ESAIM: PROCEEDINGS, January 2014, Vol. 44, p. 110-128
SMAI Groupe MAS

LONG TIME BEHAVIOR OF MARKOV PROCESSES

Patrick Cattiaux[1]

Abstract. These notes correspond to a three hours lecture given during the workshop "Metastability and Stochastic Processes" held in Marne-la-Vallée in September 21st-23rd 2011. I would like to warmly thank the organizers Tony Lelièvre and Arnaud Guillin for a very nice organization and for obliging me first to give the lecture, second to write these notes. I also want to acknowledge all the people who attended the lecture.

Résumé. Ces notes correspondent à un cours de trois heures donné durant le workshop "Metastability and Stochastic Processes" organisé à Marne-la-Vallée les 21-23 septembre 2011.

Keywords: Markov processes, convergence to equilibrium, Poincaré inequalities, Lyapunov functions, hitting times, mixing.

1. Advertisement and goals.

These notes will present a very partial overview of some techniques used to study the long time behavior of Markov processes (functional inequalities, Lyapunov functions ...). Others powerful tools (like coupling for example) are deliberately ignored mainly because of the incompetence of the author and also by lack of time or place. In addition, I will focus on simple ideas. The main consequence is the very imprecise formulation of some sentences (in general, generically ...) except for the few theorems which are explicitly stated. For these sentences to be perfectly correct, additional (but merely technical) assumptions have to be made, and so checked by the potential reader interested in using them. To worsen the situation, I have also chosen to select a very short list of references in comparison with the exponentially fast growing number of publications on this topic. I preferred to indicate in the remarks the names of some mathematicians whose contributions to the topic are important (sometimes crucial), for my own taste, letting the pleasure to the reader to discover them, for instance by giving a look at their homepage. I of course make my apologies in advance to all others contributors.

So, everybody understood that these notes have to be considered as a (potentially) pedagogical introduction to a fascinating domain at the confluence of Probability theory, Functional Analysis, P.D.E. theory, Mathematical modeling and also Geometry, even if this latest point will not be discussed here. This domain has known an amazing growing interest and continuous development during the last ten years. Some recent surveys quoted in the references and forthcoming textbooks will give a more complete and rigorous picture.

[1] Institut de Mathématiques de Toulouse. Université de Toulouse. CNRS UMR 5219.

2. Introduction.

Let $\{(X_t)_{t\in\mathbb{T}}, (\mathbb{P}_x)_{x\in E}\}$ be a (time homogeneous) strong Markov process with state space E, the time variable t belonging to $\mathbb{T}$ which is either $\mathbb{N}$ (discrete time) or $\mathbb{R}^+$ (continuous time). We introduce the associated semi-group P_t defined for bounded measurable functions f by $P_t f(x) = \mathbb{E}_x(f(X_t))$. The semi-group property is the following $P_{t+s} = P_t\, P_s$ which is a consequence of the Markov property. In the discrete time case we shall denote $P = P_1$ and $P^n = P_n$.

We shall assume that there exists a unique invariant probability measure, i.e. such that $\mathbb{E}_\mu(f(X_t)) = \int f\, d\mu$ for all t, which in addition is ergodic, i.e. such that the asymptotic (or invariant) σ-field is trivial. We shall say that μ is reversible or symmetric if

$$E_\mu(f(X_0)\, g(X_t)) = \int f\, P_t g\, d\mu = \int g\, P_t f\, d\mu = E_\mu(g(X_0)\, f(X_t))\,.$$

The semi-group easily extends to any $\mathbb{L}^p(\mu)$ for $1 \leq p \leq \infty$ as a contraction semi-group (i.e. each P_t has an operator norm equal to 1). The adjoint semi-group will be denoted by P_t^*. Probabilistically it corresponds to some time reversal. At least at a formal level, one can consider that P_t^* acts on (probability) measures, via the formula $< f, P_t^*\nu >= \int P_t f\, d\nu$ written for bounded f's. $P_t^*\nu$ is thus the law of the process at time t, when the initial distribution is ν. If $\nu = \delta_y$ we also denote by $P_t(y, .)$ this distribution.

Finally we introduce the (infinitesimal) generator of the semi-group denoted by L which is defined by $L = P - Id$ in the discrete time case, and $Lf = \lim_{t\to 0} \frac{1}{t}\,(P_t f - f)$ in the continuous time case, when this limit makes sense. The functions for which the limit exists are the elements of the domain $D(L)$ of L, this domain depends on the chosen topology (in general some $\mathbb{L}^p(\mu)$) and for $t > 0$, $P_t f \in D(L)$. We will always assume that there exists a nice core $\mathcal{C}$ in the domain. The following then holds, provided it makes sense

$$\frac{d}{ds}\, P_s f|_{s=t} = L P_t f = P_t L f\,.$$

For any probability measure ν we define the energy $\mathcal{E}_\nu(f, g) = \int (-Lf) g\, d\nu$.

Along these notes we will mainly refer to three types of generic examples (we omit regularity assumptions for the described properties to hold true):

(1) Discrete time Markov chains on finite or enumerable state space E.

(2) Non degenerate diffusion processes on $\mathbb{R}^d$ given by the solution of a s.d.e.

$$dX_t = \sqrt{2}\, dB_t - b(X_t) dt$$

where $B_.$ is a standard brownian motion. Here $L = \Delta - b\nabla$ and we assume that $d\mu = e^{-V(x)} dx$ is an invariant probability measure. μ is reversible provided $b = \nabla V$. In this case $\mathcal{E}_\mu(f, g) = \int \nabla f.\nabla g\, d\mu$. Remark that $\mathcal{E}_\mu(f, f)$ vanishes if and only if f is constant (this corresponds to ergodicity in the symmetric case).

(3) Fully degenerate diffusion processes related to kinetic (or Hamiltonian) systems, i.e. $X_t = (x_t, v_t)$ where

$$\begin{aligned} dx_t &= v_t\, dt \\ dv_t &= \sqrt{2}\, dB_t - b(x_t, v_t) dt \end{aligned}$$

where $L = \Delta_v - b\, \nabla_v + v\, \nabla_x$. L is no more elliptic but still hypoelliptic in many cases.

When $b(x_t, v_t) = v_t + \nabla F(x_t)$ the measure $\mu(dx, dv) = e^{-v^2 - F(x)} dx dv$ (assumed to be bounded) is invariant, ergodic but not reversible. In this situation $\mathcal{E}_\mu(f, g) = \int \nabla_v f.\nabla_v g\, d\mu$ so that $\mathcal{E}_\mu(f, f)$ vanishes when f only depends on (the position) x. This situation is called a hypoelliptic fully degenerate situation.

3. Ergodic theory.

According to the ergodic theorem for μ almost all x and $f \in \mathbb{L}^1(\mu)$,

$$\lim_{t\to+\infty} \frac{1}{t}\int_0^t f(X_s)\,ds \quad \left(\text{resp.} \quad \lim_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(X_j)\right) = \int f\,d\mu$$

$\mathbb{P}_x$ almost surely and in $\mathbb{L}^1(\mathbb{P}_x)$. It follows that

$$\lim_{t\to+\infty} \frac{1}{t}\int_0^t P_s f\,ds \quad \left(\text{resp.} \quad \lim_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} P_j f\right) = \int f\,d\mu$$

μ almost surely and in $\mathbb{L}^1(\mu)$. This result extends to $\mathbb{L}^2(\mu)$ (with a convergence in $\mathbb{L}^2(\mu)$ of course).

In some situations one can skip the Cesaro mean in the previous result. If it is the case, one sees that for all probability measure ν,

$$P_t^*\nu \to \mu$$

as t goes to infinity, for the weak convergence of (probability) measures. Let us describe some of these situations.

3.1. **Finite Markov chains.**

Assume that E is some finite set. $X_.$ is then a discrete time Markov chain, assumed to be irreducible, hence recurrent. In this case there exists a unique invariant probability measure μ (positive recurrence). In addition $\mu(x) > 0$ for all $x \in E$. In particular any probability measure ν on E is absolutely continuous w.r.t. μ and $\frac{d\nu}{d\mu} \in \mathbb{L}^\infty(\mu)$. The ergodic theorem holds, but if in addition the chain is aperiodic, one has the following quantitative estimate, for all ν,

$$\| P_n^*\nu - \mu \|_{TV} = \| P_n^*\,(d\nu/d\mu) - 1 \|_{\mathbb{L}^1(\mu)} \leq c\,\alpha^n \quad \text{for some } \alpha < 1. \tag{3.1}$$

Here, $\|\|_{TV}$ denotes the total variation distance. The proof of (3.1) is based on compactness arguments since $\mathbb{L}^1(\mu)$ can be identified with $\mathbb{R}^{\sharp E}$, and the set of probability densities is thus the (compact) boundary of the unit simplex in the ordinary euclidean space.

If μ is reversible, the situation reduces to the spectral theory of P considered as a symmetric matrix in the euclidean space $\mathbb{R}^{\sharp E}$ equipped with the scalar product $< u, v >= \sum_{x\in E} \mu(x)\,u_x\,v_x$. Symmetry implies that P is diagonalizable in an orthonormal basis, contraction implies that all eigenvalues belong to $[-1, 1]$, ergodicity implies that 1 is an eigenvalue of order 1, aperiodicity implies that -1 is not an eigenvalue. It immediately follows that

$$\mathrm{Var}_\mu(P^n f) \leq \alpha^{2n}\,\mathrm{Var}_\mu(f) \tag{3.2}$$

where α is the maximum of the absolute value of all the eigenvalues different from 1. This α coincides with the α in (3.1), and c can be chosen as $c = 1/(2\,\min_x \mu(x))$.

If compactness arguments do not extend to more general situations, the previous ideas in the symmetric case can be extended to more general $\mathbb{L}^2$ situations since the spectral theory of self adjoint (symmetric and closed) operators in Hilbert spaces naturally extends the one in finite dimension.

Remark 3.3. Long time behavior for finite Markov chains has been studied for a long time and is by itself a very rich theory. The arguments above are very elementary and can be (mainly) found in [7]. Much more refined results including mixing times, coupling, exact estimates etc ... have been obtained by many authors. The main author associated to this theory is obviously Persi Diaconis. Others particularly important contributors are

Laurent Saloff-Coste (who wrote in particular several surveys), Laurent Miclo, James A. Fill, and so many others. A wonderful recent (rather speaking not too old) survey on the topic is due to Ravi Montenegro and Prasad Tetali ([44]). ♢

3.2. **The symmetric $\mathbb{L}^2$ case.**

For a general nice state space, we assume that μ is reversible, and discuss first the continuous time case. The semi-group $P_.$ is not only symmetric but actually self adjoint in $\mathbb{L}^2(\mu)$. In addition remark that L is a negative operator, i.e. $\int (Lf)f\,d\mu \leq 0$ for all f.

It follows that P_t (formally equal to e^{tL}) has the following spectral representation

$$P_t f = \int_0^\infty e^{-\lambda t}\, d < E_\lambda, f > \ . \tag{3.4}$$

The meaning of such a formula is quite delicate, but can be found in superb textbooks on Functional Analysis like the one of K. Yosida. Actually, E_λ is the spectral projection (the all family is called a spectral resolution) similar to the projection onto the spectral subspaces in finite dimension. Since $-L$ is non-negative, its spectrum is included into the set of non-negative real numbers (explaining the bounds for the integral). In particular $d < E_0, f >= \int f\,d\mu\,\delta_0$, i.e. E_0 is the projection onto the eigenspace of L associated to the eigenvalue 0, which is the set of constant functions since the process is ergodic.

It follows that for μ zero mean functions f, $d < E_0, f >= 0$. Applying the bounded convergence theorem of Lebesgue, we thus immediately see that

$$P_t f \text{ goes to } 0 \text{ in } \mathbb{L}^2(\mu) \text{ as } t \to +\infty. \tag{3.5}$$

The same argument works in the discrete time setting, just replacing the previous representation by the appropriate one (using the fact that the spectrum of P is included in $[-1, 1]$ and a polynomial representation).

But, as for finite Markov chains, one can obtain a quantitative estimate for the convergence, if $E_\lambda = 0$ for all $0 < \lambda < \lambda_0$. In this case we obtain

$$\mathrm{Var}_\mu(P_t f) \leq e^{-2\,\lambda_0\, t}\, \mathrm{Var}_\mu(f)\,. \tag{3.6}$$

λ_0 is called the spectral gap.

Remark 3.7. For people who do not like spectral resolution, an elementary (but not immediate) proof of (3.5) is contained in [12]. ♢

3.3. **Additional remarks.**

In this subsection, time is either discrete or continuous, despite the notation. Actually, from the analytical point of view the semi-group to look at is not P_t but Q_t defined by $Q_t f = P_t f - \int f\,d\mu$ (this point of view is better than considering the restriction of P_t to the closed subspace of zero mean functions). In particular $\mathrm{Var}_\mu(P_t f) = \parallel Q_t f \parallel^2_{\mathbb{L}^2(\mu)}$.

Q_t is thus a bounded operator from $\mathbb{L}^p(\mu)$ to $\mathbb{L}^q(\mu)$ for $1 \leq q \leq p \leq +\infty$, with operator norm denoted by $|||\ |||_{p,q}$. Note that $|||\ Q_t\ |||_{1,1} \leq 2$ and $|||\ Q_t\ |||_{\infty,\infty} \leq 2$.

Of course due to the semi-group property

$$|||\ Q_{t+s}\ |||_{p,p} \leq |||\ Q_t\ |||_{p,p}\ |||\ Q_s\ |||_{p,p}$$

so that if $||| Q_{t_0} |||_{p,p} \leq e^{-\beta}$ for some $t_0 > 0$ and some $\beta > 0$, $||| Q_t |||_{p,p} \leq C\, e^{-\beta\, t}$ for all $t \geq 0$, for some constant $C \geq 1$. In other words, a uniform decay in $\mathbb{L}^p$ implies that this decay is exponential (or geometric).

In addition, according to Riesz-Thorin interpolation theorem and the bounds in $\mathbb{L}^1$ and $\mathbb{L}^\infty$, an exponential decay in some $\mathbb{L}^p$ space for $1 < p < +\infty$ implies exponential decay in all $\mathbb{L}^p$ spaces. For instance if

$$||| Q_t |||_{2,2} \leq C\, e^{-\beta\, t} \,,$$

it holds

$$||| Q_t |||_{p,p} \leq 2^{\frac{p-2}{p}}\, C^{2/p}\, e^{-(2/p)\, \beta\, t} \ \text{ for } p \geq 2$$

and

$$||| Q_t |||_{p,p} \leq 2^{\frac{2-p}{p}}\, C^{2(p-1)/p}\, e^{-(2(p-1)/p)\, \beta\, t} \ \text{ for } 1 \leq p \leq 2 \,.$$

The next natural question is to know what happens in case of non exponential decay. We shall see that for $||| Q_t |||_{\infty,2}$ for example (hence for many similar norms by interpolation), one can find a wide variety of decay bounds (actually any slower decay than exponential is possible). However, in the symmetric case (more generally if P_t is normal, i.e. commutes with P_t^*) exponential decay for $||| Q_t |||_{\infty,2}$ implies exponential decay for $||| Q_t |||_{2,2}$ thanks to the following nice lemma

Lemma 3.8. *Assume that μ is reversible. Assume that there exists $\lambda > 0$ such that for all f belonging to some dense subset $\mathcal{C}$ of $\mathbb{L}^2(\mu)$ there exists some c_f with*

$$Var_\mu(P_t f) \leq c_f\, e^{-2\lambda\, t} \,,$$

then for all square integrable f,

$$Var_\mu(P_t f) \leq \, e^{-2\lambda\, t}\, Var_\mu(f) \,.$$

This result seems to belong to the “folklore” of the theory. A simple (and maybe elegant) proof, is based on some convexity property for $t \mapsto \log(\mathrm{Var}_\mu(P_t f))$, available in the symmetric case. In the continuous time case see [23], for the discrete time case this proof is adapted in [24].

Another interesting fact is that the previous bound is available for all t and exact for $t = 0$ (to be more explicit, there is no constant in front of the exponential). It turns out that, in the non symmetric case, one can find examples (for instance the kinetic models described before, see [47]), for which exponential decay holds, but with an extra constant in front of the exponential. We will understand why this extra constant is necessary in the next section.

Finally one may ask whether this result extends to any $p \neq 2$. Even in the symmetric case this is not obvious (recall that interpolation has to be done for Q_t, so that some 2 appears in front of the norm). This is discussed in details in [21].

4. Exponential decay: from Poincaré to Foster-Lyapunov.

As we have seen exponential decay in $\mathbb{L}^2$ is characterized, in the symmetric case, by the existence of a spectral gap. More generally, the (exact) exponential decay is characterized by a functional inequality: Poincaré inequality. The statements for discrete time an continuous time slightly differ.

Proposition 4.1. *In the discrete time setting, the following properties are equivalent*

(1) $\forall f \in \mathbb{L}^2(\mu)$, $Var_\mu(Pf) \leq e^{-2\lambda}\, Var_\mu(f)$,

(2) $\forall f \in \mathbb{L}^2(\mu)$ *and all* n, $Var_\mu(P^n f) \leq e^{-2\lambda\, n}\, Var_\mu(f)$,

(3) *there exists* C_P *such that* $\forall f \in \mathbb{L}^2(\mu)$,

$$Var_\mu(f) \leq C_P \, \langle (Id - P^* P) f, f \rangle := C_P \int (f - P^* P f)\, f\, d\mu \,.$$

The inequality in (3) is called a Poincaré inequality, and if it holds, $e^{-2\lambda} = \frac{C_P - 1}{C_P}$.

The proof of this proposition is immediate since for a centered function f, $\langle (Id - P^*P)f, f\rangle = \mathrm{Var}_\mu(f) - \mathrm{Var}_\mu(Pf)$.

Remark that in the symmetric case, the Poincaré inequality above reads

$$\mathrm{Var}_\mu(f) \leq C_P \,\langle (Id - P^2)f, f\rangle$$

so involves the generator of the chain with transition kernel P^2, and not P. Actually, still in the symmetric case, this inequality can be rewritten as $\| \; P \; \|^2 \leq \; e^{-2\lambda}$ where P is considered as an operator on the closed hyperplane of functions with zero mean. But it is well known, thanks to symmetry, that $\| \; P \; \| = \sup\langle Pf, f\rangle$ where f describes the set of functions with variance equal to 1. So changing the constant, we may replace P^2 by P in the previous theorem.

Proposition 4.2. *In the continuous time setting, the following properties are equivalent*

(1) $\forall f \in \mathbb{L}^2(\mu)$ *and all* t, $Var_\mu(P_t f) \leq e^{-2\lambda\, t}\; Var_\mu(f)$,

(2) there exists C *such that* $\forall f \in \mathbb{L}^2(\mu)$,

$$Var_\mu(f) \leq C_P \,\langle -Lf, f\rangle = C_P \,\mathcal{E}_\mu(f,f)\,.$$

The inequality in (3) is called a Poincaré inequality, and if it holds, $\lambda = 1/C_P$.

The proof is almost immediate just differentiating the first inequality with respect to the time at time $t = 0$. When (1) holds, we shall say that the semi-group is a $\mathbb{L}^2$ *contraction.*

Remark that, in the continuous time case (symmetric or not), if a Poincaré inequality holds, any f such that $\mathcal{E}_\mu(f,f) = 0$ is necessarily constant. In particular for the kinetic models of type (3) in the introduction, the Poincaré inequality does not hold. Hence, even if there is some exponential decay, the semi-group cannot be a contraction.

The reader which is familiar with functional inequalities has to be careful here since the Poincaré inequality does not hold with the energy form we are interested in, but may hold for these models for the usual energy given by the square of the full gradient in velocity and space.

Functional inequalities are at the center of the analysis of p.d.e.'s. In this context they are generally written with the Lebesgue measure. What we called a Poincaré inequality in the previous propositions is called a weighted Poincaré-Sobolev inequality in p.d.e. theory. Terminology becomes dangerous, because weighted inequalities also exist in the area of functional inequalities, with a somewhat different meaning.

Of course Poincaré inequalities are not only useful for studying long time behavior. We shall give below some examples and some properties of such inequalities. The few properties we will recall are the most useful for our purpose.

Example 4.3. (1) Let $\mu(dx) = e^{-V(x)}\,dx$ be defined on $\mathbb{R}$. Then C_P is of order

$$\max\left\{\sup_{x\geq m} \int_m^x e^V\,dx \int_x^{+\infty} e^{-V}\,dx \;,\; \sup_{x\leq m} \int_x^m e^V\,dx \int_{-\infty}^x e^{-V}\,dx\right\}$$

where m is a median of μ. This is a consequence of Hardy's inequality (i.e. integration by parts).

(2) If $\mu(dx) = (1/Z)e^{-\gamma\,|x|}\,dx$ in $\mathbb{R}^d$, μ satisfies a Poincaré inequality with C_P only depending on the covariance matrix of μ, hence on the dimension d.

(3) If $\mu(dx) = (1/Z)e^{-\gamma\,|x|^2}\,dx$ in $\mathbb{R}^d$, μ satisfies a Poincaré inequality with C_P only depending on γ, hence independent on the dimension d.

(4) More generally, if $\mu(dx) = e^{-V(x)}\,dx$ is defined on $\mathbb{R}^d$ and if V is convex, then μ satisfies a Poincaré inequality. This result is due to Serguei Bobkov ([8]), one of the most impressive contributor to the

theory. It applies in particular to the Lebesgue measure on a convex, compact subset (for a compact interval the calculation of C_P is easily done by using Fourier series expansion).

An open question (the Kannan-Lovasz-Simonovits conjecture) is to show that the constant C_P is then dimension free if the covariance matrix is the identity.

Proposition 4.4. *(1) If μ satisfies a Poincaré inequality, μ concentrates at least as an exponential distribution. A weak formulation of this sentence is that the tails $\mu(|x| > t) \leq C\, e^{-\gamma\, t}$ for some $\gamma > 0$ (in more general metric spaces one can replace the norm by the distance to some chosen point).*

The best reference for the concentration of measure phenomenon is still the wonderful book by Michel Ledoux [39].

(2) Tensorization property, i.e. $C_P(\mu \otimes \nu) = \max(C_P(\mu), C_P(\nu))$ (of course one has to consider the generator $L_{\mu\otimes\nu} = L_\mu \otimes L_\nu$).

This property explains the gaussian example (3). It is very useful when trying to understand the behavior of Gibbs measures on infinite dimensional lattices (actually one has to use more stringent inequalities, see the papers by Boguslaw Zegarlinski and Pierre André Zitt)).

(3) Perturbation. In the situation of diffusion processes as in (2) in the introduction (i.e. the energy is given by the integral of the square of the gradient), for $\mu(dx) = e^{-V(x)}\, dx$ and $\nu(dx) = (1/Z)e^{-(V+H)(x)}\, dx$, it holds $C_P(\nu) \leq e^{OscH}\, C_P(\mu)$ where $OscH$ denotes the oscillation of H. This simple (but very useful) result is due to Richard Holley and Daniel W. Stroock. It can be generalized to more general situations (but with some care with the energy form), to Lipschitz perturbations H (Laurent Miclo) and to more general perturbations ([30], also see [13]).

Remark 4.5. Of course there is no difference between P_t and P_t^* from the point of view of $\mathbb{L}^2$ convergence. But as we said one can consider that P_t^* acts on measures. It turns out that in many cases, $P_t^*\delta_x$ is absolutely continuous with respect to μ for $t > 0$. This is typically the case in hypoelliptic situations. In the continuous time case, if in addition μ is reversible or if one prefer if the dynamics is μ reversible, the density $dP_t(x,.)/d\mu$ belongs to $\mathbb{L}^2(\mu)$ thanks to the Chapman-Kolmogorov relation (see [23] for instance).

But of course, there are also many cases (in particular in infinite dimension) where this absolute continuity does not hold. A direct treatment of infinite dimensional cases is thus delicate, we shall come later to this point. $\diamondsuit$

Remark 4.6. As shown in the previous propositions the Poincaré constant furnishes the best "contraction" decay in $\mathbb{L}^2$. According to lemma 3.8 it is also the optimal exponential decay. The symmetric case is thus the worst case. In many non symmetric situations one can improve the convergence rate in $Ce^{-\theta t}$ for some $\theta > 2\lambda$, but of course some $C > 1$. This phenomenon is known for a long time, even in the linear diffusion case (see [35]). $\diamondsuit$

4.1. **The positive spectrum.**

It is in general difficult to obtain explicit eigenfunctions associated to the spectral gap (if it is an eigenvalue). In addition, for $\lambda > 0$ any (nice) function satisfying $Lf = -\lambda f$ has μ zero mean, hence cannot be non-negative.

However, Doob's potential theory tells us that some explicit functions are almost eigenfunctions. Indeed, take some non-polar subset U of E (non-polar means that the hitting time T_U of U is finite for all (or quasi all) x). Define $W(x) = \mathbb{E}_x(e^{\lambda\, T_U})$ and assume that for some $\lambda > 0$, W is everywhere finite. Then, generically, $LW = -\lambda\, W$ on U^c. As we said in the advertisement, generically means that this statement has to be rigorously checked (in particular in which spaces does the equality hold ?) in the encountered situations.

For a recurrent Markov chain with an enumerable state space E, this statement is true and used with $U = \{y\}$. The following theorem completely describes the situation :

Theorem 4.7. *Assume that E is enumerable and that the discrete time Markov chain $X_.$ is positive recurrent and aperiodic, with unique invariant probability measure μ. Then the following statements are equivalent*

(1) *there exist $a \in E$ and $\lambda > 0$ such that for all $x \in E$, $\mathbb{E}_x\left(e^{\lambda\, T_a}\right) < +\infty$,*

(2) *there exists $0 < \theta < 1$ such that for all $x \in E$ one can find $C(x)$ with*

$$\| P^n(x,.) - \mu(.) \|_{TV} \leq C(x)\, \theta^n \,,$$

(3) *there exists a Foster-Lyapunov function, i.e. a function $W : E \to \mathbb{R}$, such that $W \geq 1$, $LW \leq -\alpha W + b\mathbb{1}_a$ for some $0 < \alpha$ and some $b \geq 0$.*

In addition if μ is symmetric, these statements are equivalent to the following additional one

(4) *there exists a constant C_P such that the Poincaré inequality*

$$Var_\mu(f) \;\leq\; C_P\, \langle (Id - P^2)f \,,\, f\rangle$$

holds for all $f \in \mathbb{L}^2(\mu)$.

The equivalence between (1) and (3) is an exercise, of course the biggest Lyapunov function is the exponential moment of the hitting time. That (3) implies (2) can be nicely shown as remarked by Martin Hairer and Jonathan C. Mattingly ([32]) even in a stronger form. The converse (2) implies (1) is more intricate, and usual proofs call upon Kendall's renewal theorem and an argument of analytic continuation. The reference certainly is the monograph by Sean Meyn and Richard Tweedie [42].

(4) implies of course (2). For the converse symmetry is required. It seems that the result is due to Mu-Fa Chen ([26] p. 221-235). A much simpler proof can be done by using lemma 3.8.

The extension to general Markov processes is not so easy, mainly because the renewal theory is much more difficult to handle with. It was at least partly done by several authors by using regeneration times, see e.g. [29] for the diffusion case. But the link with Poincaré inequality is not studied in all this literature (for short the MCMC world). The relationship between Lyapunov type conditions and functional inequalities was only recently studied (starting with [3]) and is now rather well understood.

We will try below to give a general picture, indicating the delicate points to check.

Definition 4.8. *$U \subset E$ is called a petite set if there exists a non-negative measure α on $\mathbb{R}^+$ and a non trivial non-negative measure ν on E such that*

$$\inf_{x\in U} \int_0^{+\infty} P_t(x,.)\, \alpha(dt) \;\geq\; \nu \,.$$

We shall say that W is a Foster-Lyapunov function, if $W \geq 1$ and there exist $\lambda > 0$, a closed petite set U and $b \geq 0$ such that

$$LW \leq -\,\lambda W + b\, \mathbb{1}_U \,.$$

The meaning of this inequality is that W belongs to the extended domain of the generator and the inequality is satisfied pointwise.

Up to the end of this section we assume that E is locally compact and that $X_.$ is positive Harris recurrent with unique invariant probability measure μ. Notice that the last property is automatically satisfied if there exists a Foster-Lyapunov function. Indeed recall

Remark 4.9. If there exists W as before such that $LW \leq cW$ the process is non explosive. If $LW \leq c\mathbb{1}_U$ it is recurrent, if $LW \leq -c + b\mathbb{1}_U$ for some $c > 0$ it is positive Harris recurrent. This classification can be found in [33] (also see [27]), the first statement is sometimes called Hasminski's non explosion test. ♢

Now let us state the four statements similar to the ones in theorem 4.7

(Lyap) There exists a Foster-Lyapunov function W.

(expTV) There exists $0 < \beta$ such that for all $x \in E$ one can find $C(x)$ with

$$\parallel P_t(x,.) - \mu(.) \parallel_{TV} \leq C(x)\, e^{-\beta\, t}\,.$$

(expHit) There exists a closed petite set U and $\theta > 0$ such that for all x,

$$W_\theta(x) = \mathbb{E}_x(e^{\theta\, T_U}) < +\infty\,.$$

(Poinc) The Poincaré inequality $\mathrm{Var}_\mu(f) \leq C_P\, \mathcal{E}_\mu(f,f)$ holds for some $C_P < +\infty$.

Theorem 4.10. *Let $X_.$ (continuous time) be positive Harris recurrent with unique invariant probability measure μ on the locally compact state space E. Then*

(1) (Lyap) and (expHit) are generically equivalent.
(2) (Lyap) implies (expTV) with $C(x) = CW(x)$ and β depending on U, λ and b.
(3) (Poinc) implies that W_θ is finite μ almost surely. With some extra regularity assumptions, (Poinc) will thus imply (expHit), hence (Lyap) and (expTV).
(4) If in addition μ is symmetric then (expTV) implies (Poinc).

The first point comes from the general theory of Markov processes. Some regularity of the process is necessary for W_θ to belong to the extended domain, that is why the result is generic.

The second one is mainly due to [29].

The third one is shown in [23] using results on additive functionals in [18] (themselves based on beautiful results by Liming Wu, see the reference therein).

The final one uses lemma 3.8.

The main disadvantage of this theorem is that there is no explicit expression for β in the second statement. That is why it is particularly interesting to get a direct proof of (Lyap) implies (Poinc) in the symmetric case. This is done in [2] (another estimate is derived in [23]), when $E = \mathbb{R}^d$, $d\mu = e^{-V}dx$, U is convex and relatively compact and the process possesses a carré du champ (i.e. a core which is an algebra). This is the case for non-degenerate diffusion processes of type (2) in the introduction.

The key ingredient is the following elementary inequality, which holds for any smooth an non negative W:

$$\int \frac{-LW}{W}\, f^2\, d\mu \;\leq\; \mathcal{E}_\mu(f,f)\,, \tag{4.11}$$

that holds for all f's (regular enough) without any assumption on the mean.

Another interesting feature is that one can obtain explicit bounds for the exponential moments of hitting time :

if the Poincaré inequality holds with constant C_P then for all open set U with $\mu(U) \leq \frac{1}{2}$, $\mathbb{E}_\mu(e^{\theta\, T_U}) < +\infty$ for $\theta < \mu(U)/8C_P(1-\mu(U))$.

If $\mu(U) \geq 1/2$ we may take $\theta < \mu^2(U)/2C_P$.

With extra regularity assumptions, we may replace μ by δ_x for all x in the expectation.

Example 4.12. How can we find Foster-Lyapunov functions ? This is the most important question. The answer is that we have to be a little bit lucky. If $d\mu = e^{-V}dx$, one usually tries to find W either in the form $W = \varphi(V)$ (and often $W = e^{\theta\, V}$ for θ small enough), or in the form $W(x) = \varphi(|x|)$.

One can then recover the exponential or gaussian cases discussed before. More interesting is to look at cases which are new or generalize known results. Here are three examples in this direction, all written for the diffusion case (2) in the introduction (hence the energy is the integral w.r.t. μ of the square of the usual gradient)

(1) Drift condition in the symmetric case : $< x, \nabla V(x) > \geq \alpha\, |x| - c$ for all x and some $\alpha > 0$. In particular this condition contains the case of convex V's (the log-concave situation) and we thus recover Bobkov's result, but with a worse constant.

The terminology “drift condition” is sometimes used in place of Foster-Lyapunov condition. It seems that is mainly well adapted in the diffusion case. This kind of conditions was independently used by Alexander Veretennikov and his coauthors to study the long time behavior of diffusion processes, through Sobolev estimates. It can also be used in the non-symmetric situation.

(2) If $\liminf_\infty(\beta|\nabla V|^2 - \Delta V) > 0$, for some $\beta < 1$, then there exists some Foster-Lyapunov function. This extends some previous results by Shigeo Kusuoka and Daniel W. Stroock (see [13]) and similar results obtained by the ground state transformation (that transforms the Fokker-Planck operator into a Schrödinger one, whose spectral theory is better known) yielding the so called Witten Laplacian of Bernard Helffer and Francis Nier [34].

(3) In one dimension one can replace the bounded perturbation by a super-linear perturbation (see [23]).

Example 4.13. Let us look now at kinetic models of type (3) in the introduction. As we already explained, there is no exact exponential decay, since the Poincaré inequality does not hold true. In order to show exponential convergence (called stabilization to equilibrium in p.d.e. theory) Villani uses two tools: first the regularization effect in small time due to hypo-ellipticity, second a modification of the functional space to look at (what he is calling hypo-coercivity). As a byproduct, he obtains exponential decay when μ satisfies the usual (with the usual energy) Poincaré inequality plus some technical assumptions on F (recall the model in the introduction). In [3] we obtain a slightly weaker result (the exponential decay for $|||\ .\ |||_{q,2}$ for all $q > 2$) by using an ad-hoc Foster-Lyapunov function (see also the seminal paper [50]).

The better understanding of these models is one of the most challenging problem in this area. We refer to the papers by Martin Hairer and Luc Rey-Bellet among some others. Particularly interesting are the heat baths models.

Example 4.14. The study of quasi-stationary distributions is actually strongly related to the previous theorems. Indeed, consider our positive recurrent Markov process, conditioned to not hit some subset U (considered as an extinction domain), i.e. look at $\mathbb{P}_\nu(X_t \in A | t < T_U)$. A quasi limiting distribution is a possible limit of these distributions as t goes to infinity, a quasi-stationary distribution is some ν such that the previous distributions do not depend on t, hence are all equal to ν. A general fact is that for a quasi-stationary distribution $\mathbb{P}_\nu(T_U > t)$ behaves like some $e^{-\lambda\, t}$ as t goes to infinity, i.e. T_U has some ν exponential moment. Hence the existence of a quasi-stationary distribution is linked to the Poincaré inequality (at least in the symmetric case). Moreover it can be shown that quasi-stationary distributions are given by “eigenfunctions in the positive spectrum”, i.e. are still stronger linked with Foster-Lyapunov functions (if we allow these functions to vanish at the boundary of U). Finally uniqueness of a quasi-stationary distribution is linked to a stronger property of the initial semi-group, namely ultraboundedness. For all this we refer to [15, 25] and the final section of [23].

Remark 4.15. One has to be very careful with the meaning of the exponential decay in total variation distance. It does not mean that exponential convergence holds in $\mathbb{L}^1(\mu)$. Theorem 4.10 says that convergence holds for probability densities g such that gW is integrable. Unless the process is ultrabounded, W is not bounded, so that convergence does not hold in $\mathbb{L}^1$. ♢

5. Exponential decay in $\mathbb{L}^\varphi$ spaces.

In the preceding section we have seen how to study exponential decay in $\mathbb{L}^2$ hence in all the $\mathbb{L}^p$ for $1 < p < +\infty$. The cases $p = 1$ or $p = +\infty$ are peculiar. It turns out that in many situations of physical interest, the initial law is absolutely continuous w.r.t. μ, but its density does not belong to any L^p space for $p > 1$. It is thus interesting to look at general Orlicz spaces $\mathbb{L}^\varphi$, the most famous in physically interesting situations being $\mathbb{L}_{x \ln x}$ (whether φ has to be written as a superscript or a subscript depends on the references) i.e. the set of measures with finite relative entropy.

In all this section we shall consider a C^2 function φ, defined on $\mathbb{R}^+$, sometimes written as $\varphi(u) = uF(u)$, satisfying

$$\varphi(u)/u \to +\infty \quad \text{ and } \quad \varphi(u)/u^p \to 0 \quad \text{ as } u \to +\infty \text{ and for all } p > 1\,.$$

We assume in addition that $\varphi(1) = 0$ and that φ is uniformly convex on all compact interval $[0, A]$ (but not globally on $\mathbb{R}^+$).

We shall also assume that the process admits a carré du champ, i.e. $< -Lf, f >= \Gamma(f)$ *where* Γ *is the square of a derivation. In particular we can use the usual chain rule of differential calculus for functions in the core* $\mathcal{C}$ *which is an algebra. For instance we have*

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

Define $I_\varphi(h) = \int \varphi(h)\, d\mu$ for $h \geq 0$. In particular, when $\varphi(u) = u \ln(u)$ we recover the classical relative entropy, if h is a probability density w.r.t. μ. The following lemma can be easily proved exactly as proposition 4.2

Lemma 5.1. *For all* t *and all non-negative* h*,*

$$I_\varphi(P_t h) \leq e^{-t/C_\varphi}\, I_\varphi(h)$$

if and only if the following inequality holds

$$I_\varphi(h) \;\leq\; C_\varphi \int \varphi''(h)\, \Gamma(h)\, d\mu\,.$$

Remark that, since $\int P_t h d\mu = \int h d\mu$ for all t, we may replace $P_t h$ by $P_t h / \int P_t h d\mu$ in the previous equivalence.

With this remark, if $\varphi(u) = u \ln(u)$, the above inequality becomes if we write $h = f^2$,

$$\int f^2 \, \ln\left(\frac{f^2}{\int f^2 d\mu}\right) d\mu \leq C_{LS} \int \Gamma(f)\, d\mu$$

which is exactly the logarithmic Sobolev inequality introduced by Gross and Davies (after a pioneering work by Nelson). A still very good introduction to Poincaré and log-Sobolev inequality is the book [1]. Of course this is very peculiar to the function $u \ln(u)$ and such a simplification never holds if we replace h by f^2 unless in this case.

Inequalities like in lemma 5.1 were not much studied, because they are (apparently) only useful in this framework. We refer to [19] where convergences in Orlicz spaces are extensively studied. It is known that log-Sobolev may be used for various purposes: concentration of measure, isoperimetry ... where it seems important to have the usual energy in the right hand side. They are also very useful for studying McKean-Vlasov models. The natural generalization of log-Sobolev are the so called F-Sobolev inequalities

$$\int f^2 \, F\left(\frac{f^2}{\int f^2 d\mu}\right) d\mu \leq C_F \int \Gamma(f)\, d\mu$$

studied by various authors (see in particular the book by Feng Yu Wang [48], another main contributor, and [5,6]. There also very interesting contributions due to Aida, Grigoryan, Coulhon ...). The connections between these F-Sobolev inequalities and the preceding lemma are discussed in [19] too.

It is worth noticing that the previous F-Sobolev inequality implies a Poincaré inequality, provided F is regular enough in a neighborhood of 1. To this end just choose $f = 1 + \varepsilon\, h$ for a bounded h with $\int h\, d\mu = 0$ and look at the behavior of this expression when ε goes to 0 (recall that we have chosen $F(1) = 0$).

One has to be a little bit cautious with the terminology and the properties of such inequalities. The F-Sobolev inequality we have written is "tight", i.e. is an equality for constant functions, since we have chosen $F(1) = 0$. Very often, one has to relax it into a non tight (also called defective) inequality:

$$\int f^2 \, F\left(\frac{f^2}{\int f^2 d\mu}\right) \, d\mu \leq C_F \int \Gamma(f) \, d\mu + B \int f^2 \, d\mu \,.$$

A generic result tells us that a non tight inequality plus a Poincaré inequality implies a tight one (Rothaus lemma). This is sometimes a little bit confusing in the literature. For exponential decay, tight inequalities are necessary.

As for the Poincaré inequality, it is interesting to find necessary or sufficient conditions for F-Sobolev inequalities to hold.

A first remarkable result due to F.Y. Wang is that defective F-Sobolev inequalities are equivalent to inequalities called "super Poincaré" inequalities, i.e. inequalities of the following type : there exists some function β_{SP} such that for all nice f and all $s > 1$,

$$\int f^2 \, d\mu \leq \beta_{SP}(s) \int \Gamma(f,f) \, d\mu + s \left(\int |f| \, d\mu\right)^2 \,.$$

For example, the log-Sobolev inequality is equivalent to a super Poincaré inequality with $\beta_{SP}(s) = c/\ln(s)$. Actually, one looses something with the constants, i.e. in the log-Sobolev case for example, the constant c can be bounded from above and from below by some quantity depending on C_{LS} without exact correspondence.

In this form it is easy to see that a super Poincaré inequality implies the (usual) Poincaré inequality, provided $\beta_{SP}(1)$ is finite. Note that this is not the case for log-Sobolev, i.e starting from a tight log-Sobolev inequality we obtain a super Poincaré inequality, which in turn gives us back a defective log-Sobolev inequality. This is one of the default of the theory. One also easily sees that there is a hierarchical notion for these inequalities.

Remark 5.2. The writing of the super Poincaré inequality can look strange, since in fact, it is an infinite family of inequalities satisfied altogether. Of course, the function β can be chosen non increasing, so that one can replace the all family by a single one obtained by optimizing in s. $\diamondsuit$

Now depending on the situation it may be easier to use either F-Sobolev inequalities or super Poincaré inequalities.

For example a necessary condition for such inequalities to hold true is that the measure μ satisfies some concentration property (see [5,6,16,22,48]). For instance, for $\alpha \geq 1$, $F(u) = \ln_+^{2(1-\frac{1}{\alpha})}(u)$ implies that the tails of μ behave like $e^{-|x|^\alpha}$. We recover the case of Poincaré for $\alpha = 1$, the renowned gaussian concentration property when log-Sobolev holds (presumably due to Michel Talagrand) and all the intermediate cases. Conversely one can show that the measure with density $Z_\alpha \, e^{-|x|^\alpha}$ satisfies the corresponding F-Sobolev inequality (this particular example already appeared in a paper by Jay Rosen in 1976).

Sufficient condition for super-Poincaré inequalities or F-Sobolev inequalities can be obtained via Lyapunov type conditions. We refer to [22] to details and examples. For instance for the measures $Z^{-1} \, e^{-V}$ on $\mathbb{R}^d$ a sufficient condition for the F-Sobolev inequality is

$$\alpha \, |\nabla V|^2 - \Delta V \geq \lambda \, F(e^V) - C$$

for some $0 < \alpha < 1$ and some $\lambda > 0$, provided some technical additional conditions are fulfilled (see [5, 22]). In the log-Sobolev case this was first proved by Kusuoka and Stroock ([37] also see [13]).

Similarly, drift conditions like $< x, \nabla V(x) >\geq \alpha \, |x|^\beta - c$ for all x, some $\beta > 1$ and some $\alpha > 0$ yield a F-Sobolev inequality for $F(u) = \ln_+^{2(1-\frac{1}{\beta})}(u)$ (see [22]).

Remark 5.3. In the big family of functional inequalities, there is another hierarchy interpolating between the exponential case and the gaussian one, namely Beckner inequalities (or generalized Beckner inequalities). They became popular, in particular in the community of statistics, after the work by Latala and Oleszkiewicz [38], for their relationship with concentration properties. These inequalities are generically equivalent to some F-Sobolev or some super Poincaré inequalities (see [5, 6, 48]), but one more time, depending of which problem we want to look at, they can be more useful (more precisely, easier to use). $\diamondsuit$

Remark 5.4. Another well known application of log-Sobolev inequalities is their relation with integrability improvement. Precisely, if μ satisfies a defective log-Sobolev inequality, then, for all $2 < p < +\infty$ there exists some t_p such that for $t > t_p$, P_t is a bounded operator from $\mathbb{L}^2(\mu)$ into $\mathbb{L}^p(\mu)$. This theorem is due to Gross and independently more or less at the same time to Davies. The above property is called hyperboundedness (actually, many authors use hypercontractivity instead of hyperboundedness; it seems better to keep hypercontractivity for the cases when P_t becomes a contraction in a smaller Lebesgue space). When the log-Sobolev inequality is tight, P_t is not only bounded but a contraction operator. Converse statements are also true (see e.g. [1]).

Extensions to F-Sobolev inequalities were done in [5], where it is shown (but it is quite intricate) that generically a F-Sobolev inequality implies that P_t maps $\mathbb{L}^2$ into some smaller family of Orlicz spaces.

If one can reach the case $p = +\infty$, then the semi-group is said to be ultrabounded. A beautiful result due to F.Y. Wang again, tells us that if μ satisfies a F-Sobolev with F such that $\int^{\infty} (1/uF(u))\, du < +\infty$, then the semi-group is ultrabounded. This holds in particular when $\mu(dx) = Z_\alpha\, e^{-|x|^\alpha}\, dx$ for $\alpha > 2$ (and is wrong for $\alpha \leq 2$). Since P_t and P_t^* have the same energy form, easy duality arguments show that P_t maps L^1 into L^∞. If in addition a Poincaré inequality holds, then exponential decay holds in $\mathbb{L}^\infty$, and the semi-group is then ultracontractive. $\diamondsuit$

Remark 5.5. It is natural to ask about the results that remain true if we relax the assumption on the existence of the carré du champ, in particular if we look at the discrete time case. Except for some particular models, very few has been done done in the discrete time setting, so let us start with the continuous time setting.

On the one hand, generically,

$$\frac{d}{dt}\, I_\varphi(P_t h) = -\mathcal{E}_\mu(P_t h, \varphi'(P_t h))$$

so that an exponential decay of $I_\varphi(P_t h)$ is equivalent to a modified inequality

$$I_\varphi(h) \leq C\, \mathcal{E}_\mu(h, \varphi'(h))\,.$$

Due to the fact that we can no more use time differentiation, it does not reduce to the inequality in lemma 5.1. If $\varphi(u) = u\,\ln(u)$ it reduces to $I_\varphi(h) \leq C\,\mathcal{E}_\mu(h, \ln(h))$ which is called a modified log-Sobolev inequality.

On the other hand, provided μ is reversible, Gross theorem remains true i.e. for a non-negative h, $I_\varphi(h) \leq C\,\mathcal{E}_\mu(\sqrt{h}, \sqrt{h})$ implies that the semi-group is hypercontractive. But, still in the reversible case,

$$\mathcal{E}_\mu(\sqrt{h}, \sqrt{h}) \leq \frac{1}{4}\,\mathcal{E}_\mu(h, \ln(h))$$

and equality does not hold in general (it holds when we may apply the chain rule). So there is no more equivalence between hypercontractivity and exponential decay for the entropy. The status of Gross theorem in the non reversible case is unknown. Actually some modified notion of hyperboundedness is equivalent to the modified log-Sobolev inequality, see [11].

We saw in section 4 that in the discrete time case, Poincaré inequality is related to the operator $Id - P^*P$. P^*P is thus the correct "symmetrization" of P. The log-Sobolev inequality for the same operator was studied by Miclo [43]. It is essentially not relevant unless the state space is finite. Actually the hyperboundedness property in discrete time setting is related to the log-Sobolev inequality for $\frac{1}{2}\,(L + L^*)$, see [24]. $\diamondsuit$

6. Non exponential decay. Weak Poincaré inequalities.

In this section we shall study how to control the norms $||| \, Q_t \, |||_{\infty,2}$ (recall the notation in subsection 3.3 or the total variation distance $\| \, P_t^* \nu - \mu \, \|_{TV}$, when there is no exponential decay.

In all the section we shall work with continuous time. Some aspects of what happens in the discrete time setting will be given in a final remark.

Non exponential decays have been observed by T. Liggett and J.D. Deuschel for some particular infinite dimensional systems, using Nash type inequalities. As for super Poincaré inequalities, a Nash inequality can be linearized and is then equivalent to a family of linear inequalities introduced by M. Röckner and F.Y. Wang [46]. We shall start by explaining this framework.

6.1. **Weak Poincaré inequalities.**

Definition 6.1. *We shall say that μ satisfies a weak Poincaré inequality, if there exists a (non-increasing) function β defined on $(0,+\infty)$ such that for all bounded f and all $s>0$,*

$$Var_\mu(f) \leq \beta(s)\, \mathcal{E}_\mu(f,f) + s\, Osc^2(f)\,.$$

Of course if $\beta(0)$ is finite, the usual Poincaré inequality holds true. So we will be interested in the cases when β explodes near the origin. The next result is the key:

Proposition 6.2. *If μ satisfies a weak Poincaré inequality, then*

$$Var_\mu(P_t f) \leq \eta(t)\, Osc^2(f)$$

where

$$\eta(t) = \inf \{ r>0\, ;\, \beta(r)\, \ln(1/r) \leq \, t\}\,.$$

Conversely, if μ is symmetric and $Var_\mu(P_t f) \leq \eta(t)\, Osc^2(f)$, a weak Poincaré inequality holds with

$$\beta(t) = 4t\, \inf \{ r>0\, ;\, \frac{1}{r}\, \eta^{-1}(r\, \exp(1-(r/t)))\}\,.$$

We already knows that in the symmetric case, η cannot be exponential unless the usual Poincaré inequality is satisfied. This can also be checked by using the second part of the preceding proposition. We also know that the second statement is false in the non symmetric situation.

Example 6.3. (1) Consider general Cauchy measures (Barenblatt profiles for p.d.e. specialists) i.e.

$$d\mu_\alpha(x) = Z_\alpha\, (1+|x|)^{-(d+\alpha)}\, dx$$

for $\alpha>0$. Then $\beta(s) = c/s^{2/\alpha}$ so that $\eta(t)$ behaves like $c\, \ln(t)/t^{\alpha/2}$.

(2) For the sub exponential distributions

$$d\mu_\alpha(x) = Z_\alpha\, e^{-(|x|^\alpha)}\, dx$$

with $1 \geq \alpha > 0$, $\beta(s) = c\, \ln^{2(\frac{1}{\alpha}-1)}(1/s)$ so that $\eta(t)$ behaves like $c\, e^{-\lambda\, t^{\frac{\alpha}{2-\alpha}}}$.

(3) It can be shown more generally that if V is locally bounded, $e^{-V} dx$ always satisfies some weak Poincaré inequality.

How can we prove these results. Some of them can be directly proved, but in general with worse functions β, see [46]. In one dimension, one can find analogues to (1) in example 4.3, see [4] using the so called capacity-measure criteria (a method that goes back in a different setting to Maz'ja [40]). But there are mainly Lyapunov's type conditions. The correspondence between a Lyapunov control and weak Poincaré inequalities is nevertheless quite intricate, and one has to use intermediate inequalities, Lyapunov-Poincaré inequalities in [3], weighted inequalities in [17] where the previous (optimal) results are shown (one can also look at [9, 10, 20]).

Of course, necessary conditions in terms of concentration of measure or isoperimetric profile can be written down. Most of these examples are related to the non degenerate diffusion situation. Some examples for kinetic diffusions are studied in [3].

Remark 6.4. One can extend weak Poincaré inequalities in the discrete time setting, but this situation was not yet really explored (see [24] for the analogue of proposition 6.2). ♢

Remark 6.5. One can also ask about necessary or sufficient conditions for weak Poincaré inequalities written in terms of hitting times. The situation is very much intricate than for the usual Poincaré inequality. For instance, polynomial moments for hitting times furnish some weak Poincaré inequality, which in return can give worse polynomial moments. This is partly explained in [23] and, for linear diffusions in two recent preprints by Eva Löcherbach, Oleg Loukianov and Dasha Loukianova; the last one using the Nash (optimal) formulation of weak Poincaré inequalities. ♢

6.2. Non exponential decay. Generalizing Meyn-Tweedie.

Since we mentioned Lyapunov type criteria in the previous subsection, one can hope to generalize, at least partially, theorem 4.10. This was done in a beautiful way by Randal Douc, Gersende Fort and Arnaud Guillin in [27]. Here is the main theorem therein

Theorem 6.6. *Let $X_.$ (continuous time) be positive Harris recurrent with unique invariant probability measure μ on the locally compact state space E.*

Let ϕ be a non-negative, non-decreasing concave function such that $\phi(u)/u \to 0$ as u goes to $+\infty$. Assume that there exists a ϕ-Lyapunov function: i.e. a function $W \geq 1$ a closed petite set U and $b \geq 0$ such that

$$LW \leq -\phi(W) + b\,\mathbb{1}_U\,.$$

Then for all x,

$$\parallel P_t(x,.) - \mu \parallel_{TV} \leq C\,W(x)\,\psi(t)\,,$$

where $\psi(t) = 1/\phi \circ H_\phi^{-1}(t)$ and $H_\phi(t) = \int_0^t (1/\phi(s))ds$.

Examples of applications are contained in [3, 27], including kinetic models. This theorem was also used by M. Hairer, J. Mattingly and their coauthors in various situations. Again the key is to find a ϕ-Lyapunov function and the known examples correspond to the previously mentioned ones (in the preceding subsection).

7. Fluctuations, deviation bounds and central limit type theorems.

To finish we shall briefly discuss how to use the results on the rate of convergence or the functional inequalities associated with, in the study of (exact) deviation bounds or central limit type theorem for additive functionals i.e. processes given by

$$\frac{1}{t}\int_0^t f(X_s)\,ds$$

or their discrete time analogues. We shall describe a little bit what happens in the continuous time case, and actually in the diffusion case (i.e. assuming the existence of a carré du champ.

7.1. **Deviation bounds.**

Exponential deviation inequalities

$$\mathbb{P}_\mu \left(I_f(t) = \frac{1}{t} \int_0^t f(X_s)\, ds - \int f\, d\mu \geq R \right) \leq e^{-tA(R)}$$

are shown in [18] using a beautiful result of Liming Wu in [49] telling that exact bounds, similar to the large deviations (hence asymptotic) usual bounds can be obtained, through simple results in functional analysis, here the Lumer-Philips theorem. In other words $A(R)$ behaves like the action functional of large deviations theory. For $A(R)$ not to vanish identically, some conditions are required. Generically, one needs at least a Poincaré inequality, but depending on the integrability assumptions made on f, one can require stronger inequalities of F-Sobolev type (see [18] for details).

For non exponential bounds the situation is more intricate. The bounds we obtain are then based on results for mixing processes. We will come back to the relationship with mixing a little bit later.

7.2. **Central limit theorem. Martingale approach.**

Assume that $\int f\, d\mu = 0$. The (functional) central limit theorem for additive functionals like $I_f(.)$ is a statement like

$$t \mapsto \sqrt{t}\, I_f(t) \quad \text{converges in “distribution” to a Brownian motion (with a variance } \sigma^2)$$

where convergence in distribution is understood either for any finite dimensional distribution or on a finite time interval for a natural topology on paths, under $\mathbb{P}_\mu$.

A natural idea to attack such a problem is to solve the Poisson equation $Lg = f$ and then apply Ito's formula in order to write

$$\sqrt{t}\, I_f(t) = \frac{1}{\sqrt{t}}\,(g(X_t) - g(X_0)) + \frac{1}{\sqrt{t}}\, M_t$$

where M_t is a (local) martingale with brackets $\int_0^t \Gamma(g,g)(X_s)\, ds$. The CLT problem (if g has some good behavior) thus reduces to the central limit problem for martingales which was extensively studied in particular by Rolando Rebolledo. Skipping the technical points mentioned above, the problem thus reduces to solve the Poisson equation in good enough spaces. In particular, generically $\sigma^2 = \int \Gamma(g,g)\, d\mu$, so that the latter has to be meaningful.

A natural candidate for g is given by $g = -\int_0^{+\infty} P_s f ds$ (plug L in the integral and integrate $LP_s = d/dsP_s$). For example, if f is square integrable, g will exist in $\mathbb{L}^2(\mu)$ as soon as

$$\int_0^{+\infty} \| P_s f \|_2 \, ds < +\infty \, .$$

In the reversible case it can be improved in

$$\int_0^{+\infty} \| P_s f \|_2^2 \, ds < +\infty \, .$$

One immediately sees why rate of convergence for the semi-group is useful. For the $\mathbb{L}^2$ theory, the existence of a spectral gap ensures that g exists for all square integrable f. But if f is bounded, a weak Poincaré inequality with an integrable (or a square integrable) rate is enough.

The most renowned paper (in the reversible case) on the topic is perhaps the one by Kipnis and Varadhan ([36]). The results in [36] are proved in a different way and extended with a detailed discussion in [14]. In particular a non-reversible version of Kipnis-Varadhan result is proved therein.

But, as shown in [14], results obtained via the martingale method (which can be extended to the discrete time setting, see the references in [14]) are comparable, and in general worse, than the ones we can prove using mixing results, as described in the next subsection.

7.3. **Mixing.**

Definition 7.1. *Let $\mathcal{F}_t$ (resp. $\mathcal{G}_t$) be respectively the backward (or the past) and the forward (or the future) σ-fields generated by X_s for $0 \leq s \leq t$ (resp. $t \leq s$). The strong mixing coefficient $\alpha(t)$ is defined as :*

$$\begin{aligned} \alpha(t) &= \sup_s \left\{ \sup_{A,B} \left(\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right) \,,\, A \in \mathcal{F}_s \,,\; B \in \mathcal{G}_{s+t} \right\} , \\ &= \frac{1}{4} \sup_s \left\{ \sup_{F,G} \; Cov(F,G) \,,\, F \;\mathcal{F}_s (resp. \, G \; \mathcal{G}_{s+t}) \textit{ measurable and bounded by } 1. \right\} . \end{aligned}$$

If $\lim_{t\to\infty} \alpha(t) = 0$, the process is strongly mixing.

For Markov stationary processes, one can skip the supremum over all s. The relationship between strong mixing and rate of convergence is described below.

For centered F and G as in the definition, define f (resp. g) by $\mathbb{E}_\mu(F|X_s) = f(X_s)$ (resp. $\mathbb{E}_\mu(G|X_{t+s}) = g(X_{s+t})$). Then

$$\mathrm{Cov}_\mu(F,G) = \int f\, P_t g \, d\mu = \int P^*_{t/2} f \, P_{t/2} g \, d\mu \,. \tag{7.2}$$

Since f and g are still centered, it is then easy to see that ([14])

Proposition 7.3. *For all t,*

$$||| \, Q_t \, |||^2_{\infty,2} \vee ||| \, Q^*_t \, |||^2_{\infty,2} \leq 4\alpha(t) \leq ||| \, Q_{t/2} \, |||_{\infty,2} \, ||| \, Q^*_{t/2} \, |||_{\infty,2} \; .$$

If μ is reversible, we thus have

$$||| \, Q_t \, |||^2_{\infty,2} \leq 4\alpha(t) \leq ||| \, Q_{t/2} \, |||^2_{\infty,2} = ||| \, Q_t \, |||_{\infty,2} \; .$$

If in general we do not know wether $||| \, Q_t \, |||_{\infty,2}$ and $||| \, Q^*_t \, |||_{\infty,2}$ have the same behavior at infinity, if we know that both are slowly decaying, then one can show that they behave similarly (see [14]).

Hence we may use the amazingly huge literature dealing with deviation bounds, central limit theorem, convergence to stable processes ... for mixing processes (see e.g. the books by Paul Doukhan [28], Emmanuel Rio [45] and the survey by Florence Merlevède, Magda Peligrad and Serguei Utev [41]).

This is done in [18] for deviation bounds, [14] for the central limit theorem (including anomalous rates of convergence), in [24] for convergence to stable processes (in the discrete time case using the equivalence between spectral gap and a stronger form of mixing called ρ-mixing).

8. Final comments.

We shall stop here our trip in this fascinating area. A lot of things are still not well understood and many new ideas appeared during the last five years. Among them, a new class of inequalities became rather important : transportation inequalities introduced by Michel Talagrand again. These inequalities relate Wasserstein distances (that can be useful including for infinite dimensional models) and various forms of energy. How to characterize the rate of convergence in Wasserstein distance (instead of total variation distance)? A first answer has been given very recently by Francois Bolley, Ivan Gentil and Arnaud Guillin. How are these inequalities connected with other functional inequalities ? Important contributions to this question have been done by Nathaël Gozlan (alone or with coauthors) after pioneering works by Otto, Villani, Bolley, Ledoux, Gentil, Bobkov ... among many others. These inequalities are strongly linked with large deviations theory. This was put in form by Nathaël Gozlan and Christian Léonard (see their impressive survey [31]). With no doubt, the very partial state of the art of these notes will very soon be definitely outdated.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses*. Société Mathématique de France, Paris, 2000.

[2] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin. A simple proof of the Poincaré inequality for a large class of probability measures. *Electronic Communications in Probability.*, 13:60–66, 2008.

[3] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes : Lyapunov versus Poincaré. *J. Func. Anal.*, 254:727–759, 2008.

[4] F. Barthe, P. Cattiaux, and C. Roberto. Concentration for independent random variables with heavy tails. *AMRX*, 2005(2):39–60, 2005.

[5] F. Barthe, P. Cattiaux, and C. Roberto. Interpolated inequalities between exponential and Gaussian, Orlicz hypercontractivity and isoperimetry. *Rev. Mat. Iber.*, 22(3):993–1066, 2006.

[6] F. Barthe, P. Cattiaux, and C. Roberto. Isoperimetry between exponential and Gaussian. *Electronic J. Prob.*, 12:1212–1237, 2007.

[7] M. Benaim and N. El Karoui. *Promenade aléatoire*. Mathématiques Appliquées. Les éditions de l'École Polytechnique, diffusion Ellipses, Palaiseau, 2004.

[8] S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Prob.*, 27(4):1903–1921, 1999.

[9] S. G. Bobkov. Large deviations and isoperimetry over convex probability measures. *Electr. J. Prob.*, 12:1072–1100, 2007.

[10] S.G. Bobkov and M. Ledoux. Weighted Poincaré-type inequalities for Cauchy and other convex measures. *Ann. Prob.*, 37:403–427, 2009.

[11] S.G. Bobkov and P. Tetali. Modified logarithmic sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19:289–336, 2006.

[12] P. Cattiaux. A pathwise approach of some classical inequalities. *Pot. Anal.*, 20:361–394, 2004.

[13] P. Cattiaux. Hypercontractivity for perturbed diffusion semi-groups. *Ann. Fac. des Sc. de Toulouse*, 14(4):609–628, 2005.

[14] P. Cattiaux, D. Chafai, and A. Guillin. Central Limit Theorems for additive functionals of ergodic Markov diffusions. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 9 (2):337-382, 2012.

[15] P. Cattiaux, P. Collet, A. Lambert, S. Martinez, S. Méléard, and J. San Martin. Quasi stationary distributions and diffusion models in population dynamics. *Ann. Prob.*, 37(5):1926–1969, 2009.

[16] P. Cattiaux, I. Gentil, and A. Guillin. Weak logarithmic Sobolev inequalities and entropic convergence. *Probab. Theory Related Fields*, 139(3-4):563–603, 2007.

[17] P. Cattiaux, N. Gozlan, A. Guillin, and C. Roberto. Functional inequalities for heavy tails distributions and applications to isoperimetry. *Elec. J. of Prob.*, 15:346–385, 2010.

[18] P. Cattiaux and A. Guillin. Deviation bounds for additive functionals of Markov processes. *ESAIM Probability and Statistics*, 12:12–29, 2008.

[19] P. Cattiaux and A. Guillin. Trends to equilibrium in total variation distance. *Ann. Inst. Henri Poincaré. Prob. Stat.*, 45(1):117–145, 2009.
arXiv:0703451 (2007).

[20] P. Cattiaux and A. Guillin. Functional inequalities via Lyapunov conditions. to appear in Proceedings of the summer school on Optimal Transport (Grenoble 2009). arXiv:1001.1822 (2010).

[21] P. Cattiaux, A. Guillin, and C. Roberto. Poincaré inequality and the ℓ^p convergence of semi-groups. *Elec. Comm. in Prob.*, 15:270–280, 2010.

[22] P. Cattiaux, A. Guillin, F. Y. Wang, and L. Wu. Lyapunov conditions for super Poincaré inequality. *J. Func. Anal.*, 256(6):1821–1841, 2009.

[23] P. Cattiaux, A. Guillin, and P. A. Zitt. Poincaré inequalities and hitting times. *Ann. Institut Henri Poincaré*, 49(1):95-118, 2013.

[24] P. Cattiaux and M. Manou-Abi. Limit theorems for some functionals with heavy tails of a discrete time Markov chain. To appear in ESAIM Probability and Statistics. DOI: 10.1051/ps/2013043.

[25] P. Cattiaux and S. Méléard. Competitive or weak cooperative stochastic Lotka-Voltera systems conditioned to non extinction. *J. Math. Biology.*, 60(6):797–829, 2010.

[26] M. F. Chen. *Eigenvalues, inequalities, and ergodic theory*. Probability and its Applications (New York). Springer-Verlag London Ltd., London, 2005.

[27] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stoch. Proc. Appl.*, 119(3):897–923, 2009.
[28] P. Doukhan. *Mixing. Properties and Examples*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, 1994.
[29] N. Down, S. P. Meyn, and R. L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Ann. Prob.*, 23(4):1671–1691, 1995.
[30] F. Z. Gong and L. M. Wu. Spectral gap of positive operators and its applications. *C. R. Acad. Sci. Paris, Sér. 1*, t 331:983–988, 2000.
[31] N. Gozlan and C. Léonard. Transport inequalities. a survey. *Markov Processes and Related Fields*, 16:635–736, 2010.
[32] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab. 63:109-117, 2011.
[33] R.Z. Has'minskii. *Stochastic stability of differential equations.* Sijthoff and Noordhoff, 1980.
[34] B. Helffer and F. Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten laplacians.* Number 1862 in Lect. Notes Math. Springer, 2005.
[35] C. R. Hwang, S. Y. Hwang-Ma, and S. J. Sheu. Accelerating gaussian diffusions. *Ann. Appl. Prob.*, 3(3):897–913, 1993.
[36] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusion. *Comm. Math. Phys.*, 104:1–19, 1986.
[37] S. Kusuoka and D. Stroock. Some boundedness properties of certain stationary diffusion semigroups. *J. Func. Anal.*, 60:243–264, 1985.
[38] R. Latała and K. Oleszkiewicz. Between Sobolev and Poincaré. in geometric aspects of Functional Analysis. *Lect. Notes Math.*, 1745:147–168, 2000.
[39] M. Ledoux. *The concentration of measure phenomenon.*, volume 89 of *Mathematical Surveys and Monographs.* American Mathematical Society, 2001.
[40] V. G. Maz'ja. *Sobolev spaces.* Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1985. Translated from the Russian by T. O. Shaposhnikova.
[41] F. Merlevède, M. Peligrad, and S. Utev. Recent advances in invariance principles for stationary sequences. *Probability Surveys*, 3:1–36, 2006.
[42] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.
[43] L. Miclo. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaines de Markov finies. In *Séminaire de Probabilités XXXI*, volume 1655 of *Lecture Notes in Math.*, pages 136–167. Springer, Berlin, 1997.
[44] R. Montenegro and P. Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006.
[45] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31 of *Mathématiques et Applications.* Springer-Verlag, 2000.
[46] M. Röckner and F. Y. Wang. Weak Poincaré inequalities and L^2-convergence rates of Markov semi-groups. *J. Funct. Anal.*, 185(2):564–603, 2001.
[47] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950), 2009.
[48] F. Y. Wang. *Functional inequalities, Markov processes and Spectral theory.* Science Press, Beijing, 2004.
[49] L. Wu. A deviation inequality for non-reversible Markov process. *Ann. Inst. Henri Poincaré. Prob. Stat.*, 36(4):435–445, 2000.
[50] L. Wu. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stoch. Proc. Appl.*, 91:205–238, 2001.