ON OPTIMAL MATCHING OF GAUSSIAN SAMPLES III

BY

MICHEL LEDOUX (TOULOUSE) AND **JIE-XIANG ZHU** (SHANGHAI)

Abstract. This article is a continuation of the papers [10], [11] in which the optimal matching problem and the related rates of convergence of empirical measures for Gaussian samples are addressed. A further step in both the dimensional and Kantorovich parameters is achieved here, proving that, given independent random variables $X_1, \ldots, X_n$ with common distribution the standard Gaussian measure μ on $\mathbb{R}^d$, $d \geqslant 3$, and $\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ the associated empirical measure,

$$\mathbb{E}[\mathrm{W}_p^p(\mu_n, \mu)] \approx \frac{1}{n^{p/d}}$$

for any $1 \leqslant p < d$, where W_p is the pth Kantorovich–Wasserstein metric. That is, in this range, the rates are the same as in the uniform case. The proof relies on the pde and mass transportation approach developed by L. Ambrosio, F. Stra and D. Trevisan in a compact setting.

2020 Mathematics Subject Classification: Primary 60D05, 60F25; Secondary 60H15, 49J55, 58J35.

Key words and phrases: optimal matching, empirical measure, optimal transport, Gaussian sample, Mehler kernel.

1. INTRODUCTION AND MAIN RESULTS

Given $p \geqslant 1$, the Kantorovich, or Kantorovich–Wasserstein, distance (see [6] for the terminology) between two probability measures ν and μ on the Borel sets of $\mathbb{R}^d$ with a finite pth moment is defined by

$$(1.1) \qquad \mathrm{W}_p(\nu, \mu) = \inf \left(\int\limits_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, d\pi(x, y) \right)^{1/p}$$

where the infimum is taken over all couplings π on $\mathbb{R}^d \times \mathbb{R}^d$ with respective marginals ν and μ. Here $|x - y|$ denotes the Euclidean distance between x and y in $\mathbb{R}^d$.

Denote by $X_1, \ldots, X_n$, $n \geqslant 1$, independent random variables in $\mathbb{R}^d$ with common distribution μ and let

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

be the empirical measure on the sample $(X_1, \ldots, X_n)$. The question investigated here is the order of decay in n of the expectations

(1.2) $$\mathbb{E}[W_p^p(\mu_n, \mu)]$$

when the random variables $X_1, \ldots, X_n$ are independent with the same standard Gaussian distribution $d\mu(x) = e^{-|x|^2/2} \frac{dx}{(2\pi)^{d/2}}$ on $\mathbb{R}^d$. The optimal matching problem would consist in the study of $\mathbb{E}[W_p^p(\mu_n, \nu_n)]$ for two independent samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ with $\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ and $\nu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$, which is easily compared to (1.2).

As this is a continuation of [10], we refer to that article for more background and motivation for the study of the optimal matching problem and rates of convergence in Kantorovich metrics of empirical measures of Gaussian samples. As an introduction to the results of this work, we nevertheless recall the picture for the uniform distribution on $[0,1]^d$, as well as the known results so far in the Gaussian (and more general) setting.

Throughout the paper $A \lesssim B$ between two real positive numbers A and B means that $A \leqslant CB$ where $C > 0$ is either numerical or depends on p, d, but not on anything else. In the same way, a sentence like "A is bounded from above by B" has the same meaning. The equivalence sign $A \approx B$ indicates that $A \lesssim B$ and $B \lesssim A$. In particular, these extended inequalities will then hold uniformly over $n \geqslant 1$. Actually, n may always be assumed to be larger than some fixed integer n_0, large enough for obvious inequalities to hold true.

If μ is uniform on $[0,1]^d$, then (see [1], [16], [10]), for every $1 \leqslant p < \infty$,

(1.3) $$\mathbb{E}[W_p^p(\mu_n, \mu)] \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if } d = 1, \\ \left(\frac{\log n}{n}\right)^{p/2} & \text{if } d = 2, \\ \frac{1}{n^{p/d}} & \text{if } d \geqslant 3. \end{cases}$$

The particular, and critical, case $d = 2$ is the famous Ajtai–Komlós–Tusnády theorem [1].

Before addressing the Gaussian model, it is worth mentioning that, by a simple contraction argument (see [10]), the expected cost $\mathbb{E}[W_p^p(\mu_n, \mu)]$ when μ is uniform is bounded from above by the corresponding quantity when μ is the standard Gaussian measure. Hence, the rates in the uniform case provide lower bounds for the Gaussian model. This comparison is used implicitly in the following descriptions.

Let now μ be the standard Gaussian measure on the Borel sets of $\mathbb{R}^d$. It has been shown in [6] that in dimension $d = 1$,

$$(1.4) \qquad \mathbb{E}[W_p^p(\mu_n, \mu)] \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if } 1 \leqslant p < 2, \\ \frac{\log\log n}{n} & \text{if } p = 2, \\ \frac{1}{n(\log n)^{p/2}} & \text{if } p > 2. \end{cases}$$

In view of (1.3), it therefore appears that, already in dimension 1, the rates for $p \geqslant 2$ are rather sensitive to the underlying distribution. The proof of (1.4) in [6] relies on monotone rearrangement transport and explicit one-dimensional distributional inequalities. For $p = 2$, the exact limit

$$\lim_{n\to\infty} \frac{n}{\log\log n} \mathbb{E}[W_2^2(\mu_n, \mu)] = 1$$

is established in [5] (see also [7] for numerical evidence).

In dimension $d = 2$,

$$(1.5) \qquad \mathbb{E}[W_p^p(\mu_n, \mu)] \approx \begin{cases} \left(\frac{\log n}{n}\right)^{p/2} & \text{if } 1 \leqslant p < 2, \\ \frac{(\log n)^2}{n} & \text{if } p = 2. \end{cases}$$

Again, a specific new feature appears as $p = 2$. As we will see, the case $p > 2$ is essentially open. The proof of the case $1 \leqslant p < 2$, and the upper bound for $p = 2$, given in [10] is based on the pde and mass transportation approach developed next, while the lower bound for $p = 2$ in [17] relies on the generic chaining ideas of [16] together with a scaling argument. An alternative proof of this lower bound using the pde-transportation method was presented soon after in [11]. When $p = 1$, the upper bound $\mathbb{E}[W_1(\mu_n, \mu)] \lesssim \sqrt{\frac{\log n}{n}}$ has been shown in [19] to hold for distributions μ under a mild moment assumption. While the contribution [2] establishes the exact limit

$$\lim_{n\to\infty} \frac{n}{\log n} \mathbb{E}[W_2^2(\mu_n, \mu)] = \frac{1}{4\pi}$$

in the compact uniform case, the corresponding limit in the Gaussian setting is a challenging open question.

In higher dimension $d \geqslant 3$, a general bound using dyadic decompositions in the spirit of the Ajtai–Komlós–Tusnády theorem, and actually holding for distributions μ with enough moments, has been obtained in [8], [9], stating that

$$(1.6) \qquad \mathbb{E}[W_p^p(\mu_n, \mu)] \lesssim \frac{1}{n^{p/d}}$$

whenever $1 \leqslant p < d/2$. For Gaussian samples, it is extended up to $p < 2$ in dimension 3 in [10]. It is also known from [8], [9] that

$$(1.7) \qquad \mathbb{E}[W_p^p(\mu_n, \mu)] \lesssim \frac{1}{\sqrt{n}}$$

when $p > d/2$, with an extra $\log n$ in the numerator when $p = d/2$. The methodology used, however, will never produce anything better than this $1/\sqrt{n}$ rate, and the upper bound (1.7) is actually far from the potential lower bound deduced from (1.3), and not satisfactory already when $d = 1$.

The purpose of this work is to make some progress in the understanding of the rates in the Gaussian setting when $d \geqslant 3$, with the following statements.

THEOREM 1.1. *Let* $X_1, \ldots, X_n$ *be independent with common law the standard Gaussian distribution* μ *on* $\mathbb{R}^d$, $d \geqslant 3$, *and set* $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, $n \geqslant 1$. *Then, for* $1 \leqslant p < d$,

$$\mathbb{E}[W_p^p(\mu_n, \mu)] \approx \frac{1}{n^{p/d}}.$$

In this range $1 \leqslant p < d$, $d \geqslant 3$, the rates for the Gaussian are therefore the same as for the compact uniform model. The result extends (1.6) from $p < d/2$ to $p < d$. This might look as only a small step, but it overcomes the $1/\sqrt{n}$ rate and, as the proof will amply demonstrate, the amount of work to reach this conclusion is rather significant. Due to the results in [8]–[10] when $1 \leqslant p < 2$ mentioned above, only the values $2 \leqslant p < d$ have to be considered.

As identified by (1.5) when $p = d = 2$, the case $p = d$ might be of special interest. We have been able to reach the following conclusion, not definitive however.

THEOREM 1.2. *Let* $X_1, \ldots, X_n$ *be independent with common law the standard Gaussian distribution* μ *on* $\mathbb{R}^d$, $d \geqslant 2$, *and set* $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, $n \geqslant 1$. *Then*

$$\mathbb{E}[W_d^d(\mu_n, \mu)] \lesssim \frac{(\log n)^\kappa}{n}$$

where

$$\kappa = \begin{cases} \frac{d^2 + 6d}{8} & \text{if } d = 2, 3, \\ \frac{d^2}{2} - \frac{3d}{4} & \text{if } d \geqslant 4. \end{cases}$$

The case $d = 2$ recovers the claim in (1.5) (although the proof developed here to study every $d \geqslant 2$ is more involved than the ones in [10] and [17]). The lower bound provided by the uniform model is $1/n$ when $d \geqslant 3$. A possible conjecture might be that

$$\mathbb{E}[W_d^d(\mu_n, \mu)] \approx \frac{(\log n)^{d/2}}{n}$$

for $d \geqslant 3$. This is what is suggested as a lower bound in [17].

At this point, we do not have any reasonable conjecture for $p > d \, (\geqslant 2)$.

The proof of the preceding main results relies on the pde and transportation approach developed by L. Ambrosio, F. Stra and D. Trevisan in their breakthrough article [2] towards the exact limit for the two-dimensional uniform model in W_2, already exploited in the Gaussian case in [10]. This approach relies on a heat kernel regularization argument together with transportation bounds in terms of dual

Sobolev norms. Compared to the compact case, the Gaussian model involves the so-called Mehler kernel as underlying dynamics, which is unbounded. One of the features of [10] was the introduction of a localization step in order to take into account the infinite support of the Gaussian measure and the associated Gaussian tails. A further main step is achieved here by a randomization of the regularization time, together with estimates on the Mehler kernel.

In the note [17], M. Talagrand gave a proof of the exceptional case $p = d = 2$ in (1.5) by means of a scaling argument (for more general distributions than the Gaussian). It is certainly possible that the tools that he developed therein could lead to the main results presented here, and to determining some of the left open rates, in particular in the case $p = d$. However, that note is rather difficult to grasp and we could not extract further conclusions at this stage. On the other hand, the pde and transportation method may be used to recover the main result of [17] for $p = d = 2$, for the Gaussian model, as was shown in [11].

Turning to the content of the paper, Section 2 collects several formulas and bounds on the Mehler kernel as substitutes of uniform bounds in the compact case. Section 3 briefly recalls, in the Gaussian context, the transportation inequalities needed for the proofs, already emphasized in [2], [10]. The proof of the main Theorem 1.1 is divided into the next two sections, addressing respectively the cases $p = 2$ and $p > 2$. It could have been possible to present the proof directly for $p \geqslant 2$, but the case $p = 2$ is easier to handle and provides a good warm-up for the general case. In Section 6, we discuss how the parameters may be adjusted to reach Theorem 1.2.

2. PROPERTIES OF THE MEHLER KERNEL

This short section collects basic and classical properties of the Mehler kernel and the associated Ornstein–Uhlenbeck semigroup which will be of use throughout this work. The reference [4] (and the bibliography therein) covers most of the claims emphasized below. In contrast to the compact case, the main issue here is that the Mehler kernel is spatially unbounded, and so various pointwise and integral controls have to be identified.

Let $d\mu(x) = e^{-|x|^2/2} \frac{dx}{(2\pi)^{d/2}}$ be the standard Gaussian measure on the Borel sets of $\mathbb{R}^d$. The *Mehler kernel* is given, for $t > 0$ and $x, y \in \mathbb{R}^d$, by

$$(2.1) \quad p_t(x, y) = p_t(y, x)$$

$$= \frac{1}{(1 - e^{-2t})^{d/2}} \exp\left(-\frac{e^{-2t}}{2(1 - e^{-2t})}[|x|^2 + |y|^2 - 2e^t x \cdot y]\right).$$

Here $|x|$ denotes the Euclidean length of $x \in \mathbb{R}^d$. We have

$$\int_{\mathbb{R}^d} p_t(x, y)\, d\mu(y) = 1$$

for all $t > 0$ and $x \in \mathbb{R}^d$. Moreover, the Mehler kernel satisfies the basic semigroup property with respect to μ,

$$(2.2) \qquad \int_{\mathbb{R}^d} p_s(x,z)p_t(z,y)\,d\mu(z) = p_{s+t}(x,y)$$

for all $s, t > 0$ and $x, y \in \mathbb{R}^d$.

To simplify the notation, we set $a = e^{-t} \in (0,1)$ in some of the statements. When $x = y$,

$$(2.3) \qquad p_t(x,x) = \frac{1}{(1-a^2)^{d/2}} e^{\frac{a}{1+a}|x|^2} \leqslant \frac{1}{(1-a^2)^{d/2}} e^{|x|^2/2}.$$

Further, it is easily seen that for all $x, y \in \mathbb{R}^d$,

$$(2.4) \qquad p_t(x,y) \leqslant \frac{1}{(1-a^2)^{d/2}} e^{|x|^2/2}.$$

Combining (2.2)–(2.4) also shows that for every $q \geqslant 2$ and all $x \in \mathbb{R}^d$,

$$(2.5) \qquad \int_{\mathbb{R}^d} p_t(x,y)^q\,d\mu(y) = \int_{\mathbb{R}^d} p_t(x,y)^{q-2}p_t(x,y)^2\,d\mu(y)$$
$$\leqslant \frac{1}{(1-a^2)^{(q-1)d/2}} e^{(q-1)|x|^2/2}.$$

The Mehler kernel generates the Ornstein–Uhlenbeck semigroup

$$(2.6) \qquad P_t f(x) = \int_{\mathbb{R}^d} f(y)p_t(x,y)\,d\mu(y) = \int_{\mathbb{R}^d} f(e^{-t}x + \sqrt{1-e^{-2t}}\,y)\,d\mu(y)$$

for all $t > 0$, $x \in \mathbb{R}^d$, and any suitable measurable function $f : \mathbb{R}^d \to \mathbb{R}$, with the natural extension $P_0 = \mathrm{Id}$. The family $(P_t)_{t\geqslant 0}$ defines a Markov semigroup, symmetric in $\mathrm{L}^2(\mu)$, with infinitesimal generator $\mathrm{L} = \Delta - x \cdot \nabla$ for which the integration by parts formula

$$(2.7) \qquad \int_{\mathbb{R}^d} f(-\mathrm{L}g)\,d\mu = \int_{\mathbb{R}^d} \nabla f \cdot \nabla g\,d\mu$$

holds true for any smooth functions $f, g : \mathbb{R}^d \to \mathbb{R}$. The spectrum of the operator L is $\mathbb{N}$.

The semigroup $(P_t)_{t\geqslant 0}$ consists of contractions in all $\mathrm{L}^p(\mu)$-spaces with norms $\|\cdot\|_p$, $1 \leqslant p \leqslant \infty$. The spectral gap induces exponential decay for mean zero functions f in $\mathrm{L}^2(\mu)$,

$$(2.8) \qquad \|P_t f\|_2 \leqslant e^{-t}\|f\|_2, \quad t \geqslant 0.$$

The hypercontractivity property on the other hand states that

$$(2.9) \qquad \|P_t f\|_q \leqslant \|f\|_p$$

whenever $1 < p < q < \infty$ and $e^{2t} \geqslant \frac{q-1}{p-1}$. It will be convenient later on to combine the preceding smoothing properties in the following form: for any $f : \mathbb{R}^d \to \mathbb{R}$ in $L^p(\mu)$, $p \geqslant 2$, with mean zero,

$$(2.10) \qquad \|P_t f\|_p \leqslant C e^{-t/2} \|f\|_p, \quad t \geqslant 0,$$

where $C > 0$ only depends on p. The decay $e^{-t/2}$ is far from optimal but good enough for our purpose. For the proof, let $t_0 > 0$ be such that $e^{t_0} = p - 1 > 1$, so that by (2.9), for every $t \geqslant t_0$,

$$\|P_t f\|_p = \|P_{t/2}(P_{t/2} f)\|_p \leqslant \|P_{t/2} f\|_2.$$

By the exponential decay (2.8), $\|P_{t/2} f\|_2 \leqslant e^{-t/2} \|f\|_2 \leqslant e^{-t/2} \|f\|_p$ and hence

$$\|P_t f\|_p \leqslant e^{-t/2} \|f\|_p.$$

If $t \leqslant t_0$, then $\|P_t f\|_p \leqslant \|f\|_p \leqslant C e^{-t/2} \|f\|_p$ with $C = e^{t_0/2}$. The claim (2.10) is established.

The following pseudo-Poincaré inequality is another useful tool: for any real p, $1 \leqslant p < \infty$, there exists $C = C(p, d) > 0$ such that for every smooth function $f : \mathbb{R}^d \to \mathbb{R}$ and every $0 < t \leqslant 1$,

$$(2.11) \qquad \int_{\mathbb{R}^d} |P_t f - f|^p \, d\mu \leqslant C t^{p/2} \int_{\mathbb{R}^d} |\nabla f|^p \, d\mu.$$

For $p = 2$, this inequality may be easily deduced spectrally. For a proof in the general case, let $g : \mathbb{R}^d \to \mathbb{R}$ be smooth with $\|g\|_q \leqslant 1$ such that

$$\|P_t f - f\|_p = \int_{\mathbb{R}^d} g(P_t f - f) \, d\mu$$

where $\frac{1}{p} + \frac{1}{q} = 1$. By symmetry of the semigroup $(P_t)_{t \geqslant 0}$ and integration by parts,

$$\int_{\mathbb{R}^d} g(P_t f - f) \, d\mu = \int_0^t \int_{\mathbb{R}^d} g L P_s f \, d\mu \, ds = - \int_0^t \int_{\mathbb{R}^d} \nabla P_s g \cdot \nabla f \, d\mu \, ds.$$

The integral representation (2.6) and one more integration by parts indicate that, at any point $x \in \mathbb{R}^d$,

$$\nabla P_s g(x) = e^{-s} \int_{\mathbb{R}^d} \nabla g(e^{-s} x + \sqrt{1 - e^{-2s}} \, y) \, d\mu(y)$$

$$= \frac{e^{-s}}{\sqrt{1 - e^{-2s}}} \int_{\mathbb{R}^d} y g(e^{-s} x + \sqrt{1 - e^{-2s}} \, y) \, d\mu(y).$$

Hence, by Hölder's inequality,

$$\int_{\mathbb{R}^d} \nabla P_s g \cdot \nabla f \, d\mu$$

$$= \frac{e^{-s}}{\sqrt{1 - e^{-2s}}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} y \cdot \nabla f(x) g(e^{-s}x + \sqrt{1 - e^{-2s}} \, y) \, d\mu(x) \, d\mu(y)$$

$$\leqslant \frac{e^{-s}}{\sqrt{1 - e^{-2s}}} \Big(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |y \cdot \nabla f(x)|^p \, d\mu(x) \, d\mu(y) \Big)^{1/p}$$

where it is used that

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |g(e^{-s}x + \sqrt{1 - e^{-2s}} \, y)|^q \, d\mu(x) \, d\mu(y) = \int_{\mathbb{R}^d} |g|^q \, d\mu \leqslant 1.$$

Finally, after partial integration with respect to $d\mu(y)$,

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |y \cdot \nabla f(x)|^p \, d\mu(x) \, d\mu(y) = C \int_{\mathbb{R}^d} |\nabla f|^p \, d\mu$$

where $C > 0$ only depends on p. The claim (2.11) then easily follows.

We will make use of an important and delicate property, the Riesz transform bounds. In the Gaussian setting, they were established by P. A. Meyer [12] (see also [3]) and state that, for every $1 < p < \infty$ and every suitable $f : \mathbb{R}^d \to \mathbb{R}$,

$$(2.12) \qquad \int_{\mathbb{R}^d} |\nabla f|^p \, d\mu \approx \int_{\mathbb{R}^d} |(-\mathrm{L})^{-1/2} f|^p \, d\mu$$

(with the implied constants only depending on p) where $(-\mathrm{L})^{-1/2}$ is defined spectrally on mean zero functions f, for example by the classical formula

$$(2.13) \qquad (-\mathrm{L})^{-1/2} = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{1}{\sqrt{s}} P_s \, ds.$$

Finally, some technical tools related to energy estimates will be required. The reverse Poincaré inequality for the Gaussian measure [4, Theorem 4.7.2] states that for every Borel set A in $\mathbb{R}^d$ and every $s > 0$,

$$|\nabla P_s \mathbb{1}_A|^2 \leqslant \frac{1}{e^{2s} - 1} [P_s(\mathbb{1}_A^2) - (P_s \mathbb{1}_A)^2] \leqslant \frac{1}{e^{2s} - 1}.$$

Combining this with (2.11) for $f = P_s \mathbb{1}_A$, for every $s > 0$, $t \in (0, 1)$ and $p \geqslant 1$ we get

$$(2.14) \qquad \int_{\mathbb{R}^d} |P_t(P_s \mathbb{1}_A) - P_s \mathbb{1}_A|^p \, d\mu \leqslant \frac{Ct^{p/2}}{(e^{2s} - 1)^{(p-1)/2}} \int_{\mathbb{R}^d} |\nabla P_s \mathbb{1}_A| \, d\mu$$

$$\leqslant \frac{Ct^{p/2} e^{-s}}{(e^{2s} - 1)^{(p-1)/2}} \mu(\partial A)$$

where, in the last step, it is used that $\nabla P_s = e^{-s} P_s \nabla$ and

$$\limsup_{\varepsilon \to 0} \int_{\mathbb{R}^d} |\nabla P_\varepsilon \mathbb{1}_A| \, d\mu \leqslant \mu(\partial A)$$

for a measurable subset A of $\mathbb{R}^d$ with smooth boundary ∂A.

3. MASS TRANSPORTATION BOUNDS

This short section presents the key functional-analytic tool to bound Kantorovich distances in this context. It was already put forward in [10] (see also [13], [15]).

PROPOSITION 3.1. *Let $p \geqslant 1$. For any $d\nu = f \, d\mu$ with $f - 1$ in $\mathrm{H}^{-1,p}(\mu)$, we have*

$$\mathrm{W}_p(\nu, \mu) \leqslant p \|f - 1\|_{\mathrm{H}^{-1,p}(\mu)}$$

where the $\mathrm{H}^{-1,p}(\mu)$ negative Sobolev norm is defined by

$$\|g\|_{\mathrm{H}^{-1,p}(\mu)} = \left(\int_{\mathbb{R}^d} |\nabla((-\mathrm{L})^{-1}g)|^p \, d\mu \right)^{1/p}$$

for any $g : \mathbb{R}^d \to \mathbb{R}$ with mean zero such that the right-hand side makes sense.

It may be emphasized that in the particular case $p = 2$,

$$(3.1) \qquad \|g\|_{\mathrm{H}^{-1,2}(\mu)}^2 = 2 \int_0^\infty \int_{\mathbb{R}^d} (P_s g)^2 \, d\mu \, ds.$$

Indeed, since $(-\mathrm{L})^{-1} = \int_0^\infty P_s \, ds$, by integration by parts we have

$$\int_{\mathbb{R}^d} |\nabla((-\mathrm{L})^{-1}g)|^2 \, d\mu = \int_{\mathbb{R}^d} g(-\mathrm{L})^{-1}g \, d\mu = \int_0^\infty \int_{\mathbb{R}^d} g P_s g \, d\mu \, ds,$$

from which the claim follows by the symmetry of P_s.

When $p \neq 2$, by using the Riesz transform bound (2.12), for $1 < p < \infty$ Proposition 3.1 takes the form

$$(3.2) \qquad \mathrm{W}_p^p(\nu, \mu) \lesssim \int_{\mathbb{R}^d} |(-\mathrm{L})^{-1/2}(f - 1)|^p \, d\mu.$$

4. THE CASE $2 = p < d$

Here we prove Theorem 1.1 for $2 = p < d$. Although the proof in this case is actually contained in the more general proof for $2 \leqslant p < d$, it is easier to access due to the semigroup representation (3.1) of the negative Sobolev norm.

The first step is the localization argument introduced in [10]. For $R > 0$, let $d\mu^R = \frac{1}{\mu(B_R)} \mathbb{1}_{B_R} d\mu$ where B_R is the Euclidean ball centered at 0 with radius R. Define independent random variables X_i^R, $i = 1, \ldots, n$, with common distribution μ^R by

$$X_i^R = \begin{cases} X_i & \text{if } X_i \in B_R, \\ Z_i & \text{if } X_i \notin B_R, \end{cases}$$

where $Z_1, \ldots, Z_n$ are independent with distribution μ^R, independent of the X_i's. Setting $\mu_n^R = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^R}$, by definition of the coupling we have

$$\mathrm{W}_2^2(\mu_n, \mu_n^R) \leqslant \frac{1}{n} \sum_{i=1}^n |X_i - X_i^R|^2 \leqslant \frac{4}{n} \sum_{i=1}^n |X_i|^2 \mathbb{1}_{\{|X_i| \geqslant R\}}.$$

Therefore

$$\mathbb{E}[\mathrm{W}_2^2(\mu_n, \mu_n^R)] \leqslant 4 \int\limits_{\{|x| > R\}} |x|^2 \, d\mu.$$

Since $\int_{\{|x| > R\}} |x|^2 \, d\mu = C_d \int_R^\infty r^{d+1} e^{-r^2/2} \, dr$ is of the order of $R^d e^{-R^2/2}$ as $R \to \infty$, choose $R = \sqrt{2c \log n}$ for some $c \in (2/d, 1)$ so that

$$(4.1) \qquad\qquad \mathbb{E}[\mathrm{W}_2^2(\mu_n, \mu_n^R)] \lesssim \frac{1}{n^{2/d}}.$$

As a result, we concentrate on the study of $\mathbb{E}[\mathrm{W}_2^2(\mu_n^R, \mu)]$. Note furthermore that $\mu(B_R) \geqslant 1/2$ for n large enough so that this normalization factor may essentially be neglected throughout the investigation. For further developments, it will be convenient to refer to a random variable X with distribution μ and to X^R with distribution μ^R.

A new step is the introduction of a randomized regularization time by means of the decomposition of B_R as the union of m annuli

$$D_k = \{x \in \mathbb{R}^d; \ r_{k-1} \leqslant |x| < r_k\}, \quad k = 1, \ldots, m,$$

where $0 = r_0 < r_1 < \cdots < r_m = R$ with $r_k = \sqrt{k}$, $k = 1, \ldots, m$. In particular $m = R^2 = 2c \log n$. Then define a map $T : B_R \to (0, 1)$ as $T(x) = t_k$ if $x \in D_k$ where $0 < t_1 < \cdots < t_m < 1$ will be specified later.

For this map T, consider the (random) probability density

$$f(y) = f_n^{R,T}(y) = \frac{1}{n} \sum_{i=1}^n p_{T(X_i^R)}(X_i^R, y), \quad y \in \mathbb{R}^d,$$

and set $d\mu_n^{R,T} = f_n^{R,T} d\mu$. By convexity of the Kantorovich metric W_2^2 [18, Theorem 4.8] and the representation formula for the Ornstein–Uhlenbeck semigroup

(2.6), conditionally on the X_i^R's,

$$
\begin{aligned}
\mathrm{W}_2^2(\mu_n^R, \mu_n^{R,T}) &\leqslant \frac{1}{n} \sum_{i=1}^n \mathrm{W}_2^2(\delta_{X_i^R}, p_{T(X_i^R)}(X_i^R, \cdot)\, d\mu) \\
&= \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}^d} |X_i^R - y|^2 p_{T(X_i^R)}(X_i^R, y)\, d\mu(y) \\
&= \frac{1}{n} \sum_{i=1}^n [(1 - e^{-T(X_i^R)})^2 |X_i^R|^2 + d(1 - e^{-2T(X_i^R)})] \\
&\lesssim \frac{1}{n} \sum_{i=1}^n [T(X_i^R)^2 |X_i^R|^2 + T(X_i^R)]
\end{aligned}
$$

where in the last step we have used the fact that $T(X_i^R) \in (0, 1)$. Averaging over the X_i^R's, and dropping X^R in $T(X^R)$ for simplicity, we get

$$
\begin{aligned}
(4.2) \qquad \mathbb{E}[\mathrm{W}_2^2(\mu_n^R, \mu_n^{R,T})] &\lesssim \mathbb{E}[T^2 |X^R|^2] + \mathbb{E}[T] \\
&\lesssim (t_m R^2 + 1)\mathbb{E}[T] = (t_m R^2 + 1) \sum_{k=1}^m t_k \mu(D_k).
\end{aligned}
$$

From the localization (4.1) and regularization (4.2) arguments, and the triangle inequality for W_2, we deduce

$$
(4.3) \quad \mathbb{E}[\mathrm{W}_2^2(\mu_n, \mu)] \lesssim \mathbb{E}[\mathrm{W}_2^2(\mu_n^{R,T}, \mu)] + \frac{1}{n^{2/d}} + (t_m R^2 + 1) \sum_{k=1}^m t_k \mu(D_k).
$$

From now on, we concentrate on $\mathbb{E}[\mathrm{W}_2^2(\mu_n^{R,T}, \mu)]$, for which we make use of Proposition 3.1 and (3.1) with

$$
g = g(y) = f_n^{R,T}(y) - 1 = \frac{1}{n} \sum_{i=1}^n [p_{T(X_i^R)}(X_i^R, y) - 1], \quad y \in \mathbb{R}^d,
$$

to get

$$
\mathbb{E}[\mathrm{W}_2^2(\mu_n^{R,T}, \mu)] \leqslant 4\mathbb{E}[\|f_n^{R,T} - 1\|_{\mathrm{H}^{-1,2}(\mu)}^2] = 8 \int_0^\infty \int_{\mathbb{R}^d} \mathbb{E}[(P_s g)^2]\, d\mu\, ds.
$$

In order to develop probabilistic arguments, it is convenient to center the elements $p_{T(X_i^R)}(X_i^R, y)$ in the definition of g. Write therefore, for every $y \in \mathbb{R}^d$,

$$
\begin{aligned}
g(y) &= \frac{1}{n} \sum_{i=1}^n \big[p_{T(X_i^R)}(X_i^R, y) - \mathbb{E}[p_{T(X_i^R)}(X_i^R, y)] \big] + \mathbb{E}[p_T(X^R, y)] - 1 \\
&= \widetilde{g}(y) + \phi(y)
\end{aligned}
$$

where $\phi(y) = \mathbb{E}[p_T(X^R, y)] - 1$. Recall that here $T = T(X^R)$. For every $s > 0$, $\mathbb{E}[(P_s g)^2] \leqslant 2\mathbb{E}[(P_s \widetilde{g})^2] + 2(P_s \phi)^2$ so that

$$\mathbb{E}[\mathrm{W}_2^2(\mu_n^{R,T}, \mu)] \leqslant 16 \int\limits_0^\infty \int\limits_{\mathbb{R}^d} \mathbb{E}[(P_s \widetilde{g})^2] \, d\mu \, ds + 16 \int\limits_0^\infty \int\limits_{\mathbb{R}^d} (P_s \phi)^2 \, d\mu \, ds.$$

Now, since

$$P_s \widetilde{g}(y) = \frac{1}{n} \sum_{i=1}^n \left[p_{s+T(X_i^R)}(X_i^R, y) - \mathbb{E}[p_{s+T(X_i^R)}(X_i^R, y)] \right]$$

and since the X_i^R's, $i = 1, \ldots, n$, are independent and identically distributed, for each $y \in \mathbb{R}^d$ we have

$$\mathbb{E}[(P_s \widetilde{g})^2] = \frac{1}{n} \mathbb{E}\left[(p_{s+T}(X^R, y) - \mathbb{E}[p_{s+T}(X^R, y)])^2 \right].$$

But $\mathbb{E}[p_{s+T}(X^R, y)] = P_s \phi(y) + 1$, and as we aim to control $\int_0^\infty \int_{\mathbb{R}^d} (P_s \phi)^2 \, d\mu \, ds$, we may as well replace back $\mathbb{E}[p_{T+s}(X^R, y)]$ by 1 in the latter, to obtain

$$(4.4) \qquad \mathbb{E}[\mathrm{W}_2^2(\mu_n^{R,T}, \mu)] \lesssim \frac{1}{n} \int\limits_0^\infty \int\limits_{\mathbb{R}^d} \mathbb{E}[(p_{s+T}(X^R, y) - 1)^2] \, d\mu(y) \, ds$$

$$+ \int\limits_0^\infty \int\limits_{\mathbb{R}^d} (P_s \phi)^2 \, d\mu \, ds.$$

The most important term on the right-hand side of (4.4) is the first one, on which we concentrate next. Divide the integral in s according as $s \in (0, 1)$ or $s \in (1, \infty)$ to get

$$(4.5) \qquad \frac{1}{n} \int\limits_0^\infty \int\limits_{\mathbb{R}^d} \mathbb{E}[(p_{s+T}(X^R, y) - 1)^2] \, d\mu(y) \, ds$$

$$= \frac{1}{n} \int\limits_0^1 \int\limits_{\mathbb{R}^d} \mathbb{E}[(p_{s+T}(X^R, y) - 1)^2] \, d\mu(y) \, ds$$

$$+ \frac{1}{n} \int\limits_1^\infty \int\limits_{\mathbb{R}^d} \mathbb{E}[(p_{s+T}(X^R, y) - 1)^2] \, d\mu(y) \, ds.$$

Recall in addition that by definition of T,

$$\mathbb{E}[(p_{s+T}(X^R, y) - 1)^2] = \frac{1}{\mu(B_R)} \sum_{k=1}^m \int\limits_{D_k} [p_{s+t_k}(x, y) - 1]^2 \, d\mu(x).$$

The first term on the right-hand side of (4.5) may then be simply upper bounded as

$$\frac{1}{n}\int\limits_0^1\int\limits_{\mathbb{R}^d}\mathbb{E}[(p_{s+T}(X^R,y)-1)^2]\,d\mu(y)\,ds$$

$$\lesssim \frac{1}{n}\sum_{k=1}^m \int\limits_{D_k}\int\limits_0^1\int\limits_{\mathbb{R}^d} p_{s+t_k}(x,y)^2\,d\mu(y)\,ds\,d\mu(x).$$

By (2.2), (2.3) and integration in s, for every $x\in\mathbb{R}^d$ and $k=1,\dots,m$,

$$\int\limits_0^1\int\limits_{\mathbb{R}^d} p_{s+t_k}(x,y)^2\,d\mu(y)ds=\int\limits_0^1 p_{2(s+t_k)}(x,x)\,ds \lesssim \frac{1}{t_k^{d/2-1}}e^{|x|^2/2}.$$

Hence, since $\lambda(D_k)\approx k^{d/2-1}$,

(4.6) $$\frac{1}{n}\int\limits_0^1\int\limits_{\mathbb{R}^d}\mathbb{E}[(p_{s+T}(X^R,y)-1)^2]\,d\mu(y)\,ds$$

$$\lesssim \frac{1}{n}\sum_{k=1}^m \int\limits_{D_k}\frac{1}{t_k^{d/2-1}}e^{|x|^2/2}\,d\mu(x) \lesssim \frac{1}{n}\sum_{k=1}^m \frac{\lambda(D_k)}{t_k^{d/2-1}}$$

$$\lesssim \frac{1}{n}\sum_{k=1}^m \left(\frac{k}{t_k}\right)^{d/2-1}.$$

On the other hand, for the second term on the right-hand side of (4.5), by the exponential decay (2.8) in $\mathrm{L}^2(\mu)$,

(4.7) $$\frac{1}{n}\int\limits_1^\infty\int\limits_{\mathbb{R}^d}\mathbb{E}[(p_{s+T}(X^R,y)-1)^2]\,d\mu(y)\,ds$$

$$= \frac{1}{n}\frac{1}{\mu(B_R)}\sum_{k=1}^m \int\limits_{D_k}\int\limits_1^\infty\int\limits_{\mathbb{R}^d}[p_{s+t_k}(x,y)-1]^2\,d\mu(y)\,ds\,d\mu(x)$$

$$\lesssim \frac{1}{n}\sum_{k=1}^m \int\limits_{D_k}\int\limits_0^\infty e^{-2s}\int\limits_{\mathbb{R}^d}[p_1(x,y)-1]^2\,d\mu(y)\,ds\,d\mu(x)$$

$$\lesssim \frac{1}{n}\sum_{k=1}^m \lambda(D_k) = \frac{1}{n}\lambda(B_R)$$

where it is used again that $\int_{\mathbb{R}^d} p_1(x,y)^2\,d\mu(y)=p_2(x,x)\lesssim e^{|x|^2/2}$.
    Inserting (4.6) and (4.7) in (4.5), it follows that

$$\frac{1}{n}\int\limits_0^\infty\int\limits_{\mathbb{R}^d}\mathbb{E}[(p_{s+T}(X^R,y)-1)^2]\,d\mu(y)\,ds \lesssim \frac{1}{n}\sum_{k=1}^m \left(\frac{k}{t_k}\right)^{d/2-1} + \frac{1}{n^{2/d}}$$

(recall that $d > 2$). Plugging this estimate into (4.4) and (4.3), we obtain

$$(4.8) \qquad \mathbb{E}[W_2^2(\mu_n, \mu)] \lesssim \frac{1}{n} \sum_{k=1}^{m} \left(\frac{k}{t_k} \right)^{d/2-1} + (t_m R^2 + 1) \sum_{k=1}^{m} t_k \mu(D_k)$$
$$+ \frac{1}{n^{2/d}} + \int_0^{\infty} \int_{\mathbb{R}^d} (P_s \phi)^2 \, d\mu \, ds.$$

The choice of the t_k's is now determined by optimization between the first two terms on the right-hand side of (4.8). Using $\mu(D_k) \lesssim k^{d/2-1} e^{-k/2}$, set $t_k = n^{-2/d} e^{k/d}$, $k = 1, \ldots, m$. In particular,

$$t_k \leqslant t_m = \frac{1}{n^{2(1-c)/d}}$$

for every $k = 1, \ldots, m = R^2 = 2c \log n$, $c \in (2/d, 1)$. For these values, and since $d > 2$, it follows that

$$\frac{1}{n} \sum_{k=1}^{m} \left(\frac{k}{t_k} \right)^{d/2-1} + (t_m R^2 + 1) \sum_{k=1}^{m} t_k \mu(D_k) \lesssim \frac{1}{n^{2/d}}.$$

Therefore (4.8) yields

$$(4.9) \qquad \mathbb{E}[W_2^2(\mu_n, \mu)] \lesssim \frac{1}{n^{2/d}} + \int_0^{\infty} \int_{\mathbb{R}^d} (P_s \phi)^2 \, d\mu \, ds.$$

We are left with the study of the centering term $\int_0^{\infty} \int_{\mathbb{R}^d} (P_s \phi)^2 \, d\mu \, ds$. To this end, recall that

$$P_s \phi = \mathbb{E}[p_{s+T}(X^R, \cdot) - 1]$$
$$= \frac{1}{\mu(B_R)} \sum_{k=1}^{m} [P_{s+t_k} \mathbb{1}_{D_k} - \mu(D_k)]$$
$$= \frac{1}{\mu(B_R)} \sum_{k=1}^{m} [P_{s+t_k} \mathbb{1}_{D_k} - P_s \mathbb{1}_{D_k}] + \frac{1}{\mu(B_R)} \sum_{k=1}^{m} [P_s \mathbb{1}_{D_k} - \mu(D_k)]$$
$$= \frac{1}{\mu(B_R)} \sum_{k=1}^{m} [P_{s+t_k} \mathbb{1}_{D_k} - P_s \mathbb{1}_{D_k}] + \frac{1}{\mu(B_R)} [P_s \mathbb{1}_{B_R} - \mu(B_R)].$$

Hence, for every $s > 0$,

$$\int_{\mathbb{R}^d} (P_s \phi)^2 \, d\mu \lesssim \int_{\mathbb{R}^d} \left[\sum_{k=1}^{m} [P_{s+t_k} \mathbb{1}_{D_k} - P_s \mathbb{1}_{D_k}] \right]^2 d\mu + \int_{\mathbb{R}^d} [P_s \mathbb{1}_{B_R} - \mu(B_R)]^2 \, d\mu$$

and we treat the two expressions on the right-hand side separately. The second one is easy since by the exponential decay (2.8) in $L^2(\mu)$,

$$\int_{\mathbb{R}^d} [P_s \mathbb{1}_{B_R} - \mu(B_R)]^2 \, d\mu \leqslant e^{-2s}[1 - \mu(B_R)] \lesssim \frac{e^{-2s}}{n^{2/d}}$$

by the choice of R. For the first one, by the triangle inequality and (2.14),

$$\int_{\mathbb{R}^d} \Big[\sum_{k=1}^m [P_{s+t_k}\mathbb{1}_{D_k} - P_s\mathbb{1}_{D_k}]\Big]^2 d\mu \lesssim \Big(\sum_{k=1}^m \Big[\frac{t_k e^{-s}}{\sqrt{e^{2s}-1}}\mu(\partial D_k)\Big]^{1/2}\Big)^2$$

$$\lesssim \frac{e^{-s}}{\sqrt{e^{2s}-1}}\Big(\sum_{k=1}^m \sqrt{t_k}\sqrt{\mu(\partial D_k)}\Big)^2.$$

Combining the preceding estimates gives

$$\int_0^\infty \int_{\mathbb{R}^d} (P_s\phi)^2 \, d\mu \, ds \lesssim \Big(\sum_{k=1}^m \sqrt{t_k}\sqrt{\mu(\partial D_k)}\Big)^2 + \frac{1}{n^{2/d}}.$$

Since $\mu(\partial D_k) \lesssim k^{(d-1)/2}e^{-k/2}$, the choice of $t_k = n^{-2/d}e^{k/d}$, $k = 1,\ldots,m$, easily shows that this centering contribution is at most $1/n^{2/d}$. Inserting this into (4.9) concludes the proof of the theorem for $2 = p < d$.

5. THE CASE $2 \leqslant p < d$

The pattern of the proof will be similar to the one of Section 4 but with significant increase of the technicalities since (3.1) is no more available and the arguments go through the more involved Riesz transform bound (3.2). The scheme of the proof is then similar to the one developed in [10] in the compact case (1.3), but again unboundedness of the Mehler kernel requires several delicate estimates.

As in the preceding section for $p = 2$, we truncate in the same way to a ball B_R with $R = \sqrt{2c\log n}$ for some $c \in (p/d, 1)$ for which we find similarly that

$$(5.1) \qquad \mathbb{E}[W_p^p(\mu_n, \mu_n^R)] \lesssim \frac{1}{n^{p/d}}.$$

We decompose again B_R as the union of m annuli

$$D_k = \{x \in \mathbb{R}^d;\ r_{k-1} \leqslant |x| < r_k\},$$

$k = 1,\ldots,m$, where $0 = r_0 < r_1 < \cdots < r_m = R$, and consider also the map $T : B_R \to (0,1)$ defined by $T(x) = t_k$ if $x \in D_k$. We will use the same choices $r_k = \sqrt{k}$ and $t_k = n^{-2/d}e^{k/d}$, $k = 1,\ldots,m$. In particular, for further purposes, it is important to notice again that

$$(5.2) \qquad t_k \leqslant t_m \leqslant \frac{1}{n^{2(1-c)/d}}, \quad k = 1,\ldots,m,$$

which is small enough for a number of subsequent issues (recall that $n \geqslant n_0$ is assumed large enough throughout the investigation).

For this map $T : B_R \to (0,1)$, set similarly

$$f(y) = f_n^{R,T}(y) = \frac{1}{n} \sum_{i=1}^{n} p_{T(X_i^R)}(X_i^R, y), \quad y \in \mathbb{R}^d,$$

and $d\mu_n^{R,T} = f_n^{R,T} d\mu$. By convexity [18, Theorem 4.8] and the representation formula for the Ornstein–Uhlenbeck semigroup (2.6), conditionally on the X_i^R's,

$$W_p^p(\mu_n^R, \mu_n^{R,T}) \leqslant \frac{1}{n} \sum_{i=1}^{n} W_p^p(\delta_{X_i^R}, p_{T(X_i^R)}(X_i^R, \cdot) \, d\mu)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}^d} |X_i^R - y|^p p_{T(X_i^R)}(X_i^R, y) \, d\mu(y)$$

$$\lesssim \frac{1}{n} \sum_{i=1}^{n} [(1 - e^{-T(X_i^R)})^p |X_i^R|^p + (1 - e^{-2T(X_i^R)})^{p/2}].$$

Noticing that $T(X_i^R) \in (0,1)$, it follows after taking expectation that

$$(5.3) \qquad \mathbb{E}[W_p^p(\mu_n^R, \mu_n^{R,T})] \lesssim \mathbb{E}[T^p |X^R|^p] + \mathbb{E}[T^{p/2}] \lesssim (t_m^{p/2} R^p + 1) \mathbb{E}[T^{p/2}]$$

where we recall that $T = T(X^R)$.

From (5.1), (5.3) and the triangle inequality, and since $t_m \leqslant 1/n^{2(1-c)/d}$ by (5.2), we have therefore obtained

$$(5.4) \qquad \mathbb{E}[W_p^p(\mu_n, \mu)] \lesssim \mathbb{E}[W_p^p(\mu_n^{R,T}, \mu)] + \frac{1}{n^{p/d}} + \sum_{k=1}^{m} t_k^{p/2} \mu(D_k)$$

$$\lesssim \mathbb{E}[W_p^p(\mu_n^{R,T}, \mu)] + \frac{1}{n^{p/d}}$$

where we have used $t_k = n^{-2/d} e^{k/d}$ and $\mu(D_k) \lesssim k^{d/2-1} e^{-k/2}$, $k = 1, \ldots, m$.

As in the previous section, we thus concentrate on the study of $\mathbb{E}[W_p^p(\mu_n^{R,T}, \mu)]$, which we control by applying Proposition 3.1 together with the Riesz transform bound (3.2). To this end, set

$$g = g(y) = f_n^{R,T}(y) - 1 = \frac{1}{n} \sum_{i=1}^{n} [p_{T(X_i^R)}(X_i^R, y) - 1], \quad y \in \mathbb{R}^d.$$

Therefore

$$\mathbb{E}[W_p^p(\mu_n^{R,T}, \mu)] \leqslant p^p \mathbb{E}[\|f_n^{R,T} - 1\|_{H^{-1,p}(\mu)}^p] \lesssim \mathbb{E}\left[\int_{\mathbb{R}^d} |(-L)^{-1/2} g|^p \, d\mu\right].$$

Center then the terms $p_{T(X_i^R)}(X_i^R, y)$ in the definition of g with respect to randomness in the X_i^R's. To this end, write, for every y,

$$g(y) = \frac{1}{n} \sum_{i=1}^{n} \left[p_{T(X_i^R)}(X_i^R, y) - \mathbb{E}[p_{T(X_i^R)}(X_i^R, y)]\right] + \mathbb{E}[p_T(X^R, y)] - 1$$

$$= \widetilde{g}(y) + \phi(y)$$

with $\phi(y) = \mathbb{E}[p_T(X^R, y)] - 1$ so that

$$\mathbb{E}[W_p^p(\mu_n^{R,T}, \mu)] \lesssim \mathbb{E}\Big[\int\limits_{\mathbb{R}^d} |(-L)^{-1/2}\widetilde{g}|^p\, d\mu\Big] + \int\limits_{\mathbb{R}^d} |(-L)^{-1/2}\phi|^p\, d\mu.$$

Rosenthal's inequality [14] states that for independent centered random variables $V_1, \ldots, V_n$ with a pth moment, $p \geqslant 2$, we have

$$(5.5) \qquad \mathbb{E}\Big[\Big|\sum_{i=1}^n V_i\Big|^p\Big] \leqslant C_p \sum_{i=1}^n \mathbb{E}[|V_i|^p] + C_p\Big(\sum_{i=1}^n \mathbb{E}[V_i^2]\Big)^{p/2}$$

where $C_p > 0$ only depends on p. For each fixed $y \in \mathbb{R}^d$, apply this inequality to the independent identically distributed and centered random variables

$$(-L_y)^{-1/2}\big[p_{T(X_i^R)}(X_i^R, y) - \mathbb{E}[p_{T(X_i^R)}(X_i^R, y)]\big], \quad i = 1, \ldots, n,$$

to get

$$\mathbb{E}[|(-L_y)^{-1/2}\widetilde{g}(y)|^p] \lesssim \frac{1}{n^{p-1}}\mathbb{E}\Big[\big|(-L_y)^{-1/2}\big[p_T(X^R, y) - \mathbb{E}[p_T(X^R, y)]\big]\big|^p\Big]$$
$$+ \frac{1}{n^{p/2}}\mathbb{E}\big[\big((-L_y)^{-1/2}[p_T(X^R, y) - \mathbb{E}[p_T(X^R, y)]]\big)^2\big]^{p/2}.$$

Since $\mathbb{E}[p_T(X^R, y)] = \phi(y) + 1$, and we eventually aim to control the integral $\int_{\mathbb{R}^d} |(-L)^{-1/2}\phi|^p\, d\mu$, we may replace back $\mathbb{E}[p_T(X^R, y)]$ by 1 in the preceding, that is, we get

$$(5.6) \quad \mathbb{E}[W_p^p(\mu_n^{R,T}, \mu)] \lesssim \frac{1}{n^{p-1}} \int\limits_{\mathbb{R}^d} \mathbb{E}\big[|(-L_y)^{-1/2}[p_T(X^R, y) - 1]|^p\big]\, d\mu(y)$$
$$+ \frac{1}{n^{p/2}} \int\limits_{\mathbb{R}^d} \mathbb{E}\big[((-L_y)^{-1/2}[p_T(X^R, y) - 1])^2\big]^{p/2}\, d\mu(y)$$
$$+ \int\limits_{\mathbb{R}^d} |(-L)^{-1/2}\phi|^p\, d\mu.$$

In this expression, the random variables $(-L_y)^{-1/2}[p_T(X^R, y) - 1]$ will be studied with the help of the spectral representation (2.13), that is,

$$(-L_y)^{-1/2}[p_T(X^R, y) - 1] = \frac{1}{\sqrt{\pi}} \int\limits_0^\infty \frac{1}{\sqrt{s}}[p_{s+T}(X^R, y) - 1]\, ds,$$

and similarly for $(-L)^{-1/2}\phi$.

According to (5.4) and (5.6), the proof of the theorem will therefore be complete once it is established that

$$(5.7) \qquad \frac{1}{n^{p-1}} \int\limits_{\mathbb{R}^d} \mathbb{E}\Big[\Big|\int\limits_0^\infty \frac{1}{\sqrt{s}}[p_{s+T}(X^R, y) - 1]\, ds\Big|^p\Big]\, d\mu(y) \lesssim \frac{1}{n^{p/d}},$$

$$(5.8) \qquad \frac{1}{n^{p/2}} \int\limits_{\mathbb{R}^d} \mathbb{E}\Big[\Big(\int\limits_0^\infty \frac{1}{\sqrt{s}}[p_{s+T}(X^R, y) - 1]\, ds\Big)^2\Big]^{p/2}\, d\mu(y) \lesssim \frac{1}{n^{p/d}}$$

and

$$(5.9) \qquad \int\limits_{\mathbb{R}^d} \left| \int\limits_0^\infty \frac{1}{\sqrt{s}} P_s \phi \, ds \right|^p d\mu \lesssim \frac{1}{n^{p/d}}.$$

The centering term (5.9) will be examined at the end of the proof. We concentrate on the first two terms, starting with (5.8) which is the most delicate one.

Study of (5.8). Fix $y \in \mathbb{R}^d$ to begin with; by definition of the map T,

$$\mathbb{E}\left[\left(\int\limits_0^\infty \frac{1}{\sqrt{s}} [p_{s+T}(X^R, y) - 1] \, ds \right)^2 \right]$$

$$= \frac{1}{\mu(B_R)} \sum_{k=1}^m \int\limits_{D_k} \left[\int\limits_0^\infty \frac{1}{\sqrt{s}} [p_{s+t_k}(x, y) - 1] \, ds \right]^2 d\mu(x).$$

Given $s_k \in (0,1)$, $k = 1, \ldots, m$, to be specified, we decompose the integral in s into integrals on $(0, s_k)$ and (s_k, ∞) and study separately, by the triangle inequality, the resulting two pieces in (5.8), showing that

$$(5.10) \qquad \frac{1}{n^{p/2}} \int\limits_{\mathbb{R}^d} \left(\sum_{k=1}^m \int\limits_{D_k} \left[\int\limits_0^{s_k} \frac{1}{\sqrt{s}} [p_{s+t_k}(x, y) - 1] \, ds \right]^2 d\mu(x) \right)^{p/2} d\mu(y) \lesssim \frac{1}{n^{p/d}}$$

and

$$(5.11) \qquad \frac{1}{n^{p/2}} \int\limits_{\mathbb{R}^d} \left(\sum_{k=1}^m \int\limits_{D_k} \left[\int\limits_{s_k}^\infty \frac{1}{\sqrt{s}} [p_{s+t_k}(x, y) - 1] \, ds \right]^2 d\mu(x) \right)^{p/2} d\mu(y) \lesssim \frac{1}{n^{p/d}}.$$

Concerning (5.10), for each $k = 1, \ldots, m$ and $y \in \mathbb{R}^d$, by Fubini's theorem,

$$\int\limits_{D_k} \left[\int\limits_0^{s_k} \frac{1}{\sqrt{s}} [p_{s+t_k}(x, y) - 1] \, ds \right]^2 d\mu(x)$$

$$= \int\limits_{D_k} \int\limits_0^{s_k} \int\limits_0^{s_k} \frac{1}{\sqrt{ss'}} [p_{s+t_k}(x, y) - 1][p_{s'+t_k}(x, y) - 1] \, ds \, ds' \, d\mu(x)$$

$$\lesssim \int\limits_0^{s_k} \int\limits_0^{s_k} \frac{1}{\sqrt{ss'}} \int\limits_{D_k} p_{s+t_k}(x, y) p_{s'+t_k}(x, y) \, d\mu(x) \, ds \, ds' + \mu(D_k).$$

Summing over k, it is clear that the contribution $\mu(D_k)$ will be irrelevant for the final bound and it is therefore ignored below. Now, a standard calculation with the explicit expression of the Mehler kernel $p_t(x, y)$ yields

$$(5.12) \qquad \int\limits_{D_k} p_{s+t_k}(x, y) p_{s'+t_k}(x, y) \, d\mu(x) = p_{s+s'+2t_k}(y, y) \mu(\widetilde{D}_k)$$

where $\widetilde{D}_k = -\frac{\beta}{\alpha}y + \alpha D_k$,

$$\alpha^2 = 1 + \frac{a^2}{1-a^2} + \frac{b^2}{1-b^2}, \quad \beta = \frac{a}{1-a^2} + \frac{b}{1-b^2},$$

with $a = e^{-s-t_k}, b = e^{-s'-t_k}$. For further purposes, note that

(5.13)
$$\frac{\alpha^2}{\beta} - 1 = \frac{(1-a)(1-b)}{a+b}.$$

We examine separately the expression $p_{s+s'+2t_k}(y,y)\mu(\widetilde{D}_k)$ in (5.12) according as $y \in \frac{\alpha^2}{\beta}D_k$ or not via the bound

(5.14) $\quad p_{s+s'+2t_k}(y,y)\mu(\widetilde{D}_k)$

$$\leqslant p_{s+s'+2t_k}(y,y)[\mathbb{1}_{\frac{\alpha^2}{\beta}D_k}(y) + \mathbb{1}_{(\frac{\alpha^2}{\beta}D_k)^c}(y)\mu(\widetilde{D}_k)].$$

Accordingly, the study of (5.10) is divided into two parts, and the task is to show that

(5.15)
$$\frac{1}{n^{p/2}} \int_{\mathbb{R}^d} \left[\sum_{k=1}^m \int_0^{s_k}\int_0^{s_k} \frac{1}{\sqrt{ss'}} p_{s+s'+2t_k}(y,y)\mathbb{1}_{\frac{\alpha^2}{\beta}D_k}(y)\,ds\,ds' \right]^{p/2} d\mu(y) \lesssim \frac{1}{n^{p/d}}$$

and

(5.16)
$$\frac{1}{n^{p/2}} \int_{\mathbb{R}^d} \left[\sum_{k=1}^m \int_0^{s_k}\int_0^{s_k} \frac{1}{\sqrt{ss'}} p_{s+s'+2t_k}(y,y)\mathbb{1}_{(\frac{\alpha^2}{\beta}D_k)^c}(y)\mu(\widetilde{D}_k)\,ds\,ds' \right]^{p/2} d\mu(y)$$

$$\lesssim \frac{1}{n^{p/d}}.$$

Start with (5.15) and fix now $s_k = 1/\sqrt{k}$, $k = 1, \ldots, m$. By (5.13), it may be verified that for every $0 < s, s' \leqslant s_k$,

$$\frac{\alpha^2}{\beta}D_k \subset E_k = \{x \in \mathbb{R}^d; \, r_{k-1} \leqslant |x| < (1+8/k)r_k\}.$$

Hence, together with (2.3) and a simple integration in s, s', for every $y \in \mathbb{R}^d$,

$$\int_0^{s_k}\int_0^{s_k} \frac{1}{\sqrt{ss'}} p_{s+s'+2t_k}(y,y)\mathbb{1}_{\frac{\alpha^2}{\beta}D_k}(y)\,ds\,ds' \lesssim \mathbb{1}_{E_k}(y)\frac{1}{t_k^{d/2-1}}e^{|y|^2/2}.$$

It follows that the left-hand side of (5.15) is bounded from above by

$$\frac{1}{n^{p/2}} \int_{\mathbb{R}^d} \left[\sum_{k=1}^m \mathbb{1}_{E_k}(y)\frac{1}{t_k^{d/2-1}} \right]^{p/2} e^{p|y|^2/4}\,d\mu(y).$$

Observe that for every $k = 1, \ldots, m$, $E_k \subset \bigcup_{\ell=k}^{k+80} D_\ell$ (with the obvious extension of D_k when $k \geqslant m$). As a consequence, the latter is bounded from above by

$$\frac{1}{n^{p/2}} \sum_{k=1}^{m+80} \frac{1}{t_k^{(d/2-1)p/2}} \int_{D_k} e^{p|y|^2/4} \, d\mu(y)$$

$$\lesssim \frac{1}{n^{p/2}} \sum_{k=1}^{m+80} \frac{1}{t_k^{(d/2-1)p/2}} k^{d/2-1} e^{(p/2-1)k/2}.$$

Now $t_k = n^{-2/d} e^{k/d}$, $k = 1, \ldots, m$, so that since $p < d$ the latter is of the order $\frac{1}{n^{p/d}}$, proving (5.15).

We turn next to (5.16). Fix $k = 1, \ldots, m$. If $y \notin \frac{\alpha^2}{\beta} D_k$, then (draw a picture) it is clear that

$$\inf\{|z| \in \mathbb{R}^d; \; z \in \widetilde{D}_k\} = \left| \frac{\beta}{\alpha} |y| - \alpha \widetilde{r}_k \right|,$$

where $\widetilde{r}_k = r_k$ or r_{k-1}. Thus

$$\mu(\widetilde{D}_k) \lesssim \left(1 + \left| \frac{\beta}{\alpha} |y| - \alpha \widetilde{r}_k \right|^{d-2} \right) \exp\left(-\frac{1}{2} \left(\frac{\beta}{\alpha} |y| - \alpha \widetilde{r}_k \right)^2 \right).$$

To get rid of the prefactors, let $\sigma \in (0, 1)$ (which will depend on p and d only) so that

$$\mu(\widetilde{D}_k) \lesssim \exp\left(-\frac{1-\sigma}{2} \left(\frac{\beta}{\alpha} |y| - \alpha \widetilde{r}_k \right)^2 \right).$$

Hence, together with (2.3),

$$p_{s+s'+2t_k}(y, y) \mathbb{1}_{(\frac{\alpha^2}{\beta} D_k)^c}(y) \mu(\widetilde{D}_k)$$

$$\lesssim \frac{1}{(1 - a^2 b^2)^{d/2}} \exp\left(-\frac{1-\sigma}{2} \left(\frac{\beta}{\alpha} |y| - \alpha \widetilde{r}_k \right)^2 + \frac{ab}{1 + ab} |y|^2 \right).$$

Now, since

$$\frac{\beta^2}{\alpha^2} - \frac{2ab}{1 + ab} = \alpha^2 - 1,$$

it is easily seen that

$$-\frac{1-\sigma}{2} \left(\frac{\beta}{\alpha} |y|^2 - \alpha \widetilde{r}_k \right)^2 + \frac{ab}{1 + ab} |y|^2 \leqslant (1 - \sigma) A + \frac{\sigma}{2} |y|^2$$

where

$$A = -\tfrac{1}{2}(\alpha^2 - 1)|y|^2 + \beta \widetilde{r}_k |y| - \tfrac{1}{2} \alpha^2 \widetilde{r}_k^2.$$

We would like to choose numerical non-negative constants K and L such that

$$A \leqslant A' = \frac{K}{2}|y|^2 + \frac{L}{2}\widetilde{r}_k{}^2$$

for all $y \in \mathbb{R}^d$ and $k = 1, \ldots, m$. This may be achieved provided that $K + L \geqslant 1$ and $L \leqslant 1$. Indeed, the quadratic form

$$B = A' - A = \tfrac{1}{2}(\alpha^2 - 1 + K)|y|^2 - \beta\widetilde{r}_k|y| + \tfrac{1}{2}(\alpha^2 + L)\widetilde{r}_k{}^2$$

is positive semi-definite if $Q = \beta^2 - (\alpha^2 - 1 + K)(\alpha^2 + L) \leqslant 0$. When $K + L \geqslant 1$ and $L \leqslant 1$,

$$
\begin{aligned}
Q &= -\frac{(1 - a^2 b^2)}{(1 - a^2)(1 - b^2)}\left(K + L - \frac{2ab}{1 + ab}\right) + L(1 - K) \\
&\leqslant -\frac{(1 - a^2 b^2)}{(1 - a^2)(1 - b^2)}\left(1 - \frac{2ab}{1 + ab}\right) + L^2 \\
&\leqslant -\frac{(1 - ab)^2}{(1 - a^2)(1 - b^2)} + L^2 \leqslant -1 + L^2 \leqslant 0.
\end{aligned}
$$

As a consequence of this analysis, it follows that, provided $K + L \geqslant 1$ and $L \leqslant 1$,

$$p_{s+s'+2t_k}(y, y)\mathbb{1}_{(\frac{\alpha^2}{\beta}D_k)^c}(y)\mu(\widetilde{D}_k) \lesssim \frac{1}{(1 - a^2 b^2)^{d/2}}e^{A''} \lesssim \frac{1}{(s + s' + 2t_k)^{d/2}}e^{A''}$$

uniformly over $s, s' \leqslant s_k$ and $y \in \mathbb{R}^d$, where

$$A'' = (1 - \sigma)A' + \frac{\sigma}{2}|y|^2 = \frac{(1 - \sigma)K + \sigma}{2}|y|^2 + \frac{(1 - \sigma)L}{2}\widetilde{r}_k{}^2.$$

One may now integrate in $0 < s, s' \leqslant s_k$ for every k to find that the left-hand side of (5.16) is bounded from above by

$$(5.17) \qquad \frac{1}{n^{p/2}} \int_{\mathbb{R}^d} \left[\sum_{k=1}^m \frac{1}{t_k^{d/2-1}}e^{A''}\right]^{p/2} d\mu(y),$$

still under the conditions $K + L \geqslant 1$ and $L \leqslant 1$. The definition of A'' shows after integration that

$$
\begin{aligned}
&\frac{1}{n^{p/2}} \int_{\mathbb{R}^d} \left[\sum_{k=1}^m \frac{1}{t_k^{d/2-1}}e^{A''}\right]^{p/2} d\mu(y) \\
&\qquad = \frac{1}{n^{p/2}} \int_{\mathbb{R}^d} \left[\sum_{k=1}^m \frac{1}{t_k^{d/2-1}}e^{\frac{(1-\sigma)L}{2}\widetilde{r}_k{}^2}\right]^{p/2} e^{\frac{p[(1-\sigma)K+\sigma]}{4}|y|^2} d\mu(y) \\
&\qquad \lesssim \frac{1}{n^{p/2}} \frac{1}{\left(1 - \frac{p}{2}[(1-\sigma)K + \sigma]\right)^{d/2}}\left[\sum_{k=1}^m \frac{1}{t_k^{d/2-1}}e^{\frac{(1-\sigma)L}{2}\widetilde{r}_k{}^2}\right]^{p/2}
\end{aligned}
$$

provided that $\frac{p}{2}[(1 - \sigma)K + \sigma] < 1$. Since $t_k = n^{-2/d}e^{k/d}$, $k = 1, \ldots, m$, and $\widetilde{r}_k = \sqrt{k}$ or $\sqrt{k-1}$, it is necessary in addition that $(1 - \sigma)L < 1 - 2/d$ in order that the preceding bound yields the correct rate $1/n^{p/d}$. Provided that $\sigma \in (0, 2/p)$, the preceding two conditions are indeed compatible with $K + L \geqslant 1$ and $L \leqslant 1$. As an interesting example, we can always set $K = 0$, $L = 1$ and $\sigma \in (2/d, 2/p)$, simplifying the exposition. The preceding construction will however be needed later in the case $p = d$ (Section 6). As a conclusion, (5.16) is established.

As announced, the two controls (5.15) and (5.16) together yield the expected bound (5.10).

We turn to the analysis of the second part (5.11) concerned with the values of $s \geqslant s_k$. The first step is a Minkowski integral inequality to exchange the order of integration. For this purpose, it is convenient to rewrite the left-hand side of (5.11) as

$$(5.18) \qquad \frac{1}{n^{p/2}} \int_{\mathbb{R}^d} \left(\mathbb{E}\left[\left| \int_S^\infty \frac{1}{\sqrt{s}}[p_{s+T}(X^R, y) - 1]\, ds \right|^2 \right] \right)^{p/2} d\mu(y)$$

where $S : B_R \to (0, 1)$ is defined by $S(x) = s_k$ if $x \in D_k$, $k = 1, \ldots, m$. As for $T = T(X^R)$, we write here $S = S(X^R)$ to simplify the notation. As announced, by Minkowski's inequality since $p \geqslant 2$, (5.18) is less than or equal to

$$\frac{1}{n^{p/2}} \mathbb{E}\left[\left(\int_{\mathbb{R}^d} \left| \int_S^\infty \frac{1}{\sqrt{s}}[p_{s+T}(X^R, y) - 1]\, ds \right|^p d\mu(y) \right)^{2/p} \right]^{p/2}.$$

Now, and conditionally on the randomness of X^R, also by the triangle inequality,

$$\int_{\mathbb{R}^d} \left| \int_S^\infty \frac{1}{\sqrt{s}}[p_{s+T}(X^R, y) - 1]\, ds \right|^p d\mu(y) \leqslant \left(\int_S^\infty \frac{1}{\sqrt{s}} \|p_{s+T}(X^R, \cdot) - 1\|_p\, ds \right)^p$$

where $\| \cdot \|_p$ is the $\mathrm{L}^p(d\mu(y))$ norm. By (2.10), for every $s \geqslant S$,

$$\|p_{s+T}(X^R, \cdot) - 1\|_p \lesssim e^{-s/2} \|p_S(X^R, \cdot) - 1\|_p$$

so that

$$(5.19) \qquad \left(\int_S^\infty \frac{1}{\sqrt{s}} \|p_{s+T}(X^R, \cdot) - 1\|_p\, ds \right)^p \lesssim \|p_S(X^R, \cdot) - 1\|_p^p$$

$$\lesssim \frac{1}{S^{(p-1)d/2}} e^{(p-1)|X^R|^2/2}$$

where (2.5) is used in the last step. Therefore, with the choice of $s_k = 1/\sqrt{k}$,

(5.18) is bounded from above by

$$\frac{1}{n^{p/2}}\mathbb{E}\left(\frac{1}{S^{(1-1/p)d}}e^{(1-1/p)|X^R|^2}\right)^{p/2}$$

$$\lesssim \frac{1}{n^{p/2}}\left(\sum_{k=1}^{m}k^{(1-1/p)d/2}\int_{D_k}e^{(1-1/p)|x|^2}\,d\mu(x)\right)^{p/2}$$

$$\lesssim \frac{1}{n^{p/2}}\left(\sum_{k=1}^{m}k^{(1-\frac{1}{2p})d-1}e^{(\frac{1}{2}-\frac{1}{p})k}\right)^{p/2}.$$

Recalling that $m = R^2 = 2c\log n$ with $c < 1$ yields an expression of at most the order $1/n^{p/d}$ so that (5.11) is established.

As a consequence of (5.10) and (5.11), the bound (5.8) is established.

Study of (5.7). We address here (5.7) following the steps developed for (5.8) but in a simplified way. Decomposing the integral in s simply into integrals on $(0,1)$ and $(1,\infty)$ as in the case $p = 2$, we see that (5.7) will hold provided

$$(5.20) \qquad \frac{1}{n^{p-1}}\mathbb{E}\left[\int_{\mathbb{R}^d}\left|\int_0^1\frac{1}{\sqrt{s}}[p_{s+T}(X^R,y)-1]\right|^p d\mu(y)\right] \lesssim \frac{1}{n^{p/d}}$$

and

$$(5.21) \qquad \frac{1}{n^{p-1}}\mathbb{E}\left[\int_{\mathbb{R}^d}\left|\int_1^\infty\frac{1}{\sqrt{s}}[p_{s+T}(X^R,y)-1]\right|^p d\mu(y)\right] \lesssim \frac{1}{n^{p/d}}.$$

The main simplification here with respect to (5.8) is that Fubini's theorem applies between the expectation $\mathbb{E}$ and integration against $d\mu(y)$.

Starting with (5.20), recall that by definition of the map T, for every $y \in \mathbb{R}^d$,

$$\mathbb{E}\left[\left|\int_0^1\frac{1}{\sqrt{s}}[p_{s+T}(X^R,y)-1]\,ds\right|^p\right]$$

$$= \frac{1}{\mu(B_R)}\sum_{k=1}^m\int_{D_k}\left|\int_0^1\frac{1}{\sqrt{s}}[p_{s+t_k}(x,y)-1]\,ds\right|^p d\mu(x).$$

Then, by the triangle inequality, for every $k = 1,\dots,m$,

$$\int_{D_k}\left|\int_0^1\frac{1}{\sqrt{s}}[p_{s+t_k}(x,y)-1]\,ds\right|^p d\mu(x)$$

$$\lesssim \int_{D_k}\left[\int_0^1\frac{1}{\sqrt{s}}p_{s+t_k}(x,y)\,ds\right]^p d\mu(x) + \mu(D_k),$$

and it is clear again that we may ignore the contribution $\mu(D_k)$ in what follows. By (2.4), for all $x, y \in \mathbb{R}^d$,

$$p_{s+t_k}(x,y) \leqslant \frac{1}{(1-e^{-2(s+t_k)})^{d/2}} e^{|x|^2/2} \lesssim \frac{1}{(s+t_k)^{d/2}} e^{|x|^2/2}$$

in the range $s \leqslant 1$. Hence, after integration in s, for all $x, y \in \mathbb{R}^d$,

$$\int_0^1 \frac{1}{\sqrt{s}} p_{s+t_k}(x,y)\, ds \lesssim \frac{1}{t_k^{(d-1)/2}} e^{|x|^2/2}.$$

Therefore, for every k,

$$\int_{D_k} \int_{\mathbb{R}^d} \left[\int_0^1 \frac{1}{\sqrt{s}} p_{s+t_k}(x,y)\, ds \right]^p d\mu(y)\, d\mu(x)$$

$$\lesssim \frac{1}{t_k^{(d-1)(p-2)/2}} \int_{D_k} \int_{\mathbb{R}^d} \left[\int_0^1 \frac{1}{\sqrt{s}} p_{s+t_k}(x,y)\, ds \right]^2 d\mu(y) e^{(p-2)|x|^2/2}\, d\mu(x).$$

Now

$$\int_{\mathbb{R}^d} \left[\int_0^1 \frac{1}{\sqrt{s}} p_{s+t_k}(x,y)\, ds \right]^2 d\mu(y) = \int_0^1 \int_0^1 \frac{1}{\sqrt{ss'}} p_{s+s'+2t_k}(x,x)\, ds\, ds'$$

is, by (2.3), of the order of at most $\frac{1}{t_k^{d/2-1}} e^{|x|^2/2}$. It follows that

$$\frac{1}{n^{p-1}} \sum_{k=1}^m \int_{D_k} \int_{\mathbb{R}^d} \left[\int_0^1 \frac{1}{\sqrt{s}} p_{s+t_k}(x,y)\, ds \right]^p d\mu(y)\, d\mu(x)$$

$$\lesssim \frac{1}{n^{p-1}} \sum_{k=1}^m \frac{1}{t_k^{(pd-p-d)/2}} \int_{D_k} e^{(p-1)|x|^2/2}\, d\mu(x)$$

$$\lesssim \frac{1}{n^{p-1}} \sum_{k=1}^m \frac{k^{d/2-1}}{t_k^{(pd-p-d)/2}} e^{(p-2)k/2} \lesssim \frac{1}{n^{p/d}}$$

by the choice of $t_k = n^{-2/d} e^{k/d}$, $k = 1, \ldots, m$, together with the fact that $p < d$. Hence (5.20) holds true.

Concerning (5.21), the arguments developed for (5.18) may essentially be repeated. In particular, making use of (5.19), the left-hand side of (5.21) may be seen to be bounded from above by

$$\frac{1}{n^{p-1}} \mathbb{E}(e^{(p-1)|X^R|^2/2}) \lesssim \frac{1}{n^{p-1}} \sum_{k=1}^m \int_{D_k} e^{(p-1)|x|^2/2}\, d\mu(x)$$

$$\lesssim \frac{1}{n^{p-1}} \sum_{k=1}^m k^{d/2-1} e^{(p-2)k/2}.$$

Again since $m = R^2 = 2c \log n$ with $c < 1$, this contribution is of the order at most $1/n^{p/d}$, proving (5.21).

As announced, as a consequence of (5.20) and (5.21), the bound (5.7) is established.

Study of (5.9). In the final part of the proof, we take care of the centering term

$$\int\limits_{\mathbb{R}^d}\left|\int\limits_0^\infty\frac{1}{\sqrt{s}}P_s\phi\,ds\right|^p d\mu$$

of (5.9). Recall that by definition

$$\phi = \mathbb{E}(p_T(X^R,\cdot)) - 1 = \frac{1}{\mu(B_R)}\sum_{k=1}^m[P_{t_k}(\mathbb{1}_{D_k}) - \mu(D_k)],$$

and write

$$P_s\phi = \frac{1}{\mu(B_R)}\sum_{k=1}^m[P_{s+t_k}(\mathbb{1}_{D_k}) - P_s(\mathbb{1}_{D_k})] + \frac{1}{\mu(B_R)}[P_s(\mathbb{1}_{B_R}) - \mu(B_R)]$$

$$= \frac{1}{\mu(B_R)}(\phi_{s,1} + \phi_{s,2}).$$

By means of Hölder's inequality (with respect to ds), for any $0 < \kappa < p/2 - 1$,

$$(5.22)\quad \left|\int\limits_0^\infty\frac{1}{\sqrt{s}}P_s\phi\,ds\right|^p \lesssim \int\limits_0^\infty s^{p/2-1-\kappa}e^{\kappa s}|P_s\phi|^p\,ds$$

$$\lesssim \int\limits_0^\infty s^{p/2-1-\kappa}e^{\kappa s}|\phi_{s,1}|^p\,ds + \int\limits_0^\infty s^{p/2-1-\kappa}e^{\kappa s}|\phi_{s,2}|^p\,ds.$$

We examine successively the contributions of $\phi_{s,1}$ and $\phi_{s,2}$ in the preceding. By the triangle inequality,

$$\int\limits_{\mathbb{R}^d}|\phi_{s,1}|^p\,d\mu = \int\limits_{\mathbb{R}^d}\left|\sum_{k=1}^m[P_{s+t_k}(\mathbb{1}_{D_k}) - P_s(\mathbb{1}_{D_k})]\right|^p d\mu$$

$$\leqslant \left(\sum_{k=1}^m\|P_{s+t_k}(\mathbb{1}_{D_k}) - P_s(\mathbb{1}_{D_k})\|_p\right)^p.$$

Using (2.14), for any $s > 0$ and since $t_k < 1$, we have

$$\|P_{s+t_k}(\mathbb{1}_{D_k}) - P_s(\mathbb{1}_{D_k})\|_p \lesssim \frac{e^{-s/p}}{(e^{2s} - 1)^{(p-1)/(2p)}}\sqrt{t_k}\,\mu(\partial D_k)^{1/p}.$$

For $\kappa > 0$ small enough, it follows that

$$(5.23)\quad \int\limits_0^\infty s^{p/2-1-\kappa}e^{\kappa s}\int\limits_{\mathbb{R}^d}|\phi_{s,1}|^p\,d\mu\,ds \lesssim \left(\sum_{k=1}^m\sqrt{t_k}\,\mu(\partial D_k)^{1/p}\right)^p \lesssim \frac{1}{n^{p/d}}$$

since $t_k = n^{-2/d}e^{k/d}$ and $\mu(\partial D_k) \lesssim k^{(d-1)/2}e^{-k/2}$, $k = 1,\ldots,m$.

On the other hand, it is easily seen as in the case $p = 2$ that, again for $\kappa > 0$ small enough,

$$(5.24) \quad \int_0^\infty s^{p/2-1-\kappa} e^{\kappa s} \int_{\mathbb{R}^d} |\phi_{2,s}|^p \, d\mu \, ds \lesssim \int_0^\infty s^{p/2-1-\kappa} e^{\kappa s} \int_{\mathbb{R}^d} |\phi_{2,s}|^2 \, d\mu \, ds$$

$$\lesssim \int_0^\infty s^{p/2-1-\kappa} e^{\kappa s - 2s} \mu(B_R^c) \, ds \lesssim \frac{1}{n^{p/d}}$$

by the choice of $R = \sqrt{2c \log n}$.

Together with (5.22), it follows from (5.23) and (5.24) that the bound (5.9) is established. Altogether, the proof of Theorem 1.1 is complete.

6. THE CASE $p = d$

This section addresses the proof of Theorem 1.2 for $p = d \ (\geqslant 2)$, by carefully adjusting several parameters in the various steps developed in Section 5. The proof of the case $p = d = 2$ was actually provided in [10], as a simpler version of what is developed here.

The first step is truncation to a ball B_R, this time with $R = \sqrt{2 \log n}$, for which we have similarly

$$(6.1) \qquad \mathbb{E}[\mathrm{W}_d^d(\mu_n, \mu_n^R)] \lesssim \frac{(\log n)^{d/2}}{n}.$$

The ball B_R is again decomposed as the union of m annuli

$$D_k = \{x \in \mathbb{R}^d; \ r_{k-1} \leqslant |x| < r_k\}, \quad k = 1, \ldots, m,$$

where $0 = r_0 < r_1 < \cdots < r_m = R$. Consider again the map $T : B_R \to (0,1)$ defined by $T(x) = t_k$ if $x \in D_k$. We will use the same choices $r_k = \sqrt{k}$, but modify the values of t_k as

$$t_k = \frac{e^{k/d}}{n^{2/d}\sqrt{k}}, \quad k = 1, \ldots, m.$$

Setting

$$f(y) = f_n^{R,T}(y) = \frac{1}{n} \sum_{i=1}^n p_{T(X_i^R)}(X_i^R, y), \quad y \in \mathbb{R}^d,$$

and $d\mu_n^{R,T} = f_n^{R,T} \, d\mu$, the preceding choice of the t_k's now yields in (5.3) the estimate

$$(6.2) \qquad \mathbb{E}[\mathrm{W}_d^d(\mu_n^R, \mu_n^{R,T})] \lesssim \frac{(\log n)^{\max(d/2-1, d/4)}}{n}.$$

Next, to estimate $\mathbb{E}[(\mathrm{W}_d^d(\mu_n^{R,T}, \mu)]$ as in Section 5, we need to control the terms

$$(6.3) \qquad \frac{1}{n^{d-1}} \int_{\mathbb{R}^d} \mathbb{E}\left[\left|\int_0^\infty \frac{1}{\sqrt{s}}[p_{s+T}(X^R, y) - 1]\,ds\right|^d\right] d\mu(y),$$

$$(6.4) \qquad \frac{1}{n^{d/2}} \int_{\mathbb{R}^d} \mathbb{E}\left[\left(\int_0^\infty \frac{1}{\sqrt{s}}[p_{s+T}(X^R, y) - 1]\,ds\right)^2\right]^{d/2} d\mu(y)$$

and

$$(6.5) \qquad \int_{\mathbb{R}^d} \left|\int_0^\infty \frac{1}{\sqrt{s}} P_s \phi \, ds\right|^d d\mu$$

where $\phi(y) = \mathbb{E}[p_T(X^R, y)] - 1$ for $y \in \mathbb{R}^d$.

Study of (6.4). Given $s_k = 1/\sqrt{k}$, $k = 1, \ldots, m$, it is sufficient to separately estimate

$$(6.6) \qquad \frac{1}{n^{d/2}} \int_{\mathbb{R}^d} \left(\sum_{k=1}^m \int_{D_k}\left[\int_0^{s_k} \frac{1}{\sqrt{s}}[p_{s+t_k}(x, y) - 1]\,ds\right]^2 d\mu(x)\right)^{d/2} d\mu(y)$$

and

$$(6.7) \qquad \frac{1}{n^{d/2}} \int_{\mathbb{R}^d} \left(\sum_{k=1}^m \int_{D_k}\left[\int_{s_k}^\infty \frac{1}{\sqrt{s}}[p_{s+t_k}(x, y) - 1]\,ds\right]^2 d\mu(x)\right)^{d/2} d\mu(y).$$

Concerning (6.6), with the notation of the previous section, we need to investigate

$$(6.8) \qquad \frac{1}{n^{d/2}} \int_{\mathbb{R}^d} \left[\sum_{k=1}^m \int_0^{s_k} \int_0^{s_k} \frac{1}{\sqrt{ss'}} p_{s+s'+2t_k}(y, y) \mathbb{1}_{\frac{\alpha^2}{\beta} D_k}(y)\,ds\,ds'\right]^{d/2} d\mu(y)$$

and
(6.9)
$$\frac{1}{n^{d/2}} \int_{\mathbb{R}^d} \left[\sum_{k=1}^m \int_0^{s_k} \int_0^{s_k} \frac{1}{\sqrt{ss'}} p_{s+s'+2t_k}(y, y) \mathbb{1}_{(\frac{\alpha^2}{\beta} D_k)^c}(y)\mu(\widetilde{D}_k)\,ds\,ds'\right]^{d/2} d\mu(y).$$

Arguing as for (5.15), we find that (6.8) is upper bounded by

$$\frac{1}{n^{d/2}} \sum_{k=1}^{m+80} \frac{1}{t_k^{(d/2-1)d/2}} k^{d/2-1} e^{(d/2-1)k/2} \lesssim \frac{(\log n)^{\frac{d^2+2d}{8}}}{n}.$$

Turning to (6.9), fix $k = 1, \ldots, m$. The proof proceeds as for (5.16) in Section 5 but now with a choice of $K = K_k$ and $L = L_k$ depending on k, to deduce that (6.9) is upper bounded by

$$\frac{1}{n} \Bigg[\sum_{k=1}^{m} k^{\frac{d-2}{4}} \frac{e^{[(1-\sigma)L_k - 1 + \frac{2}{d}]\frac{k}{2}}}{1 - \frac{d}{2}[(1-\sigma)K_k + \sigma]} \Bigg]^{d/2}$$

(where we recall that $\sigma \in (0, 2/d)$). Fix then $\varepsilon > 0$ small enough, and, for every $k = 1, \ldots, m$, let

$$L_k = \frac{1 - \frac{2}{d} + \frac{\varepsilon}{k}}{1 - \sigma}, \quad K_k = 1 - L_k = \frac{\frac{2}{d} - \sigma - \frac{\varepsilon}{k}}{1 - \sigma}$$

so that $K_k + L_k = 1$ and $L_k < 1$. With these choices,

$$\frac{1}{n} \Bigg[\sum_{k=1}^{m} k^{\frac{d-2}{4}} \frac{e^{[(1-\sigma)L_k - 1 + \frac{2}{d}]\frac{k}{2}}}{1 - \frac{d}{2}[(1-\sigma)K_k + \sigma]} \Bigg]^{d/2} \lesssim \frac{1}{n} \Bigg(\sum_{k=1}^{m} k^{\frac{d+2}{4}} \Bigg)^{d/2} \lesssim \frac{(\log n)^{\frac{d^2 + 6d}{8}}}{n}.$$

As a consequence of the previous analysis, (6.6) is controlled by $\frac{(\log n)^{\frac{d^2+6d}{8}}}{n}$. Concerning (6.7), it is handled as in the study of (5.11) and upper bounded by

$$\frac{1}{n^{d/2}} \Bigg(\sum_{k=1}^{m} k^{d - \frac{3}{2}} e^{(\frac{1}{2} - \frac{1}{d})k} \Bigg)^{d/2} \lesssim \frac{(\log n)^{\frac{d^2}{2} - \frac{3d}{4}}}{n}.$$

These two bounds lead to the dichotomy $d = 2, 3$ and $d \geqslant 4$ in the statement of Theorem 1.2.

Study of (6.3). This term is simpler than (6.4). Following the analysis of (5.7) in Section 5, the leading term is

$$\frac{1}{n^{d-1}} \sum_{k=1}^{m} \frac{k^{d/2-1}}{t_k^{(d^2-2d)/2}} e^{(d-2)k/2} \lesssim \frac{(\log n)^{d^2/4}}{n}$$

by the choice of $t_k = \frac{e^{k/d}}{n^{2/d}\sqrt{k}}$, $k = 1, \ldots, m$.

Study of (6.5). Using the method in the previous section, we find that (6.5) is bounded from above by

$$\Bigg(\sum_{k=1}^{m} \sqrt{t_k} \mu(\partial D_k)^{1/d} \Bigg)^{d} \lesssim \frac{(\log n)^{\frac{5d-4}{4}}}{n}.$$

since $t_k = \frac{e^{k/d}}{n^{2/d}\sqrt{k}}$ and $\mu(\partial D_k) \lesssim k^{(d-1)/2} e^{-k/2}$, $k = 1, \ldots, m$.

To conclude, it may be checked that all the logarithmic exponents are less than or equal to

$$\kappa = \max \Bigg(\frac{d^2 + 6d}{8}, \frac{d^2}{2} - \frac{3d}{4} \Bigg),$$

thereby completing the proof of Theorem 1.2.

REFERENCES

[1] M. Ajtai, J. Komlós and G. Tusnády, *On optimal matchings*, Combinatorica 4 (1984), 259–264.

[2] L. Ambrosio, F. Stra and D. Trevisan, *A PDE approach to a 2-dimensional matching problem*, Probab. Theory Related Fields 173 (2019), 433–478.

[3] D. Bakry, *Étude des transformations de Riesz sur les variétés riemanniennes à courbure de Ricci minorée*, in: Séminaire de Probabilités XXI, Lecture Notes in Math. 1247, Springer, 1987, 137–172.

[4] D. Bakry, I. Gentil and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Grundlehren Math. Wiss. 348, Springer, 2014.

[5] P. Berthet and J. Fort, *Exact rate of convergence of the expected W_2 distance between the empirical and true Gaussian distribution*, Electron. J. Probab. 25 (2020), art. 12, 16 pp.

[6] S. Bobkov and M. Ledoux, *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*, Mem. Amer. Math. Soc. 261 (2019), no. 1259, v+126 pp.

[7] S. Caracciolo, M. D'Achille and G. Sicuro, *Anomalous scaling of the optimal cost in the one-dimensional random assignment problem*, J. Statist. Phys. 174 (2019), 846–864.

[8] S. Dereich, M. Scheutzow and R. Schottstedt, *Constructive quantization: approximation by empirical measures*, Ann. Inst. H. Poincaré Probab. Statist. 49 (2013), 1183–1203.

[9] N. Fournier and A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Related Fields 162 (2015), 707–738.

[10] M. Ledoux, *On optimal matching of Gaussian samples*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklova 457 (2017), 226–264.

[11] M. Ledoux, *On optimal matching of Gaussian samples II*, 2018.

[12] P.-A. Meyer, *Transformations de Riesz pour les lois gaussiennes*, in: Séminaire de Probabilités XV, Lecture Notes in Math. 1059, Springer, 1984, 179–193.

[13] R. Peyre, *Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance*, ESAIM Control Optim. Calc. Var. 24 (2018), 1489–1501.

[14] H. P. Rosenthal, *On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables*, Israel J. Math. 8 (1970), 273–303.

[15] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Progr. Nonlinear Differential Equations Appl. 87, Birkhäuser, 2015.

[16] M. Talagrand, *Upper and Lower Bounds of Stochastic Processes*, Ergeb. Math. Grenzgeb. 60, Springer, 2014.

[17] M. Talagrand, *Scaling and non-standard matching theorems*, C. R. Math. Acad. Sci. Paris 356 (2018), 692–695.

[18] C. Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer, 2009.

[19] J. Yukich, *Some generalizations of the Euclidean two-sample matching problem*, in: Probability in Banach Spaces 8, Progr. Probab. 30, Birkhäuser, 1992, 55–66.

Michel Ledoux
Institut de Mathématiques de Toulouse
Université de Toulouse – Paul-Sabatier
and
Institut Universitaire de France
F-31062 Toulouse, France
E-mail: ledoux@math.univ-toulouse.fr

Jie-Xiang Zhu
School of Mathematical Sciences
Fudan University
Shanghai 200433, People's Republic of China
E-mail: 15110840006@fudan.edu.cn
and
Institut de Mathématiques de Toulouse
Université de Toulouse – Paul-Sabatier
E-mail: zhu@math.univ-toulouse.fr