On the robustness of certain norms

William Leeb*

Abstract

We study a family of norms defined for functions on an interval. These norms are obtained by taking the p-norm of the Volterra operator applied to the function. The corresponding distances have been previously studied in the context of comparing probability measures, and special cases include the Earth Mover's Distance and Kolmogorov Metric. We study their properties for general signals, and show that they are robust to additive noise. We also show that the norm-induced distance between a function and its perturbation is bounded by the size of the perturbation, and that the distance between one-dimensional projections of a two-dimensional function is bounded by the size of the difference in projection directions. The results are illustrated in numerical experiments.

1 Introduction

We study a certain family of norms for single-variable functions on an interval. These norms, denoted $\|f\|_p^*$, are the p-norms of the Volterra operator (i.e. the indefinite integral operator) applied to f. We also consider a discrete norm $\|\mathbf{x}\|_p^*$ for vectors $\mathbf{x}$ in $\mathbb{R}^n$, which is the p-norm of the discrete Volterra operator applied to $\mathbf{x}$. For this reason, we call the norm $\|\cdot\|_p^*$ the *Volterra p-norm*. We prove two robustness properties of the Volterra norms, neither of which hold for the ordinary p-norms:

- **Robustness to perturbations.** The norm-induced distance is robust to changes of variables/rescalings. That is, the distance $\|f - \tilde{f}\|_p^*$ 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 projections.** The norm-induced distance between two one-dimensional projections of a fixed two-dimensional function is bounded by the size of the difference in projection directions.
- **Robustness to noise.** The discrete norm $\|\mathbf{z}\|_p^*$ of a noise vector $\mathbf{z}$ in $\mathbb{R}^n$ 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.

When applied to the difference $f = P - Q$ of two probability densities P and Q, the Volterra p-norm $\|P - Q\|_p^*$ is equal to a distance between random variables introduced in [8]. Two special cases are of particular significance. When $p = \infty$, $\|P - Q\|_\infty^*$ is the Kolmogorov Metric (KM), which is the maximum absolute difference between the cumulative distribution functions [4]. The KM has been widely employed in goodness-of-fit testing, in what is called the Kolmogorov-Smirnov test [3]. When $p = 1$, $\|P-Q\|_1^*$ is the Earth Mover's Distance (EMD) between P and Q, also known as the 1-Wasserstein distance; it is characterized as the minimal cost of transforming one density into the other [13, 14]. EMD is a popular metric in machine learning and statistical applications [10, 12].

The theory we develop in this work is not restricted to measuring distances between random variables. Rather, our results are applicable to comparing two noisy, sampled signals; the values attained by each signal need not be nonnegative, and their integrals need not be equal.

The specific motivation for this work is the recent papers [11], [6], and [15], which use approximations to EMD to compare and cluster noisy images and volumes. While these works both show that EMD-type metrics perform well in the presence of additive noise, no theoretical justification is provided. The present work grew out of an attempt to address this question, albeit in the simpler one-dimensional setting. At the

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

same time, we explore additional robustness properties of EMD, and show that these properties hold for the larger class of distances induced from Volterra p-norms. In particular, the robustness to projections (in the three-dimensional setting) is proven in [11] to hold for Wasserstein distances; we prove in this work that the same property holds for the metrics induced from the Volterra p-norms, which includes 1-Wasserstein as a special case.

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. In Section 4, we prove that distances between perturbed functions are small. In Section 5, we prove robustness to projections. In Section 6, we analyze the behavior of the discrete norm. In Section 7, we present the results of numerical experiments. Section 8 contains detailed proofs of the main results. Section 9 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, [2].

2.1 The p-norms

Let $f : [0,1] \to \mathbb{R}$ be any measurable function. For any value $p \in [1,\infty)$, the p-norm is defined as follows:

$$\|f\|_p = \left(\int_0^1 |f(x)|^p dx \right)^{1/p}. \tag{1}$$

For $p = \infty$, we define

$$\|f\|_\infty = \operatorname*{ess\,sup}_{0 \le x \le 1} |f(x)|. \tag{2}$$

As is well-known, $\|f\|_p \le \|f\|_q$ if $p \le q$. We denote by L^p the set of all functions f on $[0,1]$ with $\|f\|_p < \infty$.

We define the inner product between two real-valued functions on $[0,1]$ as follows:

$$\langle f, g \rangle = \int_0^1 f(x) g(x) dx. \tag{3}$$

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

$$\|\mathbf{x}\|_p = \left(\frac{1}{n} \sum_{j=1}^n |x_j|^p \right)^{1/p}, \tag{4}$$

and when $p = \infty$,

$$\|\mathbf{x}\|_\infty = \max_{1 \le k \le n} |x_k|. \tag{5}$$

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

2.2 Absolute continuity

A function G on $[0,1]$ is said to be absolutely continuous if it can be written as

$$G(x) = G(0) + \int_0^x g(t) dt \tag{6}$$

for an integrable function g. If G is absolutely continuous, then it is differentiable almost everywhere, and $G' = g$. We denote by $\mathcal{A}_0$ the set of absolutely continuous functions G satisfying $G(1) = 0$; these functions may be written as

$$G(x) = -\int_x^1 g(t)dt \tag{7}$$

where $g = G'$ almost everywhere.

2.3 The Volterra operator

The Volterra operator $\mathcal{V}$ is defined on L^1 by

$$(\mathcal{V}f)(x) = \int_0^x f(t)dt. \tag{8}$$

We note that this is only the simplest of a large family of operators that have been widely-studied [5].

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

$$(\mathcal{V}^*f)(x) = -\int_x^1 f(t)dt. \tag{9}$$

This operator satisfies

$$\langle \mathcal{V}f, g\rangle = \langle f, \mathcal{V}^*g\rangle \tag{10}$$

where f and g are any two integrable functions on $[0, 1]$.

If $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector in $\mathbb{R}^n$, we define the discrete Volterra operator as follows:

$$(\mathcal{V}\mathbf{x})_k = \frac{1}{n}\sum_{j=1}^{k} x_k. \tag{11}$$

Note that in the normalization we adopt here, the sum is divided by n, not k.

3 The Volterra p-norms

3.1 Continuous norm

For any value $p \in [1, \infty]$, we introduce the following norm:

$$\|f\|_p^* = \|\mathcal{V}f\|_p. \tag{12}$$

Concretely, when $p < \infty$,

$$\|f\|_p^* = \left(\int_0^1 \left|\int_0^x f(t)dt\right|^p dx\right)^{1/p}, \tag{13}$$

and when $p = \infty$,

$$\|f\|_\infty^* = \operatorname*{ess\,sup}_{0\le x\le 1}\left|\int_0^x f(t)dt\right|. \tag{14}$$

We will refer to the Volterra p-norm $\|f - g\|_p^*$ of the difference $f - g$ of two functions as the *Volterra p-distance* between f and g.

Remark 1. When $p = \infty$ and P and Q are two probability densities, the Volterra ∞-distance is known as the Kolmogorov Metric between P and Q [4]: $\mathrm{KM}(P, Q) = \|P - Q\|_\infty^*$. The KM is used the Kolmogorov-Smirnov goodness-of-fit test in statistics [3].

Remark 2. When $p = 1$ and P and Q are two probability densities, the Volterra 1-distance is known as the Earth Mover's Distance between P and Q [13, 14]: $\mathrm{EMD}(P, Q) = \|P - Q\|_1^*$. The metric $\mathrm{EMD}(P, Q)$ may be defined equivalently as the solution to a transportation problem:

$$\mathrm{EMD}(P, Q) = \min_{\Pi \in \mathcal{M}(P,Q)} \int_0^1 \int_0^1 |x - y| \Pi(x, y) dx dy \tag{15}$$

where $\mathcal{M}(P, Q)$ denotes the space of all probability measures on $[0, 1] \times [0, 1]$ with marginals equal to P and Q, respectively. That is, $\Pi \in \mathcal{M}(P, Q)$ if for all x,

$$\int_0^1 \Pi(x, y) dy = P(x), \tag{16}$$

and for all y,

$$\int_0^1 \Pi(x, y) dx = Q(y). \tag{17}$$

The p-Wasserstein distance $W_p(P, Q)$ is defined as

$$W_p(P, Q) = \min_{\Pi \in \mathcal{M}(P,Q)} \left(\int_0^1 \int_0^1 |x - y|^p \Pi(x, y) dx dy \right)^{1/p}. \tag{18}$$

It is known [12] that $W_p(P, Q)$ may be written as follows:

$$W_p(P, Q) = \|(\mathcal{V}P)^{-1} - (\mathcal{V}Q)^{-1}\|_p \tag{19}$$

where $(\mathcal{V}P)^{-1}$ and $(\mathcal{V}Q)^{-1}$ denote the functional inverses of $\mathcal{V}P$ and $\mathcal{V}Q$, respectively. When $p = 1$, it is also true that $W_1(P, Q) = \|\mathcal{V}P - \mathcal{V}Q\|_p = \|P - Q\|_1^*$; when $p > 1$, however, the p-Wasserstein distance $W_p(P, Q)$ will not necessarily be equal to the Volterra p-distance $\|\mathcal{V}P - \mathcal{V}Q\|_p^*$.

3.2 Discrete norm

We define the discrete Volterra norm. If $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector in $\mathbb{R}^n$, define:

$$\|\mathbf{x}\|_p^* = \|\mathcal{V}\mathbf{x}\|_p. \tag{20}$$

Concretely, when $p < \infty$,

$$\|\mathbf{w}\|_p^* \left(\frac{1}{n} \sum_{k=1}^n \left| \frac{1}{n} \sum_{j=1}^k w_k \right|^p \right)^{1/p}. \tag{21}$$

and when $p = \infty$,

$$\|\mathbf{w}\|_\infty^* = \max_{1 \le k \le n} \left| \frac{1}{n} \sum_{j=1}^k w_k \right|. \tag{22}$$

3.3 Variational formulation of $\|f\|_p^*$

We show the following:

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

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{23}$$

Then for any bounded, measurable function f,

$$\|f\|_p^* = \sup_{G \in \mathcal{A}_0 : \|G'\|_q \le 1} \int_0^1 G(x) f(x) dx. \tag{24}$$

Remark 3. When $f = P - Q$ is the difference of two probability densities, the right side of (24) is equal to a distance introduced in [8], and Proposition 3.1 follows immediately from Theorem 1 in [8].

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

$$\|f\|_p^* = \|\mathcal{V}f\|_p = \sup_{g:\|g\|_q \le 1} \int_0^1 (\mathcal{V}f)(x)g(x)dx = \sup_{g:\|g\|_q \le 1} \langle \mathcal{V}f, g\rangle = \sup_{g:\|g\|_q \le 1} \langle f, \mathcal{V}^* g\rangle. \tag{25}$$

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:

$$\|f\|_p^* = \|\mathcal{V}f\|_p = \sup_{g:\|g\|_q \le 1} \langle f, \mathcal{V}^* g\rangle = \sup_{G \in \mathcal{A}_0 : \|G'\|_q \le 1} \langle f, G\rangle, \tag{26}$$

which completes the proof. □

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

Corollary 3.2. *Let $I \subseteq [0,1]$ be a subinterval, and let $f \in L^\infty$ be supported on I. If $\mathcal{T} : L^\infty \to L^\infty$ is a linear transformation satisfying*

$$\|(\mathcal{T}^*(G) - G)\chi_I\|_q \le \epsilon \|G'\|_q \tag{27}$$

for all functions G in $\mathcal{A}_0$, then

$$\|\mathcal{T}(f) - f\|_p^* \le \epsilon \|f\|_p. \tag{28}$$

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:

$$\begin{aligned}
\|\mathcal{T}(f) - f\|_p^* &= \sup_{G \in \mathcal{A}_0 : \|G'\|_q \le 1} \langle \mathcal{T}(f) - f, G\rangle \\
&= \sup_{G \in \mathcal{A}_0 : \|G'\|_q \le 1} \langle f, (\mathcal{T}^*(G) - G)\chi_I\rangle \\
&\le \sup_{G \in \mathcal{A}_0 : \|G'\|_q \le 1} \|f\|_p \|(\mathcal{T}^*(G) - G)\chi_I\|_q \\
&\le \epsilon \|f\|_p,
\end{aligned} \tag{29}$$

completing the proof. □

4 Distance between perturbations

As a consequence of Corollary 3.2, we will show that if f and $\tilde{f}$ are two functions on $[0,1]$ related by a perturbation, then their distance $\|f - \tilde{f}\|_p^*$ is bounded by the size of the perturbation. More precisely, we have the following result:

Theorem 4.1. *Let $f \in L^\infty$ be supported on an interval $I \subseteq [0,1]$. Suppose $\psi : I \to [0,1]$ is a continuously differentiable, monotonically increasing mapping, and $\rho : \psi(I) \to \mathbb{R}$ is a bounded, measurable function. Let $f_{\psi,\rho}$ be equal to $f_{\psi,\rho}(x) = f(\psi^{-1}(x))\rho(x)$ on $\psi(I)$, and 0 otherwise. Then*

$$\|f - f_{\psi,\rho}\|_p^* \le \epsilon \|f\|_p, \tag{30}$$

where

$$\epsilon = \max_{x \in I} |x - \psi(x)| + \max_{x \in I} |1 - \rho(\psi(x))\psi'(x)|. \tag{31}$$

Remark 4. The function ψ perturbs the domain of f, whereas ρ perturbs the range of f. The value ϵ defined by (31) is a natural measure of the "size" of the joint perturbation linking f to $f_{\psi,\rho}$. The term $\max_{x\in I}|x-\psi(x)|$ controls the amount by which the change of variables ψ can move any point in the domain of f; it will be small if ψ does not move any value x very far. The term $\max_{x\in I}|1-\rho(\psi(x))\psi'(x)|$ will be small if both ψ' and the rescaling function ρ are close to 1.

Remark 5. The case where $\rho = (\psi^{-1})'$ follows from Proposition 14 in [7]. In this setting, the integrals of f and $f_{\psi,\rho}$ are identical, $\int_0^1 f = \int_0^1 f_{\psi,\rho}$.

We note two special cases of Theorem 4.1.

Corollary 4.2. *Suppose $f:[0,1]\to\mathbb{R}$ is a bounded, measurable function supported in a subinterval $I\subset[0,1]$. Suppose that $\epsilon>0$ and $J=\{x+\epsilon : x\in I\}\subset[0,1]$, and let f_ϵ be the shift of f by ϵ; that is,*

$$f_\epsilon(x) = \begin{cases} f(x-\epsilon), & \text{if } x\in J \\ 0, & \text{otherwise} \end{cases}. \tag{32}$$

Then for all $p\in[1,\infty]$, $\|f-f_\epsilon\|_p^* \le \epsilon\|f\|_p$.

Corollary 4.3. *Suppose X is a random variable with values in $[0,1]$ and density P, and let $\psi:[0,1]\to[0,1]$ be a continuously differentiable, monotonically increasing mapping with $\psi(0)=0$ and $\psi(1)=1$. Let $Y=\psi(X)$, with density Q. Then for all $p\in[1,\infty]$,*

$$\|P-Q\|_p^* \le \max_{0\le x\le 1}|x-\psi(x)| \equiv \epsilon. \tag{33}$$

In particular, $\mathrm{EMD}(P,Q)\le\epsilon$ *and* $\mathrm{KM}(P,Q)\le\epsilon$.

Corollary 4.2 follows immediately from Theorem 4.1. Corollary 4.3 follows from the observation that Y has density $Q(y)=P(\psi^{-1}(y))(\psi^{-1})'(y)$.

5 Distance between projections

In this section we let $F:\mathbb{R}^2\to\mathbb{R}$ denote a function of two variables that is supported on $\mathbb{D}\subset\mathbb{R}^2$, the disc of radius $1/2$ and center $(0,0)$. For a given angle θ, we define the projection of F as follows:

$$f_\theta(x) = \int_{-1/2}^{1/2} F(\cos(\theta)x+\sin(\theta)y, \cos(\theta)y-\sin(\theta)x)dy \tag{34}$$

The function $f_\theta:[-1/2,1/2]\to\mathbb{R}$ is the 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 [9], and is a standard transformation in scientific imaging. We claim the following result:

Theorem 5.1. *For all angles θ and φ with $|\theta-\varphi|\le\pi/4$ and all $p\in[1,\infty]$,*

$$\|f_\theta - f_\varphi\|_p^* \le \|F\|_p\sqrt{\frac{1-\cos(\theta-\varphi)}{2}}, \tag{35}$$

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

Note that the term $\sqrt{1-\cos(\theta-\varphi)}$ may be bounded above by $|\theta-\varphi|$. Consequently, Theorem 5.1 states that the Volterra p-distance between two projections f_θ and f_φ of F may be bounded by the size of the difference in projection directions. The bound only depends on the underlying function F through its p-norm.

Remark 6. A result similar to Theorem 5.1 for p-Wasserstein distances and the three-dimensional Radon transform was proven in the paper [11]. In that work, the fact that the Wasserstein distance is insensitive to rotations of projections allows for robust clustering of images taken from different viewing directions which are not known a priori, as occurs in cryoelectron microscopy.

6 Asymptotic behavior of the discrete norm

In this section, we quantify the robustness of the discrete Volterra p-norms to additive noise. We first consider the noiseless case, where the norm is applied to a vector of samples from a function.

Theorem 6.1. *Let $f : [0,1] \to \mathbb{R}$ be a Lipschitz function, and suppose $|f(x)| \leq L$ and $|f(x)-f(y)| \leq L|x-y|$ for all x and y in $[0,1]$. Let $\mathbf{x}_n = (x_1, \ldots, x_n)$ have entries $x_j = f(j/n)$, $1 \leq j \leq n$. Then*

$$\left| \|\mathbf{x}_n\|_p^* - \|f\|_p^* \right| \leq C\frac{L}{n}, \tag{36}$$

where C does not depend on p, f, or n.

Theorem 6.1 applies when the input vector $\mathbf{x}_n$ is sampled from an underlying Lipschitz function. The next result applies to input vectors of pure noise:

Theorem 6.2. *Let $\sigma_1, \sigma_2, \ldots, \sigma_n, \ldots$ be a sequence of positive numbers, and let*

$$\overline{\sigma}_n = \sqrt{\frac{1}{n}\sum_{j=1}^{n} \sigma_j^2}. \tag{37}$$

Suppose $\mathbf{z}_n = (z_1, \ldots, z_n)$, where the z_j are independent and $z_j \sim N(0, \sigma_j^2)$. Then for all $p \in [1, \infty]$,

$$\lim_{n\to\infty} \|\mathbf{z}_n\|_p^* = 0 \tag{38}$$

holds almost surely, and

$$\mathbb{E}\|\mathbf{z}_n\|_p^* \leq C\frac{\overline{\sigma}_n}{\sqrt{n}}, \tag{39}$$

where C is a constant independent of n and p.

From Theorems 6.1 and 6.2, we immediately derive the following:

Corollary 6.3. *Let $\mathbf{x}_n$ and $\mathbf{z}_n$ be as in Theorems 6.1 and 6.2, respectively, and let $\mathbf{y}_n = \mathbf{x}_n + \mathbf{z}_n$. Then*

$$\lim_{n\to\infty} \|\mathbf{y}_n\|_p^* = \|f\|_p^* \tag{40}$$

holds almost surely, and

$$\mathbb{E}\left| \|\mathbf{y}_n\|_p^* - \|f\|_p^* \right| \leq C\frac{\overline{\sigma}_n}{\sqrt{n}}, \tag{41}$$

where C is a constant independent of n and p.

Remark 7. In the setting of Corollary 6.3, both the signal vector $\mathbf{x}_n$ and the noise vector $\mathbf{z}_n$ have comparable p-norms; consequently, $\|\mathbf{y}_n\|_p$ does not approach $\|f\|_p$ as $n \to \infty$. For example, when $p = 2$, the squared 2-norm of $\mathbf{y}_n$ converges to

$$\lim_{n\to\infty} \|\mathbf{y}_n\|_2^2 = \|f\|_2^2 + \overline{\sigma}^2 \tag{42}$$

where $\overline{\sigma} = \lim_{n\to\infty} \overline{\sigma}_n$ (assuming this limit is well-defined). By contrast, the additive noise term $\mathbf{z}_n$ has a negligible effect on the Volterra norm when n is large.

7 Numerical results

7.1 Norm under noise

To demonstrate the robustness of the Volterra norms under noise described by Corollary 6.3, we run the following experiment. We take n equispaced samples of the function f shown in the third panel of Figure 1; f is the difference of the two triangle functions shown in the upper left and upper right panels of Figure 1. Denote the vector of samples by $\mathbf{x}_n$.

We then add a vector $\mathbf{z}_n$ of iid Gaussian noise of variance 2 to $\mathbf{x}_n$, to form the vector $\mathbf{y}_n = \mathbf{x}_n + \mathbf{z}_n$. An example of $\mathbf{y}_n$, with $n = 1000$, is shown in the bottom right panel of Figure 1, with the noiseless $\mathbf{x}_n$ plotted in red for reference. For $p = 1, 2, \infty$, we evaluate the norms $\|\mathbf{y}_n\|_p^*$. For each value of n, the experiment is repeated $M = 100,000$ times. Denoting the M random signal-plus-noise vectors by $\mathbf{y}_n^{(1)}, \ldots, \mathbf{y}_n^{(M)}$, we record the average absolute error:

$$\mathrm{err}_{n,p} = \frac{1}{M} \sum_{k=1}^{M} \frac{\left| \|\mathbf{y}_n^{(k)}\|_p^* - \|f\|_p^* \right|}{\|f\|_p^*}. \tag{43}$$

Table 1 displays the values of $\mathrm{err}_{n,p}$ for each value of p and n. Figure 2 plots $\log_2(\mathrm{err}_{n,p})$ as a function of $\log_2(n)$. From both the table and the figure, we see that the average error does indeed decay like $O(1/\sqrt{n})$ as n increases, consistent with Corollary 6.3.

7.2 Distance under translation

7.2.1 Triangle function

We first illustrate Theorem 4.1, and more specifically Corollary 4.2, on the triangle functions shown in the first two panels of Figure 1. Denoting this function by f, we denote by f_ϵ the translation of f by ϵ. Figure 3, left panel, plots the normalized Volterra p-distances $\|f - f_\epsilon\|_p^*/\|f\|_p$ as functions of ϵ, for $p = 1, 2, 10, \infty$. Figure 3, right panel, plots the normalized p-norms $\|f - f_\epsilon\|_p/\|f\|_p$ as functions of ϵ.

From the left panel, we see that for each ϵ,

$$\frac{\|f - f_\epsilon\|_p^*}{\|f\|_p} \leq \epsilon. \tag{44}$$

This is predicted from Corollary 4.2. Furthermore, when $p = 1$ the inequality is an equality, but for $p > 1$ the inequality is strict.

Because the length of f's support is 0.4, when $\epsilon > 0.4$ the relative differences $\|f - f_\epsilon\|_p/\|f\|_p$ plateau and do not continue to increase with ϵ. By contrast, when $p < \infty$, $\|f - f_\epsilon\|_p^*/\|f\|_p$ grows with ϵ. Furthermore, for each value of p, the distance $\|f - f_\epsilon\|_p^*$ distinguishes a much larger range of ϵ than does the distance $\|f - f_\epsilon\|_p$.

7.2.2 Spiked function

Next, we consider the spiked function f shown in Figure 4. Again, we denote by f_ϵ the translation of f by ϵ. Figure 5, left panel, plots the normalized Volterra p-distances $\|f - f_\epsilon\|_p^*/\|f\|_p$ as functions of ϵ, for $p = 1, 2, 10, \infty$. Figure 5, right panel, plots the normalized p-norms $\|f - f_\epsilon\|_p/\|f\|_p$ as functions of ϵ.

In this example, the differences between the behaviors of the Volterra distances and the p-norm distances are even more extreme. Whereas the Volterra distances are all increasing functions of ϵ, the p-norm distances are not; their behavior depend strongly on the contours of the function f itself. Similar behavior was observed in the paper [15], which compared an approximation of EMD to Euclidean distance between a volume and its rotation; see Figure 3 and the corresponding text in [15].

7.3 Distance between projections

We illustrate the behavior described by Theorem 5.1 on the function F consisting of two Gaussian bumps

$$F(\mathbf{w}) = \exp\{-\|\mathbf{w} - \mathbf{a}\|^2/\sigma\} + \exp\{-\|\mathbf{w} - \mathbf{b}\|^2/\sigma\}, \tag{45}$$

where $\mathbf{a} = (0, 1/2)$ and $\mathbf{b} = (0, -1/2)$, and $\sigma = 1/1000$. 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 $[-1/2, 1/2] \times [-1/2, 1/2]$, along with f, $f_{\pi/16}$, and $f_{\pi/8}$.

Figure 7 plots the distances $\|f - f_\theta\|_p^*$ and $\|f - f_\theta\|_p$ for $p = 1, 2, 10, \infty$, for θ between 0 and $\pi/2$, which covers the full range of distances due to the function F's symmetry. The distances $\|f - f_\theta\|_p^*$ continue to grow with θ throughout the entire range of values, whereas $\|f - f_\theta\|_p$ plateau. Analogously to the comparison of distances between a function and its shifts, for each value of p the distance $\|f - f_\theta\|_p^*$ distinguishes a much larger range of θ than does the distance $\|f - f_\theta\|_p$.

8 Proofs

8.1 Proof of Theorem 4.1

Write $\epsilon = \epsilon_1 + \epsilon_2$, where

$$\epsilon_1 = \max_{x \in I} |x - \psi(x)|, \tag{46}$$

and

$$\epsilon_2 = \max_{x \in I} |1 - \rho(\psi(x))\psi'(x)|. \tag{47}$$

Let χ_I and χ_J be the indicator functions of I and J, respectively. Define the linear transformation $\mathcal{T} : L^\infty \to L^\infty$ by $\mathcal{T}(h)(x) = h(\psi^{-1}(x))\rho(x)$. Then the adjoint operator of $\mathcal{T}$ is $\mathcal{T}^*(G)(x) = G(\psi(x))\rho(\psi(x))\psi'(x)$.

To apply Corollary 3.2, we will show that for all $G \in \mathcal{A}_0$,

$$\|(\mathcal{T}^*(G) - G)\chi_I\|_q \leq \epsilon \|G'\|_q. \tag{48}$$

Since $G \in \mathcal{A}_0$, we may write

$$G(x) = -\int_x^1 g(t)dt \tag{49}$$

where $g = G'$ is integrable. If $\|g\|_q$ is infinite, the necessary bound is vacuous. We may therefore assume, without loss of generality, that $\|g\|_q = 1$.

We have

$$\|(\mathcal{T}^*(G) - G)\chi_I\|_q \leq \|(G \circ \psi - G)\chi_I\|_q + \|(\mathcal{T}^*(G) - G \circ \psi)\chi_I\|_q. \tag{50}$$

We will first show that

$$\|(G \circ \psi - G)\chi_I\|_q \leq \epsilon_1. \tag{51}$$

We first assume $p > 1$, i.e. $q < \infty$. Let I_x be the interval $[x, \psi(x)]$ if $x \leq \psi(x)$, or $[\psi(x), x]$ if $\psi(x) \leq x$. Let $\chi(x, t)$ be 1 if and only if $t \in I_x$, and 0 otherwise. Then:

$$\int_I |G(\psi(x)) - G(x)|^q dx = \int_I \left| \int_{I_x} g(t)dt \right|^q dx = \int_I \left| \int_0^1 g(t)\chi(x, t)dt \right|^q dx. \tag{52}$$

Lemma 8.1. *For all $x \in I$,*

$$\int_0^1 \chi(x, t)dt \leq \epsilon_1. \tag{53}$$

For all $t \in [0, 1]$,

$$\int_I \chi(x, t)dx \leq \epsilon_1. \tag{54}$$

Proof of Lemma 8.1. For the first inequality

$$\int_0^1 \chi(x,t)dt = \int_{I_x} 1dt = |I_x| = |x - \psi(x)| \leq \epsilon_1. \tag{55}$$

For the second inequality, first suppose that there is some $x \leq t$ with $t \in I_x$; note that for such x, $I_x = [x, \psi(x)]$, and so $x \leq \psi(x)$. Let x^* be the smallest such x. Then $x^* \leq t \leq \psi(x^*)$. We claim that for all $x > t$, $t \notin I_x$. Indeed, since ψ is increasing and $x > t \geq x^*$, we have $\psi(x) > \psi(x^*) \geq t$. 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 $x^* \leq t \leq \psi(x^*)$ and $|x^* - \psi(x^*)| \leq \epsilon_1$, it follows that $|t - x^*| \leq \epsilon_1$ too. Therefore,

$$\int_I \chi(x,t)dx \leq \int_{x^*}^t 1dx = |t - x^*| \leq \epsilon_1. \tag{56}$$

The same reasoning applies if there is some $x \geq t$ with $t \in I_x$. □

From Lemma 8.1 and Theorem 6.18 in [2],

$$\left(\int_0^1 |G(\psi(x)) - G(x)|^q dx\right)^{1/q} = \left(\int_0^1 \left|\int_0^1 g(t)\chi(x,t)dt\right|^q dx\right)^{1/q} \leq \epsilon_1 \|g\|_q = \epsilon_1. \tag{57}$$

which completes the proof of (51) when $p > 1$. We now handle the case $p = 1$, showing that:

$$\|(G \circ \psi - G)\chi_I\|_\infty \leq \epsilon_1. \tag{58}$$

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

$$|G(\psi(x)) - G(x)| = \left|\int_{I_x} g(t)dt\right| \leq |I_x| \|g\|_\infty = |x - \psi(x)| \leq \epsilon_1, \tag{59}$$

completing the proof of (51) for all $p \in [1, \infty]$.

Next, we will show that

$$\|(\mathcal{T}^*(G) - G \circ \psi)\chi_I\|_q \leq \epsilon_2. \tag{60}$$

When $p > 1$, i.e. $q < \infty$, we may write

$$\begin{aligned}
\|(\mathcal{T}^*(G) - G \circ \psi)\chi_I\|_q &= \left(\int_I |G(\psi(x))\rho(\psi(x))\psi'(x) - G(\psi(x))|^q dx\right)^{1/q} \\
&= \left(\int_I |G(\psi(x))|^q |\rho(\psi(x))\psi'(x) - 1|^q dx\right)^{1/q} \\
&\leq \left(\max_{x \in I} |\rho(\psi(x))\psi'(x) - 1|\right) \left(\int_0^1 |G(\psi(x))|^q dx\right)^{1/q}.
\end{aligned} \tag{61}$$

From Minkowski's Integral Inequality,

$$\begin{aligned}
\left(\int_0^1 |G(\psi(x))|^q dx\right)^{1/q} &= \left(\int_0^1 \left|\int_0^{\psi(x)} g(t)dt\right|^q dx\right)^{1/q} \\
&\leq \int_0^1 \left(\int_0^1 |g(t)|^q dt\right)^{1/q} dx \\
&= 1.
\end{aligned} \tag{62}$$

Finally, we show that $\|(\mathcal{T}^*(G) - G \circ \psi)\chi_I\|_\infty \leq \epsilon_2$. For all $x \in I$,

$$|(\mathcal{T}^*G)(x) - G(\psi(x))| = |G(\psi(x))(\rho(\psi(x)))\psi'(x) - 1)| \leq \epsilon_2 \|G\|_\infty \leq \epsilon_2 \|g\|_\infty = \epsilon_2, \tag{63}$$

completing the proof.

8.2 Proof of Theorem 5.1

Without loss of generality, we assume that $\varphi = 0$ and $c = \cos(\theta)$ and $s = \sin(\theta)$ are both positive. Denote $f = f_\varphi = f_0$. Assume first that $p > 1$, and let $q < \infty$ denote the conjugate exponent. Take any absolutely continuous function G on $[-1/2, 1/2]$ with $\|G'\|_q \leq 1$. Then we must bound the integral

$$\int_{-1/2}^{1/2} (f(x) - f_\theta(x))G(x)dx. \tag{64}$$

We have

$$\begin{aligned}\int_{-1/2}^{1/2} f_\theta(x)G(x)dx &= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} F(cx - sy, cy + sx)G(x)dxdy \\ &= \int_{\mathbb{D}} F(cx - sy, cy + sx)G(x)dxdy, \end{aligned}\tag{65}$$

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

$$\begin{aligned}\int_{-1/2}^{1/2} f_\theta(x)G(x)dx &= \int_{\mathbb{D}} F(cx - sy, cy + sx)G(x)dxdy \\ &= \int_{\mathbb{D}} F(x, y)G(cx + sy)dxdy. \end{aligned}\tag{66}$$

Consequently,

$$\begin{aligned}\int_{-1/2}^{1/2} (f(x) - f_\theta(x))G(x)dx &= \int_{\mathbb{D}} F(x, y)(G(x) - G(cx + sy))dxdy \\ &\leq \|F\|_p \left(\int_{\mathbb{D}} |G(x) - G(cx + sy)|^q dxdy\right)^{1/q}. \end{aligned}\tag{67}$$

Write $g = G'$. 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. We then have:

$$\begin{aligned}\left(\int_{\mathbb{D}} |G(x) - G(cx + sy)|^q dxdy\right)^{1/q} &= \left(\int_{\mathbb{D}} \left|\int_{I_{x,y}} g(t)dt\right|^q dxdy\right)^{1/q} \\ &= \left(\int_{\mathbb{D}} \left|\int_{-1/2}^{1/2} g(t)\chi(x, y, t)dt\right|^q dxdy\right)^{1/q}. \end{aligned}\tag{68}$$

Lemma 8.2. *For all $(x, y) \in \mathbb{D}$,*

$$\int_{-1/2}^{1/2} \chi(x, y, t)dt \leq \sqrt{\frac{1 - c}{2}}, \tag{69}$$

and for all $t \in [-1/2, 1/2]$,

$$\int_{\mathbb{D}} \chi(x, y, t)dt \leq \frac{\sqrt{1 - c}}{2c\sqrt{1 + c}}. \tag{70}$$

Assuming Lemma 8.2, we can conclude the proof of Theorem using Theorem 6.18 in [2], and the fact that

$$\frac{1}{2\cos(\theta)\sqrt{1 + \cos(\theta)}} \leq \frac{1}{\sqrt{2}} \tag{71}$$

whenever $|\theta| \leq \pi/4$. We now turn to the proof of Lemma 8.2.

Proof of Lemma 8.2. For the first inequality, we apply the Cauchy-Schwarz inequality and the fact that $\sqrt{x^2+y^2} \leq 1/2$ to get

$$\int_{-1/2}^{1/2} \chi(x,y,t)dt = |I_{x,y}| = |x - cx - sy| = |(1-c)x - sy| \leq \frac{1}{2}\sqrt{(1-c)^2+s^2} = \sqrt{\frac{1-c}{2}}. \tag{72}$$

For the second inequality, we write

$$\int_{\mathbb{D}} \chi(x,y,t)dt \leq \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} \chi(x,y,t)dt = \int_{\mathbf{R}} \chi(x,y,t)dt + \int_{\mathbf{S}} \chi(x,y,t)dt = |\mathbf{R}| + |\mathbf{S}|, \tag{73}$$

where

$$\mathbf{R} = \{(x,y) \in [-1/2,1/2] \times [-1/2,1/2] : x \leq t \leq cx + sy\}, \tag{74}$$

and

$$\mathbf{S} = \{(x,y) \in [-1/2,1/2] \times [-1/2,1/2] : cx + sy \leq t \leq x\}. \tag{75}$$

First, we bound $\int_{\mathbf{R}} \chi(x,y,t)dxdy$. If $(x,y) \in \mathbf{R}$, then $x \leq cx + sy$, and consequently

$$x \leq \frac{s}{1-c}y. \tag{76}$$

On the other hand, we must also have $t \leq cx + sy$ and consequently

$$\frac{t-sy}{c} \leq x. \tag{77}$$

Combining (76) and (77),

$$\frac{s}{1-c}y \geq \frac{t-sy}{c} \tag{78}$$

or equivalently

$$y \geq \frac{1-c}{s}t. \tag{79}$$

Therefore,

$$\begin{aligned}\int_{\mathbf{R}} \chi(x,y,t)dxdy &\leq \int_{\frac{1-c}{s}t}^{1/2}\int_{\frac{t-sy}{c}}^{t} 1dxdy \\ &= \int_{\frac{1-c}{s}t}^{1/2}\left[t - \frac{t-sy}{c}\right]dy \\ &= \frac{(1-c)^2}{2cs}t^2 - \frac{1}{2}\left(\frac{1-c}{c}\right)t + \frac{s}{8c}.\end{aligned} \tag{80}$$

Next, we bound $\int_{\mathbf{S}} \chi(x,y,t)dxdy$. If $(x,y) \in \mathbf{S}$, then $cx + sy \leq x$ and consequently

$$x \geq \frac{s}{1-c}y. \tag{81}$$

On the other hand, we must also have $t \geq cx + sy$ and consequently

$$\frac{t-sy}{c} \geq x. \tag{82}$$

Combining (81) and (82),

$$\frac{s}{1-c}y \le \frac{t-sy}{c} \tag{83}$$

or equivalently

$$y \le \frac{1-c}{s}t. \tag{84}$$

Therefore,

$$\begin{aligned}\int_{\mathbf{S}} \chi(x,y,t)dxdy &\le \int_{-1/2}^{\frac{1-c}{s}t} \int_{t}^{\frac{t-sy}{c}} dxdy \\ &= \int_{-1/2}^{\frac{1-c}{s}t} \left[\frac{t-sy}{c} - t\right] dy \\ &= \frac{(1-c)^2}{2cs}t^2 + \frac{1}{2}\left(\frac{1-c}{c}\right)t + \frac{s}{8c}. \end{aligned} \tag{85}$$

Adding the integrals over $\mathbf{R}$ and $\mathbf{S}$, and using that $|t| \le 1/2$ and $s = \sqrt{1-c^2} = \sqrt{(1-c)(1+c)}$, we obtain the following bound:

$$\begin{aligned}\int_{\mathbb{D}} \chi(x,y,t)dt &\le \int_{\mathbf{R}} \chi(x,y,t)dt + \int_{\mathbf{S}} \chi(x,y,t)dt \\ &= \frac{(1-c)^2}{2cs}t^2 - \frac{1}{2}\left(\frac{1-c}{c}\right)t + \frac{s}{8c} + \frac{(1-c)^2}{2cs}t^2 + \frac{1}{2}\left(\frac{1-c}{c}\right)t + \frac{s}{8c} \\ &= \frac{(1-c)^2}{cs}t^2 + \frac{s}{4c} \\ &\le \frac{(1-c)^2}{4cs} + \frac{s}{4c} \\ &= \frac{\sqrt{1-c}}{2c\sqrt{1+c}}. \end{aligned} \tag{86}$$

This completes the proof for $q < \infty$. The proof for $q = \infty$ follows by taking the limit as $q \to \infty$ and using $\|\cdot\|_\infty = \lim_{p\to\infty} \|\cdot\|_p$. □

8.3 Proof of Theorem 6.1

We begin with an elementary lemma, whose proof is provided for completeness.

Lemma 8.3. *Suppose f is function on the interval $[0,c]$ with Lipschitz constant L, and let $m \ge 1$ be an integer. Let $x_k = ck/m$. Then*

$$\left|\frac{c}{m}\sum_{k=1}^{m} f(x_k) - \int_0^c f(t)dt\right| \le \frac{cL}{m}. \tag{87}$$

Proof of Lemma 8.3. By the Mean Value Theorem, there is some $x_k^* \in [c(k-1)/m, ck/m]$ satisfying

$$\frac{1}{m}f(x_k^*) = \frac{1}{c}\int_{c(k-1)/m}^{ck/m} f(x)dx. \tag{88}$$

Since f has Lipschitz constant L, if $E_k = f(x_k) - f(x_k^*)$, then $|E_k| \le cL/m$. Consequently,

$$\begin{aligned}\frac{1}{m}\sum_{k=1}^{m} f(x_k) &= \frac{1}{m}\sum_{k=1}^{m} f(x_k^*) + \frac{1}{m}\sum_{k=1}^{m} E_k \\ &= \frac{1}{c}\sum_{k=1}^{m}\int_{c(k-1)/m}^{ck/m} f(x)dx + \frac{1}{m}\sum_{k=1}^{m} E_k \\ &= \int_0^c f(x)dx + \frac{1}{m}\sum_{k=1}^{m} E_k, \end{aligned} \tag{89}$$

and so

$$\left|\frac{1}{m}\sum_{k=1}^{m} f(x_k) - \frac{1}{c}\int_0^c f(t)dt\right| \le \frac{1}{m}\sum_{k=1}^{m}|E_k| \le \frac{cL}{m}. \tag{90}$$

Multiplying each side by c completes the proof. □

Now suppose f has Lipschitz constant L on $[0,1]$, and let $\mathbf{x} = \mathbf{x}_n = (x_1, \ldots, x_n)$, $x_j = f(j/n)$. Let

$$E(k,n) = \frac{1}{n}\sum_{j=1}^{k} f(j/n) - \int_0^{k/n} f(t)dt = (\mathcal{V}\mathbf{x})_k - (\mathcal{V}f)(k/n). \tag{91}$$

Applying Lemma 8.3 to the interval $[0, k/n]$,

$$|E(k,n)| \le (k/n)^2 L/k = Lk/n^2 \le L/n, \tag{92}$$

We therefore have

$$\begin{aligned}\left|\|\mathbf{x}\|_p^* - \left(\frac{1}{n}\sum_{k=1}^{n}|(\mathcal{V}f)(k/n)|^p\right)^{1/p}\right| &= \left|\left(\frac{1}{n}\sum_{k=1}^{n}|(\mathcal{V}\mathbf{x})_k|^p\right)^{1/p} - \left(\frac{1}{n}\sum_{k=1}^{n}|(\mathcal{V}f)(k/n)|^p\right)^{1/p}\right| \\ &\le \left(\frac{1}{n}\sum_{k=1}^{n}|E(k,n)|^p\right)^{1/p} \\ &\le \frac{L}{n}. \end{aligned} \tag{93}$$

Next, since $|f(x)| \le L$ for all $x \in [0,1]$, the function $|(\mathcal{V}f)(x)|^p$ has Lipschitz constant pL^p. Consequently, applying Lemma 8.3 on the interval $[0,1]$ yields:

$$\left|\frac{1}{n}\sum_{k=1}^{n}|(\mathcal{V}f)(k/n)|^p - \int_0^1 |(\mathcal{V}f)(x)|^p\,dx\right| \le \frac{pL^p}{n}, \tag{94}$$

Since $|f(x)| \le L$ for all x in $[0,1]$, it follows that $|(\mathcal{V}f)(x)| \le L$ too, and consequently:

$$\frac{1}{n}\sum_{k=1}^{n}|(\mathcal{V}f)(k/n)|^p \le L^p \tag{95}$$

and

$$\int_0^1 |(\mathcal{V}f)(x)|^p\,dx \le L^p. \tag{96}$$

Applying the Mean Value Theorem to $x \mapsto x^{1/p}$ then gives that

$$\left|\left(\frac{1}{n}\sum_{k=1}^{n}|(\mathcal{V}f)(k/n)|^p\right)^{1/p} - \left(\int_0^1 |(\mathcal{V}f)(x)|^p\,dx\right)^{1/p}\right| \le \frac{pL^p}{n}\frac{1}{p}(L^p)^{1/p-1} = \frac{L}{n}. \tag{97}$$

Combining (93) and (97) completes the proof for $p < \infty$. The corresponding result for $p = \infty$ follows by taking the limit $p \to \infty$ and using $\|\cdot\|_\infty = \lim_{p\to\infty}\|\cdot\|_p$.

8.4 Proof of Theorem 6.2

The theorem follows from the following lemma:

Lemma 8.4. *With the same notation as Theorem 6.2,*

$$\mathbb{P}\left(\|\mathbf{z}_n\|_p^* \geq t\right) \leq 2\exp(-nt^2/2\overline{\sigma}_n^2). \tag{98}$$

To see that Theorem 6.2 follows from Lemma 8.4, observe that since the right side of (98) is summable over n, it follows from the Borel-Cantelli Lemma that $\lim_{n\to\infty} \|\mathbf{z}_n\|_p^* = 0$ almost surely. Furthermore,

$$\begin{aligned}
\mathbb{E}[\|\mathbf{z}_n\|_p^*] &= \int_0^\infty \mathbb{P}\left(\|\mathbf{z}_n\|_p^* \geq t\right) dt \\
&\leq 2\int_0^\infty \exp(-nt^2/2\overline{\sigma}_n^2)dt \\
&= \frac{2\overline{\sigma}_n}{\sqrt{n}} \int_0^\infty \exp(-u^2/2)du,
\end{aligned} \tag{99}$$

completing the proof of Theorem 6.2.

We now turn to the proof of Lemma 8.4.

Proof of Lemma 8.4. Define the sum, for $k = 1, \ldots, n$.

$$S_k = \sum_{j=1}^{k} z_j. \tag{100}$$

Then the sequence $S_1, \ldots, S_n, \ldots$ is a martingale, and so for any value $\lambda > 0$,

$$X_k = \exp(\lambda S_k) \tag{101}$$

is a submartingale. By Doob's Martingale Inequality [1], for any value t,

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq k\leq n} (\mathcal{V}\mathbf{z})_k \geq t\right) &= \mathbb{P}\left(\max_{1\leq k\leq n} \frac{1}{n} S_k \geq t\right) \\
&= \mathbb{P}\left(\max_{1\leq k\leq n} X_k \geq \exp(n\lambda t)\right) \\
&\leq \mathbb{E}[X_n]\exp(-n\lambda t) \\
&= e^{n\lambda^2\overline{\sigma}_n^2/2 - n\lambda t}.
\end{aligned} \tag{102}$$

Taking $\lambda = t/\overline{\sigma}_n^2$, we get the bound

$$\mathbb{P}\left(\max_{1\leq k\leq n} \frac{1}{n} S_k \geq t\right) \leq \exp(-nt^2/2\overline{\sigma}_n^2). \tag{103}$$

By symmetry, this immediately yields the two-sided bound

$$\mathbb{P}\left(\|\mathbf{z}_n\|_\infty \geq t\right) = \mathbb{P}\left(\max_{1\leq k\leq n} \frac{1}{n}|S_k| \geq t\right) \leq 2\exp(-nt^2/2\overline{\sigma}_n^2). \tag{104}$$

Since $\|\mathbf{z}_n\|_p \leq \|\mathbf{z}_n\|_\infty$ for all p, the result follows. □

9 Conclusion

We have studied the Volterra p-norms $\|\cdot\|_p^*$ in both the discrete and continuous settings. We have shown that the distance between perturbed functions is bounded by the size of the perturbation, and that the distance between two one-dimensional projections of a two-dimensional function is bounded by the size of

the difference in projection directions. We also proved that the effect of additive noise on the discrete norm become negligible as the sampling rate increases.

The special case of Earth Mover's Distance, $\mathrm{EMD}(P,Q) = \|P-Q\|_1^*$, is of particular interest. In addition to the aforementioned papers [11], [6], and [15], EMD is being increasingly used in machine learning and statistical applications. Our results suggest that many of the favorable properties of EMD may be shared by a wider class of metrics.

Acknowledgements

I thank Joe Kileel, Amit Moscovich, Rohan Rao, and Amit Singer for stimulating discussions on the papers [11], [6], and [15]. I acknowledge support from the NSF BIGDATA award IIS 1837992 and BSF award 2018230.

References

[1] Rick Durrett. *Probability Theory and Examples*. Cambridge University Press, fifth edition, 2019.

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

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

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

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

[6] Joe Kileel, Amit Moscovich, Nathan Zelesko, and Amit Singer. Manifold learning with arbitrary norms. *arXiv preprint arXiv:2012.14172*, 2020.

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

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

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

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

[15] 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.

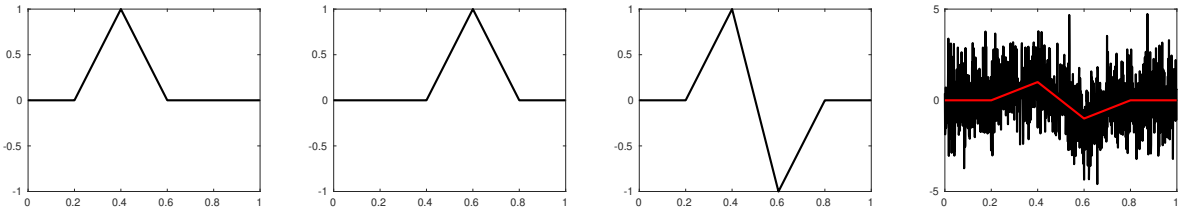

Figure 1: First and second panel: the triangle functions. Third panel: the difference of the triangle functions. Fourth panel: the difference with additive noise of variance 2; the noiseless function is plotted in red.

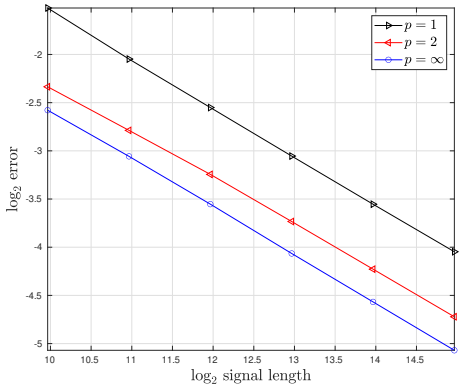

Figure 2: Plot of $\log_2(\text{err}_{n,p})$ against $\log_2(n)$, for $p = 1, 2, \infty$. The slope of each curve is approximately $1/2$, consistent with the error rate predicted by Corollary 6.3.

n	$p = 1$	$p = 2$	$p = \infty$
1000	3.492e-01	1.982e-01	1.674e-01
2000	2.419e-01	1.448e-01	1.201e-01
4000	1.707e-01	1.055e-01	8.518e-02
8000	1.204e-01	7.519e-02	5.975e-02
16000	8.517e-02	5.338e-02	4.218e-02
32000	6.050e-02	3.790e-02	2.978e-02

Table 1: Table of errors $\text{err}_{p,n}$. For each value of p, $\text{err}_{p,n}$ is approximated halved when n is quadrupled, consistent with the error rate predicted by Corollary 6.3.

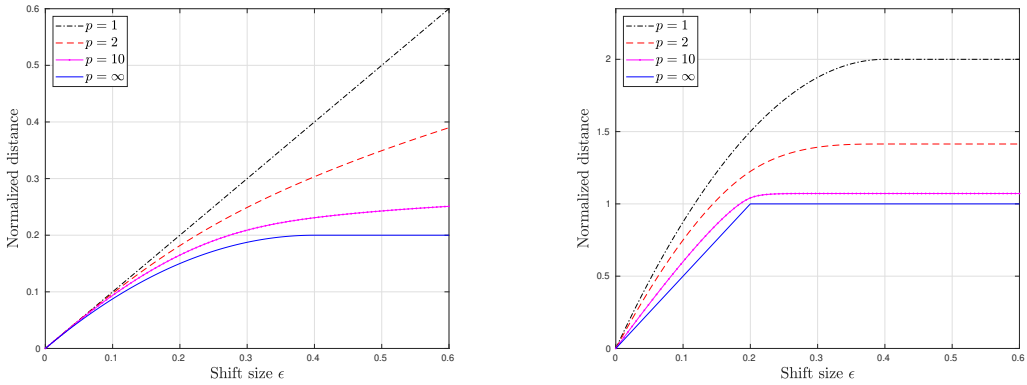

Figure 3: Shifting the triangle function. Left: Normalized Volterra p-distances. Right: Normalized p-distances.

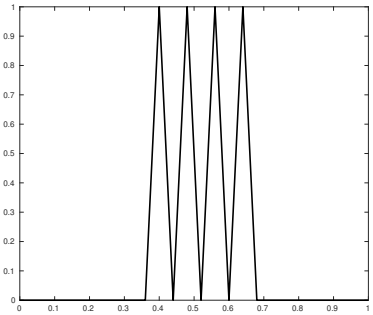

Figure 4: The spiked function.

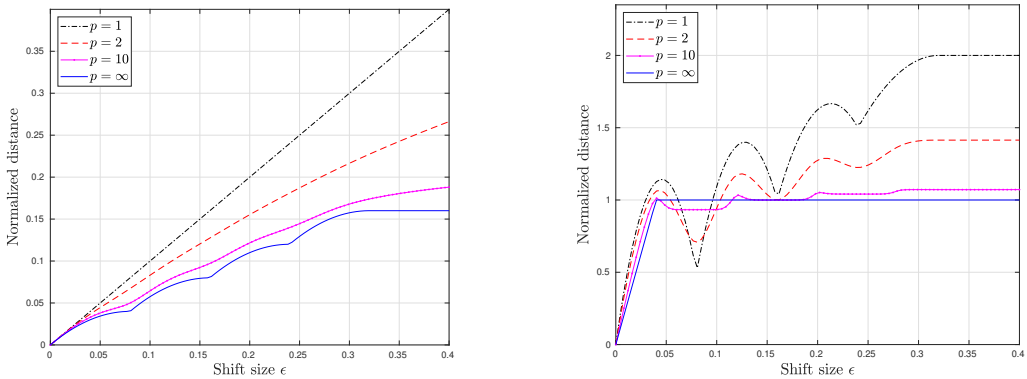

Figure 5: Shifting the spiked function. Left: Normalized Volterra p-distances. Right: Normalized p-distances.

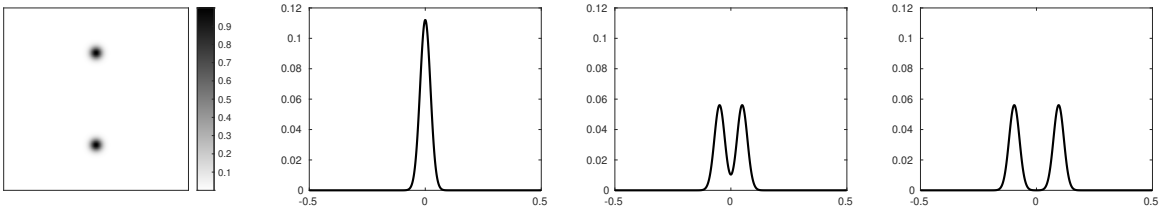

Figure 6: First panel: heat map of the function of two variables. Second panel: vertical projection. Third panel: projection at $\pi/16$ radians from y-axis. Last panel: projection at $\pi/8$ radians from y-axis.

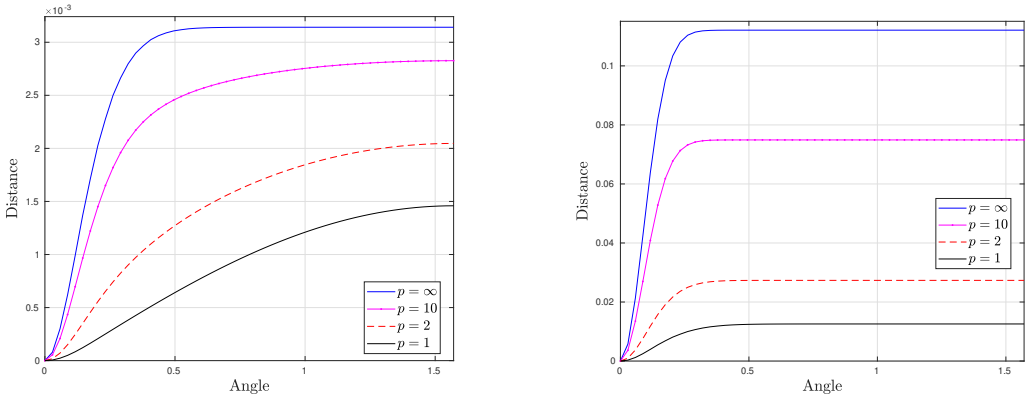

Figure 7: Distances between projections. Left: Volterra p-distances. Right: p-distances.