Optimal Matching of Random Samples and Rates of Convergence of Empirical Measures

Michel Ledoux

Personal Note

It is a pleasure to dedicate this contribution to Catriona Byrne, in recognition of her many years at the service of the scientific community, with dedication, professionalism, deep scientific knowledge and expertise, and cordiality. With special thoughts to the many rewarding and friendly exchanges over the years, as author and editor.

Optimal matching problems have been investigated from various viewpoints in computer science, algorithmic analysis and physics, while rates of convergence of empirical measures to their common distribution is a central topic in probability and mathematical statistics.

Perfect matching problems (on bipartite graphs), also called assignment problems, are combinatorial optimization problems classically studied within operation research and algorithmic, combinatorics, graph theory and mathematical physics (cf. e.g. [32, 33]). Classical applications to planning, allocation of resources, traveling salesman problems, expand nowadays to networks and complex systems. Linear programming relaxation within assignment and optimal transport problems also provide useful tools in machine learning and data science [34].

The random version of the matching problems addresses optimization of Euclidean additive functionals in geometric probability [37, 41] and rates of convergence of empirical measures. It opened recently fascinating challenges, which are active parts of current research. The close relationship with mass transportation in particular favored the novel use of tools from convex analysis, probability theory and partial differential equations (pde). This note describes a few of these stimulating

M. Ledoux (✉)
Institut de Mathématiques de Toulouse, Université de Toulouse – Paul-Sabatier, Toulouse, France
e-mail: ledoux@math.univ-toulouse.fr

J.-M. Morel, B. Teissier (eds.), *Mathematics Going Forward*, Lecture Notes in Mathematics 2313, https://doi.org/10.1007/978-3-031-12244-6_43

questions for the Euclidean random optimal matching problem and associated rates of convergence of empirical measures.

1 Euclidean Random Optimal Matching and Rates of Convergence of Empirical Measures

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

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

where the infimum runs over all permutations σ of $\{1, \dots, n\}$ (and $|\cdot|$ is, for example at this stage, the Euclidean distance on $\mathbb{R}^d$). That is, the task is to match the points of a sample $(x_1, \dots, x_n)$ with the ones of another sample $(y_1, \dots, y_n)$ minimizing a given cost function. The typical values of p are $p = 1$ and $p = 2$, also $p = \infty$ (with then $\inf_\sigma \max_{1 \leq i \leq n} |x_i - y_{\sigma(i)}|$).

The question may be formulated equivalently in the closely related mass transportation framework. Given $p \geq 1$, the Kantorovich distance (cf. [40] e.g.) between two probability measures ν and μ on the Borel sets of $\mathbb{R}^d$ with a finite p-th moment is defined by

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

where the infimum is taken over all couplings π on $\mathbb{R}^d \times \mathbb{R}^d$ with respective marginals ν and μ. The W_p metrics are monotone increasing with p. In the limit $p \to \infty$, $\mathrm{W}_\infty(\nu, \mu)$ may be understood as the infimum over all couplings π of

$$\mathrm{esssup}_\pi \Big\{ |x - y| \, ; (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \Big\}$$

(for measures ν and μ with bounded support).

Given samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_n)$ of points in $\mathbb{R}^d$, if $\nu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$ are the empirical measures on the respective samples, the right-hand side of (1) to the power p takes the form

$$\inf_\pi \sum_{i,j=1}^n |x_i - y_j|^p \pi_{ij}$$

where $\pi_{ij} = \pi(\{x_i, y_j\})$, $i, j = 1, \dots, n$. Since π has marginals ν and μ, for every i or j, $\sum_{i=1}^n \pi_{ij} = \sum_{j=1}^n \pi_{ij} = \frac{1}{n}$, and the set of those matrices π is convex and compact in $\mathbb{R}^{n^2}$, so that by the Birkhoff–von Neumann theorem the infimum is achieved on a permutation matrix $\pi_{ij} = \frac{1}{n} 1_{\{j=\sigma(i)\}}$. As a consequence

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

The matching problem is thus translated equivalently as a discrepancy problem between empirical measures in Kantorovich distances.

The random optimal matching problem deals with samples $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ of independent and identically distributed random variables in $\mathbb{R}^d$ (with a finite p-th moment), and a first order analysis aims at studying the order of growth in n of the averages

$$\mathbb{E}\bigg(\inf_\sigma \frac{1}{n} \sum_{i=1}^{n} |X_i - Y_{\sigma(i)}|^p \bigg). \tag{2}$$

If $X_1, \dots, X_n$ are independent random variables in $\mathbb{R}^d$ with common distribution μ, and if $\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$, $n \geq 1$, is the empirical measure on the sample, simple arguments from the triangle and Jensen's inequalities compare (2) to the average $\mathbb{E}(\mathrm{W}_p^p(\mu_n, \mu))$. The latter is then sometimes referred to as a semi-discrete matching as opposed to bipartite matching for the former. Almost surely, the sequence μ_n, $n \geq 1$, of empirical measures converges weakly to the common distribution μ, a central question of interest and study in probability and statistics. The strength of the approximation of μ by the empirical μ_n is indeed of basic importance in statistical applications, and orders of decay in various probabilistic distances have been considered. One of them is thus the Kantorovich distance that attracted a lot of attention (the convergence of $\mathrm{W}_p(\mu_n, \mu)$ to 0 is equivalent to the weak convergence of μ_n towards μ plus convergence of p-moments). By standard concentration tools, not developed here, rates on $\mathbb{E}(\mathrm{W}_p^p(\mu_n, \mu))$ may often be turned into bounds on $\mathrm{W}_p^p(\mu_n, \mu)$ with high probability.

More general probabilistic dependences in the random sample $(X_1, \dots, X_n)$ may be considered. Spectral measures of random matrices is one such instance, that gave rise to numerous recent contributions.

The exposition here is devoted to the Euclidean random optimal matching problem in the semi-discrete form and to the rate of convergence of the empirical measure to the reference measure in Kantorovich distance. As such, the discussion will be mostly focused on (lower and upper) bounds on the sequence of expectations

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \tag{3}$$

where $\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$, $n \geq 1$, and $X_1, \dots, X_n$ are independent identically distributed in $\mathbb{R}^d$ with common distribution μ with a finite p-th moment, as well as on possible exact (renormalized) limits (although the study of the limits for the bipartite matching problem does in general require further details). The note surveys some basic results in the area, and features challenging open questions of the current research on the asymptotic rates of (3). The basic parameters entering the discussion are $p \geq 1$, the distribution μ and the dimension d of the state space. Throughout the text, the notation $A \approx B$ expresses that $C^{-1}B \leq A \leq CB$ for some constant $C > 0$ independent of n, possibly depending on the dimension d and the parameters of the underlying distribution μ. Similarly $A \lesssim B$ and $A \gtrsim B$ indicate that $A \leq CB$ and $A \geq C^{-1}B$ respectively. The bibliography is not extensive, and often concentrated only on reference texts or articles.

2 The One-Dimensional Case

The one-dimensional case is of particular nature due to explicit representations of the Kantorovich metrics $\mathrm{W}_p(\nu, \mu)$ by monotone transport map of the distributions ν and μ on the Borel sets of $\mathbb{R}$. For example,

$$\mathrm{W}_1(\nu, \mu) = \int_{\mathbb{R}} |G(x) - F(x)| \mathrm{d}x$$

where $G(x) = \nu(\,]-\infty, x])$, $F(x) = \mu(\,]-\infty, x])$, $x \in \mathbb{R}$, are the distribution functions of ν and μ respectively. There are similar representation formulas for $\mathrm{W}_p(\nu, \mu)$, $p \geq 1$, in terms of the inverse distribution functions, quantiles or order statistics of empirical measures (cf. e.g. [11]).

On the basis of these explicit representations, rather precise descriptions of the rates of convergence of empirical measures in Kantorovich distances are available (cf. [11]). For example, $\mathbb{E}(\mathrm{W}_1(\mu_n, \mu))$ is typically of the order of $1/\sqrt{n}$ for large families of distributions μ, and a precise statement is that

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

if and only if $\int_{\mathbb{R}} \sqrt{F(x)(1 - F(x))}\, \mathrm{d}x < \infty$ (which is the case for instance if $\int_{\mathbb{R}} |x|^q \mathrm{d}\mu < \infty$ for some $q > 2$). The lower bound $\mathbb{E}(\mathrm{W}_1(\mu_n, \mu)) \gtrsim 1/\sqrt{n}$ holds true for any μ (with a first moment).

However, when $p > 1$ some differences occur already on basic examples emphasizing the size of the support of μ as influencing the rate. The standard rate $1/n^{p/2}$ is the rule for compactly supported laws, but in general it cannot be obtained under moment conditions only. For example, by comparison with W_1,

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

if and only if $\int_{\mathbb{R}}[F(x)(1 - F(x))/f(x)]\,\mathrm{d}x < \infty$, where f is the density of the absolutely continuous component of μ. Such a characterization, which admits a version for any $p > 1$, is of particular interest for log-concave measures for which two-sided comparison inequalities may be achieved. As an illustration, while $\mathbb{E}(\mathrm{W}_p^p(\mu_n, \mu))$ is of order $1/n^{p/2}$ for any $p \geq 1$ if μ is uniform on a compact interval, for μ the (standard) Gaussian distribution on $\mathbb{R}$,

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if } 1 \leq p < 2, \\ \frac{\log\log n}{n} & \text{if } p = 2, \\ \frac{1}{n(\log n)^{p/2}} & \text{if } p > 2. \end{cases} \tag{4}$$

While the rate is therefore the same as in the uniform case for $1 \leq p < 2$, two changes occur as $p = 2$ and $p > 2$. Further models are of interest, such as for instance the exponential distribution (see [11]).

Theoretical studies have been completed by various numerical simulations in the physics literature, covering related random assignment problems and their sharp asymptotic behaviors [15], such as for example the exact renormalized limit $\lim_{n\to\infty}(n/\log\log n)\,\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) = 1$ in the Gaussian case obtained in [10].

3 The Ajtai–Komlós–Tusnády Theorem in Dimension 2 and the Ultimate Matching Conjecture

In the bipartite formulation (2) and for $p = 1$, the famous Ajtai–Komlós–Tusnády theorem in dimension 2 expresses that

$$\mathbb{E}\bigg(\inf_\sigma \frac{1}{n}\sum_{i=1}^n |X_i - Y_{\sigma(i)}|\bigg) \approx \sqrt{\frac{\log n}{n}} \tag{5}$$

for samples of independent random variables uniformly distributed on the unit square $[0, 1]^2$. It has been established in [1] by the transportation method on dyadic decompositions and combinatorial arguments, then reproved and deepened by P. Shor [35] and M. Talagrand (cf. [39] and the references therein) via generic chaining tools. The point is that, from the Kantorovich dual representation (see [40]),

$$\mathrm{W}_1(\nu, \mu) = \sup\bigg[\int_{\mathbb{R}^d}\varphi\,\mathrm{d}\nu - \int_{\mathbb{R}^d}\varphi\,\mathrm{d}\mu\bigg]$$

where the supremum is taken over 1-Lipschitz maps $\varphi : \mathbb{R}^d \to \mathbb{R}$, and as such the study enters the framework of bounds on stochastic processes. Together with a Fourier representation of the parameter set as an ellipsoid, it allows indeed for the powerful use of majorizing measures and generic chaining methods (for which a complete account is the monograph [39]). The methodology covers similarly the values of $p \geq 1$, with order of growth $(\log n/n)^{p/2}$, but the limiting case $p = \infty$ shows an interesting logarithmic correction described by the Leighton–Shor grid-matching theorem [31]

$$\mathbb{E}\Big(\inf_{\sigma} \max_{1 \leq i \leq n} |X_i - Y_{\sigma(i)}|\Big) \approx \frac{(\log n)^{3/4}}{\sqrt{n}}. \tag{6}$$

This type of analysis furthermore led P. Shor [35] to a striking statement, improving upon the upper bound in the Ajtai–Komlós–Tusnády theorem, in which the coordinates of the variables (in $\mathbb{R}^2$, indicated by the superscripts 1 and 2) do not play the same role, namely

$$\mathbb{E}\bigg(\inf_{\sigma} \max\bigg(\frac{1}{n}\sum_{i=1}^{n}\big|X_i^1 - Y_{\sigma(i)}^1\big|, \max_{1\leq i\leq n}\big|X_i^2 - Y_{\sigma(i)}^2\big|\bigg)\bigg) \lesssim \sqrt{\frac{\log n}{n}}. \tag{7}$$

In this framework, the "ultimate matching conjecture" promoted by M. Talagrand [39] would be that, for every $\alpha_1, \alpha_2 > 0$ with $\frac{1}{\alpha_1} + \frac{1}{\alpha_2} = \frac{1}{2}$, there is a constant $C > 0$ such that

$$\mathbb{E}\bigg(\inf_{\sigma} \max_{j=1,2}\bigg(\sum_{i=1}^{n}\exp\bigg(\frac{1}{C}\sqrt{\frac{n}{\log n}}\big|X_i^j - Y_{\sigma(i)}^j\big|\bigg)^{\alpha_j}\bigg)\bigg) \leq Cn.$$

Using on the one hand that $e^{a^4} \geq a^4$, and on the other hand that $\sum_{i=1}^n e^{a_i^4} \geq \exp(\max_{1\leq i\leq n} a_i^4)$ together with Jensen's inequality, the case $\alpha_1 = \alpha_2 = 4$ would provide a neat common generalization of (the upper bounds in) the Ajtai–Komlós–Tusnády and Leighton–Shor theorems, and at the same time improve upon (7) corresponding to $\alpha_1 = 2$ and $\alpha_2 = \infty$ (with $\max_{1\leq i\leq n}|X_i^2 - Y_{\sigma(i)}^2|$ in the $j = 2$ coordinate). A partial version of the conjecture as well as a suitable formulation in dimension $d \geq 3$ are discussed in [39].

Turning back to rates of empirical measures (or semi-discrete matching), for μ the uniform distribution on the unit cube $[0, 1]^d$ of $\mathbb{R}^d$, for any $d \geq 1$ and $p \geq 1$,

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big) \approx \begin{cases} \frac{1}{n^{p/2}} & \text{if } d = 1, \\ \big(\frac{\log n}{n}\big)^{p/2} & \text{if } d = 2, \\ \frac{1}{n^{p/d}} & \text{if } d \geq 3. \end{cases} \tag{8}$$

The case $d = 2$ is thus the Ajtai–Komlós–Tusnády theorem, which, as for $d = 1$, develops an unusual rate with respect to the uniform spacing $1/n^{1/d}$ of n points in $[0, 1]^d$. However, this natural spacing is fully reflected in dimension $d \geq 3$, which makes this case easier than $d = 2$. Indeed, in dimension 2, there are irregularities at all scales in the distribution of a random sample $(X_1, \ldots, X_n)$ which combine to create the $\log n$ factor, while in higher dimension, there are still irregularities at many different scales but they cannot combine (see [39]). The complete range of parameters $p \geq 1$ and $d \geq 1$ in (8) is implicit in the paper [1] and the study [39]. See [27] for an independent proof relying on the mass transportation and pde methodology exposed in the subsequent Sect. 5. In the same vein, a simple Fourier analytic proof of the Ajtai–Komlós–Tusnády theorem is provided in [12] (see also [39]). The articles [18, 21] consider distributions with compact support and densities with respect to the Lebesgue measure uniformly bounded from below and above.

When $p = \infty$, the rates are respectively $1/\sqrt{n}$ in dimension 1, $(\log n)^{3/4}/\sqrt{n}$ in dimension 2 (the Leighton–Shor theorem (6)), and $(\log n/n)^{1/d}$ in dimension $d \geq 3$ [36] (and its extension [21]).

4 General Distributions and Higher Dimension

Beyond the uniform distribution, the corresponding results for more general distributions μ, in particular with unbounded support, gave rise to a number of contributions and open questions. The one-dimensional case is extensively discussed in [11], and already develops unusual phenomena as mentioned in Sect. 2.

The Ajtai–Komlós–Tusnády theorem (5) in dimension 2 for $p = 1$ extends to large families of distributions (see [39]). For example,

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \mu)\big) \lesssim \sqrt{\frac{\log n}{n}}$$

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

A non-trivial aspect of the Ajtai–Komlós–Tusnády theorem consists also in the lower bound $\mathbb{E}(\mathrm{W}_1(\mu_n, \mu)) \gtrsim \sqrt{\log n/n}$ (besides the proofs in [1] and [39], see [5] for a new proof by mass transportation-pde arguments). Lower bounds are usually not covered by general tools and for general distributions. Actually, for irregular laws, the decay can be faster, see among others [7, 8, 17].

When $p > 1$, and in higher dimension $d \geq 1$, the picture is more diversified. The general investigations of [13, 17, 20], based on couplings on dyadic decompositions together with a randomization argument by a Poisson variable, typically yield that, if for example $\int_{\mathbb{R}^d} |x|^q \mathrm{d}\mu < \infty$ for some $q > \frac{p}{1-\min(p/d,1/2)}$, then

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) \lesssim \frac{1}{n^{\min(p/d,1/2)}}. \tag{9}$$

(The case $p = d/2$ actually involves some extra logarithmic factor.) At this level of generality, these results are essentially optimal, and suitably extend the uniform example when $p < d/2$. With respect to the Ajtai–Komlós–Tusnády theorem, one structural aspect of the proof of the general bounds (9) (due in particular to the randomization step) is however that, for $d = 1$ or 2, these will never yield anything better than a rate of the order of $1/\sqrt{n}$, and are essentially restricted to $p < d/2$ in higher dimension.

The Gaussian model is a good test example to appreciate the potential range of decay. Let thus, in the following, μ be the standard Gaussian distribution on $\mathbb{R}^d$ with density $(2\pi)^{-d/2}\,\mathrm{e}^{-\frac{1}{2}|x|^2}$ with respect to the Lebesgue measure, and in particular moments of all orders. By a contraction argument, the Gaussian rates are always larger than the uniform ones from (8). The one-dimensional case is pictured in (4). In dimension $d = 2$, with respect to (9), it holds true that

$$\mathbb{E}\big(\mathrm{W}_p^p(\mu_n,\mu)\big) \approx \begin{cases} \big(\frac{\log n}{n}\big)^{p/2} & \text{if } 1 \le p < 2, \\[2ex] \frac{(\log n)^2}{n} & \text{if } p = 2, \end{cases} \tag{10}$$

which extends the uniform model for $1 \le p < 2$, while a specific new feature appears at $p = 2$ as a consequence of the unbounded support. The proof of the case $1 \le p < 2$ and of the upper bound for $p = 2$ in [27] is based on the mass transportation and pde approach presented in the next section together with a localization step, while the lower bound for $p = 2$ in [38] relies on the generic chaining ideas of [39] together with a scaling argument (an alternate proof using the transportation-pde method is presented in [28]).

In higher dimension $d \ge 3$, the general bounds (9) yield that $\mathbb{E}(\mathrm{W}_p^p(\mu_n,\mu)) \lesssim 1/n^{p/d}$ whenever $1 \le p < d/2$. This has been extended to $1 \le p < d$ in [30] by a specific Gaussian analysis of the associated Mehler kernel. In this range $1 \le p < d$, $d \ge 3$, the rates for the Gaussian are therefore the same as the ones for the compact uniform model.

As identified by (10) when $p = d = 2$, the case $p = d$ might be of special interest. A possible conjecture for $d \ge 3$ might be that

$$\mathbb{E}\big(\mathrm{W}_d^d(\mu_n,\mu)\big) \approx \frac{(\log n)^{d/2}}{n}.$$

This is suggested as a lower bound in the note [38] (upper bounds with extra logarithmic factors are obtained in [30]). It is certainly possible that the tools developed in [38] could lead to more conclusions, also for $p > d \ge 2$ (and for more general distributions than the Gaussian with exponential tail decay), but this is essentially open at this point. Actually, there is no clear conjecture for $p > d \geq 2$ (including $p = \infty$) at this point.

5 Mass Transportation, PDE, and Exact Limits and Asymptotic Expansions

On the basis of the Ajtai–Komlós–Tusnády theorem (8) for the uniform distribution μ on the unit cube $[0, 1]^d$ of $\mathbb{R}^d$, the question of the exact asymptotic behavior of $\mathbb{E}(\mathrm{W}_p^p(\mu_n, \mu))$ as $n \to \infty$ becomes natural. Again the one-dimensional case may be addressed rather simply, for example $\mathbb{E}(\mathrm{W}_2^2(\mu_n, \mu)) = 1/6n$ (for any n).

Things are much more challenging in higher dimension, and actually some deep structural issues are underlying the picture, in particular motivated by conjectures raised by S. Caracciolo et al. [16] in the physics literature. As a major recent development in this regard, answering one of these conjectures, the landmark contribution [5] by L. Ambrosio, F. Stra and D. Trevisan achieved the exact (renormalized) limit for the uniform measure μ on $[0, 1]^2$ for $p = 2$,

$$\lim_{n\to\infty} \frac{n}{\log n} \, \mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \, = \, \frac{1}{4\pi} \, . \tag{11}$$

The result actually applies to the uniform measure on a two-dimensional compact Riemannian manifold M of volume one (the results are invariant under rescaling of the measure), with the Euclidean distance in the definition of W_2 being replaced by the Riemannian distance. The factor $1/4\pi$ captures the common small time behavior of the trace of the Laplace operator Δ in the form of

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

where $p_t(x, y)$, $t > 0$, $x, y \in M$, is the associated heat kernel (generating the heat semigroup P_t, $t > 0$). The result has been extended in [4] to measures on a bounded connected domain in $\mathbb{R}^2$ with Lipschitz boundary, with Hölder continuous density uniformly strictly positive and bounded from above. The method of proof is based on a deep analysis combining mass transportation and pde tools following an Ansatz put forward in [16]. If $T = \nabla\psi$ is the optimal transport map between two probability densities ρ_0 and ρ_1 (on M), the associated Monge–Ampère equation $\rho_1(\nabla\psi) \det(\nabla^2\psi) = \rho_0$ is turned, via the linearization $\rho_j \approx 1$, into $\psi \approx \frac{1}{2}|x|^2 + f$ where f solves the Poisson equation $-\Delta f = \rho_1 - \rho_0$. In this way, the Kantorovich metric W_2 is approximated by an energy functional represented by a dual Sobolev norm through the observation that, whenever $g : M \to \mathbb{R}$ is smooth with $\int_M g \mathrm{d}\mu = 0$, by a Taylor expansion on $\mathrm{d}\nu_\varepsilon = (1+\varepsilon g)\mathrm{d}\mu$ as $\varepsilon \to 0$ (cf. [40])

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

where $\mathrm{H}^{-1,2}(\mu)$ is the dual Sobolev norm described by the trace

$$\|g\|^2_{\mathrm{H}^{-1,2}(\mu)} \;=\; \int_M \big|\nabla(-\Delta)^{-1} g\big|^2 \mathrm{d}\mu \;=\; \int_M g(-\Delta)^{-1} g \, \mathrm{d}\mu$$

with $(-\Delta)^{-1} = \int_0^\infty P_t \mathrm{d}t$. On the basis of this Ansatz, the proof of the limit (11) proceeds by regularization by the heat kernel and approximation by the energy functional, the leading term in (11), as well as the full rates in (8), reflecting the behaviour of the Green function (of the associated heat kernel) depending in particular on the dimension. More precise descriptions of the optimal map, rather than only the transport cost, are developed in [3], and in [23, 24] in connection with the behavior of the optimal transport map in the Lebesgue-to-Poisson problem together with a refined large-scale regularity theory for the Monge–Ampère equation. In case of the 2-dimensional sphere, a proof of the optimal matching rate is provided in [26] via gravitational allocation (the paper also describes related algorithmic questions of interest).

Still in dimension $d = 2$, the case $p \neq 2$ is completely open (and the value $p = 1$ should be of particular interest). For $p = 2$, the paper [16] (see also [9]) actually suggests moreover that for some value $\xi \in \mathbb{R}$,

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) \;=\; \frac{1}{4\pi} \, \frac{\log n}{n} + \frac{\xi}{n} + o\Big(\frac{1}{n}\Big). \tag{12}$$

Towards this conjecture, but still far from the answer, it is shown in [2] that

$$\bigg|\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) - \frac{1}{4\pi} \, \frac{\log n}{n}\bigg| \;\lesssim\; \frac{\sqrt{\log n \, \log\log n}}{n} \, .$$

A further conjecture in this framework would be that

$$n\big[\mathrm{W}_2^2(\mu_n, \mu) - \mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big)\big] \;\to\; \chi$$

in distribution where χ is some centered random variable with an explicit distribution as a quadratic form of a Gaussian free field (see [22, 29]). Under the conjecture (12), it would hold that

$$n\Big[\mathrm{W}_2^2(\mu_n, \mu) - \frac{1}{4\pi} \log n\Big] \;\to\; \xi + \chi$$

in distribution, which would be the ultimate description of the limiting behaviour of $\mathrm{W}_2^2(\mu_n, \mu)$ (provided the limiting value ξ is identified). For the matter of comparison, in dimension $d = 1$ for μ the Lebesgue measure on $[0, 1]$, $\mathbb{E}(\mathrm{W}_2^2(\mu_n, \mu)) = 1/6n$ while $n\,\mathrm{W}_2^2(\mu_n, \mu)$ converges in law to $\int_0^1 B(t)^2 \mathrm{d}t$ with B a Brownian bridge on $[0, 1]$ (in particular $\mathbb{E}\big(\int_0^1 B(t)^2 \mathrm{d}t\big) = 1/6$) [6].

The identification of the limits in dimension $d \geq 3$ seems to raise even higher difficulties. Let still μ denote the uniform measure on $[0, 1]^d$. In [25], M. Goldman and D. Trevisan showed that, for every $d \geq 3$ and $p \geq 1$, the limit

$$\lim_{n \to \infty} n^{p/d}\, \mathbb{E}\big(\mathrm{W}_p^p(\mu_n, \mu)\big)$$

exists and is strictly positive. The result actually extends previous works, basically covering $p < d/2$, in [7, 14, 17, 19] making use of subadditivity arguments on dyadic and combinatorial partitionings. The new ingredient in [25] is the coupling of subadditivity with the optimal transport and pde approach which has been successful in dimension 2. However, since the error in the smoothing by the heat kernel and the energy functional are of the same order in higher dimension $d \geq 3$, a delicate feature is that the leading term in the asymptotics of the Kantorovich rate might not be given anymore by the dual Sobolev norm (while higher orders are), so that identification of the limit is a serious task. Actually the prediction in [9, 16], for $p = 2 < d$, would be that

$$\mathbb{E}\big(\mathrm{W}_2^2(\mu_n, \mu)\big) = \frac{c_d}{n^{2/d}} + \frac{\xi}{4\pi^2}\,\frac{1}{n} + o\Big(\frac{1}{n}\Big),$$

but c_d is not clearly conjectured, while ξ should be explicitly given in terms of the Epstein function.

References

1. M. Ajtai, J. Komlós, G. Tusnády. On optimal matchings. *Combinatorica* **4**, 259–264 (1984).
2. L. Ambrosio, F. Glaudo. Finer estimates on the 2-dimensional matching problem. *J. Éc. polytech. Math.* **6**, 737–765 (2019).
3. L. Ambrosio, F. Glaudo, D. Trevisan. On the optimal map in the 2-dimensional random matching problem. *Discrete Contin. Dyn. Syst.* **39**, 7291–7308 (2019).
4. L. Ambrosio, M. Goldman, D. Trevisan. On the quadratic random matching problem in two-dimensional domains. *Preprint* (2021).
5. L. Ambrosio, F. Stra, D. Trevisan. A PDE approach to a 2-dimensional matching problem. *Probab. Theory Related Fields* **173**, 433–478 (2019).
6. E. del Barrio, E. Giné, F. Utzet. Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **1**, 131–189 (2005).
7. F. Barthe, C. Bordenave. Combinatorial optimization over two random point sets. *Séminaire de Probabilités XLV, Lecture Notes in Math.* **2078**, 483–535 (2013).

8. D. Benedetto, E. Caglioti. Euclidean random matching in $2d$ for non-constant densities. *J. Stat. Phys.* **181**, 854–869 (2020).
9. D. Benedetto, E. Caglioti, S. Caracciolo, M. D'Achille, G. Sicuro, A. Sportiello. Random assignment problems on $2d$ manifolds. *J. Stat. Physics* **183**, 1–40 (2021).
10. 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**, paper no. 12 (2020).
11. S. Bobkov and M. Ledoux. *One-dimensional empirical measures, order statistics, and Kantorovich transport distances*. Mem. Amer. Math. Soc. **261**, no. 1259 (2019).
12. S. Bobkov and M. Ledoux. A simple Fourier analytic proof of the AKT optimal matching theorem. *Ann. Appl. Probab.* **31**, 2567–2584 (2021).
13. 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**, 539–563 (2014).
14. J. Boutet de Monvel and O. Martin. Almost sure convergence of the minimum bipartite matching functional in Euclidean space. *Combinatorica* **22**, 523–530 (2002).
15. S. Caracciolo, M. D'Achille and G. Sicuro. Anomalous scaling of the optimal cost in the one-dimensional random assignment problem. *J. Stat. Phys.* **174**, 846–864 (2019).
16. S. Caracciolo, C. Lucibello, G. Parisi and G. Sicuro. Scaling hypothesis for the Euclidean bipartite matching problem. *Phys. Rev. E* **90**, 012118 (2014).
17. S. Dereich, M. Scheutzow and R. Schottstedt. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49**, 1183–1203 (2013).
18. V. Divol. A short proof on the rate of convergence of the empirical measure for the Wasserstein distance. *Preprint* (2021).
19. V. Dobrić and J. Yukich. Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* **8**, 97–118 (1995).
20. N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162**, 707–738 (2015).
21. N. García Trillos and D. Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canadian J. Math.* **67**, 1358–1383 (2015).
22. M. Goldman and M. Huesmann. A fluctuation result for the displacement in the optimal matching problem. *Ann. Probab.* **50**, 1446–1477 (2022).
23. M. Goldman, M. Huesmann and F. Otto. A large-scale regularity theory for the Monge–Ampère equation with rough data and application to the optimal matching problem. *Preprint* (2018).
24. M. Goldman, M. Huesmann and F. Otto. Quantitative linearization results for the Monge–Ampère equation. *Commun. Pure Appl. Math.* **74**, 2483–2560 (2021).
25. M. Goldman and D. Trevisan. Convergence of asymptotic costs for random Euclidean matching problems. *Probab. Math. Phys.* **2**, 341–362 (2021).
26. N. Holden, Y. Peres and A. Zhai. Gravitational allocation for uniform points on the sphere. *Ann. Probab.* **49**, 287–321 (2021).
27. M. Ledoux. On optimal matching of Gaussian samples. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 457, Veroyatnost' i Statistika.* **25**, 226–264 (2017).
28. M. Ledoux. On optimal matching of Gaussian samples II. *Preprint* (2018).
29. M. Ledoux. A fluctuation result in dual Sobolev norm for the optimal matching problem. *Preprint* (2019).
30. M. Ledoux and J.-X. Zhu. On optimal matching of Gaussian samples III. *Probab. Math. Statist.* **41**, 237–265 (2021).
31. T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica* **9**, 161–187 (1989).
32. L. Lovász and M.D. Plummer. *Matching theory.* American Mathematical Society, Chelsea Publishing (2009).
33. M. Mézard and A. Montanari. *Information, physics, and computation*. Oxford University Press (2009).
34. G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning* **11**, 355–607 (2019).

35. P. Shor. How to pack better than Best Fit: Tight bounds for average-case on-line bin packing. In: *Proc. 32nd Annual Symposium on Foundations of Computer Sciences*, pp. 752–759 (1991).
36. P. Shor and J. Yukich. Minimax grid matching and empirical measures. *Ann. Probab.* **19**, 1338–1348 (1991).
37. J.M. Steele. *Probability theory and combinatorial optimization.* SIAM (1997).
38. M. Talagrand. Scaling and non-standard matching theorems. *Comptes Rendus Acad. Sciences Paris, Mathématique* **356**, 692–695 (2018).
39. M. Talagrand. *Upper and lower bounds for stochastic processes. Decomposition theorems.* Second edition. Ergebnisse der Mathematik und ihrer Grenzgebiete **60**, Springer (2021).
40. C. Villani. *Topics in optimal transportation*. Graduate Studies in Mathematics **58**, American Mathematical Society (2003).
41. J. Yukich. *Probability theory of classical Euclidean optimization problems.* Lecture Notes in Mathematics **1675**, Springer (1998).