant for the Nash inequality in the Euclidean space with the Lebesgue measure.

To see this, we just apply the Euclidean Nash inequality (0.1) to $g = f\sqrt{\rho}$, where f is a smooth compactly supported function, and observe that

$$\int_{\mathbb{R}^n} |\nabla g|^2 \, dx = \int |\nabla f|^2 \rho dx + \int_{\mathbb{R}^n} \frac{LV}{V} f^2 \, d\mu = \mathcal{E}(f,f) + \int_{\mathbb{R}^n} \frac{LV}{V} f^2 \, d\mu,$$

through integration by parts. Unfortunately, this bound is not very useful since $V \notin \mathcal{L}^2(\mu)$. Nevertheless, with some care to justify the integration by parts in (2.11), (with extra hypotheses like uniform upper bounds on the Hessian of $\log \rho$), it may lead to an upper-bound on the kernel density.

Of course, this method has nothing particular to do with the Euclidean case. It extends a Nash inequality (without weight) with respect to a measure μ to a weighted Nash inequality with respect to the measure $\rho d\mu$ with weight $V = \rho^{-1/2}$, as soon as the inequality $LV \le cV$ is satisfied.

For example, one gets with this simple argument

Corollary 3.1. *In $\mathbb{R}^n$, with $\rho(x) = (1+|x|^2)^{-\beta}$ with $\beta > n$ or $\rho(x) = \exp(-(1+|x|^2)^{a/2})$ with $a > 0$, there exists a constant C such that for all $t > 0$ and $x, y \in \mathbb{R}^n$ the kernel density p_t satisfies*

$$p_t(x,y) \le \frac{C}{t^{n/2}} \, e^{Ct} \rho^{-1/2}(x) \, \rho^{-1/2}(y).$$

But since $V \notin \mathcal{L}^2(\mu)$, this may never produce any bound on the spectrum for example. So one has to look for more precise Lyapunov functions.

This is what we now perform on our model examples on the real line : we write $T(x) = \sqrt{1+x^2}$ and consider the measure

$$\mu_a(dx) = C_a \exp(-T(x)^a) dx,$$

where $a > 0$ and C_a is a normalizing constant. We denote by ρ_a the density $\exp(-T^a)$. Here, the associated operator is

$$L(f) = f'' - aT^{a-1}T' f'.$$

In this context, it is not hard to check that, for any $\beta \in \mathbb{R}$,

$$V = T^{-\beta}/\sqrt{\rho_a} \tag{3.12}$$

is a Lyapunov function. If $\beta > 1/2$, this function is in $\mathcal{L}^2(\mu_a)$. The issue is then to choose the smallest possible $V \in \mathcal{L}^2(\mu_a)$ and still have a weighted Nash inequality with rate function Φ such that $1/\Phi$ is integrable at infinity.

The main result on this example is the following

Theorem 3.2 ([2]). *If $a > 1$, then for any $\beta \in \mathbb{R}$ and V chosen as in* (3.12), *there exist constants C and $\lambda \in (0,1)$ such that*

$$\|f\|_2^2 \leq C\left[\left(\int |f|Vd\mu_a\right)^2 + \left(\int |f|Vd\mu_a\right)^{2(1-\lambda)} \mathcal{E}(f,f)^\lambda\right] \tag{3.13}$$

for all functions f. This corresponds to the rate function

$$\Phi(x) = \left(\frac{x}{C} - 1\right)^{1/\lambda},\ x > C.$$

Although tractable, the explicit value of λ in terms of the parameters a and β is not so simple. The assumption $a > 1$ is necessary, since for $a \leq 1$ the spectrum is no longer discrete (and therefore no weighted Nash inequality could occur with any $\mathcal{L}^2(\mu_a)$ weight V). What has to be underlined here is that the introduction of a weight allows us to get polynomial rate functions Φ, although we know that such polynomial growth is forbidden for non compactly supported finite measures in the absence of weights. Of course, to get these polynomial growths, one has to choose weights which are quite close to the universal weights $1/\sqrt{\rho}$ described before. If one chooses much smaller weights, the rate function will be smaller. For example, when $a > 2$, one may choose $V = 1$, and in this case one has $\Phi(x) = x(\log x)^\alpha$.

The argument of Theorem 3.2 is based on a tail estimate of the measure μ_a. If $q_a(x) = \int_x^\infty \mu_a(dy)$, then, for some constant C, one has

$$q_a(x) \leq C\frac{\rho_a(x)}{T(x)^{a-1}}. \tag{3.14}$$

One first proves a Nash inequality for smooth compactly supported functions such that $f(0) = 0$. We start with

Lemma 3.3. *Let $a \geq 1$, $\beta \in \mathbb{R}$ and V given in* (3.12). *For all smooth compactly supported functions f such that $f(0) = 0$ one has*

$$\int f^2 d\mu_a \leq C\mathcal{E}(f,f)^\gamma \left(\int |f|Vd\mu_a\right)^{2(1-\gamma)}$$

where $\gamma = 1 - 2\dfrac{a-1}{3(a-1)+2\beta} \in \left(\dfrac{1}{3}, 1\right]$.

The proof is based on cutting the integral on $[0,\infty)$ (for instance) as

$$\int_0^\infty f^2 d\mu_a = \int_0^\infty f^2 \mathbb{1}_{\left\{\frac{f}{\|f\|_2} \leq VZ^{-1/2}\right\}} d\mu_a + \int_0^\infty f^2 \mathbb{1}_{\left\{\frac{f}{\|f\|_2} > VZ^{-1/2}\right\}} d\mu_a.$$

for a suitably chosen $Z > 0$. Then both terms are controlled by the estimate (3.14), replacing f^2 by $2\int_0^x f(t)f'(t)dt$ in the second integral and using Fubini's theorem.

It remains to get rid of the assumption $f(0) = 0$. For this purpose, with the same kind of techniques one may prove the following

Lemma 3.4. *Let $a > 0$, $\beta > \frac{3-a}{2}$ and V given in* (3.12). *Then there exist $\theta \in (0,1)$ and C such that*

$$\int |f - f(0)| V d\mu_a \leq C \left[\int |f| V d\mu_a + \left(\int |f| V d\mu_a \right)^{1-\theta} \mathcal{E}(f,f)^{\theta/2} \right]$$

for all nonnegative smooth compactly supported f on $\mathbb{R}$.

Although quite similar, this lemma is more restrictive on the values of β than the previous one. Passing from functions which vanish in 0 to the general case is indeed the hard step. We refer the reader to [2] for details on the proofs. It remains to plug together those inequalities to obtain Theorem 3.2.

Corollary 3.5. *Let $a > 1$ and let $(P_t)_{t \geq 0}$ be the Markov generator on $\mathbb{R}$ with generator*

$$Lf = f'' - aT^{a-1}T'f',$$

and reversible measure $d\mu_a(x) = \rho_a(x)dx = C_a \exp(-(1+|x|^2)^{a/2})dx$.

Then for all real β there exist $\delta > 0$ and C such that, for all t, P_t has a density p_t with respect to the measure μ_a, which satisfies

$$p_t(x,y) \leq \frac{Ce^{Ct}}{t^\delta} \frac{\rho_a^{-1/2}(x)\rho_a^{-1/2}(y)}{(1+|x|^2)^{\beta/2}(1+|y|^2)^{\beta/2}}$$

for almost every $x, y \in \mathbb{R}$.

Moreover, the spectrum of $-L$ is discrete and its eigenvalues $(\lambda_n)_{n \in \mathbb{N}}$ satisfy the inequality

$$\sum_n e^{-\lambda_n t} \leq \frac{Ce^{Ct}}{t^\delta}$$

for all $t > 0$.

When $a > 2$, the same techniques also lead to a Nash inequality for μ_a with rate function $\Phi(x) = C\, x\, (\log x)^{2(1-1/a)}$, and weight $V = 1$. This recovers the ultracontractivity result mentioned earlier. Recall that when $a = 2$ the semigroup is no longer ultracontractive, but only hypercontractive, the Nash inequality with rate $\Phi(x) = x \log x$ corresponds in fact to another form of the Logarithmic Sobolev inequality.

References

[1] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Lectures on probability theory (Saint-Flour, 1992)*, Lecture Notes in Math. 1581, pages 1–114. Springer, Berlin, 1994.

[2] D. Bakry, F. Bolley, I. Gentil, and P. Maheux. Weighted Nash inequalities. *Preprint.*

[3] D. Bakry, T. Coulhon, M. Ledoux, and L. Saloff-Coste. Sobolev inequalities in disguise. *Indiana Univ. Math. J.*, 44(4):1033–1074, 1995.

[4] D. Bakry and M. Ledoux. Sobolev inequalities and Myers's diameter theorem for an abstract Markov generator. *Duke Math. J.*, 85(1):253–270, 1996.

[5] A. Bendikov, T. Coulhon, and L. Saloff-Coste. Ultracontractivity and embedding into L^∞. *Math. Ann.*, 337(4):817–853, 2007.

[6] E. A. Carlen and M. Loss. Sharp constant in Nash's inequality. *Internat. Math. Res. Notices*, 7:213–215, 1993.

[7] E. B. Davies. *Heat kernels and spectral theory*, volume 92 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1990.

[8] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[9] J. Nash. Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.*, 80:931–954, 1958.

[10] E. Nelson. The free Markoff field. *J. Functional Analysis*, 12:211–227, 1973.

[11] N. Th. Varopoulos. Hardy-Littlewood theory for semigroups. *J. Funct. Anal.*, 63(2):240–260, 1985.

[12] F.-Y. Wang. Functional inequalities for empty essential spectrum. *J. Funct. Anal.*, 170(1):219–245, 2000.

[13] F.-Y. Wang. Functional inequalities and spectrum estimates: the infinite measure case. *J. Funct. Anal.*, 194(2):288–310, 2002.

Institut de Mathématiques de Toulouse, UMR CNRS 5219
Université de Toulouse
Route de Narbonne
31062 Toulouse - France
bakry@math.univ-toulouse.fr

Ceremade, UMR CNRS 7534
Université Paris-Dauphine
Place du Maréchal De Lattre De Tassigny
75016 Paris - France
bolley@ceremade.dauphine.fr

Institut Camille Jordan, UMR CNRS 5208
Université Claude Bernard Lyon 1

43 boulevard du 11 novembre 1918
69622 Villeurbanne cedex - France
gentil@math.univ-lyon1.fr