Sharp L^q-Convergence Rate in p-Wasserstein Distance for Empirical Measures of Diffusion Processes *

Feng-Yu Wang[a)], **Bingyao Wu**[b)], **Jie-Xiang Zhu**[a)]

[a)]Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

[b)]School of Mathematics and Statistics, Fujian Normal University, Fuzhou 350007, China

wangfy@tju.edu.cn, bingyaowu@163.com, jiexiangzhu7@gmail.com

Abstract

For a class of (non-symmetric) diffusion processes on a length space, which in particular include the (reflecting) diffusion processes on a connected compact Riemannian manifold, the exact convergence rate is derived for $(\mathbb{E}[\mathbb{W}_p^q(\mu_T,\mu)])^{\frac{1}{q}} (T\to\infty)$ uniformly in $(p,q)\in[1,\infty)\times(0,\infty)$, where μ_T is the empirical measure of the diffusion process, μ is the unique invariant probability measure, and $\mathbb{W}_p$ is the p-Wasserstein distance. Moreover, when the dimension parameter is less than 4, we prove that $\mathbb{E}|T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)|^q\to 0$ as $T\to\infty$ for any $q\geq 1$, where $\Xi(T)$ is explicitly given by eigenvalues and eigenfunctions for the symmetric part of the generator.

AMS subject Classification: 60D05, 58J65.
Keywords: Empirical measure, Wasserstein distance, convergence rate, diffusion process.

1 Introduction

In recent years, the convergence rate and renormalization limit in Wasserstein distance have been intensively studied for the empirical measure of different models, among other references we would like to mention [24, 25, 26, 27] for diffusion processes on manifolds with reflected or killed boundary, [16, 17, 18, 29] for subordinated diffusion processes, and [10, 14] for McKean-Vlasov SDEs and fractional Brownian motion on torus.

*Supported in part by the National Key R&D Program of China (No. 2022YFA1006000, 2020YFA0712900) and NNSFC (11921001).

In this paper, we study the empirical measure of non-symmetric diffusion processes on a length space introduced in the recent work [28], and derive the exact convergence rate in $L^q(\mathbb{P})$ for p-Wasserstein distance uniformly in $p, q \in [1, \infty)$. To see the progress made in the present paper, in the moment we simply consider the diffusion processes on a compact Riemannian manifold.

Let M be a d-dimensional compact connected Riemannian manifold possibly with a C^2-smooth boundary ∂M, let Δ be the Laplace-Beltrami operator and b be a C^1-vector field on M. It is well known that the diffusion process X_t (with reflecting boundary if ∂M exists) generated by $L := \Delta + b$ is exponentially ergodic, with a unique invariant probability measure $\mu(\mathrm{d}x) = \mathrm{e}^{V(x)}\mathrm{d}x$ for some C^2-function V on M, where $\mathrm{d}x$ is the volume measure. By the law of large numbers, μ is approximated by the empirical measures μ_T as $T \to \infty$, where

$$\mu_T := \frac{1}{T}\int_0^T \delta_{X_t}\mathrm{d}t, \quad T > 0,$$

and δ_{X_t} is the Dirac measure at the point X_t. We investigate the convergence rate of μ_T to μ under the p-Wasserstein distance for $p \in [1, \infty)$:

$$\mathbb{W}_p(\nu, \nu') := \inf_{\pi \in \mathscr{C}(\nu,\nu')} \left(\int_{M\times M} \rho(x, y)^p \, \pi(\mathrm{d}x, \mathrm{d}y)\right)^{\frac{1}{p}}, \quad \nu, \nu' \in \mathscr{P}, \tag{1.1}$$

where $\mathscr{P}$ is the set of all probability measures on M, $\mathscr{C}(\nu, \nu')$ is the class of all couplings of ν and ν', and ρ is the Riemannian distance on M.

Write $L = \hat{L} + Z$, where $\hat{L} := \Delta + \nabla V$ is the symmetric part of L, and $Z := b - \nabla V$ is the anti-symmetric part with $\mathrm{div}_\mu Z = 0$, i.e.

$$\int_M (Zf)\mathrm{d}\mu = 0, \quad f \in C^1(M).$$

According to [31], for every $0 \neq f \in L^2(\mu)$ with $\mu(f) := \int_M f\mathrm{d}\mu = 0$, we have

$$\sqrt{T}\mu_T(f) \to N(0, 2\mathbf{V}(f)) \ \text{ weakly as } T \to \infty,$$

where $N(0, 2\mathbf{V}(f))$ is the centered Normal distribution with variance $2\mathbf{V}(f)$, $\mathbf{V}(f)$ is defined by

$$\mathbf{V}(f) := \int_0^\infty \mu(fP_tf)\mathrm{d}t \in (0, \infty),$$

and $(P_t)_{t\geq 0}$ is the Markov semigroup for X_t, i.e.

$$P_tf(x) := \mathbb{E}^x[f(X_t)], \quad t \geq 0, \ x \in M, \ f \in \mathscr{B}_b(M).$$

Here, and in what follows, $\mathbb{E}^x$ stands for the expectation taken for the underly Markov process starting from x at time 0. In general, for any probability measure ν on the state space, $\mathbb{E}^\nu$ denotes the expectation taken for the underly Markov process with initial distribution ν.

For two positive variables A and B, we denote $A \preceq B$ if $A \leq cB$ holds for some constant $c > 0$, and write $A \asymp B$ if $A \preceq B$ and $B \preceq A$. [28, Theorem 1.2] provides the following estimates on the convergence rate of $\mathbb{W}_p(\mu_T, \mu)$:

- When $d=1$, for any $(p,q)\in[1,\infty)\times[1,\infty)$, then for large $T>0$

$$(1.2)\qquad (\mathbb{E}^\nu[\mathbb{W}_p^{2q}(\mu_T,\mu)])^{\frac{1}{q}}\asymp T^{-1} \text{ uniformly in } \nu\in\mathscr{P}.$$

- When $d=2$, (1.2) holds provided $p\in[1,\infty)$ and $q\in[1,\frac{p}{p-1})$.
- When $d=3$, (1.2) holds provided $p\in[1,\frac{3}{2})$ and $q\in[1,\frac{3p}{5p-3})$.
- When $d=4$, $\mathbb{E}^\nu[\mathbb{W}_2^2(\mu_T,\mu)]\asymp T^{-1}\log T$ for large T uniformly in $\nu\in\mathscr{P}$.
- There exists a constant $c\in(1,\infty)$ such that

$$\sup_{\nu\in\mathscr{P}}\big(\mathbb{E}^\nu[\mathbb{W}_{2p}^{2q}(\mu_T,\mu)]\big)^{\frac{1}{q}}\le cT^{-\frac{2}{d(3-p^{-1}-q^{-1})-2}},\quad T\ge 1$$

  provided $d=4$ and $(p,q)\in[1,\infty)\times[1,\infty)\setminus\{(1,1)\}$, or $d\ge 5$ and $(p,q)\in[1,\infty)\times[1,\infty)$.

Moreover, when $d\le 3$, and the boundary ∂M is either empty or convex, $T\mathbb{W}_2^2(\mu_T,\mu)$ converges to $\Xi(T)$ in $L^q(\mathbb{P})$ for some $q\ge 1$ as $T\to\infty$, where

$$(1.3)\qquad \Xi(T):=\sum_{i=1}^\infty\frac{|\psi_i(T)|^2}{\lambda_i},\quad \psi_i(T):=\frac{1}{\sqrt{T}}\int_0^T\phi_i(X_t)\mathrm{d}t,$$

for $\{\lambda_i\}_{i\ge0}$ and $\{\phi_i\}_{i\ge0}$ being all eigenvalues and unitary eigenfunctions of $-\hat{L}:=-(\Delta+\nabla V)$ in $L^2(\mu)$ respectively, with Neumann boundary condition if ∂M exists.

According to [28, Theorem 1.1], we have the following assertions:

- When $d\in\{1,2\}$, for any $q\in[1,\frac{2d}{(3d-4)^+})$, where $\frac{2d}{(3d-4)^+}=\infty$ for $d=1$,

$$(1.4)\qquad \lim_{T\to\infty}\sup_{\nu\in\mathscr{P}}\mathbb{E}^\nu\big[|T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)|^q\big]=0.$$

- When $d=3$, then for any $k\in(\frac{3}{2},\infty]$, $R\in[1,\infty)$ and $q\in[1,\frac{6}{5})$,

$$(1.5)\qquad \lim_{T\to\infty}\sup_{\nu\in\mathscr{P}_{k,R}}\mathbb{E}^\nu\big[|T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)|^q\big]=0,$$

  where $\mathscr{P}_{k,R}:=\{\nu=h_\nu\mu\in\mathscr{P}:\ \|h_\nu\|_{L^k(\mu)}\le R\}$, and $\nu=h_\nu\mu$ stands for a probability measure having density h_ν with respect to μ.

These results are now considerably improved by the following Theorem 1.1, which is a consequence of the general results Theorem 2.1, Corollary 2.2, Theorem 2.3, Theorem 2.4 and Corollary 2.6 stated in Section 2, since assumptions used in these results hold for the (reflecting) diffusion generated by $L:=\Delta+\nabla V+Z$ on a compact compact Riemannian manifold, where **(A)**, (2.10) and **(C)** are well known, and **(B)** is included in [30, Lemma 5.2].

Theorem 1.1. *Let $p\in[1,\infty)$ and $q\in(0,\infty)$. Then the following assertions hold.*

(1) *For any* $(k,R)\in(1,\infty]\times[1,\infty)$,

$$(1.6)\qquad \big(\mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)]\big)^{\frac{2}{q}}\asymp\begin{cases}T^{-1}, & \text{if } d\le 3,\\ T^{-1}\log(1+T), & \text{if } d=4,\\ T^{-\frac{2}{d-2}}, & \text{if } d\ge 5\end{cases}$$

holds for large $T>0$ *uniformly in* $\nu\in\mathscr{P}_{k,R}$.

(2) *When* $p\le\frac{2d}{(d-2)^+}\vee\frac{d(d-2)}{2}$, (1.6) *holds uniformly in* $\nu\in\mathscr{P}$.

(3) *Let* ∂M *be either empty or convex. When* $d\in\{1,2\}$, (1.4) *holds for any* $q\in(0,\infty)$*; while when* $d=3$, (1.5) *holds for all* $(k,R)\in(1,\infty]\times[1,\infty)$ *and* $q\in(0,\infty)$.

(4) *Let* ∂M *be either empty or convex, let* $d\le 3$ *and* $Z=0$*. Then for any* $q\in(0,\infty)$ *and* $\nu\in\mathscr{P}$,

$$\lim_{T\to\infty}\mathbb{E}^\nu\big|T\mathbb{W}_2^2(\mu_T,\mu)\big|^q=\mathbb{E}\bigg|\sum_{i=1}^\infty\frac{2\xi_i^2}{\lambda_i^2}\bigg|^q\in(0,\infty),$$

where $\{\xi_i\}_{i\ge1}$ *are i.i.d. standard random variables.*

Besides diffusion processes on compact Riemannian manifolds considered above, results for other models presented in [28, Section 5] are also improved by using our general results stated in Section 2. To save space we do not illustrate these models in details.

The remainder of the paper is organized as follows. In Section 2, we introduce the framework and main results of the paper. In Section 3, we make some preparation estimates which will be used in Sections 4-6 to prove the main results stated in Section 2.

2 Main results under a general framework

We first recall the framework studied in [28], then state the main results derived in the present paper.

2.1 The framework

Let (M,ρ) be a bounded length space, let $\mathscr{P}$ be the set of all probability measures on M, let $\mathscr{B}_b(M)$ be the class of bounded measurable functions on M, and let $C_{b,\mathrm{L}}(M)$ be the set of all bounded Lipschitz continuous functions on M. For any $p\in[1,\infty)$, the p-Wasserstein distance $\mathbb{W}_p$ is defined as in (1.1). Let $\hat{X}_t$ be a reversible Markov process on M with the unique invariant probability measure $\mu\in\mathscr{P}$ having full support M. Throughout the paper, we simply denote $\mu(f)=\int_M f\mathrm{d}\mu$ for $f\in L^1(\mu)$.

The Markov semigroup $\hat{P}_t$ of $\hat{X}_t$ is formulated as

$$\hat{P}_tf(x)=\mathbb{E}^x[f(\hat{X}_t)],\quad t\ge0,\ x\in M,\ f\in\mathscr{B}_b(M),$$

where $\mathbb{E}^x$ stands for the expectation for the underlying Markov process starting at point x. In general, for any $\nu \in \mathscr{P}$, $\mathbb{E}^\nu$ stands for the expectation for the underlying Markov process with initial distribution ν.

Let $(\hat{\mathscr{E}}, \mathscr{D}(\hat{\mathscr{E}}))$ and $(\hat{L}, \mathscr{D}(\hat{L}))$ be, respectively, the associated symmetric Dirichlet form and self-adjoint generator in $L^2(\mu)$. We assume that $C_{b,\mathrm{L}}(M)$ is a dense subset of $\mathscr{D}(\hat{\mathscr{E}})$ under the $\hat{\mathscr{E}}_1$-norm $\|f\|_{\hat{\mathscr{E}}_1} := \sqrt{\mu(f^2) + \hat{\mathscr{E}}(f,f)}$, and

$$\hat{\mathscr{E}}(f,g) = \int_M \Gamma(f,g)\mathrm{d}\mu, \quad f,g \in C_{b,\mathrm{L}}(M)$$

holds for a symmetric square field (carré du champ operator)

$$\Gamma : C_{b,\mathrm{L}}(M) \times C_{b,\mathrm{L}}(M) \to \mathscr{B}_b(M),$$

such that for any $f,g,h \in C_{b,\mathrm{L}}(M)$ and $\phi \in C_b^1(\mathbb{R})$, we have

$$\begin{aligned} &\sqrt{\Gamma(f,f)(x)} = |\nabla f(x)| := \limsup_{y\to x} \frac{|f(y)-f(x)|}{\rho(x,y)}, \quad x \in M, \\ &\Gamma(fg,h) = f\Gamma(g,h) + g\Gamma(f,h), \quad \Gamma(\phi(f),h) = \phi'(f)\Gamma(f,h). \end{aligned}$$

Moreover, we assume that $\hat{L}$ satisfies the chain rule

$$\hat{L}\Phi(f) = \Phi'(f)\hat{L}f + \Phi''(f)|\nabla f|^2, \quad f \in \mathscr{D}(\hat{L}) \cap C_{b,\mathrm{L}}(M),\ \Phi \in C^2(\mathbb{R}).$$

Let

$$Z : C_{b,\mathrm{L}}(M) \to \mathscr{B}_b(M)$$

be a bounded vector field with $\mathrm{div}_\mu Z = 0$, i.e. it satisfies

$$\begin{aligned} &Z(fg) = fZg + gZf, \quad Z(\phi(f)) = \phi'(f)Zf, \quad f,g \in C_{b,\mathrm{L}}(M),\ \phi \in C^1(\mathbb{R}), \\ &\|Z\|_\infty := \inf\big\{K \geq 0 :\ |Zf| \leq K|\nabla f|,\ f \in C_{b,\mathrm{L}}(M)\big\} < \infty, \\ &\mu(Zf) := \int_M (Zf)\mathrm{d}\mu = 0, \quad f \in C_{b,\mathrm{L}}(M). \end{aligned}$$

Consequently, Z uniquely extends to a bounded linear operator from $\mathscr{D}(\hat{\mathscr{E}})$ to $L^2(\mu)$ satisfying

$$\mu(fZg) = -\mu(gZf), \quad f,g \in \mathscr{D}(\hat{\mathscr{E}}), \tag{2.1}$$

and

$$\mathscr{E}(f,g) := \hat{\mathscr{E}}(f,g) + \mu(fZg), \quad f,g \in \mathscr{D}(\mathscr{E}) = \mathscr{D}(\hat{\mathscr{E}})$$

is a (non-symmetric) conservative Dirichlet form with generator

$$L := \hat{L} + Z, \quad \mathscr{D}(L) = \mathscr{D}(\hat{L}),$$

which satisfies the chain rule

$$L\Phi(f) = \Phi'(f)Lf + \Phi''(f)|\nabla f|^2, \quad f \in \mathscr{D}(\hat{L}) \cap C_{b,\mathrm{L}}(M),\ \Phi \in C^2(\mathbb{R}).$$

Assume that L generates a unique diffusion process X_t on M. The associated Markov semigroup is given by

$$P_t f(x) = \mathbb{E}^x[f(X_t)], \quad x \in M, \, t \geq 0, \, f \in \mathscr{B}_b(M).$$

By Duhamel's formula, P_t and $\hat{P}_t$ have the following relation

$$P_t f = \hat{P}_t f + \int_0^t P_s\{Z\hat{P}_{t-s} f\} \mathrm{d}s, \quad f \in \mathscr{D}(\hat{\mathscr{E}}), \, t \geq 0. \tag{2.2}$$

We investigate the convergence of empirical measures $\mu_T := \frac{1}{T}\int_0^T \delta_{X_t} \mathrm{d}t$ to μ under the Wasserstein distance $\mathbb{W}_p$.

2.2 Upper bound estimate

To estimate $(\mathbb{E}[\mathbb{W}_p^q(\mu_T, \mu)])^{\frac{1}{q}}$ from above, we make the following assumption.

(A) The following conditions hold for some $d \in [1, \infty)$ and an increasing function $K : [2, \infty) \to (0, \infty)$.

- **Nash inequality.** There exists a constant $C > 0$ such that

$$\mu(f^2) \leq C\hat{\mathscr{E}}(f, f)^{\frac{d}{d+2}} \mu(|f|)^{\frac{4}{d+2}}, \quad f \in \mathscr{D}_0 := \big\{f \in \mathscr{D}(\hat{\mathscr{E}}) : \; \mu(f) = 0\big\}. \tag{2.3}$$

- **Continuity of symmetric diffusion.** For any $p \in [2, \infty)$,

$$\mathbb{E}^\mu[\rho(\hat{X}_0, \hat{X}_t)^p] = \int_{M\times M} \rho(x, y)^p \hat{p}_t(x, y)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \leq K(p)t^{\frac{p}{2}}, \quad t \in [0, 1], \tag{2.4}$$

  where $\hat{p}_t$ is the heat kernel of $\hat{P}_t$ with respect to μ.

- **Boundedness of Riesz transform.** For any $p \in [2, \infty)$,

$$\|\nabla(-\hat{L})^{-\frac{1}{2}} f\|_{L^p(\mu)} \leq K(p)\|f\|_{L^p(\mu)}, \quad f \in L^p(\mu) \text{ with } \mu(f) = 0. \tag{2.5}$$

It is well known that **(A)** holds for the (reflecting) diffusion process generated by $\hat{L} := \Delta + \nabla V$ considered in Introduction.

Besides the elliptic diffusion process on compact manifolds, some criteria on the Nash inequality (2.3) are available in [23, Section 3.4]. In general, (2.3) implies that for some constant $c_0 > 0$,

$$\|\hat{P}_t - \mu\|_{L^p(\mu)\to L^q(\mu)} \leq c_0(1 \wedge t)^{-\frac{d(q-p)}{2pq}} \mathrm{e}^{-\lambda_1 t}, \quad t > 0, \; 1 \leq p \leq q \leq \infty, \tag{2.6}$$

and that $-\hat{L}$ has purely discrete spectrum with all eigenvalues $\{\lambda_i\}_{i\geq 0}$, which are listed in the increasing order counting multiplicities, satisfy

$$\lambda_i \geq c_1 i^{\frac{2}{d}}, \quad i \geq 0, \tag{2.7}$$

for some constant $c_1 > 0$. The Markov semigroup $\hat{P}_t$ generated by $\hat{L}$ has symmetric heat kernel $\hat{p}_t$ with respect to μ formulated as

$$\hat{p}_t(x,y) = 1 + \sum_{i=1}^{\infty} \mathrm{e}^{-\lambda_i t}\phi_i(x)\phi_i(y), \quad t>0,\ x,y \in M. \tag{2.8}$$

All these assertions can be found for instance in [9].

The condition (2.4) is natural for diffusion processes due to the growth property $\mathbb{E}|B_t - B_0|^p \le ct^{\frac{p}{2}}$ for the Brownian motion B_t. There are plentiful results on the boundedness condition (2.5) for the Riesz transform, see [4, 7, 8] and references therein.

The following result shows that under assumption **(A)**, the convergence rate of $(\mathbb{E}\mathbb{W}_p^q(\mu_T,\mu))^{\frac{1}{q}}$ is given by

$$\gamma_d(T) := \begin{cases} T^{-\frac{1}{2}}, & \text{if } d \in [1,4), \\ T^{-\frac{1}{2}}\sqrt{\log T}, & \text{if } d=4, \\ T^{-\frac{1}{d-2}}, & \text{if } d \in (4,\infty). \end{cases}$$

Theorem 2.1. *Assume* **(A)**. *Then for any $(k,p,q) \in (1,\infty] \times [1,\infty) \times (0,\infty)$, there exists a constant $c \in (0,\infty)$ such that*

$$\big(\mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)]\big)^{\frac{1}{q}} \le c\|h_\nu\|_{L^k(\mu)}^{\frac{1}{q}}\gamma_d(T), \quad T \ge 2,\ \nu = h_\nu\mu \in \mathscr{P} \text{ with } h_\nu \in L^k(\mu), \tag{2.9}$$

where $\nu = h_\nu\mu$ means $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = h_\nu$.

To show that $\gamma_d(T)$ is the convergence rate of $\big(\mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)]\big)^{\frac{1}{q}}$ uniformly in $\nu \in \mathscr{P}$, we restrict p to the set

$$I(d) := [1,\infty) \cap \left[1, \frac{2d}{(d-2)^+} \vee \frac{d(d-2)}{2}\right] = \begin{cases} [1,\infty), & \text{if } d \in [1,2], \\ [1, \frac{2d}{d-2}], & \text{if } d \in (2,4], \\ [1, \frac{d(d-2)}{2}], & \text{if } d \in (4,\infty). \end{cases}$$

Corollary 2.2. *Assume* **(A)** *with* (2.4) *strengthened as*

$$\sup_{x \in M} \mathbb{E}^x[\rho(\hat{X}_0,\hat{X}_t)^p] \le K(p)t^{\frac{p}{2}}, \quad p \in [1,\infty),\ t \in [0,1]. \tag{2.10}$$

Then for any $p \in I(d)$ and $q \in (0,\infty)$, there exists a constant $c \in (0,\infty)$ such that

$$\sup_{\nu \in \mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)] \le c\gamma_d(T)^q, \quad T \ge 2. \tag{2.11}$$

When $p \notin I(d)$, for any $q \in (0,\infty)$ and $\beta \in (0,1)$, there exists a constant $c \in (0,\infty)$ such that

$$\sup_{\nu \in \mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)] \le c\gamma_d(T)^{\beta q}, \quad T \ge 2. \tag{2.12}$$

Consequently, for any $p \in [1,\infty)$ and $q \in (0,\infty)$,

$$\limsup_{T\to\infty} \frac{1}{\log(T^{-1})} \log\left[\sup_{\nu \in \mathscr{P}} \big(\mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)]\big)^{\frac{1}{q}}\right] \le \frac{1}{2 \vee (d-2)}.$$

2.3 Lower bound estimate

To derive sharp lower bound estimates, we make the following assumption **(B)** which holds in particular for the (reflecting) diffusion operator $\hat{L} := \Delta + \nabla V$ on a d-dimensional compact connected Riemannian manifold, since in this case conditions (2.13) and (2.14) are well known, and the other conditions have been verified by [30, Lemma 5.2]. For M being a smooth domain in $\mathbb{R}^d$, (2.15) is known as Sard's lemma (see [13, p130, Excercise 5.5]) and has been discussed in [5, Section 3.1.6]. The function f_ξ in (2.16) is called Lusin's approximation of h (see [2, 19]).

(B) Let $\{\lambda_i\}_{i\geq 0}$ be all eigenvalues of $-\hat{L}$ listed in the increasing order with multiplicities. There exist constants $\kappa > 0$ and $d \in [1, \infty)$ such that

$$\lambda_i \leq \kappa\, i^{\frac{2}{d}}, \quad i \geq 0, \tag{2.13}$$

$$\mathbb{W}_1(\nu \hat{P}_t, \mu) \leq \kappa\, \mathbb{W}_1(\nu, \mu), \quad t \in [0,1],\, \nu \in \mathscr{P}. \tag{2.14}$$

Moreover, for any $f \in \mathscr{D}(\hat{\mathscr{E}})$,

$$\mu\big(\{|\nabla f| > 0, f = 0\}\big) = 0, \tag{2.15}$$

and there exists a constant $c > 0$ independent of f such that

$$\mu\big(f \neq f_\xi\big) \leq \frac{c}{\xi^2} \int_M |\nabla f|^2 \mathrm{d}\mu, \quad \xi > 0 \tag{2.16}$$

holds for a family of functions $\{f_\xi : \xi > 0\}$ on M with $\|\nabla f_\xi\|_\infty \leq \xi$.

Theorem 2.3. *Assume* **(A)** *and* **(B)***. Then for any $q \in (0, \infty)$, we have for large $T > 0$,*

$$\inf_{\nu \in \mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_1^q(\mu_T, \mu)] \succeq \gamma_d(T)^q.$$

2.4 Convergence of $T\mathbb{W}_2^2(\mu_T, \mu)$ in $L^q(\mathbb{P})$

As explained in [28] that the following assumption holds for diffusion processes with Bakry-Emery curvature bounded from below.

(C) (M, ρ) is a geodesic space, there exist constants $\theta \in (0,1]$, $K > 0$ and $m \geq 1$ such that

$$|\nabla \hat{P}_t \mathrm{e}^f|^2 \leq (\hat{P}_t \mathrm{e}^f) \hat{P}_t(|\nabla f|^2 \mathrm{e}^f) + K t^\theta \|\nabla f\|_\infty^2 (\hat{P}_t \mathrm{e}^{2mf})^{\frac{1}{m}}, \quad t \in [0,1],\ f \in C_{b,\mathrm{L}}(M),$$

and there exists a function $h \in C([0,1]; [1, \infty))$ with $h(0) = 1$ such that

$$\mathbb{W}_2^2(\nu \hat{P}_r, \mu) \leq h(r) \mathbb{W}_2^2(\nu, \mu), \quad \nu \in \mathscr{P},\, r \in [0,1]. \tag{2.17}$$

Moreover, (2.3) holds.

Let $\psi_i(T)$ and $\Xi(T)$ be in (1.3). We have the following result.

Theorem 2.4. *Assume that for any $p \in (1,\infty)$ there exists a constant $k(p) \in (0,\infty)$ such that*

$$|\nabla \hat P_t f| \le k(p)(\hat P_t|\nabla f|^p)^{\frac{1}{p}}, \quad t \in [0,1],\ f \in C_{b,\mathrm{L}}(M). \tag{2.18}$$

(1) *Assume* **(A)**. *If $d \in [1,4)$, then for any $(k,R) \in (1,\infty] \times [1,\infty)$,*

$$\lim_{T\to\infty} \sup_{\nu\in\mathscr{P}_{k,R}} \mathbb{E}^\nu\Big[|\{T\mathbb{W}_2^2(\mu_T,\mu) - \Xi(T)\}^+|^q\Big] = 0, \quad q \in [1,\infty). \tag{2.19}$$

Moreover, if $d \in [1,\frac{8}{3})$, then for any $q \in [1, \frac{d}{2(d-2)^+})$, where $\frac{d}{2(d-2)^+} := \infty$ for $d \le 2$,

$$\lim_{T\to\infty} \sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\Big[|\{T\mathbb{W}_2^2(\mu_T,\mu) - \Xi(T)\}^+|^q\Big] = 0. \tag{2.20}$$

(2) *Assume* **(A)** *and* **(C)**. *If $d \in [1,4)$, then for any $(k,R) \in (1,\infty] \times [1,\infty)$,*

$$\lim_{T\to\infty} \sup_{\nu\in\mathscr{P}_{k,R}} \mathbb{E}^\nu\Big[|\{T\mathbb{W}_2^2(\mu_T,\mu) - \Xi(T)\}^-|^q\Big] = 0 \tag{2.21}$$

holds for

$$q \in \begin{cases} [1,\infty), & \text{if } d \in [1,3],\\ \big[1, \frac{d-2}{2k^*(d-3)}\big), & \text{if } d \in (3,4), \end{cases}$$

where k^* is the Hölder conjugate of k. Moreover, if $d \in [1,\frac{8}{3})$, then for any $q \in [1,\frac{d}{2(d-2)^+})$,

$$\lim_{T\to\infty} \sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\Big[|\{T\mathbb{W}_2^2(\mu_T,\mu) - \Xi(T)\}^-|^q\Big] = 0. \tag{2.22}$$

To derive the uniform convergence uniformly in $\nu \in \mathscr{P}$ also for $d \in (2,4)$, we modify $\Xi(T)$ by the following defined $\bar\Xi_r(T)$ and $\tilde\Xi_r(T)$ for a constant $r \in (0,\infty)$:

$$\begin{aligned} &\bar\psi_{i,r}(T) := \frac{1}{\sqrt{T-r}} \int_r^T \phi_i(X_t)\mathrm{d}t, \quad \bar\Xi_r(T) := \sum_{i=1}^\infty \frac{|\bar\psi_{i,r}(T)|^2}{\lambda_i}, \quad T > r,\\ &\tilde\psi_{i,r}(T) := \frac{1}{\sqrt{T}} \int_0^T \phi_i(X_{r+t})\mathrm{d}t, \quad \tilde\Xi_r(T) := \sum_{i=1}^\infty \frac{|\tilde\psi_{i,r}(T)|^2}{\lambda_i}, \quad T > 0. \end{aligned} \tag{2.23}$$

Theorem 2.5. *In the situation of Theorem 2.4, let $r \in (0,\infty)$.*

(1) *Assume* **(A)**. *If $d \in [1,4)$, then for any $q \in [1,\infty)$,*

$$\lim_{T\to\infty} \sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\Big[|\{T\mathbb{W}_2^2(\mu_T,\mu) - \bar\Xi_r(T)\}^+|^q + |\{T\mathbb{W}_2^2(\mu_T,\mu) - \tilde\Xi_r(T)\}^+|^q\Big] = 0. \tag{2.24}$$

(2) *Assume* **(A)** *and* **(C)**. *If $d \in [1,4)$, then for any*

$$q \in \begin{cases} [1,\infty), & \text{if } d \in [1,3],\\ [1, \frac{d-2}{2(d-3)}), & \text{if } d \in (3,4), \end{cases}$$

$$\lim_{T\to\infty} \sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\Big[|\{T\mathbb{W}_2^2(\mu_T,\mu) - \bar\Xi_r(T)\}^-|^q + |\{T\mathbb{W}_2^2(\mu_T,\mu) - \tilde\Xi_r(T)\}^-|^q\Big] = 0. \tag{2.25}$$

As a consequence of Theorem 2.5, we have the following result on the weak limit of $T\mathbb{W}_2^2(\mu_T,\mu)$, which leads to the limits of moments.

Corollary 2.6. *Assume* **(A)** *and* **(C)** *for* $d\in[1,4)$*, let* $Z=0$ *and* (2.18) *holds. Then for any initial distribution* ν *of* X_t*, as* $T\to\infty$*,* $T\mathbb{W}_2^2(\mu_T,\mu)$ *converges weakly to the real random variable*

$$\Xi(\infty):=\sum_{i=1}^{\infty}\frac{2\xi_i^2}{\lambda_i^2},$$

where $\{\xi_i\}_{i\ge1}$ *are i.i.d. standard Normal random variables on* $\mathbb{R}$*. Consequently, for any*

$$q\in\begin{cases}(0,\infty), & \text{if } d\in[1,3],\\ (0,\frac{d-2}{2(d-3)}), & \text{if } d\in(3,4),\end{cases}$$

$$\lim_{T\to\infty}\mathbb{E}^\nu[T^q\mathbb{W}_2^{2q}(\mu_T,\mu)]=\mathbb{E}\big|\Xi(\infty)\big|^q<\infty. \tag{2.26}$$

Proof. By (2.7), we have $\sum_{i=1}^\infty\frac{1}{\lambda_i^2}<\infty$ for $d<4$. So, as shown in the proof of [30, Lemma 2.11], we may prove the weak convergence of $T\mathbb{W}_2^2(\mu_T,\mu)$ to $\Xi(\infty)$ using Theorem 2.5. By combining this with (6.10) below, we are able to apply the dominated convergence theorem to get (2.26). □

3 Preparations

We first introduce the following deviation inequality due to [32, Theorem 1].

Lemma 3.1. *Let* X_t *be an ergodic Markov process on a Polish space* E *with generator* L *and unique invariant probability measure* μ*, and let*

$$\mathscr{E}_{\rm sym}(f,g):=\frac{1}{2}\big[\langle -Lf,g\rangle_{L^2(\mu)}+\langle -Lg,f\rangle_{L^2(\mu)}\big],\quad f,g\in\mathscr{D}(L)$$

be the associated symmetrized Dirichlet form with domain $\mathscr{D}(\mathscr{E}_{\rm sym})\subset L^2(\mu)$*. For any* $\xi\in(0,\infty)$ *and* $g\in L^1(\mu)$ *with* $\mu(g)=0$*, let*

$$I_g(\xi-):=\lim_{\varepsilon\downarrow0}\inf\left\{\mathscr{E}_{\rm sym}(h,h):\ h\in\mathscr{D}(\mathscr{E}_{\rm sym}),\ \mu(h^2)=1,\ |\mu(gh^2)|=\xi-\varepsilon,\ \mu(|g|h^2)<\infty\right\},$$

where $\inf\emptyset=\infty$ *by convention. Then*

$$\mathbb{P}^\nu\left(\left|\frac{1}{T}\int_0^T g(X_t)\mathrm{d}t\right|>\xi\right)\le 2\,\|h_\nu\|_{L^2(\mu)}\exp\Big[-TI_g(\xi-)\Big],\quad T,\xi>0,\ \nu=h_\nu\mu\in\mathscr{P}.$$

We now apply this lemma to establish a Bernstein-type inequality for the local framework introduced in Subsection 2.1. For any $g\in L^2(\mu)$ with $\mu(g)=0$, let

$$\begin{aligned}&\sigma^2(g):=2\|(-\hat L)^{-\frac12}g\|_{L^2(\mu)}^2,\\ &\mathfrak{m}(g):=\begin{cases}\|(-\hat L)^{-\frac12}g\|_{L^2(\mu)}, & \text{if } d\in[1,2),\\ \inf_{2<p<\infty}\frac{p}{p-2}\|\nabla(-\hat L)^{-1}g\|_{L^p(\mu)}, & \text{if } d=2,\\ \|\nabla(-\hat L)^{-1}g\|_{L^d(\mu)}, & \text{if } d\in(2,\infty),\end{cases}\end{aligned}\tag{3.1}$$

In the spirit of [11, Theorem 2.2], we have the following consequence of Lemma 3.1.

Corollary 3.2 (Bernstein-type inequality)**.** *Consider the framework in Subsection 2.1 such that (2.3) holds. Then there exists a constant $\alpha > 0$ such that for any $g \in \mathscr{D}((-\hat{L})^{-\frac{1}{2}})$ with $\mu(g) = 0$, any $\nu \in \mathscr{P}$ such that $h_\nu := \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ exists,*

$$\begin{aligned}(3.2)\qquad &\mathbb{P}^\nu\left(\left|\frac{1}{T}\int_0^T g(X_t)\mathrm{d}t\right| > \xi\right)\\ &\le 2\,\|h_\nu\|_{L^2(\mu)}\exp\left[-\frac{T\xi^2}{2\sigma^2(g)\,+\,\alpha\mathfrak{m}(g)\xi}\right],\quad T,\xi>0.\end{aligned}$$

Consequently, for any $p \in [1,\infty)$ there exists a constant $C(p) > 0$ such that

$$\mathbb{E}^\nu\left[\left|\frac{1}{T}\int_0^T g(X_t)\mathrm{d}t\right|^p\right] \le C(p)\,\|h_\nu\|_{L^2(\mu)}\left(\sigma(g)T^{-\frac{1}{2}} + \mathfrak{m}(g)T^{-1}\right)^p. \tag{3.3}$$

Proof. (a) We first observe that (3.2) implies (3.3). Indeed, by taking

$$r = \frac{T\xi^2}{2\sigma^2(g) + \alpha\mathfrak{m}(g)\xi},$$

we have

$$\xi = \frac{r\alpha\mathfrak{m}(g) + \sqrt{[r\alpha\mathfrak{m}(g)]^2 + 8Tr\sigma^2(g)}}{2T}.$$

So, we find constants $c_1, c_2 > 0$ depending on p such that (3.2) yields

$$\begin{aligned}&\mathbb{E}^\nu\left[\left|\frac{1}{T}\int_0^T g(X_t)\mathrm{d}t\right|^p\right] = p\int_0^\infty \xi^{p-1}\mathbb{P}^\nu\left(\left|\frac{1}{T}\int_0^T g(X_t)\mathrm{d}t\right| > \xi\right)\mathrm{d}\xi\\ &\le c_1\,\|h_\nu\|_{L^2(\mu)}\int_0^\infty\Big(\frac{\mathfrak{m}(g)r}{T} + \frac{\sigma(g)\sqrt{r}}{\sqrt{T}}\Big)^{p-1}\Big(\frac{\mathfrak{m}(g)}{T} + \frac{\sigma(g)}{\sqrt{rT}}\Big)\mathrm{e}^{-r}\mathrm{d}r\\ &\le c_2\,\|h_\nu\|_{L^2(\mu)}\left(\frac{\sigma(g)}{\sqrt{T}} + \frac{\mathfrak{m}(g)}{T}\right)^p.\end{aligned}$$

Therefore, it remains to prove (3.2).

(b) According to (2.1), the symmetrized Dirichlet form $\mathscr{E}_{\rm sym}$ in Lemma 3.1 is exactly $\hat{\mathscr{E}}$ in our setting. By Lemma 3.1, (3.2) follows if we find a constant $\alpha > 0$ such that for any $h \in \mathscr{D}(\hat{\mathscr{E}})$ with $\mu(h^2) = 1$,

$$\frac{|\mu(gh^2)|^2}{2\sigma^2(g) + \alpha\mathfrak{m}(g)|\mu(gh^2)|} \le \hat{\mathscr{E}}(h,h),$$

and this inequality is implied by

$$|\mu(gh^2)| \le \sqrt{2\sigma^2(g)\hat{\mathscr{E}}(h,h)} + \frac{\alpha}{2}\mathfrak{m}(g)\hat{\mathscr{E}}(h,h). \tag{3.4}$$

Since $C_{b,\mathrm{L}}(M)$ is dense in $\mathscr{D}(\hat{\mathscr{E}})$, we may assume that $h \in C_{b,\mathrm{L}}(M)$. To verify this estimate, let $\hat{h} = h - \mu(h)$. Then $\mu(g) = 0$ implies

$$\mu(gh^2) = 2\mu(h)\mu(g\hat{h}) + \mu\big(g\hat{h}^2\big). \tag{3.5}$$

By $|\mu(h)| \le \mu(h^2)^{\frac{1}{2}} = 1$, and noting that the Cauchy-Schwarz inequality implies

$$|\mu(g\hat{h})| = |\mu\big(\{(-\hat{L})^{-\frac{1}{2}}g\} \cdot \{(-\hat{L})^{\frac{1}{2}}\hat{h}\}\big)| \le \sqrt{\frac{\sigma^2(g)\hat{\mathscr{E}}(h,h)}{2}},$$

we obtain

$$|2\mu(h)\mu(g\hat{h})| \le \sqrt{2\sigma^2(g)\hat{\mathscr{E}}(h,h)}. \tag{3.6}$$

By (3.5) and (3.6), (3.4) follows from

$$\big|\mu\big(g\hat{h}^2\big)\big| \le \frac{\alpha}{2}\mathfrak{m}(g)\hat{\mathscr{E}}(h,h). \tag{3.7}$$

Moreover, we have

$$(-\hat{L})^{-\alpha} = \frac{1}{\Gamma(\alpha)}\int_0^\infty t^{\alpha-1}\hat{P}_t\,\mathrm{d}t, \quad \alpha > 0, \tag{3.8}$$

where Γ is the Gamma function. With (2.6) and (3.8) in hands, we prove estimate (3.7) below for $d \in [1,2), d = 2$ and $d > 2$ respectively.

(c) Let $d \in [1,2)$. By (2.3) and (2.6) we find constants $c_1, c_2 > 0$ such that

$$\begin{aligned}
&\|\hat{h}\|_{L^\infty(\mu)} = \|(-\hat{L})^{-\frac{1}{2}}(-\hat{L})^{\frac{1}{2}}\hat{h}\|_{L^\infty(\mu)} \\
&= \frac{1}{\Gamma(\frac{1}{2})}\bigg\|\int_0^\infty t^{-\frac{1}{2}}\hat{P}_t(-\hat{L})^{\frac{1}{2}}\hat{h}\,\mathrm{d}t\bigg\|_{L^\infty(\mu)} \\
&\le c_1\int_0^\infty t^{-\frac{1}{2}}(1\wedge t)^{-\frac{d}{4}}\mathrm{e}^{-\lambda_1 t}\mathrm{d}t \cdot \|(-\hat{L})^{\frac{1}{2}}\hat{h}\|_{L^2(\mu)} = c_2\sqrt{\hat{\mathscr{E}}(h,h)}.
\end{aligned}$$

Combining this with the fundamental inequalities

$$\hat{\mathscr{E}}(f,g)^2 \le \hat{\mathscr{E}}(f,f)\hat{\mathscr{E}}(g,g), \quad \hat{\mathscr{E}}(f^2,f^2) \le 4\|f\|^2_{L^\infty(\mu)}\hat{\mathscr{E}}(f,f),$$

we derive

$$\begin{aligned}
&|\mu(g\hat{h}^2)| = \big|\hat{\mathscr{E}}\big((-\hat{L})^{-1}g, \hat{h}^2\big)\big| \\
&\le \sqrt{\hat{\mathscr{E}}\big((-\hat{L})^{-1}g, (-\hat{L})^{-1}g\big)\hat{\mathscr{E}}\big(\hat{h}^2, \hat{h}^2\big)} \\
&\le 2\|(-\hat{L})^{-\frac{1}{2}}g\|_{L^2(\mu)}\|\hat{h}\|_{L^\infty(\mu)}\sqrt{\hat{\mathscr{E}}(h,h)} \le 2c_2\mathfrak{m}(g)\hat{\mathscr{E}}(h,h),
\end{aligned}$$

so that (3.7) holds for $\alpha = 4c_2$.

(d) Let $d = 2$. By the Cauchy-Schwarz inequality, the chain rule and Hölder's inequality, we have for any $p \in (2,\infty)$,

$$\begin{aligned}
&|\mu(g\hat{h}^2)| = \big|\hat{\mathscr{E}}\big((-\hat{L})^{-1}g, \hat{h}^2\big)\big| \\
&\le \mu\left(\Gamma\big((-\hat{L})^{-1}g, (-\hat{L})^{-1}g\big)^{\frac{1}{2}} \cdot \Gamma(\hat{h}^2, \hat{h}^2)^{\frac{1}{2}}\right) \\
&\le 2\mu\left(\Gamma\big((-\hat{L})^{-1}g, (-\hat{L})^{-1}g\big)^{\frac{1}{2}} \cdot |\hat{h}|\Gamma(h,h)^{\frac{1}{2}}\right) \\
&\le 2\|\nabla(-\hat{L})^{-1}g\|_{L^p(\mu)}\|\hat{h}\|_{L^{\frac{2p}{p-2}}(\mu)}\sqrt{\hat{\mathscr{E}}(h,h)}.
\end{aligned} \tag{3.9}$$

Moreover, by (2.6) for $d = 2$, we find constants $c_3, c_4 > 0$ such that for any $p \in (2, \infty)$,

$$\begin{aligned} \|\hat{h}\|_{L^{\frac{2p}{p-2}}(\mu)} &= \|(-\hat{L})^{-\frac{1}{2}}(-\hat{L})^{\frac{1}{2}}\hat{h}\|_{L^{\frac{2p}{p-2}}(\mu)} = \frac{1}{\Gamma(\frac{1}{2})}\bigg\|\int_0^\infty t^{-\frac{1}{2}}\hat{P}_t(-\hat{L})^{\frac{1}{2}}\hat{h}\,\mathrm{d}t\bigg\|_{L^{\frac{2p}{p-2}}(\mu)} \\ &\le c_3\int_0^\infty t^{-\frac{1}{2}}(1\wedge t)^{-\frac{1}{p}}\mathrm{e}^{-\lambda_1 t}\mathrm{d}t\cdot\|(-\hat{L})^{\frac{1}{2}}\hat{h}\|_{L^2(\mu)} \le \frac{c_4 p}{p-2}\sqrt{\hat{\mathscr{E}}(h,h)}. \end{aligned}$$

Combining this with (3.9) and the definition of $\mathfrak{m}(g)$ for $d = 2$, we derive (3.7) for $\alpha = 4c_4$.

(e) Finally, let $d > 2$. It is well known that when $d > 2$, the Nash inequality (2.3) implies the Sobolev inequality (cf. [9])

$$\|f\|_{L^{\frac{2d}{d-2}}(\mu)} \le C\sqrt{\hat{\mathscr{E}}(f,f)},\quad f \in \mathscr{D}(\hat{\mathscr{E}}) \text{ with } \mu(f) = 0.$$

Combining this inequality with (3.9) for $p = d$ yields

$$|\mu(g\hat{h}^2)| \le 2\|\nabla(-\hat{L})^{-1}g\|_{L^d(\mu)}\|\hat{h}\|_{L^{\frac{2d}{d-2}}(\mu)}\sqrt{\hat{\mathscr{E}}(h,h)} \le 2C\|\nabla(-\hat{L})^{-1}g\|_{L^d(\mu)}\hat{\mathscr{E}}(h,h).$$

Then (3.7) holds for $\alpha = 4C$, which completes the proof. $\square$

Remark 3.3. *Note that* (2.3) *implies that for any* $1 < p < d$ *satisfying* $\frac{1}{q} \ge \frac{1}{p} - \frac{1}{d}$*, there exists a constant* $c > 0$ *such that (see [21])*

$$\|(-\hat{L})^{-\frac{1}{2}}f\|_{L^q(\mu)} \le C_{p,q}\|f\|_{L^p(\mu)},\quad f \in L^p(\mu) \text{ with } \mu(f) = 0.$$

This together with (2.5) *yields* $\mathfrak{m}(g) < \infty$ *for*

$$g \in \begin{cases} L^1(\mu), & \text{if } d \in [1,2); \\ L^{1+0}(\mu) := \bigcup_{p>1} L^p(\mu), & \text{if } d = 2; \\ L^{\frac{d}{2}}(\mu), & \text{if } d > 2. \end{cases}$$

Recall that for every non-zero function $\phi \in L^2(\mu)$ with $\mu(\phi) = 0$,

$$\mathbf{V}(\phi) := \int_0^\infty \mu(\phi P_t\phi)\,\mathrm{d}t \in (0,\infty). \tag{3.10}$$

We have the following lemma on the second moment of $\psi_i(T)$ defined in (1.3).

Lemma 3.4. *Consider the framework in Subsection 2.1 such that* (2.3) *holds. Then there exist constants* $c, c' > 0$ *such that*

$$\Big|\mathbb{E}^\mu\big[|\psi_i(T)|^2\big] - \frac{2}{\lambda_i} + \frac{2}{\lambda_i^2}\mathbf{V}(Z\phi_i)\Big| \le \frac{c}{\lambda_i(1+T)},\quad i \ge 1,\ T > 0, \tag{3.11}$$

$$\mathbf{V}(Z\phi_i) \le c'\sqrt{\lambda_i},\quad i \ge 1. \tag{3.12}$$

Proof. (a) By Lemma 4.2(1) in [28], we have

$$\mathbf{V}(\phi_i) = \frac{1}{\lambda_i} - \frac{1}{\lambda_i^2}\mathbf{V}(Z\phi_i).$$

This together with the Markov property and the stationarity of X_t yields

$$\begin{aligned} \mathbb{E}^\mu\big[|\psi_i(T)|^2\big] &= \frac{2}{T}\int_0^T \mathrm{d}t_1 \int_{t_1}^T \mu(\phi_i P_{t_2-t_1}\phi_i)\mathrm{d}t_2 \\ &= \frac{2}{T}\int_0^T \mathrm{d}t_1 \int_0^{T-t_1} \mu(\phi_i P_t\phi_i)\mathrm{d}t \\ &= 2\Big(\frac{1}{\lambda_i} - \frac{1}{\lambda_i^2}\mathbf{V}(Z\phi_i)\Big) - \frac{2}{T}\int_0^T \mathrm{d}t_1 \int_{T-t_1}^\infty \mu(\phi_i P_t\phi_i)\mathrm{d}t. \end{aligned} \tag{3.13}$$

We next evaluate the integrand $\mu(\phi_i P_t\phi_i)$. By Duhamel's formula (2.2) and $\hat{P}_t\phi_i = \mathrm{e}^{-\lambda_i t}\phi_i$, we have

$$P_t\phi_i = \mathrm{e}^{-\lambda_i t}\phi_i + \int_0^t \mathrm{e}^{-\lambda_i(t-s)}P_s(Z\phi_i)\mathrm{d}s,$$

so that

$$\mu(\phi_i P_t\phi_i) = \mathrm{e}^{-\lambda_i t} + \int_0^t \mathrm{e}^{-\lambda_i(t-s)}\mu\big((P_s^*\phi_i)Z\phi_i\big)\mathrm{d}s,$$

where P_t^* is the semigroup with generator $L^* := \hat{L} - Z$. Noting that $\mu(\phi_i Z\phi_i) = 0$ and using Duhamel's formula (2.2) again

$$P_t^*\phi_i = \mathrm{e}^{-\lambda_i t}\phi_i - \int_0^t \mathrm{e}^{-\lambda_i(t-s)}P_s^*(Z\phi_i)\mathrm{d}s,$$

we obtain

$$\mu\big((P_s^*\phi_i)Z\phi_i\big) = -\int_0^s \mathrm{e}^{-\lambda_i(s-r)}\mu\big((Z\phi_i)P_r(Z\phi_i)\big)\mathrm{d}r.$$

Combining above identities with the fact that

$$\big|\mu\big((Z\phi_i)P_r(Z\phi_i)\big)\big| \le \mathrm{e}^{-\lambda_1 r}\|Z\phi_i\|_{L^2(\mu)}^2 \le \|Z\|_\infty^2\lambda_i\mathrm{e}^{-\lambda_1 r},$$

we find a constant $C > 0$ such that

$$\big|\mu(\phi_i P_t\phi_i) - \mathrm{e}^{-\lambda_i t}\big| \le \|Z\|_\infty^2\lambda_i\int_0^t \mathrm{e}^{-\lambda_i(t-s)}\mathrm{d}s\int_0^s \mathrm{e}^{-\lambda_i(s-r)-\lambda_1 r}\mathrm{d}r \le C\lambda_i^{-1}\mathrm{e}^{-\frac{\lambda_1}{2}t}.$$

Combining with (3.13) we finish the proof of (3.11).

(b) By the spectral representation and that $\lambda\mathrm{e}^{-\lambda t} \le t^{-1}$ for $t, \lambda > 0$, we obtain

$$\|\nabla\hat{P}_t f\|_{L^2(\mu)}^2 = \sum_{i=1}^\infty \lambda_i\mathrm{e}^{-2\lambda_i t}\mu(f\phi_i)^2 \le t^{-1}\mathrm{e}^{-\lambda_1 t}\|f\|_{L^2(\mu)}^2, \quad f \in L^2(\mu).$$

Combining with the Duhamel's formula (2.2), we derive

$$\|\nabla P_t f\|_{L^2(\mu)} \le t^{-\frac{1}{2}} \mathrm{e}^{-\frac{\lambda_1}{2}t} \|f\|_{L^2(\mu)} + \|Z\|_\infty \int_0^t s^{-\frac{1}{2}} \mathrm{e}^{-\frac{\lambda_1}{2}s} \|\nabla P_{t-s} f\|_{L^2(\mu)} \mathrm{d}s, \quad t \ge 0.$$

By the generalized Gronwall inequality, see [33], we find constants $c_0, \lambda > 0$ such that

$$\|\nabla P_t f\|_{L^2(\mu)} \le c_0 t^{-\frac{1}{2}} \mathrm{e}^{-\lambda t} \|f\|_{L^2(\mu)}, \quad t > 0,\ f \in L^2(\mu).$$

Noting that $\|\nabla \phi_i\|_{L^2(\mu)}^2 = \lambda_i$, by this and (2.1), we find a constant $c_1 > 0$ such that

$$\big|\mu((Z\phi_i)P_t(Z\phi_i))\big| = \big|\mu(\phi_i Z P_t(Z\phi_i))\big| \le \|Z\|_\infty \|\nabla P_t(Z\phi_i)\|_{L^2(\mu)} \le c_1 t^{-\frac{1}{2}} \mathrm{e}^{-\lambda t} \sqrt{\lambda_i}.$$

Therefore, for some constant $c' > 0$,

$$\mathbf{V}(Z\phi_i) = \int_0^\infty \mu((Z\phi_i)P_t(Z\phi_i)) \mathrm{d}t \le c' \sqrt{\lambda_i}.$$

Then the proof is finished. □

Finally, we present the following Lemma 3.5 on the asymptotic of $\mathbb{W}_p(\cdot, \mu)$ near by the measure μ, by using the gradient estimate

$$\|\nabla \hat{P}_t f\|_{L^p(\mu)} \le c(p) t^{-\frac{1}{2}} \|f\|_{L^p(\mu)}, \quad t \in (0,1),\ f \in C_{b,\mathrm{L}}(M). \tag{3.14}$$

for $p \in (1,\infty)$ and some constant $c(p) > 0$. This result extends [12, Proposition 2.9] and [20, Proposition 2.5] for $p > \frac{d}{d-1}$ and $M = \mathbb{R}^d$. The estimate for $p \le \frac{d}{d-1}$ is new even in the Euclidean setting.

We would like to clarify the links of conditions (3.14), (2.5) and (2.18). Firstly, (3.14) is a necessary condition for the L^p-boundedness of the Riesz transform. Indeed, by the analyticity of $(\hat{P}_t)_{t>0}$ on $L^p(\mu)$, the L^p-boundedness of the Riesz transform implies that

$$\|\nabla \hat{P}_t f\|_{L^p(\mu)} \le C_p \|(-\hat{L})^{\frac{1}{2}} \hat{P}_t f\|_{L^p(\mu)} \le C_p' \|f\|_{L^p(\mu)}, \quad t > 0,\ f \in C_{b,\mathrm{L}}(M),$$

see [3] for more relationships between (3.14) and (2.5). Secondly, according to [28, Lemma 3.1(2)], (2.18) implies (3.14) for all $p \in (1,\infty)$.

Lemma 3.5. *Consider the framework in Subsection 2.1 and let $p \in [1,\infty)$. If $p \in (1,\infty)$, assume furthermore* (2.3)*,* (2.10) *and* (3.14)*. Then there exists a constant $c > 0$ depending on p and d such that for any $\theta \in (0,1)$,*

$$\sup_{\nu \in \mathscr{P}} \mathbb{W}_p\big((1-\theta)\mu + \theta\nu, \mu\big) \le c \cdot \begin{cases} \theta, & \text{if } 1 \le p < \frac{d}{d-1}, \\ \theta \log(1+\theta^{-1}), & \text{if } p = \frac{d}{d-1}, \\ \theta^{\frac{1}{p}+\frac{1}{d}}, & \text{if } p > \frac{d}{d-1}. \end{cases}$$

Proof. We first recall that for any probability density function $f \in L^2(\mu)$,

$$\mathbb{W}_2^2(f\mu,\mu) \le \int_M \frac{|\nabla \hat{L}^{-1}(f-1)|^2}{\mathscr{M}(f)}\,\mathrm{d}\mu, \quad \mathscr{M}(f) := 1_{\{f>0\}}\frac{f-1}{\log f}, \tag{3.15}$$

$$\mathbb{W}_p(f\mu,\mu) \le p\big\|\nabla \hat{L}^{-1}(f-1)\big\|_{L^p(\mu)}. \tag{3.16}$$

These estimates have been presented in [1] and [15] respectively by using the Kantorovich dual formula and Hamilton-Jacobi equations, which are available when (M,ρ) is a length space as we assumed, see [22].

By the convexity of $\mathbb{W}_p^p$, for any $\theta \in [0,1]$ and $\mu_0,\mu_1,\nu_0,\nu_1 \in \mathscr{P}$, we have

$$\mathbb{W}_p^p((1-\theta)\mu_0+\theta\mu_1,(1-\theta)\nu_0+\theta\nu_1) \le (1-\theta)\mathbb{W}_p^p(\mu_0,\nu_0)+\theta\mathbb{W}_p^p(\mu_1,\nu_1). \tag{3.17}$$

This implies the desired estimate for $p=1$. In the following we only consider $p \in (1,\infty)$.

By an approximation argument, we may assume that ν has a density function h_ν with respect to μ, so that

$$\nu_\varepsilon := \nu\hat{P}_\varepsilon = (\hat{P}_\varepsilon h_\nu)\mu, \quad \varepsilon \in (0,1).$$

Combining this with (2.10), we find a constant $c_1>0$ independent of ε and ν such that

$$\mathbb{W}_p(\nu_\varepsilon,\nu) \le c_1\sqrt{\varepsilon}.$$

Combining this with the triangle inequality, (3.16) and (3.17), we obtain

$$\begin{aligned}
&\mathbb{W}_p\big((1-\theta)\mu+\theta\nu,\mu\big) \le \mathbb{W}_p\big((1-\theta)\mu+\theta\nu,(1-\theta)\mu+\theta\nu_\varepsilon\big)+\mathbb{W}_p\big((1-\theta)\mu+\theta\nu_\varepsilon,\mu\big)\\
&\le c_1\theta^{\frac{1}{p}}\sqrt{\varepsilon}+p\theta\big\|\nabla(-\hat{L})^{-1}\hat{P}_\varepsilon(h_\nu-1)\big\|_{L^p(\mu)}, \quad \varepsilon,\theta \in (0,1).
\end{aligned} \tag{3.18}$$

To evaluate the term $\big\|\nabla(-\hat{L})^{-1}\hat{P}_\varepsilon(h_\nu-1)\big\|_{L^p(\mu)}$, we need the gradient bound (3.14). By combining (3.14) with (2.6) for $(p,1)$ in place of (p,q), and using the semigroup property of $\hat{P}_t$, we find a constant $C_p>0$ such that

$$\|\nabla\hat{P}_t f\|_{L^p(\mu)} \le C_p(1\wedge t)^{-\frac{d(p-1)}{2p}-\frac{1}{2}}\mathrm{e}^{-\lambda_1 t}\|f\|_{L^1(\mu)}, \quad t>0,\ f\in L^1(\mu). \tag{3.19}$$

This together with (3.8) for $\alpha=1$ yields that for some constants $c_2,c_3>0$,

$$\begin{aligned}
&\big\|\nabla(-\hat{L})^{-1}\hat{P}_\varepsilon(h_\nu-1)\big\|_{L^p(\mu)} \le \int_0^\infty \|\nabla\hat{P}_{t+\varepsilon}h_\nu\|_{L^p(\mu)}\mathrm{d}t\\
&\le c_2\int_0^\infty [1\wedge(t+\varepsilon)]^{-\frac{d(p-1)}{2p}-\frac{1}{2}}\mathrm{e}^{-\lambda_1 t}\mathrm{d}t\\
&\le c_3\begin{cases} 1, & \text{if } p<\frac{d}{d-1},\\ \log(1+\varepsilon^{-1}), & \text{if } p=\frac{d}{d-1},\\ \varepsilon^{\frac{1}{2}-\frac{d(p-1)}{2p}}, & \text{if } p>\frac{d}{d-1}.\end{cases}
\end{aligned} \tag{3.20}$$

Combining this with (3.18), for $\varepsilon=\theta^2$ when $p<\frac{d}{d-1}$, and for $\varepsilon=\theta^{\frac{2}{d}}$ when $p\ge\frac{d}{d-1}$, we derive the desired estimates. $\square$

4 Proofs of Theorem 2.1 and Corollary 2.2

Proof of Theorem 2.1. Let $\nu = h_\nu \mu$ with $h_\nu \in L^k(\mu)$ and $k \in (1, \infty]$. By the Markov property and Hölder's inequality, we obtain

$$\begin{aligned}
&\mathbb{E}^\nu\big[\mathbb{W}_p^q(\mu_T, \mu)\big] = \mathbb{E}^\mu\big[h_\nu(X_0)\mathbb{W}_p^q(\mu_T, \mu)\big] \\
&\le \big(\mathbb{E}^\mu[h_\nu^k]\big)^{\frac{1}{k}}\big(\mathbb{E}^\mu[\mathbb{W}_p^{k^*q}(\mu_T, \mu)]\big)^{\frac{1}{k^*}} = \|h_\nu\|_{L^k(\mu)}\big(\mathbb{E}^\mu[\mathbb{W}_p^{k^*q}(\mu_T, \mu)]\big)^{\frac{1}{k^*}},
\end{aligned}$$

where $k^* := \frac{k}{k-1}$ is the Hölder conjugate of k. Let $p' = p \vee (k^*q) \vee 2$. By Hölder's inequality and $\mathbb{W}_p \le \mathbb{W}_{p'}$, we derive

$$\big(\mathbb{E}^\mu[\mathbb{W}_p^{k^*q}(\mu_T, \mu)]\big)^{\frac{1}{k^*}} \le \big(\mathbb{E}^\mu[\mathbb{W}_{p'}^{p'}(\mu_T, \mu)]\big)^{\frac{q}{p'}}.$$

Therefore, it suffices to prove that for any $p \in [2, \infty)$,

$$\mathbb{E}^\mu\big[\mathbb{W}_p^p(\mu_T, \mu)\big] \preceq \gamma_d(T)^p, \quad T \ge 2. \tag{4.1}$$

By (3.16) and (2.5), we obtain

$$\mathbb{W}_p^p(f\mu, \mu) \le p^p\mu\big(|\nabla \hat{L}^{-1}(f-1)|^p\big) \le (pK(p))^p\mu\big(|(-\hat{L})^{-\frac{1}{2}}(f-1)|^p\big). \tag{4.2}$$

To prove the upper bound estimate (4.1) using (4.2), we approximate μ_T by the modified empirical measures

$$\mu_{T,\varepsilon} := \mu_T \hat{P}_\varepsilon, \quad T > 0,\ \varepsilon \in (0, 1].$$

The density function of $\mu_{T,\varepsilon}$ with respect to μ is given by

$$f_{T,\varepsilon}(y) := \frac{1}{T}\int_0^T \hat{p}_\varepsilon(X_t, y)\mathrm{d}t, \quad y \in M. \tag{4.3}$$

By the convexity of $\mathbb{W}_p^p$, the stationarity of X_t and (2.4), we obtain

$$\begin{aligned}
\mathbb{E}^\mu[\mathbb{W}_p^p(\mu_T, \mu_{T,\varepsilon})] &\le \frac{1}{T}\int_0^T \mathbb{E}^\mu[\mathbb{W}_p^p(\delta_{X_t}, \delta_{X_t}\hat{P}_\varepsilon)]\mathrm{d}t \\
&\le \int_{M\times M} \rho(x,y)^p \hat{p}_\varepsilon(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \preceq \varepsilon^{\frac{p}{2}}.
\end{aligned} \tag{4.4}$$

Combining this with (4.2), we derive

$$\begin{aligned}
\mathbb{E}^\mu[\mathbb{W}_p^p(\mu_T, \mu)] &\preceq \mathbb{E}^\mu[\mathbb{W}_p^p(\mu_T, \mu_{T,\varepsilon})] + \mathbb{E}^\mu[\mathbb{W}_p^p(\mu_{T,\varepsilon}, \mu)] \\
&\preceq \varepsilon^{\frac{p}{2}} + \mathbb{E}^\mu\big[\|(-\hat{L})^{-\frac{1}{2}}(f_{T,\varepsilon} - 1)\|_{L^p(\mu)}^p\big], \quad p \in [2, \infty).
\end{aligned} \tag{4.5}$$

Noting that

$$(-\hat{L})^{-\frac{1}{2}}(f_{T,\varepsilon} - 1)(y) = \frac{1}{T}\int_0^T (-\hat{L}_y)^{-\frac{1}{2}}(\hat{p}_\varepsilon(X_t, y) - 1)\mathrm{d}t,$$

where $\hat{L}_y$ is the operator $\hat{L}$ acting on the variable y, we deduce form Corollary 3.2 that

$$\mathbb{E}^\mu\left[\|(-\hat{L})^{-\frac{1}{2}}(f_{T,\varepsilon}-1)\|_{L^p(\mu)}^p\right] \preceq T^{-\frac{p}{2}}A_\varepsilon + T^{-p}B_\varepsilon, \quad T>0, \tag{4.6}$$

where

$$A_\varepsilon := \int_M \sigma\big((-\hat{L}_y)^{-\frac{1}{2}}(\hat{p}_\varepsilon(\cdot,y)-1)\big)^p \mu(\mathrm{d}y),$$

$$B_\varepsilon := \int_M \mathfrak{m}\big((-\hat{L}_y)^{-\frac{1}{2}}(\hat{p}_\varepsilon(\cdot,y)-1)\big)^p \mu(\mathrm{d}y).$$

Estimate on A_ε. By the definition of σ, (2.8), $-\hat{L}\phi_i = \lambda_i\phi_i$, and the fact that $\{\phi_i\}_{i\geq 1}$ is orthonormal in $L^2(\mu)$, we have

$$\begin{aligned}
&\sigma\big((-\hat{L}_y)^{-\frac{1}{2}}(\hat{p}_\varepsilon(\cdot,y)-1)\big) = \sqrt{2}\big\|(-\hat{L}_y)^{-1}(\hat{p}_\varepsilon(\cdot,y)-1)\big\|_{L^2(\mu)}\\
&= \left(2\int_M \left|(-\hat{L}_y)^{-1}\sum_{i=1}^\infty \mathrm{e}^{-\lambda_i\varepsilon}\phi_i(x)\phi_i(y)\right|^2 \mu(\mathrm{d}x)\right)^{\frac{1}{2}}\\
&= \left(2\sum_{i=1}^\infty \frac{\mathrm{e}^{-2\lambda_i\varepsilon}}{\lambda_i^2}\phi_i^2(y)\right)^{\frac{1}{2}}.
\end{aligned}\tag{4.7}$$

Combining this with the identity

$$\lambda^{-\alpha} = \frac{1}{\Gamma(\alpha)}\int_0^\infty \mathrm{e}^{-\lambda s}s^{\alpha-1}\mathrm{d}s, \quad \lambda,\alpha>0,$$

and applying (2.8), we obtain that for any $\varepsilon\in(0,1)$ and $y\in M$,

$$\sum_{i=1}^\infty \frac{\mathrm{e}^{-2\lambda_i\varepsilon}}{\lambda_i^2}\phi_i^2(y) = \frac{1}{\Gamma(2)}\int_0^\infty s(\hat{p}_{s+2\varepsilon}(y,y)-1)\mathrm{d}s.$$

By (2.6) for $p=1$ and $q=\infty$ we obtain

$$|\hat{p}_{s+2\varepsilon}(y,y)-1| \preceq [1\wedge(s+2\varepsilon)]^{-\frac{d}{2}}\mathrm{e}^{-\lambda_1 s}, \quad \varepsilon,s>0.$$

Therefore, for $\varepsilon\in(0,1]$,

$$\begin{aligned}
\sum_{i=1}^\infty \frac{\mathrm{e}^{-2\lambda_i\varepsilon}}{\lambda_i^2}\phi_i^2(y) &\preceq \int_0^\infty s[1\wedge(s+2\varepsilon)]^{-\frac{d}{2}}\mathrm{e}^{-\lambda_1 s}\mathrm{d}s\\
&\preceq 1+\int_0^1 s(s+2\varepsilon)^{-\frac{d}{2}}\mathrm{d}s \preceq \begin{cases} 1, & \text{if } d\in[1,4),\\ \log(1+\varepsilon^{-1}), & \text{if } d=4,\\ \varepsilon^{-\frac{d-4}{2}}, & \text{if } d\in(4,\infty).\end{cases}
\end{aligned}$$

Consequently, for $\varepsilon \in (0,1]$,

$$A_\varepsilon \preceq \begin{cases} 1, & \text{if } d \in [1,4), \\ [\log(1+\varepsilon^{-1})]^{\frac{p}{2}}, & \text{if } d = 4, \\ \varepsilon^{-\frac{(d-4)p}{4}}, & \text{if } d \in (4,\infty). \end{cases} \tag{4.8}$$

Estimate on B_ε. Let $d \in [1,2)$. For any $y \in M$, by the same reason leading to (4.7) and (4.8) for $d < 4$, we obtain that for $\varepsilon \in (0,1)$

$$\begin{aligned} B_\varepsilon &= \int_M \mathfrak{m}\big((-\hat{L}_y)^{-\frac{1}{2}}(\hat{p}_\varepsilon(\cdot,y)-1)\big)^p \mu(\mathrm{d}y) \\ &= \int_M \big\|(-\hat{L}_y)^{-1}(\hat{p}_\varepsilon(\cdot,y)-1)\big\|_{L^2(\mu)}^p \mu(\mathrm{d}y) \preceq 1. \end{aligned}$$

Let $d = 2$. By taking $p = 4$ in the definition of $\mathfrak{m}(\cdot)$, applying (2.5) for $p = 4$, (2.6) for $p = 1$ and $q = 4$, and using (3.8) for $\alpha = 1$, we obtain that for $\varepsilon \in (0,1)$,

$$\begin{aligned} B_\varepsilon &= \int_M \mathfrak{m}\big((-\hat{L}_y)^{-\frac{1}{2}}(\hat{p}_\varepsilon(\cdot,y)-1)\big)^p \mu(\mathrm{d}y) \\ &\preceq \sup_{y\in M} \big\|(-\hat{L}_y)^{-1}(\hat{p}_\varepsilon(\cdot,y)-1)\big\|_{L^4(\mu)}^p \\ &\le \sup_{y\in M} \bigg(\int_0^\infty \|\hat{p}_{t+\varepsilon}(\cdot,y)-1\|_{L^4(\mu)}\mathrm{d}t\bigg)^p \\ &\preceq \bigg(\int_0^\infty (1\wedge t)^{-\frac{3d}{8}}\mathrm{e}^{-\lambda_1 t}\mathrm{d}t\bigg)^p = \bigg(\int_0^\infty (1\wedge t)^{-\frac{3}{4}}\mathrm{e}^{-\lambda_1 t}\mathrm{d}t\bigg)^p < \infty. \end{aligned}$$

Let $d > 2$. By the same argument used above, we have

$$\begin{aligned} B_\varepsilon &\preceq \sup_{y\in M} \big\|(-\hat{L}_y)^{-1}(\hat{p}_\varepsilon(\cdot,y)-1)\big\|_{L^d(\mu)}^p \le \sup_{y\in M} \bigg(\int_0^\infty \|\hat{p}_{t+\varepsilon}(\cdot,y)-1\|_{L^d(\mu)}\mathrm{d}t\bigg)^p \\ &\preceq \bigg(\int_0^\infty [1\wedge(t+\varepsilon)]^{-\frac{d-1}{2}}\mathrm{e}^{-\lambda_1 t}\mathrm{d}t\bigg)^p \preceq \begin{cases} 1, & \text{if } d \in (2,3), \\ \log^p(1+\varepsilon^{-1}), & \text{if } d = 3, \\ \varepsilon^{-\frac{(d-3)p}{2}}, & \text{if } d \in (3,\infty). \end{cases} \end{aligned}$$

So, in conclusion,

$$B_\varepsilon \preceq \begin{cases} 1, & \text{if } d \in [1,3), \\ \log^p(1+\varepsilon^{-1}), & \text{if } d = 3, \\ \varepsilon^{-\frac{(d-3)p}{2}}, & \text{if } d \in (3,\infty). \end{cases}$$

Combining this with (4.6) and (4.8), we derive that for $T \ge 2$ and $\varepsilon \in (0,1)$,

$$\mathbb{E}^\mu\big[\|(-\hat{L})^{-\frac{1}{2}}(f_{T,\varepsilon}-1)\|_{L^p(\mu)}^p\big] \preceq \begin{cases} T^{-\frac{p}{2}} + T^{-p}, & \text{if } d \in [1,3), \\ T^{-\frac{p}{2}} + T^{-p}\log^p(1+\varepsilon^{-1}), & \text{if } d = 3, \\ T^{-\frac{p}{2}} + T^{-p}\varepsilon^{-\frac{(d-3)p}{2}}, & \text{if } d \in (3,4), \\ T^{-\frac{p}{2}}\log^{\frac{p}{2}}(1+\varepsilon^{-1}) + T^{-p}\varepsilon^{-\frac{p}{2}}, & \text{if } d = 4, \\ T^{-\frac{p}{2}}\varepsilon^{-\frac{(d-4)p}{4}} + T^{-p}\varepsilon^{-\frac{(d-3)p}{2}}, & \text{if } d \in (4,\infty). \end{cases} \tag{4.9}$$

By taking in this estimate

$$\varepsilon = \begin{cases} T^{-1}, & \text{if } d \in [1,4), \\ T^{-\frac{2}{d-2}}, & \text{if } d \in [4,\infty), \end{cases}$$

and combining with (4.5), we derive (4.1). $\square$

Proof of Corollary 2.2. (a) By the monotonicity of $\mathbb{W}_p$ in p, it suffices to prove (2.11) for $p \in I(d) \cap [2,\infty)$. Let

$$\bar{\mu}_T := \frac{1}{T-1}\int_1^T \delta_{X_t}\mathrm{d}t, \quad T > 1.$$

For any $\nu \in \mathscr{P}$, let $\nu_1 = \nu P_1^*$. Then (2.6) implies

$$\sup_{\nu\in\mathscr{P}} \frac{\mathrm{d}\nu_1}{\mathrm{d}\mu} \le c_0 \tag{4.10}$$

for some constant $c_0 > 0$, see [28, Lemma 3.1]. Hence, by the Markov property and (2.9) for $k = \infty$, we find a constant $c_1 > 0$ such that for any $T \ge 2$,

$$\sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_p^q(\bar{\mu}_T,\mu)] = \sup_{\nu\in\mathscr{P}} \mathbb{E}^{\nu_1}[\mathbb{W}_p^q(\mu_{T-1},\mu)] \le c_1\gamma_d(T)^q. \tag{4.11}$$

Noting that noting that $\mu_T = (1-T^{-1})\bar{\mu}_T + T^{-1}\mu_1$, by the triangle inequality, (3.17) and Lemma 3.5, we obtain

$$\begin{aligned} \mathbb{W}_p(\mu_T,\mu) &\le \mathbb{W}_p\big((1-T^{-1})\mu + T^{-1}\mu_1, \mu_T\big) + \mathbb{W}_p\big((1-T^{-1})\mu + T^{-1}\mu_1,\mu\big) \\ &\le \mathbb{W}_p(\bar{\mu}_T,\mu) + c_2\gamma_{d,p}(T) \end{aligned}$$

for some constant $c_2 > 0$, where

$$\gamma_{d,p}(T) := \begin{cases} T^{-1}, & \text{if } p \in [1,\frac{d}{d-1}), \\ T^{-1}\log T, & \text{if } p = \frac{d}{d-1}, \\ T^{-(\frac{1}{p}+\frac{1}{d})}, & \text{if } p \in (\frac{d}{d-1},\infty). \end{cases}$$

Combining this with (4.11), we obtain

$$\sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)] \preceq \big(\gamma_d(T) + \gamma_{d,p}(T)\big)^q, \quad T \ge 2. \tag{4.12}$$

Noting that $\gamma_{d,p}(T) \preceq \gamma_d(T)$ holds for $p \in I(d)$ and $T \ge 2$, we derive (2.11) by (4.12).

(b) Now, let $p \notin I(d)$. For any $s \in (0,1)$, let $\nu_s = \nu P_s^*$. By [28, Lemma 3.1], (2.6) implies that for a constant $c_1 > 0$

$$\|P_t\|_{L^p(\mu)\to L^q(\mu)} \le c_1 t^{-\frac{d}{2}(p^{-1}-q^{-1})}, \quad t \in (0,1),\ 1 \le p \le q \le \infty. \tag{4.13}$$

Then

$$\sup_{\nu\in\mathscr{P}} \Big\|\frac{\mathrm{d}\nu_s}{\mathrm{d}\mu}\Big\|_{L^k(\mu)} \le \|P_s\|_{L^{\frac{k}{k-1}}(\mu)\to L^\infty(\mu)} \le c_1 s^{-\frac{d}{2}(1-k^{-1})}, \quad s \in (0,1),\ k \in [1,\infty]. \tag{4.14}$$

For $s \in (0,1)$, let

$$\tilde{\mu}_{T,s} := \frac{1}{T}\int_s^{T+s} \delta_{X_t} \mathrm{d}t. \tag{4.15}$$

By (2.9), there exists a constant $c_2 > 0$ (depending on k, p and q) such that

$$\sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_p^q(\tilde{\mu}_{T,s},\mu)] \le c_2 s^{-\frac{d(k-1)}{2k}} \gamma_d(T)^q, \quad s \in (0,1),\, T \ge 2. \tag{4.16}$$

On the other hand, (2.3) implies that $D := \sup\rho < \infty$, and it is easy to see that for $T > s$,

$$\Pi := \frac{1}{T}\int_0^s \big(\delta_{X_t} \times \delta_{X_{t+T}}\big)\mathrm{d}t + \frac{1}{T}\int_s^T \big(\delta_{X_t} \times \delta_{X_t}\big)\mathrm{d}t \in \mathscr{C}(\mu_T, \tilde{\mu}_{T,s}).$$

Therefore for any $p \in [1,\infty)$,

$$\mathbb{W}_p^p(\mu_T, \tilde{\mu}_{T,s}) \le \int_{M\times M} \rho(x,y)^p\, \Pi(\mathrm{d}x,\mathrm{d}y) \le \frac{sD^p}{T}, \quad T > s. \tag{4.17}$$

Combining this with (4.16) and using the triangle inequality, we find a constant $c_3 > 0$ such that

$$\begin{aligned}\sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)] &\le 2^q\big(c_2 s^{-\frac{d(k-1)}{2k}}\gamma_d(T)^q + D^q T^{-\frac{q}{p}} s^{\frac{q}{p}}\big)\\ &\le c_3\big(s^{-\frac{d(k-1)}{2kq}}\gamma_d(T) + \gamma_d(T)^{\frac{2\vee(d-2)}{p}} s^{\frac{1}{p}}\big)^q, \quad s \in (0,1),\, T \ge 2,\end{aligned}$$

where $k \in (1,\infty)$ to be chosen later. Since for $p \notin I(d)$ we have $\frac{2\vee(d-2)}{p} < 1$, taking $s = c_4\gamma_d(T)^{\frac{2kq[p-2\vee(d-2)]}{2kq+d(k-1)p}}$ for a sufficiently small $c_4 > 0$ yields

$$\sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_p^q(\mu_T,\mu)] \preceq \gamma_d(T)^{\frac{2kq+d(k-1)[2\vee(d-2)]}{2kq+d(k-1)p}\cdot q}, \quad T \ge 2.$$

Then the proof is finished by choosing k sufficiently close to 1, such that $\beta \le \frac{2kq+d(k-1)[2\vee(d-2)]}{2kq+d(k-1)p}$. $\square$

5 Proof of Theorem 2.3

We first deal with the stationary case with initial distribution μ. Namely, we intend to prove that for any $q \in (0,\infty)$,

$$\mathbb{E}^\mu[\mathbb{W}_1^q(\mu_T,\mu)] \succeq \gamma_d(T)^q \tag{5.1}$$

holds for large $T > 0$. The proof is organized into steps.

(a) By Hölder's inequality, if $q \in [1,\infty)$,

$$\mathbb{E}^\mu[\mathbb{W}_1^q(\mu_T,\mu)] \ge \mathbb{E}^\mu[\mathbb{W}_1(\mu_T,\mu)]^q;$$

and if $q \in (0,1)$,

$$\begin{aligned}
&\mathbb{E}^\mu[\mathbb{W}_1^q(\mu_T,\mu)] \geq \mathbb{E}^\mu[\mathbb{W}_1^2(\mu_T,\mu)]^{-(1-q)} \cdot \mathbb{E}^\mu[\mathbb{W}_1(\mu_T,\mu)]^{2-q} \\
&\succeq \gamma_d(T)^{-2(1-q)} \cdot \mathbb{E}^\mu[\mathbb{W}_1(\mu_T,\mu)]^{2-q}, \quad T \geq 2,
\end{aligned}$$

where we also apply Theorem 2.1 for $p=1$ and $q=2$. Therefore, we only need to prove (5.1) for $q=1$. Recall that $\mu_{T,\varepsilon} := \mu_T \hat{P}_\varepsilon = f_{T,\varepsilon}\mu$, where $f_{T,\varepsilon}$ is defined in (4.3). By (2.14), we have

$$\mathbb{W}_1(\mu_T,\mu) \geq \kappa^{-1}\mathbb{W}_1(\mu_{T,\varepsilon},\mu), \quad T>0,\, \varepsilon \in (0,1].$$

So, it suffices to show that for large $T>0$,

$$\sup_{\varepsilon\in(0,1]} \mathbb{E}^\mu[\mathbb{W}_1(\mu_{T,\varepsilon},\mu)] \succeq \gamma_d(T). \tag{5.2}$$

Let $g_{T,\varepsilon} := (-\hat{L})^{-1}(f_{T,\varepsilon}-1)$. By (2.16), for any $\xi>0$, there exists a ξ-Lipschitz function g_ξ such that

$$E_\xi := \{x \in M : g_{T,\varepsilon}(x) \neq g_\xi(x)\}$$

satisfies

$$\mu(E_\xi) \leq \frac{C}{\xi^2}\int_M |\nabla g_{T,\varepsilon}|^2 \mathrm{d}\mu. \tag{5.3}$$

Since $\|\nabla g_\xi\|_\infty \leq \xi$, Kantorovich's duality implies

$$\mathbb{W}_1(\mu_{T,\varepsilon},\mu) \geq \xi^{-1}|\mu_{T,\varepsilon}(g_\xi)-\mu(g_\xi)| = \xi^{-1}\left|\int_M g_\xi(f_{T,\varepsilon}-1)\mathrm{d}\mu\right| = \xi^{-1}\left|\hat{\mathscr{E}}(g_\xi, g_{T,\varepsilon})\right|.$$

By (2.15), we have μ-a.e. $1_{E_\xi^c}|\nabla(g_\xi - g_{T,\varepsilon})| = 0$, so that

$$\begin{aligned}
\mathbb{W}_1(\mu_{T,\varepsilon},\mu) &\geq \xi^{-1}\left|\hat{\mathscr{E}}(g_{T,\varepsilon},g_{T,\varepsilon}) - \hat{\mathscr{E}}(g_{T,\varepsilon},g_{T,\varepsilon}-g_\xi)\right| \\
&\geq \xi^{-1}\Big[\hat{\mathscr{E}}(g_{T,\varepsilon},g_{T,\varepsilon}) - \mu\big(1_{E_\xi}|\nabla g_{T,\varepsilon}|(|\nabla g_{T,\varepsilon}|+|\nabla g_\xi|)\big)\Big].
\end{aligned} \tag{5.4}$$

Moreover, by $\|\nabla g_\xi\|_\infty \leq \xi$, Hölder's inequality and (5.3), we find a constant $c_1>0$ such that

$$\begin{aligned}
&\mu\big(1_{E_\xi}|\nabla g_{T,\varepsilon}|(|\nabla g_{T,\varepsilon}|+|\nabla g_\xi|)\big) \leq \mu(1_{E_\xi}|\nabla g_{T,\varepsilon}|^2) + \xi\mu(1_{E_\xi}|\nabla g_{T,\varepsilon}|) \\
&\leq \mu(E_\xi)^{\frac14}\|\nabla g_{T,\varepsilon}\|_{L^4(\mu)}\|\nabla g_{T,\varepsilon}\|_{L^2(\mu)} + \xi\mu(E_\xi)^{\frac34}\|\nabla g_{T,\varepsilon}\|_{L^4(\mu)} \\
&\leq c_1\xi^{-\frac12}\|\nabla g_{T,\varepsilon}\|_{L^4(\mu)}\|\nabla g_{T,\varepsilon}\|_{L^2(\mu)}^{\frac32}.
\end{aligned}$$

This together with (5.4) yields

$$\mathbb{W}_1(\mu_{T,\varepsilon},\mu) \geq \xi^{-1}\|\nabla g_{T,\varepsilon}\|_{L^2(\mu)}^2 - c_1\xi^{-\frac32}\|\nabla g_{T,\varepsilon}\|_{L^4(\mu)}\|\nabla g_{T,\varepsilon}\|_{L^2(\mu)}^{\frac32}.$$

Taking expectation and using Hölder's inequality again, we obtain

$$\begin{aligned}
\mathbb{E}^\mu[\mathbb{W}_1(\mu_{T,\varepsilon},\mu)] &\geq \xi^{-1}\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|_{L^2(\mu)}^2\big] \\
&\quad - c_1\xi^{-\frac32}\big(\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|_{L^2(\mu)}^2\big]\big)^{\frac34}\big(\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|_{L^4(\mu)}^4\big]\big)^{\frac14}.
\end{aligned} \tag{5.5}$$

(b) To estimate $\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|^2_{L^2(\mu)}\big]$, we formulate $g_{T,\varepsilon}$ as

$$g_{T,\varepsilon} = (-\hat{L})^{-1}(f_{T,\varepsilon} - 1) = \frac{1}{\sqrt{T}} \sum_{i=1}^\infty \frac{\mathrm{e}^{-\varepsilon\lambda_i}}{\lambda_i} \psi_i(T)\phi_i,$$

where $\psi_i(T)$ is defined in (1.3). Then

$$\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|^2_{L^2(\mu)}\big] = \frac{1}{T}\sum_{i=1}^\infty \frac{\mathrm{e}^{-2\lambda_i\varepsilon}}{\lambda_i}\mathbb{E}^\mu[|\psi_i(T)|^2]. \tag{5.6}$$

Using Lemma 3.4, and choosing $m \geq 1$ such that $\lambda_m > 4c'^2$, we obtain

$$\mathbb{E}^\mu[|\psi_i(T)|^2] \geq \frac{2}{\lambda_i}\big(1 - c'\lambda_i^{-\frac{1}{2}}\big) - \frac{c}{\lambda_i T} \geq \frac{1}{\lambda_i} - \frac{c}{\lambda_i T} \geq \frac{1}{2\lambda_i}, \quad T \geq 2c,\, i \geq m.$$

Hence combining this with (5.6) and (2.13) yields

$$\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|^2_{L^2(\mu)}\big] \geq \frac{1}{2T}\sum_{i=m}^\infty \frac{\mathrm{e}^{-2\lambda_i\varepsilon}}{\lambda_i^2} \geq \frac{1}{2\kappa^2 T}\sum_{i=m}^\infty \frac{\mathrm{e}^{-2\kappa i^{\frac{2}{d}}\varepsilon}}{i^{\frac{4}{d}}}, \quad T \geq 2c. \tag{5.7}$$

By choosing

$$\varepsilon = \begin{cases} T^{-1}, & \text{if } d \in [1,4), \\ T^{-\frac{2}{d-2}}, & \text{if } d \in [4,\infty), \end{cases} \tag{5.8}$$

we derive that for large $T > 0$,

$$\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|^2_{L^2(\mu)}\big] \succeq \gamma_d(T)^2 = \begin{cases} T^{-1}, & \text{if } d \in [1,4), \\ T^{-1}\log T, & \text{if } d = 4, \\ T^{-\frac{2}{d-2}}, & \text{if } d \in (4,\infty). \end{cases} \tag{5.9}$$

(c) Noting that $g_{T,\varepsilon} = (-\hat{L})^{-1}(f_{T,\varepsilon} - 1)$, we have

$$\|\nabla g_{T,\varepsilon}\|_{L^2(\mu)} = \|\nabla(-\hat{L})^{-1}(f_{T,\varepsilon} - 1)\|_{L^2(\mu)} = \|(-\hat{L})^{-\frac{1}{2}}(f_{T,\varepsilon} - 1)\|_{L^2(\mu)}. \tag{5.10}$$

Next, by (2.5) for $p = 4$, we find a constant $c > 0$ such that

$$\|\nabla g_{T,\varepsilon}\|^4_{L^4(\mu)} \leq c\|(-\hat{L})^{-\frac{1}{2}}(f_{T,\varepsilon} - 1)\|^4_{L^4(\mu)}. \tag{5.11}$$

By combining (4.9) with (5.10) and (5.11), and using the choice of ε in (5.8), we obtain that for large $T > 0$,

$$\big(\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|^2_{L^2(\mu)}\big]\big)^{\frac{3}{4}}\big(\mathbb{E}^\mu\big[\|\nabla g_{T,\varepsilon}\|^4_{L^4(\mu)}\big]\big)^{\frac{1}{4}} \preceq \gamma_d(T)^{\frac{5}{2}} = \begin{cases} T^{-\frac{5}{4}}, & \text{if } d \in [1,4), \\ T^{-\frac{5}{4}}(\log T)^{\frac{5}{4}}, & \text{if } d = 4, \\ T^{-\frac{5}{2(d-2)}}, & \text{if } d \in (4,\infty). \end{cases} \tag{5.12}$$

Combining (5.5), (5.9) and (5.12), and choosing $\xi = K\gamma_d(T)$ for a large enough constant $K > 0$, we obtain

$$\mathbb{E}^\mu[\mathbb{W}_1(\mu_T, \mu)] \succeq \gamma_d(T),$$

which implies (5.1) for $q \geq 1$, and hence also for all $q \in (0, \infty)$ explained above.

We then aim to strengthen (5.1) uniformly in initial distribution. We first observe that for some $s > 0$,

$$\inf_{x,y \in M} p_s(x, y) \geq \frac{1}{2}. \tag{5.13}$$

According to [23, Theorem 3.3.14], (2.3) implies

$$\|P_t\|_{L^1(\mu)\to L^\infty(\mu)} < \infty, \quad t > 0.$$

Moreover, (2.3) also implies the Poincaré inequality

$$\mu(f^2) \leq \frac{1}{\lambda_1}\hat{\mathscr{E}}(f, f), \quad f \in \mathscr{D}_0,$$

where $\lambda_1 > 0$ is the spectral gap of $\hat{L}$. Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\|P_t f\|^2_{L^2(\mu)} = 2\langle P_t f, LP_t f\rangle_{L^2(\mu)} = -2\hat{\mathscr{E}}(P_t f, P_t f),$$

the Poincaré inequality implies

$$\|P_t - \mu\|_{L^2(\mu)} \leq \mathrm{e}^{-\lambda_1 t}, \quad t \geq 0.$$

Therefore, by the semigroup property, there exists a constant $c > 0$ such that for any $t \geq 2$,

$$\begin{aligned}
&\sup_{x,y\in M} |p_t(x, y) - 1| = \|P_t - \mu\|_{L^1(\mu)\to L^\infty(\mu)} \\
&\leq \|P_1 - \mu\|_{L^1(\mu)\to L^2(\mu)}\|P_{t-2} - \mu\|_{L^2(\mu)}\|P_1 - \mu\|_{L^2(\mu)\to L^\infty(\mu)} \leq c\mathrm{e}^{-\lambda_1 t}.
\end{aligned}$$

Thus, for large enough $s > 0$ such that $c\mathrm{e}^{-\lambda_1 s} \leq \frac{1}{2}$, we have

$$\inf_{x,y\in M} p_s(x, y) \geq 1 - \sup_{x,y\in M} |p_t(x, y) - 1| \geq 1 - \frac{1}{2} \geq \frac{1}{2}.$$

Hence, (5.13) holds. Consequently, for any $\nu \in \mathscr{P}$, we have

$$\nu_s := \nu P_s^* \geq \frac{1}{2}\mu. \tag{5.14}$$

By (5.1), (5.14) and the Markov property, we deduce that $\tilde{\mu}_{T,s}$, which is given in (4.15), satisfies

$$\inf_{\nu\in\mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_1^q(\tilde{\mu}_{T,s}, \mu)] = \inf_{\nu\in\mathscr{P}} \mathbb{E}^{\nu_s}[\mathbb{W}_1^q(\mu_T, \mu)] \geq \frac{1}{2}\mathbb{E}^\mu[\mathbb{W}_1^q(\mu_T, \mu)] \succeq \gamma_d(T)^q \tag{5.15}$$

for large $T > 0$. On the other hand, (4.17) implies that for $D := \sup \rho < \infty$,

$$\mathbb{W}_1(\mu_T, \tilde{\mu}_{T,s}) \le \frac{sD}{T}.$$

Combining this with (5.15) and the triangle inequality, we conclude that for large $T > 0$,

$$\begin{aligned}&\inf_{\nu\in\mathscr{P}} \mathbb{E}^\nu[\mathbb{W}_1^q(\mu_T,\mu)] \ge \inf_{\nu\in\mathscr{P}} \mathbb{E}^\nu\big[2^{-q}\mathbb{W}_1^q(\tilde{\mu}_{T,s},\mu) - \mathbb{W}_1^q(\mu_T,\tilde{\mu}_{T,s})\big]\\ &\succeq \gamma_d(T)^q - T^{-q} \asymp \gamma_d(T)^q.\end{aligned}$$

Then the proof is finished.

6 Proofs of Theorem 2.4 and Theorem 2.5

We first prove some lemmas.

Lemma 6.1. *Assume* **(A)** *and let* $d \in [1,4)$. *Then*

$$\lim_{T\to\infty} \mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_T,\mu) - \Xi(T)\}^+\big|^q\Big] = 0, \quad q \in [1,\infty). \tag{6.1}$$

Proof. For $T \ge 2$ and $\varepsilon \in (0,1)$, let $\psi_i(T)$ and $f_{T,\varepsilon}$ be defined in (2.8) and (4.3) respectively. Then

$$\begin{aligned}&\Xi_\varepsilon(T) := T\|\nabla(-\hat{L})^{-1}(f_{T,\varepsilon}-1)\|_{L^2(\mu)}^2 = \sum_{i=1}^\infty \frac{|\psi_i(T)|^2\mathrm{e}^{-2\varepsilon\lambda_i}}{\lambda_i},\\ &\mu_{T,\varepsilon,\varepsilon} := (1-\varepsilon)\mu_{T,\varepsilon} + \varepsilon\mu = [(1-\varepsilon)f_{T,\varepsilon}+\varepsilon]\mu \in \mathscr{P}.\end{aligned} \tag{6.2}$$

By (3.15), we have

$$\mathbb{W}_2^2(\mu_{T,\varepsilon,\varepsilon},\mu) \le \mu\Big(\big|\nabla(-\hat{L})^{-1}(f_{T,\varepsilon}-1)\big|^2\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}\Big). \tag{6.3}$$

For $\varepsilon \in (0,1)$, we define

$$\mu_{T,\varepsilon,\varepsilon} := (1-\varepsilon)\mu_{T,\varepsilon} + \varepsilon\mu \in \mathscr{P},$$

whose density function with respect to μ is given by $(1-\varepsilon)f_{T,\varepsilon}+\varepsilon$.

By (2.5) and (4.9), we obtain for any $q \in [1,\infty)$, $T \ge 2$ and $\varepsilon \in (0,1)$,

$$\begin{aligned}&T^q\mathbb{E}^\mu\big[\|\nabla(-\hat{L})^{-1}(f_{T,\varepsilon}-1)\|_{L^{2q}(\mu)}^{2q}\big] \preceq T^q\mathbb{E}^\mu\big[\|(-\hat{L})^{-\frac{1}{2}}(f_{T,\varepsilon}-1)\|_{L^{2q}(\mu)}^{2q}\big]\\ &\preceq \begin{cases} 1, & \text{if } d \in [1,3);\\ 1 + T^{-q}\log^{2q}(1+\varepsilon^{-1}), & \text{if } d = 3;\\ 1 + T^{-q}\varepsilon^{-(d-3)q}, & \text{if } d \in (3,4).\end{cases}\end{aligned}$$

From now on, we choose

$$\varepsilon = \varepsilon(T) = T^{-\zeta}, \tag{6.4}$$

for some $\zeta \in (1, \frac{2}{(d-2)^+})$, so that the above estimate reduces to

$$\sup_{T\geq 2} T^q \mathbb{E}^\mu\big[\|\nabla(-\hat{L})^{-1}(f_{T,\varepsilon}-1)\|_{L^{2q}(\mu)}^{2q}\big] < \infty, \quad q \in [1,\infty). \tag{6.5}$$

On the other hand, by Lemma 3.4, (3.8) for $\alpha = 1$ and (2.7), we derive that for $T \geq 2$ and $\varepsilon \in (0,1)$,

$$\begin{aligned} \mathbb{E}^\mu\big[\|f_{T,\varepsilon}-1\|_{L^2(\mu)}^2\big] &= \frac{1}{T}\sum_{i=1}^\infty \mathrm{e}^{-2\lambda_i\varepsilon}\mathbb{E}^\mu[|\psi_i(T)|^2] \preceq \frac{1}{T}\sum_{i=1}^\infty \mathrm{e}^{-2\lambda_i\varepsilon}\lambda_i^{-1} \\ &\preceq \begin{cases} T^{-1}, & \text{if } d\in[1,2); \\ T^{-1}\log(1+\varepsilon^{-1}), & \text{if } d=2; \\ T^{-1}\varepsilon^{-\frac{d-2}{2}}, & \text{if } d\in(2,4). \end{cases} \end{aligned} \tag{6.6}$$

By the argument used in [30, Lemma 3.3], for ε in (6.4) we have

$$\lim_{T\to\infty} \mathbb{E}^\mu\big[\mu\big(\big|\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}-1\big|^q\big)\big] = 0, \quad q\in[1,\infty). \tag{6.7}$$

Indeed, for each $\eta\in(0,1)$, if $|f_{T,\varepsilon}-1|\leq\eta$, then

$$\big|\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}-1\big| \leq \left|\frac{1}{\sqrt{1-\eta}}-\frac{2}{2+\eta}\right| =: \delta(\eta);$$

while if $|f_{T,\varepsilon}-1|>\eta$, for any $\theta\in(0,1)$,

$$\big|\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}-1\big| \leq 2\theta^{-1}\varepsilon^{-\frac{\theta}{2}}.$$

Combining with Chebyshev's inequality yields

$$\mathbb{E}^\mu\big[\mu\big(\big|\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}-1\big|^q\big)\big] \leq \delta(\eta)^q + 2^q\eta^{-2}\theta^{-q}\varepsilon^{-\frac{\theta q}{2}}\mathbb{E}^\mu\big[\|f_{T,\varepsilon}-1\|_{L^2(\mu)}^2\big].$$

By (6.6), for our choice of ε and a sufficiently small θ, we obtain

$$\limsup_{T\to\infty} \mathbb{E}^\mu\big[\mu\big(\big|\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}-1\big|^q\big)\big] \leq \delta(\eta)^q.$$

Then (6.7) follows by letting $\eta\to 0$.

By (6.3) and the Hölder's inequality, for any $q\in[1,\infty)$,

$$\begin{aligned} &\mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_{T,\varepsilon,\varepsilon},\mu)-(1-\varepsilon)^2\Xi_\varepsilon(T)\}^+\big|^q\Big] \\ &\leq (1-\varepsilon)^2 T^q \mathbb{E}^\mu\Big[\mu\big(|\nabla(-\hat{L})^{-1}(f_{T,\varepsilon}-1)|^2\cdot\big|\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}-1\big|\big)^q\Big] \\ &\leq T^q\mathbb{E}^\mu\Big[\mu\big(|\nabla(-\hat{L})^{-1}(f_{T,\varepsilon}-1)|^{2q}\cdot\big|\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}-1\big|^q\big)\Big] \\ &\leq \Big(T^{2q}\mathbb{E}^\mu\big[\|\nabla(-\hat{L})^{-1}(f_{T,\varepsilon}-1)\|_{L^{4q}(\mu)}^{4q}\big]\Big)^{\frac{1}{2}}\cdot\Big(\mathbb{E}^\mu\big[\mu\big(\big|\mathscr{M}\big((1-\varepsilon)f_{T,\varepsilon}+\varepsilon\big)^{-1}-1\big|^{2q}\big)\big]\Big)^{\frac{1}{2}}. \end{aligned}$$

Combining this with (6.5), (6.7) and $\Xi(T) \ge (1-\varepsilon)^2\Xi_\varepsilon(T)$, we see that for ε as in (6.4),

$$\lim_{T\to\infty} \mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2(\mu_{T,\varepsilon,\varepsilon},\mu)^2 - \Xi(T)\}^+\big|^q\Big] = 0, \quad q\in[1,\infty). \tag{6.8}$$

Moreover, by (4.4) and the convexity of $\mathbb{W}_2^2$, we derive that for ε as in (6.4),

$$\begin{aligned}
&\lim_{T\to\infty} T^q\mathbb{E}^\mu\big[\mathbb{W}_2^{2q}(\mu_T,\mu_{T,\varepsilon,\varepsilon})\big] \preceq \lim_{T\to\infty} T^q\mathbb{E}^\mu\big[\mathbb{W}_2^{2q}(\mu_T,\mu_{T,\varepsilon}) + \mathbb{W}_2^{2q}(\mu_{T,\varepsilon},\mu_{T,\varepsilon,\varepsilon})\big]\\
&\preceq \lim_{T\to\infty}\big\{T^q\mathbb{E}^\mu\big[\mathbb{W}_{2q}^{2q}(\mu_T,\mu_{T,\varepsilon})\big] + T^qD^{2q}\varepsilon^q\big\} = 0, \quad q\in[1,\infty).
\end{aligned}$$

By the triangle inequality for $\mathbb{W}_2$,

$$\begin{aligned}
&\mathbb{E}^\mu\big[\big|\mathbb{W}_2^2(\mu_T,\mu) - \mathbb{W}_2^2(\mu_{T,\varepsilon,\varepsilon},\mu)\big|^q\big]\\
&\preceq \big(\mathbb{E}^\mu\big[\mathbb{W}_2^{2q}(\mu_T,\mu)\big]\big)^{\frac12}\big(\mathbb{E}^\mu\big[\mathbb{W}_2^{2q}(\mu_T,\mu_{T,\varepsilon,\varepsilon})\big]\big)^{\frac12} + \mathbb{E}^\mu\big[\mathbb{W}_2^{2q}(\mu_T,\mu_{T,\varepsilon,\varepsilon})\big].
\end{aligned}$$

Combining these with Theorem 2.1, we deduce that

$$\lim_{T\to\infty} T^q\mathbb{E}^\mu\big[\big|\mathbb{W}_2^2(\mu_T,\mu) - \mathbb{W}_2^2(\mu_{T,\varepsilon,\varepsilon},\mu)\big|^q\big] = 0, \quad q\in[1,\infty),$$

which together with (6.8) yields (6.1). □

Lemma 6.2. *Assume* (2.3) *and either* (2.5) *or* (2.18), *and let* $d\in[1,4)$ *and*

$$q\in J(d) := \begin{cases} [1,\infty), & \text{if } d\in[1,3],\\ [1,\frac{d-2}{2(d-3)}), & \text{if } d\in(3,4),\end{cases}$$

then

$$\sum_{i=1}^\infty \frac{\sup_{T\ge1}\big(\mathbb{E}^\mu\left[|\psi_i(T)|^{2q}\right]\big)^{\frac1q}}{\lambda_i} < \infty. \tag{6.9}$$

Consequently,

$$\sup_{T\ge1,\,\varepsilon\in(0,1)} \mathbb{E}^\mu\left[|\Xi_\varepsilon(T)|^q\right] \le \sup_{T\ge1}\mathbb{E}^\mu\left[|\Xi(T)|^q\right] < \infty, \tag{6.10}$$

$$\lim_{T\to\infty}\mathbb{E}^\mu\big[|\Xi_\varepsilon(T) - \Xi(T)|^q\big] = 0 \tag{6.11}$$

holds for any $\varepsilon = \varepsilon(T)\in(0,1)$ *satisfying* $\lim_{T\to\infty}\varepsilon(T) = 0$.

Proof. (a) Let $\sigma(\cdot)$ and $\mathfrak{m}(\cdot)$ be defined in (3.1). It is easy to see that for any $i\ge1$,

$$\sigma(\phi_i) = \sqrt{2}\lambda_i^{-\frac12}.$$

By $\|\phi_i\|_{L^2(\mu)} = 1$, (2.6) and $s = \lambda_i^{-1}$, we obtain that for any $p\in[2,\infty]$,

$$\|\phi_i\|_{L^p(\mu)} = \mathrm{e}^{\lambda_i s}\|\hat{P}_s\phi_i\|_{L^p(\mu)} \preceq \lambda_i^{\frac{d(p-2)}{4p}}, \quad i\ge1. \tag{6.12}$$

If (2.18) holds, then by (2.6) we find constants $c_1, c_2 > 0$ such that for any $s \in (0, \lambda_1^{-1})$ and $p \in (2, \infty)$,

$$
\begin{aligned}
\|\nabla \hat{P}_s \phi_i\|_{L^p(\mu)} &\le c_1 \|\hat{P}_s(|\nabla \phi_i|^2)\|^{\frac{1}{2}}_{L^{\frac{p}{2}}(\mu)} \\
&\le c_1 \|\hat{P}_s\|^{\frac{1}{2}}_{L^1(\mu)\to L^{\frac{p}{2}}(\mu)} \|\nabla \phi_i\|_{L^2(\mu)} \le c_2 s^{-\frac{(p-2)d}{4p}} \lambda_i^{\frac{1}{2}}, \quad i \ge 1.
\end{aligned} \tag{6.13}
$$

By letting $s = \lambda_i^{-1}$, we obtain that for any $p \in (2, \infty)$,

$$
\|\nabla \phi_i\|_{L^p(\mu)} = \mathrm{e}^{\lambda_i s} \|\nabla \hat{P}_s \phi_i\|_{L^p(\mu)} \preceq \lambda_i^{\frac{d(p-2)}{4p} + \frac{1}{2}}.
$$

So, by taking $p > 2$ such that $\frac{p}{p-2} = \log(\mathrm{e} + \lambda_i)$ for $d = 2$ and $p = d$ for $d > 2$, we deduce from (3.1) that

$$
\mathfrak{m}(\phi_i) \preceq \begin{cases} \lambda_i^{-\frac{1}{2}}, & \text{if } d \in [1, 2), \\ \lambda_i^{-\frac{1}{2}} \log(\mathrm{e} + \lambda_i), & \text{if } d = 2, \\ \lambda_i^{-\frac{4-d}{4}}, & \text{if } d \in (2, \infty). \end{cases} \tag{6.14}
$$

If (2.5) holds, then we deduce (6.14) from (2.5) and (6.12). So, in both cases we have estimate (6.14).

Now, by combining (6.14) with Corollary 3.2, we derived that for any $q \in [1, \infty)$ and $i \ge 1$,

$$
\sup_{T \ge 1} \left(\mathbb{E}^\mu \left[|\psi_i(T)|^q\right]\right)^{\frac{1}{q}} \preceq \begin{cases} \lambda_i^{-\frac{1}{2}}, & \text{if } d \in [1, 2), \\ \lambda_i^{-\frac{1}{2}} \log(\mathrm{e} + \lambda_i), & \text{if } d = 2, \\ \lambda_i^{-\frac{4-d}{4}}, & \text{if } d \in (2, \infty). \end{cases} \tag{6.15}
$$

This together with (2.7) implies (6.9) for $d \in [1, 2]$ and all $q \in [1, \infty)$.

(b) Next, we verify (6.9) for $d \in (2, 4)$. By Lemma 3.4,

$$
\sup_{T \ge 1} \mathbb{E}^\mu \left[|\psi_i(T)|^2\right] \preceq \lambda_i^{-1}, \quad i \ge 1, \tag{6.16}
$$

while Hölder's inequality implies that for any $\tau \in (0, 1)$,

$$
\sup_{T \ge 1} \left(\mathbb{E}^\mu \left[|\psi_i(T)|^{2q}\right]\right)^{\frac{1}{q}} \le \sup_{T \ge 1} \left(\mathbb{E}^\mu \left[|\psi_i(T)|^2\right]\right)^{\frac{1-\tau}{q}} \cdot \sup_{T \ge 1} \left(\mathbb{E}^\mu \left[|\psi_i(T)|^{\frac{2(q-1+\tau)}{\tau}}\right]\right)^{\frac{\tau}{q}}.
$$

Since $\tau \in (0, 1)$ is arbitrary, this together with (6.15) and (6.16) yields that for any $q \in [1, \infty)$ and $\varsigma \in (0, 1)$,

$$
\sup_{T \ge 1} \left(\mathbb{E}^\mu \left[|\psi_i(T)|^{2q}\right]\right)^{\frac{1}{q}} \preceq \lambda_i^{-\frac{4-d}{2} - \frac{d-2}{2q} + \varsigma}, \quad i \ge 1.
$$

Consequently,

$$
\sum_{i=1}^\infty \frac{\sup_{T \ge 1} \left(\mathbb{E}^\mu \left[|\psi_i(T)|^{2q}\right]\right)^{\frac{1}{q}}}{\lambda_i} \preceq \sum_{i=1}^\infty \lambda_i^{-1 - \frac{4-d}{2} - \frac{d-2}{2q} + \varsigma}. \tag{6.17}
$$

By (2.7), the series $\sum_{i=1}^\infty \lambda_i^{-\theta}$ converges when $\theta > \frac{d}{2}$. It is easy to see that

$$1+\frac{4-d}{2}+\frac{d-2}{2q}-\varsigma>\frac{d}{2} \text{ holds for small } \varsigma\in(0,1)$$

if and only if $q \in [1, \frac{d-2}{2(d-3)^+})$, where $\frac{d-2}{2(d-3)^+} := \infty$ for $d \in (2,3]$. So, by (6.17), (6.9) holds for any $d \in (2,4)$ and $q \in J(d) = [1, \frac{d-2}{2(d-3)^+})$.

(c) Let $\Xi(T)$ be in (1.3). By the Minkowski integral inequality, for $q \in [1,\infty)$,

$$\left(\mathbb{E}^\mu\left[|\Xi(T)|^q\right]\right)^{\frac{1}{q}} \le \sum_{i=1}^\infty \frac{\left(\mathbb{E}^\mu\left[|\psi_i(T)|^{2q}\right]\right)^{\frac{1}{q}}}{\lambda_i}.$$

Then (6.10) follows from (6.9) and the fact that $\Xi_\varepsilon(T) \le \Xi(T)$. Noting that

$$\left(\mathbb{E}^\mu[|\Xi_\varepsilon(T)-\Xi(T)|^q]\right)^{\frac{1}{q}} \le \sum_{i=1}^\infty \frac{\left(\mathbb{E}^\mu\left[|\psi_i(T)|^{2q}\right]\right)^{\frac{1}{q}}}{\lambda_i}(1-\mathrm{e}^{-2\lambda_i\varepsilon}), \tag{6.18}$$

by the dominated convergence theorem, we deduce (6.11) from (6.18) and (6.10). Then the proof is finished. $\square$

For a general initial distribution, we have the following result, where Lemma 3.4 indicates that the estimate (6.19) is nearly sharp for $d \in [1,2]$.

Lemma 6.3. *Assume* **(A)** *and let $d \in [1,\infty)$. Then for any $q \in [1,\infty)$,*

$$\sup_{T\ge2,\,\nu\in\mathscr{P}} \left(\mathbb{E}^\nu\left[|\psi_i(T)|^q\right]\right)^{\frac{1}{q}} \preceq \lambda_i^{-\frac{2\wedge(4-d)}{4}+\varsigma}, \quad i\ge1, \tag{6.19}$$

where $\varsigma > 0$ is an arbitrarily small but fixed constant. Moreover, for any $\nu \in \mathscr{P}$,

$$\lim_{T\to\infty} \mathbb{E}^\nu\left[|\psi_i(T)|^q\right] = \frac{2^q\Gamma(\frac{q+1}{2})}{\sqrt{\pi}}\cdot \mathbf{V}(\phi_i)^{\frac{q}{2}}, \quad i\ge1,\ q\in(0,\infty), \tag{6.20}$$

where $\mathbf{V}(\cdot)$ is defined by (3.10).

Proof. (a) For $\nu \in \mathscr{P}$, let $\nu_\varepsilon = \nu P_\varepsilon^*$. Then by the Markov property, Hölder's inequality and (4.14), there exists a constant $c_1 > 0$ such that for any $\varepsilon \in (0,1)$ and $k \in (1,\infty)$,

$$\begin{aligned}\sup_{T\ge2,\,\nu\in\mathscr{P}} \left(\mathbb{E}^\nu\left[|\psi_i(T)|^q\right]\right)^{\frac{1}{q}} &\le \sup_{T\ge2,\,\nu\in\mathscr{P}} \mathbb{E}^\nu\left[\left|\int_0^\varepsilon \phi_i(X_t)\mathrm{d}t\right|^q\right]^{\frac{1}{q}} + \sup_{T\ge1,\,\nu\in\mathscr{P}}\left(\mathbb{E}^{\nu_\varepsilon}\left[|\psi_i(T)|^q\right]\right)^{\frac{1}{q}}\\ &\le \|\phi_i\|_\infty\varepsilon + c_1\varepsilon^{-\frac{d(k-1)}{2kq}}\sup_{T\ge1}\left(\mathbb{E}^\mu\left[|\psi_i(T)|^{\frac{kq}{k-1}}\right]\right)^{\frac{k-1}{kq}}, \quad i\ge1.\end{aligned}$$

Combining this with (6.12) for $p=\infty$ and (6.15) yields

$$\sup_{T\ge2,\,\nu\in\mathscr{P}} \left(\mathbb{E}^\nu\left[|\psi_i(T)|^q\right]\right)^{\frac{1}{q}} \preceq \lambda_i^{\frac{d}{4}}\varepsilon + \varepsilon^{-\frac{d(k-1)}{2kq}}\lambda_i^{-\frac{2\wedge(4-d)}{4}}\left(\log(\mathrm{e}+\lambda_i)1_{\{d=2\}}+1_{\{d\ne2\}}\right), \quad i\ge1.$$

By choosing k sufficiently close to 1 and optimizing in $\varepsilon \in (0,1)$, we prove (6.19).

(b) Given $q \in [1,\infty)$, by (6.19) and Chebyshev's inequality, there exists a constant $C > 0$ depending on i, such that for any $r > 0$,

$$\sup_{T\ge 2,\, \nu\in\mathscr{P}} \mathbb{P}^\nu(|\psi_i(T)| \ge r) \le C r^{-q-1}. \tag{6.21}$$

Now, we introduce a continuous indicator function $\chi \in C(\mathbb{R})$ which satisfies $0 \le \chi \le 1$, $\chi \equiv 1$ on $[-1,1]$ and $\operatorname{supp}\chi \subseteq [-2,2]$. By [31, Theorem 2.1], for any $K > 0$,

$$\lim_{T\to\infty} \mathbb{E}^\nu\big[|\psi_i(T)|^q \chi(K^{-1}\psi_i(T))\big] = \int_{\mathbb{R}} |x|^q \chi(K^{-1}x)\, N(0, 2\mathbf{V}(\phi_i))(\mathrm{d}x).$$

On the other hand, by (6.21), for any $T \ge 2$,

$$\begin{aligned} 0 \le& \mathbb{E}^\nu\big[|\psi_i(T)|^q\big] - \mathbb{E}^\nu\big[|\psi_i(T)|^q \chi(K^{-1}\psi_i(T))\big] \le \mathbb{E}^\nu\big[|\psi_i(T)|^q 1_{\{|\psi_i(T)|>K\}}\big] \\ &\le K^q \mathbb{P}^\nu(|\psi_i(T)| \ge K) + q\int_K^\infty r^{q-1}\mathbb{P}^\nu(|\psi_i(T)| \ge r)\,\mathrm{d}r \le (q+1)CK^{-1}. \end{aligned}$$

Letting $K \to \infty$, we obtain

$$\lim_{T\to\infty} \mathbb{E}^\nu\big[|\psi_i(T)|^q\big] = \int_{\mathbb{R}} |x|^q\, N(0, 2\mathbf{V}(\phi_i))(\mathrm{d}x),$$

which finishes the proof of (6.20). $\square$

Lemma 6.4. *Assume* **(A)** *and let* $d \in [1,4)$. *Then*

$$\lim_{\varepsilon\to 0} \sup_{T\ge 1,\, \nu\in\mathscr{P}} \mathbb{E}^\nu\big[|T\mathbb{W}_2^2(\mu_T,\mu) - T\mathbb{W}_2^2(\tilde\mu_{T,\varepsilon},\mu)|^q\big] = 0, \quad q \in [1,\infty). \tag{6.22}$$

Moreover, if $d \in [1, \frac{8}{3})$ *and* $q \in [1, \frac{d}{2(d-2)^+})$, *where* $\frac{d}{2(d-2)^+} := \infty$ *for* $d \le 2$, *then*

$$\lim_{\varepsilon\to 0} \sup_{T\ge 2,\, \nu\in\mathscr{P}} \mathbb{E}^\nu\big[|\Xi(T) - \tilde\Xi_\varepsilon(T)|^q\big] = 0. \tag{6.23}$$

Proof. (a) Recall that

$$\tilde\mu_{T,\varepsilon} = \frac{1}{T}\int_\varepsilon^{T+\varepsilon} \delta_{X_t}\mathrm{d}t, \quad \varepsilon \in (0,1),\, T \ge 1,$$

and for $\nu \in \mathscr{P}$, let $\nu_\varepsilon = \nu P_\varepsilon^*$. By the Markov property, Theorem 2.1 and (4.14), for any $q \in [1,\infty)$ and $k \in (1,\infty)$, we find constants $c_1, c_2 > 0$ such that

$$\begin{aligned} &\sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\big[\mathbb{W}_2^{2q}(\tilde\mu_{T,\varepsilon},\mu)\big] = \sup_{\nu\in\mathscr{P}} \mathbb{E}^{\nu_\varepsilon}\big[\mathbb{W}_2^{2q}(\mu_T,\mu)\big] \\ &\le c_1 \varepsilon^{-\frac{d(k-1)}{2k}} \mathbb{E}^\mu\Big[\mathbb{W}_2^{\frac{2kq}{k-1}}(\mu_T,\mu)\Big]^{\frac{k-1}{k}} \le c_2 \varepsilon^{-\frac{d(k-1)}{2k}} T^{-q}, \quad T \ge 1,\, \varepsilon \in (0,1). \end{aligned}$$

By combining this with the triangle inequality for $\mathbb{W}_2$, (4.17) for $p = 2$, and $D := \sup\rho < \infty$ due to the boundedness of (M,ρ), we obtain

$$\sup_{T\ge 1,\, \nu\in\mathscr{P}} \mathbb{E}^\nu\big[|T\mathbb{W}_2^2(\mu_T,\mu) - T\mathbb{W}_2^2(\tilde\mu_{T,\varepsilon},\mu)|^{2q}\big]$$

$$
\begin{aligned}
&\leq \sup_{T\geq 1,\,\nu\in\mathscr{P}} \mathbb{E}^\nu\big[|T\mathbb{W}_2^2(\mu_T,\tilde\mu_{T,\varepsilon})+2T\mathbb{W}_2(\mu_T,\tilde\mu_{T,\varepsilon})\mathbb{W}_2(\tilde\mu_{T,\varepsilon},\mu)|^{2q}\big]\\
&\preceq D^{4q}\varepsilon^{2q}+c_2D^{2q}\varepsilon^{q-\frac{d(k-1)}{2k}}.
\end{aligned}
$$

By choosing sufficiently small k such that $q-\frac{d(k-1)}{2k}>0$, and then letting $\varepsilon\to 0$, we derive (6.22).

(b) Let $(\Xi(T),\psi_i(T))$ and $(\tilde\Xi_\varepsilon(T),\tilde\psi_{i,\varepsilon}(T))$ be in (1.3) and (2.23) respectively. Then by Minkowski's integral inequality and Hölder's inequality, we obtain that for any $\nu\in\mathscr{P}$ and $\tau\in(0,1)$,

$$
\begin{aligned}
&\big(\mathbb{E}^\nu\big[|\Xi(T)-\tilde\Xi_\varepsilon(T)|^q\big]\big)^{\frac1q}\leq \sum_{i=1}^\infty\frac{1}{\lambda_i}\mathbb{E}^\nu\Big[\big||\psi_i(T)|^2-|\tilde\psi_{i,\varepsilon}(T)|^2\big|^q\Big]^{\frac1q}\\
&\leq \sum_{i=1}^\infty\frac{1}{\lambda_i}\mathbb{E}^\nu\Big[|\psi_i(T)-\tilde\psi_{i,\varepsilon}(T)|^q\cdot\big(|\psi_i(T)|+|\tilde\psi_{i,\varepsilon}(T)|\big)^q\Big]^{\frac1q}\\
&\leq \sum_{i=1}^\infty\frac{1}{\lambda_i}\mathbb{E}^\nu\big[|\psi_i(T)-\tilde\psi_{i,\varepsilon}(T)|^{\frac q\tau}\big]^{\frac\tau q}\cdot\Big(\mathbb{E}^\nu\big[|\psi_i(T)|^{\frac{q}{1-\tau}}\big]^{\frac{1-\tau}{q}}+\mathbb{E}^\nu\big[|\tilde\psi_{i,\varepsilon}(T)|^{\frac{q}{1-\tau}}\big]^{\frac{1-\tau}{q}}\Big).
\end{aligned}
$$

A direct computation gives

$$
\begin{aligned}
&\sup_{T\geq1,\,\nu\in\mathscr{P}}\mathbb{E}^\nu\big[|\psi_i(T)-\tilde\psi_{i,\varepsilon}(T)|^{\frac q\tau}\big]^{\frac\tau q}\leq \sup_{T\geq1,\,\nu\in\mathscr{P}}\mathbb{E}^\nu\Bigg[\bigg|\int_0^\varepsilon\big(\phi_i(X_t)-\phi_i(X_{T+\varepsilon})\big)\mathrm{d}t\bigg|^{\frac q\tau}\Bigg]^{\frac\tau q}\\
&\leq 2\sup_{\nu\in\mathscr{P}}\mathbb{E}^\nu\Bigg[\bigg|\int_0^\varepsilon\phi_i(X_t)\mathrm{d}t\bigg|^{\frac q\tau}\Bigg]^{\frac\tau q}\leq 2\int_0^\varepsilon\sup_{\nu\in\mathscr{P}}\mathbb{E}^\nu\big[|\phi_i(X_t)|^{\frac q\tau}\big]^{\frac\tau q}\mathrm{d}t\\
&\leq 2\int_0^\varepsilon\big\|P_t(|\phi_i|^{\frac q\tau})\big\|_\infty^{\frac\tau q}\mathrm{d}t.
\end{aligned}\tag{6.24}
$$

By (4.13) and (6.12) for $p=2\vee\frac q\tau$, we find a constant $c>0$ such that

$$
\big\|P_t(|\phi_i|^{\frac q\tau})\big\|_\infty^{\frac\tau q}\leq\|P_t\|_{L^1(\mu)\to L^\infty(\mu)}^{\frac\tau q}\cdot\|\phi_i\|_{L^{\frac q\tau}(\mu)}\leq ct^{-\frac{d\tau}{2q}}\lambda_i^{\frac{d(q-2\tau)^+}{4q}},\quad t\in(0,1).
$$

Combining this with (6.24) yields for any $\tau\in(0,1\wedge\frac{2q}{d})$,

$$
\sup_{T\geq1,\,\nu\in\mathscr{P}}\mathbb{E}^\nu\big[|\psi_i(T)-\tilde\psi_{i,\varepsilon}(T)|^{\frac q\tau}\big]^{\frac\tau q}\preceq\varepsilon^{1-\frac{d\tau}{2q}}\lambda_i^{\frac{d(q-2\tau)^+}{4q}},\quad \varepsilon\in(0,1),\ i\geq1.
$$

On the other hand, we take $k\in(1,\infty)$. By the Markov property, Hölder's inequality, (4.14) and (6.15), we obtain that for $i\geq1$,

$$
\begin{aligned}
&\sup_{T\geq1,\,\nu\in\mathscr{P}}\mathbb{E}^\nu\big[|\tilde\psi_{i,\varepsilon}(T)|^{\frac{q}{1-\tau}}\big]^{\frac{1-\tau}{q}}=\sup_{T\geq1,\,\nu_\varepsilon\in\mathscr{P}}\mathbb{E}^{\nu_\varepsilon}\big[|\psi_i(T)|^{\frac{q}{1-\tau}}\big]^{\frac{1-\tau}{q}}\\
&\leq\sup_{T\geq1,\,\nu\in\mathscr{P}}\Big\|\frac{\mathrm{d}\nu_\varepsilon}{\mathrm{d}\mu}\Big\|_{L^k(\mu)}\cdot\mathbb{E}^\mu\big[|\psi_i(T)|^{\frac{kq}{(k-1)(1-\tau)}}\big]^{\frac{(k-1)(1-\tau)}{kq}}\\
&\preceq\varepsilon^{-\frac{d(k-1)}{2k}}\lambda_i^{-\frac{2\wedge(4-d)}{4}+\varsigma},\qquad i\geq1,
\end{aligned}
$$

for any fixed constant $\varsigma > 0$. Moreover, Lemma 6.3 yields

$$\sup_{T\geq 2,\, \nu\in\mathscr{P}} \mathbb{E}^\nu \Big[|\psi_i(T)|^{\frac{q}{1-\tau}}\Big]^{\frac{1-\tau}{q}} \preceq \lambda_i^{-\frac{2\wedge(4-d)}{4}+\varsigma}, \quad i\geq 1.$$

Combining the above estimates, we conclude that for any $\tau \in (0, 1\wedge \frac{2q}{d})$, $k\in(1,\infty)$ and $\varsigma\in(0,1)$, there exists a constant $C_{\tau,k,\varsigma}\in(0,\infty)$ such that

$$\sup_{T\geq 2,\, \nu\in\mathscr{P}} \Big(\mathbb{E}^\nu\big[|\Xi(T)-\tilde{\Xi}_\varepsilon(T)|^q\big]\Big)^{\frac{1}{q}} \leq C_{\tau,k,\varsigma}\, \varepsilon^{1-\frac{d\tau}{2q}-\frac{d(k-1)}{2k}} \sum_{i=1}^\infty \lambda_i^{-1+\frac{d(q-2\tau)^+}{4q}-\frac{2\wedge(4-d)}{4}+\varsigma}. \tag{6.25}$$

Noting that when $d\in[1,\frac{8}{3})$ and $q\in[1,\frac{d}{2(d-2)^+})$, where $\frac{d}{2(d-2)^+}:=\infty$ for $d\leq 2$, we have

$$\lim_{\tau\uparrow 1\wedge\frac{2q}{d},\ \varsigma\downarrow 0} \Big[-1+\frac{d(q-2\tau)^+}{4q}-\frac{2\wedge(4-d)}{4}+\varsigma\Big] = -1+\frac{d(q-2)^+}{4q}-\frac{2\wedge(4-d)}{4} < -\frac{d}{2},$$

and

$$\lim_{k\downarrow 1}\Big[1-\frac{d\tau}{2q}-\frac{d(k-1)}{2k}\Big] = 1-\frac{d\tau}{2q} > 0, \quad \tau\in\Big(0, 1\wedge\frac{2q}{d}\Big).$$

Then we can always find $\tau\in(0,1\wedge\frac{2q}{d})$, $k\in(1,\infty)$ and $\varsigma\in(0,1)$ such that

$$-1+\frac{d(q-2\tau)}{4q}-\frac{2\wedge(4-d)}{4}+\varsigma < -\frac{d}{2} \quad\text{and}\quad 1-\frac{d\tau}{2q}-\frac{d(k-1)}{2k} > 0$$

hold. Hence, (2.7) implies

$$\sum_{i=1}^\infty \lambda_i^{-1+\frac{d(q-2\tau)^+}{4q}-\frac{2\wedge(4-d)}{4}+\varsigma} < \infty,$$

so that by letting $\varepsilon\to 0$ in (6.25) we deduce (6.23). □

Proof of Theorem 2.4(1). Let $\nu = h_\nu\mu\in\mathscr{P}_{k,R}$. We have $\mathbb{E}^\nu[\eta] = \mathbb{E}^\mu[h_\nu(X_0)\eta]$ for a nonnegative measurable functional η of $(X_\cdot)$. Let $k^* = \frac{k}{k-1}$. By Hölder's inequality we obtain

$$\begin{aligned}
\mathbb{E}^\nu\big[|\{T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)\}^+|^q\big] &= \mathbb{E}^\mu\big[h_\nu(X_0)|\{T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)\}^+|^q\big] \\
&\leq \|h_\nu\|_{L^k(\mu)}\Big(\mathbb{E}^\mu\big[|\{T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)\}^+|^{k^*q}\big]\Big)^{\frac{1}{k^*}} \\
&\leq R\Big(\mathbb{E}^\mu\big[|\{T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)\}^+|^{k^*q}\big]\Big)^{\frac{1}{k^*}}.
\end{aligned}$$

Combining this with Lemma 6.1 and letting $T\to\infty$, we derive (2.19).

Next, let $d\in[1,\frac{8}{3})$ and $q\in[1,\frac{d}{2(d-2)^+})$. For $\nu\in\mathscr{P}$, let $\nu_\varepsilon := \nu P_\varepsilon^*$ with $\varepsilon\in(0,1)$. By the fact that $\nu_\varepsilon\leq c_1\varepsilon^{-\frac{d}{2}}\mu$, (2.19) for $k=\infty$, and the Markov property, we obtain

$$\lim_{T\to\infty}\sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\big[|\{T\mathbb{W}_2^2(\tilde{\mu}_{T,\varepsilon},\mu)-\tilde{\Xi}_\varepsilon(T)\}^+|^q\big] = \lim_{T\to\infty}\sup_{\nu\in\mathscr{P}} \mathbb{E}^{\nu_\varepsilon}\big[|\{T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)\}^+|^q\big] = 0.$$

Then by triangle inequality, for any $\varepsilon \in (0,1)$,

$$
\begin{aligned}
&\limsup_{T\to\infty} \sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\big[|\{T\mathbb{W}_2^2(\mu_T,\mu) - \Xi(T)\}^+|^q\big] \\
&\preceq \limsup_{T\to\infty} \sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\big[|\{T\mathbb{W}_2^2(\tilde\mu_{T,\varepsilon},\mu) - \tilde\Xi_\varepsilon(T)\}^+|^q\big] \\
&\quad + \sup_{T\geq 2,\nu\in\mathscr{P}} \mathbb{E}^\nu\big[|T\mathbb{W}_2^2(\mu_T,\mu) - T\mathbb{W}_2^2(\tilde\mu_{T,\varepsilon},\mu)|^q + |\Xi(T) - \tilde\Xi_\varepsilon(T)|^q\big] \\
&= \sup_{T\geq 2,\nu\in\mathscr{P}} \mathbb{E}^\nu\big[|T\mathbb{W}_2^2(\mu_T,\mu) - T\mathbb{W}_2^2(\tilde\mu_{T,\varepsilon},\mu)|^q + |\Xi(T) - \tilde\Xi_\varepsilon(T)|^q\big].
\end{aligned}
$$

By combining this inequality with Lemma 6.4 and letting $\varepsilon \to 0$, we derive (2.20). □

Proof of Theorem 2.4(2). Assume **(A)** and **(C)**.

(a) As explained in the proof of Theorem 2.4(1) that, the latter assertion for $d \in [1, \frac{8}{3})$ and $q \in [1, \frac{d}{2(d-2)^+})$ is a consequence of the former assertion for $k = \infty$ and Lemma 6.4. Moreover, the argument used in the proof of (2.19) shows that (2.21) is derived from

$$
\lim_{T\to\infty} \mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_T,\mu) - \Xi(T)\}^-\big|^q\Big] = 0, \quad q \in J(d). \tag{6.26}
$$

So, it remains to prove this formula.

(b) By the triangle inequality and Minkowski's inequality, for $\varepsilon \in (0,1)$,

$$
\begin{aligned}
&\mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_T,\mu) - \Xi(T)\}^-\big|^q\Big]^{\frac{1}{q}} \\
&\leq \mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_T,\mu) - T\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu)\}^-\big|^q\Big]^{\frac{1}{q}} + \mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu) - \Xi_\varepsilon(T)\}^-\big|^q\Big]^{\frac{1}{q}} \\
&\quad + \mathbb{E}^\mu[|\Xi_\varepsilon(T) - \Xi(T)|^q]^{\frac{1}{q}}.
\end{aligned} \tag{6.27}
$$

Since $\mu_{T,\varepsilon} = \mu\hat P_\varepsilon$, it follows from (2.17) that

$$
\mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_T,\mu) - T\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu)\}^-\big|^q\Big]^{\frac{1}{q}} \leq |h(\varepsilon)-1| \cdot T\mathbb{E}^\mu\big[\mathbb{W}_2^{2q}(\mu_T,\mu)\big]^{\frac{1}{q}}.
$$

Combining this with Theorem 2.1, we derive

$$
\lim_{T\to\infty} \mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_T,\mu) - T\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu)\}^-\big|^q\Big] = 0
$$

for any $\varepsilon = \varepsilon(T) \in (0,1)$ satisfying $\lim_{T\to\infty} \varepsilon(T) = 0$. Combining this with (6.11) and (6.27), to prove (6.26), it suffices to prove that

$$
\lim_{T\to\infty} \mathbb{E}^\mu\Big[\big|\{T\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu) - \Xi_\varepsilon(T)\}^-\big|^q\Big] = 0, \quad q \in J(d), \tag{6.28}
$$

holds for some choice of $\varepsilon = \varepsilon(T) \in (0,1)$ with $\lim_{T\to\infty} \varepsilon(T) = 0$.

(c) For any $q \in J(d)$, we take a $\varsigma > 0$ sufficiently small such that $q + \varsigma \in J(d)$. Then by (6.10) and Chebyshev's inequality, there exists a constant $C > 0$ depending on $q+\varsigma$ such that

$$
\mathbb{P}^\mu\big(|\Xi_\varepsilon(T)| > r\big) \leq \mathbb{P}^\mu\big(|\Xi(T)| > r\big) \leq 1 \wedge (Cr^{-q-\varsigma}), \quad r > 0,\ T \geq 1,\ \varepsilon \in (0,1).
$$

Hence, according to the dominated convergence theorem on the measure space $(\mathbb{R}_+, r^{q-1}\mathrm{d}r)$, in order to prove (6.28), it is enough to show that for some proper choice of $\varepsilon = \varepsilon(T) \in (0,1)$ with $\lim_{T\to\infty} \varepsilon(T) = 0$,

$$\lim_{T\to\infty} \mathbb{P}^\mu\big(\{T\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu) - \Xi_\varepsilon(T)\}^- > r\big) = 0, \quad r > 0. \tag{6.29}$$

By the argument used the proof of [28, Theorem 2.4, p.35-37], under **(C)** we have

$$\limsup_{T\to\infty} \mathbb{P}^\mu\big(\{T\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu) - \Xi_\varepsilon(T)\}^- > r\big) \le \limsup_{T\to\infty} \mathbb{P}^\mu(B_{T,\varepsilon}^c), \quad \varepsilon \in (0,1), \tag{6.30}$$

where θ is the constant in the assumption **(C)**, and

$$B_{T,\varepsilon} := \big\{\|f_{T,\varepsilon} - 1\|_\infty^2 \le c_1 T^{-\frac{2+2\theta}{2+3\theta}}\big\}$$

for some constant $c_1 > 0$. By Chebyshev's inequality and (2.6) for $(2,\infty)$ in place of (p,q), we find a constant $c_2 > 0$ such that

$$\mathbb{P}^\mu(B_{T,\varepsilon}^c) \le c_1^{-1} T^{\frac{2+2\theta}{2+3\theta}} \mathbb{E}^\mu\big[\|f_{T,\varepsilon} - 1\|_\infty^2\big] \le c_2 T^{\frac{2+2\theta}{2+3\theta}} \varepsilon^{-\frac{d}{2}} \mathbb{E}^\mu\big[\|f_{T,\frac{\varepsilon}{2}} - 1\|_{L^2(\mu)}^2\big].$$

Combining this estimate with (6.6), it follows that for the choice $\varepsilon = T^{-\zeta}$ with sufficiently small $\zeta > 0$,

$$\lim_{T\to\infty} \mathbb{P}^\mu(B_{T,\varepsilon}^c) = 0.$$

By (6.30), this implies (6.29), and hence finishes the proof.

□

Proof of Theorem 2.5. We only prove the first assertion, as the second can be proved in the same way.

(a) For any $r \in (0,\infty)$, we define

$$\bar{\mu}_{T,r} := \frac{1}{T-r}\int_r^T \delta_{X_t}\mathrm{d}t, \quad T > r,$$
$$\tilde{\mu}_{T,r} := \frac{1}{T}\int_r^{T+r} \delta_{X_t}\mathrm{d}t, \quad T > 0.$$

By (4.13), there exists a constant $c(r) \in (0,\infty)$ such that

$$\nu_r := \nu P_r^* \le c(r)\mu, \quad \nu \in \mathscr{P}. \tag{6.31}$$

Then the Markov property, (2.9) and (6.1) imply

$$\begin{aligned}
&\lim_{T\to\infty} \sup_{\nu\in\mathscr{P}} \mathbb{E}^\nu\big[|\{T\mathbb{W}_2^2(\bar{\mu}_{T,r},\mu) - \bar{\Xi}_r(T)\}^+|^q\big] \\
&= \lim_{T\to\infty} \sup_{\nu\in\mathscr{P}} \mathbb{E}^{\nu_r}\big[|\{T\mathbb{W}_2^2(\mu_{T-r},\mu) - \Xi(T-r)\}^+|^q\big] \\
&\le c(r)2^{q-1} \lim_{T\to\infty} \mathbb{E}^\mu\big[|\{(T-r)\mathbb{W}_2^2(\mu_{T-r},\mu) - \Xi(T-r)\}^+|^q + r^q\mathbb{W}_2^{2q}(\mu_{T-r},\mu)\big] = 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}&\lim_{T\to\infty}\sup_{\nu\in\mathscr{P}}\mathbb{E}^{\nu}\big[|\{T\mathbb{W}_2^2(\tilde\mu_{T,r},\mu)-\tilde\Xi_r(T)\}^+|^q\big]\\&=\lim_{T\to\infty}\sup_{\nu\in\mathscr{P}}\mathbb{E}^{\nu_r}\big[|\{T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)\}^+|^q\big]\\&\le c(r)\lim_{T\to\infty}\mathbb{E}^{\mu}\big[|\{T\mathbb{W}_2^2(\mu_T,\mu)-\Xi(T)\}^+|^q=0.\end{aligned}$$

So, it remains to verify that for any $r\in(0,\infty)$ and $q\in[1,\infty)$,

$$\lim_{T\to\infty}T\sup_{\nu\in\mathscr{P}}\Big(\mathbb{E}^{\nu}\big[|\mathbb{W}_2^2(\mu_T,\mu)-\mathbb{W}_2^2(\bar\mu_{T,r},\mu)|^q\big]\Big)^{\frac{1}{q}}=0, \tag{6.32}$$

$$\lim_{T\to\infty}T\sup_{\nu\in\mathscr{P}}\Big(\mathbb{E}^{\nu}\big[|\mathbb{W}_2^2(\tilde\mu_{T,r},\mu)-\mathbb{W}_2^2(\bar\mu_{T,r},\mu)|^q\big]\Big)^{\frac{1}{q}}=0. \tag{6.33}$$

(b) Verification of (6.32). Noting that for $\mu_r:=\frac{1}{r}\int_0^r\delta_{X_t}\mathrm{d}t$, we have

$$\mu_T=(\theta-rT^{-1})\bar\mu_{T,r}+rT^{-1}\mu_r+(1-\theta)\bar\mu_{T,r},\quad\theta\in(rT^{-1},1), \tag{6.34}$$

by (3.17) and the triangle inequality, we obtain

$$\begin{aligned}&\mathbb{W}_2(\mu_T,\bar\mu_{T,r})\le\sqrt{\theta}\mathbb{W}_2\big((1-\theta^{-1}rT^{-1})\bar\mu_{T,r}+\theta^{-1}rT^{-1}\mu_r,\ \bar\mu_{T,r}\big)\\&\le\sqrt{\theta}\Big[\mathbb{W}_2\big((1-\theta^{-1}rT^{-1})\bar\mu_{T,r}+\theta^{-1}rT^{-1}\mu_r,\ (1-\theta^{-1}rT^{-1})\mu+\theta^{-1}rT^{-1}\mu_r\big)\\&\qquad+\mathbb{W}_2\big((1-\theta^{-1}rT^{-1})\mu+\theta^{-1}rT^{-1}\mu_r,\ \mu\big)+\mathbb{W}_2\big(\bar\mu_{T,r},\mu\big)\Big],\quad\theta\in(rT^{-1},1).\end{aligned} \tag{6.35}$$

By (3.17), we obtain

$$\mathbb{W}_2\big((1-\theta^{-1}rT^{-1})\bar\mu_{T,r}+\theta^{-1}rT^{-1}\mu_r,\ (1-\theta^{-1}rT^{-1})\mu+\theta^{-1}rT^{-1}\mu_r\big)\le\mathbb{W}_2(\bar\mu_{T,r},\mu), \tag{6.36}$$

and by Lemma 3.5, we find a constant $c_1\in(0,\infty)$ such that for any $T>r$ and $\theta\in(rT^{-1},1)$,

$$\mathbb{W}_2\big((1-\theta^{-1}rT^{-1})\mu+\theta^{-1}rT^{-1}\mu_r,\ \mu\big)\le c_1\cdot\begin{cases}\theta^{-1}rT^{-1}, & \text{if } d\in[1,2),\\ \theta^{-1}rT^{-1}\log(1+\theta r^{-1}T), & \text{if } d=2,\\ (\theta^{-1}rT^{-1})^{\frac12+\frac1d}, & \text{if } d\in(2,4).\end{cases} \tag{6.37}$$

Moreover, by (6.31), the Markov property and (2.9), we find a constant $c_2(r)>0$ such that

$$\sup_{\nu\in\mathscr{P}}\big(\mathbb{E}^{\nu}[\mathbb{W}_2^{2q}(\bar\mu_{T,r},\mu)]\big)^{\frac1q}=\sup_{\nu\in\mathscr{P}}\big(\mathbb{E}^{\nu_r}[\mathbb{W}_2^{2q}(\mu_{T-r},\mu)]\big)^{\frac1q}\le c_2(r)T^{-1},\quad T\ge 2r+2. \tag{6.38}$$

Combining (6.35)-(6.38) together, we find a constant $c_3>0$, depending on $r\in(0,\infty)$, such that for any $T\ge 2r+2$ and $\theta\in(rT^{-1},1)$,

$$\sup_{\nu\in\mathscr{P}}\big(\mathbb{E}^{\nu}[\mathbb{W}_2^{2q}(\mu_T,\bar\mu_{T,r})]\big)^{\frac1q}\le c_3\cdot\begin{cases}\theta^{-1}T^{-2}+\theta T^{-1}, & \text{if } d\in[1,2),\\ \theta^{-1}T^{-2}\log^2(\theta T)+\theta T^{-1}, & \text{if } d=2,\\ \theta^{-\frac2d}T^{-\frac{d+2}{d}}+\theta T^{-1}, & \text{if } d\in(2,4).\end{cases}$$

Choosing

$$\theta = \begin{cases} T^{-\frac{1}{2}}, & \text{if } d \in [1,2], \\ T^{-\frac{2}{d+2}}, & \text{if } d \in (2,4), \end{cases}$$

which is in $(rT^{-1}, 1)$ for large $T > 0$, we get

$$\sup_{\nu\in\mathscr{P}} \big(\mathbb{E}^\nu[\mathbb{W}_2^{2q}(\mu_T, \bar{\mu}_{T,r})]\big)^{\frac{1}{q}} \preceq \begin{cases} T^{-\frac{3}{2}}, & \text{if } d \in [1,2), \\ T^{-\frac{3}{2}} \log^2(T), & \text{if } d = 2, \\ T^{-\frac{d+4}{d+2}}, & \text{if } d \in (2,4), \end{cases} \qquad q \in [1,\infty).$$

Combining this with (6.38), and applying the triangle inequality and Hölder's inequality, we arrive at

$$\begin{aligned}
&T \sup_{\nu\in\mathscr{P}} \big(\mathbb{E}^\nu\big[|\mathbb{W}_2^2(\mu_T,\mu) - \mathbb{W}_2^2(\bar{\mu}_{T,r},\mu)|^q\big]\big)^{\frac{1}{q}} \\
&\le T \sup_{\nu\in\mathscr{P}} \big(\mathbb{E}^\nu\big[|\mathbb{W}_2^2(\mu_T,\bar{\mu}_{T,r}) + 2\mathbb{W}_2(\mu_T,\bar{\mu}_{T,r})\mathbb{W}_2(\bar{\mu}_{T,r},\mu)|^q\big]\big)^{\frac{1}{q}} \\
&\le T \sup_{\nu\in\mathscr{P}} \big(\mathbb{E}^\nu[\mathbb{W}_2^{2q}(\mu_T,\bar{\mu}_{T,r})]\big)^{\frac{1}{q}} + 2T \sup_{\nu\in\mathscr{P}} \big(\mathbb{E}^\nu[\mathbb{W}_2^{2q}(\mu_T,\bar{\mu}_{T,r})]\big)^{\frac{1}{2q}} \cdot \big(\mathbb{E}^\nu[\mathbb{W}_2^{2q}(\bar{\mu}_{T,r},\mu)]\big)^{\frac{1}{2q}} \\
&\preceq T^{-\frac{1}{d\vee 2+2}} \big(\log(T)1_{\{d=2\}} + 1_{\{d\neq 2\}}\big) \to 0 \quad \text{as} \quad T \to \infty.
\end{aligned}$$

So, (6.32) holds.

(c) Verification of (6.33). Let $\hat{\mu}_r := \frac{1}{r}\int_T^{T+r} \delta_{X_t} \mathrm{d}t$. We have

$$\tilde{\mu}_{T,r} = (\theta - rT^{-1})\bar{\mu}_{T,r} + rT^{-1}\hat{\mu}_r + (1-\theta)\bar{\mu}_{T,r}, \quad \theta \in (rT^{-1}, 1).$$

By using this formula replacing (6.34), the above argument leads to (6.33). Therefore, the proof is finished.

□

References

[1] L. Ambrosio, F. Stra, D. Trevisan, *A PDE approach to a 2-dimensional matching problem,* Probab. Theory Relat. Fields 173(2019), 433–477.

[2] E. Acerbi, N. Fusco, *Semicontinuity problems in the calculus of variations,* Arch. Rational Mech. Anal. 86(1984), 125–145.

[3] P. Auscher, T. Coulhon, X. T. Duong, S. Hofmann, *Riesz transform on manifolds and heat kernel regularity,* Ann. Sci. Éc. Norm. Supér. 37(2004), 911–957.

[4] D. Bakry, *Étude des transformations de Riesz dans les variétés à courbure de Ricci minorée,* Séminaire de Probabilités XXI, Lecture Notes in Math. 1247(1987), 137–172.

[5] D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators,* Grundlehren der mathematischen Wissenschaften 348. Springer, 2014.

[6] I. V. Bogachev, N. V. Krylov, M. Röckner, S. V. Shaposhnikov, *Fokker-Planck-Kolmogorov Equations,* Amer. Math. Soc. 2015.

[7] A. Carbonaro, O. Dragičević, *Bellman function and dimension-free estimates in a theorem of Bakry,* J. Funct. Anal. 265(2013), 1085–1104.

[8] L.-J. Cheng, A. Thalmaier, F.-Y. Wang, *Covariant Riesz transform on differential forms for* $1 < p \leq 2$*,* Cal. Var. Part. Diff. Equat. 62(2023), No. 245.

[9] E. B. Davies, *Heat Kernels and Spectral Theory,* Cambridge Univ. Press, 1989.

[10] K. Du, Y. Jiang, J. Li, *Empirical approximation to invariant measures for McKean-Vlasov processes: mean-field interaction vs self-interaction,* Bernoulli 29(2023), 2492–2518.

[11] F. Gao, A. Guillin, L. Wu, *Bernstein-type concentration inequalities for symmetric Markov processes*, Theory Probab. Appl. 58(2014), 358–382.

[12] M. Goldman, D. Trevisan, *Optimal transport methods for combinatorial optimization over two random point sets*, Probab. Theory Related Fields, 118(2024), 1315–1384.

[13] A. Grigory'an, *Heat Kernel and Analysis on Manifolds,* American Mathematical Society, Providence, RI; International Press, Boston, MA, 2009.

[14] M. Huesmann, F. Mattesini, D. Trevisan, *Wasserstein asymptotics for the empirical measure of fractional Brownian motion on a flat torus,* Stoch. Proc. Appl. 155(2023), 1–26.

[15] M. Ledoux, *On optimal matching of Gaussian samples*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 457, Veroyatnost' i Statistika. 25(2017), 226–264.

[16] H. Li, B. Wu, *Wasserstein convergence for empirical measures of subordinated Dirichlet diffusions on Riemannian manifolds,* arXiv:2206.03901.

[17] H. Li, B. Wu, *Wasserstein convergence for conditional empirical measures of subordinated Dirichlet diffusions on Riemannian manifolds,* Commun. Pure Appl. Anal. 23(2024), 546–576.

[18] H. Li, B. Wu, *Wasserstein convergence rates for empirical measures of subordinated processes on noncompact manifolds,* J. Theoret. Probab. 36(2023), 1243–1268.

[19] F.-C. Liu, *A Luzin type property of Sobolev functions,* Indiana Univ. Math. J. 26(1977), 645–651.

[20] M. Mariani, D. Trevisan, *Wasserstein asymptotics for Brownian motion on the flat torus and Browninan interlacements,* arXiv:2307.10325.

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

[22] C. Villani, *Optimal transport: Old and New Part 1,2,* Springer, 2009.

[23] F.-Y. Wang, *Functional Inequalities, Markov Semigroups and Spectral Theory,* 2005 Science Press.

[24] F.-Y. Wang, *Precise limit in Wasserstein distance for conditional empirical measures of Dirichlet diffusion processes,* J. Funct. Anal. 280(2021), 108998, 23pp.

[25] F.-Y. Wang, *Wasserstein convergence rate for empirical measures on noncompact manifolds,* Stoch. Proc. Appl. 144(2022), 271–287.

[26] F.-Y. Wang, *Convergence in Wasserstein distance for empirical measures of semilinear SPDEs,* Ann. Appl. Probab. 33(2023), 70–84.

[27] F.-Y. Wang, *Convergence in Wasserstein distance for empirical measures of Dirichlet diffusion processes on manifolds,* J. Eur. Math. Soc. 25(2023), 3695–3725.

[28] F.-Y. Wang, *Convergence in Wasserstein distance for empirical measures of non-symmetric subordinated diffusion processes,* arXiv:2301.08420.

[29] F.-Y. Wang, B. Wu, *Wasserstein convergence for empirical measures of subordinated diffusion processes on Riemannian manifolds,* Potential Anal. 59(2023), 933–954.

[30] F.-Y. Wang, J.-X. Zhu, *Limit theorems in Wasserstein distance for empirical measures of diffusion processes on Riemannian manifolds,* Ann. Inst. Henri Poincaré Probab. Stat. 59(2023), 437–475.

[31] L. Wu, *Moderate deviations of dependent random variables related to CLT,* Ann. Probab. 23(1995), 420–445.

[32] L. Wu, *A deviation inequality for non-reversible Markov processes,* Ann. Inst. Henri Poincaré Probab. Stat. 36(2000), 435–445.

[33] H. Ye, J. Gao, Y. Ding, *A generalized Gronwall inequality and its application to a fractional differential equation,* J. Math. Anal. Appl. 328(2007), 1075–1081.