

Article

The Wasserstein Metric between a Discrete Probability Measure and a Continuous One

Weihua Yang , Xu Zhang * and Xia Wang

School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing 100124, China

* Correspondence: zhangxu@bjut.edu.cn

Abstract: This paper examines the Wasserstein metric between the empirical probability measure of n discrete random variables and a continuous uniform measure in the d -dimensional ball, providing an asymptotic estimation of their expectations as n approaches infinity. Furthermore, we investigate this problem within a mixed process framework, where n discrete random variables are generated by the Poisson process.

Keywords: Wasserstein metric; optimal matching; random variable; Poisson process

MSC: 60B10; 60G57

1. Introduction

Article [1] investigates the Ollivier curvature of random geometric graphs, with a key step being the estimation of Wasserstein metrics between the empirical probability measure of n discrete random variables and a continuous uniform one in a d -dimensional ball. The authors applied results from [2], which are based on the interval $[0, 1]^d$, whereas Ollivier curvature is built in balls. To address this discrepancy, we aim to refine the proof based on balls in order to enhance the robustness and accuracy of the process described in [1] and to make it suitable for our purposes. Furthermore, since [2] requires $d > 2$, we extended our upper and lower bounds estimation to include the case $d = 2$, aligning our research objectives with the broader scope of the study.

Additionally, lattice methods used in statistical mechanical approaches [3] often involve similar notations and convergence from discrete physical quantities to continuous ones, suggesting potential connections with convergences from discrete probabilities to continuous ones. For instance, consider a collection of point charges denoted as $Q_i, i = 1, \dots, n$ and their corresponding locations represented by the independent and uniformly distributed random variables $X_i \in \Omega, i = 1, \dots, n$, where Ω represents a bounded region within three-dimensional space R^3 with its volume defined as $|\Omega| = 1$. Assuming an ideal scenario in implicit solvation models for biological molecules, it can be postulated that each charge possesses the value of $\frac{1}{n}$, thereby establishing a discrete charge density expressed as $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. On the other hand, we consider a continuum charge density represented by a uniform measure μ in an ideal scenario. Thus, transitioning from discrete (i.e., point) charges to a continuum charge density can be a pathway from a discrete probability measure to a continuous one in Wasserstein metrics. Consequently, we may contemplate convergences of the corresponding discrete electrostatic energies and others in terms of Wasserstein metrics or within Wasserstein spaces.

The authors in [4] have provided estimations for the convergence rate in Wasserstein metrics of empirical measures on complex computational cases, involving numerous asymptotic calculations. Our findings are consistent with their results in corresponding scenarios. In comparison, our proof primarily relies on estimating the expectations of optimal matching problems to obtain upper bounds on the expectations of Wasserstein



Citation: Yang, W.; Zhang, X.; Wang, X. The Wasserstein Metric between a Discrete Probability Measure and a Continuous One. *Mathematics* **2024**, *12*, 2320. <https://doi.org/10.3390/math12152320>

Academic Editor: Velizar Pavlov

Received: 16 June 2024

Revised: 19 July 2024

Accepted: 23 July 2024

Published: 24 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

metrics. We chose this technique because, as mentioned in [5], (i) the definition of Wasserstein metrics makes them convenient for problems involving optimal transports, such as those arising from partial differential equations; (ii) Wasserstein metrics possess a rich duality property which is particularly useful when considering (2) (in contrast to bounded Lipschitz distances), and passing back and forth from the original to the dual definition is often technically convenient; (iii) being defined by an infimum, it is often relatively straightforward to bound Wasserstein metrics from above by constructing couplings between μ_1 and μ_2 ; and (iv) Wasserstein metrics incorporate a lot of the geometry of the space. For instance, the mapping $x \rightarrow \delta_x$ is an isometric embedding of \mathcal{X} into $P_p(\mathcal{X})$ (Wasserstein space of order p), but there are much deeper links. This partly explains why a Wasserstein space of $P_p(\mathcal{X})$ is often very well adapted to statements that combine weak convergence and geometry.

Motivated by the virtues of Wasserstein spaces and these inspirations, we aim to bridge the gap between discrete probabilities and their continuous counterparts using Wasserstein metrics as measures in our study. Recently, significant advancements have been made in the research progress concerning the rate of convergence of Wasserstein metrics. For instance, in [6], the authors investigated the precise rate of convergence of the quadratic Wasserstein metric between empirical measures and uniform distributions on $[0, 1]^2$ by employing well-known techniques from partial differential equations. Additionally, in [7], researchers explored upper bounds for the mean Wasserstein metric between two probabilities on $(-\pi, \pi]^d$, where $d \geq 1$, using Fourier transformation, and subsequently applied these findings to estimate the mean Wasserstein metric between two empirical measures under certain assumptions. Furthermore, in [8], an author examined upper bounds for the mean rate within the quadratic Wasserstein metric W_2 on a d -dimensional compact Riemannian manifold where $d \geq 2$. Notably, there are also ongoing studies focusing on higher-order (p -th order) Wasserstein metrics; however, we refrain from listing them here.

2. Preliminary Estimation

Definition 1. Let $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n$ be independent and uniformly distributed random variables in a d -dimensional ball $B(0; 1) = \{x \in \mathbb{R}^d, \|x\| \leq 1\}$, $d \geq 2$, where $\|\cdot\|$ represents the Euclidean metric in \mathbb{R}^d . The random variable $M_n^d := \inf_{\sigma} \sum_{i=1}^n \|X_i - Y_{\sigma(i)}\|$ represents the optimal matching between X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_n , with σ iterating over all permutations of $\{1, 2, \dots, n\}$.

By applying the dual principle [9,10], or referring to the proof process of Lemma 1 in [2], we have

$$M_n^d = \sup_{f \in \mathcal{L}_1} \sum_{i=1}^n (f(X_i) - f(Y_i)) = \sup_{f \in \mathcal{L}_1} \left| \sum_{i=1}^n (f(X_i) - f(Y_i)) \right|,$$

where the set of Lipschitz functions $\mathcal{L}_1 = \{f : B(0; 1) \rightarrow \mathbb{R}; |f(x) - f(y)| \leq \|x - y\|, \forall x, y \in B(0; 1), f(0) = 0\}$. It is worth noting that every Lipschitz function in \mathcal{L}_1 can be extended to one in

$$\mathcal{L} = \{f : \mathbb{R}^d \rightarrow \mathbb{R}; |f(x) - f(y)| \leq \|x - y\|, \forall x, y \in \mathbb{R}^d, f(0) = 0, \|f\|_{L^\infty} \leq 1\}. \tag{1}$$

Therefore, we have $\mathcal{L}_1 = \mathcal{L}|_{B(0,1)}$. The following Lemma 1 gives an upper and lower bound estimation for the expectation $E(M_n^d)$.

Lemma 1 (Optimal matching). For the above optimal matching problem, we have

$$1 - \frac{1}{d + 1} \leq \liminf_{n \rightarrow \infty} \frac{E(M_n^d)}{n^{1-\frac{1}{d}}} \leq \limsup_{n \rightarrow \infty} \frac{E(M_n^d)}{n^{1-\frac{1}{d}}} \leq 5 + 2d + 32\sqrt{2},$$

in dimension $d \geq 3$, and

$$\liminf_{n \rightarrow \infty} \frac{E(M_n^d)}{n^{\frac{1}{2}}} \geq \frac{2}{3}, \quad \limsup_{n \rightarrow \infty} \frac{E(M_n^d)}{n^{\frac{1}{2}} \log_2 n} \leq 2^4$$

in dimension $d = 2$.

Proof. We provide a detailed proof, referring to Equation (A1) in Appendix A and Equation (A3) in Appendix B. The method employed here is essentially based on the work of [2], with several improvements and modifications made to extend its applicability to random variables within balls.

3. Main Results and Proofs

The following results present Wasserstein metrics between empirical and uniform measures in d -dimensional balls. Generally, a Wasserstein metric between two probability measures μ_1, μ_2 is defined as follows:

Definition 2. Let μ_1 and μ_2 be Borel probability measures in a compact metric space (\mathcal{X}, ρ) and let $\Gamma(\mu_1, \mu_2)$ denote the set of all couplings of μ_1 and μ_2 , i.e.,

$$\Gamma(\mu_1, \mu_2) = \{ \mu : \mu \text{ is a probability measure in } \mathcal{X} \times \mathcal{X}, \mu(A, \mathcal{X}) = \mu_1(A), \mu(\mathcal{X}, A) = \mu_2(A), A \text{ is a measurable subset of } \mathcal{X}. \}$$

A Wasserstein metric is defined as

$$W(\mu_1, \mu_2) = \inf_{\mu \in \Gamma(\mu_1, \mu_2)} \int_{\mathcal{X}^2} \rho(x, y) d\mu(x, y).$$

By applying the duality principle (Kantorovich Dual Theorem) in Chapter 6, Remark 6.5 of [5], we can express the Wasserstein metric as

$$W(\mu_1, \mu_2) = \sup_{f \in \mathcal{L}_1(\mathcal{X})} \left(\int_{\mathcal{X}} f(x) d\mu_1(x) - \int_{\mathcal{X}} f(y) d\mu_2(y) \right), \tag{2}$$

where $\mathcal{L}_1(\mathcal{X})$ denotes the set of Lipschitz functions based on the metric of \mathcal{X} with a coefficient of 1. From the duality formula, we can further assume that any function $f \in \mathcal{L}_1(\mathcal{X})$ satisfying $f(0) = 0$.

Notice: In all subsequent discussions, we will explicitly specify that the metric being considered is a Euclidean metric. Additionally, we will adopt the notation $a_n = O(n^\alpha)$ for a sequence $\{a_n\}$, where α is a constant. This notation implies the existence of positive constants C such that $a_n \leq Cn^\alpha$ as n is large enough.

Theorem 1. Let X_1, X_2, \dots, X_n be independent and uniformly distributed random variables in a d -dimensional ball $B(0; 1)$. The empirical measure m_n^d is given by

$$m_n^d(A) = \frac{1}{n} \sum_{i=1}^n 1_A(X_i),$$

which represents the proportion of points in the sample that lie in a measurable subset A of $B(0; 1)$. μ^d denotes the uniform measure in $B(0; 1)$. As the sample size n tends to infinity, it can be shown that an expected Wasserstein metric between m_n^d and μ^d , denoted as $E[W(m_n^d, \mu^d)]$, decays at a rate

$$E[W(m_n^d, \mu^d)] = \begin{cases} O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases}$$

Proof. Now, we consider a Wasserstein metric in $B(0;1)$ with $\rho(x,y) = \|x - y\|$, and then

$$\begin{aligned} W(m_n^d, \mu^d) &= \inf_{\mu \in \Gamma(m_n^d, \mu^d)} \int_{B(0;1)^2} \rho(x,y) d\mu(x,y) \\ &= \sup_{f \in \mathcal{L}_1(B(0;1))} \left(\int_{B(0;1)} f(x) dm_n^d(x) - \int_{B(0;1)} f(y) d\mu^d(y) \right). \end{aligned}$$

Let Y_1, Y_2, \dots, Y_n be independent uniformly distributed random variables in $B(0;1)$, and then

$$\int_{B(0;1)} f(y) d\mu^d(y) = E[f(Y_i)], i = 1, \dots, n.$$

So

$$\begin{aligned} W(m_n^d, \mu^d) &= \sup_{f \in \mathcal{L}_1(B(0;1))} \left(\int_{B(0;1)} f(x) dm_n^d(x) - \int_{B(0;1)} f(y) d\mu^d(y) \right) \\ &= \frac{1}{n} \sup_{f \in \mathcal{L}_1(B(0;1))} \left(\sum_{i=1}^n (f(X_i) - E[f(Y_i)]) \right) \\ &= \frac{1}{n} \sup_{f \in \mathcal{L}_1(B(0;1))} \left(\sum_{i=1}^n E[f(X_i) - f(Y_i) | X_i] \right) \\ &= \frac{1}{n} \sup_{f \in \mathcal{L}_1(B(0;1))} E \left[\sum_{i=1}^n (f(X_i) - f(Y_i)) | \mathbf{X} \right] \\ &\leq \frac{1}{n} E \left[\left(\sup_{f \in \mathcal{L}_1(B(0;1))} \sum_{i=1}^n (f(X_i) - f(Y_i)) \right) | \mathbf{X} \right] \\ &= \frac{1}{n} E [M_n^d | \mathbf{X}], \end{aligned}$$

where $\mathbf{X} = (X_1, \dots, X_n)$, and hence from Lemma 1, it has

$$E[W(m_n^d, \mu^d)] \leq \frac{1}{n} E[M_n^d] = \begin{cases} O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases}$$

□

Next, we consider an empirical measure and a uniform measure in a ball $B(0;\delta)$.

Corollary 1. In general, let X_1, X_2, \dots, X_n be independent and random variables uniformly distributed in the d -dimensional ball $B(0;\delta)$ with radius $\delta > 0$. Consider an empirical measure and a uniform measure in $B(0,\delta)$, where $m_{n,\delta}^d$ represents the empirical measure defined as

$$m_{n,\delta}^d(A) = \frac{1}{n} \sum_{i=1}^n 1_A(X_i)$$

for a measurable subset A of $B(0,\delta)$, and μ_δ^d denotes the uniform measure in $B(0;\delta)$. Then, it follows that

$$E[W(m_{n,\delta}^d, \mu_\delta^d)] = \delta \begin{cases} O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases}$$

Proof. Consider the map $\varphi : B(0;\delta) \rightarrow B(0;1)$, defined by $\varphi(x) = \frac{1}{\delta}x$, where $\varphi^{-1}(t) = \delta t$. Thus, $m_{n,\delta}^d \circ \varphi^{-1}$ and $\mu_\delta^d \circ \varphi^{-1}$ correspond to the empirical measure and the uniform measure in $B(0;1)$, respectively, which establishes a one-to-one relationship between empirical

measures in $B(0; \delta)$ and those in $B(0; 1)$, as well as between uniform measures in $B(0; \delta)$ and those in $B(0; 1)$. In particular, we can write

$$\begin{aligned} W(m_{n,\delta}^d, \mu_\delta^d) &= \inf_{\mu \in \Gamma(m_{n,\delta}^d, \mu_\delta^d)} \int_{B(0;\delta)^2} \rho(x, y) d\mu(x, y) \\ &= \inf_{\mu \in \Gamma(m_{n,\delta}^d, \mu_\delta^d)} \int_{B(0;1)^2} \rho(\varphi^{-1}(t), \varphi^{-1}(\tau)) d\mu(\varphi^{-1}(t), \varphi^{-1}(\tau)) \\ &= \inf_{\nu \in \Gamma(m_{n,\delta}^d \circ \varphi^{-1}, \mu_\delta^d \circ \varphi^{-1})} \int_{B(0;1)^2} \delta\rho(t, \tau) d\nu(t, \tau) \\ &= \delta W^d(m_{n,\delta}^d, \mu_\delta^d). \end{aligned}$$

Therefore, from Theorem 1 we obtain

$$E[W(m_{n,\delta}^d, \mu_\delta^d)] = \begin{cases} \delta O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ \delta O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases}$$

□

We next generalize Theorem 1 to the case where the number of random variables, denoted by N , follows a Poisson distribution with a parameter $(1 + \alpha_n)n$ and is independent of these random variables. This case actually corresponds to a specific spacial Poisson process \mathcal{P} in [1] with intensity measure $(1 + \alpha_n)n \frac{V_d}{|B(0;1)|}$, which describes a spatial configuration of points in the ball $B(0; 1)$. Here, V_d denotes the volume measure in d -dimensional Euclidean space. Moreover, in [1], it is also stated that $N = |\mathcal{P}|$, representing the number of random points in $B(0; 1)$, called size, and the parameter $(1 + \alpha_n)n$ of the Poisson distribution is equivalent to $(1 + \alpha_n)n \frac{V_d(B(0;1))}{|B(0;1)|}$, derived from the corresponding Poisson point process. The notation $0 \leq \alpha_n \rightarrow 0$ represents slight perturbations of n in order to observe how a expected Wasserstein metric gradually changes as n approaches infinity.

Theorem 2. Let m_N^d denote the empirical random measure with respect to independent and uniformly distributed random variables X_1, X_2, \dots, X_N in $B(0; 1)$, defined as

$$m_N^d(A) = \frac{1}{N} \sum_{i=1}^N 1_A(X_i)$$

for a measurable subset A of $B(0; 1)$. N follows a Poisson distribution with a parameter $(1 + \alpha_n)n$, which is independent of random variables X_1, X_2, \dots, X_N . Let μ^d represent the uniform measure in the aforementioned ball. Then, it follows that

$$E[W(m_N^d, \mu^d)] = \begin{cases} O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases}$$

Proof. Since the number N follows a Poisson distribution with the mean $(1 + \alpha_n)n$ and X_1, X_2, \dots, X_N are uniformly distributed random variables in $B(0; 1)$, which are independent of N , we have

$$E[W(m_N^d, \mu^d)] = \sum_{k=1}^{\infty} E[W(m_N^d, \mu^d) | N = k] \mathbb{P}(N = k).$$

According to Theorem 1, it follows that

$$E[W(m_N^d, \mu^d) | N = k] = \begin{cases} O(k^{-\frac{1}{2}} \log_2 k), & d = 2, \\ O(k^{-\frac{1}{d}}), & d \geq 3. \end{cases}$$

On the other hand, from Lemma 1.2 in [11], one can obtain

$$\begin{aligned} & \mathbb{P}(N > (1 + \alpha_n)n + c\sqrt{(1 + \alpha_n)n \log n}) \\ & \leq e^{c\sqrt{(1+\alpha_n)n \log n} + ((1+\alpha_n)n + c\sqrt{(1+\alpha_n)n \log n}) \log \frac{(1+\alpha_n)n}{(1+\alpha_n)n + c\sqrt{(1+\alpha_n)n \log n}}} \\ & \leq e^{-\frac{c^2(1+\alpha_n)n \log n}{2((1+\alpha_n)n + c\sqrt{(1+\alpha_n)n \log n})}} = O(n^{-\frac{c^2}{3}}) \end{aligned}$$

where $c > 0$ is a constant, and

$$\begin{aligned} & \mathbb{P}(N < (1 + \alpha_n)n - c\sqrt{(1 + \alpha_n)n \log n}) \\ & \leq e^{-c\sqrt{(1+\alpha_n)n \log n} + ((1+\alpha_n)n - c\sqrt{(1+\alpha_n)n \log n}) \log \frac{(1+\alpha_n)n}{(1+\alpha_n)n - c\sqrt{(1+\alpha_n)n \log n}}} \\ & = O(n^{-\frac{c^2}{3}}). \end{aligned}$$

Let us denote

$$a_n^\pm = [(1 + \alpha_n)n \pm c\sqrt{(1 + \alpha_n)n \log n}].$$

Then, we obtain an expression for the expected value as follows:

$$\begin{aligned} & E[W(m_N^d, \mu^d)] \\ & = \sum_{k=1}^{a_n^- - 1} E[W(m_N^d, \mu^d) | N = k] \mathbb{P}(N = k) \\ & \quad + \sum_{k=a_n^-}^{a_n^+} E[W(m_N^d, \mu^d) | N = k] \mathbb{P}(N = k) \\ & \quad + \sum_{k=a_n^+ + 1}^{\infty} E[W(m_N^d, \mu^d) | N = k] \mathbb{P}(N = k) \\ & := I_1 + I_2 + I_3. \end{aligned}$$

We further estimate these three terms and find that

$$I_1 = \sum_{k=1}^{a_n^- - 1} E[W(m_N^d, \mu^d) | N = k] \mathbb{P}(N = k) \leq 2\mathbb{P}(N < a_n^-) = O(n^{-\frac{c^2}{3}})$$

and

$$I_3 = \sum_{k=a_n^+ + 1}^{\infty} E[W(m_N^d, \mu^d) | N = k] \mathbb{P}(N = k) \leq 2\mathbb{P}(N > a_n^+) = O(n^{-\frac{c^2}{3}}).$$

For term I_2 , it is bounded as follows:

$$\begin{aligned} I_2 & = \sum_{k=a_n^-}^{a_n^+} E[W(m_N^d, \mu^d) | N = k] \mathbb{P}(N = k) \\ & = \begin{cases} \sum_{k=a_n^-}^{a_n^+} O(k^{-\frac{1}{2}} \log_2 k) \frac{e^{-(1+\alpha_n)n} ((1+\alpha_n)n)^k}{k!}, & d = 2, \\ \sum_{k=a_n^-}^{a_n^+} O(k^{-\frac{1}{d}}) \frac{e^{-(1+\alpha_n)n} ((1+\alpha_n)n)^k}{k!}, & d \geq 3, \end{cases} \\ & \leq \begin{cases} O((a_n^-)^{-\frac{1}{2}} \log_2(a_n^+)), & d = 2, \\ O((a_n^-)^{-\frac{1}{d}}), & d \geq 3, \end{cases} \\ & = \begin{cases} O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases} \end{aligned}$$

Since c was arbitrary, with a suitable adjustment in constant c , we conclude that

$$E[W(m_N^d, \mu^d)] = \begin{cases} O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases}$$

□

Now, we consider a d -dimensional ball $B(x; \delta)$. The number of random variables in $B(x; \delta)$, still denoted as N , follows a Poisson distribution with a parameter $(1 + \alpha_n)n\delta^d$, and N is independent of these random variables. This actually corresponds to a spatial Poisson process \mathcal{P} with intensity measure $(1 + \alpha_n)n \frac{V_d}{|B(0;1)|}$ in $B(x; \delta)$, as discussed in [1], and the parameter of Poisson distribution is equivalent to $(1 + \alpha_n)n \frac{V_d(B(x;\delta))}{|B(0;1)|}$, derived from the corresponding Poisson point process.

Corollary 2. Let $0 \leq \alpha_n \rightarrow 0$ and $x \in R^d$. We denote by $m_{x,\delta;N}^d$ the empirical measure with respect to independent and uniformly distributed random variables X_1, X_2, \dots, X_N in $B(x; \delta)$, i.e.,

$$m_{x,\delta;N}^d(A) = \frac{1}{N} \sum_{i=1}^N 1_A(X_i),$$

for a measurable subset A of $B(x; \delta)$. N follows a Poisson distribution with a parameter $(1 + \alpha_n)n\delta^d$ and is independent of random variables X_1, X_2, \dots, X_N . Let $\mu_{x,\delta}^d$ be the uniform measure in $B(x; \delta)$. Then, we have

$$E[W(m_{x,\delta;N}^d, \mu_{x,\delta}^d)] = \begin{cases} O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases}$$

Proof. Combining the proof in Theorem 2 and Corollary 1, we first note that N follows a Poisson distribution in $B(x; \delta)$ with mean value $(1 + \alpha_n)n\delta^d$. Therefore, we can obtain

$$\begin{aligned} E[W(m_{x,\delta;N}^d, \mu_{x,\delta}^d)] &= E[\delta W^d(m_N^d, \mu^d)] \\ &= \delta E[W^d(m_N^d, \mu^d)] \\ &= \delta \begin{cases} O((n\delta^2)^{-\frac{1}{2}} \log_2(n\delta^d)), & d = 2, \\ O((n\delta^d)^{-\frac{1}{d}}), & d \geq 3. \end{cases} \\ &= \begin{cases} O(n^{-\frac{1}{2}} \log_2 n), & d = 2, \\ O(n^{-\frac{1}{d}}), & d \geq 3. \end{cases} \end{aligned}$$

□

4. Conclusions

The result in Corollary 2 can be directly applied to produce Appendix A.3 in [1]. We have successfully refined the proof based on balls, thereby enhancing the robustness and accuracy of the process described in [1]. Furthermore, our study has effectively bridged the gap between discrete probabilities and their continuous counterparts by utilizing Wasserstein metrics as approach measures. Moving forward, we aim to apply our methodology to analyze lattice problems in statistical mechanical approaches that involve similar notation and convergence from discrete physical quantities to continuous ones, such as electrostatic approach problems.

We derived the upper bound for the convergence rate of the Wasserstein distance between a uniform distribution and its empirical distribution when $d \geq 2$ using the dual method. Our result is consistent with the order of convergence rate in [4], but we provide a specific constant term. Furthermore, we extended this analysis to estimate the convergence rate of random multinomial empirical distributions towards uniform distributions, yielding similar results. However, our approach does not apply to the case when $d = 1$, and we

did not obtain a lower bound estimation for the convergence rate. In real-world scenarios, connections between discrete and continuous worlds can be established through random graphs by extending mathematical concepts from manifolds to graphs. For instance, in [1], authors generalize the Ollivier graph curvature definition to enhance its versatility and prove that the Ollivier curvature of random geometric graphs in Riemannian manifolds converges to the Ricci curvature of the manifold. Additionally, Appendix C3 in [1] also provides methods for computing Wasserstein metrics through simulations.

Author Contributions: Conceptualization, W.Y., X.Z. and X.W.; methodology, W.Y. and X.Z.; validation, W.Y., X.W. and X.Z.; formal analysis, W.Y., X.W. and X.Z.; writing—original draft preparation, W.Y.; writing—review and editing, W.Y. and X.Z.; project administration, W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a school–enterprise cooperation project: Application of hyperbolic network model in data analysis.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. The Lower Bound of $E(M_n^d)$

Since

$$M_n^d = \inf_{\sigma} \sum_{i=1}^n \|X_i - Y_{\sigma(i)}\| \geq \inf_{\sigma} \sum_{i=1}^n \min_{j \leq n} \|X_i - Y_j\| = \sum_{i=1}^n \min_{j \leq n} \|X_i - Y_j\|,$$

it follows that

$$E(M_n^d | X) \geq \sum_{i=1}^n E(\min_{j \leq n} \|X_i - Y_j\| | X) \geq n \min_{x \in B(0;1)} E(\min_{j \leq n} \|x - Y_j\|).$$

Let the set of points $B(x, t) = \{y \in R^d : \|x - y\| \leq t\}$, and then

$$\frac{|B(x, t) \cap B(0;1)|}{|B(0;1)|} \leq \min\{t^d, 1\}$$

and

$$\mathbb{P}(\min_{j \leq n} \|x - Y_j\| \geq t) \geq (1 - t^d)^n, \quad t < 1.$$

Thus, we have

$$E(\min_{j \leq n} \|x - Y_j\|) = \int_0^\infty P(\min_{j \leq n} \|x - Y_j\| \geq t) dt \geq \int_0^1 (1 - t^d)^n dt \stackrel{t=n^{-1/d}u}{=} n^{-1/d} \int_0^{n^{1/d}} (1 - u^d/n)^n du.$$

Finally, by Fatou’s lemma, one has

$$\liminf_{n \rightarrow \infty} \frac{E(M_n^d)}{n^{1-\frac{1}{d}}} \geq \int_0^\infty e^{-u^d} \geq 1 - \frac{1}{d+1}. \tag{A1}$$

Appendix B. The Upper Bound of $E(M_n^d)$

Let $r = n^{-\frac{1}{d}}$, so that $r \rightarrow 0$ as $n \rightarrow \infty$, and $|B(x, r)| = \frac{1}{n} \omega_d$, where $\omega_d = |B(0;1)|$. Define

$$u(i, j) = \begin{cases} 1, & \text{if } \|X_i - Y_j\| \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

Define

$$b(x) = \frac{|B(x, r) \cap B(0; 1)|}{|B(0; 1)|}, \tag{A2}$$

, and then we have $b(x) \leq \frac{1}{n} < 1$ if $n > 1$. First decompose $\sum_{i \leq n} (f(X_i) - f(Y_i))$, as follows:

$$\begin{aligned} & \sum_{i \leq n} (f(X_i) - f(Y_i)) \\ &= \sum_{i \leq n} f(X_i) \sum_{j \leq n} u(i, j) - \sum_{j \leq n} f(Y_j) \sum_{i \leq n} u(i, j) + \sum_{i \leq n} f(X_i) (1 - \sum_{j \leq n} u(i, j)) - \sum_{j \leq n} f(Y_j) (1 - \sum_{i \leq n} u(i, j)), \end{aligned}$$

so one has the inequality

$$\begin{aligned} & \left| \sum_{i \leq n} (f(X_i) - f(Y_i)) \right| \\ & \leq \left| \sum_{i \leq n} f(X_i) \sum_{j \leq n} u(i, j) - \sum_{j \leq n} f(Y_j) \sum_{i \leq n} u(i, j) \right| + \left| \sum_{i \leq n} f(X_i) (1 - \sum_{j \leq n} u(i, j)) - \sum_{j \leq n} f(Y_j) (1 - \sum_{i \leq n} u(i, j)) \right| \\ & =: I_1 + I_2. \end{aligned}$$

Further, we will estimate two parts of $E \sup_{f \in \mathcal{L}_1} I_1$ in Appendix B.1 and $E \sup_{f \in \mathcal{L}_1} I_2$ in Appendix B.2, respectively. Combining them yields the following bound for the expectation of M_n^d :

$$\begin{aligned} E(M_n^d) &= E \sup_{f \in \mathcal{L}_1} \left| \sum_{i \leq n} (f(X_i) - f(Y_i)) \right| \\ &\leq E \sup_{f \in \mathcal{L}_1} I_1 + E \sup_{f \in \mathcal{L}_1} I_2 \\ &\leq \begin{cases} n^{\frac{1}{2}} + 4n^{\frac{1}{2}} + 4n^{\frac{1}{2}} + 2^4 n^{\frac{1}{2}} \log_2 n, & d = 2, \\ n^{1-\frac{1}{d}} + 2dn^{1-\frac{1}{d}} + 2(2 + 16\sqrt{2})n^{1-\frac{1}{d}}, & d \geq 3, \end{cases} \tag{A3} \\ &= \begin{cases} 9n^{\frac{1}{2}} + 2^4 n^{\frac{1}{2}} \log_2 n, & d = 2, \\ (5 + 2d + 32\sqrt{2})n^{1-\frac{1}{d}}, & d \geq 3. \end{cases} \end{aligned}$$

Appendix B.1. The Estimation of $E \sup_{f \in \mathcal{L}_1} I_1$

$$\begin{aligned} I_1 &= \left| \sum_{i \leq n} f(X_i) \sum_{j \leq n} u(i, j) - \sum_{j \leq n} f(Y_j) \sum_{i \leq n} u(i, j) \right| \\ &= \left| \sum_{i \leq n} \sum_{j \leq n} u(i, j) (f(X_i) - f(Y_j)) \right| \\ &\leq r \sum_{i \leq n} \sum_{j \leq n} u(i, j). \end{aligned}$$

Since f is Lipschitz, we have

$$\begin{aligned} E \sup_{f \in \mathcal{L}_1} I_1 &\leq r \sum_{i \leq n} \sum_{j \leq n} E u(i, j) \\ &= r \sum_{i \leq n} \sum_{j \leq n} E(E(u(i, j) | X_i)) \\ &= r \sum_{i \leq n} \sum_{j \leq n} E(b(X_i)) \\ &\leq r \sum_{i \leq n} \sum_{j \leq n} \frac{1}{n} \\ &= n^{1-\frac{1}{d}}. \end{aligned}$$

It should be noted that a more optimized estimation for I_1 can be found in [2].

Appendix B.2. The Estimation of $E \sup_{f \in \mathcal{L}_1} I_2$

Decompose I_2 as follows:

$$\begin{aligned}
 I_2 &= \left| \sum_{i \leq n} f(X_i)(1 - \sum_{j \leq n} u(i, j)) - \sum_{j \leq n} f(Y_j)(1 - \sum_{i \leq n} u(i, j)) \right| \\
 &\leq \left| \sum_{i \leq n} f(X_i)(1 - nb(X_i)) - \sum_{j \leq n} f(Y_j)(1 - nb(Y_j)) \right| + \left| \sum_{i \leq n} f(X_i) \left(nb(X_i) - \sum_{j \leq n} u(i, j) \right) \right| \\
 &\quad + \left| \sum_{j \leq n} f(Y_j) \left(nb(Y_j) - \sum_{i \leq n} u(i, j) \right) \right| \\
 &=: I_{21} + I_{22} + I_{23}.
 \end{aligned}$$

We will estimate these three parts separately in the following smaller subsections to obtain

$$E \sup_{f \in \mathcal{L}_1} I_2 \leq E \sup_{f \in \mathcal{L}_1} I_{21} + E \sup_{f \in \mathcal{L}_1} I_{22} + E \sup_{f \in \mathcal{L}_1} I_{23} \leq \begin{cases} 4n^{\frac{1}{2}} + 4n^{\frac{1}{2}} + 2^4 n^{\frac{1}{2}} \log_2 n, & d = 2, \\ 2dn^{1-\frac{1}{d}} + 2(2 + 16\sqrt{2})n^{1-\frac{1}{d}}, & d \geq 3. \end{cases}$$

Appendix B.2.1. The Estimation of $E \sup_{f \in \mathcal{L}_1} I_{21}$

According to $\|f\|_{L^\infty[B(0;1)]} \leq 1$ and the value (A2) of $b(X_i)$ in $B(0; 1)$, we have

$$E \sup_{f \in \mathcal{L}_1} \left(\left| \sum_{i \leq n} f(X_i)(1 - nb(X_i)) \right| \right) \leq E \left(\sum_{i \leq n} |1 - nb(X_i)| \right) = \sum_{i \leq n} E \left(|1 - nb(X_i)| \right) \leq ndr = dn^{1-\frac{1}{d}}.$$

Consequently, we can obtain

$$\begin{aligned}
 E \sup_{f \in \mathcal{L}_1} I_{21} &= E \sup_{f \in \mathcal{L}_1} \left(\left| \sum_{i \leq n} f(X_i)(1 - nb(X_i)) - \sum_{j \leq n} f(Y_j)(1 - nb(Y_j)) \right| \right) \\
 &\leq E \sup_{f \in \mathcal{L}_1} \left| \sum_{i \leq n} f(X_i)(1 - nb(X_i)) \right| + E \sup_{f \in \mathcal{L}_1} \left| \sum_{j \leq n} f(Y_j)(1 - nb(Y_j)) \right| \\
 &\leq 2dn^{1-\frac{1}{d}}.
 \end{aligned}$$

Appendix B.2.2. The Estimations of $E \sup_{f \in \mathcal{L}_1} I_{22}$ and $E \sup_{f \in \mathcal{L}_1} I_{23}$

Estimating this part is challenging, and one may employ convolution decomposition to impose f in small areas. Consequently, the following estimation holds:

$$\begin{aligned}
 E \sup_{f \in \mathcal{L}_1} I_{22} &= E \sup_{f \in \mathcal{L}_1} I_{23} \\
 &= E \sup_{f \in \mathcal{L}_1} \left(\left| \sum_{i \leq n} f(X_i)(nb(X_i) - \sum_{j \leq n} u(i, j)) \right| \right) \leq \begin{cases} 2n^{\frac{1}{2}} + 2^3 n^{\frac{1}{2}} \log_2 n, & d = 2, \\ (2 + 16\sqrt{2})n^{1-\frac{1}{d}}, & d \geq 3. \end{cases}
 \end{aligned}$$

Initially, we assume that f represents an indicator function for a set A , where A is a measurable subset of R^d , and estimate $E \left(\left| \sum_{i \leq n} 1_A(X_i)(nb(X_i) - \sum_{j \leq n} u(i, j)) \right|^2 \right)$. Thus, we have

$$\begin{aligned}
 &E \left(\left| \sum_{i \leq n} 1_A(X_i)(nb(X_i) - \sum_{j \leq n} u(i, j)) \right|^2 \right) \\
 &= E \left(\sum_{i, i' \leq n} \sum_{j, j' \leq n} 1_A(X_i)(b(X_i) - u(i, j))1_A(X_{i'})(b(X_{i'}) - u(i', j')) \right).
 \end{aligned}$$

By considering different cases of i, i' and j, j' , we obtain the inequality

$$E\left(\left|\sum_{i \leq n} \sum_{j \leq n} 1_A(X_i)(b(X_i) - u(i, j))\right|^2\right) \leq n^2(n-1) \frac{|A|}{|B(0;1)|} \frac{1}{n^2} + n^2 \frac{|A|}{|B(0;1)|} \frac{1}{n} \leq 2n \frac{|A|}{|B(0;1)|}. \tag{A4}$$

Furthermore, let us set $h(x) = c_0 1_A(x)$. By using Formula (A4), we obtain

$$\begin{aligned} & \int_{R^d} E\left(\left|\sum_{i \leq n} \sum_{j \leq n} h(X_i - t)(b(X_i) - u(i, j))\right|^2\right) dt \\ &= c_0^2 \int_{R^d} E\left(\left|\sum_{i \leq n} \sum_{j \leq n} 1_A(X_i - t)(b(X_i) - u(i, j))\right|^2\right) dt \\ &\leq 2nc_0^2|A|. \end{aligned} \tag{A5}$$

Finally, we decompose f into the sum of some well-defined convolutions to estimate these components. Since a Lipschitz function f in $B(0;1) \subset R^d$ with $f(0) = 0$ can be extended to a Lipschitz function in the entire space R^d with $\|f\|_{L^\infty} \leq 1$, we consider the function $f \in \mathcal{L}$ defined as (1). We then decompose it as follows, $f = \sum_{l=1}^{q+1} f_l$, where $f_1 = f - f * h_1, f_2 = f * h_1 - f * h_2 * h_1, \dots, f_q = f * h_{q-1} * \dots * h_1 - f * h_q * h_{q-1} * \dots * h_1, f_{q+1} = f * h_q * \dots * h_1$,

$$h_l = \begin{cases} |B(0;1)|^{-1}(2^l r)^{-d}, & x \in B(0, 2^l r) \\ 0, & otherwise \end{cases},$$

and q denoted by $2^q r < 1 \leq 2^{q+1} r$. Therefore, we have

$$\begin{aligned} E \sup_{f \in \mathcal{L}_1} I_{22} &= E \sup_{f \in \mathcal{L}} I_{22} \\ &= E\left(\sup_{f \in \mathcal{L}} \left|\sum_{i \leq n} \sum_{l=1}^{q+1} f_l(X_i) \sum_{j \leq n} (b(X_i) - u(i, j))\right|\right) \\ &\leq \sum_{l=1}^{q+1} E\left(\sup_{f \in \mathcal{L}} \left|\sum_{i \leq n} f_l(X_i) \sum_{j \leq n} (b(X_i) - u(i, j))\right|\right). \end{aligned}$$

For the first expectation mentioned above, we have

$$\begin{aligned} & E\left(\sup_{f \in \mathcal{L}} \left|\sum_{i \leq n} f_1(X_i) \sum_{j \leq n} (b(X_i) - u(i, j))\right|\right) \\ &= E\left(\sup_{f \in \mathcal{L}} \left|\sum_{i \leq n} (f - f * h_1)(X_i) \sum_{j \leq n} (b(X_i) - u(i, j))\right|\right) \\ &\leq 2r \sum_{i \leq n} E\left(\left|\sum_{j \leq n} (b(X_i) - u(i, j))\right|\right) \\ &\leq 2nr = 2n^{1-\frac{1}{d}}, \end{aligned} \tag{A6}$$

since $\|f - f * h_1\|_{L^\infty} \leq 2r$ and $E\left(\left|\sum_{j \leq n} (b(X_i) - u(i, j))\right|\right) \leq 1$.

For the expectation about $f_l = (f - f * h_l) * h_1 * \dots * h_{l-1}$ with $2 \leq l \leq q$, since $\|f - f * h_l\|_{L^\infty} \leq 2^l r$, we have

$$\begin{aligned} & E\left(\sup_{f \in \mathcal{L}} \left| \sum_{i \leq n} f_i(X_i) \sum_{j \leq n} (b(X_i) - u(i, j)) \right|\right) \\ &= E\left(\sup_{f \in \mathcal{L}} \left| \sum_{i \leq n} (f - f * h_l) * h_1 * \dots * h_{l-1}(X_i) \sum_{j \leq n} (b(X_i) - u(i, j)) \right|\right) \tag{A7} \\ &\leq 2^l r E\left(\int_{\mathbb{R}^d} \left| \sum_{i \leq n} \sum_{j \leq n} h_{l-1}(X_i - t) (b(X_i) - u(i, j)) \right| dt\right), \end{aligned}$$

and on the other hand, we have

$$\begin{aligned} & E\left(\int_{\mathbb{R}^d} \left| \sum_{i \leq n} \sum_{j \leq n} h_{l-1}(X_i - t) (b(X_i) - u(i, j)) \right| dt\right) \\ &= \int_{\mathbb{R}^d} E\left(\left| \sum_{i \leq n} \sum_{j \leq n} h_{l-1}(X_i - t) (b(X_i) - u(i, j)) \right|\right) dt \\ &\leq \left(\int_{\mathbb{R}^d} E(|\sum_{i \leq n} \sum_{j \leq n} h_{l-1}(X_i - t) (b(X_i) - u(i, j))|^2) dt\right)^{\frac{1}{2}} \tag{A8} \\ &\times \left(\int_{\mathbb{R}^d} \int_{B(0,1)^{2n}} \mathbf{1}_{\cup_{i \leq n} (x_i - \text{supp}(h_{l-1}))}(t) \frac{1}{|B(0,1)|^{2n}} dx dy dt\right)^{\frac{1}{2}} \\ &\leq (|B(0; 1 + 2^{l-1}r)|)^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} E(|\sum_{i \leq n} \sum_{j \leq n} h_{l-1}(X_i - t) (b(X_i) - u(i, j))|^2) dt\right)^{\frac{1}{2}}. \end{aligned}$$

Hence, using (A5) and combining (A7) and (A8), we have

$$\begin{aligned} & E\left(\sup_{f \in \mathcal{L}} \left| \sum_{i \leq n} f_i(X_i) \sum_{j \leq n} (b(X_i) - u(i, j)) \right|\right) \\ &\leq 2^l r E\left(\int_{\mathbb{R}^d} \left| \sum_{i \leq n} \sum_{j \leq n} h_{l-1}(X_i - t) (b(X_i) - u(i, j)) \right| dt\right) \\ &\leq 2^l r (|B(0; 1 + 2^{l-1}r)|)^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} E(|\sum_{i \leq n} \sum_{j \leq n} h_{l-1}(X_i - t) (b(X_i) - u(i, j))|^2) dt\right)^{\frac{1}{2}} \tag{A9} \\ &\leq 2^l r (|B(0; 1 + 2^{l-1}r)|)^{\frac{1}{2}} \left(2n(|B(0; 1)|^{-1} (2^{l-1}r)^{-d})^2 |\text{supp}(h_{l-1})|\right)^{\frac{1}{2}} \\ &= 2^l r \sqrt{2n} (1 + 2^{l-1}r)^{\frac{d}{2}} (2^{l-1}r)^{-\frac{d}{2}} \\ &\leq 2^{l(1-\frac{d}{2})} 2^{d+\frac{1}{2}} n^{1-\frac{1}{d}}. \end{aligned}$$

For the last expectation about $f_{q+1} = f * h_1 * \dots * h_{q-1} * h_q$, the above argument still works. Since $\|f\|_{L^\infty} \leq 1 \leq 2^{q+1}r$, we have

$$E\left(\sup_{f \in \mathcal{L}} \left| \sum_{i \leq n} f_{q+1}(X_i) \sum_{j \leq n} (b(X_i) - u(i, j)) \right|\right) \leq 2^{(q+1)(1-\frac{d}{2})} 2^{d+\frac{1}{2}} n^{1-\frac{1}{d}}. \tag{A10}$$

Summing up these estimations, (A6), (A9) and (A10), yields

$$E \sup_{f \in \mathcal{L}_1} I_{22} \leq 2n^{1-\frac{1}{d}} + \sum_{l=2}^{q+1} 2^{l(1-\frac{d}{2})} 2^{d+\frac{1}{2}} n^{1-\frac{1}{d}}.$$

If $d \geq 3$, one has $E \sup_{f \in \mathcal{L}_1} I_{22} \leq (2 + 16\sqrt{2})n^{1-\frac{1}{d}}$. If $d = 2$, we may obtain $\frac{1}{2} \log_2 n - 1 \leq q < \frac{1}{2} \log_2 n$, and hence $E \sup_{f \in \mathcal{L}_1} I_{22} \leq 2n^{\frac{1}{2}} + 2^3 n^{\frac{1}{2}} \log_2 n$.

References

1. van der Hoorn, P.; Lippner, G.; Trugenberger, C.; Krioukov, D. Ollivier-Ricci curvature convergence in random geometric graphs. *Phys. Rev. Res.* **2021**, *3*, 013211. [\[CrossRef\]](#)
2. Talagrand, M. Matching random samples in many dimensions. *Ann. Appl. Probab.* **1992**, *2*, 846–856. [\[CrossRef\]](#)
3. Kralj-Iglič, V.; Iglič, A. A Simple Statistical Mechanical Approach to the free Energy of the Electric Double Layer Including the Excluded Volume Effect. *J. Phys. II* **1996**, *6*, 477–491. [\[CrossRef\]](#)
4. Fournier, N.; Guillin, A. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields* **2015**, *162*, 707–738. [\[CrossRef\]](#)
5. Villani, C. *Optimal Transport: Old and New Part 1*; Grundlehren der Mathematischen Wissenschaften, 338. A Series of Comprehensive Studies in Mathematics; Springer: Berlin/Heidelberg, Germany, 2009.
6. Ambrosio, L.; Stra, F.; Trevisan, D. A PDE approach to a 2-dimensional matching problem. *Probab. Theory Relat. Fields* **2019**, *173*, 433–477. [\[CrossRef\]](#)
7. Bobkov, S.G.; Ledoux, M. A simple Fourier analytic proof of the AKT optimal matching theorem. *Ann. Appl. Probab.* **2021**, *31*, 2567–2584. [\[CrossRef\]](#)
8. Borda, B. Empirical measures and random walks on compact spaces in the quadratic Wasserstein metric. *Ann. Inst. Henri Poincaré Probab. Stat.* **2023**, *59*, 2017–2035. [\[CrossRef\]](#)
9. Kuhn, H.W. The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.* **1955**, *2*, 83–97. [\[CrossRef\]](#)
10. Papadimitriou, C.H.; Steiglitz, K. *Combinatorial Optimization, Algorithms and Complexity*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1982.
11. Penrose, M. *Random Geometric Graphs*; Oxford University Press: Oxford, UK, 2003.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.