QUANTIFYING THE NOISE SENSITIVITY OF THE WASSERSTEIN METRIC FOR IMAGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Wasserstein metrics are increasingly being used as similarity scores for images treated as discrete measures on a grid, yet their behavior under noise remains poorly understood. In this work, we consider the sensitivity of the signed Wasserstein distance with respect to pixel-wise additive noise and derive non-asymptotic upper bounds. Among other results, we prove that the error in the signed 2-Wasserstein distance scales with the square root of the noise standard deviation, whereas the Euclidean norm scales linearly. We present experiments that support our theoretical findings and point to a peculiar phenomenon where increasing the level of noise can decrease the Wasserstein distance. A case study on cryo-electron microscopy images demonstrates that the Wasserstein metric can preserve the geometric structure even when the Euclidean metric fails to do so.

1 INTRODUCTION

Optimal Transport (OT) provides a principled way to measure the distance between probability measures, capturing not only pointwise differences but also the underlying geometry of the data. Recent advances in computational approximation methods (Cuturi, 2013; Schmitzer, 2019) contributed greatly to the rising popularity of optimal transport across many domains, such as computer vision (Feydy et al., 2021), domain adaptation (Courty et al., 2017), and others. In imaging applications, the Wasserstein metric can be used to measure similarity by treating images as discrete measures on a grid, and assigning a point mass to every pixel, proportional to its value. One field where this approach is gaining popularity is in single-particle cryo-electron microscopy (cryo-EM), a domain characterized by extremely high noise levels, where OT-based methods have been successfully applied to fundamental tasks, including the alignment of 3D density maps (Riahi et al., 2022; Singer & Yang, 2024), the clustering 2D tomographic projections (Rao et al., 2020), and the rotational alignment of tomographic projections with heterogeneity (Shi et al., 2025). We believe that a major driver for this adoption is that, empirically, the Wasserstein metric appears more robust to noise than the standard Euclidean norm.

In machine learning, OT-based metrics have inspired approaches that exhibit improved stability in training deep neural networks, most notably Wasserstein generative adversarial networks (WGAN) (Arjovsky et al., 2017) and Wasserstein autoencoders (WAE) (Tolstikhin et al., 2017). In these methods, the exact Wasserstein distance is not calculated, and is replaced by approximations, such as adversarially learned critics in WGANs or penalized divergences in WAEs, that serve as proxies for the true metric. The suggested explanation for their success is that a noise-robust metric creates a smoother loss landscape, which prevents a model from overfitting to irrelevant perturbations in the training data (Li et al., 2018). Despite these empirical successes, the theoretical understanding of the noise sensitivity of OT-based metrics remains limited. This paper aims to be a first step to bridge that gap.

Our contribution. On the theoretical side, we provide quantitative bounds relating the signed Wasserstein cost (see equation 3) between noise-corrupted images and the signed Wasserstein cost between the clean images. Focusing on a Gaussian noise model with fixed mass and pixel-wise standard deviation proportional to σ, we show that the signed p-Wasserstein metric between a noise-corrupted $n \times n$ picture and its clean counterpart gives rise to an error term that scales like $(n\sigma)^{1/p}$ up to log terms, see Theorem 3. For the 1-Wasserstein distance, considering a similar noise model,

Theorem 4 establishes that the distance between two noisy pictures deviate at most by an order $\sigma n \log_2 n$ from the distance between the clean ones. Theorem 5 gives a bound for the case of two different measures and $p \geq 1$. We complement our theoretical results with simulations in Section 4, showcasing the properties of the signed Wasserstein distance in a variety of cases.

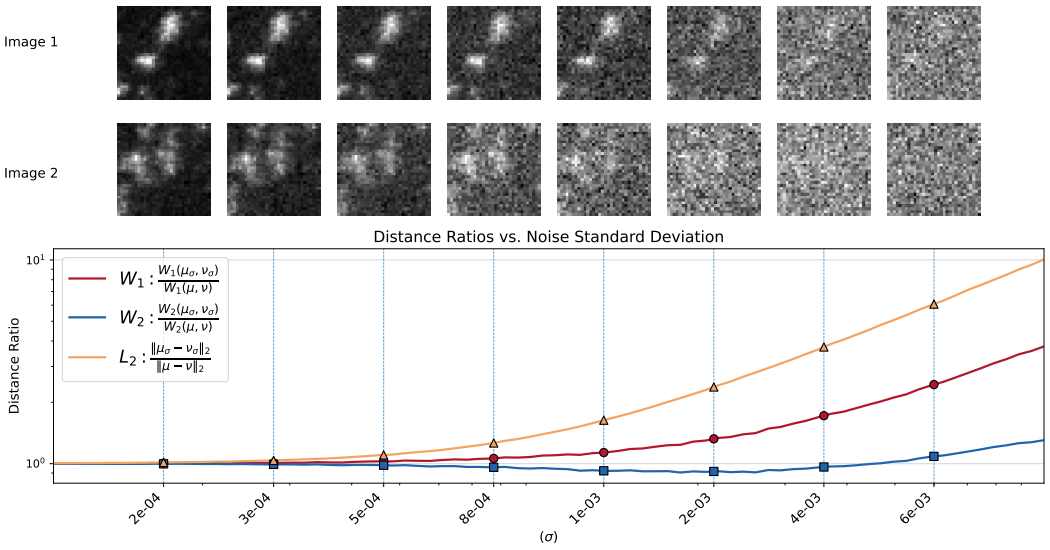

Figure 1: Distance ratios of L^2, W_1 and W_2 on a pair of noised images as a function of the noise level. L^2 diverges first, followed by W_1 and lastly, W_2 departs from the original distance between the images, exhibiting more noise robustness. Above each marker we show the pair of images that were compared using all 3 metrics. See Section 4.2 for more details.

2 Wasserstein over noised and signed measures

Wasserstein metric. Consider two probability measures $\mu, \nu \in \mathcal{P}(\mathcal{X})$. For any $p \geq 1$ and given a ground cost $\mathsf{d} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$, the Wasserstein metric between μ and ν is defined as

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} \mathsf{d}(x,y)^p \mathrm{d}\pi(x,y) \right)^{1/p}, \tag{1}$$

where $\Gamma(\mu, \nu)$ is the set of measures with respective marginals μ and ν.

Under mild conditions—see Santambrogio (2015, Theorem 1.39), the Wasserstein distance admits a dual formulation, i.e.,

$$W_p^p(\mu, \nu) = \sup_{\phi \in L^1(\mu)} \int_{\mathcal{X}} \phi(x) \mathrm{d}\mu(x) + \int_{\mathcal{X}} f^{\mathsf{d}}(y) \mathrm{d}\nu(y),$$

where $f^{\mathsf{d}}(y) := \inf_{x \in \mathcal{X}} \left(\mathsf{d}(x,y)^p - f(x) \right)$.

In the case of the 1-Wasserstein distance, the dual formulation further admits the simplified form

$$W_1(\mu, \nu) = \sup_{f \in \mathrm{Lip}_1(\mathcal{X})} \langle f, \mu - \nu \rangle, \tag{2}$$

where $\mathrm{Lip}_1(\mathcal{X})$ is the set of 1-Lipschitz functions with respect to d on $\mathcal{X}$. The dual formulations are particularly useful to study stability of optimal transport with respect to perturbations of the marginals μ and ν.

Extension to signed measures. Some image modalities (such as cryo-EM) naturally involve negative pixels, but even in modalities where all pixels are positive, once pixel-wise noise is introduced, negative pixels may appear. Since we identify pixel values with point masses, to study the effect of

pixel-wise noise we first need to explain how the Wasserstein metrics can be extended to support negative masses. The problem of generalizing OT to signed measures is not new; Mainini (2012) proposed to compute a Wasserstein-like distance between μ and ν by first constructing

$$S_{\mu,\nu} = \mu_+ + \nu_- \qquad \text{and} \qquad T_{\mu,\nu} = \nu_+ + \mu_-, \tag{3}$$

where we assume that the total mass of μ and ν are equal and μ_+ (μ_-, resp.) denotes the positive (resp. negative) part of μ. Mainini (2012) then introduced the *signed Wasserstein cost*,

$$W_p^{\pm}(\mu,\nu) := W_p(\mu_+ + \nu_-, \nu_+ + \mu_-) = W_p(S_{\mu,\nu}, T_{\mu,\nu}), \tag{4}$$

which the author denoted by $\mathbb{W}_p(\mu,\nu)$. We note that $W_p^{\pm}$ is a metric for $p = 1$ but not for $p > 1$, since it does not satisfy the triangle inequality (Mainini, 2012, Proposition 3.4). The fact that the 1-Wasserstein metric combines nicely with the positive and negative parts can be deduced from equation equation 2. The absence of triangle inequality might further lead to surprising behaviors as seen in Figure 4.

Other approaches for generalizing the Wasserstein distance to signed measures were explored by Engquist et al. (2016). Further, the usefulness of signed Wasserstein costs as defined above starts to be acknowledged in the statistics literature as the recent preprint by Groppe et al. (2025) suggests.

The issue of noise. We model images as real-valued signals on a square grid G_n of n^2 pixels which we identify with signed discrete measures. The aim of this work is to investigate how $W_p^{\pm}$ behaves when the images/measures μ and ν are corrupted. In particular, consider observing

$$\mu_\epsilon := \mu + \epsilon_\mu \quad \text{and} \quad \nu_\epsilon := \nu + \epsilon_\nu \tag{5}$$

and constructing

$$S_{\mu_\epsilon,\nu_\epsilon} := (\mu_\epsilon)_+ + (\nu_\epsilon)_- \quad \text{as well as} \quad T_{\mu_\epsilon,\nu_\epsilon} := (\nu_\epsilon)_+ + (\mu_\epsilon)_-. \tag{6}$$

Further set

$$C_S := \sum_{x\in G_n} (\mu_\epsilon)_+(x) + (\nu_\epsilon)_-(x) \quad \text{and} \quad C_T := \sum_{x\in G_n} (\nu_\epsilon)_+(x) + (\mu_\epsilon)_-(x). \tag{7}$$

Standardizing $S_{\mu_\epsilon,\nu_\epsilon}$ by C_S and $T_{\mu_\epsilon,\nu_\epsilon}$ by C_T is necessary to ensure that both measures have the same (unit) mass in the case where $\sum_{x\in G_n} \mu_\epsilon(x) \neq \sum_{x\in G_n} \nu_\epsilon(x)$. In the sequel, we will use the notation

$$\bar{S}_{\mu_\epsilon,\nu_\epsilon} := \frac{S_{\mu_\epsilon,\nu_\epsilon}}{C_S} \quad \text{and} \quad \bar{T}_{\mu_\epsilon,\nu_\epsilon} := \frac{T_{\mu_\epsilon,\nu_\epsilon}}{C_T}. \tag{8}$$

We aim at understanding the relationship between $W_p^{\pm}(\mu,\nu)$ and $W_p(\bar{S}_{\mu_\epsilon,\nu_\epsilon}, \bar{T}_{\mu_\epsilon,\nu_\epsilon})$. To put our analysis into context, consider the standard squared L^2 distance, a common metric for image comparison. In the presence of additive Gaussian noise with variance σ^2, the expected squared L^2 distance between a signal and its noisy version has a simple, direct relationship: it is exactly $n\sigma$. This metric, however, is local and insensitive to the underlying geometric structure of the signal. In contrast, the (signed) Wasserstein cost is claimed to capture this geometry, but its behavior under noise is far more complex to characterize. This paper aims to bridge that gap by providing a theoretical and empirical analysis of its robustness.

Dyadic bound on the Wasserstein distance. To get sharp estimates on the Wasserstein distance, the following proposition is particularly useful. This is Proposition 1 of Weed & Bach (2019), but on a domain with an arbitrary diameter (their formulation assumed $\operatorname{diam}(S) = 1$). The bound is based on the construction of a coupling at various scales, managing the mass imbalance in subdomains. This construction yields sharp rates in a variety of cases.

Proposition 1. *Let $\{\mathcal{Q}^k\}_{1\le k\le k^*}$ be a dyadic partition of a set S with parameter $\delta < 1$. Then, for probability measures μ and ν supported on S,*

$$W_p^p(\mu,\nu) \le \operatorname{diam}(S)^p \left(\delta^{pk^*} + \sum_{k=1}^{k^*} \delta^{p(k-1)} \sum_{Q_i^k \in \mathcal{Q}^k} |\mu(Q_i^k) - \nu(Q_i^k)| \right). \tag{9}$$

Recall that a dyadic partition of a set S with parameter $\delta < 1$ is a sequence $\{\mathcal{Q}^k\}_{1\le k\le k^*}$ possessing the following properties. First, the sets in $\mathcal{Q}^k$ form a partition of S. Further, if $Q \in \mathcal{Q}^k$, then $\operatorname{diam}(Q) \le \delta^k$. Finally, if $Q^{k+1} \in \mathcal{Q}^{k+1}$ and $Q^k \in \mathcal{Q}^k$, then either $Q^{k+1} \subset Q^k$ or $Q^{k+1} \cap Q^k = \emptyset$.

3 THEORETICAL CONTRIBUTIONS

Our main theoretical results are upper bounds in expectation on the effect that the noise has on the signed Wasserstein cost between images. To avoid boundary effects and simplify some of our analyses, we consider the pixel grid to have cyclic boundary conditions, i.e., the left–right and top–bottom edges wrap. With this choice, each pixel has the same number of neighbors.

While the Wasserstein metric naturally extends to non-probability measures, it still requires that both measures have the same mass, as described in the previous subsection. A standard i.i.d. noise model comes with the need of rescaling the pictures, which we study in the following section.

Note that all proofs of the following results are collected in Appendix A.

3.1 THE IMPACT OF RESCALING.

An important fact is that the signed Wasserstein distance, by construction, has an intricate non-linear behavior in terms of the noise when the mass of the latter is not fixed. By duality, observe that

$$W_p^p(\bar{S}_{\mu_\epsilon,\nu_\epsilon}, \bar{T}_{\mu_\epsilon,\nu_\epsilon}) = \sup_f \langle f, \bar{S}_{\mu_\epsilon,\nu_\epsilon}\rangle + \langle f^{\mathsf{d}}, \bar{T}_{\mu_\epsilon,\nu_\epsilon}\rangle \tag{10}$$

$$= \sup_f \frac{\langle f, S_{\mu_\epsilon,\nu_\epsilon}\rangle + \langle f^{\mathsf{d}}, T_{\mu_\epsilon,\nu_\epsilon}\rangle}{\sum_{x\in G_n} S_{\mu_\epsilon,\nu_\epsilon}(x)} + \left(\frac{1}{\sum_{x\in G_n} T_{\mu_\epsilon,\nu_\epsilon}(x)} - \frac{1}{\sum_{x\in G_n} S_{\mu_\epsilon,\nu_\epsilon}(x)}\right)\langle f^{\mathsf{d}}, T_{\mu_\epsilon,\nu_\epsilon}\rangle . \tag{11}$$

This decomposition shows that the optimal dual function must balance two objectives at the same time: the first one is the transport problem, and the second can be interpreted as a mass imbalance penalization.

In the case of i.i.d. Gaussian noise, the result above can be refined and yields the following theorem.

Theorem 1. *Consider two $n \times n$ images μ and ν having at least λn^2, $\lambda \in (0, 1]$ nonzero pixels. Assume that $\epsilon_\mu, \epsilon_\nu$ are $\mathcal{N}(0_{n^2}, \sigma^2 I_{n^2})$. Recall the definition of $\bar{S}_{\mu_\epsilon,\nu_\epsilon}, \bar{T}_{\mu_\epsilon,\nu_\epsilon}$ in equation 8. Then,*

$$W_1^{\pm}(\bar{S}_{\mu_\epsilon,\nu_\epsilon}, \bar{T}_{\mu_\epsilon,\nu_\epsilon}) = \frac{1}{\sum_{x\in G_n} S_{\mu_\epsilon,\nu_\epsilon}(x)} \sup_{f\in \mathrm{Lip}_1} \left\langle f, S_{\mu_\epsilon,\nu_\epsilon} - T_{\mu_\epsilon,\nu_\epsilon}\left(1 + \mathrm{O}_p\left(\frac{\sigma}{n}\right)\right)\right\rangle . \tag{12}$$

Even though the above result does not seem symmetric, we establish in the proof that

$$\sum_{x\in G_n} S_{\mu_\epsilon,\nu_\epsilon}(x) - T_{\mu_\epsilon,\nu_\epsilon}(x) = \mathrm{O}_p(\sigma n), \tag{13}$$

from which we deduce that the apparent absence of symmetry is merely an artifact of the proof.

In general, one can hope that the ratio σ/n is small, so that the result suggests that understanding the quantity $\sup_f \langle f, S_{\mu_\epsilon,\nu_\epsilon}\rangle + \langle f^{\mathsf{d}}, T_{\mu_\epsilon,\nu_\epsilon}\rangle$ under a suitable choice of noise is a first step to take towards completely characterizing the impact of the noise.

3.2 NOISE MODEL

The previous section invites us to consider a noise model for which it is not necessary to rescale the measures. To this end, we will consider slightly correlated Gaussian noise.

Assumption 1. *Consider an image modeled as an $n \times n$ grid of pixels and set $m = n^2$. Assume that the noise vector $N = (N_1, \ldots, N_m)$ is drawn from a multivariate normal distribution $\mathcal{N}(0, \Sigma)$, where the covariance matrix Σ is an $m \times m$ matrix defined as*

$$\Sigma_{ij} = \begin{cases} \sigma^2 & \text{if } i = j \\ -\frac{\sigma^2}{m-1} & \text{if } i \neq j. \end{cases} \tag{14}$$

Proposition 2 (Noise model properties)**.** *In the context of Assumption 1, the following holds.*

1. *The marginal distribution for each component is $N_i \sim \mathcal{N}(0, \sigma^2)$.*
2. *The sum of the components is zero : $\sum_{i=1}^M N_i = 0$.*

This last property is the crucial point enabling to focus on the impact of the noise, while setting aside the potential questions pertaining to rescaling the measures whose behavior was captured in Theorem 1.

3.3 Multiscale W_p bound on a single image and its noisy version

We shall begin by proving bounds in the particular case where we compare one image with a noise corrupted version of itself. We start with the case of $p = 1$.

Theorem 2. *Let $\mu : G_n \to [0, 1]$ be a probability measure on the $n \times n$ unit grid G_n with cyclic boundary conditions, and let ε satisfy Assumption 1. Then*

$$\mathbb{E}W_1^{\pm}(\mu, \mu + \varepsilon) = \mathbb{E}W_1(\varepsilon^+, \varepsilon^-) \leq \frac{2}{\sqrt{\pi}}\sigma n \log_2 n + \frac{1}{2\sqrt{\pi}}\sigma n. \tag{15}$$

It is further possible to prove a result for $p > 1$. The rates differ substantially, as is clear from the following theorem.

Theorem 3. *Let $\mu : G_n \to [0, 1]$ be a probability measure on the $n \times n$ unit grid G_n. Let ϵ be a signed measure on the grid that satisfies Assumption 1. For convenience, we again assume that n is a power-of-two. Then, for $p > 1$ with $p \in \mathbb{N}$,*

$$\mathbb{E}\left[\left(W_p^{\pm}(\mu + \epsilon, \mu)\right)^p\right] \leq \frac{4}{\sqrt{\pi}} n\sigma. \tag{16}$$

Therefore, by Jensen's inequality,

$$\mathbb{E}\left[W_p^{\pm}(\mu + \epsilon, \mu)\right] \leq \left(\frac{4}{\sqrt{\pi}} n\sigma\right)^{1/p}.$$

3.4 Multiscale W_p bound on two images and their noisy counterparts

We now consider the practically relevant setting where two different images are each corrupted by independent noise. Throughout, we assume that the noise model follows Assumption 1 and assume that both images have unit mass. Our object of interest is thus

$$W_p^{\pm}(\mu + \epsilon_\mu, \nu + \epsilon_\nu). \tag{17}$$

In the case $p = 1$, one obtains the following result.

Theorem 4. *Let $\mu, \nu : G_n \to [0, 1]$ be two probability measures on the $n \times n$ unit grid G_n with cyclic boundary conditions and let $\epsilon_\mu, \epsilon_\nu : G_n \to \mathbb{R}$ be signed noise measures that satisfy Assumption 1. For convenience we assume that n is a power-of-two. Then*

$$\mathbb{E}\left[W_1^{\pm}(\mu + \epsilon_\mu, \nu + \epsilon_\nu) - W_1^{\pm}(\mu, \nu)\right] \leq \frac{4n \log_2 n + n}{\sqrt{\pi}}\sigma + \frac{\sqrt{2}}{n}. \tag{18}$$

Even though the Wasserstein 2-distance is often used in applications and has nice theoretical properties in the continuous setting —such as the Brenier–McCann theorem (Brenier, 1991), its signed counterpart does not enjoy the same metric properties as the signed Wasserstein 1-distance, as was already hinted at in the introduction.

This absence of triangle inequality underlies the particular form of the following result.

Theorem 5. *Let $\mu, \nu : G_n \to [0, 1]$ be two probability measures on the $n \times n$ unit grid G_n with cyclic boundary conditions and let $\epsilon_\mu, \epsilon_\nu : G_n \to \mathbb{R}$ be signed noise measures that satisfy Assumption 1. For convenience we assume that n is a power-of-two. Then*

$$\mathbb{E}\big[W_p^{\pm}(\mu + \epsilon_\mu, \nu + \epsilon_\nu)\big] \leq \left(\frac{\sqrt{2}}{2}\right)^{1-\frac{1}{p}} W_1(\mu, \nu)^{\frac{1}{p}} + \frac{\sqrt{2}}{2}\left(\frac{4}{\sqrt{\pi}} n \log_2 n + \frac{2}{\sqrt{\pi}} n\right)^{\frac{1}{p}} \sigma^{\frac{1}{p}}. \tag{19}$$

4 NUMERICAL EXPERIMENTS AND RESULTS

4.1 QUANTITATIVE VALIDATION OF NOISE SCALING

The first experiment we conduct aims to quantitatively measure how the distance between an image and its noisy counterpart scales when increasing noise variance. This allows for a direct comparison between the empirical behavior of each metric and the theoretical scaling laws derived in Theorem 3. The results are reported as Figure 2. All transport costs calculated are exact and were calculated by using the POT python package Flamary et al. (2021)

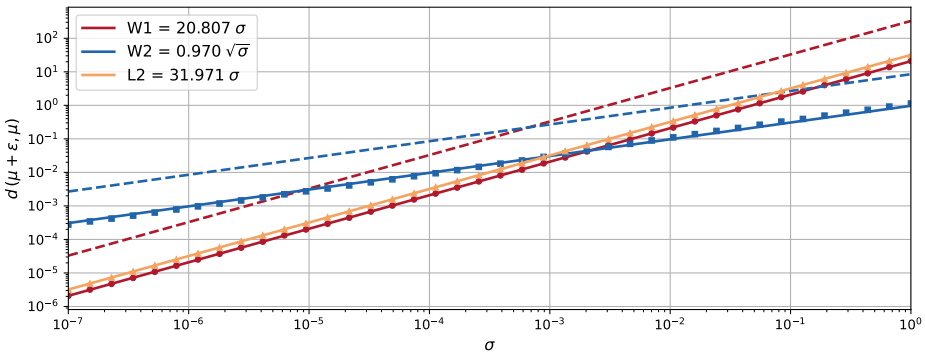

Figure 2: W_1, W_2 and L_2 (markers) distances plotted against their fits and their theoretical bounds (dashed lines)

To this end, we performed 100 independent trials, each time selecting a new, random 32x32 pixel image from the DOTMark 1.0 MicroscopyImages dataset (Schrieber et al., 2017). For each image μ, we generated a noisy version $\mu + \epsilon$ by adding zero-sum noise ϵ satisfying Assumption 1, with variances ranging from 10^{-7} to 1. We computed the difference between the original image and the noisy one for L^2, W_2 and W_1 identifying the image with the torus. This empirical result, where the W_2 distance scales with an exponent of approximately 0.5, suggests that the bound derived in Theorem 3 correctly captures the behavior of the signed 2-Wasserstein as a function of the noise variability σ.

4.2 VISUALIZING ROBUSTNESS OF INTER-IMAGE DISTANCES

We now investigate how well the different metrics preserve the original distance between two images when the latter are progressively corrupted by noise. For this experiment, we selected two distinct 32×32 pixel images from the DOTMark dataset and simultaneously corrupted them with different instances of zero-sum additive noise with a standard deviation ranging from 10^{-4} to 10^{-2}. At each noise level, we computed the W_1, W_2 and L^2 distances between the two noisy images. The results were averaged across 100 experiments. To evaluate stability, we computed a "Distance Ratio" by dividing the distance between the noisy images by the constant distance between the original, clean images. A ratio that remains close to 1 indicates that the metric's measurement reflects the underlying signals rather than the noise. The output is displayed on Figure 1, which we already exhibited in the introduction.

On that figure, the top panel visually depicts the degradation of the images as noise increases, while the main plot shows the distance ratio for each metric. The L^2 ratio (salmon-colored line) is the first to sharply diverge from 1, showing that the measured distance is quickly dominated by the noise. The W_2 ratio (blue line) is the most stable, remaining closest to the ideal ratio of 1 for the largest range of noise levels. This experiment serves as a practical illustration of the scaling laws: as the W_2 distance grows more slowly with noise, the underlying distance between the clean signals is better preserved.

Visualisation of the bound of the inter-image distance. To assess the bound established in Theorem 5, we have plotted the distance between two cryo-EM images being gradually corrupted by

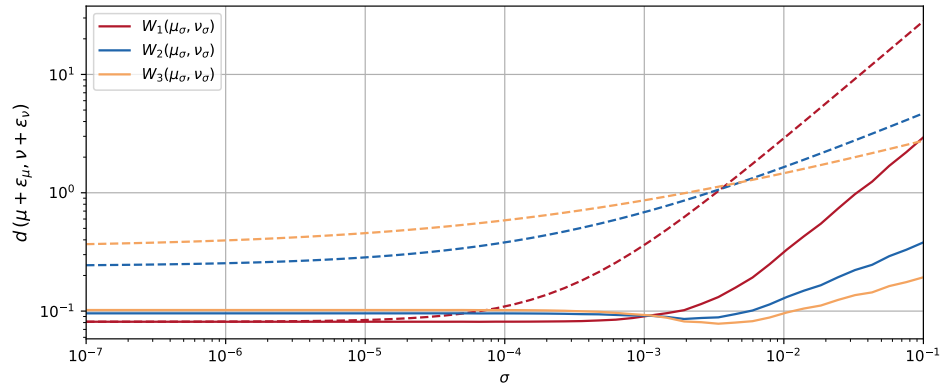

Figure 3: Distances between two randomly sampled images from the DOTMark microscopy dataset, both being noised with noise sampled from the zero-sum normal distribution, in dashed (matching colors) we have the bounds for each p from Theorem 3.

noise with the same parameter σ. We see in Figure 3 how tight the bound might be for W_1 (in the case of small noise) while it seems to not be tight for W_2 and W_3. We postulate that this is because the images used in this experiment are far from the "worst case scenario" in which where the images are very similar to each other, or very far apart. The characterization of these scenarios where the bound might be tighter comes from the analysis reported on Figure 7.

Characterizing metric behavior across image types. While W_2 is robust, its behavior is not uniform. The purpose of the next experiment is to explore how the metrics' robustness varies across different classes of images and to highlight a key nuance of the signed W_2 metric.

We repeated the distance ratio experiment from the previous section on four distinct image classes: white noise, typical cryo-EM projections, classic microscopy images, and synthetic images of two widely separated squares. These classes represent types of images which are ranging from pure noise, to having all the mass centered in a single pixel. In this experiment we aim to show how W_2 scales favorably throughout the whole range.

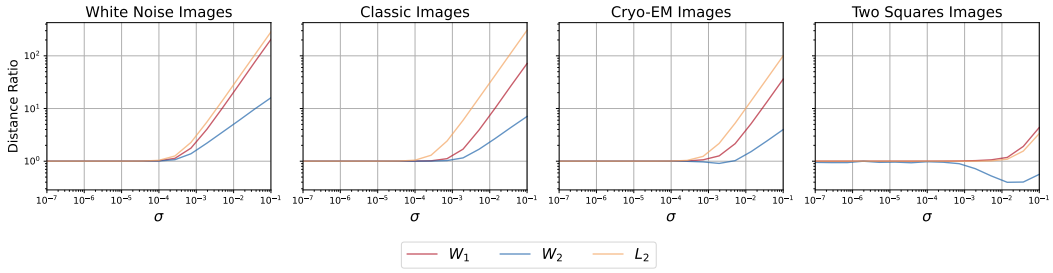

Figure 4: Ratios of the distance between the noisy images and the original images. Left: White noise images. Center left: Classic images. Center right: Images from the field of cryo-em. Right: Images of two squares which have all the mass centered in one place.

4.3 Application to Cryo-EM image alignment

For the final set of experiments, we have taken 20 different projections of the hsp90 protein in different conformational states. The image data was generated with cryojax (CryoJAX Developers, 2025), relying on the structure from the Protein Data Bank, which is based on the work of Shiau et al. (2006). The location of the protein was shifted and the pictures were corrupted with noise with $\sigma = 0.01$. Remark that the shift and very poor signal-to-noise ratio are not whimsical but accurately depict the challenges in the field of cryo-EM.

The goal is to assess how well each metric can recover known geometric relationships between particle images that undergo rotation and translation, all under a heavy-noise regime typical of real-world data.

To illustrate the difficulty of the task, Figure 5 shows a sample of the original, clean images alongside their noisy counterparts. The noise level is high enough to make the underlying molecule almost completely unrecognizable by eye, simulating the low signal-to-noise ratio of raw micrographs.

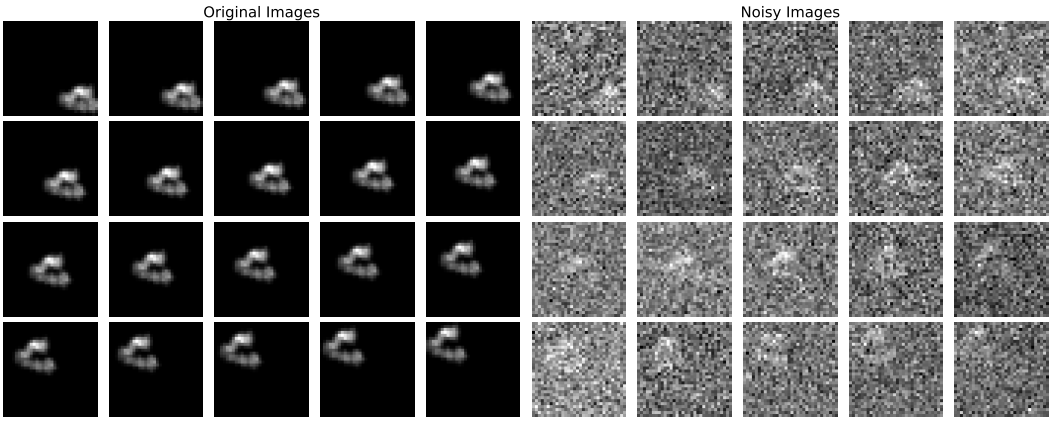

Figure 5: Original images of the molecule on the left, noised images with $\sigma = 10^{-2}$ of the molecule on the right.

For each metric, we compute all the pairwise distances, resulting in a 20x20 matrix which we can see in Figure 6. The top row shows the ground-truth distance matrices from the clean images, reflecting the structured of the transformations. The bottom row shows the matrices computed from their noisy counterparts. Under heavy noise, the L^2 distance matrix degrades into a random pattern, losing the original geometric structure. In contrast, the W_2 distance matrix preserves the global diagonal structure of the ground-truth matrix. This result is also described by our theoretical results.

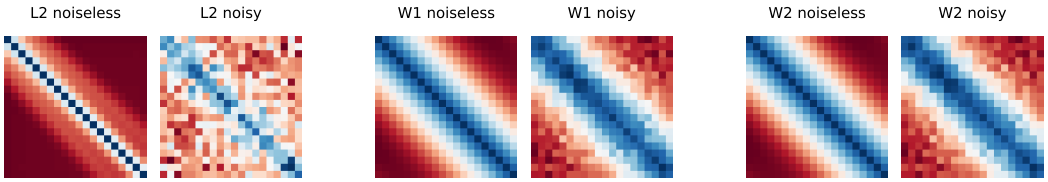

Figure 6: Top panel: Distance matrices between the different structures in the absence of noise. Bottom panel: averaged distances over 100 experiments. A more gradual blue-to-red gradient is better, W_2 performs best.

4.4 The decreasing distance phenomenon

Interestingly the estimated distance between images can even decrease when the noise increases. One can see an example in Figure 7 where for $p > 1$ we get a "dip" in the distance, showing that the images are getting closer together, similarly to the "two square images" in Figure 4.

This phenomenon, which at first sight might be surprising, can be explained by the fact that for sparse pictures, the noise appearing between two structures can be used to "bridge" the transport distance between them, like we see in Figure 8. Instead of having to transport the mass far away, a large part of it is mapped to surrounding noise, this noise is matched with noise a bit further and so on until all the mass is matched.

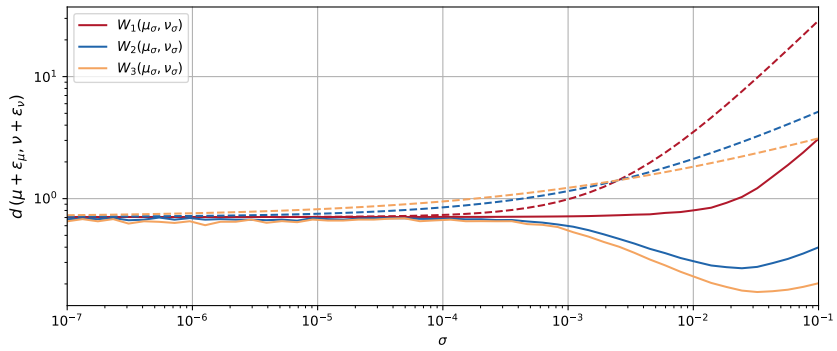

Figure 7: Distances between two images, one has a mass of 1 in (8,8) and the other in (24,24) and zero everywhere else, both being noised with noise sampled from the zero-sum normal distribution, in dashed (matching colors) we have the bounds for each p from Theorem 5.

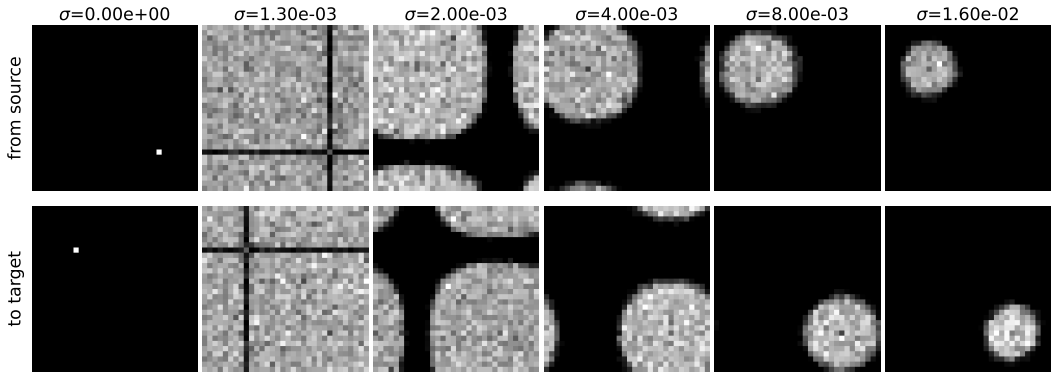

Figure 8: Top - Source image, and where the mass of the original pixel (8,8) goes. Bottom - Target image, and where the mass of the target pixel (24, 24) comes from for the optimal transport map between noised versions of the single pixel images.

5 Conclusion and Future Work

In this paper, we have investigated the behavior of the signed Wasserstein distance under noise corruption of the pictures. Our theoretical contributions provide bounds for various situations of interest. In particular, certain bounds establish a better noise robustness of the signed Wasserstein distance than the ubiquitous L^2 metric. Our numerical experiments on the DOTMark dataset corroborate these findings, with empirical results confirming that the W_2 distance is more resilient to noise than both L^2 and W_1 distances. These results make a strong case for its use in noise-plagued applications like cryo-EM.

Despite these results, there remains venue for additional work. A primary challenge would be to establish a (sharp) bound for $\mathbb{E}W_p^{\pm}(\mu_\epsilon, \nu_\epsilon) - \mathbb{E}W_p^{\pm}(\mu, \nu)$, which is hindered by the lack of triangle inequality. The numerical experiments further suggest that our bounds, despite capturing the correct behavior, are not tight. Finally, as our theory suggests that robustness increases with higher values of p, the interest of such choices of exponents for practical applications should be investigated in future works.

REFERENCES

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 214–223. PMLR, July 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, June 1991. ISSN 0010-3640, 1097-0312. doi: 10.1002/cpa.3160440402.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal Transport for Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9): 1853–1865, September 2017. ISSN 1939-3539. doi: 10.1109/TPAMI.2016.2615921.

CryoJAX Developers. Cryojax documentation, 2025. URL https://github.com/mjo22/cryojax.

Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

Bjorn Engquist, Brittany D. Froese, and Yunan Yang. Optimal Transport for Seismic Full Waveform Inversion, April 2016.

Jean Feydy, Pierre Roussillon, Alain Trouvé, and Pietro Gori. Fast and Scalable Optimal Transport for Brain Tractograms, July 2021.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T. H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python Optimal Transport. *Journal of Machine Learning Research*, 22(78): 1–8, 2021. ISSN 1533-7928. URL http://jmlr.org/papers/v22/20-451.html.

Michel Groppe, Linus Niemöller, Shayan Hundrieser, David Ventzke, Anna Blob, Sarah Köster, and Axel Munk. Optimal Transport Based Testing in Factorial Design, September 2025.

Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the Loss Landscape of Neural Nets, November 2018.

E. Mainini. A description of transport cost for signed measures. *Journal of Mathematical Sciences*, 181(6):837–855, March 2012. ISSN 1573-8795. doi: 10.1007/s10958-012-0718-2.

Rohan Rao, Amit Moscovich, and Amit Singer. Wasserstein K-Means for Clustering Tomographic Projections. In *Machine Learning for Structural Biology Workshop, Neural Information Processing Systems (NeurIPS)*, pp. 1–12, 2020. URL https://www.mlsb.io/papers/MLSB2020_Wasserstein_K-Means_for_Clustering.pdf.

A. Tajmir Riahi, G. Woollard, F. Poitevin, A. Condon, and K. Dao Duc. AlignOT: An optimal transport based algorithm for fast 3D alignment with applications to cryogenic electron microscopy density maps, October 2022. URL https://arxiv.org/abs/2210.09361v1.

Filippo Santambrogio. *Optimal transport for applied mathematicians*. Springer, 2015.

Bernhard Schmitzer. Stabilized Sparse Scaling Algorithms for Entropy Regularized Transport Problems. *SIAM Journal on Scientific Computing*, 41(3):A1443–A1481, January 2019. ISSN 1064-8275, 1095-7197. doi: 10.1137/16M1106018.

Jörn Schrieber, Dominic Schuhmacher, and Carsten Gottschlich. DOTmark - A Benchmark for Discrete Optimal Transport. *IEEE Access*, 5:271–282, 2017. ISSN 2169-3536. doi: 10.1109/ACCESS.2016.2639065.

Yunpeng Shi, Amit Singer, and Eric J. Verbeke. Fast alignment of heterogeneous images in sliced Wasserstein distance, March 2025.

Andrew K. Shiau, Seth F. Harris, Daniel R. Southworth, and David A. Agard. Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell*, 127 (2):329–340, October 2006. ISSN 0092-8674. doi: 10.1016/j.cell.2006.09.027.

Amit Singer and Ruiyi Yang. Alignment of density maps in Wasserstein distance. *Biological Imaging*, 4:e5, January 2024. ISSN 2633-903X. doi: 10.1017/S2633903X24000059.

Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein Auto-Encoders, 2017.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A), 2019. ISSN 1350-7265. doi: 10.3150/18-BEJ1065.

A Deferred proofs

A.1 Proof of Propositions

Proposition 3 (Wasserstein Distance Decomposition)**.** *Let μ and ν be two non-negative measures on a space $\mathcal{X}$ with equal total mass. It holds that*

$$W_p^p(\mu,\nu) \leq W_p^p\left((\mu-\nu)_+,(\nu-\mu)_+\right). \tag{20}$$

Proof of Proposition 3. We can decompose any two measures μ and ν into a common part and two disjoint parts. Let m be the largest measure such that for all Borel set A

$$m(A) \leq \mu(A) \text{ and } m(A) \leq \nu(A).$$

The remaining, disjoint parts of each measure are $\mu' := \mu - m = (\mu-\nu)_+$ and $\nu' := \nu - m = (\nu-\mu)_+$. Thus, we can write:

$$\mu = m + \mu' \qquad \nu = m + \nu' \tag{21}$$

Since μ and ν have the same total mass, it follows that μ' and ν' also have the same total mass.

We can then construct a valid transport plan π from μ to ν by handling the common and disjoint parts separately. For the disjoint parts, let π'_{opt} be the optimal transport plan from μ' to ν', whose cost is, by definition, $W_p^p(\mu',\nu')$. For the common part, we use the identity plan, π_{id}, which transports the mass at each point x to itself. The cost of this plan is $\int_{\mathcal{X}} d(x,x)^p \mathrm{d}\pi_{\text{id}}(x) = 0$.

Using the gluing principle, we can form a complete transport plan $\pi = \pi_{\text{id}} + \pi'_{\text{opt}}$. This is a valid plan transporting μ to ν. Its total cost is the sum of the costs of its components:

$$\text{Cost}(\pi) = \text{Cost}(\pi_{\text{id}}) + \text{Cost}(\pi'_{\text{opt}}) = 0 + W_p^p(\mu',\nu') \tag{22}$$

By the definition of the Wasserstein distance as the infimum of costs over all possible transport plans, the true optimal cost must be less than or equal to the cost of this specific plan:

$$W_p^p(\mu,\nu) \leq W_p^p(\mu',\nu') \tag{23}$$

Substituting the definitions of μ' and ν' completes the proof. □

Proposition 4. *Let $\mu,\nu : G_n \to [0,\infty)$ be images on the square grid G_n with spacing $h = 1/n$, and let $\epsilon_\mu, \epsilon_\nu$ satisfy Assumption 1. Identifying G_n with the 2-torus, let $D := \operatorname{diam}(G_n) = \sqrt{2}/2$. Then, for any $p \geq 1$,*

$$W_p^{\pm}(\mu+\epsilon_\mu,\nu+\epsilon_\nu) \leq D^{1-\frac{1}{p}}\left(W_1(\mu,\nu) + W_1(\epsilon_+^*,\epsilon_-^*)\right)^{\frac{1}{p}}, \qquad \epsilon^* := \epsilon_\mu - \epsilon_\nu \tag{24}$$

Proof. By definition of the signed distance,

$$W_p^{\pm}(\mu+\epsilon_\mu,\nu+\epsilon_\nu) = W_p\big((\mu+\epsilon_\mu)_+ + (\nu+\epsilon_\nu)_-,\ (\nu+\epsilon_\nu)_+ + (\mu+\epsilon_\mu)_-\big). \tag{25}$$

Applying the decomposition inequality of Proposition 3 (which "drops the overlap") to these non-negative arguments gives

$$W_p^{\pm}(\mu+\epsilon_\mu,\nu+\epsilon_\nu) \le W_p\big([\mu+\epsilon_\mu-\nu-\epsilon_\nu]_+,\ [\nu+\epsilon_\nu-\mu-\epsilon_\mu]_+\big). \tag{26}$$

For general $p \ge 1$ on a bounded domain of diameter D we use the standard comparison

$$W_p(\alpha,\beta) \le D^{1-\frac{1}{p}} W_1(\alpha,\beta)^{\frac{1}{p}}. \tag{27}$$

Applying this to $(\alpha,\beta) = ([\mu+\epsilon_\mu-\nu-\epsilon_\nu]_+, [\mu+\epsilon_\mu-\nu-\epsilon_\nu]_-)$ yields

$$W_p([\mu+\epsilon_\mu-\nu-\epsilon_\nu]_+, [\mu+\epsilon_\mu-\nu-\epsilon_\nu]_-) \le D^{1-\frac{1}{p}} \left(W_1([\mu+\epsilon_\mu-\nu-\epsilon_\nu]_+, [\mu+\epsilon_\mu-\nu-\epsilon_\nu]_-) \right)^{\frac{1}{p}}. \tag{28}$$

Using Proposition 6, we conclude

$$\begin{aligned} &D^{1-\frac{1}{p}} \left(W_1([\mu+\epsilon_\mu-\nu-\epsilon_\nu]_+, [\mu+\epsilon_\mu-\nu-\epsilon_\nu]_-) \right)^{\frac{1}{p}} \\ &\le D^{1-\frac{1}{p}} \left(W_1(\mu,\nu) + W_1\big((\epsilon_\mu-\epsilon_\nu)_+, (\epsilon_\mu-\epsilon_\nu)_-\big) \right)^{\frac{1}{p}}. \end{aligned}$$

Since both ϵ_μ and ϵ_ν are normally distributed, we can say that $\epsilon^* := \epsilon_\mu - \epsilon_\nu$ is also normally distributed, with $\mathrm{cov}(\epsilon^*) = 2\,\mathrm{cov}(\epsilon_\mu)$. Thus,

$$D^{1-\frac{1}{p}} \big(W_1(\mu,\nu) + W_1((\epsilon_\mu-\epsilon_\nu)_+, (\epsilon_\mu-\epsilon_\nu)_-)\big)^{\frac{1}{p}} \le D^{1-\frac{1}{p}} \big(W_1(\mu,\nu) + W_1(\epsilon_+^*, \epsilon_-^*)\big)^{\frac{1}{p}}. \tag{29}$$

□

A.2 Proof of Theorems

Proof of Proposition 2. We prove each point separately.

1. The marginal variance of each component N_i is given by the diagonal entry Σ_{ii}, which is σ^2 by definition. Since the parent distribution is a multivariate normal with a mean vector of zero, each component is marginally distributed as $\mathcal{N}(0,\sigma^2)$.

2. We compute the variance of the sum of the components:

$$\begin{aligned} \mathrm{Var}\left(\sum_{i=1}^M N_i\right) &= \sum_{i,j} \mathrm{Cov}(N_i,N_j) = \sum_{i=1}^M \sum_{j=1}^M \Sigma_{ij} && (30)\\ &= \sum_{i=1}^M \mathrm{Var}(N_i) + \sum_{i\ne j} \mathrm{Cov}(N_i,N_j) && (31)\\ &= M\cdot\sigma^2 + M(M-1)\cdot\left(-\frac{\sigma^2}{M-1}\right) && (32)\\ &= M\sigma^2 - M\sigma^2 = 0. && (33) \end{aligned}$$

   The expectation of the sum is $\mathbb{E}\left[\sum_{i=1}^M N_i\right] = \sum_{i=1}^M \mathbb{E}[N_i] = 0$. A random variable with zero mean and zero variance must be equal to zero almost surely. Therefore, $\sum_{i=1}^M N_i = 0$. □

Proposition 5. *Let $\mu : G_n \to [0,1]$ be a probability measure on the $n \times n$ unit grid G_n with cyclic boundary conditions and let $\epsilon : G_n \to \mathbb{R}$ be a signed noise measure that satisfy Assumption 1. Then, for $p > 1$,*

$$W_p^{\pm}(\mu,\mu+\epsilon) \le W_p(\epsilon_-,\epsilon_+). \tag{34}$$

Proof of proposition 5. By definition,

$$(W_p^{\pm})^p(\mu, \mu+\epsilon) = W_p^p\Big(\mu + (\mu+\epsilon)_-, (\mu+\epsilon)_+\Big). \tag{35}$$

Thus, using Proposition 3 on $\mu + (\mu+\epsilon)_-$ and $(\mu+\epsilon)_+$, we get

$$W_p^p\Big(\mu + (\mu+\epsilon)_-, (\mu+\epsilon)_+\Big) \tag{36}$$

$$\leq W_p^p\Big(\big(\mu + (\mu+\epsilon)_- - (\mu+\epsilon)_+\big)_+, \big((\mu+\epsilon)_+ - (\mu + (\mu+\epsilon)_-)\big)_+\Big) \tag{37}$$

$$= W_p^p\Big(\big(\mu - ((\mu+\epsilon)_+ - (\mu+\epsilon)_-)\big)_+, \big((\mu+\epsilon)_+ - (\mu+\epsilon)_- - \mu)\big)_+\Big) \tag{38}$$

$$= W_p^p\Big(\big(\mu - (\mu+\epsilon)\big)_+, \big(\mu+\epsilon-\mu\big)_+\Big) \tag{39}$$

$$= W_p^p\Big((-\epsilon)_+, \epsilon_+\Big) = W_p^p(\epsilon_-, \epsilon_+). \qquad \square$$

Proof of Theorem 2. Using the Kantorovich-Rubinsten duality,

$$W_1^{\pm}(\mu, \mu+\varepsilon) = \sup_{f \in \mathrm{Lip}_1} \langle f, \varepsilon \rangle = W_1(\varepsilon^+, \varepsilon^-). \tag{40}$$

The first equality is the signed dual form with $\mu - (\mu+\varepsilon) = -\varepsilon$. For the second, $\int \varepsilon = 0$ implies $\varepsilon = \varepsilon^+ - \varepsilon^-$ with equal masses, so the balanced duality gives $W_1(\varepsilon^+, \varepsilon^-) = \sup_{f \in Lip_1} \langle f, \varepsilon \rangle$

Let $m = \varepsilon^+(G_n) = \varepsilon^-(G_n)$. By homogeneity of W_1,

$$W_1(\varepsilon^+, \varepsilon^-) = m\, W_1\Big(\tfrac{\varepsilon^+}{m}, \tfrac{\varepsilon^-}{m}\Big). \tag{41}$$

Apply Proposition 1 to the probability measures ε^+/m and ε^-/m. There exists an integer k^* with $k^* \asymp \log_2 n$ such that

$$W_1(\frac{\varepsilon^+}{m}, \frac{\varepsilon^-}{m}) \leq \frac{\sqrt{2}}{2}\, 2^{-k^*} + \frac{\sqrt{2}}{2} \sum_{k=0}^{k^*} 2^{-k} \sum_{Q \in \mathcal{D}_k} |(\frac{\varepsilon^+}{m} - \frac{\varepsilon^-}{m})(Q)|. \tag{42}$$

Multiplying by m gives

$$W_1(\varepsilon^+, \varepsilon^-) \leq \frac{\sqrt{2}}{2}\, m\, 2^{-k^*} + \frac{\sqrt{2}}{2} \sum_{k=0}^{k^*} 2^{-k} \sum_{Q \in \mathcal{D}_k} \Big| \sum_{x \in Q} \varepsilon(x) \Big|. \tag{43}$$

Taking expectations and using independence and zero mean of the noise,

$$\mathbb{E} W_1(\varepsilon^+, \varepsilon^-) \leq \frac{\sqrt{2}}{2}\, 2^{-k^*}\, \mathbb{E} m + \frac{\sqrt{2}}{2} \sum_{k=0}^{k^*} 2^{-k} \sum_{Q \in \mathcal{D}_k} \mathbb{E} \Big| \sum_{x \in Q} \varepsilon(x) \Big|. \tag{44}$$

Since each $\varepsilon(x)$ is Gaussian with variance σ^2, one has $\mathbb{E}|\sum_{x \in Q} \varepsilon(x)| \leq \sigma \sqrt{|Q|}\sqrt{2/\pi}$ and $\mathbb{E} m = \sum_{x \in G_n} \mathbb{E}(\varepsilon(x))_+ = n^2\sigma/\sqrt{2\pi}$. Furthermore, the dyadic family $\mathcal{D}_k$ has $|\mathcal{D}_k| = 2^{2k}$ cubes of cardinality $|Q| = n^2/2^{2k}$. Therefore

$$\sum_{Q \in \mathcal{D}_k} \mathbb{E} \Big| \sum_{x \in Q} \varepsilon(x) \Big| \leq \sigma \sqrt{\tfrac{2}{\pi}} \sum_{Q \in \mathcal{D}_k} \sqrt{|Q|} = \sigma \sqrt{\tfrac{2}{\pi}} \cdot 2^{2k} \cdot \frac{n}{2^k} = \sigma \sqrt{\tfrac{2}{\pi}}\, n\, 2^k. \tag{45}$$

Plugging this into the multiscale sum yields

$$\frac{\sqrt{2}}{2} \sum_{k=0}^{k^*} 2^{-k} \sum_{Q \in \mathcal{D}_k} \mathbb{E} \Big| \sum_{x \in Q} \varepsilon(x) \Big| \leq \frac{\sqrt{2}}{2} \sigma \sqrt{\tfrac{2}{\pi}}\, n \sum_{k=0}^{k^*} 1 \leq \frac{\sqrt{2}}{2} \sigma \sqrt{\tfrac{2}{\pi}}\, n \big(k^* + 1\big). \tag{46}$$

With $k^* \asymp \log_2 n$ this gives the $\sigma n \log_2 n$ contribution.

For the coarse term choose k^* so that $2^{-k^*} \asymp 1/n$. Then

$$\frac{\sqrt{2}}{2} 2^{-k^*} \mathbb{E}m \asymp \frac{\sqrt{2}}{2} \frac{1}{n} \cdot \frac{n^2 \sigma}{\sqrt{2\pi}} = \frac{\sqrt{2}}{2} \frac{\sigma n}{\sqrt{2\pi}}, \tag{47}$$

which is the σn contribution.

Collecting the two contributions and absorbing absolute constants into the displayed coefficients yields

$$\mathbb{E}W_1(\varepsilon^+, \varepsilon^-) \leq \frac{2}{\sqrt{\pi}} \sigma n \log_2 n + \frac{1}{2\sqrt{\pi}} \sigma n. \tag{48}$$

In this derivation the factor m appears only in the coarse term and contributes to the σn piece after expectation. In the oscillation terms it cancels with the normalization, so no additional dependence on m remains. There is no additive grid term independent of σ, hence no $1/(\sqrt{2}n)$ tail. □

Proof of Theorem 3. By Proposition 5, we only need to upper bound $W_p(\epsilon_+, \epsilon_-)$.

By the assumption on the noise noise have total zero mass, this quantity is well defined.

Then, by the multiscale bound of Proposition 1

$$W_p^p(\epsilon_+, \epsilon_-) = 2^{-p/2} \epsilon_+(G_n) W_p^p \left(\frac{\epsilon_+}{\epsilon_+(G_n)}, \frac{\epsilon_-}{\epsilon_+(G_n)} \right) \tag{49}$$

$$\leq 2^{-pk^*-p/2} \epsilon_+(G_n) + 2^{-p/2} \sum_{k=1}^{k^*} 2^{-p(k-1)} \sum_{Q_i^k \in \mathcal{Q}^k} |\epsilon_+(Q_i^k) - \epsilon_-(Q_i^k)| \tag{50}$$

$$\leq 2^{-pk^*-1/2} \epsilon_+(G_n) + 2^{-p/2} \sum_{k=1}^{k^*} 2^{-p(k-1)} \sum_{Q_i^k \in \mathcal{Q}^k} |\epsilon(Q_i^k)|. \tag{51}$$

Now, the proof is extremely similar to the previous one and by the same argument,

$$\mathbb{E} \sum_{Q \in \mathcal{Q}_k} |\epsilon^*(Q)| \leq 4^k \sqrt{\frac{2}{\pi}} \sigma 2^{\eta - k}. \tag{52}$$

As in the previous proof,

$$\mathbb{E}\epsilon_+(G_n) = \frac{n^2}{2} \sqrt{\frac{2}{\pi}} \sigma \sqrt{1 - \frac{1}{n^2}}. \tag{53}$$

Altogether,

$$\mathbb{E}W_p^p(\epsilon_+, \epsilon_-) \leq 2^{-pk^*-1/2} \frac{4^\eta}{2} \sqrt{\frac{2}{\pi}} \sigma + 2^\eta 2^{p/2} \sum_{k=1}^{k^*} 2^{-(p-1)k} \sqrt{\frac{2}{\pi}} \sigma \tag{54}$$

$$\leq 2^{-pk^*-1/2} \frac{4^\eta}{2} \sqrt{\frac{2}{\pi}} \sigma + 2^\eta 2^{p/2} \sqrt{\frac{2}{\pi}} \sigma \frac{1 - 2^{-(p-1)k^*}}{2^{p-1} - 1}. \tag{55}$$

$$\tag{56}$$

We take $k^* = \eta$ again to get

$$\mathbb{E}W_p^p(\epsilon_+, \epsilon_-) \leq 2^{-(p-1)\eta} 2^{-1/2} \frac{2^\eta}{2} \frac{1}{\sqrt{\pi}} \sigma + 2^\eta 2^{p/2} \sqrt{\frac{2}{\pi}} \sigma \frac{1}{2^{p-1} - 3/2} \tag{57}$$

$$\leq \frac{2^\eta}{\sqrt{\pi}} \sigma \left(2^{-(p-1)\eta - 3/2} + \frac{2^{(p+1)/2}}{2^{p-1} - 1} \right). \tag{58}$$

Remark that $2^{-(p-1)\eta-3/2} \leq 1/2$ and that $\frac{2^{(p+1)/2}}{2^{p-1}-1}$ is decreasing with value $2\sqrt{2}$ at 2. The claim follows. □

Proposition 6. *For any two images $\mu, \nu : G_n \to [0, \infty)$ and independent noises $\epsilon_\mu, \epsilon_\nu$ as in Assumption 1,*

$$W_1([\mu + \epsilon_\mu - \nu - \epsilon_\nu]_+, [\nu + \epsilon_\nu - \mu - \epsilon_\mu]_+) \\ \leq W_1(\mu, \nu) + W_1\big((\epsilon_\mu - \epsilon_\nu)^+, (\epsilon_\mu - \epsilon_\nu)^-\big).$$

Proof of Proposition 6. By Kantorovich–Rubinstein duality,

$$W_1\left([\mu + \epsilon_\mu - \nu - \epsilon_\nu]_+, [\nu + \epsilon_\nu - \mu - \epsilon_\mu]_+\right) = \sup_{\|f\|_{\mathrm{Lip}} \leq 1} \int f(\mu - \nu) + \int f(\epsilon_\mu - \epsilon_\nu). \tag{59}$$

For the first term, by KR duality,

$$\sup_{\|f\|_{\mathrm{Lip}} \leq 1} \int f(\mu - \nu) \leq W_1(\mu, \nu) \tag{60}$$

For the second term, via the Jordan decomposition,

$$\sup_{\|f\|_{\mathrm{Lip}} \leq 1} \int f(\epsilon_\mu - \epsilon_\nu) \leq W_1((\epsilon_\mu - \epsilon_\nu)^+, (\epsilon_\mu - \epsilon_\nu)^-) \tag{61}$$

Adding these together, we receive the desired bound. □

Proof of Theorem 4. Recall that $W_1^\pm$ satisfies the triangle inequality, so

$$W_1^\pm(\mu + \epsilon_\mu, \nu + \epsilon_\nu) \leq W_1^\pm(\mu + \epsilon_\mu, \mu) + W_1^\pm(\mu, \nu) + W_1^\pm(\nu, \nu + \epsilon_\nu). \tag{62}$$

By symmetry

$$\mathbb{E}W_1^\pm(\mu + \epsilon_\mu, \mu) = \mathbb{E}W_1^\pm(\nu, \nu + \epsilon_\nu) \tag{63}$$

Therefore,

$$\mathbb{E}[W_1^\pm(\mu + \epsilon_\mu, \nu + \epsilon_\nu) - W_1^\pm(\mu, \nu)] \leq 2\mathbb{E}W_1^\pm(\mu, \mu + \epsilon_\mu). \tag{64}$$

We proceed to upper-bound the RHS. By the definition of the signed Wassetein metric,

$$W_1^\pm(\mu, \mu + \epsilon) = W_1(\mu_+ + (\mu + \epsilon)_-, (\mu + \epsilon)_+ + \mu_-) \tag{65}$$

$$= W_1(\mu + (\mu + \epsilon)_-, (\mu + \epsilon)_+) \qquad \text{(since } \mu_+ = \mu \text{ and } \mu_- = 0\text{).} \tag{66}$$

We now use the dyadic upper bound in equation 9. The image is partitioned into 4 quadrants recursively, thus $\delta = 1/2$. Our domain has diameter $\sqrt{2}/2$ since it is the discrete $n \times n$ unit grid $G_n \subset [0, 1] \times [0, 1] \in \mathbb{R}^2$ with cyclic boundary conditions. The inequality only holds for probability measures, so we need to rescale.

$$W_1^\pm(\mu, \mu_\epsilon) = (\mu + \epsilon)_+(G_n) W_1^\pm\left(\frac{\mu + (\mu + \epsilon)_-}{(\mu + \epsilon)_+(G_n)}, \frac{(\mu + \epsilon)_+}{(\mu + \epsilon)_+(G_n)}\right) \tag{67}$$

$$\leq \tfrac{\sqrt{2}}{2} \cdot 2^{-k^*} (\mu + \epsilon)_+(G_n) + \tfrac{\sqrt{2}}{2} \sum_{k=1}^{k^*} 2^{-(k-1)} \sum_{Q_i^k \in \mathcal{Q}^k} \left|(\mu + (\mu + \epsilon)_-)(Q_i^k) - (\mu + \epsilon)_+(Q_i^k)\right|.$$

By considering the two cases $(\mu + \epsilon)(Q_i^k) \geq 0$ and $(\mu + \epsilon)(Q_i^k) < 0$ it is easy to see that the term $(\mu + (\mu + \epsilon)_-)(Q_i^k) - (\mu + \epsilon)_+(Q_i^k)$ is equal to $-\epsilon(Q_i^k)$, so the bound above simplifies to

$$W_1^\pm(\mu, \mu_\epsilon) \leq 2^{-k^* - \frac{1}{2}} (\mu + \epsilon)_+(G_n) + \frac{\sqrt{2}}{2} \sum_{k=1}^{k^*} 2^{-(k-1)} \sum_{Q_i^k \in \mathcal{Q}^k} |\epsilon(Q_i^k)|. \tag{68}$$

Rewrite the noise as $\epsilon = \epsilon' - \bar{\epsilon}$ where ϵ' is i.i.d. $\mathcal{N}(0, \sigma^2)$ at each pixel and $\bar{\epsilon} \in \mathbb{R}$ is the mean of all ϵ' terms across the entire image. Since Q_i^k is a square region of size $2^{\eta-k} \times 2^{\eta-k}$ and $\bar{\epsilon}$ is the mean of 4^η i.i.d. Gaussian noise terms, it follows that $\epsilon'(Q_i^k) \sim \mathcal{N}(0, 4^{\eta-k}\sigma^2)$ and $\bar{\epsilon} \sim \mathcal{N}(0, \sigma^2/4^\eta) = \mathcal{N}(0, \sigma^2/n^2)$. Recall that $\mathbb{E}|X| = \sigma\sqrt{2/\pi}$ when $X \sim \mathcal{N}(0, \sigma^2)$.

Since $\epsilon^*(Q_i^k) = \sum_{x \in Q_i^k} \epsilon'(x) - 4^{\eta-k}\bar{\epsilon}$,

$$\mathrm{Var}\left(\epsilon^*(Q_i^k)\right) = \sigma^2\Big(4^{\eta-k} + \frac{4^{2(\eta-k)}}{n^2} - 2\frac{4^{2(\eta-k)}}{n^2}\Big). \tag{69}$$

Thus,

$$\mathbb{E}|\epsilon^*(Q_i^k)| = \sqrt{\frac{2}{\pi}}\sigma 2^{\eta-k}\Big(1 - n^2 4^{-k}\Big)^{1/2}. \tag{70}$$

Summing over the 4^k cells at level k,

$$\mathbb{E}\sum_{Q \in \mathcal{Q}_k} |\epsilon^*(Q)| = 4^k\sqrt{\frac{2}{\pi}}\sigma 2^{\eta-k}\Big(1 - n^2 4^{-k}\Big)^{1/2}. \tag{71}$$

Plugging this back into the RHS of equation 68 and recalling that $2^\eta = n$ gives

$$\mathbb{E}\left[\frac{\sqrt{2}}{2}\sum_{k=1}^{k^*} 2^{-(k-1)} \sum_{Q_i^k \in \mathcal{Q}^k} |\epsilon(Q_i^k)|\right] \leq \frac{\sqrt{2}}{2}\sum_{k=1}^{k^*} 2^{-(k-1)} 4^k \sqrt{\frac{2}{\pi}}\sigma 2^{\eta-k} \tag{72}$$

$$= \frac{2^{\eta+1}\sigma}{\sqrt{\pi}}k^*. \tag{73}$$

We take $k^* = \eta = \log_2 n$ to obtain the bound

$$\mathbb{E}W_1^{\pm}(\mu, \mu_\epsilon) \leq \frac{1}{\sqrt{2}n}\mathbb{E}\left[(\mu + \epsilon)_+(G_n)\right] + \frac{2n\log_2 n}{\sqrt{\pi}}\sigma. \tag{74}$$

We now bound the first term in the RHS.

$$\mathbb{E}\left[(\mu + \epsilon)_+(G_n)\right] \leq \mathbb{E}[\mu_+(G_n)] + \mathbb{E}[\epsilon_+(G_n)] \tag{75}$$

$$= 1 + \mathbb{E}[\epsilon_+(G_n)] \tag{76}$$

where the last equality follows from the fact that μ is a (non-negative) probability measure. By a symmetry argument

$$\mathbb{E}\epsilon_+(G_n) = \tfrac{1}{2}\mathbb{E}|\epsilon|(G_n). \tag{77}$$

Further set $m = \frac{1}{2}\sum_{x \in G_n} |\epsilon(x)|$ and recall that $\epsilon(x) \sim \mathcal{N}\big(0, \sigma^2(1 - 1/n^2)\big)$ to derive

$$\mathbb{E}m = \frac{n^2}{2}\sqrt{\frac{2}{\pi}}\sigma\sqrt{1 - \frac{1}{n^2}}. \tag{78}$$

Thus,

$$\frac{1}{\sqrt{2}n}\mathbb{E}\left[(\mu + \epsilon)_+(G_n)\right] \leq \frac{1}{\sqrt{2}n} + \frac{\sigma}{2\sqrt{\pi}}n. \tag{79}$$

Plugging this back into equation 74 gives

$$\mathbb{E}W_1^{\pm}(\mu, \mu_\epsilon) \leq \frac{2n\log_2 n + n/2}{\sqrt{\pi}}\sigma + \frac{1}{\sqrt{2}n}. \tag{80}$$

Note that the same bound applies to $\mathbb{E}W_1^{\pm}(\nu, \nu + \epsilon_\nu)$. By subtracting $W_1^{\pm}(\mu, \nu)$ from both sides of equation 62 and taking expectations, we have

$$\mathbb{E}\left[W_1^{\pm}(\mu_\epsilon, \nu_\epsilon) - W_1^{\pm}(\mu, \nu)\right] \leq \mathbb{E}W_1^{\pm}(\mu_\epsilon, \mu) + \mathbb{E}W_1^{\pm}(\nu, \nu_\epsilon)$$

$$\leq \frac{4n\log_2 n + n}{\sqrt{\pi}}\sigma + \frac{\sqrt{2}}{n}. \qquad \square$$

Proof of Theorem 5. Using Proposition 4

$$\mathbb{E}\big[W_p^{\pm}(\mu+\epsilon_\mu,\nu+\epsilon_\nu)\big] \leq \mathbb{E}\Big[D^{1-\frac{1}{p}}\big(W_1(\mu,\nu)+W_1(\epsilon_+^*,\epsilon_-^*)\big)^{\frac{1}{p}}\Big] \tag{81}$$

The function $t \mapsto t^{1/p}$ is concave on $[0,\infty)$, hence by Jensen:

$$\mathbb{E}\Big[D^{1-\frac{1}{p}}\big(W_1(\mu,\nu)+W_1(\epsilon_+^*,\epsilon_-^*)\big)^{\frac{1}{p}}\Big] \leq D^{1-\frac{1}{p}}\Big(\mathbb{E}\big[W_1(\mu,\nu)\ +W_1(\epsilon_+^*,\epsilon_-^*)\big]\Big)^{\frac{1}{p}} \tag{82}$$

By the linearity of expectation,

$$D^{1-\frac{1}{p}}\Big(\mathbb{E}\big[W_1(\mu,\nu)\ +W_1(\epsilon_+^*,\epsilon_-^*)\big]\Big)^{\frac{1}{p}} = D^{1-\frac{1}{p}}\Big(W_1(\mu,\nu)\ +\mathbb{E}\big[W_1(\epsilon_+^*,\epsilon_-^*)\big]\Big)^{\frac{1}{p}} \tag{83}$$

Finally, using Theorem 2 we get that

$$D^{1-\frac{1}{p}}\Big(W_1(\mu,\nu)\ +\mathbb{E}\big[W_1(\epsilon_+^*,\epsilon_-^*)\big]\Big)^{\frac{1}{p}} \leq D^{1-\frac{1}{p}}\Big(W_1(\mu,\nu)+\frac{2\sqrt{2}}{\sqrt{\pi}}\sigma n\log_2 n+\sqrt{\frac{2}{\pi}}\sigma n.\Big)^{\frac{1}{p}} \tag{84}$$

Using Jensen,

$$\mathbb{E}\big[W_p^{\pm}(\mu+\epsilon_\mu,\nu+\epsilon_\nu)\big] \leq (\frac{\sqrt{2}}{2})^{1-\frac{1}{p}}W_1(\mu,\nu)^{\frac{1}{p}}+\frac{\sqrt{2}}{2}\Big(\frac{4}{\sqrt{\pi}}n\log_2 n+\frac{2}{\sqrt{\pi}}n\Big)^{\frac{1}{p}}\sigma^{\frac{1}{p}}. \tag{85}$$

□

Proof of Theorem 1. First let us remark that

$$\sum_{x\in G_n} S_{\mu_\epsilon,\nu_\epsilon}(x)-T_{\mu_\epsilon,\nu_\epsilon}(x) = \sum_{x\in G_n}\mu(x)+\epsilon_\mu(x)-\nu(x)-\epsilon_\nu(x) \tag{86}$$

$$= 0+\sum_{x\in G_n}\epsilon_\mu(x)-\epsilon_\nu(x), \tag{87}$$

as

$$\sum_{x\in G_n}\mu_+(x)+\nu_-(x) = \sum_{x\in G_n}\nu_+(x)+\mu_-(x) \tag{88}$$

and thus

$$\sum_{x\in G_n}\mu(x)-\nu(x)=0. \tag{89}$$

Remark that under our assumptions,

$$\sum_{x\in G_n}\epsilon_\mu(x)-\epsilon_\nu(x)\sim\mathcal{N}(0,2\sigma^2N^2). \tag{90}$$

Because of this, one has that

$$\sum_{x\in G_n} S_{\mu_\epsilon,\nu_\epsilon}(x) = \sum_{x\in G_n} T_{\mu_\epsilon,\nu_\epsilon}(x)\left(1+\frac{\mathrm{O}_p\left(\sigma N\right)}{\sum_{x\in G_n}T_{\mu_\epsilon,\nu_\epsilon}(x)}\right). \tag{91}$$

Owing to our assumption on the signals, notice that

$$\sum_{x\in G_n} T_{\mu_\epsilon,\nu_\epsilon}(x) = \mathrm{O}_p(N^2). \tag{92}$$

Therefore,

$$W_1(\bar{S}_{\mu_\epsilon,\nu_\epsilon},\bar{T}_{\mu_\epsilon,\nu_\epsilon}) = \sup_{f\in\mathrm{Lip}_1}\langle \bar{S}_{\mu_\epsilon,\nu_\epsilon}-\bar{T}_{\mu_\epsilon,\nu_\epsilon},f\rangle \tag{93}$$

$$= \sup_{f\in\mathrm{Lip}_1}\left\langle \frac{S_{\mu_\epsilon,\nu_\epsilon}}{\sum_{x\in G_n}S_{\mu_\epsilon,\nu_\epsilon}(x)}-\frac{T_{\mu_\epsilon,\nu_\epsilon}}{\sum_{x\in G_n}T_{\mu_\epsilon,\nu_\epsilon}(x)},f\right\rangle \tag{94}$$

$$= \frac{1}{\sum_{x\in G_n}S_{\mu_\epsilon,\nu_\epsilon}(x)}\sup_{f\in\mathrm{Lip}_1}\left\langle S_{\mu_\epsilon,\nu_\epsilon}-T_{\mu_\epsilon,\nu_\epsilon}\left(1+\frac{\mathrm{O}_p\left(\sigma N\right)}{\sum_{x\in G_n}T_{\mu_\epsilon,\nu_\epsilon}(x)}\right),f\right\rangle. \ \square$$