Записки научных
семинаров ПОМИ
Том 457, 2017 г.

M. Ledoux

ON OPTIMAL MATCHING OF GAUSSIAN SAMPLES

Abstract. Let $X_1, \dots, X_n$ be independent random variables with common distribution the standard Gaussian measure μ on $\mathbb{R}^2$, and let $\mu_n = \frac{1}{n}\sum\limits_{i=1}^{n} \delta_{X_i}$ be the associated empirical measure. We show that, for some numerical constant $C > 0$,

$$\frac{1}{C}\,\frac{\log n}{n} \leqslant \mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \leqslant C\,\frac{(\log n)^2}{n}$$

where W_2 is the quadratic Kantorovich metric, and conjecture that the left-hand side provides the correct order. The proof is based on the recent PDE and mass transportation approach developed by L. Ambrosio, F. Stra and D. Trevisan.

To the memory of Professor V. N. Sudakov

§1. Introduction

Given $x_1, \dots, x_n$ and $y_1, \dots, y_n$ in $\mathbb{R}^d$, and $p \geqslant 1$, the optimal matching problem raises the question of controlling

$$\inf \frac{1}{n} \sum_{i=1}^{n} |x_i - y_{\sigma(i)}|^p$$

where the infimum runs over all permutations σ of $\{1, \dots, n\}$ (and $|\cdot|$ is the Euclidean distance on $\mathbb{R}^d$). The random matching problem deals with samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ of independent and identically distributed (iid) random variables in $\mathbb{R}^d$, and a first order analysis aims at studying the order of growth in n of the averages

$$\mathbb{E}\bigg(\inf \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \bigg). \tag{1}$$

Optimal matching problems have been investigated from various viewpoints in both the mathematics and physics literature, and we refer for example to the monographs [39, 34] for some account on the subject.

Key words and phrases: optimal matching, Ajtai–Komlós–Tusnády theorem, optimal transport, heat kernel, Gaussian sample.

A typical and most central instance of the optimal matching problem is provided by the example of independent X_i and Y_i uniformly distributed on the unit cube $[0,1]^d$. Since the typical distance between n points in $[0,1]^d$ is of order $\frac{1}{n^{1/d}}$, the quantities (1) are expected to be of the order $\frac{1}{n^{p/d}}$. However, this is only correct when $d \geqslant 3$. While it is of the order $\frac{1}{n^{p/2}}$ in dimension one due to the specific structure in this case, a major and groundbreaking result in this setting is the Ajtai–Komlós–Tusnády theorem [1] stating that in dimension $d = 2$,

$$\mathbb{E}\bigg(\inf \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \bigg) \approx \Big(\frac{\log n}{n} \Big)^{p/2} \tag{2}$$

where $A \approx B$ expresses that $C^{-1}A \leqslant B \leqslant CA$ for some $C > 0$ (independent of n). This two-dimensional phenomenon is one most interesting feature of the analysis due to the fact emphasized in [34] that "obstacles to matchings at different scales may combine in dimension 2 but not in dimension $d \geqslant 3$".

The preceding questions maybe addressed in the closely related transportation cost framework between an empirical and a reference measure. The question is then formulated in terms of the Kantorovich distances. Given $p \geqslant 1$, the Kantorovich distance (cf. [36] e.g.) between two probability measures μ and ν on the Borel sets of $\mathbb{R}^d$ with a finite p-th moment is defined by

$$\mathrm{W}_p(\nu, \mu) = \inf \bigg(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p d\pi(x,y) \bigg)^{1/p} \tag{3}$$

where the infimum is taken over all couplings π on $\mathbb{R}^d \times \mathbb{R}^d$ with respective marginals ν and μ. As is classical,

$$\inf \frac{1}{n} \sum_{i=1}^{n} |x_i - y_{\sigma(i)}|^p = \mathrm{W}_p^p\bigg(\frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}, \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i} \bigg). \tag{4}$$

Denote then by $X_1, \ldots, X_n$ 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 $(X_1, \ldots, X_n)$. It has been a main question of interest in probability and statistics to investigate the rate of convergence of μ_n to μ. In particular, the order of decay in Kantorovich distances has attracted a lot of attention. We discuss here some known results on the order of decay in n of the expectations

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big), \tag{5}$$

concentrating on upper bounds on these quantities. By the triangle inequality and (4), these bounds immediately transfer to the matching problem between two samples. The parameters entering the discussion are actually $p \geqslant 1$, the distribution μ and the dimension d, and it turns out, as emphasized above, that the two-dimensional case is a particular, most interesting, issue. We only highlight a few conclusions, refereeing to some relevant references for more complete descriptions and results. For the matter of comparison, it would have been perhaps more appropriate to consider the $\frac{1}{p}$-th power of (5), but to lighten the notation we leave it like that. Furthermore, under mild concentration properties (see [8]), the behaviours of $[\mathbb{E}(\mathrm{W}_p^p(\mu_n, \mu))]^{1/p}$ and $\mathbb{E}(\mathrm{W}_p(\mu_n, \mu))$ are of the same order.

The one-dimensional case is of particular nature due to explicit representations of the Kantorovich metrics $\mathrm{W}_p(\nu, \mu)$ in terms of the underlying distributions ν and μ. We refer to the recent monograph [8] for an account on this case. In particular, it follows from the analysis there that $\mathbb{E}(\mathrm{W}_1(\mu_n, \mu))$ is of the order of $\frac{1}{\sqrt{n}}$ for large families of distributions μ. For instance,

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \mu)\big) \, = \, O\Big(\frac{1}{\sqrt{n}}\Big)$$

as soon as $\int_{\mathbb{R}} |x|^q d\mu \, < \, \infty$ for some $q > 2$. However, when $p > 1$ some differences occur already on basic examples emphasizing the size of the support of μ as influencing the rate. For example, if μ is uniform on a compact interval, $\mathbb{E}(\mathrm{W}_p^p(\mu_n, \mu))$ is of order $\frac{1}{n^{p/2}}$ for any $p \geqslant 1$ while for the (standard) Gaussian distribution

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if } 1 \leqslant p < 2, \\ \frac{\log\log n}{n} & \text{if } \ p = 2, \\ \frac{1}{n(\log n)^{p/2}} & \text{if } \ p > 2. \end{cases} \tag{6}$$

While the rate is therefore the same as in the uniform case for $1 \leqslant p < 2$, two changes occur as $p = 2$ and $p > 2$. This result is achieved in [8] from a characterization of $\mathbb{E}(\mathrm{W}_p^p(\mu_n, \mu))$ when μ is log-concave in terms of its isoperimetric profile (including further models of interest such as for instance the exponential distribution).

Turning to $d \geqslant 1$, for μ the uniform distribution on $[0,1]^d$, the Ajtai–Komlós–Tusnády theorem [1], together with the corresponding result for $d \geqslant 3$ (cf. [39, 34]), therefore expresses that

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if} \quad d=1, \\ \big(\frac{\log n}{n}\big)^{p/2} & \text{if} \quad d=2, \\ \frac{1}{n^{p/d}} & \text{if} \quad d \geqslant 3. \end{cases} \tag{7}$$

(Actually, the upper bounds for $d \geqslant 3$ seem only formally established for $1 \leqslant p < \frac{d}{2}$ in the literature, but we provide below the suitable argument for the missing interval.) This result was established in [1] for $d = 2$ by combinatorial arguments, then reproved and made more precise by P. Shor [29] and M. Talagrand (cf. [33, 34]) via generic chaining. The point is that, from the Kantorovich representation (12),

$$\mathrm{W}_1(\mu_n,\mu) \;=\; \sup \frac{1}{n}\sum_{i=1}^{n}\big[\varphi(X_i) - \mathbb{E}\big(\varphi(X_i)\big)\big]$$

where the supremum is taken over 1-Lipschitz maps φ, and as such the study enters the framework of bounds on stochastic processes.

The corresponding results for distributions with unbounded support gave rise to a number of contributions. When $p = 1$, due to the works [17, 21, 38], the preceding extends to large families of distributions. For example, when $d = 2$,

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n,\mu)\big) = O\bigg(\sqrt{\frac{\log n}{n}}\bigg) \tag{8}$$

as soon as $\int_{\mathbb{R}^2} |x|^q d\mu < \infty$ for some $q > 2$.

When $p > 1$, the general investigations of [15, 9, 20] mainly based on dyadic decompositions yield the following typical conclusions. If for example $\int_{\mathbb{R}^d} |x|^q d\mu < \infty$ for some $q > \frac{p}{1-\kappa}$ where $\kappa = \min(\frac{p}{d}, \frac{1}{2})$, then

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) = O\Big(\frac{1}{n^\kappa}\Big). \tag{9}$$

(Actually, in [20], the case $p = \frac{d}{2}$ involves some extra logarithmic factor.) As discussed in [20], at this level of generality, these results are essentially optimal, and provide the correct orders for $d \geqslant 3$ (cf. (7)). Furthermore, for irregular laws, the decay can be faster (see [7, 15]), but we do not address this issue here. With respect to the Ajtai–Komlós–Tusnády theorem however, one structural aspect of the proof of the general bounds (9) is that, for $d = 1$ or 2, they will never yield anything better than a rate of the order of $\frac{1}{\sqrt{n}}$.

Some of the preceding bounds have been supplemented by the existence of the suitably renormalized quantity (5) as $n \to \infty$ (cf. [16, 11, 15, 7]). A major recent achievement in this regard is due to L. Ambrosio, F. Stra and D. Trevisan [4] who showed that for μ uniform on $[0,1]^2$,

$$\lim_{n\to\infty} \frac{n}{\log n}\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) = \frac{1}{4\pi}. \tag{10}$$

The result actually applies to the (normalized) uniform measure on a two-dimensional compact Riemannian manifold M, the factor $\frac{1}{4\pi}$ expressing the common small time behaviour of the trace of the Laplace operator in the form of

$$\lim_{t\to 0} 4\pi t \int_M p_t(x,x)d\mu(x) = 1$$

where $p_t(x,y)$, $t > 0$, $x, y \in M$, is the associated heat kernel. The methods of proof are based on a deep analysis combining PDE and mass transportation tools following an ansatz put forward in the physics literature [12]. As such, the rates in (7) and the limit in (10) do actually reflect the behaviour of the associated heat kernel depending in particular on the dimension. In case of the 2-dimensional sphere, a proof of the optimal matching rate is provided in the recent [22] via gravitational allocation.

The purpose of this work is to investigate $\mathbb{E}(\mathrm{W}_p^p(\mu_n,\mu))$ for the standard Gaussian law μ on $\mathbb{R}^d$ for $p = d = 2$ (with some additional results for $1 \leqslant p < 2$ and $d \geqslant 1$) with the methods emphasized in [4], replacing the heat kernel by the Mehler kernel. The main conclusion is the following statement.

Theorem 1. *Let $X_1, \ldots, X_n$ be independent with common law the standard normal distribution μ on $\mathbb{R}^2$, and set $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. For some numerical constant $C > 0$, and every $n \geqslant 2$,*

$$\frac{1}{C}\frac{\log n}{n} \leqslant \mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \leqslant C\frac{(\log n)^2}{n}. \tag{11}$$

For this specific model, these bounds are much more precise than the general orders (9) and are close to the example of the uniform measure on $[0,1]^2$. It may be conjectured that the correct order should be given by the left-hand side of (11). This conjecture is supported by simulations of F. Stra (communication from L. Ambrosio) which seem to indicate that the corresponding limit in (10) could be $\frac{1}{5}$. Besides, the argument when applied to $d \geqslant 3$ produces also an extra logarithmic factor which is not necessary by (9). On the other hand, it will be shown in the last part of this work that $\mathbb{E}(\mathrm{W}_p^p(\mu_n,\mu)) \approx \big(\frac{\log n}{n}\big)^{p/2}$ for any $1 \leqslant p < 2$, and it could be argued that in dimension one, there is a change of rate at $p = 2$.

To establish Theorem 1, we will first quantify some of the geometric parameters entering the asymptotic analysis of (10). These parameters will involve lower bounds on the curvature and upper bounds on the heat kernels. This will be achieved via a functional and mass transportation analysis, following the steps in [4] but only extracting the relevant information towards an upper bound. In this process, we will verify that the case $d \geqslant 3$ of (7) holds true for any $p \geqslant 1$ as conjectured in [4].

Turning to the content of this work, Section 2 presents the general transportation arguments inspired from [4] to bound Kantorovich distances by suitable (dual) Sobolev norms. These tools are then applied to the matching problem in the setting of weighted Riemannian manifolds in Section 3, while in the subsequent section the full range of (7) is detailed. The direct application of the general transportation bounds to the Gaussian model in Section 5 only yields weak bounds which have to be tightened in Section 6 via an additional localization argument, yielding Theorem 1. The last Section 7 addresses the range $1 \leqslant p < 2$.

§2. Transportation bounds

This paragraph is devoted to the transportation analysis of developed by L. Ambrosio, F. Stra and D. Trevisan in [4], on the basis of the heuristics of [12], yielding bounds on Kantorovich metrics by dual Sobolev norms. In order to cover at the same time the framework of [4] and instances with infinite support such as Gaussian measures, it is useful to consider the setting of so-called weighted Riemannian manifolds.

The definition (3) of the Kantorovich distance may be formulated for probability measures on a metric space, the Euclidean distance $|\cdot|$ being replaced by the distance ρ on the space. To describe the results of this work, it will be convenient to deal with a metric space arising from a smooth complete connected Riemannian manifold $(M, \mathfrak{g})$ without boundary of dimension $d \geqslant 1$, denoting by ρ the Riemannian distance and by dx the Riemannian volume element. To both deal with compact Riemannian manifolds, in which case the Riemannian volume element will be assumed to be normalized to a probability measure, and families of probability measures on $\mathbb{R}^d$ with unbounded support, we will consider weighted probability measures $d\mu = e^{-V}dx$ on $(M, \mathfrak{g})$, where $V : M \to \mathbb{R}$ is some smooth potential, and the resulting weighted Riemannian manifold $(M, \mathfrak{g}, \mu)$. The modern geometric analysis of weighted Riemannian manifolds $(M, \mathfrak{g}, \mu)$ is described by curvature-dimension conditions $CD(K, N)$, $K \in \mathbb{R}$, $N \geqslant 1$, involving a lower bound K on the extended Bakry–Émery–Ricci curvature and an upper bound N on the dimension (not necessarily topological) (cf. [36, 6]). Underlying these curvature-dimension conditions is the second order differential operator $\mathrm{L} = \Delta - \nabla V \cdot \nabla$, where Δ is the Laplace–Beltrami operator on $(M, \mathfrak{g})$, with invariant and symmetric measure μ. A typical and central example is of course simply the standard Gaussian measure $d\mu(x) = e^{-|x|^2/2}\frac{dx}{(2\pi)^{d/2}}$ on $\mathbb{R}^d$ with the associated Ornstein-Uhlenbeck operator $\mathrm{L} = \Delta - x \cdot \nabla$ yielding a weighted Riemannian manifold with curvature-dimension $CD(1, \infty)$. More general frameworks covering these instances are the settings of Markov triples (E, μ, Γ) of [6] and of $RCD^*(K, N)$ Riemannian metric measure spaces studied in [2, 3, 18] to which most of the conclusions developed here may be transferred.

Of particular usefulness in this study is the Kantorovich dual description of the metric $\mathrm{W}_p(\nu, \mu)$ as

$$\mathrm{W}_1(\nu, \mu) = \sup \bigg(\int_M \varphi d\nu - \int_M \varphi d\mu \bigg) \tag{12}$$

where the supremum runs over all 1-Lipschitz maps $\varphi : M \to \mathbb{R}$, and for $p > 1$,

$$\frac{1}{p}\mathrm{W}_p^p(\nu, \mu) = \sup \bigg(\int_M Q_1\varphi d\nu - \int_M \varphi d\mu \bigg) \tag{13}$$

where the supremum is taken over all bounded continuous functions $\varphi : M \to \mathbb{R}$ and where

$$Q_s\varphi(x) = \inf_{y \in M} \Big[\varphi(y) + \frac{\rho(x, y)^p}{ps^{p-1}} \Big], \quad s > 0, \quad x \in M,$$

is the infimum-convolution Hopf–Lax semigroup. It is classical (cf. e.g. [19]) that $Q_s\varphi(x)$, $s > 0$, $x \in M$, solves the Halmiton–Jacobi equation

$$\frac{d}{ds}\, Q_s\varphi = -\frac{1}{q}|\nabla Q_s\varphi|^q. \tag{14}$$

in $(0, \infty) \times M$ with initial condition φ, where $\frac{1}{p} + \frac{1}{q} = 1$.

One first result is a control of the Kantorovich metric $\mathrm{W}_p(\nu, \mu)$ by the $\mathrm{H}^{-1,p}$-Sobolev norm of the Radon–Nykodim derivative of ν with respect to μ. In the weighted Riemannian framework, recall the second order differential operator $\mathrm{L} = \Delta - \nabla V \cdot \nabla$ for which the integration by parts formula

$$\int_M \varphi(-\mathrm{L}\psi)d\mu = \int_M \nabla\varphi \cdot \nabla\psi\, d\mu \tag{15}$$

holds true for all smooth $\varphi, \psi : M \to \mathbb{R}$. Denote by $(P_t)_{t\geqslant 0}$ the Markov semigroup with infinitesimal generator L [6]. Formally the inverse $(-\mathrm{L})^{-1}$ of the non-negative operator $-\mathrm{L}$ may be described by

$$(-\mathrm{L})^{-1} = \int_0^\infty P_t\, dt$$

acting on mean zero functions in the suitable domain, a core of which being the set C_c^∞ of C^∞ compactly supported functions on M. Whenever the spectrum $\sigma(-\mathrm{L})$ of $-\mathrm{L}$ is discrete, $(-\mathrm{L})^{-1}$ can be spectrally represented on a suitable function f as

$$(-\mathrm{L})^{-1}f \;=\; \sum_{\lambda\in\sigma(-\mathrm{L})\setminus\{0\}} \frac{1}{\lambda}\, f_\lambda u_\lambda \tag{16}$$

where $(u_\lambda)_{\lambda\in\sigma(-\mathrm{L})}$ is an $\mathrm{L}^2(\mu)$ orthonormal basis of eigenvectors and $f_\lambda = \langle f, u_\lambda\rangle$. Such a picture occurs on a compact manifold for example. On $\mathbb{R}^d$ equipped with the standard Gaussian measure μ, the family of Hermite polynomials provides an orthonormal basis of eigenvectors of $\mathrm{L}^2(\mu)$ with eigenvalues $\lambda_k = k$, $k \in \mathbb{N}$ (counted with mutiplicity).

Define then, for every $p \geqslant 1$, the dual Sobolev norm $\mathrm{H}^{-1,p}(\mu)$ by

$$\|g\|_{\mathrm{H}^{-1,p}(\mu)} = \bigg(\int_M \big|\nabla((-\mathrm{L})^{-1}g)\big|^p\, d\mu\bigg)^{1/p}$$

for functions $g : M \to \mathbb{R}$ with $\int_M g d\mu = 0$ for which $\nabla((-\mathrm{L})^{-1}g)$ exists and belongs to $\mathrm{L}^p(\mu)$. In the particular case $p = 2$, the integration by parts formula (15) and the symmetry of $(P_t)_{t\geqslant 0}$ yield

$$\begin{aligned}\int_M \big|\nabla((-\mathrm{L})^{-1}g)\big|^2 d\mu &= \int_M g(-\mathrm{L})^{-1}g\, d\mu \\ &= \int_0^\infty \int_M gP_t g\, d\mu\, dt \\ &= 2\int_0^\infty \int_M (P_t g)^2 d\mu\, dt,\end{aligned} \tag{17}$$

and in particular a simpler description of the admissible functions g.

For general $p \neq 2$, a variant of the dual Sobolev norm is provided by Riesz transforms inequalities. For example on a compact manifold for the Riemannian measure $d\mu = dx$, for any $1 < p < \infty$ and any smooth $g : M \to \mathbb{R}$,

$$\int_M |\nabla g|^p\, d\mu \leqslant C\bigg(\int_M \big|(-\mathrm{L})^{1/2}g\big|^p d\mu + \int_M |g|^p d\mu)\bigg).$$

This inequality follows from the more general investigation of Riesz transforms on weighted Riemannian manifolds satisfying the curvature condition $CD(K,\infty)$ for some $K \in \mathbb{R}$ developed by D. Bakry in [5]. It is a consequence of his result that the second term $\int_M |g|^p d\mu$ on the right-hand side may be omitted when $K = 0$, including the particular example of the Gaussian model (going back in this case to [26]). It may be mentioned in addition that this term may also be omitted when g is centered, at least in the compact setting. We briefly discuss this issue in the relevant case $p > 2$. Indeed, it is enough to this purpose to ensure that

$$\|g\|_p^p = \int_M |g|^p d\mu \leqslant C\int_M \big|(-\mathrm{L})^{1/2}g\big|^p d\mu = C\big\|(-\mathrm{L})^{1/2}g\big\|_p^p.$$

But, by the Hardy–Littlewood theory for semigroups of N. Varopoulos [35],

$$\|g\|_p \leqslant C\big(\big\|(-\mathrm{L})^{1/2}g\big\|_r + \|g\|_r\big)$$

where $\frac{1}{p} = \frac{1}{r} - \frac{1}{d}$ with $1 < r < d$. Hence

$$\|g\|_p \leqslant C\big(\big\|(-\mathrm{L})^{1/2} g\big\|_p + \|g\|_r\big)$$

and after, if necessary, a finite number of iterations, it may be assumed that $r \leqslant 2$. But then, by the spectral gap or Poincaré inequality (see (19) below) – which holds true on a compact manifold –, since g is centered,

$$\|g\|_r \leqslant \|g\|_2 \leqslant C\,\big\|(-\mathrm{L})^{1/2} g\big\|_2 \leqslant C\,\big\|(-\mathrm{L})^{1/2} g\big\|_p$$

so that we indeed reach $\|g\|_p \leqslant C\|(-\mathrm{L})^{1/2} g\|_p$.

As a consequence of this discussion, on a compact manifold, for $p > 2$ and any mean zero smooth function $g : M \to \mathbb{R}$,

$$\|g\|^p_{\mathrm{H}^{-1,p}(dx)} = \int_M \big|\nabla(-\mathrm{L})^{-1} g)\big|^p \, d\mu \leqslant C \int_M \big|(-\mathrm{L})^{-1/2} g\big|^p d\mu. \tag{18}$$

This result will be used in Section 4 when extending (7) to any $p \geqslant 1$ and $d \geqslant 1$.

The following statement is the main energy estimate on the Kantorovich distance $\mathrm{W}_p(\nu,\mu)$ between two probability measures ν and μ with ν absolutely continuous with respect to μ by the dual Sobolev norm $\mathrm{H}^{-1,p}(\mu)$ of the Radon-Nikodym density $f = \frac{d\nu}{d\mu}$.

Theorem 2. *For any $1 \leqslant p < \infty$, and for all $d\nu = f d\mu$ with $f - 1$ in the domain of the dual Sobolev norm $\mathrm{H}^{-1,p}(\mu)$,*

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

When $p = 2$, Theorem 2 is closely related to Poincaré-type inequalities and their connection with transportation cost inequalities. Recall that if $d\nu = f d\mu$, the relative entropy of ν with respect to μ is given by

$$\mathrm{H}\big(\nu \,|\, \mu\big) = \int_M f \log f \, d\mu.$$

It is a standard result (cf. [36, 6]) that if μ satisfies the quadratic transportation cost inequality

$$\mathrm{W}_2^2(\nu,\mu) \leqslant 2C\mathrm{H}\big(\nu \,|\, \mu\big)$$

for some constant $C > 0$ and every ν absolutely continuous with respect to μ, then μ satisfies the Poincaré inequality

$$\int_M g^2 d\mu \leqslant C \int_M |\nabla g|^2 d\mu \tag{19}$$

for any smooth $g : M \to \mathbb{R}$ with $\int_M g d\mu = 0$. This property may be established by a Taylor expansion on $d\nu_\varepsilon = (1 + \varepsilon g) d\mu$ as $\varepsilon \to 0$ together with the limit (cf. [36])

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \mathrm{W}_2^2(\nu_\varepsilon, \mu) = \|g\|^2_{\mathrm{H}^{-1,2}(\mu)}.$$

In view of this asymptotics, the inequality of Theorem 2 is of the correct order for $p = 2$ up to a factor 4.

Proof of Theorem 2. Let first $p > 1$. By a standard regularization procedure, it may be assumed that f is smooth and that $f > 0$. Set then $g = f - 1$ so that $g > -1$ and $\int_M g d\mu = 0$. Let $\theta : [0,1] \to [0,1]$ be increasing, smooth, with $\theta(0) = 0$ and $\theta(1) = 1$. For every bounded continuous $\varphi : M \to \mathbb{R}$, by the Hamilton–Jacobi equation (14),

$$\begin{aligned}
&\int_M Q_1 \varphi \, d\nu - \int_M \varphi d\mu \\
&= \int_0^1 \frac{d}{ds} \int_M \big(1 + \theta(s) g\big) Q_s \varphi \, d\mu \, ds \\
&= \int_0^1 \int_M \Big[\theta'(s) g \, Q_s \varphi - \big(1 + \theta(s) g\big) \frac{1}{q} |\nabla Q_s \varphi|^q\Big] d\mu \, ds \\
&= \int_0^1 \int_M \Big[- \theta'(s) \nabla\big((-\mathrm{L})^{-1} g\big) \cdot \nabla Q_s \varphi - \big(1 + \theta(s) g\big) \frac{1}{q} |\nabla Q_s \varphi|^q\Big] d\mu \, ds
\end{aligned}$$

where we used integration by parts (15) in the last step.

By Young's inequality $a \cdot b \leqslant \frac{|a|^p}{p} + \frac{|b|^q}{q}$,

$$\int_M Q_1 \varphi \, d\nu - \int_M \varphi \, d\mu \leqslant \frac{1}{p} \int_0^1 \theta'(s)^p \int_M \frac{|\nabla((-\mathrm{L})^{-1} g)|^p}{[1 + \theta(s) g]^{p-1}} \, d\mu \, ds$$

and since $g > -1$,

$$\int_M Q_1\varphi\, d\nu - \int_M \varphi\, d\mu \leqslant \frac{1}{p}\int_0^1 \frac{\theta'(s)^p}{[1-\theta(s)]^{p-1}}\, ds \int_M \big|\nabla((-\mathrm{L})^{-1}g)\big|^p\, d\mu.$$

Therefore, by the Kantorovich duality formula (13),

$$\mathrm{W}_p^p(\nu,\mu) \leqslant \int_0^1 \frac{\theta'(s)^p}{[1-\theta(s)]^{p-1}}\, ds \int_M \big|\nabla((-\mathrm{L})^{-1}g)\big|^p\, d\mu.$$

The optimal choice of θ is provided by $\theta(s) = 1-(1-s)^p$ for which $\int_0^1 \frac{\theta'(s)^p}{[1-\theta(s)]^{p-1}}\, ds = p^p$. The proof is thereby completed by the Kantorovich duality formula (13). The conclusion extends (or by a direct argument) to $p=1$. Theorem 2 is established. □

§3. Application to the matching problem

In this first section on the matching problem, we address the issue of identifying the geometric features on a bound on $\mathrm{W}_p(\mu_n,\mu)$ by means of Theorem 2. We thus consider a probability measure $d\mu = e^{-V}dx$ on complete Riemannian manifold $(M,\mathfrak{g})$, invariant measure of the second order differential operator $\mathrm{L} = \Delta - \nabla V\cdot\nabla$. Let $X_1,\dots,X_n$ be a sample of independent random variables with distribution μ and $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. We follow the arguments of the investigation [4].

The first step is a (short time) regularization procedure by the heat kernel. Recall the semigroup $(P_t)_{t\geqslant 0}$ with generator L, and denote by $p_t(x,y)$, $t>0$, $x,y\in M$, the (symmetric) heat kernel such that for any suitable $\varphi : M\to\mathbb{R}$,

$$P_t\varphi(x) = \int_M \varphi(y)p_t(x,y)d\mu(y), \quad t>0,\, x\in M.$$

In particular $\int_M p_t(x,y)d\mu(y) = 1$ for every x and by the semigroup property

$$p_{s+t}(x,y) = \int_M p_s(x,z)\, p_t(z,y)d\mu(z)$$

for $s,t>0$, $x,y\in M$.

Fix $t > 0$ and set

$$f_n^t(y) = \frac{1}{n}\sum_{i=1}^n p_t(X_i, y)$$

and $d\mu_n^t = f_n^t d\mu$. By the standard convexity of the Kantorovich metrics W_p^p, $p \geqslant 1$, which is for example immediately checked on the dual representation (13) (cf. e.g. [36]),

$$\begin{aligned}\mathrm{W}_p^p(\mu_n, \mu_n^t) &\leqslant \frac{1}{n}\sum_{i=1}^n \mathrm{W}_p^p\big(\delta_{X_i}, p_t(X_i, \cdot\,)d\mu\big)\\ &= \frac{1}{n}\sum_{i=1}^n \int_M \rho(X_i, y)^p\, p_t(X_i, y)d\mu(y).\end{aligned}$$

After taking expectation in the iid X_i's,

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu_n^t)\big) \leqslant \int_M\int_M \rho(x, y)^p\, p_t(x, y)d\mu(x)d\mu(y) = D_t^p. \tag{20}$$

The quantity D_t^p is called the dispersion factor, and in various instances (see below), may be shown to be of the order of $t^{p/2}$ (for small $t > 0$).

The second step is the mere application of the energy estimate of Theorem 2. Namely, with $f = f_n^t$, and after integration with respect to the random variables X_i, $i = 1, \ldots, n$,

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n^t, \mu)\big) \leqslant p^p\, \mathbb{E}\big(\|f_n^t - 1\|_{\mathrm{H}^{-1,p}(\mu)}^p\big).$$

Since

$$\nabla(-\mathrm{L}_y)^{-1}(f_n^t - 1)(y) = \frac{1}{n}\sum_{i=1}^n \nabla_y(-\mathrm{L}_y)^{-1}\big[p_t(X_i, y) - 1\big],$$

it follows that

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n^t, \mu)\big) \leqslant p^p \int_M \mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{i=1}^n \nabla_y(-\mathrm{L}_y)^{-1}\big[p_t(X_i, y) - 1\big]\bigg|^p\bigg)d\mu(y).$$

Together with (20) and the triangle inequality for W_p, we conclude that for any $t > 0$,

$$\begin{aligned}&\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big)\\ &\qquad\leqslant C_p\bigg(D_t^p + \int_M \mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{i=1}^n \nabla_y(-\mathrm{L}_y)^{-1}\big[p_t(X_i, y) - 1\big]\bigg|^p\bigg)d\mu(y)\bigg),\end{aligned} \tag{21}$$

where $C_p > 0$ only depends on p.

Of course, the preceding formula requires that the density f_n^t, that is the heat kernels $p_t(X_i, \cdot)$, belongs to the dual Sobolev space $\mathrm{H}^{-1,p}(\mu)$. As we have seen, in the particular case $p = 2$,

$$\|f_n^t - 1\|^2_{\mathrm{H}^{-1,2}(\mu)} = 2 \int_0^\infty \int_M \big[P_s(f_n^t - 1)\big]^2 d\mu\, ds$$

so that by the semigroup property

$$\|f_n^t - 1\|^2_{\mathrm{H}^{-1,2}(\mu)} = 2 \int_t^\infty \int_M \bigg[\frac{1}{n}\sum_{i=1}^n \big[p_s(X_i, y) - 1\big]\bigg]^2 d\mu(y) ds.$$

For each $s > 0$ and $y \in M$, the random variables $p_s(X_i, y) - 1$, $i = 1, \dots, n$, are independent, centered and identically distributed, so that taking expectation in the X_i's,

$$\begin{aligned}
\mathbb{E}\big(\mathrm{W}_2^2(\mu_n^t, \mu)\big) &\leqslant 4\, \mathbb{E}\big(\|f_n^t - 1\|^2_{\mathrm{H}^{-1,2}(\mu)}\big) \\
&= 8 \int_t^\infty \int_M \mathbb{E}\bigg(\bigg[\frac{1}{n}\sum_{i=1}^n \big[p_s(X_i, y) - 1\big]\bigg]^2\bigg) d\mu(y) ds \\
&= \frac{8}{n} \int_t^\infty \int_M \mathbb{E}\big(\big[p_s(X_1, y) - 1\big]^2\big) d\mu(y)\, ds \\
&= \frac{8}{n} \int_t^\infty \int_M \int_M \big[p_s(x, y) - 1\big]^2 d\mu(x) d\mu(y)\, ds \\
&= \frac{4}{n} \int_{2t}^\infty \int_M \big[p_s(x, x) - 1\big] d\mu(x)\, ds
\end{aligned} \tag{22}$$

where the last step follows from the semigroup property.

Together with (20) and the triangle inequality for W_2, we thus end up with the following general statement for $p = 2$ which splits the control of $\mathbb{E}(\mathrm{W}_2^2(\mu_n, \mu))$ in terms of the dispersion factor and the energy functional.

Proposition 3. *In the prescribed setting and notation, for every $t > 0$,*

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \leqslant 2D_t^2 + \frac{8}{n} \int_{2t}^{\infty} \int_M \big[p_s(x,x) - 1\big] d\mu(x)\, ds. \tag{23}$$

Note that in presence of a discrete spectrum $\sigma(-\mathrm{L})$ for $-\mathrm{L}$, the trace formula provides the useful representation

$$\int_M \big[p_s(x,x) - 1\big] d\mu(x) = \sum_{\lambda \in \sigma(-\mathrm{L}) \setminus \{0\}} e^{-\lambda s}$$

and thus

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \leqslant 2D_t^2 + \frac{8}{n} \sum_{\lambda \in \sigma(-\mathrm{L}) \setminus \{0\}} \frac{1}{\lambda}\, e^{-2\lambda t}.$$

The task is now to discuss how to control the two terms on the right-hand side of (23) and to optimize in $t > 0$.

The control of the energy

$$\int_{2t}^{\infty} \int_M \big[p_s(x,x) - 1\big] d\mu(x)\, ds$$

may be achieved by splitting between the small time behaviour and the large time one. In large time, we can make use of a spectral gap or Poincaré inequality hypothesis. Assume thus that μ satisfies a Poincaré inequality (19) with constant $C_P > 0$. As is classical [6], such a Poincaré inequality ensures an exponential decay of convergence to equilibrium in the sense that for any $\varphi : M \to \mathbb{R}$ with mean zero and any $u > 0$,

$$\int_M (P_u \varphi)^2 d\mu \;\leqslant\; e^{-2u/C_P} \int_M \varphi^2 d\mu. \tag{24}$$

Apply this to $\varphi(y) = p_v(x,y) - 1$, $v > 0$, $y \in M$, for $x \in M$ fixed. Since then $P_u \varphi(y) = p_{u+v}(x,y) - 1$,

$$\int_M \big[p_{u+v}(x,y) - 1\big]^2 d\mu(y) \;\leqslant\; e^{-2u/C_P} \bigg(\int_M \big[p_v(x,y) - 1\big]^2 d\mu(y) \bigg).$$

In the applications, this inequality may be used in two ways. First, by the semigroup property and the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\big[p_{2(u+v)}(x,y)-1\big]^2 \\
&\quad = \bigg(\int\limits_M \big[p_{u+v}(x,z)-1\big]\big[p_{u+v}(z,y)-1\big]\,d\mu(z)\bigg)^2 \\
&\quad \leqslant \int\limits_M \big[p_{u+v}(x,z)-1\big]^2 d\mu(z)\int\limits_M \big[p_{u+v}(z,y)-1\big]^2 d\mu(z) \\
&\quad \leqslant e^{-4u/C_P}\int\limits_M \big[p_{v}(x,z)-1\big]^2 d\mu(z)\int\limits_M \big[p_{v}(z,y)-1\big]^2 d\mu(z).
\end{aligned} \tag{25}$$

On the other hand, after integration in $d\mu(x)$ and making use again of the semigroup property,

$$\int\limits_M \big[p_{u+v}(x,x)-1\big]\,d\mu(x) \leqslant e^{-2u/C_P}\int\limits_M \big[p_{v}(x,x)-1\big]\,d\mu(x). \tag{26}$$

Next, we assume a uniform small time bound on the heat kernel in the form of the existence of a constant $C_u > 0$ such that

$$p_s(x,y) \leqslant \frac{C_u}{s^{d/2}}, \quad 0 < s \leqslant 1, \;\; x,y \in M. \tag{27}$$

As will be clear below and in the next section, this heat kernel behaviour is actually responsible for the various rates in (7) depending on d.

We then combine the large and small time behaviours in the following way. Together with the spectral gap bound (26) and the uniform heat kernel bound (27), we get that for every $0 < t \leqslant \frac{1}{2}$,

$$
\begin{aligned}
\int_{2t}^{\infty}\int_M \big[p_s(x,x)-1\big]\,d\mu(x)\,ds &= \int_{2t}^{1}\int_M \big[p_s(x,x)-1\big]\,d\mu(x)\,ds \\
&\quad + \int_{0}^{\infty}\int_M \big[p_{s+1}(x,x)-1\big]\,d\mu(x)ds \\
&\leqslant \int_{2t}^{1} \frac{C_u}{s^{d/2}}\,ds + \int_0^{\infty} C_u\, e^{-2s/C_P}ds \\
&\leqslant C_d\Big(\frac{C_u}{t^{(d/2)-1}} + C_uC_P\Big)
\end{aligned}
\tag{28}
$$

or $C_u \log\big(\frac{1}{t}\big) + C_uC_P$ if $d = 2$, where $C_d > 0$ only depends on d.

After optimization in $t > 0$ in (23), we may therefore conclude to the following statement.

Theorem 4. *In the preceding setting, assume that the dispersion factor D_t^2 is linear in small time, that is $D_t^2 \leqslant C_D t$ for every $0 < t \leqslant 1$, that μ satisfies a Poincaré inequality with constant $C_P > 0$ and that the uniform heat kernel bound* (27) *holds true for some $C_u > 0$. Then*

$$
\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \leqslant \begin{cases} C\,\frac{\log n}{n} & \text{if} \quad d = 2, \\ C\,\frac{1}{n^{2/d}} & \text{if} \quad d \geqslant 3, \end{cases}
$$

where $C = C(d, C_D, C_P, C_u)$.

As discussed in [4], the hypotheses of the preceding statement are satisfied for the normalized Riemannian volume element μ on a compact d-dimensional Riemannian manifold $(M, \mathfrak{g})$, with constants depending on the geometry of the manifold via dimension, diameter and Ricci curvature lower bounds. We discuss here a little more precisely the geometric ingredients underlying these conditions in the extended context of weighted Riemannian manifolds $(M, \mathfrak{g})$ with weighted probability measure $d\mu = e^{-V}dx$. We may refer to the monographs [14, 13, 37, 6] for accounts and references on these standard properties.

The heat kernel bound (27) classically follows from a Sobolev-type inequality

$$\left(\int_M |\varphi|^q d\mu\right)^{2/q} \leqslant A\int_M \varphi^2 d\mu + B\int_M |\nabla\varphi|^2 d\mu \tag{29}$$

for some $A, B > 0$ and any smooth $\varphi : M \to \mathbb{R}$, where $q = \frac{2d}{d-2}$, $d > 2$. While the Sobolev inequalities device requires $d > 2$, suitable substitutes in terms of Nash-type or logarithmic Sobolev inequalities may be developed to include $d \geqslant 1$ in the heat kernel bounds (27). All these results are presented for example in the monograph [6] in the extended setting of Markov triples (E, μ, Γ) which encompasses the weighted Riemannian manifold framework.

In the weighted Riemannian framework, the Poincaré inequality (19) holds with $C_P = \frac{1}{K} > 0$ under the curvature condition $CD(K, \infty)$ with $K > 0$ while the Sobolev inequality (29) holds under the curvature-dimension condition $CD(K, d)$ for some $K > 0$ and $d > 2$. By standard elliptic theory [13], a spectral gap inequality holds on a compact Riemannian manifold $(M, \mathfrak{g})$, and a Sobolev (29) inequality holds with constants depending on dimension, diameter and lower bound on the extended curvature $CD(K, d)$ [37, 6].

The study of the dispersion factor D_t^2 is of somewhat different nature, although also connected to curvature bounds. In the non-weighted setting of a compact manifold $(M, \mathfrak{g})$ with Riemannian volume element $d\mu = dx$ and Laplace operator Δ, the classical Laplacian comparison principle [13, 37] expresses that for all $x, y \in M$,

$$\Delta\big(\rho(x,\cdot)\big)(y) \leqslant (d-1)\xi\big(\rho(x,y)\big) \tag{30}$$

where $\xi = \frac{\zeta'}{\zeta}$ and

$$\zeta(r) = \begin{cases} r & \text{if} \quad K = 0, \\ \frac{\sin(\sqrt{K}r)}{\sqrt{K}} & \text{if} \quad K > 0, \\ \frac{\sinh(\sqrt{-K}r)}{\sqrt{-K}} & \text{if} \quad K < 0, \end{cases}$$

where $\mathrm{Ric} \geqslant K(d-1)$ for some $K \in \mathbb{R}$. It thus follows that for all fixed $x \in M$,

$$\Delta\big(\rho^2(x,\cdot)\big) \leqslant C$$

for some $C > 0$ only depending on d, K and the diameter of M. Therefore,

$$\frac{d}{dt} D_t^2 = \int_M \int_M \rho(x,y)^2 \, \Delta_y p_t(x,y) d\mu(x) d\mu(y)$$
$$= \int_M \int_M \Delta_y \big(\rho(x,y)^2\big) p_t(x,y) d\mu(x) d\mu(y) \leqslant C.$$

As a consequence, $D_t^2 \leqslant Ct$, $t > 0$, for some $C > 0$ only depending on the dimension d of the manifold, the lower bound K on the Ricci curvature, and the diameter of $(M, \mathfrak{g})$.

When $1 \leqslant p \leqslant 2$, Jensen's inequality ensures that $D_t^p \leqslant Ct^{p/2}$. For $p \geqslant 2$, we first show by induction that for any integer $k \geqslant 1$, $D_t^{2k} \leqslant Ct^k$ where $C = C(k) > 0$ depending on the geometry of M may vary from line to line. To this task, since $|\nabla \rho(x, \cdot)| \leqslant 1$,

$$\frac{d}{dt} D_t^{2(k+1)} = \int_M \int_M \Delta_y \big(\rho(x,y)^{2(k+1)}\big) p_t(x,y) d\mu(x) d\mu(y)$$
$$\leqslant 2(k+1) \int_M \int_M \rho(x,y)^{2k+1} \Delta_y \rho(x,y) p_t(x,y) d\mu(x) d\mu(y)$$
$$+ 2(k+1)(2k+1) D_t^{2k}$$

and by the Laplacian comparison (30),

$$\frac{d}{dt} D_t^{2(k+1)} \leqslant C D_t^{2k}.$$

By the induction hypothesis, $\frac{d}{dt} D_t^{2(k+1)} \leqslant Ct^k$ and therefore

$$D_t^{2(k+1)} \leqslant \int_0^t Cs^k ds \leqslant C\, t^{k+1}.$$

Finally, by Hölder's inequality, for any $2k < p < 2k+2$,

$$(D_t^p)^{1/p} \leqslant (D_t^{2k})^{\theta/2k} (D_t^{2k+2})^{(1-\theta)/(2k+2)}$$

where $\theta \in (0,1)$. Hence again $(D_t^p)^{1/p} \leqslant Ct^{1/2}$, $t > 0$.

We thus conclude that

$$D_t^p \leqslant Ct^{p/2} \tag{31}$$

for all $p \geqslant 1$ and $t > 0$ where $C > 0$ depends on the compact manifold $(M, \mathfrak{g})$. Extensions may be provided in weighted manifolds, but for simplicity we only consider one such example, namely the Gaussian measure setting addressed in the next sections, in which the explicit semigroup description allows for a simple argument towards the control of the dispersion factor.

§4. The Ajtai–Komlós–Túsnady upper bounds for all parameters

As a consequence of the preceding analysis, Theorem 4 covers the case $p = 2$ of the Ajtai–Komlós–Tusnády upper bounds (7) for the (normalized) uniform measure μ on a compact Riemannian manifold $(M, \mathfrak{g})$ by the methodology of [4]. In this section, we address the upper bounds in the missing range $p \geqslant \frac{d}{2}$, $d \geqslant 3$, in (7), that is

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) = \begin{cases} O\big(\frac{1}{n^{p/2}}\big) & \text{if} \quad d = 1, \\ O\big(\big(\frac{\log n}{n}\big)^{p/2}\big) & \text{if} \quad d = 2, \\ O\big(\frac{1}{n^{p/d}}\big) & \text{if} \quad d \geqslant 3. \end{cases} \tag{32}$$

(In this setting, lower bounds might followed from the strategy developed in [4] for $p = d = 2$ but further details are certainly necessary in this regard.) In the compact Riemannian manifold framework, all the necessary smoothness, curvature and heat kernel bounds are satisfied. In addition, the Riesz transform bounds (18) may also be used. The provided treatment actually includes all values of $p \geqslant 1$ and $d \geqslant 1$.

Assume first that $d \geqslant 3$ to simplify some expressions. We start from (21) together therefore with the Riesz transform bound (18), yielding

$$\begin{aligned} &\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \\ &\quad \leqslant C\bigg(D_t^p + \int_M \mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{i=1}^n (-\mathrm{L}_y)^{-1/2}\big[p_t(X_i, y) - 1\big]\bigg|^p\bigg) d\mu(y)\bigg). \end{aligned} \tag{33}$$

Here and below, $C > 0$ denotes a constant, depending on p and the underlying geometric structure but not of n, and possibly varying from line to line.

For each $y \in M$, the random variables

$$(-\mathrm{L}_y)^{-(1/2)}[p_t(X_i, y) - 1], \quad i = 1, \ldots, n,$$

are independent, centered and identically distributed. By Rosenthal's inequality [28],

$$\begin{aligned} \mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{i=1}^{n}(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_i,y)-1\big]\bigg|^p\bigg) \\ \leqslant C_p\bigg(\frac{1}{n^{p-1}}\,\mathbb{E}\Big(\big|(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big|^p\Big) \\ +\frac{1}{n^{p/2}}\,\mathbb{E}\Big(\big[(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big]^2\Big)^{p/2}\bigg) \end{aligned} \tag{34}$$

for some constant $C_p > 0$ only depending on $p \geqslant 2$. When $1 \leqslant p \leqslant 2$, it is enough to keep the second piece on the right-hand side.

By spectral analysis,

$$(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big] = \frac{\sqrt{\pi}}{2}\int_0^\infty \frac{1}{\sqrt{s}}\,\big[p_{t+s}(X_1,y)-1\big]ds.$$

The uniform bound on the heat kernel (27) ensures that for all $x, y \in M$ and $0 < t \leqslant \frac{1}{2}$,

$$\bigg|\int_0^{1/2}\frac{1}{\sqrt{s}}\,\big[p_{t+s}(x,y)-1\big]ds\bigg| \leqslant \sqrt{2} + \int_0^{1/2}\frac{1}{\sqrt{s}}\cdot\frac{C}{(t+s)^{d/2}}\,ds \leqslant \frac{C}{t^{(d-1)/2}}.$$

On the other hand,

$$\bigg|\int_{1/2}^\infty \frac{1}{\sqrt{s}}\,\big[p_{t+s}(x,y)-1\big]ds\bigg| \leqslant \sqrt{2}\int_{1/2}^\infty \big|p_{t+s}(x,y)-1\big|\,ds.$$

By the exponential decay (25), and again (27), uniformly in $s \geqslant \frac{1}{2}$ and $x, y \in M$,

$$\big|p_{t+s}(x,y)-1\big| \leqslant C\,e^{-s/C}.$$

Summarizing the preceding steps,

$$\big|(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big| \,\leqslant\, \frac{C}{t^{(d-1)/2}}$$

for every $0 < t \leqslant \frac{1}{2}$. Hence,

$$\int_M \mathbb{E}\Big(\big|(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big|^p\Big)d\mu(y) \leqslant \frac{C}{t^{(d-1)(p-2)/2}} \int_M \mathbb{E}\Big(\big[(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big]^2\Big)d\mu(y).$$

Now, by the representation $(-\mathrm{L})^{-1} = \int_0^\infty P_s ds$,

$$\begin{aligned}\int_M \mathbb{E}\Big(\big[(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big]^2\Big)d\mu(y) &= \int_M\int_M p_t(x,y)(-\mathrm{L}_y)^{-1}p_t(x,y)d\mu(x)d\mu(y) \\ &= \int_0^\infty\int_M\int_M p_t(x,y)\,p_{t+s}(x,y)d\mu(x)d\mu(y)ds \\ &= \int_{2t}^\infty\int_M \big[p_s(x,x)-1\big]d\mu(x)ds\end{aligned}$$

which is bounded by $\frac{C}{t^{(d/2)-1}}$ for every $0 < t \leqslant \frac{1}{2}$ by (28). Therefore, in the range $0 < t \leqslant \frac{1}{2}$,

$$\int_M \mathbb{E}\Big(\big|(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big|^p\Big)d\mu(y) \leqslant \frac{C}{t^{(d-1)(p-2)/2}}\cdot\frac{C}{t^{(d/2)-1}}.$$

We next investigate the second term on the right-hand of (34). By symmetry, for every y,

$$\begin{aligned}\mathbb{E}\Big(\big[(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big]^2\Big) &= \mathbb{E}\Big(\big[(-\mathrm{L}_x)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big]^2\Big) \\ &= \int_{2t}^\infty \big[p_s(y,y)-1\big]\,ds.\end{aligned}$$

An analysis similar to the preceding one for both small and large values of s shows that

$$\mathbb{E}\Big(\big[(-\mathrm{L}_y)^{-(1/2)}\big[p_t(X_1,y)-1\big]\big]^2\Big) \leqslant \frac{C}{t^{(d/2)-1}}$$

for every $0 < t \leqslant \frac{1}{2}$ and $y \in M$.

These estimates in Rosenthal's inequality (34) therefore lead to

$$\int_M \mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{i=1}^n (-\mathrm{L}_y)^{-1/2}\big[p_t(X_i,y)-1\big]\bigg|^p\bigg)d\mu(y)$$
$$\leqslant C\bigg(\frac{1}{n^{p-1}}\cdot\frac{1}{t^{(d-1)(p-2)/2+(d/2)-1}}+\frac{1}{n^{p/2}}\cdot\frac{1}{t^{(p/2)((d/2)-1)}}\bigg)$$

for $0 < t \leqslant \frac{1}{2}$. Together with the dispersion rate (31), we thus get that

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big)\leqslant C\bigg(t^{p/2}+\frac{1}{n^{p-1}}\cdot\frac{1}{t^{(d-1)(p-2)/2+(d/2)-1}}+\frac{1}{n^{p/2}}\cdot\frac{1}{t^{(p/2)((d/2)-1)}}\bigg)$$

for $0 < t \leqslant \frac{1}{2}$. Optimizing in $t > 0$, namely choosing $t \sim \frac{1}{n^{2/d}}$, thus yields

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) = O\Big(\frac{1}{n^{p/d}}\Big).$$

This is the announced claim when $d \geqslant 3$. Obvious modifications for $d = 1$ and 2 yield similarly the corresponding conclusions, completing the picture in (32).

It could be noted that the dependence in $p \geqslant 2$ in Rosenthal's inequality (34) is of order $\big(\frac{p}{\log p}\big)^p$ [23]. Following this dependence throughout the various computations shows that, in case the $d = 2$ for example,

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) \leqslant C^p\Big(\frac{p}{\log p}\Big)^p\Big(\frac{\log n}{n}\Big)^{p/2}.$$

Therefore, by (4), for an independent copy $(Y_1,\dots,Y_n)$ of the sample $(X_1,\dots,X_n)$,

$$\mathbb{E}\bigg(\inf\frac{1}{n}\sum_{i=1}^n |X_i - Y_{\sigma(i)}|^p\bigg) \leqslant C^p\Big(\frac{p}{\log p}\Big)^p\Big(\frac{\log n}{n}\Big)^{p/2}$$

where we recall that σ runs over all permutations of $\{1,\dots,n\}$. Hence

$$\mathbb{E}\Big(\inf\max_{1\leqslant i\leqslant n}|X_i - Y_{\sigma(i)}|\Big) \leqslant n^{1/p}C\,\frac{p}{\log p}\Big(\frac{\log n}{n}\Big)^{1/2}.$$

Optimizing in $p \geqslant 2$ depending on n ($p \sim \log n$) shows that

$$\mathbb{E}\Big(\inf\max_{1\leqslant i\leqslant n}|X_i - Y_{\sigma(i)}|\Big) \leqslant C\,\frac{(\log n)^{3/2}}{\sqrt{n}}$$

which gets close to the Leighton–Shor theorem [25] (stating the result with $(\log n)^{3/4}$).

Since we only considered, for simplicity, manifolds without boundary, some further details might be necessary in order to cover the uniform distribution on the unit cube $[0,1]^d$. We only detail the argument for $d=2$ and $p=2$. By the preceding analysis, the result holds true on the two-dimensional torus. Endowing the latter with the (equivalent) induced Euclidean metric, the issue is to take care of periodicity. That is, in the Kantorovich dual description (13)

$$\frac{1}{2}\,\mathrm{W}_2^2(\mu_n,\mu) = \sup\bigg(\sum_{i=1}^n Q_1\varphi(X_i) - \int_{[0,1]^2} \varphi\,dx\bigg),$$

the supremum is taken over (continuous, bounded) coordinate-wise periodic functions $\varphi : [0,1]^2 \to \mathbb{R}$ for the torus case while it is taken over all φ for the cube case. We use a simple symmetrization trick to operate the comparison.

Divide $[-1,1]^2$ into the four (disjoint) regions

$$\begin{aligned}
A_1 &= \big\{(x_1,x_2);\, x_1 \in [0,1], x_2 \in [0,1]\big\},\\
A_2 &= \big\{(x_1,x_2);\, x_1 \in [0,1], x_2 \in [-1,0)\big\},\\
A_3 &= \big\{(x_1,x_2);\, x_1 \in [-1,0), x_2 \in [-1,0)\big\},\\
A_4 &= \big\{(x_1,x_2);\, x_1 \in [-1,0), x_2 \in [0,1]\big\}.
\end{aligned}$$

Given $\varphi : [0,1]^2 \to \mathbb{R}$, define $\psi = \psi_\varphi$ on $[-1,1]^2$ by

$$\psi(x_1,x_2) = \begin{cases}
\varphi(x_1,x_2) & \text{if } (x_1,x_2) \in A_1,\\
\varphi(x_1,-x_2) & \text{if } (x_1,x_2) \in A_2,\\
\varphi(-x_1,-x_2) & \text{if } (x_1,x_2) \in A_3,\\
\varphi(-x_1,x_2) & \text{if } (x_1,x_2) \in A_4.
\end{cases}$$

The point is that ψ is periodic on $[-1,1]^2$ in the sense that $\psi(1,x_2) = \psi(-1,x_2)$ and $\psi(x_1,1) = \psi(x_1,-1)$ for all $x_1,x_2 \in [-1,1]$.

A simple exercise shows that for every $(x_1,x_2) \in [-1,1]^2$,

$$\begin{aligned}
&Q_1\psi(x_1,x_2)\\
&\quad = Q_1\varphi(x_1,x_2)\mathbb{1}_{\{(x_1,x_2)\in A_1\}} + Q_1\varphi(x_1,-x_2)\mathbb{1}_{\{(x_1,x_2)\in A_2\}}\\
&\qquad + Q_1\varphi(-x_1,-x_2)\mathbb{1}_{\{(x_1,x_2)\in A_3\}} + Q_1\varphi(-x_1,x_2)\mathbb{1}_{\{(x_1,x_2)\in A_4\}}.
\end{aligned}$$

Let $Z_1, \dots, Z_n$ be independent with common uniform distribution $\tilde{\mu}$ on $[-1,1]^2$. Then, by symmetry and exchangeability,

$$\mathbb{E}\bigg(\sup_{\psi_\varphi}\bigg[\frac{1}{n}\sum_{i=1}^n Q_1\psi_\varphi(Z_i) - \frac{1}{4}\int_{[-1,1]^2}\psi_\varphi\,dx\bigg]\bigg)$$
$$= \mathbb{E}\bigg(\sup_{\varphi}\bigg[\frac{1}{n}\sum_{i=1}^n Q_1\varphi(X_i) - \int_{[0,1]^2}\varphi\,dx\bigg]\bigg).$$

Therefore, if $\tilde{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$, then

$$\mathrm{W}_2(\tilde{\mu}_n, \tilde{\mu}) \geqslant \mathrm{W}_2(\mu_n, \mu).$$

The upper bounds on the torus thus transfer into bounds for the cube, justifying the claim.

§5. The Gaussian case

The aim of this section and the next ones is to investigate the optimal matching problem for Gaussian samples and to prove Theorem 1. Therefore, $X_1, \dots, X_n$ are here independent with common standard normal distribution $d\mu(x) = e^{-|x|^2/2}\frac{dx}{(2\pi)^{d/2}}$ on $\mathbb{R}^d$. Set as before $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. We investigate

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big)$$

as a function of n, $d \geqslant 1$ and $1 \leqslant p < \infty$.

Before starting the study itself, it is worthwhile mentioning that by a simple comparison, the rates in this Gaussian setting are at least the ones of the uniform case. Denote indeed by λ the uniform distribution on $[0,1]^d$, let $U_1, \dots, U_n$ be independent and distributed as λ, and set $\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{U_i}$. Now λ is the image of the standard Gaussian distribution μ on $\mathbb{R}^d$ by the map $\Phi^{\otimes d}$ where

$$\Phi(x) = \int_{-\infty}^x e^{-u^2/2}\frac{du}{\sqrt{2\pi}}, \quad x \in \mathbb{R},$$

which satisfies $\|\Phi\|_{\mathrm{Lip}} \leqslant \frac{1}{\sqrt{2\pi}} \leqslant 1$. Therefore, if $X_1, \dots, X_n$ are independent distributed as μ, then $\Phi(X_1), \dots, \Phi(X_n)$ are independent distributed as λ and, given $\varphi : \mathbb{R}^d \to \mathbb{R}$ 1-Lipschitz,

$$\frac{1}{n}\sum_{i=1}^{n}\varphi\big(\Phi^{\otimes d}(X_i)\big) - \int_{\mathbb{R}^d}\varphi\,d\lambda = \frac{1}{n}\sum_{i=1}^{n}\varphi\circ\Phi^{\otimes d}(X_i) - \int_{\mathbb{R}^d}\varphi\circ\Phi^{\otimes n}\,d\mu.$$

Since $\varphi\circ\Phi^{\otimes d}$ is 1-Lipschitz, it follows from the Kantorovich duality (12) that

$$\mathbb{E}\big(\mathrm{W}_1(\nu_n,\lambda)\big) \leqslant \mathbb{E}\big(\mathrm{W}_1(\mu_n,\mu)\big). \tag{35}$$

Since for the uniform distribution $\mathbb{E}(\mathrm{W}_p^p(\nu_n,\lambda)) \approx [\mathbb{E}(\mathrm{W}_1(\nu_n,\lambda))]^p$ for any $p \geqslant 1$ (cf. (7)), the known lower bounds on λ transfer to μ.

Turning to the upper bounds, the one-dimensional study is described by (6), while the general estimates (9) provide the optimal rate

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

when $1 < p \leqslant \frac{d}{2}$. We will actually be mostly interested in the two-dimensional setting $d = 2$ for which a satisfactory answer only holds for $p = 1$ by (8). In the spirit of the Ajtai–Komlós–Tusnády theorem, the main concern will be $p = d = 2$.

The study of $\mathbb{E}(\mathrm{W}_2^2(\mu_n,\mu))$ in this Gaussian setting will follow the transportation approach of [4] presented in Sections 2 and 3. Denote by $p_t(x,y)$, $t > 0$, $x, y \in \mathbb{R}^d$, the Mehler kernel on $\mathbb{R}^d$ defined by the integral representation

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

The family $(P_t)_{t\geqslant 0}$ defines the Ornstein-Uhlenbeck semigroup with generator $\mathrm{L} = \Delta - x\cdot\nabla$. The explicit form of $p_t(x,y)$ ensures that for $t > 0$ and $x \in \mathbb{R}^d$,

$$p_t(x,x) = \frac{1}{(1-e^{-2t})^{d/2}}\,\exp\Big(\frac{e^{-t}}{1+e^{-t}}\,|x|^2\Big).$$

Following thus the steps in Sections 2 and 3 and Proposition 3, the dispersion factor satisfies

$$\begin{aligned} D_t^2 &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|x-y|^2\,p_t(x,y)d\mu(x)d\mu(y) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\big|x - e^{-t}x + \sqrt{1-e^{-2t}}\,y|^2 d\mu(x)d\mu(y) \leqslant 2dt \end{aligned}$$

for any $0 < t \leqslant 1$. On the other hand, for any $s > 0$,

$$\int_{\mathbb{R}^d} p_s(x,x) d\mu(x) = \frac{1}{(1-e^{-s})^d}.$$

Hence, for $0 < t \leqslant 1$, the energy is controlled as

$$\int_t^\infty \int_{\mathbb{R}^d} \big[p_s(x,x) - 1\big] d\mu(x) \leqslant \int_t^\infty \frac{d\, e^{-s}}{(1-e^{-s})^d}\, ds$$
$$\leqslant \begin{cases} C \log\big(\frac{1}{t}\big) & \text{if } \ d = 1, \\ C\, \frac{1}{t^{d-1}} & \text{if } \ d \geqslant 2, \end{cases}$$

for some $C > 0$ only depending on d. Together with the dispersion estimate $D_t^2 \leqslant 2dt$, Proposition 3 yields after optimization

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) = \begin{cases} O\big(\frac{1}{n^{1/2}}\big) & \text{if } \ d = 1, \\ O\big(\frac{1}{n^{1/4}}\big) & \text{if } \ d \geqslant 2. \end{cases} \tag{36}$$

According to the known result when $d = 1$ and to the comparison with the uniform case, these bounds are not of the expected order. It will be the purpose of the next section to suitably improve upon this result.

§6. The Gaussian case revisited: localization

In order to improve upon the crude estimates (36), we make use of a standard localization argument (cf. e.g. [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

$$X_i^R = \begin{cases} X_i & \text{if } \ |X_i| \leqslant R, \\ Z_i & \text{if } \ |X_i| > R, \end{cases}$$

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

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

Therefore, after taking expectation,

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

As a bound,

$$\int\limits_{\{|x|>R\}} |x|^2 d\mu = C_d \int\limits_R^\infty r^{d+1} e^{-r^2/2}\, dr$$

is of the order of $R^{d+1}e^{-R^2/2}$ as $R \to \infty$, a natural choice is $R = c\sqrt{\log n}$ for some $c > 0$ large enough so that

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu_n^R)\big) = O\Big(\frac{1}{n^{c'}}\Big) \tag{37}$$

for some $c' > 1$.

Rather than μ_n, we now work with μ_n^R following the investigation of Section 3 with the separate control of the dispersion factor and the energy functional. The short time regularization is performed similarly. Fix $t > 0$ and set

$$f(y) = f_n^{R,t}(y) = \frac{1}{n}\sum_{i=1}^n p_t(X_i^R, y)$$

and $d\mu_n^{R,t} = f_n^{R,t} d\mu$. By convexity,

$$\begin{aligned}\mathrm{W}_2^2(\mu_n^R, \mu_n^{R,t}) &\leqslant \frac{1}{n}\sum_{i=1}^n \mathrm{W}_2^2\big(\delta_{X_i^R}, p_t(X_i^R, \cdot)d\mu\big)\\ &= \frac{1}{n}\sum_{i=1}^n \int\limits_{\mathbb{R}^d} |X_i^R - y|^2\, p_t(X_i^R, y) d\mu(y).\end{aligned}$$

After expectation $\mathbb{E} = d\mu^R$ in the iid X_i^R's,

$$\begin{aligned}\mathbb{E}\big(\mathrm{W}_2^2(\mu_n^R, \mu_n^{R,t})\big) &\leqslant \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} |x-y|^2\, p_t(x,y) d\mu^R(x) d\mu(y)\\ &\leqslant \frac{1}{m(B_R)} \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} |x-y|^2\, p_t(x,y) d\mu(x) d\mu(y)\\ &\leqslant \frac{2d(1-e^{-t})}{\mu(B_R)}.\end{aligned}$$

In the application $R \to \infty$ so that we may assume that $\mu(B_R) \geqslant \frac{1}{2}$ and thus

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n^R, \mu_n^{R,t})\big) \leqslant 4dt. \tag{38}$$

We next evaluate $\mathbb{E}(\mathrm{W}_2^2(\mu_n^{R,t}, \mu))$ from Theorem 2. Therefore

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n^{R,t}, \mu)\big) \leqslant 4\,\mathbb{E}\big(\|f_n^{R,t} - 1\|^2_{\mathrm{H}^{-1,2}(\mu)}\big) = 8 \int_0^\infty \int_{\mathbb{R}^d} \mathbb{E}\big((P_s g)^2\big) d\mu\, ds \tag{39}$$

where $g = g(y) = f_n^{R,t}(y) - 1 = \frac{1}{n} \sum_{i=1}^n [p_t(X_i^R, y) - 1]$. To integrate with respect to the X_i^R's, it is convenient to center g. To this task, write

$$g(y) = \frac{1}{n} \sum_{i=1}^n \big[p_t(X_i^R, y) - \mathbb{E}\big(p_t(X_i^R, y)\big)\big] + \mathbb{E}\big(p_t(X_1^R, y)\big) - 1 \;=\; \tilde{g}(y) + \phi(y)$$

so that

$$\mathbb{E}\big((P_s g)^2\big) = \mathbb{E}\big((P_s \tilde{g})^2\big) + (P_s \phi)^2. \tag{40}$$

Now

$$\begin{aligned}
\phi(y) = \mathbb{E}\big(p_t(X_1^R, y)\big) - 1 &= \int_{\mathbb{R}^d} p_t(x,y) d\mu^R(x) - 1 \\
&= \frac{1}{\mu(B_R)} \int_{\mathbb{R}^d} \big(\mathbb{1}_{B_R}(x) - \mu(B_R)\big) p_t(x,y) d\mu(x) \\
&= \frac{1}{\mu(B_R)}\, P_t\big(\mathbb{1}_{B_R} - \mu(B_R)\big)(y).
\end{aligned}$$

On the other hand, for each $y \in \mathbb{R}^d$,

$$\begin{aligned}
\mathbb{E}\big(P_s \tilde{g}(y)^2\big) &= \mathbb{E}\bigg(\bigg[\frac{1}{n} \sum_{i=1}^n \big[p_{t+s}(X_i^R, y) - \mathbb{E}\big(p_{t+s}(X_i^R, y)\big)\big]\bigg]^2\bigg) \\
&= \frac{1}{n}\, \mathbb{E}\Big(\big[p_{t+s}(X_1^R, y) - \mathbb{E}\big(p_{t+s}(X_1^R, y)\big)\big]^2\Big) \\
&= \frac{1}{n}\bigg[\int_{\mathbb{R}^d} p_{t+s}(x,y)^2 d\mu^R(x) - \bigg(\int_{\mathbb{R}^d} p_{t+s}(x,y) d\mu^R(x)\bigg)^2\bigg].
\end{aligned}$$

After integration in $d\mu(y)$, by the heat kernel semigroup property,

$$\begin{aligned}
&\int_{\mathbb{R}^d} \mathbb{E}\big(P_s\tilde{g}(y)^2\big)d\mu(y) \\
&\quad = \frac{1}{n}\bigg[\int_{\mathbb{R}^d} p_{2(t+s)}(x,x)d\mu^R(x) - \frac{1}{\mu(B_R)^2}\int_{\mathbb{R}^d}(P_{t+s}\mathbb{1}_{B_R})^2 d\mu\bigg] \\
&\quad = \frac{1}{n}\int_{\mathbb{R}^d}\big[p_{2(t+s)}(x,x)-1\big]d\mu^R(x) - \frac{1}{n}\int_{\mathbb{R}^d}(P_s\phi)^2 d\mu.
\end{aligned}$$

Collecting the preceding steps in (40),

$$\int_{\mathbb{R}^d}\mathbb{E}\big((P_sg)^2\big)d\mu = \frac{1}{n}\int_{\mathbb{R}^d}\big[p_{2(t+s)}(x,x)-1\big]d\mu^R(x) + \Big(1-\frac{1}{n}\Big)\int_{\mathbb{R}^d}(P_s\phi)^2 d\mu.$$

Since the Gaussian measure μ satisfies a Poincaré inequality with constant $C_P = 1$, by (24),

$$\int_{\mathbb{R}^d}(P_s\phi)^2 d\mu \leqslant e^{-2s}\int_{\mathbb{R}^d}\phi^2 d\mu \leqslant e^{-2s}\,\frac{1-\mu(B_R)}{\mu(B_R)}.$$

As above, if $R = c\sqrt{\log n}$ for some large $c > 0$, then $1-\mu(B_R) = O\big(\frac{1}{n^{c'}}\big)$ for some $c' > 1$. Summarizing this step, with $R \sim \sqrt{\log n}$,

$$\int_0^\infty \mathbb{E}\bigg(\int_{\mathbb{R}^d}(P_sg)^2\,d\mu\bigg)ds = \frac{1}{2n}\int_{2t}^\infty\int_{\mathbb{R}^d}\big[p_s(x,x)-1\big]d\mu^R(x)\,ds + O\Big(\frac{1}{n^{c'}}\Big).$$

Therefore, together with (37), (38) and (39), we have obtained at this stage that

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \leqslant C\bigg(\frac{1}{n^{c'}} + t + \frac{1}{2n}\int_{2t}^\infty\int_{\mathbb{R}^d}\big[p_s(x,x)-1\big]d\mu^R(x)\,ds\bigg) \tag{41}$$

for $R \sim \sqrt{\log n}$ and some $C > 0$, $c' > 1$ only depending on d.

In the following, we evaluate the (non-negative) energy

$$\int_{\mathbb{R}^d}\big[p_s(x,x)-1\big]d\mu^R(x) = \frac{1}{\mu(B_R)}\int_{B_R}p_s(x,x)dm(x) - 1$$

of (41) as a function of s and R. First, with $a = e^{-s} \in (0,1)$,

$$p_s(x,x) = \frac{1}{(1-a^2)^{d/2}} \exp\Big(\frac{a}{1+a}\,|x|^2\Big)$$

so that

$$\int_{B_R} p_s(x,x)d\mu(x) = \frac{1}{(1-a^2)^{d/2}\,\theta^d}\,\mu(B_{\theta R})$$

where $\theta = \sqrt{\frac{1-a}{1+a}}$. Hence

$$\int_{\mathbb{R}^d} \big[p_s(x,x) - 1\big] d\mu^R(x) = \frac{1}{(1-a^2)^{d/2}\,\theta^d}\,\frac{\mu(B_{\theta R})}{\mu(B_R)} - 1.$$

We distinguish between two cases. If $\theta R \leqslant 1$, that is $e^{-s} = a \geqslant \frac{R^2-1}{R^2+1}$, $s \leqslant \log\big(\frac{R^2-1}{R^2+1}\big) \sim \frac{2}{R^2} \to 0$,

$$\frac{1}{(1-a^2)^{d/2}\,\theta^d}\,\frac{\mu(B_{\theta R})}{m(B_R)} - 1 \leqslant \frac{C_d\,\theta^d R^d}{(1-a^2)^{d/2}\,\theta^d} \leqslant \frac{C_d R^d}{(1-e^{-s})^{d/2}} \leqslant \frac{C_d R^d}{s^{d/2}}$$

with a constant $C_d > 0$ only depending on d and possibly changing from place to place. On the other hand, for every $s > 0$,

$$\frac{1}{(1-a^2)^{d/2}\,\theta^d}\,\frac{\mu(B_{\theta R})}{\mu(B_R)} - 1 = \frac{\mu(B_{\theta R}) - (1-a)^d\,\mu(B_R)}{(1-a)^d\,\mu(B_R)}$$

and

$$\begin{aligned}\mu(B_{\theta R}) - (1-a)^d\,\mu(B_R) &= \mu(B_R)\big[1-(1-a)^d\big] - \big[\mu\big(B_R\big) - \mu\big(B_{\theta R}\big)\big]\\ &\leqslant \mu(B_R)\big[1-(1-a)^d\big]\\ &\leqslant \mu(B_R)da.\end{aligned}$$

Hence

$$\frac{1}{(1-a^2)^{d/2}\theta^{d/2}}\,\frac{\mu(B_{\theta R})}{\mu(B_R)} - 1 \leqslant \frac{2da}{(1-a)^d}\,.$$

We use the latter in the range $\theta R \geqslant 1$, that is $e^{-s} = a \leqslant \frac{R^2-1}{R^2+1} < 1$.

Combining the bounds, with $T = \log\big(\frac{R^2+1}{R^2-1}\big) \sim \frac{2}{R^2} \to 0$, but $T \gg t$,

$$\begin{aligned}
&\int_{2t}^{\infty}\int_{\mathbb{R}^d} \big[p_s(x,x) - 1\big]\,d\mu^R(x)\,ds \\
&= \int_{2t}^{T}\int_{\mathbb{R}^d} \big[p_s(x,x) - 1\big]\,d\mu^R(x)\,ds + \int_{T}^{\infty}\int_{\mathbb{R}^d} \big[p_s(x,x) - 1\big]\,d\mu^R(x)\,ds \\
&\leqslant \int_{2t}^{T} \frac{C_d R^d}{s^{d/2}}\,ds + \int_{T}^{\infty} \frac{2de^{-s}}{(1-e^{-s})^d}\,ds \\
&\leqslant \frac{C_d R^d}{t^{(d/2)-1}} + C_d R^{2(d-1)}
\end{aligned}$$

when $d \geqslant 3$, while when $d = 2$ the upper bound reads

$$C_2 R^2 \log\Big(\frac{1}{t}\Big) + C_2 R^2$$

and when $d = 1$,

$$C_1 R\sqrt{T} + C_1 \log\Big(\frac{1}{T}\Big) \,\leqslant\, C_1'(1 + \log R^2).$$

Collecting these kernel estimates in (41) with $R \sim \sqrt{\log n}$,

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \,\leqslant\, C\bigg(\frac{1}{n^{c'}} + t + \frac{1}{n}\bigg(\frac{R^d}{t^{(d/2)-1}} + R^{2(d-1)}\bigg)\bigg)$$

with the adaptations when $d = 1, 2$. Optimization in $t > 0$ then yields

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n,\mu)\big) \,=\, \begin{cases} O\big(\frac{\log\log n}{n}\big) & \text{if} \quad d = 1, \\ O\big(\frac{(\log n)^2}{n}\big) & \text{if} \quad d = 2, \\ O\big(\frac{\log n}{n^{2/d}}\big) & \text{if} \quad d \geqslant 3. \end{cases} \tag{42}$$

Together with the lower bound (35), we therefore reach Theorem 1 for $d = 2$. We note that the technology is refined enough to reach the one-dimensional case (6) but unfortunately produces an extra $\log n$ with respect to (9) when $d \geqslant 3$, and thus probably also when $d = 2$. As discussed in the introduction, it may indeed be conjectured that the correct order when $d = 2$ should be $\frac{\log n}{n}$.

§7. The Gaussian case: bounds for $\mathrm{W}_p, 1 \leqslant p < 2$

In this last section, we provide optimal rates for the optimal matching of Gaussian samples in the Kantorovich metrics W_p, $1 \leqslant p < 2$, in any dimension $d \geqslant 1$. As in the preceding sections, $X_1, \ldots, X_n$ are independent with standard normal distribution on $\mathbb{R}^d$ and $\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$.

Theorem 5. *For every* $1 \leqslant p < 2$,

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if } \ d = 1, \\ \big(\frac{\log n}{n}\big)^{p/2} & \text{if } \ d = 2, \\ \frac{1}{n^{p/d}} & \text{if } \ d \geqslant 3. \end{cases} \tag{43}$$

Proof. The lower bounds follow from the comparison (35) with the uniform distribution. Towards the upper bounds, we apply (21) to the Gaussian model with the Mehler kernel $p_t(x,y)$. First, as for $p = 2$,

$$\begin{aligned} D_t^p &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |x - y|^p \, p_t(x,y) d\mu(x) d\mu(y) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \big|x - e^{-t}x + \sqrt{1 - e^{-2t}}\, y\big|^p d\mu(x) d\mu(y) \leqslant C\, t^{p/2} \end{aligned} \tag{44}$$

for every $0 < t \leqslant 1$, where $C > 0$ only depends on d and p.

Next, for each y fixed, the random vectors $\nabla_y (-\mathrm{L}_y)^{-1} [p_s(X_i, y) - 1]$, $i = 1, \ldots, n$, are independent, centered and identically distributed. Hence

$$\begin{aligned} \mathbb{E}\bigg(\bigg|\frac{1}{n} \sum_{i=1}^{n} \nabla_y (-\mathrm{L}_y)^{-1} \big[p_t(X_i, y) - 1\big]\bigg|^2\bigg) & \\ &= \frac{1}{n}\, \mathbb{E}\Big(\big|\nabla_y (-\mathrm{L}_y)^{-1} \big[p_t(X_1, y) - 1\big]\big|^2\Big) \\ &= \frac{1}{n} \int_{\mathbb{R}^d} \big|\nabla_y (-\mathrm{L}_y)^{-1} \big[p_t(x, y) - 1\big]\big|^2 d\mu(x). \end{aligned}$$

Since $(-\mathrm{L})^{-1} = \int_0^\infty P_s ds$,

$$\nabla_y (-\mathrm{L}_y)^{-1} \big[p_t(x,y) - 1\big] = \int_0^\infty \nabla_y\, p_{t+s}(x,y) ds = \int_t^\infty \nabla_y\, p_s(x,y) ds.$$

Therefore

$$\int_{\mathbb{R}^d} \big|\nabla_y (-\mathrm{L}_y)^{-1}\big[p_t(x,y) - 1\big]\big|^2 d\mu(x) = \int_t^\infty \int_t^\infty \int_{\mathbb{R}^d} \nabla_y\, p_s(x,y) \cdot \nabla_y\, p_{s'}(x,y) d\mu(x)\, ds\, ds'.$$

It is immediate to check on the explicit expression of $p_s(x,y)$ that for each $s > 0$ and $x, y \in \mathbb{R}^d$,

$$e^s\, \nabla_y\, p_s(x,y) = -\nabla_x\, p_s(x,y) + x p_s(x,y).$$

Therefore, for each $y \in \mathbb{R}^d$,

$$\begin{aligned} e^{s+s'} \int_{\mathbb{R}^d} & \nabla_y\, p_s(x,y) \cdot \nabla_y\, p_{s'}(x,y) d\mu(x) \\ &= \int_{\mathbb{R}^d} \nabla_x\, p_s(x,y) \cdot \nabla_x\, p_{s'}(x,y) d\mu(x) + \int_{\mathbb{R}^d} |x|^2 p_s(x,y) p_{s'}(x,y) d\mu(x) \\ &\quad - \int_{\mathbb{R}^d} x \cdot \nabla_x\, p_s(x,y) p_{s'}(x,y) d\mu(x) - \int_{\mathbb{R}^d} x \cdot \nabla_x\, p_{s'}(x,y) p_s(x,y) d\mu(x) \\ &= \int_{\mathbb{R}^d} \nabla_x\, p_s(x,y) \cdot \nabla_x\, p_{s'}(x,y) d\mu(x) + d \int_{\mathbb{R}^d} p_s(x,y)\, p_{s'}(x,y) d\mu(x) \end{aligned}$$

by integration by parts. Next, by the semigroup property,

$$e^{s+s'} \int_{\mathbb{R}^d} \nabla_y\, p_s(x,y) \cdot \nabla_y p_{s'}(x,y) d\mu(x) = \int_{\mathbb{R}^d} \nabla_x p_s(x,y) \cdot \nabla_x p_{s'}(x,y) d\mu(x) + d\, p_{s+s'}(y,y).$$

Now,

$$\begin{aligned}\int_{\mathbb{R}^d} \nabla_x\, p_s(x,y)\cdot \nabla_x\, p_{s'}(x,y)d\mu(x) &= -\int_{\mathbb{R}^d} p_s(x,y)\,\mathrm{L}_x\, p_{s'}(x,y)d\mu(x)\\ &= -\int_{\mathbb{R}^d} p_s(x,y)\partial_{s'} p_{s'}(x,y)d\mu(x)\\ &= -\partial_{s'}\bigg(\int_{\mathbb{R}^d} p_s(x,y)p_{s'}(x,y)d\mu(x)\bigg)\\ &= -\,\partial_{s'}p_{s+s'}(y,y).\end{aligned}$$

Hence we have obtained that

$$\begin{aligned}&\int_{\mathbb{R}^d} \big|\nabla_y(-L_y)^{-1}\big[p_t(x,y)-1\big]\big|^2 d\mu(x)\\ &\qquad= \int_t^\infty\!\!\int_t^\infty e^{-(s+s')}\big[-\,\partial_{s'}p_{s+s'}(y,y)+d\,p_{s+s'}(y,y)\big]\,ds\,ds'.\end{aligned}$$

Finally, after integration by parts in s',

$$\begin{aligned}&\int_{\mathbb{R}^d} \big|\nabla_y(-\mathrm{L}_y)^{-1}\big[p_t(x,y)-1\big]\big|^2 d\mu(x)\\ &\quad= \int_t^\infty e^{-(t+s)}p_{t+s}(y,y)\,ds+\int_t^\infty\!\!\int_t^\infty e^{-(s+s')}(d-1)p_{s+s'}(y,y)\big]dsds'\\ &\quad= \int_{2t}^\infty e^{-s}\big[1+(d-1)(s-2t)\big]p_s(y,y)ds.\end{aligned}$$

By Jensen's inequality, we thus conclude from the preceding analysis that

$$\begin{aligned}&\int_{\mathbb{R}^d} \mathbb{E}\bigg(\bigg|\frac{1}{n}\sum_{i=1}^n \nabla_y(-\mathrm{L}_y)^{-1}\big[p_t(X_i,y)-1\big]\bigg|^p\bigg)d\mu(y)\\ &\qquad\leqslant \frac{1}{n^{p/2}}\int_{\mathbb{R}^d}\bigg(\int_{2t}^\infty e^{-s}\big[1+(d-1)(s-2t)\big]p_s(y,y)\,ds\bigg)^{p/2}d\mu(y)\end{aligned}$$

and thus, together with (21) and (44),

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) \leqslant C\bigg(t^{p/2} + \frac{1}{n^{p/2}} \int_{\mathbb{R}^d} \bigg(\int_{2t}^{\infty} e^{-s}\big[1+(d-1)(s-2t)\big]p_s(y,y)ds\bigg)^{p/2} d\mu(y)\bigg)$$

where $C > 0$ only depends on p and d. Now, recall that

$$p_s(y,y) = \frac{1}{(1-a^2)^{d/2}} \exp\Big(\frac{a}{1+a}\,|y|^2\Big)$$

where $a = e^{-s}$. Note that $\frac{a}{1+a} \leqslant \frac{1}{2}$. We examine separately small values and large values of s (recalling that $t \to 0$). First

$$\int_{2t}^{1} e^{-s}\big[1+(d-1)(s-2t)\big]p_s(y,y)ds \leqslant C_d\, e^{|y|^2/2} \int_{2t}^{1} s^{-d/2} ds.$$

On the other hand,

$$\int_{1}^{\infty} e^{-s}\big[1+(d-1)(s-2t)\big]p_s(y,y)\, ds \leqslant C_d\, e^{|y|^2/2}.$$

Hence

$$\int_{2t}^{\infty} e^{-s}\big[1+(d-1)(s-2t)\big]p_s(y,y)\, ds \leqslant C_d\Big(1 + \frac{1}{t^{(d/2)-1}}\Big)\, e^{|y|^2/2}.$$

with the last parenthesis replaced by $1 + \log(\frac{1}{t})$ when $d = 2$.

Since $p < 2$, after integration in $d\mu(y)$ in the preceding bound, we finally reach that for every $0 < t \leqslant 1$,

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,m)\big) \leqslant C_{p,d}\bigg(t^{p/2} + \frac{1}{n^{p/2}} \cdot \frac{1}{t^{(d-2)p/4}}\bigg)$$

with the last term replaced by $(\log(\frac{1}{t}))^{p/2}$ when $d = 2$. Note that $C_{p,d} \to \infty$ as $p \to 2$. Optimization in $t > 0$ then yields the various conclusions as $d = 1$, $d = 2$ or $d \geqslant 3$, establishing Theorem 5. □

Acknowledgements. I warmly thank L. Ambrosio for raising the question of the optimal matching for Gaussian samples and for numerous exchanges on the topic of this note.

References

1. M. Ajtai, J. Komlós, G. Tusnády, *On optimal matchings.* — Combinatorica **4** (1984), 259–264.
2. L. Ambrosio, N. Gigli, G. Savaré, *Bakry–Émery curvature-dimension condition and Riemannian Ricci curvature bounds.* — Ann. Probab. **43** (2015), 339–404.
3. L. Ambrosio, A. Mondino, G. Savaré, *Nonlinear diffusion equations and curvature conditions in metric measure spaces* (2015).
4. L. Ambrosio, F. Stra, D. Trevisan. *A PDE approach to a 2-dimensional matching problem* (2016). To appear in Probab. Theory Related Fields.
5. D. Bakry, *Étude des transformations de Riesz sur les variétés riemanniennes à courbure de Ricci minorée.* — Séminaire de Probabilités XXI, Lecture Notes in Math. **1247**, 137–172. Springer (1987).
6. D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators.* Grundlehren der mathematischen Wissenschaften 348. Springer (2014).
7. F. Barthe, C. Bordenave, *Combinatorial optimization over two random point sets.* — Séminaire de Probabilités XLV, Lecture Notes Math. **2078**, 483–535. Springer (2013).
8. S. Bobkov, M. Ledoux, *One-dimensional empirical measures, order statistics, and Kantorovich transport distances* (2016). To appear in Memoirs Amer. Math. Soc.
9. E. Boissard, T. Le Gouic, *On the mean speed of convergence of empirical and occupation measures in Wasserstein distance.* — Ann. Inst. Henri Poincaré Probab. Stat. **50** (2014), 539–563.
10. F. Bolley, A. Guillin, C. Villani, *Quantitative concentration inequalities for empirical measures on non-compact spaces.* — Probab. Theory Related Fields **137** (2007), 541–593.
11. J. Boutet de Monvel, O. Martin, *Almost sure convergence of the minimum bipartite matching functional in Euclidean space.* — Combinatorica **22** (2002), 523–530.
12. S. Caracciolo, C. Lucibello, G. Parisi, G. Sicuro, *Scaling hypothesis for the Euclidean bipartite matching problem.* — Physical Review E **90** (2014), 012118.
13. I. Chavel, *A modern introduction in Riemannian Geometry.* Cambridge Studies in Advanced Mathematics 98, 2nd edn. Cambridge University Press (2006).
14. E. B. Davies, *Heat kernels and spectral theory.* Cambridge Tracts in Mathematics 92. Cambridge University Press (1989).
15. S. Dereich, M. Scheutzow, R. Schottstedt, *Constructive quantization: approximation by empirical measures.* — Ann. Inst. Henri Poincaré Probab. Stat. **49** (2013), 1183–1203.
16. V. Dobrić, J. Yukich, *Asymptotics for transportation cost in high dimensions.* — J. Theoret. Probab. **8** (1995), 97–118.
17. R. Dudley, *The speed of mean Glivenko-Cantelli convergence.* — Ann. Math. Statist. **40** (1968), 40–50.

18. M. Erbar, K. Kuwada, K.-T. Sturm, *On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces.* — Invent. Math. **201** (2015), 993–1071.
19. L. Evans, *Partial Differential Equations.* Graduate Studies in Mathematics 19. American Mathematical Society (1998).
20. N. Fournier, A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure.* — Probab. Theory Related Fields **162** (2015), 707–738.
21. M. Hahn, Y. Shao, *An exposition of Talagrand's mini-course on matching theorems.* Probability in Banach spaces **8**, Progr. Probab. 30, 3–38. Birkhäuser (1992).
22. N. Holden, Y. Peres, A. Zhai, *Gravitational allocation for uniform points on the sphere* (2017).
23. W. Johnson, G. Schechtman, J. Zinn, *Best constants in moment inequalities for linear combinations of independent and exchangeable random variables.* — Ann. Probab. **13** (1985), 234–253.
24. K. Kuwada, *Duality on gradient estimates and Wasserstein controls.*— J. Funct. Anal. **258** (2010), 3758–3774.
25. T. Leighton, P. Shor, *Tight bounds for minimax grid matching with applications to the average case analysis of algorithms.* — Combinatorica **9** (1989), 161–187.
26. P.-A. Meyer, *Transformations de Riesz pour les lois gaussiennes.* — Séminaire de Probabilités XV, Lecture Notes in Math. 1059, 179–193. Springer (1984).
27. F. Otto, C. Villani, *Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality.* — J. Funct. Anal. **173** (2000), 361–400.
28. 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.
29. P. Shor, *How to pack better than Best Fit: Tight bounds for average-case on-line bin packing.* Proc. 32nd Annual Symposium on Foundations of Computer Sciences, 752–759 (1991).
30. P. Shor, J. Yukich, *Minimax grid matching and empirical measures.* — Ann. Probab. **19** (1991), 1338–1348.
31. M. Talagrand, *Matching random samples in many dimensions.* — Ann. App. Probab. **2** (1992), 846–856.
32. M. Talagrand, *The transportation cost from the uniform measure to the empirical measure in dimension* $\geqslant 3$. — Ann. Probab. **22** (1994), 919–959.
33. M. Talagrand, *Matching theorems and discrepency computations using majorising measures.* — J. Amer. Math. Soc. **17**, 455–537 (1994).
34. M. Talagrand, *Upper and Lower Bounds of Stochastic Processes.* Modern Surveys in Mathematics 60. Springer-Verlag (2014).
35. N. Varopoulos, *Hardy–Littlewood theory for semi-groups.* — J. Funct. Anal. **11** (1985), 240–260.
36. C. Villani, *Optimal Transport. Old and New.* Grundlehren der mathematischen Wissenschaften 338. Springer (2009).
37. F.-Y. Wang, *Analysis for Diffusion Processes on Riemannian Manifolds.* World Scientific (2014).

38. J. Yukich, *Some generalizations of the Euclidean two-sample matching problem.* Probability in Banach spaces 8, Progr. Probab. 30, 55–66. Birkhäuser (1992).
39. J. Yukich, *Probability theory of classical Euclidean optimization problems.* Lecture Notes in Mathematics 1675. Springer (1998).

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

Поступило 20 сентября 2017 г.