Fourier-Based Bounds for Wasserstein Distances and Their Implications in Computational Inversion

Wanli Hong[*] Vladimir A. Kobzar[†] Kui Ren[‡]

December 9, 2023

Abstract

Computational inverse problems entail fitting a mathematical model to data. These problems are often solved numerically, by minimizing the mismatch between the model and the data using an appropriate metric. We focus on the case when this metric is the Wasserstein-p (W_p) distance between probability measures as well as its generalizations by Piccoli et al., for unbalanced measures, including the Kantorovich-Rubinstein norm. The recent work of Niles-Weed and Berthet established that W_p is bounded from below and above by weighted ℓ_p norms of the wavelet coefficients of the mismatch, among other things, relying on the fluid dynamics formulation of W_p. Building on this research, we establish lower and upper bounds on W_p on the hypercube and flat torus in terms of weighted ℓ_q norms of the Fourier coefficients of the mismatch. In this setting, for measures uniformly bounded above, the lower bound increases as p increases. Based on that fact, in our setting, the lower bound resolves the open problem posed by Steinerberger to prove the existence of a Fourier-based lower bound on W_p that grows with p. When W_p is used as the mismatch metric in computational inversion, these bounds allow us to analyze the effects of stopping early the computational minimization of the mismatch on the resolution of frequencies, and the dependence of the resolution of frequencies on p. Since the W_p distance is used in a broad range of other problems in mathematics and computational sciences, we expect that our bounds will also be of interest beyond inverse problems.

1 Introduction

Optimal transport (OT) is a fundamental problem in mathematics with growing and promising applications to computational inverse problems. While extensive connections have been established between OT and many areas of analysis, connections between OT and Fourier

[*]Center for Data Science, New York University; wh992@nyu.edu
[†]Department of Applied Physics and Applied Mathematics, Columbia University; vak2116@columbia.edu
[‡]Department of Applied Physics and Applied Mathematics, Columbia University; kr2002@columbia.edu

analysis specifically are still relatively unexplored. We consider the Wasserstein-p distance, or $W_p(\mu,\nu)$, which represents the minimum transportation cost between a pair of probability measures μ and ν using the p-th moment of a distance function (focusing specifically on the Euclidean distance) as the underlying transportation cost between points in $\mathbb{R}^d$. We assume that the measures μ and ν as absolutely continuous with respect to the Lebesgue measure and, therefore, are associated with a pair of probability densities f and g. Accordingly, we will denote $W_p(\mu,\nu)$ as $W_p(f,g)$ by reference to the corresponding densities. We also assume that these densities have a Fourier basis expansion on $[0,1)^d$. Building on the recent work applying wavelet analysis of W_p in nonparametric statistics [48], as well as the recent work applying Fourier analysis to W_p on the circle in the context of measure-theoretic discrepancy theory [57], we establish upper and lower bounds on $W_p(f,g)$ on the hypercube $\mathbb{H}^d$ and flat torus $\mathbb{T}^d$ in terms of the weighted ℓ_q norm $\|\hat{f}-\hat{g}\|_{q,w^r}$ of the Fourier coefficients of the mismatch between f and g. This norm $\|\cdot\|_{q,w^r}$ is given by

$$\|\lambda\|_{q,w^r} = \begin{cases} \left(\sum_k |w_k^r \lambda_k|^q\right)^{\frac{1}{q}} & \text{if } q<\infty \\ \sup_k |w_k^r\lambda_k| & \text{if } q=\infty \end{cases} \tag{1}$$

where for $k \in \mathbb{Z}^d \setminus 0$, the weights w^r are[1]

$$w_k^r = 1/(2\pi\|k\|_r). \tag{2}$$

	$p\in[1,\infty),\ s=1$	$p\in(1,2],\ s\in(1,\infty]$	$p\in(2,\infty),\ s\in(1,\frac{2p-2}{p-2}]$
$\|f\|_{L_s}, \|g\|_{L_s} \le M$	$d^{-\frac{1}{2}}\|\hat{f}-\hat{g}\|_{\infty,w^1}$	$d^{\frac{1}{q}-\frac{1}{2}}M^{-\frac{1}{p'}}\|\hat{f}-\hat{g}\|_{q,w^{q'}}$	

Table 1: The lower bounds on $W_p^{\mathbb{T}^d}(f,g)$ and $W_p^{\mathbb{H}^d}(f,g)$ given by Theorem 4.5 in terms of the weighted ℓ_q norms of the Fourier coefficients of $f-g$ where $q=p's/(s-1)$ and p' and q' are the Hölder conjugates of p and q.

Table 1 summarizes our lower bounds for $W_p^{\mathbb{T}^d}$ and $W_p^{\mathbb{H}^d}$ and Table 2 summarizes our upper bounds for $W_p^{\mathbb{T}^d}$. (On $\mathbb{H}^d$, the corresponding upper bounds contain an additional term, specified in Lemma 4.9, due to the absence of transport across the boundary.) We also establish similar bounds for the metrics developed in [50, 51] that generalize W_p to unbalanced measures; for $p=1$ this metric is the classic Kantorovich-Rubinstein norm.

Reference [57] posed proving the existence of a Fourier-based lower bound on W_p that grows with p, as an open problem. Our lower bound resolves this open problem on $\mathbb{H}^d$ and $\mathbb{T}^d$ for measures that are absolutely continuous with respect to the Lebesque measure and are uniformly bounded above a.e. Moreover, in the context of computational inversion using W_p as the mismatch metric, these bounds allow us to analyze the resolution of frequencies

[1]The amplitude of the zero frequency $\hat{f}_0-\hat{g}_0$ is zero because the densities have the same mass. Therefore, in the case of balanced transport, we do not need to specify w_0^r. However, we will specify it in the unbalanced transport setting.

<table>
<tr><th></th><th>$p = 1,\ \xi = 0$</th><th>$p \in (1, 2],\ \xi > 0$</th><th>$p \in (2, \infty),\ \xi > 0$</th></tr>
<tr><td>$f \wedge g \geq \xi$</td><td>$\|\hat{f} - \hat{g}\|_{2,w^2}$</td><td>$O\left(\xi^{-\frac{1}{p'}} \|\hat{f} - \hat{g}\|_{2,w^2}\right)$</td><td rowspan="2">$O\left(d^{\frac{1}{2}-\frac{1}{p}} p\ \xi^{-\frac{1}{p'}} \|\hat{f} - \hat{g}\|_{p',w^{p'}}\right)$</td></tr>
<tr><td>$f \wedge g \geq \xi$
$\|f - g\|_{\dot{\mathcal{H}}^\beta} \leq z$</td><td>$O\left(\sqrt{z}\|\hat{f} - \hat{g}\|^{\frac{1}{2}}_{\infty,w^2}\right)$</td><td>$O\left(\xi^{-\frac{1}{p'}} \sqrt{z}\|\hat{f} - \hat{g}\|^{\frac{1}{2}}_{p',w^2}\right)$</td></tr>
</table>

Table 2: The upper bounds on $W_p^{\mathbb{T}^d}(f, g)$ given by Theorem 4.7 and Proposition 4.8 in terms of the weighted ℓ_q norms of the Fourier coefficients of $f - g$ where $\beta > d/p - d/2 - 1$, and p' is again the Hölder conjugate of p. (The homogeneous Sobolev norm $\dot{\mathcal{H}}^\beta$ also has a Fourier-based representation given by (9).)

and the effects of stopping the minimization process early. As discussed in Section 6.2, recent work [48] established upper and lower bounds on $W_p(\mu, \nu)$ in terms of weighted ℓ_p norms of the wavelet coefficients of the mismatch of the corresponding densities. To establish our Fourier-based lower and upper bounds, in this work we generalize the proof techniques from reference [48]. Furthermore, we apply these wavelet-based bounds to analyze the resolution in computational inversion in the wavelet domain.

We consider the classic *inverse problem* setting: let the function $f : \mathcal{M} \times \Omega \to \mathbb{R}$ represent a model of a given phenomenon (the *forward model*) where $\mathcal{M}$ is the space of the model parameters, and Ω is the spatial domain, $\mathbb{T}^d$ or $\mathbb{H}^d$ in our case. Accordingly, if we fix a model $m \in \mathcal{M}$, $f(m)$ is a function from Ω to $\mathbb{R}$. The *inverse problem* entails reconstructing m from the observed data $g : \Omega \to \mathbb{R}$, i.e., in the appropriate sense solving

$$f(m) = g \tag{3}$$

for m. Even if f is invertible with respect to m, an analytic expression for its inverse typically does exist. Therefore, this problem is usually solved computationally by minimizing the mismatch between the model and the data using an appropriate mismatch functional Φ:

$$m^* \in \arg\min_{m \in \mathcal{M}} \Phi(f(m), g). \tag{4}$$

If g is not corrupted with noise and is in the range of $f(m)$, then all metrics will lead to the same optimal value in (4). However, the metric will make a difference when the data g is noisy and/or the minimization problem can not be solved accurately [6].

Historically, the L_2 norm has been used to minimize the mismatch in computational inversion. Reference [24] showed that using a Sobolev norm $\mathcal{H}^s$ with $s < 0$ (which is a weaker norm than L_2) as the mismatch functional Φ would lead to a smoother optimal solution in the presence of noise. As discussed more fully in Appendix A, W_2 is asymptotically equivalent to a weighted negative order Sobolev norm $\dot{\mathcal{H}}^{-1}$, which has a Fourier-based representation facilitating the analysis of the frequency content of the inversion using the W_2 distance; this equivalence leads to nonasymptotic Fourier-based upper and lower bounds.

As further discussed in Section 2.1 below, using the W_p distance with other values of p, most notably $p = 1$, and its generalizations to unbalanced measures, empirically revealed a number of attractive features in the context of computational inversion and other applications. However, there exists limited analysis of computational inversion using W_p for $p \neq 2$, and there do not appear to be any other previously known Fourier-based bounds for W_p, except for measures supported on a finite grid of points when $p = 1$ and 2 in [5] and measures supported on a circle $\mathbb{T}$ in [57]. These bounds are also discussed in Appendix A.

Standard optimal transport definitions and results: We will now review the standard optimal transport definitions and results used in this work. For a given cost function $c : \Omega \times \Omega \to \mathbb{R}_+$, the optimal transport cost represents the minimal cost of transporting one probability measure $\mu \in \mathcal{P}(\Omega)$ to another $\nu \in \mathcal{P}(\Omega)$.[2] This cost is formulated as the *Kantorovich problem*

$$KP(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} c(x, y) d\pi(x, y) \tag{5}$$

where $\Pi(\mu, \nu)$ is the space of all joint distributions $\pi \in \mathcal{P}(\Omega \times \Omega)$ (*transport plans*) with the marginals matching μ and ν:

$$\int_\Omega \pi(x, \cdot) = \mu(x), \quad \text{and} \quad \int_\Omega \pi(\cdot, y) = \nu(y).$$

When Ω is a *Polish space*, i.e., complete and separable metric space, and c is a lower semi-continuous function, KP admits a solution [53]. The corresponding dual problem is

$$DP(\mu, \nu) := \max_{\substack{\phi,\psi \in C(\Omega) \\ \phi \oplus \psi \leq c}} \int_\Omega \phi d\mu + \int_\Omega \psi d\nu = \max_{\phi \in c-\text{conc}} \int_\Omega \phi d\mu + \int_\Omega \phi^c d\nu \tag{6}$$

where $c-\text{conc}$ refers to the set of c-concave functions and ϕ^c refers to a c-transform of ϕ. If c is uniformly continuous and bounded on Ω, strong duality $KP = DP$ holds and DP admits a solution (ϕ, ϕ^c) referred to as *potentials* or *Kantorovich potentials*.[3]

When the cost function c is the p-th moment of a distance function d, the optimal transportation cost is also a distance, referred to as the *Wasserstein-p (W_p) distance*:[4]

$$W_p^p(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} d^p(x, y) d\pi(x, y). \tag{7}$$

We consider two metric spaces: the first is the hypercube in $\mathbb{R}^d$:

$$\mathbb{H}^d := [0, 1]^d \tag{8}$$

[2]It is possible to formulate transport over when μ and ν are defined over different domains, but we will not require this generalization in the present paper.

[3]A solution to KP may also exist, and strong duality may also hold under weaker assumptions than the ones set forth in the text accompanying this footnote. See, e.g., [53] for details.

[4]We use the terms "metric" and "distance" interchangeably, referring to the same mathematical object.

with $d_{\mathbb{H}^d}(x,y) := \|x-y\|$ where $\|\cdot\|$ is the standard Euclidean norm. The second space is the *flat d-dimensional torus*:

$$\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d.$$

The elements of $\mathbb{T}^d$ are equivalence classes $[x] = \{x+k \mid k \in \mathbb{Z}^d\}$ where $x \in [0,1)^d$. For for simplicity, we will denote $[x]$ by x. The metric space over $\mathbb{T}^d$ is equipped with the distance $d_{\mathbb{T}^d}(x,y) = \min_{k\in\mathbb{Z}^d} |x-y+k|$, which makes $\mathbb{T}^d$ a Polish space (see, e.g., [27]). Accordingly, the domain of integration over $\mathbb{T}^d$ is $[0,1)^d$, and the set of probability measures $\mathcal{P}(\mathbb{T}^d)$ is given by $\mathbb{Z}^d$ periodic measures μ on $\mathbb{R}^d$ such that each μ is a probability measure when restricted to $[0,1)^d$. Similarly each $f : \mathbb{T}^d \to \mathbb{R}$ is identified with $\mathbb{Z}^d$ periodic function $f : \mathbb{R}^d \to \mathbb{R}$ (see, e.g., [39]).

Notation: We will refer to $\mathbb{H}^d$ and $\mathbb{T}^d$ collectively as $\Omega := \{\mathbb{H}^d, \mathbb{T}^d\}$. We will denote the Wasserstein distance over $\mathbb{H}^d$ and $\mathbb{T}^d$ by $W_p^{\mathbb{H}^d}$ and $W_p^{\mathbb{T}^d}$ respectively, and when certain results holds for both $W_p^{\mathbb{H}^d}$ and $W_p^{\mathbb{T}^d}$, we will write W_p^{Ω}. We will omit these superscripts whenever the relevant metric space or spaces are clear from the context.

We will use $\|\cdot\|_p$ to denote the ℓ_p norm when applied to vectors or sequences (which we may also denote by $\|\cdot\|_{\ell_p}$ for further clarity), the induced 2-norm when applied to a matrix, or the L_p norm when applied to a function (which we may also denote by $\|\cdot\|_{L_p}$). When $p=2$, we may denote the relevant norm by $\|\cdot\|$ and omit the subscript. To match the integrand of the Benamou-Brenier formulation of W_p (Theorem 4.1) we also define the L_p norm of a vector field $F : \mathbb{R}^d \to \mathbb{R}^d$ as follows:[5]

$$\|F\|_{L_p} = \Big(\int_\Omega \|F(x)\|_2^p dx\Big)^{\frac{1}{p}}.$$

The symbol $\circledast$ represents a convolution, and $\hat{f}$ represents the Fourier series or Fourier transform of f, whichever is appropriate. For functions on Ω, we will denote the inhomogeneous and homogeneous Sobolev norms $\mathcal{H}^\beta$ and $\dot{\mathcal{H}}^\beta$ by, respectively:

$$\|f\|_{\mathcal{H}^\beta} = \Big(\sum_k \big(1+(2\pi\|k\|_2)^2\big)^\beta \hat{f}_k^2\Big)^{\frac{1}{2}} \text{ and } \|f\|_{\dot{\mathcal{H}}^\beta} = \Big(\sum_k (2\pi\|k\|_2)^{2\beta} \hat{f}_k^2\Big)^{\frac{1}{2}}. \tag{9}$$

$\mathcal{P}(\Omega)$ represents the set of probability measures over Ω, and $\mathcal{P}_2(\Omega)$ shall refer to such probability measures with finite second moments. $C^m(\Omega)$ refers to the set of continuous functions with m continuous derivatives on Ω, $C_c^m(\Omega)$ shall refer to such functions with compact support. We will also use $C := C^0$ to denote the set of continuous functions. The notation $a \lesssim b$ indicates that there exists a positive constant M for which $a \leq Mb$ holds, and $a \asymp b$ indicates that $a \lesssim b$ and $a \gtrsim b$. $\lesssim_p$ indicates that the constant M may depend on p. In

[5]Since all norms are equivalent in a finite-dimensional vector space, this norm is equivalent up to a constant pre-factor to the more standard definition:

$$\|F\|_{L_p} = \Big(\int_\Omega \sum_i |F_i(x)|^p dx\Big)^{\frac{1}{p}}.$$

the wavelet discussion, this constant may also depend on the specific choice of wavelets and dimension, which we will not indicate by a subscript. In all other cases, if there is no subscript in $\lesssim$, $\asymp$ or $\gtrsim$, this will indicate that the constant is uniform in all parameters. If the region of integration over a d-dimensional domain is omitted, then we will assume that it is $[0,1)^d$. $\mathbb{R}_+$ refers to the half-line $[0,\infty)$ and I refers to the interval $[0,1]$. $a \wedge b$ and $a \vee b$ refers to, respectively, the minimum and maximum of a and b.

The rest of the paper is organized as follows. In Section 2, we review related work, including applications of the W_p distance in computational problems and existing Fourier and wavelet-based bounds on this distance, as well as the fluid dynamics formulation of W_p. In Section 3, we introduce our approach by establishing elementary Fourier-based bounds for W_1 in one dimension. In Section 4, we establish our Fourier-based upper and lower bounds on W_p in higher dimension. In Section 5, we extend these bounds to generalized W_p for unbalanced measures. Using these bounds, we develop wavelet and Fourier-based resolution analysis in Section 6, and concluding remarks and suggestions for future work are offered in Section 7.

2 Related work

In this section, we review related work, including applications of the W_p distance in inverse and certain other computational problems, as well as existing Fourier and wavelet-based bounds on this distance.

2.1 W_p in computational problems

The W_p distance metricizes weak convergence of probability measures and has a number of attractive features for inverse and other computational problems. For example, $W_p(f,g)$ depends continuously on $f-g$ even when the densities f and g have non-overlapping support. Therefore, the W_p metric allows us to compare such densities more meaningfully than other popular distances, like the L^p and total variation norms. Currently, W_p is used extensively in machine learning and computer science, such as generative modeling [4] and robust estimation [47, 46] (see also overview of applications in [32, 56, 7]).

In statistics, the process of fitting a parametric model to data using W_p (instead of, for example, the Kullback-Leibler divergence used in likelihood maximization) is known as *minimum Kantorovich distance estimation* [9]. The W_p metric is also used in Bayesian statistics for likelihood-free inference [12, 14] and parameter estimation [8, 55, 13, 45]. However, in parametric statistics, the underlying problems tend to be low-dimensional. Such problems are intrinsically different from the high-dimensional inverse problems where reconstructing high-resolution information of m is a critical objective of the inversion process. On the other hand, recent work [48] in nonparametric statistics determined the minmax estimation rates when the error between the target density and its empirical distribution is measured in W_p using a characterization of Wasserstein distance in terms of weighted ℓ_p norms of the wavelet coefficients. This problem is similar to the high-dimensional inverse problems studied in this paper, and we use extensively the ideas and methods from that work.

While the L^2 norm, as well as other L^p norms, have been historically used in computational inverse problems, as noted previously, they do not provide a meaningful comparison whenever the support of the model and the data do not overlap, which can happen when the data is shifted relative to the model or when the data lies on a low dimensional manifold. Also, when the signals are wavelike, using the L^2 norm leads to incorrect matching when the model and the data have a significant phase mismatch (this phenomenon is referred to as *cycle skipping*).

The W_2 distance was applied and/or analyzed in the context of various inverse problems, such as the earthquake location problem [18], full waveform inversion [23, 25], and tomographic reconstruction [1]. In another work, a loss function based on the W_4 metric (which penalizes the outliers more heavily than the W_2 metric) was used in the context of computerized tomography (CT) [2]. Reference [22] introduced the W_2 distance as the mismatch functional in the context of seismic inverse problems. This reference showed that this distance is convex with respect to translations and dilations, which addresses the cycle-skipping issue mentioned above; see also [25]. The frequency content of computational solutions to inverse problems, as well the convexity of the optimization problems, based on the W_2 metric, have been recently analyzed in [21, 24]. The Fourier-based bounds on W_2 in [49] (discussed in Appendix A) were used in [24] to analyze the frequency content of computational solutions to inverse problems using this metric.

The W_1 distance was used in image and language processing [30] and Wasserstein GANs [4]. In the context of inverse problems, the generalization of the W_1 distance to general signed measures with different mass (Kantorovich-Rubinstein norm) was empirically shown to have attractive properties in the context of inverse problems [34, 41, 42]; see also [44] for a survey of results and numerical experiments indicating the attractive properties of this norm specifically in the full waveform inversion setting. A number of fast algorithms were developed to solve optimal transport based on the W_1 distance [37] and its variants such as entropy regularized W_1 [35, 36] and unbalanced W_1 [33]. Lastly, [29] developed a data-driven denoiser for inverse problems related to the W_1 metric; see also [60]. We refer interested readers to reference [17] for various mathematical properties on the W_1 metric and to references [52, 54] and references therein for the development of fast computational algorithms to evaluate the metric. However, we are not aware of any existing analysis of computational inversion using W_1 or, more generally, W_p for $p \neq 2$ as the mismatch functional.

3 Elementary Fourier-based bounds for W_1 in 1D

Before considering the multi-dimensional case, as a simple exercise to build intuition, let us develop Fourier-based bounds on W_1 in on the interval and circle. We consider the classic divergence formulation of W_1 due to Beckmann (which applies in 1D and higher dimension) [10], [53, Theorem 4.6]:

$$W_1(\mu,\nu) = \inf_{V \in \mathcal{M}^d_{\mathrm{div}}} \left\{ \int_\Omega \|V(x)\| dx \ \mid \ \nabla \cdot V = F \right\}, \tag{10}$$

where $F := \mu - \nu$ and $\mathcal{M}^d_{\mathrm{div}}$ denotes the space of vector measures on Ω with divergence which is a scalar measure. In the case of transport on $\mathbb{H}^d$, we additionally impose the boundary condition $V \cdot n|_{\partial\Omega} = 0$, which reduces to $V(0) = V(1) = 0$ on the interval $[0,1]$.

W_1 on an interval: In the case of transport in 1D, the divergence is simply the derivative. If f and g, the probability densities associated with μ and ν, have a Fourier expansion (16), then the Fourier coefficients of V are given by $\hat{V}_k = \hat{F}_k/(2\pi i k)$ for $k \neq 0$ where $\hat{F}_k = \hat{f}_k - \hat{g}_k$. The constant frequency coefficient

$$\hat{V}_0 = -\sum_{k\neq 0} \frac{1}{2\pi i k} \hat{F}_k$$

is determined by the above-mentioned boundary conditions, which require that there shall be no transport at the endpoints of the interval. Accordingly, this coefficient can be expressed in terms of the centers of mass of the measures. Let the *the center of mass* of μ be given by

$$m_\mu = \int_0^1 x \ f(x) dx. \tag{11}$$

Then, the constant frequency coefficient $\hat{V}_0$ of the transport field reflects the signed distance between the centers of mass of μ and ν:

$$m_\mu - m_\nu = \int_0^1 xF(x)dx = \sum_{k\neq 0} \hat{F}_k \int_0^1 x e^{2\pi i k x} dx = \sum_{k\neq 0} \frac{1}{2\pi i k} \hat{F}_k = -\hat{V}_0.$$

Accordingly, in the case of W_1 on the interval, the optimal V can be determined explicitly in terms of the Fourier coefficients of F. Then, by Hölder's inequality and Parseval's identity:

$$W_1(f,g) = \int_0^1 |V(x)| dx \leq \Big(\int_0^1 |V(x)|^2 dx \Big)^{\frac{1}{2}} = \big(|\hat{V}_0|^2 + \|\hat{F}\|^2_{2,w^1} \big)^{\frac{1}{2}}. \tag{12}$$

where w^1 is given by (2). A lower bound follows from another elementary computation: for any k

$$|\hat{V}(k)| = \Big| \int_0^1 V(x) e^{-i2\pi k x} dx \Big| \leq \|V\|_1 = W_1(f,g). \tag{13}$$

Therefore

$$\max \big(|\hat{V}_0|, \|\hat{F}\|_{\infty, w^1} \big) \leq W_1(f,g) \tag{14}$$

In higher dimensions, the divergence constraint in (10) implies that $2\pi i \langle k, \hat{V}_k \rangle = \hat{F}_k$. Therefore, the divergence operator has a nontrivial kernel, and the projection of V onto the kernel is determined as a result of the optimization in (10). Specifically,

$$\widehat{V}_k = \begin{cases} \frac{k}{2\pi i \|k\|^2} \widehat{F}_k + \widehat{Q}_k & \text{if } k \neq 0 \\ \widehat{Q}_0 & \text{if } k = 0 \end{cases} \tag{15}$$

where $\widehat{Q}_k \in \mathbb{C}^d$, and, for $k \neq 0$, $\widehat{Q}_k \in k^\perp$; these vectors $\widehat{Q}_k$ parametrize the kernel of the divergence operator in the Fourier domain (subject to V being a real-valued vector field, and in the case of transport on $\mathbb{H}^d$ also subject to the boundary condition). We are not aware of an explicit representation of the optimal $\widehat{Q}_k$'s solving (10) in dimension higher than 1.

4 Fourier-based bounds for W_p

In this section, we establish our main results: lower and upper bounds on $W_p(f,g)$ on $\mathbb{H}^d$ and $\mathbb{T}^d$, expressed in terms of the weighted ℓ_q norms for the Fourier coefficients of $f-g$. As noted previously, we assume that μ and ν are absolutely continuous with respect to the Lebesgue measure and the associated densities f and g have a Fourier series expansion in $L_2(\Omega)$

$$f = \sum_k \hat{f}_k \psi_k \text{ and } g = \sum_k \hat{g}_k \psi_k \tag{16}$$

where the Fourier basis functions are

$$\psi_k(x) := e^{2\pi i \langle k, x\rangle} = \prod_{i=1}^{d} \psi_{k_i}(x_i) \tag{17}$$

and $\psi_{k_i}(x_i) := e^{2\pi i k_i x_i}$ for $k \in \mathbb{Z}^d$. We will refer by Ψ to the set of the Fourier basis functions ψ_k for $k \in \mathbb{Z}^d$.

Note that the Fourier basis functions do not satisfy all the wavelet assumptions in Section 6.2; in particular, they fail Assumptions 4 (Locality) and 5 (Norm). Nevertheless, we can generalize the proofs in reference [48] to obtain Fourier-based bounds on W_p, as discussed below.

For $p=1$, we will use the Beckman formulation (10) to obtain the upper bound and the dual formulation of W_1

$$W_1 = \sup_{\|\nabla h\|_\infty \le 1} \int_\Omega h d(\mu - \nu) \tag{18}$$

to obtain the lower bound. Similarly to [48] for $p > 1$, we will use the following fluid-dynamics characterization of W^p due to [11],[16] to obtain lower and upper bounds. We denote by K_Ω the set of pairs of measures (ρ, E) on $\Omega \times [0,1]$ where ρ is scalar-valued and E is vector-valued. In the case of transport on $\mathbb{H}^d$, E must also satisfy the boundary condition $E \cdot n = 0$ on $\partial \mathbb{H}^d \times [0,1]$.

Theorem 4.1 (Benamou-Brenier)**.** *For any measures μ and ν on $\mathbb{H}^d$ and $p \in (1, \infty)$*

$$W_p^p(\mu,\nu) = \inf_{(\rho,E)\in K_\Omega} \{\mathcal{B}_p(\rho,E) : \rho(\cdot,1) = \mu, \rho(\cdot,0) = \nu, \partial_t \rho + \nabla_x \cdot E = 0\} \tag{19}$$

where

$$\mathcal{B}_p(\rho, E) := \begin{cases} \int_{\Omega\times[0,1]} \|\frac{dE}{d\rho}(x,t)\|^p d\rho(x,t) & \text{if } E \ll \rho \\ +\infty & \text{otherwise.} \end{cases}$$

Remark 4.2. *The preceding theorem has been generalized to the transport on $\mathbb{T}^d$ (Theorem 1.6.4 in [15] and references cited therein).*

4.1 Lower bounds

We will use the conjugate exponent p' of p given by

$$\frac{1}{p} + \frac{1}{p'} = 1. \tag{20}$$

When $p \in [1,2]$, we can control the $L_{p'}$ norm by the ℓ_p norm of the Fourier coefficients using the Hausdorff-Young inequality; see, e.g., Theorem 4.27 in [26].

Theorem 4.3 (Hausdorff-Young inequality)**.** *If $p \in [1,2]$ and $\hat{f} \in \ell_p$, then for a function f on $[0,1)^d$ represented by the Fourier series (16),*

$$\Big\| \sum_k \hat{f}_k \psi_k \Big\|_{L_{p'}} \leq \|\hat{f}\|_p. \tag{21}$$

We use the following lemma from reference [40] (Remark following Lemma 3.5). This lemma is based on the divergence formulation of W_1 in (18) when $p = 1$ and fluid dynamic formulation (19) of W_p for $p > 1$, which as discussed in Remark 4.2, extends to $W_p^{\mathbb{T}^d}$. Therefore, the lemma holds for W_p on $\mathbb{T}^d$ as well $\mathbb{H}^d$.

Lemma 4.4. *For all $h \in W^{1,q}(\Omega)$,[6] if μ and $\nu \in \mathcal{P}(\Omega) \cap L^s(\Omega)$ and $\|\nu\|_{L^s}, \|\nu\|_{L^s} \leq M$ and*

$$\frac{1}{q} + \frac{1}{p} + \frac{1}{s} = 1 + \frac{1}{ps}, \tag{22}$$

then

$$\int_\Omega h \ d(\mu - \nu) \leq M^{1/p'} \|\nabla h\|_{L_q(\Omega)} W_p(\mu, \nu). \tag{23}$$

To prove our lower bound, we construct a function h from the Fourier coefficients of $f - g$. We will control $\|\nabla h\|_{L_q(\Omega)}$ using the Hausdorff-Young inequality, which requires that $q \in [2, \infty]$. Accordingly, if $p = 1$, by (18), $q = \infty$. If $p > 1$, then by (22) $q = \frac{ps}{(p-1)(s-1)} = \frac{p's}{(s-1)}$. If $1 < p < 2$, then for all $s \geq 1$, we have $q \geq 2$. On the other hand, if $p > 2$, then $q \geq 2$ for all $s \in [1, \frac{2p-2}{p-2}]$. Based on this calculation, in Appendix B, we establish the following lower bound.

Theorem 4.5. *If f and $g \in \mathcal{P}(\Omega)$ have a Fourier expansion (16), then for $p \in [1, \infty)$,*

$$W_p(f, g) \geq d^{-\frac{1}{2}} \|\hat{f} - \hat{g}\|_{\infty, w^1}.$$

Furthermore, if $1 < p \leq 2$ and $\|f\|_{L^s}, \|g\|_{L^s} \leq M$ for any $s \in (1, \infty]$, or if $2 < p$ and $\|f\|_{L^s}, \|g\|_{L^s} \leq M$ for $s \in (1, \frac{2p-2}{p-2}]$, then

$$W_p(f, g) \geq d^{\frac{1}{q} - \frac{1}{2}} M^{-\frac{1}{p'}} \|\hat{f} - \hat{g}\|_{q, w^{q'}}$$

where $q = \frac{p's}{s-1}$. In each case, the bounds hold on $\mathbb{H}^d$ and $\mathbb{T}^d$, p' and q' are the conjugate exponents of p and q respectively given by (20) and the weights $w^{q'}$ are given by (2).

[6] $W^{1,q}(\Omega)$ denotes the Sobolev space of L^q functions whose gradient is also in $L^q(\Omega)$.

As a shorthand, the preceding bounds can be combined into a single expression:

$$W_p(f,g) \geq d^{\frac{1}{q}-\frac{1}{2}} M^{-\frac{1}{p'}} \|\hat{f}-\hat{g}\|_{q,w^{q'}}$$

where

$$q := \begin{cases} \infty & \text{if } p \in [1,\infty) \\ \frac{p's}{s-1} & \text{if } p \in (1,2] \text{ and } s \in (1,\infty] \\ \frac{p's}{s-1} & \text{if } p \in (2,\infty) \text{ and } s \in (1, \frac{2p-2}{p-2}]. \end{cases} \tag{24}$$

Note that $d^{-\frac{1}{2}} \leq d^{\frac{1}{q}-\frac{1}{2}} \leq 1$ for $q \geq 2$. Therefore, if necessary, we can eliminate the dependence on q from the lower bound.

4.2 Upper bounds

We construct a vector field $V : \Omega \to \mathbb{R}^d$

$$V := \sum_k (\hat{f}_k - \hat{g}_k) V_k.$$

satisfying $\nabla \cdot V = f - g$. This field is used to construct a feasible point for the divergence formulation (10) of W_1, as well as, together with ρ defined in the proof of Theorem 4.7 in Appendix D, for the fluid dynamics formulation (19) of W_p when $p > 1$.

On the torus: On $\mathbb{T}^d$, we take

$$(V_k)_j := \begin{cases} \frac{c_{j,k}}{i2\pi k_j}\psi_k & \text{if } k_j \neq 0 \\ 0 & \text{if } k_j = 0 \end{cases} \tag{25}$$

for the standard Fourier basis functions ψ_k in (17). For each Fourier frequency $k \in \mathbb{Z}^d \setminus 0$, taking $c_{1,k}, \dots, c_{d,k}$ such that $\sum_{j:k_j \neq 0} c_{j,k} = 1$, ensures that

$$\nabla \cdot V_k = \psi_k \text{ on } \Omega \tag{26}$$

since

$$\frac{\partial}{\partial j}(V_k)_j = \begin{cases} c_{j,k}\psi_k & \text{if } k_j \neq 0 \\ 0 & \text{if } k_j = 0 \end{cases}$$

Specifically, for $c_{j,k} = \frac{k_j^2}{\|k\|_2^2}$, the Fourier coefficients $\hat{V}_k$ match those of the inverse (15) of the divergence operator acting on $f-g$ with the projection on the operator's kernel (represented by $\hat{Q}_k$'s) equal to zero. In Appendix C, we prove the following upper bound on the L_p norm of V.

Lemma 4.6. *For $p \in [1, \infty)$ and the weights w^ζ given by (2):*

$$\|V\|_{L_p} \leq C_{d,p} \|\hat{f} - \hat{g}\|_{\zeta, w^\zeta} \tag{27}$$

where the Hölder conjugate p' is given by (20),

$$\zeta = \begin{cases} 2 & \text{if } p \in [1,2] \\ p' & \text{if } p \in (2, \infty) \end{cases} \qquad \text{and} \qquad C_{d,p} = \begin{cases} 2^{-\frac{1}{p'}} & \text{if } p \in [1,2] \\ 2^{-\frac{1}{p'}} d^{\frac{1}{2}-\frac{1}{p}} & \text{if } p \in (2, \infty) \end{cases} \tag{28}$$

Based on this bound, in Appendix D, we establish a Fourier-based upper bound on W_p (that parallels Proposition 1 in [48] establishing a similar wavelet-based bound).

Theorem 4.7. *For $p = 1$, we have*

$$W_1^{\mathbb{T}^d}(f, g) \leq \|\hat{f} - \hat{g}\|_{2, w^2}$$

and for $p > 1$, if for almost every $x \in \mathbb{T}^d$,

$$f(x) \wedge g(x) \geq \xi > 0 \tag{29}$$

then

$$W_p^{\mathbb{T}^d}(f, g) \leq C_{d,p}\ p\ \xi^{-1/p'} \|\hat{f} - \hat{g}\|_{\zeta, w^\zeta}$$

where ζ and $C_{d,p}$ are given by (28).

Again, as a shorthand, the preceding bounds can be combined as

$$W_p^{\mathbb{T}^d}(f, g) \leq C_{d,p}\ p\ \xi^{-1/p'} \|\hat{f} - \hat{g}\|_{\zeta, w^\zeta} \tag{30}$$

If $f - g$ belongs to the Sobolev space $\dot{\mathcal{H}}^\beta$ for sufficiently large β, then (27) can be upper bounded using the square root of product of this norm and the weighted $\ell_{p'}$ for $p' \in (2, \infty]$. We prove the following result in Appendix E.

Proposition 4.8. *If $f - g \in \dot{\mathcal{H}}^\beta$ and*

$$\beta > \frac{d}{p} - \frac{d}{2} - 1, \tag{31}$$

for $p \in [1, 2)$, then

$$\|\hat{f} - \hat{g}\|_{2, w^2} \lesssim \|f - g\|_{\dot{\mathcal{H}}^\beta}^{\frac{1}{2}} \|\hat{f} - \hat{g}\|_{p', w^2}^{\frac{1}{2}} \tag{32}$$

where the weights w^2 are given by (2).

If $\|f(m) - g\|_{\dot{\mathcal{H}}^\beta}$ is bounded uniformly in m over the feasible set of parameters of the forward model, this proposition leads to an upper bound on W_p in terms of the weighted $\ell_{p'}$ norm instead of the weighted ℓ_2 norm upper bound. The latter norm matches the norm in a lower bound given by Theorem 4.5 when $p = 1$ or $p \in (1, 2]$.

On the hypercube: In the case of transport on $\mathbb{H}^d$, the vector field V also needs to satisfy the boundary condition $V \cdot n = 0$ on $\partial\Omega$. Let $\tilde{V}_k$ be given by V_k in (25) except that each $(V_k)_j$ is evaluated on the boundary $x_j = 0$:

$$(\tilde{V}_k)_j(x) := (V_k)_j(x_1, \ldots, x_{j-1}, 0, x_{j-1}, \ldots, x_d) = 0 \tag{33}$$

We also let V_k' denote V_k modified by subtracting its boundary value:

$$\begin{aligned}(V_k')_j(x) &:= (V_k)_j(x) - (\tilde{V}_k)_j(x)\\ &= \begin{cases} \frac{c_{j,k}}{i2\pi k_j}(\psi_{k_j}(x_j) - 1) \prod_{m\neq j} \psi_{k_m}(x_m)) & \text{if } k_j \neq 0 \\ 0 & \text{if } k_j = 0\end{cases}\end{aligned}$$

Note that $\tilde{V}_k$ is in the kernel of the divergence operator, and therefore, (26) holds with respect to V_k' instead of V_k. If we take $c_{j,k} = k_j^2/\|k\|_2^2$, as we did previously section, enforcing the boundary condition will lead to a mixing of frequencies that appears difficult to analyze. However, if we take $c_{k,j} = 1$ if j is the index of the component of k with the largest absolute value, i.e. $|k_j| = \|k\|_\infty$, and zero otherwise, the analysis becomes more tractable. If there are multiple such $c_{j,k}$, we can set any one of them, e.g., the smallest one, to 1 and set the remaining ones to zero. Specifically,

$$c_{k,j} = \begin{cases} 1 & \text{if } j = \arg\min_{i \text{ s.t. } |k_i| = \|k\|_\infty} i \\ 0 & \text{otherwise}\end{cases}.$$

Given $(k_1, \ldots, k_{j-1}, k_{j+1}, \ldots, k_d)$, let k^j denote a set of all frequencies k_j such that $|k_j| = \|k\|_\infty$ for each $k \in k^j$:

$$k^j = \Big\{(k_1, \ldots, k_{j-1}, k_j, k_{j+1}, \ldots k_d) \mid k_j \in \mathbb{Z} \text{ s.t. } |k_j| \geq \max_{m\neq j} |k_m|\Big\}.$$

We let $\hat{f}_{k^j}$ denote a sequence of Fourier coefficients corresponding to each $k_j \in k^j$. Accordingly, in the present discussion, summation over k^j denotes summation over all possible combinations of $k_1, \ldots, k_{j-1}$ and $k_{j+1}, \ldots, k_d$ while summation over k_j entails summation over the elements of k^j given a specific $k_1, \ldots, k_{j-1}$ and $k_{j+1}, \ldots, k_d$. Based on that, we can obtain an upper bound for $W_p^{\mathbb{H}^d}$ that contains an additional term attributable to the boundary conditions relative to the upper bound for $W_p^{\mathbb{T}^d}$ in (30).

Lemma 4.9. *For $p \in [1, \infty)$,*

$$\|V'\|_{L_p} \leq C_{d,p}\Big(\|\hat{f} - \hat{g}\|_{\zeta, w^\zeta} + \Big(\sum_j \sum_{k^j} \|\hat{f}_{k^j} - \hat{g}_{k^j}\|_{1,w^\infty}^\zeta\Big)^{\frac{1}{\zeta}}\Big) \tag{34}$$

where ζ and $C_{d,p}$ are given by (28) and the weights w^ζ and w^∞ are given by (2).

Remark 4.10. *Theorem 4.7 shall also apply to $W_p^{\mathbb{H}^d}(f, g)$ as modified by replacing the right-hand side of (30) with*

$$C_{d,p}\ p\ \xi^{-\frac{1}{p'}}\Big(\|\hat{f} - \hat{g}\|_{\zeta, w^\zeta} + \Big(\sum_j \sum_{k^j} \|\hat{f}_{k^j} - \hat{g}_{k^j}\|_{1,w^\infty}^\zeta\Big)^{\frac{1}{\zeta}}\Big)\Big).$$

5 Unbalanced transport

In this section, we extend the foregoing bounds to generalized W_p for unbalanced positive measures and generalized W_1 for signed measures.

5.1 Generalized W_p for positive measures

We consider the unbalanced $W_p^{a,b}$ metric introduced in [50]:

$$W_p^{a,b}(\mu,\nu) := \left(T_p^{a,b}(\mu,\nu)\right)^{\frac{1}{p}}$$

where

$$T_p^{a,b}(\mu,\nu) := \inf_{\substack{\tilde{\mu},\tilde{\nu}\in\mathcal{M}(\Omega)\\ |\tilde{\mu}|=|\tilde{\nu}|}} a^p(|\mu-\tilde{\mu}|+|\nu-\tilde{\nu}|)^p + b^p\ W_p^p(\tilde{\mu},\tilde{\nu}),$$

$\mathcal{M}(\Omega)$ is the space of positive Borel regular measures on Ω with finite mass, and $a, b > 0$.[7] In this section, we develop Fourier-based upper and lower bounds for $W_p^{a,b}$ on $\mathbb{T}^d$ and $\mathbb{H}^d$.

Proposition 2 in [50] guarantees that there exist optimal $\tilde{\mu}$ and $\tilde{\nu}$ such that $\tilde{\mu} \le \mu$ and $\tilde{\nu} \le \nu$. Therefore, $\|\tilde{\mu}\|_{L^s} \le \|\mu\|_{L^s}$ and $\|\tilde{\nu}\|_{L^s} \le \|\nu\|_{L^s}$. Based on this, in Appendix G, we generalize Lemma 4.4 to the unbalanced metric.

Lemma 5.1. *For all $h \in W^{1,q}(\Omega)$, if μ and $\nu \in \mathcal{M}(\Omega)\cap L^s(\Omega)$, $\|\mu\|_{L^s}, \|\nu\|_{L^s} \le M$, q satisfies (22), and*

$$\|h\|_\infty \le a \text{ and } d^{\frac{1}{2}-\frac{1}{q}} M^{1/p'}\|\nabla h\|_q \le b,$$

then

$$\int_\Omega h\ d(\mu-\nu) \le 2^{\frac{p-1}{p}} W_p^{a,b}(\mu,\nu).$$

We again assume that the measures μ and ν are absolutely continuous with respect to the Lebesgue measure and, therefore, are associated with a pair of positive densities f and g. We will denote $W_p^{a,b}(\mu,\nu)$ as $W_p^{a,b}(f,g)$ by reference to the corresponding density. In Appendix H we use this lemma to adjust the construction of the test function h from the proof of Theorem 4.5 in Appendix B. This generalizes our previous lower bound to the unbalanced setting.

Theorem 5.2. *If f and g are positive functions that have a Fourier expansion (16), then for $p \in [1,\infty)$,*

$$W_p^{a,b}(f,g) \ge d^{-\frac{1}{2}} b\|\hat{f}-\hat{g}\|_{\infty,w^1} \wedge a.$$

Furthermore, if $p \in (1,2]$ and $\|f\|_{L^s}, \|g\|_{L^s} \le M$ for $s \in (1,\infty]$, or if $p > 2$ and $\|f\|_{L^s}, \|g\|_{L^s} \le M$ for $s \in (1, \frac{2p-2}{p-2}]$, then

$$W_p^{a,b}(f,g) \ge 2^{(1-p)/p} \left(\left(d^{\frac{1}{q}-\frac{1}{2}} M^{-1/p'} b\right) \wedge \left(\frac{a}{\|(\hat{f}-\hat{g})^{\frac{1}{p}}\|_{q,w^{q'}}}\right)\right) \|\hat{f}-\hat{g}\|^q_{q,w^{q'}}$$

[7] The Wasserstein metric is a well-defined distance for an arbitrary pair of positive measures f and g of equal mass since it is homogeneous under scalar multiplication: $W_p(\mu,\nu) := cW_p\left(\frac{1}{c}\mu, \frac{1}{c}\nu\right)$.

and $q = \frac{p's}{s-1}$. *In each case, the bounds hold on* $\mathbb{H}^d$ *and* $\mathbb{T}^d$, p' *and* q' *are the conjugate exponents of* p *and* q *respectively given by* (20) *and the weights* w^p *and* $w^{q'}$ *are given by* (2) *for* $k \in \mathbb{Z}^d \setminus 0$ *while the weight associated with the zero's frequency* $\hat{f}_0 - \hat{g}_0$ *is* $w_0^{q'} = 1$.

As a shorthand, the preceding bounds can be combined into a single expression:

$$W_p^{a,b}(f,g) \geq 2^{(1-p)/p} \left(\left(d^{\frac{1}{q}-\frac{1}{2}} M^{-1/p'} b \right) \wedge \left(\frac{a}{\|(\hat{f}-\hat{g})^{\frac{1}{p}}\|_{q,w^{q'}}} \right) \right) \|\hat{f}-\hat{g}\|_{q,w^{q'}}^q$$

where q is given by (24).

For purposes of an upper bound, we assume, without loss of generality, that $\hat{f}_0 - \hat{g}_0 = \int_\Omega (f-g) \leq 0$. To preserve the positivity of the measures, we take $\tilde{f} = f - \hat{f}_0 + \hat{g}_0$ and $\tilde{g} = g$. Note that all the Fourier amplitudes of $\tilde{f}$ and f are the same except for the one corresponding to the frequency $k = 0$. Therefore, by a standard result Equation (76)

$$W_p^{a,b}(f,g) \leq \left(a^p |f - \tilde{f}|^p + b^p W_p^p(\tilde{f}, g) \right)^{\frac{1}{p}} \tag{35}$$

$$\leq 2^{\frac{1}{p}-\frac{1}{\zeta}} \left(a^\zeta |\hat{f}_0 - \hat{g}_0|^\zeta + b^\zeta W_p(\tilde{f}, g)^\zeta \right)^{\frac{1}{\zeta}}. \tag{36}$$

Based on that, we extend our previous upper bounds to the unbalanced case as follows.

Remark 5.3. *Theorem 4.7 applies to* $W_p^{a,b}(f,g)$ *on* $\mathbb{T}^d$ *provided that the bound* (30) *shall be replaced with*

$$W_p^{a,b}(f,g) \leq 2^{\frac{1}{p}-\frac{1}{\zeta}} C_{d,p} p \; b \; \xi^{-\frac{1}{p'}} \|\hat{f}-\hat{g}\|_{\zeta,w^\zeta}$$

where ζ *is given by* (28) *and the weighted* ℓ_ζ *norm shall include the weight* $w_0^\zeta = a/(bC_{d,p} p \, \xi^{-\frac{1}{p'}})$ *associated with the zero's frequency* $\hat{f}_0 - \hat{g}_0$.

Remark 5.4. *Proposition 4.8 extends to* $W_p^{a,b}(f,g)$ *provided that in such case,* (32) *shall be replaced with*

$$\|\hat{f}-\hat{g}\|_{2,w^2} \lesssim 2^{\frac{1}{p}-\frac{1}{\zeta}} \left(a^\zeta |\hat{f}_0 - \hat{g}_0|^\zeta + b^\zeta \|f-g\|_{\dot{\mathcal{H}}^\beta}^{\frac{\zeta}{2}} \|\hat{f}-\hat{g}\|_{p',w^2}^{\frac{\zeta}{2}} \right)^{\frac{1}{\zeta}}$$

where for purposes of computing the weighted $\ell_{p'}$ *norm we exclude the zero's frequency.*

In the case of $p = 1$, on $\mathbb{T}^d$, we have

$$W_1^{a,b}(f,g) \leq \sqrt{2} b \|\hat{f}-\hat{g}\|_{2,w^2} \tag{37}$$

where for purposes of computing the weighted ℓ_2 norm, we include the zero's frequency $\hat{f}_0 - \hat{g}_0$ with the weight a/b. If $f - g \in \dot{\mathcal{H}}^\beta$ for $\beta > d/2 - 1$, then also

$$W_1^{a,b}(f,g) \lesssim a|\hat{f}_0 - \hat{g}_0| + b\|f-g\|_{\dot{\mathcal{H}}^\beta}^{\frac{1}{2}} \|\hat{f}-\hat{g}\|_{\infty,w^2}^{\frac{1}{2}} \tag{38}$$

where for purposes of computing the weighted ℓ_∞ norm we exclude the zero's frequency.

5.2 Generalized W_1 for signed measures

For general (potentially unbalanced) signed measures μ and ν that can be decomposed into positive measures:

$$\mu = \mu^+ - \mu^- \text{ and } \nu = \nu^+ - \nu^-,$$

references [3] and [51] generalized the W_1 metric as follows:[8]

$$\mathcal{W}_1^{a,b}(\mu, \nu) := W_1^{a,b}(\mu^+ + \nu^-, \mu^- + \nu^+)$$

Reference [51] also showed that for $a = b = 1$, this metric is equivalent to the Kantorovich-Rubinstein norm [28], also called bounded Lipschitz distance or Fortet–Mourier distance [59]:

$$W_1^{1,1}(\mu, \nu) = \sup \left\{ \int \phi \; d(\mu - \nu) : \phi \in \mathcal{C}^0, \|\phi\|_\infty \leq 1, \|\phi\|_{Lip} \leq 1 \right\}$$

where $\mathcal{C}^0$ is a set of continuous real-valued functions on $\mathbb{R}^d$; see also [50]. The Kantorovich-Rubinstein norm was used in seismic inversion in references [34, 41, 42, 44] mentioned previously. Accordingly, the Fourier-based bounds in the previous section for $W_1^{a,b}(f^+ + g^-, f^- + g^+)$ apply immediately to $\mathcal{W}_1^{a,b}(f, g)$.

6 Resolution in computational inversion

In this section, we study the implication of the Fourier-based bounds, as well as the existing wavelet-based bounds established in [48], in the resolution analysis of computational inversion methods based on the W_p metrics. We focus on the case when the forward operator is linear (when f is nonlinear, our case can represent the linearization of f around some estimated solution m_0). Our analysis mirrors the resolution analysis in [6] and [24] with respect to the L^2 and $\mathcal{H}^s$ norms, respectively.

Let $\mathcal{M}$ and $\mathcal{G}$ be two function spaces and $A : \mathcal{M} \mapsto \mathcal{G}$ a linear map between them. We are interested in solving the linear problem of the form

$$Am = g. \tag{39}$$

We assume that A is invertible and solve this problem by finding m_δ such that $W_p(Am_\delta, g) \leq \delta$. We denote the noisy signal by $g^\delta := Am$ and the noise by the function $n : \Omega \to \mathbb{R}$

$$n := g^\delta - g. \tag{40}$$

In the settings we consider in this paper, the noise is either attributed to the early stopping of the minimization of the W_p distance (optimization error) or inaccurate measurements (estimation error). We assume that n is in the range of A and

$$\int_\Omega n(x)dx = 0. \tag{41}$$

[8]Reference [3] showed that this procedure fails to yield a distance for $p \neq 1$ since the resulting mismatch functional fails to satisfy the triangular inequality.

Since A is invertible, we denote the solution recovered from the noisy data by

$$m_\delta = A^{-1} g^\delta.$$

6.1 Resolution of frequencies

Let the function n defined by (40) be represented by a Fourier series $\widehat{n}$, and denote its weighted ℓ_q norm by

$$\delta_q := \|\widehat{n}\|_{q,w^2} \tag{42}$$

where q will be specified later. (Equation (41) guarantees that $\widehat{n}_0 = 0$.) Let B represent a bandwidth-limited approximation of A^{-1} given in the Fourier domain by

$$\widehat{(Bg)}_k \asymp \begin{cases} \widehat{(A^{-1})g}_k & \text{if } \|k\|_2 \le k_c \\ 0 & \text{if } \|k\|_2 > k_c \end{cases}$$

and let m_δ^c represent the bandwidth-limited approximation of m_δ given by

$$m_\delta^c := B g_\delta.$$

The following elementary theorem, proved in Appendix I, provides an upper bound on the reconstruction error of B in terms of δ_q.

Theorem 6.1. *For m, m_δ^c and n, as defined above, if $m \in \dot{\mathcal{H}}^r$,*

$$\|m - m_\delta^c\|_{L^2} \le (2\pi k_c)^{-r} \|m\|_{\dot{\mathcal{H}}^r} + \|Bn\|_{L^2} \tag{43}$$

De-smoothing inversion: We assume that $\|B\|_{\ell_{q,w^2} \to L^2} \lesssim k_c^\alpha$ for some $\alpha > 0$ where $\|\cdot\|_{\ell_{q,w^2} \to L^2}$ is the operator norm of a map from the weighted ℓ_q space of Fourier coefficients to L^2. We shall refer to the corresponding B as a *de-smoothing* inverse operator.[9] Then, if $r > 0$, the upper bound (43) is minimized by

$$k_c^{\alpha + r} \lesssim \frac{(2\pi)^{-r} r}{\alpha} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{\delta_q}.$$

On the other hand, if $r \le 0$, then the optimal cut-off frequency $k_c = 0$, i.e., no recovery is possible. Also, if $\|B\|_{\ell_{q,w^2} \to L^2}$ is bounded from above uniformly in k_c, then, if $r \ge 0$ these upper bounds are minimized by setting $k_c \to \infty$ and reconstructing all the frequencies, while if $r < 0$, then the optimal cut-off frequency is again $k_c = 0$.

If $\|n\|_{\dot{\mathcal{H}}^\beta} \le z$, then one can consider the operator norm of B as a map to L^2 from the weighted ℓ_q space of Fourier coefficients, restricted to coefficients of functions whose $\dot{\mathcal{H}}^\beta$ norm is bounded by z. In this case let us assume that that $\|Bn\|_{L^2} \asymp k_c^{\alpha'} h(z) \delta_q^\varepsilon$ for some constants $\alpha', \varepsilon > 0$ and function h. Then, if $r > 0$, the upper bound (43) is minimized by

$$k_c^{\alpha' + r} \lesssim \frac{(2\pi)^{-r} r}{\alpha'} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{h(z)\delta_q^\varepsilon}.$$

[9] It is also possible to find optimal k_c by a similar computation if $r > 0$ and $\|B\|_{\ell_{q,w^2} \to L^2}$ is proportional to another increasing function of k_c, e.g. if $\|B\|_{\ell_{q,w^2} \to L^2} \asymp \log k_c$, then the upper bound (43) is minimized when $k_c^r \asymp r\|m\|_{\dot{\mathcal{H}}^r}/(\delta_q)$. On the other hand, if $r \le 0$, then the optimal $k_c = 0$.

Early stopping of W_p minimization: Since A is invertible and g is in its range, the same exact solution m is obtained by minimizing to zero the mismatch between Am and g using W_p or any other metric. However, the present model allows us to specifically analyze computational inversion when the minimization

$$\begin{aligned} &\min_m W_p(Am, g) \\ &\text{s.t. } Am \in K_i \end{aligned} \tag{44}$$

is subject to the *early stopping condition* $W_p(Am, g) \le \delta$. In such case, let m_δ denote the approximate solution; we denote the exact solution by m^*, such that $g = Am^*$. We define the noisy data by reference to the early stopping solution $g^\delta := Am_\delta$. Then, the "noise" attributable to the early stopping is

$$n = g^\delta - g = Am_\delta - g.$$

p	$[1, \infty)$	$(1, 2]$	$(2, \infty)$
s	1	$(1, \infty]$	$\left(1, \frac{2p-2}{p-2}\right]$
$K_s : \|Am\|_{L_s}, \|g\|_{L_s} \le M$	$k_c^{\alpha+r} \lesssim d^{\frac{1}{2}} \frac{(2\pi)^{-r} r}{\alpha} \cdot \frac{\|m\|_{\mathcal{H}^r}}{\delta}$	$k_c^{\alpha+r} \lesssim_M d^{\frac{1}{2}-\frac{1}{q}} \frac{(2\pi)^{-r} r}{\alpha} \cdot \frac{\|m\|_{\mathcal{H}^r}}{\delta}$	

Table 3: The upper bounds on the resolution of frequencies k^c when the minimization in (44) is subject to the early stopping threshold δ, α is given by (47) and the optimization constraint is K_s, which holds automatically for $s = 1$ since we assume that Am and g are probability densities.

We assume that the data g is probability density $\mathcal{P}(\mathbb{T}^d)$ or $\mathcal{P}(\mathbb{H}^d)$ and the constraint K_i also requires Am to be a probability density. For purposes of the upper bound on k_c, we can consider transport on the flat torus or hypercube (where that the Fourier-based W_p lower bounds are the same). For purposes of the lower bound on k_c, to simplify the calculations, we consider the torus only. If g belongs to L_s for $s \in (1, \infty]$ when $p \in [1, \infty)$ or $s \in (1, (2p-2)/(p-2)]$ when $p \in (2, \infty)$, then we can include a constraint requiring that Am shall belong to

$$K_s := \{f \in \mathcal{P}(\Omega) \mid \|f\|_{L_s} \le M\}$$

where $M := \|g\|_{L_s}$. ($K_1 = \mathcal{P}(\Omega)$ holds trivially.) For purposes of the lower bound we include the constraint requiring that Am shall belong to:

$$K_\xi := \{f \in \mathcal{P}(\mathbb{T}^d) \mid f \ge \xi \text{ a.e. on } \mathbb{T}^d\}$$

where ξ satisfies

$$g \ge \xi \text{ a.e. in } \mathbb{T}^d. \tag{45}$$

(We can always take $\xi = 0$ since $g \in \mathcal{P}(\mathbb{T}^d)$.) Moreover, if $p \in [1,2]$ and $g \in \dot{\mathcal{H}}^\beta$ where β satisfies the hypothesis of Proposition 4.8, then we can take

$$K_{\xi,z} := \{f \in \mathcal{P}(\mathbb{T}^d) \mid f \geq \xi \text{ a.e. on } \mathbb{T}^d,\ \|f - g\|_{\dot{\mathcal{H}}^s} \leq z\}$$

where $z > 0$ is some constant and ξ is given by (45). Using the lower and upper bounds in Theorem 4.5 and Theorem 4.7, we obtain upper and lower bounds on the resolution k_c corresponding to each function class. These upper and lower bounds are set forth in Table 3 and Table 4, respectively. In the remainder of this discussion, we assume that $\alpha > 0$ and $\alpha' > 0$ and $r \leq 0$. We will refer to $\|m\|_{\dot{\mathcal{H}}^r}/\delta_q$ and $\|m\|_{\dot{\mathcal{H}}^r}/\sqrt{z\delta_q}$ as the *signal-to-noise* ratios. We observe that, holding the early stopping threshold δ constant, for the sufficiently large signal-to-noise ratio, the leading order behavior of k_c as a function of p, s and d will be determined by $\alpha + r$ and, if applicable, $\alpha' + r$, h and ε.

	$p = 1,\ \xi = 0$	$p \in (1,2],\ \xi > 0$	$p \in (2,\infty),\ \xi > 0$
$K_\xi : Am \wedge g \geq \xi$	$k_c^{\alpha+r} \gtrsim \frac{(2\pi)^{-r}r}{\alpha} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{\delta}$	$k_c^{\alpha+r} \gtrsim_\xi \frac{(2\pi)^{-r}r}{\alpha} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{\delta}$	$k_c^{\alpha+r} \gtrsim_\xi \frac{1}{p} d^{\frac{1}{p'}-\frac{1}{2}} \frac{(2\pi)^{-r}r}{\alpha} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{\delta}$
$K_{\xi,z} : Am \wedge g \geq \xi$ $\|Am - g\|_{\dot{\mathcal{H}}^\beta} \leq z$	$k_c^{\alpha'+r}$ $\gtrsim \frac{(2\pi)^{-r}r}{\alpha'} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{h(z)\delta^\varepsilon}$	$k_c^{\alpha'+r}$ $\gtrsim_\xi \frac{(2\pi)^{-r}r}{\alpha'} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{h(z)\delta^\varepsilon}$	

Table 4: The lower bounds on the resolution of frequencies k^c when the minimization in (44) is subject to the early stopping threshold δ.

Diagonal operators in the Fourier domain: The exponent α may depend on the dimension d and the exponent q, in addition to the intrinsic properties of the forward map A. To illustrate this, consider A that is diagonal in the Fourier domain and decays algebraically:

$$\widehat{A}_k = \|k\|_2^{-\gamma}\,. \tag{46}$$

While it might be too simplistic to assume that the operator is diagonal in the Fourier domain, i.e, that there is no mixing of frequencies, the decay behavior is universal in many inverse problems for physical models, such as inverse coefficients problems for partial differential equations [31]. In such problems, the forward operators are often smoothing operators with the degree of smoothing parameterized by γ. Moreover, unless the forward model has a truly significant mixing across a wide band of frequencies, the asymptotic resolution analysis we perform here (in the high-frequency regime, as the noise is mainly assumed to be of high frequency) should still provide useful insight in the context of more complicated forward operators.

Our bandwidth-limited approximation B of A^{-1} is given in the Fourier domain by

$$\widehat{B}_k = \begin{cases} \|k\|_2^\gamma & \text{if } \|k\|_2 \leq k_c \\ 0 & \text{if } \|k\|_2 > k_c \end{cases}$$

Then we obtain the following bound, shown in Appendix J.

Remark 6.2. *When A is diagonal in the Fourier domain, as described above, $q \in [1, \infty]$, and α given by*

$$\alpha = \begin{cases} 1 + \gamma + \frac{d}{2} - \frac{d}{q} & \text{if } \gamma > -1 \\ \frac{d}{2} - \frac{d}{q} & \text{if } \gamma \leq -1 \end{cases}. \tag{47}$$

is strictly positive, we have

$$\|B\|_{\ell_{q,w^2} \to L^2} \lesssim k_c^{\alpha}$$

Furthermore if $\|n\|_{\dot{\mathcal{H}}_\beta} \leq z$, then

$$\|Bn\|_{\ell_{1,w^2} \to L^2} \lesssim k_c^{\alpha'} \sqrt{z \delta_q}.$$

where

$$\alpha' = \begin{cases} 1 + 2\gamma - \beta + \frac{d}{2} - \frac{d}{q} & \text{if } \gamma > \beta/2 - 1/2 \\ \frac{d}{2} - \frac{d}{q} & \text{if } \gamma \leq \beta/2 - 1/2 \end{cases} \tag{48}$$

We assume that γ, d and q and if applicable β are such that $\alpha > 0$ or $\alpha' > 0$. Plugging these expressions of α and α' in Table 3 and Table 4, respectively, we observe that, holding the early stopping threshold δ constant, for the sufficiently large signal-to-noise ratio:

- For all p in W_p, the bounds on the resolution k_c of m_δ^c will either remain constant or increase if we increase p (depending on regularity of g);
- For $p \in [1, 2]$, the bounds on the resolution k_c will either remain constant or decrease if we increase the dimension d; and
- For $p \in (2, \infty)$, the bounds on the resolution k_c will increase if we increase the dimension d.

Optimal resolution with fixed noise: In contrast to the previous discussion where we held the early stopping threshold δ constant as we changed p, if we hold the noise n constant as we decrease p, by the monotonicity of the ℓ_q norm, δ_q and $\delta_{p'}$ given by (53) will decrease. In this setting, we show that the resolution may increase when we decrease p by constructing a specific example of high-frequency noise.

We focus on the case when $\|g\|_\infty = M < \infty$, and $g \geq 2\xi > 0$ a.e. on $\mathbb{T}^d$. In this setting, the observed data $g_\delta = g + \tilde{n}$ is corrupted with high-frequency bandwidth-limited noise $\tilde{n}$ constructed as follows. We start with the function n given by

$$\hat{n}_k = \begin{cases} 0 & \text{if } \|k\|_2 < k_n \\ \|k\|_2^{-\eta} & \text{if } k_n \leq \|k\|_2 \leq b k_n \\ 0 & \text{if } \|k\|_2 > b k_n \end{cases} \tag{49}$$

for some constant $b > 1$. As shown in Appendix K taking $\eta > 0$ satisfying

$$d < \eta \tag{50}$$

guarantees $\|n\|_{L_\infty} \leq C_{\eta,d}$ uniformly in k_n for a fixed b. Then, we rescale the noise

$$\tilde{n} := \frac{\xi}{C_{\eta,d}} n \tag{51}$$

which ensures that $\|\tilde{n}\|_{L_\infty} \leq \xi$ and therefore $g_\delta = g + \tilde{n} \geq \xi$ a.e. on $\mathbb{T}^d$. By the monotonicity of $L^p(\mathbb{T}^d)$, $\|\tilde{n}\|_{L_s} \leq \xi$ for all $s \in [1, \infty)$. When $p \in [1, 2)$, for $\beta > \frac{d}{p} - \frac{d}{2} - 1$, we also bound $\|\tilde{n}\|_{\dot{\mathcal{H}}^\beta} \leq C_{\beta,d,\eta}$ uniformly in k_n. As shown in Appendix K, this result is guaranteed by (50) as long as $\frac{d}{p} - \frac{d}{2} \leq \beta$. The following proposition allows us to determine different regimes when resolution k_c may increase or decrease when we decrease p. For simplicity, we assume that the exponent γ in the forward operator equals to the exponent β in the $\mathcal{H}^\beta$ of the norm of the noise, and therefore $1 + 2\gamma - \beta = 1 + \gamma$. Let us also assume that $\gamma > -1$, and therefore the bracketed terms will be present according to (47) and (48).

Proposition 6.3. *In the setting described above, if $p \in [1, 2]$,*

$$\delta_{p'} = \|\hat{n}\|_{p',w^2} \asymp k_n^{-\eta-1+d-\frac{d}{p}} \tag{52}$$

and the resolution k_c corresponding to (44) with the hypothesis classes K_s for $s = \infty$ and $K_{\xi,z}$ is bounded as follows

$$k_n^{\frac{1+\eta+\frac{d}{p}-d}{2(1+\gamma+\frac{d}{p}-\frac{d}{2}+r)}} \lesssim_{\xi,z} k_c \lesssim_{M,d} k_n^{\frac{1+\eta+\frac{d}{p}-d}{1+\gamma+\frac{d}{p}-\frac{d}{2}+r}}.$$

If, on the other hand, $p \in (2, \infty)$, then, for $q = p's/(s-1)$ and $s = \frac{2p-2}{p-2}$, we have $q = 2$ for all p. This leads to the following bounds

$$\delta_2 := \|\hat{n}\|_{2,w^2} \asymp k_n^{-\eta-1+\frac{d}{2}} \tag{53}$$

and the resolution k_c corresponding to (44), with the hypothesis classes K_s for the value of s specified above and K_ξ, is bounded as follows:

$$k_n^{\frac{1+\eta-\frac{d}{2}}{1+\gamma+\frac{d}{p}-\frac{d}{2}+r}} \lesssim_{\xi} k_c \lesssim_{M,d} k_n^{\frac{1+\eta-\frac{d}{2}}{1+\gamma+r}}.$$

Note that the sign of

$$\frac{\partial}{\partial p}\Big(\frac{a + \frac{d}{p}}{b + \frac{c}{d}}\Big)$$

is given by the sign $ac - bd$. Accordingly, to determine the relationship between p and increase/decrease of the lower bounds on k_c for all p and the upper bound when $p \in [1, 2]$ we need to consider compare η with

$$t := r + \gamma + \frac{d}{2}.$$

Therefore, for sufficiently large k_n, as p increases,

- If $p \in [1,2]$, and $t < \eta$, then both upper and lower bounds on k_c will increase, and otherwise $t > \eta$, these bounds will decrease.
- If $p \in (2,\infty)$, then the *lower* bound on k_c will increase.
- If $p \in (2,\infty)$ then the *upper* bound on k_c will remain constant.

For these purposes, when $p \in [1,2]$, we set the early stopping threshold to be $\delta_{p'}$ in (52). When $p \in (2,\infty)$, for purposes of the lower bound, the same threshold applies, while for purposes of the upper bound, this threshold is δ_2 in (53).

Generalization of Kantorovich-Rubinstein norm for unbalanced signed measures: In this section we consider reconstruction using $\mathcal{W}_1^{a,b}(f,g) = W_1^{a,b}(f^+ + g^-, f^- + g^+)$, which generalizes Kantorovich-Rubinstein norm.

$$\begin{aligned} &\min_m \mathcal{W}_p^{a,b}(Am, g) \\ &\text{s.t. } Am \in K_i \end{aligned} \tag{54}$$

is subject to the *early stopping condition* $W_p^{a,b}(Am,g) \le \delta$.

First, we observe that our upper and lower bounds on this metric are not conditioned on any lower or upper bounds on f and g, or their norms. For simplicity, we can assume that the forward operator A does not change the zero's frequency, i.e., $\hat{A}_0 = 1$. In this setting, Theorem 6.1, and Remark 6.2 still hold if we define the bandwidth-limited inverse B such that it never cuts off the zero's frequency. Accordingly, it will be possible to bound the resolution of frequencies k_c from above and below if $\alpha > 0$ or $\alpha' > 0$ where α and α' are given by (47) and (48).

If the feasible set is $K_1 = \mathcal{P}(\mathbb{T}^d)$ and the early stopping condition $\delta \le a/b\sqrt{d}$, then

$$k_c^{[1+\gamma]+\frac{d}{2}+r} \lesssim \sqrt{d}\frac{(2\pi)^{-r} r}{([1+\gamma]+\frac{d}{2})b} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{\delta}.$$

On the other hand, if $\delta > a/b\sqrt{d}$, then

$$k_c^{[1+\gamma]+\frac{d}{2}+r} \lesssim \frac{(2\pi)^{-r} r}{[1+\gamma]+\frac{d}{2}} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{a}.$$

In this setting, we have the following lower bound:

$$\frac{(2\pi)^{-r} r}{1+\gamma} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{\delta} \lesssim k_c^{1+\gamma+r}$$

If $g \in \dot{H}^\beta$ for $\beta > \frac{d}{2} - 1$, we also obtain the following lower bound:

$$\frac{(2\pi)^{-r} r}{[1+2\gamma-\beta]+\frac{d}{2}} \cdot \frac{\|m\|_{\dot{\mathcal{H}}^r}}{\delta^{\frac{1}{2}}} \lesssim_z k_c^{[1+2\gamma-\beta]+\frac{d}{2}}.$$

The bracketed terms $1+\gamma$ and $1+2\gamma-\beta$ will appear if $\gamma > -1$ or $\gamma > \beta/2 - 1/2$ respectively, see Remark 6.2. If $1+2\gamma-\delta = 1+\gamma$, then the upper and lower bounds will match up modulo a square root.

6.2 Existing wavelet-based bounds for W_p

A significant body of literature bounding $W_p(f,g)$ by a weighted ℓ_p norm of wavelet or similar multiresolution coefficients of $f-g$. For example, reference [30] embeds a discrete distribution supported on a finite number of points in $\mathbb{R}^d$ in W^1 into sparse vectors in a higher dimensional $\ell^1(\mathbb{R}^{\Delta^d})$ where Δ is the level of the finest grid. That reference uses this approach to develop a fast sparse optimization algorithm to match images.

The wavelet-based bounds that are most relevant for our purposes appear in [48] on the W_p distance given by a weighted ℓ_p norm of the wavelet expansion of $f-g$. They consider the hypercube $\mathbb{H}^d$ in (8) with the Euclidean distance and assume the existence of basis sets Φ and Ψ_j for $j \geq 0$ satisfying the following standard assumptions of a wavelet basis of functions in $L_2(\mathbb{H}^d)$:

1. (Basis) $\Phi \cup \{\cup_{j\geq 0}\Psi_j\}$ form an orthonormal basis for $L_2(\mathbb{H}^d)$;
2. (Regularity) The functions in Φ and Ψ_j for $j \geq 0$ all lie in $C^r(\mathbb{H}^d)$ and polynomials of degree at most r lie in span(Φ).
3. (Tensor construction) Each ψ in Ψ_j can be expressed as $\psi(\mathbf{x}) = \prod_{i=1}^d \psi_i(x_i)$ for some univariate functions ψ_i.
4. (Locality) For each ψ in Ψ_j, there exists a rectangle $I_\psi \subseteq \mathbb{H}^d$ such that $\mathrm{supp}(\psi) \subseteq I_\psi$, $\mathrm{diam}(I_\psi) \lesssim 2^{-j}$, and $\|\sum_{\psi\in\Psi_j} \mathbb{1}\{x \in I_\psi\}\|_\infty \lesssim 1$.
5. (Norm) $\|\psi\|_{L_p(\mathbb{H}^d)} \asymp 2^{dj(\frac{1}{2}-\frac{1}{p})}$ for all $\psi \in \Psi_j$.
6. (Bernstein estimate) $\|\nabla f\|_{L_p(\mathbb{H}^d)} \lesssim 2^j \|f\|_{L_p(\mathbb{H}^d)}$ for any $f \in \mathrm{span}\Big(\Phi \cup \{\cup_{j\geq k\geq 0}\Psi_j\}\Big)$.

See Appendix E in [48] and also [20] and Chapter 2.12 in [19] for additional details regarding this classic wavelet construction.

Reference [48] assumes following wavelet expansions of the probability densities $f, g \in L_p(\Omega)$ for $p \in [1,\infty)$:

$$f = \sum_{\phi\in\Phi} \alpha_\phi \phi + \sum_{j\geq 0}\sum_{\psi\in\Psi_j} \beta_\psi \psi \text{ and } g = \sum_{\phi\in\Phi} \alpha'_\phi \phi + \sum_{j\geq 0}\sum_{\psi\in\Psi_j} \beta'_\psi \psi. \tag{55}$$

where Φ and Ψ are sets of functions satisfying the wavelet assumptions above. The following upper bound holds if constant functions lie in the span of Φ (Assumption 2 holds with $r=0$).

Proposition 6.4 (Prop. 1 in [48]). *For $p \in [1,\infty)$, if for almost every $x \in \Omega$, we have*

$$f(x) \wedge g(x) \geq \xi > 0, \tag{56}$$

then

$$W_p(f,g) \lesssim \xi^{-1/p'} \delta_p^u(f-g)$$

where

$$\delta_p^u(f) := \Big(\|\alpha\|_{\ell_p} + \sum_{j\geq 0} 2^{-j} 2^{dj(\frac{1}{2}-\frac{1}{p})} \|\beta_j\|_{\ell_p} \Big) \tag{57}$$

and the Hölder conjugates p and p' are given by (20) below.

If the wavelets have at least one continuous derivative (Assumption 2 holds with $r = 1$), then the following lower bound holds.

Proposition 6.5 (Prop. 3 in [48]). *For $p \in [1, \infty)$, if for almost every $x \in [0,1]^d$, we have*

$$f(x) \vee g(x) \leq M, \tag{58}$$

then

$$W_p(f, g) \gtrsim M^{-1/p'} \delta_p^l(f - g')$$

where

$$\delta_p^l(f) := \Big(\|\alpha\|_{\ell_p} + \sup_{j\geq 0} \Big\{ 2^{-j} 2^{dj(\frac{1}{2}-\frac{1}{p})} \|\beta_j\|_{\ell_p} \Big\} \Big) \tag{59}$$

6.3 Resolution of wavelets

In this section, we present a wavelet-based resolution analysis that parallels the Fourier-based analysis above. We are interested in solving the problem (39) from the noisy data g_δ when the noise n has the wavelet expansion

$$n = \sum_{\phi\in\Phi} \alpha_\phi \phi + \sum_{j\geq 0} \sum_{\psi\in\Psi_j} \beta_\psi \psi.$$

Now B represents a limited-resolution approximation of A^{-1} given in the wavelet domain by

$$(Bg)_\psi = \begin{cases} (A^{-1}g)_\psi & \text{if } \psi \in \Psi_j \text{ for } j \leq j_c \\ 0 & \text{if } \ \psi \in \Psi_j \text{ for } j > j_c. \end{cases}$$
$$(Bg)_\phi = (A^{-1}g)_\phi \text{ for all } \phi \in \Phi$$

We also let $m_\delta^c := Bg_\delta^c$. We consider m given by (39) in $\mathcal{H}^r(\mathbb{H}^d)$ and assume that our wavelet basis has regularity greater than $|r|$, as specified in Section 6.2. According to Theorem 4, Chapter 3 in [43], given wavelet coefficients α^m and β^m of m, its Sobolev H^r norm is given by

$$\|m\|_{\mathcal{H}^r}^2 \asymp \sum_{\phi\in\Phi} |\alpha_\phi^m|^2 + \sum_{j\geq 0} \sum_{\psi\in\Psi_j} 4^{jr} |\beta_\psi^m|^2. \tag{60}$$

We can decompose

$$m = m_\Phi + m_\Psi$$

into orthogonal vectors in the span of the father wavelets Φ and mother wavelets Ψ:

$$m_\Phi = \sum_{\phi\in\Phi} \alpha_\phi^m \phi \text{ and } m_\Psi = \sum_{j\geq 0}\sum_{\psi\in\Psi_j} \beta_\psi^m \psi.$$

The same decomposition will apply to the $\mathcal{H}^r$ norm of m:

$$\|m\|_{\mathcal{H}^r}^2 \asymp \|\alpha^m\|_2^2 + \|m_\Psi\|_{\mathcal{H}^r}^2. \tag{61}$$

The following theorem bounds the reconstruction error from above in terms of the weighted wavelet-based norm of the noise.

Theorem 6.6. *Under the assumptions on the wavelet basis in Section 6.2, we have*

$$\|m - m_\delta^c\|_{L^2} \lesssim \|m_\Phi\|_2 + 4^{-j_c r}\|m_\Psi\|_{\mathcal{H}^r} + \|B\|_{\delta_p\to L^2}\delta_q(n) \tag{62}$$

where $\delta_p(n)$ represents the weighted ℓ_p norm of wavelet coefficients of the noise n given by (57) or (59), and $\|\cdot\|_{\delta_p\to L^2}$ refers to the operator norm of a map from the normed space of such coefficients to L^2.

De-smoothing inversion (wavelet domain): We assume that $\|B\|_{\delta_p\to L^2} \asymp_{d,p} 4^{j_c h}$ for some $h > 0$, and we shall refer to the corresponding B as a *de-smoothing* inverse operator.[10] Then, if $r > 0$, the upper bound (43) is minimized when

$$4^{j_c(h+r)} \asymp_{d,p} \left(\frac{\|m\|_{\mathcal{H}^r}}{\delta_q}\right).$$

.

On the other hand, if $r \leq 0$, then the optimal $k_c = 0$, i.e., no recovery is possible in the de-smoothing inversion setting. Also, if $\|B\|_{\delta_p\to L^2}$ is upper bounded uniformly in j_c, then for $r \geq 0$ the upper bound (62) is minimized by eliminating the cut-off resolution threshold and reconstructing all scales, while if $r < 0$, the optimal $j_c = 0$.

Diagonal operators in the wavelet domain: Similarly to the Fourier case, in the present case, h may also depend on the exponent p and dimension d, as well as the intrinsic properties of the map A. To illustrate this dependence, we assume the forward operator A is diagonal in the wavelet domain:

$$\begin{cases} (A_\psi m)_\psi = 4^{-\gamma j} m_\psi & \text{if } \psi \in \Psi_j \text{ for all } j \geq 0 \\ (A_\phi m)_\phi = m_\phi & \text{for all } \phi \in \Phi \end{cases}$$

[10] It is possible to perform a similar analysis of $\|B\|_{\delta_p\to L^2} \asymp \log(2^j)$.

for some $\gamma > 0$. Therefore, B is given by

$$\begin{cases} (B_\psi g)_\psi = 4^{\gamma j} g_\psi & \text{if } \psi \in \Psi_j \text{ for all} j \leq j_c \\ (B_\psi g)_\psi = 0 & \text{if } \psi \in \Psi_j \text{ for all} j > j_c \\ (B_\phi g)_\phi = g_\phi & \text{for all } \phi \in \Phi \end{cases}$$

According to Lemma 1 in [48], the set Φ has finite cardinality $c_\Phi := |\Phi| \lesssim 1$. Also, by that Lemma the cardinality of $\Psi_j \lesssim 2^{dj}$. By (76), the operator norm from δ_p^u given (57) space to L_2 is computed as follows.

$$\begin{aligned} \|Bn\|_{L^2}^2 &= \sum_{\phi\in\Phi} \alpha_\phi^2 + \sum_{j_c \geq j \geq 0} 4^{j(\gamma+1)} 4^{-j} \sum_{\psi\in\Psi_j} \beta_\psi^2 \\ &\leq (c_\Phi^{\frac{1}{2}-\frac{1}{p}} \|\alpha\|_p)^2 + \sum_{j_c\geq j\geq 0} 4^{j(\gamma+1)} \left(2^{-j} 2^{dj(\frac{1}{2}-\frac{1}{p})} \|\beta_j\|_p\right)^2 \\ &\leq (j_c+1)^{-1} \left(c_\Phi^{\frac{1}{2}-\frac{1}{p}} \|\alpha\|_p + \sum_{j_c\geq j\geq 0} 2^{j(\gamma+1)} 2^{-j} 2^{dj(\frac{1}{2}-\frac{1}{p})} \|\beta_j\|_p \right)^2 \\ &= (j_c+1)^{-1} C_w^2 (\delta_p^u(n))^2 \end{aligned} \tag{63}$$

where

$$C_w = \begin{cases} \max(c_\Phi^{\frac{1}{2}-\frac{1}{p}}, 2^{j_c(\gamma+1)}) & \text{if } \gamma \geq -1 \\ \max(c_\Phi^{\frac{1}{2}-\frac{1}{p}}, 1) & \text{if } \gamma < -1 \end{cases} \tag{64}$$

On the other hand, the operator norm from δ_p^l given (59) space to L_2 is computed as follows.

$$\begin{aligned} \|Bn\|_{L^2}^2 &= \sum_{\phi\in\Phi} \alpha_\phi^2 + \sum_{j_c \geq j \geq 0} 4^{j(\gamma+1)} 4^{-j} \sum_{\psi\in\Psi_j} \beta_\psi^2 \\ &\leq (c_\Phi^{\frac{1}{2}-\frac{1}{p}} \|\alpha\|_p)^2 + \sup_{j_c\geq j\geq 0} \left\{ 4^{j(\gamma+1)} \left(2^{-j} 2^{dj(\frac{1}{2}-\frac{1}{p})} \|\beta_j\|_p\right)^2 \right\} \\ &\leq \frac{1}{2} \left(c_\Phi^{\frac{1}{2}-\frac{1}{p}} \|\alpha\|_p + \sup_{j_c\geq j\geq 0} \left\{ 4^{j(\gamma+1)} \left(2^{-j} 2^{dj(\frac{1}{2}-\frac{1}{p})}\right)^2 \right\} \right)^{\frac{1}{2}} \\ &= \frac{1}{2} C_w^2 (\delta_p^l(n))^2 \end{aligned} \tag{65}$$

Accordingly, if $\gamma \geq -1$ for j_c enough, the reconstruction error bound is minimized when the cutoff scale j_c is bounded as follows

$$\frac{(j_c+1)^{\frac{1}{2}}}{C_w} \cdot \frac{\|m\|_{\mathcal{H}_\Psi^r}}{\delta_q^u} \lesssim 4_{j_c}^{j_c(1+\gamma+r)} \lesssim \frac{1}{C_w} \cdot \frac{\|m\|_{\mathcal{H}_\Psi^r}}{\delta_q}.$$

This indicates that, in contrast to the resolution of frequencies in computational inversion using the W_p metric, the resolution of wavelets in such an inversion does not exhibit a complicated dependence on p and the dimension d.

6.4 Relationship to existing work

When $p = 2$, the upper bound (43) is minimized when

$$k_c \asymp \left(\delta_2^{-1}\|m\|_{\mathcal{H}^r}\right)^{\frac{1}{1+r+\gamma}}.$$

This resolution matches the optimal resolution in [24] for inverse matching using the $\mathcal{H}^{-1}$ norm. More generally, this reference showed that in the context of using the $\mathcal{H}^s$ norm

$$k_c \asymp \left(\delta_s^{-1}\|m\|_{\mathcal{H}^r}\right)^{\frac{1}{1+r+\gamma-s}}.$$

where δ_s is the early stopping threshold (corresponding to $\mathcal{H}^s$ norm of the "noise" representing the difference between the Am_δ at early stopping and the data g). Taking $s = 0$, corresponds to the L_2 norm matching, leading to lower resolution (smoother reconstruction) than the W_p metric minimization; on the other hand, using the $\mathcal{H}^s$ with the negative $s = \frac{d}{p'} - \frac{d}{2} = \frac{d}{2} - \frac{d}{p}$ leads to the same resolution in $\mathcal{H}^s$ and W_p matching for $p \in [1, 2]$ (in each case, assuming that the early stopping thresholds are the same: $\delta_s = \delta_{p'}$). However, the resolution analysis of W_2 was based on its equivalence with the weighted homogeneous Sobolev norm $\mathcal{H}^{-1}$. The latter norms entail mixing between different frequencies of the signal (66), which somewhat complicates the resolution analysis. Since our Fourier-based bounds do not entail such mixing, they lead to a more straightforward analysis.

7 Conclusion and Future Directions

The present work develops Fourier-based norm bounds on the W_p distance and makes progress towards understanding the effects of using the W_p metric and its generalizations in the context of computational inverse problems by applying these bounds to determine the optimal resolution of frequencies in the context of solving such problems using this metric.

In addition to the resolution analysis, previous work [21, 24] analyzed the convexity of the objective function, as well as the regularity of iterative solutions, in computational inversion using the W_2 mismatch functional. Accordingly, a potential future direction would be to generalize this convexity and regularity analysis to other values of p. As the Fourier-based bounds do not entail mixing between different frequencies, our results may also lead to provable frequency matching algorithms for W_p minimization in the context of computational inverse problems and other settings; we leave this intriguing research directions to future work. Moreover, since the W_p distance is used for a broad range of problems in mathematics and computational sciences, we expect that our Fourier-based norm bounds will be of interest beyond inverse problems.

Acknowledgments

We would like to thank the anonymous referees for their useful comments. This work is partially supported by the National Science Foundation through grants DMS-1937254 and

DMS-2309802. V.A.K. is grateful to Robert Kohn for his fruitful suggestion to explore wavelet analysis in the context of the present problem. V.A.K. thanks Sinan Gunturk, James Scott and Stefan Steinerberger for helpful input at critical points in this project.

References

[1] I. Abraham, R. Abraham, M. Bergounioux, and G. Carlier, *Tomographic reconstruction from a few views: A multi-marginal optimal transport approach*, Applied Mathematics & Optimization, 75 (2017), pp. 55–73.

[2] J. Adler, A. Ringh, O. Öktem, and J. Karlsson, *Learning to solve inverse problems using wasserstein loss*, arXiv:1710.10898, (2017).

[3] L. Ambrosio, E. Mainini, and S. Serfaty, *Gradient flow of the Chapman- Rubinstein-Schatzman model for signed vortices*, Ann. Inst. H. Poincare Anal. Non Lineaire, 28 (2011), pp. 217–246.

[4] M. Arjovsky, S. Chintala, and L. Bottou, *Wasserstein generative adversarial networks*, in Proceedings of the 34th International Conference on Machine Learning, D. Precup and Y. W. Teh, eds., vol. 70 of Proceedings of Machine Learning Research, PMLR, 06–11 Aug 2017, pp. 214–223.

[5] G. Auricchio, A. CodeGoni, S. Gualandi, G. Toscani, and M. Veneroni, *The equivalence of Fourier-based and Wasserstein metrics on imaging problems*, arXiv:2005.06530v1, (2020).

[6] G. Bal and K. Ren, *Physics-based models for measurement correlations. application to an inverse Sturm-Liouville problem*, Inverse Problems, 25 (2009). 055006.

[7] M. Ballu, Q. Berthet, and F. Bach, *Stochastic optimization for regularized wasserstein estimators*, in Proceedings of the 37th International Conference on Machine Learning, A. S. Hal Daumé III, ed., vol. 119 of Proceedings of Machine Learning Research, PMLR, 2020, pp. 602–612.

[8] ——, *Stochastic optimization for regularized wasserstein estimators*, arXiv:2002.08695v1, (2020).

[9] F. Bassetti, A. Bodini, and E. Regazzini, *On minimum kantorovich distance estimators*, Statistics & Probability Letters, 76 (2006), pp. 1298–1302.

[10] M. Beckmann, *A continuous model of transportation*, Econometrica, 20 (1952), pp. 643–660.

[11] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the monge-kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[12] E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert, *Approximate bayesian computation with the wasserstein distance.*, Journal of the Royal Statistical Society: Series B, 81(2) (2019), pp. 235–269.

[13] E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert, *Approximate Bayesian computation with the Wasserstein distance*, arXiv:1905.03747, (2019).

[14] E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert, *On parameter estimation with the Wasserstein distance*, Information and Inference: A Journal of the IMA, 8 (2019), pp. 657–676.

[15] N. Bonnotte, *Unidimensional and Evolution Methods for Optimal Transportation*, PhD thesis, Université Paris Sud - Paris XI; Scuola normale superiore (Pise, Italie), 2013.

[16] Y. Brenier, *Extended Monge-Kantorovich theory*, in Optimal transportation and applications, M. Franca, ed., vol. 1813 of Lecture Notes in Math., Springer, Berlin, 2003, pp. 91–121.

[17] D. Chafa'́i and F. Malrieu, *On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities*, Annales de l'I.H.P., 46 (2010), pp. 72–96.

[18] J. Chen, Y. Chen, H. Wu, and D. Yang, *The quadratic Wasserstein metric for earthquake location*, J. Comput. Phys., 373 (2018), pp. 188–209.

[19] A. Cohen, *Numerical analysis of wavelet methods*, volume 32 of Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam, 2003.

[20] A. Cohen, I. Daubechies, and P. Vial, *Wavelets on the interval and fast wavelet transforms*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 54–81.

[21] W. Ding, *Computational Inversion with Wasserstein Distances and Neural Network Induced Loss Functions*, PhD thesis, Columbia University, 2022.

[22] B. Engquist and B. D. Froese, *Application of the Wasserstein metric to seismic signals*, Commun. Math. Sci., 12 (2014), pp. 979–988.

[23] B. Engquist, B. D. Froese, and Y. Yang, *Optimal transport for seismic full waveform inversion*, Commun. Math. Sci., 14 (2016), pp. 2309–2330.

[24] B. Engquist, K. Ren, and Y. Yang, *The quadratic Wasserstein metric for inverse data matching*, Inverse Problems, 36 (2020), p. 055001. arXiv:1911.06911.

[25] B. Engquist and Y. Yang, *Optimal transport based seismic inversion: Beyond cycle skipping*, Commun. Pure Appl. Math., 75 (2022), pp. 2201–2244.

[26] G. B. Folland, *A Course on Abstract Harmonic Analysis*, CRC Press, Inc., 1995.

[27] J. González-Delgado, A. González-Sanz, J. Cortés, and P. Neuvial, *Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology*, Electronic Journal of Statistics, 17 (2023), pp. 1547 – 1586.

[28] L. G. Hanin, *Kantorovich-rubinstein norm and its application in the theory of lipschitz spaces*, Proceedings of the American Mathematical Society, 115 (1992), pp. 345–352.

[29] H. Heaton, S. W. Fung, A. T. Lin, S. Osher, and W. Yin, *Wasserstein-based projections with applications to inverse problems*, SIAM Journal on Mathematics of Data Science, 4 (2022), pp. 581–603.

[30] P. Indyk and N. Thaper, *Fast image retrieval via embeddings*, 3rd Intl. Workshop on Statistical and Computational Theories of Vision, 2 (2003).

[31] V. Isakov, *Inverse Problems for Partial Differential Equations*, Springer-Verlag, New York, second ed., 2006.

[32] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, *Optimal mass transport: signal processing and machine-learning applications*, IEEE Signal Processing Magazine, 34 (2017), pp. 43–59.

[33] W. Lee, R. Lai, W. Li, and S. Osher, *Generalized unnormalized optimal transport and its fast algorithms*, J. Comput. Phys., 436 (2021). 110041.

[34] J. Lellmann, D. Lorenz, C. Schönlieb, and T. Valkonen, *Imaging with Kantorovich-Rubinstein discrepancy*, SIAM J. Imag. Sci., 7 (2014), pp. 2833–2859.

[35] Q. Liao, J. Chen, Z. Wang, B. Bai, S. Jin, and H. Wu, *Fast Sinkhorn I: An O(N) algorithm for the Wasserstein-1 metric*, Commun. Math. Sci., 20:7 (2022), p. 2053–2067.

[36] Q. Liao, Z. Wang, J. Chen, B. Bai, S. Jin, and H. Wu, *Fast Sinkhorn II: Collinear triangular matrix and linear time accurate computation of optimal transport*, 2022.

[37] J. Liu, W. Yin, W. Li, and Y.-T. Chow, *Multilevel optimal transport: a fast approximation of Wasserstein-1 distance*, SIAM J. Sci. Comput., 43 (2021), pp. A193–A220.

[38] G. Loeper, *Uniqueness of the solution to the Vlasov-Poisson system with bounded density*, J. Math. Pures Appl., 86 (2006), pp. 68–79.

[39] T. Manole, S. Balakrishnan, J. Niles-Weed, and L. Wasserman, *Plugin estimation of smooth optimal transport maps*, 2022.

[40] B. Maury, A. Roudneff-Chupin, and F. Santambrogio, *A macroscopic crowd motion model of gradient flow type*, Math. Models Methods Appl. Sci., 10 (2010), pp. 1787—-1821.

[41] L. Métivier, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, *Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion*, Geophys. J. Int., 205 (2016), pp. 345–377.

[42] ——, *An optimal transport approach for seismic tomography: application to 3D full waveform inversion*, Inverse Probl., 32 (2016). 115008.

[43] Y. Meyer and D. Salinger, *Wavelets and Operators*, Cambridge University Press, 2011.

[44] L. Métivier, R. Brossier, F. Kpadonou, J. Messud, and A. Pladys, *A review of the use of optimal transport distances for high resolution seismic imaging based on the full waveform*, 2022.

[45] K. Nadjahi, V. De Bortoli, A. Durmus, R. Badeau, and U. Simsekli, *Approximate bayesian computation with the sliced-Wasserstein distance*, arXiv:1910.12815, (2019).

[46] S. Nietert, R. Cummings, and Z. Goldfeld, *Robust estimation under the wasserstein distance.*, 2023.

[47] S. Nietert, Z. Goldfeld, and R. Cummings, *Outlier-robust optimal transport: Duality, structure, and statistical analysis*, in Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, G. Camps-Valls, F. J. R. Ruiz, and I. Valera, eds., vol. 151 of Proceedings of Machine Learning Research, PMLR, 28–30 Mar 2022, pp. 11691–11719.

[48] J. Niles-Weed and Q. Berthet, *Minimax estimation of smooth densities in Wasserstein distance*, Annals of Statistics, (2022), pp. 1519–1540.

[49] R. Peyre, *Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of wasserstein distance*, ESAIM: COCV, 24 (2018), pp. 1489–1501.

[50] B. Piccoli and F. Rossi, *On properties of the generalized Wasserstein distance*, Archive for Rational Mechanics and Analysis, 222 (2000), pp. 1339– 1365.

[51] B. Piccoli, F. Rossi, and M. Tournus, *A Wasserstein norm for signed measures, with application to non-local transport equation with source term*, 2023.

[52] E. K. Ryu, W. Li, P. Yin, and S. Osher, *Unbalanced and partial l1 Monge-Kantorovich problem: A scalable parallel first-order method*, J. Sci. Comput., 75 (2018), pp. 1596–1613.

[53] F. Santambrogio, *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, Birkhäuser, 2015.

[54] B. Schmitzer and B. Wirth, *Dynamic models of Wasserstein-1-type unbalanced transport*, ESAIM: COCV, 25 (2019). 23.

[55] C. Scricciolo, *Bayes and maximum likelihood for L1-Wasserstein deconvolution of Laplace mixtures*, Statistical Methods & Applications, 27 (2018), pp. 333–362.

[56] J. Solomon, *Optimal transport on discrete domains*, in Proceedings of Symposia in Applied Mathematics, vol. 76, 2020, p. 103.

[57] S. Steinerberger, *Wasserstein distance, Fourier series and applications*, Monatshefte für Mathematik, 194 (2021), pp. 305–338.

[58] C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, Providence, RI., 2003.

[59] ——, *Optimal Transport: Old and New*, Springer Science & Business Media, 2008.

[60] P. Yong, W. Liao, J. Huang, Z. Li, and Y. Lin, *Misfit function for full waveform inversion based on the wasserstein metric with dynamic formulation*, Journal of Computational Physics, 399 (2019), p. 108911.

A Existing Fourier-based bounds for W_p

W_p on the circle: In the case of W_p on the circle $\mathbb{T}$ for $1 \leq p < \infty$, [57] showed that

$$W_p(f,g) \lesssim_p \Big(\sum_{k=1}^{\infty} \frac{|\hat{f}(k) - \hat{g}(k)|^2}{k^{2p-2}} \Big)^{\frac{1}{2p}}.$$

When g is a Lebesgue measure λ, [57] proved the following lower bound,

$$W_1(f,\lambda) \gtrsim M^{-1} \sum_{\substack{k\in\mathbb{Z} \\ k\neq 0}}^{\infty} \frac{1+\log|k|}{k^2} |\hat{f}(k)|^2$$

where $M = \|f\|_\infty$.

W_2 **on** $\mathbb{R}^d$**:** Given a positive measure μ on M and $F : M \to \mathbb{R}$, the weighted Sobolev $\dot{\mathcal{H}}^s(\mu)$ semi-norm of F is given by

$$\|F\|^2_{\dot{\mathcal{H}}^s(\mu)} := \int_M |\xi|^{2s} |\hat{F}(\xi)|^2 d\mu(\xi). \tag{66}$$

We will omit the parenthesis in $\dot{\mathcal{H}}^s(\mu)$ if μ is the *Lebesgue measure, which we will denote by* λ. If μ is absolutely continuous with respect to λ, then it can be associated with a density g. If $\omega = 1/\sqrt{g}$ has a Fourier expansion, then the preceding norm becomes

$$\|F\|^2_{\dot{\mathcal{H}}^s(\mu)} := \int_M |\xi|^{2s} |\hat{F} \circledast \hat{\omega}(\xi)|^2 d\xi. \tag{67}$$

When μ is close to ν, referred to as *linearized or asymptotic*, in the case of transport on $\mathbb{R}^n$, W_2 is equivalent to the weighted homogeneous Sobolev $\dot{\mathcal{H}}^{-1}(\mu)$ norm (e.g., Theorem 7.26 in [58]). Based on this equivalence, reference [49] established

$$\frac{1}{c}\|\mu - \nu\|_{\dot{\mathcal{H}}^{-1}} \leq W_2(\mu, \nu) \leq 2\|\mu - \nu\|_{\dot{\mathcal{H}}^{-1}(\mu)} \tag{68}$$

where the lower bound holds only if $\mu \leq \rho_0\lambda$ and $\nu \leq \rho_1\lambda$ with

$$c = \frac{2(\rho_0^{\frac{1}{2}} - \rho_1^{\frac{1}{2}})}{\ln(\rho_0/\rho_1)}$$

(For $\rho_0 = \rho_1$, we take $c = \rho_0^{\frac{1}{2}}$ by continuity.) Note that in the lower bound we have an unweighted Sobolev norm while in the upper bound we have a weighted one. The constant c is bounded from above by the prefactor $\max(\rho_0, \rho_1)^{\frac{1}{2}}$ that appears in a similar lower bound in [38]; the two lower bounds coalesce when $\rho_0 = \rho_1$. If the probability measures μ and ν are absolutely continuous with respect to the Lebesgue measure and have densities f and g, which have Fourier transforms, then

$$\|f - g\|^2_{\dot{\mathcal{H}}^{-1}(\nu)} = \int_{\mathbb{R}^n} \left| |\xi|^{-1}((\hat{f} - \hat{g}) \circledast \hat{\omega}(\xi)) \right|^2 d\xi. \tag{69}$$

where $\omega = 1/\sqrt{g}$. Reference [24] used the relationship between $\dot{\mathcal{H}}^{-1}$ and W_2 to analyze the frequency content of W_2 inverse matching in the Fourier domain in the asymptotic regime. In this setting, it observed that the weighting leads to mixing between different modes of g in the Fourier domain, which prevents matching mode by mode.

W_1 **and** W_2 **on a finite grid:** Reference [5] analysed the equivalence of a Fourier-based metric and W_p for $p = 1$ and 2 for discrete measures supported on a finite grid. Specifically in the case of a discrete measure μ supported on a regular grid G_N (of N^d points) in $[0, 1)^d$ given by

$$G_N := \{x \in \mathbb{R}^d : Nx \in \mathbb{Z}^d \cap [0, N)^d\}$$

its discrete Fourier transform

$$\hat{\mu}(k) = \sum_{x \in G_n} \mu_x e^{-i\langle k, x\rangle}$$

is $2\pi N$-periodic in each coordinate of x, and therefore it suffices to consider $k \in [0, 2\pi N]^d$. Letting $T = 2\pi N$, this reference showed the equivalence of the metrics $f_{1,2}$ and $\mathcal{F}_{2,2}$ based on the Fourier transform $\hat{\mu}$, as defined below, and W_p for $p = 1$ and 2:

$$f_{1,2}(\mu, \nu) \leq W_1(\mu, \nu) \leq \frac{T^2}{2\pi} f_{1,2}(\mu, \nu) \text{ and } \frac{1}{2\sqrt{2}} \mathcal{F}_{2,2}(\mu, \nu) \leq W_2(\mu, \nu) \leq \frac{T^3}{\pi} \mathcal{F}_{2,2}(\mu, \nu)$$

where

$$f_{p,2}(\mu, \nu) = \Big(\frac{1}{|T|^d} \int_{[0,T]^d} \frac{|\hat{\mu}(k) - \hat{\nu}(k)|^2}{|k|^{2p}} dk\Big)^{\frac{1}{2}}, \ \mathcal{F}_{2,2}(\mu, \nu) = \sqrt{f_{p,2}(\mu, \nu_{m_\mu - m_\nu}) + |m_\mu - m_\nu|^2}$$

and $\nu_{m_\mu - m_\nu}$ is the translation of ν by $m_\mu - m_\nu$ so that μ and $\nu_{m_\mu - m_\nu}$ have the same center of mass given by (11).

B Proof of Theorem 4.5

Proof. Let h be a function given by

$$h = \sum_k \lambda_k \psi_k$$

for some sequence of Fourier coefficients λ that we will determine later. Since Assumption 3 (Locality) is not satisfied with respect to Fourier basis function, Lemma E.2 and consequently Lemma 7 in [48] would not generalize to the present setting. However, the following bound holds instead.

Lemma B.1. *If $q \in (2, \infty]$, q' is its conjugate exponent given by (20), and*

$$d^{\frac{1}{2} - \frac{1}{q}} \|\lambda\|_{q', w^{q'}} \leq 1 \tag{70}$$

where the weights $w^{q'}$ are given by (2), then $\|\nabla h\|_{L_q} \leq 1$.

Proof. We have

$$
\begin{aligned}
\|\nabla h\|_{L_q}^q &= \Big\| \sum_k \lambda_k \nabla \psi_k \Big\|_{L_q}^q \\
&= \int_\Omega \Big(\sum_{j=1}^d \Big| \sum_k 2\pi i \lambda_k k_j \psi_k(x) \Big|^2 \Big)^{\frac{q}{2}} dx \\
&\le d^{\frac{q}{2}-1} \int_\Omega \sum_{j=1}^d \Big| \sum_k 2\pi i \lambda_k k_j \psi_k(x) \Big|^q dx \\
&\le d^{\frac{q}{2}-1} \sum_{j=1}^d \Big(\sum_k |2\pi \lambda_k k_j|^{q'} \Big)^{\frac{q}{q'}} \\
&\le d^{\frac{q}{2}-1} \Big(\sum_{j=1}^d \sum_k |2\pi \lambda_k k_j|^{q'} \Big)^{\frac{q}{q'}} \\
&= d^{\frac{q}{2}-1} \Big(\sum_k (2\pi \|k\|_{q'})^{q'} |\lambda_k|^{q'} \Big)^{\frac{q}{q'}} \\
&\le 1
\end{aligned}
$$

Specifically, the first inequality follows from the fact that $\|\cdot\|_2 \le d^{\frac{1}{2}-\frac{1}{q}} \|\cdot\|_q$. The second inequality follows from the Hausdorff-Young inequality (21); the third inequality follows from the fact that $1 \le \frac{q}{q'}$, and the last inequality follows from the weighted norm bound (70). ∎

Assuming (70) holds, by Lemma 4.4 and the preceding lemma,

$$W_p(f, g) \ge M^{-1/p'} \langle \lambda, \hat{f} - \hat{g} \rangle \tag{72}$$

We now optimize λ to maximize this lower bound subject to the constraint (70). Note that the Hölder's inequality holds with equality when

$$|\langle a, b \rangle| = \|a\|^{q'} \|b\|^q$$

which holds when $|a_k|^{q'} = \alpha |b_k|^q$. Accordingly, for the weights $w_k^{q'} = 1/(2\pi \|k\|_{q'})$, the value λ guaranteeing that

$$\langle \lambda, (\hat{f} - \hat{g}) \rangle = \sum_k \lambda_k (\hat{f}_k - \hat{g}_k) = \sum_k \lambda_k / w_k (\hat{f}_k - \hat{g}_k) w_k \tag{73}$$

is equal to

$$\|\lambda\|_{q', 1/w^{q'}} \|\lambda\|_{q, w^{q'}},$$

is given by

$$\lambda_k = w_k^{q'} \alpha \frac{(\hat{f}_k - \hat{g}_k)^*}{|\hat{f}_k - \hat{g}_k|} \Big(w_k^{q'} |\hat{f}_k - \hat{g}_k| \Big)^{\frac{q}{q'}} \tag{74}$$

for some $\alpha > 0$ that we will determine next (the superscript * denotes the complex conjugate). Also observe that $\frac{q}{q'} = q - 1$. Accordingly,

$$\|\lambda\|_{q',w^{q'}} \leq \alpha \Big\| \Big(|\hat{f} - \hat{g}| w^{q'} \Big)^{\frac{q}{q'}} \Big\|_{q'} = \alpha \|\hat{f} - \hat{g}\|_{q,w^{q'}}^{\frac{q}{q'}} = \alpha \|\hat{f} - \hat{g}\|_{q,w^{q'}}^{q-1} \tag{75}$$

Now, setting

$$\alpha = 1 \Big/ (d^{\frac{1}{2}-\frac{1}{q}} \|\hat{f} - \hat{g}\|_{q,w^{q'}}^{q-1})$$

guarantees (70). Now evaluating

$$\langle \lambda, (\hat{f} - \hat{g}) \rangle = \alpha \sum_k \frac{|\hat{f}_k - \hat{g}_k|}{w_k} \Big(\frac{|\hat{f}_k - \hat{g}_k|}{w_k} \Big)^{q-1} = \alpha \|\hat{f}_k - \hat{g}_k\|_{q,v}^q = d^{\frac{1}{q}-\frac{1}{2}} \|\hat{f}_k - \hat{g}_k\|_{q,v^{q'}}$$

completes the proof of Theorem 4.5. $\square$

C Proof of Lemma 4.6

When $p \in [1, 2]$, we use Hölder's inequality and Parseval's identity:

$$\begin{aligned}
\|V\|_{L_p}^p &= \int_{\mathbb{T}^d} \|V(x)\|_2^p dx \\
&= \int_{\mathbb{T}^d} \Big(\sum_j \Big(\sum_k (\hat{f}_k - \hat{g}_k)(V_k)_j(x) \Big)^2 \Big)^{\frac{p}{2}} dx \\
&= \int_{\mathbb{T}^d} \Big(\sum_j \Big(\sum_k \frac{k_j}{i2\pi \|k\|_2^2} (\hat{f}_k - \hat{g}_k) \psi_k(x) \Big)^2 \Big)^{\frac{p}{2}} dx \\
&\leq \Big(\int_{\mathbb{T}^d} dx \Big)^{\frac{2}{2-p}} \Big(\int_{\mathbb{T}^d} \sum_j \Big(\sum_k \frac{k_j}{i2\pi \|k\|_2^2} (\hat{f}_k - \hat{g}_k) \psi_k(x) \Big)^2 dx \Big)^{\frac{p}{2}} \\
&= \Big(\sum_j \sum_k \frac{k_j^2}{i2\pi \|k\|_2^4} (\hat{f}_k - \hat{g}_k)^2 \Big)^{\frac{p}{2}} \\
&= \|\hat{f}_k - \hat{g}_k\|_{2,w^2}^p
\end{aligned}$$

where the sequence of weights w^2 is given by (2) and the inequality follows from the application of Hölder's inequality with conjugate exponents $2/p$ and $2/(2-p)$; afterwards we use Parseval's identity and the fact that the volume of $\mathbb{T}^d$ is 1.

When $p \in (2, \infty)$,we use the Hausdorff-Young inequality. Taking $c_{j,k} = k^p / \|k\|_p^p$ in the definition of V_k, we have

$$\begin{aligned}
\|V\|_{L_p}^p &\le d^{\frac{p}{2}-1}\int_{\mathbb{T}^d}\|V(x)\|_p^p dx \\
&= d^{\frac{p}{2}-1}\sum_j \int_{\mathbb{T}^d}\Big(\sum_k \frac{k_j^{p-1}}{i2\pi\|k\|_p^p}(\hat f_k-\hat g_k)\psi_k(x)\Big)^p dx \\
&\le d^{\frac{p}{2}-1}\sum_j\sum_k \frac{k_j^{(p-1)p'}}{i2\pi\|k\|_2^{pp'}}(\hat f_k-\hat g_k)^{p'} \\
&= d^{\frac{p}{2}-1}\sum_k \frac{1}{i2\pi\|k\|_2^{pp'-p}}(\hat f_k-\hat g_k)^{p'} \\
&= d^{\frac{p}{2}-1}\|\hat f_k-\hat g_k\|_{p',w^{p'}}^p
\end{aligned}$$

The first inequality above follows from the standard result for vectors in $\mathbb{R}^d$,

$$\|x\|_p \le d^{\frac{1}{p}-\frac{1}{q}}\|x\|_q. \tag{76}$$

The second inequality follows from the application of Hausdorff-Young inequality and the sequence of weights $w^{p'}$ is again given by (2).

D Proof of Proposition 4.7

Proof. When $p>1$, following [48], we use the fluid-dynamics characterization of W^p in (19). If E and ρ are absolutely continuous with respect to the Lebesgue measure on $\Omega\times[0,1]$, we identify them with their densities as follows. We set

$$\rho(x,t) = (1-\lambda(t))f(x)+\lambda(t)g(x)$$

where

$$\lambda(t) = \begin{cases}\frac{1}{2}(2t)^p & \text{if } t\le 1/2\\ 1-\frac{1}{2}(2-2t)^p & \text{if } t>1/2\end{cases}$$

Then the lower bound (29) leads to

$$\rho(x,t) \ge \begin{cases}\frac{1}{2}(2t)^p\xi & \text{if } t<1/2\\ \frac{1}{2}(2-2t)^p\xi & \text{if } t>1/2\end{cases}$$

a.e. on Ω. Moreover, since

$$\lambda'(t) = \begin{cases} p(2t)^{p-1} & \text{if } t<1/2 \text{ and } p>1\\ p(2-2t)^{p-1} & \text{if } t>1/2 \text{ and } p>1\\ 1 & \text{if } p=1\end{cases}$$

we have

$$\frac{\lambda'(t)^p}{\rho(x,t)^{p-1}} \leq \begin{cases} 2^{p-1}p^p\xi^{1-p} & \text{if } t \neq \frac{1}{2} \text{ and } p > 1 \\ 1 & \text{if } t \neq \frac{1}{2} \text{ and } p = 1 \end{cases}$$

Also similarly to [48], we let

$$E(x,t) = \lambda'(t)V(x) \text{ for } t \in [0,1] \setminus \{1/2\}$$

It can be verified by differentiation that the pair of ρ and E defined above satisfies the PDE in (19). Therefore, since our choices of E and ρ are feasible for the optimization problem and (19):

$$\begin{aligned} W_p(f,g) &\leq \Big(\int_{\Omega\times[0,1]} \Big\| \frac{dE}{d\rho}(x,t) \Big\|^p d\rho(x,t) \Big)^{\frac{1}{p}} \leq \Big(\int_{\Omega\times[0,1]} \|V(x)\|^p \frac{\lambda'(t)^p}{\rho(x,t)^{p-1}} \Big)^{\frac{1}{p}} \\ &\leq 2^{1/p-1}p\xi^{1/p-1}\|V\|_p = 2^{-\frac{1}{p'}}p\xi^{-1/p'}\|V\|_p \end{aligned}$$

and the result follows from Lemma 4.6.

When $p = 1$, the divergence formulation provides

$$W_1(f,g) \leq \|V\|_1$$

since V is a feasible point for (10). $\square$

E Proof of Proposition 4.8

Proof. We will use the generalized Hölder's inequality in the ℓ_p spaces and

$$\sum_i |a_i b_i c_i| \leq \|a\|_p \|b\|_q \|c\|_r$$

for $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, and the Fourier representation of the homogeneous Sobolev norm (9).

$$\|\hat{f} - \hat{g}\|_{2,w}^2 = (w_0(\hat{f} - \hat{g}))^2 + \sum_k \frac{1}{(2\pi\|k\|_2)^2}(\hat{f}_k - \hat{g}_k)^2 \tag{77}$$

$$= (w_0(\hat{f}_0 - \hat{g}_0))^2 + \sum_k \frac{1}{(2\pi\|k\|_2)^{\beta+1}} \frac{(\hat{f}_k - \hat{g}_k)}{2\pi\|k\|_2}(2\pi\|k\|_2)^\beta(\hat{f}_k - \hat{g}_k) \tag{78}$$

$$\leq (w_0(\hat{f}_0 - \hat{g}_0))^2 + C_{p,d,\beta}^2\|\hat{f}_0 - \hat{g}_0\|_{p',w^2}\|f - g\|_{\dot{\mathcal{H}}^s} \tag{79}$$

where

$$C_{p,d,\beta}^2 = \Big(\sum_k \frac{1}{(2\pi\|k\|_2)^{(\beta+1)t}} \Big)^{\frac{1}{t}}$$

and

$$\frac{1}{t} + \frac{1}{p'} + \frac{1}{2} = 1$$

leading to $t = 2p/(2-p)$. Since the summand above is a decreasing function of $\|k\|$ (whenever the sum converges), we can bound $C_{p,d,\beta}$ by an integral with a radially symmetric integrand plus constant discretization error:

$$1 \leq C_{p,d,\beta} \leq \Big(\omega_{d-1} \int_1^\infty \frac{1}{r^{(\beta+1)t}} r^{d-1} dr + O(1)\Big)^{\frac{1}{2t}} \leq \Big(\omega_{d-1} \int_1^\infty \frac{1}{r^{(\beta+1)t}} r^{d-1} dr + O(1)\Big)^{\frac{1}{2}}$$

where ω_{d-1} is the surface area of the $d-1$ unit sphere. If

$$\beta > \frac{d}{p} - \frac{d}{2} - 1,$$

this integral converges and is equal to 1. Accordingly, $1 \leq C_{p,d,\beta} \leq \Big(\omega_{d-1} + O(1)\Big)^{\frac{1}{2t}} = O(1)$. □

F Proof of Lemma 4.9

Let $\tilde{V}$ denote the vector field V with the j component evaluated on $x_j = 0$

$$\tilde{V} := \sum_k (\hat{f}_k - \hat{g}_k)\tilde{V}_k.$$

where V_k is as defined in (33). Similarly to the proof of Lemma 4.6 in Appendix C, for $p \in [1,2]$,

$$\begin{aligned}
\|\tilde{V}\|_{L_p}^p &= \int_{\mathbb{T}^d} \|\tilde{V}(x)\|_2^p dx \\
&= \int_{\mathbb{T}^d} \Big(\sum_j \Big(\sum_{k^j} \Big(\sum_{k_j : k_j = |k|_\infty} \frac{1}{i2\pi k_j} (\hat{f}_k - \hat{g}_k) \Big) \prod_{m \neq j} \psi_{k_m}(x_m) \Big)^2 \Big)^{\frac{p}{2}} dx \\
&\leq \Big(\int_{\mathbb{T}^d} dx \Big)^{\frac{2}{2-p}} \Big(\int_{\mathbb{T}^d} \sum_j \Big(\sum_{k^j} \Big(\sum_{k_j : k_j = |k|_\infty} \frac{1}{i2\pi k_j} (\hat{f}_{k^j} - \hat{g}_{k^j}) \Big) \prod_{m \neq j} \psi_{k_m}(x_m) \Big)^2 \Big)^{\frac{p}{2}} \\
&= \Big(\sum_j \sum_{k^j} \Big(\sum_{k_j : k_j = |k|_\infty} \frac{1}{i2\pi k_j} (\hat{f}_k - \hat{g}_k) \Big)^2 \Big)^{\frac{p}{2}} \\
&= \Big(\sum_j \sum_{k^j} \|\hat{f}_{k^j} - \hat{g}_{k^j}\|_{1,w^\infty}^2 \Big)^{\frac{p}{2}}
\end{aligned}$$

where the sequence of weights w is given by (2). Also for $p \in (2, \infty)$, by Hausdorff-Young inequality

$$\begin{aligned}
\|\tilde{V}\|_{L_p}^p &= \int_{\mathbb{T}^d} \|\tilde{V}(x)\|_2^p dx \\
&\leq d^{\frac{p}{2}-1} \int_{\mathbb{T}^d} \|\tilde{V}(x)\|_p^p dx \\
&= d^{\frac{p}{2}-1} \int_{\mathbb{T}^d} \sum_j \Big(\sum_{k^j} \Big(\sum_{k_j : k_j = |k|_\infty} \frac{1}{i2\pi k_j} (\hat{f}_k - \hat{g}_k) \Big) \prod_{m \neq j} \psi_{k_m}(x_m) \Big)^p dx \\
&\leq d^{\frac{p}{2}-1} \sum_j \sum_{k^j} \Big(\sum_{k_j : k_j = |k|_\infty} \frac{1}{i2\pi k_j} (\hat{f}_k - \hat{g}_k) \Big)^{p'} \\
&= d^{\frac{p}{2}-1} \sum_j \Big(\sum_{k^j} \|\hat{f}_{k^j} - \hat{g}_{k^j}\|_{1,w^\infty}^{p'} \Big)^{p'}
\end{aligned}$$

G Proof of Lemma 5.1

Proof. By Lemma B.1 and Hölder's inequality

$$\begin{aligned}
\int h(f-g)dx &= \int h(f - \tilde{f})dx + \int h(\tilde{f} - \tilde{g})dx + \int h(g - \tilde{g})dx \\
&\leq \|h\|_\infty \|f - \tilde{f}\|_1 + M^{1/p'} \|\nabla h\|_{L_q(\Omega)} W_p(\tilde{f}, \tilde{g}) + \|h\|_\infty \|g - \tilde{g}\|_1 \\
&= \Big(\Big(\|h\|_\infty (\|f - \tilde{f}\|_1 + \|g - \tilde{g}\|_1) + M^{1/p'} \|\nabla h\|_{L_q(\Omega)} W_p(\tilde{f}, \tilde{g}) \Big)^p \Big)^{\frac{1}{p}} \\
&\leq \Big(2^{p-1} \Big(\|h\|_\infty^p (\|f - \tilde{f}\|_1 + \|g - \tilde{g}\|_1)^p + (M^{1/p'} \|\nabla h\|_{L_q(\Omega)} W_p(\tilde{f}, \tilde{g}))^p \Big) \Big)^{\frac{1}{p}} \\
&\leq 2^{\frac{p-1}{p}} W_p^{a,b}(f, g)
\end{aligned}$$

where in the next to last inequality, we have used the standard inequality $(x+y)^p \leq 2^{p-1}(x^p + y^p)$ for all $x, y > 0$. □

H Proof of Theorem 5.2

Proof. We modify the proof of Theorem 4.5 starting after (75). Note that if

$$\|\lambda\|_1 \leq a \tag{80}$$

then $\|h\|_\infty \leq a$, and we obtain the following result. We have

$$\|\lambda\|_1 = \alpha \Big\| |\hat{f} - \hat{g}|^{\frac{q}{q'}} w^{\frac{q}{q'}+1} \Big\|_1 = \alpha \Big\| |\hat{f} - \hat{g}|^{\frac{q}{q'}} w^{q'} \Big\|_1 = \alpha \Big\| (\hat{f} - \hat{g})^{\frac{1}{q'}} \Big\|_{q, w^{q'}}^q. \tag{81}$$

Therefore, setting

$$\alpha = \min\left(M^{-1/p'} d^{\frac{1}{q}-\frac{1}{2}} b \Big/ \|\hat{f} - \hat{g}\|_{q,w^{q'}}^{q-1}, a \left\| (\hat{f} - \hat{g})^{\frac{1}{q'}} \right\|_{q,w^{q'}}^{q} \right)$$

guarantees (70) and (80). Now evaluating

$$\begin{aligned}\langle \lambda, (\hat{f} - \hat{g}) \rangle &= \alpha \sum_k |\hat{f}_k - \hat{g}_k| w_k \Big(|\hat{f}_k - \hat{g}_k| w_k \Big)^{q-1} = \alpha \|\hat{f} - \hat{g}\|_{q,w^{q'}}^q \\ &= \left(\left(M^{-1/p'} d^{\frac{1}{q}-\frac{1}{2}} b \right) \wedge \left(\frac{a}{\|(\hat{f} - \hat{g})^{\frac{1}{q'}}\|_{q,w^{q'}}^q} \right) \right) \|\hat{f} - \hat{g}\|_{q,w^{q'}}^q\end{aligned}$$

completes this proof. ∎

I Proof of Theorem 6.1

Proof. By a standard decomposition,

$$\|m - m_\delta^c\|_{L^2} = \|m - BAm + BAm - Bg^\delta\|_{L^2} \le \|(\mathcal{I} - BA)m\|_{L^2} + \|Bn\|_{L^2}$$

where $\mathcal{I}$ denotes the identity operator. From the definition of $\dot{\mathcal{H}}^r$, the operator $\mathcal{I} - BA$ has the following norm

$$\|\mathcal{I} - BA\|_{\dot{\mathcal{H}}^r \to L^2} = (2\pi k_c)^{-r}.$$

Therefore,

$$\|(\mathcal{I} - BA)m\|_{L^2} \le (2\pi k_c)^{-r} \|m\|_{\dot{\mathcal{H}}^r}. \tag{82}$$

∎

J Proof of Remark 6.2

Proof. We use the standard result for vectors in $\mathbb{R}^d$ (76). Also the cardinality of $\{k \in \mathbb{Z}^d \ : \ 1 \le \|k\| \le k_c\}$ is proportional to k_c^d.

$$\begin{aligned}\|Bn\|_{L^2} &= \left(\sum_{1 \le \|k\| \le k_c} |\widehat{n}_k|^2 \|k\|^{2\gamma} \right)^{\frac{1}{2}} \\ &= \left(\sum_{1 \le \|k\| \le k_c} \|k\|^{2\gamma+2} \frac{|\widehat{n}_k|^2}{\|k\|^2} \right)^{\frac{1}{2}} \\ &\lesssim \begin{cases} k_c^{1+\gamma+\frac{d}{2}-\frac{d}{q}} \delta_q & \text{if } \gamma > -1 \\ k_c^{\frac{d}{2}-\frac{d}{q}} \delta_q & \text{if } \gamma \le -1 \end{cases}\end{aligned}$$

When $\|n\|_{\dot{\mathcal{H}}_\beta} \leq z$,

$$\begin{aligned}
\|Bn\|_{L^2} &= \left(\sum_{1 \leq \|k\| \leq k_c} |\widehat{n}_k|^2 \|k\|^{2\gamma} \right)^{\frac{1}{2}} \\
&= \left(\sum_{1 \leq \|k\| \leq k_c} \|k\|^{2\gamma+1-\beta} \frac{|\widehat{n}_k|}{\|k\|} \|k\|^\beta |\widehat{n}_k| \right)^{\frac{1}{2}} \\
&\leq \sqrt{z} \left(\sum_{1 \leq \|k\| \leq k_c} \|k\|^{4\gamma+2-2\beta} \frac{|\widehat{n}_k|^2}{\|k\|^2} \right)^{\frac{1}{2}} \\
&\lesssim \sqrt{z} \begin{cases} k_c^{1+2\gamma-beta+\frac{d}{2}-\frac{d}{q}} \sqrt{\delta_q} & \text{if } \gamma > \beta/2 - 1/2 \\ k_c^{\frac{d}{2}-\frac{d}{q}} \sqrt{\delta_q} & \text{if } \gamma \leq \beta/2 - 1/2 \end{cases}
\end{aligned}$$

□

K Proof of Proposition 6.3

Proof. If $d < \eta$, then

$$\begin{aligned}
\|n\|_{L_\infty}^2 &\leq \sum_{k_n \leq \|k\|_2 \leq ck_n} \|k\|^{-\eta} && (83) \\
&\approx \int_1^\infty r^{-\eta+d-1} dr && (84) \\
&\leq C_n && (85)
\end{aligned}$$

uniformly in k_n and b. Therefore, we can rescale n by taking

$$\tilde{n} = \frac{\xi}{C_n} n$$

to ensure that $\|\tilde{n}\|_{L_\infty} \leq \xi$ and therefore $g_\delta = g + n \geq \xi$ a.e. on $\mathbb{T}^d$. B monotonicity of L^p on $\mathbb{T}^d$ with bounded Lebesgue measure, $\|\tilde{n}\|_{L_s} \leq \xi$ for all $s \in [1, \infty)$. Also

$$\begin{aligned}
\|\tilde{n}\|_{\dot{\mathcal{H}}^s}^2 &= (2\pi\|k\|_2)^{2\beta} \hat{\tilde{n}}_k^2 && (86) \\
&\approx \int_{k_n}^{bk^n} r^{2(\beta-\eta)+d-1} dr && (87)
\end{aligned}$$

This integral converges of $2(\beta - \eta) + d < 0$. Accordingly, it suffices to show that equivalently $2\beta + d < 2\eta$ if $d < \eta$. Let us rewrite the inequality (31) as $s = \frac{d}{p} - \frac{d}{2} - 1 + \varepsilon_2$ for some $\varepsilon_2 > 0$. For $p \in [1, 2)$, we have

$$2\eta > 2d \geq 2d \left(\frac{1}{p} - \frac{1}{2} \right) + d = 2(\beta + 1 - \varepsilon_2) + d \geq 2\beta + d$$

where the last inequality holds if $\varepsilon_2 \leq 1$ or equivalently $\frac{d}{p} - \frac{d}{2} \geq \beta$. We estimate

$$\delta_q^q = \|\hat{n}\|_{q,w^2}^q \asymp \sum_k \frac{\hat{n}_k^q}{(2\pi\|k\|_2)^q}$$

by the integral

$$\int_{k_n}^{bk_n} r^{q(-\eta-1)+d-1} dr \asymp (bk_n)^{q(-\eta-1)+d} - k_n^{q(-\eta-1)+d} \asymp bk_n^{q(-\eta-1)+d}$$

Therefore,

$$\delta_q \asymp (bk_n)^{(-\eta-1)+\frac{d}{q}}. \tag{88}$$

and the result follows by application of the upper and lower bounds on k_c. □

L Proof of Theorem 6.6

Proof. By the standard error decomposition:

$$\|m - m_\delta^c\|_{L^2} = \|m - BAm + BAm - Bg^\delta\|_{L^2} \leq \|(\mathcal{I} - BA)m\|_{L^2} + \|Bn\|_{L^2(\Omega)}$$

where $\mathcal{I}$ denotes the identity operator. Using (60),

$$\|\mathcal{I} - BA\|_{L^2(\Omega)} \leq \|\alpha\|_2 + 4^{jr}\|n\|_{\mathcal{H}_\Psi^r}^2$$

□