On metrics robust to noise, perturbations, and changes in projection angle

William Leeb

School of Mathematics
University of Minnesota, Twin Cities
Minneapolis, MN

Abstract

We study a family of distances between functions of a single variable. These distances are examples of integral probability metrics, and have been used previously for comparing probability measures. Special cases include the Earth Mover's Distance and the Kolmogorov Metric. We examine their properties for general signals, proving that they are robust to a broad class of perturbations and that the distance between one-dimensional tomographic projections of a two-dimensional function is bounded by the size of the difference in projection angles. We also establish error bounds for approximating the metric from finite samples, and prove that these approximations are robust to additive Gaussian noise. The results are illustrated in numerical experiments.

1 Introduction

Many tasks in statistics and machine learning require specification of a metric to measure the distance between two data vectors. For example, typical methods for clustering will attempt to group points together that are close and separate points that are far, where "close" and "far" are determined by a certain metric or similarity measure [48, 56, 2]. The specification of the metric can alter the effectiveness of such methods. A good metric will be robust to noise and to irrelevant perturbations of the input data, so that only the "meaningful" characteristics of each data vector inform the distance.

A popular class of metrics, typically used for comparing two probability measures, are called *integral probability metrics*, also known as *maximum mean discrepancies* [26, 14, 5, 41, 38, 3, 54]; if g and h are defined on a measure space $\mathcal{X}$, these distances are given by

$$d_{\mathcal{F}}(g,h) = \sup_{F \in \mathcal{F}} \int_{\mathcal{X}} F(x)(g(x) - h(x))dx, \tag{1}$$

where $\mathcal{F}$ is a specified class of test functions. (Of course, this may also be well-defined if g and h are not probability measures, which is the perspective we will take in this paper.)

One of the most widely-used examples of a metric of the form (1) is the Earth Mover's Distance (EMD) for comparing probability distributions g and h, applicable when $\mathcal{X}$ is a metric space. Informally, the EMD between g and h is equal to the minimal cost of transforming g into h by rearranging the mass, where the cost is determined by the metric on $\mathcal{X}$ [57, 58]. EMD and other related distances, such as the p-Wasserstein distance, are popular metrics in machine learning and statistical applications [44, 50, 13, 45, 49, 34, 10, 47, 40, 15, 39]. It is a consequence of the Kantorovich-Rubinstein Theorem [27, 18, 36, 37, 20] that the EMD between g and h is equal to the distance (1) when $\mathcal{F}$ is the set of 1-Lipschitz functions.

Despite their applicability to a wide range of problems, many fundamental properties of these metrics are not well understood. The present work is an attempt to further elucidate their properties. We are most directly inspired by two results that are known to hold for Wasserstein distances. The first result, which is proven in [33], states that if there is a smooth bijection $\varphi : \mathcal{X} \to \mathcal{X}$ and

$$g(x) = h(\varphi(x))\frac{d\varphi}{dx}(x), \tag{2}$$

(where $\frac{d\varphi}{dx}(x)$ is the Radon-Nikodym derivative of φ), then

$$W_1(g, h) \leq \sup_{x\in\mathcal{X}} d_{\mathcal{X}}(x, \varphi(x)), \tag{3}$$

where W_1 denotes the EMD, or 1-Wasserstein distance. Informally, (3) tells us that the EMD is robust to "small" perturbations of the data, where the "size" of a perturbation is the maximum distance that any point in the domain may be moved.

A result in the recent paper [46] suggests that the p-Wasserstein distance W_p is natural to use for clustering tomographic projection images that arise from cryo-electron microscopy (cryo-EM), a technique for molecular reconstruction that is increasingly used in structural biology [52, 7, 17]. Suppose g and h are functions of two variables that are each projections of a common three-dimensional volume f; that is,

$$g(x, y) = \int_{\mathbb{R}} f(R_g(x, y, z))dz, \tag{4}$$

where R_g is an orthogonal transformation; and similarly for h. Then [46] proves that

$$\min_{R\in SO(2)} W_p(g, h \circ R) \leq \theta(R_g, R_h), \tag{5}$$

where $\theta(R_g, R_h)$ is the angle between the projection directions of R_g and R_h.

These results raise two natural sets of questions, which motivate the content of this paper. First, are the stability properties (3) and (5) unique to Wasserstein distances, or are there other families of metrics that exhibit similar properties?

Second, are there additional stability properties that have not yet been identified? There are two particularly interesting questions that arise in the context of cryo-EM. The first of these is robustness to *heterogeneity*. Cryo-EM datasets may contain projection images from multiple 3D volumes, either from different molecules or a single molecule in different conformations [60, 22, 35, 4, 31, 51, 55, 32]. When comparing two such projection images, it would be desirable for the distance metric to be bounded by the size of the distortion between the volumes.

Another potentially desirable property is robustness to additive noise. Again, this is of particular significance in the tomographic applications described in [46], since images from cryo-EM are typically very noisy. To even make sense of the question of noise, however, one must move away from comparing only probability measures, and allow for signals with unequal mass and negative values.

In this paper, we provide precise answers to these questions, in the simple setting of functions of a single variable. We study a certain family of metrics between single-variable functions on an interval. These metrics are induced by norms, denoted by $\|f\|_{V^p}$, which are the p-norms of the Volterra operator (i.e. the indefinite integral operator) applied to f. As is known, these metrics may be written in the form (1) by taking $\mathcal{F}$ to be the space of functions whose derivative is in $L^{p/(p-1)}$; see Proposition 3.1 below. We call the norm $\|\cdot\|_{V^p}$ the *Volterra p-norm*, and its induced metric the *Volterra p-distance*. The Volterra distances have been used previously for comparing probability measures on the real line [41, 38]. We also consider a discrete approximation $\|\mathbf{f}\|_{\nu_p}$ for vectors $\mathbf{f}$ of samples of f on an equispaced grid.

We show that the Volterra p-distances exhibit robustness properties analogous to (3) and (5), while also being robust to 2D perturbations and additive noise. We remark that, while our results on distances between projections are applicable to 2D tomography, they may still give insights into 3D tomography, analogous to those gleaned from the study of multi-reference alignment [9, 8, 6, 43, 1].

The main results of the paper are summarized as follows:

- **Robustness to 1D perturbations.** The Volterra distances are robust to integral-preserving changes of variable. That is, the distance $\|f - \tilde{f}\|_{V^p}$ between a function f and a perturbation $\tilde{f}$ of f is bounded by a certain measure of the size of the perturbation.

- **Robustness to changes in projection angle.** The Volterra distance between two one-dimensional projections of a two-dimensional function is bounded by the size of the difference in projection angles.
- **Robustness to 2D perturbations.** The Volterra distance between two one-dimensional projections of a two-dimensional function and its perturbation is bounded by the size of the perturbation.
- **Convergence of the discrete norm.** For a broad class of functions f, the discrete norm $\|\mathbf{f}\|_{\nu_p}$ of a vector $\mathbf{f}$ of samples of f taken on an equispaced grid converges to $\|f\|_{V^p}$.
- **Robustness to noise.** The discrete norm $\|Z\|_{\nu_p}$ of a noise vector Z vanishes as $n \to \infty$, whereas the discrete norm of a signal vector converges to the corresponding continuous norm. In particular, the discrete norm of a noisy, sampled function converges to the norm of the noiseless function.

The remainder of the paper is structured as follows. In Section 2, we review basic terminology and notation. In Section 3, we formally define the Volterra p-norms and prove a general stability result that will be useful in the proofs of the main results. In Section 4, we study the robustness of the Volterra norms. More specifically, in Section 4.1, we prove robustness to one variable perturbations; in Section 4.2 we prove robustness to changes in tomographic projection direction; in Section 4.3, we prove robustness to two variable perturbations; in Section 4.4, we analyze the approximation error of the approximate norms based on function samples; in Section 4.5, we analyze the behavior of the discrete norm in the presence of additive noise. In Section 5, we present the results of numerical experiments. Section 6 contains detailed proofs of the main results. Section 7 provides a brief summary and conclusion.

2 Preliminaries

This section introduces the basic definitions and notation that will be used in the rest of the paper. Proofs of most of the results stated here may be found in, for example, [21]. Familiarity with basic concepts of measure and integration will be assumed. Throughout, $a < b$ will denote arbitrary real numbers.

2.1 The Lebesgue p-norms

Let $f : [a, b] \to \mathbb{R}$ be any measurable function. For any value $p \in [1, \infty)$, the p-norm is defined as follows:

$$\|f\|_{L^p([a,b])} = \left(\int_a^b |f(x)|^p dx \right)^{1/p}. \tag{6}$$

For $p = \infty$, we define

$$\|f\|_{L^\infty([a,b])} = \operatorname*{ess\,sup}_{a \le x \le b} |f(x)|. \tag{7}$$

When a and b are clear from the context, we will just write $\|f\|_{L^p}$ for brevity. We denote by $L^p = L^p([a, b])$ the set of all functions f on $[a, b]$ with $\|f\|_{L^p([a,b])} < \infty$. As is well-known, $\|f\|_{L^p([a,b])} \le \|f\|_{L^q([a,b])}(b-a)^{1/p-1/q}$ if $p \le q$; in particular, $L^p([a, b]) \subset L^q([a, b])$ for $p \le q$.

We define the inner product between two real-valued functions on $[a, b]$ as follows:

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx. \tag{8}$$

We also define the p-norm $\|\mathbf{x}\|_{\ell_p}$ for vectors $\mathbf{x}$ in $\mathbb{R}^n$. When $p < \infty$,

$$\|\mathbf{x}\|_{\ell_p} = \left(\frac{1}{n} \sum_{j=1}^n |x_j|^p \right)^{1/p}, \tag{9}$$

and when $p = \infty$,

$$\|\mathbf{x}\|_{\ell_\infty} = \max_{1 \le k \le n} |x_k|. \tag{10}$$

Note the normalization by n when $p < \infty$. With this convention, $\|\mathbf{x}\|_{\ell_p} \le \|\mathbf{x}\|_{\ell_q}$ whenever $p \le q$.

2.2 Absolute continuity

A function G on $[a, b]$ is said to be absolutely continuous if it can be written as

$$G(x) = G(a) + \int_a^x g(t)\,dt \tag{11}$$

for a function g in $L^1([a, b])$. If G is absolutely continuous, then it is differentiable almost everywhere, and $G' = g$ where the derivative exists. We denote by $\mathcal{A}_0$ the set of absolutely continuous functions G satisfying $G(b) = 0$; these functions may be written as

$$G(x) = -\int_x^b g(t)\,dt \tag{12}$$

where $g = G'$ almost everywhere. For brevity, whenever G is in $\mathcal{A}_0$, G' will denote any function such that $G(x) = -\int_x^b G'(t)\,dt$.

The following result is standard (e.g., see Section 3.5 of [21]):

Theorem 2.1 (Integration-by-parts)**.** *If F and G are absolutely continuous functions on $[a, b]$, then*

$$\int_a^b (F'(x)G(x) + F(x)G'(x))dx = F(b)G(b) - F(a)G(a). \tag{13}$$

2.3 The Volterra operator

The Volterra operator $\mathcal{V}$ is defined on $L^1([a, b])$ by

$$(\mathcal{V}f)(x) = \int_a^x f(t)dt. \tag{14}$$

We note that this is only the simplest of a large family of operators that have been widely studied [25]. Note that if f is in $L^1([a, b])$, $\mathcal{V}f$ is in L^∞, with $\|\mathcal{V}f\|_{L^\infty} \leq \|f\|_{L^1}$; furthermore, $\mathcal{V}f$ is, by definition, absolutely continuous when f is in $L^1([a, b])$.

The adjoint transform $\mathcal{V}^*$ is given by

$$(\mathcal{V}^*f)(x) = \int_x^b f(t)dt. \tag{15}$$

This operator satisfies

$$\langle \mathcal{V}f, g\rangle = \langle f, \mathcal{V}^*g\rangle \tag{16}$$

where f and g are two functions in $L^1([a, b])$.

3 The Volterra p-norms

3.1 The continuous Volterra norm

Let f be in $L^1([a, b])$. For any value $p \in [1, \infty]$, we define the following norm, which we will refer to as the *Volterra p-norm*:

$$\|f\|_{V^p} = \|\mathcal{V}f\|_{L^p}. \tag{17}$$

Concretely, when $p < \infty$,

$$\|f\|_{V^p} = \left(\int_a^b \left|\int_a^x f(t)dt\right|^p dx\right)^{1/p}, \tag{18}$$

and when $p = \infty$,

$$\|f\|_{V^\infty} = \operatorname*{ess\,sup}_{a \le x \le b} \left| \int_a^x f(t)dt \right|. \tag{19}$$

Note that, because $\mathcal{V}f$ is in $L^p([a,b])$, the Volterra p-norm of f is finite for any function f in $L^1([a,b])$. If f and g are two functions in $L^1([a,b])$, we will refer to $\|f - g\|_{V^p}$ as the *Volterra p-distance* between f and g.

Remark 1. When $p = \infty$ and f and g are two probability densities, the Volterra ∞-distance is known as the Kolmogorov Metric between f and g [24]: $\mathrm{KM}(f,g) = \|f - g\|_{V^\infty}$. The KM arises in the context of goodness-of-fit testing in statistics [23].

Remark 2. When $p = 1$ and f and g are two probability densities, the Volterra 1-distance is known as the Earth Mover's Distance between f and g [57, 58]: $\mathrm{EMD}(f,g) = \|f - g\|_{V^1}$. The metric $\mathrm{EMD}(f,g)$ may be defined equivalently as the solution to a transportation problem:

$$\mathrm{EMD}(f,g) = \min_{\Pi \in \mathcal{M}(f,g)} \int_a^b \int_a^b |x - y| \Pi(x,y) dx dy \tag{20}$$

where $\mathcal{M}(f,g)$ denotes the space of all probability measures on $[a,b] \times [a,b]$ with marginals equal to f and g, respectively. That is, $\Pi \in \mathcal{M}(f,g)$ if for all x,

$$\int_a^b \Pi(x,y) dy = f(x), \tag{21}$$

and for all y,

$$\int_a^b \Pi(x,y) dx = g(y). \tag{22}$$

The p-Wasserstein distance $W_p(f,g)$ is defined as

$$W_p(f,g) = \min_{\Pi \in \mathcal{M}(f,g)} \left(\int_a^b \int_a^b |x - y|^p \Pi(x,y) dx dy \right)^{1/p}. \tag{23}$$

It is known [50] that $W_p(f,g)$ may be written as follows:

$$W_p(f,g) = \|(\mathcal{V}f)^{-1} - (\mathcal{V}g)^{-1}\|_{L^p} \tag{24}$$

where $(\mathcal{V}f)^{-1}$ and $(\mathcal{V}g)^{-1}$ denote the functional inverses of $\mathcal{V}f$ and $\mathcal{V}g$, respectively. When $p = 1$, it is also true that $W_1(f,g) = \|\mathcal{V}f - \mathcal{V}g\|_{L^1} = \|f - g\|_{V^1}$; when $p > 1$, however, the p-Wasserstein distance $W_p(f,g)$ is generally *not* equal to the Volterra p-distance $\|\mathcal{V}f - \mathcal{V}g\|_{V^p}$.

3.2 Variational formulation of $\|f\|_{V^p}$

The following result is an alternate formulation of the Volterra norm that will be useful for analysis. It essentially appears as Theorem 1 in [38]; we provide a self-contained proof for the reader's convenience.

Proposition 3.1. *Let $p \in [1, \infty]$ and let q be the conjugate exponent:*

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{25}$$

Then for any function f in $L^1([a,b])$,

$$\|f\|_{V^p} = \sup_{G \in \mathcal{A}_0 : \|G'\|_{L^q} \le 1} \langle f, G \rangle. \tag{26}$$

Proof of Proposition 3.1. By duality of L^p and L^q, we have:

$$\|f\|_{V^p} = \|\mathcal{V}f\|_{L^p} = \sup_{g:\|g\|_{L^q}\leq 1} \int_a^b (\mathcal{V}f)(x)g(x)dx = \sup_{g:\|g\|_{L^q}\leq 1} \langle \mathcal{V}f, g\rangle = \sup_{g:\|g\|_{L^q}\leq 1} \langle f, \mathcal{V}^* g\rangle. \tag{27}$$

Any function of the form $\mathcal{V}^*g$ is contained in $\mathcal{A}_0$, and any function G in $\mathcal{A}_0$ is of the form $G = \mathcal{V}^*g$ where $g = G'$ almost everywhere. Consequently:

$$\|f\|_{V^p} = \|\mathcal{V}f\|_{L^p} = \sup_{g:\|g\|_{L^q}\leq 1} \langle f, \mathcal{V}^* g\rangle = \sup_{G\in\mathcal{A}_0:\|G'\|_{L^q}\leq 1} \langle f, G\rangle, \tag{28}$$

which completes the proof. □

As a corollary, we derive the following general stability result:

Corollary 3.2. *Let $p \geq 1$. Let $I \subseteq [a, b]$ be a subinterval, and suppose $\mathcal{T} : L^\infty([a, b]) \to L^\infty([a, b])$ is a linear transformation that maps $L^1([a, b])$ to $L^1([a, b])$ and satisfies*

$$\|(\mathcal{T}^*(G) - G)\chi_I\|_{L^q} \leq \epsilon\|G'\|_{L^q} \tag{29}$$

for all functions G in $\mathcal{A}_0$. Then for all functions f in $L^p([a, b])$ supported on I,

$$\|\mathcal{T}(f) - f\|_{V^p} \leq \epsilon\|f\|_{L^p}. \tag{30}$$

Proof. Since $\langle\mathcal{T}(f) - f, G\rangle = \langle f, \mathcal{T}^*(G) - G\rangle = \langle f\chi_I, \mathcal{T}^*(G) - G\rangle = \langle f, (\mathcal{T}^*(G) - G)\chi_I\rangle$, from Hölder's inequality we have:

$$\begin{aligned}
\|\mathcal{T}(f) - f\|_{V^p} &= \sup_{G\in\mathcal{A}_0:\|G'\|_{L^q}\leq 1} \langle\mathcal{T}(f) - f, G\rangle \\
&= \sup_{G\in\mathcal{A}_0:\|G'\|_{L^q}\leq 1} \langle f, (\mathcal{T}^*(G) - G)\chi_I\rangle \\
&\leq \sup_{G\in\mathcal{A}_0:\|G'\|_{L^q}\leq 1} \|f\|_{L^p}\|(\mathcal{T}^*(G) - G)\chi_I\|_{L^q} \\
&\leq \epsilon\|f\|_{L^p},
\end{aligned} \tag{31}$$

completing the proof. □

3.3 Trapezoidal rule approximation to $\|f\|_{V^p}$

Suppose f is a function on $[a, b]$, and we are given samples of f on an equispaced grid of points in $[a, b]$, from which we wish to approximate $\|f\|_{V^p}$. That is, let $a_0 < a_1 < \cdots < a_n$ be equispaced points in $[a, b]$ defined by

$$a_k = a + \frac{k}{n}(b - a), \quad 0 \leq k \leq n. \tag{32}$$

Note that $a_0 = a$ and $a_n = b$. We suppose we are given the values of $f(a_j)$, $0 \leq j \leq n$, or possibly noisy estimates of these, and wish to approximate $\|f\|_{V^p}$.

To this end, we introduce some convenient notation. If $\mathbf{v}$ is a vector in $\mathbb{R}^{n+1}$ and $1 \leq p < \infty$, we define the norm

$$\|\mathbf{v}\|_{\tau_p} = \left(\frac{b-a}{n}\sum_{k=1}^{n-1} |\mathbf{v}[k]|^p + \frac{b-a}{2n}|\mathbf{v}[0]|^p + \frac{b-a}{2n}|\mathbf{v}[n]|^p\right)^{1/p}. \tag{33}$$

When $p = \infty$, we define

$$\|\mathbf{v}\|_{\tau_\infty} = \max_{0\leq k\leq n} |\mathbf{v}[k]|. \tag{34}$$

If F is a function on $[a,b]$ and $\mathbf{v}$ is a vector with entries $\mathbf{v}[k] = F(a_k)$, $0 \le k \le n$, then $\|\mathbf{v}\|_{\tau_p}$ is the trapezoidal rule approximation to $\|F\|_{L^p}$ when $1 \le p < \infty$, and $\|\mathbf{v}\|_{\tau_\infty}$ is an approximation to $\|F\|_{L^\infty}$.

It will be convenient to introduce the following notation for working with the trapezoidal rule. If $\mathbf{v}$ if a vector in $\mathbb{R}^{n+1}$, we define

$$\sum_{0\le k\le n}^{\text{trap}} \mathbf{v}[k] \equiv \sum_{k=1}^{n-1} \mathbf{v}[k] + \frac{1}{2}\mathbf{v}[0] + \frac{1}{2}\mathbf{v}[n]. \tag{35}$$

For instance, when $1 \le p < \infty$, we may write

$$\|\mathbf{v}\|_{\tau_p} = \left(\frac{b-a}{n} \sum_{0\le k\le n}^{\text{trap}} |\mathbf{v}[k]|^p \right)^{1/p}. \tag{36}$$

Suppose n is a positive integer. We define the following discrete Volterra operator $\mathbf{V} : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ on a vector $\mathbf{x}$ by $(\mathbf{V}\mathbf{x})[0] = 0$, and

$$(\mathbf{V}\mathbf{x})[k] = \frac{b-a}{n} \sum_{j=1}^{k-1} \mathbf{x}[j] + \frac{b-a}{2n}\mathbf{x}[0] + \frac{b-a}{2n}\mathbf{x}[k], \quad 1 \le k \le n. \tag{37}$$

We then define the discrete Volterra p-norm of $\mathbf{x}$ as

$$\|\mathbf{x}\|_{\nu_p} = \|\mathbf{V}\mathbf{x}\|_{\tau_p}. \tag{38}$$

The interpretation of this quantity may be understood as follows. Suppose f is a function on $[a,b]$, and $a_0 < a_1 < \cdots < a_n$ are equispaced points in $[a,b]$, as in (32). Let $\mathbf{f}$ be the vector in $\mathbb{R}^{n+1}$ with entries $\mathbf{f}[k] = f(a_k)$, for $0 \le k \le n$. Then $(\mathbf{V}\mathbf{f})[k]$ is the trapezoidal rule approximation to $(\mathcal{V}f)(a_k)$, and $\|\mathbf{f}\|_{\nu_p}$ approximates $\|f\|_{V_p}$. The approximation error will be described in Section 4.4.

4 Properties of the Volterra p-norms

4.1 One-dimensional perturbations

As a consequence of Corollary 3.2, we will show that if f and $\tilde{f}$ are two functions on $[a,b]$ related by a perturbation, then their distance $\|f - \tilde{f}\|_{V^p}$ is bounded by the size of the perturbation. More precisely, we have the following result:

Theorem 4.1. *Suppose $f \in L^p$ is supported on an interval I. Let $\varphi : J \to I$ be continuously differentiable and monotonically increasing, with $I = \varphi(J)$. Let $f_\varphi(x) = f(\varphi(x))\varphi'(x)$ on J, and 0 otherwise. Suppose $I \cup J \subset [a,b]$. Then*

$$\|f - f_\varphi\|_{V^p} \le \min\left\{\epsilon \cdot C(f,\varphi,p),\, \epsilon^{1/p} \cdot \|f\|_{L^1}\right\}, \tag{39}$$

where

$$\epsilon = \max_{x\in J} |x - \varphi(x)|, \tag{40}$$

and

$$C(f,\varphi,p) = \min\{\|f\|_{L^p}, \|f_\varphi\|_{L^p}\}. \tag{41}$$

The proof of Theorem 4.1 is contained in Section 6.1.

Remark 3. The special case of Theorem 4.1 in which $p = 1$ follows from Proposition 14 in [33].

Remark 4. A special case of Theorem refthm:perturbation may be described as follows. Suppose f in $L^p([a,b])$ is supported on a subinterval $I \subset [a,b]$. Let $\epsilon > 0$ and $J = \{x + \epsilon : x \in I\} \subset [a,b]$, and let f_ϵ be the shift of f by ϵ; that is,

$$f_\epsilon(x) = \begin{cases} f(x-\epsilon), & \text{if } x \in J \\ 0, & \text{otherwise} \end{cases}. \tag{42}$$

Then for all $p \in [1,\infty]$, $\|f - f_\epsilon\|_{V^p} \le \min\left\{\epsilon \cdot \|f\|_{L^p},\, \epsilon^{1/p} \cdot \|f\|_{L^1}\right\}$.

Remark 5. Another special case of Theorem 4.1 is as follows. Suppose X is a bounded random variable with density f_X supported in $I \subset [a,b]$, and let $\varphi : J \to I$ be a continuously differentiable, one-to-one mapping. Let $Y = \varphi^{-1}(X)$, with density f_Y (supported on J). Then for all $p \in [1,\infty]$,

$$\|f_X - f_Y\|_{V^p} \le \min\left\{\epsilon,\, \epsilon^{1/p} \cdot C\right\}, \tag{43}$$

where $C = \min\{\|f_X\|_{L^p}, \|f_Y\|_{L^p}\}$; in particular, $\mathrm{EMD}(f_X, f_Y) \le \epsilon$. This follows from the fact that Y has density $f_Y(y) = f_X(\varphi(y))\varphi'(y)$.

4.2 Changes in projection angle

In this section we let $F : \mathbb{R}^2 \to \mathbb{R}$ denote a function of two variables supported on $\mathbb{D}_R \subset \mathbb{R}^2$, the disc of radius R and center $(0,0)$. For a given angle θ, we define the projection of F as follows:

$$f_\theta(x) = \int_{\mathbb{R}} F(\cos(\theta)x + \sin(\theta)y, \cos(\theta)y - \sin(\theta)x)dy \tag{44}$$

The function $f_\theta : [-R,R] \to \mathbb{R}$ is the tomographic projection of F onto the line passing through $(0,0)$, making angle θ with the x-axis. Such a transformation is known as the two-dimensional Radon transform with parallel beam geometry [42], and is a standard transformation in scientific imaging [53, 16]. We claim the following result:

Theorem 4.2. *For all angles θ and φ with $|\theta - \varphi| < \pi$ and all $p \in [1,\infty]$,*

$$\begin{aligned} \|f_\theta - f_\varphi\|_{V^p} &\le (2\sin(|\theta-\varphi|/2))^{1/p} \cdot |\theta-\varphi|^{1-1/p} \cdot \|F\|_{L^p} \cdot R^{2-1/p} \\ &\le |\theta - \varphi| \cdot \|F\|_{L^p} \cdot R^{2-1/p}, \end{aligned} \tag{45}$$

where $\|F\|_{L^p}$ denotes the p-norm of F over $\mathbb{D}$.

If the problem is renormalized so that $R = 1$ and $\|F\|_{L^p} = 1$, the bound may be stated more simply as

$$\|f_\theta - f_\varphi\|_{V^p} \le (2\sin(|\theta-\varphi|/2))^{1/p} \cdot |\theta-\varphi|^{1-1/p} \le |\theta-\varphi|. \tag{46}$$

The proof of Theorem 4.2 may be found in Section 6.2.

Remark 6. As described in the introduction, a result similar to Theorem 4.2 for p-Wasserstein distances and the three-dimensional Radon transform was proven in the paper [46]. In that work, the fact that the Wasserstein distance is insensitive to changes in projection angle allows for robust clustering of images taken from different viewing directions which are not known a priori, as occurs in cryo-electron microscopy [52, 7, 17].

4.3 Two-dimensional perturbations

In this section, we consider the distance between the tomographic projection of a function, and the tomographic projection of a perturbation of the same function. Unlike the setting of Section 4.2, both projections are taken along the same direction; however, the function F of two variables may change. Of course, when the perturbation is a rotation, then one may view the setting of Section 4.2 as a special case of this more general setting; however, the estimate from Theorem 4.2 is sharper than the more general bound that we will prove here, in Theorem 4.3 below.

One of the motivations for this model is the problem of heterogeneity in cryo-electron microscopy. This refers to the setting where the set of projection images come from a molecule in multiple conformations. It is natural to model each conformation as an L^1-preserving perturbation of the type considered here.

Theorem 4.3. *Suppose $F : \mathbb{R}^2 \to \mathbb{R}$ is in L^p, and is supported on a compact set A. Let $\Phi : B \to A$ be a continuously differentiable, one-to-one mapping, and let $F_\Phi(x, y) = F(\Phi(x, y))|J_\Phi(x, y)|$ on B, and 0 elsewhere, where*

$$J_\Phi(x, y) = \det \begin{bmatrix} \partial_x \varphi_1(x, y) & \partial_y \varphi_1(x, y) \\ \partial_x \varphi_2(x, y) & \partial_y \varphi_2(x, y) \end{bmatrix} \tag{47}$$

is the Jacobian of $\Phi = (\varphi_1, \varphi_2)$ at (x, y). Suppose $A \cup B \subset \mathbb{D}_R$, the disc of radius R centered at $(0, 0)$. Let f and f_Φ denote the tomographic projections onto the x-axis of F and F_Φ, respectively. Then

$$\|f - f_\Phi\|_{V^p} \le \epsilon \cdot (4R)^{1-1/p} \cdot C(F, \Phi, p), \tag{48}$$

where

$$\epsilon = \max_{(x,y)\in B} \|(x, y) - \Phi(x, y)\|, \tag{49}$$

and

$$C(F, \Phi, p) = \min\{\|F\|_{L^p}, \|F_\Phi\|_{L^p}\}. \tag{50}$$

The proof of Theorem 4.3 may be found in Section 6.3.

Remark 7. Let $\theta \in [0, \pi]$, $c = \cos(\theta)$ and $s = \sin(\theta)$, and $\Phi(x, y) = (cx - sy, sx + cy)$. Then Theorem 4.3 gives the same bound as Theorem 4.2 when $p = 1$, but a weaker bound otherwise. For example, when $R = 1$, $p = \infty$, and $\|F\|_{L^\infty} = 1$, the bound of Theorem 4.2 is θ, whereas the bound from Theorem 4.3 is $8 \sin(\theta/2)$. However, Theorem 4.3 to a vastly larger set of functions Φ.

4.4 Convergence of the discrete norms

In this section, we consider the behavior of the discrete Volterra norms, defined in Section 3.3, for vectors consisting of samples of a function f on $[a, b]$ from an equispaced grid. Let n be a positive integer, and define $a_0 < a_1 < \cdots < a_n$ as in (32), namely

$$a_k = a + \frac{k}{n}(b - a), \quad 0 \le k \le n. \tag{51}$$

Note that $a_0 = a$ and $a_n = b$. Let $\mathbf{f}$ be the vector in $\mathbb{R}^{n+1}$ with entries $\mathbf{f}[k] = f(a_k)$, for $0 \le k \le n$.

Denote the mean of f on $[a, b]$ by

$$\mu(f) = \frac{1}{b - a} \int_a^b f(t)dt, \tag{52}$$

and let $f_{\text{cen}}(x) = f(x) - \mu(f)$.

If $\mathbf{w}$ is a vector in $\mathbb{R}^{n+1}$, let

$$\mu(\mathbf{w}) = \frac{1}{n} \sum_{k=1}^{n-1} \mathbf{w}[k] + \frac{1}{2n}\mathbf{w}[0] + \frac{1}{2n}\mathbf{w}[n] = \frac{1}{n} \sum_{0\le k\le n}^{\text{trap}} \mathbf{w}[k], \tag{53}$$

and let $\mathbf{w}_{\text{cen}}$ in $\mathbb{R}^{n+1}$ have entries $\mathbf{w}_{\text{cen}}[k] = \mathbf{w}[k] - \mu(\mathbf{w})$.

Theorem 4.4. *Suppose $a = c_0 < c_1 < \cdots < c_r = b$, and suppose f is Lipschitz continuous on each interval (c_j, c_{j+1}), $0 \le j \le r-1$, and continuous from either the right or left at each c_j, $0 \le j \le r$. Then for all $1 \le p \le \infty$, and*

$$\left| \|\mathbf{f}\|_{\nu_p} - \|f\|_{V^p} \right| \le \frac{C}{n}, \tag{54}$$

where $C > 0$ does not depend on n or p. The same bound holds by replacing f with f_{cen} and $\mathbf{f}$ with $\mathbf{f}_{\mathrm{cen}}$.

The proof of Theorem 4.4 may be found in Section 6.4. If instead of being merely piecewise Lipschitz, the function f is C^2 and not too oscillatory, then the discrete Volterra norms give a higher order approximation to the Volterra norms of f:

Theorem 4.5. *Suppose f is a two times continuously differentiable function on $[a, b]$, and has only finitely many roots. Then for all $1 \le p \le \infty$,*

$$\left| \|\mathbf{f}\|_{\nu_p} - \|f\|_{V^p} \right| \le \frac{C}{n^2}, \tag{55}$$

where $C > 0$ does not depend on n or p. The same bound holds by replacing f with f_{cen} and $\mathbf{f}$ with $\mathbf{f}_{\mathrm{cen}}$.

The proof of Theorem 4.5 may be found in Section 6.5.

4.5 Gaussian noise

In this section, we show that the discrete Volterra metrics are robust to additive noise. More precisely, as the number n of subintervals on which samples are taken grows, the effects of additive Gaussian noise with finite energy vanish at a predictable rate.

Theorem 4.6. *Let $\sigma_0, \sigma_2, \dots, \sigma_n, \dots$ be a bounded sequence of positive numbers, and let $Z = (Z[0], \dots, Z[n])$ where $Z[0], Z[1], \dots, Z[n], \dots$ are independent with $Z[j] \sim N(0, \sigma_j^2)$. Suppose too that $\sigma > 0$ satisfies*

$$\frac{1}{n} \sum_{j=1}^{n} \sigma_j^2 \le \sigma^2. \tag{56}$$

Let $t > 0$. Then for all $1 \le p \le \infty$,

$$\mathbb{P}\left\{ \|Z\|_{\nu_p} \ge t \right\} \le A e^{-Bt^2 n/\sigma^2}, \tag{57}$$

where $A > 0$ and $B > 0$ are constants independent of t, n, or p;

$$\lim_{n\to\infty} \|Z\|_{\nu_p} = 0 \tag{58}$$

almost surely; and

$$\mathbb{E}\|Z\|_{\nu_p} \le C \frac{\sigma}{\sqrt{n}}, \tag{59}$$

where $C > 0$ is a constant independent of n and p. Furthermore, (57), (58) *and* (59) *hold with Z replaced by Z_{cen}.*

Corollary 4.7. *Suppose f satisfies the conditions of Theorem 4.4, Z satisfies the conditions of Theorem 4.6, and $Y = \mathbf{f} + Z$. Let $t > 0$. Then for all $1 \le p \le \infty$,*

$$\mathbb{P}\left\{ \left| \|Y\|_{\nu_p} - \|f\|_{V^p} \right| \ge t \right\} \le A e^{-Bt^2 n/\sigma^2}, \tag{60}$$

where $A > 0$ and $B > 0$ are constants independent of t, n, or p;

$$\lim_{n\to\infty} \|Y\|_{\nu_p} = \|f\|_{V^p} \tag{61}$$

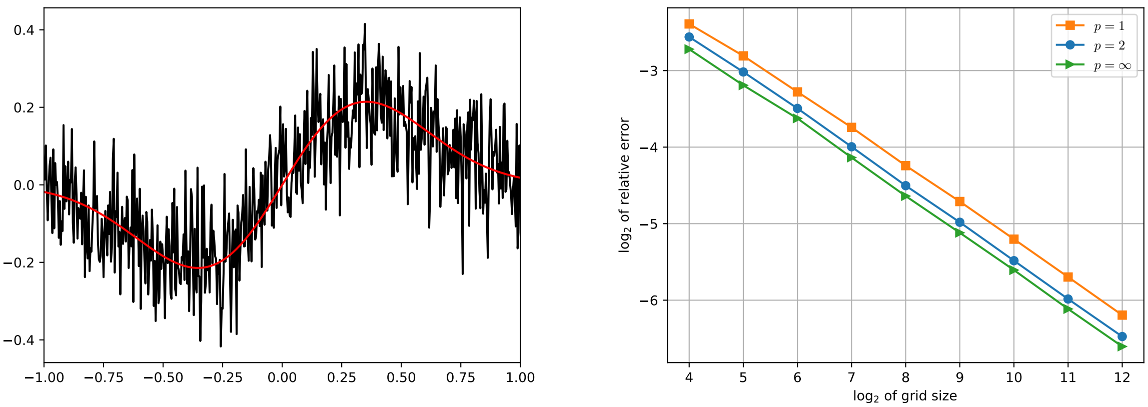

Figure 1: The first panel shows a realization of the noisy draws when $n = 512$, with the noiseless curve graphed in red. The second panel plots $\log_2(\text{err}_{n,p})$ against $\log_2(n)$, for $p = 1, 2, \infty$, where the number of draws is 5000. The slope of each curve is approximately $-1/2$, consistent with the error rate predicted by Corollary 4.7.

almost surely; and

$$\mathbb{E}\left|\|Y\|_{\nu_p} - \|f\|_{V^p}\right| \le C\frac{\sigma}{\sqrt{n}}, \tag{62}$$

where C is a constant independent of n and p. Furthermore, (60), (61) *and* (62) *hold with f replaced by f_{cen} and Y replaced by Y_{cen}.*

The proofs of Theorem 4.6 and Corollary 4.7 are provided in Section 6.6.

Remark 8. In the setting of Corollary 4.7, both the signal vector $\mathbf{f}$ and the noise vector Z have comparable p-norms; consequently, $\|Y\|_{\ell_p}$ does not approach $\|f\|_{L^p}$ as $n \to \infty$. For example, if $\sigma_j = \sigma$ for all j, then almost surely

$$\lim_{n\to\infty} \|Y\|_{\ell_2}^2 = \|f\|_{L^2}^2 + \sigma^2. \tag{63}$$

By contrast, (61) states the effect of the additive noise term Z on the Volterra norm is negligible when n is large.

5 Numerical results

5.1 Norm under noise

To demonstrate the robustness of the Volterra norms under noise described by Corollary 4.7, we run the following experiment. For different values of n, we take a vector $\mathbf{f}$ of $n+1$ equispaced samples from the function f on $[-1, 1]$ defined by

$$f(x) = xe^{-x^2/4}; \tag{64}$$

A vector Z of iid Gaussian noise with variance .01 is then added to each sample; let $Y = \mathbf{f} + Z$. A plot of a realization of Y, when $n = 512$, is shown in the left panel of Figure 1.

For $p = 1, 2, \infty$, we evaluate the norms $\|Y\|_{\nu_p}$. For each value of n, the experiment is repeated $M = 5000$ times. Denoting the M random signal-plus-noise vectors by $Y_1, \dots, Y_M$, we record the average absolute error:

$$\text{err}_{n,p} = \frac{1}{M}\sum_{k=1}^{M} \frac{\left|\|Y_k\|_{\nu_p} - \|f\|_{V^p}\right|}{\|f\|_{V^p}}, \tag{65}$$

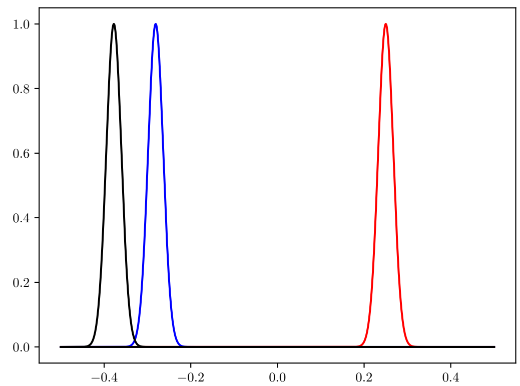

Figure 2: Three translations of the bump function (66) used in the experiment from Section 5.2.

where $\|f\|_{V^p}$ is computed using adaptive numerical integration. The right panel of Figure 1 plots $\log_2(\text{err}_{n,p})$ as a function of $\log_2(n)$. The average error decays like $O(1/\sqrt{n})$ as n increases, consistent with Corollary 4.7.

5.2 Distance under translation

We illustrate Theorem 4.1 on the functions shown in Figure 2; these are translations of the function f on $[-1/2, 1/2]$ defined by

$$f(x) = \cos(5x/2)e^{-10x^2}. \tag{66}$$

The top row of Figure 3 plots the estimated Volterra p-distances between the function and its translations as a function of the translation size for $p = 1, 2, \infty$, based on the trapezoidal rule approximation with $n = 500$ subintervals. The bottom row of Figure 3 plots the Lebesgue p-distances, $p = 1, 2, \infty$, between the function and its translations as a function of the translation size, using the same samples.

The Volterra distances exhibit the behavior described by the bound in Theorem 4.1, namely, the distances grow as concave functions of the translation size (the parameter ϵ). When $p = 1$ and $p = 2$, the distances continue to grow as ϵ grows, whereas when $p = \infty$ the distances level off, consistent with the upper bound from Theorem 4.1. By contrast, all of the Lebesgue distances quickly saturate to a constant value, independent of ϵ, as soon as the translation is big enough so that the numerical supports of the translated functions do not overlap.

5.3 Distance under dilation

We illustrate Theorem 4.1 on the function f defined on $[0, 2]$ by

$$f(x) = x(2 - x)e^{-10(2-x)^2}. \tag{67}$$

We consider the family of dilations of f parameterized by $\eta \geq 1$; these are the functions f_η defined by $f_\eta(x) = f(\eta x)\eta$ on $[0, 1/\eta]$, and $f_\eta(x) = 0$ elsewhere. The size ϵ of the dilation is

$$\epsilon = 2\left(1 - \frac{1}{\eta}\right). \tag{68}$$

Figure 4 shows the function f and two of its dilates.

The top row of Figure 5 plots the estimated Volterra p-distances between f and its dilates as a function of the dilation size for $p = 1, 2, \infty$, based on the trapezoidal rule approximation with $n = 5000$ subintervals. The

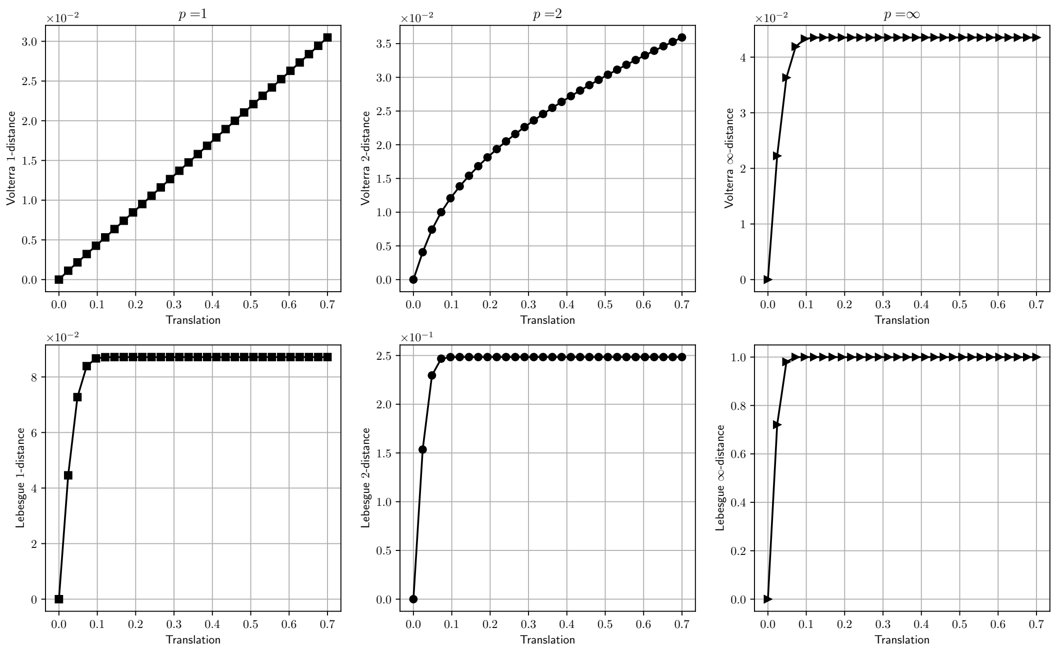

Figure 3: The first row shows the approximated Volterra distances (based on $n = 500$ subintervals) between the function (66) and its translates, as a function of the translation size. The second row shows the approximated Lebesgue distances between the function (66) and its translates. The values of p (from left to right) are $p = 1, 2, \infty$.

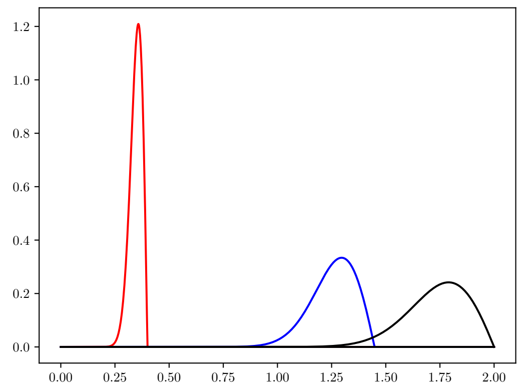

Figure 4: The function (67) is shown on the right, in black, and two of its dilates; these are used in the experiment from Section 5.3.

bottom row of Figure 5 plots the Lebesgue p-distances, $p = 1, 2, \infty$, between the function and its translations as a function of the translation size, using the same samples.

The Volterra distances exhibit the behavior described by the bound in Theorem 4.1: the distances grow as concave functions of the dilation size. Because the transformation preserves the integral of f, the L^1 distance levels off when the dilation size is big, since the supports of the function and its dilate are almost disjoint. By contrast, the L^2 and L^∞ distances grow rapidly for large dilation sizes (that is, they are convex functions of the dilation size). This is because the L^2 and L^∞ norms grow with the dilation size, and hence these distances reflect the size of the individual functions and not the relationship between the functions.

5.4 Distance between rotated projections

We illustrate the behavior described by Theorem 4.2 on the function F consisting of two Gaussian bumps

$$F(\mathbf{w}) = \exp\{-\|\mathbf{w} - \mathbf{a}\|^2/\sigma\} + \exp\{-\|\mathbf{w} - \mathbf{b}\|^2/\sigma\}, \tag{69}$$

where $\mathbf{a} = (0, 2/5)$ and $\mathbf{b} = (0, -2/5)$, and $\sigma = 1/2000$. Numerically, this function is supported in the disc $\mathbb{D}$ centered at $(0, 0)$ of radius $1/2$. We denote by f the projection of F onto the x-axis, and f_θ the projection of F after rotation by θ radians. Figure 6 shows a heatmap of F on the square $[-1/2, 1/2] \times [-1/2, 1/2]$.

Figure 7 plots the estimated Volterra and Lebesgue distances between f_θ and f, for $p = 1, 2, \infty$, as functions of θ between 0 and $\pi/2$, which covers the full range of distances due to F's symmetry. The distances are evaluated using $n = 5000$ subintervals. The Volterra 1 and 2 distances grow monotonically with θ throughout the entire range of values, where the Volterra ∞ distance and the Lebesgue distances plateau when θ is big enough so that the numerical supports of the projected bumps are disjoint.

5.5 Shrinking rings

The next experiment illustrates the behavior described by Theorem 4.3 on the function F consisting of Gaussian bumps

$$F(\mathbf{w}) = \sum_{k=0}^{6} h_k \exp\{-\|\mathbf{w} - \mathbf{a}_k\|^2/\sigma\}, \tag{70}$$

where $\sigma = 1/4000$, the centers $\mathbf{a}_k$ are equally spaced along a ring of radius $1/2$ and make angle $(2k/7 + \sqrt{2} + \sqrt{3} + \sqrt{5})\pi$ with the x-axis, and $h_k = (k+4)/7$. We plot the distance of F to the "shrunken" function

$$F_s(\mathbf{w}) = \sum_{k=0}^{6} h_k \exp\{-\|\mathbf{w} - (1-s)\mathbf{a}_k\|^2/\sigma\}, \tag{71}$$

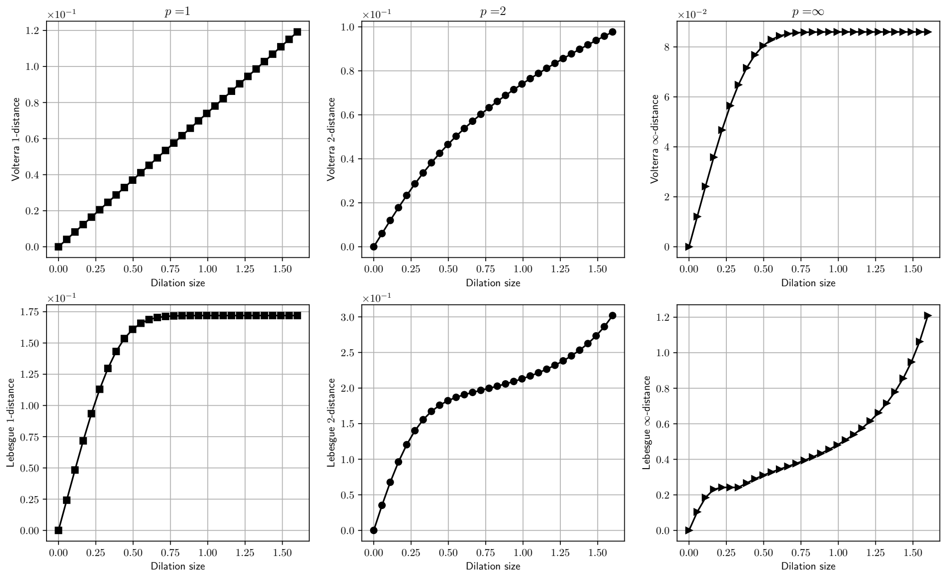

Figure 5: The first row shows the approximated Volterra distances (based on $n = 5000$ subintervals) between the function (66) and its dilates, as a function of the dilation size. The second row shows the approximated Lebesgue distances between the function (66) and its dilates. The values of p (from left to right) are $p = 1, 2, \infty$.

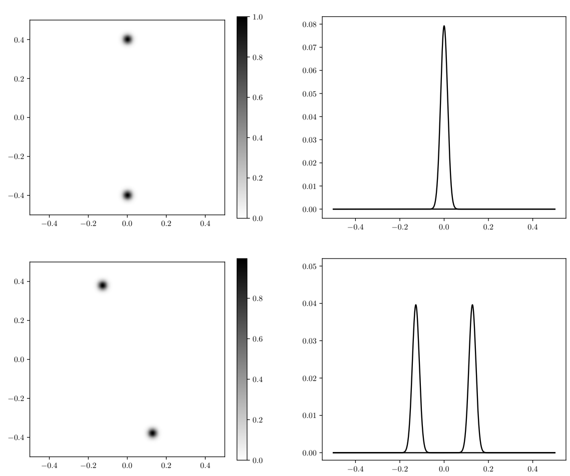

Figure 6: The function F (top left) and a rotation (bottom left), with their respective projections onto the x-axis. These are used in the experiment from Section 5.4.

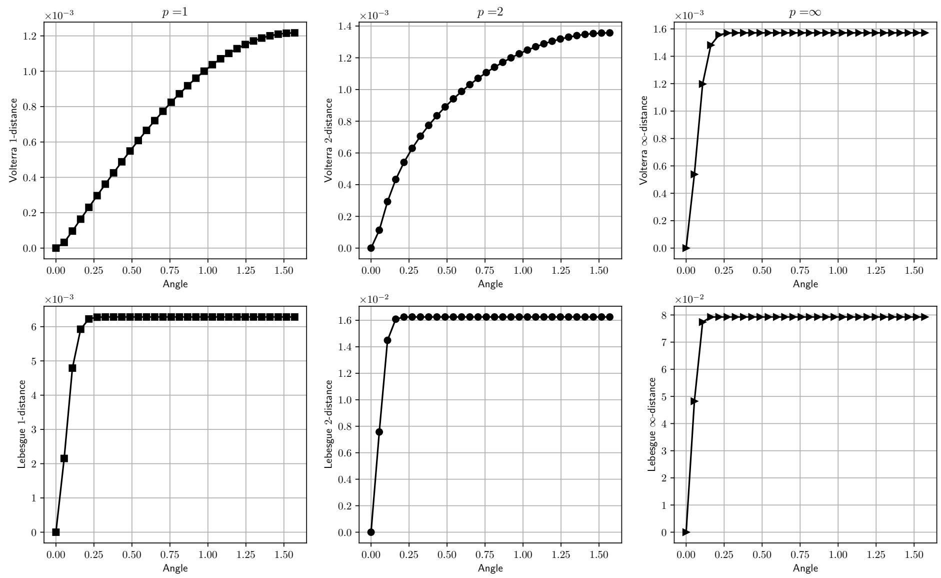

Figure 7: The first row shows the approximated Volterra distances (based on $n = 5000$ subintervals) between the projection of the two Gaussians at zero radians and at θ radians, as a function of θ. The second row shows the approximated Lebesgue distances. The values of p (from left to right) are $p = 1, 2, \infty$.

where s is a parameter between 0 and 1. Since the approximate support of F is the union of seven disjoint discs, the numerical support of F_s is related to that of F by moving of each of these discs towards the origin. Figure 8 shows F and an example of F_s, and their respective projections onto the x-axis.

Figure 9 plots the estimated Volterra and Lebesgue distances between F and F_s, for $p = 1, 2, \infty$, as functions of s. The distances are evaluated using $n = 5000$ subintervals. The Volterra 1 and 2 distances grow monotonically with s; the Volterra ∞ distance exhibits more irregular, though still monotonic behavior; and the Lebesgue distances are irregular and do not grow monotonically with the shrinkage parameter.

6 Proofs of the main results

6.1 Proof of Theorem 4.1

In this section, we will prove Theorem 4.1. We recall the statement of the theorem:

Theorem 6.1. *Suppose $f \in L^p$ is supported on an interval I. Let $\varphi : J \to I$ be continuously differentiable and monotonically increasing, with $I = \varphi(J)$. Let $f_\varphi(x) = f(\varphi(x))\varphi'(x)$ on J, and 0 otherwise. Suppose $I \cup J \subset [a, b]$. Then*

$$\|f - f_\varphi\|_{V^p} \leq \min\left\{\epsilon \cdot C(f, \varphi, p),\, \epsilon^{1/p} \cdot \|f\|_{L^1}\right\}, \tag{72}$$

where

$$\epsilon = \max_{x \in I} |x - \varphi(x)|, \tag{73}$$

and

$$C(f, \varphi, p) = \min\{\|f\|_{L^p}, \|f_\varphi\|_{L^p}\}. \tag{74}$$

We will use the following theorem, which is a special case of Theorem 6.18 in [21]:

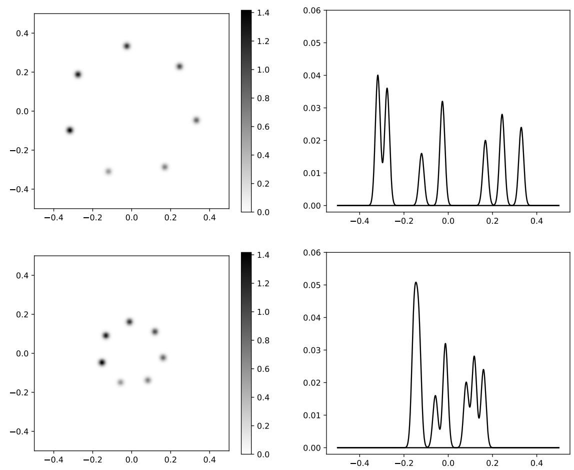

Figure 8: The function F (top left) and a shrunken version (bottom left), with their respective projections onto the x-axis. These are used in the experiment from Section 5.5.

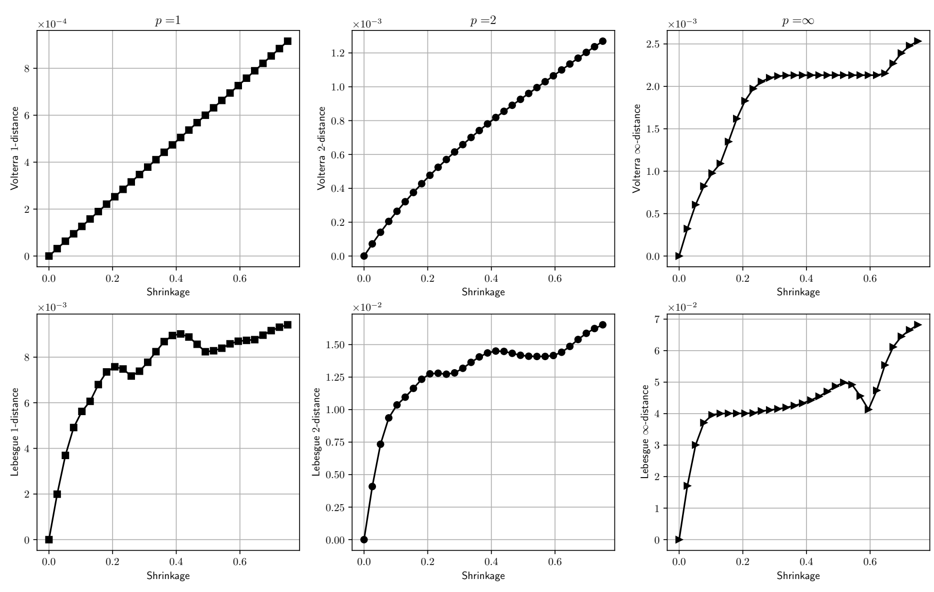

Figure 9: The first row shows the approximated Volterra distances (based on $n = 5000$ subintervals) between the projection of the function F from (70) and the shrunken function F_s from (71). The second row shows the approximated Lebesgue distances. The values of p (from left to right) are $p = 1, 2, \infty$.

Theorem 6.2. *Let $K = K(x,t)$ be a non-negative, integrable function on $I \times [a,b]$, that satisfies*

$$\int_I K(x,t)dx \le C \tag{75}$$

and

$$\int_a^b K(x,t)dt \le C, \tag{76}$$

where $C > 0$ is a constant. For a function g in $L^q([a,b])$, let

$$(Tg)(x) = \int_a^b K(x,t)g(t)dt. \tag{77}$$

Then $\|Tg\|_{L^q(I)} \le C \cdot \|g\|_{L^q([a,b])}$.

Let I_x be the interval $[x, \psi(x)]$ if $x \le \psi(x)$, or $[\psi(x), x]$ if $\psi(x) \le x$. Let $\chi(x,t)$ be 1 if and only if $t \in I_x$, and 0 otherwise. We then have the following lemma:

Lemma 6.3. *For all $x \in I$,*

$$\int_a^b \chi(x,t)dt \le \epsilon. \tag{78}$$

For all $t \in [a,b]$,

$$\int_I \chi(x,t)dx \le \epsilon. \tag{79}$$

Proof. For the first inequality

$$\int_a^b \chi(x,t)dt = \int_{I_x} 1dt = |I_x| = |x - \psi(x)| \le \epsilon. \tag{80}$$

For the second inequality, first suppose that there is some $x \le t$ with $t \in I_x$; note that for such x, $I_x = [x, \psi(x)]$, and so $x \le \psi(x)$. Let x^* be the smallest such x. Then $x^* \le t \le \psi(x^*)$. We claim that for all $x > t$, $t \notin I_x$. Indeed, since ψ is increasing and $x > t \ge x^*$, we have $\psi(x) > \psi(x^*) \ge t$. Since both $x > t$ and $\psi(x) > t$, t does not lie in I_x, as claimed.

Consequently, all x for which t lies in I_x are contained inside the interval $[x^*, t]$. Since $x^* \le t \le \psi(x^*)$ and $|x^* - \psi(x^*)| \le \epsilon$, it follows that $|t - x^*| \le \epsilon$ too. Furthermore, if $x > t$, then $\chi(x,t) = 0$ since $t \notin I_x$; and since x^* is the smallest x for which $t \in I_x$, if $x < x^*$ then $t \notin I_x$, hence $\chi(x,t) = 0$. Therefore,

$$\int_I \chi(x,t)dx \le \int_{x^*}^t 1dx = |t - x^*| \le \epsilon. \tag{81}$$

Analogous reasoning yields the same bound in the case that there exists $x \ge t$ with $t \in I_x$. □

Let $\psi : I \to J$ be the inverse of φ, $\psi = \varphi^{-1}$. We first remark that $(f_\varphi)_\psi = f$, and so without loss of generality we may assume that $C(f, \varphi, p) = \|f\|_{L^p}$, since we can then switch the roles of f and f_φ.

Let χ_I and χ_J be the indicator functions of I and J, respectively. Take any function G in $\mathcal{A}_0$, with $\|G\|_{L^q} = 1$; and let $g = G'$, so that

$$G(x) = -\int_x^b g(t)dt. \tag{82}$$

Then from Proposition 3.1, it is enough to show that

$$|\langle f_\varphi - f, G\rangle| \le \min\left\{\epsilon \cdot \|f\|_{L^p}, \epsilon^{1/p} \cdot \|f\|_{L^1}\right\}. \tag{83}$$

To that end, by performing a change of variables we may write

$$
\begin{aligned}
|\langle f_\varphi - f, G\rangle| &= \left|\int_a^b (f(\varphi(x))\varphi'(x) - f(x))G(x)\,dx\right| \\
&= \left|\int_a^b f(x)(G(\psi(x)) - G(x))\,dx\right| \\
&= \left|\int_I f(x)(G(\psi(x)) - G(x))\,dx\right|. \qquad (84)
\end{aligned}
$$

We will first prove

$$\left|\int_I f(x)(G(\psi(x)) - G(x))\,dx\right| \le \|f\|_{L^p} \cdot \epsilon, \qquad (85)$$

and then prove

$$\left|\int_I f(x)(G(\psi(x)) - G(x))\,dx\right| \le \|f\|_{L^1} \cdot \epsilon^{1/p}. \qquad (86)$$

6.1.1 Proof of (85)

Let q be the conjugate exponent of p, that is, $1/p + 1/q = 1$. By Hölder's inequality, we may bound the integral as follows:

$$\left|\int_I f(x)(G(\psi(x)) - G(x))\,dx\right| \le \|f\|_{L^p}\|(G\circ\psi - G)\chi_I\|_{L^q}. \qquad (87)$$

We will show that

$$\|(G\circ\psi - G)\chi_I\|_{L^q} \le \epsilon. \qquad (88)$$

Recall that I_x denotes the interval $[x, \psi(x)]$ if $x \le \psi(x)$, or $[\psi(x), x]$ if $\psi(x) \le x$; and $\chi(x,t)$ is 1 if and only if $t \in I_x$, and 0 otherwise. Then:

$$\left(\int_I |G(\psi(x)) - G(x)|^q dx\right)^{1/q} = \left(\int_I \left|\int_{I_x} g(t)dt\right|^q dx\right)^{1/q} = \left(\int_I \left|\int_a^b g(t)\chi(x,t)dt\right|^q dx\right)^{1/q}. \qquad (89)$$

We first assume $p > 1$, i.e. $q < \infty$. From Lemma 6.3 and Theorem 6.2,

$$\left(\int_I |G(\psi(x)) - G(x)|^q dx\right)^{1/q} = \left(\int_I \left|\int_a^b g(t)\chi(x,t)dt\right|^q dx\right)^{1/q} \le \epsilon\|g\|_{L^q} = \epsilon. \qquad (90)$$

which completes the proof of (88) when $p > 1$. We now handle the case $p = 1$, showing that:

$$\|(G\circ\psi - G)\chi_I\|_{L^\infty} \le \epsilon. \qquad (91)$$

For all $x \in I$ we have, with the same definition of I_x used previously,

$$|G(\psi(x)) - G(x)| = \left|\int_{I_x} g(t)dt\right| \le |I_x|\|g\|_{L^\infty} = |x - \psi(x)| \le \epsilon, \qquad (92)$$

completing the proof of (88), and hence of (85), for all $p \in [1, \infty]$.

6.1.2 Proof of (86)

We will now prove the bound (86), namely that for any G in $\mathcal{A}_0$ with $\|G'\|_{L^1} \leq 1$,

$$\int_I f(x)(G(\psi(x)) - G(x))dx \leq \|f\|_{L^1} \cdot \epsilon^{1/p}. \tag{93}$$

We have:

$$\int_I f(x)(G(\psi(x)) - G(x))dx \leq \|f\|_{L^1} \|(G \circ \psi - G)\chi_I\|_{L^\infty}, \tag{94}$$

and so it is enough to show

$$\|(G \circ \psi - G)\chi_I\|_{L^\infty} \leq \epsilon^{1/p}. \tag{95}$$

As before, let I_x be the interval $[x, \psi(x)]$ if $x \leq \psi(x)$, or $[\psi(x), x]$ if $\psi(x) \leq x$, and let $\chi(x,t)$ be 1 if and only if $t \in I_x$, and 0 otherwise. Using that $G(x) = -\int_x^b g(t)dt$, we may write

$$\begin{aligned} \|(G \circ \psi - G)\chi_I\|_{L^\infty} &= \max_{x \in I} |G(\psi(x)) - G(x)| \\ &= \max_{x \in I} \left| \int_{I_x} g(t)dt \right|, \end{aligned} \tag{96}$$

and for all x in I, Hölder's inequality yields

$$\begin{aligned} \left| \int_{I_x} g(t)dt \right| &= \left| \int_a^b g(t)\chi(x,t)dt \right| \\ &\leq \|g\|_{L^q} \left(\int_a^b \chi(x,t)^p dt \right)^{1/p} \\ &= \left(\int_a^b \chi(x,t)dt \right)^{1/p} \\ &\leq \epsilon^{1/p}, \end{aligned} \tag{97}$$

where the last inequality follows from Lemma 6.3. This completes the proof.

6.2 Proof of Theorem 4.2

We recall the statement of Theorem 4.2:

Theorem 6.4. *For all angles θ and φ with $|\theta - \varphi| < \pi$ and all $p \in [1, \infty]$,*

$$\begin{aligned} \|f_\theta - f_\varphi\|_{V^p} &\leq (2\sin(|\theta - \varphi|/2))^{1/p} \cdot |\theta - \varphi|^{1-1/p} \cdot \|F\|_{L^p} \cdot R^{2-1/p} \\ &\leq |\theta - \varphi| \cdot \|F\|_{L^p} \cdot R^{2-1/p}, \end{aligned} \tag{98}$$

where $\|F\|_{L^p}$ denotes the p-norm of F over $\mathbb{D}$.

Without loss of generality, we assume that $\varphi = 0$, and that $c = \cos(\theta)$ and $s = \sin(\theta)$ are both positive. For brevity, let $f = f_\varphi = f_0$.

6.2.1 Proof of Theorem 4.2 for general R from $R = 1$

We will prove the result for any $R > 0$ assuming that it is true when $R = 1$. To see how the general result follows from this case, given $F : \mathbb{D}_R \to \mathbb{R}$ define $\widetilde{F} : \mathbb{D}_1 \to \mathbb{R}$ by $\widetilde{F}(x,y) = F(Rx, Ry)$, and $\widetilde{f}_\theta : [-1,1] \to \mathbb{R}$ by

$$\widetilde{f}_\theta(x) = \int_{-1}^1 \widetilde{F}(cx + sy, cy - sx)dy, \tag{99}$$

and $\widetilde{f} = \widetilde{f}_0$. Then the result for $R = 1$ states that

$$\|\widetilde{f} - \widetilde{f}_\theta\|_{V^p([-1,1])} \le C_{\theta,p}\|\widetilde{F}\|_{L^p(\mathbb{D}_1)}, \tag{100}$$

where $C_{\theta,p} = (2\sin(\theta/2))^{1/p}\theta^{1-1/p}$. Performing a change of variables $u = Ry$, we find

$$\begin{aligned}
\widetilde{f}_\theta(x) &= \int_{-1}^{1} \widetilde{F}(cx + sy, cy - sx)dy \\
&= \frac{1}{R}\int_{-R}^{R} \widetilde{F}(cx + s(u/R), c(u/R) - sx)du \\
&= \frac{1}{R}\int_{-R}^{R} \widetilde{F}((cRx + su)/R, (cu - sRx)/R)du \\
&= \frac{1}{R}\int_{-R}^{R} F(Rx + su, cu - sRx)du \\
&= \frac{1}{R} f_\theta(Rx). \qquad (101)
\end{aligned}$$

Consequently, by another change of variables $u = Rt$,

$$\begin{aligned}
\int_{-1}^{x} (\widetilde{f}(t) - \widetilde{f}_\theta(t))dt &= \frac{1}{R}\int_{-R}^{Rx} (\widetilde{f}(u/R) - \widetilde{f}_\theta(u/R))du \\
&= \frac{1}{R^2}\int_{-R}^{Rx} (f(u) - f_\theta(u))du. \qquad (102)
\end{aligned}$$

Therefore, for all $1 \le p < \infty$, letting $v = Rx$,

$$\begin{aligned}
\|\widetilde{f} - \widetilde{f}_\theta\|^p_{V^p([-1,1])} &= \int_{-1}^{1} \left|\int_{-1}^{x} (\widetilde{f}(t) - \widetilde{f}_\theta(t))dt\right|^p dx \\
&= \int_{-1}^{1} \left|\frac{1}{R^2}\int_{-R}^{Rx} (f(u) - f_\theta(u))du\right|^p dx \\
&= \frac{1}{R}\int_{-R}^{R} \left|\frac{1}{R^2}\int_{-R}^{v} (f(u) - f_\theta(u))du\right|^p dv \\
&= \frac{1}{R^{2p+1}}\int_{-R}^{R} \left|\int_{-R}^{v} (f(u) - f_\theta(u))du\right|^p dv \\
&= \frac{1}{R^{2p+1}}\|f - f_\theta\|^p_{V^p([-R,R])}, \qquad (103)
\end{aligned}$$

and so

$$\|\widetilde{f} - \widetilde{f}_\theta\|_{V^p([-1,1])} = \frac{1}{R^{2+1/p}}\|f - f_\theta\|_{V^p([-R,R])}. \tag{104}$$

On the other hand,

$$\begin{aligned}
\|\widetilde{F}\|^p_{L^p(\mathbb{D}_1)} &= \int_{\mathbb{D}_1} |\widetilde{F}(x,y)|^p dxdy \qquad (105) \\
&= \int_{\mathbb{D}_1} |F(Rx,Ry)|^p dxdy \qquad (106) \\
&= \frac{1}{R^2}\int_{\mathbb{D}_R} |F(u,v)|^p dudv \qquad (107) \\
&= \frac{1}{R^2}\|F\|^p_{L^p(\mathbb{D}_R)}, \qquad (108)
\end{aligned}$$

and so

$$\|\widetilde{F}\|_{L^p(\mathbb{D}_1)} = \frac{1}{R^{2/p}}\|F\|_{L^p(\mathbb{D}_R)}. \tag{109}$$

Combining (100), (104) and (109) gives

$$\frac{1}{R^{2+1/p}}\|f - f_\theta\|_{V^p([-R,R])} \leq C_{\theta,p}\frac{1}{R^{2/p}}\|F\|_{L^p(\mathbb{D}_R)}, \tag{110}$$

or equivalently,

$$\|f - f_\theta\|_{V^p([-R,R])} \leq (2\sin(\theta/2))^{1/p}\theta^{1-1/p}\|F\|_{L^p(\mathbb{D}_R)}R^{2-1/p}, \tag{111}$$

as claimed. The proof for $p = \infty$ can be proved similarly, or by taking the limit as $p \to \infty$ and using that $\|\cdot\|_{L^\infty} = \lim_{p\to\infty}\|\cdot\|_{L^p}$.

6.2.2 Proof of Theorem 4.2 when $R = 1$

We start with two lemmas. Let $I_{x,y}$ be the interval $[x, cx+sy]$ when $x \leq cx+sy$, and the interval $[cx+sy, x]$ when $cx + sy \leq x$; and let $\chi(x, y, t)$ be 1 if $t \in I_{x,y}$ and 0 otherwise. Let $\mathbb{D} = \mathbb{D}_1$.

Lemma 6.5. *For all $(x, y) \in \mathbb{D}$,*

$$\int_{-1}^{1} \chi(x, y, t)dt \leq 2\sin(\theta/2). \tag{112}$$

Proof. We apply the Cauchy-Schwarz inequality, the fact that $\sqrt{x^2 + y^2} \leq 1$, and the identity $c = \cos(\theta) = \cos^2(\theta/2) - \sin^2(\theta/2)$, to get the following:

$$\begin{aligned}\int_{-1}^{1} \chi(x, y, t)dt &= |I_{x,y}| \\ &= |x - cx - sy| \\ &= |(1 - c)x - sy| \\ &\leq \sqrt{(1 - c)^2 + s^2} \\ &= \sqrt{2}\sqrt{1 - c} \\ &= \sqrt{2}\sqrt{1 - \cos^2(\theta/2) + \sin^2(\theta/2)} \\ &= \sqrt{2}\sqrt{2\sin^2(\theta/2)} \\ &= 2\sin(\theta/2), \end{aligned} \tag{113}$$

as claimed. □

Lemma 6.6. *For all $t \in [-1, 1]$,*

$$\int_{\mathbb{D}} \chi(x, y, t)\, dx\, dy \leq \theta. \tag{114}$$

We defer the proof of Lemma 6.6 to Section 6.2.3. Assuming Lemma 6.6, let us see how to prove the result. First, suppose $p = 1$. Take any absolutely continuous function G on $[-1, 1]$ with $\|G'\|_{L^\infty} \leq 1$. Then we must bound the integral

$$\int_{-1}^{1} (f(x) - f_\theta(x))G(x)\, dx. \tag{115}$$

We have

$$\int_{-1}^{1} f_\theta(x)G(x)\,dx = \int_{-1}^{1}\int_{-1}^{1} F(cx-sy, cy+sx)G(x)\,dx\,dy$$
$$= \int_{\mathbb{D}} F(cx-sy, cy+sx)G(x)\,dx\,dy, \tag{116}$$

since F is supported on $\mathbb{D}$. Applying a change of variables gives:

$$\int_{-1}^{1} f_\theta(x)G(x)\,dx = \int_{\mathbb{D}} F(cx-sy, cy+sx)G(x)\,dx\,dy$$
$$= \int_{\mathbb{D}} F(x,y)G(cx+sy)\,dx\,dy. \tag{117}$$

Consequently,

$$\int_{-1}^{1} (f(x)-f_\theta(x))G(x)\,dx = \int_{\mathbb{D}} F(x,y)(G(x)-G(cx+sy))\,dx\,dy$$
$$\leq \|F\|_{L^1} \sup_{(x,y)\in\mathbb{D}} |G(x)-G(cx+sy)|. \tag{118}$$

Let $g = G'$; then $\|g\|_{L^\infty} \leq 1$, and we have:

$$\sup_{(x,y)\in\mathbb{D}} |G(x)-G(cx+sy)| = \sup_{(x,y)\in\mathbb{D}} \left|\int_{I_{x,y}} g(t)\,dt\right|$$
$$= \sup_{(x,y)\in\mathbb{D}} \left|\int_{-1}^{1} g(t)\chi(x,y,t)\,dt\right|.$$
$$\leq \|g\|_{L^\infty} \sup_{(x,y)\in\mathbb{D}} \left|\int_{-1}^{1} \chi(x,y,t)\,dt\right|.$$
$$\leq 2\sin(\theta/2), \tag{119}$$

where we have used Lemma 6.5. Consequently,

$$\|f - f_\theta\|_{V^1} \leq 2\sin(\theta/2)\|F\|_{L^1}. \tag{120}$$

Next, we prove the result when $p = \infty$. Take any absolutely continuous function G on $[-1,1]$ with $\|G'\|_{L^1} \leq 1$. Take $g = G'$, so that $\|g\|_{L^1} \leq 1$. Then, just as before, we have

$$\int_{-1}^{1} (f(x)-f_\theta(x))G(x)\,dx = \int_{\mathbb{D}} F(x,y)(G(x)-G(cx+sy))\,dx\,dy$$
$$\leq \|F\|_{L^\infty} \int_{\mathbb{D}} |G(x)-G(cx+sy)|\,dx\,dy$$
$$= \int_{\mathbb{D}} \left|\int_{I_{x,y}} g(t)dt\right|\,dx\,dy$$
$$= \int_{\mathbb{D}} \left|\int_{-1}^{1} g(t)\chi(x,y,t)dt\right|\,dx\,dy$$
$$\leq \int_{-1}^{1} |g(t)| \int_{\mathbb{D}} \chi(x,y,t)\,dx\,dy\,dt$$
$$\leq \|g\|_{L^1} \sup_{|t|\leq 1} \int_{\mathbb{D}} \chi(x,y,t)\,dx\,dy$$
$$\leq \theta, \tag{121}$$

where we have used Lemma 6.6. Consequently,

$$\|f - f_\theta\|_{V^\infty} \leq \theta \|F\|_{L^\infty}. \tag{122}$$

Since the mapping from F to $\int_{-1}^{x}(f - f_\theta)$ is linear, the Riesz-Thorin Interpolation Theorem (see, e.g. Theorem 6.27 in [21]) then implies that if F is in $L^p(\mathbb{D})$,

$$\|f - f_\theta\|_{V^p} \leq (2\sin(\theta/2))^{1/p}\theta^{1-1/p}\|F\|_{L^p}, \tag{123}$$

as claimed.

We now turn to the proof of Lemma 6.6.

6.2.3 Proof of Lemma 6.6

First, observe that

$$\int_{\mathbb{D}} \chi(x, y, t)dxdy = 2|\mathbf{S}_{t,(1,0),(x,y)} \cup \mathbf{S}_{t,(x,y),(1,0)}|, \tag{124}$$

where, for unit vectors $\mathbf{v}$ and $\mathbf{w}$, $\mathbf{S}_{t,\mathbf{v},\mathbf{w}}$ is the region defined by

$$\mathbf{S}_{t,\mathbf{v},\mathbf{w}} = \{\mathbf{u} \in \mathbb{D} : \langle \mathbf{u}, \mathbf{v}\rangle \leq t \leq \langle \mathbf{u}, \mathbf{w}\rangle\}. \tag{125}$$

By rotational symmetry, the following lemma is immediate:

Lemma 6.7. *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are any unit vectors with angle* θ*, then* $|\mathbf{S}_{t,(1,0),(x,y)}| = |\mathbf{S}_{t,\mathbf{a},\mathbf{b}}|$*. Furthermore,* $|\mathbf{S}_{t,\mathbf{a},\mathbf{b}}| = |\mathbf{S}_{-t,\mathbf{a},\mathbf{b}}|$*, and* $|\mathbf{S}_{t,\mathbf{a},\mathbf{b}} \cap \mathbf{S}_{-t,\mathbf{a},\mathbf{b}}| = 0$*.*

By this lemma, it follows that

$$\int_{\mathbb{D}} \chi(x, y, t)dxdy = 2|\mathbf{S}_{t,\mathbf{v},\mathbf{w}}| = 2\left|\{\mathbf{u} \in \mathbb{D} : \langle \mathbf{u}, \mathbf{v}\rangle \leq t \leq \langle \mathbf{u}, \mathbf{w}\rangle\}\right|, \tag{126}$$

where $\mathbf{w} = (\cos(\theta/2), \sin(\theta/2))$ and $\mathbf{v} = (\cos(\theta/2), -\sin(\theta/2))$. It will be convenient to refer to Figure 10, where $\mathbf{w}$ corresponds to the point labeled B, and $\mathbf{v}$ corresponds to the point labeled E. In the figure, the line AD is perpendicular to OB, and intersects OB at distance t from the origin; consequently, the set of all vectors $\mathbf{u}$ in $\mathbb{D}$ with $\langle \mathbf{u}, \mathbf{w}\rangle \geq t$ is the circular segment through the points A, B and D. Similarly; the line CF is perpendicular to OE, and intersects OE at distance t from the origin; consequently, the set of all vectors $\mathbf{u}$ in $\mathbb{D}$ with $\langle \mathbf{u}, \mathbf{v}\rangle \leq t$ is the circular segment through the points C, A and F. The intersection of these two circular segments is the region bounded by A, C and G.

To evaluate the area of this region, we will first find the area of the full circular segment through A, B and D, and then subtract off the area of the region bounded by C, G and D.

Lemma 6.8. *The area of the circular segment through* A*,* B *and* D *is*

$$\arccos(t) - t\sqrt{1 - t^2}, \tag{127}$$

where arccos *takes values in* $[0, \pi]$*.*

Proof. This is immediate from the well-known formula for the area of a circular segment, and the fact that the line segment from O to H has length t. ☐

The next lemma is also elementary, and likely known already; however, since we could not find the exact identity in the literature, we provide a self-contained proof.

Lemma 6.9. *When* $t \leq \cos(\theta/2)$*, the intersection between the circular segment bounded by* A*,* B *and* D *and the circular segment bounded by* C*,* E *and* F *has area*

$$\arcsin\left(\sqrt{1 - t^2}\cos(\theta/2) - t\sin(\theta/2)\right) - t\sqrt{1 - t^2} + t^2\tan(\theta/2), \tag{128}$$

where arcsin *takes values in* $[-\pi/2, \pi/2]$*. When* $t > \cos(\theta/2)$*, the two circular segments are disjoint.*

Proof. We begin by showing the second part, namely that when $t > \cos(\theta/2)$, the circular segments are disjoint, or equivalently that the point G lies outside of the circle. Indeed, it is straightforward to show that G is located at the point $(t/\cos(\theta/2), 0)$; hence, G is inside the circle so long as $t/\cos(\theta/2) \leq 1$, or equivalently, $t \leq \cos(\theta/2)$, as desired.

Let us now suppose that $t \leq \cos(\theta/2)$, and evaluate the area of the region bounded by C, G, and D. The line segment from G to D has arc-length parameterization

$$\alpha(s) = t(\cos(\theta/2), \sin(\theta/2)) + s(\sin(\theta/2), -\cos(\theta/2)), \tag{129}$$

and the line segment from C to G has arc-length parameterization

$$\beta(s) = t(\cos(\theta/2), -\sin\theta/2) + (\sqrt{1-t^2} + t\cdot\tan(\theta/2) - s)(\sin(\theta/2), \cos(\theta)/2), \tag{130}$$

where

$$t\cdot\tan(\theta/2) \leq s \leq \sqrt{1-t^2}. \tag{131}$$

The counterclockwise arc from D to C has arc-length parameterization

$$\gamma(\varphi) = (\cos(\varphi), \sin(\varphi)), \tag{132}$$

where

$$-\arcsin\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right) \leq \varphi \leq \arcsin\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right). \tag{133}$$

When $t \leq \cos(\theta/2)$, we will evaluate the area using Green's Theorem, by computing $\frac{1}{2}\oint(xdy - ydx)$ over each curve. For α, we have

$$\begin{aligned}\frac{1}{2}\oint_\alpha xdy &= \frac{1}{2}\int_{t\cdot\tan(\theta/2)}^{\sqrt{1-t^2}} \left[(t\cdot\cos(\theta/2) + s\cdot\sin(\theta/2))(-\cos(\theta/2))\right] ds \\ &= \frac{t\cdot\cos^2(\theta/2)}{2}\left(t\cdot\tan(\theta/2) - \sqrt{1-t^2}\right) + \frac{\sin(\theta/2)\cos(\theta/2)}{4}\left(t^2\cdot\tan^2(\theta/2) - 1 + t^2\right),\end{aligned} \tag{134}$$

and

$$\begin{aligned}\frac{1}{2}\oint_\alpha ydx &= \frac{1}{2}\int_{t\cdot\tan(\theta/2)}^{\sqrt{1-t^2}} \left[(t\cdot\sin(\theta/2) - s\cdot\cos(\theta/2))(\sin(\theta/2))\right] ds \\ &= \frac{t\cdot\sin^2(\theta/2)}{2}\left(\sqrt{1-t^2} - t\cdot\tan(\theta/2)\right) + \frac{\sin(\theta/2)\cos(\theta/2)}{4}(t^2\cdot\tan^2(\theta/2) - 1 + t^2),\end{aligned} \tag{135}$$

and hence

$$\frac{1}{2}\oint_\alpha(xdy - ydx) = \frac{t}{2}\cdot\left(t\cdot\tan(\theta/2) - \sqrt{1-t^2}\right). \tag{136}$$

Similarly,

$$\frac{1}{2}\oint_\beta(xdy - ydx) = \frac{t}{2}\cdot\left(t\cdot\tan(\theta/2) - \sqrt{1-t^2}\right). \tag{137}$$

Finally, it is straightforward to check that

$$\frac{1}{2}\oint_\gamma(xdy - ydx) = \arcsin\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right). \tag{138}$$

Adding all three integrals together, we find that the area of the region is

$$\frac{1}{2}\oint_\gamma(xdy - ydx) = \arcsin\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right) - t\sqrt{1-t^2} + t^2\tan(\theta/2), \tag{139}$$

as claimed. □

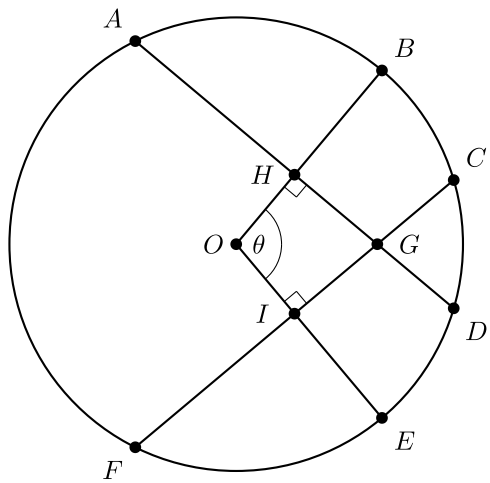

Figure 10: Diagram for the proof of Lemma 6.6. The points labeled B and C are located at $(\cos(\theta/2), \sin(\theta/2))$ and $(\cos(\theta/2), -\sin(\theta/2))$, respectively. The point labeled O is the origin, $(0, 0)$. The line segment OB is orthogonal to the line AD, and the line segment OE is orthogonal to the line FC. The line segments OH and OI each have length t.

From Lemmas 6.8 and 6.9, we find

$$\begin{aligned}&\frac{1}{2}\int_{\mathbb{D}} \chi(x, y, t)dxdy \\ &= \begin{cases} \arccos(t) - t\sqrt{1-t^2} & \text{if } t > \cos(\theta/2); \\ \arccos(t) - \arcsin\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right) - t^2\tan(\theta/2), & \text{if } t \le \cos(\theta/2). \end{cases}\end{aligned} \tag{140}$$

To conclude the proof, we must show that this expression is bounded above by $\theta/2$ for all values of t between 0 and 1. In fact, we will show that (140) is a decreasing function of t, and hence is maximized at $t = 0$. It is immediately apparent that the expression is decreasing in t when $t > \cos(\theta/2)$, since this is the area of the circular segment with chord at distance t from the origin. When $t \le \cos(\theta)$, we first observe that

$$\begin{aligned}&\frac{d}{dt}\arcsin\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right) \\ &= \frac{\frac{d}{dt}\left[\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right]}{\sqrt{1-(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2))^2}} \\ &= \frac{-t(1-t^2)^{-1/2}\cos(\theta/2) - \sin(\theta/2)}{\sqrt{1-\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right)^2}},\end{aligned} \tag{141}$$

and the square of the denominator may be written more simply as

$$\begin{aligned}&1-(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2))^2 \\ &= 1-(1-t^2)\cos^2(\theta/2) - t^2\sin^2(\theta/2) + 2t\sqrt{1-t^2}\cos(\theta/2)\sin(\theta/2) \\ &= 1-\cos^2(\theta/2) + t^2\cos^2(\theta/2) - t^2\sin^2(\theta/2) + 2t\sqrt{1-t^2}\cos(\theta/2)\sin(\theta/2) \\ &= \sin^2(\theta/2) + t^2\cos^2(\theta/2) - t^2\sin^2(\theta/2) + 2t\sqrt{1-t^2}\cos(\theta/2)\sin(\theta/2) \\ &= t^2\cos^2(\theta/2) + (1-t^2)\sin^2(\theta/2) + 2t\sqrt{1-t^2}\cos(\theta/2)\sin(\theta/2) \\ &= \left(t\cdot\cos(\theta/2) + \sqrt{1-t^2}\sin(\theta/2)\right)^2;\end{aligned} \tag{142}$$

consequently, because $t \cdot \cos(\theta/2) + \sqrt{1-t^2}\sin(\theta/2) > 0$,

$$\begin{aligned}\frac{d}{dt}\arcsin\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right) &= \frac{-t(1-t^2)^{-1/2}\cos(\theta/2) - \sin(\theta/2)}{t\cdot\cos(\theta/2) + \sqrt{1-t^2}\sin(\theta/2)} \\ &= \frac{-t(1-t^2)^{-1/2}\cos(\theta/2) - \sin(\theta/2)}{t\cdot\cos(\theta/2) + \sqrt{1-t^2}\sin(\theta/2)} \\ &= \frac{-t}{\sqrt{1-t^2}}. \end{aligned} \tag{143}$$

Therefore,

$$\begin{aligned}&\frac{d}{dt}\left[\arccos(t) - \arcsin\left(\sqrt{1-t^2}\cos(\theta/2) - t\cdot\sin(\theta/2)\right) - t^2\tan(\theta/2)\right] \\ =& \frac{-1}{\sqrt{1-t^2}} + \frac{1}{\sqrt{1-t^2}} - 2t\tan(\theta/2) \\ =& -2t\tan(\theta/2), \end{aligned} \tag{144}$$

which is negative. Therefore, the maximum value of $\int_{\mathbb{D}} \chi(x,y,t)dxdy$ occurs when $t = 0$, where the value is

$$\begin{aligned} 2\arccos(0) - 2\arcsin(\cos(\theta/2)) &= \pi - 2\arcsin(\sin(\pi/2 + \theta/2)) \\ &= \pi - 2\arcsin(\sin(\pi/2 - \theta/2)) \\ &= \pi - 2\left(\frac{\pi}{2} - \frac{\theta}{2}\right) \\ &= \theta; \end{aligned} \tag{145}$$

note that arcsin takes values in $[-\pi/2, \pi/2]$, and $\pi/2 - \theta/2$ lies between 0 and $\pi/2$ since θ is between 0 and π. This completes the proof.

6.3 Proof of Theorem 4.3

We recall the statement of the theorem:

Theorem 6.10. *Suppose $F : \mathbb{R}^2 \to \mathbb{R}$ is in L^p, and is supported on a compact set A. Let $\Phi : B \to A$ be a continuously differentiable, one-to-one mapping, and let $F_\Phi(x,y) = F(\Phi(x,y))|J_\Phi(x,y)|$ on B, and 0 elsewhere, where*

$$J_\Phi(x,y) = \det\begin{bmatrix} \partial_x\varphi_1(x,y) & \partial_y\varphi_1(x,y) \\ \partial_x\varphi_2(x,y) & \partial_y\varphi_2(x,y) \end{bmatrix} \tag{146}$$

is the Jacobian of $\Phi = (\varphi_1, \varphi_2)$ at (x,y). Suppose $A \cup B \subset \mathbb{D}_R$, the disc of radius R centered at $(0,0)$. Let f and f_Φ denote the tomographic projections onto the x-axis of F and F_Φ, respectively. Then

$$\|f - f_\Phi\|_{V^p} \le \epsilon \cdot (4R)^{1-1/p} \cdot C(F,\Phi,p), \tag{147}$$

where

$$\epsilon = \max_{(x,y)\in B} \|(x,y) - \Phi(x,y)\|, \tag{148}$$

and

$$C(F,\Phi,p) = \min\{\|F\|_{L^p}, \|F_\Phi\|_{L^p}\}. \tag{149}$$

6.3.1 Proof of Theorem 4.3 for general R from $R = 1$

We will first prove the result for any $R > 0$, assuming that it is true when $R = 1$. Given $F : \mathbb{D}_R \to \mathbb{R}$, define $\widetilde{F} : \mathbb{D}_1 \to \mathbb{R}$ by $\widetilde{F}(x, y) = F(Rx, Ry)$, and $\widetilde{f} : [-1, 1] \to \mathbb{R}$ by

$$\widetilde{f}(x) = \int_{-1}^{1} \widetilde{F}(x, y)\, dy, \tag{150}$$

and similarly define $\widetilde{f}_\Phi$ by

$$\widetilde{f}_\Phi(x) = \int_{-1}^{1} \widetilde{F}_\Phi(x, y)\, dy. \tag{151}$$

Let $\widetilde{A} = \{(x/R, y/R) : (x, y) \in A\}$, $\widetilde{B} = \{(x/R, y/R) : (x, y) \in B\}$, and define $\widetilde{\Phi} : \widetilde{B} \to \widetilde{A}$ by $\widetilde{\Phi}(x, y) = \Phi(Rx, Ry)/R$ and $\widetilde{\Psi} : \widetilde{A} \to \widetilde{B}$ by $\widetilde{\Psi} = \widetilde{\Phi}^{-1}$.

Note that, if $\widetilde{\Phi} = (\widetilde{\varphi}_1, \widetilde{\varphi}_2)$, so that $\widetilde{\varphi}_j(x, y) = \varphi_j(Rx, Ry)/R$, $j = 1, 2$, then

$$\begin{aligned}
J_{\widetilde{\Phi}}(x, y) &= \partial_x \widetilde{\varphi}_1(x, y)\partial_y \widetilde{\varphi}_2(x, y) - \partial_y \widetilde{\varphi}_1(x, y)\partial_x \widetilde{\varphi}_2(x, y) \\
&= \frac{1}{R^2}\left[\partial_x \varphi_1(Rx, Ry)\partial_y \varphi_2(Rx, Ry) - \partial_y \varphi_1(Rx, Ry)\partial_x \varphi_2(Rx, Ry)\right] \\
&= \frac{1}{R^2}\left[R(\partial_x \varphi_1)(Rx, Ry) \cdot R(\partial_y \varphi_2)(Rx, Ry) - R(\partial_y \varphi_1)(Rx, Ry) \cdot R(\partial_x \varphi_2)(Rx, Ry)\right] \\
&= \left[(\partial_x \varphi_1)(Rx, Ry) \cdot (\partial_y \varphi_2)(Rx, Ry) - (\partial_y \varphi_1)(Rx, Ry) \cdot (\partial_x \varphi_2)(Rx, Ry)\right] \\
&= J_\Phi(Rx, Ry).
\end{aligned} \tag{152}$$

Consequently,

$$\begin{aligned}
\widetilde{F}_\Phi(x, y) &= F_\Phi(Rx, Ry) \\
&= F(\Phi(Rx, Ry))|J_\Phi(Rx, Ry)| \\
&= F(R\Phi(Rx, Ry)/R)|J_{\widetilde{\Phi}}(x, y)| \\
&= F(R\widetilde{\Phi}(x, y))|J_{\widetilde{\Phi}}(x, y)| \\
&= \widetilde{F}(\widetilde{\Phi}(x, y))|J_{\widetilde{\Phi}}(x, y)|.
\end{aligned} \tag{153}$$

Also, let

$$\begin{aligned}
\widetilde{\epsilon} &= \max_{(x,y)\in \widetilde{B}} \|(x, y) - \widetilde{\Phi}(x, y)\| \\
&= \max_{(x,y)\in \widetilde{B}} \|(Rx, Ry)/R - \Phi(Rx, Ry)/R\| \\
&= \frac{1}{R} \max_{(x,y)\in B} \|(x, y)/R - \Phi(x, y)\| \\
&= \frac{\epsilon}{R}.
\end{aligned} \tag{154}$$

Then the result for $R = 1$ states that

$$\|\widetilde{f} - \widetilde{f}_\Phi\|_{V^p([-1,1])} \leq \widetilde{\epsilon} \cdot 4^{1-1/p} \cdot C(\widetilde{F}, \widetilde{\Phi}, p), \tag{155}$$

where

$$C(\widetilde{F}, \widetilde{\Phi}, p) = \min\{\|\widetilde{F}\|_{L^p}, \|\widetilde{F}_\Phi\|_{L^p}\}. \tag{156}$$

Performing a change of variables $u = Ry$, we find

$$\begin{aligned}\widetilde{f}_\Phi(x) &= \int_{-1}^{1} \widetilde{F}_\Phi(x,y)\,dy \\ &= \frac{1}{R}\int_{-R}^{R} \widetilde{F}_\Phi(x,u/R)\,du \\ &= \frac{1}{R}\int_{-R}^{R} \widetilde{F}_\Phi((Rx,u)/R)\,du \\ &= \frac{1}{R}\int_{-R}^{R} F_\Phi(Rx,u)\,du \\ &= \frac{1}{R} f_\Phi(Rx), \end{aligned} \tag{157}$$

and similarly, $\widetilde{f}(x) = f(Rx)/R$. Consequently, by another change of variables $u = Rt$,

$$\begin{aligned}\int_{-1}^{x} (\widetilde{f}(t) - \widetilde{f}_\Phi(t))\,dt &= \frac{1}{R}\int_{-R}^{Rx} (\widetilde{f}(u/R) - \widetilde{f}_\Phi(u/R))\,du \\ &= \frac{1}{R^2}\int_{-R}^{Rx} (f(u) - f_\Phi(u))\,du. \end{aligned} \tag{158}$$

Therefore, for all $1 \le p < \infty$, letting $v = Rx$,

$$\begin{aligned}\|\widetilde{f} - \widetilde{f}_\Phi\|_{V^p([-1,1])}^p &= \int_{-1}^{1} \left| \int_{-1}^{x} (\widetilde{f}(t) - \widetilde{f}_\Phi(t))\,dt \right|^p dx \\ &= \int_{-1}^{1} \left| \frac{1}{R^2}\int_{-R}^{Rx} (f(u) - f_\Phi(u))\,du \right|^p dx \\ &= \frac{1}{R}\int_{-R}^{R} \left| \frac{1}{R^2}\int_{-R}^{v} (f(u) - f_\Phi(u))\,du \right|^p dv \\ &= \frac{1}{R^{2p+1}}\int_{-R}^{R} \left| \int_{-R}^{v} (f(u) - f_\Phi(u))\,du \right|^p dv \\ &= \frac{1}{R^{2p+1}} \|f - f_\Phi\|_{V^p([-R,R])}^p, \end{aligned} \tag{159}$$

and so

$$\|\widetilde{f} - \widetilde{f}_\Phi\|_{V^p([-1,1])} = \frac{1}{R^{2+1/p}} \|f - f_\Phi\|_{V^p([-R,R])}. \tag{160}$$

On the other hand,

$$\begin{aligned}\|\widetilde{F}\|_{L^p(\mathbb{D}_1)}^p &= \int_{\mathbb{D}_1} |\widetilde{F}(x,y)|^p\,dx\,dy \\ &= \int_{\mathbb{D}_1} |F(Rx,Ry)|^p\,dx\,dy \\ &= \frac{1}{R^2}\int_{\mathbb{D}_R} |F(u,v)|^p\,du\,dv \\ &= \frac{1}{R^2}\|F\|_{L^p(\mathbb{D}_R)}^p, \end{aligned} \tag{161}$$

and so

$$\|\widetilde{F}\|_{L^p(\mathbb{D}_1)} = \frac{1}{R^{2/p}} \|F\|_{L^p(\mathbb{D}_R)}. \tag{162}$$

Similarly,

$$\|\widetilde{F}_\Phi\|_{L^p(\mathbb{D}_1)} = \frac{1}{R^{2/p}}\|F_\Phi\|_{L^p(\mathbb{D}_R)}. \tag{163}$$

From (162) and (163), we see that

$$C(\widetilde{F}, \widetilde{\Phi}, p) = \frac{1}{R^{2/p}} C(F, \Phi, p). \tag{164}$$

Combining (154), (155), (160), and (164) gives

$$\frac{1}{R^{2+1/p}}\|f - f_\Phi\|_{V^p([-R,R])} \leq \widetilde{\epsilon} \cdot 4^{1-1/p} \cdot \frac{1}{R^{2/p}} \cdot C(F, \Phi, p) = \epsilon \cdot \frac{1}{R^{2/p+1}} \cdot 4^{1-1/p} \cdot C(F, \Phi, p), \tag{165}$$

or equivalently,

$$\|f - f_\Phi\|_{V^p([-R,R])} \leq \epsilon \cdot (4R)^{1-1/p} \cdot C(F, \Phi, p), \tag{166}$$

as claimed. The proof for $p = \infty$ can be proved similarly, or by taking the limit as $p \to \infty$ and using that $\|\cdot\|_{L^\infty} = \lim_{p\to\infty} \|\cdot\|_{L^p}$.

6.3.2 Proof of Theorem 4.3 for $R = 1$

First, suppose $p = 1$. Let $\mathbb{D} = \mathbb{D}_1$. Let $\Psi = \Phi^{-1}$ with $\Psi(u, v) = (\psi_1(u, v), \psi_2(u, v))$. By definition,

$$f(x) = \int_{\mathbb{R}} F(x, y)\, dy, \tag{167}$$

and

$$\begin{aligned} f_\Phi(x) &= \int_{\mathbb{R}} F_\Phi(x, y)\, dy \\ &= \int_{y:(x,y)\in B} F(\Phi(x, y))|J_\Phi(x, y)|\, dy. \end{aligned} \tag{168}$$

Let G on $[-1, 1]$ be absolutely continuous, whose derivative $g = G'$ has $\|g\|_{L^\infty} \leq 1$.

Performing a change of variables gives

$$\begin{aligned} \int_{-1}^{1} G(x) f_\Phi(x)\, dx &= \int_{-1}^{1} G(x) \int_{y:(x,y)\in B} F(\Phi(x, y))|J_\Phi(x, y)|\, dy\, dx \\ &= \int_B G(x) F(\Phi(x, y))|J_\Phi(x, y)|\, dy\, dx \\ &= \int_A G(\psi(u, v)) F(u, v)\, du\, dv, \end{aligned} \tag{169}$$

where we have let $\psi = \psi_1$. Similarly,

$$\int_{-1}^{1} G(x) f(x)\, dx = \int_A G(x) F(x, y)\, dx\, dy. \tag{170}$$

We then have

$$\begin{aligned} \int_{-1}^{1} G(x)(f(x) - f_\Phi(x))\, dx &= \int_A G(x) F(x, y)\, dx\, dy - \int_A G(\psi(x, y)) F(x, y)\, dx\, dy \\ &= \int_A (G(x) - G(\psi(x, y))) F(x, y)\, dx\, dy \\ &\leq \|F\|_{L^1} \max_{(x,y)\in A} |G(x) - G(\psi(x, y))|. \end{aligned} \tag{171}$$

Now, because $g = G'$ satisfies $\|g\|_{L^\infty} \leq 1$, we have

$$\begin{aligned} |G(x) - G(\psi(x,y))| &= \left| \int_x^{\psi(x,y)} g(t)\,dt \right| \\ &\leq \|g\|_{L^\infty} |x - \psi(x,y)| \\ &\leq \|(x,y) - (\psi_1(x,y), \psi_2(x,y))\| \\ &\leq \epsilon, \end{aligned} \tag{172}$$

and therefore, taking the supremum over all such G and using Proposition 3.1 shows that

$$\|f - f_\Phi\|_{V^1} \leq \epsilon \cdot \|F\|_{L^1}. \tag{173}$$

We now prove the result for $p = \infty$, again assuming that $R = 1$. Let $I_{x,y}$ be the interval $[x, \psi(x,y)]$ when $x \leq \psi(x,y)$, and $[\psi(x,y), x]$ when $x > \psi(x,y)$; and let $\chi(x,y,t)$ be 1 if $t \in I_{x,y}$, and 0 otherwise. We begin with the following lemma:

Lemma 6.11. *For all* $|t| \leq 1$*,*

$$\int_A \chi(x,y,t)\,dx\,dy \leq 4\epsilon. \tag{174}$$

Proof. Let $\mathbf{S}_1 = \{(x,y) \in A : x \leq t \leq \psi(x,y)\}$, and let $\mathbf{S}_2 = \{(x,y) \in A : \psi(x,y) \leq t \leq x\}$. Then

$$\int_A \chi(x,y,t)\,dx\,dy = |\mathbf{S}_1 \cup \mathbf{S}_2|. \tag{175}$$

To bound the area of $\mathbf{S}_1$, observe first that any (x,y) contained in $\mathbf{S}_1$ must satisfy $t - \epsilon \leq x \leq t$. Indeed, since, by assumption, $\psi(x,y) - x \leq \epsilon$, we have

$$x \geq \psi(x,y) - \epsilon \geq t - \epsilon, \tag{176}$$

as claimed. Consequently, since $A \subset \mathbb{D}$ and the radius of $\mathbb{D}$ is 1,

$$|\mathbf{S}_1| \leq |\{(x,y) \in \mathbb{D} \;:\; t - \epsilon \leq x \leq t\}| \leq 2\epsilon. \tag{177}$$

Similarly, $|\mathbf{S}_2| \leq 2\epsilon$, and hence

$$\int_A \chi(x,y,t)\,dx\,dy = |\mathbf{S}_1 \cup \mathbf{S}_2| \leq 4\epsilon, \tag{178}$$

as claimed. □

Now, take an absolutely continuous G on $[-1,1]$ whose derivative $g = G'$ satisfies $\|g\|_{L^1} = 1$. Using (169) and (170) as before, we have

$$\begin{aligned} \int_{-1}^1 G(x)(f(x) - f_\Phi(x))\,dx &= \int_{-1}^1 G(x) \int_{\mathbb{R}} F(x,y)\,dy\,dx - \int_{-1}^1 G(x) f_\Phi(x)\,dx \\ &= \int_A (G(x) - G(\psi(x,y)))F(x,y)\,dx\,dy \\ &\leq \|F\|_{L^\infty} \int_A |G(x) - G(\psi(x,y))|\,dx\,dy. \end{aligned} \tag{179}$$

Then, using $g = G'$, we have

$$\begin{aligned}\int_A |G(x) - G(\psi(x,y))|\,dx\,dy &= \int_A \left|\int_{I_{x,y}} g(t)dt\right| dx\,dy \\ &= \int_A \left|\int_{-1}^{-1} g(t)\chi(x,y,t)\,dt\right| dx\,dy \\ &\le \int_A \int_{-1}^{1} |g(t)|\chi(x,y,t)\,dt\,dx\,dy \\ &= \int_{-1}^{1} |g(t)| \int_A \chi(x,y,t)\,dx\,dy\,dt \\ &\le \|g\|_{L^1} \sup_{|t|\le 1} \int_A \chi(x,y,t)\,dx\,dy. \\ &= \sup_{|t|\le 1} \int_A \chi(x,y,t)\,dx\,dy. \end{aligned} \tag{180}$$

Invoking Lemma 6.11 and Proposition 3.1 then shows

$$\|f - f_\Phi\|_{V^\infty} \le 4 \cdot \epsilon \cdot \|F\|_{L^\infty}. \tag{181}$$

Since the mapping $F \mapsto \int_{-1}^{x}(f - f_\Phi)$ is linear, we may now combine the bounds (173) and (181) using the Riesz-Thorin Interpolation Theorem (see, e.g. Theorem 6.27 in [21]) to complete the proof.

6.4 Proof of Theorem 4.4

We recall the statement of the theorem:

Theorem 6.12. *Suppose $a = c_0 < c_1 < \cdots < c_r = b$, and suppose f is Lipschitz continuous on each interval (c_j, c_{j+1}), $0 \le j \le r-1$, and continuous from either the right or left at each c_j, $0 \le j \le r$. Then for all $1 \le p \le \infty$, and*

$$\left| \|\mathbf{f}\|_{\nu_p} - \|f\|_{V^p} \right| \le \frac{C}{n}, \tag{182}$$

where $C > 0$ does not depend on n or p. The same bound holds by replacing f with f_{cen} and $\mathbf{f}$ with $\mathbf{f}_{\text{cen}}$.

For simplicity, we will suppose that f is continuous from the left. Take a constant L so that f has Lipschitz constant not exceeding L on each interval $(c_k, c_{k+1}]$, $0 \le k \le r-1$, such that $|f(t)| \le L$ for all $t \in [a, b]$. To begin, suppose $1 \le p < \infty$.

The following lemmas are standard bounds on the error of the trapezoidal approximation to an integral; their proofs are included for completeness.

Lemma 6.13. *Fix an integer m, $1 \le m \le n$, and suppose that $c_{k_m - 1} \le a_m < c_{k_m}$. Then*

$$\left| \frac{b-a}{n} \sum_{0 \le j \le m}^{\text{trap}} f(a_j) - \int_a^{a_m} f(t)dt \right| \le \frac{mL(b-a)^2}{2n^2} + \frac{2k_m(b-a)}{n}. \tag{183}$$

Proof. Fix $0 \le j \le m$. Because f is bounded by L,

$$\left| \frac{b-a}{2n} \left(f(a_j) + f(a_{j+1})\right) - \int_{a_j}^{a_{j+1}} f(t)dt \right| \le \frac{2L(b-a)}{n}. \tag{184}$$

On the other hand, if there are no c_ℓ between a_j and a_{j+1}, then f is L-Lipschitz on $(a_j, a_{j+1}]$, and so

$$\begin{aligned}\left|\frac{b-a}{2n}\left(f(a_j)+f(a_{j+1})\right)-\int_{a_j}^{a_{j+1}} f(t)dt\right| &= \frac{1}{2}\left|\int_{a_j}^{a_{j+1}} (f(a_j)-f(t))+(f(a_{j+1})-f(t))dt\right| \\ &\leq \int_{a_j}^{a_{j+1}} \frac{L(b-a)}{2n}dt \\ &= \frac{L(b-a)^2}{2n^2}. \end{aligned} \tag{185}$$

Since there are at most k_m intervals $(a_j, a_{j+1}]$ containing a value from among $c_0, \ldots, c_{k_m-1}$, and at most m subintervals not containing any such value, we have

$$\begin{aligned}\left|\frac{b-a}{n}\sum_{0\leq j\leq m}^{\text{trap}} f(a_j)-\int_a^{a_m} f(t)dt\right| &= \left|\frac{b-a}{2n}\sum_{j=0}^{m-1}(f(a_j)+f(a_{j+1}))-\int_a^{a_m} f(t)dt\right| \\ &\leq \sum_{j=0}^{m-1}\left|\frac{b-a}{2n}\left(f(a_j)+f(a_{j+1})\right)-\int_{a_j}^{a_{j+1}} f(t)dt\right| \\ &\leq \frac{mL(b-a)^2}{2n^2}+\frac{2k_m(b-a)}{n}, \end{aligned} \tag{186}$$

as claimed. ☐

Lemma 6.14. *Suppose G is a function on $[a, b]$ that is Lipschitz continuous, with Lipschitz constant bounded above by A. Then*

$$\left|\frac{b-a}{n}\sum_{0\leq j\leq m}^{\text{trap}} G(a_j)-\int_a^{a_m} G(t)dt\right| \leq \frac{mA(b-a)^2}{2n^2}. \tag{187}$$

Proof. Apply (185) to G, and conclude as in (186). ☐

The next lemma bounds the Lipschitz norm of $|\mathcal{V}f(x)|^p$:

Lemma 6.15. *The function $|\mathcal{V}f(x)|^p$ is continuous on $[a, b]$, with Lipschitz constant bounded above by $pL^p(b-a)^{p-1}$.*

Proof. Let $G(x) = |\mathcal{V}f(x)|$. Since $\|f\|_\infty \leq L$, G is Lipschitz continuous with Lipschitz constant L:

$$G(x)-G(y) = \left|\int_a^x f(t)dt\right| - \left|\int_a^y f(t)dt\right| \leq \left|\int_x^y f(t)dt\right| \leq L|x-y|. \tag{188}$$

Furthermore, G is bounded above:

$$|G(x)| \leq \int_a^b |f(t)|dt \leq L(b-a). \tag{189}$$

Since the derivative of $y \mapsto y^p$ is py^{p-1}, which is bounded above by $pL^{p-1}(b-a)^{p-1}$ on $[0, L(b-a)]$, we have

$$\begin{aligned}|\mathcal{V}f(y)|^p - |\mathcal{V}f(x)|^p &= G(y)^p - G(x)^p \\ &\leq pL^{p-1}(b-a)^{p-1}|G(y)-G(x)| \\ &\leq pL^{p-1}(b-a)^{p-1}L|x-y| \\ &= pL^p(b-a)^{p-1}|x-y|, \end{aligned} \tag{190}$$

as claimed. ☐

Now, for $0 \le m \le n$, define $E(m,n)$ by

$$E(m,n) = \frac{b-a}{n} \sum_{0 \le j \le m}^{\text{trap}} f(a_j) - \int_a^{a_m} f(t)dt = (\mathbf{Vf})[m] - (\mathcal{V}f)(a_m). \tag{191}$$

Applying Lemma 6.13, since $k_m \le r$ and $m \le n$, we have

$$|E(m,n)| \le \frac{mL(b-a)^2}{2n^2} + \frac{2k_m(b-a)}{n} \le \frac{C}{n}, \tag{192}$$

where C is independent of n. Consequently,

$$\begin{aligned} \left| \|\mathbf{f}\|_{\nu_p} - \left(\frac{b-a}{n} \sum_{0 \le m \le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p \right)^{1/p} \right| &= \left| \left(\frac{b-a}{n} \sum_{0 \le m \le n}^{\text{trap}} |(\mathbf{Vf})[m]|^p \right)^{1/p} - \left(\frac{b-a}{n} \sum_{0 \le m \le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p \right)^{1/p} \right| \\ &\le \left(\frac{b-a}{n} \sum_{0 \le m \le n}^{\text{trap}} |E(m,n)|^p \right)^{1/p} \\ &\le \frac{C}{n}, \end{aligned} \tag{193}$$

where C may have changed, but is still independent of n and p.

By Lemma 6.15, the Lipschitz constant of $|\mathcal{V}f(x)|^p$ is $pL^p(b-a)^{p-1}$. Applying Lemma 6.13 then yields

$$\left| \frac{b-a}{n} \sum_{0 \le m \le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p - \int_a^b |(\mathcal{V}f)(x)|^p \, dx \right| \le \frac{n(pL^p(b-a)^{p-1})(b-a)^2}{2n^2} = \frac{pL^p(b-a)^{p+1}}{2n}. \tag{194}$$

Since $|f(t)| \le L$ for all t in $[a,b]$, it follows that $|(\mathcal{V}f)(x)| \le L(b-a)$ too, and consequently:

$$\frac{b-a}{n} \sum_{0 \le m \le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p \le L^p(b-a)^{p+1} \tag{195}$$

and

$$\int_a^b |(\mathcal{V}f)(x)|^p \, dx \le L^p(b-a)^{p+1}. \tag{196}$$

The function $y \mapsto y^{1/p}$ has derivative $y^{1/p-1}/p$, which on $[0, L^p(b-a)^{p+1}]$ has maximum value

$$\frac{(L^p(b-a)^{p+1})^{1/p-1}}{p} = \frac{L^{1-p}(b-a)^{1/p-p}}{p}, \tag{197}$$

and consequently

$$\begin{aligned} \left| \left(\frac{b-a}{n} \sum_{0 \le m \le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p \right)^{1/p} - \left(\int_a^b |(\mathcal{V}f)(x)|^p \, dx \right)^{1/p} \right| &\le \frac{L^{1-p}(b-a)^{1/p-p}}{p} \cdot \frac{pL^p(b-a)^{p+1}}{2n} \\ &= \frac{L(b-a)^{1/p+1}}{2n} \\ &\le \frac{C}{n}. \end{aligned} \tag{198}$$

Combining (193) and (198) completes the proof of (182) for $p < \infty$. The corresponding result for $p = \infty$ follows by taking the limit $p \to \infty$ and using the convergence of the p-norm to the ∞-norm.

To prove the result for f_{cen} and $\mathbf{f}_{\text{cen}}$, suppose first that $p < \infty$. From Lemma 6.13, we have

$$|\mu(f) - \mu(\mathbf{f})| = \left| \frac{1}{b-a} \int_a^b f(x)dx - \frac{1}{n} \sum_{0 \le j \le n}^{\text{trap}} f(a_j) \right| \le \frac{C}{n}, \tag{199}$$

where $C > 0$ is a constant independent of n. Letting $\widetilde{\mathbf{f}}$ be the vector in $\mathbb{R}^{n+1}$ with entries $\widetilde{\mathbf{f}}[k] = f_{\text{cen}}(a_k) = f(a_k) - \mu(f)$, $0 \le k \le n$, applying (182) to f_{cen} gives

$$\left| \|\widetilde{\mathbf{f}}\|_{\nu_p} - \|f_{\text{cen}}\|_{V^p} \right| \le \frac{C}{n}. \tag{200}$$

Furthermore, for all $0 \le k \le n$,

$$\mathbf{f}_{\text{cen}}[k] - \widetilde{\mathbf{f}}[k] = \mu(f) - \mu(\mathbf{f}), \tag{201}$$

and so

$$\begin{aligned} \left| \|\widetilde{\mathbf{f}}\|_{\nu_p} - \|\mathbf{f}_{\text{cen}}\|_{\nu_p} \right| &\le \|\widetilde{\mathbf{f}} - \mathbf{f}_{\text{cen}}\|_{\nu_p} \\ &= \|(\mu(f) - \mu(\mathbf{f}))\mathbf{1}\|_{\nu_p} \\ &= \left(\frac{b-a}{n} \sum_{k=0}^{n} \left| \frac{b-a}{n} \sum_{j=0}^{n} (\mu(f) - \mu(\mathbf{f})) \right|^p \right)^{1/p} \\ &\le \frac{C}{n}. \end{aligned} \tag{202}$$

The result now follows by combining (200) and (202). As before, the result for $p = \infty$ follows by taking the limit $p \to \infty$ and using the convergence of the p-norm to the ∞-norm.

6.5 Proof of Theorem 4.5

We recall the statement of the theorem:

Theorem 6.16. *Suppose f is a two times continuously differentiable function on $[a, b]$, and has only finitely many roots. The for all $1 \le p \le \infty$,*

$$\left| \|\mathbf{f}\|_{\nu_p} - \|f\|_{V^p} \right| \le \frac{C}{n^2}, \tag{203}$$

where $C > 0$ does not depend on n or p. The same bound holds by replacing f with f_{cen} and $\mathbf{f}$ with $\mathbf{f}_{\text{cen}}$.

The proof makes use of the following version of the Euler-Maclaurin formula, which is Theorem 3 in [11] (see also Theorem 2 in [12], which appears to correct an error in the definition of the parameters t_j appearing in the statement below):

Theorem 6.17 (Euler-Maclaurin formula with jumps)**.** *Let $a = c_0 < c_1 < \cdots < c_{r-1} < c_r = b$, and let $q \ge 2$ be an integer. Suppose that, for every $0 \le j \le r$, H is in $C^{q-1}([c_j, c_{j+1}])$ and $H^{(q)}$ is absolutely integrable on (c_j, c_{j+1}). Suppose too that $H(c_j) = (H(c_j+) + H(c_j-))/2$ for all $1 \le j \le r-1$, and $H(c_0) = (H(c_0+) + H(c_r-))/2$. Let n be a positive integer, and $\delta = (b-a)/n$. For $0 \le t < 1$, define*

$$\widehat{I} = \delta \sum_{k=0}^{n-1} H\left(a + (k+t)\delta\right) \tag{204}$$

For integer $\ell \ge 1$, let P_ℓ denote the ℓ-th Bernoulli polynomial on $[0, 1)$, extended 1-periodically to the whole real line (so that $P_\ell(x + k) = P_\ell(x)$ for all integers k). Let $t_j = (a/\delta + t - c_j/\delta) \bmod 1$, $0 \le j \le r-1$.

Furthermore, for $0 \leq j \leq r-1$, let

$$\gamma_j = \begin{cases} 0, & \text{if } t_j = 0, \\ 1, & \text{if } t_j \neq 0; \end{cases} \tag{205}$$

let

$$\alpha_1 = \sum_{j=0}^{r-1} \gamma_j P_1(t_j - a)(H(c_j-) - H(c_j+)); \tag{206}$$

and, for $2 \leq \ell \leq q$, let

$$\alpha_\ell = \sum_{j=0}^{r-1} \frac{P_\ell(t_j - a)}{\ell!}(H^{(\ell-1)}(c_j-) - H^{(\ell-1)}(c_j+)). \tag{207}$$

Then

$$\widehat{I} - \int_a^b H(x)dx = \alpha_1 \delta + \sum_{\ell=2}^{q} \alpha_\ell \delta^\ell - \frac{\delta^q}{q!}\int_a^b H^{(q)}(x) \sum_{j=0}^{r-1} P_q(t_j - x/\delta - a)dx. \tag{208}$$

The next lemma is standard:

Lemma 6.18. *Suppose f is a C^2 function on $[a, b]$. Then*

$$\left| \frac{b-a}{2}(f(a) + f(b)) - \int_a^b f(t)dt \right| \leq C\|f''\|_{L^\infty}(b-a)^3, \tag{209}$$

where $C > 0$ is a universal constant.

Corollary 6.19. *Let $0 \leq m \leq n$. Then*

$$|(\mathcal{V}f)(a_m) - (\mathbf{Vf})[m]| \leq C\frac{m}{n^3}, \tag{210}$$

where the constant $C > 0$ does not depend on m or n.

Proof. When $m = 0$, the left side is 0. When $m \geq 1$,

$$\begin{aligned} |(\mathcal{V}f)(a_m) - (\mathbf{Vf})[m]| &= \left| \int_a^{a_m} f(t)dt - (\mathbf{Vf})[m] \right| \\ &= \left| \sum_{j=0}^{m-1} \int_{a_j}^{a_{j+1}} f(t)dt - \frac{b-a}{2n}\sum_{j=0}^{m-1}(f(a_j) + f(a_{j+1})) \right| \\ &= \left| \sum_{j=0}^{m-1} \left(\int_{a_j}^{a_{j+1}} f(t)dt - \frac{b-a}{2n}(f(a_j) + f(a_{j+1})) \right) \right| \\ &\leq C\frac{m}{n^3}, \end{aligned} \tag{211}$$

as claimed. □

Now, for $0 \leq m \leq n$, define $E(m, n)$ by

$$E(m, n) = (\mathcal{V}f)(a_m) - (\mathbf{Vf})[m]. \tag{212}$$

Applying Lemma 6.19, we have

$$|E(m, n)| \leq C\frac{m}{n^3} \leq \frac{C}{n^2}, \tag{213}$$

where C is independent of n. Consequently,

$$\left| \|\mathbf{f}\|_{\nu_p} - \left(\frac{b-a}{n} \sum_{0\le m\le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p \right)^{1/p} \right| = \left| \left(\frac{b-a}{n} \sum_{0\le m\le n}^{\text{trap}} |(\mathbf{Vf})[m]|^p \right)^{1/p} - \left(\frac{b-a}{n} \sum_{0\le m\le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p \right)^{1/p} \right|$$
$$\le \left(\frac{b-a}{n} \sum_{0\le m\le n}^{\text{trap}} |E(m,n)|^p \right)^{1/p}$$
$$\le \frac{C}{n^2}. \tag{214}$$

We will show that there are constants A and B, not depending on n or p, such that

$$\left| \frac{b-a}{n} \sum_{0\le m\le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p - \int_a^b |(\mathcal{V}f)(x)|^p dx \right| \le A \cdot p \cdot \frac{B^p}{n^2}. \tag{215}$$

Let $G(x) = (\mathcal{V}f)(x)$, and $H(x) = |G(x)|^p = |(\mathcal{V}f)(x)|^p$. Since $\int_a^b f = 0$, $G(a) = G(b) = 0$. Furthermore, because f is C^2, G is C^3. Let $a = c_0 < c_1 < \cdots < c_{r-1} < c_r = b$ be points such that $\operatorname{sign}(G(x))$ is constant on each subinterval (c_j, c_{j+1}); note that $G(c_j) = 0$ for all $0 \le j \le r$.

Lemma 6.20. *There are constants A and B, depending on f but not on n or p, such that, in the notation of Theorem 6.17,*

$$\left| \int_a^b H''(x) \sum_{j=0}^{r-1} P_2(t_j - x/\delta - a)dx \right| \le A \cdot p \cdot B^p. \tag{216}$$

Proof. Let us first suppose that $p > 1$. Fix a value ℓ such that $G(x) > 0$ on $(c_\ell, c_{\ell+1})$; then $H(x) = G(x)^p$ on $(c_\ell, c_{\ell+1})$,

$$H'(x) = pG(x)^{p-1}G'(x) = pG(x)^{p-1}f(x), \tag{217}$$

and

$$H''(x) = p(p-1)G(x)^{p-2}f(x)^2 + pG(x)^{p-1}f'(x). \tag{218}$$

Take points $c_\ell = b_1 < b_2 < \cdots < b_s = c_{\ell+1}$ such that the sign of f is constant on each subinterval (b_i, b_{i+1}). Then using the notation of Theorem 6.17,

$$\sum_{j=1}^{r-1} \int_{c_\ell}^{c_{\ell+1}} pG(x)^{p-1} f'(x) P_2(t_j - x/\delta - a)dx \le p\|G\|_{L^\infty}^{p-1} \|f'\|_{L^\infty} \sum_{j=1}^{r-1} \int_{c_\ell}^{c_{\ell+1}} P_2(t_j - x/\delta - a)dx,$$
$$\le pC_1C_2^p, \tag{219}$$

where C_1 and C_2 do not depend on n or p. An identical argument works if $G(x) < 0$ on $(c_\ell, c_{\ell+1})$, and the estimate is trivial if $G(x) = 0$ on $(c_\ell, c_{\ell+1})$.

Next, every $1 \le m \le s$, since $f = G'$ has constant sign in (b_m, b_{m+1}), we can perform the change of variables $u = G(x)$ and get

$$\begin{aligned}\int_{b_m}^{b_{m+1}} H''(x)P_q(t_j - x/\delta - a)dx &= p(p-1)\int_{b_m}^{b_{m+1}} G(x)^{p-2}f(x)^2P_2(t_j - x/\delta - a)dx\\ &\le p(p-1)\|P_2\|_{L^\infty}\int_{b_m}^{b_{m+1}} G(x)^{p-2}f(x)^2dx\\ &\le p(p-1)\|P_2\|_{L^\infty}\int_{b_m}^{b_{m+1}} u^{p-2}f(G^{-1}(u))du\\ &\le p(p-1)\|P_2\|_{L^\infty}\|f\|_{L^\infty}\int_{G(b_m)}^{G(b_{m+1})} u^{p-2}du\\ &= p\|P_2\|_{L^\infty}\|f\|_{L^\infty}\left(G(b_{m+1})^{p-1} - G(b_m)^{p-1}\right)\\ &\le pC_3C_4^p, \end{aligned}\tag{220}$$

where C_3 and C_4 do not depend on n or p. An identical argument works if $G(x) < 0$ on $(c_\ell, c_{\ell+1})$, and the estimate is trivial if $G(x) = 0$ on $(c_\ell, c_{\ell+1})$. Combining the bounds (219) and (220) completes the proof when $p > 1$.

If $p = 1$, suppose $G(x) > 0$ on $(c_\ell, c_{\ell+1})$; then $H(x) = G(x)$ on $(c_\ell, c_{\ell+1})$, and consequently $H' = G'$ and $H'' = G'' = f'$. Therefore,

$$\begin{aligned}\left|\int_a^b H''(x)\sum_{j=0}^{r-1}P_2(t_j - x/\delta - a)dx\right| &= \left|\int_a^b f'(x)\sum_{j=0}^{r-1}P_2(t_j - x/\delta - a)dx\right|\\ &\le \|f\|_{L^\infty}\left|\int_a^b \sum_{j=0}^{r-1}P_2(t_j - x/\delta - a)dx\right|, \end{aligned}\tag{221}$$

which is a constant that does not depend on n. The same argument applies if $G(x) < 0$ on $(c_\ell, c_{\ell+1})$, and the estimate is trivial if $G(x) = 0$ on $(c_\ell, c_{\ell+1})$. □

Lemma 6.21. *In the notation of Theorem 6.17, $\alpha_1 = 0$ and $|\alpha_2| \le A \cdot p \cdot B^p$ for all $p \ge 1$, where A and B are constants not depending on n or p.*

Proof. since H is continuous in the interior of $[a, b]$, $H(c_j-) = H(c_j+)$ when $1 \le j \le r-1$. Furthermore, when $j = 0$, $t_0 = 0$, and hence $\gamma_0 = 0$; consequently, all terms in the sum (206) are 0, hence $\alpha_1 = 0$.

If $p > 1$, the parameter α_2 defined in (207) may be bounded by

$$\begin{aligned}|\alpha_2| &= \left|\frac{1}{2}\sum_{j=0}^{r-1}P_2(t_j - a)(H'(c_j-) - H'(c_j+))\right|\\ &\le p\sum_{j=0}^{r-1}P_2(t_j - a)\|G\|_{L^\infty}^{p-1}\|f\|_{L^\infty}\\ &\le Ap\|f\|_{L^\infty}^p\\ &= ApB^p, \end{aligned}\tag{222}$$

where A and B are independent of n or p, and where we have used the bound $\|G\|_{L^\infty} \le L \cdot \|f\|_{L^\infty}$. If $p = 1$, then $H' = \pm f$, and so

$$|\alpha_2| = \left|\frac{1}{2}\sum_{j=0}^{r-1}P_2(t_j - a)(H'(c_j-) - H'(c_j+))\right| \le \sum_{j=0}^{r-1}P_2(t_j - a)\|f\|_{L^\infty},\tag{223}$$

which is bounded independently of n. So the bound $|\alpha_2| \le ApB^p$ is valid for all $p \ge 1$, for A and B independent of n or p. □

Combining Lemma 6.20 and Lemma 6.21, we can apply Theorem 6.17 with $t = 0$:

$$\begin{aligned}\left|\frac{b-a}{n}\sum_{0\le k\le n}^{\text{trap}}|(\mathcal{V}f)(a_k)|^p - \int_a^b |(\mathcal{V}f)(x)|^p dx\right| &= \left|\delta\sum_{k=0}^{n-1} H((k+1/2)\delta) - \int_a^b H(x)dx\right| \\ &= \left|\alpha_2\delta^2 - \frac{\delta^2}{2}\sum_{j=0}^{r-1}\int_a^b H''(x)P_2(t_j - x/\delta - a)dx\right| \\ &\le A\cdot p\cdot \frac{B^p}{n^2}, \end{aligned} \tag{224}$$

where A and B are independent of n or p.

Since $|(\mathcal{V}f)(x)| \le (b-a)\cdot \|f\|_{L^\infty}$,

$$\frac{b-a}{n}\sum_{0\le m\le n}^{\text{trap}} |(\mathcal{V}f)(a_m)|^p \le (b-a)^{p+1}\|f\|_{L^\infty}^p, \tag{225}$$

and

$$\int_a^b |(\mathcal{V}f)(x)|^p\, dx \le (b-a)^{p+1}\|f\|_{L^\infty}^p; \tag{226}$$

consequently, by making A and B larger if necessary, we can assume that both $\frac{L}{n}\sum_{m=1}^n |(\mathcal{V}f)(a_m)|^p$ and $\int_a^b |(\mathcal{V}f)(x)|^p\, dx$ lie in the interval $[0, A\cdot B^p]$, where A and B are from (224).

The function $y \mapsto y^{1/p}$ has derivative $y^{1/p-1}/p$, which on $[0, A\cdot B^p]$ has maximum value

$$\frac{(A\cdot B^p)^{1/p-1}}{p} = \frac{A^{1/p-1}B^{1-p}}{p}, \tag{227}$$

and consequently

$$\begin{aligned}\left|\left(\frac{b-a}{n}\sum_{0\le m\le n}^{\text{trap}}|(\mathcal{V}f)(a_m)|^p\right)^{1/p} - \left(\int_a^b |(\mathcal{V}f)(x)|^p\, dx\right)^{1/p}\right| &\le \frac{A^{1/p-1}B^{1-p}}{p}\cdot A\cdot p\cdot\frac{B^p}{n^2} \\ &= \frac{A^{1/p}B}{n^2} \\ &\le \frac{C}{n^2}, \end{aligned} \tag{228}$$

where $C > 0$ does not depend on n or p. Combining (193) and (198) completes the proof for $p < \infty$. The corresponding result for $p = \infty$ follows by taking the limit $p \to \infty$ and using the convergence of the p-norm to the ∞-norm.

To prove the result for f_{cen} and $\mathbf{f}_{\text{cen}}$, suppose first that $p < \infty$. From Corollary 6.19 (applied when $m = n$), we have

$$|\mu(f) - \mu(\mathbf{f})| = \left|\frac{1}{b-a}\int_a^b f(x)dx - \frac{1}{n}\sum_{0\le j\le n}^{\text{trap}} f(a_j)\right| \le \frac{C}{n^2}, \tag{229}$$

where $C > 0$ is a constant independent of n. Letting $\widetilde{\mathbf{f}}$ be the vector with entries $\widetilde{\mathbf{f}}[k] = f_{\text{cen}}(a_k) = f(a_k) - \mu(f)$, the first part of the theorem implies that

$$\left|\|\widetilde{\mathbf{f}}\|_{\nu_p} - \|f_{\text{cen}}\|_{V^p}\right| \le \frac{C}{n^2}. \tag{230}$$

Furthermore, for all $0 \le k \le n$,

$$\mathbf{f}_{\text{cen}}[k] - \widetilde{\mathbf{f}}[k] = \mu(f) - \mu(\mathbf{f}), \tag{231}$$

and so

$$\begin{aligned}\left| \|\widetilde{\mathbf{f}}\|_{\nu_p} - \|\mathbf{f}_{\text{cen}}\|_{\nu_p} \right| &\le \|\widetilde{\mathbf{f}} - \mathbf{f}_{\text{cen}}\|_{\nu_p} \\ &= \|(\mu(f) - \mu(\mathbf{f}))\mathbf{1}\|_{\nu_p} \\ &= \left(\frac{b-a}{n} \sum_{k=0}^{n} \left| \frac{b-a}{n} \sum_{j=0}^{n} (\mu(f) - \mu(\mathbf{f})) \right|^p \right)^{1/p} \\ &\le \frac{C}{n^2}. \end{aligned} \tag{232}$$

The result now follows by combining (230) and (232). As before, the result for $p = \infty$ follows by taking the limit $p \to \infty$ and using the convergence of the p-norm to the ∞-norm.

6.6 Proof of Theorem 4.6 and Corollary 4.7

We recall the statement of the theorem and its corollary:

Theorem 6.22. *Let $\sigma_0, \sigma_2, \ldots, \sigma_n, \ldots$ be a bounded sequence of positive numbers, and let $Z = (Z[0], \ldots, Z[n])$ where $Z[0], Z[1], \ldots, Z[n], \ldots$ are independent with $Z[j] \sim N(0, \sigma_j^2)$. Suppose too that $\sigma > 0$ satisfies*

$$\frac{1}{n} \sum_{j=1}^{n} \sigma_j^2 \le \sigma^2. \tag{233}$$

Let $t > 0$. Then for all $1 \le p \le \infty$,

$$\mathbb{P}\left\{ \|Z\|_{\nu_p} \ge t \right\} \le A e^{-Bt^2 n/\sigma^2}, \tag{234}$$

where $A > 0$ and $B > 0$ are constants independent of t, n, or p;

$$\lim_{n\to\infty} \|Z\|_{\nu_p} = 0 \tag{235}$$

almost surely; and

$$\mathbb{E}\|Z\|_{\nu_p} \le C \frac{\sigma}{\sqrt{n}}, \tag{236}$$

where $C > 0$ is a constant independent of n and p. Furthermore, (234), (235) *and* (236) *hold with Z replaced by Z_{cen}.*

Corollary 6.23. *Suppose f satisfies the conditions of Theorem 4.4, Z satisfies the conditions of Theorem 4.6, and $Y = \mathbf{f} + Z$. Let $t > 0$. Then for all $1 \le p \le \infty$,*

$$\mathbb{P}\left\{ \left| \|Y\|_{\nu_p} - \|f\|_{V^p} \right| \ge t \right\} \le A e^{-Bt^2 n/\sigma^2}, \tag{237}$$

where $A > 0$ and $B > 0$ are constants independent of t, n, or p;

$$\lim_{n\to\infty} \|Y\|_{\nu_p} = \|f\|_{V^p} \tag{238}$$

almost surely; and

$$\mathbb{E}\left| \|Y\|_{\nu_p} - \|f\|_{V^p} \right| \le C \frac{\sigma}{\sqrt{n}}, \tag{239}$$

where C is a constant independent of n and p. Furthermore, (237), (238) *and* (239) *hold with f replaced by f_{cen} and Y replaced by Y_{cen}.*

First, define $S_0[0] = S_1[0] = 0$, and for $k = 1, \ldots, n$, let

$$S_0[k] = \frac{b-a}{n} \sum_{j=0}^{k-1} Z[j], \tag{240}$$

and

$$S_1[k] = \frac{b-a}{n} \sum_{j=1}^{k} Z[j]. \tag{241}$$

For $0 \leq k \leq n$, let

$$T[k] = \frac{1}{2}(S_0[k] + S_1[k]). \tag{242}$$

Note that when $k \geq 1$,

$$T[k] = \frac{b-a}{n} \sum_{0 \leq j \leq k}^{\text{trap}} Z[j]. \tag{243}$$

Note too that S_0 and S_1 are martingales; that is, $\mathbb{E}[S_\ell[k] \mid Z[0], \ldots, Z[k-1]] = S_\ell[k-1]$, for $\ell = 0, 1$.

Lemma 6.24. *For any $t > 0$,*

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_\ell[k]| \geq t \right) \leq 2 \exp(-nt^2/2\sigma^2), \tag{244}$$

for $\ell = 0, 1$.

Proof. Let $S = S_0$. Let $\lambda > 0$, and define

$$X[k] = \exp(\lambda S[k]), \quad 0 \leq k \leq n. \tag{245}$$

Then X is a submartingale, i.e. $\mathbb{E}[X[k] \mid Z[0], \ldots, Z[k-1]] \geq X[k-1]$. Observe that $S[n]$ is normally distributed with mean zero and with variance

$$\overline{\sigma}^2 = \frac{(b-a)^2}{n^2} \sum_{j=0}^{n-1} \sigma_j^2 \leq \frac{(b-a)^2}{n} \sigma^2. \tag{246}$$

Consequently, using the standard formula for the Gaussian moment generating function,

$$\mathbb{E}[X[n]] = e^{\lambda^2 \overline{\sigma}^2/2} \leq e^{\lambda^2 (b-a)^2 \sigma^2/2n}. \tag{247}$$

By Doob's Inequality (e.g. see Theorem 5.4.2 in [19]), for any real number t,

$$\begin{aligned} \mathbb{P}\left(\max_{0 \leq k \leq n} S[k] \geq t \right) &= \mathbb{P}\left(\max_{0 \leq k \leq n} X[k] \geq \exp(\lambda t) \right) \\ &\leq \mathbb{E}[X[n]] \exp(-\lambda t) \\ &\leq e^{\lambda^2 (b-a)^2 \sigma^2/2n - \lambda t}. \end{aligned} \tag{248}$$

Taking $\lambda = tn/\sigma^2(b-a)^2$ yields the bound

$$\mathbb{P}\left(\max_{0 \leq k \leq n} S[k] \geq t \right) \leq \exp(-nt^2/2(b-a)^2\sigma^2). \tag{249}$$

By symmetry and the union bound, and the fact that $S = S_0$, this immediately gives the bound

$$\mathbb{P}\left(\max_{0 \leq k \leq n} |S_0[k]| \geq t \right) \leq 2 \exp(-nt^2/2(b-a)^2\sigma^2). \tag{250}$$

An identical argument holds for S_1, completing the proof. □

Since $T[k] = (S_0[k] + S_1[k])/2$, and since

$$\|Z\|_{\nu_\infty} = \max_{0 \le k \le n} |T[k]|, \tag{251}$$

the union bound shows

$$\mathbb{P}\left(\|Z\|_{\nu_\infty} \ge t\right) = \mathbb{P}\left(\max_{0\le k\le n} |T[k]| \ge t\right) \le 4\exp(-nt^2/2(b-a)^2\sigma^2). \tag{252}$$

This establishes (234) for $p = \infty$; the result then follows for all $p \ge 1$ since $\|Z\|_{\nu_p} \le \|Z\|_{\nu_\infty}$.

To see that the rest of the theorem follows from (234), observe that since the right side of (234) is summable over n, it follows from the Borel-Cantelli Lemma that $\lim_{n\to\infty} \|Z\|_{\nu_p} = 0$ almost surely, establishing (235). Furthermore,

$$\begin{aligned}
\mathbb{E}[\|Z\|_{\nu_p}] &= \int_0^\infty \mathbb{P}\left(\|Z\|_{\nu_p} \ge t\right) dt \\
&\le 2\int_0^\infty \exp(-nt^2/2(b-a)^2\sigma^2)dt \\
&= \frac{2\sigma}{\sqrt{n}} \int_0^\infty \exp(-u^2/2(b-a)^2)du,
\end{aligned} \tag{253}$$

which establishes (236) and completes the proof of the theorem for Z. The corresponding results for Z_{cen} may be deduced from those of Z and the standard concentration bound

$$\mathbb{P}\{|\mu(Z)| > t\} \le Ce^{-Dt^2n/\sigma^2}, \tag{254}$$

where C and D are constants independent of n or t; this follows from the fact that $\mu(Z) \sim N(0, \sigma^2/n)$. This completes the proof of Theorem 4.6.

To prove (237), recall that Theorem 4.4 gives the bound

$$\left|\|f\|_{V^p} - \|\mathbf{f}\|_{\nu_p}\right| \le \frac{C}{n}, \tag{255}$$

where C is a constant not depending on p, t or n. From the triangle inequality we have

$$\begin{aligned}
\|Y\|_{\nu_p} - \|f\|_{V^p} &= \|\mathbf{f} + Z\|_{\nu_p} - \|f\|_{V^p} \\
&\le \|\mathbf{f}\|_{\nu_p} + \|Z\|_{\nu_p} - \|f\|_{V^p} \\
&\le \frac{C}{n} + \|Z\|_{\nu_p},
\end{aligned} \tag{256}$$

and similarly, since $\|\mathbf{f}\|_{\nu_p} - \|Z\|_{\nu_p} \le \|Y\|_{\nu_p}$,

$$\begin{aligned}
\|f\|_{V^p} - \|Y\|_{\nu^p} &\le \|f\|_{V^p} - \|\mathbf{f}\|_{\nu^p} + \|Z\|_{\nu_p} \\
&\le \frac{C}{n} + \|Z\|_{\nu_p}.
\end{aligned} \tag{257}$$

Combining (256) and (257) shows

$$\left|\|Y\|_{\nu_p} - \|f\|_{V^p}\right| \le \frac{C}{n} + \|Z\|_{\nu_p}, \tag{258}$$

If $t - C/n \ge t/2$, then from Theorem 4.6,

$$\begin{aligned}
\mathbb{P}\left\{\left|\|f\|_{V^p} - \|Y\|_{\nu_p}\right| \ge t\right\} &\le \mathbb{P}\{\|Z\|_{\nu_p} \ge t - C/n\} \\
&\le \mathbb{P}\{\|Z\|_{\nu_p} \ge t/2\} \\
&\le Ae^{-Bn(t/2)^2/\sigma^2},
\end{aligned} \tag{259}$$

which is a bound of the desired form. On the other hand, if $t - C/n < t/2$, then $t \leq 2C/n$, and so

$$\begin{aligned}\mathbb{P}\left\{\left|\|f\|_{V^p} - \|Y\|_{\nu_p}\right| \geq t\right\} &\leq \mathbb{P}\{\|Z\|_{\nu_p} \geq t - C/n\} \\ &\leq Ae^{-Bn(t-C/n)^2/\sigma^2} \\ &= Ae^{-Bn(t^2-2Ct/n+C^2/n^2)/\sigma^2} \\ &= Ae^{-Bnt^2/\sigma^2}e^{2BCt/\sigma^2}e^{-BC^2/n\sigma^2} \\ &\leq Ae^{-Bnt^2/\sigma^2}e^{3BC^2/n\sigma^2}, \end{aligned} \tag{260}$$

which is bounded above by an expression of the desired form as well, since $e^{3BC^2/n\sigma^2}$ is bounded. This completes the proof of (237).

The limit (238) follows immediately from (258) and the fact that $\|Z\|_{\nu_p} \to 0$ almost surely. To prove (239), take expectations of each side of (258) and apply (236).

7 Conclusion

This paper has proven a number of robustness properties of the Volterra p-distances for functions of a single variable. These results extend previous results on Earth Mover's Distance, an increasingly popular metric in machine learning and statistical applications. Our results indicate that many of the favorable properties of EMD are shared by a wider class of metrics, which may be better suited for certain applications; for instance, it might be preferable in certain contexts to use a metric that is embeddable into ℓ_2. Of course, the current results are limited to functions of one variable and tomographic projections of functions of two variables. It is natural to consider whether the present results may be used to compare higher dimensional functions by, for instance, looking at multiple one-dimensional projections; this idea has been used previously in the context of approximating Wasserstein distances, under the name of "sliced" Wasserstein distances [13, 29, 30]. In forthcoming work, we consider a class of rapidly computable metrics between functions of two variables that exhibit a set of properties analogous to those of the Volterra distances reported here.

Acknowledgements

I thank Joe Kileel, Amit Moscovich, Rohan Rao, and Amit Singer for stimulating discussions on the papers [46], [28], and [59]. I give additional thanks to Amit Singer for helpful suggestions on an earlier version of the manuscript. I acknowledge support from NSF BIGDATA award IIS-1837992, BSF award 2018230, and NSF CAREER award DMS-2238821.

References

[1] Emmanuel Abbe, Tamir Bendory, William Leeb, João M. Pereira, Nir Sharon, and Amit Singer. Multireference alignment is easier with an aperiodic translation distribution. *IEEE Transactions on Information Theory*, 65(6):3565–3584, 2018.

[2] Charu C. Aggarwal and Chandan K. Reddy, editors. *Data Clustering: Algorithms and Applications.* CRC Press, 2014.

[3] Rohit Agrawal and Thibaut Horel. Optimal bounds between f-divergences and integral probability metrics. *The Journal of Machine Learning Research*, 22(1):5662–5720, 2021.

[4] Yariv Aizenbud and Yoel Shkolnisky. A max-cut approach to heterogeneity in cryo-electron microscopy. *Journal of Mathematical Analysis and Applications*, 479(1):1004–1029, 2019.

[5] Michael Arbel, Anna Korba, Adil Salim, and Arthur Gretton. Maximum mean discrepancy gradient flow. *Advances in Neural Information Processing Systems*, 32, 2019.

[6] Afonso S. Bandeira, Ben Blum-Smith, , Joe Kileel, Jonathan Niles-Weed, Amelia Perry, and Alexander S. Wein. Estimation under group actions: recovering orbits from invariants. *Applied and Computational Harmonic Analysis*, 2023.

[7] Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Processing Magazine*, 37(2):58–76, 2020.

[8] Tamir Bendory, Dan Edidin, William Leeb, and Nir Sharon. Dihedral multi-reference alignment. *IEEE Transactions on Information Theory*, 68(5):3489–3499, 2022.

[9] Tamir Bendory, Ariel Jaffe, William Leeb, Nir Sharon, and Amit Singer. Super-resolution multi-reference alignment. *Information and Inference: A Journal of the IMA*, 11(2):533–555, 2022.

[10] Espen Bernton, Pierre E. Jacob, Mathieu Gerber, and Christian P. Robert. On parameter estimation with the Wasserstein distance. *Information and Inference: A Journal of the IMA*, 8(4):657–676, 2019.

[11] Jean-Paul Berrut. A circular interpretation of the Euler-Maclaurin formula. *Journal of Computational and Applied Mathematics*, 189:375–386, 2006.

[12] Jean-Paul Berrut and Manfred R. Trummer. Extrapolation quadrature from equispaced samples of functions with jumps. *Numerical Algorithms*, 92:65–88, 2023.

[13] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51:22–45, 2015.

[14] Badr-Eddine Chérief-Abdellatif and Pierre Alquier. MMD-Bayes: Robust Bayesian estimation via maximum mean discrepancy. In *Symposium on Advances in Approximate Bayesian Inference*, pages 1–21, 2020.

[15] Alexander Cloninger, Brita Roy, Carley Riley, and Harlan M. Krumholz. People mover's distance: Class level geometry using fast pairwise data adaptive transportation costs. *Applied and Computational Harmonic Analysis*, 47(1):248–257, 2019.

[16] Ronald R. Coifman, Yoel Shkolnisky, Fred J. Sigworth, and Amit Singer. Graph Laplacian tomography from unknown random projections. *IEEE Transactions on Image Processing*, 17(10):1891–1899, 2008.

[17] Allison Doerr. Single-particle cryo-electron microscopy. *Nature Methods*, 13(1):23, 2016.

[18] Richard M. Dudley. *Real Analysis and Probability.* CRC Press, 2018.

[19] Rick Durrett. *Probability Theory and Examples.* Cambridge University Press, fourth edition, 2010.

[20] D.A. Edwards. On the Kantorovich-Rubinstein theorem. *Expositiones Mathematicae*, 29:387–398, 2011.

[21] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications.* John Wiley and Sons, second edition, 1999.

[22] Joachim Frank and Abbas Ourmazd. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods*, 100:61–67, 2016.

[23] Frank J. Massey, Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[24] Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.

[25] Israel Gohberg and Mark Grigorevich Krein. *Theory and Applications of Volterra Operators in Hilbert Space.* AMS, 1970.

[26] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

[27] H.G. Kellerer. Duality theorems and probability metrics. In *Proceedings of the Seventh Conference on Probability Theory*, pages 211–220. VNU Press, 1985.

[28] Joe Kileel, Amit Moscovich, Nathan Zelesko, and Amit Singer. Manifold learning with arbitrary norms. *Journal of Fourier Analysis and Applications*, 27(5):82, 2021.

[29] Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde. Generalized sliced Wasserstein distances. *Advances in Neural Information Processing Systems*, 32, 2019.

[30] Soheil Kolouri, Gustavo K. Rohde, and Heiko Hoffmann. Sliced Wasserstein distance for learning Gaussian mixture models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3427–3436, 2018.

[31] Roy R. Lederman, Joakim Andén, and Amit Singer. Hyper-molecules: on the representation and recovery of dynamical structures for applications in flexible macro-molecules in cryo-EM. *Inverse Problems*, 36(4):044005, 2020.

[32] Roy R. Lederman and Amit Singer. A representation theory perspective on simultaneous alignment and classification. *Applied and Computational Harmonic Analysis*, 49(3):1001–1024, 2020.

[33] William Leeb and Ronald Coifman. Hölder-Lipschitz norms and their duals on spaces with semigroups, with applications to Earth Mover's Distance. *Journal of Fourier Analysis and Applications*, 22(4):910–953, 2016.

[34] Elizaveta Levina and Peter Bickel. The Earth Mover's Distance is the Mallows Distance: Some insights from statistics. In *Proceedings of the Eighth IEEE International Conference on Computer Vision*, volume 2, pages 251–256. IEEE, 2001.

[35] Hstau Y. Liao, Yaser Hashem, and Joachim Frank. Efficient estimation of three-dimensional covariance and its application in the analysis of heterogeneous samples in cryo-electron microscopy. *Structure*, 23(6):1129–1137, 2015.

[36] G. Rubinstein L.V. Kantorovich. On a space of completely additive functions. *Vestnik Leningradskogo Universiteta*, 13(7):5259, 1958.

[37] G.P. Akilov L.V. Kantorovich. *Functional Analysis*. 2nd edition, 1982.

[38] Makoto Maejima and Svetlozar T. Rachev. An ideal metric and the rate of convergence to a self-similar process. *The Annals of Probability*, 15(2):708–727, 1987.

[39] Gal Mishne, Ronen Talmon, Ron Meir, Jackie Schiller, Maria Lavzin, Uri Dubin, and Ronald R. Coifman. Hierarchical coupled-geometry analysis for neuronal structure and activity pattern discovery. *IEEE Journal of Selected Topics in Signal Processing*, 10(7):1238–1253, 2016.

[40] Caroline Moosmüller and Alexander Cloninger. Linear Optimal Transport Embedding: Provable fast Wasserstein distance computation and classification for nonlinear problems. *Information and Inference: A Journal of the IMA*, 12(1):363–389, 2023.

[41] Alfred Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29:429–443, 1997.

[42] Frank Natterer. *The Mathematics of Computerized Tomography*. Wiley, 1986.

[43] Amelia Perry, Jonathan Weed, Afonso Bandeira, Philippe Rigollet, and Amit Singer. The sample complexity of multi-reference alignment. *SIAM Journal on Mathematics of Data Science*, 1(3):497–517, 2019.

[44] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[45] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.

[46] Rohan Rao, Amit Moscovich, and Amit Singer. Wasserstein k-means for clustering tomographic projections. In *NeurIPS 2020, Machine Learning for Structural Biology (MLSB) Workshop*, 2020.

[47] Philippe Rigollet and Jonathan Weed. Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Information and Inference: A Journal of the IMA*, 8(4):691717, 2019.

[48] Lior Rokach. A survey of clustering algorithms. In *Data Mining and Knowledge Discovery Handbook*, pages 269–298. Springer, 2009.

[49] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The Earth Mover's Distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

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

[51] Sjors H. W. Scheres. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods in Enzymology*, 579:125–157, 2016.

[52] Amit Singer and Fred J. Sigworth. Computational methods for single-particle electron cryomicroscopy. *Annual Review of Biomedical Data Science*, 3:163–190, 2020.

[53] Amit Singer and Hau-Tieng Wu. Two-dimensional tomography from noisy projections taken at unknown random directions. *SIAM Journal on Imaging Sciences*, 6(1):136–175, 2013.

[54] Bharath K. Sriperumbudur, Kenji Fukumizu, Arthur Gretton, Bernhard Schölkopf, and Gert R.G. Lanckriet. On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6:1550–1599, 2012.

[55] Bogdan Toader, Fred J. Sigworth, and Roy R. Lederman. Methods for Cryo-EM single particle reconstruction of macromolecules having continuous heterogeneity. *Journal of Molecular Biology*, 435(9):168020, 2023.

[56] Rosanna Verde and Antonio Irpino. Dynamic clustering of histogram data: using the right metric. In *Selected Contributions in Data Analysis and Classification*, pages 123–134. Springer, 2007.

[57] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. AMS, 2001.

[58] Cédric Villani. *Optimal Transport: Old and New*. Springer, 2008.

[59] Nathan Zelesko, Amit Moscovich, Joe Kileel, and Amit Singer. Earthmover-based manifold learning for analyzing molecular conformation spaces. In *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1715–1719, 2020.

[60] Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18(2):176–185, 2021.