Advanced Studies in Pure Mathematics 39, 2004
Stochastic Analysis on Large Scale Interacting Systems
pp. 89–114

Ten Explicit Criteria of One-Dimensional Processes

Mu-Fa Chen

Abstract.

The traditional ergodicity consists a crucial part in the theory of stochastic processes, plays a key role in practical applications. The ergodicity has much refined recently, due to the study on some inequalities, which are especially powerful in the infinite dimensional situation. The explicit criteria for various types of ergodicity for birth-death processes and one-dimensional diffusions are collected in Tables 8.1 and 8.2, respectively. In particular, an interesting story about how to obtain one of the criteria for birth-death processes is explained in details. Besides, a diagram for various types of ergodicity for general reversible Markov processes is presented.

The paper is organized as follows. First, we recall the study on an exponential convergence from different point of view in different subjects: probability theory, spectral theory and harmonic analysis. Then we show by examples the difficulties of the study and introduce the explicit criterion for the convergence, the variational formulas and explicit estimates for the convergence rates. Some comparison with the known results and an application are included. Next, we present ten (eleven) criteria for the two classes of processes, respectively, with some remarks. In particular, a diagram of various types of ergodicity for general reversible Markov processes is presented. For which, partial proofs are included in Appendix. Finally, we indicate a generalization to Banach spaces, this enables us to cover a large class of inequalities (equivalently, various types of ergodicity).

Let us begin with the paper by recalling the three traditional types of ergodicity.

Received November 21, 2002.

Revised December 20, 2002.

Research supported in part by NSFC (No. 10121101), RFDP and 973 Project.

§1. Three traditional types of ergodicity

Let $Q = (q_{ij})$ be a regular Q-matrix on a countable set $E = \{i, j, k, \cdots\}$. That is, $q_{ij} \geq 0$ for all $i \neq j$, $q_i := -q_{ii} = \sum_{j \neq i} q_{ij} < \infty$ for all $i \in E$ and Q determines uniquely a transition probability matrix $P_t = (p_{ij}(t))$ (which is also called a Q-process or a Markov chain). Denote by $\pi = (\pi_i)$ a stationary distribution of P_t: $\pi P_t = \pi$ for all $t \geq 0$. From now on, assume that the Q-matrix is irreducible and hence the stationary distribution π is unique. Then, the three types of ergodicity are defined respectively as follows.

$$\text{(1.1)} \qquad \textit{Ordinary ergodicity:} \qquad \lim_{t\to\infty} |p_{ij}(t) - \pi_j| = 0$$

$$\text{(1.2)} \qquad \textit{Exponential ergodicity:} \qquad \lim_{t\to\infty} e^{\hat{\alpha} t} |p_{ij}(t) - \pi_j| = 0$$

$$\text{(1.3)} \qquad \textit{Strong ergodicity:} \qquad \lim_{t\to\infty} \sup_i |p_{ij}(t) - \pi_j| = 0$$

$$\iff \lim_{t\to\infty} e^{\hat{\beta} t} \sup_i |p_{ij}(t) - \pi_j| = 0,$$

where $\hat{\alpha}$ and $\hat{\beta}$ are (the largest) positive constants and i, j varies over whole E. The equivalence in (1.3) is well known but one may refer to Proof (b) in the Appendix of this paper. These definitions are meaningful for general Markov processes once the pointwise convergence is replaced by the convergence in total variation norm. The three types of ergodicity were studied in a great deal during 1953–1981. Especially, it was proved that

strong ergodicity $\Longrightarrow$ exponential ergodicity $\Longrightarrow$ ordinary ergodicity.

Refer to Anderson (1991), Chen (1992, Chapter 4) and Meyn and Tweedie (1993) for details and related references. The study is quite complete in the sense that we have the following criteria which are described by the Q-matrix plus a test sequence (y_i) only, except the exponential ergodicity for which one requires an additional parameter λ.

Theorem 1.1 (Criteria). *Let $H \neq \emptyset$ be an arbitrary but fixed finite subset of E. Then the following conclusions hold.*

(1) *The process P_t is ergodic iff the system of inequalities*

$$\text{(1.4)} \qquad \begin{cases} \sum_j q_{ij} y_j \leq -1, & i \notin H \\ \sum_{i \in H} \sum_{j \neq i} q_{ij} y_j < \infty \end{cases}$$

has a nonnegative finite solution (y_i).

(2) *The process P_t is exponentially ergodic iff for some $\lambda > 0$ with $\lambda < q_i$ for all i, the system of inequalities*

$$(1.5)\qquad \begin{cases} \sum_j q_{ij} y_j \le -\lambda y_i - 1, & i \notin H \\ \sum_{i\in H} \sum_{j\neq i} q_{ij} y_j < \infty \end{cases}$$

has a nonnegative finite solution (y_i).

(3) *The process P_t is strongly ergodic iff the system (1.4) of inequalities has a bounded nonnegative solution (y_i).*

The probabilistic meaning of the criteria reads respectively as follows:

$$\max_{i\in H} \mathbb{E}_i \sigma_H < \infty, \quad \max_{i\in H} \mathbb{E}_i e^{\lambda \sigma_H} < \infty \quad \text{and} \quad \sup_{i\in E} \mathbb{E}_i \sigma_H < \infty,$$

where $\sigma_H = \inf\{t \ge \text{the first jumping time} : X_t \in H\}$ and λ is the same as in (1.5). The criteria are not completely explicit since they depend on the test sequences (y_i) and in general it is often non-trivial to solve a system of infinite inequalities. Hence, one expects to find out some explicit criteria for some specific processes. Clearly, for this, the first candidate should be the birth-death process. Recall that for a birth-death process with state space $E = \mathbb{Z}_+ = \{0, 1, 2, \cdots\}$, its Q-matrix has the form: $q_{i,i+1} = b_i > 0$ for all $i \ge 0$, $q_{i,i-1} = a_i > 0$ for all $i \ge 1$ and $q_{ij} = 0$ for all other $i \neq j$. Along this line, it was proved by Tweedie (1981)(see also Anderson (1991) or Chen (1992)) that

$$(1.6)\qquad S := \sum_{n\ge 1} \mu_n \sum_{j\le n-1} \frac{1}{\mu_j b_j} < \infty \Longrightarrow \text{Exponential ergodicity},$$

where $\mu_0 = 1$ and $\mu_n = b_0 \cdots b_{n-1}/a_1 \cdots a_n$ for all $n \ge 1$. Refer to Wang (1980), Yang (1986) or Hou et al (2000) for the probabilistic meaning of S. The condition is explicit since it depends only on the rates a_i and b_i. However, the condition is not necessary. A simple example is as follows. Let $a_i = b_i = i^\gamma\ (i \ge 1)$ and $b_0 = 1$. Then the process is exponential ergodic iff $\gamma \ge 2$ (see Chen (1996)) but $S < \infty$ iff $\gamma > 2$. Surprisingly, the condition is correct for strong ergodicity.

Theorem 1.2 (Zhang, Lin and Hou (2000)).

$$S < \infty \Longleftrightarrow \textit{Strong ergodicity.}$$

Refer to Hou et al (2000). With a different proof, the result is extended by Y. H. Zhang (2001) to the single-birth processes with state space $\mathbb{Z}_+$. Here, the term "single birth" means that $q_{i,i+1} > 0$ for all $i \ge 0$

but $q_{ij} \geq 0$ can be arbitrary for $j < i$. Introducing this class of Q-processes is due to the following observation: If the first inequality in (1.4) is replaced by equality, then we get a recursion formula for (y_i) with one parameter only. Hence, there should exist an explicit criterion for the ergodicity (resp. uniqueness, recurrence and strong ergodicity). For (1.5), there is also a recursion formula but now two parameters are involved and so it is unclear whether there exists an explicit criterion or not for the exponential ergodicity.

Note that the criteria are not enough to estimate the convergence rate $\hat{\alpha}$ or $\hat{\beta}$ (cf. Chen (2000a)). It is the main reason why we have to come back to study the well-developed theory of Markov chains. For birth-death processes, the estimation of $\hat{\alpha}$ was studied by Doorn in a book (1981) and in a series of papers (1985, 1987, 1991). He proved, for instance, the following lower bound

$$\hat{\alpha} \geq \inf_{i \geq 0} \left\{a_{i+1} + b_i - \sqrt{a_i b_i} - \sqrt{a_{i+1} b_{i+1}}\right\},$$

which is exact when a_i and b_i are constant. The following formula for the lower bounds was implicated in his papers and rediscovered in a different point of view (in the study on spetral gap) by Chen (1996):

$$\hat{\alpha} = \sup_{v>0} \inf_{i \geq 0} \{a_{i+1} + b_i - a_i/v_{i-1} - b_{i+1} v_i\}.$$

Besides, the precise $\hat{\alpha}$ was determined by Doorn for four practical models. The main tool used in Doorn's study is the Karlin-Mcgregor's representation theorem, a specific spectral representation, involving heavy techniques. There is no explicit criterion for $\hat{\alpha} > 0$ ever appeared so far.

§2. The first (non-trivial) eigenvalue (spectral gap)

The birth-death processes have a nice property—symmetrizability: $\mu_i p_{ij}(t) = \mu_j p_{ji}(t)$ for all i, j and $t \geq 0$. Then, the matrix Q can be regarded as a self-adjoint operator on the real L^2-space $L^2(\mu)$ with norm $\|\cdot\|$. In other words, one can use the well-developed L^2-theory. For instance, one can study the L^2-exponential convergence given below. Assuming that $Z = \sum_i \mu_i < \infty$ and then setting $\pi_i = \mu_i / Z$. Then, the convergence means that

$$\|P_t f - \pi(f)\| \leq \|f - \pi(f)\| \leq e^{-\lambda_1 t} \tag{2.1}$$

for all $t \geq 0$, where $\pi(f) = \int f \mathrm{d}\pi$ and λ_1 is the first non-trivial eigenvalue (more precisely, the spectral gap) of $(-Q)$ (cf. Chen (1992, Chapter 9)).

The estimation of λ_1 for birth-death processes was studied by Sullivan (1984), Liggett (1989) and Landim, Sethuraman and Varadhan (1996) (see also Kipnis & Lamdin (1999)). It was used as a comparison tool to handle the convergence rate for some interacting particle systems, which are infinite-dimensional Markov processes. Here we recall three results as follows.

Theorem 2.1 (Sullivan (1984)). *Let c_1 and c_2 be two constants satisfying*

$$c_1 \geq \sup_{i\geq 1} \frac{\sum_{j\geq i} \mu_j}{\mu_i}, \qquad c_2 \geq \sup_{i\geq 1} \frac{\mu_i}{\mu_i a_i}.$$

Then $\lambda_1 \geq 1/4c_1^2 c_2$.

Theorem 2.2 (Liggett (1989)). *Let c_1 and c_2 be two constants satisfying*

$$c_1 \geq \sup_{i\geq 1} \frac{\sum_{j\geq i} \mu_j}{\mu_i a_i}, \qquad c_2 \geq \sup_{i\geq 1} \frac{\sum_{j\geq i} \mu_j a_j}{\mu_i a_i}.$$

Then $\lambda_1 \geq 1/4c_1 c_2$.

Theorem 2.3 (Liggett (1989)). *For bounded a_i and b_i, $\lambda_1 > 0$ iff (μ_i) has an exponential tail.*

The reason we are mainly interested in the lower bounds is that on the one hand, they are more useful in practice and on the other hand, the upper bounds are usually easier to obtain from the following classical variational formula.

$$\lambda_1 = \inf\big\{D(f) : \mu(f) = 0, \mu\big(f^2\big) = 1\big\},$$

where

$$D(f) = \frac{1}{2}\sum_{i,j} \mu_i q_{ij}(f_j - f_i)^2, \qquad \mathscr{D}(D) = \{f \in L^2(\mu) : D(f) < \infty\}$$

and $\mu(f) = \int f \mathrm{d}\mu$.

Let us now leave Markov chains for a while and turn to diffusions.

§3. One-dimensional diffusions

As a parallel of birth-death process, we now consider an elliptic operator $L = a(x)\mathrm{d}^2/\mathrm{d}x^2 + b(x)\mathrm{d}/\mathrm{d}x$ on the half line $[0, \infty)$ with $a(x) > 0$ everywhere. Again, we are interested in estimation of the principle eigenvalues, which consist of the typical, well-known Sturm-Liouville eigenvalue problem in the spectral theory. Refer to Egorov & Kondratiev (1996) for the present status of the study and references. Here, we mention two results, which are the most general ones we have ever known before.

Theorem 3.1. *Let $b(x) \equiv 0$ (which corresponds to the birth-death process with $a_i = b_i$ for all $i \geq 1$) and set $\delta = \sup_{x>0} x \int_x^\infty a^{-1}$. Here we omit the integration variable when it is integrated with respect to the Lebesgue measure. Then, we have*

1. *Kac & Krein (1958): $\delta^{-1} \geq \lambda_0 \geq (4\delta)^{-1}$, here λ_0 is the first eigenvalue corresponding to the Dirichlet boundary $f(0) = 0$.*
2. *Kotani & Watanabe (1982): $\delta^{-1} \geq \lambda_1 \geq (4\delta)^{-1}$.*

It is simple matter to rewrite the classical variational formula as (3.1) below. Similarly, we have (3.2) for λ_0.

Poincaré inequalities.

$$\lambda_1 : \quad \|f - \pi(f)\|^2 \leq \lambda_1^{-1} D(f) \tag{3.1}$$

$$\lambda_0 : \quad \|f\|^2 \leq \lambda_0^{-1} D(f), \quad f(0) = 0. \tag{3.2}$$

It is interesting that inequality (3.2) is a special but typical case of the weighted Hardy inequality discussed in the next section.

§4. Weighted Hardy inequality

The classical Hardy inequality goes back to Hardy (1920):

$$\int_0^\infty \left(\frac{f}{x}\right)^p \leq \left(\frac{p}{p-1}\right)^p \int_0^\infty f'^p, \quad f(0) = 0, f' \geq 0,$$

where the optimal constant was determined by Landau (1926). After a long period of efforts by analysts, the inequality was finally extended to the following form, called weighted Hardy inequality (Muckenhoupt (1972))

$$\int_0^\infty f^2 \mathrm{d}\nu \leq A \int_0^\infty f'^2 \mathrm{d}\lambda, \quad f \in C^1, f(0) = 0, \tag{4.1}$$

where ν and λ be nonnegative Borel measures.

The Hardy-type inequalities play a very important role in the study of harmonic analysis and have been treated in many publications. Refer to the books: Opic & Kufner (1990), Dynkin (1990), Mazya (1985) and the survey article Davies (1999) for more details. We will come back this inequality soon.

We have finished the overview of the study on the exponential convergence (equivalently, the Poincaré inequality) in the different subjects. In order to have a more concrete feeling about the the difficulties of the topic, we now introduce some simple examples.

§5. Difficulties

First, consider the birth-death processes with finite state space E.

When $E = \{0, 1\}$, the Q-matrix becomes $Q = \begin{pmatrix} -b_0 & b_0 \\ a_1 & -a_1 \end{pmatrix}$. Then, it is trivial that $\lambda_1 = a_1 + b_0$. The result is nice since either a_1 or b_0 increases, so does λ_1. If we go one more step, $E = \{0, 1, 2\}$, then we have four parameters b_0, b_1 and a_1, a_2 and

$$\lambda_1 = 2^{-1}\big[a_1 + a_2 + b_0 + b_1 - \sqrt{(a_1 - a_2 + b_0 - b_1)^2 + 4a_1b_1}\big].$$

Now, the role for λ_1 played by the parameters becomes ambiguous. When $E = \{0, 1, 2, 3\}$, we have six parameters: $b_0, b_1, b_2, a_1, a_2, a_3$. Then

$$\lambda_1 = \frac{D}{3} - \frac{C}{3 \cdot 2^{1/3}} + \frac{2^{1/3}\left(3B - D^2\right)}{3C},$$

where the quantities D, B and C are not too complicated:

$$\begin{aligned} D &= a_1 + a_2 + a_3 + b_0 + b_1 + b_2, \\ B &= a_3\, b_0 + a_2\, (a_3 + b_0) + a_3\, b_1 \\ &\qquad + b_0\, b_1 + b_0\, b_2 + b_1\, b_2 + a_1\, (a_2 + a_3 + b_2), \\ C &= \left(A + \sqrt{4(3B - D^2)^3 + A^2}\right)^{1/3}. \end{aligned}$$

However, in the last expression, another quantity is involved:

$$\begin{aligned} A = &-2\,a_1^3 - 2\,a_2^3 - 2\,a_3^3 + 3\,a_3^2\,b_0 + 3\,a_3\,b_0^2 - 2\,b_0^3 + 3\,a_3^2\,b_1 \\ &- 12\,a_3\,b_0\,b_1 + 3\,b_0^2\,b_1 + 3\,a_3\,b_1^2 + 3\,b_0\,b_1^2 - 2\,b_1^3 - 6\,a_3^2\,b_2 + 6\,a_3\,b_0\,b_2 \\ &+ 3\,b_0^2\,b_2 + 6\,a_3\,b_1\,b_2 - 12\,b_0\,b_1\,b_2 3\,b_1^2\,b_2 - 6\,a_3\,b_2^2 + 3\,b_0\,b_2^2 + 3\,b_1\,b_2^2 \\ &- 2\,b_2^3 + 3\,a_1^2\,(a_2 + a_3 - 2\,b_0 - 2\,b_1 + b_2) \\ &+ 3\,a_2^2\,[a_3 + b_0 - 2\,(b_1 + b_2)] \\ &+ 3\,a_2\,\big[a_3^2 + b_0^2 - 2\,b_1^2 - b_1\,b_2 - 2\,b_2^2 \\ &\qquad - a_3(4\,b_0 - 2\,b_1 + b_2) + 2\,b_0(b_1 + b_2)\big] \\ &+ 3\,a_1\,\big[a_2^2 + a_3^2 - 2\,b_0^2 - b_0\,b_1 - 2\,b_1^2 - a_2(4\,a_3 - 2\,b_0 + b_1 - 2\,b_2) \\ &\qquad + 2\,b_0\,b_2 + 2\,b_1\,b_2 + b_2^2 + 2\,a_3(b_0 + b_1 + b_2)\big]. \end{aligned}$$

Thus, the roles of the parameters are completely mazed! Of course, it is impossible to compute λ_1 explicitly when the size of the matrix is greater than five!

Next, we go to the estimation of λ_1. Consider the infinite state space $E = \{0, 1, 2, \cdots\}$. Denote by g and $D(g)$, respectively, the eigenfunction of λ_1 and the degree of g when g is polynomial. Three examples of the perturbation of λ_1 and $D(g)$ are listed in Table 1.1.

$b_i\ (i \geq 0)$	$a_i\ (i \geq 1)$	λ_1	$D(g)$
$i + c\,(c > 0)$	$2i$	1	1
$i + 1$	$2i + 3$	2	2
$i + 1$	$2i + \left(4 + \sqrt{2}\right)$	3	3

Table 1.1 Three examples of the perturbation of λ_1 and $D(g)$

The first line is the well known linear model, for which $\lambda_1 = 1$, independent of the constant $c > 0$, and g is linear. Next, keeping the same birth rate, $b_i = i + 1$, changes the death rate a_i from $2i$ to $2i + 3$ (resp. $2i + 4 + \sqrt{2}$), which leads to the change of λ_1 from one to two (resp. three). More surprisingly, the eigenfunction g is changed from linear to quadratic (resp. triple). For the other values of a_i between $2i$, $2i + 3$ and $2i + 4 + \sqrt{2}$, λ_1 is unknown since g is non-polynomial. As seen from these examples, the first eigenvalue is very sensitive. Hence, in general, it is very hard to estimate λ_1.

Hopefully, I have presented enough examples to show the difficulties of the topic.

§6. Results about λ_1, $\hat{\alpha}$ and λ_0

It is position to state our results. To do so, define

$$\mathscr{W} = \{w : w_i \uparrow\uparrow, \pi(w) \geq 0\}, \quad Z = \sum_i \mu_i,$$

$$\delta = \sup_{i>0} \sum_{j \leq i-1} \frac{1}{\mu_j b_j} \sum_{j \geq i} \mu_j,$$

where $\uparrow\uparrow$ means strictly increasing. By suitable modification, we can define $\mathscr{W}'$ and explicit sequences δ_n and δ_n'. Refer to Chen (2001a) for details.

The next result provides a complete answer to the question proposed in Section 1.

Theorem 6.1. *For birth-death processes, the following assertions hold.*

(1) *Dual variational formulas:*

$$\text{(6.1)}\quad \lambda_1 = \sup_{w\in \mathscr{W}} \inf_{i\ge 0} \mu_i b_i (w_{i+1}-w_i) \Big/ \sum_{j\ge i+1} \mu_j w_j \qquad [\textit{Chen}\ (1996)]$$

$$\text{(6.2)}\quad = \inf_{w\in \mathscr{W}'} \sup_{i\ge 0} \mu_i b_i (w_{i+1}-w_i) \Big/ \sum_{j\ge i+1} \mu_j w_j \qquad [\textit{Chen}\ (2001a)]$$

(2) *Appoximating procedure and explicit bounds:*

$$Z\delta^{-1} \ge {\delta'_n}^{-1} \ge \lambda_1 \ge \delta_n^{-1} \ge (4\delta)^{-1} \textit{ for all } n \qquad [\textit{Chen}\ (2000b, 2001a)].$$

(3) *Explicit criterion:* $\lambda_1 > 0$ *iff* $\delta < \infty$ [*Miclo* (1999), *Chen* (2000*b*)].
(4) *Relation:* $\hat{\alpha} = \lambda_1$ [*Chen* (1991)].

In (6.1), only two notations are used: the sets $\mathscr{W}$ and $\mathscr{W}'$ of test functions (sequences). Clearly, for each test function, (6.1) gives us a lower bound of λ_1. This explains the meaning of "variational". Because of (6.1), it is now easy to obtain some lower estimates of λ_1, and in particular, one obtains all the lower bounds mentioned above. Next, by exchanging the orders of "sup" and "inf", we get (6.2) from (6.1), ignoring a slight modification of $\mathscr{W}$. In other words, (6.1) and (6.2) are dual of one to the other. For the explicit estimates "$\delta^{-1} \ge \lambda_0 \ge (4\delta)^{-1}$" and in particular for the criterion, one needs to find out a representative test function w among all $w \in \mathscr{W}$. This is certainly not obvious, because the test function w used in the formula is indeed a mimic of the eigenfunction (eigenvector) of λ_1, and in general, the eigenvalues and the corresponding eigenfunctions can be very sensitive, as we have seen from the above examples. Fortunately, there exists such a representative function with a simple form. We will illustrate the function in the context of diffusions in the second to the last paragraph of this section.

In parallel, for diffusions on $[0,\infty]$, define

$$C(x) = \int_0^x b/a, \qquad \delta = \sup_{x>0} \int_0^x e^{-C} \int_x^\infty e^C/a,$$
$$\mathscr{F} = \big\{ f \in C[0,\infty) \cap C^1(0,\infty) : f(0) = 0 \text{ and } f'|_{(0,\infty)} > 0 \big\}.$$

Theorem 6.2 (Chen (1999a, 2000b, 2001a)). *For diffusion on* $[0,\infty)$*, the following assertions hold.*

(1) *Dual variational formulas:*

$$\lambda_0 \geq \sup_{f\in\mathscr{F}} \inf_{x>0} e^{C(x)} f'(x) \Big/ \int_x^\infty f e^C / a \tag{6.3}$$

$$\lambda_0 \leq \inf_{f\in\mathscr{F}'} \sup_{x>0} e^{C(x)} f'(x) \Big/ \int_x^\infty f e^C / a \tag{6.4}$$

Furthermore, the signs of the equality in (6.3) *and* (6.4) *hold if both a and b are continuous on* $[0,\infty)$.

(2) *Appoximating procedure and explicit bounds: A decreasing sequence* $\{\delta_n\}$ *and an increasing sequence* $\{\delta_n'\}$ *are constructed explicitly such that*

$$\delta^{-1} \geq \delta_n'^{\,-1} \geq \lambda_0 \geq \delta_n^{-1} \geq (4\delta)^{-1} \qquad \text{for all } n.$$

(3) *Explicit criterion:* λ_0 *(resp.* λ_1*)* > 0 *iff* $\delta < \infty$.

We mention that the above two results are also based on Chen and Wang (1997a).

To see the power of the dual variational formulas, let us return to the weighted Hardy's inequality.

Theorem 6.3 (Muckenhoupt (1972)). *The optimal constant A in the inequality*

$$\int_0^\infty f^2 \mathrm{d}\nu \leq A \int_0^\infty f'^2 \mathrm{d}\lambda, \quad f \in C^1, f(0) = 0, \tag{6.5}$$

satisfies $B \leq A \leq 4B$, *where* $B = \sup_{x>0} \nu[x,\infty] \int_x^\infty (\mathrm{d}\lambda_{\text{abs}}/\mathrm{dLeb})^{-1}$ *and* $\mathrm{d}\lambda_{\text{abs}}/\mathrm{dLeb}$ *is the derivative of the absolutely continuous part of* λ *with respect to the Lebesgue measure.*

By setting $\nu = \pi$ and $\lambda = e^C \mathrm{d}x$, it follows that the criterion in Theorem 6.2 is a consequence of the Muckenhoupt's Theorem. Along this line, the criteria in Theorems 6.1 and 6.2 for a typical class of the processes were also obtained by Bobkov and Götze (1999a, b), in which, the contribution of an earlier paper by Luo (1992) was noted.

We now point out that the explicit estimates "$\delta^{-1} \geq \lambda_0 \geq (4\delta)^{-1}$" in Theorems 6.2 or 6.3 follow from our variational formulas immediately. Here we consider the lower bound "$(4\delta)^{-1}$" only, the proof for the upper bound "δ^{-1}" is also easy, in terms of (6.4).

Recall that $\delta = \sup_{x>0} \int_0^x e^{-C} \int_x^\infty e^C/a$. Set $\varphi(x) = \int_0^x e^{-C}$. By using the integration by parts formula, it follows that

$$\int_x^\infty \frac{\sqrt{\varphi}\, e^C}{a} = -\int_x^\infty \sqrt{\varphi}\, \mathrm{d}\left(\int_\bullet^\infty \frac{e^C}{a}\right)$$
$$\leq \frac{\delta}{\sqrt{\varphi(x)}} + \frac{\delta}{2}\int_x^\infty \frac{\varphi'}{\varphi^{3/2}} \leq \frac{2\delta}{\sqrt{\varphi(x)}}.$$

Hence

$$I\big(\sqrt{\varphi}\,\big)(x) = \frac{e^{-C(x)}}{\big(\sqrt{\varphi}\big)'(x)}\int_x^\infty \frac{\sqrt{\varphi} e^C}{a} \leq \frac{e^{-C(x)}\sqrt{\varphi(x)}}{(1/2)e^{-C(x)}} \cdot \frac{2\delta}{\sqrt{\varphi(x)}} = 4\delta.$$

This gives us the required bound by (6.3).

Theorem 6.2 can be immediately applied to the whole line or higher-dimensional situation. For instance, for Laplacian on compact Riemannian manifolds, it was proved by Chen & Wang (1997b) that

$$\lambda_1 \geq \sup_{f\in\mathscr{F}} \inf_{r\in(0,D)} I(f)(r)^{-1} =: \xi_1,$$

where $I(f)$ is the same as before but for some specific function $C(x)$. Thanks are given to the coupling technique which reduces the higher dimensional case to dimension one. We now have $\delta^{-1} \geq {\delta'_n}^{-1} \downarrow \geq \xi_1 \geq \delta_n^{-1} \uparrow \geq (4\delta)^{-1}$, similar to Theorem 6.2. Refer to Chen (2000b, 2001a) for details. As we mentioned before, the use of the test functions is necessary for producing sharp estimates. Actually, the variational formula enables us to improve a number of best known estimates obtained previously by geometers, but none of them can be deduced from the estimates "$\delta^{-1} \geq \xi_1 \geq (4\delta)^{-1}$". Besides, the approximating procedure enables us to determine the optimal linear approximation of ξ_1 in K:

$$\xi_1 \geq \frac{\pi^2}{D^2} + \frac{K}{2},$$

where D is the diameter of the manifold and K is the lower bound of Ricci curvature (cf., Chen, Scacciatelli and Yao (2001)). We have thus shown the value of our dual variational formulas.

§7. Three basic inequalities

Up to now, we have mainly studied the Poincaré inequality, i.e., (7.1) below. Naturally, one may study other inequalities, for instance, the logarithmic Sobolev inequality or the Nash inequality listed below.

(7.1)

Poincaré inequality: $\|f-\pi(f)\|^2 \le \lambda_1^{-1} D(f)$

(7.2)

Logarithmic Sobolev inequality: $\int f^2 \log(|f|/\|f\|) \mathrm{d}\pi \le \sigma^{-1} D(f)$

(7.3)

Nash inequality: $\|f-\pi(f)\|^{2+4/\nu} \le \eta^{-1} D(f) \|f\|_1^{4/\nu}$

(for some $\nu > 0$).

Here, to save notation, σ (resp. η) denotes the largest constant so that (7.2) (resp. (7.3)) holds.

The importance of these inequalities is due to the fact that each inequality describes a type of ergodicity. First, (7.1) $\Longleftrightarrow$ (2.1). Next, the logarithmic Sobolev inequality implies (is indeed equivalent to, in the context of diffusions) the decay of the semigroup P_t to π exponentially in relative entropy with rate σ and the Nash inequality is equivalent to $\|P_t f - \pi(f)\| \le C\|f\|_1/t^{\nu/2}$.

§8. Criteria

Recently, the criteria for the last two inequalities as well as for the discrete spectrum (which means that there is no continuous spectrum and moreover, all eigenvalues have finite multiplicity) are obtained by Mao (2000, 2002a, b), based on the weighted Hardy's inequality. On the other hand, the main parts of Theorems 6.1 and 6.2 are extended to a general class of Banach spaces in Chen (2002a, d, e), which unify a large class inequalities and provide a unified criterion in particular. We can now summarize the results in Table 8.1. The table is arranged in such order that the property in the latter line is stranger than the former one, the only exception is that even though the strong ergodicity is often stronger than the logarithmic Sobolev inequality but they are not comparable in general (Chen (2002b)).

Birth-death processes

Transition intensity:

$$\begin{aligned} i \to i+1 \quad & \text{at rate} \quad b_i = q_{i,i+1} > 0 \\ \to i-1 \quad & \text{at rate} \quad a_i = q_{i,i-1} > 0. \end{aligned}$$

Define

$$\mu_0 = 1, \quad \mu_n = \frac{b_0 \cdots b_{n-1}}{a_1 \cdots a_n}, \ n \geq 1; \quad \mu[i,k] = \sum_{i \leq j \leq k} \mu_j.$$

Property	Criterion
Uniqueness	$\sum_{n\geq 0} \frac{1}{\mu_n b_n} \mu[0,n] = \infty \quad (*)$
Recurrence	$\sum_{n\geq 0} \frac{1}{\mu_n b_n} = \infty$
Ergodicity	$(*)$ & $\mu[0,\infty) < \infty$
Exponential ergodicity L^2-exponential convergence	$(*)$ & $\sup_{n\geq 1} \mu[n,\infty) \sum_{j\leq n-1} \frac{1}{\mu_j b_j} < \infty$
Logarithmic Sobolev inequality	$(*)$ & $\sup_{n\geq 1} \mu[n,\infty) \log[\mu[n,\infty)^{-1}] \sum_{j\leq n-1} \frac{1}{\mu_j b_j} < \infty$
Strong ergodicity L^1-exponential convergence	$(*)$ & $\sum_{n\geq 0} \frac{1}{\mu_n b_n} \mu[n+1,\infty) = \sum_{n\geq 1} \mu_n \sum_{j\leq n-1} \frac{1}{\mu_j b_j} < \infty$
Nash inequality	$(*)$ & $\sup_{n\geq 1} \mu[n,\infty)^{(\nu-2)/\nu} \sum_{j\leq n-1} \frac{1}{\mu_j b_j} < \infty \ (\varepsilon)$

Table 8.1. Ten criteria for birth-death processes

Here, "$(*)$ & $\cdots$" means that one requires the uniqueness condition in the first line plus the condition "$\cdots$". The "(ε)" in the last line means that there is still a small gap from being necessary. In other words, when $\nu \in (0,2]$, there is still no criterion for the Nash inequality.

Diffusion processes on $[0,\infty)$ with reflecting boundary
Operator:

$$L = a(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + b(x)\frac{\mathrm{d}}{\mathrm{d}x}.$$

Define

$$C(x) = \int_0^x b/a, \qquad \mu[x,y] = \int_x^y e^C/a.$$

For the Nash inequality, we have the same remark as before. The reason we have one more criterion here is due to the equivalence of the logarithmic Sobolev inequality and the exponential convergence in entropy. However, this is no longer true in the discrete case. In general, the logarithmic Sobolev inequality is stronger than the exponential convergence in entropy. A criterion for the exponential convergence in entropy for birth-death processes remains open (cf., Zhang and Mao (2000) and Mao and Zhang (2000)). The two equivalences in the tables come from the next diagram.

Property	Criterion
Uniqueness	$\int_0^\infty \mu[0,x]e^{-C(x)} = \infty \quad (*)$
Recurrence	$\int_0^\infty e^{-C(x)} = \infty$
Ergodicity	$(*)$ & $\mu[0,\infty) < \infty$
Exponential ergodicity L^2-exponential convergence	$(*)$ & $\sup_{x>0} \mu[x,\infty) \int_0^x e^{-C} < \infty$
Discrete spectrum	$(*)$ & $\lim_{n\to\infty} \sup_{x>n} \mu[x,\infty) \int_n^x e^{-C} = 0$
Logarithmic Sobolev inequality Exponential convergence in entropy	$(*)$ & $\sup_{x>0} \mu[x,\infty)\log[\mu[x,\infty)^{-1}] \int_0^x e^{-C} < \infty$
Strong ergodicity L^1-exponential convergence	$(*)$ & $\int_0^\infty \mu[x,\infty)e^{-C(x)} < \infty$?
Nash inequality	$(*)$ & $\sup_{x>0} \mu[x,\infty)^{(\nu-2)/\nu} \int_0^x e^{-C} < \infty(\varepsilon)$

Table 8.2. Eleven criteria for one-dimensional diffusions

§9. New picture of ergodic theory

Theorem 9.1. *Let $(E,\mathscr{E})$ be a measurable space with countably generated $\mathscr{E}$. Then, for a Markov processes with state space $(E,\mathscr{E})$, reversible and having transition probability densities with respect to a probability measure π, we have the diagram shown in Figure 9.1.*

Nash inequality

$\swarrow$ $\searrow$

Logarithmic Sobolev inequality — *L^1-exponential convergence*

$\Downarrow$ $\Vert$

Exponential convergence in entropy — *π-a.s. Strong ergodicity*

$\Downarrow$ $\Downarrow$

Poincaré inequality $\Longleftrightarrow$ *π-a.s. Exponential ergodicity*

$\Downarrow$

L^2-algebraic ergodicity

$\Downarrow$

Ordinary ergodicity

Fig. 9.1. Diagram of nine types of ergodicity

Here are some remarks about Figure 9.1.

(1) The importance of the diagram is obvious. For instance, by using the estimates obtained from the study on Poincaré inequality, based on the advantage on the analytic approach — the L^2-theory and the equivalence in the diagram, one can estimate exponentially ergodic convergence rates, for which, the known knowledge is still very limited. Actually, these two convergence rates are often coincided (cf. the proofs given in Appendix). In

particular, one obtains a criterion for the exponential ergodicity in dimension one, which has been opened for a long period. Conversely, one obtains immediately some criteria, which are indeed new, for Poincaré inequality to be held from the well-known criteria for the exponential ergodicity. Next, there is still very limited known knowledge about the L^1-spectrum, due to the structure of the L^1-space, which is only a Banach but not Hilbert space. Based on the probabilistic advantage and the identity in the diagram, from the study on the strong ergodicity, one learns a lot about the L^1-spectral gap of the generator.

(2) The L^2-*algebraic ergodicity* means that $\text{Var}(P_t f) \le CV(f)t^{1-q}$ $(t > 0)$ holds for some V having the properties: V is homogeneous of degree two (in the sense that $V(cf+d) = c^2V(f)$ for any constants c and d) and $V(f) < \infty$ for all functions f with finite support (cf. Liggett (1991)). Refer to Chen and Wang (2000), Röckner and Wang (2001) for the study on the L^2-algebraic convergence.

(3) The diagram is complete in the following sense: each single-directed implication can not be replaced by double-directed one. Moreover, the L^1-exponential convergence (resp., the strong ergodicity) and the logarithmic Sobolev inequality (resp., the exponential convergence in entropy) are not comparable.

(4) The reversibility is used in both of the identity and the equivalence. Without the reversibility, the L^2-exponential convergence still implies π-a.s. exponentially ergodic convergence.

(5) An important fact is that the condition "having densities" is used only in the identity of L^1-exponential convergence and π-a.s. strong ergodicity, without this condition, L^1-exponential convergence still implies π-a.s. strong ergodicity, and so the diagram needs only a little change (However, the reversibility is still required here). Thus, it is a natural open problem to remove this "density's condition".

(6) Except the identity and the equivalence, all the implications in the diagram are suitable for general Markov processes, not necessarily reversible, even though the inequalities are mainly valuable in the reversible situation. Clearly, the diagram extends the ergodic theory of Markov processes.

The diagram was presented in Chen (1999c, 2002b), originally stated mainly for Markov chains. Recently, the identity of L^1-exponential convergence and the π-a.s. strong ergodicity is proven by Mao (2002c). A counter-example of diffusion was constructed by Wang (2001) to show that the strong ergodicity does not imply the exponential convergence in entropy. Partial proofs of the diagram are given in Appendix.

§10. Go to Banach spaces

To conclude this paper, we indicate an idea to show the reason why we should go to the Banach spaces.

Theorem 10.1 (Varopoulos, N. (1985); Carlen, E. A., Kusuoka, S., Stroock, D. W. (1987); Bakry, D., Coulhon, T., Ledoux, M. and Saloff-Coste, L. (1995)). *When $\nu > 2$, the Nash inequality*

$$\|f - \pi(f)\|^{2+4/\nu} \leq C_1 D(f) \|f\|_1^{4/\nu}$$

is equivalent to the Sobolev-type inequality

$$\|f - \pi(f)\|_{\nu/(\nu-2)}^2 \leq C_2 D(f),$$

where $\|\cdot\|_p$ is the $L^p(\mu)$-norm.

In view of Theorem 10.1, it is natural to study the inequality

$$\big\|(f - \pi(f))^2\big\|_{\mathbb{B}} \leq AD(f)$$

for a general Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}}, \mu)$. It is interesting that even for the general setup, we still have quite satisfactory results. Refer to Bobkov and Götze (1999a, b) and Chen (2002a, d, e) for details.

§11. Appendix: Partial proofs of Theorem 9.1

The detailed proofs and some necessary counterexamples were presented in Chen (1999c, 2002b) for reversible Markov processes, except the identity of the L^1-exponential convergence and π-a.s. strong ergodicity. Note that for discrete state spaces, one can rule out "a.s." used in the diagram. Here, we prove the new identity and introduce some more careful estimates for the general state spaces. The author would like to acknowledge Y. H. Mao for his nice ideas which are included in this appendix. The steps of the proofs are listed as follows.

(a) Nash inequality $\Longrightarrow$ L^1-exponential convergence and π-a.s. Strong ergodicity.
(b) L^1-exponential convergence $\Longleftrightarrow$ π-a.s. Strong ergodicity.
(c) Nash inequality $\Longrightarrow$ Logarithmic Sobolev inequality.
(d) L^2-exponential convergence $\Longrightarrow$ π-a.s. Exponential ergodicity.
(e) Exponential ergodicity $\Longrightarrow$ L^2-exponential convergence.

(a) Nash inequality $\Longrightarrow L^1$-exponential convergence and π-a.s. Strong ergodicity [Chen (1999b)]. Denote by $\|\cdot\|_{p\to q}$ the operator's norm from $L^p(\pi)$ to $L^q(\pi)$. Note that

$$\begin{aligned}
\text{Nash inequality} &\iff \operatorname{Var}(P_t(f)) = \|P_t f - \pi(f)\|_2^2 \le C^2\|f\|_1^2/t^{q-1} \\
&\qquad\qquad\qquad\qquad\qquad\qquad (q := \nu/2+1) \\
&\iff \|(P_t-\pi)f\|_2 \le C\|f\|_1/t^{(q-1)/2}. \\
&\iff \|P_t-\pi\|_{1\to 2} \le C/t^{(q-1)/2}.
\end{aligned}$$

Since $\|P_t-\pi\|_{1\to 1} \le \|P_t-\pi\|_{1\to 2}$, we have

$$\text{Nash inequality} \Longrightarrow L^1\text{-algebraic convergence.}$$

Furthermore, because of the semigroup property, the convergence of $\|\cdot\|_{1\to 1}$ must be exponential, we indeed have

$$\text{Nash inequality} \Longrightarrow L^1\text{-exponential convergence.}$$

In the symmetric case: $P_t-\pi = (P_t-\pi)^*$, and so

$$\|P_{2t}-\pi\|_{1\to\infty} \le \|P_t-\pi\|_{1\to 2}\|P_t-\pi\|_{2\to\infty} = \|P_t-\pi\|_{1\to 2}^2.$$

Hence, $\|P_t-\pi\|_{1\to\infty} \le C/t^{q-1}$. Thus,

$$\begin{aligned}
&\operatorname{ess\,sup}_x\|P_t(x,\cdot)-\pi\|_{\text{Var}} = \operatorname{ess\,sup}_x \sup_{|f|\le 1} |(P_t(x,\cdot)-\pi)f| \\
&\quad\le \operatorname{ess\,sup}_x \sup_{\|f\|_1\le 1} |(P_t(x,\cdot)-\pi)f| = \sup_{\|f\|_1\le 1} \operatorname{ess\,sup}_x|(P_t(x,\cdot)-\pi)f| \\
&\quad= \|P_t-\pi\|_{1\to\infty} \le C/t^{q-1} \to 0, \qquad \text{as } t\to\infty.
\end{aligned}$$

This gives us the π-a.s. strong ergodicity.

(b) L^1-exponential convergence $\iff \pi$-a.s. Strong ergodicity [Mao (2002c)]. Since $(L^1)^* = L^\infty \Longrightarrow \|P_t-\pi\|_{1\to 1} = \|P_t^*-\pi\|_{\infty\to\infty}$ and $P_t^*(x,\cdot) \ll \pi$, we have

$$\begin{aligned}
\|P_t^*-\pi\|_{\infty\to\infty} &= \operatorname{ess\,sup}_x \sup_{\|f\|_\infty=1} |(P_t^*-\pi)f(x)| \\
&= \operatorname{ess\,sup}_x \sup_{\sup|f|=1} |(P_t^*-\pi)f(x)| \\
&= \operatorname{ess\,sup}_x\|P_t^*(x,\cdot)-\pi\|_{\text{Var}}.
\end{aligned}$$

Hence, π-a.s. strong ergodicity is exactly the same as the L^1-exponential convergence. Without condition "$P_t^*(x,\cdot) \ll \pi$", the second equality becomes "$\geq$", and so we have in the general reversible case that

$$L^1\text{-exponential convergence} \Longrightarrow \pi\text{-a.s. Strong ergodicity.}$$

(c) Nash inequality $\Longrightarrow$ Logarithmic Sobolev inequality [Chen (1999b)]. Because $\|f\|_1 \leq \|f\|_p$ for all $p \geq 1$, we have $\|\cdot\|_{2\to 2} \leq \|\cdot\|_{1\to 2} \leq C/t^{(q-1)/2}$, and so

$$\text{Nash inequality} \Longrightarrow \text{Poincaré inequality} \Longleftrightarrow \lambda_1 > 0.$$
$$\|P_t\|_{p\to 2} \leq \|P_t\|_{1\to 2} \leq \|P_t - \pi\|_{1\to 2} + \|\pi\|_{1\to 2} < \infty, \qquad p \in (1,2).$$

The assertion now follows from [Bakry (1992); Theorem 3.6 and Proposition 3.9].

The remainder of the Appendix is devoted to the proof of the assertion:

$$(A1)\quad L^2\text{-exponential convergence} \Longleftrightarrow \pi\text{-a.s. Exponential ergodicity.}$$

Actually, this is done by Chen (2000a). Because, by assumption, the process is reversible and $P_t(x,\cdot) \ll \pi$. Set $p_t(x,y) = \frac{\mathrm{d}P_t(x,\cdot)}{\mathrm{d}\pi}(y)$. Then we have $p_t(x,y) = p_t(y,x)$, $\pi \times \pi$-a.s. (x,y). Hence

$$(A2)\qquad \int p_s(x,y)^2\pi(\mathrm{d}y) = \int p_s(x,y)p_s(y,x)\pi(\mathrm{d}y) = p_{2s}(x,x) < \infty$$

(Carlen et al (1987)).

This means that $p_t(x,\cdot) \in L^2(\pi)$ for all $t > 0$ and π-a.s. $x \in E$. Thus, by [Chen (2000a); Theorem 1.2] and the remarks right after the theorem, $(A1)$ holds.

The proof above is mainly based on the time-discrete analog result by Roberts and Rosenthal (1997). Here, we present a more direct proof of $(A2)$ as follows.

(d) L^2-exponential convergence $\Longrightarrow$ π-a.s. Exponential ergodicity [Chen (1991, 1998, 2000a)]. Let $\mu \ll \pi$. Then

$$(A3)\qquad \begin{aligned} \|\mu P_t - \pi\|_{\text{Var}} &= \sup_{|f|\le 1} |(\mu P_t - \pi) f| = \sup_{|f|\le 1} \left|\pi\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi} P_t f - f\right)\right| \\ &= \sup_{|f|\le 1} \left|\pi\left(f P_t^*\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right) - f\right)\right| \\ &= \sup_{|f|\le 1} \left|\pi\left[f\left(P_t^*\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\right)\right)\right]\right| \\ &\le \left\|P_t^*\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\right)\right\|_1 \le \left\|\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\right\|_2 e^{-t\,\text{gap}(L^*)} \\ &= \left\|\frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1\right\|_2 e^{-t\,\text{gap}(L)}. \end{aligned}$$

We now consider two cases separately.

In the reversible case with $P_t(x,\cdot) \ll \pi$, by $(A2)$, we have

$$(A4)\qquad \begin{aligned} \|P_t(x,\cdot) - \pi\|_{\text{Var}} &\le \left\|P_{t-s}\left(\frac{\mathrm{d}P_s(x,\cdot)}{\mathrm{d}\pi} - 1\right)\right\|_1 \\ &\le \|p_s(x,\cdot) - 1\|_2 e^{-(t-s)\,\text{gap}(L)} \\ &= \left[\sqrt{p_{2s}(x,x) - 1}\, e^{s\,\text{gap}(L)}\right] e^{-t\,\text{gap}(L)}, \quad t \ge s. \end{aligned}$$

Therefore, there exists $C(x) < \infty$ such that

$$(A5)\qquad \|P_t(x,\cdot) - \pi\|_{\text{Var}} \le C(x) e^{-t\,\text{gap}(L)}, \qquad t \ge 0, \quad \pi\text{-a.s. } (x).$$

Denote by ε_1 be the largest ε such that $\|P_t(x,\cdot) - \pi\|_{\text{Var}} \le C(x)e^{-\varepsilon t}$ for all t. Then $\varepsilon_1 \ge \text{gap}(L) = \lambda_1$.

In the φ-irreducible case, without using the reversibility and transition density, from $(A3)$, one can still derive π-a.s. exponential ergodicity (but may have different rates). Refer to Roberts and Tweedie (2001) for a proof in the time-discrete situation (the title of the quoted paper is confused, where the term "L^1-convergence" is used for the π-a.s. exponentially ergodic convergence, rather than the standard meaning of L^1-exponential convergence used in this paper. These two types of convergence are essentially different as shown in Theorem 9.1). In other words, the reversibility and the existence of the transition density are not essential in this implication.

(e) π-a.s. Exponential ergodicity $\Longrightarrow$ L^2-exponential convergence [Chen (2000a), Mao (2002c)]. In the time-discrete case, a similar assertion was proved by Roberts and Rosenthal(1997) and so can be extended to the time-continuous case by using the standard technique [cf., Chen (1992), §4.4]. The proof given below provides more precise estimates. Let the σ-algebra $\mathscr{E}$ be countably generated. By Numemelin and P. Tuominen (1982) or [Numemelin (1984); Theorem 6.14 (iii)], we have in the time-discrete case that

$$\begin{aligned}&\pi\text{-a.s. geometrically ergodic convergence}\\&\qquad\iff \|\|P^n(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1 \text{ geometric convergence,}\end{aligned}$$

here and in what follows, the L^1-norm is taken with respect to the variable "$\bullet$". This implies in the time-continuous case that

$$\begin{aligned}&\pi\text{-a.s. exponentially ergodic convergence}\\&\qquad\iff \|\|P_t(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1 \text{ exponential convergence.}\end{aligned}$$

Assume that $\|\|P_t(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1 \le Ce^{-\varepsilon_2 t}$ with largest ε_2.

We now prove that $\|\|P_t(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1 \ge \|P_t-\pi\|_{\infty\to 1}$. Let $\|f\|_\infty = 1$. Then

$$\begin{aligned}\|(P_t-\pi)f\|_1 &= \int \pi(\mathrm{d}x)\Big|\int\Big[P_t(x,\mathrm{d}y)-\pi(\mathrm{d}y)\Big]f(y)\Big|\\&\le \int \pi(\mathrm{d}x)\sup_{\|g\|_\infty\le 1}\Big|\int\Big[P_t(x,\mathrm{d}y)-\pi(\mathrm{d}y)\Big]g(y)\Big|\\&= \|\|P_t(\bullet,\cdot)-\pi\|_{\text{Var}}\|_1\\&\qquad\qquad(\text{Need } P_t(x,\cdot)\ll\pi \text{ or reversibility!}).\end{aligned}$$

Next, we prove that $\|P_{2t}-\pi\|_{\infty\to 1} = \|P_t-\pi\|_{\infty\to 2}^2$ in the reversible case. We have

$$\begin{aligned}\|(P_t-\pi)f\|_2^2 &= ((P_t-\pi)f,(P_t-\pi)f) = (f,(P_t-\pi)^2 f)\\&= (f,(P_{2t}-\pi)f) \le \|f\|_\infty\|(P_{2t}-\pi)f\|_1\\&\le \|f\|_\infty^2\|P_{2t}-\pi\|_{\infty\to 1}.\end{aligned}$$

Hence $\|P_{2t}-\pi\|_{\infty\to 1} \ge \|P_t-\pi\|_{\infty\to 2}^2$. The inverse inequality is obvious by using the semigroup property and symmetry: $\|P_{2t}-\pi\|_{\infty\to 1} \le \|P_t-\pi\|_{\infty\to 2}\|P_t-\pi\|_{2\to 1} = \|P_t-\pi\|_{\infty\to 2}^2$.

We remark that in general case, without reversibility, we have $\|P_t-\pi\|_{\infty\to 1} \ge \|P_t-\pi\|_{\infty\to 2}^2/2$. Actually,

$$\begin{aligned}\|(P_t-\pi)f\|_2^2 &\le \int |(P_t-\pi)f|^2\mathrm{d}\pi \le 2\|f\|_\infty\int|(P_t-\pi)f|\mathrm{d}\pi\\&\le 2\|f\|_\infty^2\|P_t-\pi\|_{\infty\to 1},\qquad f\in L^\infty(\pi).\end{aligned}$$

Finally, assume that the process is reversible. We prove that $\lambda_1 = \text{gap}(L) \geq \varepsilon_2$. We have just proved that for every f with $\pi(f) = 0$ and $\|f\|_2 = 1$, $\|P_t f\|_2^2 \leq C\|f\|_\infty^2 e^{-2\varepsilon_2 t}$. Following [Wang (2000; Lemma 2.2), or Röckner and Wang (2001)], by the spectral representation theorem, we have

$$\begin{aligned}\|P_t f\|_2^2 &= \int_0^\infty e^{-2\lambda t} \mathrm{d}(E_\lambda f, f)\\ &\geq \left[\int_0^\infty e^{-2\lambda s} \mathrm{d}(E_\lambda f, f)\right]^{t/s} \qquad \text{(by Jensen inequality)}\\ &= \|P_s f\|_2^{2t/s}, \quad t \geq s.\end{aligned}$$

Thus, $\|P_s f\|_2^2 \leq \left[C\|f\|_\infty^2\right]^{s/t} e^{-2\varepsilon_2 s}$. Letting $t \to \infty$, we get

$$\|P_s f\|_2^2 \leq e^{-2\varepsilon_2 s}, \qquad \pi(f) = 0,\ \|f\|_2 = 1,\ f \in L^\infty(\pi).$$

Since $L^\infty(\pi)$ is dense in $L^2(\pi)$, we have

$$\|P_s f\|_2^2 \leq e^{-2\varepsilon_2 s}, \quad s \geq 0, \quad \pi(f) = 0,\ \|f\|_2 = 1.$$

Therefore, $\lambda_1 \geq \varepsilon_2$. Q.E.D.

Remark A1. Note that when $p_{2s}(\cdot, \cdot) \in L^{1/2}(\pi)$ (in particular, when $p_{2s}(x, x)$ is bounded in x) for some $s > 0$, from $(A4)$, it follows that there exists a constant C such that $\| \|P_t(\bullet, \cdot) - \pi\|_{\text{Var}}\|_1 \leq Ce^{-\lambda_1 t}$. Then, we have $\varepsilon_2 \geq \lambda_1$. Combining this with (e), we indeed have $\lambda_1 = \varepsilon_2$.

Remark A2. It is proved by Hwang et al (2002) that under mild condition, in the reversible case, $\lambda_1 = \varepsilon_1$. Refer also to Wang (2002) for related estimates.

Final remark. The main body of this paper is an updated version of Chen (2001c), which was written at the beginning stage of the study on seeking explicit criteria. The resulting picture is now quite complete and so the most parts of the original paper has to be changed, except the first section. This paper also refines a part of Chen (2002c).

Acknowledgement. This paper is based on the talks given at "Stochastic analysis on large scale interacting systems", Shonan Village Center, Hayama, Japan (July 17–26, 2002), "Stochastic analysis and statistical mechanics", Yukawa Institute, Kyoto University, Japan (July 29–30, 2002), the "First Sino-German Conference on Stochastic Analysis—A Satellite Conference of ICM 2002", The Sino-German Center, Beijing (August 29–September 3, 2002), and "Stochastic analysis in infinite dimensional spaces", RIMS, Kyoto University, Japan (November 6–8, 2002). The author is grateful for the kind invitation, financial support and the warm hospitality made by the organization committees: Profs. T. Funaki, H. Osada, N. Yosida, T. Kumagai and their colleagues and students; Profs. S. Albeverio, Z. M. Ma and M. Röckner; Profs. S. Aida, I. Shigekawa, Y. Takahashi and their colleagues.

The materials of the paper are included in a mini-course presented at the "Second Sino-French Colloquium in Probability and Applications", Wuhan, China (April 9–19, 2001), and included in the talks presented at Institute of Mathematics, Academy Sinica, Taipei (April 19–May 1, 2001), and in Department of Mathematics, Kyoto University, Kyoto, Japan (October 17, 2002–January 17, 2003). The author would like to acknowledge Prof. L. M. Wu; Profs. C. R. Hwang, T. S. Chiang, Y. S. Chow and S. J. Sheu; Profs. I. Shigekawa and Y. Takahashi for the financial support and the warm hospitality.

References

[1] Anderson, W. J. (1991), *Continuous-Time Markov Chains*, Springer Series in Statistics, Springer, Berlin.

[2] Bakry, D. (1992), *L'hypercontractivité et son utilisation en théorie des semigroupes*, LNM, **1581** Springer.

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

[4] Bobkov, S. G. and Götze, F. (1999a), *Discrete isoperimetric and Poincaré inequalities*, Prob. Th. Rel. Fields **114**, 245–277.

[5] Bobkov, S. G. and Götze, F. (1999b), *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163**, 1–28.

[6] Carlen, E. A., Kusuoka, S., Stroock, D. W. (1987), *Upper bounds for symmetric Markov transition functions*, Ann. Inst. Henri Poincaré **2**, 245–287.

[7] Chen, M. F. (1991), *Exponential L^2-convergence and L^2-spectral gap for Markov processes*, Acta Math. Sin. New Ser. **7**(1), 19–37.

[8] Chen, M. F. (1992), *From Markov Chains to Non-Equilibrium Particle Systems*, World Scientific, Singapore.

[9] Chen, M. F. (1996), *Estimation of spectral gap for Markov chains*, Acta Math. Sin. New Ser. **12**(4), 337-360.

[10] Chen, M. F. (1998), *Estimate of exponential convergence rate in total variation by spectral gap*, Acta Math. Sin. Ser. (A) 41:1 (Chinese Ed.), 1–6; Acta Math. Sin. New Ser. **14**(1), 9–16.

[11] Chen, M. F. (1999a), *Analytic proof of dual variational formula for the first eigenvalue in dimension one*, Sci. Sin. (A) **42**:8, 805–815.

[12] Chen, M. F. (1999b), *Nash inequalities for general symmetric forms*, Acta Math. Sin. Eng. Ser. **15**:3, 353–370.

[13] Chen, M. F. (1999c), *Eigenvalues, inequalities and ergodic theory* (II) , Advances in Math. **28**:6, 481–505.

[14] Chen, M. F. (2000a), *Equivalence of exponential ergodicity and L^2-exponential convergence for Markov chains*, Stoch. Proc. Appl. **87**, 281–297.

[15] Chen, M. F. (2000b), *Explicit bounds of the first eigenvalue*, Sci. in China, Ser. A, **43**(10), 1051–1059.

[16] Chen, M. F. (2001a), *Variational formulas and approximation theorems for the first eigenvalue in dimension one*, Sci. Chin. (A) **44**(4), 409–418.

[17] Chen, M. F. (2001b), *Ergodic Convergence Rates of Markov Processes — Eigenvalues, Inequalities and Ergodic Theory* [*Collection of papers,* 1993—2001], http://www.bnu.edu.cn/~chenmf/main_eng.htm.

[18] Chen, M. F. (2001c), *Explicit criteria for several types of ergodicity*, Chin. J. Appl. Prob. Stat. **17**(2), 1–8.

[19] Chen, M. F. (2002a), *Variational formulas of Poincaré-type inequalities in Banach spaces of functions on the line*, Acta Math. Sin. Eng. Ser. **18**(3), 417–436.

[20] Chen, M. F. (2002b), *A new story of ergodic theory*, in "Applied Probability", 25-34, eds. R. Chan et al., AMS/IP Studies in Advanced Mathematics, **26**, 2002.

[21] Chen, M. F. (2002c), *Ergodic Convergence Rates of Markov Processes — Eigenvalues, Inequalities and Ergodic Theory*, in Proceedings of "ICM 2002", **III**, 25–40, Higher Education Press, Beijing.

[22] Chen, M. F. (2002d), *Variational formulas of Poincaré-type inequalities for birth-death processes*, preprint, submitted to Acta Math. Sin. Eng. Ser.

[23] Chen, M. F. (2002e), *Variational formulas of Poincaré-type inequalities for one-dimensional processes*, to appear in "IMS volume dedicated to Rabi Bhattacharya".

[24] Chen, M. F., Scacciatelli, E. and Yao, L. (2002), *Linear approximation of the first eigenvalue on compact manifolds*, Sci. Sin. (A) **45**(4), 450–461.

[25] Chen, M. F. and Wang, F. Y. (1997a), *Estimation of spectral gap for elliptic operators*, Trans. Amer. Math. Soc., **349**(3), 1239-1267.

[26] Chen, M. F. and Wang, F. Y. (1997b), *General formula for lower bound of the first eigenvalue*, Sci. Sin. **40**(4), 384–394.

[27] Chen, M. F. and Wang, Y. Z. (2000), *Algebraic Convergence of Markov Chains*, to appear in Ann. Appl. Probab.

[28] Davies, E. B. (1999), *A review of Hardy inequality*, Operator Theory: Adv. & Appl. **110**, 55–67.

[29] Van Doorn, E. (1981), *Stochastic Monotonicity and Queuing Applications of Birth-Death Processes*, Lecture Notes in Statistics **4**, Springer.

[30] Van Doorn, E. (1985), *Conditions for exponential ergodicity and bounds for the decay parameter of a birth-death process*, Adv. Appl. Prob. **17**, 514–530.

[31] Van Doorn, E. (1987), *Representations and bounds for zeros of orthogonal polynomials and eigenvalues of sign-symmetric tri-diagonal matrices*, J. Approx. Th. **51**, 254–266.

[32] Van Doorn, E. (1991), *Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes*, Adv. Appl. Prob. **23**, 683–700.

[33] Dynkin, E. M. (1990), *EMS*, Springer-Valerg (1991), Berlin.

[34] Egorov, Y. and Kondratiev, V. (1996), *On Spectral Theory of Elliptic Operators*, Birkbäuser, Berlin.

[35] Hwang, C. R., Hwang-Ma, S. Y. and Sheu, S. J. (2002), *Accelerating diffusions*, preprint.

[36] Hardy, G. H. (1920), *Note on a theorem of Hilbert*, Math. Zeitschr. **6**, 314–317.

[37] Hou, Z. T. et al (2000) *Birth-death Processes*, Hunan Sci. Press, Hunan.

[38] Kac, I. S. and Krein, M. G. (1958), *Criteria for discreteness of the spectrum of a singular string*, Izv. Vyss. Učebn. Zaved. Mat. **2**, 136–153 (In Russian).

[39] Kipnis, C. and Landim, C. (1999), *Scaling Limits of Interacting Particle Systems*, Springer-Verlag, Berlin.

[40] Kotani, S. and Watanabe, S. (1982), *Krein's spectral theory of strings and generalized diffusion processes*, Lecture Notes in Math. **923**, 235–259.

[41] Landau, E. (1926), *A note on a theorem concerning series of positive term*, J. London Math. Soc. **1**, 38–39.

[42] Landim, C., Sethuraman, S. and Varadhan, S. R. S. (1996) , *Spectral gap for zero range dynamics*, Ann. Prob. **24**, 1871–1902.

[43] Liggett, T. M. (1989), *Exponential L^2 convergence of attractive reversible nearest particle systems*, Ann. Prob. **17**, 403–432.

[44] Liggett, T. M. (1991), *L^2 Rates of convergence for attractive reversible nearest particle systems: the critical case*, Ann. Prob. **19**(3), 935–959.

[45] Luo, J. H. (1992), *On discrete analog of Poincaré-type inequalities and density representation*, preprint (unpublished).

[46] Mao, Y. H. (2000), *On empty essential spectrum for Markov processes in dimension one*, preprint.

[47] Mao, Y. H. (2001), *Strong ergodicity for Markov processes by coupling methods*, preprint.

[48] Mao, Y. H. (2002a), *The logarithmic Sobolev inequalities for birth-death process and diffusion process on the line*, Chin. J. Appl. Prob. Statis. **18**(1), 94–100.

[49] Mao, Y. H. (2002b), *Nash inequalities for Markov processes in dimension one*, Acta. Math. Sin. Eng. Ser. **18**(1), 147–156.

[50] Mao, Y. H. (2002c), *In preparation.*

[51] Mao, Y. H. and Zhang, S. Y. (2000), *Comparison of some convergence rates for Markov process* (In Chinese), Acta. Math. Sin. **43**(6), 1019–1026.

[52] Mazya, V. G. (1985), *Sobolev Spaces*, Springer-Valerg.

[53] Meyn, S. P. and Tweedie, R. L. (1993), *Markov Chains and Stochastic Stability*, Springer, London.

[54] Miclo, L. (1999a), *Relations entre isopérimétrie et trou spectral pour les chaînes de Markov finies*, Prob. Th. Rel. Fields **114**, 431–485.

[55] Miclo, L. (1999b), *An example of application of discrete Hardy's inequalities*, Markov Processes Relat. Fields **5**, 319–330.

[56] Muckenhoupt, B. (1972), *Hardy's inequality with weights* Studia Math. **XLIV**, 31–38.

[57] Nummelin, E. (1984), *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge Univ. Press.

[58] Nummelin, E. and Tuominen, P. (1982), *Geometric ergodicity of Harris recurrent chains with applications to renewal theory*, Stoch. Proc. Appl. **12**, 187-202.

[59] Opic, B. and Kufner, A. (1990), *Hardy-type Inequalities*, Longman, New York.

[60] Roberts, G. O. and Rosenthal, J. S. (1997), *Geometric ergodicity and hybrid Markov chains*, Electron. Comm. Probab. **2**, 13–25.

[61] Roberts, G. O. and Tweedie, R. L. (2001), *Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains*, J. Appl. Probab. **38**(A), 37–41.

[62] Röckner, M. and Wang, F. Y. (2001), *Weak Poincaré inequalities and L^2-convergence rates of Markov semigroups*, J. Funct. Anal. **185**(2), 564–603.

[63] Sullivan, W. G. (1984), *The L^2 spectral gap of certain positive recurrent Markov chains and jump processes* Z. Wahrs. **67**, 387–398.

[64] Tweedie, R. L. (1981), *Criteria for ergodicity, exponential ergodicity and strong ergodicity of Markov processes*, J. Appl. Prob. **18**, 122–130.

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

[66] Wang, F. Y. (2000), *Functional inequalities, symmegroup properties and spectrum estimates*, Infinite Dim. Anal., Quantum Probab. and related Topics, **3**:2, 263–295.

[67] Wang, F. Y. (2001), *Convergence rates of Markov semigroups in probability distances*, preprint.

[68] Wang, F. Y. (2002), *Coupling,convergence rate of Markov processes and weak Poincaré inequalities*, Sci. in China (A), **45**:8, 975–983.

[69] Wang, Z. K. (1980), *Birth-death Processes and Markov Chains*, Science Press, Beijing (In Chinese).

[70] Yang, X. Q. (1986), *Constructions of Time-homogeneous Markov Processes with Denumerable States*, Hunan Sci. Press, Hunan.

[71] Zhang, S. Y. and Mao, Y. H. (2000), *Exponential convergence rate in Boltzman-Shannon entropy*, Sci. Sin. (*A*) **44**(3), 280–285.
[72] Zhang, Y. H. (2001), *Strong ergodicity for single birth processes*, J. Appl. Prob. **38**, 270–277.

Department of Mathematics
Beijing Normal University
Beijing 100875
The People's Republic of China
E-mail address: `mfchen@bnu.edu.cn`