Wasserstein asymptotics for empirical measures of diffusions on four dimensional closed manifolds

Dario Trevisan[a) **Feng-Yu Wang**[b) **Jie-Xiang Zhu**[c)

[a) Dipartimento di Matematica, Università di Pisa, Italy

[b) Center for Applied Mathematics, Tianjin University, China

[c) Department of Mathematics, Shanghai Normal University, China

dario.trevisan@unipi.it, wangfy@tju.edu.cn, jiexiangzhu7@gmail.com

October 30, 2024

Abstract

We identify the leading term in the asymptotics of the quadratic Wasserstein distance between the invariant measure and empirical measures for diffusion processes on closed weighted four-dimensional Riemannian manifolds. Unlike results in lower dimensions, our analysis shows that this term depends solely on the Riemannian volume of the manifold, remaining unaffected by the potential and vector field in the diffusion generator.

AMS subject Classification: 49Q22, 60B1.
Keywords: Empirical measures; Diffusion processes; Optimal transport.

1 Introduction and main result

On a d-dimensional closed Riemannian manifold M, let $L := \Delta + \nabla V \nabla + Z$,

$$L(f) = \Delta f + \nabla V\, \nabla f + Zf, \quad \forall f \in C^2(M),$$

where Δ is the Laplacian, $V \in C^2(M)$ such that for the Riemannian volume measure vol

$$\mu(\mathrm{d}x) := \mathrm{e}^{V(x)}\mathrm{vol}(\mathrm{d}x)$$

is a probability measure, and Z is a C^1-vector field (i.e., a derivation) with $\mathrm{div}_\mu(Z) = 0$:

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

Let $X := (X_t)_{t\geq 0}$ be the diffusion process generated by L and write P_t for its transition semigroup. It is well-known that X is exponentially ergodic with μ as its unique invariant probability measure. Consider then the empirical measure

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

where for any $x \in M$, δ_x denotes the Dirac measure at point x. By ergodicity, a.s. the weak convergence of probabilities $\mu_T \to \mu$ holds as $T \to \infty$. It is of interest to establish quantitative convergence results in terms of suitable metrics on the space of probabilities on M. A natural choice is provided here by the Wasserstein distance $\mathbb{W}_p$ (for any $p \geq 1$) induced by the Riemannian distance ρ on M. The distance $\mathbb{W}_p$ is the p-th root of the optimal cost required to transport μ_T into μ, where the displacement cost from x to y is given by the p-th power of the Riemannian distance $\rho(x,y)^p$. In this context, it has been deeply investigated in a series of recent works [11, 12, 14] and the following asymptotic behavior holds [13, theorem 1.1], given $1 \leq p \leq \max\left\{\frac{2d}{(d-2)^+}, \frac{d(d-2)}{2}\right\}$, for every $x \in M$:

$$\mathbb{E}^x\left[\mathbb{W}_p^p(\mu_T,\mu)\right] \sim^{1} \begin{cases} T^{-p/2} & \text{if } d \leq 3, \\ ((\log T)/T)^{p/2} & \text{if } d = 4, \\ T^{-p/(d-2)} & \text{if } d \geq 5, \end{cases} \tag{1.1}$$

where $\mathbb{E}^x$ denotes the expectation with respect to the probability $\mathbb{P}^x$ under which the diffusion process has initial condition $X_0 = x$. It is conjectured that (1.1) can be extended to all $1 \leq p < \infty$.

If one denotes with $R_{p,d}(T)$ the right hand side in (1.1), the existence of the limit

$$\lim_{T\to\infty} \frac{\mathbb{E}^x\left[\mathbb{W}_p^p(\mu_T,\mu)\right]}{R_{p,d}(T)} =: \mathsf{c}(L,p) \tag{1.2}$$

is also naturally conjectured, although it is only proved so far in the case $d \leq 3$ in [12, 13]. For $p = 2$, it reads

$$\lim_{T\to\infty} T\mathbb{E}^x\left[\mathbb{W}_2^2(\mu_T,\mu)\right] = \sum_{i=1}^{\infty}\frac{2}{\lambda_i^2}\left(1 - \frac{1}{\lambda_i}\mathbf{V}(Z\phi_i)\right), \tag{1.3}$$

where $\{\lambda_i\}_{i\geq 1}$ are the strictly positive eigenvalues of $-(\Delta + \nabla V)$ in $L^2(\mu)$, $\{\phi_i\}_{i\geq 1}$ are associated unit $L^2(\mu)$-norm eigenfunctions and $\mathbf{V}$ denotes the quadratic form

$$\mathbf{V}(\phi) := \int_0^\infty \int_M \phi P_t\phi \,\mathrm{d}\mu\,\mathrm{d}t. \tag{1.4}$$

[1] The notation $f(T) \sim g(T)$ means that it holds $c^{-1} \leq f(T)/g(T) \leq c$ for T sufficiently large, for some constant $c \in (0,\infty)$ possibly depending on M, L and other parameters but not T. We also collect for later use the notation $f(T) \lesssim g(T)$, when it holds $f(T) \leq cg(T)$ for T sufficiently large, for some constant $c < \infty$.

The appearance of $\mathbf{V}$ is ultimately due to the central limit theorem in this setting [15]: for every $\phi \in L^2(\mu)$ with $\int_M \phi \, \mathrm{d}\mu = 0$, one has convergence in law, as $T \to \infty$,

$$\frac{1}{\sqrt{T}} \int \phi \mathrm{d}\mu_T = \frac{1}{\sqrt{T}} \int_0^T \phi(X_t) \mathrm{d}t \to \mathcal{N}(0, 2\mathbf{V}(\phi)). \tag{1.5}$$

Let us notice that, since $\mathbf{V}(\phi) \geq 0$ for every ϕ, the limit (1.3) yields that $\mathsf{c}(L, 2) \leq \mathsf{c}(\Delta + \nabla V \nabla, 2)$, i.e., convergence is faster (although with the same asymptotic rate) in the non-symmetric case, i.e., if $Z \neq 0$.

If the dimension of M is larger than 3, existence of the limit (1.2) is currently an entirely open problem. In [10, section 1.3], it is conjectured that for $d \geq 5$, the constant $\mathsf{c}(\Delta + \nabla V \nabla, p)$ could be given by an expression in terms of the corresponding limiting constant for the Brownian interlacement occupation measure, although only an upper bound for $M = \mathbb{T}^d$ the flat torus and $V = 0$ is established [10, theorem 1.2].

In this communication we show the validity of (1.2) for $d = 4$ and $p = 2$.

Theorem 1.1. *With the notation introduced above, for a closed Riemannian manifold M with dimension $d = 4$, weighted volume measure μ, and the occupation measure μ_T of the diffusion process with generator $L = \Delta + \nabla V \nabla + Z$, it holds*

$$\sup_{x \in M} \left| \frac{T}{\log T} \mathbb{E}^x \left[\mathbb{W}_2^2 \left(\mu_T, \mu \right) \right] - \frac{\mathrm{vol}(M)}{8\pi^2} \right| \lesssim \sqrt{\frac{\log \log T}{\log T}}. \tag{1.6}$$

Thus, we explicitly compute that $\mathsf{c}(L, 2) = \mathrm{vol}(M)/(8\pi^2)$ for any 4-dimensional closed Riemannian manifold. In particular, we show that the leading term in the Wasserstein asymptotics does not depend on V, nor Z. In particular, the independence from the field Z can be explained as follows: in (1.3), the series of the terms $1/\lambda_i^2$ diverges logaritmically when $d = 4$, but the series of terms $\mathbf{V}(Z\phi_i)/\lambda_i^3$ is still convergent (see (2.8) below). This phenomenon, although novel in this setting, is not completely unexpected, for in the transport of i.i.d. samples on weighted two-dimensional manifolds [1–4] the leading term in the asymptotics also depends on the volume only.

The overall structure of the proof borrows from the literature of transport of i.i.d. samples, and in particular [1]: in Section 2, we prove Theorem 2.1 concerning the asymptotics for the transport cost between a smoothed empirical measure $\mu_{T,\varepsilon} := \hat{P}_\varepsilon^* \mu_T$ for $\varepsilon > 0$, where $\hat{P}_\varepsilon$ is the symmetric diffusion semigroup generated by $\hat{L} := \Delta + \nabla V \nabla$, and $\hat{P}_\varepsilon^*$ denotes its dual action on measures. Next, in Section 3, we provide an estimate on the expectation of $\mathbb{W}_2^2(\mu_T, \mu_{T,\varepsilon})$, which refines the simpler contraction estimate

$$\mathbb{W}_2^2(\mu_T, \mu_{T,\varepsilon}) \leq c\varepsilon, \quad T > 0, \, \varepsilon \in (0, 1) \tag{1.7}$$

valid for some constant $c = c(M) > 0$. We acknowledge that both these facts parallel the argument from [1], but they require new considerations and the introduction of different tools than the case of transport of i.i.d. points. Finally, in Section 4, we combine them to establish our main result.

Unless otherwise stated, we always take in what follows $d = 4$ and set $\varepsilon := (\log T)^\gamma / T$ for some fixed constant $\gamma > 3$ and consider the case of X being stationary, i.e., any marginal law of X equals μ. To keep notation simple we write $\mathbb{P}$ and $\mathbb{E}$ for expectation with respect to this law.

2 Asymptotics for the smoothed empirical measures

Theorem 2.1. *With the notation introduced above, on a four dimensional closed Riemannian manifold M, it holds*

$$\left|\frac{T}{\log T}\mathbb{E}[\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu)] - \frac{\mathrm{vol}(M)}{8\pi^2}\right| \lesssim \frac{\log\log T}{\log T}.$$

To prove this result, we rely on a well-known bound for $\mathbb{W}_2$ in terms of the H^{-1} norm of the density $u_{T,\varepsilon}$ of the smoothed empirical measure $\mu_{T,\varepsilon}$ with respect to μ.

We introduce first some notation. Let $\hat{p}_t$ be the (symmetric) heat kernel of $\hat{P}_t$ with respect to μ. Then, we have

$$\mu_{T,\varepsilon} = \hat{P}_\varepsilon^* \mu_T =: u_{T,\varepsilon}\mu, \quad u_{T,\varepsilon} := \frac{1}{T}\int_0^T \hat{p}_\varepsilon(X_t,\cdot)\mathrm{d}t. \tag{2.1}$$

Next, we consider the Poisson kernel

$$q_\varepsilon(x,y) := \int_0^\infty [\hat{p}_{t+\varepsilon}(x,y) - 1]\mathrm{d}t, \quad x,y \in M,$$

so that

$$f_{T,\varepsilon} := (-\hat{L})^{-1}(u_{T,\varepsilon} - 1) = \frac{1}{T}\int_0^T q_\varepsilon(X_t,\cdot)\mathrm{d}t. \tag{2.2}$$

In the following, we first estimate the expectation of $\mu(|\nabla f_{T,\varepsilon}|^2) := \int_M |\nabla f_{T,\varepsilon}|^2 \mathrm{d}\mu$, then prove the above theorem by comparing it with $\mathbb{E}[\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu)]$.

2.1 Estimate of $\mathbb{E}\left[\mu(|\nabla f_{T,\varepsilon}|^2)\right]$

Recall that, for $i \geq 1$, ϕ_i denotes a (zero mean) unit norm eigenfunction (for $\hat{L}$) with eigenvalue $-\lambda_i < 0$, i.e. $\|\phi_i\|_{L^2(\mu)} = 1$ and $\hat{L}\phi_i = -\lambda_i\phi_i$. We define

$$\psi_i(T) := \frac{1}{\sqrt{T}}\int_M \phi_i \mathrm{d}\mu_T = \frac{1}{\sqrt{T}}\int_0^T \phi_i(X_t)\mathrm{d}t, \quad T > 0,$$

where the factor $1/\sqrt{T}$ arises from the central limit theorem (1.5). By (2.1) and the spectral representation $\hat{p}_\varepsilon(x,y) - 1 = \sum_{i=1}^\infty \mathrm{e}^{-\lambda_i\varepsilon}\phi_i(x)\phi_i(y)$, we have

$$u_{T,\varepsilon} - 1 = \frac{1}{\sqrt{T}}\sum_{i=1}^\infty \mathrm{e}^{-\lambda_i\varepsilon}\psi_i(T)\phi_i. \tag{2.3}$$

Therefore,

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

Since $(\phi_i)_{i\geq 1}$ is an orthonormal sequence in $L^2(\mu)$, we find after an integration by parts

$$\mathbb{E}\left[\mu\left(|\nabla f_{T,\varepsilon}|^2\right)\right] = \mathbb{E}\left[\int_M f_{T,\varepsilon}(-\hat{L})f_{T,\varepsilon}\mathrm{d}\mu\right] = \frac{1}{T}\sum_{i=1}^{\infty}\frac{\mathrm{e}^{-2\lambda_i\varepsilon}}{\lambda_i}\mathbb{E}[|\psi_i(T)|^2]. \tag{2.4}$$

We claim that the following expansion for $\mathbb{E}\big[|\psi_i(T)|^2\big]$ holds (we keep the dimension d general, for possible future reference).

Proposition 2.2. *For any closed Riemannian manifold M with dimension $d \geq 1$, there exists a constant $c = c(M, L) > 0$ such that*

$$\left|\mathbb{E}\big[|\psi_i(T)|^2\big] - \frac{2}{\lambda_i} + \frac{2}{\lambda_i^2}\mathbf{V}(Z\phi_i)\right| \leq \frac{c}{\lambda_i(1+T)}, \quad i \geq 1,\, T > 0. \tag{2.5}$$

Using this in our four-dimensional setting we deduce the main bound for this part:

$$\left|\frac{T}{\log T}\mathbb{E}\left[\mu\left(|\nabla f_{T,\varepsilon}|^2\right)\right] - \frac{\mathrm{vol}(M)}{8\pi^2}\right| \lesssim \frac{\log\log T}{\log T}. \tag{2.6}$$

Indeed, combining $\|\nabla\phi_i\|_{L^2(\mu)} = \sqrt{\lambda_i}$ with the gradient estimate from [12, Lemma 3.1], we find that for some constants $c = c(M, L) < \infty$ and $\lambda = \lambda(M, L) > 0$, it holds

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

and we have, integrating by parts and using that $\mathrm{div}_\mu Z = 0$,

$$\big|\mu((Z\phi_i)P_t(Z\phi_i))\big| = \big|\mu(\phi_i Z P_t(Z\phi_i))\big| \leq c\|Z\|_\infty^2(1\wedge t)^{-\frac{1}{2}}\mathrm{e}^{-\lambda t}\sqrt{\lambda_i}.$$

So, $\mathbf{V}(Z\phi_i) \leq c_1\sqrt{\lambda_i}$ for some constant $c_1 = c_1(M, L) < \infty$. Next, we recall the small time asymptotics for the heat trace [5, Corollary 3.2.]:

$$\sum_{i=1}^{\infty}\mathrm{e}^{-t\lambda_i} = \mathrm{tr}\,\mathrm{e}^{t\hat{L}} - 1 = \frac{\mathrm{vol}(M)}{16\pi^2 t^2} + O(t^{-1}), \qquad \text{as } t \to 0. \tag{2.7}$$

By standard Tauberian arguments we find for the eigenvalue counting function $N(\lambda) := \sum_{i=1}^{\infty} 1_{\{\lambda_i\leq\lambda\}}$ the asymptotics $\lambda^{-2}N(\lambda) = \mathrm{vol}(M)/(32\pi^2) + o(\lambda^{-2})$, which yields convergence of the series

$$\sum_{i=1}^{\infty}\frac{1}{\lambda_i^3}\mathbf{V}(Z\phi_i) \leq c_1\sum_{i=1}^{\infty}\lambda_i^{-5/2} = \frac{5c_1}{2}\int_0^{\infty}\lambda^{-5/2-1}N(\lambda)\mathrm{d}\lambda < \infty; \tag{2.8}$$

as well as the asymptotics

$$\sum_{i=1}^{\infty}\frac{\mathrm{e}^{-2\varepsilon\lambda_i}}{\lambda_i^2} = \frac{\mathrm{vol}(M)}{16\pi^2}\log(\varepsilon^{-1}) + O(1), \quad \text{as } \varepsilon \to 0, \tag{2.9}$$

which can be also obtained directly from (2.7) and integrating by parts. Indeed,

$$\sum_{i=1}^{\infty}\frac{\mathrm{e}^{-\lambda_i s}}{\lambda_i}=\int_s^1\sum_{i=1}^{\infty}\mathrm{e}^{-\lambda_i t}\mathrm{d}t+O(1)=\frac{\mathrm{vol}(M)}{16\pi^2 s}+O(\log s^{-1})\quad\text{as } s\to 0, \tag{2.10}$$

and

$$\begin{aligned}\sum_{i=1}^{\infty}\frac{\mathrm{e}^{-2\varepsilon\lambda_i}}{\lambda_i^2}&=\int_{2\varepsilon}^{\infty}\sum_{i=1}^{\infty}\frac{\mathrm{e}^{-\lambda_i s}}{\lambda_i}\mathrm{d}s=\int_{2\varepsilon}^{1}\sum_{i=1}^{\infty}\frac{\mathrm{e}^{-\lambda_i s}}{\lambda_i}\mathrm{d}s+O(1)\\&=\int_{2\varepsilon}^{1}\Big[\frac{\mathrm{vol}(M)}{16\pi^2 s}+O(\log s^{-1})\Big]\mathrm{d}s+O(1)=\frac{\mathrm{vol}(M)}{16\pi^2}\log(\varepsilon^{-1})+O(1),\quad\text{as } \varepsilon\to 0.\end{aligned}$$

Combining the claim (2.5) with these estimates in (2.5) yields (2.6).

Proof of Proposition 2.2. By [12, Lemma 4.2(1)], we have the identity

$$\mathbf{V}(\phi_i)=\frac{1}{\lambda_i}-\frac{1}{\lambda_i^2}\mathbf{V}(Z\phi_i). \tag{2.11}$$

By the Markov property and stationarity of X, it follows that (writing $\mu(g)=\int_M g\mathrm{d}\mu$)

$$\begin{aligned}\mathbb{E}\big[|\psi_i(T)|^2\big]&=\frac{2}{T}\int_0^T\mathrm{d}t_1\int_{t_1}^T\mu(\phi_i P_{t_2-t_1}\phi_i)\mathrm{d}t_2=\frac{2}{T}\int_0^T\mathrm{d}t_1\int_0^{T-t_1}\mu(\phi_i P_t\phi_i)\mathrm{d}t\\&=2\Big(\frac{1}{\lambda_i}-\frac{1}{\lambda_i^2}\mathbf{V}(Z\phi_i)\Big)-\frac{2}{T}\int_0^T\mathrm{d}t_1\int_{T-t_1}^{\infty}\mu(\phi_i P_t\phi_i)\mathrm{d}t.\end{aligned} \tag{2.12}$$

We next evaluate the integrand $\mu(\phi_i P_t\phi_i)$. By Duhamel's formula

$$P_t f=\hat{P}_t f+\int_0^t P_s(Z\hat{P}_{t-s}f)\mathrm{d}s,\quad t\geq 0 \tag{2.13}$$

and ϕ_i being an eigenfunction for $\hat{L}$, so that $\hat{P}_t\phi_i=\mathrm{e}^{-\lambda_i t}\phi_i$, we have

$$P_t\phi_i=\mathrm{e}^{-\lambda_i t}\phi_i+\int_0^t\mathrm{e}^{-\lambda_i(t-s)}P_s(Z\phi_i)\mathrm{d}s, \tag{2.14}$$

and therefore

$$\mu(\phi_i P_t\phi_i)=\mathrm{e}^{-\lambda_i t}+\int_0^t\mathrm{e}^{-\lambda_i(t-s)}\mu\big((P_s^*\phi_i)Z\phi_i\big)\mathrm{d}s, \tag{2.15}$$

where we write P_t^* for the semigroup with generator $L^*:=\hat{L}-Z$. Noting that $\mu(\phi_i Z\phi_i)=0$ and using Duhamel's formula (2.13) again

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

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 derive

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

Combining with (2.12) we finish the proof. □

2.2 Estimate on $\mathbb{E}[|\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu) - \mu(|\nabla f_{T,\varepsilon}|^2)|]$

The main result in this part is the following (with the assumptions $d = 4$ and on ε).

Proposition 2.3. *It holds*

$$\mathbb{E}\left[\left|\mathbb{W}_2\left(\mu_{T,\varepsilon},\mu\right) - \sqrt{\mu(|\nabla f_{T,\varepsilon}|^2)}\right|\right] \lesssim \sqrt{\frac{1}{T\log T}}, \tag{2.16}$$

and

$$\mathbb{E}\left[\left|\mathbb{W}_2^2\left(\mu_{T,\varepsilon},\mu\right) - \mu(|\nabla f_{T,\varepsilon}|^2)\right|\right] \lesssim \frac{1}{T}. \tag{2.17}$$

To prove this result, we need some lemmas. Let $\hat{L}_x$ and ∇_y^n, for $n \in \mathbb{N}$ stand for the corresponding operators acting on variables $x, y \in M$ respectively (let ∇^0 be the identity operator). We state and prove the following result for general dimension d and $\varepsilon \in (0,1)$, for possible future reference.

Lemma 2.4. *For any $n \in \mathbb{N}$ and $p \in (1,\infty)$, there exists a constant $c = (n,p,M,V) < \infty$ such that, for any $\varepsilon \in (0,1)$,*

$$\sup_{y\in M}\int_M \big|\nabla_x\hat{L}_x^{-1}\nabla_y^n q_\varepsilon(x,y)\big|^p\mu(\mathrm{d}x) \le c\begin{cases}\varepsilon^{-\frac{(d+n-3)p-d}{2}}, & \text{if } (d+n-3)p > d;\\ \{\log(1+\varepsilon^{-1})\}^p, & \text{if } (d+n-3)p = d;\\ 1, & \text{if } (d+n-3)p < d.\end{cases} \tag{2.18}$$

and

$$\sup_{y\in M}\int_M \big|\nabla_y^n q_\varepsilon(x,y)\big|^p\mu(\mathrm{d}x) \le c\begin{cases}\varepsilon^{-\frac{(d+n-2)p-d}{2}}, & \text{if } (d+n-2)p > d;\\ \{\log(1+\varepsilon^{-1})\}^p, & \text{if } (d+n-2)p = d;\\ 1, & \text{if } (d+n-2)p < d.\end{cases} \tag{2.19}$$

Proof. These bounds could be established by the same argument in [1, Corollary 3.13] using pointwise upper bound of $\big|\nabla_x \hat{L}_x^{-1}\nabla_y^n q_\varepsilon\big|$. Here we provide an alternative approach. Let ρ be the Riemannian distance on M, so that the standard heat kernel bounds give, for some constants $c\in(1,\infty)$ and $\lambda>0$,

$$\big|\nabla_y^n\big(\hat{p}_t(x,y)-1\big)\big| \le ct^{-\frac{d+n}{2}}\mathrm{e}^{-\lambda t-\frac{\rho(x,y)^2}{ct}},\quad x,y\in M,\, t>0. \tag{2.20}$$

Consequently, given $1<p<\infty$, for every $y\in M$ and $t>0$,

$$\begin{aligned}&\int_M |\nabla_y^n\big(\hat{p}_t(x,y)-1\big)|^p\mu(\mathrm{d}x)\\ &\le c_1 t^{-\frac{(d+n)p}{2}}\bigg(\int_{\{x\,;\,\rho(x,y)\le\sqrt{t}\}}+\sum_{k=1}^\infty\int_{\{x\,;\,k\sqrt{t}<\rho(x,y)\le(k+1)\sqrt{t}\}}\bigg)\mathrm{e}^{-p\lambda t-\frac{p\rho(x,y)^2}{ct}}\mu(\mathrm{d}x)\\ &\le ct^{-\frac{(d+n)p}{2}+\frac{d}{2}}\mathrm{e}^{-p\lambda t}\Big(1+\sum_{k=1}^\infty(k+1)^d\mathrm{e}^{-\frac{pk^2}{c}}\Big)\le ct^{-\frac{(d+n)p}{2}+\frac{d}{2}}\mathrm{e}^{-p\lambda t},\end{aligned}$$

where we conventionally keep denoting with c possibly different constants, and we also used the fact that $\sup_{y\in M}\mu(\{x\,;\,\rho(x,y)\le r\})\le cr^d$ for some (possibly different) constant $c<\infty$. Then,

$$\sup_{y\in M}\big\|\nabla_y^n\big(\hat{p}_t(\cdot,y)-1\big)\big\|_{L^p(\mu)}\le ct^{-\frac{d+n}{2}+\frac{d}{2p}}\mathrm{e}^{-\lambda t}. \tag{2.21}$$

On the other hand, notice that

$$\nabla_x(-\hat{L}_x)^{-1}\nabla_y^n\hat{p}_t(x,y)=\nabla_x(-\hat{L}_x)^{-\frac12}\nabla_y^n(-\hat{L}_x)^{-\frac12}\hat{p}_t(x,y)=\nabla_x(-\hat{L}_x)^{-\frac12}\nabla_y^n(-\hat{L}_y)^{-\frac12}\hat{p}_t(x,y).$$

By the definition of q_ε and the L^p-boundedness of the Riesz transform, we find a constant c such that

$$I:=\big\|\nabla(-\hat{L})^{-1}\nabla_y^n q_\varepsilon(\cdot,y)\big\|_{L^p(\mu)}\le c\big\|\nabla_y^n(-\hat{L}_y)^{-\frac12}q_\varepsilon(\cdot,y)\big\|_{L^p(\mu)}. \tag{2.22}$$

Since

$$(-\hat{L}_y)^{-\frac12}q_\varepsilon(x,y)=\frac{1}{\sqrt{\pi}}\int_0^\infty\frac{1}{\sqrt{s}}\hat{P}_t q_\varepsilon(x,\cdot)(y)\mathrm{d}t=\frac{1}{\sqrt{\pi}}\int_0^\infty\mathrm{d}t\int_0^\infty\frac{\hat{p}_{t+s+\varepsilon}(x,y)-1}{\sqrt{s}}\mathrm{d}s,$$

(2.21) and (2.22) yield

$$\begin{aligned}I&\le c\int_0^\infty\mathrm{d}t\int_0^\infty\frac{\big\|\nabla_y^n\big(\hat{p}_{t+s+\varepsilon}(\cdot,y)-1\big)\big\|_{L^p(\mu)}}{\sqrt{s}}\mathrm{d}s\\ &\le c\int_0^\infty s^{-\frac12}\mathrm{d}s\int_0^\infty\mathrm{e}^{-\lambda(t+s+\varepsilon)}(t+s+\varepsilon)^{-\frac{d+n}{2}+\frac{d}{2p}}\mathrm{d}t\\ &\le c\int_0^\infty s^{-\frac12}\mathrm{e}^{-\lambda s}\Big\{(s+\varepsilon)^{-\frac{[(d+n-2)p-d]^+}{2p}}+1_{\{(d+n-2)p=d\}}\log\{1+(s+\varepsilon)^{-1}\}\Big\}\mathrm{d}s.\end{aligned}$$

This implies the desired estimate (2.18). Similarly,

$$\left\|\nabla_y^n q_\varepsilon\right\|_{L^p(\mu)} \leq \int_\varepsilon^\infty \left\|\nabla_y^n\big(\hat{p}_t(\cdot,y)-1\big)\right\|_{L^p(\mu)}\mathrm{d}t.$$

Combined this with (2.21), the proof of (2.19) is complete. □

Remark 2.5. *It is easy to see that* (2.20) *also implies that for any* $n\in\mathbb{N}$ *such that* $d+n>2$,

$$\left\|\nabla^n f_{T,\varepsilon}\right\|_\infty \leq c\int_\varepsilon^\infty \sup_{x,y\in M}\left|\nabla_y^n\big(\hat{p}_t(x,y)-1\big)\right|\mathrm{d}t \lesssim \varepsilon^{-\frac{d+n-2}{2}}. \tag{2.23}$$

The second step towards the proof of Proposition 2.3 is to evaluate the probability of the event

$$A_{T,\varepsilon}^\xi := \left\{\left\|\nabla^2 f_{T,\varepsilon}\right\|_\infty \leq \xi\right\}, \tag{2.24}$$

for $\xi>0$. To this aim, we collect the following concentration inequality for diffusion processes, see also [13, Corollary 3.2].

Lemma 2.6. *Assume that the dimension of* M *is* $d\geq 3$. *Then, there exists a constant* $c=c(M,L)\in(0,\infty)$ *such that, for every* $g\in L^{d/2}(M)$ *with zero mean (i.e.,* $\mu(g)=0$*) and* $T,\xi>0$,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{T}\int_0^T g(X_t)\mathrm{d}t\right|>\xi\right) &\leq 2\exp\left(-\frac{2T\xi^2}{\sigma^2(g)\big(\sqrt{1+2c\|g\|_{L^{d/2}(\mu)}\xi/\sigma^2(g)}+1\big)^2}\right)\\ &\leq 2\exp\left(-\frac{T\xi^2}{2\big(\sigma^2(g)\,+\,c\|g\|_{L^{d/2}(\mu)}\xi\big)}\right),\end{aligned} \tag{2.25}$$

where

$$\sigma^2(g) := 2\int_M g(-\hat{L})^{-1}g\,\mathrm{d}\mu = \int_M\left|\nabla\int_0^\infty \hat{P}_t g\right|^2\mathrm{d}\mu.$$

Proof. By [16, theorem 1] we have

$$\mathbb{P}\left(\left|\frac{1}{T}\int_0^T g(X_t)\mathrm{d}t\right|>\xi\right) \leq 2\exp\big(-TI_g(\xi-)\big), \quad \forall\, T,\xi>0, \tag{2.26}$$

where

$$I_g(\xi) := \inf\left\{\mu(|\nabla h|^2)\,;\, h\in W^{2,1}(\mu), \mu(h^2)=1,\, |\mu(h^2g)|=\xi,\, \mu(h^2|g|)<\infty\right\}$$

and

$$I_g(\xi-) := \lim_{\varepsilon\to 0+} I_g(\xi-\varepsilon).$$

Following to the argument used in [7, theorem 2.2], for every $h\in W^{2,1}(\mu)$ with $\mu(h^2)=1$, we notice that

$$\frac{2|\mu(gh^2)|^2}{\sigma^2(g)\big(\sqrt{1+2c\|g\|_{L^{d/2}(\mu)}|\mu(gh^2)|/\sigma^2(g)}+1\big)^2} \leq \mu(|\nabla h|^2)$$

is equivalent to

$$|\mu(gh^2)| \le \sqrt{2\sigma^2(g)\mu(|\nabla h|^2)} + c\|g\|_{L^{d/2}(\mu)}\mu(|\nabla h|^2). \tag{2.27}$$

So, by (2.26), it suffices to verify (2.27) for some constant $c > 0$. We write

$$\mu(gh^2) = 2\bar{h}\mu(gh) + \mu\big(g(h-\bar{h})^2\big),$$

where $\bar{h} := \mu(h)$. Since $|\bar{h}| \le \mu(h^2)^{1/2} = 1$ and by Cauchy-Schwarz inequality

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

we obtain

$$|2\bar{h}\mu(hg)| \le \sqrt{2\sigma^2(g)\mu(|\nabla h|^2)}.$$

On the other hand, by the Sobolev-Poincaré inequality on M, it follows that $h \in L^{\frac{2d}{d-2}}(\mu)$ and there exists $c < \infty$ such that

$$\|h - \bar{h}\|_{L^{\frac{2d}{d-2}}(\mu)} \le c\mu(|\nabla h|^2)^{1/2}.$$

Combined with Hölder's inequality, it follows that

$$\big|\mu\big(g(h-\bar{h})^2\big)\big| \le \|g\|_{L^{d/2}(\mu)}\|h-\bar{h}\|^2_{L^{\frac{2d}{d-2}}(\mu)} \le c^2\|g\|_{L^{d/2}(\mu)}\mu(|\nabla h|^2),$$

which implies (2.27) up to replacing c with c^2. □

Back to our four-dimensional setting, we apply the concentration inequality to estimate the probability of $A^{\xi}_{T,\varepsilon}$ with $\xi = 1/(\log T)$, which is sufficient for our purposes.

Lemma 2.7. *There exists a constant $C = C(M, L, \gamma) > 0$ such that, for $\xi = 1/\log T$, it holds*

$$\mathbb{P}\big((A^{\xi}_{T,\varepsilon})^{\mathsf{c}}\big) \lesssim \exp\big(-C(\log T)^{\gamma-2}\big).$$

Proof. For fixed $y \in M$, applying (2.25) with $g = \nabla^2_y q_\varepsilon(\cdot, y)$, and using Lemma 2.4 with $p = n = 2$ and $d = 4$, we find a constant $C > 0$ such that

$$\begin{aligned}\mathbb{P}\left(\left|\nabla^2_y f_{T,\varepsilon}(y)\right| > \xi/2\right) &\lesssim \exp\left(-\frac{T\xi^2}{2\int_M |\nabla_x \hat{L}_x^{-1}\nabla^2_y q_\varepsilon(x,y)|^2\mu(\mathrm{d}x) + c\|\nabla^2_y q_\varepsilon(\cdot,y)\|_{L^2(\mu)}\xi}\right)\\ &\lesssim \exp\big(-C(\log T)^{\gamma-2}\big).\end{aligned}$$

Furthermore, by (2.23) with $n = 3$, we can always bound the Lipschitz constant of $y \mapsto |\nabla^2 f_{T,\varepsilon}|(y)$ in terms of

$$K := \left\|\nabla^3 f_{T,\varepsilon}\right\|_\infty \lesssim \varepsilon^{-\frac{5}{2}}.$$

Thus, choosing a suitable ℓ-net with $K \cdot \ell = \xi/2$, hence with $N(\ell) \lesssim (K/\xi)^4$ elements, we obtain that

$$\mathbb{P}\left(\sup_{y\in M}\left|\nabla^2_y f_{T,\varepsilon}(y)\right| > \xi\right) \lesssim N(\ell)\cdot\exp(-C(\log T)^{\gamma-2}) \lesssim \varepsilon^{-10}\xi^{-4}\exp(-C(\log T)^{\gamma-2}).$$

This implies the desired estimate, for a smaller constant $C > 0$. □

Proof of Proposition 2.3. We set $\xi = 1/\log T$ and consider the event $A^{\xi}_{T,\varepsilon}$. By Lemma 2.7, (2.16) follows if

$$\mathbb{E}\left[1_{A^{\xi}_{T,\varepsilon}}\left|\mathbb{W}_2\left(\mu_{T,\varepsilon},\mu\right)-\sqrt{\mu(|\nabla f_{T,\varepsilon}|^2)}\right|\right]\lesssim\sqrt{\frac{1}{T\log T}}. \tag{2.28}$$

To prove this estimate, we introduce the probability measure $\hat{\mu}_{T,\varepsilon} := \exp(\nabla f_{T,\varepsilon})_{\#}\mu$. By [8, theorem 1.1] on $A^{\xi}_{T,\varepsilon}$ the map $\nabla f_{T,\varepsilon}$ is the optimal map transforming from μ to $\hat{\mu}_{T,\varepsilon}$, so that

$$\mathbb{W}_2^2\left(\hat{\mu}_{T,\varepsilon},\mu\right)=\mu(|\nabla f_{T,\varepsilon}|^2). \tag{2.29}$$

Next, we argue that, still on $A^{\xi}_{T,\varepsilon}$

$$\mathbb{W}_2^2\left(\mu_{T,\varepsilon},\hat{\mu}_{T,\varepsilon}\right)\lesssim\xi^2\mu(|\nabla f_{T,\varepsilon}|^2). \tag{2.30}$$

This is a consequence of the Dacorogna-Moser interpolation scheme: since $f_{T,\varepsilon} = (-\hat{L})^{-1}(u_{T,\varepsilon} - 1)$ and $\mathrm{div}_\mu \circ \nabla = \hat{L}$, where, the function $u_s := (1-s) + su_{T,\varepsilon}$ and the time-dependent vector field

$$Y_s := \frac{\nabla f_{T,\varepsilon}}{u_s},\quad s\in[0,1]$$

satisfy the equation

$$\frac{\mathrm{d}}{\mathrm{d}s}u_s+\mathrm{div}_\mu\big(u_sY_s\big)=0.$$

Then, by [1, Proposition A.1.], one obtains (2.30).

By the triangle inequality, we derive

$$\mathbb{E}\left[1_{A^{\xi}_{T,\varepsilon}}\left|\mathbb{W}_2\left(\mu_{T,\varepsilon},\mu\right)-\sqrt{\mu(|\nabla f_{T,\varepsilon}|^2)}\right|\right]\lesssim\xi\mathbb{E}\left[\mu(|\nabla f_{T,\varepsilon}|^2)\right]^{1/2}. \tag{2.31}$$

This together with (2.6) implies (2.28), and hence (2.16) is proved.

To prove the other estimate, write

$$\mathbb{E}\left[\left|\mathbb{W}_2^2\left(\mu_{T,\varepsilon},\mu\right)-\mu(|\nabla f_{T,\varepsilon}|^2)\right|\right]\le\big(\mathbb{E}\left[A^2\right]\big)^{1/2}\Big(\big(\mathbb{E}\left[A^2\right]\big)^{1/2}+2\big(\mathbb{E}\left[B^2\right]\big)^{1/2}\Big),$$

where

$$A=\mathbb{W}_2\left(\mu_{T,\varepsilon},\mu\right)-\sqrt{\mu(|\nabla f_{T,\varepsilon}|^2)},\quad B=\sqrt{\mu(|\nabla f_{T,\varepsilon}|^2)}.$$

Noting that (2.29) and the triangle inequality imply

$$\begin{aligned}\mathbb{E}\left[A^2\right]&=\mathbb{E}\left[\left|\mathbb{W}_2(\mu_{T,\varepsilon},\mu)-\sqrt{\mu(|\nabla f_{T,\varepsilon}|^2)}\right|^2\right]\\&\le\mathbb{E}\left[\mathbb{W}_2^2\big(\mu_{T,\varepsilon},\hat{\mu}_{T,\varepsilon}\big)\right]\\&\le D^2\mathbb{P}\big((A^{\xi}_{T,\varepsilon})^c\big)+\mathbb{E}\big[1_{A^{\xi}_{T,\varepsilon}}\mathbb{W}_2^2\big(\mu_{T,\varepsilon},\hat{\mu}_{T,\varepsilon}\big)\big]\\&\lesssim\xi^2\mu(|\nabla f_{T,\varepsilon}|^2)\lesssim\frac{1}{T\log T},\end{aligned}$$

where D is the diameter of M. Then (2.17) follows from (2.6), Lemma 2.7 and (2.30).

□

3 Improved contractivity estimate

Aim of this section is to establish the following improved version of (1.7).

Lemma 3.1. *With the notation introduced above, on a four dimensional closed Riemannian manifold M, it holds*

$$\mathbb{E}\left[\mathbb{W}_2^2\left(\mu_T,\mu_{T,\varepsilon}\right)\right]\lesssim\frac{\log\log T}{T}. \tag{3.1}$$

Proof. Let $\xi=1/\log T$. Then on the event $A^{\xi}_{T,\varepsilon}$ introduced in (2.24), we have

$$\|u_{T,\varepsilon}-1\|_\infty=\|\hat{L}f_{T,\varepsilon}\|_\infty\lesssim\|\nabla^2 f_{T,\varepsilon}\|_\infty\le\xi=1/(\log T)$$

where we have used the fact that $|\nabla V\nabla f_{T,\varepsilon}|\lesssim\|\nabla V\|_\infty\|\nabla^2 f_{T,\varepsilon}\|_\infty$. Combining this with Ledoux's upper bounds for $\mathbb{W}_2$ (see [1,9] or [11, Lemma A.1]), we have for large $T>0$ and any $\varepsilon'\in(0,\varepsilon)$,

$$\mathbb{W}_2^2(\mu_{T,\varepsilon'},\mu_{T,\varepsilon})\le 4\int_M\frac{|\nabla(-\hat{L})^{-1}(u_{T,\varepsilon'}-u_{T,\varepsilon})|^2}{u_{T,\varepsilon}}\mathrm{d}\mu\le 8\int_M|\nabla(-\hat{L})^{-1}(u_{T,\varepsilon'}-u_{T,\varepsilon})|^2\mathrm{d}\mu.$$

Then, on $A^{\xi}_{T,\varepsilon}$ we find

$$\begin{aligned}\mathbb{W}_2^2(\mu_{T,\varepsilon'},\mu_{T,\varepsilon})&\lesssim\int_M|\nabla(-\hat{L})^{-1}(u_{T,\varepsilon'}-u_{T,\varepsilon})|^2\mathrm{d}\mu=\frac{1}{T}\sum_{i=1}^\infty\frac{(\mathrm{e}^{-\lambda_i\varepsilon'}-\mathrm{e}^{-\lambda_i\varepsilon})^2}{\lambda_i}|\psi_i(T)|^2\\&\le\frac{1}{T}\sum_{i=1}^\infty\frac{\mathrm{e}^{-2\lambda_i\varepsilon'}-\mathrm{e}^{-2\lambda_i\varepsilon}}{\lambda_i}|\psi_i(T)|^2=2\int_{\varepsilon'}^{\varepsilon}\|u_{T,s}-1\|^2_{L^2(\mu)}\mathrm{d}s.\end{aligned}\tag{3.2}$$

Next, by (2.3), Lemma 2.2 and (2.10), we have

$$\mathbb{E}\left[\|u_{T,s}-1\|^2_{L^2(\mu)}\right]=\frac{1}{T}\sum_{i=1}^\infty\mathrm{e}^{-2\lambda_i s}\mathbb{E}\left[|\psi_i(T)|^2\right]\lesssim T^{-1}\sum_{i=1}^\infty\lambda_i^{-1}\mathrm{e}^{-2\lambda_i s}\lesssim T^{-1}s^{-1}.$$

By (3.2), this gives

$$\mathbb{E}\big[1_{A^{\xi}_{T,\varepsilon}}\mathbb{W}_2^2(\mu_{T,\varepsilon'},\mu_{T,\varepsilon})\big]\lesssim T^{-1}\log(\varepsilon/\varepsilon').$$

Then, using the triangle inequality and the fact that $\mathbb{E}\big[\mathbb{W}_2^2(\mu_T,\mu_{T,\varepsilon'})\big]\lesssim\varepsilon'$ by (1.7), we derive

$$\begin{aligned}\mathbb{E}\left[1_{A^{\xi}_{T,\varepsilon}}\mathbb{W}_2^2\left(\mu_T,\mu_{T,\varepsilon}\right)\right]&\lesssim\mathbb{E}\big[\mathbb{W}_2^2(\mu_T,\mu_{T,\varepsilon'})\big]+\mathbb{E}\big[1_{A^{\xi}_{T,\varepsilon}}\mathbb{W}_2^2(\mu_{T,\varepsilon'},\mu_{T,\varepsilon})\big]\\&\lesssim\varepsilon'+T^{-1}\log(\varepsilon/\varepsilon').\end{aligned}$$

Choosing finally $\varepsilon'=\frac{\log\log T}{T}$ and combining with Proposition 2.7, we finish the proof. ☐

4 Proof of theorem 1.1

We consider first the case of X being stationary. In this situation, we argue that

$$\left|\frac{T}{\log T}\mathbb{E}[\mathbb{W}_2^2(\mu_T,\mu)]-\frac{\mathrm{vol}(M)}{8\pi^2}\right|\lesssim\sqrt{\frac{\log\log T}{\log T}}. \tag{4.1}$$

Indeed, by (2.6) and Proposition 2.3 we have

$$\left|\frac{T}{\log T}\mathbb{E}[\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu)]-\frac{\mathrm{vol}(M)}{8\pi^2}\right|\lesssim\frac{\log\log T}{\log T}.$$

Combining this with Lemma 3.1 implies

$$\begin{aligned}
&\left|\frac{T}{\log T}\mathbb{E}[\mathbb{W}_2^2(\mu_T,\mu)]-\frac{\mathrm{vol}(M)}{8\pi^2}\right|\\
&\lesssim\frac{\log\log T}{\log T}+\frac{T}{\log T}\Big|\mathbb{E}[\mathbb{W}_2^2(\mu_T,\mu)-\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu)]\Big|\\
&\le\frac{\log\log T}{\log T}+\frac{T}{\log T}\Big(\mathbb{E}[\mathbb{W}_2^2(\mu_T,\mu_{T,\varepsilon})]+2\sqrt{\mathbb{E}[\mathbb{W}_2^2(\mu_T,\mu_{T,\varepsilon})]\cdot\mathbb{E}[\mathbb{W}_2^2(\mu_T,\mu)]}\Big)\\
&\lesssim\frac{\log\log T}{\log T}+\sqrt{\frac{\log\log T}{\log T}}\lesssim\sqrt{\frac{\log\log T}{\log T}}.
\end{aligned}$$

So, (4.1) holds in the case of X stationary.

To address the general case, for any $x\in M$, let (X_t^x,X_t^μ) be the coupling by reflection for the diffusions generated by L with initial distributions δ_x and μ respectively (so that X^μ is stationary). According to the proof of [6, Theorem 1], there exists an increasing function $g:[0,D]\to[0,\infty)$, where D is the diameter of M, such that

$$c_1r\le g(r)\le c_2r,\quad r\in[0,D]$$

holds for some constants $c_2>c_1>0$ (independent of x) and that

$$\mathrm{d}g(\rho(X_t^x,X_t^\mu))\le \mathrm{d}M_t-\delta g(\rho(X_t^x,X_t^\mu))\mathrm{d}t$$

holds for some martingale M_t and a constant $\delta>0$. Therefore by taking expectation, we obtain

$$\mathbb{E}[g(\rho(X_t^x,X_t^\mu))]\le\mu\big(g(\rho(x,\cdot))\big)\mathrm{e}^{-\delta t}.$$

Consequently, writing

$$\mu_T^x:=\frac{1}{T}\int_0^T\delta_{X_t^x}\mathrm{d}t\quad\text{ and }\quad\mu_T^\mu:=\frac{1}{T}\int_0^T\delta_{X_t^\mu}\mathrm{d}t$$

satisfy

$$\mathbb{E}[\mathbb{W}_2^2(\mu_T^x,\mu_T^\mu)]\le\frac{1}{T}\int_0^T\mathbb{E}[\rho^2(X_t^x,X_t^\mu)]\mathrm{d}t\le\frac{c_1^{-1}c_2D^2}{T}\int_0^T\mathrm{e}^{-\delta t}\mathrm{d}t\le\frac{c_1^{-1}c_2D^2}{\delta T}.$$

We thus derive from (4.1) that

$$\begin{aligned}
&\sup_{x\in M}\big|\mathbb{E}^x[\mathbb{W}_2^2(\mu_T,\mu)]-\mathbb{E}[\mathbb{W}_2^2(\mu_T^\mu,\mu)]\big| = \sup_{x\in M}\big|\mathbb{E}[\mathbb{W}_2^2(\mu_t^x,\mu)-\mathbb{W}_2^2(\mu_T^\mu,\mu)]\big| \\
&\le \sup_{x\in M}\Big(\mathbb{E}[\mathbb{W}_2^2(\mu_T^x,\mu_T^\mu)]+2\mathbb{E}\big[\mathbb{W}_2(\mu_T^x,\mu_T^\mu)\mathbb{W}_2(\mu_T^\mu,\mu)\big]\Big) \\
&\lesssim \frac{1}{T}+\frac{1}{\sqrt{T}}\sqrt{\mathbb{E}[\mathbb{W}_2^2(\mu_T^\mu,\mu)]}\lesssim \frac{\sqrt{\log T}}{T}.
\end{aligned}$$

Applying (4.1) again, we derive that for large T,

$$\sup_{x\in M}\left|\frac{T}{\log T}\mathbb{E}^x[\mathbb{W}_2^2(\mu_{T,\varepsilon},\mu)]-\frac{\mathrm{vol}(M)}{8\pi^2}\right|\lesssim\sqrt{\frac{\log\log T}{\log T}}+\frac{1}{\sqrt{\log T}}\lesssim\sqrt{\frac{\log\log T}{\log T}},$$

and the proof is completed.

References

[1] Luigi Ambrosio and Federico Glaudo, *Finer estimates on the 2-dimensional matching problem*, J. Éc. polytech. Math. **6** (2019), 737–765.

[2] Luigi Ambrosio, Michael Goldman, and Dario Trevisan, *On the quadratic random matching problem in two-dimensional domains*, Electronic Journal of Probability **27** (2022), 1–35.

[3] Luigi Ambrosio, Federico Stra, and Dario Trevisan, *A PDE approach to a 2-dimensional matching problem*, Probab. Theory Relat. Fields **173** (2019), no. 1-2, 433–477.

[4] Dario Benedetto, Emanuele Caglioti, Sergio Caracciolo, Matteo d'Achille, Gabriele Sicuro, and Andrea Sportiello, *Random assignment problems on 2d manifolds*, Journal of Stfatistical Physics **183** (2021), no. 2, 1–40.

[5] Nelia Charalambous and Julie Rowlett, *The heat trace for the drifting laplacian and schrödinger operators on manifolds*, Asian Journal of Mathematics **23** (2019), no. 4, 539–560.

[6] Mu-Fa Chen and Feng-Yu Wang, *General formula for lower bound of the first eigenvalue on riemannian manifolds*, Science in China Series A: Mathematics **40** (1997), no. 4, 384–394.

[7] Fuqing Gao, Arnaud Guillin, and Liming Wu, *Bernstein-type concentration inequalities for symmetric markov processes*, Theory of Probability & Its Applications **58** (2014), no. 3, 358–382.

[8] Federico Glaudo, *On the c-concavity with respect to the quadratic cost on a manifold*, Nonlinear Analysis **178** (2019), 145–151.

[9] Michel Ledoux, *On optimal matching of Gaussian samples*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **457** (2017), no. Veroyatnost' i Statistika. 25, 226–264.

[10] Mauro Mariani and Dario Trevisan, *Wasserstein asymptotics for brownian motion on the flat torus and brownian interlacements*, arXiv preprint arXiv:2307.10325 (2023).

[11] Feng-Yu Wang, *Wasserstein convergence rate for empirical measures on noncompact manifolds*, Stochastic Processes and their Applications **144** (2022), 271–287.

[12] ______, *Convergence in wasserstein distance for empirical measures of non-symmetric subordinated diffusion processes*, arXiv preprint arXiv:2301.08420 (2023).

[13] Feng-Yu Wang, Bingyao Wu, and Jie-Xiang Zhu, *Sharp l^q-convergence rate in p-wasserstein distance for empirical measures of diffusion processes*, arXiv preprint arXiv:2408.09116 (2024).

[14] Feng-Yu Wang and Jie-Xiang Zhu, *Limit theorems in wasserstein distance for empirical measures of diffusion processes on riemannian manifolds*, Annales de l'Institut Henri Poincare (B) Probabilites et statistiques, vol. 59, Institut Henri Poincaré, 2023, pp. 437–475.

[15] Liming Wu, *Moderate deviations of dependent random variables related to clt*, The Annals of Probability (1995), 420–445.

[16] ______, *A deviation inequality for non-reversible markov processes*, Annales de l'Institut Henri Poincare (B) Probability and Statistics, vol. 36, Elsevier, 2000, pp. 435–445.

Acknowledgements.

D.T. acknowledges the MUR Excellence Department Project awarded to the Department of Mathematics, University of Pisa, CUP I57G22000700001, the HPC Italian National Centre for HPC, Big Data and Quantum Computing - Proposal code CN1 CN00000013, CUP I53C22000690001, the PRIN 2022 Italian grant 2022WHZ5XH - "understanding the LEarning process of QUantum Neural networks (LeQun)", CUP J53D23003890006, the INdAM-GNAMPA project 2024 "Tecniche analitiche e probabilistiche in informazione quantistica" and the project G24-202 "Variational methods for geometric and optimal matching problems" funded by Università Italo Francese. Research also partly funded by PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Commission under the NextGeneration EU programme.

F.-Y.W. and J.-X.Z. acknowledge the National Key R&D Program of China (No. 2022YFA1006000, 2020YFA0712900) and NNSFC (11921001).