CONVERGENCE OF EMPIRICAL MEASURES FOR I.I.D. SAMPLES IN $W^{-\alpha,p}$

GAUTAM IYER AND RAGHAVENDRA VENKATRAMAN

ABSTRACT. Given N i.i.d. samples from a probability measure μ on $\mathbb{R}^d$, we study the rate of convergence of the empirical measure $\mu_N \to \mu$ in the negative Sobolev space $W^{-\alpha,p}$. When $W^{-\alpha,p}$ contains point measures (i.e. when $\alpha p > (p-1)d$), we show $\boldsymbol{E}\|\mu_N - \mu\|^p_{W^{-\alpha,p}} \leqslant C_d/N^{p/2}$ for an explicit dimensional constant C_d, and obtain a Gaussian tail bound. When $0 < \alpha p \leqslant d(p-1)$, we prove a similar result for Gaussian regularizations.

1. Introduction.

1.1. **Main Results.** Let μ be a probability measure on $\mathbb{R}^d$, and $X_1, \dots, X_N$ be N i.i.d. samples from μ. Let μ_N be the empirical measure

$$\mu_N \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{X_i}\,,$$

where δ_x denotes the Dirac δ-measure at x. The aim of this paper is to study the convergence of $\mu_N \to \mu$ in negative Sobolev spaces. For L^2 based Sobolev spaces, $\boldsymbol{E}\|\mu_N - \mu\|^2_{H^{-s}}$ can be computed exactly (see Proposition 2.2, below). The main result of this paper bounds the convergence rate of $\mu_N \to \mu$ in L^p based Sobolev spaces for all $p \in (1,\infty)$.

To state our result, let $\alpha > 0$, $p \in (1,\infty)$ and $W^{-\alpha,p} = W^{-\alpha,p}(\mathbb{R}^d)$ be the L^p based Sobolev space with index $-\alpha$. While there are many equivalent definitions of a norm in $W^{-\alpha,p}$, the version that is most convenient for us is based on the heat kernel, and stated precisely in Section 2, below. We show that when $\alpha p > (p-1)d$ (i.e. when the space $W^{-\alpha,p}$ contains point measures), we have $\boldsymbol{E}\|\mu_N - \mu\|^p_{W^{-\alpha,p}} \leqslant O(1/N^{p/2})$ and a Gaussian tail bound.

Theorem 1.1. *Let $p \in (1,\infty)$, $q = p/(p-1)$ be its Hölder conjugate, and suppose $\alpha > d/q = \alpha p/(p-1)$. There exists an absolute constant C (independent of μ, p and d) such that for every $N \in \mathbb{N}$ we have*

$$\left(\boldsymbol{E}\|\mu_N - \mu\|^p_{W^{-\alpha,p}}\right)^{1/p} \leqslant \frac{Cd^{1/p}}{\sqrt{N}}\|\delta_0\|_{W^{-\alpha,p}}\,. \tag{1.1}$$

Here δ_0 is the Dirac δ-distribution at 0. Additionally, for every $N \in \mathbb{N}$, the random variable $\|\mu_N - \mu\|_{W^{-\alpha,p}}$ satisfies the Gaussian tail bound

$$\boldsymbol{P}\left(\|\mu_N - \mu\|_{W^{-\alpha,p}} > \lambda\right) \leqslant 2\exp\left(\frac{-N\lambda^2}{Cd^{2/p}p\|\delta_0\|^2_{W^{-\alpha,p}}}\right),$$

for every $\lambda > 0$.

Notice that when $\alpha \leqslant d/q$, the empirical measure $\mu_N \notin W^{-\alpha,p}$, and so both the left and right hand side of (1.1) are infinite. In this case it is natural to study the convergence of the Gaussian regularized point measures instead. For any $\varepsilon > 0$, define $\mu^\varepsilon \stackrel{\text{def}}{=} \mu * \Phi_\varepsilon$ and $\mu^\varepsilon_N \stackrel{\text{def}}{=} \mu_N * \Phi_\varepsilon$, where Φ_ε is the density of a centered radial Gaussian with variance 2ε (see (2.3), below). When $\alpha \leqslant d/q$ we bound $\boldsymbol{E}\|\mu^\varepsilon_N - \mu\|^p_{W^{-\alpha,p}}$, and obtain a Gaussian tail bound.

Theorem 1.2. *Let $p \in (1,\infty)$, $\alpha > 0$ and suppose $\mu \in W^{-\alpha,p}$. There exists an absolute constant C (independent of μ, p and d) such that for every $N \in \mathbb{N}$, we have*

$$\left(\boldsymbol{E}\|\mu^\varepsilon_N - \mu\|^p_{W^{-\alpha,p}}\right)^{1/p} \leqslant \frac{Cd^{1/p}}{\sqrt{N}}\|\Phi_\varepsilon\|_{W^{-\alpha,p}} + \|\mu^\varepsilon - \mu\|_{W^{-\alpha,p}}\,. \tag{1.2}$$

Additionally, for every $N \in \mathbb{N}$, $\lambda > 0$ we have the Gaussian tail bound

$$\boldsymbol{P}\left(\|\mu^\varepsilon_N - \mu\|_{W^{-\alpha,p}} > \lambda\right) \leqslant 2\exp\left(\frac{-N\left(\lambda - \|\mu^\varepsilon - \mu\|_{W^{-\alpha,p}}\right)^2_+}{Cpd^{2/p}\|\Phi_\varepsilon\|^2_{W^{-\alpha,p}}}\right). \tag{1.3}$$

In order to use Theorem 1.2 one needs to understand the behavior of $\|\Phi_\varepsilon\|_{W^{-\alpha,p}}$ and $\|\mu^\varepsilon - \mu\|_{W^{-\alpha,p}}$ as $\varepsilon \to 0$. These are both well known and stated here as remarks for easy reference.

Remark 1.3. For every $\mu \in W^{-\alpha,p}$, we know $\|\mu^\varepsilon - \mu\|_{W^{-\alpha,p}}$ vanishes with ε, but the rate may be arbitrarily slow in ε. If, however, μ was more regular than $W^{-\alpha,p}$, then one has an explicit rate. For instance, if $\mu \in W^{-\alpha',p}$ for some $\alpha' < \alpha$, then standard regularity estimates imply

$$\|\mu^\varepsilon - \mu\|_{W^{-\alpha,p}} \leqslant C_{\alpha,d}\varepsilon^{\frac{\alpha-\alpha'}{2}\wedge 1}\,.$$

Remark 1.4. It is well known that

$$\|\Phi_\varepsilon\|_{W^{-\alpha,p}} \approx \begin{cases} |\ln\varepsilon|^{1/p} & \alpha = \frac{d}{q} \\ \dfrac{1}{\varepsilon^{\frac{d}{q}-\alpha}} & \alpha < \frac{d}{q}\,. \end{cases}$$

with constants that depend on α, p, and d. For convenience, we state this precisely in the appendix (see Proposition A.1), and present a proof with explicit bounds on the constants.

The constant $d^{1/p}$ appearing in (1.1) can be replaced with the p-th root of the best weak L^1 bound for the Hardy–Littlewood maximal function. The standard proof using the Vitali covering lemma shows this constant is at most 3^d, and a

2020 *Mathematics Subject Classification.* Primary: 60B10. Secondary: 60G50, 46E35.

Key words and phrases. empirical measure convergence, negative Sobolev space.

This work has been partially supported by the National Science Foundation under grants DMS-2342349, DMS-2406853, DMS-2407592. The authors also thank the SIAM Northern States Section and the Center for Nonlinear Analysis.

classical result of Stein and Strömberg [SS83] improves this to $O(d)$. To the best of our knowledge the sharp constant for this weak L^1 bound is not known.

Finally we mention that our results do not require any regularity or moment conditions on the probability measure μ, nor any bounds from below for μ on its support. Moreover a standard localization argument can be used to obtain similar rates of convergence when $\mathbb{R}^d$ is replaced by a closed manifold. We also note that without further assumptions matching lower bounds for (1.1) can not be obtained, as can be seen by taking $\mu = \delta_0$.

1.2. **Motivation and literature review.** The convergence of the empirical measure $\mu_N \to \mu$ is a central question in probability and statistics, and has been extensively studied. The law of large numbers shows that almost surely, the sequence $\mu_N \to \mu$ weakly in the space of probability measures.

There are a number of results quantifying the weak convergence of $\mu_N \to \mu$. One popular approach is to use the *Wasserstein distance*, which is natural as it measures the transportation cost of moving μ to ν. Explicitly for $p \in [1, \infty)$, the *Wasserstein distance* between two probability measures μ, ν (see for instance [Vil09]) is defined by

$$W_p(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \left(\boldsymbol{E}_{(X,Y)\sim\gamma} |X - Y|^p\right)^{1/p},$$

where $\Gamma(\mu, \nu)$ is the set of all couplings of μ and ν.

Seminal work of Ajtai, Kolmós, Tusnády [AKT84] studies the rate of convergence of $W_p(\mu_N, \mu)$ in the context of an optimal matching problem when $\mu = \mathrm{Unif}([0,1]^d)$ is the uniform distribution. It has since been studied by several authors [Tal21, Led23], and recent work of Goldman and Trevisan [GT21] shows

$$\lim_{N\to\infty} N^{1/d} \left(\boldsymbol{E} W_p^p(\mu_N, \mu)\right)^{1/p} \in (0, \infty), \quad \text{when } \mu = \mathrm{Unif}([0,1]^d),\ d \geqslant 3.$$

This gives both an upper and lower bound showing $(\boldsymbol{E} W_p^p(\mu_N, \mu))^{1/p} \approx 1/N^{p/d}$.

For general distributions, to the best of our knowledge, the sharp rate of convergence is not known. Under a suitable moment condition, one can show [DSS13, BLG14, FG15] the upper bound

$$(\boldsymbol{E} W_p^p(\mu_N, \mu))^{1/p} \leqslant \frac{C}{N^{1/d}} \quad \text{when } 1 \leqslant p < \frac{d}{2}.$$

This has been extended in [LZ21] for Gaussians for all $p \in [1, d)$, and by Divol [Div21] for all $p \in [1, \infty)$ for compactly supported distributions with strictly positive densities (see [LS89, GTS15] for the $p = \infty$ case). More generally Weed and Bach [WB19] show the bounds

$$\boldsymbol{E} W_p(\mu_N, \mu) \leqslant \frac{C}{N^{1/t}} \quad \text{and} \quad W_p(\mu_N, \mu) \geqslant \frac{C}{N^{1/s}}$$

for every $t > d_p^*(\mu)$ and $s < d_*(\mu)$. Here $d_p^*(\mu)$ and $d_*(\mu)$ are the upper and lower Wasserstein dimensions respectively (see Definition 4 in [WB19]). (Note the lower bound holds surely, and not on average.)

The Wasserstein bounds above suffer from the curse of dimensionality – the convergence rate becomes extremely slow when the dimension is large. Our aim in this paper is to look for results that allow for faster convergence in high dimensions. To hint at why this is to be expected, we note that for any (nice enough) test function f, we have the *dimension independent* identity

$$\boldsymbol{E}\left(\int_{\mathbb{R}^d} f \, d(\mu_N - \mu)\right)^2 = \frac{\mathrm{Var}_\mu(f)}{N},$$

which is the foundation of several Monte Carlo methods used today (see for instance [Liu08]). Of course, for the above to be defined, the function f needs to be regular enough to allow for evaluation at points. This suggests dimension independent bounds may require the use of spaces which discount fine scale structure.

One such example is a maximum mean discrepancy (MMD) distance, which is known to quantify weak convergence [SGBSM23]. With this approach, the probability measures are embedded into a Hilbert space and the norm in this space is used to measure distance. In this setting the expected value of the MMD distance squared can be computed exactly using the modulated energy [HCRS25] (see also Proposition 2.2, below).

The present paper aims to study convergence rates outside the Hilbert space setting, where an exact calculation is not available. The simplest and most natural example is to consider *negative Sobolev spaces*, which have been used by many authors to quantify weak convergence [Thi12]. Here we obtain rates of convergence with a dimension independent exponent. More precisely, Theorem 1.1 shows the convergence rate in L^p based Sobolev spacesis $O(C_d/\sqrt{N})$, for an explicit dimensional constant C_d. The disadvantage of using negative Sobolev spaces is that for some indices the space may not contain point measure μ_N, or even the target measure μ. We address this by studying the convergence of Gaussian regularized point measures instead, as stated in Theorem 1.2.

Plan of this paper. In Section 2 we state the main result of this paper (Theorem 2.1) which yields both Theorem 1.1 and 1.2. We prove Theorem 2.1 in Section 3 when $\alpha \notin \mathbb{N}$, and in Section 4 when $\alpha \in \mathbb{N}$. In Section 5 we prove Proposition 2.2, which is a stronger version of Theorem 2.1 for $p = 2$. Finally in the appendix we compute the negative Sobolev norms of Gaussians, as stated in Remark 1.4

Acknowledgements. The authors thank Giovanni Leoni, Dejan Slepčev, and Matthew Rosenzweig for helpful discussions.

2. Main results and proofs.

Before proving Theorems 1.1 and 1.2, we define the $W^{-\alpha,p}$ norm that is used in the statement of the theorem. There are several equivalent definitions of the norm in $W^{-\alpha,p}$ (see for instance [Tri92]), but the version that is most convenient for us is the one from Appendix D in [AKM19] based on heat kernels. For $\alpha > 0$, $p \in (1, \infty)$ we define the $W^{-\alpha,p}$ norm by

$$\|u\|_{W^{-\alpha,p}} = \begin{cases} \|u\|_{\mathcal{W}^{-\alpha,p}} & \alpha \notin \mathbb{N}, \\ \|u\|_{\mathscr{W}^{-\alpha,p}} & \alpha \in \mathbb{N}. \end{cases} \tag{2.1}$$

Here the norms $\|\cdot\|_{\mathcal{W}^{-\alpha,p}}$ and $\|\cdot\|_{\mathscr{W}^{-\alpha,p}}$ are defined by

$$\|u\|_{\mathcal{W}^{-\alpha,p}} \stackrel{\text{def}}{=} \left(\int_0^1 \left(t^{\frac{\alpha}{2}}\|u*\Phi_t\|_{L^p(\mathbb{R}^d)}\right)^p \frac{dt}{t}\right)^{\frac{1}{p}}, \tag{2.2a}$$

$$\|u\|_{\mathscr{W}^{-\alpha,p}} \stackrel{\text{def}}{=} \left(\int_{\mathbb{R}^d}\left(\int_0^1 \left(t^{\frac{\alpha}{2}}|u*\Phi_t|\right)^2 \frac{dt}{t}\right)^{\frac{p}{2}} dx\right)^{\frac{1}{p}}, \tag{2.2b}$$

and Φ is the standard Euclidean heat kernel

$$\Phi_t(x) \stackrel{\text{def}}{=} \frac{1}{(4\pi t)^{d/2}}\exp\Big(-\frac{|x|^2}{4t}\Big), \quad x\in\mathbb{R}^d,\ t>0, \tag{2.3}$$

For convenience, when $t=0$ we will use Φ_0 to denote the Dirac δ distribution at 0. Appendix D in [AKM19] provides a detailed proof showing (2.1) is equivalent to the standard $W^{-\alpha,p}$ norm based on duality.

Both Theorems 1.1 and 1.2 follow immediately from the following result for Gaussian regularizations.

Theorem 2.1. *There exists an absolute constant C, independent of d and μ, such that for any $p\in(1,\infty)$ and $\varepsilon\geqslant 0$ we have the moment bound*

$$\big(\boldsymbol{E}\|\mu_N^\varepsilon-\mu^\varepsilon\|_{W^{-\alpha,p}}^p\big)^{1/p} \leqslant \frac{Cd^{1/p}\sqrt{p}}{\sqrt{N}}\|\Phi_\varepsilon\|_{W^{-\alpha,p}}, \tag{2.4}$$

and the Gaussian tail bound

$$\boldsymbol{P}\big(\|\mu_N^\varepsilon-\mu^\varepsilon\|_{W^{-\alpha,p}}>\lambda\big)\leqslant 2\exp\left(\frac{-N\lambda^2}{Cd^{2/p}p\|\Phi_\varepsilon\|_{W^{-\alpha,p}}^2}\right), \quad \text{for every } \lambda>0. \tag{2.5}$$

If $\alpha p>(p-1)d$ and $\mu\in W^{-\alpha,p}$, then both (2.4) and (2.5) also hold for $\varepsilon=0$.

Theorems 1.1 and 1.2 follow immediately from Theorem 2.1.

Proof of Theorems 1.1 and 1.2. When $\alpha p>d/q$, Theorem 2.1 with $\varepsilon=0$ reduces to Theorem 1.1. For Theorem 1.2, the triangle inequality implies

$$\big(\boldsymbol{E}\|\mu_N^\varepsilon-\mu\|_{W^{-\alpha,p}}^p\big)^{1/p} \leqslant \big(\boldsymbol{E}\|\mu_N^\varepsilon-\mu^\varepsilon\|_{W^{-\alpha,p}}^p\big)^{1/p} + \|\mu^\varepsilon-\mu\|_{W^{-\alpha,p}}.$$

Now (1.2) and (1.3) follow from (2.4) and (2.5) respectively. □

We prove Theorem 2.1 in Section 3 when $\alpha\notin\mathbb{N}$, and in Section 4 when $\alpha\in\mathbb{N}$, and these are the main contributions of this paper. To briefly comment on the proof, we note that the standard dual norm on $W^{-\alpha,p}$ is not convenient to work with in this context, which is why we use the heat kernel based bounds (2.2a)–(2.2b). As we will see shortly, the main tool used in the proof is a bound on the *subgaussian norm* of the random variable $\Phi_t(x-X_1)$, treating $t\in(0,1)$ and $x\in\mathbb{R}^d$ as parameters. Of course, since $\Phi_t\in L^\infty$, the subgaussian norm is easily bounded by $\|\Phi_t\|_{L^\infty}$. However, this bound is too crude to yield Theorem 2.1. What we need is a bound that decays nicely in *both* the parameters x and t. We prove this bound in Lemma 3.2 (below) and control the subgaussian norm in terms of the Hardy–Littlewood maximal function of μ. We then use the weak-L^1 bound on the maximal function to prove Theorem 2.1.

Before delving into the proof, we briefly digress and state the analog of Theorem 2.1 when $p=2$. In this case the expressions in both (2.2a) and (2.2b) agree, and there is a slightly stronger result that can be obtained with a simpler proof. To simplify notation let $H^{-\alpha}$ denote the L^2-based Sobolev space with index $-\alpha$, and define

$$\|u\|_{H^{-\alpha}} \stackrel{\text{def}}{=} \|u\|_{\mathcal{W}^{-\alpha,2}} = \|u\|_{\mathscr{W}^{-\alpha,2}} = \left(\int_0^1 \left(t^{\frac{\alpha}{2}}\|u*\Phi_t\|_{L^2(\mathbb{R}^d)}\right)^2 \frac{dt}{t}\right)^{\frac{1}{2}},$$

whether or not $\alpha\in\mathbb{N}$. Our next result shows that when $p=2$ the bound (2.4) can be made exact equality, and (2.5) can be proved without any dimensionally dependent constants.

Proposition 2.2. *For every $\alpha>0$, $\varepsilon>0$ and every $N\in\mathbb{N}$ we have*

$$\boldsymbol{E}\|\mu_N^\varepsilon-\mu^\varepsilon\|_{H^{-\alpha}}^2 = \frac{1}{N}\Big(\|\Phi_\varepsilon\|_{H^{-\alpha}}^2-\|\mu_\varepsilon\|_{H^{-\alpha}}^2\Big). \tag{2.6}$$

Additionally, we have the Gaussian tail bound

$$\boldsymbol{P}\big(\|\mu_N^\varepsilon-\mu^\varepsilon\|_{H^{-\alpha}}>\lambda\big)\leqslant 2\exp\left(\frac{-N\lambda^2}{C\|\Phi_\varepsilon\|_{H^{-\alpha}}^2}\right), \quad \text{for every } \lambda>0. \tag{2.7}$$

If $\alpha>\frac{d}{2}$ and $\mu\in H^{-\alpha}$ then both (2.6) and (2.7) also hold for $\varepsilon=0$.

3. Convergence in $\mathcal{W}^{-\alpha,p}$.

The aim of this section is to prove Theorem 2.1 when $\alpha\notin\mathbb{N}$. We do this by proving a slightly more general result. Let $\mathcal{W}^{-\alpha,p}=\mathcal{W}^{-\alpha,p}(\mathbb{R}^d)$ be the space of all tempered distributions for which the norm in (2.2a) is finite. We will now show that Theorem 2.1 holds with $\|\cdot\|_{W^{-\alpha,p}}$ replaced with $\|\cdot\|_{\mathcal{W}^{-\alpha,p}}$, whether or not $\alpha\in\mathbb{N}$. Explicitly, the result we prove is as follows.

Proposition 3.1. *There exists an absolute constant C, independent of d and μ, such that for any $p\in(1,\infty)$ and $\varepsilon\geqslant 0$ we have the moment bound*

$$\big(\boldsymbol{E}\|\mu_N^\varepsilon-\mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}^p\big)^{1/p} \leqslant \frac{Cd^{1/p}\sqrt{p}}{\sqrt{N}}\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}, \tag{3.1}$$

and the Gaussian tail bound

$$\boldsymbol{P}\big(\|\mu_N^\varepsilon-\mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}>\lambda\big)\leqslant 2\exp\left(\frac{-N\lambda^2}{Cd^{2/p}p\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}^2}\right), \quad \text{for every } \lambda>0. \tag{3.2}$$

If $\alpha p>(p-1)d$ and $\mu\in\mathcal{W}^{-\alpha,p}$, then both (3.1) and (3.2) also hold for $\varepsilon=0$.

We state this version here for convenience, and note that it immediately implies Theorem 2.1 for $\alpha\notin\mathbb{N}$.

3.1. Proof of Proposition 3.1.

As mentioned earlier, the main idea behind the proof of Proposition 3.1 is to consider the random variable $\Phi_t(x-X_1)$, treating $x\in\mathbb{R}^d$ and $t\in(0,1)$ as parameters, and show that the *subgaussian norm* decays suitably in both x and t. Recall (see for instance Section 2.5 in [Ver18]) that if Z is a one dimensional random variable its subgaussian norm, denoted by $\|Z\|_{\psi_2}$ is defined by

$$\|Z\|_{\psi_2} = \inf\Big\{c>0\ \Big|\ \boldsymbol{E}\exp\Big(\frac{Z^2}{c^2}\leqslant 2\Big)\Big\}.$$

What we need to prove Theorem 2.1 is to show that $\|\Phi_t(x-X_1)\|_{\psi_2}$ essentially decays like a Gaussian in x,t. This is our first lemma.

Lemma 3.2. *For any $t \in (0,1), x \in \mathbb{R}^d$ we have*

$$\sigma(x,t) \stackrel{\text{def}}{=} \|\Phi_t(x - X_1)\|_{\Psi_2} \leqslant \frac{C}{(4\pi t)^{d/2}} \exp\left(\frac{-1}{4t\big(V_d M\mu(x)\big)^{2/d}}\right), \tag{3.3}$$

where and $M\mu$ denotes the Hardy–Littlewood maximal function of μ. Here V_d denotes the volume of the unit ball in $\mathbb{R}^d$.

Here clarify $\|\Phi_t(x - X_1)\|_{\psi_2}$ denotes the subgaussian norm of the random variable $\Phi_t(x - X_1)$, where x, t are fixed, deterministic, parameters. We prove Lemma 3.2 in Section 3.2, below.

To see how (3.3) decays in x, t, we note that the maximal function of any probability measure "usually" decays like $1/|x|^d$ as $|x| \to \infty$. When this happens, $\sigma(x,t)$ will satisfy a Gaussian upper bound. Of course, the maximal function does not always decay like $1/|x|^d$, and obtaining the decay we need from (3.3) requires a little work. We do this by combining (3.3) with the weak L^1 bound for the maximal function. Explicitly, what we need is the following lemma.

Lemma 3.3. *There exists an absolute constant C, independent of d, μ, such that for every $p \in (1,\infty)$, $\alpha > 0$, $\varepsilon \geqslant 0$ we have*

$$\left(\int_0^1 t^{\frac{\alpha p}{2}-1} \|\sigma(\cdot, t+\varepsilon)\|_{L^p}^p \, dt\right)^{1/p} \leqslant C d^{1/p} \|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}. \tag{3.4}$$

Remark 3.4. Note that when $\alpha > d/q$ (i.e. when $\alpha p > (p-1)d$), the right hand side of (3.4) is finite when $\varepsilon = 0$.

Momentarily postponing the proofs of Lemmas 3.2 and 3.3 we now prove Proposition 3.1. Here and throughout this paper we use $C > 0$ to denote an absolute constant that does not depend on the measure μ, or the parameters $p > 1, d \in \mathbb{N}, \alpha > 0$, and may increase from line to line.

Proof of Proposition 3.1. For any $x \in \mathbb{R}^d$, $t \in \mathbb{R}$, define the random variable $S_N(x,t)$ by

$$S_N(x,t) \stackrel{\text{def}}{=} (\mu_N - \mu) * \Phi_t(x) = \frac{1}{N}\sum_{i=1}^N \big(\Phi_t(x - X_i) - \mu_t(x)\big). \tag{3.5}$$

Now we compute

$$\boldsymbol{E}\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}^p = \int_{t=0}^1 \int_{x\in\mathbb{R}^d} t^{\frac{\alpha p}{2}-1} \boldsymbol{E}|S_N(x, t+\varepsilon)|^p \, dx\, dt. \tag{3.6}$$

Since the p-th moment of a random variable is controlled by $\sqrt{p}$ times its subgaussian norm, this implies

$$\boldsymbol{E}\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}^p \leqslant C^p p^{p/2} \int_{t=0}^1 \int_{x\in\mathbb{R}^d} t^{\frac{\alpha p}{2}-1} \|S_N(x, t+\varepsilon)\|_{\psi_2}^p \, dx\, dt. \tag{3.7}$$

Now by Lemma 3.2 and Hoeffding's inequality the random variable $S_N(x, t+\varepsilon)$ is subgaussian with norm controlled by

$$\|S_N(x, t+\varepsilon)\|_{\psi_2} \leqslant \frac{C\sigma(x, t+\varepsilon)}{\sqrt{N}}. \tag{3.8}$$

Using this in (3.7) implies

$$\big(\boldsymbol{E}\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}^p\big)^{1/p} \leqslant \frac{C\sqrt{p}}{\sqrt{N}}\Big(\int_0^1 t^{\frac{\alpha p}{2}-1} \|\sigma(\cdot, t+\varepsilon)\|_{L^p}^p \, dt\Big)^{1/p}.$$

Combining this with Lemma 3.3 yields (3.1) as desired.

We now prove the Gaussian tail bound (3.2). Notice

$$\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}} = f_N(X_1, \dots, X_N)$$

where

$$f_N(y_1, \dots, y_N) \stackrel{\text{def}}{=} \|g_N(\cdot, \cdot; y_1, \dots, y_N)\|_*$$

$$g_N(x, t; y_1, \dots, y_N) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \Phi_{t+\varepsilon}(x - y_i) - \mu_{t+\varepsilon}(x), \tag{3.9}$$

and $\|\cdot\|_*$ is the weighted L^p norm

$$\|h\|_* \stackrel{\text{def}}{=} \left(\int_0^1 \int_{\mathbb{R}^d} t^{\frac{\alpha p}{2}-1} |h(x,t)|^p \, dx\, dt\right)^{1/p}.$$

We now claim for any $i \in \{1, \dots, N\}$, $y_i, y_i' \in \mathbb{R}^d$ the function f_N satisfies

$$|f_N(y_1, \dots, y_N) - f_N(y_1, \dots, y_{i-1}, y_i', y_{i+1}, \dots, y_N)| \leqslant \frac{2}{N}\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}. \tag{3.10}$$

To prove this we note that by symmetry it suffices to prove (3.10) for $i = 1$. Let $y_1, y_1' \in \mathbb{R}^d$, and $\tilde{y} = (y_2, \dots, y_N)$. Since $\|\cdot\|_*$ is a norm, the reverse triangle inequality implies

$$|f_N(y_1, \tilde{y}) - f_N(y_1', \tilde{y})| \leqslant \|h(\cdot, \cdot; y_1, y_1')\|_*,$$

where

$$h(x, t; y_1, y_1') \stackrel{\text{def}}{=} \frac{1}{N}\big(\Phi_{t+\varepsilon}(x - y_1) - \Phi_{t+\varepsilon}(x - y_1')\big). \tag{3.11}$$

Clearly

$$\|h(\cdot, \cdot; y_1, y_1')\|_* \leqslant \frac{2}{N}\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}$$

from which (3.10) follows.

Since the bounded differences property (3.10) holds, we may now apply McDiarmid's inequality to show $\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}$ is subgaussian with

$$\Big\|\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}} - \boldsymbol{E}\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}\Big\|_{\psi_2} \leqslant \frac{C}{\sqrt{N}}\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}.$$

Thus

$$\begin{aligned}\Big\|\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}\Big\|_{\psi_2} &\leqslant \frac{C}{\sqrt{N}}\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}} + C\boldsymbol{E}\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}} \\ &\leqslant \frac{C}{\sqrt{N}}\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}} + C\big(\boldsymbol{E}\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}^p\big)^{1/p}.\end{aligned}$$

Using (3.1) this implies $\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}}$ is subgaussian with

$$\Big\| \|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathcal{W}^{-\alpha,p}} \Big\|_{\psi_2} \leqslant \frac{Cd^{1/p}\sqrt{p}}{\sqrt{N}} \|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}$$

This immediately implies (3.2), concluding the proof. □

It remains to prove Lemmas 3.2 and 3.3, which we do in subsequent sections.

3.2. **The subgaussian norm of $\Phi_t(x - X_1)$ (Lemma 3.2).** The key step in the proof of Proposition 3.1 was the bound on the subgaussian norm of $\Phi_t(x - X_1)$. We now present its proof.

Proof of Lemma 3.2. We know (see for instance [Ver18, Proposition 2.6.6 (i)]) that if for some $K_1(x,t) > 0$ and all $\lambda > 0$ we have

$$\boldsymbol{P}\big(|\Phi_t(x - X_1)| > \lambda\big) \leqslant 2\exp\Big(-\frac{\lambda^2}{K_1(x,t)^2}\Big)\,, \tag{3.12}$$

then there exists an absolute constant C such that

$$\sigma(x,t) \stackrel{\text{def}}{=} \|\Phi_t(x - X_1)\|_{\psi_2} \leqslant CK_1(x,t)\,. \tag{3.13}$$

Next we note that

$$\lambda \leqslant K_1(x,t)\sqrt{\ln 2} \implies 2\exp\Big(-\frac{\lambda^2}{K_1(x,t)^2}\Big) \geqslant 1 \geqslant \boldsymbol{P}\big(|\Phi_t(x - X_1)| > \lambda\big)\,.$$

Also,

$$\lambda \geqslant (4\pi t)^{-d/2} \implies \boldsymbol{P}\big(|\Phi_t(x - X_1)| > \lambda\big) = 0 \leqslant 2\exp\Big(-\frac{\lambda^2}{K_1(x,t)^2}\Big)\,,$$

which again verifies (3.12). Thus we only need to ensure that (3.12) holds for all λ such that

$$K_1\sqrt{\ln 2} < \lambda < \frac{1}{(4\pi t)^{d/2}}\,. \tag{3.14}$$

Now define $r = r(t,\lambda)$ by

$$r \stackrel{\text{def}}{=} \sqrt{4t\ln\Big(\frac{1}{(4\pi t)^{\frac{d}{2}}\lambda}\Big)}\,,$$

and note

$$\boldsymbol{P}\big(\Phi_t(x - X_1) > \lambda\big) = \boldsymbol{P}\big(X_1 \in B(x,r)\big) = \mu\big(B(x,r)\big) \leqslant V_d M\mu(x) r^d\,.$$

Thus (3.12) will hold for all λ satisfying (3.14) provided we can ensure

$$V_d M\mu(x) r^d \leqslant 2\exp\Big(-\frac{\lambda^2}{K_1(x,t)^2}\Big)\,.$$

Clearly this is equivalent to showing that for all

$$0 < r < \sqrt{4t\ln\Big(\frac{1}{(4\pi t)^{\frac{d}{2}}K_1(x,t)(\ln 2)^{1/2}}\Big)}\,, \tag{3.15}$$

we have

$$dK_1^2 \ln\Big(\frac{\eta(x)}{r}\Big) \geqslant \frac{1}{(4\pi t)^d} e^{-r^2/2t} \quad \text{where} \quad \eta(x) = \Big(\frac{2}{V_d M\mu(x)}\Big)^{1/d}\,. \tag{3.16}$$

Choose

$$K_1(x,t) = \frac{1}{(4\pi t)^{d/2}} e^{-\gamma^2\eta(x)^2/4t}\,, \tag{3.17}$$

for some small constant $\gamma > 0$ that will be chosen shortly. Note that both the left and right hand side of (3.16) are decreasing functions of r. Moreover, the inequality (3.16) clearly holds for $r = 0$. Thus (3.16) will hold for all r in the interval (3.15), provided it holds for $r = r_1$, where

$$r_1 \stackrel{\text{def}}{=} \sqrt{4t\ln\Big(\frac{1}{(4\pi t)^{\frac{d}{2}}K_1(x,t)(\ln 2)^{1/2}}\Big)} = \sqrt{\gamma^2\eta(x)^2 + 2t\ln\Big(\frac{1}{\ln 2}\Big)}\,.$$

When $r = r_1$, and K_1 is defined by (3.17), the inequality (3.16) is equivalent to

$$de^{-\gamma^2\eta(x)^2/2t}\ln\Big(\frac{\eta(x)}{r_1}\Big) \geqslant e^{-r_1^2/2t}\,.$$

Through a direct calculation, this in turn is equivalent to having $\gamma \leqslant \gamma_1$, where

$$\gamma_1 \stackrel{\text{def}}{=} \sqrt{\frac{1}{2^{1/d}} - \frac{2t\ln\big(\frac{1}{\ln 2}\big)}{\eta(x)^2}}\,. \tag{3.18}$$

We now divide the analysis into two cases.

Case I: $\eta(x)^2 > 2^{1+1/d}t|\ln\ln 2|$. In this case the quantity under the square-root in (3.18) is positive and we choose $\gamma = \gamma_1 > 0$. Using the definition of η (from (3.16)) this implies

$$K_1(x,t) = \frac{1}{(4\pi t)^{d/2}\sqrt{\ln 2}}\exp\Big(\frac{-2^{3/d}}{4t\big(V_d M\mu(x)\big)^{2/d}}\Big)\,.$$

Now the argument above shows (3.12) holds for all $\lambda > 0$, and so using (3.13) we obtain (3.3) as desired.

Case II: $\eta(x)^2 \leqslant 2^{1+1/d}t|\ln\ln 2|$. In this case the definition of η (in (3.16)) implies

$$\exp\Big(\frac{-1}{4t\big(V_d M\mu(x)\big)^{2/d}}\Big) \geqslant \frac{1}{C_1'(d)} \quad \text{where} \quad C_1'(d) \stackrel{\text{def}}{=} \exp\Big(\frac{|\ln\ln 2|}{2^{1+1/d}}\Big)\,.$$

Now

$$\begin{aligned}\|\Phi_t(x - X_1)\|_{\psi_2} &\leqslant C\|\Phi_t(x - X_1)\|_{L^\infty} \leqslant \frac{C}{(4\pi t)^{d/2}} \\ &\leqslant \frac{CC_1'(d)}{(4\pi t)^{d/2}}\exp\Big(\frac{-1}{4t\big(V_d M\mu(x)\big)^{2/d}}\Big)\,,\end{aligned}$$

which immediately implies (3.3) as desired. □

3.3. **Decay of σ (Lemma 3.3).** Given Lemma 3.2, one can prove Lemma 3.3 quickly using the weak L^1 bound on the Hardy–Littlewood maximal function. We present the proof below.

Proof of Lemma 3.3. Define the function F by

$$F(\lambda,t) = \frac{1}{(4\pi t)^{d/2}} \exp\left(\frac{-1}{4t(V_d\lambda)^{2/d}}\right), . \tag{3.19}$$

Using (3.3) and the layer-cake formula we see

$$\|\sigma(\cdot,t)\|_{L^p}^p \leqslant C^p \int_{\lambda=0}^{\infty} \partial_\lambda(F^p)\left|\{M\mu>\lambda\}\right| d\lambda \leqslant C^p d \int_{\lambda=0}^{\infty} \partial_\lambda(F^p)\,\frac{d\lambda}{\lambda}\,.$$

The last inequality above followed from the Stein–Strömberg weak L^1 bound for the maximal function (see Theorem 3 in [SS83]). Substituting $r = 1/(V_d\lambda)^{1/d}$ and using

$$\frac{d\lambda}{\lambda} = (-d)\frac{dr}{r} \qquad \partial_\lambda = \frac{-V_d}{d} r^{d+1}\partial_r \tag{3.20}$$

we obtain

$$\|\sigma(\cdot,t)\|_{L^p}^p \leqslant -C^p d V_d \int_0^\infty \partial_r(F^p)\, r^d\, dr = C^p d^2 V_d \int_0^\infty F^p\, r^{d-1}\, dr\,. \tag{3.21}$$

The last equality above was obtained by using the fact that $F(r)r^d \to 0$ as both $r \to 0$ and $r \to \infty$, and integrating by parts.

Next we notice that using radial coordinates to compute $\|\Phi_t\|_{L^p}^p$ gives

$$\|\Phi_t\|_{L^p}^p = \frac{1}{(4\pi t)^{pd/2}} \int_{\mathbb{R}^d} \exp\Big(\frac{-pr^2}{4t}\Big) A_{d-1} r^{d-1}\, dr\,,$$

where A_{d-1} is the surface area of the unit sphere in $\mathbb{R}^d$. Using the identity $dV_d = A_{d-1}$ and (3.19) this implies

$$\|\Phi_t\|_{L^p}^p = dV_d \int_0^\infty F^p\, r^{d-1}\, dr\,.$$

Substituting this in (3.21) implies

$$\|\sigma(\cdot,t)\|_{L^p}^p \leqslant C^p d \|\Phi_t\|_{L^p}^p\,,$$

which immediately yields (3.4), concluding the proof. □

4. Convergence in $\mathscr{W}^{-\alpha,p}$.

We now prove Theorem 2.1 when $\alpha \in \mathbb{N}$. As before, we do this by proving a slightly more general result. Let $\mathscr{W}^{-\alpha,p} = \mathscr{W}^{-\alpha,p}(\mathbb{R}^d)$ be the space of all tempered distributions u for which the norm in (2.2b) is finite. We now show that Theorem 2.1 holds with $\|\cdot\|_{W^{-\alpha,p}}$ replaced with $\|\cdot\|_{\mathscr{W}^{-\alpha,p}}$, whether or not $\alpha \in \mathbb{N}$. This will immediately imply Theorem 2.1 for $\alpha \in \mathbb{N}$, and hence conclude the proof of Theorem 2.1.

Proposition 4.1. *There exists an absolute constant C, independent of d and μ, such that for any $p \in (1,\infty)$ and $\varepsilon \geqslant 0$ we have the moment bound*

$$\big(\boldsymbol{E}\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathscr{W}^{-\alpha,p}}^p\big)^{1/p} \leqslant \frac{Cd^{1/p}\sqrt{p}}{\sqrt{N}}\|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}\,, \tag{4.1}$$

and the Gaussian tail bound

$$\boldsymbol{P}\big(\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathscr{W}^{-\alpha,p}} > \lambda\big) \leqslant 2\exp\left(\frac{-N\lambda^2}{Cd^{2/p}p\|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}^2}\right), \quad \text{for every } \lambda>0\,. \tag{4.2}$$

If $\alpha p > (p-1)d$ and $\mu \in \mathscr{W}^{-\alpha,p}$, then both (4.1) and (4.2) also hold for $\varepsilon = 0$.

While the proof of Proposition 4.1 uses the same strategy as that of Proposition 3.1, there is one added difficulty. When working in $\mathcal{W}^{-\alpha,p}$, we note $\boldsymbol{E}\|\mu_N - \mu\|_{\mathcal{W}^{-\alpha,p}}^p$ can be written in terms of $\boldsymbol{E}S_N$ as in (3.6), after which we use Hoeffding's inequality to proceed. In the space $\mathscr{W}^{-\alpha,p}$ this is no longer possible. We work around this using a Minkowski type inequality and bound the p-th moment $\boldsymbol{E}\|\mu_N - \mu\|_{\mathscr{W}^{-\alpha,p}}^p$ in terms of an integral of σ. For clarity of presentation, we first state a lemma bounding $\|\sigma\|_{\mathscr{W}^{-\alpha,p}}$.

Lemma 4.2. *There exists an absolute constant C, independent of d, μ, such that for every $p \in (1,\infty)$, $\alpha > 0$, $\varepsilon \geqslant 0$ we have*

$$\left(\int_{\mathbb{R}^d}\left(\int_0^1 t^{\alpha-1}\sigma(x,t+\varepsilon)^2\,dt\right)^{p/2} dx\right)^{1/p} \leqslant Cd^{1/p}\|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}\,. \tag{4.3}$$

Remark 4.3. As with Remark 3.4, note that when $\alpha > d/q$ (i.e. when $\alpha p > (p-1)d$), the right hand side of (4.3) is finite when $\varepsilon = 0$.

Momentarily postponing the proof of Lemma 4.2, we now prove Proposition 4.1.

Proof of Proposition 4.1. Notice

$$\boldsymbol{E}\|\mu_N^\varepsilon - \mu\|_{\mathscr{W}^{-\alpha,p}}^p = \int_{\mathbb{R}^d} \boldsymbol{E}J_{N,\alpha,\varepsilon}(x)^p\,dx \tag{4.4}$$

where $J_{N,\alpha,\varepsilon}(x)$ is the random variable

$$J_{N,\alpha,\varepsilon}(x) \stackrel{\text{def}}{=} \left(\int_0^1 t^{\alpha-1}S_N(x,t+\varepsilon)^2\,dt\right)^{1/2}$$

and S_N is defined in (3.5). Now let $\|\cdot\|_{\psi_1}$ denote the subexponential norm, and recall (see for instance [Ver18], Lemma 2.7.6) that for any random variable Z we have

$$\|Z^2\|_{\psi_1} = \|Z\|_{\psi_2}^2\,.$$

Since $\|\cdot\|_{\psi_1}$ is a norm, the above identity and Minkowski's inequality imply

$$\begin{aligned}\|J_{N,\alpha,\varepsilon}(x)\|_{\psi_2} = \|J_{N,\alpha,\varepsilon}(x)^2\|_{\psi_1}^{1/2} &\leqslant \left(\int_0^1 t^{\alpha-1}\|S_N(x,t+\varepsilon)^2\|_{\psi_1}\,dt\right)^{1/2}\\ &= \left(\int_0^1 t^{\alpha-1}\|S_N(x,t+\varepsilon)\|_{\psi_2}^2\,dt\right)^{1/2}\\ &\stackrel{(3.8)}{\leqslant} \frac{C}{\sqrt{N}}\left(\int_0^1 t^{\alpha-1}\sigma(x,t+\varepsilon)^2\,dt\right)^{1/2}.\end{aligned} \tag{4.5}$$

Substituting the above in (4.4) implies

$$\big(\boldsymbol{E}\|\mu_N^\varepsilon - \mu\|_{\mathscr{W}^{-\alpha,p}}^p\big)^{1/p} \leqslant C\sqrt{p}\Big\|\|J_{N,\alpha,\varepsilon}\|_{\psi_2}\Big\|_{L^p}$$

$$\overset{(4.5)}{\leqslant} \frac{C\sqrt{p}}{\sqrt{N}} \left(\int_{\mathbb{R}^d} \left(\int_0^1 t^{\alpha-1} \sigma(x, t+\varepsilon)^2 \, dt \right)^{p/2} dx \right)^{1/p} .$$

Combined with Lemma 4.2 this yields (4.1) as desired.

Next we turn to the Gaussian tail bound (4.2). The argument is almost identical to the proof of (3.2). As with the proof of (3.2), the starting point is to notice that

$$\|\mu_N^\varepsilon - \mu^\varepsilon\|_{\mathscr{W}^{-\alpha,p}} = f_N(X_1, \cdots, X_N),$$

where

$$f_N(y_1, \cdots, y_N) \overset{\text{def}}{=} \|g_N(\cdot, \cdot; y_1, \cdots, y_N)\|_\star .$$

Here g_N is the same function defined in (3.9), and $\|\cdot\|_\star$ is the weighted norm

$$\|h\|_\star \overset{\text{def}}{=} \left(\int_{\mathbb{R}^d} \left(\int_0^1 t^{\alpha-1} h(x,t)^2 \, dt \right)^{\frac{p}{2}} dx \right)^{\frac{1}{p}} .$$

We now claim for any $i \in \{1, \dots, N\}$, $y_i, y_i' \in \mathbb{R}^d$ the function f_N satisfies

$$|f_N(y_1, \dots, y_N) - f_N(y_1, \dots, y_{i-1}, y_i', y_{i+1}, \dots, y_N)| \leqslant \frac{2}{N} \|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}} . \tag{4.6}$$

The proof of (3.10) is very similar to that of (3.10). Indeed, by symmetry it suffices to prove (4.6) for $i = 1$. For this choose any $y_1, y_1' \in \mathbb{R}^d$ and $\tilde{y} = (y_2, \cdots, y_N) \in (\mathbb{R}^d)^{N-1}$. The reverse triangle inequality implies

$$|f_N(y_1, \tilde{y}) - f_N(y_1', \tilde{y})| \leqslant \|h(\cdot, \cdot; y_1, y_1')\|_\star ,$$

where h is the same function defined in (3.11). This immediately implies

$$\|h(\cdot, \cdot; y_1, y_1')\|_\star \leqslant \frac{2}{N} \|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}} ,$$

from which (4.6) follows.

Given (4.6) we obtain (4.2) using McDiarmid's inequality and the moment bound (4.1) exactly as in the proof of (3.2). □

It remains to prove Lemma 4.2. The proof is a direct consequence of Lemma 3.2 and the Stein–Strömberg weak L^1 bound on the Hardy–Littlewood maximal function (Theorem 3 in [SS83]).

Proof of Lemma 4.2. Define the function F by

$$F(\lambda) \overset{\text{def}}{=} \left(\int_0^1 t^{\alpha-1} \frac{1}{(4\pi(t+\varepsilon)^d)} \exp\left(\frac{-1}{2(t+\varepsilon)(V_d \lambda)^{2/d}} \right) dt \right)^{\frac{p}{2}} ,$$

and observe

$$\|\sigma_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}^p \overset{(3.3)}{\leqslant} C^p \int_{\mathbb{R}^d} F(M\mu(x)) \, dx = C^p \int_0^\infty \partial_\lambda F(\lambda) \big| \{ M\mu(x) > \lambda \} \big| \, d\lambda$$

$$\leqslant C^p d \int_0^\infty \partial_\lambda F \, \frac{d\lambda}{\lambda} . \tag{4.7}$$

Making the same change of variables $r = 1/(V_d \lambda)^{1/d}$, and using (3.20) shows

$$\int_0^\infty \partial_\lambda F(\lambda) \, \frac{d\lambda}{\lambda} = -V_d \int_0^\infty \partial_r F(r) \, r^d \, dr = dV_d \int_0^\infty F(r) \, r^{d-1} \, dr . \tag{4.8}$$

To obtain the last equality above we integrated by parts and used the fact that $r^d F(r)$ vanishes at both 0 and ∞. Computing $\|\Phi_\varepsilon\|$ using radial coordinates and using the identity $dV_d = A_{d-1}$ we see

$$dV_d \int_0^\infty F(r) \, r^{d-1} \, dr = \|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}^p$$

Combining this with (4.7) and (4.8) yields (4.3) as desired. □

5. Exact second moments in $H^{-\alpha}$ (Proposition 2.2).

The proof of Proposition 2.2 is a short and elementary calculation.

Proof of Proposition 2.2. For any $\varepsilon > 0$ and $N \in \mathbb{N}$ we have

$$\|\mu_N^\varepsilon - \mu^\varepsilon\|_{H^{-\alpha}}^2 = \int_0^1 \int_{\mathbb{R}^d} t^{\alpha-1} \Big(\frac{1}{N} \sum_{i=1}^N \Phi_{t+\varepsilon}(x - X_i) - \mu_{t+\varepsilon}(x) \Big)^2 \, dx \, dt .$$

Notice that for every x and t the summands $\Phi_{t+\varepsilon}(x - X_i) - \mu_{t+\varepsilon}(x)$ are centered i.i.d. random variables. Thus

$$\boldsymbol{E} \|\mu_N^\varepsilon - \mu^\varepsilon\|_{H^{-\alpha}}^2 = \frac{1}{N} \int_0^1 \int_{\mathbb{R}^d} t^{\alpha-1} \big(\boldsymbol{E} \Phi_{t+\varepsilon}(x - X_1)^2 - \mu_{t+\varepsilon}(x)^2 \big) \, dx \, dt . \tag{5.1}$$

Notice

$$\int_0^1 \int_{\mathbb{R}^d} t^{\alpha-1} \boldsymbol{E} \Phi_{t+\varepsilon}(x - X_1)^2 \, dx \, dt = \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} t^{\alpha-1} \Phi_{t+\varepsilon}(x-y)^2 \, d\mu(y) \, dx \, dt$$

$$= \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} t^{\alpha-1} \Phi_{t+\varepsilon}(z)^2 \, dz \, d\mu(y) \, dt = \int_{\mathbb{R}^d} \|\Phi_\varepsilon\|_{H^{-\alpha}}^2 \, d\mu(y) = \|\Phi_\varepsilon\|_{H^{-\alpha}}^2 .$$

Also, by definition,

$$\int_0^1 \int_{\mathbb{R}^d} t^{\alpha-1} \mu_{t+\varepsilon}(x)^2 \, dx \, dt = \|\mu^\varepsilon\|_{H^{-\alpha}}^2 .$$

Thus (5.1) reduces to (2.6) as desired. When $\alpha > d/2$, all terms in (5.1) remain finite when $\varepsilon = 0$, and in this case (5.1) also holds when $\varepsilon = 0$.

The proof of (2.7) is almost identical to that of (3.2) and (4.2). In the proofs of both (3.2) and (4.2) we note that the dimensional factor $d^{1/p}$ only arises through the bound on the mean (inequalities (3.1) and (4.1) respectively). In our case the mean is the exactly given by (2.6), which doesn't have the $d^{1/p}$ factor. As a result this factor is also absent from the Gaussian tail bound (2.7). □

Appendix A. The $\mathcal{W}^{-\alpha,p}$ and $\mathscr{W}^{-\alpha,p}$ norms of Gaussians.

Since the bounds in Propositions 3.1 and 4.1 involve $\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}$ and $\|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}$, we now compute them explicitly. A scaling argument quickly shows their behavior in ε as $\varepsilon \to 0$. However, their explicit dimensional dependence is useful and isn't easily available in the existing literature. We present it here for reference.

Proposition A.1. *For every $\varepsilon > 0$ we have*

$$\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}} = \frac{\varepsilon^{\frac{1}{2}(\alpha-\frac{d}{q})}}{p^{\frac{d}{2p}}(4\pi)^{\frac{d}{2q}}}\mathcal{B}_0(\varepsilon,\alpha,p,d)^{1/p}\,, \tag{A.1}$$

$$\|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}} = \frac{\varepsilon^{\frac{1}{2}(\alpha-\frac{d}{q})}}{(4\pi)^{\frac{d}{2}}}\mathscr{B}_0(\varepsilon,\alpha,p,d)^{1/p}\,, \tag{A.2}$$

where $\mathcal{B}_0$ and $\mathscr{B}_0$ are given by

$$\mathcal{B}_0(\varepsilon) = \mathcal{B}_0(\varepsilon,\alpha,p,d) \stackrel{\text{def}}{=} \int_0^{1/\varepsilon} \frac{s^{\frac{\alpha p}{2}-1}}{(1+s)^{(p-1)\frac{d}{2}}}\,ds\,,$$

$$\mathscr{B}_0(\varepsilon) = \mathscr{B}_0(\varepsilon,\alpha,p,d) = \int_{\mathbb{R}^d}\left(\int_0^{\frac{1}{\varepsilon}} \frac{s^{\alpha-1}}{(1+s)^d}\exp\left(-\frac{|y|^2}{2(1+s)}\right)ds\right)^{\frac{p}{2}} dy\,.$$

(1) When $\alpha > d/q$

$$\lim_{\varepsilon\to 0}\varepsilon^{\frac{1}{2}(\alpha-\frac{d}{q})}\mathcal{B}_0(\varepsilon)^{1/p} = \left(\frac{2}{\alpha p - d(p-1)}\right)^{\frac{1}{p}}\,, \tag{A.3}$$

$$\lim_{\varepsilon\to 0}\varepsilon^{\frac{1}{2}(\alpha-\frac{d}{q})}\mathscr{B}_0(\varepsilon)^{1/p} \in (0,\infty)\,, \tag{A.4}$$

and $\|\delta_0\|_{\mathcal{W}^{-\alpha,p}}$, $\|\delta_0\|_{\mathscr{W}^{-\alpha,p}}$ are the limits as $\varepsilon\to 0$ of the expressions in (A.1) and (A.2) respectively.

(2) When $\alpha = d/q$

$$\lim_{\varepsilon\to 0}\left(\frac{\mathcal{B}_0(\varepsilon)}{|\ln\varepsilon|}\right)^{1/p} = 1\,, \quad \text{and} \quad 0 < \lim_{\varepsilon\to 0}\left(\frac{\mathscr{B}_0(\varepsilon)}{|\ln\varepsilon|}\right)^{1/p} \leqslant A_{d-1}^{\frac{1}{p}}\Gamma\left(\frac{d}{p}\right)^{\frac{1}{p}}.$$

(3) When $0 < \alpha < d/q$, both $\mathcal{B}_0(\varepsilon)$ and $\mathscr{B}_0(\varepsilon)$ converge to finite, non-zero constants as $\varepsilon\to 0$. Hence both norms $\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}$ and $\|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}$ diverge like $1/\varepsilon^{\frac{1}{2}(\frac{d}{q}-\alpha)}$ as $\varepsilon\to 0$.

For clarity of presentation we divide the proof into two parts, each addressing bounds in $\mathcal{W}^{-\alpha,p}$ and $\mathscr{W}^{-\alpha,p}$ respectively.

A.1. Computation of $\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}$.

Proof of Proposition A.1 for $\mathcal{W}^{-\alpha,p}$. Using the semigroup property of the heat kernel we compute

$$\begin{aligned}
\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}^p &= \int_0^1\int_{\mathbb{R}^d} t^{\frac{\alpha p}{2}-1}\Phi_{t+\varepsilon}(x)^p\,dx\,dt \\
&= \int_0^1\int_{\mathbb{R}^d}\frac{t^{\frac{\alpha p}{2}-1}}{(4\pi(t+\varepsilon))^{\frac{dp}{2}}}\exp\left(-\frac{p|x|^2}{4(t+\varepsilon)}\right)dx\,dt \\
&= \int_0^1 \frac{t^{\frac{\alpha p}{2}-1}}{(4\pi(t+\varepsilon))^{\frac{dp}{2}}}\left(\frac{4\pi(t+\varepsilon)}{p}\right)^{\frac{d}{2}} \\
&= \frac{1}{p^{d/2}(4\pi)^{\frac{d}{2}(p-1)}}\int_0^1 \frac{t^{\frac{\alpha p}{2}-1}}{(t+\varepsilon)^{\frac{d(p-1)}{2}}}\,dt = \frac{\varepsilon^{\frac{\alpha p}{2}-\frac{d(p-1)}{2}}}{p^{d/2}(4\pi)^{\frac{d}{2}(p-1)}}\mathcal{B}_0(\varepsilon,\alpha,p,d)\,,
\end{aligned}$$

where the last equality follows from a change of variables. Taking the $p-$th root yields (A.1).

We now prove items (1)–(3) for the quantity $\mathcal{B}_0(\varepsilon)$. First, when $\alpha > d/q$, the dominated convergence theorem yields

$$\lim_{\varepsilon\to 0^+}\left(\varepsilon^{\frac{\alpha p}{2}-\frac{d(p-1)}{2}}\mathcal{B}_0(\varepsilon,\alpha,p,d)\right)^{1/p} = \left(\frac{2}{\alpha p - d(p-1)}\right)^{\frac{1}{p}}.$$

This proves (A.3) as desired. Moreover, in this regime, the Sobolev embedding theorem implies $\Phi_\varepsilon \to \delta_0$ in $W^{-\alpha,p}$ as $\varepsilon\to 0$ and hence

$$\|\delta_0\|_{\mathcal{W}^{-\alpha,p}} = \lim_{\varepsilon\to 0^+}\|\Phi_\varepsilon\|_{\mathcal{W}^{-\alpha,p}}\,,$$

as claimed.

Turning to item (2) for the quantity $\mathcal{B}_0(\varepsilon)$, we compute directly by the l'Hôpital rule that

$$\begin{aligned}
\lim_{\varepsilon\to 0^+}\frac{\mathcal{B}_0(\varepsilon)}{|\ln\varepsilon|} &= -\lim_{\varepsilon\to 0^+}\frac{1}{\ln\varepsilon}\int_0^{1/\varepsilon}\frac{s^{\frac{\alpha p}{2}-1}}{(1+s)^{(p-1)\frac{d}{2}}}\,ds \\
&= -\lim_{\varepsilon\to 0^+}\varepsilon\frac{(1/\varepsilon)^{\frac{\alpha p}{2}-1}}{(1+1/\varepsilon)^{(p-1)\frac{d}{2}}}(-\frac{1}{\varepsilon^2}) = \lim_{\varepsilon\to 0^+}\varepsilon^{\frac{(p-1)d-\alpha p}{2}} = 1\,.
\end{aligned}$$

Finally, to prove item (3), we now assume $0 < \alpha < d/q$. In this case

$$\begin{aligned}
\mathcal{B}_0(\varepsilon) &\leqslant \int_0^1 s^{\frac{\alpha p}{2}-1}\,ds + \int_1^{\frac{1}{\varepsilon}} s^{\frac{\alpha p}{2}-(p-1)\frac{d}{2}-1}\,ds \\
&\leqslant \frac{2}{\alpha p} + \frac{2}{|\alpha p-(p-1)d|}\left(1-\varepsilon^{(p-1)\frac{d}{2}-\frac{\alpha p}{2}}\right),
\end{aligned}$$

and so monotone convergence theorem completes item (3) for the quantity $\mathcal{B}_0(\varepsilon)$, yielding that

$$\lim_{\varepsilon\to 0^+}\mathcal{B}_0(\varepsilon) = \int_0^\infty \frac{s^{\frac{\alpha p}{2}-1}}{(1+s)^{\frac{(p-1)d}{2}}} \in (0,\infty)\,.$$

□

A.2. Computation of $\|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}$.

Before we can estimate $\|\Phi\|_{\mathscr{W}^{-\alpha,p}}$, we need a calculus lemma concerning asymptotics of the inner integral in (2.2b).

Lemma A.2. *For any $r \geqslant 0$ let*

$$I_\varepsilon(r) := \int_0^{\frac{1}{\varepsilon}}\frac{s^{\alpha-1}}{(1+s)^d}\exp\left(-\frac{r^2}{2(1+s)}\right)ds\,.$$

For all $r\geqslant 0$, then $I_\varepsilon(r)$ satisfies the bounds

$$I_\varepsilon(r) \leqslant \begin{cases} \dfrac{1}{\alpha} + \dfrac{1}{\alpha-d}\left(\dfrac{1}{\varepsilon^{\alpha-d}}-1\right) & \alpha \neq d \\ \dfrac{1}{\alpha} + \log\dfrac{1}{\varepsilon} & \alpha = d\,. \end{cases} \tag{A.5}$$

Additionally, for every $r > 0$, we have

$$\frac{C_\alpha}{r^{2(d-\alpha)}} \int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1} e^{-t}\, dt \leqslant I_\varepsilon(r) \leqslant \frac{1}{r^{2(d-\alpha)}} \int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1} e^{-t}\, dt\,. \tag{A.6}$$

where $C_\alpha \stackrel{\text{def}}{=} \min(1, \frac{1}{2^{\alpha-1}})$.

Proof. We first note the elementary bound

$$I_\varepsilon(r) \leqslant \int_0^{\frac{1}{\varepsilon}} \frac{s^{\alpha-1}}{(1+s)^d}\, ds \leqslant \int_0^1 s^{\alpha-1}\, ds + \int_1^{\frac{1}{\varepsilon}} s^{\alpha-d-1}\, ds = \frac{1}{\alpha} + \frac{1}{\alpha-d}\Big(\frac{1}{\varepsilon^{\alpha-d}} - 1\Big)\,,$$

if $\alpha \neq d$, and

$$I_\varepsilon(r) \leqslant \int_0^{\frac{1}{\varepsilon}} \frac{s^{\alpha-1}}{(1+s)^d}\, ds \leqslant \int_0^1 s^{\alpha-1}\, ds + \int_1^{\frac{1}{\varepsilon}} s^{-1}\, ds = \frac{1}{\alpha} + \log\frac{1}{\varepsilon}\,.$$

when $\alpha = d$. This is the proof of (A.5).

We also need a different bound that captures decay in r for r large. For this, the change of variables $\frac{r^2}{2(1+s)} = t$, gives

$$I_\varepsilon(r) = \int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} \left(\frac{r^2}{t} - 1\right)^{\alpha-1} \frac{t^{d-1}}{r^{2(d-1)}} e^{-t} \frac{dt}{t}\,. \tag{A.7}$$

Since $(r^2 - t)^{\alpha-1} \leqslant r^{2(\alpha-1)}$ on the interval of integration, we obtain

$$I_\varepsilon(r) \leqslant \frac{1}{r^{2(d-\alpha)}} \int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1} e^{-t}\, dt\,.$$

This yields the upper bound in (A.6).

For the lower bound, using (A.7) one more time and that on the interval of integration, we have $r^2 \geqslant r^2\big(1 - \frac{\varepsilon}{2(1+\varepsilon)}\big) \geqslant r^2 - t \geqslant \frac{r^2}{2}$, we compute

$$I_\varepsilon(r) \geqslant \frac{C_\alpha}{r^{2(d-\alpha)}} \int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1} e^{-t}\, dt\,,$$

with $C_\alpha = 1$ if $\alpha \leqslant 1$ and $C_\alpha = \frac{1}{2^{\alpha-1}}$ if $\alpha \geqslant 1$. This yields the lower bound in (A.6) □

With these observations about I_ε, we are ready to prove Lemma A.1 for $\mathscr{W}^{-\alpha,p}$.

Proof of Proposition A.1 for $\mathscr{W}^{-\alpha,p}$. We compute

$$\begin{aligned}\|\Phi_\varepsilon\|^p_{\mathscr{W}^{-\alpha,p}} &= \int_{\mathbb{R}^d} \left(\int_0^1 t^{\alpha-1} \Phi_{t+\varepsilon}(x)^2\, dt\right)^{\frac{p}{2}} dx \\ &= \int_{\mathbb{R}^d} \left(\int_0^1 \frac{t^{\alpha-1}}{(4\pi(t+\varepsilon))^d} \exp\Big(-\frac{|x|^2}{2(t+\varepsilon)}\Big)\, dt\right)^{\frac{p}{2}} dx\end{aligned}$$

Here, make the change of variables $t = \varepsilon s, y = \frac{x}{\sqrt{\varepsilon}}$, to obtain

$$\|\Phi_\varepsilon\|^p_{\mathscr{W}^{-\alpha,p}} = \varepsilon^{\frac{\alpha p}{2} - (p-1)\frac{d}{2}} \int_{\mathbb{R}^d} \left(\int_0^{\frac{1}{\varepsilon}} \frac{s^{\alpha-1}}{(4\pi(1+s))^d} \exp\Big(-\frac{|y|^2}{2(1+s)}\Big)\, ds\right)^{\frac{p}{2}} dy$$

$$\stackrel{\text{def}}{=} \varepsilon^{\frac{\alpha p}{2} - (p-1)\frac{d}{2}} \mathscr{B}_0(\varepsilon, \alpha, d, p)\,.$$

Upon taking the p-th root of both sides, we get (A.2). We turn to analyzing the $\varepsilon \to 0^+$ behavior of $\mathscr{B}_0$, thereby verifying items (1)-(3) in the different regimes.

We must divide our analysis in a few cases.

Case I: $\alpha > d > (p-1)d/p$. Lemma A.2 show that for small ε we have

$$\begin{aligned}\mathscr{B}_0(\varepsilon, \alpha, d, p) &= A_{d-1} \int_0^\infty I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\, dr \\ &= A_{d-1}\left(\int_0^{\sqrt{\varepsilon}} + \int_{\sqrt{\varepsilon}}^\infty\right) I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\, dr = A_{d-1}\big(J_1 + J_2\big)\,.\end{aligned}$$

Using (A.5) we see

$$J_1 \leqslant \left(\int_0^{\sqrt{\varepsilon}} \frac{2}{(\alpha-d)^{\frac{p}{2}}} \varepsilon^{(d-\alpha)p/2} r^{d-1}\, dr\right) = \frac{2}{(\alpha-d)^{p/2}} \varepsilon^{\frac{1}{2}((p-1)d - \alpha p)}\,.$$

For J_2 we use (A.6) to obtain

$$J_2 = \int_{\sqrt{\varepsilon}}^\infty I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\, dr \leqslant \int_{\sqrt{\varepsilon}}^\infty r^{\alpha p - (p-1)d - 1} \left(\int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1} e^{-t}\, dt\right)^{p/2} dr\,.$$

Making the change of variables $s = \sqrt{\varepsilon} r$ in the last integral, we obtain

$$\begin{aligned}J_2 &\leqslant \varepsilon^{\frac{1}{2}((p-1)d-\alpha p)} \int_1^\infty s^{\alpha p - (p-1)d - 1} \left(\int_{\frac{s^2}{4}}^\infty t^{d-\alpha-1} e^{-t}\, dt\right)^{p/2} ds \\ &\leqslant \varepsilon^{\frac{1}{2}((p-1)d-\alpha p)} \int_1^\infty s^{\alpha p - (p-1)d - 1} e^{-\frac{ps^2}{8}} \frac{1}{(\alpha-d)^{\frac{p}{2}}} s^{p(d-\alpha)}\, ds \\ &= \frac{\varepsilon^{\frac{1}{2}((p-1)d-\alpha p)}}{(\alpha-d)^{p/2}} \int_1^\infty s^{d-1} e^{-\frac{ps^2}{8}}\, ds\,.\end{aligned}$$

Combining upper bounds for J_1 and J_2 we find

$$\mathscr{B}_0(\varepsilon, \alpha, d, p) \leqslant \frac{A_{d-1}}{(\alpha-d)^{p/2}} \left(2 + \int_1^\infty s^{d-1} e^{-\frac{ps^2}{8}}\, ds\right) \varepsilon^{\frac{d(p-1)-\alpha p}{2}}\,, \tag{A.8}$$

so that in this regimes, the gaussians $\{\Phi_\varepsilon\}_{\varepsilon>0}$ have uniformly bounded $\mathscr{W}^{-\alpha,p}$ norms.

In order to complete the proof of (A.4), we must show a matching lower bound on $\mathscr{B}_0(\varepsilon)$. For this, using the lower bound portion of (A.6) we compute, for ε small enough (so that $e^{-\frac{p\varepsilon^2}{4}} \geqslant \frac{1}{2}$),

$$\begin{aligned}\mathscr{B}_0(\varepsilon) &\geqslant \int_{|y| \leqslant \sqrt{\varepsilon}} I_\varepsilon(|y|)^{\frac{p}{2}}\, dy \geqslant C_\alpha A_{d-1} \int_0^{\sqrt{\varepsilon}} \frac{r^{d-1}}{r^{(d-\alpha)p}} \left(\int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1} e^{-t}\, dt\right)^{\frac{p}{2}} dr \\ &\geqslant \frac{A_{d-1}}{(\alpha-d)^p} \int_0^{\sqrt{\varepsilon}} \frac{r^{d-1}}{r^{(d-\alpha)p}} e^{-\frac{pr^2}{4}} \left(\Big(\frac{\varepsilon r^2}{2(1+\varepsilon)}\Big)^{d-\alpha} - \Big(\frac{r^2}{2}\Big)^{d-\alpha}\right)^{\frac{p}{2}} dr \\ &\geqslant \frac{A_{d-1}}{2^{1+\frac{(d-\alpha)p}{2}}(\alpha-d)^p} \int_0^{\sqrt{\varepsilon}} r^{d-1} \frac{\varepsilon^{\frac{(d-\alpha)p}{2}}}{2}\, dr = \frac{A_{d-1}}{2^{2+\frac{(d-\alpha)p}{2}}(\alpha-d)^p d} \varepsilon^{\frac{d(p-1)-\alpha p}{2}}\,,\end{aligned}$$

which matches the scaling of the upper bound from (A.8). Combining the previous two displays completes the proof of (A.4).

Case II: $\frac{p-1}{p}d \leqslant \alpha \leqslant d$. In this case, we compute

$$\begin{aligned}\mathscr{B}_0(\varepsilon,\alpha,d,p) &= A_{d-1}\int_0^\infty I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr \\ &= A_{d-1}\Big(\int_0^{\sqrt{\varepsilon}} + \int_{\sqrt{\varepsilon}}^{\frac{1}{\sqrt{\varepsilon}}} + \int_{\frac{1}{\sqrt{\varepsilon}}}^\infty\Big) I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr\,.\end{aligned} \tag{A.9}$$

Let us analyze each of the pieces in turn, starting first with the middle piece (which will turn out to be most important). To analyze this middle term we must observe that when $d > \alpha$, for $r > 0$, then

$$I_\varepsilon(r) \leqslant \frac{1}{r^{2(d-\alpha)}}\int_0^\infty t^{d-\alpha-1}e^{-t}\,dt = \frac{\Gamma(d-\alpha)}{r^{2(d-\alpha)}}\,. \tag{A.10}$$

Inserting (A.10) we obtain

$$\begin{aligned}\int_{\sqrt{\varepsilon}}^{\frac{1}{\sqrt{\varepsilon}}} I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr &\leqslant \Gamma(d-\alpha)^{\frac{p}{2}}\int_{\sqrt{\varepsilon}}^{\frac{1}{\sqrt{\varepsilon}}} \frac{1}{r^{p(d-\alpha)}} r^{d-1}\,dr \\ &\leqslant \Gamma(d-\alpha)^{\frac{p}{2}} \times \begin{cases} \ln\frac{1}{\varepsilon} & \text{if } \alpha p = d(p-1) \\ \frac{1}{|\alpha p-(p-1)d|}\varepsilon^{\frac{d(p-1)-\alpha p}{2}} & \text{if } d(p-1) < \alpha p \leqslant dp\,.\end{cases}\end{aligned} \tag{A.11}$$

Now we turning to the piece on $[0,\sqrt{\varepsilon}]$ we compute directly that when $\alpha \leqslant d$, by inserting the uniform bound (A.5) that

$$\int_0^{\sqrt{\varepsilon}} I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr \leqslant C(\alpha,d,p)\varepsilon^{\frac{d}{2}}\big(1+(\log^{\frac{p}{2}}\frac{1}{\varepsilon})1_{\alpha=d}\big)\,, \tag{A.12}$$

where the notation $1_{\alpha=d}$ serves to indicate that this logarithmic correction is only present when $\alpha = d$.

Finally, the contribution of the piece $\big[\frac{1}{\sqrt{\varepsilon}},\infty\big)$, when $\alpha \leqslant d$, is

$$\begin{aligned}\int_{\frac{1}{\sqrt{\varepsilon}}}^\infty I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr &\leqslant \int_{\frac{1}{\sqrt{\varepsilon}}}^\infty r^{d-1-(d-\alpha)p}\Big(\int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1}e^{-t}\,dt\Big)^{\frac{p}{2}}\,dr \\ &\leqslant \frac{1}{(\alpha-d)^{\frac{p}{2}}}\int_{\frac{1}{\sqrt{\varepsilon}}}^\infty r^{d-1-(d-\alpha)p}\exp\Big(-\frac{pr^2}{4}\Big) r^{(d-\alpha)p}\,dr = O\Big(e^{-\frac{p}{\varepsilon}}\Big)\,,\end{aligned} \tag{A.13}$$

which is, evidently, an exponentially small contribution. Inserting the computations from (A.11),(A.12) and (A.13) into (A.9) we have shown that for ε small enough,

$$\|\Phi_\varepsilon\|_{\mathscr{W}^{-\alpha,p}}^p \leqslant 2A_{d-1}\Gamma(d-\alpha)^{\frac{p}{2}} \begin{cases} \ln\Big(\frac{1}{\varepsilon}\Big) & \text{if } \alpha p = d(p-1) \\ \frac{1}{\alpha p - d(p-1)} & \text{if } \alpha p > d(p-1)\,.\end{cases}$$

As in Case 1, we must show a matching lower bound which demonstrates that these rates are asymptotically (for small ε) optimal. To do this, we use (A.6) to compute in spherical coordinates, when $\alpha p > d(p-1)$, that when ε is small enough,

$$\begin{aligned}\mathscr{B}_0(\varepsilon) &\geqslant A_{d-1}\int_{\sqrt{\varepsilon}\leqslant r\leqslant\frac{1}{\sqrt{\varepsilon}}} I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr \\ &\geqslant A_{d-1}\int_{\sqrt{\varepsilon}}^{\frac{1}{\sqrt{\varepsilon}}} \frac{r^{d-1}}{r^{p(d-\alpha)}}\Big(\int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1}e^{-t}\,dt\Big)^{\frac{p}{2}}\,dr \\ &= A_{d-1}\varepsilon^{\frac{1}{2}(\alpha p-(p-1)d)}\int_1^{\frac{1}{\varepsilon}} s^{\alpha p-(p-1)d-1}\Big(\int_{\frac{\varepsilon^2 s^2}{2(1+\varepsilon)}}^{\frac{s^2}{2}} t^{d-\alpha-1}e^{-t}\,dt\Big)^{\frac{p}{2}}\,ds \\ &\geqslant A_{d-1}\varepsilon^{\frac{1}{2}(\alpha p-(p-1)d)}\int_1^{\frac{1}{\varepsilon}} s^{\alpha p-(p-1)d-1}e^{-\frac{ps^2}{4}}\frac{1}{2(d-\alpha)^{p/2}}s^{(d-\alpha)p}\,ds \\ &= \frac{A_{d-1}}{2(d-\alpha)^{p/2}}\varepsilon^{\frac{1}{2}(\alpha p-(p-1)d)}\int_1^{\frac{1}{\varepsilon}} s^{d-1}e^{-\frac{ps^2}{4}}\,ds \\ &\geqslant \frac{A_{d-1}}{4(d-\alpha)^{p/2}}\varepsilon^{\frac{1}{2}(\alpha p-(p-1)d)}\int_1^\infty s^{d-1}e^{-\frac{ps^2}{4}}\,ds\,.\end{aligned}$$

In the previous calculation we made the change of variables $r = \sqrt{\varepsilon}s$.

In the critical case, that is, when $\alpha p = d(p-1)$, we estimate slightly differently, arguing that, if $d \geqslant p$, then

$$\begin{aligned}\mathscr{B}_0(\varepsilon) &\geqslant A_{d-1}\int_{\sqrt{\varepsilon}\leqslant r\leqslant\frac{1}{\sqrt{\varepsilon}}} I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr \\ &\geqslant A_{d-1}\int_{\sqrt{\varepsilon}}^{\frac{1}{\sqrt{\varepsilon}}} \frac{r^{d-1}}{r^{p(d-\alpha)}}\Big(\int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1}e^{-t}\,dt\Big)^{\frac{p}{2}}\,dr \\ &= A_{d-1}\int_1^{\frac{1}{\varepsilon}} s^{-1}\Big(\int_{\frac{\varepsilon^2 s^2}{2(1+\varepsilon)}}^{\frac{s^2}{2}} t^{\frac{d}{p}-1}e^{-t}\,dt\Big)^{\frac{p}{2}}\,ds \\ &\geqslant A_{d-1}\int_1^{\frac{1}{\varepsilon}} s^{-1}\Big(\int_1^{\frac{s^2}{2}} e^{-t}\,dt\Big)^{\frac{p}{2}}\,ds \\ &\geqslant A_{d-1}\int_1^{\frac{1}{\varepsilon}} s^{-1}\big(1-e^{-\frac{s^2}{2}}\big)^{\frac{p}{2}}\,ds \\ &\geqslant \big(1-e^{-\frac{1}{2}}\big)^{\frac{p}{2}}A_{d-1}\log\frac{1}{\varepsilon}\,,\end{aligned}$$

whereas, when $d < p$,

$$\begin{aligned}\mathscr{B}_0(\varepsilon) &\geqslant A_{d-1}\int_{\sqrt{\varepsilon}\leqslant r\leqslant\frac{1}{\sqrt{\varepsilon}}} I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr \\ &\geqslant A_{d-1}\int_{\sqrt{\varepsilon}}^{\frac{1}{\sqrt{\varepsilon}}} \frac{r^{d-1}}{r^{p(d-\alpha)}}\Big(\int_{\frac{\varepsilon r^2}{2(1+\varepsilon)}}^{\frac{r^2}{2}} t^{d-\alpha-1}e^{-t}\,dt\Big)^{\frac{p}{2}}\,dr\end{aligned}$$

$$
\begin{aligned}
&= A_{d-1}\int_1^{\frac{1}{\varepsilon}} s^{-1}\left(\int_{\frac{\varepsilon^2 s^2}{2(1+\varepsilon)}}^{\frac{s^2}{2}} t^{\frac{d}{p}-1}e^{-t}\,dt\right)^{\frac{p}{2}} ds\\
&\geqslant A_{d-1}\int_1^{\frac{1}{\varepsilon}} s^{-1}\left(\int_{\frac{\varepsilon^2 s^2}{2(1+\varepsilon)}}^{1} e^{-t}\,dt\right)^{\frac{p}{2}} ds\\
&\geqslant A_{d-1}\int_1^{\frac{1}{\varepsilon}} s^{-1}\left(\int_{\frac{1}{2}}^{1} e^{-t}\,dt\right)^{\frac{p}{2}} ds\\
&\geqslant \left(1-e^{-\frac{1}{2}}\right)^{\frac{p}{2}} A_{d-1}\int_1^{\frac{1}{\varepsilon}} s^{-1}\,ds\\
&\geqslant \left(1-e^{-\frac{1}{2}}\right)^{\frac{p}{2}} A_{d-1}\log\frac{1}{\varepsilon}\,.
\end{aligned}
$$

Cases 1 and 2 together complete the proof of items (1) and (2) in Proposition A.1 for the quantity $\mathscr{B}_0$.

Case III: $\frac{p-1}{p}d > \alpha$. Finally, in this case, we compute

$$
\begin{aligned}
\mathscr{B}_0(\varepsilon,\alpha,d,p) &= A_{d-1}\int_0^{\infty} I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr\\
&= A_{d-1}\left(\int_0^1 + \int_1^{\frac{1}{\sqrt{\varepsilon}}} + \int_{\frac{1}{\sqrt{\varepsilon}}}^{\infty}\right) I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr\,.
\end{aligned}
$$

As in Case 1, we estimate these term-by-term. For the first piece we compute

$$
\int_0^1 I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr \leqslant \frac{1}{d}\left(\frac{1}{\alpha}+\frac{1}{d-\alpha}\right)^{\frac{p}{2}},
$$

whereas for the middle piece we compute by plugging in (A.10) we obtain

$$
\int_1^{\frac{1}{\sqrt{\varepsilon}}} I_\varepsilon(r)^{\frac{p}{2}} r^{d-1}\,dr \quad \leqslant \Gamma(d-\alpha)^{\frac{p}{2}}\int_1^{\frac{1}{\sqrt{\varepsilon}}} \frac{1}{r^{p(d-\alpha)}} r^{d-1}\,dr \leqslant \frac{\Gamma(d-\alpha)^{\frac{p}{2}}}{d(p-1)-\alpha p}\,.
$$

The estimate for the last piece is exactly the same as (A.13). To conclude, by the monotone convergence theorem we obtain that in this sub-critical case,

$$
\lim_{\varepsilon\to 0^+}\mathscr{B}_0(\varepsilon,\alpha,d,p) = \int_{\mathbb{R}^d}\left(\int_0^\infty \frac{s^{\alpha-1}}{(1+s)^d}\exp\left(-\frac{|y|^2}{2(1+s)}\right)ds\right)^{\frac{p}{2}} dy \in (0,\infty)\,. \qquad \square
$$

References

[AKM19] S. Armstrong, T. Kuusi, and J.-C. Mourrat. *Quantitative stochastic homogenization and large-scale regularity*, volume 352 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2019. doi:10.1007/978-3-030-15545-2.

[AKT84] M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984. doi:10.1007/BF02579135.

[BLG14] E. Boissard and T. Le Gouic. On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(2):539–563, 2014. doi:10.1214/12-AIHP517.

[Div21] V. Divol. A short proof on the rate of convergence of the empirical measure for the Wasserstein distance, 2021, 2101.08126. URL https://arxiv.org/abs/2101.08126.

[DSS13] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(4):1183–1203, 2013. doi:10.1214/12-AIHP489.

[FG15] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015. doi:10.1007/s00440-014-0583-7.

[GT21] M. Goldman and D. Trevisan. Convergence of asymptotic costs for random Euclidean matching problems. *Probab. Math. Phys.*, 2(2):341–362, 2021. doi:10.2140/pmp.2021.2.341.

[GTS15] N. García Trillos and D. Slepčev. On the rate of convergence of empirical measures in ∞-transportation distance. *Canad. J. Math.*, 67(6):1358–1383, 2015. doi:10.4153/CJM-2014-044-6.

[HCRS25] E. Hess-Childs, M. Rosenzweig, and S. Serfaty. A sharp commutator estimate for all Riesz modulated energies, 2025, 2511.13461. URL https://arxiv.org/abs/2511.13461.

[Led23] M. Ledoux. Optimal matching of random samples and rates of convergence of empirical measures. In *Mathematics going forward—collected mathematical brushstrokes*, volume 2313 of *Lecture Notes in Math.*, pages 615–627. Springer, Cham, [2023] ©2023. doi:10.1007/978-3-031-12244-6_43.

[Liu08] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York, 2008.

[LS89] T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica*, 9(2):161–187, 1989. doi:10.1007/BF02124678.

[LZ21] M. Ledoux and J.-X. Zhu. On optimal matching of Gaussian samples III. *Probab. Math. Statist.*, 41(2):237–265, 2021. doi:10.37190/0208-4147.41.2.3.

[SGBSM23] C.-J. Simon-Gabriel, A. Barp, B. Schölkopf, and L. Mackey. Metrizing weak convergence with maximum mean discrepancies. *Journal of Machine Learning Research*, 24(184):1–20, 2023. URL http://jmlr.org/papers/v24/21-0599.html.

[SS83] E. M. Stein and J.-O. Strömberg. Behavior of maximal functions in $\mathbf{R}^n$ for large n. *Ark. Mat.*, 21(2):259–269, 1983. doi:10.1007/BF02384314.

[Tal21] M. Talagrand. *Upper and lower bounds for stochastic processes—decomposition theorems*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Cham, second edition, [2021] ©2021. doi:10.1007/978-3-030-82595-9.

[Thi12] J.-L. Thiffeault. Using multiscale norms to quantify mixing and transport. *Nonlinearity*, 25(2):R1–R44, 2012. doi:10.1088/0951-7715/25/2/R1.

[Tri92] H. Triebel. *Theory of function spaces. II*, volume 84 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1992. doi:10.1007/978-3-0346-0419-2.

[Ver18] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. doi:10.1017/9781108231596. An introduction with applications in data science, With a foreword by Sara van de Geer.

[Vil09] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. doi:10.1007/978-3-540-71050-9. Old and new.

[WB19] J. Weed and F. Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019. doi:10.3150/18-BEJ1065.

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213.
Email address: gautam@math.cmu.edu

Department of Mathematics, University of Utah, Salt Lake City, UT 84112.
Email address: raghav@math.utah.edu