Convergence in total variation distance for (in)homogeneous Markov processes

Yong-Hua Mao, Liping Xu, Ming Zhang, Yu-Hui Zhang

School of Mathematical Sciences, Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing Normal University, Beijing, 100875, PR China
Laboratoire de Probabilité et Modèles Aléatoires, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris cedex 05, France
Department of Science and Technology, China University of Political Science and Law, Beijing, 102249, PR China

A R T I C L E   I N F O

Article history:
Received 6 January 2017
Received in revised form 6 January 2018
Accepted 11 January 2018
Available online 2 February 2018

M S C:
26D10
47D07
60J25
60J60

Keywords:
(In)homogeneous Markov processes
Functional inequalities
Rate of convergence
Total variation distance

A B S T R A C T

In this paper, we study the rate of convergence in total variation distance for time continuous Markov processes, by using some $I^\psi$ and $I^{\psi, t}$-inequalities. For homogeneous reversible process, we use some homogeneous inequalities, including the Poincaré and relative entropy inequalities. For the time-inhomogeneous diffusion process, we use some inhomogeneous inequalities, including the time-dependent Poincaré and Log-Sobolev inequalities. This extends some results for the time-homogeneous diffusion processes in Cattiaux and Guillin (2009).

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

In the recent work (Cattiaux and Guillin, 2009), Cattiaux and Guillin obtained, for homogeneous continuous Markov processes, the rate of convergence to the invariant measure in total variation distance, by using functional inequalities. The authors considered the ergodic diffusion processes, which means that the infinitesimal generator $L$ satisfies the hypotheses: for $\Psi \in C^2$ and smooth $f$,

$$L \Psi(f) = \frac{\partial \Psi}{\partial x}(f) L f + \frac{1}{2} \frac{\partial^2 \Psi}{\partial x^2}(f) \Gamma(f), \quad \text{and} \quad \Gamma(f) = (\Psi'(f))^2 \Gamma(f),$$

where $\Gamma$ is the corresponding carré du champ operator associated to $L$, see Definition 2.2.

In the present paper, we extend their results in two directions. Firstly, we consider the general but reversible Markov processes. Secondly, we consider the inhomogeneous diffusion process. The inhomogeneity brings more flexibility and complex behaviors might appear consequently, see Saloff-Coste and Zúñiga (2007, 2009, 2011) for inhomogeneous finite Markov chains, and Arnaudon et al. (2008), Cheng and Zhang (2016) for inhomogeneous diffusion. In both cases, we give the
First, let us introduce some basic notations and definitions, used throughout this paper. We use $C$ or $C'$ for a positive constant, which may vary from line to line, and will indicate (in the subscript) the dependent parameters. We denote the entropy and oscillation of a (probability density) function $f$, respectively by

$$\text{Ent}_\mu(f) = \int f \log f \, d\mu \quad \text{and} \quad \text{Osc}(f) = \sup_{x,y} |f(x) - f(y)|.$$  

We always let $\psi$ be a fixed convex function in $C^2(\mathbb{R}^+)$, which satisfies

- $\psi$ is locally uniform convex, i.e. for any $A > 0$, $\inf_{|A|} \psi'' > 0$.
- $\psi(1) = 0$ and $\lim_{x \to \infty} \psi(x)/x = \infty$.

**Definition 2.1.** Given a measurable space $(E, \mathcal{E})$, let $\mathcal{P}(E)$ be the total of probability measures. For any $\mu_1, \mu_2 \in \mathcal{P}(E)$, we define the total variation distance between $\mu_1$ and $\mu_2$ by

$$\|\mu_1 - \mu_2\|_{TV} = \sup_{A \in \mathcal{E}} \{ |\mu_1(A) - \mu_2(A)| \}.$$  

Our results below differ from that in Cattiaux and Guillin (2009) in two folds. Firstly, in Section 2.1, we give the convergence rate in the total variation distance for time-homogeneous reversible Markov processes, in contrast the diffusion processes in Cattiaux and Guillin (2009). Secondly, in Section 2.2, we give the merging in the total variation distance for time-inhomogeneous diffusion processes. Here the merging means the convergence from different initial laws, and we refer to Saloff-Coste and Zúñiga (2007, 2009, 2011) for the case of finite Markov chains.

### 2.1. Time-homogeneous reversible Markov processes

Assume the time-homogeneous Markov process $(X_t, \mathbb{P}_x)$ is reversible with respect to its stationary distribution $\mu$, for more details, see Wang (2004). Let $(P_t)_{t \geq 0}$ be the associated semi-group with the infinitesimal generator $(L, D(L))$ in $L^2(\pi)$, that is

$$\int_E P_t g d\mu = \int_E g P_t d\mu, \quad f, g \in L^2(\mu) \quad \text{and} \quad \int_E f L g d\mu = \int_E g L f d\mu, \quad f, g \in D(L).$$

For any probabilistic density $h$ w.r.t. $\mu$, the law of $X_t$ with initial distribution $hd\mu$ is given by $P_t hd\mu$. We also use notations $I_\psi(t, h) = \int \psi(P_t h) d\mu$ and $I_\psi(h) = I_\psi(0, h) = \int \psi(h) d\mu$. The typical examples of reversible Markov processes include the ergodic reversible diffusion processes in Wang (2004), reversible Markov jump processes in Chen (2004). We also refer to these references for the functional inequalities.

**Definition 2.2.** We define a bilinear form

$$\Gamma(f, g) = \frac{1}{2} \left( Lfg - gLf - fg \right), \quad f, g \in D(L),$$

the homogeneous carré du champ operator associated to the semigroup or the infinitesimal generator. We use abbreviation $\Gamma'(f) = \Gamma(f, f)$.

The carré du champ operator plays a crucial role in the analysis and geometry of Markov processes. The simplest example is that the Euclidean Laplacian $\Delta$ gives rise to the standard carré du champ operator $\Gamma'(f, g) = \nabla f \cdot \nabla g$, the usual scalar product of the gradients. For further information, we refer to the original paper (Bakry and Émery, 1984) and a recent monograph (Bakry et al., 2014).

Now, we can state our first main theorem for the time-homogeneous reversible Markov processes.
**Theorem 2.3.** The following two statements are equivalent.

1. For all probabilistic density $h$ with $I_\psi(h) < \infty$ and all time $t$, we have exponential decay
   $$I_\psi(t, h) \leq e^{-t/C_\psi} I_\psi(h).$$

2. For all “nice” probabilistic density $h$, in the sense of Cattiaux et al. (2007) and Bakry et al. (2014), the following $I_\psi$-inequality holds,
   $$\int \psi(h) d\mu \leq C_\psi \int \Gamma(\psi(h), h) d\mu. \quad (1)$$

Moreover, if any of the conditions is satisfied, we have
   $$\|P_t h \mu - \mu\|_{TV} \leq C_\psi e^{-t/2C_\psi} \sqrt{I_\psi(h)}, \quad (2)$$

In order to get the convergence for fixed $h$, condition (II) can be relaxed to hold for all $P_t h$, not necessary for all probability densities; The $I_\psi$-inequality is just the Poincaré inequality when $\psi(x) = (x - 1)^2$, and is the entropy inequality studied in Zhang and Mao (2001) when $\psi(x) = x \log x$. These special cases of Theorem 2.3 are then the generalization of Cattiaux and Guillin (2009, Theorem 2.1 and 2.7).

We also consider the weak $I_\psi$-inequality condition as in Röckner and Wang (2001) and Cattiaux and Guillin (2009), by using the truncation argument. Then the total variation can be controlled by two terms, with the first one controlled by $\sqrt{I_\psi(t, h \wedge K)}$ for $K > 0$, and the second one controlled from the De La Vallée-Poussin theorem. Finally we have the following theorem.

**Theorem 2.4.** If there exist a non-increasing positive function $\beta_\psi$ and a positive functional $N$ with $N(P_t^* h) \leq N(h)$, such that the following weak $I_\psi$-inequality
   $$\int \psi(h) d\mu \leq \beta_\psi(s) \int \Gamma(\psi(h), h) d\mu + sN(h) \quad (3)$$

holds for all probability density $h$ and $s > 0$, then we have
   $$I_\psi(t, h) \leq \xi(t) (I_\psi(h) + N(h)), \quad (4)$$

where
   $$\xi(t) := \inf\{s > 0, \beta_\psi(s) \log(1/s) \leq t\}. \quad (5)$$

Moreover, we have the total variation estimate
   $$\|P_t h \mu - \mu\|_{TV} \leq C_\psi \sqrt{\xi(t)} (I_\psi(h \wedge K) + N(h \wedge K)) + \frac{2 \int h\psi(h) d\mu}{\psi(K)}, \forall K > 0, \quad (6)$$

for some nonnegative $\psi$ growing to infinity.

Similarly, when $\psi(x) = (x - 1)^2$, the weak $I_\psi$-inequality becomes the weak Poincaré inequality, while the case $\psi(x) = x \log x$ corresponds to the weak relative entropy inequality. In the later case with $N(h) = \text{Osc}^2(\sqrt{h})$, we also give the $\varepsilon$-decay of entropy from the weak relative entropy inequality.

**Theorem 2.5.** If there exists a non-increasing positive $\beta$ such that the weak relative entropy inequality
   $$\text{Ent}_\mu(h) \leq \beta(s) \int \Gamma(\log h, h) d\mu + s \text{Osc}^2(\sqrt{h})$$

holds for all probability density $h$ and $s > 0$, then we have $\varepsilon$-decay of entropy
   $$\text{Ent}_\mu(P_t h) \leq \left(\frac{1}{\varepsilon} + \varepsilon\right) \xi(\varepsilon, t) \text{Osc}^2(\sqrt{h}), \quad \forall \varepsilon > 0, \quad (7)$$

where
   $$\xi(\varepsilon, t) := \inf\{s > 0, \beta(s) \log(\frac{\varepsilon}{s}) \leq 2t\}.$$

Moreover, we have the total variation estimate
   $$\|P_t h \mu - \mu\|_{TV} \leq \frac{4 \int h\psi(h) d\mu}{(\varphi \circ \tilde{\varphi})^{-1} \left(\sqrt{2 \int h\psi(h) d\mu} \sqrt{\xi(\varepsilon, t)}\right)}, \quad \text{with } \tilde{\varphi}(x) = \sqrt{x}\varphi(x), \quad (8)$$

for some nonnegative $\varphi$ growing to infinity.
2.2. Time-inhomogeneous diffusion processes

We study in this subsection the merging in the total variation distance for inhomogeneous diffusion processes with different initial laws. Following Collet and Malrieu (2008), on \( \mathbb{R}^d \), we consider the following in-homogeneous SDE:

\[
X_{t,r} = x + \int_r^t b(s, X_{s,r}) \, ds + \sqrt{2} \int_r^t \sigma(s, X_{s,r}) \, dB_s, \quad t \geq r \geq 0,
\]

where \( (B_t)_{t \geq 0} \) is a standard Brownian motion on \( \mathbb{R}^d \). Let \( L_t \) be the in-homogeneous generator:

\[
L_tf(x) = \sum_{i,j=1}^d a_{ij}(t, x) \partial_{ij} f(x) + \sum_{i=1}^d b_i(t, x) \partial_i f(x),
\]

where \( a = \sigma \sigma^T \) and \( b \) are smooth functions of \( (t, x) \), and let \( (P_{s,t})_{0 \leq s < t} \) be the associated semigroup, which satisfies the semigroup property: for any \( s \leq r \leq t \), \( P_{s,r} = P_{s,t} P_{t,r} \), \( P_{s,t} \equiv I \), and the Kolmogorov equations:

\[
\partial_t P_{s,t} = -L_s P_{s,t}, \quad \partial_t P_{s,t} = P_{s,t} L_t f.
\]

We will always denote by \( \mu_t = \mu_0 P_{0,t} = \mu_t^* \), the law of \( X_t \).

In Collet and Malrieu (2008), the log-Sobolev inequality is studied for the in-homogeneous diffusion processes, and we also refer to Cheng (2017) for a recent study on diffusions on manifolds with time-depending metrics. For more examples of in-homogeneous diffusion processes see Collet and Malrieu (2008), Cheng (2017) and references therein.

**Definition 2.6 (Collet and Malrieu, 2008).** Define a bilinear form

\[
\Gamma_i(f, g) = \frac{1}{2} \left( L_i(fg) - fL_i(g) - gL_i(f) \right), \quad f, g \in D(L_i),
\]

the inhomogeneous carré du champ operator associated to the semigroup or infinitesimal generator. We use abbreviation \( \Gamma_i(f) = \Gamma_i(f, f) \).

Assume that \( L_i \) is a diffusion operator, that is, for any \( \Phi \in C^2 \),

\[
L_i(\Phi(f)) = \Phi'(f) L_i(f) + \Phi''(f) \Gamma_i(f), \quad f \in D(L_i)
\]

and

\[
\Gamma_i(\Phi(f), g) = \Phi'(f) \Gamma_i(f, g), \quad f, g \in D(L_i).
\]

Given two different initial laws \( \mu \) and \( \nu \), we assume \( \mu \) and \( \mu_t \) are absolutely continuous w.r.t \( \nu \) and \( \nu_t \), with Radon–Nikodym derivatives

\[
f_0 = \frac{d\mu}{dv}, \quad \text{and} \quad f_t = \frac{d\mu_t}{dv}.
\]

Note that if \( \mu \) is absolutely continuous w.r.t \( \nu \), then \( \mu P_t \) is absolutely continuous w.r.t \( \nu P_t \). Indeed, let \( P_t(x, dy) \) be the probability transition function. By assuming that \( \nu P_t(A) = \int_A \nu(dx) P_t(x, A) = 0 \) for some measurable set \( A \), we see \( P_t(x, A) = 0 \), \( \nu \)-a.s., so that \( \mu P_t(A) = \int_A f_0(x) \nu(dx) P_t(x, A) = 0 \). We also use notations

\[
I_{\psi,1} = \int f \psi f \, dv, \quad \text{and} \quad I_{\psi} = I_{\psi,0}.
\]

Now, we can state our second main result for continuous inhomogeneous diffusion Markov processes.

**Theorem 2.7.** Fix \( t > 0 \). If the following \( I_{\psi,s} \)-inequality

\[
\int \psi(f_s) \, dv_s \leq C(s) \int \psi''(f_s) F_s(f_s) \, dv_s
\]

holds for \( 0 \leq s \leq t \) and \( \mu_s \) and \( \nu_s \), then we have exponential-like decay

\[
I_{\psi,s} \leq e^{-\frac{s}{t}} \frac{d\mu}{d\nu} \psi_t
\]

holds for \( 0 \leq s \leq t \). Moreover, we have the total variation estimate

\[
\left\| \mu_t - \nu_t \right\|_{TV} \leq C_{\psi} e^{-\frac{s}{t}} \frac{d\mu}{d\nu} \sqrt{I_{\psi,s}}.
\]

The constant \( C(s) \) is dependent of \( s \) and \( \psi \). Similar to the homogeneous case in Cattiaux and Guillin (2009), when \( \psi(x) = (x-1)^2 \), the \( I_{\psi,s} \)-inequality is just the Poincaré inequality while the case that \( \psi(x) = x \log x \) corresponds to the Log-Sobolev inequality. We give these two special cases in the following two corollaries without proof.
Corollary 2.8. Fix \( t > 0 \). If Poincaré inequality
\[
v_s(f_s^2) - (v_s f_s)^2 \leq C(s) v_s (\Gamma_s(f_s))
\]
holds for \( 0 \leq s \leq t \) and \( \mu_s \) and \( v_s \), then
\[
v_s(f_s^2) - (v_s f_s)^2 \leq e^{-\int_0^s \frac{\Gamma_s}{C(s)} (v_s^2 - 1) ds}, \quad s \leq t.
\]
Moreover, we have the total variation estimate
\[
\|\mu_s - v_s\|_{TV} \leq e^{-\int_0^s \frac{\Gamma_s}{C(s)} (v_s^2 - 1) ds}.
\]

Corollary 2.9. Fix \( t > 0 \). If Log-Sobolev inequality
\[
\text{Ent}_{v_s}(f_s) \leq C(s) v_s (\Gamma_s(f_s) f_s)
\]
holds for \( 0 \leq s \leq t \) and \( \mu_s \) and \( v_s \), then
\[
\text{Ent}_{v_s}(f_s) \leq e^{-\int_0^s \frac{\Gamma_s}{C(s)} (v_s^2 - 1) ds}, \quad s \leq t.
\]
Moreover, we have the total variation estimate
\[
\|\mu_s - v_s\|_{TV} \leq \sqrt{2} e^{-\frac{1}{2} \int_0^s \frac{\Gamma_s}{C(s)} (v_s^2 - 1) ds} \text{Ent}_{v_s}(d\nu/d\mu).
\]

Similar to the time-homogeneous case (Theorem 2.3), we can also rewrite Theorem 2.7 in the following equivalence form.

Theorem 2.10. Given \( t > 0 \), the following are equivalent.

(I) For all \( 0 \leq s_0 \leq s \leq t \), we have exponential decay
\[
I_{\psi, s} \leq e^{-\int_0^s \frac{\Gamma_s}{C(s)} (v_s^2 - 1) ds}.
\]

(II) For all \( 0 \leq s \leq t \), \( I_{\psi, s} \)-inequality holds for \( \mu_s \) and \( v_s \), i.e.
\[
\int \psi(f_s) d\nu_s \leq C(s) \int \psi''(f_s) \Gamma_s(f_s) d\nu_s.
\]
Moreover, if any of the above two conditions is satisfied, then we have the total variation estimate
\[
\|\mu_s - v_s\|_{TV} \leq C_{\psi} e^{-\frac{1}{2} \int_0^s \frac{\Gamma_s}{C(s)} (v_s^2 - 1) ds}, \quad \forall \ 0 \leq s_0 \leq s \leq t.
\]

3. Proofs of the main results

In this section, we will give the proofs of our main results in Section 2. In Section 3.1, we consider the time-homogeneous case, and in Section 3.2, we consider the time-inhomogeneous case. We recall the function \( \psi \) being a fixed convex function in \( C^2(\mathbb{R}^+) \), which satisfies

- \( \psi \) is locally uniform convex, i.e. for any \( A > 0 \), \( \inf_{[0,A]} \psi'' > 0 \).
- \( \psi(1) = 0 \) and \( \lim_{x \to \infty} \psi(x)/x = \infty \).

3.1. Time-homogeneous case

Lemma 3.1 (Pinsker Inequality (Pinsker, 1964; Cattiaux and Guillin, 2009)). Given a pair of probability measures \((\mathbb{P}, \mathbb{Q})\), there exists \( C_\psi > 0 \), such that
\[
\|\mathbb{P} - \mathbb{Q}\|_{TV} \leq C_\psi \sqrt{\int \psi \left( \frac{d\mathbb{Q}}{d\mathbb{P}} \right) d\mathbb{P}}.
\]

Lemma 3.2.
\[
\frac{d}{dt} I_{\psi}(t, h) = -\int \Gamma'(\psi'(P_t h), P_t h) d\mu.
\]
Proof. By definition, we have
\[
\frac{d}{dt} l_\psi(t, h) = \int \psi'(P_t h) \frac{d}{dt} P_t h \, d\mu = \int \psi'(P_t h) L P_t h \, d\mu.
\]
Using the symmetry of $L$, we obtain
\[
\frac{d}{dt} l_\psi(t, h) = \int (\psi'(P_t h) P_t h - P_t h L (\psi'(P_t h)) - 2 \Gamma'(P_t h), P_t h) \, d\mu
\]
\[
= \int \psi'(P_t h) P_t h L^* 1 \, d\mu - \int \psi'(P_t h) L^* (P_t h) \, d\mu - 2 \int \Gamma'(P_t h), P_t h) \, d\mu
\]
\[
= \int \psi'(P_t h) L P_t h \, d\mu - \int \psi'(P_t h) L (P_t h) \, d\mu - 2 \int \Gamma'(P_t h), P_t h) \, d\mu
\]
\[
= - \frac{d}{dt} l_\psi(t, h) - 2 \int \Gamma'(P_t h), P_t h) \, d\mu,
\]
which gives the conclusion. □

Now, we can proceed to prove the main theorems in Section 2.1.

Proof of Theorem 2.3. The total variation estimate (2) comes immediately from the Pinsker inequality (Lemma 3.1) and condition (I). Then it suffices to prove the equivalence of the two conditions.

(I) $\Rightarrow$ (II): If (I) holds, then
\[
\frac{l_\psi(t, h) - l_\psi(h)}{t} \leq \frac{e^{-t/C_\psi} - 1}{t} l_\psi(h).
\]
Taking $t \to 0$ yields
\[
\frac{d}{dt} \bigg|_{t=0} l_\psi(t, h) \leq -\frac{l_\psi(h)}{C_\psi},
\]
which gives by Lemma 3.2 that
\[
\int \psi(h) d\mu \leq C_\psi \int \Gamma'(\psi'(h), P_t h) \, d\mu.
\]

(II) $\Rightarrow$ (I): If (II) holds, then replacing $h$ by $P_t h$, we have
\[
l_\psi(t, h) = \int \psi(P_t h) d\mu \leq C_\psi \int \Gamma'(\psi'(P_t h), P_t h) d\mu,
\]
which gives by Lemma 3.2 that
\[
l_\psi(t, h) \leq -C_\psi \frac{d}{dt} l_\psi(t, h).
\]
By using the Gronwall inequality,
\[
l_\psi(t, h) \leq e^{-t/C_\psi} l_\psi(h).
\]

Proof of Theorem 2.4. If the weak $l_\psi$-inequality (3) holds, then replacing $h$ by $P_t h$, we have
\[
l_\psi(t, h) \leq \beta_\psi(s) \int \Gamma'(\psi'(P_t h), P_t h) d\mu + s N(P_t h)
\]
\[
= -\beta_\psi(s) \frac{d}{dt} l_\psi(t, h) + s N(P_t h)
\]
\[
\leq -\beta_\psi(s) \frac{d}{dt} l_\psi(t, h) + s N(h),
\]
and therefore
\[
\frac{1}{\beta_\psi(s)} l_\psi(t, h) + \frac{d}{dt} l_\psi(t, h) \leq \frac{s}{\beta_\psi(s)} N(h).
\]
Multiplying both sides with $e^{t/\beta_\psi(s)}$ and integrating on $t$ over $[0, t]$, we have
\[
e^{t/\beta_\psi(s)} l_\psi(t, h) - l_\psi(h) \leq s N(h)(e^{t/\beta_\psi(s)} - 1).
\]
That is,
\[ I_\phi(t, h) \leq e^{-t/\beta_\phi(s)} I_\phi(h) + sN(h), \quad \forall s > 0. \]

By the definition (4) of \( \xi(t) \) and minimizing on the right-hand side for \( e^{-t/\beta_\phi(s)} \leq s \), we get the decay estimate
\[ I_\phi(t, h) \leq \xi(t)(I_\phi(h) + N(h)). \]

By the truncation argument, Pinsker inequality (Lemma 3.1) and De La Vallée-Poussin theorem, we finally get the total variation estimate (5):
\[
\|P_t h \mu - \mu\|_{TV} = \|P_t h - 1\|_{TV} \\
= \int |P_t(h \wedge K) - \int (h \wedge K) \mu|d\mu + 2 \int (h - K) 1_{h\geq K} d\mu \\
\leq C_\phi \sqrt{I_\phi(t, h \wedge K)} + \frac{2}{\phi(K)} \int h\phi(h)d\mu \\
\leq C_\phi \sqrt{\xi(t)(I_\phi(h \wedge K) + N(h \wedge K))} + \frac{2}{\phi(K)} \int h\phi(h)d\mu,
\]
for all \( K > 0 \) and some non-negative \( \phi \) growing to infinity. \( \square \)

**Proof of Theorem 2.5.** The proof of the first part is similar to that of Cattiaux et al. (2007, Proposition 4.1). So, it suffices to prove the total variation estimate (7). Again by using the truncation argument, Pinsker inequality (Lemma 3.1), De La Vallée-Poussin theorem, and the entropy decay (6) in the first part, we get
\[
\|P_t h \mu - \mu\|_{TV} = \|P_t h - 1\|_{TV} \\
= \int |P_t(h \wedge K) - \int (h \wedge K) \mu|d\mu + 2 \int (h - K) 1_{h\geq K} d\mu \\
\leq \sqrt{2} \int (h \wedge K) d\mu \sqrt{\int P_t(\frac{h \wedge K}{\mu(h \wedge K)}) \log(P_t(\frac{h \wedge K}{\mu(h \wedge K)})) d\mu} \\
+ 2 \int (h - K) 1_{h\geq K} d\mu \\
\leq \sqrt{2} (\frac{1}{\epsilon} + \epsilon)^{\frac{1}{2}} \sqrt{\xi(\epsilon, t)} \int (h \wedge K) d\mu \text{ Osc} \left( \sqrt{P_t(\frac{h \wedge K}{\mu(h \wedge K)})} \right) \\
+ 2 \int h\phi(h)d\mu \sqrt{\phi(K)} \\
\leq \sqrt{2} (\frac{1}{\epsilon} + \epsilon)^{\frac{1}{2}} \sqrt{\xi(\epsilon, t)} \sqrt{K} + 2 \int h\phi(h)d\mu \sqrt{\phi(K)}.
\]

On the right-hand side by letting the two terms equal, i.e. taking
\[
K = \phi^{-1}\left( \frac{\sqrt{2} \int h\phi(h)d\mu}{(\frac{1}{\epsilon} + \epsilon)^{\frac{1}{2}} \sqrt{\xi(\epsilon, t)}} \right),
\]
we finally get the desired total variation estimate (7). \( \square \)

3.2. Time-inhomogeneous case

In this section we will give the proof of our second main result in this paper, for the inhomogeneous case in Section 2.2. We focus on the proof of Theorem 2.7, and at first we give the following lemma concerning the time derivative.

**Lemma 3.3.** For any measurable function \( h \),
\[
\nu_t \left( \int \frac{d}{dt} f_t \right) = \nu_t (f_t L_t h - L_t (h f_t)).
\]

**Proof.** By Radon–Nikodym theorem and differentiation on time, we have that
\[
\frac{d}{dt} \int h d\mu_t = \frac{d}{dt} \int P_{0,t} h d\mu_t = \int P_{0,t} L_t h d\mu_t = \int L_t h d\mu_t = \int f_t L_t h \nu_t.
\]
Proof of Theorem 2.7. We have
\[ \frac{d}{dt} \int h f_t dv_t = \frac{d}{dt} \int P_{0,t}(h f_t) dv = \int (P_{0,t} L_t(h f_t) + P_{0,t}(h \frac{d}{dt} f_t)) dv = \int (L_t(h f_t) + h \frac{d}{dt} f_t) dv_t, \]
which gives finally
\[ \int f_t L_t h dv_t = \int (L_t(h f_t) + h \frac{d}{dt} f_t) dv_t. \]
The formula is then proved by definition. □

This lemma is elementary, but critical to handle the inhomogeneous case. So far, we have not seen such kind of lemma used before to study the inhomogeneous asymptotic behavior. We remark that in the homogeneous case, Lemma 3.3 is reduced into the trivial one:
\[ v(h \frac{d}{dt} f_t) = v(f_t L h). \]
This is always true by letting \( v_t \equiv v \) and \( L_t = L \) by homogeneity. On the other hand, even if \( L_t \) admits a reversible probability measure, say \( \tilde{v}_t \), what we need is an equation above concerning \( v_t \), rather than \( \tilde{v}_t \).

Proof of Theorem 2.10. By differentiation, we have
\[ \frac{d}{dt} \int \psi(f_t) dv_t = \frac{d}{dt} \int P_{0,t} \psi(f_t) dv \\
= \int (P_{0,t} L_t \psi(f_t) + P_{0,t} \psi'(f_t) \frac{d}{dt} f_t) dv \\
= v_t(L_t \psi(f_t) + \psi'(f_t) \frac{d}{dt} f_t). \]
Taking \( h = \psi'(f_t) \) in Lemma 3.3, we have
\[ v_t(\psi'(f_t) \frac{d}{dt} f_t) = v_t(f_t L_t \psi(f_t) - L_t(\psi'(f_t) f_t)). \]
After putting this into (14) and using the diffusion condition (8), (9) and the \( l_{\psi,t} \)-inequality (10), we have
\[ \frac{d}{dt} \int \psi(f_t) dv_t = v_t(L_t \psi(f_t) + f_t L_t \psi'(f_t) - L_t(\psi'(f_t) f_t)) \\
= v_t(\psi'(f_t) L_t f_t + f_t L_t \psi'(f_t) - L_t(\psi'(f_t) f_t) + \psi''(f_t) \Gamma_t(f_t)) \\
= v_t(-2 \Gamma_t(\psi'(f_t), f_t) + \psi''(f_t) \Gamma_t(f_t)) \\
\leq - \frac{1}{C(t)} \int \psi(f_t) dv_t. \]
By using the Gronwall inequality, we finally get the exponential decay
\[ \int \psi(f_t) dv_t \leq e^{-\int_0^t \frac{1}{C(t')} dt'} \int \psi(f_0) dv. \]
The total variation control (11) is again an obvious application of the Pinsker inequality (Lemma 3.1). □

The proof of the equivalent form, Theorem 2.10, is similar. We make the differential in the exponential decay to get \( l_{\psi,t} \)-inequality.

Proof of Theorem 2.10. Similar to the proof of Theorem 2.3, we now prove the other side: (I) \( \implies \) (II). In the previous proof of Theorem 2.7, we know that
\[ \frac{d}{dt} l_{\psi,t} = -v_t(\psi''(f_t) \Gamma_t(f_t)). \]
From the decay inequality (12), we have
\[ \frac{l_{\psi,s} - l_{\psi,s_0}}{s - s_0} \leq \frac{1}{C(s_0)} \frac{d}{ds} l_{\psi,s_0} \]
and after taking \( s \to s_0 \), we have
\[ \int \psi''(f_0) \Gamma_0(f_0) dv_{s_0} = - \frac{d}{ds} l_{\psi,s} \geq \frac{1}{C(s_0)} l_{\psi,s_0}. \]
This gives the inequality (13). □
Acknowledgments

Research was supported in part by NSFC (No. 11131003, 11571043, 11626245, 11771047). We thank a lot for a referee’s constructive suggestions and valuable comments. The second author is also very grateful to the School of Mathematical Sciences, Beijing Normal University, for providing so nice work environment during this project.

References