The Annals of Applied Probability
2024, Vol. 34, No. 2, 2447–2487
https://doi.org/10.1214/23-AAP2028

LIMIT DISTRIBUTION THEORY FOR SMOOTH *p*-WASSERSTEIN DISTANCES

BY ZIV GOLDFELD[1,a], KENGO KATO[2,b], SLOAN NIETERT[3,c] AND GABRIEL RIOUX[4,d]

[1]*School of Electrical and Computer Engineering, Cornell University,* [a]*goldfeld@cornell.edu*
[2]*Department of Statistics and Data Science, Cornell University,* [b]*kk976@cornell.edu*
[3]*Department of Computer Science, Cornell University,* [c]*nietert@cs.cornell.edu*
[4]*Center for Applied Mathematics, Cornell University,* [d]*ger84@cornell.edu*

The Wasserstein distance is a metric on a space of probability measures that has seen a surge of applications in statistics, machine learning, and applied mathematics. However, statistical aspects of Wasserstein distances are bottlenecked by the curse of dimensionality, whereby the number of data points needed to accurately estimate them grows exponentially with dimension. Gaussian smoothing was recently introduced as a means to alleviate the curse of dimensionality, giving rise to a parametric convergence rate in any dimension, while preserving the Wasserstein metric and topological structure. To facilitate valid statistical inference, in this work, we develop a comprehensive limit distribution theory for the empirical smooth Wasserstein distance. The limit distribution results leverage the functional delta method after embedding the domain of the Wasserstein distance into a certain dual Sobolev space, characterizing its Hadamard directional derivative for the dual Sobolev norm, and establishing weak convergence of the smooth empirical process in the dual space. To estimate the distributional limits, we also establish consistency of the nonparametric bootstrap. Finally, we use the limit distribution theory to study applications to generative modeling via minimum distance estimation with the smooth Wasserstein distance, showing asymptotic normality of optimal solutions for the quadratic cost.

1. Introduction.

1.1. *Overview.* The Wasserstein distance is an instance of the Kantorovich optimal transport problem [63], which defines a metric on a space of probability measures. Specifically, for $1 \le p < \infty$, the p-Wasserstein distance between two Borel probability measures μ and ν on $\mathbb{R}^d$ with finite pth moments is defined by

$$\mathsf{W}_p(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \left[\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \, d\pi(x,y)\right]^{1/p}, \tag{1}$$

where $\Pi(\mu,\nu)$ is the set of couplings (or transportation plans) of μ and ν. The Wasserstein distance has seen a surge of applications in statistics, machine learning, and applied mathematics, ranging from generative modeling [6, 59, 96], image recognition [84, 86], and domain adaptation [25, 26] to robust optimization [12, 48, 75] and partial differential equations [62, 88]. The widespread applicability of the Wasserstein distance is driven by an array of desirable properties, including its metric structure (W_p metrizes weak convergence plus convergence of pth moments), a convenient dual form, robustness to support mismatch, and a rich geometry it induces on a space of probability measures. We refer to [4, 87, 100, 101] as standard references on optimal transport theory.

Received December 2021; revised April 2023.
MSC2020 subject classifications. Primary 60F05, 62E20; secondary 62G09.
Key words and phrases. Dual Sobolev space, Gaussian smoothing, functional delta method, limit distribution theory, Wasserstein distance.

However, statistical aspects of Wasserstein distances are bottlenecked by the curse of dimensionality, whereby the number of data points needed to accurately estimate them grows exponentially with dimension. Specifically, for the empirical distribution $\hat{\mu}_n$ of n independent observations from a distribution μ on $\mathbb{R}^d$, it is known that $\mathbb{E}[\mathsf{W}_p(\hat{\mu}_n, \mu)]$ scales as $n^{-1/d}$ for $d > 2p$ under moment conditions [16, 37, 46, 69, 102]. This slow rate renders performance guarantees in terms of W_p all but vacuous when d is large. It is also a roadblock towards a refined statistical analysis concerning limit distributions, bootstrap, and valid inference.

Gaussian smoothing was recently introduced as a means to alleviate the curse of dimensionality of empirical W_p [51–53, 77, 85]. For $\sigma > 0$, the smooth p-Wasserstein distance is defined as $\mathsf{W}_p^{(\sigma)}(\mu, \nu) := \mathsf{W}_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma)$, where $*$ denotes convolution and $\gamma_\sigma = N(0, \sigma^2 I_d)$ is the isotropic Gaussian distribution with variance parameter σ^2. For sufficiently sub-Gaussian μ, [53] showed that the expected smooth distance between $\hat{\mu}_n$ and μ exhibits the parametric convergence rate, that is, $\mathbb{E}[\mathsf{W}_1^{(\sigma)}(\hat{\mu}_n, \mu)] = O(n^{-1/2})$ in any dimension. This is a significant departure from the $n^{-1/d}$ rate in the unsmoothed case. [51] further showed that $\mathsf{W}_1^{(\sigma)}$ maintains the metric and topological structure of W_1 and is able to approximate it within a $\sigma\sqrt{d}$ gap. The structural properties and fast empirical convergence rates were later extended to $p > 1$ in [77]. Other follow-up works explored relations between $\mathsf{W}_p^{(\sigma)}$ and maximum mean discrepancies [105], analyzed its rate of decay as $\sigma \to \infty$ [20], and adopted it as a performance metric for nonparametric mixture model estimation [60].

A limit distribution theory for $\mathsf{W}_1^{(\sigma)}$ was developed in [52, 85], where the scaled empirical distance $\sqrt{n}\mathsf{W}_1^{(\sigma)}(\hat{\mu}_n, \mu)$ was shown to converge in distribution to the supremum of a tight Gaussian process in every dimension d under mild moment conditions. This result relies on the dual formulation of W_1 as an integral probability metric (IPM) over the class of 1-Lipschitz functions. Gaussian smoothing shrinks the function class to that of 1-Lipschitz functions convolved with a Gaussian density, which is shown to be μ-Donsker in every dimension, thereby yielding the limit distribution. Extending these results to empirical $\mathsf{W}_p^{(\sigma)}$ with $p > 1$, however, requires substantially new ideas due to the lack of an IPM structure. Consequently, works exploring $\mathsf{W}_p^{(\sigma)}$ with $p > 1$, such as [77, 105], did not contain limit distribution results for it and this question remained largely open.

The present paper closes this gap and provides a comprehensive limit distribution theory for empirical $\mathsf{W}_p^{(\sigma)}$ with $p > 1$. Our main limit distribution results are summarized in the following theorem, where the "null" refers to when $\mu = \nu$, while "alternative" corresponds to $\mu \neq \nu$. In what follows, the dimension $d \geq 1$ is arbitrary.

THEOREM 1.1 (Main results). *Let $1 < p < \infty$, and μ, ν be Borel probability measures on $\mathbb{R}^d$ with finite pth moments. Let $\hat{\mu}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ and $\hat{\nu}_n = n^{-1}\sum_{i=1}^n \delta_{Y_i}$ be the empirical distributions of independent observations $X_1, \ldots, X_n \sim \mu$ and $Y_1, \ldots, Y_n \sim \nu$. Suppose that μ satisfies Condition* (4) *ahead* (*which requires μ to be sub-Gaussian*).

(i) (*One-sample null case*) *We have*

$$\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu) \xrightarrow{d} \sup_{\substack{\varphi \in C_0^\infty: \\ \|\varphi\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)} \leq 1}} G_\mu(\varphi),$$

where $G_\mu = (G_\mu(\varphi))_{\varphi \in C_0^\infty}$ is a centered Gaussian process whose paths are linear and continuous with respect to (*w.r.t.*) *the Sobolev seminorm $\|\varphi\|_{\dot{H}^{1,q}(\mu*\gamma_\sigma)} := \|\nabla\varphi\|_{L^q(\mu*\gamma_\sigma;\mathbb{R}^d)}$. Here q is the conjugate index of p, that is, $1/p + 1/q = 1$.*

(ii) (*Two-sample null case*) *If* $\nu = \mu$, *then we have*

$$\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) \xrightarrow{d} \sup_{\substack{\varphi \in C_0^\infty: \\ \|\varphi\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)} \le 1}} [G_\mu(\varphi) - G'_\mu(\varphi)],$$

where G'_μ *is an independent copy of* G_μ.

(iii) (*One-sample alternative case*) *If* $\nu \neq \mu$ *and* ν *is sub-Weibull, then we have*

$$\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu) - \mathsf{W}_p^{(\sigma)}(\mu, \nu)) \xrightarrow{d} N\left(0, \frac{\mathrm{Var}_\mu(g * \phi_\sigma)}{p^2[\mathsf{W}_p^{(\sigma)}(\mu, \nu)]^{2(p-1)}}\right),$$

where g *is an optimal transport potential from* $\mu * \gamma_\sigma$ *to* $\nu * \gamma_\sigma$ *for* W_p^p, *and* $\phi_\sigma(x) = (2\pi\sigma^2)^{-d/2}e^{-|x|^2/(2\sigma^2)}$ *is the Gaussian density.*

(iv) (*Two-sample alternative case*) *If* $\nu \neq \mu$ *and* ν *satisfies Condition* (4), *then we have*

$$\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) - \mathsf{W}_p^{(\sigma)}(\mu, \nu)) \xrightarrow{d} N\left(0, \frac{\mathrm{Var}_\mu(g * \phi_\sigma) + \mathrm{Var}_\nu(g^c * \phi_\sigma)}{p^2[\mathsf{W}_p^{(\sigma)}(\mu, \nu)]^{2(p-1)}}\right),$$

where g^c *is the c-transform of g for the cost function* $c(x, y) = |x - y|^p$.

Parts (i) and (ii) show that the null limit distributions are non-Gaussian. On the other hand, Parts (iii) and (iv) establish asymptotic normality of empirical $\mathsf{W}_p^{(\sigma)}$ under the alternative. Notably, these result have the correct centering, $\mathsf{W}_p^{(\sigma)}(\mu, \nu)$, which enables us to construct confidence intervals for $\mathsf{W}_p^{(\sigma)}(\mu, \nu)$.

The proof strategy for Theorem 1.1 differs from existing approaches to limit distribution theory for empirical W_p for general distributions. In fact, an analog of Theorem 1.1 is not known to hold for classic W_p in this generality, except for the special case where μ, ν are discrete (see a literature review below for details). The key insight is to regard $\mathsf{W}_p^{(\sigma)}$ as a functional defined on a subset of a certain dual Sobolev space. We show that the smooth empirical process converges weakly in the dual Sobolev space and that $\mathsf{W}_p^{(\sigma)}$ is Hadamard (directionally) differentiable w.r.t. the dual Sobolev norm. We then employ the extended functional delta method [83, 90] to obtain the limit distribution of one- and two-sample empirical $\mathsf{W}_p^{(\sigma)}$ under both the null and the alternative. The derivation of the alternative limit distribution requires $p > 1$ since we rely on uniqueness (up to additive constants) of OT dual potentials, which does not hold for $p = 1$. As aforementioned, limit distributions for $p = 1$ were derived in [52, 85] via a markedly different proof technique that hinges on the IPM structure of $\mathsf{W}_1^{(\sigma)}$.

The limit distributions in Theorem 1.1 are nonpivotal in the sense that they depend on the population distributions μ and ν, which are unknown in practice. To facilitate statistical inference using $\mathsf{W}_p^{(\sigma)}$, we employ the bootstrap to estimate the limit distributions and prove its consistency for each case of Theorem 1.1. Under the alternative, the consistency follows from the linearity of the Hadamard derivative. Under the null, where the Hadamard (directional) derivative is nonlinear, the bootstrap consistency is not obvious but still holds. This is somewhat surprising in light of [43, 45], where it is demonstrated that the bootstrap, in general, fails to be consistent for functionals whose Hadamard directional derivatives are nonlinear (cf. Proposition 1 in [43] or Corollary 3.1 in [45]). Nevertheless, our application of the bootstrap differs from [43, 45] so there is no contradiction, and the specific structure of the Hadamard derivative of $\mathsf{W}_p^{(\sigma)}$ allows to establish consistency under the null (see the discussion after Proposition 3.8 for more details). These bootstrap consistency results enable constructing confidence intervals for $\mathsf{W}_p^{(\sigma)}$ and using it to test the equality of distributions.

As an application of the limit theory, we study implicit generative modeling under the minimum distance estimation (MDE) framework [78, 80, 104]. MDE extends the maximum-likelihood principle beyond the KL divergence and applies to models supported on low-dimensional manifolds [6] (whence the KL divergence is not well-defined), as well as to cases when the likelihood function is intractable [58]. For MDE with $\mathsf{W}_p^{(\sigma)}$, we establish limit distribution results for the optimal solution and the smooth p-Wasserstein error. Our results hold for arbitrary dimension, again contrasting the classic case where analogous distributional limits for MDE with W_p are known only for $p = d = 1$ [10]. Remarkably, when $p = 2$, the Hilbertian structure of the underlying dual Sobolev space allows showing asymptotic normality of the MDE solution.

1.2. *Literature review.* Analysis of empirical Wasserstein distances, or more generally empirical optimal transport distances, has been an active research area in the statistics and probability theory literature. In particular, significant attention was devoted to rates of convergence and exact asymptotics [2, 5, 8, 14, 16–18, 23, 29, 37, 38, 40, 46, 67, 69, 71, 72, 93, 94, 102]. As noted before, the empirical Wasserstein distance suffers from the curse of dimensionality, namely, $\mathbb{E}[\mathsf{W}_p(\hat{\mu}_n, \mu)] = O(n^{-1/d})$ whenever $d > 2p$. This rate is known to be sharp in general [40]. The recent work by [23, 72] discovered that the rate can be improved under the alternative, namely, $\mathbb{E}[|\mathsf{W}_p(\hat{\mu}_n, \nu) - \mathsf{W}_p(\mu, \nu)|] = O(n^{-\alpha/d})$ for $d \geq 5$ if $\nu \neq \mu$, where $\alpha = p$ for $1 \leq p < 2$ and $\alpha = 2$ for $2 \leq p < \infty$. Their insight is to use the duality formula for W_p^p and exploit regularity of optimal transport potentials. [72] also derive matching minimax lower bounds up to log factors under some technical conditions.

Another central problem that has seen a rapid development is limit distribution theory for empirical Wasserstein distances. However, except for the two special cases discussed next, to the best of our knowledge, there is no proven analog of our Theorem 1.1 for classic Wasserstein distances, that is, a comprehensive limit distribution theory for empirical W_p that holds for general d and p. The first case for which the limit distribution is well understood is when $d = 1$. Then, W_p reduces to the L^p distance between quantile functions for $1 \leq p < \infty$, and further simplifies to the L^1 distance between distribution functions when $p = 1$. Building on such explicit expressions, [30] and [31] derived null limit distributions in $d = 1$ for $p = 1$ and $p = 2$, respectively. More recently, under the alternative ($\mu \neq \nu$), [35] derived a central limit theorem (CLT) when $d = 1$ and $p \geq 2$. The second case where a limit distribution theory for empirical W_p is available is when μ, ν are discrete. If the distributions are finitely discrete, that is, $\mu = \sum_{j=1}^m r_j \delta_{x_j}$ and $\nu = \sum_{j=1}^k s_j \delta_{y_j}$ for two simplex vectors $r = (r_1, \dots, r_m)$ and $s = (s_1, \dots, s_k)$, then $\mathsf{W}_p(\mu, \nu)$ can be seen as a function of those simplex vectors r and s. Leveraging this, [92] applied the delta method to obtain limit distributions for empirical W_p in the finitely discrete case. An extension to countably infinite supports was provided in [95], while [32] treated the semidiscrete case where μ is finitely discrete but ν is general.

Except for these two special cases, limit distributions for Wasserstein distances are less understood. To avoid repetitions, we focus here our discussion on the one sample case. In [36], a CLT for $\sqrt{n}(\mathsf{W}_2^2(\hat{\mu}_n, \nu) - \mathbb{E}[\mathsf{W}_2^2(\hat{\mu}_n, \nu)])$ is derived in any dimension, but the limit Gaussian distribution degenerates to 0 when $\mu = \nu$; see also [34] for an extension to general $1 < p < \infty$. Notably, the centering constant there is the expected empirical Wasserstein distance $\mathbb{E}[\mathsf{W}_2^2(\hat{\mu}_n, \nu)]$, which in general can not be replaced with the (more natural) population distance $\mathsf{W}_2^2(\mu, \nu)$. The recent preprint [71] addressed this gap and established a CLT for $\sqrt{n}(\mathsf{W}_2^2(\tilde{\mu}_n, \nu) - \mathsf{W}_2^2(\mu, \nu))$ for a wavelet-based estimator $\tilde{\mu}_n$ of μ, assuming that the ambient space is $[0, 1]^d$ and that μ, ν are absolutely continuous w.r.t. the Lebesgue measure with smooth and strictly positive densities. Following arguments similar to [36], they first derive a CLT for $\sqrt{n}(\mathsf{W}_2^2(\tilde{\mu}_n, \nu) - \mathbb{E}[\mathsf{W}_2^2(\tilde{\mu}_n, \nu)])$ and then use the strict positivity of the densities and higher order regularity of optimal transport potentials to control the bias term as $\mathbb{E}[\mathsf{W}_2^2(\tilde{\mu}_n, \nu)] - \mathsf{W}_2^2(\mu, \nu) = o(n^{-1/2})$.

Our proof techniques differ from the aforementioned arguments for classic W_p. Specifically, as opposed to the two-step approach of [71] described above, we directly prove asymptotic normality for $\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n,\nu)-\mathsf{W}_p^{(\sigma)}(\mu,\nu))$ under the alternative. Their derivation does not apply to our case even when $p=2$ since their bias bound requires that the densities of μ and ν be bounded away from zero on their (compact) supports, which fails to hold after the Gaussian convolution. Our argument also differs from that of [92, 95], even though they also rely on the functional delta method. Specifically, since we do not assume that μ, ν are discrete, $\mathsf{W}_p^{(\sigma)}$ can not be parameterized by simplex vectors, and hence the application of the functional delta method is nontrivial. Very recently, an independent work [61] used the extended functional delta method for the supremum functional [19] to derive limit distributions for classic W_p, with $p\geq 2$, for compactly supported distributions under the alternative in dimensions $d\leq 3$.[1]

Finally, we briefly compare the smooth Wasserstein distance with entropic regularized OT (EOT) [3, 27]. EOT enjoys fast computational methods and a similar statistical profile to that of $\mathsf{W}_p^{(\sigma)}$, in terms of parametric convergence rates [49, 73] and limit distributions [11, 33, 55–57, 65, 73], but it forfeits the Wasserstein metric and topological structure. Indeed, EOT is not a metric even for distance-like costs $c(x,y)=|x-y|^p$ for $p\in[1,\infty)$, which makes it less compatible for applications like testing or MDE.[2]

1.3. *Organization.* The rest of the paper is organized as follows. In Section 2, we collect background material on Wasserstein distances, smooth Wasserstein distances, and dual Sobolev spaces. In Section 3, we prove Theorem 1.1 and explore the validity of the bootstrap for empirical $\mathsf{W}_p^{(\sigma)}$. Section 4 presents applications of our limit distribution theory to MDE with $\mathsf{W}_p^{(\sigma)}$. Proofs for Section 3 and 4 can be found in Section 5. Section 6 provides concluding remarks and discusses future research directions. Finally, the supplemental material contains additional proofs.

1.4. *Notation.* Let $|\cdot|$ and $\langle\cdot,\cdot\rangle$ denote the Euclidean norm and inner product, respectively. Let $B(x,r)=\{y\in\mathbb{R}^d:|y-x|\leq r\}$ denote the closed ball with center x and radius r. Given a finite signed Borel measure ℓ on $\mathbb{R}^d$, we identify ℓ with the linear functional $f\mapsto \ell(f):=\int f\,d\ell$. Let $\lesssim$ denote inequalities up to some numerical constants. For any $a,b\in\mathbb{R}$, we use the shorthands $a\vee b=\max\{a,b\}$ and $a\wedge b=\min\{a,b\}$.

For a topological space S, $\mathcal{B}(S)$ and $\mathcal{P}(S)$ denote, respectively, the Borel σ-field on S and the class of Borel probability measures on S. We write $\mathcal{P}:=\mathcal{P}(\mathbb{R}^d)$ and for $1\leq p<\infty$, use $\mathcal{P}_p$ to denote the subset of $\mu\in\mathcal{P}$ with finite pth moment $\int_{\mathbb{R}^d}|x|^p\,d\mu(x)<\infty$. We use $*$ to denote the convolution. Let $\stackrel{w}{\to}$, $\stackrel{d}{\to}$, and $\stackrel{\mathbb{P}}{\to}$ denote weak convergence of probability measures, convergence in distribution of random variables, and convergence in probability, respectively. When necessary, convergence in distribution is understood in the sense of Hoffmann–Jørgensen (cf. Chapter 2 in [99]).

Throughout, we assume that $(X_1,Y_1),(X_2,Y_2),\ldots$ are the coordinate projections of the product probability space $\prod_{i=1}^{\infty}(\mathbb{R}^{2d},\mathcal{B}(\mathbb{R}^{2d}),\mu\otimes\nu)$. To generate auxiliary random variables, we extend the probability space as $(\Omega,\mathcal{A},\mathbb{P})=[\prod_{i=1}^{\infty}(\mathbb{R}^{2d},\mathcal{B}(\mathbb{R}^{2d}),\mu\otimes\nu)]\times([0,1],\mathcal{B}([0,1]),\mathrm{Leb})$, where Leb denotes the Lebesgue measure on $[0,1]$. For $\beta\in(0,2]$, let $\psi_\beta(t)=e^{t^\beta}-1$ for $t\geq 0$, and recall that the corresponding Orlicz (quasi-)norm of a real-valued random variable ξ is defined as $\|\xi\|_{\psi_\beta}:=\inf\{C>0:\mathbb{E}[\psi_\beta(|\xi|/C)]\leq 1\}$. A Borel

[1][61] was posted on arXiv after the present paper was submitted to the journal.

[2]EOT between μ and itself does not nullify. While this issue can be corrected by considering the (centered) Sinkhorn divergence, it still is not a metric since it lacks the triangle inequality [11].

probability measure $\mu \in \mathcal{P}$ is called *β-sub-Weibull* if $\|\|X\|\|_{\psi_\beta} < \infty$ for $X \sim \mu$. We say that μ is sub-Weibull if it is β-sub-Weibull for some $\beta \in (0, 2]$. Finally, μ is *sub-Gaussian* if it is 2-sub-Weibull.

For an open set $\mathcal{O}$ in a Euclidean space, $C_0^\infty(\mathcal{O})$ denotes the space of compactly supported, infinitely differentiable, real functions on $\mathcal{O}$. We write $C_0^\infty = C_0^\infty(\mathbb{R}^d)$ and define $\dot{C}_0^\infty = \{f + a : f \in C_0^\infty, a \in \mathbb{R}\}$. For any $p \in [1, \infty)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$, let $L^p(\mu; \mathbb{R}^k)$ denote the space of measurable maps $f : \mathbb{R}^d \to \mathbb{R}^k$ such that $\|f\|_{L^p(\mu;\mathbb{R}^k)} = (\int_{\mathbb{R}^d} |f|^p \, d\mu)^{1/p} < \infty$; when $d = 1$ we use the shorthand $L^p(\mu) = L^p(\mu; \mathbb{R}^1)$. Recall that $(L^p(\mu; \mathbb{R}^k), \|\cdot\|_{L^p(\mu;\mathbb{R}^k)})$ is a Banach space. Finally, for a subset A of a topological space S, let $\overline{A}^S$ denote the closure of A; if the space S is clear from the context, then we simply write $\overline{A}$ for the closure.

2. Background.

2.1. *Wasserstein distances and their smooth variants.* Recall that, for $1 \le p < \infty$, the p-Wasserstein distance $\mathsf{W}_p(\mu, \nu)$ between $\mu, \nu \in \mathcal{P}_p$ is defined in (1). Some basic properties of W_p are (cf. e.g., [4, 87, 100, 101]): (i) the inf is attained in the definition of W_p, that is, there exists a coupling $\pi^\star \in \Pi(\mu, \nu)$ such that $\mathsf{W}_p^p(\mu, \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, d\pi^\star(x, y)$, and the optimal coupling $\pi^\star$ is unique when $p > 1$ and $\mu \ll dx$; (ii) W_p is a metric on $\mathcal{P}_p$; and (iii) convergence in W_p is equivalent to weak convergence plus convergence of pth moments: $\mathsf{W}_p(\mu_n, \mu) \to 0$ if and only if $\mu_n \stackrel{w}{\to} \mu$ and $\int |x|^p \, d\mu_n(x) \to \int |x|^p \, d\mu(x)$.

The proof of the limit distribution for empirical $\mathsf{W}_p^{(\sigma)}$ under the alternative hinges on duality theory for W_p, which we summarize below. For a function $g : \mathbb{R}^d \to [-\infty, \infty)$ and a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, the *$c$-transform* of g is defined by

$$g^c(y) = \inf_{x \in \mathbb{R}^d} \big[c(x, y) - g(x)\big], \quad y \in \mathbb{R}^d.$$

A function $g : \mathbb{R}^d \to [-\infty, \infty)$, not identically $-\infty$, is called *c-concave* if $g = f^c$ for some function $f : \mathbb{R}^d \to [-\infty, \infty)$.

LEMMA 2.1 (Duality for W_p). *Let $1 \le p < \infty$, $\mu, \nu \in \mathcal{P}_p$, and set the cost function to $c(x, y) = |x - y|^p$.*

(i) (*Theorem* 5.9 *in* [101]; *Theorem* 6.1.5 *in* [4]) *The following duality holds,*

$$\mathsf{W}_p^p(\mu, \nu) = \sup_{g \in L^1(\mu)} \left[\int_{\mathbb{R}^d} g \, d\mu + \int_{\mathbb{R}^d} g^c \, d\nu\right], \tag{2}$$

and there is at least one c-concave function $g \in L^1(\mu)$ that attains the supremum in (2); *we call this g an optimal transport potential from μ to ν for W_p^p.*

(ii) (*Theorem* 3.3 *in* [47]) *Let $1 < p < \infty$, suppose that $g : \mathbb{R}^d \to [-\infty, \infty)$ is c-concave, and take K as the convex hull of $\{x : g(x) > -\infty\}$. Then g is locally Lipschitz on the interior of K.*

(iii) (*Corollary* 2.7 *in* [34]) *If $1 < p < \infty$ and $\mu \ll dx$ is supported on an open connected set A, then the optimal transport potential from μ to ν for W_p^p is unique on A up to additive constants, that is, if g_1 and g_2 are optimal transport potentials, then there exists $C \in \mathbb{R}$ such that $g_1(x) = g_2(x) + C$ for all $x \in A$.*

The *smooth Wasserstein distance* convolves the distributions with an isotropic Gaussian kernel. Gaussian convolution levels out local irregularities in the distributions, while largely preserving the structure of classic W_p. Recalling that $\gamma_\sigma = N(0, \sigma^2 I_d)$, the smooth p-Wasserstein distance is defined as follows.

DEFINITION 2.1 (Smooth Wasserstein distance). Let $1 \leq p < \infty$ and $\sigma \geq 0$. For $\mu, \nu \in \mathcal{P}_p$, the *smooth p-Wasserstein distance* between μ and ν with smoothing parameter σ is

$$\mathsf{W}_p^{(\sigma)}(\mu, \nu) := \mathsf{W}_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma).$$

The smooth Wasserstein distance was studied in [51–53, 77, 85] for structural properties and empirical convergence rates. We recall two basic properties: (i) $\mathsf{W}_p^{(\sigma)}$ is a metric on $\mathcal{P}_p$ that generates the same topology as classic W_p; (ii) for $\mu, \nu \in \mathcal{P}_p$ and $0 \leq \sigma_1 \leq \sigma_2 < \infty$, we have $\mathsf{W}_p^{(\sigma_2)}(\mu, \nu) \leq \mathsf{W}_p^{(\sigma_1)}(\mu, \nu) \leq \mathsf{W}_p^{(\sigma_2)}(\mu, \nu) + C_{p,d}\sqrt{\sigma_2^2 - \sigma_1^2}$ for a constant $C_{p,d}$ that depends only on p, d. In particular, $\mathsf{W}_p^{(\sigma)}(\mu, \nu)$ is continuous and monotonically nonincreasing in $\sigma \in [0, +\infty)$ with $\lim_{\sigma \downarrow 0} \mathsf{W}_p^{(\sigma)}(\mu, \nu) = \mathsf{W}_p(\mu, \nu)$. See [77] for additional structural results, including an explicit expression for $C_{p,d}$ and weak convergence of smooth optimal couplings. For empirical convergence, it was shown in [77] that under appropriate moment assumptions $\mathbb{E}[\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu)] = O(n^{-1/2})$ for $p > 1$ in any dimension d. Versions of this result for $p = 1$ and $p = 2$ were derived earlier in [52, 53, 85].

2.2. *Sobolev spaces and their duals.* Our proof strategy for the limit distribution results is to regard W_p as a functional defined on a subset of a certain dual Sobolev space. We will show that the smooth empirical process converges weakly in the dual Sobolev space and that W_p is Hadamard (directionally) differentiable w.r.t. the dual Sobolev norm. Given these, the limit distributions in Theorem 1.1 follow via the functional delta method. Here we briefly discuss (homogeneous) Sobolev spaces and their duals.

DEFINITION 2.2 (Sobolev spaces and their duals). Let $\rho \in \mathcal{P}$ and $1 \leq p < \infty$.

(i) For a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, let

$$\|f\|_{\dot{H}^{1,p}(\rho)} := \|\nabla f\|_{L^p(\rho;\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} |\nabla f|^p \, d\rho\right)^{1/p}$$

be the *Sobolev seminorm*. We define the *homogeneous Sobolev space* $\dot{H}^{1,p}(\rho)$ by the completion of $\dot{C}_0^\infty$ w.r.t. $\|\cdot\|_{\dot{H}^{1,p}(\rho)}$.

(ii) Let q be the conjugate index of p, that is, $1/p + 1/q = 1$. Let $\dot{H}^{-1,p}(\rho)$ denote the topological dual of $\dot{H}^{1,q}(\rho)$. The *dual Sobolev norm* $\|\cdot\|_{\dot{H}^{-1,p}(\rho)}$ (dual to $\|\cdot\|_{\dot{H}^{1,q}(\rho)}$) of a continuous linear functional $\ell : \dot{H}^{1,q}(\rho) \to \mathbb{R}$ is defined by

$$\|\ell\|_{\dot{H}^{-1,p}(\rho)} = \sup\{\ell(f) : f \in \dot{C}_0^\infty, \|f\|_{\dot{H}^{1,q}(\rho)} \leq 1\}.$$

The restriction $f \in \dot{C}_0^\infty$ can be replaced with $f \in C_0^\infty$ in the definition of the dual norm $\|\cdot\|_{\dot{H}^{-1,p}(\rho)}$ since $\ell(f + a) = \ell(f)$ for any $\ell \in \dot{H}^{-1,p}(\rho)$.

We have defined the homogeneous Sobolev space $\dot{H}^{1,p}(\rho)$ as the completion of $\dot{C}_0^\infty$ w.r.t. $\|\cdot\|_{\dot{H}^{1,p}(\rho)}$. It is not immediately clear that the so-constructed space is a function space over $\mathbb{R}^d$. Below we present an explicit construction of $\dot{H}^{1,p}(\rho)$ when $d\rho/d\kappa$ is bounded away from zero for some reference measure $\kappa \gg dx$ satisfying the p-Poincaré inequality. To that end, we first define the Poincaré inequality.

DEFINITION 2.3 (Poincaré inequality). For $1 \leq p < \infty$, a probability measure $\kappa \in \mathcal{P}$ is said to satisfy the *p-Poincaré inequality* if there exists a finite constant C such that

$$\|\varphi - \kappa(\varphi)\|_{L^p(\kappa)} \leq \mathsf{C}\|\nabla\varphi\|_{L^p(\kappa;\mathbb{R}^d)}, \quad \forall \varphi \in C_0^\infty.$$

The smallest constant satisfying the above is denoted by $\mathsf{C}_p(\kappa)$.

The standard Poincaré inequality refers to the 2-Poincaré inequality. It is known that any log-concave distribution (i.e., a distribution κ of the form $d\kappa = e^{-V}\, dx$ for some convex function $V : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$; cf. [70, 89]) satisfies the p-Poincaré inequality for any $1 \le p < \infty$ [13, 74]. In particular, the Gaussian distribution γ_σ satisfies every p-Poincaré inequality (see also [15], Corollary 1.7.3).

REMARK 2.1 (Explicit construction of $\dot{H}^{1,p}(\rho)$). Suppose that there exists a reference measure $\kappa \in \mathcal{P}$, with $\kappa \gg dx$, that satisfies the p-Poincaré inequality. Assume that $d\rho/d\kappa \ge c$ for some constant $c > 0$ (in our applications, $\rho = \gamma_\sigma$ or $\mu * \gamma_\sigma$ for some $\mu \in \mathcal{P}_p$; in either case, the stated assumption is satisfied with $\kappa = \gamma_\sigma$ or $\gamma_{\sigma/\sqrt{2}}$). Let $\mathcal{C} = \{f \in \dot{C}_0^\infty : \kappa(f) = 0\}$. Then, $\|\cdot\|_{\dot{H}^{1,p}(\rho)}$ is a proper norm on $\mathcal{C}$, and the map $\iota : f \mapsto \nabla f$ is an isometry from $(\mathcal{C}, \|\cdot\|_{\dot{H}^{1,p}(\rho)})$ into $(L^p(\rho; \mathbb{R}^d), \|\cdot\|_{L^p(\rho;\mathbb{R}^d)})$. Let V be the closure of $\iota\mathcal{C}$ in $L^p(\rho; \mathbb{R}^d)$ under $\|\cdot\|_{L^p(\rho;\mathbb{R}^d)}$. The inverse map $\iota^{-1} : \iota\mathcal{C} \to \mathcal{C}$ can be extended to V as follows. For any $g \in V$, choose $f_n \in \mathcal{C}$ such that $\|\nabla f_n - g\|_{L^p(\rho;\mathbb{R}^d)} \to 0$. Since ∇f_n is Cauchy in $L^p(\rho; \mathbb{R}^d)$ and thus in $L^p(\kappa; \mathbb{R}^d)$ (as $\|\cdot\|_{L^p(\kappa;\mathbb{R}^d)} \lesssim \|\cdot\|_{L^p(\rho;\mathbb{R}^d)}$), f_n is Cauchy in $L^p(\kappa)$ by the p-Poincaré inequality, so $\|f_n - f\|_{L^p(\kappa)} \to 0$ for some $f \in L^p(\kappa)$. Set $\iota^{-1} g = f$ and extend $\|\cdot\|_{\dot{H}^{1,p}(\rho)}$ by $\|f\|_{\dot{H}^{1,p}(\rho)} = \lim_{n\to\infty} \|f_n\|_{\dot{H}^{1,p}(\rho)} = \|g\|_{L^p(\rho;\mathbb{R}^d)}$. The space $(\iota^{-1}V, \|\cdot\|_{\dot{H}^{1,p}(\rho)})$ is a Banach space of functions over $\mathbb{R}^d$. Finally, the homogeneous Sobolev space $\dot{H}^{1,p}(\rho)$ can be constructed as $\dot{H}^{1,p}(\rho) = \{f + a : a \in \mathbb{R}, f \in \iota^{-1}V\}$ with $\|f + a\|_{\dot{H}^{1,p}(\rho)} = \|f\|_{\dot{H}^{1,p}(\rho)}$.

The next lemma summarizes some basic results about the space $\dot{H}^{-1,p}(\rho)$ and $\dot{H}^{-1,p}(\rho)$-valued random variables that we use in the sequel. The proof can be found in Section 1 of the Supplementary Material [54].

LEMMA 2.2. *Let $1 < p < \infty$ and $\rho \in \mathcal{P}$. The dual space $\dot{H}^{-1,p}(\rho)$ is a separable Banach space. The Borel σ-field on $\dot{H}^{-1,p}(\rho)$ coincides with the cylinder σ-field (the smallest σ-field that makes the coordinate projections, $\dot{H}^{-1,p}(\rho) \ni \ell \mapsto \ell(f) \in \mathbb{R}$, measurable).*

Consider a stochastic process $Y = (Y(f))_{f\in\dot{H}^{1,q}(\rho)}$ indexed by $\dot{H}^{1,q}(\rho)$, that is, $\omega \mapsto Y(f, \omega)$ is measurable for each $f \in \dot{H}^{1,q}(\rho)$. The process can be thought of as a map from Ω into $\dot{H}^{-1,p}(\rho)$ as long as Y has paths in $\dot{H}^{-1,p}(\rho)$, that is, for each fixed $\omega \in \Omega$, the map $f \mapsto Y(f, \omega)$ is continuous and linear. The fact that the Borel σ-field on $\dot{H}^{-1,p}(\rho)$ coincides with the cylinder σ-field guarantees that a stochastic process indexed by $\dot{H}^{1,q}(\rho)$ with paths in $\dot{H}^{-1,p}(\rho)$ is Borel measurable as a map from Ω into $\dot{H}^{-1,p}(\rho)$.

2.3. W_p *and dual Sobolev norm.* In Section 3, we will explore limit distributions for empirical $\mathsf{W}_p^{(\sigma)}$. One of the key technical ingredients there is a comparison of the Wasserstein distance with a certain dual Sobolev norm, which we present next.

PROPOSITION 2.1 (Comparison between W_p and dual Sobolev norm; Theorem 5.26 in [39]). *Let $1 < p < \infty$, and suppose that $\mu_0, \mu_1 \in \mathcal{P}_p$ with $\mu_0, \mu_1 \ll \rho$ for some reference measure $\rho \in \mathcal{P}$. Denote their respective densities by $f_i = d\mu_i/d\rho$, $i = 0, 1$. If f_0 or f_1 is bounded from below by some $c > 0$, then*

$$\mathsf{W}_p(\mu_0, \mu_1) \le p c^{-1/q} \|\mu_1 - \mu_0\|_{\dot{H}^{-1,p}(\rho)}. \tag{3}$$

REMARK 2.2. If ρ satisfies the q-Poincaré inequality, then for every $\varphi \in C_0^\infty$ with $\|\varphi\|_{\dot{H}^{1,q}(\rho)} \leq 1$, we have

$$\begin{aligned}\int \varphi(f_1 - f_0)\, d\rho &= \int (\varphi - \rho(\varphi))(f_1 - f_0)\, d\rho \\ &\leq \|\varphi - \rho(\varphi)\|_{L^q(\rho)} \|f_1 - f_0\|_{L^p(\rho)} \\ &\leq \mathsf{C}_q(\rho)\|\nabla\varphi\|_{L^q(\rho;\mathbb{R}^d)} \|f_1 - f_0\|_{L^p(\rho)} \leq \mathsf{C}_q(\rho)\|f_1 - f_0\|_{L^p(\rho)},\end{aligned}$$

so that $\mathsf{W}_p(\mu_0, \mu_1) \leq pc^{-1/q}\mathsf{C}_q(\rho)\|f_1 - f_0\|_{L^p(\rho)}$.

Proposition 2.1 follows directly from Theorem 5.26 of [39]. Similar comparison inequalities appear in [67, 79, 103]. We include a self-contained proof of Proposition 2.1 in Section 2 of the supplementary material [54] as some elements of the proof are key to our derivation of the null limit distribution for empirical $\mathsf{W}_p^{(\sigma)}$. The proof builds on the Benamou–Brenier dynamic formulation of optimal transport [9], which shows that $\mathsf{W}_p(\mu_0, \mu_1)$ is bounded from above by the length of any absolutely continuous path from μ_0 to μ_1 in $(\mathcal{P}_p, \mathsf{W}_p)$. The dual Sobolev norm emerges as a bound on the length of the linear interpolation $t \mapsto t\mu_1 + (1-t)\mu_0$.

3. Limit distribution theory. The goal of this section is to establish Theorem 1.1. The proof relies on two key steps: (i) establish weak convergence of the smooth empirical process $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ in the dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$; and (ii) regard $\mathsf{W}_p^{(\sigma)}$ as a functional defined on a subset of $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ and characterize its Hadamard directional derivative w.r.t. the corresponding dual Sobolev norm. Given (i) and (ii), the limit distribution results follow from the functional delta method, and the asymptotic normality under the alternative further follows from linearity of the Hadamard directional derivative.

3.1. *Preliminaries.* Throughout this section, we fix $1 < p < \infty$, take q as the conjugate index of p, and let $\sigma > 0$. For $\mu, \nu \in \mathcal{P}_p$, let $X_1, \dots, X_n \sim \mu$ and $Y_1, \dots, Y_n \sim \nu$ be independent observations and denote the associated empirical distributions by $\hat{\mu}_n := n^{-1}\sum_{i=1}^n \delta_{X_i}$ and $\hat{\nu}_n := n^{-1}\sum_{i=1}^n \delta_{Y_i}$, respectively.

3.1.1. *Weak convergence of smooth empirical process in dual Sobolev spaces.* The first building block of our limit distribution results is the following weak convergence of the smoothed empirical process $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$.

PROPOSITION 3.1 (Weak convergence of smooth empirical process). *Suppose that $X \sim \mu$ satisfies*

$$\int_0^\infty e^{\frac{\theta r^2}{2\sigma^2}} \sqrt{\mathbb{P}(|X| > r)}\, dr < \infty \quad \text{for some } \theta > p - 1. \tag{4}$$

Then, the smoothed empirical process $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ converges in distribution in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ as $n \to \infty$. The limit process is a centered Gaussian process indexed by $\dot{H}^{1,q}(\mu * \gamma_\sigma)$ with covariance function $(f_1, f_2) \mapsto \mathrm{Cov}_\mu(f_1 * \phi_\sigma, f_2 * \phi_\sigma)$. Here Cov_μ denotes the covariance under μ.

The proof of Proposition 3.1 relies on the prior work [77] by a subset of the authors, where it was shown that the smoothed function class $\mathcal{F} * \phi_\sigma = \{f * \phi_\sigma : f \in \mathcal{F}\}$ with $\mathcal{F} = \{f \in \dot{C}_0^\infty : \|f\|_{\dot{H}^{1,q}(\gamma_\sigma)} \leq 1\}$ is μ-Donsker. We then prove the weak convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$ following a similar argument to Lemma 1 in [76]. This, in turn, implies weak convergence in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ when μ has mean zero, since in that case $\dot{H}^{-1,p}(\gamma_\sigma)$ is continuously embedded into $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$. To account for non-centered distributions, we use a reduction to the mean zero case via translation. See also Remark 5.2 for an alternative proof for $p = 2$ that relies on the CLT in the Hilbert space.

Inspection of the proof of Proposition 3.1 shows that Condition (4) implies

$$\int_{\mathbb{R}^d} e^{(p-1)|x|^2/\sigma^2}\, d\mu(x) < \infty, \tag{5}$$

which requires μ to be sub-Gaussian; see Remark 5.1 for details. It is not difficult to see that Condition (4) is satisfied if μ is compactly supported or sub-Gaussian with $\||X\||_{\psi_2} < \sigma/\sqrt{p-1}$ for $X \sim \mu$, and that the condition fails to hold for $X \sim \gamma_{\sigma/\sqrt{p-1}}$ (which instead clearly satisfies (5)).

A natural question is whether a condition in the spirit of (4) is necessary for the conclusion of Proposition 3.1 to hold. Indeed, we show that some form of sub-Gaussianity is necessary for the smooth empirical process to converge to zero in $\dot{H}^{-1,p}(\gamma_\sigma)$.

PROPOSITION 3.2 (Necessity of sub-Gaussian condition). *The following hold.*

(i) *If* $(\hat{\mu}_n - \mu) * \gamma_\sigma \to 0$ *in* $\dot{H}^{-1,p}(\gamma_\sigma)$ *as* $n \to \infty$ *a.s., then* $\int_{\mathbb{R}^d} e^{\theta|x|^2/(2\sigma^2)}\, d\mu(x) < \infty$ *for any* $\theta < p - 1$.

(ii) *Conversely, if* $\int_{\mathbb{R}^d} e^{(p-1)|x|^2/(2\sigma^2)}\, d\mu(x) < \infty$*, then* $(\hat{\mu}_n - \mu) * \gamma_\sigma \to 0$ *in* $\dot{H}^{-1,p}(\gamma_\sigma)$ *as* $n \to \infty$ *a.s.*

3.1.2. *Functional delta method.* Another ingredient of our limit distribution results is the (extended) functional delta method [43, 45, 83, 91]. Let D be a normed space and $\Phi : \Xi \subset D \to \mathbb{R}$ be a function. Following [83, 90], we say that Φ is *Hadamard directionally differentiable* at $\theta \in \Xi$ if there exists a map $\Phi'_\theta : T_\Xi(\theta) \to \mathbb{R}$ such that

$$\lim_{n\to\infty} \frac{\Phi(\theta + t_n h_n) - \Phi(\theta)}{t_n} = \Phi'_\theta(h)$$

for any $h \in T_\Xi(\theta)$, $t_n \downarrow 0$, and $h_n \to h$ in D such that $\theta + t_n h_n \in \Xi$. Here $T_\Xi(\theta)$ is the *tangent cone* to Ξ at θ defined as

$$T_\Xi(\theta) := \left\{ h \in D : h = \lim_{n\to\infty} \frac{\theta_n - \theta}{t_n} \text{ for some } \theta_n \to \theta \text{ in } \Xi \text{ and } t_n \downarrow 0 \right\}.$$

The tangent cone $T_\Xi(\theta)$ is closed, and if Ξ is convex, then $T_\Xi(\theta)$ coincides with the closure in D of $\{(\tilde{\theta} - \theta)/t : \tilde{\theta} \in \Xi, t > 0\}$ (cf. Proposition 4.2.1 in [7]). The derivative Φ'_θ is positively homogeneous (i.e., $\Phi'_\theta(ch) = c\Phi'_\theta(h)$ for any $c \geq 0$) and continuous, but need not be linear.

LEMMA 3.1 (Extended functional delta method [43, 45, 83, 91]). *Let D be a normed space and $\Phi : \Xi \subset D \to \mathbb{R}$ be a function that is Hadamard directionally differentiable at $\theta \in \Xi$ with derivative $\Phi'_\theta : T_\Xi(\theta) \to \mathbb{R}$. Let $T_n : \Omega \to \Xi$ be maps such that $r_n(T_n - \theta) \stackrel{d}{\to} T$ for some $r_n \to \infty$ and Borel measurable map $T : \Omega \to D$ with values in $T_\Xi(\theta)$. Then, $r_n(\Phi(T_n) - \Phi(\theta)) \stackrel{d}{\to} \Phi'_\theta(T)$. Further, if Ξ is convex, then we have the expansion $r_n(\Phi(T_n) - \Phi(\theta)) = \Phi'_\theta(r_n(T_n - \theta)) + o_{\mathbb{P}}(1)$.*

REMARK 3.1 (Choice of domain Ξ). The domain Ξ is arbitrary as long as it contains the ranges of T_n for all n, and the tangent cone $T_\Xi(\theta)$ contains the range of the limit variable T.

3.2. *Limit distributions under the null* ($\mu = \nu$). We shall apply the extended functional delta method to derive the limit distributions of $\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu)$ and $\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n)$ as $n \to \infty$, namely, proving Parts (i) and (ii) of Theorem 1.1. To set up the problem over a (real) vector space, we regard $\rho \mapsto \mathsf{W}_p^{(\sigma)}(\rho, \mu) = \mathsf{W}_p(\rho * \gamma_\sigma, \mu * \gamma_\sigma)$ as a map $h \mapsto \mathsf{W}_p(\mu * \gamma_\sigma + h, \mu * \gamma_\sigma)$ defined on a set of finite signed Borel measures. The comparison result from Proposition 2.1 implies that the latter map is Lipschitz in $\|\cdot\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$, and Proposition 3.1 shows that $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ is weakly convergent in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$. These suggest choosing the ambient space to be $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$.

To cover the one- and two-sample cases in a unified manner, consider the same map but in two variables. Take $D_\mu = \dot{H}^{-1,p}(\mu * \gamma_\sigma)$, set $\Xi_\mu := D_\mu \cap \{h = (\rho - \mu) * \gamma_\sigma : \rho \in \mathcal{P}_p\}$, and define the function $\Phi : \Xi_\mu \times \Xi_\mu \subset D_\mu \times D_\mu \to \mathbb{R}$ as

$$\Phi(h_1, h_2) := \mathsf{W}_p(\mu * \gamma_\sigma + h_1, \mu * \gamma_\sigma + h_2), \quad (h_1, h_2) \in \Xi_\mu \times \Xi_\mu.$$

We endow $D_\mu \times D_\mu$ with a product norm (e.g., $\|h_1\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + \|h_2\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$). Since the set Ξ_μ (and thus $\Xi_\mu \times \Xi_\mu$) is convex, the tangent cone $T_{\Xi_\mu \times \Xi_\mu}(0, 0)$ coincides with the closure in $D_\mu \times D_\mu$ of $\{(h_1, h_2)/t : (h_1, h_2) \in \Xi_\mu \times \Xi_\mu, t > 0\}$. We next verify that Φ is Hadamard directionally differentiable at $(0, 0)$.

PROPOSITION 3.3 (Hadamard directional derivative of W_p under the null). *Let $1 < p < \infty$ and $\mu \in \mathcal{P}_p$. Then, the map $\Phi : (h_1, h_2) \mapsto \mathsf{W}_p(\mu * \gamma_\sigma + h_1, \mu * \gamma_\sigma + h_2)$, $\Xi_\mu \times \Xi_\mu \subset D_\mu \times D_\mu \to \mathbb{R}$, is Hadamard directionally differentiable at $(h_1, h_2) = (0, 0)$ with derivative $\Phi'_{(0,0)}(h_1, h_2) = \|h_1 - h_2\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$, that is, for any $(h_1, h_2) \in T_{\Xi_\mu \times \Xi_\mu}(0)$, $t_n \downarrow 0$ and $(h_{n,1}, h_{n,2}) \to (h_1, h_2)$ in $D_\mu \times D_\mu$ such that $(t_n h_{n,1}, t_n h_{n,2}) \in \Xi_\mu \times \Xi_\mu$, we have*

$$\lim_{n\to\infty} \frac{\Phi(t_n h_{n,1}, t_n h_{n,2})}{t_n} = \|h_1 - h_2\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}.$$

Proposition 3.3 follows from the next Gâteaux differentiability result for W_p, which may be of independent interest, combined with Lipschitz continuity of Φ w.r.t. $\|\cdot\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ (cf. Proposition 2.1).

LEMMA 3.2 (Gâteaux directional derivative of W_p). *Let $\mu \in \mathcal{P}_p$ and $h_i \in \dot{H}^{-1,p}(\mu)$, $i = 1, 2$ be finite signed Borel measures with total mass 0 such that $h_i \ll \mu$ and $\mu + h_i \in \mathcal{P}_p$. Then,*

$$\frac{d}{dt^+}\mathsf{W}_p(\mu + th_1, \mu + th_2)|_{t=0} = \|h_1 - h_2\|_{\dot{H}^{-1,p}(\mu)},$$

where d/dt^+ denotes the right derivative.

REMARK 3.2 (Comparison with Exercise 22.20 in [101]). Exercise 22.20 in [101] states that (in our notation)

$$\lim_{\epsilon\downarrow 0} \frac{\mathsf{W}_2((1 + \epsilon h)\mu, \mu)}{\epsilon} = \|h\mu\|_{\dot{H}^{-1,2}(\mu)}, \tag{6}$$

for any sufficiently regular function h with $\int h\, d\mu = 0$ ($h\mu$ is understood as a signed measure $h\, d\mu$). Theorem 7.26 in [100] provides a proof of the one-sided inequality that the liminf of the left-hand side above is at least $\|h\mu\|_{\dot{H}^{-1,2}(\mu)}$, when $\mu \in \mathcal{P}_2$ satisfies $\mu \ll dx$ and h is bounded. The subsequent Remark 7.27 states that "We shall not consider the converse of this inequality, which requires more assumptions and more effort." However, we could not find references that establish rigorous conditions applicable to our problem under which the derivative formula (6) holds. Lemma 3.2 provides a rigorous justification for this formula and extends it to general $p > 1$.

Given these preparations, the proof of Theorem 1.1 Parts (i) and (ii) is immediate.

PROOF OF THEOREM 1.1, PARTS (I) AND (II). Let G_μ denote the weak limit of $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$; cf. Proposition 3.1. Recall that $D_\mu = \dot{H}^{-1,p}(\mu * \gamma_\sigma)$ is separable (cf. Lemma 2.2), so $(T_{n,1}, T_{n,2}) := ((\hat{\mu}_n - \mu) * \gamma_\sigma, (\hat{\nu}_n - \mu) * \gamma_\sigma)$ is a Borel measurable map from Ω into the product space $D_\mu \times D_\mu$ [99], Lemma 1.4.1. Since $T_{n,1}$ and $T_{n,2}$ are independent, by Example 1.4.6 in [99] and Proposition 3.1, $(T_{n,1}, T_{n,2}) \xrightarrow{d} (G_\mu, G'_\mu)$ in $D_\mu \times D_\mu$, where G'_μ is an independent copy of G_μ. Since $(T_{n,1}, T_{n,2}) \in T_{\Xi_\mu \times \Xi_\mu}(0,0)$ and $T_{\Xi_\mu \times \Xi_\mu}(0,0)$ is closed in $D_\mu \times D_\mu$, we see that $(G_\mu, G'_\mu) \in T_{\Xi_\mu \times \Xi_\mu}(0,0)$ by the portmanteau theorem.

Applying the functional delta method (Lemma 3.1) and Proposition 3.3, we conclude that

$$\begin{aligned}\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) &= \sqrt{n}\big(\Phi(T_{n,1}, T_{n,2}) - \Phi(0,0)\big)\\ &\xrightarrow{d} \Phi'_{(0,0)}(G_\mu, G'_\mu)\\ &= \|G_\mu - G'_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}.\end{aligned}$$

Likewise, we also have

$$\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu) = \sqrt{n}\big(\Phi(T_{n,1}, 0) - \Phi(0,0)\big) \xrightarrow{d} \Phi'_{(0,0)}(G_\mu, 0) = \|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}.$$

This completes the proof. □

3.3. *Limit distributions under the alternative* ($\mu \neq \nu$).

3.3.1. *One-sample case.* We start from the simpler situation where ν is known and prove Part (iii) of Theorem 1.1. Our proof strategy is to first establish asymptotic normality of the pth power of $\mathsf{W}_p^{(\sigma)}$, from which Part (iii) follows by applying the delta method for $s \mapsto s^{1/p}$. For notational convenience, define

$$\mathsf{S}_p^{(\sigma)}(\mu, \nu) := \big[\mathsf{W}_p^{(\sigma)}(\mu, \nu)\big]^p,$$

for which one-sample asymptotic normality under the alternative is stated next.

PROPOSITION 3.4. *Suppose that $\mu \in \mathcal{P}$ satisfies Condition (4), $\nu \in \mathcal{P}$ is sub-Weibull, and $\mu \neq \nu$. Let g be an optimal transport potential from $\mu * \gamma_\sigma$ to $\nu * \gamma_\sigma$ for W_p^p. Then, we have*

$$\sqrt{n}\big(\mathsf{S}_p^{(\sigma)}(\hat{\mu}_n, \nu) - \mathsf{S}_p^{(\sigma)}(\mu, \nu)\big) \xrightarrow{d} N\big(0, \mathrm{Var}_\mu(g * \phi_\sigma)\big).$$

We again use the functional delta method to prove this proposition, but with a slightly different setting. Set $D_\mu = \dot{H}^{-1,p}(\mu * \gamma_\sigma)$ as before, and consider the function $\Psi : \Lambda_\mu \subset D_\mu \to \mathbb{R}$ defined by

$$\Psi(h) := \mathsf{W}_p^p(\mu * \gamma_\sigma + h, \nu * \gamma_\sigma), \quad h \in \Lambda_\mu,$$

where

$$\Lambda_\mu := D_\mu \cap \big\{h = (\rho - \mu) * \gamma_\sigma : \rho \in \mathcal{P} \text{ is sub-Weibull}\big\}. \tag{7}$$

As long as μ is sub-Weibull (recall that Condition (4) requires μ to be sub-Gaussian), the set Λ_μ contains 0. This set is also convex, and so the tangent cone $T_{\Lambda_\mu}(0)$ coincides with the closure in D_μ of $\{h/t : h \in \Lambda_\mu, t > 0\}$. The corresponding Hadamard directional derivative of W_p^p is given next.

PROPOSITION 3.5 (Hadamard directional derivative of W_p^p w.r.t. one argument). *Let $1 < p < \infty$, and suppose that $\mu, \nu \in \mathcal{P}$ are sub-Weibull. Let g be an optimal transport potential from $\mu * \gamma_\sigma$ to $\nu * \gamma_\sigma$ for W_p^p, which is uniquely determined up to additive constants* (*see Lemma* 2.1(iii)). *Then*:

(i) *$g \in \dot{H}^{1,q}(\mu * \gamma_\sigma)$, where q is the conjugate index of p; and*

(ii) *the map $\Psi : \Lambda_\mu \subset D_\mu \to \mathbb{R}$, $h \mapsto \mathsf{W}_p^p(\mu * \gamma_\sigma + h, \nu * \gamma_\sigma)$, is Hadamard directionally differentiable at $h = 0$ with derivative $\Psi_0'(h) = h(g)$, that is, for any $h \in T_{\Lambda_\mu}(0)$, $t_n \downarrow 0$, and $h_n \to h$ in D_μ such that $t_n h_n \in \Lambda_\mu$, we have*

$$\lim_{n\to\infty} \frac{\Psi(t_n h_n) - \Psi(0)}{t_n} = h(g). \tag{8}$$

As in the null case, Part (ii) of Proposition 3.5 follows from the following Gâteaux differentiability result for W_p^p, combined with local Lipschitz continuity of Ψ w.r.t. $\|\cdot\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$.

LEMMA 3.3 (Gâteaux directional derivative of W_p^p w.r.t. one argument). *Let $1 < p < \infty$ and $\mu, \nu, \rho \in \mathcal{P}$ be sub-Weibull. Let g be an optimal transport potential from $\mu * \gamma_\sigma$ to ν. Then*

$$\frac{d}{dt^+}\mathsf{W}_p^p\big((\mu + t(\rho - \mu)) * \gamma_\sigma, \nu\big)\big|_{t=0} = \int_{\mathbb{R}^d} g\, d\big((\rho - \mu) * \gamma_\sigma\big),$$

where the integral on the right-hand side is well defined and finite.

REMARK 3.3 (Comparisons with Theorem 8.4.7 in [4] and Theorem 5.24 in [87]). Theorem 8.4.7 in [4] derives the following differentiabiliy result for W_p^p. Let $\mu_t : I \to (\mathcal{P}_p, \mathsf{W}_p)$ be an absolutely continuous curve for some open interval I, and let v_t be an "optimal" velocity field satisfying the continuity equation for μ_t (see Theorem 8.4.7 in [4] for the precise meaning). Then, for any $\nu \in \mathcal{P}_p$, we have that

$$\frac{d}{dt}\mathsf{W}_p^p(\mu_t, \nu) = \int_{\mathbb{R}^d\times\mathbb{R}^d} p|x-y|^{p-2}\langle x - y, v_t(x)\rangle\, d\pi_t(x, y) \tag{9}$$

for almost every (a.e.) $t \in I$, where $\pi_t \in \Pi(\mu_t, \nu)$ is an optimal coupling for $\mathsf{W}_p(\mu_t, \nu)$. See also Theorem 5.24 in [87]. Since (9) only holds for a.e. $t \in I$, while we need the (right) differentiability at a specific point, the result of [4], Theorem 8.4.7, (or [87], Theorem 5.24) does not directly apply to our problem. We overcome this difficulty by establishing regularity of optimal transport potentials (see Lemma 5.3 ahead), for which Gaussian smoothing plays an essential role.

We are now ready to prove Proposition 3.4 and obtain Part (iii) of Theorem 1.1 combined with the delta method for the map $s \mapsto s^{1/p}$.

PROOF OF PROPOSITION 3.4. By Proposition 3.1, $T_n := (\hat{\mu}_n - \mu) * \gamma_\sigma \in \Lambda_\mu$ and $\sqrt{n}T_n \xrightarrow{d} G_\mu$ in D_μ. Also $G_\mu \in T_{\Lambda_\mu}(0)$ with probability one by the portmanteau theorem. Applying the functional delta method (Lemma 3.1) and Proposition 3.5, we have

$$\sqrt{n}\big(\mathsf{S}_p^{(\sigma)}(\hat{\mu}_n, \nu) - \mathsf{S}_p^{(\sigma)}(\mu, \nu)\big) = \sqrt{n}\big(\Psi(T_n) - \Psi(0)\big) \xrightarrow{d} G_\mu(g) \sim N\big(0, \mathrm{Var}_\mu(g * \phi_\sigma)\big),$$

as desired. $\square$

3.3.2. *Two-sample case.* Finally, we consider the two-sample case and prove the following, from which Part (iv) of Theorem 1.1 follows.

PROPOSITION 3.6. *Let $1 < p < \infty$. Suppose that $\mu, \nu \in \mathcal{P}$ satisfy Condition (4) and $\nu \neq \mu$. Let g be an optimal transport potential from $\mu * \gamma_\sigma$ to $\nu * \gamma_\sigma$ for W_p^p. Then, we have*

$$\sqrt{n}\big(\mathsf{S}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) - \mathsf{S}_p^{(\sigma)}(\mu, \nu)\big) \xrightarrow{d} N\big(0, \mathrm{Var}_\mu(g * \phi_\sigma) + \mathrm{Var}_\nu(g^c * \phi_\sigma)\big).$$

Set $D_\mu = \dot{H}^{-1,p}(\mu * \gamma_\sigma)$ and $D_\nu = \dot{H}^{-1,p}(\nu * \gamma_\sigma)$. Consider the function $\Upsilon : \Lambda_\mu \times \Lambda_\nu \subset D_\mu \times D_\nu \to \mathbb{R}$ defined by

$$\Upsilon(h_1, h_2) := \mathsf{W}_p^p(\mu * \gamma_\sigma + h_1, \nu * \gamma_\sigma + h_2), \quad (h_1, h_2) \in \Lambda_\mu \times \Lambda_\nu,$$

where Λ_μ is given in (7) and Λ_ν is defined analogously. Here we endow $D_\mu \times D_\nu$ with a product norm (e.g., $\|h_1\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + \|h_2\|_{\dot{H}^{-1,p}(\nu*\gamma_\sigma)}$).

We note that if g is an optimal transport potential from $\mu * \gamma_\sigma$ to $\nu * \gamma_\sigma$, then g^c is an optimal transport potential from $\nu * \gamma_\sigma$ to $\mu * \gamma_\sigma$, as $g^{cc} = g$. With this in mind, Proposition 3.5 immediately yields the following proposition.

PROPOSITION 3.7 (Hadamard directional derivative of W_p^p w.r.t. two arguments). *Let $1 < p < \infty$, and suppose that $\mu, \nu \in \mathcal{P}$ are sub-Weibull. Let g be an optimal transport potential from $\mu * \gamma_\sigma$ to $\nu * \gamma_\sigma$ for W_p^p. Then, $(g, g^c) \in \dot{H}^{1,q}(\mu * \gamma_\sigma) \times \dot{H}^{1,q}(\nu * \gamma_\sigma)$, and the map $\Upsilon : \Lambda_\mu \times \Lambda_\nu \subset D_\mu \times D_\nu \to \mathbb{R}$, $(h_1, h_2) \mapsto \mathsf{W}_p^p(\mu * \gamma_\sigma + h_1, \nu * \gamma_\sigma + h_2)$, is Hadamard directionally differentiable at $(h_1, h_2) = (0, 0)$ with derivative $\Upsilon'_{(0,0)}(h_1, h_2) = h_1(g) + h_2(g^c)$ for $(h_1, h_2) \in T_{\Lambda_\mu \times \Lambda_\nu}(0, 0)$.*

Given Proposition 3.7, the proof of Proposition 3.6 is analogous to that of Proposition 3.4, and is thus omitted for brevity. As before, Part (iv) of Theorem 1.1 follows via the delta method for $s \mapsto s^{1/p}$.

3.4. *Bootstrap.* The limit distributions in Theorem 1.1 are nonpivotal, as they depend on the population distributions μ and/or ν, which are unknown in practice. To overcome this and facilitate statistical inference using $\mathsf{W}_p^{(\sigma)}$, we apply the bootstrap to estimate the limit distributions of empirical $\mathsf{W}_p^{(\sigma)}$.

We start from the one-sample case. Given the data $X_1, \dots, X_n$, let $X_1^B, \dots, X_n^B$ be an independent sample from $\hat{\mu}_n$, and set $\hat{\mu}_n^B := n^{-1} \sum_{i=1}^n \delta_{X_i^B}$ as the bootstrap empirical distribution. Let $\mathbb{P}^B$ denote the conditional probability given $X_1, X_2, \dots$. The next proposition shows that the bootstrap consistently estimates the limit distribution of empirical $\mathsf{W}_p^{(\sigma)}$ under both the null and the alternative.

PROPOSITION 3.8 (Bootstrap consistency: one-sample case). *Suppose that μ satisfies Condition (4).*

(i) (*Null case*) *We have*

$$\sup_{t \geq 0} \big|\mathbb{P}^B\big(\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\mu}_n) \leq t\big) - \mathbb{P}\big(\|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \leq t\big)\big| \xrightarrow{\mathbb{P}} 0. \tag{10}$$

(ii) (*Alternative case*) *Assume in addition that ν is sub-Weibull with $\nu \neq \mu$. Let $\mathfrak{v}_1^2$ denote the asymptotic variance of $\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu) - \mathsf{W}_p^{(\sigma)}(\mu, \nu))$ given in Part (iii) of Theorem 1.1 and assume $\mathfrak{v}_1^2 > 0$. Then, we have*

$$\sup_{t \in \mathbb{R}} \big|\mathbb{P}^B\big(\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \nu) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu)) \leq t\big) - \mathbb{P}\big(N(0, \mathfrak{v}_1^2) \leq t\big)\big| \xrightarrow{\mathbb{P}} 0.$$

Part (ii) of the proposition is not surprising given that the Haramard directional derivative of the function Ψ in Proposition 3.5 is $\Psi_0'(h) = h(g)$, which is linear in h. Part (i) is less obvious since the function $h_1 \mapsto \Phi(h_1, 0)$ from Proposition 3.3 has a nonlinear Hadamard directional derivative, $\Psi_{(0,0)}'(h_1, 0) = \|h_1\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$. Recall that [43], Proposition 1, or [45], Corollary 3.1, show that the bootstrap is inconsistent for functionals with nonlinear derivatives, but these results do not collide with Part (i) of Proposition 3.8 since our application of the bootstrap differs from theirs. For instance, [43], Proposition 1, specialized to our setting states that the conditional law of $\sqrt{n}(\Phi(\hat{\mu}_n^B - \mu, 0) - \Phi(\hat{\mu}_n - \mu, 0)) = \sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \mu) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu))$ does not converge weakly to $\|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ in probability. Heuristically, $\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu)$ is nonnegative while $\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \mu) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu))$ can be negative, so the conditional law of the latter cannot mimic the distribution of the former. Further, when μ is unknown, the conditional law of $\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \mu) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu))$ is infeasible. The correct bootstrap analog for $\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu)$ is $\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\mu}_n) = \Phi(\hat{\mu}_n^B - \mu, \hat{\mu}_n - \mu)$, and the proof of Proposition 3.8 shows that it can be approximated by $\|\hat{\mu}_n^B - \mu - (\hat{\mu}_n - \mu)\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} = \|\hat{\mu}_n^B - \hat{\mu}_n\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$, whose conditional law (after scaling) converges weakly to $\|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ in probability.

Next, consider the two-sample case. In addition to $X_1^B, \dots, X_n^B$ and $\hat{\mu}_n^B$, given $Y_1, \dots, Y_n$, let $Y_1^B, \dots, Y_n^B$ be an independent sample from $\hat{\nu}_n$, and set $\hat{\nu}_n^B := n^{-1}\sum_{i=1}^n \delta_{Y_i^B}$. With a slight abuse of notation, we reuse $\mathbb{P}^B$ for the conditional probability given $(X_1, Y_1), (X_2, Y_2), \dots$.

PROPOSITION 3.9 (Bootstrap consistency: two-sample under the alternative). *Suppose that μ and ν satisfy Condition (4) and $\mu \neq \nu$. Let $\mathfrak{v}_2^2$ denote the asymptotic variance of $\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) - \mathsf{W}_p^{(\sigma)}(\mu, \nu))$ given in Part (iv) of Theorem 1.1 and assume $\mathfrak{v}_2^2 > 0$. Then, we have*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}^B\left(\sqrt{n}\left(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\nu}_n^B) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n)\right) \le t\right) - \mathbb{P}\left(N\left(0, \mathfrak{v}_2^2\right) \le t\right)\right| \xrightarrow{\mathbb{P}} 0.$$

EXAMPLE 3.1 (Confidence interval for $\mathsf{W}_p^{(\sigma)}$). Consider constructing confidence intervals for $\mathsf{W}_p^{(\sigma)}(\mu, \nu)$. For $\alpha \in (0, 1)$, let $\hat{\xi}_\alpha$ denote the conditional α-quantile of $\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\nu}_n^B)$ given the data. Then, by Proposition 3.9 above and Lemma 23.3 in [98], the interval

$$\left[2\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) - \hat{\xi}_{1-\alpha/2}, 2\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) - \hat{\xi}_{\alpha/2}\right],$$

contains $\mathsf{W}_p^{(\sigma)}(\mu, \nu)$ with probability approaching $1 - \alpha$.

For the two-sample case under the null, instead of separately sampling bootstrap draws from $\hat{\mu}_n$ and $\hat{\nu}_n$ (see Remark 3.4 below), we use the pooled empirical distribution $\hat{\rho}_n = (2n)^{-1}\sum_{i=1}^n(\delta_{X_i} + \delta_{Y_i})$ (cf. Chapter 3.7 in [99]). Given $(X_1, Y_1), \dots, (X_n, Y_n)$, let $Z_1^B, \dots, Z_{2n}^B$ be an independent sample from $\hat{\rho}_n$, and set

$$\hat{\rho}_{n,1}^B = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i^B} \quad \text{and} \quad \hat{\rho}_{n,2}^B = \frac{1}{n}\sum_{i=n+1}^{2n} \delta_{Z_i^B}.$$

The following proposition shows that this two-sample bootstrap is consistent for the null limit distribution of empirical $\mathsf{W}_p^{(\sigma)}$.

PROPOSITION 3.10 (Bootstrap consistency: two-sample under the null). *Suppose that μ and ν satisfy Condition (4). Then, for $\rho = (\mu + \nu)/2$, we have*

$$\sup_{t\ge 0}\left|\mathbb{P}^B\left(\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\rho}_{n,1}^B, \hat{\rho}_{n,2}^B) \le t\right) - \mathbb{P}\left(\|G_\rho - G_\rho'\|_{\dot{H}^{-1,p}(\rho*\gamma_\sigma)} \le t\right)\right| \xrightarrow{\mathbb{P}} 0,$$

where G'_ρ is an independent copy of G_ρ. In particular, if $\mu = \nu$, then

$$\sup_{t\geq 0}\left|\mathbb{P}^B\left(\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\rho}_{n,1}^B, \hat{\rho}_{n,2}^B) \leq t\right) - \mathbb{P}\left(\|G_\mu - G'_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \leq t\right)\right| \xrightarrow{\mathbb{P}} 0.$$

REMARK 3.4 (Inconsistency of naive bootstrap). One may consider using $\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\nu}_n^B)$ (rather than $\mathsf{W}_p^{(\sigma)}(\hat{\rho}_{n,1}^B, \hat{\rho}_{n,2}^B)$) to approximate the distribution of $\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n)$, but this bootstrap is not consistent. Indeed, from the proof of Proposition 3.10, we may deduce that, if $\mu = \nu$, then $\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\nu}_n^B)$ is expanded as

$$\left\|\sqrt{n}(\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma - \sqrt{n}(\hat{\nu}_n^B - \hat{\nu}_n) * \gamma_\sigma + \sqrt{n}(\hat{\mu}_n - \hat{\nu}_n) * \gamma_\sigma\right\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + o_{\mathbb{P}}(1),$$

which converges in distribution to $\|G_\mu^1 - G_\mu^2 + G_\mu^3 - G_\mu^4\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ unconditionally, where $G_\mu^1, \dots, G_\mu^4$ are independent copies of G_μ. Hence, the conditional law of $\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\nu}_n^B)$ does not converge weakly to the law of $\|G_\mu - G'_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ in probability.

EXAMPLE 3.2 (Testing the equality of distributions). Consider testing the equality of distributions, that is, $H_0 : \mu = \nu$ against $H_1 : \mu \neq \nu$. We shall use $\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n)$ as a test statistic and reject H_0 if $\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) > c$ for some critical value c. Proposition 3.10 implies that, if we choose $c = \hat{c}_{1-\alpha}$ to be the conditional $(1-\alpha)$-quantile of $\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\rho}_{n,1}^B, \hat{\rho}_{n,2}^B)$ given the data, then the resulting test is asymptotically of level α,

$$\lim_{n\to\infty} \mathbb{P}\left(\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) > \hat{c}_{1-\alpha}\right) = \alpha \quad \text{if } \mu = \nu.$$

Here $\alpha \in (0, 1)$ is the nominal level. To see that the test is consistent, note that if $\mu \neq \nu$, then $\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n) \geq \mathsf{W}_p^{(\sigma)}(\mu, \nu) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu) - \mathsf{W}_p^{(\sigma)}(\hat{\nu}_n, \nu) \geq \mathsf{W}_p^{(\sigma)}(\mu, \nu)/2$ with probability approaching one, while $\hat{c}_{1-\alpha} = O_{\mathbb{P}}(1)$ by Proposition 3.10.

Testing the equality of distributions using Wasserstein distances was considered in [82], but their theoretical analysis is focused on the $d = 1$ case, partly because of the lack of null limit distribution results for empirical W_p in higher dimensions. We overcome this obstacle by using the smooth Wasserstein distance.

4. Minimum distance estimation with $\mathsf{W}_p^{(\sigma)}$. We consider the application of our limit distribution theory to MDE with $\mathsf{W}_p^{(\sigma)}$. Given an independent sample $X_1, \dots, X_n$ from a distribution $\mu \in \mathcal{P}$, MDE aims to learn a generative model from a parametric family $\{\nu_\theta\}_{\theta\in\Theta} \subset \mathcal{P}$ that approximates μ under some statistical divergence. We use $\mathsf{W}_p^{(\sigma)}$ as the proximity measure and the empirical distribution $\hat{\mu}_n$ as an estimate for μ, which leads to the following MDE problem

$$\inf_{\theta\in\Theta} \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta).$$

MDE with classic W_1 is called the Wasserstein GAN, which continues to underlie state-of-the-art methods in generative modeling [6, 59]. MDE with $\mathsf{W}_p^{(\sigma)}$ was previously examined for $p = 1$ in [52] and for $p > 1$ in [77]. Specifically, [77] established measurability, consistency, and parametric convergence rates for MDE with $\mathsf{W}_p^{(\sigma)}$ for $p > 1$, but did not derive limit distribution results. We will expand on this prior work by providing limit distributions for the $\mathsf{W}_p^{(\sigma)}$ MDE problem.

Analogously to the conditions of Theorem 4 in [52], we assume the following.

ASSUMPTION 1. Let $1 < p < \infty$, and assume that the following conditions hold. (i) The distribution μ satisfies Condition (4). (ii) The parameter space $\Theta \subset \mathbb{R}^{d_0}$ is compact with nonempty interior. (iii) The map $\theta \mapsto \nu_\theta$ is continuous w.r.t. the weak topology. (iv) There exists a unique $\theta^\star$ in the interior of Θ such that $\nu_{\theta^\star} = \mu$. (v) There exists a neighborhood N_0 of $\theta^\star$ such that $(\nu_\theta - \nu_{\theta^\star}) * \gamma_\sigma \in \dot{H}^{-1,p}(\mu * \gamma_\sigma)$ for every $\theta \in N_0$. (vi) The map $N_0 \ni \theta \mapsto (\nu_\theta - \nu_{\theta^\star}) * \gamma_\sigma \in \dot{H}^{-1,p}(\mu * \gamma_\sigma)$ is norm differentiable with nonsingular derivative $\mathfrak{D}$ at $\theta^\star$. That is, there exists $\mathfrak{D} = (\mathfrak{D}_1, \dots, \mathfrak{D}_{d_0}) \in (\dot{H}^{-1,p}(\mu * \gamma_\sigma))^{d_0}$, where $\mathfrak{D}_1, \dots, \mathfrak{D}_{d_0}$ are linearly independent elements of $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$, such that

$$\big\|(\nu_\theta - \nu_{\theta^\star}) * \gamma_\sigma - \langle \theta - \theta^\star, \mathfrak{D}\rangle\big\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} = o\big(|\theta - \theta^\star|\big),$$

as $\theta \to \theta^\star$ in N_0, where $\langle t, \mathfrak{D}\rangle = \sum_{i=1}^{d_0} t_i \mathfrak{D}_i$ for $t = (t_1, \dots, t_{d_0}) \in \mathbb{R}^{d_0}$.

REMARK 4.1. Conditions (v) and (vi) are high-level conditions that warrant a discussion. Since $\mathsf{W}_p^{(\sigma)}$ is invariant under a common location shift ($\mathsf{W}_p^{(\sigma)}(\mu, \nu) = \mathsf{W}_p^{(\sigma)}(\mu * \delta_a, \nu * \delta_a)$ for every $a \in \mathbb{R}^d$), we may assume without loss of generality that μ has mean zero, for which $\|\cdot\|_{\dot{H}^{1,q}(\gamma_\sigma)} \lesssim \|\cdot\|_{\dot{H}^{1,q}(\mu*\gamma_\sigma)}$ (as $d(\mu * \gamma_\sigma)/d\gamma_\sigma \ge e^{-\mathbb{E}_\mu[|X|^2]/(2\sigma^2)}$ by Jensen's inequality). Assume that $\{\nu_\theta\}_{\theta\in\Theta}$ is dominated by a common Borel measure ρ on $\mathbb{R}^d$ and denote by f_θ the density of ν_θ w.r.t. ρ. Then $\nu_\theta * \gamma_\sigma$ has Lebesgue density $\int \phi_\sigma(\cdot - y) f_\theta(y)\, d\rho(y)$, so for every $\varphi \in \dot{C}_0^\infty$ with γ_σ-mean zero, we have

$$\begin{aligned}\big((\nu_\theta - \nu_{\theta^\star}) * \gamma_\sigma\big)(\varphi) &= \int (\varphi * \phi_\sigma)(y)\big(f_\theta(y) - f_{\theta^\star}(y)\big)\, d\rho(y) \\ &\le \mathsf{C}_q(\gamma_\sigma)\|\varphi\|_{\dot{H}^{1,q}(\gamma_\sigma)} \int \big|f_\theta(y) - f_{\theta^\star}(y)\big| e^{\frac{(p-1)|y|^2}{2\sigma^2}}\, d\rho(y),\end{aligned}$$

where we use the fact that $(\varphi * \phi_\sigma)(y) \le \mathsf{C}_q(\gamma_\sigma)\|\varphi\|_{\dot{H}^{1,q}(\gamma_\sigma)} e^{\frac{(p-1)|y|^2}{2\sigma^2}}$; see (15). Hence, Condition (v) is satisfied if $\int |f_\theta(y) - f_{\theta^\star}(y)| e^{\frac{(p-1)|y|^2}{2\sigma^2}}\, d\rho(y) < \infty$ for every θ in a neighborhood of $\theta^\star$. Next, assume that f_θ admits the Taylor expansion $f_\theta(y) = f_{\theta^\star}(y) + \langle \dot{f}_{\theta^\star}(y), \theta - \theta^\star\rangle + \langle r_\theta(y), \theta - \theta^\star\rangle$ with $r_\theta(y) = o(1)$ as $\theta \to \theta^\star$. Then Condition (vi) holds with $\mathfrak{D}(\varphi) = \int \varphi(x) \int \phi_\sigma(x - y) \dot{f}_{\theta^\star}(y)\, d\rho(y)\, dx = \int (\varphi * \phi_\sigma)(y) \dot{f}_{\theta^\star}(y)\, d\rho(y)$ for $\varphi \in C_0^\infty$, provided that $\int |\dot{f}_{\theta^\star}(y)| e^{\frac{(p-1)|y|^2}{2\sigma^2}}\, d\rho(y) < \infty$ and

$$\int \big|r_\theta(y)\big| e^{\frac{(p-1)|y|^2}{2\sigma^2}}\, d\rho(y) = o(1), \quad \theta \to \theta^\star.$$

We derive limit distributions for the optimal value function and MDE solution, following the methodology of [10, 52, 80].

THEOREM 4.1 (Limit distributions for MDE with $\mathsf{W}_p^{(\sigma)}$). *Suppose that Assumption 1 holds. Let $\mathbb{G}_n^{(\sigma)} := \sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ be the smooth empirical process, and G_μ its weak limit in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$; cf. Proposition 3.1. Then, the following hold.*

(i) *We have* $\inf_{\theta\in\Theta} \sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) \xrightarrow{d} \inf_{t\in\mathbb{R}^{d_0}} \|G_\mu - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$.

(ii) *Let $(\hat{\theta}_n)_{n\in\mathbb{N}}$ be a sequence of measurable estimators satisfying*

$$\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_{\hat{\theta}_n}) \le \inf_{\theta\in\Theta} \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) + o_{\mathbb{P}}\big(n^{-1/2}\big).$$

Then, provided that $\operatorname{argmin}_{t\in\mathbb{R}^{d_0}} \|G_\mu - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ *is almost surely unique, we have* $\sqrt{n}(\hat{\theta}_n - \theta^\star) \xrightarrow{d} \operatorname{argmin}_{t\in\mathbb{R}^{d_0}} \|G_\mu - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$.

In general, it is nontrivial to verify that $\operatorname{argmin}_{t\in\mathbb{R}^{d_0}} \|G_\mu - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ is almost surely unique. However, for $p=2$, the Hilbertian structure of $\dot{H}^{-1,2}(\mu*\gamma_\sigma)$ guarantees this uniqueness. Indeed, Lemma 5.1 below (or an application of the Lax–Milgram theorem) shows that $\dot{H}^{-1,2}(\mu*\gamma_\sigma)$ is isometrically isomorphic to a closed subspace of $L^2(\mu*\gamma_\sigma;\mathbb{R}^d)$. Denote by E the corresponding isometric isomorphism. Setting $\underline{G}_\mu := E(G_\mu)$ and $\underline{\mathfrak{D}} = (\underline{\mathfrak{D}}_1,\dots,\underline{\mathfrak{D}}_{d_0}) := (E(\mathfrak{D}_1),\dots,E(\mathfrak{D}_{d_0}))$, we have

$$\|G_\mu - \langle t,\mathfrak{D}\rangle\|_{\dot{H}^{-1,2}(\mu*\gamma_\sigma)} = \|\underline{G}_\mu - \langle t,\underline{\mathfrak{D}}\rangle\|_{L^2(\mu*\gamma_\sigma;\mathbb{R}^d)}.$$

The unique minimizer in t of the above display is given by

$$\hat{t}_\mu = \big[\big(\langle \underline{\mathfrak{D}}_j, \underline{\mathfrak{D}}_k\rangle_{L^2(\mu*\gamma_\sigma;\mathbb{R}^d)}\big)_{1\le j,k\le d_0}\big]^{-1}\big(\langle \underline{G}_\mu, \underline{\mathfrak{D}}_j\rangle_{L^2(\mu*\gamma_\sigma;\mathbb{R}^d)}\big)_{j=1}^{d_0}. \tag{11}$$

Since $\underline{G}_\mu$ is a centered Gaussian random variable in $L^2(\mu*\gamma_\sigma;\mathbb{R}^d)$, $\hat{t}_\mu$ is a mean–zero Gaussian vector in $\mathbb{R}^{d_0}$.

COROLLARY 4.1 (Asymptotic normality for MDE solutions when $p=2$). *Consider the setting of Theorem 4.1 Part* (ii) *and let $p=2$. Then $\sqrt{n}(\hat{\theta}_n - \theta^\star) \xrightarrow{d} \hat{t}_\mu$, the mean–zero Gaussian vector in* (11).

Without assuming the uniqueness of $\operatorname{argmin}_{t\in\mathbb{R}^{d_0}} \|G_\mu - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$, limit distributions for MDE solutions can be stated in terms of set-valued random variables. Consider the set of approximate minimizers

$$\hat{\Theta}_n := \Big\{\theta\in\Theta : \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n,\nu_\theta) \le \inf_{\theta'\in\Theta} \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n,\nu_{\theta'}) + n^{-1/2}\lambda_n\Big\}, \tag{12}$$

where λ_n is any nonnegative sequence with $\lambda_n = o_{\mathbb{P}}(1)$. We will show that $\hat{\Theta}_n \subset \theta^\star + n^{-1/2}K_n$ with inner probability approaching one for some sequence K_n of random, convex, and compact sets; cf. [80], Section 2. To describe the sets K_n, for any $\beta\ge 0$ and $h\in\dot{H}^{-1,p}(\mu*\gamma_\sigma)$, define

$$K(h,\beta) := \Big\{t\in\mathbb{R}^{d_0} : \|h - \langle t,\mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \le \inf_{t'\in\mathbb{R}^{d_0}} \|h - \langle t',\mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + \beta\Big\} \in \mathfrak{K},$$

where $\mathfrak{K}$ is the class of compact, convex, and nonempty subsets of $\mathbb{R}^{d_0}$ endowed with the Hausdorff topology. That is, the topology induced by the Hausdorff metric $d_H(K_1,K_2) := \inf\{\delta>0 : K_2\subset K_1^\delta, K_1\subset K_2^\delta\}$, where $K^\delta := \bigcup_{x\in K}\{y\in\mathbb{R}^{d_0} : \|x-y\|\le\delta\}$. Lemma 7.1 in [80] shows that $h\mapsto K(h,\beta)$ is measurable from $\dot{H}^{-1,p}(\mu*\gamma_\sigma)$ into $\mathfrak{K}$ for any $\beta\ge 0$.

PROPOSITION 4.1 (Limit distribution for set of approximate minimizers). *Under Assumption 1, there exists a sequence of nonnegative real numbers $\beta_n \downarrow 0$ such that* (i) $\mathbb{P}_*(\hat{\Theta}_n \subset \theta^\star + n^{-1/2}K(\mathbb{G}_n^{(\sigma)},\beta_n)) \to 1$, *where $\mathbb{P}_*$ denotes inner probability*; *and* (ii) $K(\mathbb{G}_n^{(\sigma)},\beta_n) \xrightarrow{d} K(G_\mu,0)$ *as $\mathfrak{K}$–valued random variables.*

The proof of this proposition is an adaptation of that of Theorem 7.2 in [80]. A self-contained argument is provided in Section 3 of the Supplementary Material [54].

5. Remaining proofs.

5.1. *Proofs for Section 3.1.1.* We fix some notation. For a nonempty set S, let $\ell^\infty(S)$ denote the space of bounded real functions on S endowed with the sup-norm $\|\cdot\|_{\infty,S} = \sup_{s\in S}|\cdot|$. The space $(\ell^\infty(S), \|\cdot\|_{\infty,S})$ is a Banach space.

5.1.1. *Proof of Proposition* 3.1. We divide the proof into three steps. In Steps 1 and 2, we will establish weak convergence of $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ in $\dot{H}^{-1,p}(\gamma_\sigma)$. Step 3 is devoted to weak convergence of $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$.

Step 1. Observe that

$$\big((\hat{\mu}_n - \mu) * \gamma_\sigma\big)(f) = (\hat{\mu}_n - \mu)(f * \phi_\sigma). \tag{13}$$

Consider the function classes

$$\mathcal{F} = \big\{f \in \dot{C}_0^\infty : \|f\|_{\dot{H}^{1,q}(\gamma_\sigma)} \le 1\big\} \quad \text{and} \quad \mathcal{F} * \phi_\sigma = \{f * \phi_\sigma : f \in \mathcal{F}\}.$$

The proof of Theorem 3 in [77] shows that the function class $\mathcal{F} * \phi_\sigma$ is μ-Donsker. For completeness, we provide an outline of the argument. Since for any constant $a \in \mathbb{R}$ and any function $f \in \mathcal{F}$, $(\hat{\mu}_n - \mu)(f * \phi_\sigma) = (\hat{\mu}_n - \mu)((f - a) * \phi_\sigma)$, it suffices to show that $\mathcal{F}_0 * \phi_\sigma$ with $\mathcal{F}_0 := \{f \in \mathcal{F} : \gamma_\sigma(f) = 0\}$ is μ-Donsker. To this end, we will apply Theorem 1 in [97] or its simple adaptation, Lemma 8 in [77].

Fix any $\eta \in (0, 1)$. We first observe that, for any $f \in \mathcal{F}_0$ and any multi-index $k = (k_1, \dots, k_d) \in \mathbb{N}_0^d$, we have

$$\big|\partial^k (f * \phi_\sigma)(x)\big| \lesssim \big(\mathsf{C}_q(\gamma_\sigma) \vee \sigma^{-\bar{k}+1}\big) \exp\left(\frac{(p-1)|x|^2}{2\sigma^2(1-\eta)}\right) \tag{14}$$

up to constants independent of f, x, and σ, where $\bar{k} = \sum_{j=1}^d k_j$. Here $\partial^k = \partial_1^{k_1} \cdots \partial_d^{k_d}$ is the differential operator and $\mathsf{C}_q(\gamma_\sigma)$ is the q-Poincaré constant for the Gaussian measure γ_σ. To see this, observe that

$$(f * \phi_\sigma)(x) = \int_{\mathbb{R}^d} \frac{\phi_\sigma(x-y)}{\phi_\sigma(y)} f(y)\phi_\sigma(y)\, dy.$$

Applying Hölder's inequality and using the fact that $\|f\|_{L^q(\gamma_\sigma)} \le \mathsf{C}_q(\gamma_\sigma)\|f\|_{\dot{H}^{1,q}(\gamma_\sigma)} \le \mathsf{C}_q(\gamma_\sigma)$ (recall that $\gamma_\sigma(f) = 0$), we obtain

$$\big|(f * \phi_\sigma)(x)\big| \le \mathsf{C}_q(\gamma_\sigma)\left[\int_{\mathbb{R}^d} \frac{\phi_\sigma^p(x-y)}{\phi_\sigma^{p-1}(y)}\, dy\right]^{1/p}.$$

A direct calculation further shows that

$$\int_{\mathbb{R}^d} \frac{\phi_\sigma^p(x-y)}{\phi_\sigma^{p-1}(y)}\, dy = \exp\left(\frac{p(p-1)|x|^2}{2\sigma^2}\right),$$

which implies

$$\big|(f * \phi_\sigma)(x)\big| \le \mathsf{C}_q(\gamma_\sigma) \exp\left(\frac{(p-1)|x|^2}{2\sigma^2}\right), \tag{15}$$

establishing (14) when $\bar{k} = 0$. Derivative bounds follow similarly; see [77] for details.

Next, we construct a cover $\{\mathcal{X}_j\}_{j=1}^\infty$ of $\mathbb{R}^d$. Let $B_r = B(0, r)$. For $\delta > 0$ fixed and $r = 2, 3, \dots$, let $\{x_1^{(r)}, \dots, x_{N_r}^{(r)}\}$ be a minimal δ-net of $B_{r\delta} \setminus B_{(r-1)\delta}$. Set $x_1^{(1)} = 0$ with $N_1 = 1$. It is not difficult to see from a volumetric argument that $N_r = O(r^{d-1})$. Set $\mathcal{X}_j = B(x_j^{(r)}, \delta)$ for $j = \sum_{k=1}^{r-1} N_k + 1, \dots, \sum_{k=1}^r N_k$. By construction, $\{\mathcal{X}_j\}_{j=1}^\infty$ forms a cover of $\mathbb{R}^d$ with diameter 2δ. Set $\alpha = \lfloor d/2 \rfloor + 1$ and $M_j = \sup_{f \in \mathcal{F}_0} \max_{\bar{k} \le \alpha} \sup_{x \in \mathrm{int}(\mathcal{X}_j)} |\partial^k (f * \phi_\sigma)(x)|$. By Theorem 1 in [97] combined with Theorem 2.7.1 in [99] (or their simple adaptation; cf. Lemma 8 in [77]), $\mathcal{F}_0 * \phi_\sigma$ is μ-Donsker if $\sum_{j=1}^\infty M_j \mu(\mathcal{X}_j)^{1/2} < \infty$. By inequality (14),

$$\max_{\sum_{k=1}^{r-1} N_k + 1 \le j \le \sum_{k=1}^r N_j} M_j \lesssim \sigma^{-\lfloor d/2 \rfloor} \exp\left(\frac{(p-1)r^2\delta^2}{2\sigma^2(1-\eta)}\right)$$

up to constants independent of r and σ. Hence, $\sum_{j=1}^{\infty} M_j \mu(\mathcal{X}_j)^{1/2}$ is finite if

$$\sum_{r=1}^{\infty} r^{d-1} \exp\left(\frac{(p-1)r^2\delta^2}{2\sigma^2(1-\eta)}\right)\sqrt{\mathbb{P}(|X| > (r-1)\delta)} < \infty.$$

By Riemann approximation, the sum on the the left-hand side above can be bounded by

$$\delta^{-d-1} \int_1^{\infty} t^{d-1} \exp\left(\frac{(p-1)t^2}{2\sigma^2(1-\eta)}\right)\sqrt{\mathbb{P}(|X| > t - 2\delta)}\, dt,$$

which is finite under our assumption by choosing η and δ sufficiently small, and absorbing t^{d-1} into the exponential function.

Step 2. Let $\mathcal{U} = \{f \in \dot{H}^{1,q}(\gamma_\sigma) : \|f\|_{\dot{H}^{1,q}(\gamma_\sigma)} \le 1\}$. Recall from Remark 2.1 that $\dot{H}^{1,q}(\gamma_\sigma) \subset L^q(\gamma_\sigma)$. From Step 1, we know that $\mathcal{F} * \phi_\sigma$ is μ-Donsker. The same conclusion holds with $\mathcal{F}$ replaced by $\mathcal{U}$. This can be verified as follows. From the proof of (14) when $\bar{k} = 0$, we see that for $f_1, f_2 \in \dot{H}^{1,q}(\gamma_\sigma)$ with γ_σ-mean zero,

$$\left|(f_1 * \phi_\sigma)(x) - (f_2 * \phi_\sigma)(x)\right| \le \mathsf{C}_q(\gamma_\sigma)\|f_1 - f_2\|_{\dot{H}^{1,q}(\gamma_\sigma)} \exp\left(\frac{(p-1)|x|^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}^d.$$

Since the exponential function on the right-hand side is square-integrable w.r.t. μ under Condition (4) and $\mathcal{F}_0$ is dense in $\mathcal{U}_0 := \{f \in \mathcal{U} : \gamma_\sigma(f) = 0\}$ for $\|\cdot\|_{\dot{H}^{1,q}(\gamma_\sigma)}$ by construction (cf. Remark 2.1), we see that

$$\mathcal{U}_0 * \phi_\sigma \subset \{g : \exists g_m \in \mathcal{F}_0 * \phi_\sigma \text{ such that } g_m \to g \text{ poinwise and in } L^2(\mu)\}.$$

Thus, by Theorem 2.10.2 in [97], $\mathcal{U}_0 * \phi_\sigma$ (or equivalently, $\mathcal{U} * \phi_\sigma$) is μ-Donsker. Since the map $\ell^\infty(\mathcal{U} * \phi_\sigma) \ni L \mapsto (L(f * \phi_\sigma))_{f \in \mathcal{U}} \in \ell^\infty(\mathcal{U})$ is isometric, in view of (13), we have $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma \xrightarrow{d} G_\mu^\circ$ in $\ell^\infty(\mathcal{U})$ for some tight Gaussian process G_μ°.

Let $\mathrm{lin}^\infty(\mathcal{U})$ denote all bounded real functionals L on $\mathcal{U}$, such that $L(0) = 0$ and

$$L(\alpha f + (1-\alpha)g) = \alpha L(f) + (1-\alpha)L(g), \quad 0 \le \alpha \le 1,\ f, g \in \mathcal{U}. \tag{16}$$

Equip $\mathrm{lin}^\infty(\mathcal{U})$ with the norm $\|\cdot\|_{\infty,\mathcal{U}} = \sup_{f\in\mathcal{U}}|\cdot|$. Each element in $\mathrm{lin}^\infty(\mathcal{U})$ extends uniquely to the corresponding element in $\dot{H}^{-1,p}(\gamma_\sigma)$, and the extension, denoted by $\iota : \mathrm{lin}^\infty(\mathcal{U}) \to \dot{H}^{-1,p}(\gamma_\sigma)$, is isometrically isomorphic. This follows from an argument similar to the proof of Lemma 1 in [76]. Indeed, it is not difficult to verify that each element L in $\mathrm{lin}^\infty(\mathcal{U})$ is *prelinear*, that is, for every $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ and $f_1, \dots, f_m \in \mathcal{U}$, whenever $\alpha_1 f_1 + \cdots + \alpha_m f_m = 0$, we have $\alpha_1 L(f_1) + \cdots + \alpha_m L(f_m) = 0$ (use the fact that $\mathcal{U}$ is centrally symmetric, that is, $-f \in \mathcal{U}$ whenever $f \in \mathcal{U}$, and $L(-f) = -L(f)$, which follows by taking $\alpha = 1/2$ and $g = -f$ in (16)). By Lemma 2.3.5 in [41], the function T_L defined by

$$T_L(\alpha_1 f_1 + \cdots + \alpha_m f_m) = \alpha_1 L(f_1) + \cdots + \alpha_m L(f_m), \quad \alpha_1, \dots, \alpha_m \in \mathbb{R},\ f_1, \dots, f_m \in \mathcal{U}$$

is well defined and linear on the linear span of $\mathcal{U}$, that is, $\dot{H}^{1,q}(\gamma_\sigma)$. Further, as $\|T_L\|_{\dot{H}^{-1,p}(\gamma_\sigma)} = \|L\|_{\infty,\mathcal{U}}$ by construction, $\iota : L \mapsto T_L$ is a linear isometry from $\mathrm{lin}^\infty(\mathcal{U})$ onto $\dot{H}^{-1,p}(\gamma_\sigma)$.

Since $\mathrm{lin}^\infty(\mathcal{U})$ is a closed subspace of $\ell^\infty(\mathcal{U})$ and $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ has paths in $\mathrm{lin}^\infty(\mathcal{U})$, we see that $G_\mu^\circ \in \mathrm{lin}^\infty(\mathcal{U})$ with probability one by the portmanteau theorem and $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma \xrightarrow{d} G_\mu^\circ$ in $\mathrm{lin}^\infty(\mathcal{U})$. Now, since $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ is a (random) signed measure that is bounded on $\mathcal{U}$ with probability one, we can regard $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma$ as a random variable with values in $\dot{H}^{-1,p}(\gamma_\sigma)$. Conclude that $\sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma \xrightarrow{d} \iota \circ G_\mu^\circ$ in $\dot{H}^{-1,p}(\gamma_\sigma)$ by the continuous mapping theorem. For notational convenience, redefine G_μ° by $\iota \circ G_\mu^\circ$. The limit variable $G_\mu^\circ = (G_\mu^\circ(f))_{f\in \dot H^{1,q}(\gamma_\sigma)}$ is a centered Gaussian process with covariance function $\mathrm{Cov}(G_\mu^\circ(f), G_\mu^\circ(g)) = \mathrm{Cov}_\mu(f * \phi_\sigma, g * \phi_\sigma)$.

Step 3. We will show that $\sqrt{n}(\hat\mu_n - \mu) * \gamma_\sigma$ converges in distribution to a centered Gaussian process in $\dot H^{-1,p}(\mu * \gamma_\sigma)$. For $X \sim \mu$ with $a = \mathbb{E}[X]$, let μ^{-a} denote the distribution of $X - a$, and let $\hat\mu_n^{-a} = n^{-1}\sum_{i=1}^n \delta_{(X_i - a)}$. It is not difficult to see that μ^{-a} satisfies Condition (4). Applying the result of Step 2 with μ replaced by μ^{-a}, we have $\sqrt{n}(\hat\mu_n^{-a} - \mu^{-a}) * \gamma_\sigma \xrightarrow{d} G^\circ_{\mu^{-a}}$ in $\dot H^{-1,p}(\gamma_\sigma)$. Since $\|\cdot\|_{\dot H^{1,q}(\gamma_\sigma)} \lesssim \|\cdot\|_{\dot H^{1,q}(\mu^{-a}*\gamma_\sigma)}$ (as $d(\mu^{-a} * \gamma_\sigma)/d\gamma_\sigma \geq e^{-\mathbb{E}_\mu[|X-a|^2]/(2\sigma^2)}$ by Jensen's inequality), we have $\|\cdot\|_{\dot H^{-1,p}(\mu^{-a}*\gamma_\sigma)} \lesssim \|\cdot\|_{\dot H^{-1,p}(\gamma_\sigma)}$, that is, the continuous embedding $\dot H^{-1,p}(\gamma_\sigma) \hookrightarrow \dot H^{-1,p}(\mu^{-a} * \gamma_\sigma)$ holds. Thus $\sqrt{n}(\hat\mu_n^{-a} - \mu^{-a}) * \gamma_\sigma \xrightarrow{d} (G^\circ_{\mu^{-a}}(f))_{f\in\dot H^{1,q}(\mu^{-a}*\gamma_\sigma)}$ in $\dot H^{-1,p}(\mu^{-a} * \gamma_\sigma)$.

Observe that for $\varphi \in C_0^\infty$,

$$\begin{aligned}\|\varphi(\cdot + a)\|^q_{\dot H^{1,q}(\mu^{-a}*\gamma_\sigma)} &= \int_{\mathbb{R}^d} |\nabla\varphi(\cdot + a)|^q \, d(\mu^{-a} * \gamma_\sigma) \\ &= \int_{\mathbb{R}^d} |\nabla\varphi|^q \, d(\mu * \gamma_\sigma) = \|\varphi\|^q_{\dot H^{1,q}(\mu*\gamma_\sigma)}.\end{aligned}$$

Thus, the map $\tau_a : \dot H^{-1,p}(\mu^{-a} * \gamma_\sigma) \to \dot H^{-1,p}(\mu * \gamma)$, defined by $\tau_a(h)(f) = h(f(\cdot + a))$, is continuous (indeed, isometrically isomorphic). Conclude that

$$\sqrt{n}(\hat\mu_n - \mu) * \gamma_\sigma = \tau_a\big(\sqrt{n}(\hat\mu_n^{-a} - \mu^{-a}) * \gamma_\sigma\big) \xrightarrow{d} \tau_a G^\circ_{\mu^{-a}} =: G_\mu \quad \text{in } \dot H^{-1,p}(\mu * \gamma_\sigma).$$

The limit variable $G_\mu = (G_\mu(f))_{f\in\dot H^{1,q}(\mu*\gamma_\sigma)} = (G^\circ_{\mu^{-a}}(f(\cdot + a)))_{f\in\dot H^{1,q}(\mu*\gamma_\sigma)}$ is a centered Gaussian process with covariance function

$$\begin{aligned}\mathrm{Cov}\big(G_\mu(f), G_\mu(g)\big) &= \mathrm{Cov}\big(G^\circ_{\mu^{-a}}\big(f(\cdot + a), G^\circ_{\mu^{-a}}\big(g(\cdot + a)\big)\big) \\ &= \mathrm{Cov}_{\mu^{-a}}\big(f(\cdot + a) * \phi_\sigma, g(\cdot + a) * \phi_\sigma\big) \\ &= \mathrm{Cov}_{\mu^{-a}}\big(f * \phi_\sigma(\cdot + a), g * \phi_\sigma(\cdot + a)\big) \\ &= \mathrm{Cov}_\mu(f * \phi_\sigma, g * \phi_\sigma).\end{aligned}$$

This completes the proof.

REMARK 5.1 (Proof of: (4) $\Rightarrow$ (5)). Follow the notation that appeared in Step 1 in the proof above. Set $\mathcal{X}_1' = \mathcal{X}_1$ and $\mathcal{X}_j' = \mathcal{X}_j \setminus \bigcup_{i=1}^j \mathcal{X}_i$ for $j \geq 2$. The collection $\{\mathcal{X}_j'\}_{j=1}^\infty$ forms a partition of $\mathbb{R}^d$. Observe that $e^{\frac{(p-1)|x|^2}{2\sigma^2}} \leq \sum_{j=1}^\infty M_j \mathbb{1}_{\mathcal{X}_j'}(x)$, so that

$$\int e^{\frac{(p-1)|x|^2}{\sigma^2}} \, d\mu(x) \leq \sum_{j=1}^\infty M_j^2 \mu(\mathcal{X}_j') \leq \sum_{j=1}^\infty M_j^2 \mu(\mathcal{X}_j).$$

For sufficiently small η and δ, Condition (4) ensures $\sum_{j=1}^\infty M_j \mu(\mathcal{X}_j)^{1/2} < \infty$, which implies $\sum_{j=1}^\infty M_j^2 \mu(\mathcal{X}_j) < \infty$ as $M_j \mu(\mathcal{X}_j)^{1/2} \to 0 (j \to \infty)$. Conclude that $\int e^{\frac{(p-1)|x|^2}{\sigma^2}} \, d\mu(x) < \infty$.

REMARK 5.2 (Alternative proof for $p = 2$). Observe that $(\hat\mu_n - \mu) * \gamma_\sigma = n^{-1} \times \sum_{i=1}^n (\delta_{X_i} - \mu) * \gamma_\sigma = n^{-1}\sum_{i=1}^n Z_i$ with $Z_i = (\delta_{X_i} - \mu) * \gamma_\sigma$, and that $Z_1, Z_2, \ldots$ are i.i.d. random variables with values in $\dot H^{-1,p}(\gamma_\sigma)$ (cf. (14)). Since $\dot H^{-1,2}(\gamma_\sigma)$ is isometrically isomorphic to a closed subspace of $L^2(\gamma_\sigma; \mathbb{R}^d)$ (see Lemma 5.1 ahead), we may apply the CLT in the Hilbert space to derive a limit distribution for $\sqrt{n}(\hat\mu_n - \mu) * \gamma_\sigma = n^{-1/2}\sum_{i=1}^n Z_i$ in $\dot H^{-1,2}(\gamma_\sigma)$. Let $E : \dot H^{-1,2}(\gamma_\sigma) \to L^p(\gamma_\sigma; \mathbb{R}^d)$ be the linear isometry given in Lemma 5.1 ahead and $\underline{Z}_i = E(Z_i)$ be the corresponding $L^2(\gamma_\sigma;\mathbb{R}^d)$-valued random variables. Since $L^2(\gamma_\sigma;\mathbb{R}^d)$ is a Hilbert space, $n^{-1/2}\sum_{i=1}^n Z_i$ obeys the CLT if $\mathbb{E}[\|\underline{Z}_1\|^2_{L^2(\gamma_\sigma;\mathbb{R}^d)}] = \mathbb{E}[\|Z_1\|^2_{\dot{H}^{-1,2}(\gamma_\sigma)}] < \infty$, which is satisfied under Condition (4). Indeed, for $p=2$, it is not difficult to see that the CLT in $\dot{H}^{-1,2}(\gamma_\sigma)$ holds for $n^{-1/2}\sum_{i=1}^n Z_i$ under a slightly weaker moment condition, namely, $\int_{\mathbb{R}^d} e^{|x|^2/\sigma^2}\,d\mu(x) < \infty$.

5.1.2. *Proof of Proposition* 3.2. *Part* (i). Let

$$\mathfrak{F} = \{f * \phi_\sigma : f \in \dot{H}^{1,q}(\gamma_\sigma), \|f\|_{L^q(\gamma_\sigma)} \le 1, \|f\|_{\dot{H}^{1,q}(\gamma_\sigma)} \le C\}$$

for some sufficiently large but fixed constant C. It is not difficult to see that $\lim_{n\to\infty}\|(\hat{\mu}_n - \mu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)} = 0$ a.s. if and only if $\mathfrak{F}$ is μ-Glivenko–Cantelli.

Suppose first that $\mathfrak{F}$ is μ-Glivenko–Cantelli. Let $F_{\mathfrak{F}}$ denote the minimal envelope for $\mathfrak{F}$, that is, $F_{\mathfrak{F}}(x) = \sup_{f\in\mathfrak{F}}|f(x)|$. By Theorem 3.7.14 in [50], $F_{\mathfrak{F}}$ must be μ-integrable. We shall bound $F_{\mathfrak{F}}$ from below. Fix any $x \in \mathbb{R}^d$. Consider

$$\varphi_x(y) = \frac{g_x^{p-1}(y)}{\|g_x\|^{p-1}_{L^p(\gamma_\sigma)}} \quad \text{with} \quad g_x(y) = \frac{\phi_\sigma(x-y)}{\phi_\sigma(y)} = e^{-|x|^2/(2\sigma^2)+\langle x,y\rangle/\sigma^2}, \qquad y \in \mathbb{R}^d.$$

Observe that $\nabla_y\varphi_x(y) = ((p-1)x/\sigma^2)\varphi_x(y)$ and thus $\|\varphi_x\|_{\dot{H}^{1,q}(\gamma_\sigma)} = (p-1)|x|/\sigma^2$. Thus, for $\tilde{\varphi}_x = \varphi_x/(1+|x|)$, we have $\|\tilde{\varphi}_x\|_{L^q(\gamma_\sigma)} \le 1$, $\|\tilde{\varphi}_x\|_{\dot{H}^{1,q}(\gamma_\sigma)} \le (p-1)/\sigma^2$, and

$$(\tilde{\varphi}_x * \phi_\sigma)(x) = \frac{1}{1+|x|}\|g_x\|_{L^p(\gamma_\sigma)} = \frac{1}{1+|x|}e^{(p-1)|x|^2/(2\sigma^2)}.$$

Also, from Proposition 1.5.2 in [15], we see that $\tilde{\varphi}_x \in \dot{H}^{1,q}(\gamma_\sigma)$. Conclude that, as long as $C \ge (p-1)/\sigma^2$,

$$F_{\mathfrak{F}}(x) \ge \frac{1}{1+|x|}e^{(p-1)|x|^2/(2\sigma^2)}.$$

Now, the left-hand side is μ-integrable, so that $\int_{\mathbb{R}^d} e^{\theta|x|^2/(2\sigma^2)}\,d\mu(x) < \infty$ for any $\theta < p-1$.

Part (ii). Conversely, suppose that $\int_{\mathbb{R}^d} e^{(p-1)|x|^2/(2\sigma^2)}\,d\mu(x) < \infty$, which ensures that $F_{\mathfrak{F}}$ is μ-integrable from (14). From the proof of Proposition 3.1, for any $M > 0$, we see that the restricted function class $\{f\mathbb{1}_{F_{\mathfrak{F}}\le M} : f \in \mathfrak{F}\}$ is μ-Donsker and thus μ-Glivenko–Cantelli (cf. Theorem 3.7.14 in [50]). Since the envelope function $F_{\mathfrak{F}}$ is μ-integrable, we conclude that $\mathfrak{F}$ is μ-Glivenko–Cantelli; cf. the proof of Theorem 3.7.14 in [50].

5.2. *Proofs for Section* 3.2. Recall that $1 < p < \infty$ and q is its conjugate index, that is, $1/p + 1/q = 1$.

5.2.1. *Proof of Lemma* 3.2. One of the main ingredients of the proof of Lemma 3.2 is Theorem 8.3.1 in [4], which is stated next (see also the Benamou–Brenier formula [9]).

THEOREM 5.1 (Theorem 8.3.1 in [4]). *Let I be an open interval, and let $I \ni t \mapsto \mu_t$ be a continuous curve in $\mathcal{P}_p(\mathbb{R}^d)$ (equipped with W_p) such that for some Borel vector field $\mathbb{R}^d \times I \ni (x,t) \mapsto v_t(x) \in \mathbb{R}^d$, the continuity equation*

$$\partial_t\mu_t + \nabla\cdot(v_t\mu_t) = 0 \quad \text{in } \mathbb{R}^d \times I \tag{17}$$

holds in the distributional sense, that is,

$$\int_I\int_{\mathbb{R}^d}\big(\partial_t\varphi(x,t) + \langle v_t(x), \nabla_x\varphi(x,t)\rangle\big)\,d\mu_t(x)\,dt = 0, \quad \forall\varphi \in C_0^\infty(\mathbb{R}^d \times I).$$

If $\|v_t\|_{L^p(\mu_t;\mathbb{R}^d)} \in L^1(I)$, then $\mathsf{W}_p(\mu_a,\mu_b) \le \int_a^b \|v_t\|_{L^p(\mu_t;\mathbb{R}^d)}\,dt$ for all $a < b$ with $a, b \in I$.

For a vector field $v : \mathbb{R}^d \to \mathbb{R}^d$, define

$$j_p(v) := \begin{cases} |v|^{p-2}v & \text{if } v \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $w = j_p(v)$ if and only if $v = j_q(w)$, and for any $\rho \in \mathcal{P}$,

$$\|j_p(v)\|^q_{L^q(\rho;\mathbb{R}^d)} = \|v\|^p_{L^p(\rho;\mathbb{R}^d)} = \int_{\mathbb{R}^d} \langle j_p(v), v\rangle \, d\rho.$$

We will also use the following lemma.

LEMMA 5.1. *Let $\rho \in \mathcal{P}$ be a reference measure. For any $h \in \dot{H}^{-1,p}(\rho)$, there exists a unique vector field $E = E(h) \in L^p(\rho; \mathbb{R}^d)$ such that*

$$\begin{cases} \int_{\mathbb{R}^d} \langle \nabla\varphi, E\rangle \, d\rho = h(\varphi) & \forall \varphi \in C_0^\infty, \\ j_p(E) \in \overline{\{\nabla\varphi : \varphi \in C_0^\infty\}}^{L^q(\rho;\mathbb{R}^d)}. \end{cases} \tag{18}$$

The map $h \mapsto E(h)$ is homogeneous (i.e., $E(ah) = aE(h)$ for all $a \in \mathbb{R}$ and $h \in \dot{H}^{-1,p}(\rho)$) and such that $\|E(h)\|_{L^p(\rho;\mathbb{R}^d)} = \|h\|_{\dot{H}^{-1,p}(\rho)}$ for all $h \in \dot{H}^{-1,p}(\rho)$. If $p = 2$, then the map $h \mapsto E(h)$ is a linear isometry from $\dot{H}^{-1,2}(\rho)$ into $L^2(\rho; \mathbb{R}^d)$.

The proof of Lemma 5.1 in turn relies on the following existence result of optimal solutions in Banach spaces. We provide its proof for the sake of completeness.

LEMMA 5.2. *Let $(V, \|\cdot\|)$ be a reflexive real Banach space, and let $J : V \to \mathbb{R} \cup \{+\infty\}$ ($J \not\equiv +\infty$) be weakly lower semicontinuous (i.e., $J(v) \le \liminf_n J(v_n)$ for any $v_n \to v$ weakly) and coercive (i.e., $J(v) \to \infty$ as $\|v\| \to \infty$). Then there exists $v_0 \in V$ such that $J(v_0) = \inf_{v\in V} J(v)$.*

PROOF OF LEMMA 5.2. Let $v_n \in V$ be such that $J(v_n) \to \inf_{v\in V} J(v) =: \underline{J}$. By coercivity, v_n is bounded, so by reflexivity and the Banach–Alaoglu theorem, there exists a weakly convergent subsequence v_{n_k} such that $v_{n_k} \to v_0$ weakly. Since J is weakly lower semicontinuous, we conclude $J(v_0) \le \liminf_k J(v_{n_k}) = \underline{J}$. □

We turn to the proof of Lemma 5.1, which is inspired by the first part of the proof of Theorem 8.3.1 in [4].

PROOF OF LEMMA 5.1. Let V denote the closure in $L^q(\rho; \mathbb{R}^d)$ of the subspace $V_0 = \{\nabla\varphi : \varphi \in C_0^\infty\}$. Endowing V with $\|\cdot\|_{L^q(\rho;\mathbb{R}^d)}$ gives a reflexive Banach space because any closed subspace of a reflexive Banach space is reflexive. Define the linear functional $L : V_0 \to \mathbb{R}$ by $L(\nabla\varphi) := h(\varphi)$. To see that L is well defined, observe that

$$\begin{aligned} |h(\varphi)| &\le \|\varphi\|_{\dot{H}^{1,q}(\rho)} \|h\|_{\dot{H}^{-1,p}(\rho)} \\ &= \|\nabla\varphi\|_{L^q(\rho;\mathbb{R}^d)} \|h\|_{\dot{H}^{-1,p}(\rho)}. \end{aligned}$$

This also shows that L can be extended to a bounded linear functional on V.

Consider the optimization problem

$$\min_{v\in V} J(v) \quad \text{with } J(v) := \frac{1}{q}\int_{\mathbb{R}^d} |v|^q \, d\rho - L(v). \tag{19}$$

The functional J is finite, weakly lower semicontinuous, and coercive. By Lemma 5.2 there exists a solution v_0 to the optimization problem (19). Further, the functional J is Gâteaux differentiable with derivative

$$J'(v; w) := \lim_{t \to 0} \frac{J(v + tw) - J(v)}{t} = \int_{\mathbb{R}^d} \langle w, j_q(v) \rangle \, d\rho - L(w).$$

Thus, for $E = j_q(v_0)$, we have $\int_{\mathbb{R}^d} \langle \nabla\varphi, E\rangle \, d\rho = L(\nabla\varphi)$ for all $\varphi \in C_0^\infty$ and $j_p(E) = v_0 \in V$.

To show uniqueness of E, pick another vector field $E' \in L^p(\rho; \mathbb{R}^d)$ satisfying (18). Then, $j_p(E') \in V$ satisfies $J'(j_p(E'); w) = 0$ for all $w \in V$, so by convexity of J, $j_p(E')$ is another optimal solution to (19). However, since J is strictly convex, the optimal solution to (19) is unique, so that $j_p(E') = j_p(E)$, that is, $E' = E$.

Now, the map $h \mapsto E(h)$ is homogeneous, as $aE(h)$ clearly satisfies the first equation in (18) for h replaced with ah and $j_p(aE(h)) = |a|^{p-2} a j_p(E(h)) \in V$. Further, as $j_p(E(h)) \in \overline{\{\nabla\varphi : \varphi \in C_0^\infty\}}^{L^q(\rho;\mathbb{R}^d)}$ by construction, it also satisfies

$$\begin{aligned} \|E(h)\|_{L^p(\rho;\mathbb{R}^d)} &= \sup\left\{ \int_{\mathbb{R}^d} \langle \nabla\varphi, E(h) \rangle \, d\rho : \varphi \in C_0^\infty, \|\nabla\varphi\|_{L^q(\rho;\mathbb{R}^d)} \le 1 \right\} \\ &= \|h\|_{\dot{H}^{-1,p}(\rho)}. \end{aligned}$$

Finally, if $p = 2$, then $j_2(v) = v$, so it is clear that the map $h \mapsto E(h)$ is linear. □

We are now ready to prove Lemma 3.2.

PROOF OF LEMMA 3.2. Let $\mu_t = \mu + th_1$ and $\nu_t = \mu + th_2$ for $t \in [0, 1]$. For notational convenience, let $h = h_1 - h_2 \in D_\mu \cap \{\text{finite signed Borel measures}\}$. We will first show that

$$\liminf_{t \downarrow 0} \frac{\mathsf{W}_p(\mu_t, \nu_t)}{t} \ge \|h_1 - h_2\|_{\dot{H}^{-1,p}(\mu_0)}.$$

The proof is inspired by Theorem 7.26 in [100]. Observe that for any $\varphi \in C_0^\infty$ and $t > 0$,

$$h(\varphi) = \int_{\mathbb{R}^d} \varphi \, dh = \int_{\mathbb{R}^d} \varphi \, d\left(\frac{\mu_t - \nu_t}{t} \right) = \frac{1}{t} \int_{\mathbb{R}^d} \varphi \, d(\mu_t - \nu_t).$$

Let $\pi_t \in \Pi(\mu_t, \nu_t)$ be an optimal coupling for $\mathsf{W}_p^p(\mu_t, \nu_t)$, that is, $\mathsf{W}_p^p(\mu_t, \nu_t) = \iint |x - y|^p \, d\pi_t(x, y)$. Then

$$\frac{1}{t} \int_{\mathbb{R}^d} \varphi \, d(\mu_t - \nu_t) = \frac{1}{t} \iint_{\mathbb{R}^d \times \mathbb{R}^d} \{\varphi(x) - \varphi(y)\} \, d\pi_t(x, y).$$

Since φ is smooth and compactly supported, there exists a constant $C = C_{\varphi,p} < \infty$ such that

$$\varphi(x) - \varphi(y) \le \langle \nabla\varphi(y), x - y \rangle + C|x - y|^{2 \wedge p}, \quad \forall x, y \in \mathbb{R}^d.$$

Indeed, for $p \ge 2$, we can take $C = C_1 := \sup_{x \in \mathbb{R}^d} \|\nabla^2\varphi(x)\|_{\mathrm{op}}/2$ (here $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm for matrices). For $1 < p < 2$, we have

$$\varphi(x) - \varphi(y) \le \langle \nabla\varphi(y), x - y \rangle + C_1 C_2^{2-p} |x - y|^p, \quad \forall x, y \in S := \mathrm{supp}(\varphi)$$

with $C_2 := \sup\{|x - y| : x, y \in S\}$. Here $\mathrm{supp}(\varphi)$ denotes the support of φ, $\mathrm{supp}(\varphi) := \overline{\{\varphi \ne 0\}}$. If $x \in S$ and $d(y, S) := \inf\{|y - z| : z \in S\} > 1$, then $\varphi(x)/|x - y|^p \le \|\varphi\|_\infty$, so that we have

$$\varphi(x) - \varphi(y) = \varphi(x) \le \left(\|\varphi\|_\infty \vee C_1 C_3^{2-p} \right) |x - y|^p, \quad \forall x \in S, y \in S^c$$

with $C_3 := \sup\{|x-y| : x \in S, d(y,S) \le 1\} < \infty$. Finally, if $d(x,S) > 1$ and $y \in S$, then

$$\begin{aligned}\frac{-\varphi(y) - \langle \nabla\varphi(y), x-y\rangle}{|x-y|^p} &\le \|\varphi\|_\infty |x-y|^{-p} + \|\nabla\varphi\|_\infty |x-y|^{1-p} \\ &\le \|\varphi\|_\infty + \|\nabla\varphi\|_\infty,\end{aligned}$$

so that we have

$$\begin{aligned}\varphi(x) - \varphi(y) &= -\varphi(y) \\ &\le \langle \nabla\varphi(y), x-y\rangle \\ &\quad + \Big((\|\varphi\|_\infty + \|\nabla\varphi\|_\infty) \vee C_1 C_4^{2-p}\Big)|x-y|^p, \quad \forall x \in S^c, y \in S\end{aligned}$$

with $C_4 := \sup\{|x-y| : d(x,S) \le 1, y \in S\} < \infty$.

Now, we have

$$\begin{aligned}&\frac{1}{t}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \{\varphi(x) - \varphi(y)\}\, d\pi_t(x,y) \\ &\quad \le \frac{1}{t}\bigg\{\iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle \nabla\varphi(y), x-y\rangle\, d\pi_t(x,y) + C\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^{2\wedge p}\, d\pi_t(x,y)\bigg\} \\ &\quad \le \frac{1}{t}\bigg[\iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle \nabla\varphi(y), x-y\rangle\, d\pi_t(x,y) + C\bigg\{\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^{p}\, d\pi_t(x,y)\bigg\}^{2/(2\vee p)}\bigg] \\ &\quad = \frac{1}{t}\bigg\{\iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle \nabla\varphi(y), x-y\rangle\, d\pi_t(x,y) + C\mathsf{W}_p^{2\wedge p}(\mu_t,\nu_t)\bigg\}.\end{aligned}$$

Applying Proposition 2.1 with $\rho = \mu$, we know that $\mathsf{W}_p(\mu_t,\nu_t) \le \mathsf{W}_p(\mu_t,\mu) + \mathsf{W}_p(\mu,\nu_t) \le pt(\|h_1\|_{\dot H^{-1,p}(\mu)} + \|h_2\|_{\dot H^{-1,p}(\mu)}) = O(t)$ as $t \downarrow 0$, so that $\mathsf{W}_p^{2\wedge p}(\mu_t,\nu_t) = O(t^{2\wedge p}) = o(t)$ as $t \downarrow 0$. Further, by Hölder's inequality, with q being the conjugate index of p, we have

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle \nabla\varphi(y), x-y\rangle\, d\pi_t(x,y) \le \|\nabla\varphi\|_{L^q(\nu_t;\mathbb{R}^d)} \underbrace{\bigg\{\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p\, d\pi_t(x,y)\bigg\}^{1/p}}_{=\mathsf{W}_p(\mu_t,\nu_t)}.$$

Here

$$\begin{aligned}\|\nabla\varphi\|^q_{L^q(\nu_t;\mathbb{R}^d)} &= \int_{\mathbb{R}^d} |\nabla\varphi|^q\, d\mu + t\int_{\mathbb{R}^d} |\nabla\varphi|^q\, dh_2 \\ &= \|\nabla\varphi\|^q_{L^q(\mu;\mathbb{R}^d)} + O(t), t \downarrow 0.\end{aligned}$$

Conclude that

$$h(\varphi) \le \|\nabla\varphi\|_{L^q(\mu;\mathbb{R}^d)} \liminf_{t\downarrow 0} \frac{\mathsf{W}_p(\mu_t,\nu_t)}{t},$$

that is,

$$\liminf_{t\downarrow 0} \frac{\mathsf{W}_p(\mu_t,\nu_t)}{t} \ge \sup\{h(\varphi) : \varphi \in C_0^\infty, \|\nabla\varphi\|_{L^q(\mu;\mathbb{R}^d)} \le 1\} = \|h\|_{\dot H^{-1,p}(\mu)}.$$

To prove the reverse inequality, let $\mathfrak{h} \mapsto E(\mathfrak{h})$ be the map from $\dot H^{-1,p}(\mu)$ into $L^p(\mu;\mathbb{R}^d)$ given in Lemma 5.1. Let $f_t^1 = d\mu_t/d\mu = 1 + t\,dh_1/d\mu$. Since $\mu_1 = \mu + h_1$ is a probability measure, we have $1 + dh_1/d\mu \ge 0$, that is, $dh_1/d\mu \ge -1$, so that $f_t^1 \ge 1/2$ for $t \in [0,1/2]$. Likewise, $f_t^2 := d\nu_t/d\mu \ge 1/2$ for $t \in [0,1/2]$.

Fix $t \in [0, 1/2]$ and consider the curve $\rho_s = (1-s)\mu_t + s\nu_t = \mu_t - sth$ for $s \in [0, 1]$. Then ρ_s satisfies the continuity equation (17) with $v_s = E(-th)/((1-s)f_t^1 + sf_t^2))$. By Theorem 5.1 (Theorem 8.3.1 in [4]), we have

$$\mathsf{W}_p(\mu_t, \nu_t) \le \int_0^1 \|v_s\|_{L^p(\rho_s;\mathbb{R}^d)}\, ds = \int_0^1 \left(\int_{\mathbb{R}^d} \frac{|E(-th)|^p}{[(1-s)f_t^1 + sf_t^2]^{p-1}}\, d\mu \right)^{1/p} ds.$$

Since $E(-th) = -tE(h)$ by homogeneity and $1/2 \le f_t^i \to 1$ as $t \downarrow 0$, the dominated convergence theorem yields that, as $t \downarrow 0$,

$$\begin{aligned}\frac{\mathsf{W}_p(\mu_t, \nu_t)}{t} &\le \int_0^1 \left(\int_{\mathbb{R}^d} \frac{|E(h)|^p}{[(1-s)f_t^1 + sf_t^2]^{p-1}}\, d\mu \right)^{1/p} ds \\ &= \|E(h)\|_{L^p(\mu;\mathbb{R}^d)} + o(1) \\ &= \|h\|_{\dot{H}^{-1,p}(\mu)} + o(1).\end{aligned}$$

This completes the proof. $\square$

5.2.2. *Proof of Proposition* 3.3. Pick arbitrary $(h_1, h_2) \in T_{\Xi_\mu \times \Xi_\mu}(0, 0)$, $t_n \downarrow 0$, and $(h_{n,1}, h_{n,2}) \to (h_1, h_2)$ in $D_\mu \times D_\mu$ such that $(t_n h_{n,1}, t_n h_{n,2}) \in \Xi_\mu \times \Xi_\mu$. By density, for any $\epsilon > 0$, there exist $c > 0$ and $\rho_i \in \mathcal{P}_p$ for $i = 1, 2$ such that $\|h_i - \tilde{h}_i\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} < \epsilon$ for $\tilde{h}_i = c(\rho_i - \mu) * \gamma_\sigma$. By scaling, Lemma 3.2 holds with (h_1, h_2) replaced by $(\tilde{h}_1, \tilde{h}_2)$. Assume without loss of generality that n is large enough such that $\|h_{n,i} - h_i\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} < \epsilon$ for $i = 1, 2$ and $ct_n \le 1/2$. The density of $\mu * \gamma_\sigma + t_n\tilde{h}_i = ((1 - ct_n)\mu + ct_n\rho_i) * \gamma_\sigma$ w.r.t. $\mu * \gamma_\sigma$ is

$$\frac{d(\mu * \gamma_\sigma + t_n\tilde{h}_i)}{d(\mu * \gamma_\sigma)} \ge (1 - ct_n) \ge \frac{1}{2}, \quad i = 1, 2.$$

Thus, by Proposition 2.1, we have

$$\begin{aligned}&\left| \frac{\Phi(t_n h_{n,1}, t_n h_{n,2})}{t_n} - \frac{\Phi(t_n\tilde{h}_1, t_n\tilde{h}_2)}{t_n} \right| \\ &\quad \le \sum_{i=1}^2 \frac{\mathsf{W}_p(\mu * \gamma_\sigma + t_n h_{n,i}, \mu * \gamma_\sigma + t_n\tilde{h}_i)}{t_n} \\ &\quad \lesssim \sum_{i=1}^2 \|h_{n,i} - \tilde{h}_i\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \\ &\quad \le \sum_{i=1}^2 \bigl(\|h_{n,i} - h_i\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + \|h_i - \tilde{h}_i\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}\bigr) \\ &\quad < 4\epsilon.\end{aligned}$$

Further,

$$\bigl| \|h_1 - h_2\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} - \|\tilde{h}_1 - \tilde{h}_2\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \bigr| \le \sum_{i=1}^2 \|h_i - \tilde{h}_i\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} < 2\epsilon.$$

Thus, using the result of Lemma 3.2, we conclude that

$$\limsup_{n\to\infty} \left| \frac{\Phi(t_n h_{n,1}, t_n h_{n,2})}{t_n} - \|h_1 - h_2\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \right| \lesssim \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we obtain the desired conclusion.

5.3. *Proofs for Section* 3.3.

5.3.1. *Proof of Lemma* 3.3. The proof of Lemma 3.3 relies on the following technical lemma concerning regularity of optimal transport potentials, which could be of independent interest. Recall that any locally Lipschitz function on $\mathbb{R}^d$ is differentiable a.e. by the Rademacher theorem (cf. [44]). Here and in what follows a.e. is taken w.r.t. the Lebesgue measure.

LEMMA 5.3 (Regularity of optimal transport potential). *Let* $1 < p < \infty$. *Suppose that* $\mu \in \mathcal{P}_p$ *and* $\nu \in \mathcal{P}$ *is* β*-sub-Weibull for some* $\beta \in (0, 2]$. *Let* g *be an optimal transport potential from* $\mu * \gamma_\sigma$ *to* ν *for* W_p^p. *Then there exists a constant* C *that depends only on* p, d, σ, β, *upper bounds on* $\mathbb{E}_\mu[|X|]$ *and* $\|\|Y\|\|_{\psi_\beta}$ *for* $Y \sim \nu$, *and a lower bound on* $\int \phi_\sigma \, d\mu$, *such that*

$$\begin{cases} g & \textit{is locally Lipschitz}, \\ \left|g(x) - g(0)\right| \le C\left(1 + |x|^{\frac{2p}{\beta}}\right)|x| & \forall x \in \mathbb{R}^d, \\ \left|\nabla g(x)\right| \le C\left(1 + |x|^{\frac{2p}{\beta}}\right) & \textit{for a.e. } x \in \mathbb{R}^d. \end{cases}$$

The proof of Lemma 5.3 borrows ideas from Lemmas 9 and 10 and Theorem 11 in the recent work by [72], which in turn build on [24, 47].

PROOF OF LEMMA 5.3. By Theorem 11 in [72], there exists a constant C_1 depending only on p, d, β, and an upper bound on $\|\|Y\|\|_{\psi_\beta}$ for $Y \sim \nu$, such that

$$\sup_{y \in \partial^c g(x)} |y| \le C_1 \left\{ \left(|x| + 1\right)^{\frac{p}{p-1}} \vee \sup_{y : |x-y| \le 2} \left[\log\left(\frac{1}{(\mu * \gamma_\sigma)(B_y)}\right)\right]^{\frac{p}{\beta(p-1)}} \right\}, \quad x \in \mathbb{R}^d,$$

where $\partial^c g(x) = \{y \in \mathbb{R}^d : c(z, y) - g(z) \ge c(x, y) - g(x), \forall z \in \mathbb{R}^d\}$ is the c-superdifferential of g at x for the cost function $c(x, y) = |x - y|^p$, and $B_y = B(y, 1) = \{x \in \mathbb{R}^d : |x - y| \le 1\}$.

Next, by Proposition 2 in [81], $\mu * \gamma_\sigma$ has Lebesgue density f_μ that is, (c_1, c_2)-regular with $c_1 = 3/\sigma^2$ and $c_2 = 4\,\mathbb{E}_\mu[|X|]/\sigma^2$, that is,

$$\left|\nabla \log f_\mu(x)\right| \le c_1|x| + c_2, \quad \forall x \in \mathbb{R}^d.$$

From the proof of Lemma 10 in [72], we have

$$f_\mu(x) \ge e^{-c_2^2} f_\mu(0) e^{-(1+c_1)|x|^2}, \quad \forall x \in \mathbb{R}^d. \tag{20}$$

Thus, whenever $|x - y| \le 2$,

$$(\mu * \gamma_\sigma)(B_y) = \int_{B_y} f_\mu(z)\, dz \ge \inf_{z \in B_y} f_\mu(z) \times \int_{B_y} dz \ge c_3 e^{-c_2^2} f_\mu(0) e^{-2(1+c_1)(|x|^2+9)},$$

where c_3 is a constant that depends only on d. Conclude that there exists a constant C_2 depending only on p, d, σ, β, upper bounds on $\mathbb{E}_\mu[|X|]$ and $\|\|Y\|\|_{\psi_\beta}$ for $Y \sim \nu$, and a lower bound on $f_\mu(0)$, such that

$$\sup_{y \in \partial^c g(x)} |y| \le C_2\left(1 + |x|^{\frac{2p}{\beta(p-1)}}\right), \quad \forall x \in \mathbb{R}^d.$$

The rest of the proof mirrors the latter half of the proof of Lemma 9 in [72]. Since $g \in L^1(\mu * \gamma_\sigma)$ and $\mu * \gamma_\sigma$ is equivalent to the Lebesgue measure (i.e., $\mu * \gamma_\sigma \ll dx$ and $dx \ll \mu * \gamma_\sigma$), $g(x) > -\infty$ for a.e. $x \in \mathbb{R}^d$. Since any open convex set in $\mathbb{R}^d$ agrees with the interior of its closure (cf. Proposition 6.2.10 in [42]), the convex hull of $\{x : g(x) > -\infty\}$ agrees with

$\mathbb{R}^d$. Thus, by Lemma 2.1(ii) (Theorem 3.3 in [47]), g is locally Lipschitz on $\mathbb{R}^d$. Further, by Propositon C.4 in [47], $\partial^c g(x)$ is nonempty for all $x \in \mathbb{R}^d$. For any $x \in \mathbb{R}^d$ and $y \in \partial^c g(x)$,

$$g(x) = c(x, y) - g^c(y).$$

Thus, for any $x' \in \mathbb{R}^d$,

$$\begin{aligned} g(x') - g(x) &\le c(x', y) - g^c(y) - [c(x, y) - g^c(y)] \\ &= c(x', y) - c(x, y) \\ &= |x' - y|^p - |x - y|^p \\ &\le p(|x - y|^{p-1} \vee |x' - y|^{p-1})|x - x'| \\ &\le C_3[1 + (|x| \vee |x'|)^{\frac{2p}{\beta}}]|x - x'|, \end{aligned}$$

where C_3 depends only on p, β, C_2. Interchanging x and x', we conclude that

$$|g(x) - g(x')| \le C_3[1 + (|x| \vee |x'|)^{\frac{2p}{\beta}}]|x - x'|, \quad x, x' \in \mathbb{R}^d, \tag{21}$$

which implies the desired conclusion. □

PROOF OF LEMMA 3.3. Let $\mu_t = (\mu + t(\rho - \mu)) * \gamma_\sigma = (1 - t)\mu * \gamma_\sigma + t\rho * \gamma_\sigma$ for $t \in [0, 1]$, and let g_t be an optimal transport potential from μ_t to ν. Without loss of generality, we may normalize g_t in such a way that $g_t(0) = 0$ for $t \in [0, 1]$.

We will apply Lemma 5.3 with (μ, ν) replaced with $((1 - t)\mu + t\rho, \nu)$ for $t \in [0, 1/2]$. It is not difficult to see that, as long as $t \in [0, 1/2]$,

$$\mathbb{E}_{(1-t)\mu+t\rho}[|X|] \le \mathbb{E}_\mu[|X|] + \mathbb{E}_\rho[|X|] \quad \text{and} \quad \int_{\mathbb{R}^d} \phi_\sigma \, d((1 - t)\mu + t\rho) \ge \frac{1}{2} \int_{\mathbb{R}^d} \phi_\sigma \, d\mu.$$

Thus, by Lemma 5.3, there exist constants C and K independent of t such that for every $t \in [0, 1/2]$,

$$\begin{cases} g_t & \text{is locally Lipschitz,} \\ |g_t(x)| \le C(1 + |x|^K)|x| & \forall x \in \mathbb{R}^d, \\ |\nabla g_t(x)| \le C(1 + |x|^K) & \text{for a.e. } x \in \mathbb{R}^d. \end{cases}$$

By duality (Lemma 2.1(i)), we have with $h = (\rho - \mu) * \gamma_\sigma$,

$$\begin{aligned} \mathsf{W}_p^p(\mu_t, \nu) &\ge \int_{\mathbb{R}^d} g_0 \, d\mu_t + \int_{\mathbb{R}^d} g_0^c \, d\nu \\ &= \int_{\mathbb{R}^d} g_0 \, d\mu_0 + \int_{\mathbb{R}^d} g_0^c \, d\nu + t \int_{\mathbb{R}^d} g_0 \, dh \\ &= \mathsf{W}_p^p(\mu_0, \nu) + t \int_{\mathbb{R}^d} g_0 \, dh, \end{aligned}$$

so that

$$\liminf_{t \downarrow 0} \frac{\mathsf{W}_p^p(\mu_t, \nu) - \mathsf{W}_p^p(\mu_0, \nu)}{t} \ge \int_{\mathbb{R}^d} g_0 \, dh.$$

Second, by construction,

$$\begin{aligned}
\mathsf{W}_p^p(\mu_t,\nu) &= \int_{\mathbb{R}^d} g_t \, d\mu_t + \int_{\mathbb{R}^d} g_t^c \, d\nu \\
&= \int_{\mathbb{R}^d} g_t \, d\mu_0 + \int_{\mathbb{R}^d} g_t^c \, d\nu + t\int_{\mathbb{R}^d} g_t \, dh \\
&\le \int_{\mathbb{R}^d} g_0 \, d\mu_0 + \int_{\mathbb{R}^d} g_0^c \, d\nu + t\int_{\mathbb{R}^d} g_t \, dh \\
&= \mathsf{W}_p^p(\mu_0,\nu) + t\int_{\mathbb{R}^d} g_t \, dh.
\end{aligned}$$

Pick any $t_n \downarrow 0$. Since $\mu_0 = \mu * \gamma_\sigma \ll dx$, μ_0 has full support $\mathbb{R}^d$, and $\mu_{t_n} \stackrel{w}{\to} \mu_0$, we have by Theorem 3.4 in [34] that there exists some sequence of constants a_n such that $g_{t_n} - a_n \to g_0$ pointwise. Since we have normalized g_t in such a way that $g_t(0) = 0$, we have $a_n \to 0$, that is, $g_{t_n} \to g_0$ pointwise. Further, since $|g_t(x)| \le C(1+|x|^K)|x|$ for all $t \in [0,1/2]$, the dominated convergence theorem yields that

$$\int_{\mathbb{R}^d} g_{t_n} \, dh \to \int_{\mathbb{R}^d} g_0 \, dh.$$

Conclude that

$$\limsup_{n\to\infty} \frac{\mathsf{W}_p^p(\mu_{t_n},\nu) - \mathsf{W}_p^p(\mu_0,\nu)}{t_n} \le \int_{\mathbb{R}^d} g_0 \, dh.$$

This completes the proof $\square$

5.3.2. *Proof of Proposition* 3.5. *Part* (i). We first note that $\dot{H}^{1,q}(\mu * \gamma_\sigma)$ is a function space over $\mathbb{R}^d$. To see this, observe that if we choose a reference measure κ to be an isotropic Gaussian distribution with sufficiently small variance parameter, then the relative density $d(\mu * \gamma_\sigma)/d\kappa$ is bounded away from zero. Indeed, for $\kappa = \gamma_{\sigma/\sqrt{2}}$, we have

$$\begin{aligned}
\frac{d(\mu * \gamma_\sigma)}{d\gamma_{\sigma/\sqrt{2}}}(x) &= 2^{-d/2} \int_{\mathbb{R}^d} e^{-|x-y|^2/(2\sigma^2)+|x|^2/\sigma^2} \, d\mu(y) \\
&= 2^{-d/2} \int_{\mathbb{R}^d} e^{|x+y|^2/(2\sigma^2)-|y|^2/\sigma^2} \, d\mu(y) \\
&\ge 2^{-d/2} e^{-\mathbb{E}_\mu[|X|^2]/\sigma^2}
\end{aligned}$$

by Jensen's inequality, which guarantees that $\dot{H}^{1,q}(\mu * \gamma_\sigma)$ is a function space over $\mathbb{R}^d$ in view of Remark 2.1.

By regularity of g from Lemma 5.3, we know that g is locally Lipschitz and $\|g\|_{L^q(\mu*\gamma_\sigma)} \vee \|\nabla g\|_{L^q(\mu*\gamma_\sigma;\mathbb{R}^d)} < \infty$ (the latter alone does not automatically guarantee $g \in \dot{H}^{1,q}(\mu * \gamma_\sigma)$). As in Proposition 1.5.2 in [15], choose a sequence $\zeta_j \in C_0^\infty$ with the following property:

$$0 \le \zeta_j \le 1, \quad \zeta_j(x) = 1 \quad \text{if } |x| \le j, \sup_{j,x} \big|\nabla \zeta_j(x)\big| < \infty.$$

Let $\varphi_j = \zeta_j g$. Each φ_j belongs to the ordinary Sobolev $(1,q)$-space w.r.t. the Lebesgue measure, so $\nabla \varphi_j$ can be approximated by gradients of C_0^∞ functions under $\|\cdot\|_{L^q(dx;\mathbb{R}^d)}$ (cf. [1], Corollary 3.23). Since $\mu * \gamma_\sigma$ has a bounded Lebesgue density, this shows that $\varphi_j \in \dot{H}^{1,q}(\mu * \gamma_\sigma)$. Now,

$$\|\nabla \varphi_j - \nabla g\|_{L^p(\mu*\gamma_\sigma;\mathbb{R}^d)} \le \big\|(\nabla \zeta_j) g\big\|_{L^q(\mu*\gamma_\sigma;\mathbb{R}^d)} + \big\|(\zeta_j - 1)\nabla g\big\|_{L^q(\mu*\gamma_\sigma;\mathbb{R}^d)} \to 0$$

as $j \to \infty$, implying that $g \in \dot{H}^{1,q}(\mu * \gamma_\sigma)$.

Part (ii). Pick any $h \in T_{\Lambda_\mu}(0)$, $t_n \downarrow 0$, and $h_n \to h$ in D_μ such that $t_n h_n \in \Lambda_\mu$. For any $\epsilon > 0$, there exist some constant $c > 0$ and sub-Weibull $\rho \in \mathcal{P}$ such that $\|h - \tilde{h}\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)} < \epsilon$ for $\tilde{h} = c(\rho - \mu) * \gamma_\sigma$.

Observe that

$$\begin{aligned}\left|\Psi(t_n h_n) - \Psi(t_n \tilde{h})\right| &= \left|\mathsf{W}_p^p(\mu * \gamma_\sigma + t_n h_n, \nu * \gamma_\sigma) - \mathsf{W}_p^p(\mu * \gamma_\sigma + t_n \tilde{h}, \nu * \gamma_\sigma)\right| \\ &\le p\big(\mathsf{W}_p^{p-1}(\mu * \gamma_\sigma + t_n h_n, \nu * \gamma_\sigma) \vee \mathsf{W}_p^{p-1}(\mu * \gamma_\sigma + t_n \tilde{h}, \nu * \gamma_\sigma)\big) \\ &\quad \times \mathsf{W}_p(\mu * \gamma_\sigma + t_n h_n, \mu * \gamma_\sigma + t_n \tilde{h}).\end{aligned}$$

Assume that n is large enough so that $ct_n \le 1/2$ and $\|h_n - h\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)} < \epsilon$. The density of $\mu * \gamma_\sigma + t_n \tilde{h} = ((1 - ct_n)\mu + ct_n \rho) * \gamma_\sigma$ w.r.t. $\mu * \gamma_\sigma$ is

$$\frac{d(\mu * \gamma_\sigma + t_n \tilde{h})}{d(\mu * \gamma_\sigma)} \ge 1 - ct_n \ge \frac{1}{2}.$$

Thus, by Proposition 2.1,

$$\mathsf{W}_p(\mu * \gamma_\sigma + t_n h_n, \mu * \gamma_\sigma + t_n \tilde{h}) \lesssim t_n \|h_n - \tilde{h}\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)} < 2t_n \epsilon.$$

Also, by Proposition 2.1,

$$\begin{aligned}\mathsf{W}_p(\mu * \gamma_\sigma + t_n \tilde{h}, \nu * \gamma_\sigma) &\le \mathsf{W}_p(\mu * \gamma_\sigma + t_n \tilde{h}, \mu * \gamma_\sigma) + \mathsf{W}_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma) \\ &\lesssim t_n \|\tilde{h}\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)} + \mathsf{W}_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma) = O(1).\end{aligned}$$

Likewise, $\mathsf{W}_p(\mu * \gamma_\sigma + t_n h_n, \nu * \gamma_\sigma) = O(1)$. Conclude that

$$\limsup_{n \to \infty} \left|\Psi(t_n h_n) - \Psi(t_n \tilde{h})\right| / t_n \lesssim \epsilon.$$

Further, $|h(g) - \tilde{h}(g)| \le \|g\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)} \|h - \tilde{h}\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)} \lesssim \epsilon$. Combining Lemma 3.3, we conclude that

$$\limsup_{n \to \infty} \left|\frac{\Psi(t_n h_n) - \Psi(0)}{t_n} - h(g)\right| \lesssim \epsilon.$$

This completes the proof.

5.4. *Proofs for Section* 3.4.

5.4.1. *Proof of Proposition* 3.8. We first prove the following lemma. We note that the empirical distributions $\hat{\mu}_n^B$ and $\hat{\mu}_n$ are finitely discrete, so $\sqrt{n}(\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma$ defines a random variable with values in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ (cf. (14) and Step 3 of the proof of Proposition 3.1). Let $\mathfrak{L}_n^B = \mathfrak{L}_n^B(X_1, \dots, X_n)$ denote its (regular) conditional law given the data (which exists as $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ is a separable Banach space; cf. Chapter 11 in [42]).

LEMMA 5.4. *If μ satisfies Condition* (4), *then $\mathfrak{L}_n^B \stackrel{w}{\to} \mathbb{P} \circ G_\mu^{-1}$ almost surely.*

PROOF OF LEMMA 5.4. From the proof of Proposition 3.1, the function class $\mathcal{U} * \phi_\sigma$ is μ-Donsker with a μ-square integrable envelope. The rest of the proof follows from the Giné–Zinn theorem for the bootstrap (cf. Theorem 3.6.2 in [99]) and repeating the arguments in Steps 2 and 3 in the proof of Proposition 3.1. □

The proof of Proposition 3.8 Part (i) relies on the following technical lemmas.

LEMMA 5.5. *Let V be a real seminormed space and V^* be its topological dual with dual norm $\|v^*\|_{V^*} = \sup_{v:\|v\|_V \le 1} v^*(v)$. If $G = (G(v))_{v \in V}$ is a Gaussian process with paths in V^* and a tight measurable map into V^*, then it is Gaussian in the Banach space sense, that is, for every $f \in V^{**}$ (the second dual of V), $f(G)$ is a univariate Gaussian random variable. If G is centered, that is, $\mathbb{E}[G(v)] = 0$ for all $v \in V$, then G has zero Bochner mean in V^*.*

PROOF. Let $V_1 = \{v \in V : \|v\|_V \le 1\}$. Then the map $\iota : V^* \to \ell^\infty(V_1)$ defined by $\iota : v^* \mapsto v^*|_{V_1}$ is a linear isometry. By assumption, $\iota G = (G(v))_{v \in V_1}$ is a Gaussian process and a tight measurable map into $\ell^\infty(V_1)$, so by Lemma 3.9.8 in [97], it is Gaussian in the Banach space sense, that is, for every $F \in (\ell^\infty(V_1))^*$, $F(\iota G)$ is Gaussian. Pick any $f \in V^{**}$. Then $f \circ \iota^{-1}$ is continuous and linear on the vector subspace ιV^* in $\ell^\infty(V_1)$. Let F denote the Hahn–Banach extension of $f \circ \iota^{-1}$; then $F \in (\ell^\infty(V_1))^*$ and $f(G) = (f \circ \iota^{-1})(\iota G) = F(\iota G)$ is Gaussian. Finally, if G is centered, then the proof of Lemma 3.9.8 in [97] shows that $f(G)$ has mean zero for every $f \in V^{**}$, which implies that G has zero Bochner mean by the definition of the Bochner integral (recall from the Fernique theorem that $\mathbb{E}[\|G\|_{V^*}] < \infty$, so the Bochner expectation exists). □

LEMMA 5.6. *Suppose that B is a real separable Banach space with norm $\|\cdot\|$ and G is a B-valued Gaussian random variable with zero Bochner mean. Then, unless G degenerates to zero, $\|G\|$ has a continuous distribution function.*

PROOF. Let F denote the distribution function of $\|G\|$. Set $r_0 = \inf\{r \ge 0 : F(r) > 0\}$, the left endpoint of the support of $\|G\|$. By log-concavity of the Gaussian measure, $\log F$ is concave on (r_0, ∞), which implies that F is (absolutely) continuous on (r_0, ∞); see Theorem 11.1 in [28]. The function F may have a jump at r_0. So it remains to verify that, unless G degenerates to zero, F has no jump at r_0. Indeed, the argument on Pages 60–61 in [68] shows that $r_0 = 0$. But $F(0) - F(0-) = \mathbb{P}(\|G\| = 0) = 0$, so F has no jump at $r_0 = 0$. □

We are now ready to prove Proposition 3.8.

PROOF OF PROPOSITION 3.8. *Part* (i). Assume without loss of generality that μ is not a point mass (otherwise $\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\mu}_n) = \|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} = 0$ and the result trivially follows).

We first verify that the limit variable $\|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma)}$ has a continuous distribution function. Recall that $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ is a separable Banach space. In view of Lemmas 5.5 and 5.6 above, it suffices to verify that G_μ does not degenerate to zero. Since μ is not a point mass, for every $x_0 \in \mathbb{R}^d$, it holds that $0 < W_p(\mu * \gamma_\sigma, \delta_{x_0} * \gamma_\sigma) \lesssim \|(\mu - \delta_{x_0}) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$. This implies that there exists at least one function $f \in C_0^\infty$ such that $\int f d(\mu - \delta_{x_0}) * \gamma_\sigma = \int f * \phi_\sigma d(\mu - \delta_{x_0}) = \int f * \phi_\sigma \, d\mu - f * \phi_\sigma(x_0) > 0$. Since $x \mapsto f * \phi_\sigma(x)$ is continuous, $\int f * \phi_\sigma \, d\mu - f * \phi_\sigma(x)$ is strictly positive in a neighborhood of x_0. Choosing x_0 from the support of μ, we see that $\mathrm{Var}_\mu(f * \phi_\sigma) > 0$.

Hence, we have verified that the limit variable $\|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma)}$ has a continuous distribution function. In view of Problem 23.1 in [98], it suffices to prove the convergence in probability (10) for each fixed $t \ge 0$. Let $T_n = (\hat{\mu}_n - \mu) * \gamma_\sigma$ and $T_n^B = (\hat{\mu}_n^B - \mu) * \gamma_\sigma$. By Proposition 3.1 and Lemma 5.4, we know that

$$
\begin{aligned}
(\sqrt{n} T_n^B, \sqrt{n} T_n) &= (\sqrt{n}(\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma + \sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma, \sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma) \\
&\xrightarrow{d} (G'_\mu + G_\mu, G_\mu) \quad \text{in } \dot{H}^{-1,p}(\mu * \gamma_\sigma) \times \dot{H}^{-1,p}(\mu * \gamma_\sigma)
\end{aligned}
$$

unconditionally, where G'_μ is an independent copy of G_μ (cf. Theorem 2.2 in [66]). Thus, by Proposition 3.5 and the second claim of the functional delta method (Lemma 3.1),

$$\begin{aligned}
\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\mu}_n) &= \sqrt{n}\Phi(T_n^B, T_n) \\
&= \Phi'_{(0,0)}(\sqrt{n}T_n^B, \sqrt{n}T_n) + R_n \\
&= \|\sqrt{n}(T_n^B - T_n)\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + R_n \\
&= \|\sqrt{n}(\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + R_n.
\end{aligned}$$

Here $R_n = o_{\mathbb{P}}(1)$ unconditonally. Choose $\epsilon_n \to 0$ such that $\mathbb{P}(|R_n| > \epsilon_n) \to 0$. By Markov's inequality, we have $\mathbb{P}^B(|R_n| > \epsilon_n) \xrightarrow{\mathbb{P}} 0$. By Lemma 5.4 and the continuous mapping theorem, we also have

$$\sup_{t\geq 0}\left|\mathbb{P}^B(\|\sqrt{n}(\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \leq t) - \mathbb{P}(\|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \leq t)\right| \xrightarrow{\mathbb{P}} 0.$$

Thus, for each $t \geq 0$,

$$\begin{aligned}
&\mathbb{P}^B(\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\mu}_n) \leq t) \\
&\leq \mathbb{P}^B(\|\sqrt{n}(\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \leq t + \epsilon_n) + \mathbb{P}^B(|R_n| > \epsilon_n) \\
&= \mathbb{P}(\|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma)} \leq t + \epsilon_n) + o_{\mathbb{P}}(1) \\
&= \mathbb{P}(\|G_\mu\|_{\dot{H}^{-1,p}(\mu*\gamma)} \leq t) + o_{\mathbb{P}}(1).
\end{aligned}$$

The reverse inequality follows similarly.

Part (ii). The argument is analogous to Part (i). Observe that, by Proposition 3.5,

$$\begin{aligned}
\sqrt{n}(\mathsf{S}_p^{(\sigma)}(\hat{\mu}_n^B, \nu) - \mathsf{S}_p^{(\sigma)}(\hat{\mu}_n, \nu)) &= \sqrt{n}(\Psi(T_n^B) - \Psi(T_n)) \\
&= \Psi'_0(\sqrt{n}T_n^B) - \Psi'_0(\sqrt{n}T_n) + o_{\mathbb{P}}(1) \\
&= \sqrt{n}(T_n^B - T_n)(g) + o_{\mathbb{P}}(1) \\
&= \sqrt{n}((\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma)(g) + o_{\mathbb{P}}(1).
\end{aligned}$$

Taking pth root and applying the delta method, we have

$$\sqrt{n}(\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n^B, \nu) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu)) = \frac{1}{p[\mathsf{W}_p^{(\sigma)}(\mu, \nu)]^{p-1}} \cdot \sqrt{n}((\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma)(g) + o_{\mathbb{P}}(1).$$

The rest of the proof is completely analogous to Part (i). □

PROOF OF PROPOSITION 3.9. By Lemma 5.4 and Example 1.4.6 in [99], the conditional law of $(\sqrt{n}(\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma, \sqrt{n}(\hat{\nu}_n^B - \hat{\nu}_n) * \gamma_\sigma)$ given the data converges weakly to the law of (G_μ, G_ν) in $\dot{H}^{-1,p}(\mu * \gamma_\sigma) \times \dot{H}^{-1,p}(\nu * \gamma_\sigma)$ almost surely, where G_μ and G_ν are independent. By Theorem 2.2 in [66], for $T_{n,1}^B = (\hat{\mu}_n^B - \mu) * \gamma_\sigma$ and $T_{n,2}^B = (\hat{\nu}_n^B - \nu) * \gamma_\sigma$, we have

$$\begin{aligned}
&(\sqrt{n}T_{n,1}^B, \sqrt{n}T_{n,2}^B) \\
&\quad = (\sqrt{n}(\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma + \sqrt{n}(\hat{\mu}_n - \mu) * \gamma_\sigma, \sqrt{n}(\hat{\nu}_n^B - \hat{\nu}_n) * \gamma_\sigma + \sqrt{n}(\hat{\nu}_n - \nu) * \gamma_\sigma) \\
&\quad \xrightarrow{d} (G_\mu + G'_\mu, G_\nu + G'_\nu) \text{in} \dot{H}^{-1,p}(\mu * \gamma_\sigma) \times \dot{H}^{-1,p}(\nu * \gamma_\sigma)
\end{aligned}$$

unconditionally, where G'_μ, G'_ν are copies of G_μ, G_ν, respectively, and G_μ, G'_μ, G_ν, G'_ν are independent. Thus, by Proposition 3.7 and Lemma 3.1, for $T_{n,1} = (\hat{\mu}_n - \mu) * \gamma_\sigma$ and $T_{n,2} = (\hat{\nu}_n - \nu) * \gamma_\sigma$, we have

$$\begin{aligned}
&\sqrt{n}\big(\mathsf{S}_p^{(\sigma)}(\hat{\mu}_n^B, \hat{\nu}_n^B) - \mathsf{S}_p^{(\sigma)}(\hat{\mu}_n, \hat{\nu}_n)\big) \\
&\quad = \sqrt{n}\big(\Upsilon(T_{n,1}^B, T_{n,2}^B) - \Upsilon(T_{n,1}, T_{n,2})\big) \\
&\quad = \Upsilon'_{(0,0)}\big(\sqrt{n}T_{n,1}^B, \sqrt{n}T_{n,2}^B\big) - \Upsilon'_{(0,0)}(\sqrt{n}T_{n,1}, \sqrt{n}T_{n,2}) + o_{\mathbb{P}}(1) \\
&\quad = \sqrt{n}\big(T_{n,1}^B - T_{n,1}\big)(g) + \sqrt{n}\big(T_{n,2}^B - T_{n,2}\big)\big(g^c\big) + o_{\mathbb{P}}(1) \\
&\quad = \sqrt{n}\big((\hat{\mu}_n^B - \hat{\mu}_n) * \gamma_\sigma\big)(g) + \sqrt{n}\big((\hat{\nu}_n^B - \hat{\nu}_n) * \gamma_\sigma\big)\big(g^c\big) + o_{\mathbb{P}}(1).
\end{aligned}$$

The rest of the proof is analogous to Proposition 3.8 Part (ii). □

PROOF OF PROPOSITION 3.10. It is not difficult to see that $\sqrt{2n}(\hat{\rho}_n - \rho) * \gamma_\sigma \xrightarrow{d} G_\rho$ in $\dot{H}^{-1,p}(\rho * \gamma_\sigma)$. By Theorem 3.7.7 and Example 1.4.6 in [99], the conditional law of $(\sqrt{n}(\hat{\rho}_{n,1}^B - \hat{\rho}_n) * \gamma_\sigma, \sqrt{n}(\hat{\rho}_{n,2}^B - \hat{\rho}_n) * \gamma_\sigma)$ given the data converges weakly to the law of (G_ρ, G'_ρ) in $\dot{H}^{-1,p}(\rho * \gamma_\sigma) \times \dot{H}^{-1,p}(\rho * \gamma_\sigma)$ almost surely, where G'_ρ is an independent copy of G_ρ. Thus, arguing as in the proof of Proposition 3.9, for $T_{n,j}^B = (\hat{\rho}_{n,j}^B - \rho) * \gamma_\sigma$ $(j = 1, 2)$, we have

$$\big(\sqrt{n}T_{n,1}^B, \sqrt{n}T_{n,2}^B\big) \xrightarrow{d} \big(G_\rho^1 + G_\rho^2/\sqrt{2}, G_\rho^3 + G_\rho^4/\sqrt{2}\big) \quad \text{in } \dot{H}^{-1,p}(\rho * \gamma_\sigma) \times \dot{H}^{-1,p}(\rho * \gamma_\sigma)$$

unconditionally, where $G_\rho^1, \dots, G_\rho^4$ are independent copies of G_ρ. Define Φ by replacing μ with ρ in Section 3.2. Then, by Proposition 3.3 and the second claim of the functional delta method (Lemma 3.1), we see that

$$\begin{aligned}
\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\rho}_{n,1}^B, \hat{\rho}_{n,2}^B) &= \sqrt{n}\Phi(T_{n,1}^B, T_{n,2}^B) \\
&= \Phi'_{(0,0)}\big(\sqrt{n}T_{n,1}^B, \sqrt{n}T_{n,2}^B\big) + o_{\mathbb{P}}(1) \\
&= \big\|\sqrt{n}\big(T_{n,1}^B - T_{n,2}^B\big)\big\|_{\dot{H}^{-1,p}(\rho * \gamma_\sigma)} + o_{\mathbb{P}}(1) \\
&= \big\|\sqrt{n}\big(\hat{\rho}_{n,1}^B - \hat{\rho}_n\big) * \gamma_\sigma - \sqrt{n}\big(\hat{\rho}_{n,2}^B - \hat{\rho}_n\big) * \gamma_\sigma\big\|_{\dot{H}^{-1,p}(\rho * \gamma_\sigma)} + o_{\mathbb{P}}(1).
\end{aligned}$$

The rest of the proof is analogous to Proposition 3.8 Part (i). □

5.5. *Proof of Theorem* 4.1.

5.5.1. *Preliminary lemmas.* Recall the notation Ξ_μ and D_μ from Section 3.2.

LEMMA 5.7. *Let $\mu \in \mathcal{P}_p$ for $1 < p < \infty$. Under Assumption 1, the map*

$$(h, \theta) \in \Xi_\mu \times N_0 \mapsto \mathsf{W}_p(\mu * \gamma_\sigma + h, \nu_\theta * \gamma_\sigma)$$

is Hadamard directionally differentiable at $(0, \theta^\star)$ with derivative

$$(h, \theta) \in \Xi_\mu \times N_0 \mapsto \|h - \langle \theta, \mathfrak{D} \rangle\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)}.$$

Furthermore, the expansion

$$\mathsf{W}_p(\mu * \gamma_\sigma + h, \nu_\theta * \gamma_\sigma) = \big\|h - \big\langle \theta - \theta^\star, \mathfrak{D}\big\rangle\big\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)} + r\big(h, \theta - \theta^\star\big),$$

holds, with remainder r satisfying $r(th, t(\theta - \theta^\star)) = o(t)$ as $t \downarrow 0$ uniformly w.r.t. (h, θ) varying in $K \subset \Xi_\mu \times N_0$, a compact subset of $D_\mu \times \mathbb{R}^{d_0}$.

PROOF. Consider the map $\psi : (h, \theta) \in \Xi_\mu \times N_0 \mapsto (h, (\nu_\theta - \nu_{\theta^\star}) * \gamma_\sigma) \in \Xi_\mu \times \Xi_\mu$. The norm differentiability condition, Assumption 1 (vi), establishes Fréchet (hence Hadamard) directional differentiability of ψ at $(0, \theta^\star)$ with

$$\psi'_{(0,\theta^\star)}(h, \theta) = \big(h, \langle \theta, \mathfrak{D}\rangle\big) \in T_{\Xi_\mu \times \Xi_\mu}(0, 0).$$

The chain rule for Hadamard directional derivatives paired with Proposition 3.3 yields

$$\begin{aligned}(\Phi \circ \psi)'_{(0,\theta^\star)}(h, \theta) &= \Phi'_{\psi(0,\theta^\star)} \circ \psi'_{(0,\theta^\star)}(h, \theta) \\ &= \Phi'_{(0,0)}\big(h, \langle\theta, \mathfrak{D}\rangle\big) \\ &= \big\|h - \langle \theta, \mathfrak{D}\rangle\big\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}.\end{aligned}$$

The final assertion follows from compact directional differentiability of the composition [90]. $\square$

LEMMA 5.8. *Assume the setting of Lemma 5.7.*

(i) *There exists a neighborhood N_1 of $\theta^\star$ with $\overline{N_1} \subset N_0$ such that*

$$\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) \geq \frac{C}{2}|\theta - \theta^\star| - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu), \quad \forall \theta \in \overline{N_1},$$

where $C > 0$ is such that $\|\langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \geq C|t|$ for every $t \in \mathbb{R}^{d_0}$.

(ii) *Let $\xi_n = O_{\mathbb{P}}(1)$ and $\Theta_n := \{\theta \in \overline{N_1} : \sqrt{n}|\theta - \theta^\star| \leq \xi_n\}$; then, uniformly in $\theta \in \Theta_n$,*

$$\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) = \big\|\mathbb{G}_n^{(\sigma)} - \sqrt{n}\langle\theta - \theta^\star, \mathfrak{D}\rangle\big\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + o_{\mathbb{P}}(1).$$

PROOF. *Part* (i). Assumption 1 (vi) guarantees that there exists a constant $C > 0$ such that $\|\langle\theta - \theta^\star, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} \geq C|\theta - \theta^\star|$ for every $\theta \in N_0$. Let N_2 be an open ball of radius $\bar{r}$ centered at $\theta^\star$ whose closure is contained in N_0; then there exists $t_0 > 0$ such that, for every $0 \leq t \leq t_0$, the remainder term r of Lemma 5.7 satisfies $t^{-1}|r(0, t(\bar{\theta} - \theta^\star))| \leq C\bar{r}/2$ for every $\bar{\theta} \in \partial N_2$. Hence, $|r(0, \theta - \theta^\star)| \leq (C/2)|\theta - \theta^\star|$ for every $\theta \in t_0\overline{N_2} =: \overline{N_1}$ as $\theta - \theta^\star = t(\bar{\theta} - \theta^\star)$ for some $\bar{\theta} \in \partial N_2$ and $0 \leq t \leq t_0$. The triangle inequality yields, for any $\theta \in \overline{N}_1$,

$$\begin{aligned}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) &\geq \mathsf{W}_p^{(\sigma)}(\mu, \nu_\theta) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu), \\ &= \big\|\langle\theta - \theta^\star, \mathfrak{D}\rangle\big\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + r\big(0, \theta - \theta^\star\big) - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu), \\ &\geq \frac{C}{2}|\theta - \theta^\star| - \mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu).\end{aligned}$$

Part (ii). Since $\mathbb{G}_n^{(\sigma)} \stackrel{d}{\to} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ and G_μ is tight, the sequence $\mathbb{G}_n^{(\sigma)}$ is uniformly tight by Lemma 1.3.8 and Problem 1.3.9 in [99]. Pick any $\epsilon, \delta > 0$. By uniform tightness, there exists a compact set $K_\epsilon \subset \dot{H}^{-1,p}(\mu * \gamma_\sigma)$ such that $\mathbb{P}(\mathbb{G}_n^{(\sigma)} \in K_\epsilon) \geq 1 - \epsilon/2$ for every $n \in \mathbb{N}$. Further, since $\xi_n = O_{\mathbb{P}}(1)$, there exists $M_\epsilon > 0$ such that $\mathbb{P}(|\xi_n| \leq M_\epsilon) \geq 1 - \epsilon/2$ for every $n \in \mathbb{N}$. Define the event $A_{n,\epsilon} = \{\mathbb{G}_n^{(\sigma)} \in K_\epsilon\} \cap \{|\xi_n| \leq M_\epsilon\}$. Observe that $\mathbb{P}(A_{n,\epsilon}) \geq 1 - \epsilon$ for every $n \in \mathbb{N}$. Then, on this event $A_{n,\epsilon}$, it holds that $\Theta_n \subset \Theta_{n,\epsilon} := \{\theta \in \overline{N_1} : \sqrt{n}|\theta - \theta^\star| \leq M_\epsilon\}$. Since $\Theta_{n,\epsilon}$ is compact, we have, for every $\theta \in \Theta_{n,\epsilon}$,

$$\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) = \big\|\mathbb{G}_n^{(\sigma)} - \sqrt{n}\langle\theta - \theta^\star, \mathfrak{D}\rangle\big\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + \sqrt{n}r\big(n^{-1/2}\mathbb{G}_n^{(\sigma)}, \theta - \theta^\star\big).$$

Set $\zeta_n := \sup_{\theta \in \Theta_n} |\sqrt{n}r(n^{-1/2}\mathbb{G}_n^{(\sigma)}, \theta - \theta^\star)|$. Then, on the event $A_{n,\epsilon}$,

$$|\zeta_n| \leq \sup_{h \in K_\epsilon, |u| \leq M_\epsilon} \sqrt{n}\big|r\big(n^{-1/2}h, n^{-1/2}u\big)\big|,$$

and the right hand side can be made less than δ for n sufficiently large. Hence, for every sufficiently large n,

$$\begin{aligned}\mathbb{P}(|\zeta_n|>\delta)&=\mathbb{P}(\{|\zeta_n|>\delta\}\cap A_{n,\epsilon})+\mathbb{P}(\{|\zeta_n|>\delta\}\cap A_{n,\epsilon}^c)\\&\le\mathbb{P}(\{|\zeta_n|>\delta\}\cap A_{n,\epsilon})+\mathbb{P}(A_{n,\epsilon}^c)\\&\le 0+\epsilon,\end{aligned}$$

that is, $\zeta_n=o_{\mathbb{P}}(1)$. This implies the desired result. □

LEMMA 5.9. *Under the setting of Lemma 5.7, there exists a sequence of measurable estimators $\hat\theta_n$ satisfying $\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_{\hat\theta_n})=\inf_{\theta\in\Theta}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)$ and $\hat\theta_n\overset{a.s.}{\to}\theta^\star$.*

The proof of Lemma 5.9 follows from a small modification to the proof of Theorems 2 and 3 in [52], see Section 4 of the Supplementary Material [54] for complete details.

5.5.2. *Proof of Theorem 4.1.* *Part* (i). Given the above lemmas, the proof follows closely [80], Theorem 4.2, or [52], Appendix B.4. Let $\hat\theta_n$ be the sequence of measurable estimators afforded by Lemma 5.9. For any neighborhood N of $\theta^\star$,

$$\inf_{\theta\in\Theta}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)=\inf_{\theta\in N}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)$$

with probability approaching one.

By Assumption 1 (vi), there exists $C>0$ such that $\|\langle t,\mathfrak{D}\rangle\|_{\dot H^{-1,p}(\mu*\gamma_\sigma)}\ge C|t|$ for every $t\in\mathbb{R}^{d_0}$. Thus, by Lemma 5.8(i), there exists a neighborhood N_1 of $\theta^\star$ with $\overline{N_1}\subset N_0$ such that

$$\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)\ge\frac{C}{2}|\theta-\theta^\star|-\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\mu),\quad\forall\theta\in\overline{N_1}.$$

Set $\overline{\Theta}_n:=\{\theta\in\Theta:\sqrt{n}|\theta-\theta^\star|\le\xi_n\}$ with $\xi_n:=(4/C)\|\mathbb{G}_n^{(\sigma)}\|_{\dot H^{-1,p}(\mu*\gamma_\sigma)}=O_{\mathbb{P}}(1)$. By Lemma 5.8(ii), the expansion

$$\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)=\|\mathbb{G}_n^{(\sigma)}-\sqrt{n}\langle\theta-\theta^\star,\mathfrak{D}\rangle\|_{\dot H^{-1,p}(\mu*\gamma_\sigma)}+o_{\mathbb{P}}(1),\tag{22}$$

holds uniformly in $\theta\in N_1\cap\overline{\Theta}_n$. Then, for arbitrary $\theta\in N_1\cap\overline{\Theta}_n^c$,

$$\begin{aligned}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)&>\frac{C}{2}\frac{\xi_n}{\sqrt{n}}-\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\mu)\\&=\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\mu)+o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

so that

$$\begin{aligned}\inf_{\theta\in N_1\cap\overline{\Theta}_n^c}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)&>\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\mu)+o_{\mathbb{P}}(n^{-1/2})\\&\ge\inf_{\theta\in N_1\cap\overline{\Theta}_n}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)+o_{\mathbb{P}}(n^{-1/2}).\end{aligned}$$

This shows that $\inf_{\theta\in\Theta}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)=\inf_{\theta\in N_1\cap\overline{\Theta}_n}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)+o_{\mathbb{P}}(n^{-1/2})$.

Now, reparametrizing by $t=\sqrt{n}(\theta-\theta^\star)$ and setting $T_n:=\{t\in\mathbb{R}^{d_0}:|t|\le\xi_n,\theta^\star+t/\sqrt{n}\in N_1\}$ in (22), we have

$$\inf_{\theta\in N_1\cap\overline{\Theta}_n}\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat\mu_n,\nu_\theta)=\inf_{t\in T_n}\|\mathbb{G}_n^{(\sigma)}-\langle t,\mathfrak{D}\rangle\|_{\dot H^{-1,p}(\mu*\gamma_\sigma)}+o_{\mathbb{P}}(1).$$

Set $\mathfrak{g}_n := \|\mathbb{G}_n^{(\sigma)} - \langle \cdot, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$. For any $t \in \mathbb{R}^{d_0}$ such that $|t| > \xi_n$, we have

$$\mathfrak{g}_n(t) \geq C|t| - \|\mathbb{G}_n^{(\sigma)}\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} > 3\mathfrak{g}_n(0) \geq 3 \inf_{|t'|\leq \xi_n} \mathfrak{g}_n(t').$$

Since $\{t \in \mathbb{R}^{d_0} : |t| \leq \xi_n\} \subset T_n$ with probability approaching one (as $\xi_n = O_{\mathbb{P}}(1)$), we have $\inf_{t\in T_n} \mathfrak{g}_n(t) = \inf_{t\in\mathbb{R}^{d_0}} \mathfrak{g}_n(t)$ with probability approaching one. Conclude that

$$\inf_{\theta\in\Theta} \sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) = \inf_{t\in\mathbb{R}^{d_0}} \|\mathbb{G}_n^{(\sigma)} - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)} + o_{\mathbb{P}}(1).$$

Finally, since the map $h \in \dot{H}^{-1,p}(\mu * \gamma_\sigma) \mapsto \inf_{t\in\mathbb{R}^{d_0}} \|h - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ is continuous, the continuous mapping theorem yields

$$\inf_{\theta\in\Theta} \sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) \xrightarrow{d} \inf_{t\in\mathbb{R}^{d_0}} \|G_\mu - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}.$$

This completes the proof of Part (i).

Part (ii). Let N_1 be as in the proof of Part (i) and recall that $\hat{\theta}_n \in N_1$ with probability approaching one. By the definition of $\hat{\theta}_n$ and Lemma 5.8(ii),

$$\begin{aligned} \underbrace{\inf_{\theta\in\Theta} \sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta)}_{=O_{\mathbb{P}}(1)} + o_{\mathbb{P}}(1) &\geq \sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_{\hat{\theta}_n}) \\ &\geq \frac{C}{2}\sqrt{n}|\hat{\theta}_n - \theta^\star| - \underbrace{\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \mu)}_{=O_{\mathbb{P}}(1)}, \end{aligned} \tag{23}$$

with probability tending to one. This implies that $\sqrt{n}|\hat{\theta}_n - \theta^\star| = O_{\mathbb{P}}(1)$. Let $\mathbb{M}_n(t) := \|\mathbb{G}_n^{(\sigma)} - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$ and $\mathbb{M}(t) := \|G_\mu - \langle t, \mathfrak{D}\rangle\|_{\dot{H}^{-1,p}(\mu*\gamma_\sigma)}$. Observe that $\mathbb{M}_n$ and $\mathbb{M}$ are convex in t. Again, from the proof of Part (i), for $\hat{t}_n := \sqrt{n}(\hat{\theta}_n - \theta^\star) = O_{\mathbb{P}}(1)$, we have

$$\sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_{\hat{\theta}_n}) = \mathbb{M}_n(\hat{t}_n) + o_{\mathbb{P}}(1).$$

Hence,

$$\begin{aligned} \mathbb{M}_n(\hat{t}_n) &= \sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_{\hat{\theta}_n}) + o_{\mathbb{P}}(1) \\ &\leq \inf_{\theta\in\Theta} \sqrt{n}\mathsf{W}_p^{(\sigma)}(\hat{\mu}_n, \nu_\theta) + o_{\mathbb{P}}(1) \\ &= \inf_{t\in\mathbb{R}^{d_0}} \mathbb{M}_n(t) + o_{\mathbb{P}}(1). \end{aligned}$$

Since $\mathbb{G}_n^{(\sigma)} \xrightarrow{d} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$, $(\mathbb{M}_n(t_1), \ldots, \mathbb{M}_n(t_k)) \xrightarrow{d} (\mathbb{M}(t_1), \ldots, \mathbb{M}(t_k))$ for any finite set of points $(t_i)_{i=1}^k \subset \mathbb{R}^{d_0}$ by the continuous mapping theorem. Applying Theorem 1 in [64] (or Lemma 6 in [52]) yields $\hat{t}_n \xrightarrow{d} \operatorname{argmin}_{t\in\mathbb{R}^{d_0}} \mathbb{M}(t)$.

6. Concluding remarks. In this paper, we have developed a comprehensive limit distribution theory for empirical $\mathsf{W}_p^{(\sigma)}$ that covers general $1 < p < \infty$ and $d \geq 1$, under both the null and the alternative. Our proof technique leveraged the extended functional delta method, which required two main ingredients: (i) convergence of the smooth empirical process in an appropriate normed vector space; and (ii) characterization of the Hadamard directional derivative of $\mathsf{W}_p^{(\sigma)}$ w.r.t. the norm. We have identified the dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma)$ as the normed space of interest and established the items above to obtain the limit distribution results. Linearity of the Hadamard directional derivative under the alternative enabled establishing the asymptotic normality of the empirical (scaled) $\mathsf{W}_p^{(\sigma)}$.

To facilitate statistical inference using $\mathsf{W}_p^{(\sigma)}$, we have established the consistency of the nonparametric bootstrap. The limit distribution theory was used to study generative modeling via $\mathsf{W}_p^{(\sigma)}$ MDE. We have derived limit distributions for the optimal solutions and the corresponding smooth Wasserstein error, and obtained Gaussian limits when $p = 2$ by leveraging the Hilbertian structure of the corresponding dual Sobolev space. Our statistical study, together with the appealing metric and topological structure of $\mathsf{W}_p^{(\sigma)}$ [51, 77], suggest that the smooth Wasserstein framework is compatible with high-dimensional learning and inference.

An important direction for future research is the efficient computation of $\mathsf{W}_p^{(\sigma)}$. While standard methods for computing W_p are applicable in the smooth case (by sampling the Gaussian noise), it is desirable to find computational techniques that make use of the structure induced by the convolution with a known smooth kernel. Another appealing direction is to establish Berry–Esseen type bounds for the limit distributions in Theorem 1.1. Of particular interest is to explore how parameters such as d and σ affect the accuracy of the limit distributions in Theorem 1.1. [85] addressed a similar problem for empirical $\mathsf{W}_1^{(\sigma)}$ under the one-sample null case, but their proof relies substantially on the IPM structure of W_1 and finite sample Gaussian approximation techniques developed by [21, 22]. These techniques do not apply to $p > 1$, and thus new ideas, such as the linearization arguments herein, are required to develop Berry–Esseen type bounds for $p > 1$.

Funding. Z. Goldfeld is supported by NSF Grants CCF-1947801, CCF-2046018, DMS-2210368, and the 2020 IBM Academic Award.

K. Kato is partially supported by NSF Grants DMS-1952306, DMS-2014636, and DMS-2210368.

S. Nietert is supported by the NSF Graduate Research Fellowship under Grant DGE-1650441.

G. Rioux is partially supported by the NSERC postgraduate fellowship PGSD-567921-2022.

SUPPLEMENTARY MATERIAL

Additional proofs (DOI: 10.1214/23-AAP2028SUPP; .pdf). The Supplementary Material [54] contains proofs of Lemmas 2.2 and 5.9, and Propositions 2.1 and 4.1.

REFERENCES

[1] ADAMS, R. A. (1975). *Sobolev Spaces*. Academic Press, New York–London. MR0450957

[2] AJTAI, M., KOMLÓS, J. and TUSNÁDY, G. (1984). On optimal matchings. *Combinatorica* **4** 259–264. MR0779885 https://doi.org/10.1007/BF02579135

[3] ALTSCHULER, J., WEED, J. and RIGOLLET, P. (2017). Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Proceedings of the* 31*st International Conference on Neural Information Processing Systems* 1961–1971.

[4] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2401600

[5] AMBROSIO, L., STRA, F. and TREVISAN, D. (2019). A PDE approach to a 2-dimensional matching problem. *Probab. Theory Related Fields* **173** 433–477. MR3916111 https://doi.org/10.1007/s00440-018-0837-x

[6] ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein generative adversarial networks. In *International Conference on Machine Learning*.

[7] AUBIN, J.-P. and FRANKOWSKA, H. (2009). *Set-Valued Analysis*. *Modern Birkhäuser Classics*. Birkhäuser, Boston, MA. MR2458436 https://doi.org/10.1007/978-0-8176-4848-0

[8] BARTHE, F. and BORDENAVE, C. (2013). Combinatorial optimization over two random point sets. In *Séminaire de Probabilités XLV*. *Lecture Notes in Math.* **2078** 483–535. Springer, Cham. MR3185927 https://doi.org/10.1007/978-3-319-00321-4_19

[9] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. MR1738163 https://doi.org/10.1007/s002110050002

[10] BERNTON, E., JACOB, P. E., GERBER, M. and ROBERT, C. P. (2019). On parameter estimation with the Wasserstein distance. *Inf. Inference* **8** 657–676. MR4049226 https://doi.org/10.1093/imaiai/iaz003

[11] BIGOT, J., CAZELLES, E. and PAPADAKIS, N. (2019). Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electron. J. Stat.* **13** 5120–5150. MR4041704 https://doi.org/10.1214/19-EJS1637

[12] BLANCHET, J., MURTHY, K. and ZHANG, F. (2022). Optimal transport-based distributionally robust optimization: Structural properties and iterative schemes. *Math. Oper. Res.* **47** 1500–1529. MR4435024 https://doi.org/10.1287/moor.2021.1178

[13] BOBKOV, S. G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* **27** 1903–1921. MR1742893 https://doi.org/10.1214/aop/1022874820

[14] BOBKOV, S. G. and LEDOUX, M. (2021). A simple Fourier analytic proof of the AKT optimal matching theorem. *Ann. Appl. Probab.* **31** 2567–2584. MR4350968 https://doi.org/10.1214/20-aap1656

[15] BOGACHEV, V. I. (1998). *Gaussian Measures*. *Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. MR1642391 https://doi.org/10.1090/surv/062

[16] BOISSARD, E. and LE GOUIC, T. (2014). On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 539–563. MR3189084 https://doi.org/10.1214/12-AIHP517

[17] BOLLEY, F., GUILLIN, A. and VILLANI, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* **137** 541–593. MR2280433 https://doi.org/10.1007/s00440-006-0004-7

[18] BOUTET DE MONVEL, J. H. and MARTIN, O. C. (2002). Almost sure convergence of the minimum bipartite matching functional in Euclidean space. *Combinatorica* **22** 523–530. MR1956991 https://doi.org/10.1007/s00493-002-0004-x

[19] CÁRCAMO, J., CUEVAS, A. and RODRÍGUEZ, L.-A. (2020). Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli* **26** 2143–2175. MR4091104 https://doi.org/10.3150/19-BEJ1188

[20] CHEN, H.-B. and NILES-WEED, J. (2022). Asymptotics of smoothed Wasserstein distances. *Potential Anal.* **56** 571–595. MR4396827 https://doi.org/10.1007/s11118-020-09895-9

[21] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

[22] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.* **126** 3632–3651. MR3565470 https://doi.org/10.1016/j.spa.2016.04.009

[23] CHIZAT, L., ROUSSILLON, P., LÉGER, F., VIALARD, F.-X. and PEYRÉ, G. (2020). Faster Wasserstein distance estimation with the Sinkhorn divergence. In *International Conference on Neural Information Processing Systems*.

[24] COLOMBO, M. and FATHI, M. (2021). Bounds on optimal transport maps onto log-concave measures. *J. Differ. Equ.* **271** 1007–1022. MR4154935 https://doi.org/10.1016/j.jde.2020.09.032

[25] COURTY, N., FLAMARY, R. and TUIA, D. (2014). Domain adaptation with regularized optimal transport. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 274–289. Springer, Berlin.

[26] COURTY, N., FLAMARY, R., TUIA, D. and RAKOTOMAMONJY, A. (2016). Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 1853–1865.

[27] CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* (*NeurIPS*-2013) 2292–2300.

[28] DAVYDOV, Y. A., LIFSHITS, M. A. and SMORODINA, N. V. (1998). *Local Properties of Distributions of Stochastic Functionals*. *Translations of Mathematical Monographs* **173**. Amer. Math. Soc., Providence, RI. MR1604537 https://doi.org/10.1090/mmono/173

[29] DEB, N., GHOSAL, P. and SEN, B. (2021). Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. Available at arXiv:2107.01718.

[30] DEL BARRIO, E., GINÉ, E. and MATRÁN, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* **27** 1009–1071. MR1698999 https://doi.org/10.1214/aop/1022677394

[31] DEL BARRIO, E., GINÉ, E. and UTZET, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** 131–189. MR2121458 https://doi.org/10.3150/bj/1110228245

[32] DEL BARRIO, E., GONZÁLEZ SANZ, A. and LOUBES, J.-M. (2024). Central limit theorems for semi-discrete Wasserstein distances. *Bernoulli* **30** 554–580. MR4665589 https://doi.org/10.3150/23-bej1608

[33] DEL BARRIO, E., GONZÁLEZ SANZ, A., LOUBES, J.-M. and NILES-WEED, J. (2023). An improved central limit theorem and fast convergence rates for entropic transportation costs. *SIAM J. Math. Data Sci.* **5** 639–669. MR4616887 https://doi.org/10.1137/22M149260X

[34] DEL BARRIO, E., GONZÁLEZ-SANZ, A. and LOUBES, J.-M. (2021). Central limit theorems for general transportation costs. Available at arXiv:2102.06379.

[35] DEL BARRIO, E., GORDALIZA, P. and LOUBES, J.-M. (2019). A central limit theorem for L_p transportation cost on the real line with application to fairness assessment in machine learning. *Inf. Inference* **8** 817–849. MR4045479 https://doi.org/10.1093/imaiai/iaz016

[36] DEL BARRIO, E. and LOUBES, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.* **47** 926–951. MR3916938 https://doi.org/10.1214/18-AOP1275

[37] DEREICH, S., SCHEUTZOW, M. and SCHOTTSTEDT, R. (2013). Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1183–1203. MR3127919 https://doi.org/10.1214/12-AIHP489

[38] DOBRIĆ, V. and YUKICH, J. E. (1995). Asymptotics for transportation cost in high dimensions. *J. Theoret. Probab.* **8** 97–118. MR1308672 https://doi.org/10.1007/BF02213456

[39] DOLBEAULT, J., NAZARET, B. and SAVARÉ, G. (2009). A new class of transport distances between measures. *Calc. Var. Partial Differential Equations* **34** 193–231. MR2448650 https://doi.org/10.1007/s00526-008-0182-5

[40] DUDLEY, R. M. (1969). The speed of mean Glivenko–Cantelli convergence. *Ann. Math. Stat.* **40** 40–50.

[41] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. MR1720712 https://doi.org/10.1017/CBO9780511665622

[42] DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

[43] DÜMBGEN, L. (1993). On nondifferentiable functions and the bootstrap. *Probab. Theory Related Fields* **95** 125–140. MR1207311 https://doi.org/10.1007/BF01197342

[44] EVANS, L. C. and GARIEPY, R. F. (2018). *Measure Theory and Fine Properties of Functions*. Routledge, London.

[45] FANG, Z. and SANTOS, A. (2019). Inference on directionally differentiable functions. *Rev. Econ. Stud.* **86** 377–412. MR3936869 https://doi.org/10.1093/restud/rdy049

[46] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

[47] GANGBO, W. and MCCANN, R. J. (1996). The geometry of optimal transportation. *Acta Math.* **177** 113–161. MR1440931 https://doi.org/10.1007/BF02392620

[48] GAO, R. and KLEYWEGT, A. (2023). Distributionally robust stochastic optimization with Wasserstein distance. *Math. Oper. Res.* **48** 603–655. MR4588934

[49] GENEVAY, A., CHIZAT, L., BACH, F., CUTURI, M. and PEYRÉ, G. (2019). Sample complexity of Sinkhorn divergences. In *Proceedings of the* 22*nd International Conference on Artificial Intelligence and Statistics* 1574–1583.

[50] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862

[51] GOLDFELD, Z. and GREENEWALD, K. (2020). Gaussian-smoothed optimal transport: Metric structure and statistical efficiency. In *International Conference on Artificial Intelligence and Statistics*.

[52] GOLDFELD, Z., GREENEWALD, K. and KATO, K. (2020). Asymptotic guarantees for generative modeling based on the smooth Wasserstein distance. In *International Conference on Neural Information Processing Systems*.

[53] GOLDFELD, Z., GREENEWALD, K. H., NILES-WEED, J. and POLYANSKIY, Y. (2020). Convergence of smoothed empirical measures with applications to entropy estimation. *IEEE Trans. Inf. Theory* **66** 1489–1501.

[54] GOLDFELD, Z., KATO, K., NIETERT, S. and RIOUX, G. (2024). Supplement to "Limit distribution theory for smooth p-Wasserstein distances." https://doi.org/10.1214/23-AAP2028SUPP

[55] GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2022). Statistical inference with regularized optimal transport. ArXiv preprint. Available at arXiv:2205.04283.

[56] GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2022). Limit theorems for entropic optimal transport maps and the Sinkhorn divergence. ArXiv preprint. Available at arXiv:2207.08683.

[57] GONZALEZ-SANZ, A., LOUBES, J.-M. and NILES-WEED, J. (2022). Weak limits of entropy regularized optimal transport; potentials, plans and divergences. ArXiv preprint. Available at arXiv:2207.07427.

[58] GOURIEROUX, C., MONFORT, A. and RENAULT, E. (1993). Indirect inference. *J. Appl. Econometrics* **8** S85–S118.

[59] GULRAJANI, I., AHMED, F., ARJOVSKY, M., DUMOULIN, V. and COURVILLE, A. C. (2017). Improved training of Wasserstein GANs. In *International Conference on Neural Information Processing Systems*.

[60] HAN, F., MIAO, Z. and SHEN, Y. (2021). Nonparametric mixture MLEs under Gaussian-smoothed optimal transport distance. ArXiv preprint. Available at arXiv:2112.02421.

[61] HUNDRIESER, S., KLATT, M., STAUDT, T. and MUNK, A. (2022). A unifying approach to distributional limits for empirical optimal transport. ArXiv preprint. Available at arXiv:2202.12790.

[62] JORDAN, R., KINDERLEHRER, D. and OTTO, F. (1998). The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **29** 1–17. MR1617171 https://doi.org/10.1137/S0036141096303359

[63] KANTOROVITCH, L. (1942). On the translocation of masses. *C. R.* (*Dokl.*) *Acad. Sci. URSS* **37** 199–201. MR0009619

[64] KATO, K. (2009). Asymptotics for argmin processes: Convexity arguments. *J. Multivariate Anal.* **100** 1816–1829. MR2535389 https://doi.org/10.1016/j.jmva.2009.02.008

[65] KLATT, M., TAMELING, C. and MUNK, A. (2020). Empirical regularized optimal transport: Statistical theory and applications. *SIAM J. Math. Data Sci.* **2** 419–443. MR4105566 https://doi.org/10.1137/19M1278788

[66] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen*. *Inst. Math. Stat.* (*IMS*) *Collect.* **1** 282–292. IMS, Beachwood, OH. MR2462212 https://doi.org/10.1214/193940307000000202

[67] LEDOUX, M. (2019). On optimal matching of Gaussian samples. *J. Math. Sci.* **238** 495–522.

[68] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015 https://doi.org/10.1007/978-3-642-20212-4

[69] LEI, J. (2020). Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* **26** 767–798. MR4036051 https://doi.org/10.3150/19-BEJ1151

[70] LOVÁSZ, L. and VEMPALA, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30** 307–358. MR2309621 https://doi.org/10.1002/rsa.20135

[71] MANOLE, T., BALAKRISHNAN, S., NILES-WEED, J. and WASSERMAN, L. (2021). Plugin estimation of smooth optimal transport maps. Available at arXiv:2107.12364.

[72] MANOLE, T. and NILES-WEED, J. (2021). Sharp convergence rates for empirical optimal transport with smooth costs. Available at arXiv:2106.13181.

[73] MENA, G. and NILES-WEED, J. (2019). Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. *Adv. Neural Inf. Process. Syst.* **32**.

[74] MILMAN, E. (2009). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** 1–43. MR2507637 https://doi.org/10.1007/s00222-009-0175-9

[75] MOHAJERIN ESFAHANI, P. and KUHN, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Program.* **171** 115–166. MR3844536 https://doi.org/10.1007/s10107-017-1172-1

[76] NICKL, R. (2009). On convergence and convolutions of random signed measures. *J. Theoret. Probab.* **22** 38–56. MR2472004 https://doi.org/10.1007/s10959-008-0177-3

[77] NIETERT, S., GOLDFELD, Z. and KATO, K. (2021). Smooth p-Wasserstein distance: Structure, empirical approximation, and statistical applications. In *International Conference on Machine Learning*.

[78] PARR, W. C. and SCHUCANY, W. R. (1980). Minimum distance and robust estimation. *J. Amer. Statist. Assoc.* **75** 616–624. MR0590691

[79] PEYRE, R. (2018). Comparison between W_2 distance and $\dot{\mathrm{H}}^{-1}$ norm, and localization of Wasserstein distance. *ESAIM Control Optim. Calc. Var.* **24** 1489–1501. MR3922440 https://doi.org/10.1051/cocv/2017050

[80] POLLARD, D. (1980). The minimum distance method of testing. *Metrika* **27** 43–70. MR0563412 https://doi.org/10.1007/BF01893576

[81] POLYANSKIY, Y. and WU, Y. (2016). Wasserstein continuity of entropy and outer bounds for interference channels. *IEEE Trans. Inf. Theory* **62** 3992–4002. MR3515734 https://doi.org/10.1109/TIT.2016.2562630

[82] RAMDAS, A., TRILLOS, N. G. and CUTURI, M. (2017). On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy* **19** 47. MR3608466 https://doi.org/10.3390/e19020047

[83] RÖMISCH, W. (2004). Delta method, infinite dimensional. In *Encyclopedia of Statistical Sciences* Wiley, New York.

[84] RUBNER, Y., TOMASI, C. and GUIBAS, L. J. (2000). The Earth mover's distance as a metric for image retrieval. *Int*. *J*. *Comput*. *Vis*. **40** 99–121.

[85] SADHU, R., GOLDFELD, Z. and KATO, K. (2021). Limit distribution theory for the smooth 1-Wasserstein distance with applications. ArXiv preprint. Available at arXiv:2107.13494.

[86] SANDLER, R. and LINDENBAUM, M. (2011). Nonnegative matrix factorization with Earth mover's distance metric for image analysis. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **33** 1590–1602. https://doi.org/10.1109/TPAMI.2011.18

[87] SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser/Springer, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

[88] SANTAMBROGIO, F. (2017). {Euclidean, metric, and Wasserstein} gradient flows: An overview. *Bull*. *Math*. *Sci*. **7** 87–154. MR3625852 https://doi.org/10.1007/s13373-017-0101-1

[89] SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Stat*. *Surv*. **8** 45–114. MR3290441 https://doi.org/10.1214/14-SS107

[90] SHAPIRO, A. (1990). On concepts of directional differentiability. *J*. *Optim*. *Theory Appl*. **66** 477–487. MR1080259 https://doi.org/10.1007/BF00940933

[91] SHAPIRO, A. (1991). Asymptotic analysis of stochastic programs. *Ann*. *Oper*. *Res*. **30** 169–186.

[92] SOMMERFELD, M. and MUNK, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **80** 219–238. MR3744719 https://doi.org/10.1111/rssb.12236

[93] TALAGRAND, M. (1992). Matching random samples in many dimensions. *Ann*. *Appl*. *Probab*. **2** 846–856. MR1189420

[94] TALAGRAND, M. (1994). The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann*. *Probab*. **22** 919–959. MR1288137

[95] TAMELING, C., SOMMERFELD, M. and MUNK, A. (2019). Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *Ann*. *Appl*. *Probab*. **29** 2744–2781. MR4019874 https://doi.org/10.1214/19-AAP1463

[96] TOLSTIKHIN, I., BOUSQUET, O., GELLY, S. and SCHOELKOPF, B. (2018). Wasserstein auto-encoders. In *International Conference on Learning Representations*.

[97] VAN DER VAART, A. (1996). New Donsker classes. *Ann*. *Probab*. **24** 2128–2140. MR1415244 https://doi.org/10.1214/aop/1041903221

[98] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[99] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. Springer, Berlin.

[100] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

[101] VILLANI, C. (2008). *Optimal Transport*: *Old and New*. Springer, Berlin.

[102] WEED, J. and BACH, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. MR4003560 https://doi.org/10.3150/18-BEJ1065

[103] WEED, J. and BERTHET, Q. (2019). Estimation of smooth densities in Wasserstein distance. In *Conference on Learning Theory*.

[104] WOLFOWITZ, J. (1957). The minimum distance method. *Ann*. *Math*. *Stat*. **28** 75–88. MR0088126 https://doi.org/10.1214/aoms/1177707038

[105] ZHANG, Y., CHENG, X. and REEVES, G. (2021). Convergence of Gaussian-smoothed optimal transport distance with sub-gamma distributions and dependent samples. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* 2422–2430.